Phylogeographic and quantitative trait locus analysis of the ability of Aegilops tauschii Coss., the D genome progenitor of common wheat, to cause genome doubling in the F1 hybrids wit…

Matsuoka, Yoshihiro

Title

Phylogeographic and quantitative trait locus analysis of the ability of *Aegilops tauschii* Coss., the D genome progenitor of common wheat, to cause genome doubling in the F_1 hybrids with *Triticum turgidum* L., the AB genome progenitor

Author information

Yoshihiro Matsuoka

Graduate School of Agricultural Science, Kobe University, Kobe, Japan

Email: pinehill@port.kobe-u.ac.jp

Abstract

Aegilops tauschii Coss. (DD genome) is a wild paternal progenitor of common wheat (*Triticum aestivum* L.; AABBDD genome). This species has genetically distinctive intraspecific lineages that differ in their patterns of involvement in allopolyploid speciation of common wheat. *Ae. tauschii* accessions can cause genome doubling at variable frequencies depending on their genotypes via unreduced gamete production and fusion in the trihaploid F_1 hybrids (ABD genome) with *Triticum turgidum* L. (AABB genome), the maternal progenitor. In this study, we examined the variation patterns of *Ae. tauschii*'s ability to cause hybrid genome doubling based on an artificial cross experiment and attempted to improve on a previous linkage map of loci that control the expression of this ability by using an increased number of anchor markers. According to the results, this ability was genealogically and geographically widespread within the species, suggesting that it might not have been critically involved in shaping common wheat speciation patterns. The weak phylogeographic structure of the trait variation is consistent with the idea that the genes for hybrid genome doubling have some function (most likely, meiotic) in *Ae. tauschii* and are maintained because of their adaptive importance, whereas genes may accumulate non-deleterious mutations that could positively or negatively influence the expression of genome doubling when placed in the hybrid genome background. The linkage analysis used 1,035 anchor markers and identified five loci on chromosomes 2D, 3D, 6D, and 7D that significantly influenced the expression of hybrid genome doubling.

Keywords

allopolyploid speciation, hybrid genome doubling, intraspecific variation, natural hybridization, unreduced gametes, wheat evolution

Acknowledgements

This paper is dedicated to the memory of Shigeo Takumi, my old friend and long-term collaborator. I thank Taihachi Kawahara, Kenji Kato, Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK), International Center for Agricultural Research in the Dry Areas (ICARDA), Plant Germ-plasm Institute of Kyoto University, National BioResources Project, and U. S. Department of Agriculture for the seeds of *Ae. tauschii* accessions. This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP 20K20720, JP 19H02935, and JP 19KK0157.

Introduction

Aegilops tauschii Coss. (formerly known as *Aegilops squarrosa* L.; DD genome) is the wild progenitor of common wheat (*Triticum aestivum* L.; AABBDD genome) (Kihara 1944; McFadden and Sears 1944). This species occurs in a variety of habitats, including seashores, wastelands, steppes, roadsides, and temperate forests in the central Eurasia from central Syria to western China, and is often found growing as a weed in wheat and barley fields (van Slageren 1994; Kilian et al. 2011). *Ae. tauschii* displays diverse phenotypes of morphological, phenological, and physiological traits (Matsuoka et al. 2015; Saisho et al. 2016). Distinctive genetic lineages are known within this species: two large lineages (TauL1 and TauL2), and one small lineage (TauL3) (Mizuno et al. 2010; Matsuoka et al. 2013; Gogniashvili et al. 2016). TauL1 occurs throughout the species distribution range, and TauL2 and TauL3 are endemic to the Transcaucasus-Middle East and Georgia regions, respectively. TauL1 appears to be reproductively isolated from TauL2 and TauL3 because the intermediate genotypes are very rare in its natural habitats (Wang et al. 2013).

In the early stages of the evolution of common wheat, *Ae. tauschii* is thought to have naturally hybridized as a male with a cultivar of tetraploid wheat (*Triticum turgidum* L.; AABB genome) to produce trihaploid F_1 hybrids (ABD genome). The direct ancestors of common wheat, i.e., allohexaploid F_2 plants with the AABBDD genome, arose through hybrid genome doubling via the fusion of unreduced gametes produced by the F_1 hybrids (Kihara 1946). The identity of the ancestral *Ae. tauschii*, which participated in a critical hybrid cross ca. 8,000 years ago, remains largely unclear. Of the three intraspecific lineages, TauL2 and TauL3 are more closely related to the D genome of common wheat than is TauL1. Furthermore, the TauL2 accessions that have a high potential for natural hybridization with the *T. turgidum* cluster in the southern coastal Caspian region (Matsuoka and Takumi 2017). Therefore, TauL2 and TauL3 may represent lineages that are closely related to the ancestral *Ae. tauschii* (Matsuoka et al. 2013; Gaurav et al. 2021).

Hybrid genome doubling has played a key role in the evolution of common wheat. Understanding its genetic mechanism may help clarify the origin of common wheat and also elucidate the patterns of plant polyploid evolution, as almost all polyploids arise by way of unreduced gametes (Harlan and DeWet 1975). The progenitors of common wheat provide suitable models for studying these mechanisms. Embryo-rescue-free artificial cross experiments have generated *T. turgidum–Ae. tauschii* F_1 hybrids that produce unreduced gametes and set allohexaploid F_2 seeds via fusion (Kihara 1946). The parental accession's ability to cause hybrid genome doubling, which

varies depending on the genotype, can be measured as the selfed seedset rate (i.e., the number of seeds set per floret) in F_1 hybrids (Kihara et al. 1965).

Previous artificial cross experiments have shed some light on the genetic underpinning of hybrid genome doubling via unreduced gamete fusion in *T. turgidum*–*Ae. tauschii* F_1 hybrids. First, the natural variation pattern of the ability to cause hybrid genome doubling has little association with the intraspecific lineage structure in *Ae. tauschii*: large variation exists within each TauL1 and TauL2, but not between these lineages (Matsuoka et al. 2013). The variation pattern of the small TauL3 lineage remains to be determined. Interestingly, the natural variation pattern for the ability is clearly structured in *T. turgidum* (Matsuoka and Mori 2020). Second, the ability to cause hybrid genome doubling is a complex trait controlled by several quantitative trait loci (QTLs) for unreduced gamete production. Six and four QTLs have been identified in the genomes of *Ae. tauschii* and *T. turgidum*, respectively (Matsuoka et al. 2013; Hao et al. 2014; Matsuoka and Mori 2020). The present work aimed to further elucidate the natural variation patterns and QTLs for the ability to cause hybrid genome doubling in *Ae. tauschii*. We estimated the ability to cause hybrid genome doubling (measured as selfed seedset rates in the F_1 hybrids with a tester *T. turgidum* accession) for 60 accessions using a generalized linear mixed model (GLMM) approach and analyzed the structure of the trait variation. QTL analysis was performed to map the loci associated with hybrid genome doubling using a comparatively high number of anchor markers for segregant genotyping. The objectives of the study were to (1) examine the phylogeographic variation pattern for the ability to cause hybrid genome doubling and (2) provide an improved linkage map for loci that control the ability to cause hybrid genome doubling in *Ae. tauschii*.

