Plastid genome sequencing reveals biogeographical structure and extensive population genetic variation in wild populations of *Phalaris arundinacea* L. in north-western Europe

AUDE PERDEREAU¹, MANFRED KLAAS², SUSANNE BARTH² and TREVOR R. HODKINSON¹

¹Botany School of Natural Sciences, Trinity College Dublin, The University of Dublin, Dublin 2, Ireland, ²Crops Environment and Land Use Programme, Teagasc, Oak Park, Carlow, Ireland

Abstract

New and comprehensive collections of the perennial rhizomatous reed canary grass (*Phalaris arundinacea*) were made in NW Europe along north-to-south and east-to-west clines from Denmark, Germany, Ireland, Poland, Sweden and the United Kingdom. Rhizome, seed and leaf samples were taken for analysis and genetic resource conservation. A subsample covering the geographical range was characterized using plastid genome sequencing and SNP discovery generated using a long-read PCR and MiSeq sequencing approach. Samples were also subject to flow cytometry and all found to be tetraploid. New sequences were assembled against a *Lolium perenne* (perennial ryegrass) reference genome, and an average of approximately 60% of each genome was aligned (81 064 bp). Genetic variation was high among the 48 sequenced genotypes with a total of 1793 SNPs, equating to 23 SNPs per kbp. SNPs were subject to principal coordinate and Structure analyses to detect population genetic groupings and to examine phylogeographical pattern. Results indicate substantial genetic variation and population genetic structuring of this allogamous species at a broad geographical scale in NW Europe with plastid genetic diversity organized more across an east-to-west than a north-to-south cline.

Keywords: energy crops, marginal land, perennial rhizomatous grassesphylogeography, plastid genome, SNPs

Received 26 January 2016 and accepted 29 February 2016

Introduction

*Phalaris arundinacea*, or reed canary grass, is a C₃ photosynthetic perennial patch-forming species with circum-polar boreotemperate distribution (Anderson, 1961; Casler et al., 2009a). It is important for grazing but has also come under scrutiny as a biomass and bioenergy crop (Gyulai et al., 2003; Lewandowski et al., 2003; Sanderson & Adler, 2008; Wrobel et al., 2009; Kukk et al., 2011). It is native to Asia, Europe and North America and is believed to have been cultivated in Europe since the mid-18th century, starting in Scandinavia (Alway, 1931; Sahramaa, 2004). It is also widely naturalized outside its native range (Jakubowski et al., 2013; Voshell & Hilu, 2014) and populations from Europe were introduced to North America after European colonization and have mixed with native populations (Casler et al., 2009a; Waggy, 2010). In fact, agricultural expansion combined with hybridization of multiply introduced Eurasian genotypes has increased genetic variability of the species in North America (Lavergne & Molofsky, 2007; Lavergne et al., 2010).

Reed canary grass is an outbreeding species with good dispersal ability (Nelson et al., 2013; Voshell & Hilu, 2014). In addition to seed dispersal, it can also spread via extensive rhizome systems and can dominate large areas with a dense broad-leaved canopy. It is also very deep rooted, a trait that allows it to occupy apparently dry areas in open ground. It tolerates flooding and can grow in shallow water due to increased aerenchyma in its roots (Cope & Gray, 2009; Jørgensen et al., 2012).

Its closest relatives are *P. caesia*, *P. rotgesii* and *P. lindigii* (which are sometimes considered conspecific), but it is also closely allied with *P. aquatica*, *P. minor*, *P. maderensis*, *P. coerulescens*, *P. appendiculata* and *P. paradoxa* in a lineage with *x*=7 as its basic chromosome number compared with some other *Phalaris* that have *x*=6 (Ferreira et al., 2002; Quintanar et al., 2007; Voshell et al., 2011; Voshell & Hilu, 2014).

*Phalaris arundinacea* has traditionally been used for grazing, hay production and soil conservation and has a popular ornamental variegated variety (var. *picta*). It has attracted attention as a bioenergy feedstock because of its high biomass yields and broad adaptation to a wide...
range of habitats including wetlands, stream banks, pastures, damp woodland, forest margins and disturbed areas. It thrives in a range of soils but particularly in clays and is rarely found in soils below a pH of 5 (Cope & Gray, 2009). It is therefore a prime candidate species for growth on marginal land that is unsuitable or unproductive for other agricultural uses. The challenge is to develop new high yielding genotypes suitable for growth in a range of habitats and on marginal land (Jing et al., 2012; Nijsen et al., 2012; Hodkinson et al., 2015).

Biomass yield and stress tolerance are key limiting factors for the production of P. arundinacea as a bioenergy feedstock (Casler et al., 2009b; Clifton-Brown et al., 2011; Jones et al., 2015). It generally has better cold and water-logging tolerance than some other potential bioenergy crops such as Miscanthus and has been utilized in Scandinavia for bioenergy and fodder (Heinson et al., 2011). One limitation is that it is considered invasive in some countries (Galatowitsch et al., 1999; Green & Galatowitsch, 2001; Lavergne & Molofsky, 2004; Barney & DiTomaso, 2011; He et al., 2011). However, there is huge potential to utilize natural genetic variation of this phenotypically variable species in breeding programmes (Brodersen et al., 2008; Cogliatti et al., 2011; Olmstead et al., 2013; Ramstein et al., 2015).

