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

[Jefferson Paril](https://orcid.org/0000-0002-5693-4123)1, [Gunjan Pandey](https://orcid.org/0000-0002-3514-4624)2, Emma B. Barnett1, [Rahul V. Rane](https://orcid.org/0000-0003-4616-6244)3, [Leon Court](https://orcid.org/%200000-0003-3673-9504)2, Thomas Walsh2, and [Alexandre Fournier-Level](https://orcid.org/0000-0002-6047-7164)1\*

1 School of BioSciences, University of Melbourne, Parkville, Australia

2 CSIRO Land and Water, Acton, Australia

3 CSIRO Health and Biosecurity, Parkville, Australia

\* Corresponding author: [alexandre.fournier@unimelb.edu.au](mailto:alexandre.fournier@unimelb.edu.au)

# Abstract

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 2.44Gb with N50=361.79Mb 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, 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 et al. 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).

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, USA) and sequenced using HiSeq X platform (Illumina, Inc., San Diego, USA) 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 and PromethION 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 et al, 2020) with the minimum overlap parameter set to 6,000. 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), and used in three rounds of contig polishing using *Racon* (v1.4.22; Vaser et al, 2017). 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. This assembly was assessed using *BUSCO* (v5; Simão et al, 2015) against the Viridiplantae and Poales lineages’ gene sets.

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, USA) 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 polished using *3d-dna* (Dudchenko et al, 2017). Genome size was estimated based on the kmer distribution of the Illumina sequences using Jellyfish 2.3.0 (Marcais and Kingsford, 2011) using kmer=19bp as parameter.

The assembly is available on the National Center for Biotechnology Information of the United States (NCBI) database under the accession number ([SAMN25144995, JAKKIG000000000](https://www.ncbi.nlm.nih.gov/nuccore/2206643664#sequence_JAKKIG000000000.1)). 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 <https://adaptive-evolution.biosciences.unimelb.edu.au/projects/ryegrass_genome_browser.html>.

## 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, UK). 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 et al, 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, 2012) using Bowtie2 (v2.3; Langmead & Salzberg, 2012). After filtering, ~197 million reads were used for *de novo* transcriptome assembly using the De novo RNA-Seq Assembly Pipeline (Cabau et al, 2017) including the rice protein sequences (release 51) 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. 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](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/001/735/GCF_000001735.4_TAIR10.1/)), rice (*Oryza sativa*; [IRGSP v1](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/433/935/GCF_001433935.1_IRGSP-1.0/)), sorghum (*Sorghum bicolor*; [NCBI v3](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/195/GCF_000003195.3_Sorghum_bicolor_NCBIv3/)), maize (*Zea mays*; [B73 Reference NAM v5](https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/902/167/145/GCF_902167145.1_Zm-B73-REFERENCE-NAM-5.0/)), and perennial ryegrass (*Lolium perenne*; [Kyuss v1](https://datacommons.cyverse.org/browse/iplant/home/shared/commons_repo/curated/Copetti_Kyuss_assembly_annotation_March_2021)) 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](https://github.com/davidemms/OrthoFinder/releases/tag/2.5.4); Emms and Keyll 2018) was used to cluster the CDS of all 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;](http://data.pantherdb.org/ftp/panther_library/current_release/) Mi et al, 2019). Significant gene family contraction and expansion in each of the six genomes were determined using CAFE ([v5](https://github.com/hahnlab/CAFE5/releases/download/v5.0/CAFE5-5.0.0.tar.gz); 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](http://geneontology.org/)” (The UniProt Consortium, 2019).

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 and to estimate the pairwise sequence divergence times between species by empirical Bayesian approach. These single-copy gene orthogroups were aligned using MACSE ([v2.06](https://bioweb.supagro.inra.fr/macse/releases/macse_v2.06.jar); Ranwez et al, 2011). The phylogenetic tree was generated using IQ-TREE ([v2.0.7](https://github.com/Cibiv/IQ-TREE/releases/download/v2.0.7/iqtree-2.0.7-Linux.tar.gz); Minh et al, 2020) and [TimeTree.org](http://timetree.org/) fossil record estimates of median divergence times between *A. thaliana* & rice estimated to 160 million years ago (MYA), sorghum & 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.

## Herbicide resistance genes

The use of a glyphosate-resistance 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 Resource (UniProt) database](https://www.uniprot.org/help/uniprotkb). 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](https://github.com/hahnlab/CAFE5/releases/download/v5.0/CAFE5-5.0.0.tar.gz); 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).

