Integrating a newly developed BAC-based physical mapping resource for *Lolium perenne* with a genome-wide association study across a *L. perenne* European ecotype collection identifies genomic contexts associated with agriculturally important traits

J. Harper¹, J. De Vega², S. Swain¹, D. Heavens², D. Gasior¹, A. Thomas¹, C. Evans¹, A. Lovatt¹, S. Lister¹, D. Thorogood¹, L. Skøt¹, M. Hegarty¹, T. Blackmore¹, D. Kudrna³, S. Byrne⁴, T. Asp⁵, W. Powell⁶, N. Fernandez-Fuentes¹ and I. Armstead¹,*

¹Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK, ²Earlham Institute, Norwich Research Park, Norwich NR4 7UH, UK, ³Arizona Genomics Institute, School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA, ⁴Teagasc, Department of Crop Science, Carlow, Ireland, ⁵Department of Molecular Biology and Genetics, Crop Genetics and Biotechnology, Aarhus University, Slagelse, Denmark and ⁶Scotland’s Rural College, Edinburgh, UK

*For correspondence. E-mail: ipa@aber.ac.uk

Received: 25 June 2018 Returned for revision: 17 September 2018 Editorial decision: 21 November 2018 Accepted: 28 November 2018

- Background and Aims *Lolium perenne* (perennial ryegrass) is the most widely cultivated forage and amenity grass species in temperate areas worldwide and there is a need to understand the genetic architectures of key agricultural traits and crop characteristics that deliver wider environmental services. Our aim was to identify genomic regions associated with agriculturally important traits by integrating a bacterial artificial chromosome (BAC)-based physical map with a genome-wide association study (GWAS).
- Methods BAC-based physical maps for *L. perenne* were constructed from ~212 000 high-information-content fingerprints using Fingerprint Contig and Linear Topology Contig software. BAC clones were associated with both BAC-end sequences and a partial minimum tiling path sequence. A panel of 716 *L. perenne* diploid genotypes from 90 European accessions was assayed in the field over 2 years, and genotyped using a *Lolium* Infinium SNP array. The GWAS was carried out using a linear mixed model implemented in TASSEL, and extended genomic regions associated with significant markers were identified through integration with the physical map.
- Key Results Between ~3600 and 7500 physical map contigs were derived, depending on the software and probability thresholds used, and integrated with ~35 k sequenced BAC clones to develop a resource predicted to span the majority of the *L. perenne* genome. From the GWAS, eight different loci were significantly associated with heading date, plant width, plant biomass and water-soluble carbohydrate accumulation, seven of which could be associated with physical map contigs. This allowed the identification of a number of candidate genes.
- Conclusions Combining the physical mapping resource with the GWAS has allowed us to extend the search for candidate genes across larger regions of the *L. perenne* genome and identified a number of interesting gene model annotations. These physical maps will aid in validating future sequence-based assemblies of the *L. perenne* genome.

Key words: *Lolium perenne*, BAC-based physical map, FPC, LTC, GWAS, ecotypes, candidate genes, flowering, heading date, water-soluble carbohydrate, plant width.

INTRODUCTION

*Lolium perenne* (perennial ryegrass) is widely grown in Northern Europe and in temperate areas worldwide as a forage and amenity grass. For a number of decades there have been public and private plant breeding programmes for *L. perenne*, and the *L. perenne* seed constituent is both the largest and most commercially valuable component within many marketed grass seed mixes. Due to this wide cultivation, in addition to its economic and agricultural value, it contributes to the environmental footprint of large areas of grasslands. Thus, the trait biology of this grass can have major effects on landscapes, soil structures, hydrology, water quality and carbon sequestration, in addition to the ‘intended’ qualities that recommend it for on-farm use.

*Lolium perenne* is a diploid species that contains a two-locus gametophytic self-incompatibility system (Thorogood et al., 2002, 2017). As a consequence, both commercial and *in situ* ecotype populations are often highly heterozygous and variable both within and between populations (Blackmore et al., 2015, 2016). One advantage of this is that there is considerable phenotypic variation across the gene pool that is available to plant breeders for targeted crop improvement. The disadvantage is that delimiting and controlling this variation, in order to define a variety that meets the requirements for distinctness, uniformity and stability, can be a major challenge.

© The Author(s) 2019. Published by Oxford University Press on behalf of the Annals of Botany Company. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.
Historically, in order to understand the complexity of the genetic control of target traits in *L. perenne* and related species, there has been a major focus on the analysis of quantitative trait loci (QTL) using genetic maps and trait data developed from tightly defined families (for an overview see Shinozuka et al., 2012). These have been derived from two-way pseudo-test-crosses (i.e. a pair cross of two heterozygous individuals; e.g. Inoue et al., 2004; Ergon et al., 2006; Byrne et al., 2009; Brown et al., 2010; Hegarty et al., 2013; Paina et al., 2016) or, where available, using inbred lines or doubled haploid genotypes to derive *F*₂ and *BC*₁₅-type populations (e.g. Yamada et al., 2004; Turner et al., 2006; Kobayashi et al., 2011; Foito et al., 2015) and *Lolium*/*Festuca* introgression populations (Humphreys et al., 2005; Moore et al., 2005; Armstead et al., 2006a, b). More recent developments for rapid generation of high-density genetic maps, such as the *Lolium*/*Festuca* DAfT array (Kopecky et al., 20 09), a *Lolium* Infinium SNP-genotyping array (Blackmore et al., 2015) and genotyping by sequencing (Ashraf et al., 2016; Pemberton et al., 2016; Velmurugan et al., 2016), have enabled a number of genome-wide association studies (GWAS) and germplasm sampling studies to be conducted across *L. perenne* populations, often in the context of advanced breeding populations, aiming at defining marker/trait associations with greater resolution for both candidate gene identification and genomic selection. Across these studies, associations and genomic predictions have been developed particularly for heading date, but also for crown rust resistance, seed yield, biomass, water-soluble carbohydrate (WSC) and dry matter digestibility (Blackmore et al., 2015, 2016; Fè et al., 2015; Ashraf et al., 2016; Grinberg et al., 2016; Byrne et al., 2017; Aroju et al., 2018; Cericola et al., 2018).

Concurrent with the development of new methods of genetic analysis, there have been remarkable advances in the technologies and software available for generating genome sequences and assembling long-distance contiguous DNA sequence reads, which can, or are approaching the ability to, approximate complete chromosome (pseudomolecule) sequences. In-depth transcriptome sequencing and assembly have also guided detailed annotations of many genomes that both describe gene structure and predict function. Recent published advances in this area include two initial genome assemblies for *L. perenne* (Byrne et al., 2015; Velmurugan et al., 2016) and a high-quality reference genome assembly for *Hordeum vulgare* (Mascher et al., 2017). These have built on the more established reference assemblies for rice (International Rice Genome Sequencing Project, 2005) and *Brachypodium distachyon* (International Brachypodium Initiative, 2010) amongst grass species. Along with direct sequence assembly methods, other approaches that rely on different metrics have been used in developing and validating sequence-based contigs, including bacterial artificial chromosome (BAC)-based physical mapping (e.g. Gu et al., 2009; Zhou et al., 2009; Fleury et al., 2010; Ariyadasa et al., 2014; Varshney et al., 2014). This technique relies on developing ‘fingerprints’ of individual BACs based on restriction enzyme fragment sizes, and identifying overlapping BAC clones through fingerprint matching (Sulston et al., 1988; Luo et al., 2003). Extended regions of contiguity can then be developed using software approaches implemented in FingerPrint Contig (Soderlund et al., 1997, 2000) and Linear Topology Contig (Frenkel et al., 2010). Physical map assemblies can then be integrated with sequence-based assemblies through BAC-end sequencing (BES) and selective BAC sequencing, often based on projecting a minimum tiling path (MTP) of BACs through each physical contig. This has proved to be an important adjunct in constructing and validating sequence scaffolds in genome assemblies (Mascher et al., 2017).