Genetic diversity of Phalaris has been studied using several DNA marker systems including microsatellites (Li et al., 2011; Barth et al., 2016) minisatellites (Gabor et al., 2014), ISSRs (McRoberts et al., 2005), AFLP (Casler et al., 2009a) and SNPs (Ramstein et al., 2015). Phalaris has also been assessed for its biochemical and morphological variation (Knowles, 1987; Østrem, 1987, 1988; Gifford et al., 2002). Voshell & Hilu (2014) used plastid DNA sequences in phylogenetic studies of Phalaris but few, if any, have examined extensive plastid genome SNP data for population and phylogeographic analyses until now. Plastid genome SNPs are particularly informative for phylogeographic studies because of their uniparental inheritance and their ability to trace seed-mediated gene flow (McGrath et al., 2007; Diekmann et al., 2012).

Several methods exist for plastid genome sequencing (Diekmann et al., 2009; Stull et al., 2013). However, we examined the utility of a long-read PCR and MiSeq DNA sequencing approach developed by Uribe-Convers et al. (2014) to sequence a large proportion of the plastid genome in the sampled populations of Phalaris included here as part of an EU funded GrassMargins project (grassmargins.com) that aimed to develop new bioenergy crops suitable for marginal lands. No reference plastid genome for Phalaris is available so we aligned the data to the related grass Lolium perenne (Diekmann et al., 2009). We conducted extensive sampling of new germplasm across north-west Europe over latitudinal and longitudinal distances of 3000 and 2500 km, respectively. Populations were sampled for ex situ conservation from a broad range of habitats and assessed for plastid genome variation, diversity, phylogeographic pattern and gene pool structure.

Materials and methods

Sample collection and ploidy measurement

Samples of P. arundinacea were collected in Denmark, Germany, Ireland, Poland, Sweden and the United Kingdom (Table 1, Fig. 1) over a distance of 3000 km along a latitudinal gradient (51°37′41.92”N to 65°59′58.04”N) and over a distance of 2500 km along a longitudinal gradient (6°45′42.29”W to 23°11′13.20”E). Ten populations per country were collected, each comprising 30 accessions. Seed from 30 plants, rhizome from 15 plants and representative herbarium specimens were taken at each site. Seed was stored in paper envelopes and rhizomes kept fresh in approximately 50 mL of potting compost in an aerated ziplock bag until they could be grown up from rhizome. Plants were screened by flow cytometry for ploidy estimation compared to P. arundinacea genotypes with known tetraploidy determined by standard Feulgen staining of mitotic root tip meristems (following Hodkinson et al., 2002) and light microscopy at ×1000 magnification (Philipot, 2011). Fresh leaves of 80 genotypes and four Phalaris standards were sent by courier to a commercial company, Plant Cytometry Services, the Netherlands (www.plantcytometry.nl), for flow cytometry analyses. DNA content of samples was compared to known internal controls (Vinca major) with DAPI fluorochrome staining and the resulting ratios used to determine ploidy with reference to Phalaris standards with known ploidy. Because of time and money constraints, not all samples were sequenced to assess plastid DNA variation. Instead, a subset of 48 accessions grown from rhizome were used for sequencing with approximately one to two plants per sampled population based on survival and health of the rhizome grown material while maximizing population number (Table 1).

Long-read PCRs and next-generation sequencing

DNA was extracted from fresh or dried leaf tissue with a DNeasy Plant Extraction kit (Qiagen, Valencia, CA, USA), and quantity and quality of DNA of each sample was checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and diluted to between 30 and 100 ng mL⁻¹. The primers and the long-read PCR protocol were as in Appendix 1 of Uribe-Convers et al. (2014). All regions (named 1–16; Fig. S1) described in the paper were tested and amplified on four samples of P. arundinacea. After two rounds of optimization (increasing/decreasing the number of cycles, the annealing temperature and/or the time of the elongation step), it was decided to leave out regions 2, 5, 11 and 14 because of inconsistent amplification. The other regions were successfully amplified in the 48 samples with or without optimization (Table S1).