Materials and methods

Plant materials

Sixty *T. turigidum–Ae. tauschii* F_1 hybrid genotypes, each derived from one of the 60 *Ae. tauschii* accessions (24 belonging to TauL1, 33 to TauL2, and three to TauL3), were used to evaluate of the ability to cause hybrid genome doubling (Table 1). These *Ae. tauschii* accessions represent the entire natural habitat range of the species in the Eurasia. F_1 hybrid genotypes were produced through an artificial cross experiment using a single durum wheat tester (*Triticum turgidum* L. subsp. *durum* cv. ‘Langdon’) as the female parent. Durum wheat is a candidate female progenitor of common wheat and was chosen as the tester to mimic the genetic conditions for early stage common wheat evolution (Matsuoka and Nasuda 2004).

The *Ae. tauschii* KU-2103 accession that generated an F_1 genotype having a high selfed seedset rate when crossed with the tester was used as the “high accession” in the QTL analysis, whereas the KU-2080 accession that generated an F_1 genotype having a low genome doubling frequency was used as the “low accession.” The QTL mapping population (279 trihaploid plants), in which only the loci on the D genome were segregated, was previously produced by artificially crossing the F_1 plants between the high and low accessions with a tester (Matsuoka et al. 2013).

Genetic relationships between the *Ae. tauschii* accessions

A probabilistic principal component analysis of the binary genotype dataset that was obtained previously using 169 Diversity Arrarys Technology (DArT) markers (Matsuoka et al. 2013) was done to analyze genetic relationships between the *Ae. tauschii* accessions. The pcaMethods package (Stacklies et al. 2007) for R ver. 4.0 (R Core Team 2021) was used with the “ppca” option in the probabilistic principal component analysis. Principal component plot was drawn using the ggplot2 package (Wickham 2016).

Artificial cross experiment

Plants were cultivated in individual pots in a greenhouse from early winter to late spring. Spikes of tester plants were hand-emasculated before anthesis by removing anthers from the first and second florets. The third and fourth florets were entirely removed. Each emasculated spike was bagged until further use. Two days after emasculation, pollen collected from an *Ae. tauschii* plant was delivered to the pistils of an emasculated spike. Pollinated spikes were bagged through the grain filling stage until harvest.

Estimates of the ability to cause hybrid genome doubling

F_1 hybrid seeds were germinated in Petri dishes at 20˚C and transplanted into individual pots. The plants were grown in a greenhouse from early winter to late spring. For each F_1 hybrid genotype, cultivation was done in a single or multiple (up to three) seasons using one to seven plants in total. In each plant, except for the plant that was derived from *Ae. tauschii* accession PI 486274, multiple early spikes were bagged before anthesis for selfing. After the harvest, the numbers of seeds set and associated florets were counted for each spike. The count dataset was compiled from published and unpublished data (Table S1). The occurrence of hybrid genome doubling was confirmed for 27 F_1 genotypes by chromosome observation of the offspring in the previous study

(Takumi et al. 2009).

For each parental *Ae. tauschii* accession, the ability to cause hybrid genome doubling (i.e., the selfed seedset rate of the F_1 hybrid genotype) was estimated by generalized linear mixed model (GLMM) analysis using the *glmer* function (with the bobyqa optimizer option) of the lme4 package (Bates et al., 2015) for R ver. 4.0 (R Core Team 2021). The *dispersion_glmer* function of the blmeco package (Korner-Nievergelt et al. 2015) was used to estimate overdispersion, whereas the *r.squarredGLMM* function of the MuMIn package (Bartoń, 2019; Nakagawa and Schielzeth, 2013) was used to calculate the theoretical marginal R^2 (the proportion of variance explained by the fixed effect) and conditional R^2 (the proportion of variance explained by the fixed effect and random effects) values. Based on the fixed effect coefficients, the selfed seedset rates of the F_1 genotypes, standard errors, and asymptotic 95% confidence intervals were calculated using the *lsmeans* function of the lsmeans package (Lenth 2016). Graphs were drawn using the ggplot2 package (Wickham 2016).

Geographic patterns for the ability to cause hybrid genome doubling were examined by creating a map based on the coordinates of the *Ae. tauschii* accessions using a spatial dataset obtained from DIVA-GIS ver. 7.5.0 (Hijimans et al. 2012) and the ggplot2 package (Wickham 2016).

QTL analysis

QTL analysis was performed to map *Ae. tauschii* loci that influence the ability to cause genome doubling in the F_1 hybrids with the tester. Cultivation of and DNA extraction from the segregants, the tester-KU-2103 F_1 hybrid (one individual), and the tester-KU-2080 F_1 hybrid (one individual) were conducted as previously described (Matsuoka et al. 2013). Each plant was haplotyped using D-genome specific microsatellite markers and DArTseq, which was performed at Diversity Arrays Technology Pty. Ltd. The microsatellite haplotype dataset was obtained from a previous work (Matsuoka et al. 2013). The DArTseq single nucleotide polymorphisms (SNPs) were screened based on the following criteria: they must be physically mapped to the D genome of the *T. aestivum* reference genome sequence wheat_ChineseSpring04 (Diversity Arrays Technology); (b) they must be homozygous because the individual plants are haploids; (c) there are no missing values in the tester-KU-2103 and tester-KU-2080 F_1 hybrids; and (d) each SNP had a the missing data percentage of less than 5%.

The MSTmap algorithm (the *mstmap* function with the “bychr” and “anchor” options) implemented in the ASMap package and the Kosambi genetic map distance function were used to construct a linkage map based on the microsatellite markers and

qualified SNPs (Kosambi 1944; Wu et al. 2008; Taylar and Butler 2017). The R/qtl package based on the model for recombinant inbred lines was used to carry out QTL analysis with the *calc.genoprob* function (density, 1 cM; genotyping error rate 0.001) and the *sim.geno* function (256 draws) (Broman et al. 2003). The phenotype dataset (i.e., selfed seedset rates of the segregants) was obtained from a previous work (Matsuoka et al. 2013). Single-QTL analysis was performed using the *scanone* function, and the statistical significance of putative QTLs was estimated using 100,000-fold permutation tests. A multiple QTL model was constructed consecutively using the *makeqtl*, *fitqtl*, *addqtl*, *refineqtl*, and *addint* functions. The *bayesint* function was used to calculate the approximate 95% Bayesian credible intervals for the QTL locations with the "expandtomarkers" option.

Results

Estimates of the ability to cause hybrid genome doubling

Seedset count data were obtained from 15,714 florets in 789 spikes from 189 F_1 hybrid plants (Table S1). To estimate the ability to cause hybrid genome doubling in the *Ae. tauschii* accessions, we performed a GLMM analysis of the spike-wise seedset counts as a binominal response variable that consisted of the numbers of successful (i.e., a seed set) and unsuccessful (i.e., no seed set) events and generated a model that had "*Ae. tauschii* accession" as the fixed effect and "plant" (189 plants), "spike" (789 spikes in total), and "season" (four seasons in total) as the random effects. Variances for the random effects were estimated as 0.06 (plant), 0.09 (spike), and 0.10 (season). The estimated dispersion value (1.04) was within the rage indicative of adequate fit of the data to the model (0.75-1.40). The marginal and conditional R^2 values of the model were 0.65 and 0.86, respectively. Using the fixed effect coefficient estimated by the GLMM, the estimated value for the ability to cause hybrid genome doubling (i.e., the selfed seedset rate) was calculated for each of the 60 *Ae. tauschii* accessions: the overall median, mean, and standard deviation (SD) of the values were 0.32, 0.32, and 0.14, respectively (Tables 1 and 2).