# Results

## Genome assembly

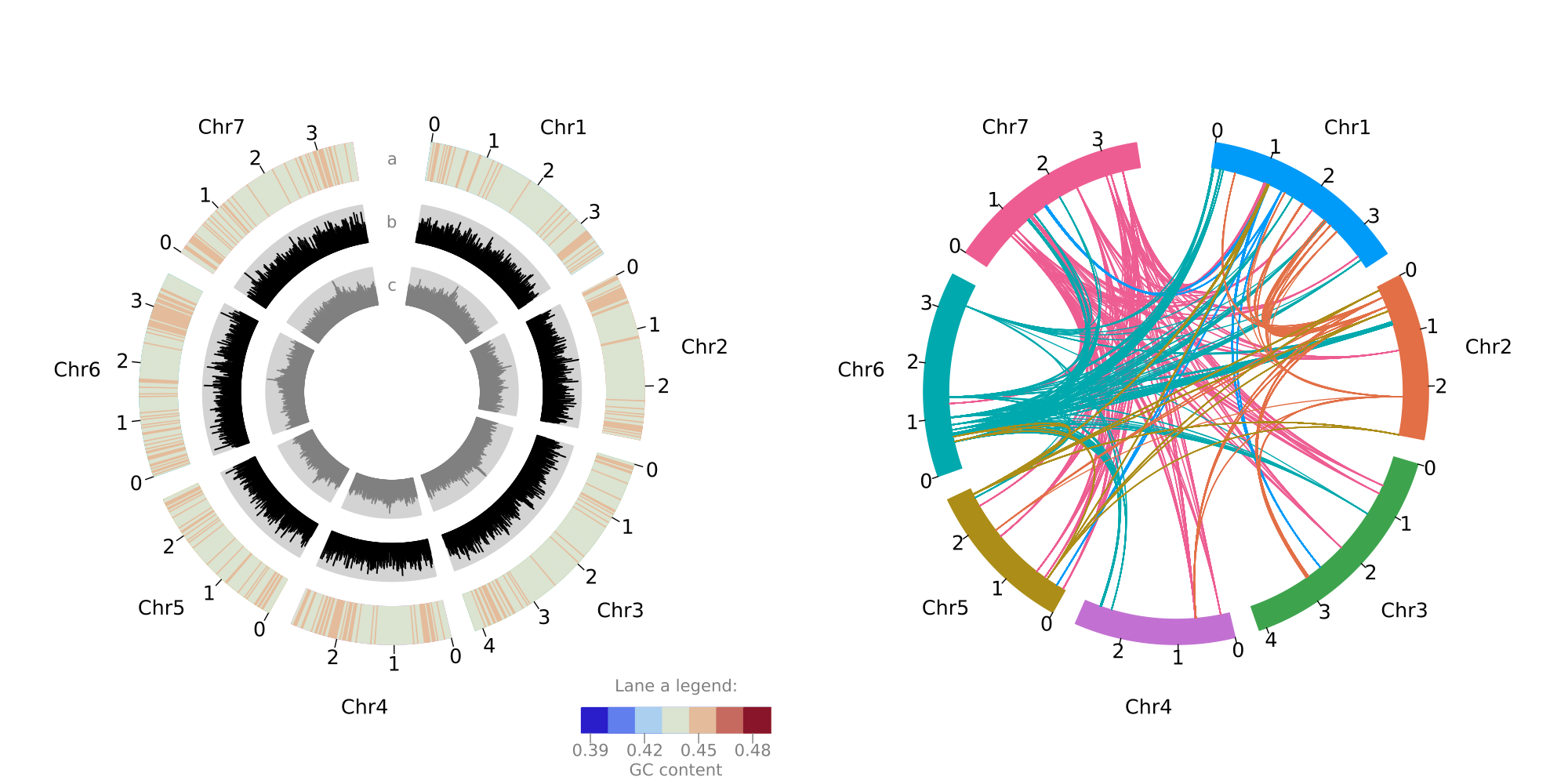
A total of 294.8 Gb from short- and long-read whole-genome sequencing was generated. Illumina sequencing accounted for 68.69% of this output (202.5 Gb across 675 million 150-bp paired-end reads). Oxford Nanopore sequencing using MinION and Promethion platforms accounted for 31.31% of the sequencing output (92.3 Gb across 9.04 million reads with N50=16.2kb). The Hi-C library generated 66 Gb of raw sequence data. We estimated the genome size to be 2.40 Gb based on k-mer frequency analysis (k=19bp).

The final assembly reached a chromosomal level resolution with a total length 2.44Gb (N50=361.79Mb) over 1,764 scaffolds where the 7 longest sequences constitute 96% of the assembly and correspond to the 7 expected chromosomes (Table 1 and Supplementary Figure 1). This assembly is 99.8% complete based on BUSCO analysis using the Viridiplantae gene set (i.e. Complete & single-copy [S]: 29.4%, Complete & 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%, n:4896).