PCR products were checked on a 1% agarose gel and purified by precipitation in a 20% polyethylene glycol 8000 (PEG)/
| Country  | Accession number* | Sample sequenced | Latitude | Longitude | Details of the site                                                                 |
|----------|-------------------|------------------|----------|-----------|-------------------------------------------------------------------------------------|
| Poland   | 1–30              | 54°40'48.51"N    | 17°41'11.40"E | East of Wicko, wet sandy meadow with sedges, large stand                            |
| Poland   | 31–60             | 54°45'16.95"N    | 17°33'49.44"E | Leba, wet sandy ground, near town centre                                            |
| Poland   | 61–90             | 54°33'2.78"N     | 18°25'38.91"E | Rumia, Cisowa waste ground. Sandy moist soil                                     |
| Poland   | 91–120            | 54°23'21.42"N    | 18°28'42.18"E | Near Gdansk airport, wet sandy and gravelly soil, wet                            |
| Poland   | 121–150           | 53°41'58.62"N    | 20°5'18.30"E  | Near Ostroda                                                                       |
| Poland   | 151–180           | 53°50'2.83"N     | 21°3'30.36"E  | Along road near Borki Wielkie. Between road and edge of wheat fields              |
| Poland   | 181–210           | 53°50'36.21"N    | 21°37'26.03"E | North Konin, large stand, near water pipe. Wet dip in field not used for agriculture |
| Germany  | 1–30              | 52°14'35.24"N    | 14°25'49.56"E | Mullrose, in clearing on open land near but not on canal in town                  |
| Germany  | 31–60             | 52°4'55.62"N     | 14°9'47.04"E  | South of Trebatsch, ditch and field edge                                          |
| Germany  | 61–90             | 52°13'9.87"N     | 11°42'25.14"E | Flooded major river bank/plain near Hohenwarthe, wet edge – large stands          |
| Germany  | 91–120            | 52°40'25.74"N    | 9°43'45.00"E  | North east of Merkendorf, edge of forest, wetland                                |
| Germany  | 121–150           | 52°58'30.54"N    | 9°50'46.62"E  | Near Saltau, large stand on floodplain with sedges and along stream edges         |
| Germany  | 151–180           | 53°25'51.60"N    | 9°46'39.78"E  | Train Elsdorf                                                                    |
| Germany  | 181–210           | 54°2'42.55"N     | 9°5'13.34"E   | Train crossing west of Windbergen, wet edge of agricultural field                |
| Germany  | 211–240           | 54°8'35.12"N     | 8°50'20.58"E  | North of Busum                                                                    |
| Denmark  | 1–30              | 54°32'44.89"N    | 9°9'31.39"E   | North of Immenstedt, edge of agricultural fields (wheat)                          |
| Denmark  | 31–60             | 55°53'19.21"N    | 8°56'16.51"E  | Ellholt, on border of Germany and Denmark, large stands                           |
| Denmark  | 61–90             | 55°21'21.06"N    | 8°40'2.52"E   | South of Esbjerg, North of Ribe, near coast                                       |
| Denmark  | 91–120            | 56°54'44.50"N    | 8°13'48.57"E  | Ringkobing, edge of fjord, saline, with Phragmites                               |
| Denmark  | 121–150           | 56°12'13.48"N    | 8°9'39.07"E   | West of Stadil, Brunbjerg, along edge of cycle way, in dip and on bank near edge of fjord |
| Denmark  | 151–180           | 56°17'52.74"N    | 8°52'18.66"E  | Near Hallundbaek, acidic wetland below heath, near willow stand, floodplain, large stand |
| Denmark  | 181–210           | 56°26'55.95"N    | 9°36'44.33"E  | South of Viborg and Foulum, large stands along floodplain. Very wet and waterlogged. |
| Denmark  | 211–240           | 56°2'42.06"N     | 9°40'59.13"E  | Near Old Rye, Mossso, good stand on small river, agricultural land but along river and drainage channels |
| Denmark  | 241–270           | 55°51'15.77"N    | 12°33'8.31"E  | Maglemose, large stands                                                          |
| Denmark  | 271–300           | 55°56'16.98"N    | 12°28'35.70"E | West of Niverod, large stand                                                      |
| Sweden   | 1–30              | 56°10'47.02"N    | 13°1'25.70"E  | Tranarp, near bridge -river floodplain                                            |
| Sweden   | 31–60             | 56°46'57.22"N    | 13°53'55.22"E | North of Kanna, edge of river. Various small streams                             |
| Sweden   | 61–90             | 57°46'49.32"N    | 14°10'17.33"E | Jonkoping edge of water in town. Disturbed, growing well and setting seed. Recently reclaimed land |
| Sweden   | 91–120            | 58°52'20.04"N    | 14°53'56.79"E | South of Askersund.                                                              |
| Sweden   | 121–150           | 59°14'33.04"N    | 15°0'34.71"E  | Near Vintrosa/Orebro, wet river bank and floodplain                               |
| Sweden   | 151–180           | 59°30'48.14"N    | 15°57'38.31"E | Near Koping, gravel and moist soil not waterlogged                               |
| Sweden   | 181–210           | 63°45'3.39"N     | 20°12'5.63"E  | Umea south. Near Lake and in clearing. Waste sandy ground among pine forest      |
| Sweden   | 211–240           | 64°36'27.31"N    | 20°57'4.32"E  | North of Hjorgbole, large stand on field                                         |
| Sweden   | 241–270           | 65°59'58.04"N    | 21°12'28.88"E | Nedre Svartha, on edge of field on waterlogged soil                               |
amplicons of each individual were pooled together and quantified with Qubit dsDNA BR assay kit (Invitrogen, Paisley, UK) and then diluted down to 2 ng μL⁻¹. Sequencing libraries were constructed using the Illumina Nexerta XT library preparation kit and protocol (Illumina, San Diego, CA, USA). The resulting libraries were multiplexed in one Illumina MiSeq Reagent Nano kit version 3 and 150 cycles single-end reads were performed at the Trinity Genome Sequencing Laboratory, Institute of Molecular Medicine, St James Hospital, Trinity College Dublin, Ireland.