Structure of variation for the ability to cause hybrid genome doubling

We examined the genealogical variation patterns of *Ae. tauschii*'s ability to cause hybrid genome doubling using the accession-wise GLMM estimates. The probabilistic principal component analysis confirmed that the 60 *Ae. tauschii* accessions were genetically grouped into three distinct lineages: TauL1, TauL2, and TauL3 (Fig. 1a). In TauL1 (24 accessions), the ability values varied from 0.08 to 0.56 (median = 0.34,

mean= 0.35, SD = 0.13); in TauL2 (33 accessions), from 0.04 to 0.58 (median = 0.33, mean = 0.33, SD = 0.15); in TauL3 (three accessions), from 0.04 to 0.18 (median = 0.13, mean = 0.11, SD = 0.07) (Fig. 1b; Table 2). The observed lineage-wise variation patterns were consistent with the previous results and indicated that the ability was quite comparable between the TauL1and TauL2 accessions (Matsuoka et al. 2013). Furthermore, the geographic distribution of the TauL1, TauL2, and TauL3 accessions showed that the ability to cause hybrid genome doubling was a common trait (Fig. 2). The accessions that had high degrees of the ability (ability values > ca. 0.3) were found to occur widely in the Transcaucasus–Middle Eastern and central–southern Asian regions, whereas some accessions derived from the Caspian coastal region and Transcaucasia had particularly high degrees of the ability (ability values > ca. 0.5).

QTLs for the ability to cause hybrid genome doubling

The MSTmap algorithm generated 48 linkage groups of the microsatellite markers and qualified SNPs. After removing those with five or fewer microsatellite markers and/or SNPs, a reasonable genetic map of the seven chromosomes was obtained. The map contained 69 microsatellite markers and 966 SNPs. The number of microsatellite markers/SNPs that anchored to each chromosome was 126 (1D), 180 (2D), 182 (3D), 96 (4D), 168 (5D), 114 (6D), and 169 (7D). The total map length was 938.3 cM. The average spacing and maximum spacing were 0.9 cM and 19.4 cM, respectively.

Single-QTL analysis performed using the selfed seedset rates of the segregants as the phenotype data (Matsuoka et al. 2013) revealed four significant QTLs located on chromosomes 2D, 3D, 6D, and 7D and a scan for additional QTLs using the *addqtl* function detected one significant QTL located on chromosome 3D (permutation test, $p < 0.05$) (Fig. 3). Based on these QTLs, an additive multiple-QTL model had a logarithm of odds (LOD) score (relative to the no QTL model) of 27.8 and an estimated proportion of variance explained (PVE) value of 36.8%. The *p*-value, based on the LOD score, was 0.00. All of these QTLs had negative estimated effects, indicating that the alleles of the high accession KU-2103 positively influenced the phenotype expression (Table 3). Pairwise loci interactions had relatively small effects (PVE < 1.0%), but the interaction between the 6D and 7D loci was significant (PVE = 1.8%, add-one-pairwise-interaction-at-a-time test, $p = 0.00$).

Discussion

The present study showed that (1) the variation for *Ae. tauschii*'s ability to cause hybrid genome doubling has a weak phylogeographic structure and (2) *Ae. tauschii* had

five QTLs that influenced the expression of hybrid genome doubling. Together with the previous studies, these findings have immediate implications for the role of hybrid genome doubling in the evolution of common wheat.

In hybridizing with *T. turgidum*, the TauL1 accessions expressed pre-pollination and post-pollination barriers much stronger than the TauL2 accessions, suggesting that the ability of *Ae. tauschii* to cause these reproductive barriers might have markedly influenced common wheat's speciation by inducing lineage-associated patterns of gene flow (Matsuoka and Takumi 2017). In the southern coastal Caspian region, the TauL2 accessions that have a high potential for natural hybridization (i.e., formation of viable F_1 hybrids) with *T. turgidum* occur (Matsuoka and Takumi 2017). Therefore, the pattern of common wheat's speciation might have been deeply structured by the genealogy and geography of *Ae. tauschii*'s ability to cause pre- and post-pollination reproductive barriers. In contrast, the ability to cause hybrid genome doubling is widely distributed within the species. This finding suggests that this ability might have less critically been involved in shaping the pattern of common wheat's speciation than the pre- and post-pollination reproductive barriers, despite hybrid genome doubling being an important mechanism for the formation of the common wheat allohexaploid genome. An important caveat is that our present and previous artificial cross experiments relied on a single *T. turgidum* cultivar as the tester. Further studies are required to clarify the roles of reproductive barriers and hybrid genome doubling in allopolyploid speciation of common wheat.

The *Ae. tauschii* accession KU-2103 (the "high accession" in the QTL analysis) was one of the accessions with the shortest genetic distance to the D genome of common wheat (Wang et al. 2013). In a previous artificial cross experiment study that used 31 accessions of wild *T. turgidu*m (*Triticum turgidum* L. subsp. *dicoccoides*), virtually no genome doubling was observed in their F_1 hybrids with KU-2103 (Matsuoka and Mori 2020). Therefore, *Ae tauschii* is distinctive in that its ability is genealogically and geographically widespread within the species. This finding is consistent with the idea that, in *Ae. tauschii*, the genes for hybrid genome doubling have some function (most likely, meiotic), are inherited from a common ancestor of the intraspecific lineages, and are maintained because of their adaptive importance (Matsuoka et al. 2013). The observed variability in the ability might have resulted from non-deleterious mutations in the genes involved in hybrid genome doubling. Alleles that arose from such mutations may remain phenotypically cryptic within the species but have a positive or negative influence on the expression of genome doubling when placed in the hybrid genome background.

The reason why the ability is common in *Ae. tauschii* but virtually null in wild accessions of *T. turgidum* is not clear. In wild *T. turgidum*, non-deleterious alleles that

have a strong negative effect on the expression of hybrid genome doubling might have arisen through mutations and then increased in frequency to near fixation through genetic drift. Importantly, *T. turgidum* cultivars have variable degrees of the ability: some accessions showed high selfed seedset rates (> 50%) in the F_1 hybrids with KU-2103, whereas other accessions had moderate or low selfed seedset rates (Matsuoka and Mori 2020). This suggests that wild *T. turgidum* populations with considerable degrees of the ability exist or existed but are unfound. Another possible interpretation would be that the ability was null in wild *T. turgidum* but arose in the cultivars through mutation after domestication. At any rate, the contrasting variation patterns make *Ae. tauschii* and *T. turgidum* suitable models to study the evolution of the hybrid genome doubling ability.

