Rounding up the annual ryegrass genome: High-quality reference genome of *Lolium rigidum*

Jefferson Paril[1], Gunjan Pandey[2], Emma M. Barnett[1], Rahul V. Rane[3], Leon Court[2], Thomas Walsh[2] and Alexandre Fournier-Level[1]*

[1]School of BioSciences, University of Melbourne, Parkville, VIC, Australia, [2]CSIRO Land and Water, Acton, ACT, Australia, [3]CSIRO Health and Biosecurity, Parkville, VIC, Australia

The genome of the major agricultural weed species, annual ryegrass (*Lolium rigidum*) was assembled, annotated and analysed. Annual ryegrass is a major weed in grain cropping, and has the remarkable capacity to evolve resistance to herbicides with various modes of action. The chromosome-level assembly was achieved using short- and long-read sequencing in combination with Hi-C mapping. The assembly size is 2.44 Gb with N_{50} = 361.79 Mb across 1,764 scaffolds where the seven longest sequences correspond to the seven chromosomes. Genome completeness assessed through BUSCO returned a 99.8% score for complete (unique and duplicated) and fragmented genes using the Viridiplantae set. We found evidence for the expansion of herbicide resistance-related gene families including detoxification genes. The reference genome of *L. rigidum* is a critical asset for leveraging genetic information for the management of this highly problematic weed species.

Introduction

Lolium rigidum (Gaudin and Paschoud 1811) also known as annual ryegrass, rigid ryegrass, or Wimmera grass, is the world's most herbicide resistant weed species. It has developed resistance to over a dozen different modes of action across a number of herbicides and has the highest incidence of resistance in any weed species (Heap 2022). In particular, it is the first weed species reported to have evolved resistance to glyphosate (Powles et al., 1998).

L. rigidum is a diploid grass species with a chromosome number of 2n = 2x = 14 (Terrell 1966; Monaghan 1980) and an estimated genome size of ~2Gb, similar to that of the closely-related forage crop *Lolium perenne* (Byrne et al., 2015; Frei et al., 2021). This species is known to hybridise with other members of the *Lolium* genus such as *L. multiflorum* and *L. perenne* (Kloot 1983). This genus is thus a complex of cross-compatible species which can produce fertile hybrids and makes species boundaries ambiguous (Naylor 1960; Terrell 1966; Kloot 1983).

L. rigidum is a highly-competitive, self-incompatible, wind-pollinated, annual, C3 weed species (Monaghan 1980; Mccraw and Spoor, 1983), which can produce up to 45,000 seeds per square metre in wheat fields where it can achieve high densities (33%–67% abundance in agricultural field conditions; Rerkasem et al., 1980). The combination of its high fecundity and outcrossing reproduction regime results in large and genetically diverse populations with high adaptive potential. The seeds have varying levels of dormancy ensuring their persistence in the soil seedbank (Goggin et al., 2012). Ryegrass infestation causes significant yield reduction in rapeseed and cereal crops (Lemerle et al., 1995) and its seeds can get infected with *Clavibacter toxicus* causing livestock poisoning (Riley and McKay 1991; Ophel et al., 1993).

L. rigidum is native to the Mediterranean region and was widely introduced around the world as a pasture crop. In the 19th century, it was introduced to Australia (Kloot 1983) where it successfully adapted through a combination of artificial and natural selection. It is now the major weed in the wheat-growing regions of Australia (Reeves 1976; Medd et al., 1985; Powles and Matthews 1992). We selected a glyphosate-resistant plant from Australia as the source of the reference genome to represent the remarkable capacity of this weed species to evolve resistance to herbicides.

In this paper, we report a reference, chromosomal-level genome assembly of *Lolium rigidum*. This information is a valuable resource towards genomically-informed management of this major agricultural weed species with a particular emphasis on the issue of herbicide resistance evolution.

Materials and methods

Plant sampling, tissue culture, and DNA extraction

A single glyphosate-resistant plant from Wagga Wagga (NSW, Australia) was selected as the reference genotype for *Lolium rigidum*. This individual was tissue-cultured to induce embryogenic calli for clonal multiplication and maintenance following the protocol for *Lolium* spp. by Creemers-Molenaar and Beerepoot. (1992). DNA was extracted using Qiagen DNeasy plant mini kit (QIAGEN N.V., Venlo, Netherlands) following manufacturer's instructions.