Data analyses
The Galaxy online public server was used to analyse the MiSeq data (http://usegalaxy.org/) (Giardine et al., 2005; Blankenberg et al., 2010; Goecks et al., 2010). FastQ files were imported into the server. The read quality was checked using FastQC (Andrews, 2010). Reads were then cleaned at high stringency (minimum quality = 30, maximum number of bases allowed outside of quality range = 5) and assembled against a reference genome (L. perenne L., GenBank accession no. NC009950, Diekmann et al. (2009)) using BWA with default parameters (Li & Durbin, 2009). SNP and INDELs were called using MPileup tool (Li et al., 2009), visually checked on IGV v2.3 (Broad Institute, www.broadinstitute.org) then filtered (minimum coverage depth = 5) using SnpSift filter tool (Cingolani et al., 2012). The individual files were combined, filtered again (maximum-likelihood estimate of the 1st alternative allele frequency AF1 = 1) and then recoded into Excel.

Genetic structure was investigated using STRUCTURE v2.3.4. (Pritchard et al., 2000; Falush et al., 2003), which applies the Markov chain Monte Carlo (MCMC) algorithm. This procedure clusters individuals into populations and estimates the proportion of membership in each population for each individual. An admixture model with correlated allele frequencies was used, the K value was set from 2 to 15, and 10 runs were performed.
for each value of $K$. The length of the burn-in period was set to 3000, and the MCMC chains after burn-in were run for an additional 2000 times. The optimal value of $K$ was determined by examination of the $D_k$ statistic (Evanno et al., 2005) using Structure Harvester (Earl & vonHoldt, 2012). The frequency distribution for the most probable number of $K$ clusters was mapped into ARCGIS 10.1 (ESRI) for each sample. Finally, principal coordinate analyses were run into GENALEX 6.5 (Peakall & Smouse, 2012).

Results

All genotypes screened for DNA content by flow cytometry were uniformly tetraploid (Table S2). DNA content ratios relative to the internal standard were 1.41, 1.42, 1.44 and 1.46 for the known tetraploid plants and ranged from 1.42 to 1.50 for the other 80 Phalaris samples included here. All samples screened for ploidy and collected in NW Europe were therefore confirmed as tetraploid.

Long-read PCR was successful for most primer combinations of Uribe-Convers et al. (2014). With just 10 PCRs per sample, it is possible to amplify approximately two-thirds of the plastid genome. The amplicons were quantified, combined and sequenced using an Illumina MiSeq. Of 10 106 895 reads, 9 762 130 passed the filter and 97.75% of the reads were assigned to an individual, for an average of 198 802 reads per individual. The reads were not trimmed due to excellent per base sequence quality (from the FastQC reports). Over 89% of the reads passed the high-stringency filter (8 509 126 reads). Reads were assembled to the reference genome, and on average, 57.9% of the genome (78 294 bp) was aligned (ranging from 41 981 to 97 555 bp, Table 2). The resulting assemblies had an average depth of coverage of 97× (minimum 30×–maximum: 241×). Sequences can be found in GenBank at Bioproject ID PRJNA301092.

A total of 93.3% of the SNPs passed the filter for a total of 2385 SNPs and INDELS. After removing the SNPs only found between Phalaris and the reference, 1793 SNPs were discovered among the Phalaris samples (Table 3), with an average of 23 SNPs per kbp. The average number of SNPs in coding regions per kbp was 13. SNP content per gene was highly variable ranging from 0 in the photosystem II reaction centre protein J

Table 2 Mapped length of Phalaris arundinacea plastid genome sequenced

| Study | Complete genome | accession NC009950 (bp) |
|-------|-----------------|------------------------|
| Max    | 135 282         | 113 900                |
| % of Max| 61.23%          | 32.86%                 |
| Min    | 78 294          | 41 981                 |
| % of Min| 87.65%          | 76.00%                 |
| Max    | 97 555          | 97 555                 |
| % of Max| 97.93%          | 97.93%                 |

© 2016 The Authors. Global Change Biology Bioenergy Published by John Wiley & Sons Ltd., 9, 46–56
The ribosomal protein L2 (rpl2), the ribosomal protein S18 (rps18) and the transfer RNA gene trnL-GAU to 113 in the ribosomal protein S12 (rps12).

The SNP data for *Phalaris* genotypes were subject to principal coordinate and Structure analyses to define cytoplasmic gene pools and examine clustering of individuals in a phylogeographic context. Axes 1 and 2 of the principal coordinate analysis (PC1 and PC2) explained 19.8% and 7.1% of the variation, respectively, and, although not entirely discrete (separated), a number of general groupings can be determined. The samples from the eastern part of the sampling range (Poland and Germany) are clustered to the right part of the plot and the western samples from Ireland and Britain towards the left part of the plot. Some rough groupings by country were also resolved. Structuring is more evident across an east-to-west than north-to-south cline.

