Genetic characterization of wild barley populations (*Hordeum vulgare* ssp. *spontaneum*) from Kazakhstan based on genome wide SNP analysis

Yerlan Turuspekov1), Saule Abugalieva1), Kanat Ermekbayev1) and Kazuhiro Sato*2)

1) Institute of Plant Biology and Biotechnology, Timiryazev street 45, Almaty 050040, Kazakhstan
2) Institute of Plant Science and Resources, Okayama University, Kurashiki, Okayama 710-0046, Japan

The wild ancestral form of barley, *Hordeum vulgare* ssp. *spontaneum*, is a valuable source for gene enrichment of cultivated barley. The purpose of this work was to study the area of distribution as well as the extent and structure of genetic variation of wild barley populations grown in Kazakhstan. It was found that distribution of wild barley populations in Kazakhstan is restricted to the most southern province. A genome wide single nucleotide polymorphism (SNP) analysis was performed in order to study the level of the genetic diversity in 96 accessions representing 14 wild barley populations from Kazakhstan and 25 accessions from the Middle East which is the center of diversity of this subspecies. The oligonucleotide pooled assay was used to genotype 384 SNPs distributed throughout the genome. In total 233 polymorphic SNPs were selected for further statistical analysis. The level of genetic diversity of wild barley populations from Kazakhstan was predictably narrower (He = 0.19 ± 0.01) in comparison with wild barley samples from the Middle East (He = 0.29 ± 0.01). The results suggested that *H. vulgare* ssp. *spontaneum* populations in Kazakhstan probably represent a recent spread of a limited number of plants from the primary distribution area and might be well adapted to winter low temperature.

**Key Words:** wild barley, genetic diversity, single nucleotide polymorphism, genome.

Introduction

Wild barley, *Hordeum vulgare* ssp. *spontaneum*, is a direct progenitor of cultivated barley and valuable genetic resource for its improvement for resistance to biotic and abiotic stresses and productivity. There are multiple lines of evidence suggesting that most genetic variation of wild barley is concentrated in populations that grow in the Middle East (Nevo 1992). There are several reports describing the extent and structure of the genetic variation of wild barley populations in some parts of Central Asia (Turuspekov et al. 1996, Volis et al. 2001), which is the marginal distribution area of this subspecies (Bothmer et al. 2003). At the same time, despite the facts that Kazakhstan is the largest country in Central Asia and barley is second important cereal crop after wheat in Kazakhstan, there were no reports on characterization of the genetic diversity of wild barley grown in Kazakhstan.

The genetic diversity in wild barley populations has been studied by using a number of different genetic markers, including storage proteins (Doll and Brown 1979, Nevo et al. 1983), isozymes (Brown et al. 1978, Nevo et al. 1986), and PCR based DNA markers (Baum et al. 1997, Nevo et al. 2005, Owuor et al. 1999, Turpeinen et al. 2003). Of these technologies, the automated genome-wide profiling of plants with single nucleotide polymorphism (SNP) markers is increasingly used for evaluation of genetic resources, including barley (Close et al. 2009).

The objectives of this work were to (i) describe the area of distribution for wild barley populations growing in Kazakhstan and (ii) assess the level and structure of genetic diversity in these populations as compared to accessions from the Middle East.

Materials and Methods

The plant material from southern Kazakhstan consisted of 96 accessions of wild barley that represent 14 different populations (Table 1, Fig. 1). The plants were collected by the expedition team of Kazakhstan with distances at least 10 m apart, and locations were recorded by GPS device (Table 1). In order to make a small subset of *H. vulgare* ssp. *spontaneum* collection from International Barley Core Collection, 36 accessions were selected to reflect the structure of country of origin in total 150 accessions. Of the 36 accessions, 25 were collected from the Middle East and used for
All wild barley accessions from Kazakhstan and the Middle East were grown in individual pots in a glasshouse. DNA samples were extracted and purified by using commercial kits (Qiagen, CA, USA). The DNA concentration for each sample was adjusted to 50 ng/μl. Selected 384 SNPs from the Illumina oligonucleotide pool assay (OPA) with known genetic positions from a consensus barley genetic map (Close et al. 2009) were used in this study. The map location of each SNP was given according to Sato et al. (2011). PCR, hybridization, and scanning were performed according to the GoldenGate genotyping assay protocol (Illumina Inc.; Fan et al. 2006) at the Institute of Plant Science and Resources, Okayama University, Japan. SNP basecalling was performed using GenomeStudio software (Illumina Inc.).