**Table 1**. *Lolium rigidum* genome assembly ([NCBI RefSeq: GCF\_022539505.1](https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_022539505.1/)) statistics.

| **Estimated genome size** | 2.40 Gbp |
| --- | --- |
| **Total assembly size** | 2.44 Gbp |
| **Total scaffolds** | 1,764 |
| **N50** | 0.36 Gbp |
| **L50** | 4 |
| **Longest scaffold** | 0.41 Gbp |
| **Number of chromosomes** | 7 |
| **Total chromosome length** | 2.35 Gbp |
| **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 |

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.



**Figure 1**. Features of the *Lolium rigidum* genome (each tick is ×100Mb). **Left panel - lane a**: GC content heatmap of mean GC content per 2.35Mb 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. **Right panel**: chord diagram shows the syntenic relationships within the top 5 orthogroups with the most paralogs in the genome, where the colours match the colours of the chromosome most of the paralogs per orthogroup are located.

## 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* (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 2 panel b). 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). 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 2 panel C 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.



**Figure 2**. *Lolium rigidum* comparative genomics. **a** **(left)**: phylogeny based on single-copy gene orthologs; **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*.

GO term enrichment analysis of significantly expanded gene families in *L. rigidum* reveals herbicide resistance-related biological functions are significantly enriched. The 15 most significantly enriched GO terms are presented in Table 2.

**Table 2**. List of the top 15 significantly enriched gene ontology terms from the significantly expanded *Lolium rigidum* gene families.

| **Rank** | **Biological Process** | **GO ID** | **Fold enrichment (p-value)** |
| --- | --- | --- | --- |
| 1 | xylan biosynthetic process | 0045492 | 2.89 (1.20✕10-2) |
| 2 | xenobiotic detoxification by transmembrane export across the plasma membrane | 1990961 | 2.81 (2.08✕10-2) |
| 3 | xenobiotic export from cell | 0046618 | 2.81 (2.08✕10-2) |
| 4 | lignin metabolic process | 0009808 | 2.78 (4.28✕10-4) |
| 5 | xyloglucan metabolic process | 0010411 | 2.78 (8.30✕10-4) |
| 6 | cell wall polysaccharide biosynthetic process | 0070592 | 2.77 (5.36✕10-6) |
| 7 | hydrogen peroxide catabolic process | 0042744 | 2.75 (2.55✕10-13) |
| 8 | hydrogen peroxide metabolic process | 0042743 | 2.75 (2.55✕10-13) |
| 9 | cellulose metabolic process | 0030243 | 2.73 (1.25✕10-5) |
| 10 | xenobiotic transport | 0042908 | 2.71 (3.90✕10-3) |
| 11 | cellular component macromolecule biosynthetic process | 0070589 | 2.66 (1.51✕10-5) |
| 12 | cell wall macromolecule biosynthetic process | 0044038 | 2.66 (1.51✕10-5) |
| 13 | reactive oxygen species metabolic process | 0072593 | 2.65 (1.74✕10-13) |
| 14 | cellulose biosynthetic process | 0030244 | 2.64 (2.58✕10-2) |
| 15 | cellular response to ethylene stimulus | 0071369 | 2.62 (4.55✕10-4) |

## 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 2). 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. Individual chromosome sequencing can be performed, in addition to assembling the mitochondria and chloroplast genomes. This will consolidate the scaffolds excluded from the 7 chromosomes.

## 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

Raw Illumina, MinION, PromethION, and Hi-C reads are available under the NCBI Bioproject PRJNA799061. The reference genome assembly and the genome annotations can be found in NCBI’s database at [https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF\_022539505.1](https://www.ncbi.nlm.nih.gov/data-hub/genome/GCF_022539505.1/). A web-based genome browser is available at <https://adaptive-evolution.biosciences.unimelb.edu.au/projects/ryegrass_genome_browser.html>.

The comparative genomics, plotting, and statistical analyses pipelines are described in the github repository:

<https://github.com/jeffersonfparil/Lolium_rigidum_genome_assembly_and_annotation>.

# Author Contributions

AFL, JP, and EBB conceived the project and designed the experiments. EBB performed the nucleic acid extractions and short-read sequencing. EBB and JP performed the MinION sequencing. LC performed the PromethION and Hi-C sequencing. GP and RVR 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

Funding for this study was partially provided by the Department of Agriculture, Water and the Environment (DAWE) and the Grains Research and Development Corporation (GRDC) (grant ID: 4-FY9JQPE), Australian Research Data Commons (grant ID: DP727) as well as by the Commonwealth Scientific and Industrial Research Organisation (CSIRO) (Applied Genomics Initiative).