Genome sequencing and assembly

Short- and long-read DNA sequence data were generated and scaffolded using Hi-C sequence information. Short-read sequencing libraries were constructed using NEBNext Ultra II DNA Library Prep kit for Illumina (NEB, United States) and sequenced using HiSeq X platform (Illumina, Inc., San Diego, United States) ran in 150-bp paired-end mode. Adapter sequences were removed from the resulting reads using TrimGalore (v 0.6.6). Long-read sequencing

was carried out on MinION (2 libraries using SQK-LSK109 kit and sequenced on FLO-MIN106D flowcell) and PromethION (1 library using SQK-LSK109 kit and sequenced on a FLO-PRO002 flowcell) platforms. Basecalling was performed using *guppy* (v5.1; Wick et al., 2019) under the *dna_r9.4.1_450bps_sup.cfg* model. The long-read sequences were trimmed using *Porechop* (v0.2.4; Wick et al., 2017) and filtered using *filtlong* (v0.2.1) to obtain high quality reads. The long-reads were assembled using *Flye* (v2.9; Kolmogorov and Derek, 2020) with the minimum overlap parameter set to 6,000, kmer size of 17, genome size of 2.25 Gb, and with no scaffolding. Duplicate contigs were purged using *purge_dups* (v1.2.5; Guan et al., 2020) with the default settings. The long-reads were error-corrected and trimmed using *Canu* (v2.2; Koren et al., 2017) under default settings, and used in three rounds of contig polishing using *Racon* (v1.4.22; Vaser et al., 2017) under default settings. This was followed by three rounds of short-read-based polishing using *Polca* (*MaSURCA* v4.0.7; Zimin et al., 2013) to obtain the final contig assembly using default settings. This assembly was assessed using *BUSCO* (v5.2.2; Simão et al., 2015) against the Viridiplantae and Poales lineages' gene sets (i.e. viriplantae_obd10 and poales_odb10).

A Hi-C library was prepared using 20 mg of leaf tissue and the Arima HiC kit following the manufacturer's instructions. The library was sequenced on NovaSeq 6000 platform (Illumina, Inc., San Diego, United States) to generate 500 million reads. The final contig assembly was scaffolded based on the genomic topological information using *ALLHiC* (v1; Zhang et al., 2019) and manually curated using JuiceBox (v1.9.8; Dudchenko et al., 2017). The assembly was checked by NCBI's GenBank decontamination pipeline, and the contaminating sequences were removed. Genome size was estimated based on the kmer distribution of the Illumina sequences with Jellyfish 2.3.0 (Marçais and Kingsford, 2011) and GenomeScope (v1.0.0; Vurture et al., 2017) with kmer ranging from 15 bp to 25 bp and the kmer with the best model fit was used. Genome assembly completeness using k-mer spectrum was automatically assessed during the Polca (*MaSURCA* v4.0.7; Zimin et al., 2013) run. LTR_retriever (Ou et al., 2018) in tandem with GenomeTools (i.e. LTR harvest; Gremme et al., 2013) were used to assess the assembly contiguity.

The assembly is available on the National Center for Biotechnology Information of the United States (NCBI) database under the accession number (SAMN25144995, JAKKIG000000000). Raw Illumina, MinION, PromethION, and Hi-C reads are available under the NCBI Bioproject PRJNA799061. The genome assembly and annotations are available to browse at http://traitnet.adaptive-evolution.org/jbrowse/JBrowse-1.16.11/.

Transcriptome sequencing, assembly, and genome annotation

Clones from the reference plant established through tissue culture were grown under greenhouse conditions. Two independent samples each of whole seedlings, roots, stems,

leaves, inflorescence and meristem tissue were snap-frozen and total RNA was extracted using Isolate II RNA plant kit (Bioline, United Kingdom). RNA sequencing libraries were synthesised for each sample using NEBNext Ultra II stranded RNA library synthesis kits, indexed using the NEBNext Multiplex Oligos for Illumina barcode kit. Libraries were quantified using NEBNext Library Quant KIt for Illumina, normalised, pooled, and sequenced on an Illumina HiSeq X Ten platform to generate ~257 million 150-bp paired-end reads. The reads were demultiplexed and error-corrected using Rcorrector (v1.0.4; Song and Florea, 2015). Adapters and low quality base pairs were trimmed using TrimGalore (v0.6.0). Ribosomal RNA sequences were discarded when one of the paired-end reads mapped to the sequences present in the SILVA database (v138.1; Quast et al., 2013) using Bowtie2 (v2.3; Langmead & Salzberg, 2012). After filtering, ~197 million reads were used for *de novo* transcriptome assembly including the rice protein sequences (release 51 Os-Nipponbare-Reference-IRGSP-1.0) as guide and using both Trinity (v2.8.4; Haas et al., 2013) and Oases (v0.2.09; Schulz et al., 2012) as assemblers. The resulting two assemblies were merged into a single compacted meta-assembly using the *De novo* RNA-Seq Assembly Pipeline (Cabau et al., 2017). The filtered reads were re-mapped against the meta-assembly and transcripts with FPKM>1 were included in the transcriptome.

The genome was annotated using NCBI's genome annotation pipeline using the *de novo* assembled transcriptome. Transposable elements were identified using RepeatMasker and RepeatModeller (v4.1.2 and v2.0.3, respectively; Flynn et al., 2020).