The statistical analyses of population genetics parameters were performed using GenAIEx (Peakall and Smouse 2006, 2012) and Popgene (Yeh and Boyle 1997). The phylogenetic trees were constructed using neighbor joining method by 1000 bootstrap replications in MEGA6 (Tamura et al. 2013).

Table 1. Collection sites of 14 wild barley populations in southern Kazakhstan

| Population | ID in Fig. 2 | Longitude | Latitude | Altitude (m) | Location |
|------------|--------------|-----------|----------|--------------|----------|
| 1          | 1.1–1.7      | E42 23 56 | N069 38 27 | 620          | Suburb of Shymkent |
| 2          | 2.1–2.7      | E42 09 28 | N069 45 30 | 762          | Near Akbastau    |
| 3          | 3.1–3.7      | E41 57 95 | N069 29 55 | 715          | Near Kazygurt hils |
| 4          | 4.1–4.7      | E41 46 15 | N069 32 12 | 699          | Near Turbat      |
| 5          | 5.1–5.7      | E41 52 34 | N069 28 15 | 520          | Near Sharapkhana |
| 6          | 6.1–6.7      | E41 29 16 | N069 26 09 | 621          | Near Zhbek Zholy |
| 7          | 7.1–7.7      | E41 29 15 | N069 26 08 | 623          | Near Zhbek Zholy |
| 8          | 8.1–8.7      | E41 29 15 | N069 26 09 | 626          | Near Zhbek Zholy |
| 9          | 9.1–9.7      | E41 29 17 | N069 26 09 | 623          | Near Zhbek Zholy |
| 10         | 10.1–10.7    | E41 26 20 | N069 07 18 | 416          | Near Saryagash   |
| 11         | 11.1–11.7    | E41 24 37 | N069 03 54 | 400          | Near Abai        |
| 12         | 12.1–12.7    | E41 12 48 | N069 35 59 | 270          | Near Birlik      |
| 13         | 13.1–13.7    | E41 53 02 | N069 35 48 | 780          | Near Altyntobe   |
| 14         | 14.1–14.5    | E41 29 18 | N069 26 08 | 624          | Near Krasnovodopad breeding station |

Results and Discussion

Geographic locations of wild barley population in southern Kazakhstan

Collecting trips in 2008–2009 showed that wild barley (H. vulgare ssp. spontaneum) distribution in Kazakhstan is restricted to the most southern province. The region is considered as one of the most northern margin of distribution for this wild barley by Bothmer et al. (2003) but not described in an earlier study of Harlan and Zohary (1966). Within the province the typical area of growth is in valleys located close to mountain ranges with some sporadic appearance on roadsides of deserted areas, such as population 12 (Table 1, Fig. 1). The most northern successful collecting location was the regional capital Shymkent (population 1). The 14 populations collected were situated within the province as follows: north-east side (populations 1 and 2); south side (populations 3–11, 13 and 14); and far west-south side (population 12). Searches for additional populations to the north, east, and west from Shymkent were not successful. Therefore, assuming that wild barley migrated from the Middle East via other countries of Central Asia, e.g. Uzbekistan, Shymkent might be the most northern location of wild barley in Kazakhstan. The elevation of collection sites varied from 270 m (deserted area, population 12) to 780 m (population 13) above sea level (Table 1). Plants have winter growth habit and seed matures at the end of May under local growing conditions. The major limitation for expansion of wild populations to the north appears to be winter low temperature. Therefore, the accessions which were studied might have undergone natural selection for winter survival.

The extent and structure of wild barley populations from Kazakhstan

The genetic diversity of wild barley populations was studied based on the analysis of 384 SNP genotypes using the GoldenGate assay (Illumina Inc.). Of these SNPs, 278 with excellent base call rates for all accessions were selected for further analysis. Forty-five SNPs monomorphic across
all samples were removed from the analysis. Therefore, a total of 233 SNP markers were used in the genetic analysis.

