Annals of Botany **118**: 941–955, 2016
doi:10.1093/aob/mcw137, available online at www.aob.oxfordjournals.org

ANNALS OF BOTANY
Founded 1887

Ecological characteristics and *in situ* genetic associations for yield-component traits of wild *Miscanthus* from eastern Russia

Lindsay V. Clark[1], Elena Dzyubenko[2], Nikolay Dzyubenko[2], Larisa Bagmet[2], Andrey Sabitov[2], Pavel Chebukin[2], Douglas A. Johnson[3], Jens Bonderup Kjeldsen[4], Karen Koefoed Petersen[5], Uffe Jørgensen[4], Ji Hye Yoo[6], Kweon Heo[6], Chang Yeon Yu[6], Hua Zhao[7], Xiaoli Jin[8], Junhua Peng[9], Toshihiko Yamada[10] and Erik J. Sacks[1,*]

[1]Department of Crop Sciences, University of Illinois at Urbana-Champaign, 1201 W. Gregory Drive, Urbana, IL 61802, USA, [2]Vavilov All-Russian Institute of Plant Genetic Resources, 42–44 Bolshaya Morskaya Street, 190000 St. Petersburg, Russia, [3]USDA-ARS Forage and Range Research Lab, Utah State University, Logan, UT 84322-6300, USA, [4]Department of Agroecology, Aarhus University, Blichers Allé 20, DK-8830 Tjele, Denmark, [5]Department of Food Science, Aarhus University, Kirstinebjergvej 10, DK-5792 Årslev, Denmark, [6]Kangwon National University, Chuncheon, Gangwon 200-701, South Korea, [7]College of Horticulture and Forestry Science, Huazhong Agricultural University, Wuhan, Hubei 430070, China, [8]Agronomy Department, Key Laboratory of Crop Germplasm Resource of Zhejiang Province, Zhejiang University, Hangzhou 310058, China, [9]Science and Technology Center, China Seed Group Co. Ltd, Wuhan, Hubei 430040, China and [10]Field Science Center for Northern Biosphere, Hokkaido University, Sapporo, Hokkaido 060-0810, Japan
**For correspondence. E-mail esacks@illinois.edu*

Received: 13 February 2016 Returned for revision: 31 March 2016 Accepted: 24 May 2016 Published electronically: 24 July 2016

• **Background and aims** *Miscanthus* is a genus of perennial C_4 grasses native to East Asia. It includes the emerging ligno-cellulosic biomass crop *M.* ×*giganteus*, a hybrid between *M. sinensis* and *M. sacchariflorus*. Biomass yield and cold tolerance are of particular interest in *Miscanthus*, given that this crop is more temperate adapted than its C_4 relatives maize, sorghum and sugarcane.
• **Methods** A plant exploration was conducted in eastern Russia, at the northern extreme of the native range for *Miscanthus*, with collections including 174 clonal germplasm accessions (160 *M. sacchariflorus* and 14 *M. sinensis*) from 47 sites. Accessions were genotyped by restriction site-associated DNA sequencing (RAD-seq) and plastid microsatellites.
• **Key Results** *Miscanthus sinensis* was found in maritime climates near Vladivostok (43·6°N) and on southern Sakhalin Island (46·6°N). *Miscanthus sacchariflorus* was found inland at latitudes as high as 49·3°N, where *M. sinensis* was absent. Most *M. sacchariflorus* accessions were diploid, but approx. 2 % were tetraploids. Molecular markers revealed little population structure (Jost's $D < 0{\cdot}007$ among diploid groups) but high genetic diversity (expected heterozygosity = 0·14) within the collection of Russian *M. sacchariflorus*. Genome-wide association (GWA) analysis for traits measured at the collection sites revealed three *M. sacchariflorus* single nucleotide polymorphisms (SNPs) significantly associated with the number of stems per unit area, one with height and one with basal stem diameter; three were near or within previously described sorghum quantitative trait loci for related traits.
• **Conclusions** This new *Miscanthus* germplasm collection from eastern Russia will be useful for breeding *Miscanthus* and sugarcane cultivars with improved adaptation to cold. Moreover, a strategy is proposed to facilitate the rapid utilization of new germplasm collections: by implementing low-cost SNP genotyping to conduct GWA studies of phenotypic data obtained at collection sites, plant breeders can be provided with actionable information on which accessions have desirable traits and alleles.

Key words: Chloroplast, genome-wide association analysis (GWAS), germplasm, *Miscanthus sacchariflorus*, *Miscanthus sinensis*, population genetics, restriction site-associated DNA sequencing (RAD-seq), Russia, single nucleotide polymorphism (SNP).

INTRODUCTION

Miscanthus is a genus of perennial C_4 East Asian grasses that has attracted considerable recent interest as a biomass crop for energy, heat and fibre. The most commonly grown clone for biomass has been distributed under many cultivar names, but is genetically identical to the *M.* ×*giganteus* type specimen described by Hodkinson and Renvoize (2001; Głowacka *et al*., 2015). We will refer to this clone as *M.* ×*giganteus* '1993-1780' in reference to its accession number in the Kew Living Collection. *Miscanthus* is more temperate adapted (growing as far as 50°N in Russia; Hodkinson *et al*., 2015) than its C_4 relatives sugarcane, sorghum and maize, yet *M.* ×*giganteus* '1993-1780' originated from sub-tropical southern Japan (35°N, assuming origination from Yokohama; Greef *et al*., 1997) and is therefore unlikely to represent the maximum cold tolerance of the genus (Głowacka *et al*., 2014). Insufficient winter hardiness has been documented for first-year plantings

of *M*. ×*giganteus* '1993-1780' in parts of northern Europe, whereas better winter hardiness was observed in other *Miscanthus* accessions (Clifton-Brown and Lewandowski, 2000; Farrell *et al*., 2006). In Urbana, Illinois, we also observed severe damage to first-year plantings of *M*. ×*giganteus* '1993-1780' in 2014 after an especially cold winter, whereas other, recently bred, *M*. ×*giganteus* genotypes in the same trial were not so adversely affected (unpubl. data). Several studies have also shown that, although *M*. ×*giganteus* '1993-1780' has higher growth and photosynthesis at temperatures below 14 °C than other species in the tribe Andropogoneae, there is variation for these traits among and within *Miscanthus* species (Purdy *et al*., 2013; Friesen *et al*., 2014; Głowacka *et al*., 2014). Thus, a major goal of *Miscanthus* breeding is the production of biomass cultivars with improved cold hardiness and chilling-tolerant photosynthesis. A large public gene bank for *Miscanthus* does not yet exist, but several academic research groups have recently imported germplasm to Europe and North America from the native range, characterized those collections with genotyping by sequencing (GBS) or restriction site-associated DNA sequencing (RAD-seq) approaches (Slavov *et al*., 2013; Barth *et al*., 2014; Clark *et al*., 2014, 2015) and performed genome-wide association studies (GWAS) using phenotypic data from replicated field trials (Slavov *et al*., 2014).

The two most widely distributed species of *Miscanthus*, *M. sinensis* and *M. sacchariflorus*, are also the two parent species of the hybrid *M*. ×*giganteus*. *Miscanthus sinensis* and *M. sacchariflorus* are phylogenetically distinct from each other within *Miscanthus sensu stricto* (Hodkinson *et al*., 2002), and differ in their growth habits, with *M. sinensis* being caespitose (tufted), and *M. sacchariflorus* having a spreading habit due to long rhizomes (Chae *et al*., 2014). *Miscanthus sacchariflorus* is typically found in riparian and wetland habitats, unlike *M. sinensis* and other *Miscanthus* species, which are more commonly found on hilly sites with good drainage (Sacks *et al*., 2013). Although *M. sinensis* and *M. sacchariflorus* exhibit considerable overlap in their geographic distribution, the range of *M. sinensis* extends further south (Clifton-Brown *et al*., 2008; Sacks *et al*., 2013).

The most promising region from which to collect cold-adapted *Miscanthus* germplasm is expected to be at the northernmost extent of its native range, in eastern Russia. However, previous collections have been limited; Yook *et al*. (2014) included three Russian *M. sacchariflorus* accessions from 43·3–43·6°N, 131·4–132·1°E and four Russian *M. sinensis* accessions from 42·4–43·4°N, 130·4–132·0°E in a phylogenetic study that was focused on *Miscanthus* from Korea. Reports from botanical surveys, however, have indicated the presence of *M. sacchariflorus* as far north as Lake Bolon (49·8°N, 136·4°E) and Blagoveshchensk (50·3°N, 127·3°E), and *M. sinensis* primarily in maritime regions such as southern Sakhalin (to 47·4°N) and near Vladivostok (43·1°N, 131·9°E), with rare reports near Lake Khanka (45°N, 132°E) and Khabarovsk (48·5°N, 135·1°E) (Herbarium of the Komarov Botanical Institute; Harkevich, 1985).

Given the limited number of *Miscanthus* accessions from Russia's Far East that have been conserved in germplasm collections previously, and the potential utility of these populations for improving cold tolerance, we conducted a plant exploration as a joint effort of the NI Vavilov Research Institute of Plant Industry, the USDA-ARS and the University of Illinois. Materials from our exploration are being maintained by the US National Plant Germplasm System and will be made available as they are propagated. To facilitate rapid use of the accessions by plant breeders, we also recorded phenotypic characteristics at the collection sites and subsequently used RAD-seq to obtain single nucleotide polymorphism (SNP) genotypes for each accession. Analyses of phenotypic, ecological and genetic data were conducted with the following objectives: (1) to identify potential geographic, climatic, ecological and anthropogenic influences on the ranges of *M. sinensis* and *M. sacchariflorus*; (2) to quantify phenotypic variation for biomass traits; (3) to understand population structure of Russian *M. sinensis* and *M. sacchariflorus* in order to identify genetic groups for germplasm conservation, association analysis and potential sources of heterosis; (4) to compare genetic diversity of Russian *Miscanthus* with previously characterized *Miscanthus* populations in order to assess its relative utility for breeding; and (5) to investigate the potential to identify quantitative trait loci (QTLs) for traits of agronomic interest via GWAS of *in situ* phenotypic data obtained during germplasm collection. Although it is currently unusual to perform GWAS for crop germplasm without phenotypic data from replicated field trials, there are previous examples of successful association studies using phenotypic data from natural populations, particularly in forestry (Parchman *et al*., 2012; Budde *et al*., 2014) and animal ecology (Johnston *et al*., 2011, 2014; Narum *et al*., 2013). Here we demonstrate that value can be added to a crop germplasm collection by combining *in situ* phenotype data and inexpensive genotyping data in a GWAS.

MATERIALS AND METHODS

Field collections and observations

From 3 to 29 September 2012, a plant exploration for *Miscanthus* germplasm was conducted in eastern Russia (Fig. 1). Additional details of our sampling route are provided in Supplementary Data Materials and Methods. An initial collection near Nevelsk on Sakhalin Island was followed by collections on the mainland. On the mainland, we explored the following areas: (1) the north–south corridor between Khabarovsk and Vladivostok along highway M60, which, north of Lake Khanka, was near and parallel to the Ussuri River; (2) north-east from Khabarovsk along the Amur River on its eastern side via highway P454 until the Gur River; (3) west from Khabarovsk on the M58 to Birobidzhan and from Birobidzhan south along the Bira River until near to the Amur River, which is the border with China; (4) the area south of Lake Khanka to the city of Ussuriysk; and (5) on Russky Island just south of Vladivostok. On four sections of highway, we noted the number of distinct stands of *Miscanthus* at least 100 m apart from each other, and calculated the average number of stands per kilometre (Fig. 1A, inset).