Comparative genomics

The reference genomes, annotations and coding DNA sequences (CDS) of *Arabidopsis thaliana* (TAIR10 v1), rice (*Oryza sativa*; IRGSP v1), sorghum (*Sorghum bicolor*; NCBI v3), maize (*Zea mays*; B73 Reference NAM v5), and perennial ryegrass (*Lolium perenne*; Kyuss v1) were used for comparative genomics analyses. The well-curated genome of *A. thaliana* was used as the outgroup. Rice and maize genomes represent well-annotated grass genomes. Perennial ryegrass is a closely related species. Sorghum is an additional grass crop species.

OrthoFinder (v2.5.4; Emms and Keyll 2019) was used to cluster all the CDS of the six species into orthogroups which includes paralogs within species and orthologs among species. The resulting orthogroups were assigned to gene families they most likely belong to using HMMER (v3.3.2; Mistry et al., 2013) and PantherHMM gene family models (v17; Mi et al., 2019). Significant gene family contraction and expansion in each of the six genomes were determined using CAFE (v5; De Bie et al., 2006) with a *p*-value<0.01. The significantly expanded gene families were used for gene ontology (GO) enrichment analysis using the GO consortium's web tool "Gene ontology enrichment analysis tool" (The UniProt Consortium, 2019) with a *p*-value<0.05. An additional GO term enrichment analysis was performed with two random sets of 500 randomly sampled genes from the full set of significantly enriched genes to test the consistency of the analysis result.

Orthogroups consisting of a single gene in each of the six genomes, i.e. single-copy gene orthogroups were used to generate a phylogenetic tree by maximum likelihood. These single-copy gene orthogroups were aligned using MACSE (v2.06; Ranwez et al., 2011) and concatenated into a single alignment. The phylogenetic time tree was generated using IQ-TREE (v2.0.7; Minh et al., 2020), where the best fitting substitution model per orthogroup was selected using ModelFinder (Kalyaanamoorthy et al., 2017), and dates of ancestral nodes were obtained from TimeTree.org (i.e. fossil record estimates of median divergence times between *A. thaliana* and rice estimated to 160 million years ago (MYA), sorghum and perennial ryegrass: 62 MYA, and perennial ryegrass & annual ryegrass: 2.74 MYA).

The rate of transversions at four-fold degenerate sites (4DTv) for each pair of sequences across paralogs within species and orthologs across species was calculated to estimate relative divergence times and identify whole genome duplication (WGD) events. For computational efficiency, only the paralogs and orthologs with 2 to 5 members were included in 4DTv calculations.

Herbicide resistance genes

The use of a glyphosate-resistant plant as reference genome allowed the investigation of the potential genomic basis of herbicide resistance. The resistance-conferring genomic features may be point mutations in genes coding for essential enzymes targeted by herbicides or in detoxification genes. These mutations can be detected using pairwise rates of synonymous and non-synonymous substitutions, i.e. Ka/Ks ratio (1: neutral, >1:positive selection, <1: stabilising selection), estimated between homologous pairs of protein coding sequences. Additionally, the resistance-conferring changes may be structural variants leading to gene loss or duplication. These can be detected by assessing the patterns of expansion and contraction in the genes coding for the target of glyphosate (i.e. enolpyruvylshikimate phosphate synthase; EPSPS indispensable for aromatic amino acid synthesis), and detoxification-related gene families.

To perform these analyses, 8 enzymes known to confer herbicide resistance were selected: enolpyruvylshikimate phosphate synthase (EPSPS), superoxide dismutase (SOD), ascorbate peroxidase (APX), glutathione S-transferase (GST), monodehydroascorbate reductase (MDAR), glutathione peroxidase (GPX), cytochrome P450 (CYP450), and ATP-binding cassette transporter (ABC). The sequences of these proteins were downloaded from the Universal Protein