The parameters from the genetic diversity analysis of 96 accessions from 14 populations indicated a relatively low level of genetic diversity (He) with the range from 0.002 in population 6 to 0.160 in population 12 (Table 3). The highest within-population heterogeneity was observed for the most southern population (12), which was collected on the roadside of deserted area close proximity to the border with Uzbekistan (Fig. 1). Population 6, which was collected near the Krasnovodopad breeding station in southern Kazakhstan had the lowest genetic diversity. Phylogenetic analysis of accessions from all the 14 populations revealed only one cluster with five sub-clusters (Ia–Ie) as shown in Fig. 2. Although population 6 showed distribution within a sub-cluster, most of the accessions form a population were in multiple sub-clusters and the populations might share genetically close accessions. Therefore, there was a lack of correlation between geographic and genetic distances.

Comparative assessment of genetic variation in wild barley populations from the Middle East and Kazakhstan

In order to compare the diversity level of wild barley populations from Kazakhstan with that of germplasm from the center of diversity, 25 accessions from the Middle East (accessions of International Barley Core Collection) were analyzed using the same set of 233 SNP markers which were well spread on barley genome. The results of this comparative genetic description of the two germplasm arrays is summarized in Table 4. All parameters, including effective number of alleles and heterozygosity index, indicate higher levels of genetic variation for the Middle East germplasm than for the Kazakhstan germplasm. In particular, the heterozygosity parameter (He) is much higher for the Middle East germplasm (0.29) than for the Kazakhstan germplasm (0.19). The analysis of variance including all the accessions from the Middle East and Kazakhstan arrays revealed that 23% of the total variation can be attributed to the variation between arrays, while 77% of the total variation is within...
arrays. On the other hand, the neighbor joining phylogenetic tree clearly differentiates the accessions from the two geographic areas, although accession 5.4 from Kazakhstan population was closer to the Middle East cluster II (Fig. 2).

An advantage of this study is that selected SNP markers were previously well-characterized and information on EST markers is available in barley databases (Close et al. 2009). However, the selected 384 SNP markers were developed from EST sequences of several standard cultivars, including Haruna Nijo, Barke, and Morex (Sato et al. 2011). The only wild barley accession that used as a source for EST sequences was line H602 (*H. vulgare* ssp. *spontaneum* var. *transcaspicum*). Therefore, polymorphisms detected in the wild barley population in this study may be biased. Nevertheless, the SNPs represent the best currently available tool for characterizing genetic diversity in barley. The high winter hardiness of wild barleys in this area (Kazakhstan and Uzbekistan) were observed at the winter barley screening nursery at Oregon State University, USA (Hayes, personal communication). Wild barley accessions in Kazakhstan probably represents a recent spread of a limited number of plants by human activity from the primary distribution area that were reported previously (Harlan and Zohary 1966).

| Population | Middle East | Kazakhstan |
|------------|-------------|------------|
| Number of accessions | 25 | 96 |
| *Na*<sup>a</sup> | 2.19 ± 0.03 | 1.65 ± 0.04 |
| *Ne*<sup>b</sup> | 1.48 ± 0.02 | 1.32 ± 0.03 |
| *I*<sup>c</sup> | 0.47 ± 0.01 | 0.28 ± 0.02 |
| *He*<sup>d</sup> | 0.29 ± 0.01 | 0.19 ± 0.01 |

<sup>a</sup> Mean allele number.  
<sup>b</sup> Effective number of alleles.  
<sup>c</sup> Shannon-Weave information index.  
<sup>d</sup> Expected heterozygosity.

Fig. 2. The consensus phylogenetic tree for accessions from Kazakhstan (96) and the Middle East (25) constructed using the circle style of the neighbor joining method. Numerals with period indicate population and plant numbers of accessions from Kazakhstan. Accessions of the Middle East are shown in country code and numerals as designated in Table 2. Bootstrap values with 1000 replications are shown in italics.
Therefore, wild barley in Kazakhstan may be one of examples for expansion of its secondary distribution area.