Collections were made at 48 sites. At 47 sites, from one to 13 plants were measured and rhizomes collected. We endeavoured to obtain both rhizomes (live plants) and seed at each site. Seed was collected from 41 sites; the number of plants per accession that contributed seed ranged from one to 100. Material from the seed collections was not included in the genetic studies reported here. For each plant from which rhizomes were collected, we measured the height of the tallest stem

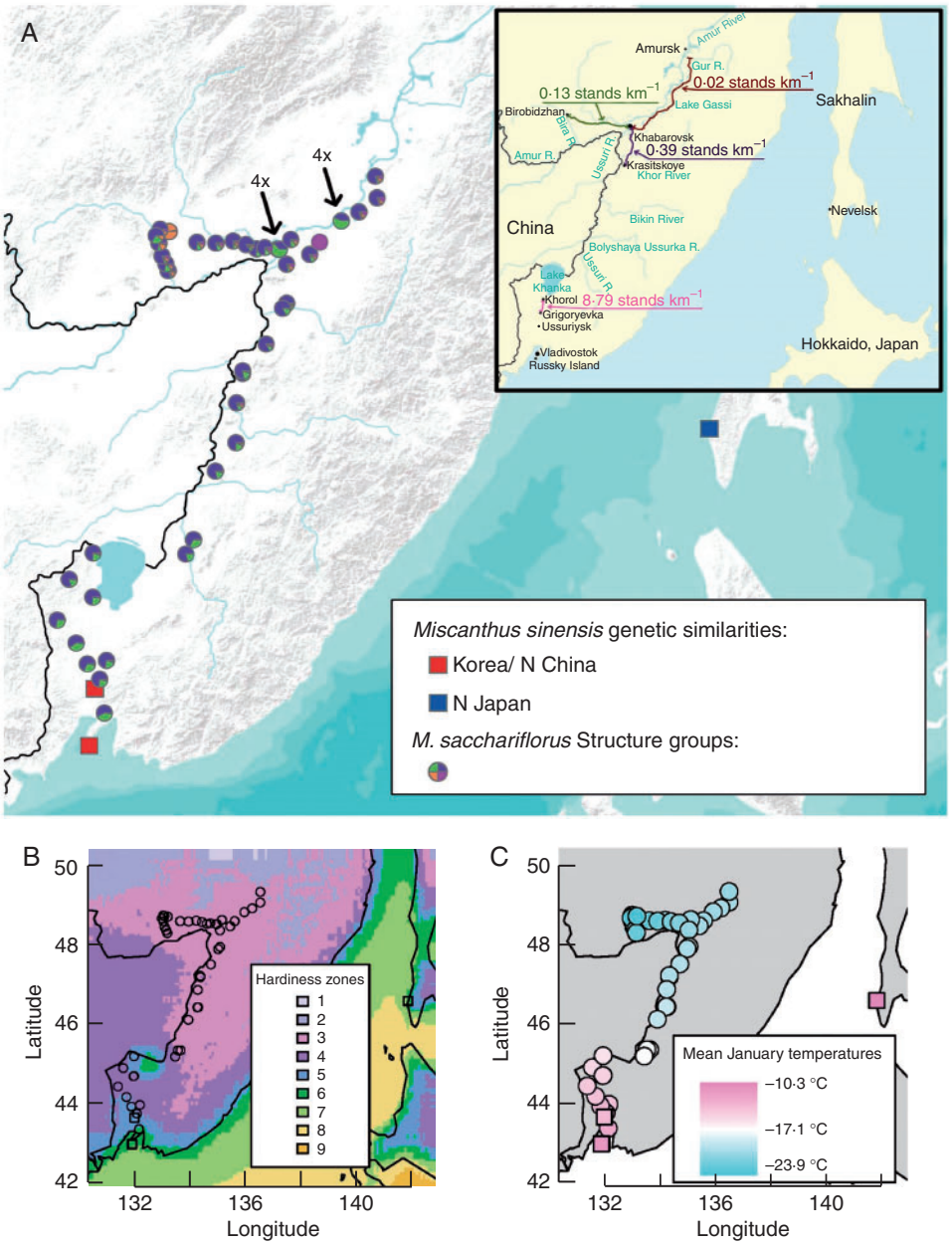

FIG. 1. Collection locations, temperature data and population structure of *Miscanthus* in eastern Russia. (A) *Miscanthus sacchariflorus* collection sites are indicated with pies, with colours representing assignment (Q values) to four genetic clusters determined by Structure, using 29 260 RAD-seq SNPs across 160 individuals. Q values were highly similar among individuals within each site. *Miscanthus sinensis* individuals are indicated by squares, and colours indicate assignment to previously identified populations in East Asia (Clark *et al*., 2014) by discriminant analysis of principal components using 24 132 RAD-seq SNPs. The inset shows frequency of stands of *M. sacchariflorus* observed at least 100 m apart along four stretches of highway. (B) USDA plant hardiness zones, based on data from 1982–2011 from NAPPFAST (2012). (C) Mean temperatures during January based on data from 1950–2000 available at WorldClim (Hijmans *et al*., 2005).

(stem height), the diameter of the internode at the base of the tallest stem (stem diameter) and the number of stems in a 0·25 m^2 area (stem density). We also recorded the approximate number of plants found and sampled at the collection site, the site size (in m^2), the frequency of *Miscanthus* at the site, aspect (compass direction of slope), slope, and site physical and vegetative descriptions (Supplementary Data Dataset S1). Rhizomes were collected from 182 plants and were shipped to USDA-APHIS-PPQ in Beltsville, MD for US quarantine, and a backup set was also sent to Aarhus University in Foulum, Denmark. At the time of publication of this paper, 158 clones of *M. sacchariflorus* (out of 165 collected) and 14 of *M. sinensis* (out of 17 collected) have been released from US quarantine. Material from seed increase will be made available from USDA (http://www.ars-grin.gov; accessions W6 49502–49505; Dataset S1). Leaf tissue from each plant at Foulum was freeze-dried and shipped to the University of Illinois for DNA extraction.

Flow cytometry

Flow cytometry was performed at Aarhus University using a protocol modified from Petersen *et al*. (2003). Samples of *M. sinensis* [MS-104 (Jørgensen, 1997) or MS-110 (Petersen *et al*., 2003)] that were known to be diploid were used as an internal standard. Additional details are provided in Supplementary Data Materials and Methods.

Genotyping

A total of 174 plants (160 *M. sacchariflorus* and 14 *M. sinensis*) from 47 sites (44 *M. sacchariflorus* and three *M. sinensis*) were included in the genotyping analysis. Nuclear SNPs were obtained by RAD-seq (Poland *et al*., 2012*a*). Entries were also screened with ten plastid microsatellites (de Cesare *et al*., 2010; Jiang *et al*., 2012) using previously described protocols (Clark *et al*., 2014). Additional details are provided in Supplementary Data Materials and Methods. Sequences from RAD-seq have been deposited at the NCBI Sequence Read Archive (accession SRP063572).

Data analysis

To obtain SNP genotypes for 160 Russian *M. sacchariflorus* individuals, we used the UNEAK pipeline in TASSEL 3·0·162 (Lu *et al*., 2013). Filtering for SNPs with a minimum call rate of 0·5 and a minimum minor allele frequency of 0·01 produced 29 260 SNPs (Supplementary Data Dataset S2). To obtain SNP genotypes for *M. sinensis*, we ran the UNEAK pipeline under the same conditions on the 14 Russian *M. sinensis* individuals and 595 individuals from a previous study (576 *M. sinensis* individuals plus 19 *M. sacchariflorus* and hybrid individuals; Clark *et al*., 2014), and removed SNPs that appeared heterozygous in double haploid lines, yielding 24 132 SNPs.

Russian *M. sacchariflorus* genotypes were analysed in Structure 2·3·4 (Falush *et al*., 2003) to assess population structure. Three runs each at $K = 1$ through $K = 10$ were performed under default conditions with a burn-in of 10 000 reps, followed by 50 000 reps. The number of clusters (K) was selected by analysing the results in Structure Harvester (Earl and VonHoldt, 2011) using the Evanno method (Evanno *et al*., 2005). Mean Structure results (Q values) for each collection site were plotted using ArcGIS 10·1 (ESRI, Redlands, CA, USA).

Given the weak geographic structuring in *M. sacchariflorus* detected by Structure, Mantel tests were used to assess the hypothesis of genetic isolation by distance among Russian *M. sacchariflorus* individuals. Geographic distances between collection sites were calculated using the R (R Core Team, 2014) package geosphere (Hijmans, 2014). SNP data were imported in numeric (0 or 2 for homozygotes, 1 for heterozygotes) format, with individuals as rows and markers as columns ('hapMap2genlight' function; Clark *et al*., 2014), and genetic distances between individuals were calculated as Euclidian distances between rows using the 'dist' function in R. Correlation between the geographic and genetic distance matrices was then tested using the 'mantel.rtest' function in the R package ade4 (Chessel *et al*., 2004) with 999 permutations. To make a matrix of genetic distances based on plastid haplotypes, plastid microsatellite data were imported into the R package polysat (Clark and Jasieniuk, 2011) and dissimilarities were calculated with the 'meandistance.matrix' and 'Lynch.distance' functions. The genetic distance matrix based on plastid data was then used in a Mantel test, comparing it with the geographic distance matrix as was done with SNP data.

Suggested seed increase groups for *M. sacchariflorus* were selected based on geography and ploidy. Genetic differentiation between seed increase groups was estimated with Jost's *D* across 29 260 RAD-seq SNPs using the 'D_Jost' function in the R package mmod (Jost, 2008; Winter, 2012).

To construct a plastid haplotype network for Russian *M. sacchariflorus*, the distances between haplotypes that were calculated with polysat were analysed with a modified source code from the R package pegas (Paradis, 2010; Clark *et al*., 2014). Additional connections were added to the network where two haplotypes only differed at one locus.

To assess the relationship between the Russian *M. sinensis* individuals and previously identified genetic clusters (Clark *et al*., 2014), and determine hybrid status, *M. sinensis* SNP data were evaluated by principal component analysis (PCA) and discriminant analysis of principal components (DAPC) using the R package adegenet (Jombart *et al*., 2010; Jombart and Ahmed, 2011). A DAPC was also performed on the *M. sacchariflorus* SNP dataset but was uninformative (see the Results).