Estimated genome size	2.40 Gbp
Total assembly size	2.44 Gbp
Total scaffolds	1,764
N_{50}	0.36 Gbp
L_{50}	4
Longest scaffold	0.41 Gbp
Number of chromosomes	7
Total chromosome length	2.35 Gbp
Gapped length (N's)	0.13%
GC content	44.76%
Total length of retroelements	0.82 Gb (33.61%)
Total length of DNA transposons	0.09 Gb (3.84%)
Number of gene models	57,529
Mean gene length	3,624 bp
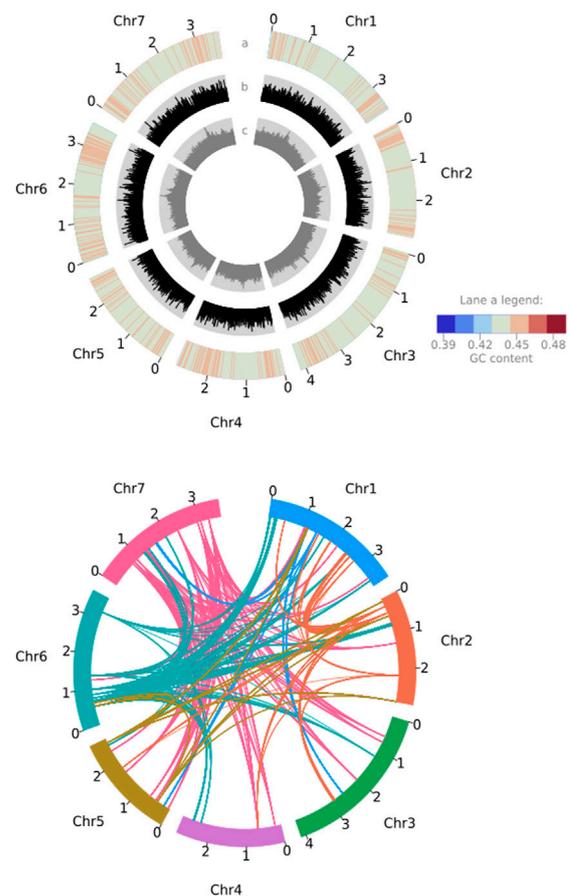
FIGURE 1
Left panel: Assembly statistics, and right panel: features of the *Lolium rigidum* genome (each tick is ×100 Mb). Top right panel—lane (a) GC content heatmap of mean GC content per 2.35 Mb window (ranging from 42% to 47%); lane (b) distribution of Copia long terminal repeat (LTR) retrotransposon family; lane (c) distribution of Gypsy LTR retrotransposon family. Bottom right panel: chord diagram shows the syntenic relationships within the top 5 orthogroups with the most paralogs in the genome (i.e. F-box domain-containing protein family, F-box protein interaction domain protein-related family, auxin-responsive protein family, and two orthogroups with unknown function: OS06G0725500 and PTHR32141), where the colours match the colours of the chromosome most of the paralogs per orthogroup are located.

Resource (UniProt) database. Protein sequences specific to *L. rigidum*, *L. multiflorum*, *A. thaliana*, *O. sativa*, and *Z. mays* were used because of the high quality of the gene annotation in these species. The predicted protein sequences of the six genomes were queried against the herbicide-resistance-related protein sequences and the best matching encoding gene was identified (E-value≤1 ×10⁻¹⁰). Significantly expanded and contracted gene families using all six species were identified using CAFE (v5; De Bie et al., 2006) with a *p*-value<0.01. The coding sequences of EPSPS gene paralogs within the annual ryegrass genome and homologs in the other five genomes were further analysed using Ka/Ks ratio across 15-bp non-overlapping sliding windows using KaKs_calculator (version 2; Wang et al., 2009).

The full workflow including the scripts and links to the datasets used for the comparative genomics analysis are found in the README.md file of https://github.com/jeffersonfparil/Lolium_rigidum_genome_assembly_and_annotation.

Results

Genome assembly

A total of 294.8 Gb from short- and long-read whole-genome sequencing was generated. Illumina sequencing accounted for 68.69% of this output, Oxford Nanopore sequencing using MinION and Promethion platforms accounted for 31.31%, and the Hi-C library generated 66 Gb of raw sequence data. We estimated the genome size to be 2.26 Gb based on k-mer frequency analysis (kmer = 16 bp found to generate the best

FIGURE 2
Lolium rigidum comparative genomics. **(A)** (left): phylogeny based on single-copy gene orthologs (100% bootstrap support across all clades); **(A)** (centre): number of significantly expanded and contracted gene families; **(A)** (right): distribution of genes with multiple orthologs, single-copy orthologs and unique orthologs, **(B)** Venn diagram of shared gene families between *L. rigidum*, *L. perenne*, *O. sativa*, and *Z. mays*; **(C)** distribution of the transversion rates in four-fold degenerate sites (4DTv) within orthogroups in *L. rigidum*, *L. perenne* and *Z. mays*.

model fit according to GenomeScope v1.0.0). The sequencing output information is summarised in Supplementary Table S1.

The final assembly reached a chromosomal level resolution with a total length 2.44 Gb (N_{50} = 361.79 Mb) over 7 scaffolds and 1,757 unplaced contigs where the 7 scaffolds constitute 96% of the assembly and correspond to the 7 expected chromosomes (Figure 1 left panel and Supplementary Figure S1). This assembly is 99.99% k-mer complete with 39.28 consensus quality score (QV) according to Polca (*MaSURCA* v4.0.7). The average mapping rates of the genomic and transcriptomic sequencing reads to the assembly are 97.48% and 90.92%, respectively. This assembly is 99.8% complete based on BUSCO analysis using the Viridiplantae gene set (i.e. Complete and single-copy [S]: 29.4%, Complete and duplicated [D]: 70.4%, Fragmented [F]: 0.2%, Missing [M]: 0.0%, Total [n]: 425) and 97.2% complete using the Poales gene set (S:32.4%, D: 64.8%, F:0.9%, M:1.9%, and n:4,896).

The *L. rigidum* genome assembly mainly consists of interspersed repeats (72.44%), transposable elements and

repetitive sequences accounting for 33.61% and 34.99% of the genome, respectively. Among the transposable elements, long terminal repeat (LTR) sequences were predominant (30.91% of the genome), mostly composed of Copia (24.51%) and Gyspy (6.40%) LTRs. The intactness of these LTR retrotransposons (LTR-RTs) which is proportional to the contiguity of the assembly was measured with the LTR assembly index (LAI). The average genome-wide raw LAI was 13.04 (standardised LAI = 7.41) which is above the minimum threshold 2 (0.1% intact LTR-RTs divided by 5% total LTR-RTs; Ou et al., 2018). The distribution of LAI across the genome assembly is shown in Supplementary Figure S2.

