

Benchmarking methods for genome annotation using nanopore direct RNA in a non-model crop plant

Jade M. Davis, Kristina K. Gagalova, Lilian M.V.P. Sanglard, Sabrina Cuellar, Mark R. Gibberd, Fatima Naim*, 

Centre for Crop and Disease Management, School of Molecular and Life Sciences, Curtin University, Bentley, WA 6102, Australia

*Corresponding author. Centre for Crop and Disease Management, 304 Brand Dr, Curtin University, Bentley, WA 6102, Australia.

E-mail: fatima.naim@curtin.edu.au.

Associate Editor: Aida Ouangraoua

Abstract

Motivation: High-quality genome annotations are essential for transcriptomic analyses investigating plant responses to environmental stress. While nanopore long-read direct RNA sequencing offers a powerful approach for improving genome annotations, studies benchmarking optimal tools for this process have primarily focused on animal models. In this study, we benchmarked five annotation tools: StringTie3, IsoQuant, Bambu, FLAIR, and FLAMES, using direct RNA data from barley infected with Net Form Net Blotch disease.

Results: We observed substantial variation across tools in isoform detection, structural completeness, splicing classification, and handling of 5' read truncation. Several tools successfully identified novel transcripts, with the two top-performing reference-guided approaches both detecting over 700 previously unannotated transcripts, including candidates with predicted roles in disease response. Our results highlight the importance of plant-specific benchmarking of bioinformatic tools and demonstrate the utility of direct RNA sequencing for improving genome annotations, supporting ongoing efforts to enhance reference resources for non-model plant species.

Availability and implementation: Benchmarking code is available at https://github.com/jadedavis5/benchmarking_paper. Datasets are described in the 'Data availability' section.

1 Introduction

High-quality genome annotations that capture transcriptomic diversity are fundamental to many bioinformatic analyses, including differential gene expression and alternative splicing (AS) studies. While reference annotations are relatively well-curated for the model plant *Arabidopsis thaliana* (Reiser *et al.* 2024), they remain limited for many non-model plant species (Anjanappa and Grissem 2021, Vuruputoor *et al.* 2023). Improving annotations is particularly challenging due to the inherent complexity of plant genomes, stemming from factors including polyploidy, large genome size, and high repeat content (Claros *et al.* 2012, Mehrotra and Goyal 2014, Kress *et al.* 2022, Ranawaka *et al.* 2023). These factors necessitate substantial sequencing and computational effort to produce annotations that accurately define genic features, such as intron/exon boundaries and regulatory regions. Comprehensive annotations are especially critical for investigating AS, a key mechanism driving transcript and protein diversity in multicellular organisms, which has been increasingly linked to evolutionary adaptation (Wright *et al.* 2022). Alternative splicing modulates gene function and enhances

phenotypic plasticity, shaping plant responses to both developmental cues and environmental stressors (Liu *et al.* 2022). It has been implicated in responses to abiotic stressors such as drought, salinity, and low temperature (Amin *et al.* 2016, Calixto *et al.* 2018, Chong *et al.* 2019, Zhang *et al.* 2020, Zhang *et al.* 2022, Alhabsi *et al.* 2025), as well as to biotic challenges including viral and fungal infections (Bedre *et al.* 2019, Dinesh-Kumar and Baker 2000, Huang *et al.* 2020, Wang *et al.* 2020).

Several computational strategies can be employed to improve plant genome annotations (Dominguez Del Angel *et al.* 2018), with transcriptome-based approaches playing a central role in structural annotation. By capturing actively expressed transcripts, RNA sequencing (RNA-seq) enables annotation refinement that reflects transcriptomic variation across specific conditions or tissue types. For example, in cassava (*Manihot esculenta*), annotation enrichment using RNA-seq under cold stress facilitated the identification of 323 additional differentially expressed genes compared to using the reference Ensembl annotation (Chenna *et al.* 2024). Other studies have enhanced genome annotations for various plants, including *Brassica oleracea* (wild cabbage), *Zea mays* (maize), *Papaver somniferum* (opium poppy), and

Amaranthus hypochondriacus (amaranth), using RNA-seq (Wang et al. 2016, Xu et al. 2022, Winkler et al. 2024, Yang et al. 2024).