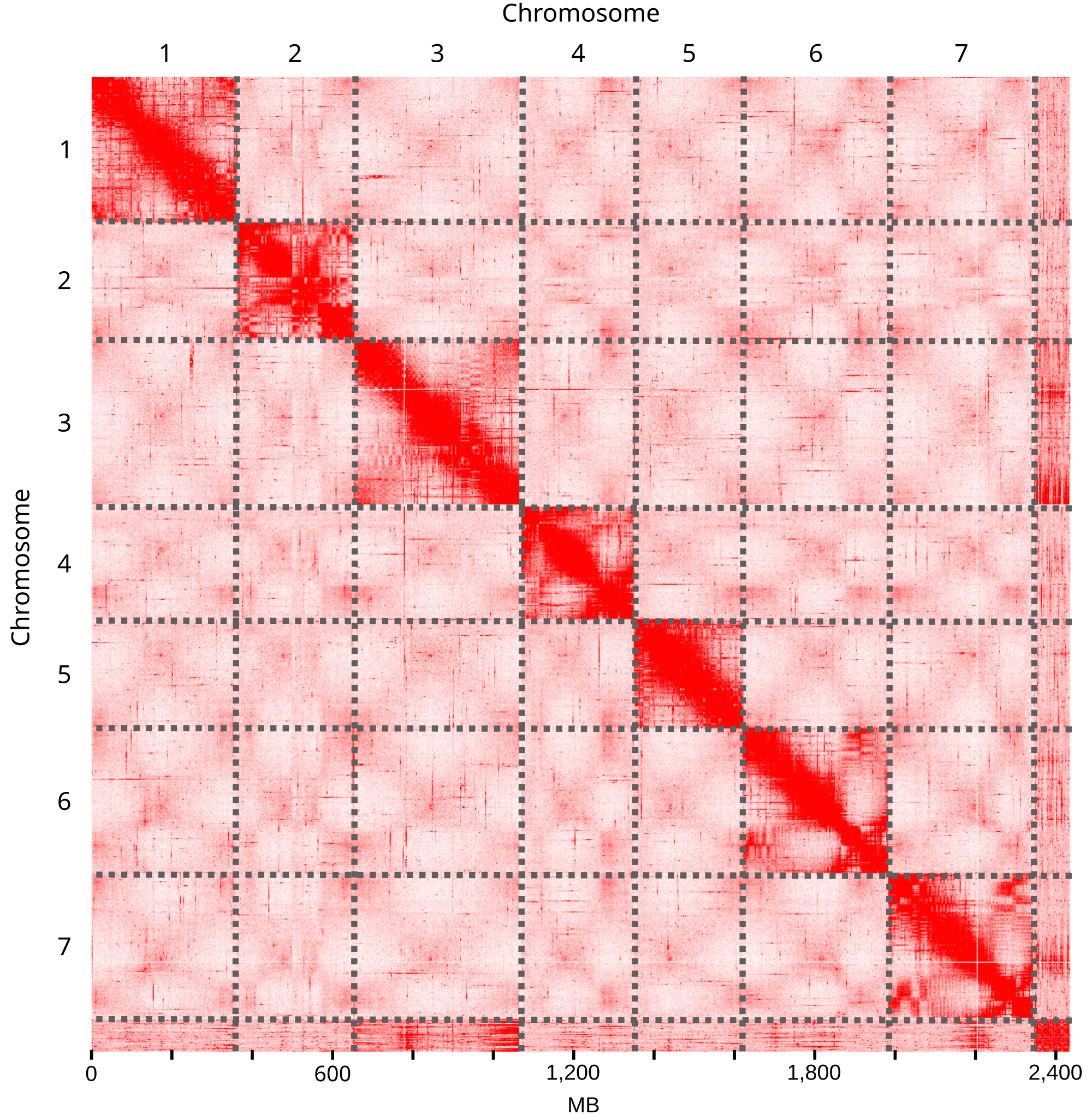
# Conflict of Interest Statement

The authors declare that they have no conflicting interest.