A previous study identified six QTLs for the ability to cause hybrid genome doubling in the genome of *Ae. tauschii*: QTL1 on chromosome 1D (position 19.8 cM), QTL2 on 2D (24.3 cM), QTL3 on 3D (50.7 cM), QTL4 on 3D (94.4 cM), QTL5 on 6D (118.0 cM), and QTL6 on 7D (94.4 cM). In addition, a significant pairwise loci interaction was detected between QTL2 and QTL7 (Matsuoka et al. 2013). The linkage map constructed in the present study had 1,035 markers: therefore, it was much improved relative to the previous map (77 markers). The present study found no QTL on chromosome 1D; however, the results of the present and previous QTL analyses were largely consistent (Table 3). The QTLs on the 2D, 3D, 6D, and 7D chromosomes require further investigation to clarify their roles in the expression of hybrid genome doubling, whereas the significance of the QTL on chromosome 1D must be re-evaluated. Examples of the functionally characterized meiotic genes of common wheat include *TaMSH7* (mismatch repair gene) on homoeologous group 3 chromosomes and *TaRAD51* (recombination gene) and *TaPHS1* (homologous chromosome pairing and recombination regulator gene) on homoeologous group 7 chromosomes (Lloyd et al. 2007; Devisetty et al. 2010; Khoo et al. 2012). Further research is required to address whether the D genome homoeologs of these genes are QTL candidates for chromosome 3D and 7D.

References

Bartoń K (2019) MuMIn: Multi-Model Inference. R package version 1.43.6. https://CRAN.R-project.org/package=MuMIn. Accessed 16 March 2020

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48. https://www.jstatsoft.org/article/view/v067i01

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics 19:889–890. https://academic.oup.com/bioinformatics/article/19/7/889/197785

Devisetty UK, Mayes K, Mayes S. (2010) The *RAD51* and *DMC1* homoeologous genes of bread wheat: cloning, molecular characterization and expression analysis. BMC Research Notes 3:245. https://bmcresnotes.biomedcentral.com/articles/10.1186/1756-0500-3-245

Gaurav K, Arora S, Silva P, Sanchez-Martin J, Horsnell R, Gao L, Brar GS, Widrig V, Raupp J, Singh N, et al. (2021) Evolution of the bread wheat D-subgenome and enriching it with diversity from *Aegilops tauschii*. bioRxiv. https://www.biorxiv.org/content/10.1101/2021.01.31.428788v2

Gogniashvili M, Jinjikhadze T, Maisaia I, Akhalkatsi M, Kotorashvili A, Kotaria N, Beridze T, Dudnikov A. Ju. (2016) Complete chloroplast genomes of *Aegilops tauschii* Coss. and *Ae. cylindrica* Host sheds light on plasmon D evolution. Current Genetics 62:791–798. https://link.springer.com/article/10.1007/s00294-016-0583-5

Hao M, Luo J, Zeng D, Zhang L, Ning S, Yuan Z, Yan Z, Zhang H, Zheng Y, Feuillet C, *et al.* (2014) *QTug.sau-3B* is a major quantitative trait locus for wheat hexaploidization. G3 Genes Genomes Genetics 4:1943–1953. https://academic.oup.com/g3journal/article/4/10/1943/6025612

Harlan JR, deWet JMJ (1975) On Ö. Winge and a Prayer: The origins of polyploidy. The Botanical Review 41:361–390

Hijmans RJ, Guarino L, Mathur P (2012) DIVA-GIS. http://www.diva-gis.org/Data. Accessed 13 Mar 2022.

Kihara H (1944) Discovery of the DD-analyser, one of the ancestors of *Triticum vulgare*. Agriculture and Horticulture 19:889–890. (in Japanese)

Kihara H (1946) Maturation division in F_1 hybrids between *Triticum dicoccoides* × *Aegilops squarrosa*. La Kromosomo 1:6–11. (in Japanese with English summary)

Kihara H, Yamashita K, Tanaka M. (1965) Morphological, physiological, genetical and cytological studies in *Aegilops* and *Triticum* collected from Pakistan, Afghanistan and Iran. In: Yamashita K (ed) Results of the Kyoto University Scientific Expedition to the Karakoram and Hindukush, 1955, Vol. I, Cultivated plants and their relatives. Kyoto University, Kyoto, pp 1–118

Kilian B, Mammen K, Millet E, Sharma R, Graner A, Salamini F, Hammer K, Özkan H. 2011. *Aegilops*. In: Kole C (ed) Wild Crop Relatives: Genomic and Breeding Resources, Cereals. Springer-Verlag, Berlin, pp 1–76

Korner-Nievergelt F, Roth T, von Felten S, Guelat J, Almasi B, Korner-Nievergelt P (2015) Bayesian Data Analysis in Ecology using Linear Models with R, BUGS and Stan. Elsevier, London

Kosambi DD. (1944) The estimation of map distance from recombination values. Annals of Eugenics 12:172–175

Khoo KHP, Able AJ, Able JA (2012) Poor Homologous Synapsis 1 interacts with chromatin but does not colocalise with ASYnapsis 1 during early meiosis in bread wheat. International Journal of Plant Genomics 2012:514398. https://downloads.hindawi.com/archive/2012/514398.pdf

Lenth RV (2016) Least-squares means: the R package lsmeans. Journal of Statistical Software 69:1–33. https://www.jstatsoft.org/article/view/v069i01

Lloyd AH, Milligan AS, Langridge P, Able JA (2007) *TaMSH7*: a cereal mismatch repair gene that affects fertility in transgenic barley (*Hordeum vulgare* L.) BMC Plant Biology 7:67. https://bmcplantbiol.biomedcentral.com/articles/10.1186/1471-2229-7-67

Matsuoka Y, Nasuda S, Ashida Y, Nitta M, Tsujimoto H, Takumi S, Kawahara T (2013) Genetic basis for spontaneous hybrid genome doubling during allopolyploid speciation of common wheat shown by natural variation analyses of the paternal species. PLoS ONE 8:e68310. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0068310

Matsuoka Y, Takumi S, Kawahara T (2015) Intraspecific lineage divergence and its association with reproductive trait change during species range expansion in central Eurasian wild wheat *Aegilops tauschii* Coss. (Poaceae). BMC Evolutionary Biology 15: 213. https://bmcecolevol.biomedcentral.com/articles/10.1186/s12862-015-0496-9

Matsuoka Y, Mori N (2020) Reproductive and genetic roles of the maternal progenitor in the origin of common wheat (*Triticum aestivum* L.). Ecology and Evolution 10: 13926–13937. https://onlinelibrary.wiley.com/doi/full/10.1002/ece3.6985

Matsuoka Y, Nasuda S (2004) Durum wheat as a candidate for the unknown female progenitor of bread wheat: an empirical study with a highly fertile F_1 hybrid with *Aegilops tauschii* Coss. Theoretical and Applied Genetics 109: 1710–1717. https://link.springer.com/article/10.1007/s00122-004-1806-6

Matsuoka Y, Takumi S (2017) The role of reproductive isolation in allopolyploid speciation patterns: empirical insights from the progenitors of common wheat. Scientific Reports 7: 16004. https://www.nature.com/articles/s41598-017-15919-z

McFadden ES, Sears ER. 1944. The artificial synthesis of *Triticum spelta*. Records of the Genetics Society of America 13:26–27