Third-generation long-read RNA-seq technologies, such as those developed by Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio), offer significant advantages for RNA-seq-based annotations (Zmienko et al. 2021). These technologies can generate reads spanning thousands of bases (Amarasinghe et al. 2020), enabling full-length transcript capture. In contrast, conventional short-read platforms, which typically produce reads ≤ 300 bp (Gupta and Verma 2019), often lack the resolution necessary to reconstruct complete transcripts (Steijger et al. 2013). Among long-read methods, ONT's direct RNA (dRNA) sequencing, which selectively captures native polyadenylated mRNA without PCR amplification, is particularly promising for the detection of full-length and low-abundance transcripts (Sun et al. 2025). This approach has uncovered novel isoforms in rice, bamboo, and *Arabidopsis thaliana* (Zhang et al. 2020, Li et al. 2021, Li et al. 2023), including $\sim 38\,500$ previously unannotated transcripts in the latter, proving a powerful tool for enriching plant genome annotations.

A growing number of bioinformatic tools support dRNA-directed annotation, including StringTie3, IsoQuant, Bambu, FLAIR, and FLAMES. Each employs distinct algorithms for transcriptome assembly, from splice or intron graphs to read clustering and correction workflows (Table 1) (Kovaka et al. 2019, Tian et al. 2021, Chen et al. 2023, Prjibelski et al. 2023, Tang et al. 2024, Shinder et al. 2025). The diversity of approaches creates uncertainty about which tools are best suited to specific datasets or annotation goals. Accordingly, several recent benchmarking efforts have evaluated long-read annotation tools. Most notably, the Long-read RNA-Seq Genome Annotation Assessment Project (LRGASP) (Pardo-Palacios et al. 2024b) study compared reference-guided and *de novo* annotation strategies across real and simulated datasets from human, mouse, and manatee. Other studies have also benchmarked tools using simulated, synthetic, and/or real datasets from humans and model organisms, including mouse, *Drosophila*, and *Caenorhabditis elegans* (Sagniez et al. 2024, Su et al. 2024). Tool performance varied across sequencing technologies, reference quality, and annotation strategies,

highlighting a potential need for plant-specific benchmarking (Zhu et al. 2024).

Barley is an important global cereal crop (Verstegen et al. 2014) with a large diploid genome of 5.1 Gbp, of which $>80\%$ is composed of repetitive elements (Sato 2020). Despite its genomic complexity, barley is often considered a model species for other *Triticeae* cereals, such as wheat and rye. This is partly due to its well-curated genomic resources and the availability of diverse germplasms (Jiang et al. 2025). Therefore, barley serves as an ideal system for benchmarking dRNA annotation tools. While recent efforts have vastly improved barley annotations (Jayakodi et al. 2020, Mascher et al. 2021, Milne et al. 2021, Coulter et al. 2022, Jayakodi et al. 2024), such as PanBaRT20, which captures 79 600 genes and 582 000 transcripts across multiple tissues and cultivars (Guo et al. 2025), existing resources mainly reflect developmental and abiotic stress conditions. Host responses to biotic stressors, including Net Form Net Blotch (NFB) disease of barley, caused by *Pyrenophora teres* f. *teres*, remain underrepresented. Although NFB-induced gene expression has been studied (Moolhuijzen et al. 2021), standard barley genome reference annotations were used. Alternative splicing under disease pressure is also poorly characterized, primarily due to inadequate annotations. Accordingly, long-read dRNA sequencing could be applied to enrich annotations, although plant-specific benchmarking of annotation tools is still needed. This study addresses the gap by benchmarking reference-guided and *de novo* dRNA annotation tools in barley under NFB disease pressure and assessing the biological relevance of novel transcripts identified.

2 Methods

2.1 Plant and fungal propagation and inoculation of leaves

Methodology for fungal inoculation and leaf tissue harvest was adapted from Naim et al. (2021) as follows. Barley (cv. RGT Planet) was grown for 3 weeks under natural light in glasshouse conditions (Curtin University). A fungal spore solution was

Table 1 Comparison of long-read transcriptome annotation tools.