For many decades, plant breeders and others have been collecting perennial ryegrass and related germplasm from diverse locations in order to preserve and curate the range of genetic variation available within this species group; the gene bank at Aberystwyth University alone contains in excess of 6000 different accessions of *Lolium* spp. However, characterizing such accessions genotypically or phenotypically is a major challenge, but is also fundamental to their future exploitation. With climate uncertainty and changing patterns of land use, it is pressing that we develop a greater understanding of the potential of these germplasm collections for both enhancing the agricultural relevance of forage grass varieties and addressing new challenges that may relate more directly to enhancing environmental services, such as water and nutrient use efficiency and carbon sequestration. The work we present here is part of this effort to exploit new germplasm resources through linking new genetic and genomic approaches with comprehensive phenotyping studies. Specifically, the aim of the present study has been to integrate a newly developed genomics resource for *L. perenne* with the results of a GWAS of a broad selection of European *L. perenne* ecotypes in order to define the genomic contexts of significant GWAS markers and to suggest candidate genes. This will enable the prioritization of particular *L. perenne* accessions and genotypes for further evaluation in terms of broadening the range of germplasm available and deepening our understanding of the biology that underpins addressing future challenges.

**MATERIALS AND METHODS**

**Figure 1** illustrates the workflow described in the following sections and the Supplementary Data.

**Development of a BAC-based *L. perenne* physical map**

The development of the *L. perenne* BAC-based physical map is described in more detail in Supplementary Data Methods S1. Briefly, two BAC libraries, each consisting of 120 960 clones, were developed from an inbred *L. perenne* genotype using HindIII or BstY1 partial restriction and all clones were fingerprinted using high-information-content fingerprinting (HICF; Luo et al., 2003). Fingerprint data derived from the individual BACs were used to construct physical maps using FingerPrint Contig (FPC) and Linear Topology Contig (LTC) software. We obtained *Fpc* v. 9.4 from [http://www.agcol.arizona.edu/software/fpc/](http://www.agcol.arizona.edu/software/fpc/) with HICF data analysed largely as described in Kim et al. (2007). The LTC software was obtained from MultiQTL Ltd (Haifa, Israel). Initial LTC physical map construction was performed using LTCbeta 2.1 [physical map version LTC-18(2s)] and subsequent LTC physical maps were constructed using LTC 1.4.6, using a range of different stringencies from
$P = 1 \times 10^{-12}$ to $1 \times 10^{-27}$ (physical map versions LTC-12 to LTC-27) for the construction of networks of significant clone overlaps, from which were derived the BAC contigs constituting each version of the physical map.

**Sequence tagging of the L. perenne physical map**

**Shotgun clone sequencing.** An MTP was predicted from the physical map version LTC-18(2s), and 39 202 BAC clones were shotgun-sequenced, largely using the methods of Mascher et al. (2017). Each BAC was assembled individually using ABYSS with Knr length 71 (v. 1.5.2; Simpson et al., 2009) and BACs with a total assembly length <30 or >200 kb were re-sequenced. Methods employed in the further concatenation of the assemblies, alignment with the L. perenne draft genome assembly of Byrne et al. (2015) and further characterization of the sequence database used for anchoring GWAS markers to the physical maps (designated the LpBAC5000 database) are provided in Supplementary Data Methods S2.

**BAC-end sequences.** BAC-end sequencing of all 24 1920 clones was carried out following standard protocols using M13 primers, Sanger sequencing and capillary fragment separation and detection (Kim et al., 2007).

**Three-dimensional BAC pool screening with PCR-based markers**

Previously genetically mapped markers were associated with BACs using 3-D BAC pool screening. BAC pools were constructed from both BAC libraries by Amplicon Express (Pullman, WA, USA) according to a ‘superpool/matrix pool’ approach (Bouzidi et al., 2006). PCR-based markers were
screened across the BAC pools using either agarose gel electrophoresis or KASP technology (LGC, Teddington, UK). Further details are provided in Supplementary Data Methods S3 [which also describes a restriction site associated DNA (RAD) sequencing pilot study for assigning BES with genetic positions] with additional marker sequences provided in Supplementary Data Table S1.

Population genotyping and phenotyping

The development and genotyping of the ecotype population has been described previously (Blackmore et al., 2015). Briefly, 716 genotypes from a European ecotype collection sampled from 90 locations ($n = 8$ for 89 locations, $n = 4$ for one location; Supplementary Data Table S2) were maintained and grown as spaced plants in a field site at Aberystwyth, UK, over a period of 2 years. Plants were genotyped using a custom Illumina Infinium iSelect array across 3425 SNPs and 2034 informative markers (minimum minor allele frequency set at 5 %) spanning the L. perenne genome identified.

Plants were grown as spaced plants in three replicated blocks, with management of plants following standard procedures for perennial ryegrass national testing (Supplementary Data Methods S4). Plants were evaluated for a range of growth-related phenotypes and biochemical measures as described in Table 1. Field-based measurements and analytical chemistry measurements for WSC, neutral and acid detergent fibre (ADF/ NDF), total nitrogen (N) and dry matter digestibility were conducted according to standard practices used in perennial ryegrass national variety assessments (Supplementary Data Methods S5).

GWAS

The GWAS was implemented using TASSEL (Trait Analysis by aSSociation, Evolution, and Linkage) 5.2.4.1 (https://bitbucket.org/tasseladmin/tassel-5-source/downloads/; Bradbury et al., 2007) using the mixed linear model MLM (PCA+K) workflow, which genetic population structure is estimated using principal components analysis and fitted as a fixed effect, and genotype relatedness is estimated through a kinship matrix and fitted as a random effect (Zhang et al., 2010). TASSEL default parameters were used throughout (the first five principal components were used for population structure estimation; minimum minor allele frequency was set at 5 %). Dry weight data were square-root-transformed prior to analysis. False discovery rates were controlled using the Benjamini–Hochberg procedure at a level of 10 %.

RESULTS

Physical map construction

Summary statistics for the two BAC libraries and their subsequent processing are described in Supplementary Data Table S3. The details of the derived physical maps assembled using FPC and LTC are summarized in Table 2. In comparing the different versions of the physical map, LTC-15 and particularly LTC-12 incorporated fewer clones than the other physical maps. The main reason for this was that at the lower stringency levels more clones had >500 significant overlaps and so were excluded from the subsequent assemblies. Both FPC physical maps incorporated more clones than any of the LTC physical maps, presumably as a consequence of the different methods employed by the two softwares for establishing contig reliability. Overall, 29 482 clones were excluded from all of the physical map versions, 125 321 clones were common to all of the LTC physical maps from LTC-18 to LTC-27, and 121 842 of these were also common to FPC1.03. Where clones were incorporated in contigs across cut-off thresholds and assembly methods, contig composition and clone order within contigs tended to be highly conserved (for a list of the physical maps contigs generated, see Supplementary Data Table S4, available from https://doi.org/10.20391/bb56e6d7-8913-4bd7-8167-2b7e4c01382b).

Direct anchoring of physical maps to genetic/genomic positions

Of the 1240 markers that were screened on the superpools and matrix pools, 887 could be assigned to contigs (excluding contigs with >1000 constituent BACs) across the FPC and LTC methods and stringencies with the requirement that the marker was positive for a minimum of two different BACs in the same contig. Overall, this resulted in between 522 and 711 contigs being assigned genetic/genomic positions, depending on the assembly software and stringency used. Marker/BAC/physical contig associations are given in Supplementary Data Table S4.

BAC-end sequences

Summary statistics for the BES of 241920 clones are given in Supplementary Data Table S3 (BES data are available from NCBI under GenBank accession numbers MJ032229–MJ424519 and can also be downloaded from https://doi.org/10.20391/61921116-dd04-4d85-b0fd-e0d734bc63c8). For the clones that were incorporated in the versions of the physical maps, ~71 % had BES for both ends of the clone, 21 % had BES for one end of the clone and 8 % had no associated BES information.
Table 2. Summary statistics for L. perenne physical maps

| Version | FPC-1.02 | FPC-1.03 | LTC-12 | LTC-15 | LTC-18 | LTC-21 | LTC-24 | LTC-27 | LTC-18(2a) |
|---------|----------|----------|--------|--------|--------|--------|--------|--------|------------|
| HICF BACS* | 211 880 | 211 880 | 211 880 | 211 880 | 211 880 | 211 880 | 211 880 | 211 880 | 154 836 |
| Contigs (ctgs) | 7472 | 4029 | 3765 | 3695 | 4110 | 4745 | 5937 | 6780 | 4048 |
| Number of BACS in ctgs | 163 865 | 163 112 | 107 280 | 132 519 | 145 994 | 151 793 | 152 579 | 147 193 | 142 887 |
| ctgs ≤1000* | 45 052 | 45 805 | – | – | – | – | – | – | – |
| ctgs >1000† | 2963 | 2963 | 104 600 | 79 361 | 65 886 | 60 087 | 59 301 | 64 687 | 11 949 |
| Singletons BACS/ctg | 101–1000 | 91 | 349 | 211 880 | 3765 | 211 880 | 3765 | 211 880 | 11 949 |

*Number of BACs associated with HICFs.
†Total number of physical contigs consisting of either ≤1000 or >1000 BACs.