The Structure analysis revealed four clusters (Fig. S2) with the $D_{K}$ statistic (Evanno et al., 2005), and membership of genotypes to these clusters follows some but not strong geographical pattern (Figs 3 and 4). One cluster type is more common in the east (blue in Fig. 3) and another more common in the west and Sweden (red in Fig. 3). Some structure can be seen when cluster membership is mapped in ARCGIS (Fig. 4) especially for the Polish and Swedish material. The yellow and blue cluster types are less common in the west and the blue type particularly prevalent in the east.

**Discussion**

The utility of the long-read PCR MiSeq whole-genome plastid DNA sequencing method for *Phalaris*

A number of methods exist for whole-genome sequencing of plant plastid DNA (Diekmann et al., 2009; Uribe-Convers et al., 2014). We applied an amplification-based method of

---

**Table 3** SNPs and INDELs recorded in *Phalaris arundinacea*

| Name            | Count |
|-----------------|-------|
| Coding region (CDS + rRNA + tRNA + STS) | 1021  |
| Intron          | 243   |
| Unknown (Intergenic DNA + repeat regions) | 529   |
| Total           | 1793  |
| SNP per kbp (out of the average number of bp mapped) | 22.90 |
| SNP in coding regions per kbp | 13.04 |

**Table 3 (continued)**

| Name            | Count |
|-----------------|-------|
| Coding region (CDS + rRNA + tRNA + STS) | 13    |
| Intron          | 12    |
| Unknown         | 87    |
| Total           | 112   |
| SNP count per gene |       |
| atpB            | 58    |
| atpE            | 14    |
| atpI            | 12    |
| cemA            | 16    |
| clpP            | 21    |
| infA            | 11    |
| matK            | 41    |
| ndhA            | 86    |
| ndhC            | 9     |
| ndhH            | 57    |
| ndhI            | 5     |
| ndhJ            | 7     |
| ndhK            | 22    |
| petA            | 18    |
| petB            | 24    |
| petD            | 25    |
| psaA            | 58    |
| psaB            | 71    |
| psbA            | 12    |
| psbB            | 34    |
| psbC            | 0     |
| psbH            | 4     |
| psf             | 0     |
| psfK            | 12    |
| psfN            | 3     |
| rbcL            | 17    |
| Unknown sequenced region | 393  |
| rpl14           | 16    |
| rpl16           | 36    |
| rpl2            | 0     |
| rpl20           | 6     |
| rpl23           | 8     |
| rpl33           | 3     |
| rpoA            | 46    |
| rpoB            | 64    |
| rpoC1           | 31    |
| rpoC2           | 102   |
| rps11           | 15    |
| rps12           | 113   |
| rps14           | 8     |
| rps15           | 6     |

(continued)
plastid DNA isolation and enrichment using the long-read PCR primers of Uribe-Convers et al. (2014). We found that amplification and MiSeq sequencing was highly effective using the protocols and ‘universal’ primers described in Uribe-Convers et al. (2014) with nearly 90% of reads passing the high-stringency filter. The universal primers were designed for the eudicot and showed good application to the monocot grass species studied here.

In addition, when the MiSeq reads were assembled to the reference genome for each genotype, an average of 57.9% of the genome (78 294 bp) was aligned (ranging from 41 981 to 97 555 bp, Table 3). The resulting assemblies had an average depth of coverage of 97× (minimum 30×, maximum: 241×). Previous studies have used molecular markers to study Phalaris diversity including microsatellites (Li et al., 2011), minisatellites (Gabor et al., 2014), ISSRs (McRoberts et al., 2005), and AFLP (Casler et al., 2009a), but these are not as easily transferrable among laboratories as the SNP markers developed here. We have submitted the sequences to GenBank with the Bioproject ID PRJNA301092, and the data are available for future comparisons and analyses. It would, for example, be useful to extend the sampling to southern and eastern Europe, and the data can be used directly for that purpose. No complete plastid genome has been published for P. arundinacea, and we had to align our data to a related grass species L. perenne (Diekmann et al., 2009). However, we have sequenced a maximum of 85% of the plastid genome of Phalaris in this manuscript. Such data will be highly valuable for various aspects of comparative genomics in the future with the added value of having population level data (not just a single genotype).

Nuclear SNP markers have also been developed and used for P. arundinacea by Ramstein et al. (2015) and M. Klaas, S. Barth & T.R. Hodkinson (in preparation) using genotyping by sequencing (GBS). Ramstein et al. (2015) used GBS to generate 5228 nuclear SNP markers and applied them in a genomewide association study with 35 traits, but they did not assess plastid genome variation nor assess population genetic variation at broad geographical scale. Populations of P. arundinacea have been assessed by M. Klaas, S. Barth & T.R. Hodkinson (in preparation) with approximately the same population of plants included here. It will be useful to compare and interpret our plastid DNA results with those of nuclear SNP analysis in future studies. Combined they will provide an excellent resource to organize and manage the genetic resources of Phalaris and to study its molecular ecology including its invasion dynamics.