Gene family contraction and expansion

The comparison between the genomes of *L. rigidum*, *A. thaliana*, and the four grass crop species is summarised in

Figure 2. We confirmed that *L. rigidum* is very closely related to *L. perenne* and that the five grass species share a common ancestor around 65 MYA (Figure 2 panel a). The four grass species, *L. rigidum*, *L. perenne*, *O. sativa,* and *Z. mays* have more shared gene families than species-specific gene families, and *L. rigidum* shares more gene families with *L. perenne* than with *O. sativa* and *Z. mays* (Figure 2B). Gene families are, on average, ~14 times more expanded than contracted in *L. rigidum*. This is in striking opposition to the *L. perenne* genome where gene families are ~19 times more contracted than expanded (Figure 2 panel A central area). The *S. bicolor* genome also exhibited more contraction than expansion, but this is only slight compared with that of *L. perenne*. Also surprisingly, the distribution of genes with multiple orthologs, unique paralogs, and single-copy orthologs in *L. rigidum* is more similar to that of *Z. mays* than *L. perenne*. The distribution of 4DTv in Figure 2C shows two things: first, *L. rigidum* expectedly diverged more recently from *L. perenne* than from *Z. mays*; and second, *L. rigidum* experienced a recent WGD event while *L. perenne* experienced repeated and older WGD events as evidenced by the comparatively flatter 4DTv distribution.

GO term enrichment analysis of significantly expanded gene families in *L. rigidum* reveals herbicide resistance-related biological functions are significantly enriched (p-value<0.05; 368 significantly enriched GO terms from 15,192 gene names). The 15 most significantly enriched GO terms are presented in Supplementary Table S3. The following five GO terms showed consistent enrichment across two independent sets of 500 randomly sampled gene names from the full set of significantly expanded gene families: cellular response to stimulus (GO:0051716), response to stimulus (GO:0050896), regulation of metabolic process (GO:0051252), regulation of biosynthetic process (GO: 2001141), and organic substance metabolic process (GO: 0071704).

Herbicide resistance genes

There is statistically significant evidence for the expansion of the detoxification gene families tested, except for the monodehydroascorbate reductase (MDAR). Interestingly, there was no evidence for significant expansion of the EPSPS gene family. Instead, there is evidence for positive selection on one EPSPS gene, specifically at a site located between position 556 and 570 bp of the consensus CDS (Supplementary Figure S2). This position is conserved between *L. rigidum*, and *L. perenne*, but not with *A. thaliana* and *S. bicolor*. This putative target-site mutation, in addition to the expansion of detoxification genes, hints at the possible basis of glyphosate resistance in the reference genotype.

Discussion

Genome assembly

A combination of long- and short-read sequence assembly scaffolded with Hi-C proved to be sufficient to obtain a high-quality, chromosome-level 2.4-Gb reference genome of *L. rigidum*. The very high proportion of duplicated genes is common in plant genomes (Panchy et al., 2016) but the genome was relatively unambiguous with no evidence of polyploidisation according to the distribution of 4DTv. The simple diploid nature of annual ryegrass made it easier to resolve contig placements compared to polyploid species (Kyriakidou et al., 2018). The reference genome is mostly repetitive, consisting of long terminal repeat (LTR) families. Such expansion of LTRs in genomes has been linked to crop domestication (Qin et al., 2014; Huang et al., 2017) which annual ryegrass is, before becoming a noxious weed. Despite the high-quality of this genome assembly, additional sequencing efforts could improve it further, i.e. individual chromosome sequencing can be performed in the future.

Comparative genomics and herbicide resistance genes

The phylogeny inferred using our assembly and the reference genomes of five other plants matched the expected relationships and divergence times. *L. rigidum* indeed diverged later from and shares more gene families with the other grass species than with *A. thaliana*. Despite being closely related and having similar genome sizes, the patterns of gene expansion and contraction in *L. rigidum* and *L. perenne* are the opposite of each other. This suggests that *L. rigidum* underwent recent single-gene duplication events which is further supported by the distribution of 4DTv. These single-gene duplication events may have been mediated by tandem duplication (gene duplication resulting in multiple paralogous genes adjacent to each other) which is supported by the proximity of the expanded detoxification genes.

The expansion of herbicide resistance-related gene families is another interesting finding, with six out of the seven detoxification gene families tested showing significant expansion. This, in conjunction with evidence for positive selection in one of the EPSPS genes without expansion of the whole family, suggests that the mechanism of glyphosate resistance in this specific plant genotype is already multifactorial. Glyphosate resistance here is likely achieved through a combination of intensified neutralisation of reactive oxygen species (ROS) by the increased number of detoxification enzymes, and possibly by rendering the EPSPS enzyme resistant to disruption by glyphosate molecules. Given that we have stronger evidence for the former rather than the latter, we suggest that ROS scavenging by detoxification genes

may be more important than preventing the disruption of EPSPS activity in aromatic amino acid synthesis in the reference genotype sequenced.