Tool	Approach	Key features	Reference annotation
StringTie3 (Kovaka et al. 2019, Shinder et al. 2025)	Splice graph	Builds exon-splice graphs; selects paths with the highest read support	Optional
IsoQuant (Prjibelski et al. 2023)	Intron graph	Simplifies intron-based graphs; finds best-supported transcript paths	Optional
Bambu (Chen et al. 2023)	Read clustering + novel discovery rate (NDR)	Clusters reads; uses NDR to filter novel transcripts, pretrained model fallback	Optional
FLAIR (Tang et al. 2024)	Align–Correct–Collapse	Corrects junctions with short-read or reference data; iterative filtering	Recommended for correction
FLAMES (Tian et al. 2021)	Read clustering + collapsing	Collapses similar/truncated isoforms; optimized for single-cell, but supports bulk	Required

prepared using frozen plugs from *Pyrenophora teres* f. *teres* [NB29 isolate (Syme *et al.* 2018)]. A 10 μ l spore solution droplet (2000 spores/ml) was placed 10 and 15 cm from the tip of the second mature leaf. The detailed protocol for spot inoculation is available online (Sanglard *et al.* 2025).

2.2 Sample collection, RNA extraction, and sequencing

RNA extraction was performed as described in Moolhuijzen *et al.* (2023). Three biological replicates for control and NFNB-infected leaf samples were harvested 3 days post-inoculation (dpi). Infected leaves were harvested using a 2-mm biopsy punch, with three punches collected at observed disease lesions (above, within, and below the lesion) in a line parallel to the midrib. Control samples were harvested by targeting a similar leaf area. Each replicate consisted of three punches from two lesions per leaf, across a total of four leaves, for a total of 24 punches. Total RNA was extracted using the Plant/Fungi Total RNA Purification Kit (Norgen Biotek, Canada) and eluted in 40 μ l. Samples were quantified using a Nanodrop spectrophotometer (NanoDrop Technologies, USA) and a Qubit RNA High Sensitivity (HS) kit run on a QuantiFluor RNA system E3310 (Qubit, UK). RNA quality was assessed by Genomics WA (Perth, Australia) using the TapeStation 4200 platform (Agilent, USA).

RNA samples were sequenced using both Illumina short-read and Oxford Nanopore Technologies (ONT) long-read direct RNA (dRNA) sequencing. Illumina library preparation and sequencing were performed at Genomics WA (Perth, Australia) using the SureSelect XT HS2 mRNA Library Preparation Kit (Agilent, USA). Paired-end (2 $\times$ 150 bp) sequencing was performed on the NovaSeq 6000 platform (Illumina, USA) and generated ~32–37 million raw read pairs per replicate (Table 1, available as [supplementary data](#) at *Bioinformatics Advances* online). Direct RNA sequencing was performed in-house (Centre for Crop and Disease Management, Australia), by pooling equal amounts of RNA from biological replicates for each treatment to create one representative sample. An aliquot of this sample containing 1000 ng of RNA was used for library preparation with the ONT Direct RNA Sequencing Kit (SQK-RNA004, ONT, UK). To increase transcript diversity in the control barley dRNA library, three additional control leaf RNA samples were included; these were generated using the same preparation protocol. The only modification to the manufacturer's protocol was reducing the RNA Calibration Strand (RCS) solution volume to 0.1 μ l. Libraries were loaded onto two FLO-PRO004RA flow cells and sequenced using the ONT PromethION 2 Solo platform, which resulted in 5.5 and 8.4 million raw reads for the control and NFNB-infected samples with estimated N50 values of 1.21 and 0.98 kb, respectively (Table 2, available as [supplementary data](#) at *Bioinformatics Advances* online).

2.3 Direct RNA read pre-processing and mapping

Direct RNA reads were used to generate annotations for tool comparison (Fig. 1). The majority of bioinformatic tools were executed using Singularity (Kurtzer *et al.* 2017) containers

retrieved from the Galaxy depot (Abueg *et al.* 2024) (<https://depot.galaxyproject.org/singularity/>). Unless otherwise specified, tools were executed with default parameters. Raw pod5 files were basecalled using Dorado v0.7.3 (Oxford Nanopore Technologies 2024) with the `rna004_130bps_sup@v5.0.0` model. Raw read length distribution was analysed using SeqKit fx2tab v2.9.0 (Shen *et al.* 2016) with ‘-name-length’ parameters (Fig. 1, available as [supplementary data](#) at *Bioinformatics Advances* online).