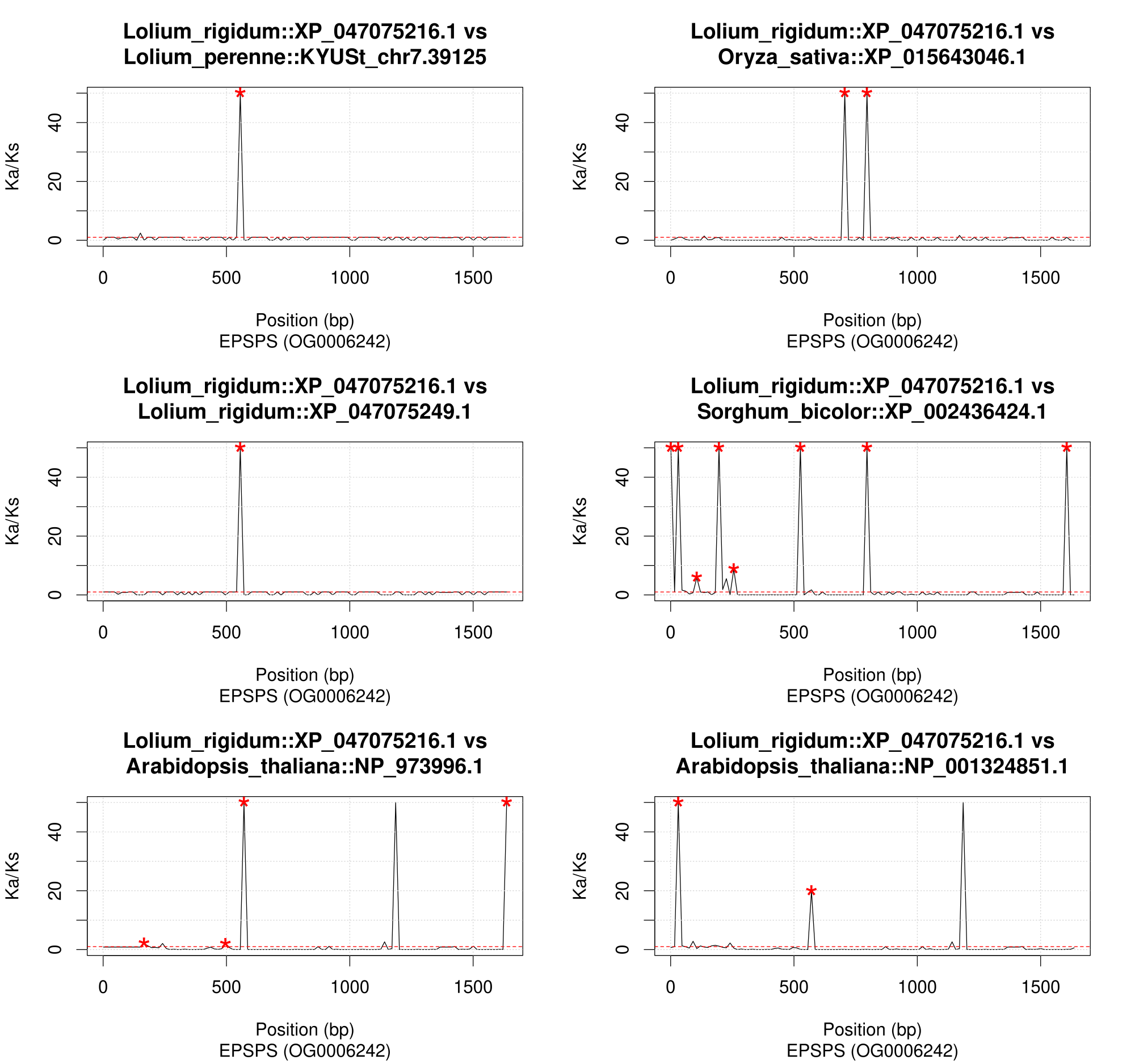
# Acknowledgments

We thank DAWE, GRDC, ARDC and CSIRO for the funding, and the University of Melbourne for hosting this study.

# Supplementary Figures



**Supplementary Figure 1**. Hi-C interaction heatmap highlighting the 7 *Lolium rigidum* chromosomes.



**Supplementary Figure 2**. 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≤0.001.

# 

# References

Byrne, Stephen L., Istvan Nagy, Matthias Pfeifer, Ian Armstead, Suresh Swain, Bruno Studer, Klaus Mayer, et al. “A Synteny-Based Draft Genome Sequence of the Forage Grass Lolium Perenne.” The Plant Journal 84, no. 4 (November 2015): 816–26. https://doi.org/10.1111/tpj.13037.

Cabau, Cédric, Frédéric Escudié, Anis Djari, Yann Guiguen, Julien Bobe, and Christophe Klopp. “Compacting and Correcting Trinity and Oases RNA-Seq de Novo Assemblies.” PeerJ 5 (February 16, 2017): e2988. https://doi.org/10.7717/peerj.2988.