Conclusion

We have assembled the first reference genome of the agriculturally important and noxious weed species, *Lolium rigidum*, at a high-quality and at chromosome level. This reference genome is pivotal in deciphering the genetic bases of new and emerging herbicide resistances, and the development of modern molecular tools for the management of this highly herbicide-resistant weed species. Upon analysing this reference genome representing only a single genotype, we were able to gather some evidence for the multifactorial bases of glyphosate resistance, i.e. target site resistance conferred by single point mutations within the gene, and non-target site resistance through the extensive duplication of detoxification genes. Hence, it is doubtless that this reference genome will be crucial for the genetic mapping of herbicide resistance, making use of more genotypes in more sophisticated experimental designs. It will also be instrumental in the development of new and novel genomically informed weed and herbicide resistance control strategies including genomic prediction models which will improve the speed and cost-effectiveness of herbicide resistance assays.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

AF-L, JP, and EB conceived the project and designed the experiments. EB performed the nucleic acid extractions and short-read sequencing. EB and JP performed the MinION sequencing. LC performed the PromethION and Hi-C sequencing. GP and RR assembled the genome, submitted the assembly to NCBI, requested for the annotation, and deposited the raw data to NCBI. JP performed the comparative genomics and analysed the assembly and annotations. JP drafted the manuscript. All authors edited and contributed to the article.

Funding

Acknowledgments

Conflict of interest

JP and AF-L have received funds from the Grain Research and Development Corporation.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2022.1012694/full#supplementary-material.

SUPPLEMENTARY FIGURE S1
Panel A: Lolium rigidum plant. Panel B: Hi-C interaction heatmap highlighting the 7 Lolium rigidum chromosomes. The unlabelled partition at the bottom-right corner of the heatmap refers to 1,757 unplaced contigs with some correlation with other regions coinciding with high repeat contents.

SUPPLEMENTARY FIGURE S2
Distribution of long-terminal repeat retrotransposon assembly index (LAI; measure of the intactness of these LTR retrotransposons (LTR-RTs) which is proportional to the contiguity of the assembly) across the Lolium rigidum genome assembly.

SUPPLEMENTARY FIGURE S3
Ka/Ks (ratio of the number of nonsynonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site per unit time) across non-overlapping 15-bp sliding windows comparing an enolpyruvylshikimate phosphate synthase (EPSPS) gene of L. rigidum (i.e. XP_047075216.1) to the EPSPS genes of O. sativa, S. bicolor, a paralog in L. rigidum, and two homologs in A. thaliana. Red asterisks show significant peaks at $p \leq 0.001$.

References

Byrne, S. L., Nagy, I., Pfeifer, M., Armstead, I., Swain, S., Studer, B., et al. (2015). A synteny-based draft genome sequence of the forage grass Lolium perenne. *Plant J.* 84 (4), 816–826. doi:10.1111/tpj.13037

Cabau, C., Escudié, F., Djari, A., Guiguen, Y., Bobe, J., and Klopp, C. (2017). Compacting and correcting trinity and Oases RNA-seq de Novo assemblies. *PeerJ* 5, e2988. doi:10.7717/peerj.2988

Creemers-Molenaar, J., and Beerepoot, L. J. (1992). "*In vitro* culture and micropropagation of ryegrass (Lolium spp.)," in *Biotechnology in agriculture and forestry*. Editor Y. P. S. Bajaj (Berlin, Heidelberg: Springer), 549–575. doi:10.1007/978-3-662-07770-2_33

De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). Cafe: A computational tool for the study of gene family evolution. *Bioinformatics* 22, 1269–1271. doi:10.1093/bioinformatics/btl097

Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., and Durand, N. C. (2017). De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* 356, 92–95. doi:10.1126/science.aal3327

Emms, D. M., and Kelly., S. (2019). OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20 (1), 238–251. doi:10.1186/s13059-019-1832-y

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* 117, 9451–9457. 17 (April 28, 2020): 9451–57. doi:10.1073/pnas.1921046117

Frei, Daniel, Veekman, Elisabeth, Grogg, Daniel, Stoffel-Studer, Ingrid, Aki, Morishima, Shimizu-Inatsugi, Rie, et al. (2021). Ultralong Oxford Nanopore reads enable the development of a reference-grade perennial ryegrass genome assembly. *Genome Biol. Evol.* 13, evab159–8. doi:10.1093/gbe/evab159

Goggin, D. E., Powles, S. B., and Steadman, K. J. (2012). Understanding Lolium rigidum seeds: The key to managing a problem weed? *Agronomy* 2 (3), 222–239. doi:10.3390/agronomy2030222

Gremme, G., Steinbiss, S., and Kurtz, S. (2013). GenomeTools: A comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans. Comput. Biol. Bioinform.* 10 (3), 645–656. doi:10.1109/TCBB.2013.68

Guan, D., McCarthy, S. A., Wood, J., Howe, K., Wang, Y., and Durbin, R. (2020). Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* 36 (9), 2896–2898. doi:10.1093/bioinformatics/btaa025

Haas, B. J., Papanicolaou, A., Moran, Y., Grabherr, M., Philip, D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8 (8), 1494–1512. doi:10.1038/nprot.2013.084

Heap, I. M. (2022). The international survey of herbicide resistant weeds. Available Www.Weedscience.Org (accesss June 3, 2022).