For genome-wide association analysis of the phenotypic traits in *M. sacchariflorus*, 64-nucleotide RAD tags output by the UNEAK pipeline were aligned to the *Sorghum bicolor* genome version 2·0 (Paterson *et al*., 2009; available at http://phytozome.jgi.doe.gov) using Bowtie2 (Langmead and Salzberg, 2012) with parameters set for high sensitivity (-D 20 -R 3 -N 1 -L 18 -i S,1,0·50 –local), given that we were making cross-genus sequence alignments. A total of 16 137 pairs of RAD tags (55 % of the total) both aligned to the same location in the sorghum genome and therefore were retained at this step. Two individuals out of 160 were removed for having >50 % missing data, and three tetraploid individuals were removed, leaving 155 individuals for analysis. SNPs were then filtered to have ≤50 % missing data, minor allele frequency ≥0·05 and observed heterozygosity ≤50 %, leaving 5971 SNPs, which were analysed to detect SNP–trait associations. To reduce environmental error in traits that were significantly correlated with latitude, linear models were fit with the trait (after log transformation if applicable) as the dependent variable and latitude as the independent variable. For each individual, the predicted phenotype based on latitude was subtracted from the actual phenotype value to produce the adjusted phenotype value, which was then evaluated by GWAS. Traits evaluated were stem height (untransformed and adjusted by latitude), stem diameter (log transformed and adjusted by latitude) and number of stems per 0·25 m^2 (log transformed). Of the 155 accessions that were analysed in GWAS, 91 accessions each fell into one of 30 groups of related individuals, leaving 94 unrelated individuals and groups of individuals (Supplementary Data Fig. S1; relatedness was included in the GWAS model). GWAS using the Q–K mixed model method (Yu *et al*. 2006), with the relationship matrix and the first four principal components included in the model, was performed both in rrBLUP (Endelman, 2011) and in TASSEL5 (Bradbury *et al*., 2007). In

rrBLUP, the additive relationship matrix was calculated using the EM imputation method (Poland *et al*., 2012*b*), and in TASSEL5 the kinship matrix was calculated using scaled identity-by-state. In both programs, the P3D (population parameters previously determined) method was not used, given the relatively small number of SNPs being evaluated, and the K matrix was not compressed, given the small number of individuals in the study. The FDR (false discovery rate) method (Benjamini and Hochberg, 1995) was used to correct *P*-values for multiple testing. SNP–trait associations were considered significant if *P*-values were $<0{\cdot}05$ after FDR correction. To estimate heritability based on the SNP data, the 'kin.blup' function of rrBLUP was used to fit each trait to the additive relationship matrix and perform REML (restricted maximum likelihood) estimations of genetic variance and error variance. The genetic variance was then divided by the sum of the genetic and error variances.

RESULTS

Ecology, ploidy and genetics of M. sinensis

We encountered *M. sinensis* at only three sites in eastern Russia (Fig. 1A): southern Sakhalin Island (46·58°N, 141·84°E), near highway M60 between Vladivostok and Ussuriysk (43·65°N, 132·00°E) and on Russky Island just south of Vladivostok (42·98°N, 131·91°E). On Russky Island, *M. sinensis* was especially abundant, and a sample was taken from a stand of >1000 individuals (Fig. 2A). All three of the sites where *M. sinensis* was found have a maritime-influenced climate with relatively mild winters (USDA hardiness zones 5 and 6; average annual extreme minimum temperature of –29 to –18 °C), which probably contributed to the success of *M. sinensis* at these sites, in contrast to regions further inland (USDA hardiness zones 3 and 4; –40 to –29 °C) where we only found *M. sacchariflorus* (Fig 1B).

Using nuclear SNPs on our 14 Russian *M. sinensis* and 595 previously genotyped *Miscanthus* individuals from East Asia (Clark *et al*., 2014), PCA and DAPC revealed that the *M. sinensis* individuals from near Vladivostok on the mainland and from Russky Island were most similar to individuals from northern China (above 36°N) and Korea, and that the *M. sinensis* individuals from Sakhalin were most similar to individuals from northern Japan (Hokkaido and northern Honshu; Fig. 1A; Supplementary Data Fig. S2). None of the Russian *M. sinensis* that were collected showed evidence of hybridization with *M. sacchariflorus* (Fig. S2). Thus, the *M. sinensis* individuals that we found in eastern Russia were most genetically similar to geographically adjacent *M. sinensis* in neighbouring countries. Analysis of the maternally inherited plastid SSR (simple sequence repeat) alleles was consistent with the results of the nuclear SNP data. The *M. sinensis* individuals from Sakhalin had the most common plastid haplotype found in Hokkaido, Japan (haplotype C in Clark *et al*., 2014). Two out of the three *M. sinensis* individuals from north of Vladivostok also had haplotype C, which is also common in South Korea. However, the third *M. sinensis* individual from north of Vladivostok, as well as the individual from Russky Island, had haplotype G, which is uncommon in *M. sinensis* and previously found only in the population from northern China (Clark *et al*., 2014). Of the 13 *M. sinensis* individuals that were tested by flow cytometry, all were found to be diploid. *Miscanthus sinensis* plants from Sakhalin were shorter but had thicker stems than those near Vladivostok (Fig. 3A, B, G, H).

Ecology of M. sacchariflorus

Miscanthus sacchariflorus was found throughout most of the area that we explored. However, we did not observe any *M. sacchariflorus* north and east of the Anyuy River (49·33°N, 136·52°E) as we travelled along the eastern side of the Amur River on road P454 from Khabarovsk to the Gur River. Along the Amur River watershed, we found *M. sacchariflorus* as far inland (west) as Birobidzhan (48·67°N, 132·98°E); there may have been more *M. sacchariflorus* further west along the Bira River, but time constraints prevented us from travelling any further. The frequency of *M. sacchariflorus* also varied considerably across our study area (Fig. 1A, inset). The greatest frequency of *M. sacchariflorus* was observed near the south and west of Lake Khanka, with 8·79 stands km^{-1} (Figs 1A and 2B, C), whereas the lowest frequency was from Khabarovsk to the Gur River along the Amur River basin, with only 0·02 stands km^{-1} observed. West of Khabarovsk to Birobidzhan, the frequency of stands was higher ($0{\cdot}13\ km^{-1}$) than east of Khabarovsk. However, just south of Khabarovsk, near the Ussuri River, the stand density of *M. sacchariflorus* was higher still at 0·39 stands km^{-1}. Though the common name of *M. sacchariflorus* is Amur silvergrass, we observed that it was more common near the Ussuri River and Lake Khanka than along the Amur River. In the Lake Khanka region, we also observed rectangular hayfields that had high population densities of flowering *M. sacchariflorus* (Fig. 2D). Land disturbance in the Lake Khanka area, especially deforestation to make pasture and haylands for raising cows (Fig. 2D, E), may have contributed to the large population of *M. sacchariflorus* observed. Around 160 years ago, Maximowicz (1859) observed that *M. sacchariflorus* growing in the prairies of the Russian Far East and north-east China was used for hay and grazing; thus, there has been a longstanding interaction between livestock farmers in this region and indigenous populations of *M. sacchariflorus*. East of Khabarovsk in the Amur River basin, *M. sacchariflorus* was absent near the villages we visited; however, we saw large populations on nearby islands surrounded by the Amur River and without access by road (Fig. 2F), suggesting that human disturbance may have negatively impacted *M. sacchariflorus* in adjacent populated areas.

Miscanthus sacchariflorus was commonly found in relatively flat areas near surface water but not on flooded land. Less frequently, we found *M. sacchariflorus* on land with $\geq 10°$ slope, or not near surface water. Soil texture where *M. sacchariflorus* grew was typically loam or sandy loam, and less frequently clay or clay–loam. *Miscanthus sacchariflorus* was found in open areas, in close association with other grasses and forbs, and sometimes adjacent to forest, but not under the forest canopy (Fig. 2G). A rhizomatous grass that was found growing near *M. sacchariflorus* at more than half (52 %) of the collection sites was identified as *Hemarthria sibirica* using ITS (internal transcribed spacer) sequencing (identical to NCBI accession KF163639·1 at 622 out of 623 nucleotides; Supplementary

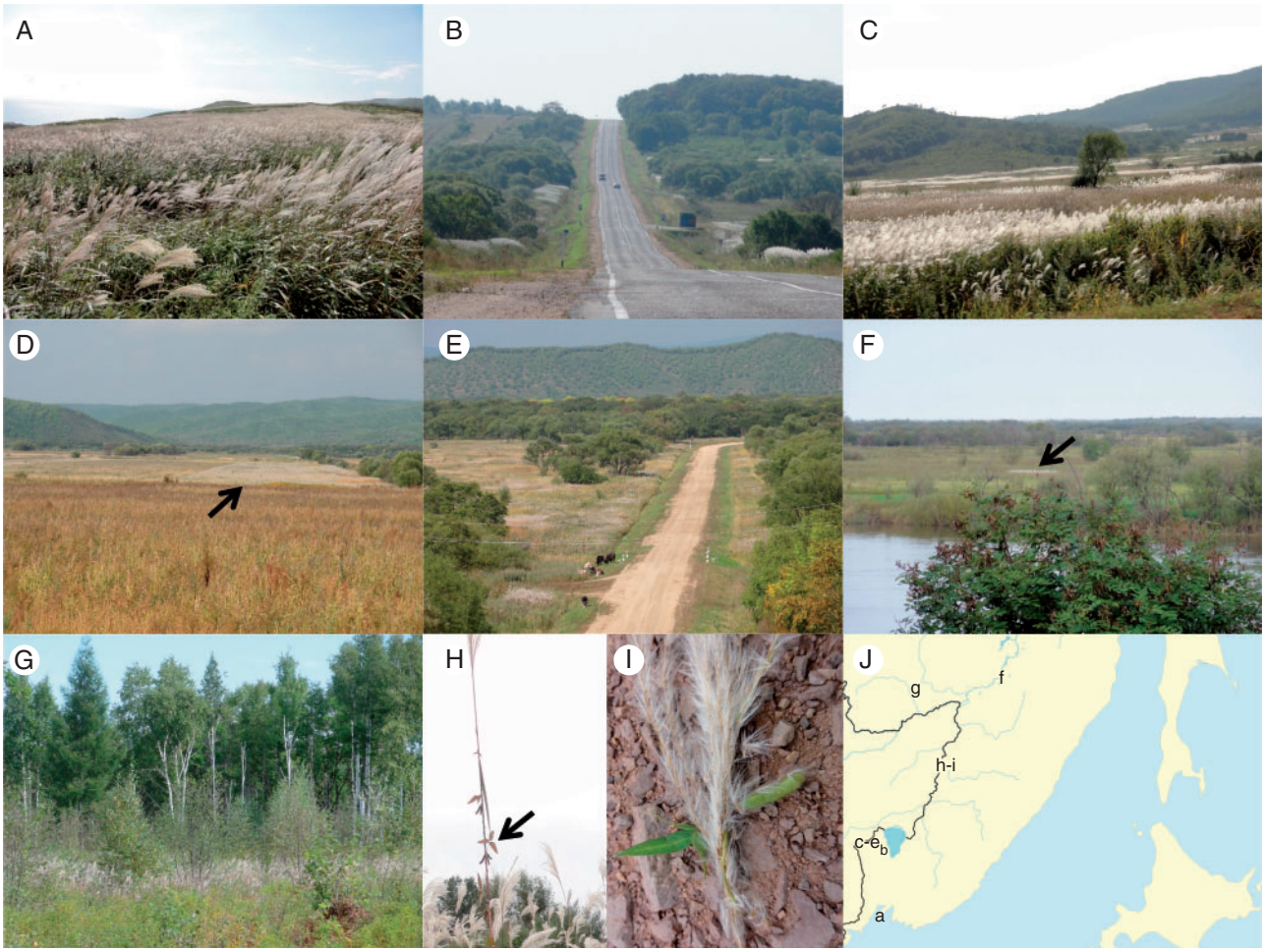

FIG. 2. Photos of *Miscanthus* in eastern Russia. (A) Large population of *M. sinensis* on Russky Island, south of Vladivostok facing the Sea of Japan. (B and C) Stands of *M. sacchariflorus* (white patches) growing in high frequency along roads near Lake Khanka, on land that is sloped and well drained. (D) Two rectangular hayfields near Lake Khanka with a large population of *M. sacchariflorus* growing in the field on the right (indicated with an arrow) and the field on the left was recently cut. (E) Near Lake Khanka, *M. sacchariflorus* (white patches) growing in lowlands that have recently been deforested by cutting and burning, with cows grazing in the foreground, and recently logged mountainside in the background. (F) *Miscanthus sacchariflorus* (indicated with an arrow) growing on an island in the Amur River near the village of Sinda, Khabarovsk Krai. (G) *Miscanthus sacchariflorus* growing in an open area next to a birch (*Betula* sp.) forest. (H) Wild soybean, *Glycine max* (synonym = *Soja ussuriensis;* indicated with an arrow) twining up a stem of *M. sacchariflorus*. (I) Close-up of wild soybean leaf and seed pod growing on an inflorescence of *M. sacchariflorus*. (J) Map of locations where photos were taken.