Creemers-Molenaar, J., and L. J. Beerepoot. “In Vitro Culture and Micropropagation of Ryegrass (Lolium Spp.).” In High-Tech and Micropropagation III, edited by Y. P. S. Bajaj, 549–75. Biotechnology in Agriculture and Forestry. Berlin, Heidelberg: Springer, 1992. https://doi.org/10.1007/978-3-662-07770-2\_33.

De Bie, Tijl, Nello Cristianini, Jeffery P. Demuth, and Matthew W. Hahn. “CAFE: A Computational Tool for the Study of Gene Family Evolution.” Bioinformatics 22, no. 10 (May 15, 2006): 1269–71. https://doi.org/10.1093/bioinformatics/btl097.

“De Novo Assembly of the Aedes Aegypti Genome Using Hi-C Yields Chromosome-Length Scaffolds.” Accessed June 3, 2022. https://www.science.org/doi/10.1126/science.aal3327.

Dudchenko, Olga, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger, Neva C. Durand, Muhammad S. Shamim, et al. “De Novo Assembly of the Aedes Aegypti Genome Using Hi-C Yields Chromosome-Length Scaffolds.” Science 356, no. 6333 (April 7, 2017): 92–95. https://doi.org/10.1126/science.aal3327.

Emms, David M., and Steven Kelly. “OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics.” Genome Biology 20, no. 1 (November 14, 2019): 238. https://doi.org/10.1186/s13059-019-1832-y.

Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. “RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families.” Proceedings of the National Academy of Sciences 117, no. 17 (April 28, 2020): 9451–57. https://doi.org/10.1073/pnas.1921046117.

Frei, Daniel, Elisabeth Veekman, Daniel Grogg, Ingrid Stoffel-Studer, Aki Morishima, Rie Shimizu-Inatsugi, Steven Yates, et al. “Ultralong Oxford Nanopore Reads Enable the Development of a Reference-Grade Perennial Ryegrass Genome Assembly.” Genome Biology and Evolution 13, no. 8 (July 10, 2021): evab159. https://doi.org/10.1093/gbe/evab159.

Gill, Gurjeet. “Why Annual Ryegrass Is a Problem in Australian Agriculture.” Plant Protection Quarterly 11, no. Supplement 1 (1996): 193–95.

Goggin, Danica E., Stephen B. Powles, and Kathryn J. Steadman. “Understanding Lolium Rigidum Seeds: The Key to Managing a Problem Weed?” Agronomy 2, no. 3 (September 2012): 222–39. https://doi.org/10.3390/agronomy2030222.

Guan, Dengfeng, Shane A McCarthy, Jonathan Wood, Kerstin Howe, Yadong Wang, and Richard Durbin. “Identifying and Removing Haplotypic Duplication in Primary Genome Assemblies.” Bioinformatics 36, no. 9 (May 1, 2020): 2896–98. https://doi.org/10.1093/bioinformatics/btaa025.

Haas, Brian J., Alexie Papanicolaou, Moran Yassour, Manfred Grabherr, Philip D. Blood, Joshua Bowden, Matthew Brian Couger, et al. “De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis.” Nature Protocols 8, no. 8 (August 2013): 1494–1512. https://doi.org/10.1038/nprot.2013.084.

Heap, Ian M. “The International Survey of Herbicide Resistant Weeds. Online. Internet. Friday, June 3, 2022.” Available Www.Weedscience.Org, 2022.

Huang, Cong, Xinhui Nie, Chao Shen, Chunyuan You, Wu Li, Wenxia Zhao, Xianlong Zhang, and Zhongxu Lin. “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 Biotechnology Journal 15, no. 11 (2017): 1374–86. https://doi.org/10.1111/pbi.12722.

Kloot, Pm. “The Genus Lolium in Australia.” Australian Journal of Botany 31, no. 4 (1983): 421. https://doi.org/10.1071/BT9830421.

Kolmogorov, Mikhail, Derek M. Bickhart, Bahar Behsaz, Alexey Gurevich, Mikhail Rayko, Sung Bong Shin, Kristen Kuhn, et al. “MetaFlye: Scalable Long-Read Metagenome Assembly Using Repeat Graphs.” Nature Methods 17, no. 11 (November 2020): 1103–10. https://doi.org/10.1038/s41592-020-00971-x.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive k-Mer Weighting and Repeat Separation.” Genome Research 27, no. 5 (May 1, 2017): 722–36. https://doi.org/10.1101/gr.215087.116.

Krueger, Felix, Frankie James, Phil Ewels, Ebrahim Afyounian, and Benjamin Schuster-Boeckler. FelixKrueger/TrimGalore: V0.6.7 - DOI via Zenodo. Zenodo, 2021. https://doi.org/10.5281/zenodo.5127899.