Huang, C., Nie, X., Shen, C., You, C., Wu, L., Zhao, W., et al. (2017). Population structure and genetic basis of the agronomic traits of upland cotton in China revealed by a genome-wide association study using high-density SNPs. *Plant Biotechnol. J.* 15 (11), 1374–1386. doi:10.1111/pbi.12722

Jean, G., and Paschoud, J. J. (1811). *Agrostologia helvetica: Definitionem descriptionemque graminum et plantarum eis affinium in helvetia sponte nascentium complectens*. Parisiis: Apud J.J. Paschoud. doi:10.5962/bhl.title.159420

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Arndt, V. H., and Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi:10.1038/nmeth.4285

Kloot, P. (1983). The genus Lolium in Australia. *Aust. J. Bot.* 31 (4), 421. doi:10.1071/BT9830421

Kolmogorov, M., Derek, M., Behsaz, B., Gurevich, A., Rayko, M., Shin, S. B., et al. (2020). metaFlye: scalable long-read metagenome assembly using repeat graphs. *Nat. Methods* 17, 1103–1110. doi:10.1038/s41592-020-00971-x

Koren, S., Walenz, B. P., Berlin, K., Jason, R., Miller, N. H., and Adam, B. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi:10.1101/gr.215087.116

Kyriakidou, M., Tai, H. H., Anglin, N. L., Ellis, D., and Strömvik, M. V. (2018). Current strategies of polyploid plant genome sequence assembly. *Front. Plant Sci.* 9, 1660. doi:10.3389/fpls.2018.01660

Langmead, Ben, and Salzberg, Steven L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9 (4), 357–359. March 4, 2012): 357–59. doi:10.1038/nmeth.1923

Lemerle, D., Verbeek, B., and Coombes, N. (1995). Losses in grain yield of winter crops from Lolium rigidum competition depend on crop species, cultivar and season. *Weed Res.* 35, 503–509. doi:10.1111/j.1365-3180.1995.tb01648.x

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-Mers. *Bioinformatics* 27 (6), 764–770. doi:10.1093/bioinformatics/btr011

Mccraw, J. M., and Spoor, W. (1983). Self-incompatibility in Lolium species. 1. Lolium rigidum gaud. And L. Multiflorum L. *Hered. (Edinb).* 50, 21–27. doi:10.1038/hdy.1983.3

Medd, R. W., Auld, B. A., Kemp, D. R., and Murison, D. (1985). The influence of wheat density and spatial arrangement on annual ryegrass, Lolium rigidum Gaudin, competition. *Aust. J. Agric. Res.* 36, 361–371. doi:10.1071/ar9850361, no

Mi, Huaiyu, Muruganujan, Anushya, Ebert, Dustin, Huang, Xiaosong, and Thomas, P. D. (2019). PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 47, D419–D426. D1 (January 8, 2019): D419–26. doi:10.1093/nar/gky1038

Minh, B Q, Schmidt, Heiko A., Chernomor, Olga, Schrempf, Dominik, Woodhams, Michael D, Haeseler, A, et al. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. 5 (May 1, 2020): 1530–34. doi:10.1093/molbev/msaa015

Mistry, Jaina, Finn, Robert D., Eddy, Sean R., Bateman, A., and Punta, M. (2013). "Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions." *Nucleic Acids Res.* 41 (12), e121. doi:10.1093/nar/gkt263

Monaghan, N. M. (1980). The biology and control of Lolium rigidum as a weed of wheat. *Weed Res.* 20, 117–121. doi:10.1111/j.1365-3180.1980.tb00054.x

Naylor, B. (1960). Species differentiation in the genus Lolium. *Heredity* 15, 219–233. doi:10.1038/hdy.1960.78

Ophel, Kathy M., Bird, Alan F., and Allen, Kerr. (1993). Association of bacteriophage particles with toxin production by clavibacter toxicus, the causal agent of annual ryegrass toxicity. *Phytopathology* 83 (6), 676–681. doi:10.1094/phyto-83-676

Ou, Shujun, Chen, Jinfeng, and Jiang, Ning (2018). Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res.* 46 (21), e126. doi:10.1093/nar/gky730

Panchy, Nicholas, Lehti-Shiu, Melissa, and Shiu, Shin-Han (2016). Evolution of gene duplication in plants. *Plant Physiol.* 171 (4), 2294–2316. doi:10.1104/pp.16.00523