Diversity and plastid DNA variability

The extent of SNP discovery of the 48 sampled plants from the European populations indicates that it is not necessary to amplify and sequence the entire plastid genome of each P. arundinacea genotype because sufficient diversity is recorded in the partial genomes sequenced in this study (approximately 60% of the genome). Plastid DNA is maternally inherited and each locus would be expected to be linked and provide similar population genetic and phylogeographic signal, apart from some differences that could be attributed to variation in mutation rate. In fact, mutation rate is high and not uniform among genes (Table 3). The samples of P. arundinacea were collected in a broad range of habitats separated by large geographical distances. Genotypes were sampled from Poland, Germany, Denmark, Sweden, Ireland and the United Kingdom over a distance of 3000 km along a latitudinal gradient and over a distance of 2500 km along a longitudinal gradient (Fig. 1; Table 1). The sampling strategy aimed to encompass maximum geographical variation with plants collected from waterlogged and dry well-drained sites, from freshwater and saline habitats and from large and small stands (Table 1). The results of our plastid DNA study indicate that a high genetic diversity of Phalaris has been collected.

High levels of plastid variability have also been recorded in other grasses such as L. perenne using nuclear SSR markers by McGrath et al. (2007). Lolium is also an outbreeding wind-pollinated species but differs in its habitat preferences to P. arundinacea. High variability was found in genic as well as nongenic regions. It is not known what adaptive significance these variants might have in European ecosystems, but it is clear that the breeders have a high genetic variability of cytotypes to work from. The plastid types will, but not always because of uniparental inheritance, plastid capture and lower mutation rates (Hodkinson et al., 2002), reflect patterns of nuclear genetic variability.

Population structure and differentiation

The principal coordinate analysis (Fig. 2) showed some geographical patterning of the genotypes despite the percentage of variation explaining each axis being low (PC1 20% and PC2 7%). Groupings can be seen in the PC1 vs. PC2 plot, but there is also considerable overlap among country groups in ordination space reflecting a relatively high degree of gene flow. The samples from the eastern part of the sampling range (Poland and Germany) are grouped to the right-hand part of the plot. Structuring is more evident across an east-to-west than north-to-south cline. This might reflect the high proportion of more northern samples that had an oceanic (Britain, Ireland, Denmark,
Sweden) rather than a more continental climate of eastern Germany and Poland.

The structure analysis identified four major clusters (Figs 3 and S2), and similar pattern to the principal coordinate analyses is recovered (Figs 3 and 4). One cluster type is more common in the east including Poland and Germany (shown in blue in Fig. 3) and another is more common in the north Europe including Ireland, Sweden and the United Kingdom (red in Fig. 3). Considerable admixture is also evident in some genotypes but not in others. The third and fourth clusters (green and yellow) do not relate to geographical distribution and are relatively randomly distributed across the range (Fig. 3) except that yellow is rarer in Ireland and the United Kingdom.

Dilution of population genetic structuring relative to geographical proximity is expected because reed canary grass is a common outbreeding, wind-pollinated species with a gametophytic self-incompatibility system (Nelson et al., 2013). It also has good dispersal ability via its

---

Fig. 2  Principal coordinates analysis of Phalaris arundinacea plastid DNA SNP data. Axis one accounts for 19.8% of the variation and axis two 7.1% of the variation. Ellipses show some grouping of genotypes relative to geography. Polish and German samples are more common in the right ellipse. Irish and British samples are mostly in the lower left ellipse and Danish samples in upper left ellipse. Swedish samples are found in the second ellipse and also form a subgroup in lower left of the plot.

Fig. 3  Estimated population structure of Phalaris arundinacea using plastid SNP data and Structure analysis. Each individual is represented by a vertical bar. SNP variation partitioned into coloured segments that represent the individual’s estimated membership fractions in each of the $K = 4$ clusters (cluster 1 green; cluster 2 red; cluster 3 blue; cluster 4 yellow) determined as optimal by the $\Delta K$ statistic of Evanno et al. (2005). Samples arranged according to country.
seeds that are produced in profusion and are dispersed over the water and germinate in spring (Cope & Gray, 2009; Voshell & Hilu, 2014). It can also regenerate and spread via extensive rhizome systems over large areas. It is also possible that genotypes have been deliberately or inadvertently moved by humans for forage or habitat restoration. Unlike many other perennial rhizomatous grasses such as Dactylis, Lolium and Festuca, P. arundinacea is not extensively planted and moved around Europe by seed. We sampled populations with no known history of cultivation and do not expect to have sampled commercial cultivars or material introgressed via cultivar/wild population hybridization. In a separate study as part of the grassmargins.com project, we sampled Dactylis glomerata across the same geographical range and did not detect the same extent of population genetic structuring relating to geographical location (data not shown) possibly due to the higher level of planting and human-mediated seed movement in that forage species.