RNA Control Strand (RCS) sequences were removed by aligning reads to the ENO2 reference [Ensembl transcript ID YHR174W_mRNA, accessed August 2024 (Harrison *et al.* 2024)] using Minimap2 v2.28-r120955 (Li 2018) with recommended ONT dRNA settings (‘-ax splice -uf -k14’). Unmapped reads were extracted with Samtools fastq v1.21 (Li *et al.* 2009) using the ‘-f 4’ flag setting and filtered using Chopper v0.9.0 (De Coster and Rademakers 2023) applying minimum thresholds of Phred 10 and 100 bp (‘-q 10 -l 100’). Read quality was assessed using FastQC v0.12.1 (Andrews 2010) and summarized using MultiQC v1.25.2 (Ewels *et al.* 2016). Cleaned reads were aligned to the RGT Planet reference genome (Jayakodi *et al.* 2024) using Minimap2 v2.28-r120955 with recommended ONT dRNA settings, and the resulting alignments were sorted and indexed using Samtools v1.21. A summary of raw and processed read counts, along with alignment rates, is provided in Table 2, available as [supplementary data](#) at *Bioinformatics Advances* online. Gene body coverage was calculated relative to the RGT Planet reference annotation (Jayakodi *et al.* 2024), converted to BED format using AGAT `agat_convert_sp_gff2bed.pl` v1.4.1 (Dainat 2025), with RSeQC v5.0.4 `geneBody_coverage.py` (Wang *et al.* 2012) (Fig. 2, available as [supplementary data](#) at *Bioinformatics Advances* online).

2.4 Transcript discovery and annotation

Five tools were evaluated for reference-guided annotation: StringTie3, FLAIR, IsoQuant, Bambu and FLAMES (Kovaka *et al.* 2019, Tian *et al.* 2021, Chen *et al.* 2023, Prjibelski *et al.* 2023, Tang *et al.* 2024, Shinder *et al.* 2025). All analyses used the RGT Planet annotation in the respective input parameter. StringTie3 v3.0.0 was run with the long-read (‘-L’) parameter and the output for both samples were combined with the reference annotation using ‘-merge’. For FLAIR v2.0.0, BAM files were converted to BED format using FLAIR `bam2Bed12` and refined using FLAIR `correct` with the reference annotation. Refined BED files were concatenated before processing with FLAIR `collapse`, applying the ‘-annotation_reliant’ flag. Bambu v3.8.3 was run with ‘quant = FALSE’ to omit quantification. IsoQuant v3.6.3 was run with ‘-data_type nanopore’, and the SAMPLE_ID.extended_annotation.gtf output was selected for further analysis. The 925331f GitHub commit of FLAMES (<https://github.com/mritchielab/FLAMES>; accessed February 2025) was run using the ‘bulk_long_pipeline’ workflow. Configuration followed the example documentation (https://mritchielab.github.io/FLAMES/reference/bulk_long_pipeline.html, accessed February 2025), except barcode demultiplexing and quantification were omitted and ‘no_flank’ was set to false. For *de novo* annotation, Bambu, StringTie3, and IsoQuant were executed as above without reference-guided parameters. Bambu was tested with novel discovery rate (NDR) thresholds of 0.1, 0.5, 0.75, and 1. NDRs