Shotgun clone sequencing

We assembled 35 434 BACs individually. The average BAC assembly length was 77.5 kb, the median number of sequences was 9 and N50 contig length was 10.3 ± 4.8 kb (Supplementary Data Table SI, B). From these, 33 480 BACs were extracted with total contig sequence length in the size range 30–200 kb, which formed the basis of the LpBAC5000 database. Further analysis details are provided in Supplementary Data Results S1. This sequencing project files are deposited online under Bioproject PRJNA475227, BioSample:SAMN09382314, SRA Sample:SRS3412769 and SRA Study:SRP150420.

LpBAC5000 database

Overall, LpBAC5000 consisted of 207 958 sequences totaling ∼2634 Mb extracted from 33 480 clones (~85 %) of the predicted MTP. The average contig size was ~12.7 kb and the average sequence length per BAC was ~78.7 kb. In order to estimate the coverage of the gene space within LpBAC5000, the alignment lengths of LpBAC5000 sequence contigs to gene models within the lope_transcripts.V1.0.fasta, Bdistachyon_314_v3.1.cds.fasta and Hv_IBSC_PGSB_r1_CDS_HighConf.fa databases were estimated. For the L. perenne database, ~89 % of the gene space was represented, tagging ~97 % of the gene models. For the B. distachyon and H. vulgare databases, the equivalent figures were ~73 and ~78 % for the gene space and ~74 and ~85 % for the tagged gene models, respectively. Summary statistics of the LpBAC5000 sequence database are given in Supplementary Data Table S5A, B. The LpBAC5000 sequence file can be downloaded from https://doi.org/10.20391/dbf05330-7485-444f-a475-8310bee5d510.

Integration of the GWAS with the sequence-tagged physical map

BLAST searches of significant marker sequences against the LpBAC5000 database were used to identify associated physical map contigs (Table 4, Fig. 3, Supplementary Data Fig. S3). For six of the identified physical contigs there was a strong indication of conserved synteny, with both B. distachyon and H. vulgare with physical contigs encompassing between eight and 32 B. distachyon gene models. The remaining physical contig, 1969c-18, identified by marker ctg20671, could be associated with three gene models in total; two were physically close on the physical map version. This compares with the Byrne et al. (2015) estimate of 2.07 Gb from their draft genome assembly (Supplementary Data Table S6).
**DISCUSSION**

The aim of this paper has been to explore the targeted use of *L. perenne* BAC-based physical maps to suggest broader genomic regions, validated through the physical contig structures, that could be associated with markers found to be significant in a GWAS analysis derived from the field performance of a selection of European *L. perenne* ecotypes. To this end, the physical map assemblies and associated sequence tags presented in the paper encompass the majority of the *L. perenne* genome and extend the ‘genomic reach’ of the current published sequence-based assemblies for *L. perenne* (Byrne et al., 2015; Velmurugan et al., 2016). As a result, we have been able to identify both genomic regions and relevant gene models contained within these regions, which are associated with the GWAS markers and phenotypes. As a significance level we chose to use a 10 % FDR as, in taking this work forward, this represents a good balance between the likelihood of identifying useful variation and the resources required for screening plant material in follow-up evaluations. In terms of candidate genes, as always for non-model species (and particularly so for non-cereal grass species) there is a frustrating gap between our ability to identify candidate genes and our subsequent ability to validate these candidate genes experimentally. However, even in the absence of direct functional validation of candidates, knowledge of the regions of genomes that contribute to trait performances are valuable in identifying potentially useful genotypes/phenotypes present in germplasm collections for further experimentation.

Blackmore et al. (2015, 2016) reported the establishment of this panel of 716 *L. perenne* ecotypes sampled from 90 locations across Europe and described the genetic relationships between population and patterns of linkage disequilibrium on the basis of the genotype scores from the Infinium marker array. The present study, in part, represents a follow-up analysis incorporating marker and phenotype data, but also integrates this within a genomic context. As is common to a number of population genetics analyses within the ryegrasses, flowering time features strongly in terms of traits that can be associated with genetic markers and genomic regions, and four of the eight loci within the 10 % FDR were associated with heading date, one from year 1 and three from year 2. As an outcome this is not particularly surprising, as the phenology of flowering is known to be a key determinant of many growth, physiological and biochemical parameters in the annual cycle of perennial grasses. Commercially, the selection of genotypes that have co-ordinated flowering times represents a primary route for controlling phenotypic variation. In addition to flowering time, significant marker/trait associations were identified for plant width, plant biomass and WSC content. The integration of the physical mapping with the GWAS is discussed below on a marker-by-marker/contig-by-contig basis.

**TABLE 3. Significant marker/trait associations from the GWAS**

| Trait                  | Year | Month | Marker          | Genetic map position | $P^0$ | $F_{1}^1$ | Ab × Au$^2$ |
|------------------------|------|-------|-----------------|----------------------|------|----------|-----------|
|                        |      |       |                 |                       |      | LG       | cM        | LG       | cM       |
| Plant width            | 1    | May   | ctg8394-538     |                       | 1.66 × 10^-4 | 7     | 68       | -         | -        |
| Heading date           | 1    | –     | ctg50617-428    |                       | 2.41 × 10^-4 | -     | -        | 2         | 35       |
| WSC$^3$                | 1    | June  | ctg35543-1175   |                       | 1.86 × 10^-4 | -     | -        | 5         | 41       |
| WSC                    | 1    | June  | ctg41386-226    |                       | 9.62 × 10^-4 | 6     | 83       | 6         | 43       |
| Heading date           | 2    | –     | ctg8613-723     |                       | 1.02 × 10^-4 | -     | -        | -         | -        |
| Plant biomass          | 1    | June  | ctg54379-73     |                       | 3.71 × 10^-4 | 3     | 55       | 3         | 50       |
| Heading date           | 2    | –     | ctg9479-1216    |                       | 4.17 × 10^-4 | 4     | 56       | 4         | 43       |
| Heading date           | 2    | –     | ctg20671-156    |                       | 5.13 × 10^-3 | -     | -        | 7         | 50       |

1. *L. perenne* $F_{1}$ genetic map.
2. *L. perenne* AberMagic × Aurora genetic map.
3. Significant with FDR controlled within 10 %.
4. Markers ctg35543-503/641/365/281 (number after ‘−’ is the position of the SNP within the contig) were also significant within 10 % FDR ($P = 2.79 \times 10^{-5}$ to $8.41 \times 10^{-6}$).
5. Markers ctg40624-321/549 were also significantly associated with WSC ($P = 1.52 \times 10^{-5}$). Ctg40624 genetically maps to the same position on LG5 and is likely to be derived from a different exon from the same gene as ctg35543.

*B. distachyon* chromosome 3 and the third was from chromosome 1. This last gene model had no *H. vulgare* orthologue at the stringencies tested. Marker ctg41386-226 was significantly associated with WSC, but could not be aligned with any of the physical contigs through the LpBAC5000 database.

BLAST searches of LpBAC5000 against the published *L. perenne* draft genome sequence were also used to suggest scaffolds from this draft genome assembly that were contiguous according to the physical map contigs. The numbers of draft genome scaffolds aligned varied from nine for 4073c-18 to one for 1969c-18 (Fig. 3, Supplementary Data Fig. S3).