Mizuno N, Yamazaki M, Matsuoka Y, Kawahara T, Takumi S (2010) Population structure of wild wheat D - genome progenitor *Aegilops tauschii* Coss.: implications for intraspecific lineage diversification and evolution of common wheat. Molecular Ecology 19: 999–1013. https://onlinelibrary.wiley.com/doi/10.1111/j.1365-294X.2010.04537.x

Nakagawa S, Schielzeth H. (2013) A general and simple method for obtaining R^2 from generalized linear mixed - effects models. Methods in Ecology and Evolution 4: 133–142. https://besjournals.onlinelibrary.wiley.com/doi/10.1111/j.2041-210x.2012.00261.x

R Core Team (2021) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 5 October 2021

Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier M-H, Leroy P, Ganal MW (1998) A Microsatellite Map of Wheat. Genetics 149: 2007–2023.

https://academic.oup.com/genetics/article/149/4/2007/6034278?login=true

Saisho D, Takumi S, Matsuoka Y (2016) Salt tolerance during germination and seedling growth of wild wheat *Aegilops tauschii* and its impact on the species range expansion. Scientific Reports 6: 38554. https://www.nature.com/articles/srep38554

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods a Bioconductor package providing PCA methods for incomplete data. Bioinformatics 23: 1164-1167. https://academic.oup.com/bioinformatics/article/23/9/1164/272597

Takumi S, Naka Y, Morihiro H, Matsuoka Y (2009) Expression of morphological and flowering time variation through allopolyploidization: an empirical study with 27 wheat synthetics and their parental *Aegilops tauschii* accessions. Plant Breeding 128: 585–590. https://onlinelibrary.wiley.com/doi/full/10.1111/j.1439-0523.2009.01630.x

Taylor J, Butler D. (2017) R package ASMap: efficient genetic linkage map construction and diagnosis. Journal of Statistical Software 79: 1–29. https://www.jstatsoft.org/article/view/v079i06

van Slageren MW (1994) Wild wheats: a monograph of *Aegilops* L. and *Amblyopyrum* (Jaub. & Spach) Eig (Poaceae). Wageningen Agricultural University, Wageningen

Wang J, Luo M-CC, Chen Z, You FM, Wei Y, Zheng Y, Dvorak J (2013) *Aegilops tauschii* single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. New Phytologist 198: 925–937. https://nph.onlinelibrary.wiley.com/doi/10.1111/nph.12164

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag, New York

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genet. 4: e1000212. https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1000212

Statements and Declarations

Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP 20K20720, JP 19H02935, and JP 19KK0157.

Competing Interests

The author has no relevant financial or non-financial interests to disclose.

Author Contributions

The author contributed to the study conception and design, material preparation, data collection, analysis, and draft writing.

Data Availability

The data that supports the findings of this study are available in the supporting information of this article.

Table 1 *Ae. tauschii* accessions used and the estimated values for their ability to cause hybrid genome doubling

No.	Accession	Lineage	Country of origin	Source [a)]	Estimated value for the ability to cause hybrid genome doubling	Standard error	Asymptotic lower confidence level	Asymptotic upper confidence level
1	AE1090	TauL1	Kazakhstan	IPK	0.42	0.08	0.28	0.57
2	AT55	TauL1	China	OKAYAMA	0.23	0.06	0.14	0.36
3	AT76	TauL1	China	OKAYAMA	0.12	0.04	0.07	0.21
4	AT80	TauL1	China	OKAYAMA	0.24	0.06	0.14	0.38
5	IG126387	TauL1	Turkmenistan	ICARDA	0.31	0.08	0.18	0.48
6	IG127015	TauL1	Armenia	ICARDA	0.32	0.12	0.14	0.58
7	IG131606	TauL1	Kyrgyzstan	ICARDA	0.36	0.08	0.22	0.54
8	IG47259	TauL1	Syria	ICARDA	0.24	0.10	0.09	0.49
9	IG48042	TauL1	India	ICARDA	0.28	0.09	0.14	0.48
10	KU-2001	TauL1	Pakistan	KYOTO/NBRP	0.56	0.09	0.38	0.72
11	KU-2012	TauL1	Afghanistan	KYOTO/NBRP	0.37	0.09	0.21	0.56
12	KU-2025	TauL1	Afghanistan	KYOTO/NBRP	0.21	0.06	0.12	0.36
13	KU-2039	TauL1	Afghanistan	KYOTO/NBRP	0.46	0.09	0.30	0.64

14	KU-2068	TauL1	Iran	KYOTO/NBRP	0.08	0.05	0.02	0.25
15	KU-2132	TauL1	Turkey	KYOTO/NBRP	0.53	0.13	0.28	0.76
16	KU-2136	TauL1	Turkey	KYOTO/NBRP	0.40	0.13	0.19	0.65
17	KU-2144	TauL1	Iran	KYOTO/NBRP	0.32	0.07	0.20	0.48
18	KU-2816	TauL1	Armenia	KYOTO/NBRP	0.56	0.09	0.37	0.73
19	KU-2826	TauL1	Georgia	KYOTO/NBRP	0.49	0.12	0.27	0.72
20	KU-2828	TauL1	Georgia	KYOTO/NBRP	0.50	0.09	0.32	0.68
21	PI476874	TauL1	Afghanistan	USDA	0.29	0.07	0.18	0.43
22	PI486274	TauL1	Turkey	USDA	0.37	0.16	0.14	0.69
23	PI499262	TauL1	China	USDA	0.28	0.08	0.16	0.45
24	PI508262	TauL1	China	USDA	0.35	0.12	0.16	0.60
25	IG47202	TauL2	Azerbaijan	ICARDA	0.24	0.06	0.15	0.38
26	KU-20-10	TauL2	Iran	KYOTO/NBRP	0.34	0.08	0.20	0.52
27	KU-20-8	TauL2	Iran	KYOTO/NBRP	0.10	0.03	0.06	0.18
28	KU-20-9	TauL2	Iran	KYOTO/NBRP	0.40	0.08	0.26	0.55
29	KU-2069	TauL2	Iran	KYOTO/NBRP	0.13	0.06	0.06	0.28
30	KU-2074	TauL2	Iran	KYOTO/NBRP	0.35	0.09	0.20	0.54
31	KU-2075	TauL2	Iran	KYOTO/NBRP	0.30	0.08	0.18	0.47
32	KU-2076	TauL2	Iran	KYOTO/NBRP	0.33	0.07	0.21	0.47
33	KU-2078	TauL2	Iran	KYOTO/NBRP	0.48	0.09	0.32	0.65
34	KU-2079	TauL2	Iran	KYOTO/NBRP	0.25	0.06	0.15	0.37