Kyriakidou, Maria, Helen H. Tai, Noelle L. Anglin, David Ellis, and Martina V. Strömvik. “Current Strategies of Polyploid Plant Genome Sequence Assembly.” Frontiers in Plant Science 9 (2018). https://www.frontiersin.org/article/10.3389/fpls.2018.01660.

Langmead, Ben, and Steven L Salzberg. “Fast Gapped-Read Alignment with Bowtie 2.” Nature Methods 9, no. 4 (March 4, 2012): 357–59. https://doi.org/10.1038/nmeth.1923.

Lemerle, D., B. Verbeek, and N. Coombes. “Losses in Grain Yield of Winter Crops from Lolium Rigidum Competition Depend on Crop Species, Cultivar and Season.” Weed Research 35, no. 6 (1995): 503–9. https://doi.org/10.1111/j.1365-3180.1995.tb01648.x.

Mccraw, J M, W Spoor, and West Mains Road. “Self-Incompatibility in Lolium Species. 1. Lolium Rigidum Gaud. and L. Multiflorum L.” Heredity 50 (1983): 21–27.

Medd, RW, BA Auld, DR Kemp, and D Murison. “The Influence of Wheat Density and Spatial Arrangement on Annual Ryegrass, Lolium Rigidum Gaudin, Competition.,” 3, Australian Journal of Agricultural Research, no. 36 (1985): 361–71.

Mi, Huaiyu, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D Thomas. “PANTHER Version 14: More Genomes, a New PANTHER GO-Slim and Improvements in Enrichment Analysis Tools.” Nucleic Acids Research 47, no. D1 (January 8, 2019): D419–26. https://doi.org/10.1093/nar/gky1038.

Minh, Bui Quang, Heiko A Schmidt, Olga Chernomor, Dominik Schrempf, Michael D Woodhams, Arndt von Haeseler, and Robert Lanfear. “IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era.” Molecular Biology and Evolution 37, no. 5 (May 1, 2020): 1530–34. https://doi.org/10.1093/molbev/msaa015.

Mistry, Jaina, Robert D. Finn, Sean R. Eddy, Alex Bateman, and Marco Punta. “Challenges in Homology Search: HMMER3 and Convergent Evolution of Coiled-Coil Regions.” Nucleic Acids Research 41, no. 12 (July 1, 2013): e121. https://doi.org/10.1093/nar/gkt263.

Monaghan, N. M. “The Biology and Control of Lolium Rigidum as a Weed of Wheat.” Weed Research 20, no. 2 (1980): 117–21. https://doi.org/10.1111/j.1365-3180.1980.tb00054.x.

Naylor, B. “Species differentiation in the genus Lolium.” Heredity 15 (1960): 219–33.

Ophel, Kathy M., Alan F. Bird, and Allen Kerr. “Association of Bacteriophage Particles with Toxin Production by Clavibacter Toxicus, the Causal Agent of Annual Ryegrass Toxicity.” Phytopathology 83, no. 6 (1993): 676–81.

Panchy, Nicholas, Melissa Lehti-Shiu, and Shin-Han Shiu. “Evolution of Gene Duplication in Plants.” Plant Physiology 171, no. 4 (August 2016): 2294–2316. https://doi.org/10.1104/pp.16.00523.

Powles, Stephen B, Debrah F Lorraine-Colwill, James J Dellow, and Christopher Preston. “Evolved Resistance to Glyphosate in Rigid Ryegrass (Lolium Rigidum) in Australia.” Weed Science 46, no. 5 (1998): 604–7.

Powles, Stephen B., and John M. Matthews. “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, edited by Ian Denholm, Alan L. Devonshire, and Derek W. Hollomon, 75–87. Dordrecht: Springer Netherlands, 1992. https://doi.org/10.1007/978-94-011-2862-9\_7.

Qin, Cheng, Changshui Yu, Yaou Shen, Xiaodong Fang, Lang Chen, Jiumeng Min, Jiaowen Cheng, et al. “Whole-Genome Sequencing of Cultivated and Wild Peppers Provides Insights into Capsicum Domestication and Specialization.” Proceedings of the National Academy of Sciences 111, no. 14 (April 8, 2014): 5135–40. https://doi.org/10.1073/pnas.1400975111.

Quast, Christian, Elmar Pruesse, Pelin Yilmaz, Jan Gerken, Timmy Schweer, Pablo Yarza, Jörg Peplies, and Frank Oliver Glöckner. “The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools.” Nucleic Acids Research 41, no. D1 (January 1, 2013): D590–96. https://doi.org/10.1093/nar/gks1219.