Marker ctg8613: physical contig 4073c-18; trait, heading date year 2 (Fig. 3A)

Marker ctg8613 had its closest BLAST alignment with gene model Bradi2g51467, which associated it with physical contig 4073c-18. This was a relatively large physical contig consisting of 138 BACs and spanning 2227 high-information-content (HIC) fingerprints. In terms of conserved synteny with *B. distachyon*, this contig spanned a region containing...
36 gene models from Bradi2g51255 to Bradi2g51550, 32 of which were tagged within LpBAC5000 (Table 4). Twenty-seven of these gene models were associated with annotations, one of which (Bradi2g51370/HORVU3Hr1G075540) represents a member of the Casein Kinase 1 gene family (CK1). Members of this gene family have been directly associated with the modulation of flowering time in both Arabidopsis thaliana (hereafter referred to as arabidopsis) and rice (Tan et al., 2013; Kwon et al., 2015; Zhang et al., 2016) and, in the latter case, a CK1 gene underlies the Hd16/Early Flowering 1 heading date QTL (Dai and Xue, 2010; Hori et al., 2013). Marker ctg8613 has not previously been genetically mapped in L. perenne but superpool/matrix pool screens of a different marker, R1_302, identified three independent BACs from contig 4073c-18. This would place this contig on chromosome 3 of the Lolium/Festuca introgression map (King et al., 2013). Additionally, marker ctg41342 could also be aligned with physical contig 4073c-18 and this has been previously mapped in the AberMagic × Aurora genetic mapping population to chromosome 3 at 49 cM (marker ctg41342 was 79% homozygous for one of the alleles in the present study and was not significantly associated with heading date). This placing of physical contig 4073c-18 on chromosome 3 would not suggest that this CK1 is a direct orthologue of the gene underlying the Hd16 QTL in rice (Os03g0793500/LOC_Os03g57940), but is consistent with orthology with Os01g0772600/LOC_Os01g56580. While we are not aware of any rice heading date QTL directly associated with
Table 4. Brachypodium distachyon (Bd) and Hordeum vulgare (Hv) gene models associated with physical contigs identified by significant markers from the GWAS

| Bd Marker | Hv Marker | Functional annotations |
|-----------|-----------|------------------------|
| ctg8613   | ctg8613   | K11838 - ubiquitin carboxyl-terminal hydrolase 7 (USP7, UBP15) (1 of 4) [RICE]ubiquitin carboxyl-terminal hydrolase, putative, expressed [ARAB]ubiquitin-specific protease 13 |
Table 4. Continued

| Bd1 | Hv1 | Functional annotations12 |
|-----|-----|--------------------------|
| 1_58210 | 2_060710 | 2_060730 | aldol-keto reductase activity (Blast2GO) [RICE] | oxidoreductase, aldol/keto reductase family protein, putative, expressed [ARAB] | NAD(P)-linked oxidoreductase superfamily protein |
| 1_58220 | 2_060710 | 2_060730 | aldol-keto reductase activity (Blast2GO) [RICE] | oxidoreductase, aldol/keto reductase family protein, putative, expressed [ARAB] | NAD(P)-linked oxidoreductase superfamily protein |
| 1_58230 | 2_060680 | aka phytochrome interacting factor like [RICE] | helix-loop-helix DNA-binding domain containing protein, expressed [ARAB] | phytochrome interacting factor 3 |
| 1_58240 | 2_060630 | /60650 | ATF-dependent RNA helicase activity (Blast2GO) [RICE] | DEAD-box ATP-dependent RNA helicase, putative, expressed [ARAB] | putative mitochondrial RNA helicase 2 |
| 1_58250 | 2_060620 | /60640 | – |
| 1_58251 | – | – |
| 1_58252 | – | – |
| 1_58254 | – | – |
| Marker = ctg35543 Trait = WSC P = 2.66 × 10−4 Physical ctg = 4029c-18 |
| 4_31640 | 5_066810 | /66820 | microtubule minus-end binding (Blast2GO) [RICE] | Spc97/Spc98 family protein, putative, expressed [ARAB] | spindle body pole component 98 |
| 4_31645 | 5_066850 | – |
| 4_31650 | – | – |
| 4_31660 | 5_066860 | – |
| 4_31670 | 5_066860 | /692066950 | – |
| 4_31680 | – | – |
| Marker = ctg8394 Trait = plant width P = 1.66 × 10−4 Physical ctg = 2309c-18 |
| 1_34850 | 7_095230 | PF02458 | - transferase family (transferase) (1 of 99) [RICE] | transferase family protein, putative, expressed [ARAB] |
| 1_34857 | 7_095190 | PF01535//PF12854//PF13041//PF13812 | - PPR repeat (PPR) | PPR repeat family (PPR_2) |
| 1_34867 | 7_095200 | – |
| 1_34878 | – | – |
| 1_34890 | 7_095220 | – |
| 1_31050 | – | – |
| 1_34900 | 7_095170 | /95180 | PF20879 - WRC (WRC) (1 of 21) [RICE] | ATP-dependent RNA helicase activity (Blast2GO) [RICE] | DEAD-box ATP-dependent RNA helicase, putative, expressed [ARAB] |
| 1_30026 | 7_119150 | PTHR22835//PTHR22763:SF116 | - ZINC FINGER FYVE DOMAIN CONTAINING PROTEIN | SUBFAMILY NOT NAMED (1 of 2) [RICE] |
| Marker = ctg9479 Trait = heading date (Y2) P = 4.17 × 10−4 Physical ctg = 2356c-18 |
| 4_26270 | 4_026720 | glycosyl hydrolase (GH), subfamily GHS1 [RICE] | alpha-N-arabinofuranosidase A, putative, expressed [ARAB]| alpha-L-arabinofuranosidase 1 |
| 4_26275 | – | PF02365 | - no apical meristem (NAM) protein (NAM) (1 of 133) [RICE] | no apical meristem protein, putative, expressed [ARAB] | NAC domain containing protein 52 |
| 4_26280 | – | – |
| 4_26287 | 4_026690 | pectin methylesterase inhibitor (PMEI) [RICE] | PME/invertase inhibitor, putative, expressed [ARAB] |
| 4_26300 | 4_026680 | PINFORMED-like auxin efflux carrier, syntetic to Sb05g002150, Os11g04190, Os12g04000 [RICE] | auxin efflux carrier component, putative, expressed [ARAB] | auxin efflux carrier protein |
| 4_26310 | 4_026630 | – |
| 4_26317 | 4_026640 | PTHR22439//PTHR22439:SF130 - SERINE/THREONINE-PROTEIN KINASE/SUBFAMILY NOT NAMED (1 of 2) [RICE] | CAMK_CAMK-like | 42 CAMK includes calcium/calmodulin dependent protein kinases, expressed [ARAB] | calcium-dependent protein kinase 17 |
Harper et al. — L. perenne BAC-based physical map integrated with European ecotype GWAS

Table 4. Continued

| Bd¹ | Ht² | Functional annotations³,⁴ |
|-----|-----|---------------------------|
| 4_26330 | 0_026660 | holo-[acyl-carrier-protein] synthase activity ([Blast2GO]_RICE)dehydrogenase-phosphopantetheinytransferase, putative, expressed, [ARAB]—phosphopantetheinyl transferase superfamily |
| 4_26342 | 0_026570 | KOG2615 - permease of the major facilitator superfamily (1 of 10), [RICE]major facilitator superfamily antipporter, putative, expressed, [ARAB]—phosphatidylinositol transferase superfamily |
| 4_26354 | 0_026570 | KOG2615 - permease of the major facilitator superfamily (1 of 10), [RICE]major facilitator superfamily antipporter, putative, expressed, [ARAB]—phosphatidylinositol transferase superfamily |
| 4_26366 | 0_026570 | KOG2615 - permease of the major facilitator superfamily (1 of 10), [RICE]major facilitator superfamily antipporter, putative, expressed, [ARAB]—phosphatidylinositol transferase superfamily |
| 4_26369 | 0_026340 | — |
| 4_26380 | 0_026390 | — |

Marker = ctg54379 Trait = biomass $P = 3.71 \times 10^{-4}$ Physical ctg = 896c-18

| 2_50230 | 0_027830 | PTHR22792:SF58 - LA-RELATED PROTEIN 6A (1 of 1), [RICE]la domain containing protein, expressed, [ARAB]－RNA-binding protein |
| 2_50237 | 0_027800 | PTHR22792:SF58 - LA-RELATED PROTEIN 6A (1 of 1), [RICE]la domain containing protein, expressed, [ARAB]－RNA-binding protein |
| 2_50250 | 0_027860 | PF03740 - protein of unknown function (DUF624) (1 of 9), [RICE]expressed protein, [ARAB]—membrane protein |
| 2_50260 | 0_027850 | intracellular cyclic nucleotide activated cation channel activity, voltage-gated potassium channel activity (Blast2GO), [RICE]potassium channel KAT1, putative, expressed, [ARAB]—potassium channel in Arabidopsis thaliana 1 |
| 2_50270 | 0_027810 | PTHR10209:SF171 - GIBBERELLIN 2-BETA-DIOXYGENASE 2-RELATED (1 of 1), [RICE]gibberellin 2-beta-dioxygenase, putative, expressed, [ARAB]—gibberellin 2-beta-dioxygenase |
| 2_50280 | 0_027810 | PTHR10209:SF171 - GIBBERELLIN 2-BETA-DIOXYGENASE 2-RELATED (1 of 1), [RICE]gibberellin 2-beta-dioxygenase, putative, expressed, [ARAB]—gibberellin 2-beta-dioxygenase |