35	KU-2080	TauL2	Iran	KYOTO/NBRP	0.04	0.01	0.02	0.08
36	KU-2088	TauL2	Iran	KYOTO/NBRP	0.47	0.08	0.32	0.62
37	KU-2090	TauL2	Iran	KYOTO/NBRP	0.29	0.07	0.17	0.45
38	KU-2091	TauL2	Iran	KYOTO/NBRP	0.30	0.09	0.16	0.50
39	KU-2092	TauL2	Iran	KYOTO/NBRP	0.54	0.08	0.39	0.69
40	KU-2093	TauL2	Iran	KYOTO/NBRP	0.34	0.07	0.21	0.48
41	KU-2096	TauL2	Iran	KYOTO/NBRP	0.24	0.07	0.13	0.39
42	KU-2097	TauL2	Iran	KYOTO/NBRP	0.05	0.02	0.03	0.09
43	KU-2098	TauL2	Iran	KYOTO/NBRP	0.20	0.05	0.12	0.33
44	KU-2100	TauL2	Iran	KYOTO/NBRP	0.46	0.10	0.29	0.65
45	KU-2103	TauL2	Iran	KYOTO/NBRP	0.58	0.08	0.43	0.72
46	KU-2104	TauL2	Iran	KYOTO/NBRP	0.57	0.08	0.41	0.73
47	KU-2105	TauL2	Iran	KYOTO/NBRP	0.38	0.08	0.25	0.54
48	KU-2106	TauL2	Iran	KYOTO/NBRP	0.39	0.08	0.25	0.54
49	KU-2109	TauL2	Iran	KYOTO/NBRP	0.32	0.12	0.14	0.58
50	KU-2111	TauL2	Iran	KYOTO/NBRP	0.51	0.12	0.29	0.72
51	KU-2124	TauL2	Iran	KYOTO/NBRP	0.14	0.05	0.07	0.26
52	KU-2126	TauL2	Iran	KYOTO/NBRP	0.24	0.10	0.09	0.49
53	KU-2155	TauL2	Iran	KYOTO/NBRP	0.18	0.06	0.09	0.33
54	KU-2156	TauL2	Iran	KYOTO/NBRP	0.30	0.07	0.18	0.45
55	KU-2158	TauL2	Iran	KYOTO/NBRP	0.39	0.07	0.26	0.54

56	KU-2159	TauL2	Iran	KYOTO/NBRP	0.58	0.09	0.41	0.74
57	KU-2160	TauL2	Iran	KYOTO/NBRP	0.36	0.08	0.23	0.52
58	AE454	TauL3	Georgia	IPK	0.18	0.05	0.10	0.29
59	AE929	TauL3	Georgia	IPK	0.04	0.01	0.02	0.08
60	KU-2829A	TauL3	Georgia	KYOTO/NBRP	0.13	0.04	0.07	0.22

[a)] IPK, Institut für Pflanzengenetik und Kulturpflanzenforschung: ICARDA, International Center for Agricultural Research in the Dry Areas: KYOTO, Plant Germ-plasm Institute of Kyoto University: NBRP, National BioResources Project: OKAYAMA, Kenji Kato, Okayama University: USDA, U. S. Department of Agriculture.

Table 2 Summary of the values for the ability to cause hybrid genome doubling estimated for the 60 *Ae. tauschii* accessions

Category	No. of accessions	Median	Mean	Standard deviation	Range
Overall	60	0.32	0.32	0.14	0.04–0.58
TauL1	24	0.34	0.35	0.13	0.08–0.56
TauL2	33	0.33	0.33	0.15	0.04–0.58
TauL3	3	0.13	0.11	0.07	0.04–0.18

Table 3 Additive multiple-QTL model [a] for the ability to cause hybrid genome doubling in *Ae. tauschii*

QTL name	Chromosome	Position (cM)	LOD score	*P* [b]	%var [c]	Estimated effect (standard error)	Approximate 95% Bayesian credible interval (cM) [flanking marker name [d]]
2D@42.9	2D	42.9	8.34	0.00	9.3	-0.04 (0.01)	37.4 [6040147] 70.9 [12747359]
3D@54.3	3D	54.3	5.98	0.00	6.6	-0.03 (0.01)	43.7 [2257776] 59.4 [gwm52]
3D@105.8	3D	105.8	3.36	0.00	3.6	-0.02 (0.01)	94.2 [7345387] 140.3 [1106895]
6D@86.6	6D	86.6	3.34	0.00	3.6	-0.02 (0.01)	72.4 [1220491] 97.5 [1228345]
7D@161.0	7D	161.0	5.39	0.00	5.9	-0.03 (0.01)	142.7 [4540106] 164.8 [3955347]

[a] LOD score (relative to the no QTL model), %var by all terms in the model, and *P*-value based on the LOD score are 27.8, 36.8, and 0, respectively.

[b] *P* denotes the drop-one-QTL-at-a-time analysis of variance *P*-values for the LOD peaks.

[c] %var denotes the estimated proportion of the phenotype variance that is explained.

[d] The numbers indicate the DArTseq maker IDs. The microsatellite marker gwm52 is described in Röder et al. (1998).

Figure Captions

Fig. 1

a Plot of the first two components from the probabilistic principal component analysis. The first component PC1 (*x*) and second component (PC2) (*y*) account for 32.4% and 9.1% of the total variance, respectively. Circles, crosses, and squares indicate the TauL1, TauL2, and TauL3 accessions, respectively. **b** *Ae. tauschii* lineage-wise box and dot plots of the estimated ability to cause hybrid genome doubling