Polyploidy does not explain any of the witnessed structuring of genotypes as all samples were recorded as tetraploid (Table S2). A range of ploidy levels have been recorded in European P. arundinacea, but the most common type is tetraploid. Our wide sampling of north-western European material indicates that tetraploid genotypes are indeed the dominant type and perhaps the only type. Phalaris is variable morphologically in its vegetative and reproductive characters, and because of this, some taxonomists recognize infraspecific categories, including subspecies and varieties (Anderson, 1961). Cytologically, diploid (2n = 2x = 14), tetraploid (2n = 4x = 28) and hexaploid (2n = 6x = 42) cytotypes have been recorded (Anderson, 1961; McWilliam & Neal-Smith, 1962; Ferreira et al., 2002). These three cytotypes were recognized as different taxa (Baldini, 1993, 1995) on the basis of morphological differences, chromosome number, ecology and geography: P. rotgesii (Husn.) Baldini (2x), restricted to Corsica and Sardinia, P. arundinacea L. (4x), with holarctic distribution, and P. caesia Nees (6x), originated in the Mediterranean area and spread to Eastern and Southern Africa. Also, several studies indicate the presence of aneuploids and other meiotic irregularities (Ferreira et al., 2002).

Conclusions

Our analysis of near-whole-plastid genome data in P. arundinacea clearly defines groupings for genetic resource characterization and has helped define useful
breeding material for future evaluation. It has provided valuable data on plastid genome variation of *P. arundinacea* across north-west Europe that can be compared to nuclear GBS data in future studies. Plastid genetic SNP variation is high and genotypes show some broad-scale population genetic structuring particularly in an east-to-west cline despite the good dispersal ability of *P. arundinacea* via seed and its allogamous wind-pollinated breeding system.

**Acknowledgements**

We thank Ruby Prickett for collecting some *Phalaris* plant material in Ireland and Valeria Mattiangelii from the Smurfit Institute of Genetics, Trinity College Dublin and Elaine Kenny from the Trinity Genome Sequencing Laboratory Dublin for technical assistance with the MiSeq sequencing. This work was supported by EU FP7 KBBE.2011.3.1-02 grant number 289461 (GrassMargins).

**References**

Alway FJ (1931) Early trials and use of reed canary grass as a forage plant. *Journal of the American Society of Agronomy*, 23, 64–66.

Andersen D (1961) Taxonomy and distribution of the genus *Phalaris*. *Irish State Journal of Science*, 36, 1–96.

Andrews S (2010) FastQC a quality control tool for high throughput sequence data [Online]. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed March 2013).

Baldini RM (1993) The genus *Phalaris* (Gramineae) in Italy. *Wettbl.*, 47, 1–53.

Baldini RM (1995) Revision of the genus *Phalaris* L. (Gramineae). *Wettbl.*, 49, 265–329.

Barney JN, Ditomaso JM (2011) Global climate niche estimates for bioenergy crops and invasive species of agronomic origin: potential problems and opportunities. *PLoS One*, 6, e17222.

Barth S, Jankowska MJ, Hedkinson TR, Vellanti T, Klaas M (2016) Variation in sequences containing microsatellite motifs in the perennial bison grass and forage grass, *Phalaris arundinacea* (Poaceae). *BMC Research Notes*, 9, 184, doi: 10.1186/s13104-016-1994-6.

Blankenbengel D, Kuster GV, Coraor N et al. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Current Protocols in Molecular Biology*, 89, 19.10.1–19.10.21.

Bredeson C, Laverenge S, Molofsky J (2000) Genetic variation in photosynthetic characteristics among invasive and native populations of reed canarygrass (*Phalaris arundinacea*). *Biological Invasions*, 10, 1317–1325.

Casler MD, Phillips MM, Krohn AL (2009a) DNA polymorphisms reveal geographic characteristics among invasive and native populations of reed canarygrass (*Phalaris arundinacea*). *American Journal of Botany*, 96, 238–248.

Casler MD, Cherney JH, Brummer EC (2009b) Biomass yield of naturalized populations and cultivars of reed canary grass. *Biomass Research*, 2, 165–173.

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X (2012) Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, Snpsift. *Frontiers in Genetics*, 3, 35.

Clifton-Brown J, Ebson P, Sanderson R, Hastings A, Valentine J, Donnison I (2011) Thermal requirements for seed germination in *Miscanthus* compared with Switchgrass (*Panicum virgatum*), Reed canary grass (*Phalaris arundinacea*), Maize (*Zea mays*) and perennial ryegrass (*Lolium perenne*). *GCB Biennurs*, 3, 375–386.

Cogliatti M, Bongiorno F, Dalla Valle H, Rogers WJ (2011) Canaryseed (*Phalaris aquatica*) accessions from nineteen countries show useful genetic variation for agronomic traits. *Canadian Journal of Plant Science*, 91, 37–48.

Cope T, Gray A (2009) *Grasses of the British Isles*. B.S.B.I. Handbook No. 13. Botanical Society of the British Isles, London.