Marker = ctg20671 Trait = heading date (Y2) $P = 5.15 \times 10^{-4}$ Physical ctg = 1969c-18

| 1_14590 | 0_027940 | N-acetyl-gamma-glutamyl-p-phosphate reductase activity ([Blast2GO]_RICE)seryldehydehyde dehydrogenase, NAD binding domain containing protein, putative, expressed, [ARAB]—oxidoreductases, acting on the aldehyde or oxo group of donors, NAD or NADP as acceptor; copper ion binding |
| 3_15607 | 0_027940 | DNA-directed DNA polymerase activity (Blast2GO)_RICE]DNA-directed polymerase, putative, expressed, [ARAB]—Y-family DNA polymerase H |
| 3_15620 | 0_027940 | DNA-directed DNA polymerase activity, 3'-5' exonuclease activity ([Blast2GO]_RICE)Pol-I-like DNA polymerase, putative, expressed, [ARAB]—DNA polymerase gamma 2 |

¹Gene model identifiers are abbreviated as C_nnnnn(n), where C is the chromosome and nnnnn(n) is the identifier number. Gene models highlighted in bold are as photoregulatory protein kinase 3 (PPK3). The PPK protein family has been shown in arabidopsis to interact in the blue-light-dependent phosphorylation of cryptochrome 2, a protein involved in photoperiod-induced flowering, and ppk mutants are associated with delayed flowering (Tan et al., 2013). Thus, the CKI gene model identified on physical contig 4073c-18 is a good candidate for influencing flowering time in L. perenne.
Harper et al. — L. perenne BAC-based physical map integrated with European ecotype GWAS

Fig. 3. Diagrammatic representations of physical map contigs 4073c-18 (A) and 2356c-18 (B) in relation to conserved syntenic regions in the B. distachyon genome. Horizontal bars at the top of each panel represent the genomic region containing the B. distachyon gene models indicated. Yellow or purple rectangles on the blue background illustrate the relative positions of the B. distachyon gene models within the defined region; the purple rectangles indicate the positions of the candidate genes identified for both physical contigs in Table 4 and discussed in the text. Dark grey rectangles on the light grey background illustrate the positions of B. distachyon gene models just outside of the defined region. Wide horizontal red bars illustrate the B. distachyon gene models tagged within LpBAC5000 or within the L. perenne genomic scaffolds from Byrne et al. (2015), above and below the contig illustration, respectively. Narrow horizontal red bars indicate where gene space sequences are contiguous. Orange filled boxes indicate the aligned positions of the marker sequences. Red BACs within the contig illustrations indicate BACs for which sequence information is available. In (A), the BACs from which the aligned LpBAC5000 contigs were derived are number-coded and detailed in the key. In (B) the BAC names are given directly.
A second physical contig associated with heading date consisted of 65 BACs spanning 980 HIC fingerprints that showed conserved synteny with a region of the B. distachyon genome from Bradi1g58190 to Bradi1g58254, including ten gene models, eight of which were tagged in LpBAC5000. Of the eight gene models in this region with annotations, one, Bradi1g58230/HORVU2Hr1G060680, is predicted to code for a phytochrome-interacting factor (PIF). The PIFs are a family of transcription factors that interact with the red/far-red light receptors, phytochromes, and are so involved in light perception and photomorphogenesis (Pham et al., 2018). In arabidopsis, PIF3 has been associated particularly with seedling de-etiolation, chlorophyll and anthocyanin biosynthesis and, through interaction with circadian clock genes, diurnal physiological responses (Soy et al., 2016). It has also been shown in arabidopsis that both constitutive and tissue-specific expression of PIF3 in the plant vasculature can promote flowering; pif3 mutants can also show delayed flowering. In rice, the putative PIF3 orthologue Os07g0143200/LOC_Os07g05010, known as OsPIF14 (Cordeiro et al., 2016) or Phytochrome Interacting Factor 3-Like 14 (OsPIL14; Nakamura et al., 2007), has been suggested as being involved in the interaction between light and stress signalling through repression of Dehydration-Responsive Element-Binding 1/C-Repeat-Binding Factor (DREB1/CBF) genes (Cordeiro et al., 2016). This represents a possible route through which the PIF3 gene model within this region in L. perenne could have a significant effect on flowering time in this ecotype population.

A third physical contig associated with heading date consisted of 54 BACs spanning 977 HIC fingerprints and showed conserved synteny with a region of the B. distachyon genome from Bradi4g26270 to Bradi4g26410. The region contained 16 gene models, 13 of which were tagged within LpBAC5000. Among these, gene model Bradi4g26300/HORVU4Hr1G026680 is annotated as a PIN-FORMED auxin efflux carrier protein. Polar auxin transport is established as being a key process in plant morphogenesis in which differential auxin gradients determine the developmental differentiation of meristems (Balzan et al., 2014). Bradi4g26300 itself belongs to the Sister-of-PIN1 (SoPIN1) clade and sopin1 mutants in B. distachyon have been associated with abnormal spikelet development (O’Connor, 2014), and it is not unlikely that the orthologue of this gene in L. perenne is also involved in determining floral meristem development. Expressions of PIN genes are known to be influenced by abiotic stresses, implying that there can be an environmentally induced response in terms of auxin gradients within plant tissues (Wang et al., 2015). Particularly, for such a diverse pan-European ecotype collection grown in a single (UK) environment, a proportion of the genotypes are likely to have been under some degree of environmental stress simply due to the contrasts between the collection and evaluation sites. Thus, environmentally induced alterations in the expression patterns at the floral meristem of the L. perenne orthologue of Bradi4g26300 could result in changes in flowering patterns.

This physical contig consisted of 58 BACs spanning 978 HIC fingerprints. In terms of conserved synteny with B. distachyon, 4029c-18 included gene models from Bradi4g31640 to Bradi4g31720. Only six of the ten gene models within this region were tagged within LpBAC5000, partly due to lack of MTP coverage at one end of this contig. However, one of the BES within this region could be aligned with Bradi4g31720 (and Scaffold_3258_ref0017503, from Byrne et al., 2015), thus establishing the region of conserved synteny at the end of the contig that lacked MTP coverage. The most interesting annotated gene model within this region was Bradi4g31680, to which the most similar gene model in arabidopsis is AT1G50480, itself annotated as 10-formyltetrahydrofolate synthetase (THFS). THFS catalyses the conversion of formate and tetrahydrofolate to 10-formyl tetrahydrofolate and is an important enzyme in one-carbon metabolism (Hanson and Roje, 2001). The reverse of this reaction is catalysed by 10-formyl tetrahydrofolate deformylases, which have been established as being essential enzymes in photorespiration in arabidopsis under ambient CO2 concentration. Knockouts of 10-formyl tetrahydrofolate deformylases result in accumulation of glycine and 5-and 10-formyl tetrahydrofolate (Collakova et al., 2008). While THFS is directly metabolically linked through to methionine/serine/glycine/purine metabolism, in general terms it is an enzyme that affects the partitioning of carbon between different pools in plants. Additionally, it could also influence photosynthetic efficiency through metabolic feedback on photorespiration. Thus, variation in accumulation of WSCs across this ecotype collection could be a function of overall photosynthetic efficiency and assimilate partitioning influenced by differential THFS activity. Interestingly, Blackmore et al. (2015) identified marker ctg35543 as being one of the top 50 markers contributing to the first principal component (associated with east–west geographical distribution) in their analysis of genetic diversity. This is a further indication of the potential significance of this gene.