Data Fig. S3) and morphology. Other plants commonly associated with *M. sacchariflorus* at the collection sites included: *Allium* sp., *Artemisia* sp., *Aster* sp., *Betula* sp., *Iris* sp., *Juncus* sp., *Lespedeza* sp., *Phragmites* sp., *Quercus* sp., *Salix* sp. and the wild soybean *Glycine max* (synonyms *Soja ussuriensis*, *Glycine soja*) (Fig. 2G–I). Knowledge of which species are frequently associated with *M. sacchariflorus* is useful for understanding their in-common environmental adaptation and the complex ecological communities to which they belong. Herbarium specimens that were collected during the expedition are described in Supplementary Data Table S1 and maintained at the Vavilov Institute.

Stem height of the *M. sacchariflorus* accessions at the collection sites ranged from 1·2 to 2·6 m (mean 2·0; Fig 3D, G). Most accessions had thin stems (mean diameter of 3·8 mm), but one plant (RU2012-182) had stems as thick as 8 mm (Fig. 3E, H; Dataset S1). We observed a strong correlation between latitude and phenotype of *M. sacchariflorus* at the collection sites. Relative to the mean, stem height decreased modestly with increasing latitude (0·24 m in 5·6°of latitude change), and stem diameter decreased substantially (1·4 mm in 5·6°of latitude change; Fig. 3G, H; Supplementary Data Fig. S4). Stem height and diameter were also significantly correlated with elevation, but more weakly than they were correlated with latitude (Fig. S4). The effect of elevation on stem height and diameter was non-significant when latitude was included as an effect in the model (data not shown), suggesting that the strong correlation between latitude and elevation (Fig. S4) drove the correlation between elevation and stem height and diameter. The number of stems per unit area, a key component of biomass yield for *Miscanthus* (Matumura *et al*., 1985, 1986, 1987), varied greatly among accessions (from 5 to 112 per 0·25 m^2; Fig. 3F) but no correlation with latitude or elevation was observed for this trait (Fig. 3I). One individual in particular, RU2012-169, had above

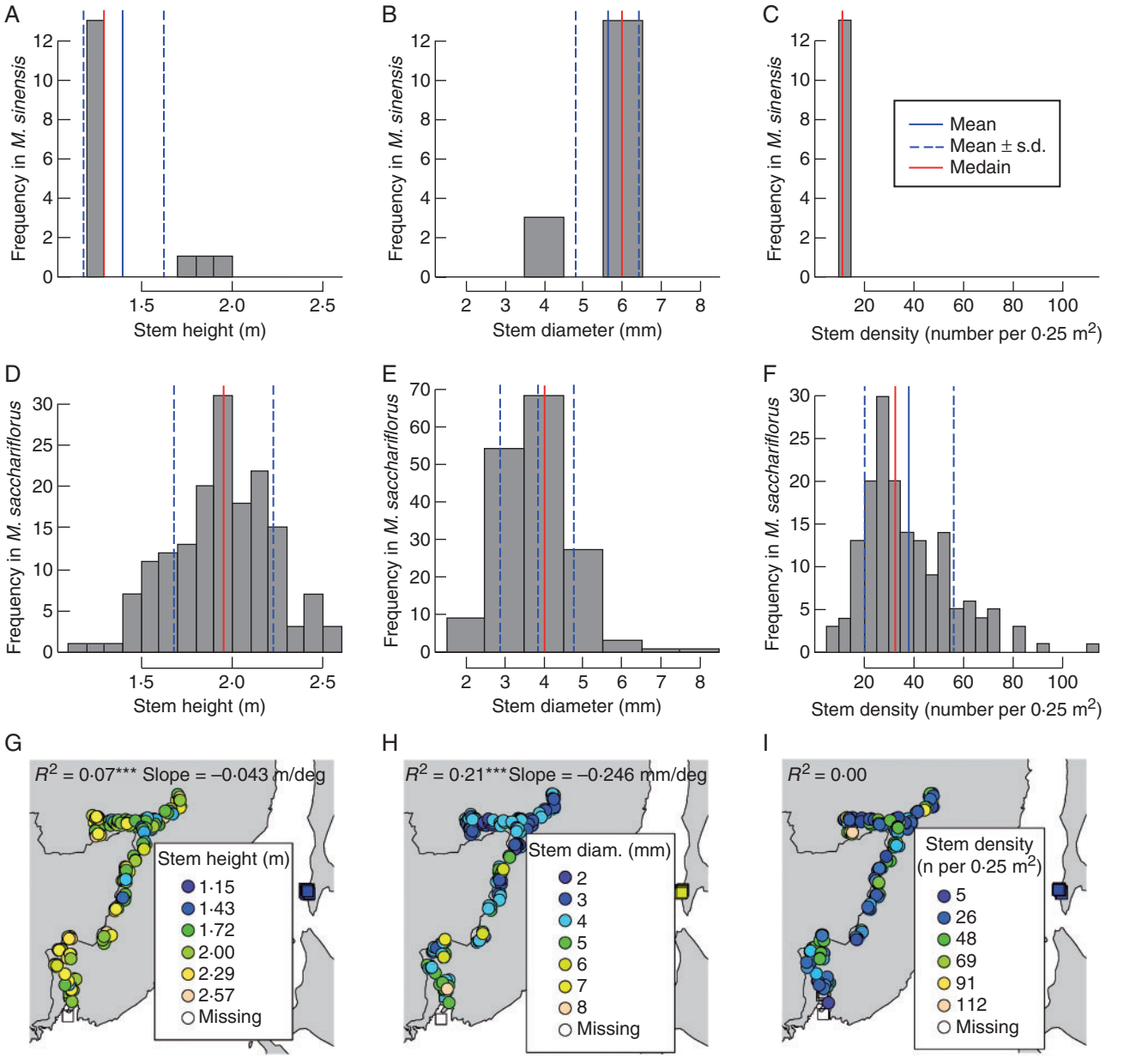

FIG. 3. Phenotypes of *Miscanthus*, recorded at collection sites. Seventeen *M. sinensis* individuals across three sites and 165 *M. sacchariflorus* individuals across 44 sites were measured. Stem diameter was measured at the base of the stem. (A–F) Histograms of phenotype distributions for *M. sinensis* (A–C) and *M. sacchariflorus* (D–F). (G, H) Maps of phenotype distributions. R^2 and slope values were calculated by fitting linear models relating phenotype to latitude in *M. sacchariflorus* only. Circles indicate *M. sacchariflorus*, and squares indicate *M. sinensis*

average values for height (2·42 m), stem diameter (6 mm) and stems per area (51 in 0·25 m^2) and is therefore an especially good candidate for breeding biomass cultivars (Fig. S4).

Population structure, ploidy and diversity of M. sacchariflorus

Statistical analyses of the RAD-seq SNP data for 160 *M. sacchariflorus* individuals revealed weak population structure. The Bayesian Information Criterion produced by the 'find.clusters' function in the R package adegenet was at its minimum when one cluster was assumed, suggesting that it would not be meaningful to split the set into distinct groups of individuals using DAPC (Supplementary Data Fig. S5). A Neighbor–Joining tree calculated from Euclidian distances between genotypes also did not suggest clear groupings, apart from groups of closely related individuals collected at the same sites (Supplementary Data Fig. S6). Using Structure, the Evanno *et al*. (2005) method suggested that five clusters (K = 5) would be ideal, primarily because likelihood values were highly variable from run to run at K = 6 (Supplementary Data Fig. S7). However, we found that ancestry assignments (Q values) were not reproducible from run to run at K = 5, so we instead present the results at K = 4 (Fig. 1A) and K = 2 (Supplementary Data

Fig. S8). Sets of individuals from two collection sites were placed into two respective clusters by Structure at K = 4, and Q values at K = 4 and K = 2 otherwise followed a gradient suggesting isolation by distance without clear geographic breaks (Fig. 1A; Fig. S8). Since adegenet, Structure and Neighbor–Joining indicated that population structure was absent or weak, a Mantel test was performed to determine whether a gradient of population structure was significantly associated with geography. The Mantel test confirmed that geographic distance was correlated with nuclear genetic distance at $P < 0{\cdot}001$ (Fig. 4A).

Out of 163 *M. sacchariflorus* that were screened by flow cytometry, 160 were diploid and three were tetraploid. The tetraploids, which were collected at two sites in the Amur River basin and included in the SNP analyses, were genetically similar to the diploid accessions found further south in our collection range (Fig. 1A; Fig. S8).

RAD-seq SNPs indicated high genetic diversity for the Russian *M. sacchariflorus* (Table 1). We compared SNP diversity statistics obtained from our 160 Russian *M. sacchariflorus* individuals with those obtained from previous studies of *Miscanthus* using an identical RAD-seq protocol. In a large collection of *M. sinensis* that was representative of geographic diversity across most of East Asia (Clark *et al.*, 2014), the south-east China population had the greatest number of SNPs with a minor allele frequency >0·05 (9262 SNPs), consistent with our finding that south-east China was the centre of radiation for *M. sinensis* after the last glacial maximum. The two most isolated populations of *M. sinensis*, in North Japan and the Sichuan Basin, had the lowest number of SNPs with a minor allele frequency >0·05 (6942 and 7531, respectively). In comparison, Russian *M. sacchariflorus* were highly diverse, with 12 265 SNPs at a minor allele frequency >0·05. However, the Japanese *M. sacchariflorus–M. ×giganteus* complex had 17 909 SNPs with a minor allele frequency >0·05, probably due to its interspecific nature (Clark *et al.*, 2015). Mean expected heterozygosity across all SNPs followed a similar pattern (Table 1).