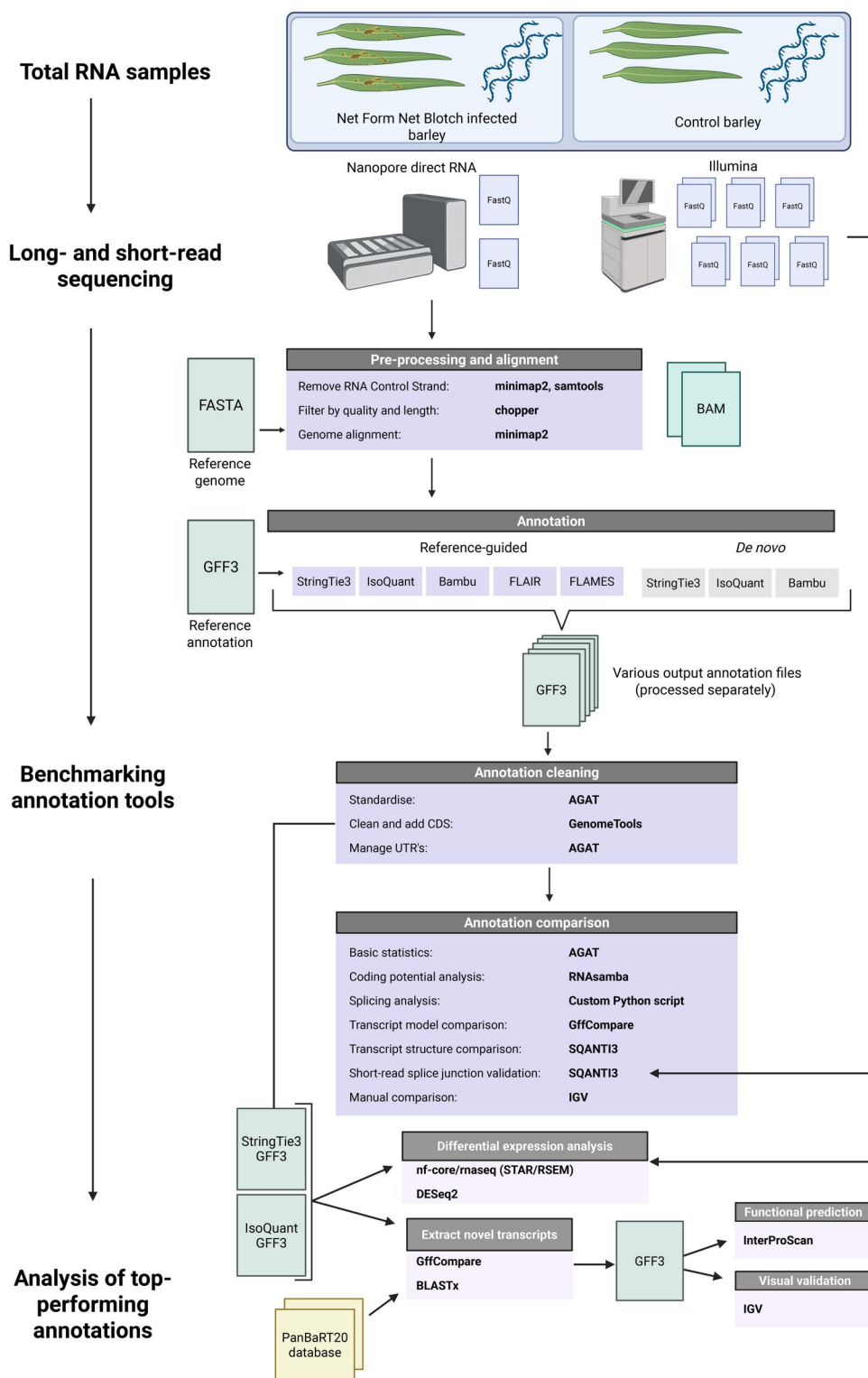


Figure 1 Flowchart summarizing the study workflow. RNA samples extracted from control and Net Form Net Blotch (NFNB) infected barley leaves underwent Oxford Nanopore direct RNA sequencing and Illumina short-read sequencing. Direct RNA reads were pre-processed and aligned to the reference genome before annotation using both reference-guided and *de novo* tools. Resulting annotations were standardized and benchmarked across multiple metrics, and the annotations generated by the two top-performing reference-guided tools were selected for further analysis. Figure created in BioRender <https://BioRender.com/hjwuyoj>.

between 0.1 and 0.75 produced errors to increase NDR, meaning a value of 1 was selected.

Annotations were standardized using AGAT `agat_convert_sp_gxf2gxf.pl` v1.4.1 (Dainat 2025) and cleaned with GenomeTools GFF3 v1.6.5 (Gremme et al. 2013) with ‘-sort -tidy’ options. Coding sequences (CDS) were annotated with GenomeTools CDS v1.6.5 with ‘-startcodon -matchdescstart’ flags, and untranslated regions (UTRs) were managed using AGAT `agat_sp_manage_UTRs.pl` v1.4.1.