Physical contig 896c-18 consisted of 60 BACs and spanned 1180 HIC fingerprints. It contained a conserved syntenic region covering Bradi2g50230 to Bradi2g50340 but also included a not-immediately-co-linear gene model at the end of the contig, Bradi2g35187. The region contained 14 gene models, 11 of which were tagged within LpBAC5000. Among the gene models, Bradi2g50280 was annotated as a gibberellin 2-oxidase (GA2ox). The closest orthologue in rice to Bradi2g50280 is LOC_Os01g55240/Os01g0757200, which is identified as gibberellin 2-oxidase 3 (GA2ox3; Lo et al., 2008). The GA2ox3 family in rice is responsible for regulating the levels of biologically active gibberellins in tissues by metabolizing active gibberellin (GA) forms to inactive ones. Specifically, GA2ox3
is an important enzyme in the conversion of biologically active GA1 and GA4 to the inactive GA34 and GA8 (Lo et al., 2008). In general, GA deficiency in rice is associated with dwarf phenotypes, but there can also be effects on seed dormancy and shoot and root growth, including tillering, flowering and seed-set. Specifically, a GA2ox3 T-DNA insertion mutant in rice was found to have a severe dwarf phenotype with no seed production (Lo et al., 2008) and a rice GWAS study identified SNP markers close to GA2ox3 that were significantly associated with seed dormancy (Magwa et al., 2016). Marker ctg54379 was significantly associated with vegetative biomass measured in June of year 1. Biomass in ryegrass is a function of overall leaf growth, but also of tiller number. The GA2ox family will play an important role in the regulation of growth and architectural parameters in ryegrasses through their role in gibberellin metabolism, as they do in other plant species. However, the Q-Q plot for this trait (Fig. 2) does not obviously show the ‘elbow’ shape associated with the presence of unusually significant markers. Therefore, while there is close proximity between the GA2ox3 gene and the significant marker for biomass, the interpretation of this marker/trait association has to be treated with caution.

**Marker ctg8394 and marker ctg20671**

Marker ctg8394 and marker ctg20671 were associated with physical contigs 2309c-18 (Supplementary Data Fig. S3D) and 1969c-18 (Supplementary Data Fig. S3E), respectively. Both physical contigs had some inconsistencies in terms of conserved synteny with B. distachyon, and 2309c-18 showed inconsistent clone order and composition across the various physical map assembly parameters (Supplementary Data Table S4). Physical contig 1969c-18 was, seemingly, a reliable contig; however, gene models could only be tagged within LpBAC5000 at one end of the contig. None of the gene model annotations could be obviously related to possible functions in moderating the associated traits for either physical contig. These contigs are not discussed further.

**QTL and GWAS studies in L. perenne**

**Heading date.** There have been a considerable number of published studies of the genetic control of flowering in L. perenne and related grasses, mostly based upon QTL analyses in biparental genetic mapping families (see Shinozuka et al., 2012, for a comprehensive overview). Heading date QTL have been identified on all seven L. perenne linkage groups (LGs) and, in the present study, significant markers were associated with LGs 2, 3, 4 and 7. Often, the lack of common markers across studies and the low resolution of many QTL analyses presents problems in terms of identifying equivalent QTL, and this is certainly true when trying to put the results of the present study in a broader context. We can say that the assumed genetic position of marker ctg8613-723 on LG3 is not incompatible with the position of the heading-date QTL identified on the same LG associated with marker LpRGA5 in the study of Studer et al. (2008). Similarly, the genetic position of ctg20671-156 on LG7 is not incompatible with the minor heading-date QTL identified on LG7 associated with markers LTCoa/b in the study of Armstead et al. (2004). More certainly, we can say that none of the significant heading-date markers from the present study can be associated, in terms of genetic/genomic position, with the major heading-date QTL that have been identified on LG4 and LG7, for which there are good candidate genes, i.e. the L. perenne orthologues of Vernalisation 1 and Flowering Locus T, respectively (Armstead et al., 2004; Jensen et al., 2005; Skot et al., 2011). More recent studies have analysed GWAS data for significant associations with flowering time. Arjoju et al. (2016) used six full-sib families derived from crosses between perennial ryegrass varieties and did not identify any significant marker/trait associations in the GWAS. However, using a single-marker analysis approach, they did identify significant associations on LG2, 4 and 7. Similarly, Fe et al. (2015) presented a GWAS study using a mixture of F2 families and synthetic populations, both derived from breeding material, and identified 19 SNPs significantly (5 % FDR) associated with heading date. In the study of Fe et al. (2015) the relationship of these SNPs to the published L. perenne sequence scaffolds (Byrne et al., 2015) was provided. However, none of these scaffolds could be associated with the physical contigs identified in the present study.

**Plant width, biomass and WSC.** Marker ctg8394-538 was associated with plant width, which is a function of both tiller number and growth habit (i.e. prostrate versus erect). QTL for tiller number have been identified in different studies on all LGs of L. perenne and Lolium multiflorum (Inoue et al., 2004; Yamada et al., 2004; Kobayashi et al., 2011; Sartie et al., 2011), but the lack of common markers means that direct comparisons with the present study have very little resolution. However, in addition to tiller number, Yamada et al. (2004) also recorded prostrate versus erect growth habit and identified QTL for this trait on LG4 and 7. The QTL on LG7 had the highest LOD score of any QTL in that study and was also the only QTL that was non-co-incident with any of the other morphological traits measured. The position of this QTL [derived from one RFLP marker, RZ144, that can be linked through from the study of Yamada et al. (2004) to a previous version of the F2 genetic map] is not incompatible with the genetic position of marker ctg8394-538, though no obvious candidate gene was identified on the associated physical contig, 2309c-18.

Marker ctg54379-73 was significantly associated with plant biomass, though the Q-Q plot indicates that this may have been a chance outcome. Biomass is a trait that is a fundamental productivity measure for forage grasses and, not surprisingly, QTL have been identified on a number of different L. perenne LGs, including LG3. Two studies (Turner et al., 2008; Anhalt et al., 2009) associated the QTL peak for biomass with SSR marker rv1133. In a combined F2 genetic map that integrated some of the Infinium iSelect markers with previously genetically mapped markers (unpublished), rv1133 and ctg54379-73 genetically map ~20 cM apart. Therefore, the previously identified biomass QTL are probably not co-incident with any effect associated with marker ctg54379-73. Sartie et al. (2011) identified a QTL for tiller number (which is a component of biomass) on
LG3, but there are no common markers that cross-reference to this study.

Ctg35543-1175 and ctc41386-226 identified significant associations with WSC on LG5 and LG6 respectively. Foito et al. (2015) and Turner et al. (2006) conducted a QTL analysis for WSC accumulation and polar metabolite concentrations (including soluble sugars) in L. perenne genetic mapping populations and identified QTL for various WSC fractions on LG1, 2, 3, 5, 6 and 7, which, specifically, included a QTL on LG5 for leaf glucose content in the autumn and QTL for total WSC, fructan, glucose and fructose on LG6. However, as for a number of the studies referred to in this discussion, lack of common markers makes establishing co-incidence of markers and QTL across studies problematic.

Future directions

In this study we have described the development and deployment of a BAC-based physical mapping approach in order to inform and increase the resolution of a GWAS in terms of identifying potentially useful germplasm accessions, candidate genomic regions and genes. Clearly this is not an end-point for L. perenne genomics, the ultimate goal being to be able to define an accurate genomic sequence for L. perenne in the form of seven pseudomolecules and to integrate this with both comprehensive gene annotations and descriptions of physical genetic variation. Work is currently in progress to reach these ends at a number of research centres and it is hoped that the resources reported here will aid in the achievement of these aims. Beyond genomics, major challenges also exist for grassland geneticists and biotechnologists in understanding and exploiting the considerable germplasm resources that are available. These challenges include both the establishment of protocols and, hopefully, pipelines for higher-throughput phenotyping of L. perenne accessions and, at the molecular level, methods for validating candidate genes. In the latter context, while genetic transformation has been available for many years for L. perenne and related species and so the potential to knock out or otherwise modify the expression of specific genes is there, this is often quite a blunt tool, particularly if one is trying to understand the contribution of allelic differences to trait variation. For this, the application of Crispr Cas-9 and similar genome editing technologies (Yin et al., 2017) is likely to be the biotechnological route to a greater understanding. Thus, it is hoped that in the not too distant future we will be able to integrate phenotyping, genetics, genomics and biotechnology to more fully understand the biology of these grasses and so to contribute germplasm-based solutions to the various societal and environmental challenges and opportunities associated with grasslands.