Complete plastid haplotypes across ten microsatellite loci were obtained for 159 *M. sacchariflorus* individuals, yielding 13 unique haplotypes (Supplementary Data Table S2). Twelve of these could be connected to each other by single mutations in a haplotype network, whereas the remaining haplotype (S) only shared alleles at three out of ten loci with the most closely related haplotype (Fig. 5). Amplicon size of haplotype S was more similar to typical *M. sacchariflorus* haplotypes than it was to *M. sinensis* haplotypes. None of the Russian *M. sacchariflorus* individuals had plastid haplotypes from *M. sinensis*. Six out of the seven most common haplotypes (Z, Y, W, V, U and T) had also been found in *M. sacchariflorus* from China and/or Japan in our previous studies (Clark *et al.*, 2014, 2015). Haplotype S was also observed in an individual of *M. sacchariflorus* 'Robustus' (98m0002) obtained from M. Deuter of Tinplant (Germany) and in a diploid accession of *M. sacchariflorus* obtained from a US nursery (UI10-00008). Two of the tetraploid individuals had haplotype U, and the third tetraploid had a unique haplotype in the dataset that differed from haplotype U by one mutation. Overall, the presence of multiple common haplotypes in the Russian *M. sacchariflorus* accessions resulted in a high Gini–Simpson index (probability of getting two different haplotypes if two individuals were chosen at random) of 0·82, which was also high in comparison with typical *M. sinensis* populations (Table 1). However, we observed considerable overlap in the geographic distributions of plastid haplotypes. A Mantel test comparing haplotype dissimilarity with geographic distance did not find significant correlation (Fig. 4B).

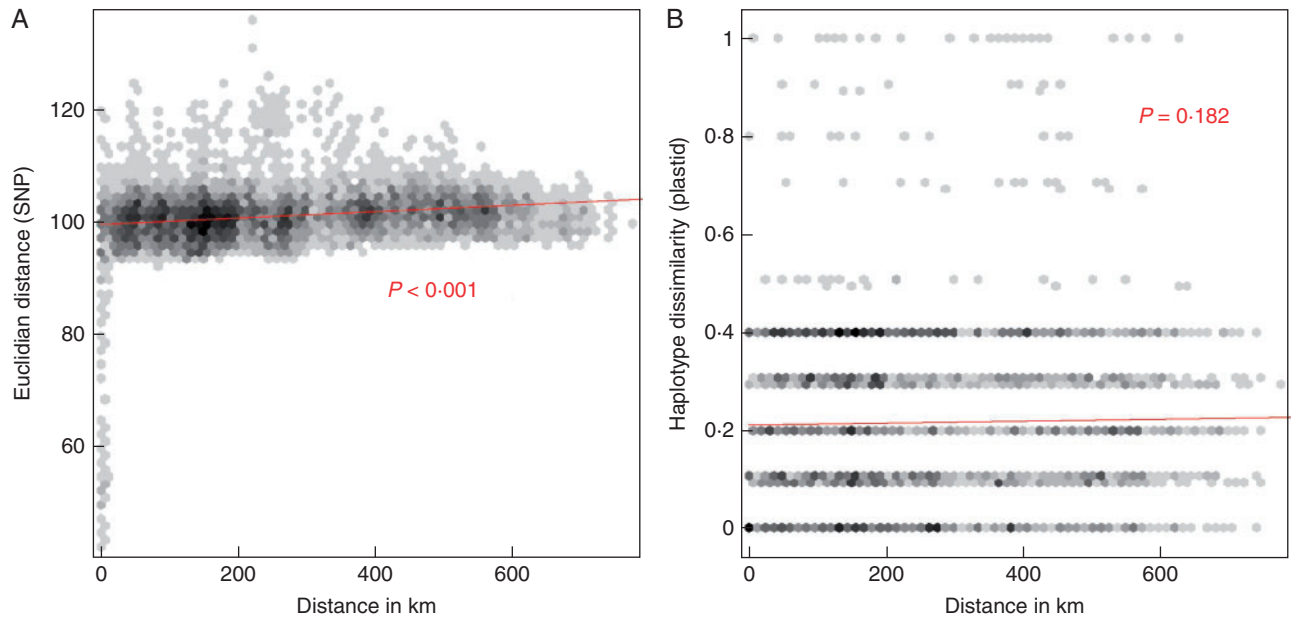

FIG. 4. Relationship between genetic distance and geographic distance for *M. sacchariflorus* individuals, shown as density plots. Significance of correlation (*P*) was determined by a Mantel test. Red lines were fit with linear regression. (A) Euclidian genetic distance based on 29 260 RAD-seq SNPs for 160 individuals vs. geographic distance in kilometres. (B) Proportion of plastid microsatellite alleles shared for 159 individuals vs. geographic distance.

TABLE 1. *Genetic diversity of Russian* Miscanthus sacchariflorus *compared with previously studied populations (Clark et al., 2014, 2015) of* Miscanthus

Population	No. of individuals	No. of plastid haplotypes	Gini–Simpson index of plastid haplotypes	No. of nuclear SNPs with minor allele frequency >0·05*	Mean expected heterozygosity across all nuclear SNPs
Russia Msa[†]	160	13	0·82 ± 0·01	12 265	0·14
Japan Msa–Mxg[‡]	78	19	0·83 ± 0·04	17 909	0·19
SE China Msi[§]	125	22	0·75 ± 0·04	9262	0·13
Yangtze-Qinling Msi	114	18	0·57 ± 0·06	7902	0·13
Sichuan Msi	58	9	0·74 ± 0·04	7531	0·12
Korea, N China Msi	195	12	0·48 ± 0·04	8596	0·13
N Japan Msi	96	6	0·25 ± 0·06	6942	0·11
S Japan Msi	32	11	0·87 ± 0·04	8077	0·12

*SNPs appearing heterozygous in any one of three doubled haploid *M. sinensis* lines were removed from all SNP datasets for the calculation of diversity metrics.
[†]Msa, *M. sacchariflorus*.
[‡]Msa-Mxg, *M. sacchariflorus–M. ×giganteus* complex.
[§]Msi, *M. sinensis*.

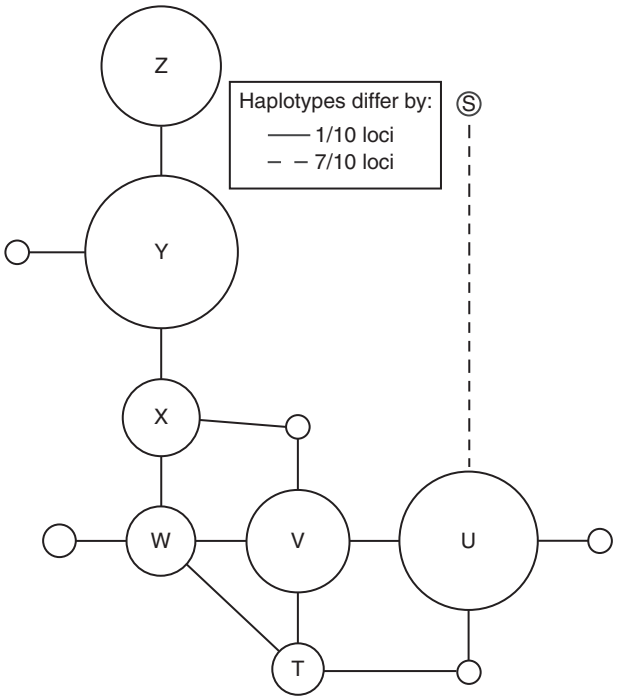

FIG. 5. Plastid haplotype network of *M. sacchariflorus* based on ten microsatellite markers. A total of 159 individuals from eastern Russia were evaluated. Haplotypes found in more than two individuals, as well as one unusual haplotype that was dissimilar from all others, are indicated by letters. Amplicon sizes for all haplotypes are listed in Table S2.

SNP–trait associations

Using the 5971 SNPs that met our criteria for GWAS, and phenotypic data recorded at the collection sites, we identified five significant marker–trait associations. We identified three significant associations with number of stems per unit area, including two SNPs aligning to *S. bicolor* chromosome 2 and one SNP aligning to *S. bicolor* chromosome 6 (Fig. 6A). The two SNPs on chromosome 2 were both in the untranslated regions of transcribed genes, while the SNP on chromosome 6 coded for an amino acid substitution (Table 2). Intriguingly, one of the tagged genes on chromosome 2 coded for a protein kinase, and was also <1 cM from the peak of a QTL region on *M. sacchariflorus* linkage group 4 for the ratio of compressed circumference to basal circumference, a trait heavily dependent on stems per unit area (Hongxu Dong, unpubl. res.). Moreover, the SNP on chromosome 6 fell within a QTL for vegetative branching in sorghum that contains several homologues of rice genes also known to be involved in branching (Kong *et al*., 2014), and within a QTL for number of tillers per plant in sorghum (Shiringani *et al*., 2010; Table 2). For all three of the significant stem density SNPs in *M. sacchariflorus*, the minor allele was associated with fewer stems per unit area (Fig. 6B). Additionally, we identified one SNP aligning to sorghum chromosome 6 that was significantly associated with stem diameter when we corrected for the environmental influence of latitude. The stem diameter SNP was near two previously identified QTLs for stem diameter in sorghum (Shiringani *et al*., 2010; Phuong *et al*., 2013). We also identified one SNP on chromosome 3 that was significantly associated with stem height (corrected for latitude), and was near a previously identified sorghum QTL for plant height (Phuong *et al*., 2013). Heritability as estimated from SNP data was 0·33, 0·81 and 0·71 for stems per area, stem height and stem diameter, respectively. The latter two are similar to heritabilities of 0·88 and 0·60 found for stem height and stem diameter, respectively, in *M. sinensis* by Slavov *et al*. (2014). Individual significantly associated SNPs accounted for between 5 and 18 % of the variation in their respective traits under an additive model (Fig. 6D). The three SNPs associated with stems per area accounted for 27 % of the variation in this trait when taken together.

DISCUSSION

Ecology and population genetics of M. sacchariflorus *and* M. sinensis

Our results further clarify the native range of *M. sinensis* and *M. sacchariflorus* in Russia, which is useful for understanding the extent of environmental adaptation in these species. We found *M. sinensis* in eastern Russia restricted to maritime