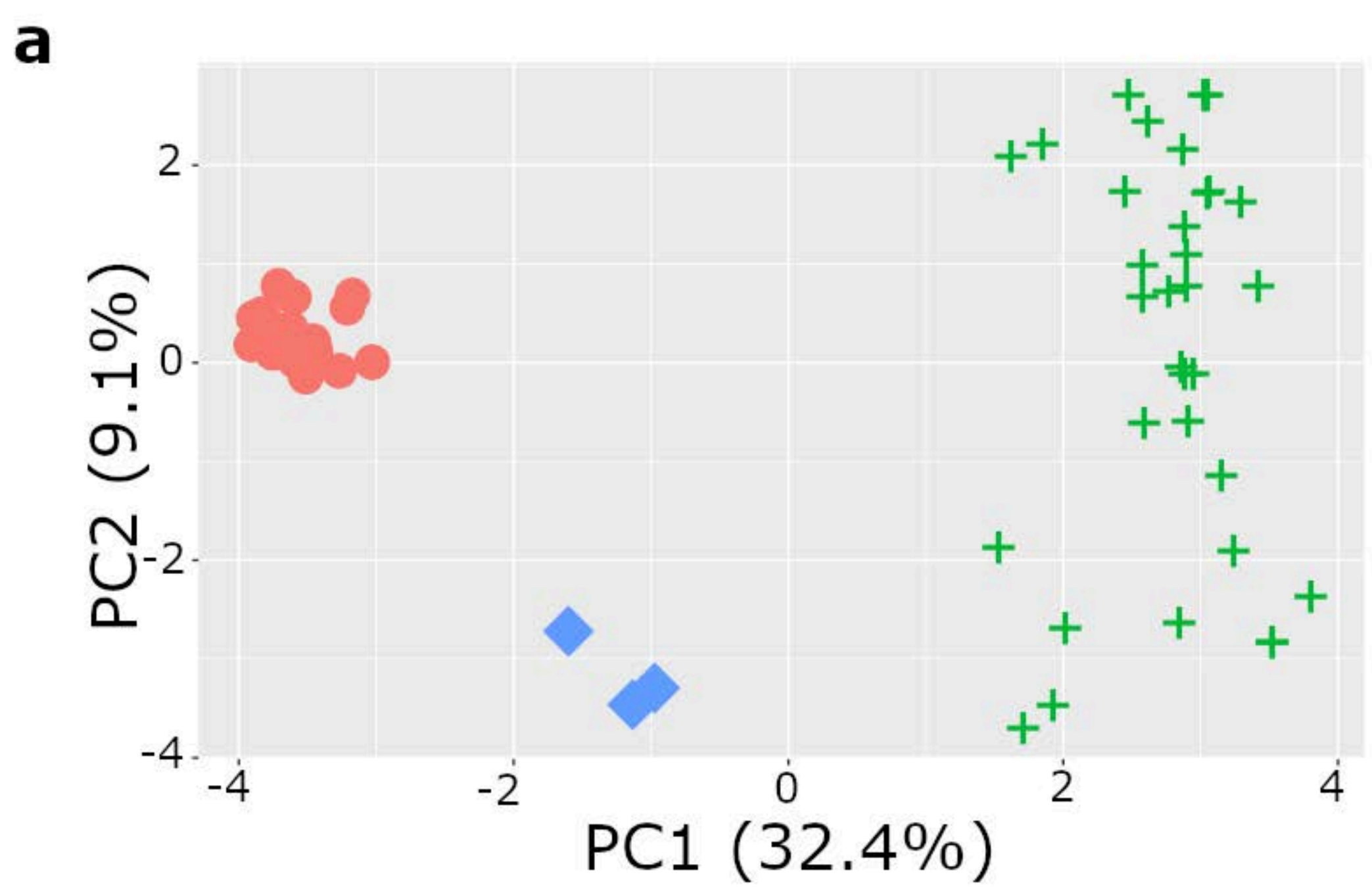
Fig. 2

Geographic patterns for the ability to cause hybrid genome doubling in *Ae. tauschii*. Circles, crosses, and squares denote TauL1, TauL2, and TauL3, respectively. Each accession is colored according to the ability value key. The three TauL1 accessions (AT55, AT76, and AT80) from central China are not shown