SUPPLEMENTARY DATA

Supplementary data are available online at https://academic.oup.com/aob and consist of the following. Fig. S1: distribution of assembly lengths and N50s for sequenced BACs in the MTP. Fig. S2: QQ plots for non-significant marker/trait associations from the GWAS. Fig. S3: diagrammatic representations of physical map cts 62852c-18, 4029c-18, 896c-18, 2309c-18 and 1969c-18 in relation to conserved syntenic regions in the B. distachyon genome. Methods S1: details of physical map construction using FPC and LTC. Methods S2: methods used for concatenation of BAC assemblies for comparison with the published assembly of Byrne et al. (2015) and for screening of MTP clones for potential cross-contamination. Methods S3: construction of BAC superpools and matrix pools for marker screening and details of the RAD sequencing pilot study. Methods S4 and S5: field and analytical chemistry measurements and protocols used in the evaluation of the ecotype family. Results S1: concatenation of BAC assemblies, identification and removal of potential cross-contaminating sequences and RAD sequencing pilot study. Table S1: map positions, marker sequences and SNP positions for markers assayed on the BAC superpools and matrix pools using KASP. Table S2: geographical collection sites of L. perenne accessions used in the GWAS analysis. Tables S3, S5 and S6: summary statistics for BAC library production and fingerprinting, sequence size range in the LpBAC5000 database and estimates of the L. perenne genome coverage within the physical maps. Supplementary Data available via doi consist of the following. Table S4: L. perenne physical maps generated using FPC and LTC assembly softwares from BAC HICF data, https://doi.org/10.20391/bb56ed67-8913-4bd7-8167-2b7e4013828b; LpBAC5000 sequence database, https://doi.org/10.20391/dfb05330-7485-44f-a475-8310be5d510; BAC-end sequence database, https://doi.org/10.20391/61921116-d3d-4d85-b0fd-e0734abe63c8.

ACKNOWLEDGEMENTS

We would like to dedicate this paper to the memory of Dr Suresh Swain, who was responsible for developing the initial versions of the L. perenne physical map and, sadly, passed away during the course of this programme. This work was supported by the UK Biotechnology and Biological Sciences Research Council through Responsive Mode, Institute Strategic Programme and Core Strategic Programme grants (grants BB/J004405/1, BB/CSP1730/1 and BB/G012342/1) with financial contributions from Germinal Holdings (UK), Syngenta (UK) and Vialactia Biosciences (NZ).

LITERATURE CITED

Anhalt UCM, Heslop-Harrison JS, Phipho HP, Byrne S, Barth S. 2009. Quantitative trait loci mapping for biomass yield traits in a Lolium intrad line derived F-2 population. Euphytica 170: 99–107.

Ariyadasa R, Mascher M, Nussbaumer T, et al. 2014. A sequence-ready physical map of barley anchored genetically by two million single-nucleotide polymorphisms. Plant Physiology 164: 412–423.

Armstead IP, Turner LB, Farrell M, et al. 2004. Synteny between a major heading-date QTL in perennial ryegrass (Lolium perenne L.) and the H3 heading-date locus in rice. Theoretical and Applied Genetics 108: 822–828.

Armstead I, Donnison I, Aubry S, et al. 2006a. From crop to model to crop: identifying the genetic basis of the staygreen mutation in the Lolium/ Festuca forage and amenity grasses. New Phytologist 172: 592–597.

Armstead IP, Harper JA, Turner LB, et al. 2006b. Introgression of crown rust (Puccinia coronata) resistance from meadow fescue (Festuca pratensis) into Italian ryegrass (Lolium multiflorum): genetic mapping and identification of associated molecular markers. Plant Pathology 55: 62–67.
Aroju SK, Barth S, Milbourne D, et al. 2016. Markers associated with heading and aftermat heading in perennial ryegrass full-sib families. BMC Plant Biology 16: 160.

Aroju SK, Conaghan P, Barth S, et al. 2018. Genomic prediction of crown rust resistance in Lolium perenne. BMC Genetics 19: 35.

Ashraf BH, Byrne S, Fe D, et al. 2016. Estimating genomic heritabilities at the level of family-pool samples of perennial ryegrass using genotyping-by-sequencing. Theoretical and Applied Genetics 129: 45–52.

Balzan S, Johal GS, Carraro N. 2014. The role of auxin transporters in monocots development. Frontiers in Plant Science 5: 393.

Blackmore T, Thomas I, McMahon R, Powell W, Hegarty M. 2015. Genetic-geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (Lolium perenne) using array-based SNP genotyping. Theoretical and Applied Genetics 128: 1917–1932.

Blackmore T, Thorogood D, Skot L, McMahon R, Powell W, Hegarty M. 2016. Gernplasm dynamics: the role of ecotypic diversity in shaping the patterns of genetic variation in Lolium perenne. Scientific Reports 6: 26500.

Bouzid MF, Franchel J, Tao Q, et al. 2006. A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions. Theoretical and Applied Genetics 113: 81–89.

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.

Brown RN, Barker RE, Warnke SE, et al. 2010. Identification of quantitative trait loci for seed traits and floral morphology in a field-grown Lolium perenne x Lolium multiflorum mapping population. Plant Breeding 129: 29–34.

Byrne S, Guiney E, Barth S, Donnison I, Mur LAJ, Milbourne D. 2009. Identification of coincident QTL for days to heading, spike length and spikelets per spike in Lolium perenne. Euphytica 166: 61–70.

Byrne SL, Nagy I, Pfeifer M, et al. 2015. A syntax-based draft genome sequence of the forage grass Lolium perenne. Plant Journal 84: 816–826.

Byrne SL, Conaghan P, Barth S, et al. 2017. Using variable importance measures to identify a small set of SNPs to predict heading date in perennial ryegrass. Scientific Reports 7: 3566.

Cericola F, Lenk I, Fe D, et al. 2018. Optimized use of low-depth genotyping-by-sequencing for genomic prediction among multi-parental family pools and single plants in perennial ryegrass (Lolium perenne L.). Frontiers in Plant Science 9: 369.

Collakova E, Goyer A, Naponelli V, et al. 2008. Arabidopsis 10-formyl tetrahydrofolate deformylases are essential for photospiration. Plant Cell 20: 1818–1832.

Cordeiro AM, Figueiredo DD, Tepperman J, et al. 2016. Rice phytochrome-interacting factor protein OsPIF14 represses OsDREB1B gene expression through an extended N-box and interacts preferentially with the active form of phytochrome B. Biochimica et Biophysica Acta. Gene Regulatory Mechanisms 1859: 393–404.

Dai C, Xue HW. 2010. Rice early flowering1, a CKI, phosphorylates DELLA protein SLR1 to negatively regulate gibberellin signalling. EMBO Journal 29: 927–939.

Ergon A, Fang C, Jorgensen O, Aamld L, Rognli OA. 2006. Quantitative trait loci controlling vernalisation requirement, heading time and number of panicles in meadow fescue (Festuca pratensis Huds.). Theoretical and Applied Genetics 112: 232–242.

Fe D, Cericola F, Byrne S, et al. 2015. Genomic dissection and prediction of heading date in perennial ryegrass. BMC Genomics 16: 921.

Fleury D, Luo MC, Dvorak J, et al. 2010. Physical mapping of a large plant genome using global high-information-content-fingerprinting: the distal region of the wheat ancestor Aegilops tauschii chromosome 3DS. BMC Genomics 11: 382.

Foito A, Hackett CA, Byrne SL, Stewart D, Barth S. 2015. Quantitative trait loci analysis to study the genetic regulation of non-polar metabolites in perennial ryegrass. Plant Science 11: 412–424.

Frøkenfelt U, Paux E, Mester D, Feuillet C, Korol A. 2010. LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Bioinformatics 11: 584.

Grinberg NF, Lovatt A, Hegarty M, et al. 2016. Implementation of genomic prediction in Lolium perenne (L.) breeding populations. Frontiers in Plant Science 7: 133.

Go YQ, Ma YQ, Huo NX, et al. 2009. A BAC-based physical map of Brachypodium distachyon and its comparative analysis with rice and wheat. BMC Genomics 10: 496.

Hanson AD, Roje S. 2001. One-carbon metabolism in higher plants. Annual Review of Plant Physiology and Plant Molecular Biology 52: 119–137.

Hegarty M, Yadav R, Lee M, et al. 2013. Genotyping by RAD sequencing enables mapping of fatty acid composition traits in perennial ryegrass (Lolium perenne (L.)). Plant Biotechnology Journal 11: 572–581.

Hori K, Ogiso-Tanaka E, Matusbara K, Yamanouchi U, Ebana K, Yano M. 2015. HD16, a gene for casein kinase 1, is involved in the control of rice flowering time by modulating the day-length response. Plant Journal 76: 36–46.