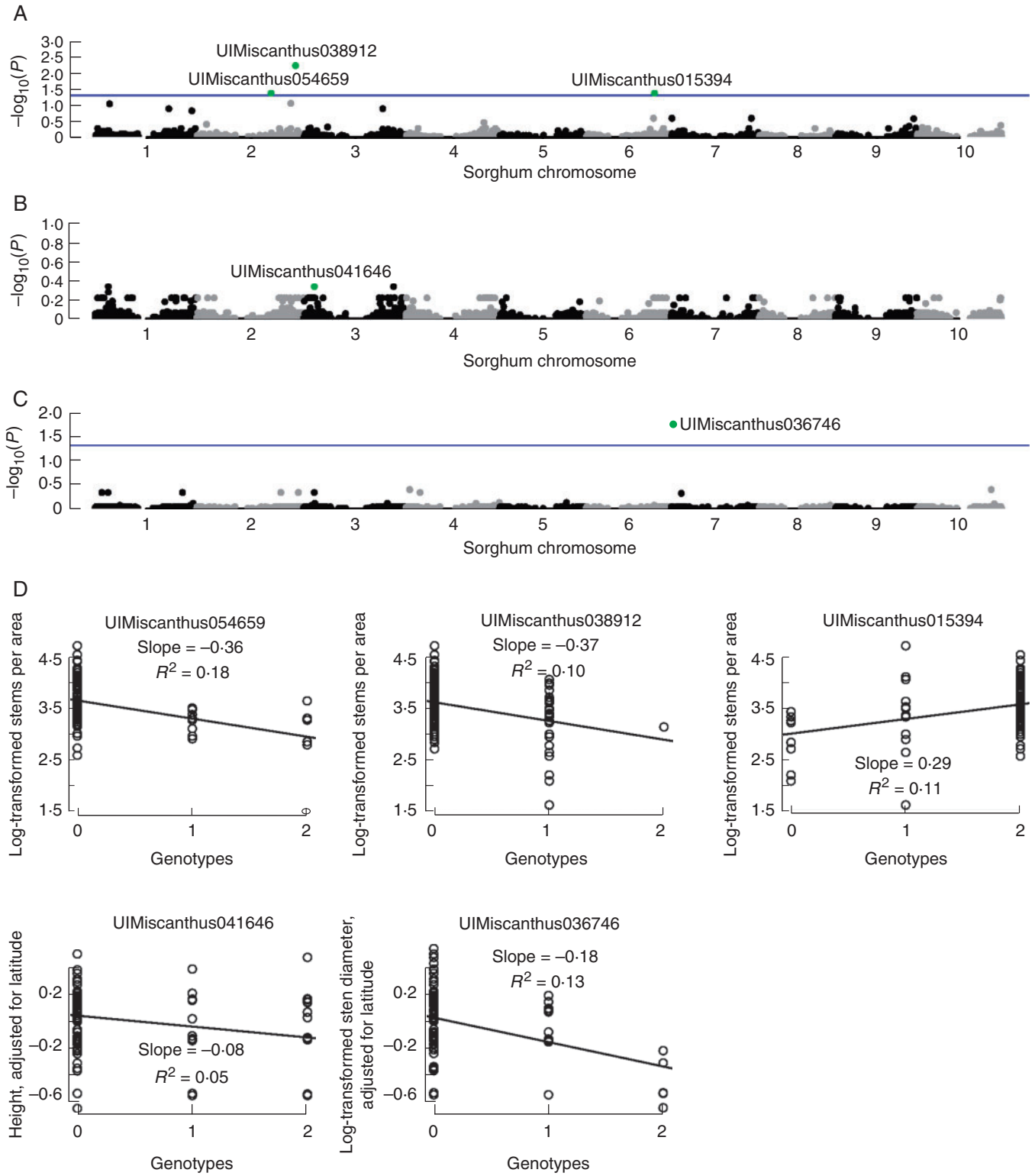

FIG. 6. Results of Q–K mixed model analysis to detect SNP–trait associations in Russian *Miscanthus sacchariflorus*. (A–C) Manhattan plots of log-transformed *P*-values vs. aligned position to the *Sorghum bicolor* 2.0 reference genome. *P*-values were calculated in rrBLUP and corrected using the method of Benjamini and Hochberg (1995). The blue line indicates $P = 0{\cdot}05$ after correction. Significant SNPs are highlighted in green. (A) Stems per area. The phenotype was measured as the number of stems in 0·25 m^2 at the collection site, and was log transformed. (B) Stem height. The phenotype was adjusted for latitude. (C) Stem diameter at base. The phenotype was log transformed, then adjusted for latitude. (D) Linear models of trait vs. genotype for the five significant SNPs. A genotype of zero or two indicates a homozygote, whereas a genotype of one indicates a heterozygote.

TABLE 2. *SNPs significantly associated with number of stems per unit area and stem diameter in Russian* Miscanthus sacchariflorus

Trait	Marker	*Sorghum bicolor* chromosome	*S. bicolor* position*	Nearest gene	Position relative to gene	Arabidopsis gene	Gene ontology	Sorghum QTL
Stems per area	UIMiscanthus054659[†]	2	53 659 234	Sobic.002G169700	In 5′ UTR		Protein kinase	
	UIMiscanthus038912[†,‡]	2	71 469 182	Sobic.002G350900	In 3′ UTR		None	
	UIMiscanthus015394[†]	6	50 139 247	Sobic.006G127200	In protein-coding region, Arg→Gln	AT1G51540	Galactose oxidase/kelch repeat	qM1_6·1[§], tillers per plant**
Height	UIMiscanthus041646[‡]	3	7 815 391	Sobic.003G089400	Upstream	AT5G37870	Ubiquitin protein ligase	PHE[¶]
Diameter	UIMiscanthus036746[†,‡]	6	57 761 452	Sobic.006G221700	In 5′ UTR		Protein binding	SDI[¶], stem diameter**

*Positions are in base pairs, from alignments to the *S. bicolor* genome version 2.0. Positions are given for the SNPs themselves, as opposed to the start of the RAD tags.
[†]Significant association using rrBLUP.
[‡]Significant association using TASSEL5.
[§]QTL from Kong et al. (2014).
[¶]QTL from Phuong et al. (2013).
**QTL from Shiringani et al. (2010).

regions with USDA hardiness zones 5 and 6, whereas *M. sacchariflorus* was also found further inland in hardiness zones 3–6 (Fig. 1B). Although *M. sinensis* has been reported further inland near Lake Khanka (Herbarium at the Komarov Botanical Institute; Harkevich, 1985), we were unable to find it in that region despite deliberate searching. Hodkinson *et al.* (2016) also did not find *M. sinensis* in the Lake Khanka region. The previous report of *M. sinensis* near Khabarovsk (Harkevich, 1985) was probably an error, given that this was very far outside of the range in which we observed *M. sinensis*. However, as the global climate continues to warm, we predict that the large population of *M. sinensis* on Russky Island, as well as coastal populations south-west of Vladivostok identified by Hodkinson *et al.* (2016), will migrate northward into the mainland of eastern Russia and interact more with the currently well-established populations of *M. sacchariflorus* there. In China and Japan, where *M. sacchariflorus* and *M. sinensis* are sympatric, we have observed that *M. sacchariflorus* is typically restricted to riparian and wetland environments, whereas *M. sinensis* occupies upland sites. However, in eastern Russia, we observed *M. sacchariflorus* in both riparian and upland environments (especially near Lake Khanka; Fig. 2B, C), suggesting that the presence or absence of interspecies competition affects the distribution of *M. sacchariflorus*.

One hypothesis for why *M. sacchariflorus* is adapted to environments with colder winters (e.g. inland eastern Russia) than *M. sinensis* is that the rhizomatous growth habit of *M. sacchariflorus* allows for better winter survival via cold avoidance of buds deep in the soil compared with the caespitose habit of *M. sinensis*. Rhizomatous growth habit has been associated with cold adaptation among C4 grasses in Canada due to rhizome buds being deeper underground than crown buds (Schwarz and Reaney, 1989). However, during our collecting, we observed that most of the *M. sacchariflorus* rhizomes were in the top 15 cm of soil, suggesting that *M. sacchariflorus* from eastern Russia may have physiological tolerance to cold that allows its rhizomes to overwinter in frozen soil. Future experiments in controlled environments will be needed to test this hypothesis.

We encountered an apparent north-eastern edge to the range of *M. sacchariflorus* at 49·3°N, 136·5°E. The absence of *M. sacchariflorus* north-east of 49·3°N, 136·5°E corresponds to the transition from hardiness zone 3 to 2 (Fig. 1B); thus, extreme winter temperatures may limit the range of *M. sacchariflorus*, although other factors such as lack of disturbed open habitat cannot be ruled out. West of Khabarovsk along the northern side of the Amur River watershed, we travelled as far inland as 48·7°N, 133·0°E without finding a western edge to the range of *M. sacchariflorus*. At the most extreme, we found *M. sacchariflorus* in areas with a mean January temperature of –24 °C (Fig. 1C). Based on our observations and the hardiness zone map (Fig. 1B), it would be worthwhile for future expeditions to explore if the western and northern boundaries of *M. sacchariflorus* are the Zeya and Selemdzha Rivers of eastern Russia, respectively.

In contrast to previous studies that found substantial population structure in *M. sinensis* (Chou *et al.*, 2000; Iwata *et al.*, 2005; Slavov *et al.*, 2013; Zhang *et al.*, 2013; Zhao *et al.*, 2013; Clark *et al.*, 2014, 2015; Nie *et al.*, 2014), relatively weak population structure was observed in the present study for Russian *M. sacchariflorus*. Although nuclear SNPs gave a significant signal of isolation by distance (Fig. 4A), neither DAPC (Fig. S5), Structure (Fig. 1A) nor Neighbor–Joining (Fig. S6) revealed geographically distinct genetic clusters (except for groups of closely related, possibly clonal, individuals found at the same collection sites; Fig. S6), and there was no significant isolation by distance found among chloroplast haplotypes (Fig. 4B). In Japan, all or nearly all *M. sacchariflorus* are tetraploid (Hirayoshi *et al.*, 1957) and many have introgressions from diploid *M. sinensis* (Clark *et al.*, 2015), whereas in Russia diploidy predominates and there was no evidence of introgression with *M. sinensis* in our sampling range, although Hodkinson *et al.* (2016) found putative hybrids further south in Russia. In north-east China, Jiang *et al.* (2013) similarly found that diploid populations of *M. sacchariflorus* did not form an introgressed hybrid swarm with *M. sinensis*, though F_1 hybrids were observed there.

It is possible that with a broader geographic sample of *M. sacchariflorus*, genetic patterns will emerge that explain the species' history similarly to how the migration history of *M. sinensis* has been elucidated (Clark *et al.*, 2014). On the other hand, the lack of population structure in *M. sacchariflorus* may reflect fundamental biological differences between *M. sinensis* and *M. sacchariflorus*. For example, *M. sacchariflorus* may have higher gene flow than *M. sinensis*, perhaps due to rhizome pieces being dispersed along waterways or by plowing (Deng *et al.*, 2013). Moreover, being more cold-tolerant, *M. sacchariflorus* was probably not restricted in range to the same extent as *M. sinensis* during the last glacial maximum, and so the magnitude of founder effects from recolonization of Asia may not be as pronounced. Lack of recent founder effects is also consistent with the greater genetic diversity that we observed in *M. sacchariflorus* relative to *M. sinensis* (Table 1). The high levels of genetic diversity in the accessions of *M. sacchariflorus* from eastern Russia suggest that this germplasm collection will be especially useful for breeding.

To facilitate the long-term preservation of genes from the Russian *M. sacchariflorus* genotypes that we collected as clonal divisions, the best option will be to develop seed-based germplasm pools by inter-mating the clones, because seed can be kept viable and safe in cold storage for many decades, whereas loss of vegetative stock plants is a substantial risk for clonal collections. The optimal number of germplasm pools for preserving the Russian *M. sacchariflorus* genes should balance the desire to preserve genetic differences associated with geographic distance with the expense of maintaining and increasing seed stocks of multiple accessions for an obligate outcrossing species. At a minimum, there would need to be two seed-based germplasm pools for the Russian *M. sacchariflorus* accessions: one for the tetraploids and one for the diploids. More conservatively, we would recommend splitting the diploid *M. sacchariflorus* accessions into three germplasm pools: (1) for accessions from Vladivostok in the south to the south-west shore of Lake Khanka in the north; (2) along the Ussuri River watershed from north-east of Lake Khanka in the south to near Ussuri's confluence with the Amur River at Khabarovsk in the north; and (3) along the Amur River watershed from near Birobidzhan and the Bira River in the west to near the village of Naykhin and the confluence of the Amur and Anyuy Rivers in the east (Supplementary Data Fig. S9). However, genetic differentiation between these Russian *M. sacchariflorus* groups (Supplementary Data Table S3) is an order of magnitude lower than that between the six major genetic groups of *M. sinensis* identified by Clark *et al.* (2014).