2.5 Annotation quality analysis and benchmarking

Gene and transcript counts were obtained using AGAT `agat_sp_statistics.pl` v1.4.1. Coding potential was assessed by extracting transcripts with `gffread` v0.12.7 (Pertea and Pertea 2020) and classifying them using RNAsamba `classify` v0.2.5 (Camargo et al. 2020) with the ‘full_length_weights.hdf5’ model. Transcript splicing was classified using a custom Python script (`calculate_ss_stats.py` in the provided Nextflow-based (Di Tommaso et al. 2017) pipeline). Transcript-level precision and sensitivity were calculated against RGT Planet annotations using GffCompare v0.12.6 (Pertea and Pertea 2020), with ‘-R’ to skip reference annotations not overlapped by query models. Precision was calculated as the proportion of transcripts in the query annotation that matched reference transcripts (true positives), relative to the total number of query transcripts. Sensitivity was calculated as the proportion of true positives among all transcripts in the reference annotation. Novel splice junctions and transcript structures were analysed with SQANTI3 v5.3.6 `sqanti3_qc.py` (Pardo-Palacios et al. 2024a) using ‘-short_reads’ with raw Illumina reads. Following Dong et al. (2023), junctions with >10 uniquely mapping reads were considered validated. GffCompare v0.12.6 was run to assess transcript discovery across tools and the ‘tracking’ output file was used to identify isoforms with shared intron chains. Visualizations were generated in RStudio 2024.12.0+467 (RStudio Team 2022) using ggplot2 v3.5.1 (Wickham 2016), ComplexUpset v1.3.5 (Krassowski et al. 2022) and RColorBrewer v1.1–3 palettes (Neuwirth 2022).

2.6 Visualization of transcript models

To assess annotation quality in *de novo* approaches, 10 transcripts associated with disease resistance expressed in the NFNB-infected direct RNA data were visualized. Transcripts were selected by querying ‘disease resistance’ in the functional annotation search of the PanBARLEX portal (<https://panbarlex.ipk-gatersleben.de/>, accessed February 2025) and results for RGT Planet were extracted. Expression was assessed using IsoQuant v3.6.3 with ‘-no_model_construction -data_type nanopore’ parameters and the RGT Planet annotation. Ten transcripts with >20 mapped reads were randomly selected for visualization.

To assess annotation quality in reference-guided approaches, five of the disease resistance-associated transcripts and four novel transcripts were visualized. Novel transcripts were identified by querying annotations against RGT Planet using GffCompare v0.12.6. For each tool, a random transcript labelled ‘unknown’ (intergenic to all reference annotations) was

selected, except for FLAMES, which reported none. Selected annotations from both *de novo* and reference-guided approaches were visualized alongside direct RNA alignments using IGV v2.17.3 (Robinson et al. 2011).

2.7 Functional characterization, visualization, and expression analysis of novel transcripts

Novel transcripts identified by IsoQuant and StringTie3 reference-guided approaches were functionally characterized. For each annotation, transcripts labelled ‘unknown’ by GffCompare v0.12.6 against RGT Planet were extracted using AGAT `agat_sp_extract_sequences.pl` v1.4.1, run with ‘-merge’, and queried against the PanBaRT20 protein database (Guo et al. 2025) (https://ics.hutton.ac.uk/panbart20/downloads/PanBaRT20_transuite_transfeat_pep_renamed.fasta.gz, accessed March 2025) using BLASTx v2.12.0 (Camacho et al. 2009). Transcripts with >95% percentage identity and query coverage were classified as known. A custom Bash script was used to extract annotations for novel transcripts, followed by conversion into a protein FASTA using AGAT `agat_sp_extract_sequences.pl` v1.4.1 with ‘-t cds -p -cfs -cis’ parameters. Functional domains were predicted with InterProScan v5.76–107.0 (Jones et al. 2014, Blum et al. 2025) with the v5.76–107.0 data release using ‘-disable-precalc -goterms’ parameters. Gene Ontology (GO) terms were extracted, counted and then summarized using REVIGO v1.8.1 (Supek et al. 2011) (<http://revigo.irb.hr/>, accessed October 2025) with the ‘higher value is better’ setting, a small list and *Triticum aestivum* set as the closest reference. The biological processes treemap was exported and then refined using ggplot2 v3.5.1 and treemap v2.4–4 in RStudio 2024.12.0+467. Novel transcripts with predicted functions were visualized with direct RNA alignments in IGV v2.17.3 to manually assess transcript structure. Transcripts were subjectively classified as true if they were supported by ≥3 splice-aligned reads and false if not. Transcripts substantially shorter than alignments were also classified as false (Fig. 3, available as supplementary data at Bioinformatics Advances online).