Ranwez, Vincent, Sébastien Harispe, Frédéric Delsuc, and Emmanuel J. P. Douzery. “MACSE: Multiple Alignment of Coding SEquences Accounting for Frameshifts and Stop Codons.” PLOS ONE 6, no. 9 (September 16, 2011): e22594. https://doi.org/10.1371/journal.pone.0022594.

Reeves, T. G. “Effect of Annual Ryegrass (Lolium Rigidum Gaud.) on Yield of Wheat.” Weed Research 16, no. 1 (1976): 57–63. https://doi.org/10.1111/j.1365-3180.1976.tb00379.x.

Rerkasem, K., W. R. Stern, and N. A. Goodchild. “Associated Growth of Wheat and Annual Ryegrass. 1. Effect of Varying Total Density and Proportion in Mixtures of Wheat and Annual Ryegrass.” Australian Journal of Agricultural Research 31, no. 4 (1980): 649–58. https://doi.org/10.1071/ar9800649.

Riley, I. T., and A. C. McKay. “Inoculation of Lolium Rigidum with Clavibacter Sp., the Toxigenic Bacteria Associated with Annual Ryegrass Toxicity.” Journal of Applied Bacteriology 71, no. 4 (1991): 302–6. https://doi.org/10.1111/j.1365-2672.1991.tb03794.x.

Schulz, Marcel H., Daniel R. Zerbino, Martin Vingron, and Ewan Birney. “Oases: Robust de Novo RNA-Seq Assembly across the Dynamic Range of Expression Levels.” Bioinformatics 28, no. 8 (April 15, 2012): 1086–92. https://doi.org/10.1093/bioinformatics/bts094.

Simão, Felipe A., Robert M. Waterhouse, Panagiotis Ioannidis, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. “BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs.” Bioinformatics 31, no. 19 (October 1, 2015): 3210–12. https://doi.org/10.1093/bioinformatics/btv351.

Song, Li, and Liliana Florea. “Rcorrector: Efficient and Accurate Error Correction for Illumina RNA-Seq Reads.” GigaScience 4, no. 1 (October 19, 2015): 48. https://doi.org/10.1186/s13742-015-0089-y.

Terrell, Edward E. “Taxonomic Implications of Genetics in Ryegrasses (Lolium).” The Botanical Review 32, no. 2 (April 1966): 138–64. https://doi.org/10.1007/BF02858658.

The UniProt Consortium. “UniProt: A Worldwide Hub of Protein Knowledge.” Nucleic Acids Research 47, no. D1 (January 8, 2019): D506–15. https://doi.org/10.1093/nar/gky1049.

Vaser, Robert, Ivan Sović, Niranjan Nagarajan, and Mile Šikić. “Fast and Accurate de Novo Genome Assembly from Long Uncorrected Reads.” Genome Research 27, no. 5 (May 1, 2017): 737–46. https://doi.org/10.1101/gr.214270.116.

Wang, Da-Peng, Hao-Lei Wan, Song Zhang, and Jun Yu. “γ-MYN: A New Algorithm for Estimating Ka and Ks with Consideration of Variable Substitution Rates.” Biology Direct 4, no. 1 (June 16, 2009): 20. https://doi.org/10.1186/1745-6150-4-20.

Wick, Ryan R., Louise M. Judd, Claire L. Gorrie, and Kathryn E.YR 2017 Holt. “Completing Bacterial Genome Assemblies with Multiplex MinION Sequencing.” Microbial Genomics 3, no. 10 (n.d.): e000132. https://doi.org/10.1099/mgen.0.000132.

Wick, Ryan R., Louise M. Judd, and Kathryn E. Holt. “Performance of Neural Network Basecalling Tools for Oxford Nanopore Sequencing.” Genome Biology 20, no. 1 (June 24, 2019): 129. https://doi.org/10.1186/s13059-019-1727-y.

Zhang, Xingtan, Shengcheng Zhang, Qian Zhao, Ray Ming, and Haibao Tang. “Assembly of Allele-Aware, Chromosomal-Scale Autopolyploid Genomes Based on Hi-C Data.” Nature Plants 5, no. 8 (August 2019): 833–45. https://doi.org/10.1038/s41477-019-0487-8.

Zimin, Aleksey V., Guillaume Marçais, Daniela Puiu, Michael Roberts, Steven L. Salzberg, and James A. Yorke. “The MaSuRCA Genome Assembler.” Bioinformatics 29, no. 21 (November 1, 2013): 2669–77. https://doi.org/10.1093/bioinformatics/btt476.