Genome-wide association analysis using field observations of M. sacchariflorus

Using an inexpensive genotyping method and phenotypic data obtained at the time of germplasm collection for the Russian *M. sacchariflorus*, we were able to identify three SNPs that were significantly associated with the number of stems per unit area, one with height and one with stem diameter. Four of these SNPs were in the transcribed regions of genes; one coded for an amino acid substitution, and three were near (within 10 Mb) or within sorghum QTLs for related traits (Table 2), suggesting that the associations were not spurious. The number of stems per unit area is not only important for the ecology of grasses (Harper, 1985), but is also a useful predictor of biomass yield in *Miscanthus* (Gauder *et al.*, 2012). In a previous agronomic study, yield of *M. sacchariflorus* was negatively impacted by low number of stems per unit area, but the effect was not observed until the third year (Clifton-Brown and Lewandowski, 2002), highlighting the utility of measuring this trait during germplasm collection.

Our results indicate that value can be added to a germplasm collection if phenotyping is done at the collection site and if low-cost, high-density genotyping is performed via high-throughput sequencing. Plant breeders are unlikely to use accessions in germplasm collections until phenotypic and/or genotypic data are available to predict the value of accessions for traits of interest. If phenotypic data are not obtained by the germplasm collectors at the collection sites, then typically none will be available until the germplasm repository receives grant funding for replicated testing (McCouch *et al.*, 2012). Although phenotypic data from collection sites are unreplicated and confounded by site to site environmental variation, there is value in collecting such data for relatively stable traits because potential users of the germplasm will find the data useful and better than no data at all. Moreover, our study and others (Parchman *et al.*, 2012; Budde *et al.*, 2014) have shown that although lack of replication for *in situ* phenotypic data would be expected to reduce the statistical power of association analyses as compared with analyses performed using replicated field trial data, it is still possible to identify useful significant SNP–trait associations with standard GWAS methodology (Q–K mixed model analysis and multiple testing correction using FDR). The high genetic diversity and absence of strong population structure within diploid Russian *M. sacchariflorus* also enhanced our power to detect SNP–trait associations. Overall, these results suggest that, for at least some important traits, plant phenotypic data measured in the wild can be sufficiently robust to be useful to plant breeders, particularly when adjusted for known environmental covariates, e.g. latitude in our study.

The use of *in situ* phenotypic data for GWAS will help to accelerate plant breeding, given that many years typically pass between when germplasm is collected and when replicated field trials are conducted and high quality phenotypic data are available. GWAS results are especially useful for the efficient utilization of germplasm collections, given that they can be used to make predictions about the performance of individual accessions, thereby aiding in the selection of accessions for further evaluation or breeding (McCouch *et al.*, 2012). Moreover, at approx. US$20 per sample for GBS or RAD-seq, it may be less expensive to genotype an accession than to collect and import it. SNP data that are associated with DNA sequence tags can be cross-referenced to any sequence data that are generated in the future, making them much more valuable long term than older marker technologies such as RFLPs (restriction fragment length polymorphisms) and microsatellites (Kilian and Graner, 2012). Using an inexpensive sequencing-based genotyping method (RAD-seq), we were able to identify SNPs associated with traits of agronomic interest several years in advance of when we would be able to obtain data from replicated field trials. In fact, it has been possible for us to initiate replicated trials of these accessions only as early as spring of 2015, and replicated data

from mature plants will not be available until the end of 2017, >5 years after the accessions were collected. We expect that many breeders will be willing to accept the risk of performing marker-assisted selection using QTLs from preliminary studies such as ours, given the potential reward of increasing yields as quickly as possible. We hope that the present study will encourage future germplasm collectors to take the time to record phenotypic data *in situ* and conduct genomic analyses that can rapidly provide end-users with useful guidance on how best to use the collected materials. By helping prospective end-users of the germplasm avoid a data desert, it should be possible to circumvent this common barrier to the use of germplasm collections, especially for newly collected accessions.

Applications for breeding Miscanthus *and* Saccharum

This collection of *Miscanthus* from eastern Russia, which is being curated by the US National Plant Germplasm System, along with the collection of Hodkinson *et al*. (2016), which is being kept at Trinity College and Teagasc, Ireland and at the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, probably represents the maximum winter hardiness for both *M. sinensis* and *M. sacchariflorus*. Additionally, the collection includes individuals with desirable biomass traits. Prior to the establishment of this collection, the National Plant Germplasm System did not have any *M. sacchariflorus* accessions, and the one *M. sacchariflorus* genotype available from commercial horticulture nurseries in North America is short (<1·5 m) and has thin stems, which are undesirable traits for biomass production. In contrast, we collected many accessions that are tall (up to 2·6 m) and have thick stems (up to 8 mm diameter). Moreover, the three tetraploid accessions of *M. sacchariflorus* that we collected will be especially valuable for breeding sterile triploid *M.* ×*giganteus* (tetraploid *M. sacchariflorus* × diploid *M. sinensis*) that is more winter hardy than the currently predominant cultivar used for biomass production in North America and Europe. In addition to being useful sources of genes for improving *Miscanthus*, the accessions of *M. sinensis* and *M. sacchariflorus* that we collected in eastern Russia can also be crossed to *Saccharum* to introgress cold hardiness into sugarcane, thereby extending its range of cultivation into higher latitudes and elevations than is currently possible for this tropical crop. Previous efforts to cross *Miscanthus* and *Saccharum* have been successful, but have focused on introgressing disease resistance, rather than cold hardiness, into *Saccharum* (Sacks *et al*., 2013). Furthermore, our GWAS results may facilitate marker-assisted selection of *M. sacchariflorus* accessions for high numbers of stems per area, plant height and large stem diameter, which should be useful for breeding cultivars with improved biomass yield and resistance to lodging. Due to the small number of individuals included in our GWAS, it is possible that we have overestimated QTL effects (Beavis, 1998; Ioannidis, 2008); however, they represent testable hypotheses, and we intend to perform crosses to validate these QTLs so that they can potentially be used in breeding as quickly as possible. Given that both winter hardiness and stems per area of *M. sacchariflorus* are proabably related to the plant's underground architecture, it will be important to determine if these two traits are correlated with each other and controlled by the same genes, because such knowledge would facilitate the breeding of cultivars that are both more winter hardy and higher yielding than current *Miscanthus* cultivars. With the recent interest in *Miscanthus* for bioenergy, we expect that this germplasm collection and its associated data will be highly requested resources.

SUPPLEMENTARY DATA

Supplementary data are available at http://www.aob.oxfordjournals.org and consist of the following. Table S1: list of herbarium specimens collected. Table S2: *Miscanthus sacchariflorus* plastid haplotypes found in eastern Russia. Table S3: Jost's *D* statistic showing differentiation between *Miscanthus sacchariflorus* seed increase groups from Russia. Figure S1: relationship matrix of 155 diploid Russian *M. sacchariflorus* accessions used in genome-wide association analysis (GWAS). Figure S2: principal component analysis of *M. sinensis* SNP data. Figure S3: ITS DNA sequence used for identifying *Hemarthria sibirica*. Figure S4: associations between phenotypes in *M. sacchariflorus*. Figure S5: Bayesian Information Criterion for selecting the number of *M. sacchariflorus* clusters in DAPC. Figure S6: Neighbor–Joining tree *of M. sacchariflorus*. Figure S7: choice of number of *M. sacchariflorus* clusters in Structure analysis. Figure S8: Structure results for Russian *M. sacchariflorus* when two clusters are assumed (K = 2). Figure S9: recommendations for seed increase groups in *M. sacchariflorus* collected in eastern Russia. Supplementary Materials and Methods: additional details on exploration route, flow cytometry and DNA protocols. Dataset S1: 2012 Collection of Miscanthus Germplasm in Eastern Russia. Dataset S2: DNA sequences, allele frequencies, alignment positions and GWAS results for *M. sacchariflorus* SNPs.

ACKNOWLEDGEMENTS

We gratefully acknowledge Inga Kotlyarskaya and Andrey Kseudz for their expert guiding and driving during the collection expedition, and Melina Salgado for assisting with DNA extraction and PCR. Karen Williams, Melanie Harrison, the editor Susanne Barth and two anonymous reviewers provided valuable feedback on earlier drafts of this manuscript. This work was supported by the USDA-ARS NPGS Plant Exploration Program [Cooperative Agreement #59-1275-1-338] and the DOE Office of Science, Office of Biological and Environmental Research (BER) [grant nos DE-SC0006634 and DE-SC0012379].

LITERATURE CITED

Barth S, Cormican P, Michel T, Velmurugan J, Klaas M. 2014. Genotyping by sequencing of a collection of Miscanthus spp. accessions. In: *Plant and Animal Genome XXII Conference*. San Diego, CA.

Beavis WD. 1998. QTL analyses: power, precision, and accuracy. In: Paterson AH, ed. *Molecular dissection of complex traits*. Boca Raton, FL: CRC Press, 145–162.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B* **57**: 289–300.

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**: 2633–2635.

Budde KB, Heuertz M, Hernández-Serrano A, *et al.* 2014. *In situ* genetic association for serotiny, a fire-related trait, in Mediterranean maritime pine (*Pinus pinaster*). *New Phytologist* **201**: 230–241.

de Cesare M, Hodkinson TR, Barth S. 2010. Chloroplast DNA markers (cpSSRs, SNPs) for *Miscanthus*, *Saccharum* and related grasses (Panicoideae, Poaceae). *Molecular Breeding* **26**: 539–544.

Chae WB, Hong SJ, Gifford JM, Lane Rayburn A, Sacks EJ, Juvik JA. 2014. Plant morphology, genome size, and SSR markers differentiate five distinct taxonomic groups among accessions in the genus *Miscanthus*. *GCB Bioenergy* **6**: 646–660.

Chessel D, Dufour AB, Thioulouse J. 2004. The ade4 package – I: one-table methods. *R News* **4**: 5–10.

Chou C-H, Chiang Y-C, Chiang T-Y. 2000. Genetic variability and phytogeography of *Miscanthus sinensis* var. *condensatus*, an apomictic grass, based on RAPD fingerprints. *Canadian Journal of Botany* **78**: 1262–1268.

Clark LV, Jasieniuk M. 2011. POLYSAT: an R package for polyploid microsatellite analysis. *Molecular Ecology Resources* **11**: 562–566.

Clark LV, Brummer JE, Głowacka K, *et al.* 2014. A footprint of past climate change on the diversity and population structure of *Miscanthus sinensis*. *Annals of Botany* **114**: 97–107.

Clark LV, Stewart JR, Nishiwaki A, *et al.* 2015. Genetic structure of *Miscanthus sinensis* and *Miscanthus sacchariflorus* in Japan indicates a gradient of bidirectional but asymmetric introgression. *Journal of Experimental Botany* **66**: 4213–4225.

Clifton-Brown JC, Lewandowski I. 2000. Overwintering problems of newly established *Miscanthus* plantations can be overcome by identifying genotypes with improved rhizome cold tolerance. *New Phytologist* **148**: 287–294.