The fingerprinted wild barley germplasm can be a source of alleles for improvement of cultivated barley. The procedure is clearly shown by Hori et al. (2005) who backcrossed a wild barley segment into an elite cultivar background. SNP marker information is the key for accomplishing efficient introgression. The barley genome sequence data (The International Barley Genome Sequencing Consortium 2012) will provide a larger catalog of SNPs for introducing genes from wild barley more precisely and can help to eliminate deleterious segments efficiently.

Acknowledgements

We would like to thank Dr. Patrick M. Hayes (Oregon State University, USA) for his critical reading of the manuscript.

Literature Cited

Baum, B.R., E. Nevo, D.A. Johnson and A. Beiles (1997) Genetic diversity in wild barley (Hordeum spontaneum Koch.) in the Near East: A molecular analysis using random amplified polymorphic DNA (RAPD). Genet. Resour. Crop Evol. 44: 147–157.

Bothmer, R.v., K. Sato, T. Komatsuda, S. Yasuda and G. Fischbeck (2003) The domestication of cultivated barley. In: Bothmer, R.v., T.v. Hintum, H. Knüppfer and K. Sato (eds.) Diversity in Barley (Hordeum vulgare), Elsevier, Amsterdam, pp. 9–27.

Brown, A.H.D., D. Zohary and E. Nevo (1978) Outcrossing rates and heterozygosity in natural populations of Hordeum spontaneum Koch. in Israel. Heredity 41: 49–62.

Close, T.J., P.R. Bhat, S. Lonardi, Y. Wu, N. Rostoks, L. Ramsay, A. Druka, N. Stein, J.T. Svensson, S. Wanamaker et al. (2009) Developing recombinant chromosome substitution lines and doubled haploid lines derived from a cross between Hordeum vulgare ssp. vulgare and Hordeum vulgare ssp. spontaneum. Mol. Breed. 16: 295–311.

Nevo, E., A. Beilis and N. Storch (1983) Microgeographic edaphic differentiation in hordein polymorphisms of wild barley. Theor. Appl. Genet. 64: 123–132.

Nevo, E., A. Beilis and D. Zohary (1986) Genetic resources of wild barley in the Near East: structure, evolution and application in breeding. Biol. J. Linn. Soc. 27: 355–380.

Nevo, E. (1992) Origin, evolution, population genetics and resources for breeding of wild barley, Hordeum spontaneum in the Fertile Crescent. In: Shewry, P.R. (ed.) Barley: Genetics, Biochemistry, Molecular Biology and Biotechnology, CAB International, Wallingford, UK, pp. 19–43.

Owuor, E.D., T. Fahima, A. Beharav, A. Korol and E. Nevo (1999) RAPD divergence caused by microsite edaphic selection in wild barley. Genetica 105: 177–192.

Peakall, R. and P.E. Smouse (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Notes 6: 288–295.

Peakall, R. and P.E. Smouse (2012) GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics 28: 2537–2539.

Sato, K., T.J. Close, P. Bhat, M. Muñoz-Amatriain and G.J. Muehlbauer (2011) Single nucleotide polymorphism mapping and alignment of recombinant chromosome substitution lines in barley. Plant Cell Physiol. 52: 728–737.

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. 30: 2725–2729.

The International Barley Genome Sequencing Consortium (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature 491: 711–716.

Turpcein, T., V. Vanhala, E. Nevo and E. Niissila (2003) AFLP genetic polymorphism in wild barley (Hordeum spontaneum) populations in Israel. Theor. Appl. Genet. 106: 1333–1339.

Turuspekov, Y., S. Abugalieva, S. Mendlinger and S. Volis (1996) Polymorphism of wild barley Hordeum spontaneum K. from Turkmenistan. Russ. J. Genet. 32: 767–773.

Volis, S., S. Mendlinger, Y. Turuspekov, U. Esnazarov, S. Abugalieva and N. Orlovsky (2001) Allozyme variation in Turkmenistan population of wild barley, Hordeum spontaneum Koch. Ann. Bot. 87: 435–446.

Yeh, F.C. and T.J.B. Boyle (1997) Population genetic analysis of co-dominant and dominant markers and quantitative traits. Belg. J. Bot. 129: 157.