Diekmann K, Hedkinson TR, Wolfe KH, van den Berkrom R, Dix PJ, Barth S (2009) Complete chloroplast genome sequence of a major allomagous forage species, perennial ryegrass (*Lolium perenne*) L. *DNA Research*, 16, 165–176.

Diekmann K, Hedkinson TR, Barth S (2012) New chloroplast microsatellite markers suitable for assessing genetic diversity of *Lolium perenne* and other related grass species. *Annu. of Botany*, 110, 1327–1339.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanvo method. *Conservation Genetics Resources*, 4, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 6611–6620.

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.

Ferreira V, Reynoso L, Sepiniak B, Grass E (2002) Cytological analysis of the *Phalaris arundinacea* (L.) × *Phalaris aquatica* (L.) amphiploid. *Carpologia*, 55, 151–160.

Gabor C, Oana-maria B, Samfira I, Sorina F (2014) Evaluation of genetic diversity of some *Phalaris* sp. using minisatellites. *Journal of Horticulture, Forestry and Biotechnology*, 18, 141–146.

Galatowsich SM, Anderson NO, Ascher PD, Hall A (1999) Invasiveness in wetland plants in temperate North America. *Wetlands*, 19, 733–755.

Gardine B, Riemer C, Hardison R (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Research*, 15, 1451–1457.

Gifford ALS, Ferdy JB, Molofsky J (2002) Genetic composition and morphological variation among populations of the invasive grass, *Phalaris arundinacea*. *Botany*, 80, 779–785.

Goeccks J, Nekrutenko A, Taylor J (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology*, 11, 806.

Green EK, Galatowsich SM (2001) Differences in wetland plant community establishment with additions of nitrate-N and invasive species (*Phalaris arundinacea* and *Typha* sp. glauca). *Canadian Journal of Botany*, 79, 170–178.

Gyulai G, Mester Z, Kiss J, Szeman L, Idrumn A, Hészky L (2003) Somaclonal breeding of reed canarygrass (*Phalaris arundinacea* L.) and *Grass and Forage Science*, 58, 210–214.

He Z, Bentley LP, Holaday AS (2011) Greater seasonal carbon gain across a broad temperature range contributes to the invasive potential of *Phalaris arundinacea* (Poaceae: reed canary grass) over the native sedge *Carex stricta* (Cyperaceae). *American Journal of Botany*, 98, 20–30.

Heinsoo K, Hein K, Melts I, Holm B, Ivask M (2011) Reed canary grass yield and fuel quality in Estonian farmers’ fields. *Biomass and Bioenergy*, 35, 617–625.

Hodkinson TR, Chase MW, Takahashi C, Leitch I, Bennett MD, Renzo A (2002) The use of DNA sequencing (ITS and *trnL*-F), AFLP, and fluorescent in situ hybridization to study allopolyploid *Miscanthus* (Poaceae). *American Journal of Botany*, 89, 279–286.

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.

Jakubowski AR, Casler MD, Jackson RD (2013) Genetic evidence suggests a widespread distribution of native North American populations of reed canarygrass. *Biological Invasions*, 15, 261–266.

Jing Q, Conijn SJC, Jongschappe RE, Bindraban PS (2012) Modeling the productivity of energy crops in different agro-ecological environments. *Biomass and Bioenergy*, 46, 618–633.

Jones MB, Finnan J, Hedkinson TR (2015) Morphological and physiological traits for higher biomass production in perennial rhizomatous grasses grown on marginal land. *GCB Bioenergy*, 7, 375–385.

Jørgensen CJ, Struwe S, Elsberling B (2012) Temporal trends in *N₂O* flux dynamics in a Danish wetland – effects of plant-mediated gas transport of *N₂O* and *O₂* following changes in water level and soil mineral-N availability. *Global Change Biology*, 18, 210–222.

Knowles RP (1987) Genetics of seed color in reed canarygrass, *Phalaris arundinacea* L. *Canadian Journal of Plant Sciences*, 67, 1051–1055.

Kool K, Roostalu H, Suuser E, Roosser H, Shanskiy M, Astover A (2011) Reed canary grass biomass yield and energy use efficiency in Northern European pedoclimatic conditions. *Biomass and Bioenergy*, 35, 4407–4416.

Laverenge S, Molofsky J (2004) Reed canary grass (*Phalaris arundinacea*) as a model species in the study of plant invasions. *Critical Reviews in Plant Sciences*, 23, 415–429.

Laverenge S, Molofsky J (2007) Increased genetic variation and evolutionary potential drive the success of an invasive grass. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 3883–3888.

Laverenge S, Muenke NJ, Molofsky J (2010) Genome size reduction can trigger rapid phenotypic evolution in invasive plants. *Annuals of Botany*, 105, 109–116.

Lewandowski I, Sculock JMO, Lindvall E, Christou M (2003) The development and current status of perennial rhizomatous grasses as energy crops in the US and Europe. *Biomass and Bioenergy*, 25, 335–361.