Humphreys J, Harper JA, Armstead JP, Humphreys MW. 2005. Introgen-mapping of genes for drought resistance transferred from Festuca arundinacea var. glaucescens into Lolium multiflorum. Theoretical and Applied Genetics 110: 579–587.

Inoue M, Gao ZS, Cai HW. 2004. QTL analysis of lodging resistance and related traits in Italian ryegrass (Lolium multiflorum Lam.). Theoretical and Applied Genetics 109: 1576–1585.

International Rice Genome Sequencing Project. 2005. The mapped-base sequence of the rice genome. Nature 436: 793–800.

International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass Brachypodium distachyon. Nature 463: 763–768.

Jensen LB, Andersen JR, Frei U, et al. 2005. QTL mapping of vernalization response in perennial ryegrass (Lolium perenne L.) reveals co-location with an orthologue of wheat VRN1. Theoretical and Applied Genetics 110: 527–536.

Kim H, San Miguel P, Nelson W, et al. 2007. Comparative physical mapping of rice (Oryza sativa (AA genome type) and O. punctata (BB genome type). Genetics 176: 379–390.

King J, Armstead I, Harper J, et al. 2013. Exploitation of interspecific diversity for monocot crop improvement. Heredity 110: 475–483.

Kobayashi S, Humphreys MO, Tase K, Sanada Y, Yamada T. 2011. Molecular marker dissection of ryegrass plant development and its response to growth environments and foliage cuts. Crop Science 51: 600–611.

Kopecky D, Bartos J, Lukaszewski AJ, et al. 2009. Development and mapping of DAfT markers within the Festuca-Lolium complex. BMC Genomics 10: 473.

Kwon CT, Koo BH, Kim D, Yoo SC, Paek NC. 2009. Casein kinases 1 and 2 alpha phosphorylate Oryza sativa pseudo-response regulator 37 (OsPRR37) in photoperiodic flowering in rice. Molecules and Cells 38: 81–88.

Lo SF, Yang SY, Chen KT, et al. 2008. A novel class of gibberellin 2-oxidades control semidwarfism, tillering, and root development in rice. Plant Cell 20: 2603–2618.

Luo MC, Thomas C, You FM, et al. 2003. High-throughput fingerprinting of barley artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82: 378–389.

Magwa RA, Zhao H, Xing YZ. 2016. Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (Oryza sativa L.). BMC Genetics 17: 228.

Mascher M, Gundlach H, Himmelbach A, et al. 2017. A chromosome conformation capture ordered sequence of the barley genome. Nature 544: 427–433.

Moore BJ, Donnison IS, Harper JA, et al. 2005. Molecular tagging of a senescence gene by introgression mapping of a stay-green mutation from Festuca pratensis. New Phytologist 165: 801–806.

Nakamura Y, Kato T, Immitio T, et al. 2010. Genmi M, Mizuno T. 2007. Characterization of a set of phytochrome-interacting factor-like bHLH proteins in Oryza sativa. Bioscience Biotechnology and Biochemistry 71: 1183–1191.

O’Connor DL, Runions A, Sluis A, et al. 2014. A division in pin-mediated auxin patterning during organ initiation in grasses. PLoS Computational Biology 10: e1003447.

Paina C, Byrne SL, Studer B, Rognli OA, Asp T. 2016. Using a candidate gene-based genetic linkage map to identify QTL for winter survival in perennial ryegrass. PLOS One 11: e0152004.

Pemberton LW, Drayton MC, Bain M, et al. 2016. Targeted genotyping-by-sequencing permits cost-effective identification and discrimination of pasture grass species and cultivars. Theoretical and Applied Genetics 129: 991–1005.

Pham VN, Kathare PK, Hug E. 2018. Phytochromes and phytochrome interacting factors. Plant Physiology 176: 1025–1038.
Sartie AM, Matthew C, Easton HS, Faville MJ. 2011. Phenotypic and QTL analyses of herbage production-related traits in perennial ryegrass (Lolium perenne L.). Euphytica 182: 295–315.

Shinozuka H, Cogan NOI, Spangenberg GC, Forster JW. 2012. Quantitative trait locus (QTL) meta-analysis and comparative genomics for candidate gene prediction in perennial ryegrass (Lolium perenne L.). BMC Genetics 13: 101.

Simpson JP, Wong K, Jackman SD, Schein JE, Jones SJM, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. Genome Research 19: 1117–1123.

Skot I, Sanderson R, Thomas A, et al. 2011. Allelic variation in the perennial ryegrass Flowering Locus T gene is associated with changes in flowering time across a range of populations. Plant Physiology 155: 1013–1022.

Soderlund C, Longden I, Mott R. 1997. FPC: a system for building contigs from restriction fingerprinted clones. Computer Applications in the Biosciences 13: 523–535.

Soderlund C, Humphray S, Dunham A, French L. 2000. Contigs built with fingerprints, markers, and FPCV4.7. Genome Research 10: 1772–1787.

Soj J, Leivar P, Gonzalez-Schain N, et al. 2016. Molecular convergence of clock and photosensory pathways through PIF3-TOC1 interaction and co-occupancy of target promoters. Proceedings of the National Academy of Sciences of the USA 113: 4870–4875.

Studer B, Jensen LB, Hentrup S, Brazauskas G, Kolliker R, Lubberstedt T. 2008. Genetic characterisation of seed yield and fertility traits in perennial ryegrass (Lolium perenne L.). Theoretical and Applied Genetics 117: 781–791.

Sulston J, Mallett F, Staden R, Durbin R, Hornsell T, Coulson A. 1988. Software for genome mapping by fingerprinting techniques. Computer Applications in the Biosciences 4: 125–132.

Tan ST, Dai C, Liu HT, Xue HW. 2013. Arabidopsis casein kinase1 proteins CK1.3 and CK1.4 phosphorylate cryptochrome2 to regulate blue light signaling. Plant Cell 25: 2618–2632.

Thorogood D, Kaiser WJ, Jones JG, Armstead I. 2002. Self-incompatibility in ryegrass 12. Genotyping and mapping the S and Z loci of Lolium perenne L. Heredity 88: 385–390.

Thorogood D, Yates S, Manzanares C, et al. 2017. A novel multivariate approach to phenotyping and association mapping of multi-locus gametophytic self-incompatibility reveals S, Z, and other loci in a perennial ryegrass (Poaceae) population. Frontiers in Plant Science 8: 1331.

Turner LB, Cairns AJ, Armstead IP, et al. 2006. Dissecting the regulation of fructan metabolism in perennial ryegrass (Lolium perenne) with quantitative trait locus mapping. New Phytologist 169: 45–57.

Turner LB, Cairns AJ, Armstead IP, Thomas H, Humphreys MW, Humphreys MO. 2008. Does fructan have a functional role in physiological traits? Investigation by quantitative trait locus mapping. New Phytologist 179: 765–775.

Varshney RK, Mir RR, Bhatia S, et al. 2014. Integrated physical, genetic and genome map of chickpea (Cicer arietinum L.). Functional & Integrative Genomics 14: 59–73.

Velmurugan J, Mollison E, Barth S, et al. 2016. An ultra-high density genetic linkage map of perennial ryegrass (Lolium perenne) using genotyping by sequencing (GBS) based on a reference shotgun genome assembly. Annals of Botany 118: 71–87.

Wang YQ, Chai CL, Valliyodan B, Maupin C, Annen B, Nguyen HT. 2015. Genome-wide analysis and expression profiling of the PIN auxin transporter gene family in soybean (Glycine max). BMC Genomics 16: 951.

Yamada T, Jones ES, Cogan NOI, et al. 2004. QTL analysis of morphological, developmental, and winter hardiness-associated traits in perennial ryegrass. Crop Science 44: 925–935.

Yin K, Gao C, Qiu J-L. 2017. Progress and prospects in plant genome editing. Nature Plants 3: 17107.

Zhang JH, Sun HL, Zhao XY, Liu XM. 2016. Arabidopsis Casein Kinase 1-Like 8 enhances NaCl tolerance, early flowering, and the expression of flowering-related genes. Journal of Plant Interactions 11: 138–145.

Zhang ZW, Ersoz E, Lair CQ, et al. 2010. Mixed linear model approach adapted for genome-wide association studies. Nature Genetics 42: 355–360.

Zhou SG, Wei FS, Nguyen J, et al. 2009. A single molecule scaffold for the maize genome. PLoS Genetics 5: e1000711.