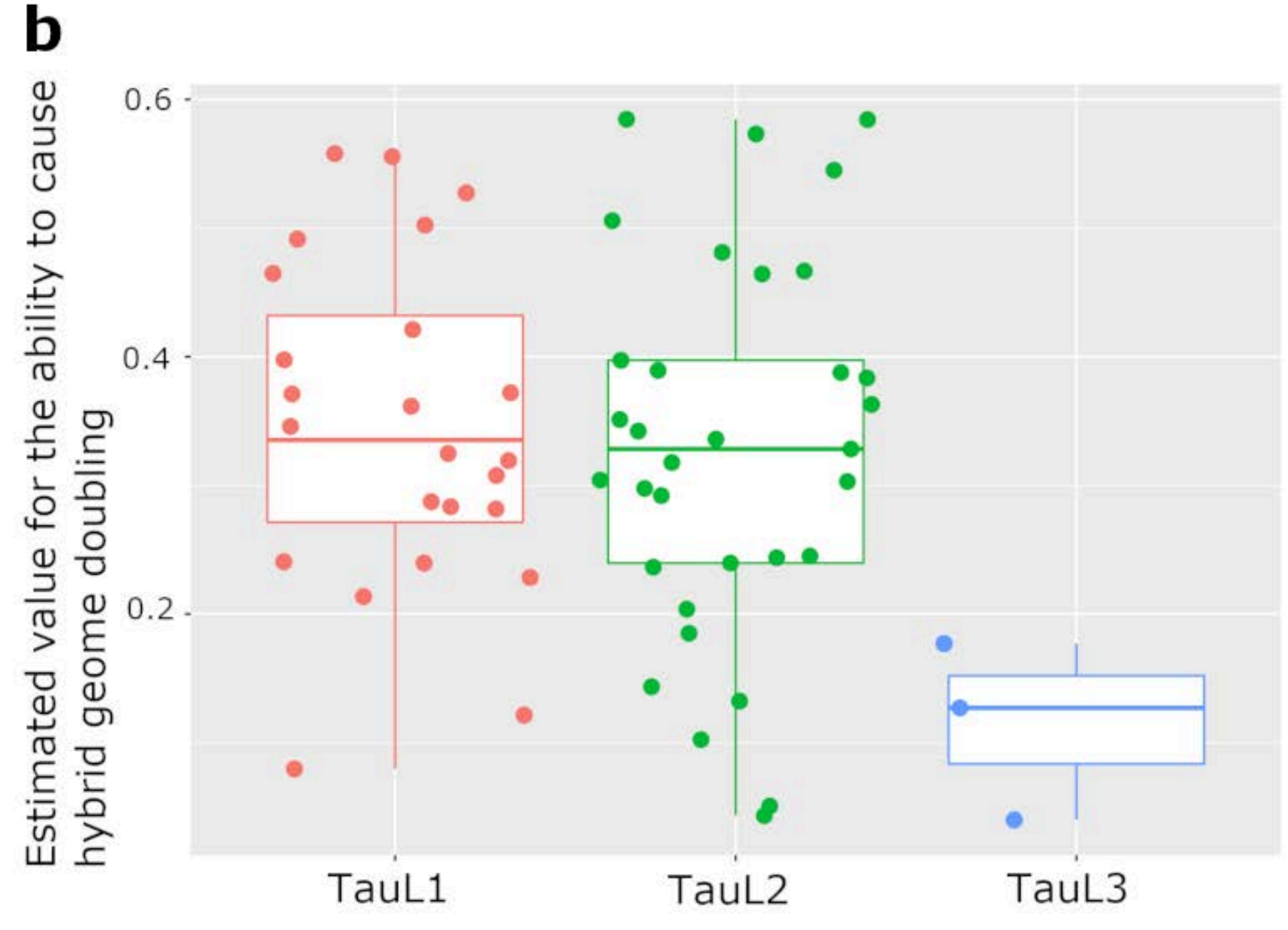
Fig. 3

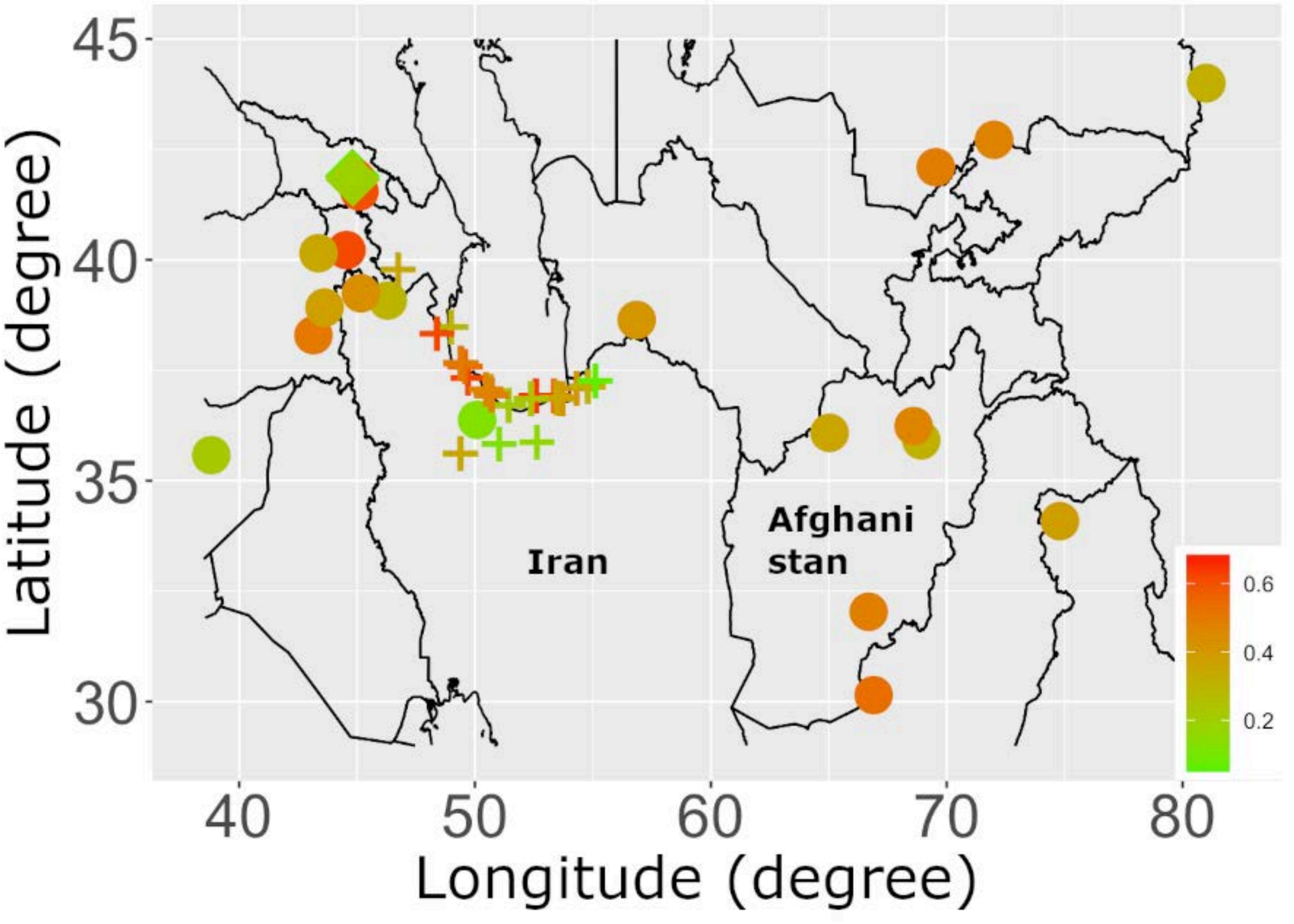
QTL analysis of the ability to cause hybrid genome doubling in *Ae. tauschii*. The horizontal dashed line represents a significant LOD score determined by permutation analysis

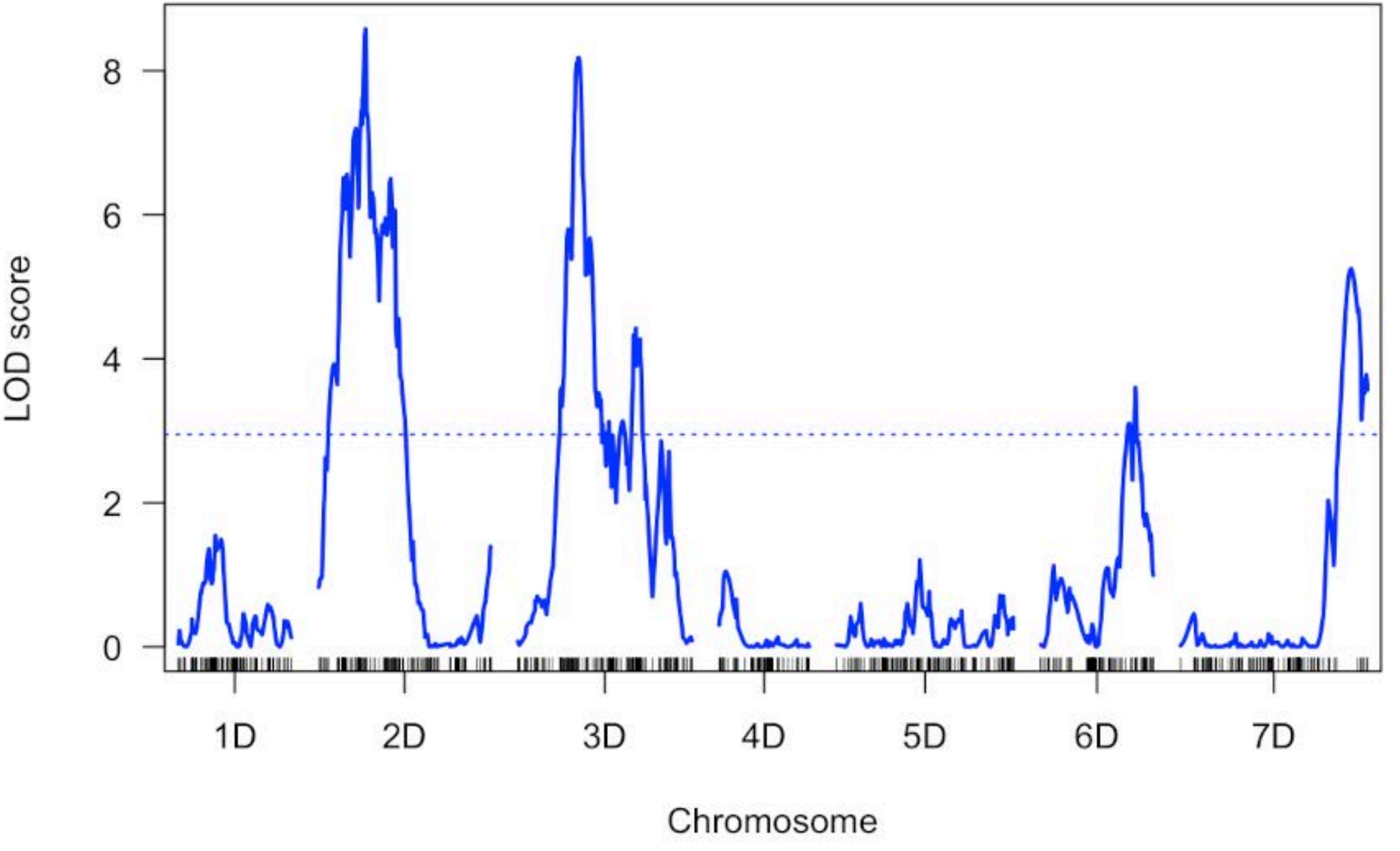
a


2
0
-2
-4
PC2 (9.1%)
-4
-2
0
2
4
PC1 (32.4%)

b


Estimated value for the ability to cause hybrid geome doubling
0.6
0.4
0.2
TauL1
TauL2
TauL3


45
40
35
30
Latitude (degree)
40
50
60
70
80
Longitude (degree)
Iran
Afghani
stan
0.6
0.4
0.2


LOD score
0
2
4
6
8
1D
2D
3D
4D
5D
6D
7D
Chromosome