Clifton-Brown JC, Lewandowski I. 2002. Screening *Miscanthus* genotypes in field trials to optimise biomass yield and quality in Southern Germany. *European Journal of Agronomy* **16**: 97–110.

Clifton-Brown J, Chiang Y-C, Hodkinson TR. 2008. *Miscanthus*: genetic resources and breeding potential to enhance bioenergy production. In: Vermerris W, ed. *Genetic improvement of bioenergy crops*. Berlin: Springer, 273–294.

Deng Z, Chen X, Xie Y, Li X, Pan Y, Li F. 2013. Effects of size and vertical distribution of buds on sprouting and plant growth of the clonal emergent macrophyte *Miscanthus sacchariflorus* (Poaceae). *Aquatic Botany* **104**: 121–126.

Earl DA, VonHoldt BM. 2011. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4**: 359–361.

Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome Journal* **4**: 250–255.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**: 2611–2620.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Farrell AD, Clifton-Brown JC, Lewandowski I, Jones MB. 2006. Genotypic variation in cold tolerance influences the yield of *Miscanthus*. *Annals of Applied Biology* **149**: 337–345.

Friesen PC, Peixoto MM, Busch FA, Johnson DC, Sage RF. 2014. Chilling and frost tolerance in *Miscanthus* and *Saccharum* genotypes bred for cool temperate climates. *Journal of Experimental Botany* **65**: 3749–3758.

Gauder M, Graeff-Hönninger S, Lewandowski I, Claupein W. 2012. Long-term yield and performance of 15 different *Miscanthus* genotypes in southwest Germany. *Annals of Applied Biology* **160**: 126–136.

Głowacka K, Adhikari S, Peng J, *et al.* 2014. Variation in chilling tolerance for photosynthesis and leaf extension growth among genotypes related to the C4 grass *Miscanthus*×*giganteus*. *Journal of Experimental Botany* **65**: 5267–5278.

Głowacka K, Clark LV, Adhikari S, *et al.* 2015. Genetic variation in *Miscanthus* × *giganteus* and the importance of estimating genetic distance thresholds for differentiating clones. *Global Change Biology Bioenergy* **7**: 386–404.

Greef J, Deuter M, Jung C, Schondelmaier J. 1997. Genetic diversity of European *Miscanthus* species revealed by AFLP fingerprinting. *Genetic Resources and Crop Evolution* **44**: 185–195.

Harkevich SS. 1985. *The vascular plants of Soviet Far East*. Nauka Publisher.

Harper JL. 1985. Modules, branches, and the capture of resources. In: Jackson JBC, Buss LW, Cook RE, eds. *Population biology and evolution of clonal organisms*. New Haven, CT: Yale University Press, 1–34.

Hijmans RJ. 2014. *geosphere: spherical trigonometry*. R package version 1.3-11.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978.

Hirayoshi I, Nishikawa K, Kubono M, Murase T. 1957. Cyto-genetical studies on forage plants (VI): On the chromosome number of Ogi (*Miscanthus sacchariflorus*). *Research Bulletin of the Faculty of Agriculture, Gifu University* **8**: 8–13.

Hodkinson TR, Renvoize S. 2001. Nomenclature of *Miscanthus*×*giganteus* (Poaceae). *Kew Bulletin* **56**: 759–760.

Hodkinson T, Chase M, Lledó D, Salamin N, Renvoize SA. 2002. Phylogenetics of *Miscanthus*, *Saccharum* and related genera (Saccharinae, Andropogoneae, Poaceae) based on DNA sequences from ITS nuclear ribosomal DNA and plastid *trnL* intron and *trnL-F* intergeneic spacers. *Journal of Plant Research* **115**: 381–392.

Hodkinson TR, Klaas M, Jones MB, Prickett R, Barth S. 2015. *Miscanthus*: a case study for the utilization of natural genetic variation. *Plant Genetic Resources* **13**: 219–237.

Hodkinson TR, Petrunenko E, Klaas M *et al.* 2016. New breeding collections of *Miscanthus sinensis*, *M. sacchariflorus* and hybrids from Primorsky Krai, Far Eastern Russia. In: *Proceedings of the Biomass 2015: Perennial biomass crops for a resource constrained world*. Springer.

Ioannidis JPA. 2008. Why most discovered true associations are inflated. *Epidemiology* **19**: 640–648.

Iwata H, Kamijo T, Tsumura Y. 2005. Genetic structure of *Miscanthus sinensis* ssp. *condensatus* (Poaceae) on Miyake Island: implications for revegetation of volcanically devastated sites. *Ecological Research* **20**: 233–238.

Jiang J-X, Wang Z-H, Tang B-R, Xiao L, Ai X, Yi Z-L. 2012. Development of novel chloroplast microsatellite markers for *Miscanthus* species (Poaceae). *American Journal of Botany* **99**: e230–e233.

Jiang J, Zhu M, Ai X, Xiao L, Deng G, Yi Z. 2013. Molecular evidence for a natural diploid hybrid between *Miscanthus sinensis* (Poaceae) and *M. sacchariflorus*. *Plant Systematics and Evolution* **299**: 1367–1377.

Johnston SE, McEwan JC, Pickering NK, *et al.* 2011. Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Molecular Ecology* **20**: 2555–2566.

Johnston SE, Orell P, Pritchard VL, *et al.* 2014. Genome-wide SNP analysis reveals a genetic basis for sea-age variation in a wild population of Atlantic salmon (*Salmo salar*). *Molecular Ecology* **23**: 3452–3468.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics* **27**: 3070–3071.

Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**: 94.

Jørgensen U. 1997. Genotypic variation in dry matter accumulation and content of N, K and Cl in *Miscanthus* in Denmark. *Biomass and Bioenergy* **12**: 155–169.

Jost L. 2008. G_{ST} and its relatives do not measure differentiation. *Molecular Ecology* **17**: 4015–4026.

Kilian B, Graner A. 2012. NGS technologies for analyzing germplasm diversity in genebanks. *Briefings in Functional Genomics* **11**: 38–50.

Kong W, Guo H, Goff VH, Lee T-H, Kim C, Paterson AH. 2014. Genetic analysis of vegetative branching in sorghum. *Theoretical and Applied Genetics* **127**: 2387–2403.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**: 357–359.

Lu F, Lipka AE, Glaubitz J, *et al.* 2013. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genetics* **9**: e1003215. doi:10.1371/journal.pgen.1003215.

Matumura M, Hasegawa T, Saijoh Y. 1985. Ecological aspects of *Miscanthus sinensis* var. *condensatus*, *M. sacchariflorus*, and their 3×-, 4×-hybrids (1) Process of vegetative spread. *Research Bulletin of the Faculty of Agriculture, Gifu University* **50**: 423–433.

Matumura M, Hasegawa T, Saijoh Y. 1986. Ecological aspects of *Miscanthus sinensis* var. *condensatus*, *M. sacchariflorus* and their 3×-, 4×-hybrids (2) Growth behaviour of the current year's rhizomes. *Research Bulletin of the Faculty of Agriculture, Gifu University* **51**: 347–362.

Matumura M, Hasegawa T, Saijoh Y. 1987. Ecological aspects of *Miscanthus sinensis* var. *condensatus*, *M. sacchariflorus* and their 3×-, 4×-hybrids. (3) Above ground standing crop and response cutting. *Research Bulletin of the Faculty of Agriculture, Gifu University* **52**: 315–324.

Maximowicz CJ. 1859. *Primitiae Florae Amurensis.* St. Petersburg: Buchdruckerei der Kaiserlichen Akademie der Wissenschaften.

McCouch SR, McNally KL, Wang W, Hamilton RS. 2012. Genomics of gene banks: a case study in rice. *American Journal of Botany* **99**: 407–423.

NAPPFAST. 2012. NAPPFAST Global Plant Hardiness Zones based on CFSR base data. Last accessed February 17, 2015.

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA. 2013. Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology* **22**: 2841–2847.

Nie G, Zhang X-Q, Huang L-K, *et al.* 2014. Genetic variability and population structure of the potential bioenergy crop *Miscanthus sinensis* (Poaceae) in southwest China based on SRAP markers. *Molecules* **19**: 12881–12897.

Paradis E. 2010. pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**: 419–420.

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA. 2012. Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**: 2991–3005.

Paterson AH, Bowers JE, Bruggmann R, *et al.* 2009. The *Sorghum bicolor* genome and the diversification of grasses. *Nature* **457**: 551–556.

Petersen KK, Hagberg P, Kristiansen K. 2003. Colchicine and oryzalin mediated chromosome doubling in different genotypes of *Miscanthus sinensis*. *Plant Cell, Tissue and Organ Culture* **73**: 137–146.

Phuong N, Stützel H, Uptmoor R. 2013. Quantitative trait loci associated to agronomic traits and yield components in a *Sorghum bicolor* L. Moench RIL population cultivated under pre-flowering drought and well-watered conditions. *Agricultural Sciences* **4**: 781–791.

Poland JA, Brown PJ, Sorrells ME, Jannink J-L. 2012*a*. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One* **7**: e32253. doi:10.1371/journal.pone.0032253.

Poland J, Endelman J, Dawson J, *et al.* 2012*b*. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome Journal* **5**: 103.

Purdy SJ, Maddison AL, Jones LE, *et al.* 2013. Characterization of chilling-shock responses in four genotypes of *Miscanthus* reveals the superior tolerance of *M.* ×*giganteus* compared with *M. sinensis* and *M. sacchariflorus*. *Annals of Botany* **111**: 999–1013.

R Core Team. 2014. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Sacks EJ, Juvik JA, Lin Q, Stewart JR, Yamada T. 2013. The gene pool of *Miscanthus* species and its improvement. In: Paterson AH, ed. *Genomics of the Saccharinae*. New York: Springer, 73–101.

Schwarz AG, Reaney MJT. 1989. Perennating structures and freezing tolerance of northern and southern populations of C4 grasses. *Botanical Gazette* **150**: 239–246.

Shiringani AL, Frisch M, Friedt W. 2010. Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. *Theoretical and Applied Genetics* **121**: 323–336.

Slavov G, Robson P, Jensen E, *et al.* 2013. Contrasting geographic patterns of genetic variation for molecular markers vs. phenotypic traits in the energy grass *Miscanthus sinensis*. *GCB Bioenergy* **5**: 562–571.

Slavov GT, Nipper R, Robson P, *et al.* 2014. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass *Miscanthus sinensis*. *New Phytologist* **201**: 1227–1239.

Winter D. 2012. MMOD: an R library for the calculation of population differentiation statistics. *Molecular Ecology Resources* **12**: 1158–1160.

Yook MJ, Lim S-H, Song J-S, *et al.* 2014. Assessment of genetic diversity of Korean *Miscanthus* using morphological traits and SSR markers. *Biomass and Bioenergy* **66**: 81–92.

Yu J, Pressoir G, Briggs WH, *et al.* 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* **38**: 203–208.

Zhang QX, Shen YK, Shao RX, *et al.* 2013. Genetic diversity of natural *Miscanthus sinensis* populations in China revealed by ISSR markers. *Biochemical Systematics and Ecology* **48**: 248–256.

Zhao H, Wang B, He J, *et al.* 2013. Genetic diversity and population structure of *Miscanthus sinensis* germplasm in China. *PLoS One* **8**: e75672. doi: 10.1371/journal.pone.0075672.