Powles, S. B., Lorraine-Colwill, D. F., Dellow, J. J., and Preston, C. (1998). Evolved resistance to glyphosate in rigid ryegrass (Lolium rigidum) in Australia. *Weed Sci.* 46 (5), 604–607. doi:10.1017/s0043174500091165

Powles, S. B., and Matthews, J. M. (1992). "Multiple herbicide resistance in annual ryegrass (Lolium rigidum): A driving force for the adoption of integrated weed management," in *Resistance '91: Achievements and developments in combating pesticide resistance*. Editors Ian Denholm, Alan L. Devonshire, and Derek W. Hollomon (Dordrecht: Springer Netherlands), 75–87. doi:10.1007/978-94-011-2862-9_7

Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., et al. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into capsicum domestication and specialization. *Proc. Natl. Acad. Sci. U. S. A.* 111, 5135–5140. 14 (April 8, 2014): 5135–40. doi:10.1073/pnas.1400975111

Quast, C., Pruesse, E., Yilmaz, P., Jan, G., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* 41, D590–D596. D1 (January 1, 2013): D590–96. doi:10.1093/nar/gks1219

Ranwez, V., Harispe, S., Delsuc, F., Emmanuel, J., and Douzery, P. (2011). Macse: Multiple alignment of coding SEquences accounting for frameshifts and stop codons. *PLOS ONE* 6 (9), e22594. doi:10.1371/journal.pone.0022594

Reeves, T. G. (1976). Effect of annual ryegrass (Lolium rigidum gaud.) on yield of wheat. *Weed Res.* 16, 57–63. doi:10.1111/j.1365-3180.1976.tb00379.x

Rerkasem, K., Stern, W. R., and Goodchild, N. A. (1980). Associated growth of wheat and annual ryegrass. 1. Effect of varying total density and proportion in mixtures of wheat and annual ryegrass. *Aust. J. Agric. Res.* 31 (4), 649–658. doi:10.1071/ar9800649

Riley, I. T., and McKay, A. C. (1991). Inoculation of Lolium rigidum with clavibacter sp., the toxigenic bacteria associated with annual ryegrass toxicity. *J. Appl. Bacteriol.* 71 (4), 302–306. doi:10.1111/j.1365-2672.1991.tb03794.x

Schulz, M. H., Zerbino, D. R., Martin, V., and Oases, E. B. (2012). Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086–1092. doi:10.1093/bioinformatics/bts094

Simão, Felipe A., Waterhouse, Robert M., Ioannidis, Panagiotis, Kriventseva, Evgenia V., and Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and

annotation completeness with single-copy orthologs. *Bioinformatics* 31, 3210–3212. (October 1, 2015): 3210–12. doi:10.1093/bioinformatics/btv351

Song, Li, and Florea, Liliana (2015). Rcorrector: Efficient and accurate error correction for Illumina RNA-seq reads. *GigaScience* 4, 48. doi:10.1186/s13742-015-0089-y

Terrell, Edward E. (1966). Taxonomic implications of genetics in ryegrasses (Lolium). *Bot. Rev.* 32, 138–164. doi:10.1007/BF02858658

The UniProt Consortium (2019). UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* 47, D506–D515. D1 (January 8, 2019): D506–15. doi:10.1093/nar/gky1049

Vaser, R., Sović, I., Nagarajan, N., and Šikić, M. (2017). Fast and accurate de Novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746. doi:10.1101/gr.214270.116

Vurture, Gregory W., Sedlazeck, Fritz J., Nattestad, Maria, Underwood, Charles J., Han, Fang, Gurtowski, James, et al. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* 33, 2202–2204. doi:10.1093/bioinformatics/btx153

Wang, Da-Peng, Wan, Hao-Lei, Zhang, Song, and γ-Myn, Jun Yu. (2009). Gamma-MYN: A new algorithm for estimating Ka and ks with consideration of variable substitution rates. *Biol. Direct* 4, 20. doi:10.1186/1745-6150-4-20

Wick, Ryan R., Judd, Louise M., Gorrie, Claire L., and Kathryn, E. Y. R. (2017). Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb. Genom.* 3, e000132. doi:10.1099/mgen.0.000132

Wick, Ryan R., Judd, Louise M., and Holt, Kathryn E. (2019). Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* 20, 129. doi:10.1186/s13059-019-1727-y

Zhang, Xingtan, Zhang, Shengcheng, Zhao, Qian, Ray, Ming, and Tang, Haibao (2019). Assembly of allele-aware, chromosomal-scale Autopolyploid genomes based on Hi-C data. *Nat. Plants* 5 (8), 833–845. doi:10.1038/s41477-019-0487-8

Zimin, Aleksey V., Marçais, Guillaume, Puiu, Daniela, Roberts, Michael, Salzberg, Steven L., and Yorke, James A. (2013). The MaSuRCA genome assembler. *Bioinformatics* 29, 2669–2677. doi:10.1093/bioinformatics/btt476