Expression of novel transcripts was assessed against the IsoQuant and StringTie3 annotations using short-read Illumina data. Raw reads were processed with nf-core/rnaseq v3.18.0 (Patel et al. 2024) using STAR/RSEM (Li and Dewey 2011, Dobin et al. 2013), CSI indexing and the StringTie3 or IsoQuant reference-guided annotation (‘-aligner star_rsem -bam_csi_index’). Transcript counts from the ‘rsem.merged.transcript_counts.-tsv’ output files were analysed in DESeq2 v1.46.0 (Love et al. 2014) using a Wald test. Transcripts with log2-fold change ≥2 and adjusted *P* value ≤.05 were considered significantly differentially expressed. Finally, dplyr v1.1.4 (Wickham et al. 2023) was used to identify novel transcripts that were significantly differentially expressed. The nf-core/rnaseq v3.18.0 pipeline was also run with the original RGT Planet reference annotation to perform a principal component analysis of the short-read RNA-seq data (Fig. 4, available as supplementary data at Bioinformatics Advances online).

3 Results

3.1 Annotation tools performed variably

In this study, long-read dRNA sequencing data from barley (cv. RGT Planet) infected with NFNB were used to compare the performance of various annotation tools. Initially, dRNA read quality, length distribution, and gene body coverage were assessed (Table 2 and Figs. 1 and 2, available as [supplementary data](#) at *Bioinformatics Advances* online), confirming that the sequencing data were consistent with previous reports (Soneson *et al.* 2019, Sun *et al.* 2020, Gleeson *et al.* 2022). Several long-read annotation tools were then trialled, with outputs showing substantial variation between tools. Computationally, StringTie3 ran fastest and consumed the least memory among both reference-guided and *de novo* approaches (Table 3, available as [supplementary data](#) at *Bioinformatics Advances* online).

Comparison of annotation outputs revealed substantial variation between tools. Among reference-guided methods, StringTie3 produced the highest number of annotations, identifying 53 929 genes and 58 119 transcripts, followed closely by IsoQuant (53 444 genes and 57 825 transcripts) and Bambu (52 081 genes and 54 695 transcripts), all of which exceeded the RGT Planet reference annotations. FLAIR and FLAMES produced markedly fewer annotations, reporting 28 073/28 979 and 11 694/20 041 genes/transcripts, respectively. For *de novo* approaches, Bambu annotated the highest number of genes (16 463), followed closely by StringTie3 (16 027) and IsoQuant (13 661). Bambu reported a substantial number of transcripts (50 587), which were 3.3× and 2.8× higher than IsoQuant (15 105) and StringTie3 (18 046), respectively (Fig. 2A). FLAMES and StringTie3 had the longest median transcript length of 3004 bp and 3734 bp for reference-guided and *de novo* assemblies, respectively (Fig. 5,

available as [supplementary data](#) at *Bioinformatics Advances* online).

Splicing profiles differed between tools, especially for mono-exonic transcripts, which are rare in eukaryotic genomes and often indicative of pseudogenes (Vuruputoor *et al.* 2023). Reference-guided annotations from StringTie3, IsoQuant, and Bambu showed splicing profiles similar to reference annotations. FLAMES produced a lower percentage of mono-exonic transcripts and a higher percentage of canonically spliced transcripts, while FLAIR annotated a higher proportion of mono-exonic transcripts. Among *de novo* methods, IsoQuant and Bambu produced no mono-exonic transcripts, whereas StringTie3 contained 6% (Fig. 2B).

Predicted coding potential for reference-guided methods was consistent with the reference annotation for StringTie3, IsoQuant, Bambu, and FLAMES (<3% non-coding), while 20% of FLAIR transcripts were non-coding. In *de novo* mode, IsoQuant and StringTie3 maintained <3% non-coding transcripts, while Bambu reached 7% (Fig. 2C).

Transcript-level precision and sensitivity also varied. In reference-guided methods, Bambu, StringTie3, IsoQuant, and FLAIR achieved high sensitivity (>98%), while FLAMES ranked lowest (77%). More variation was observed in precision, with Bambu ranking highest (95%), followed by IsoQuant (90%) and StringTie3 (89%), with FLAIR (70%) and FLAMES (46%) much lower. In *de novo* methods, IsoQuant and Bambu had similar sensitivities (~71%), while StringTie3 was slightly lower (63%). Precision was highest for IsoQuant and StringTie3 (~52%), while Bambu was lowest (21%) (Fig. 2D).

Analysis of transcript discovery across tools revealed shared and tool-specific transcripts. Reference-guided tools annotated a combined total of 72 309 unique transcripts. The largest intersection represented 32 549 transcripts annotated by StringTie3,

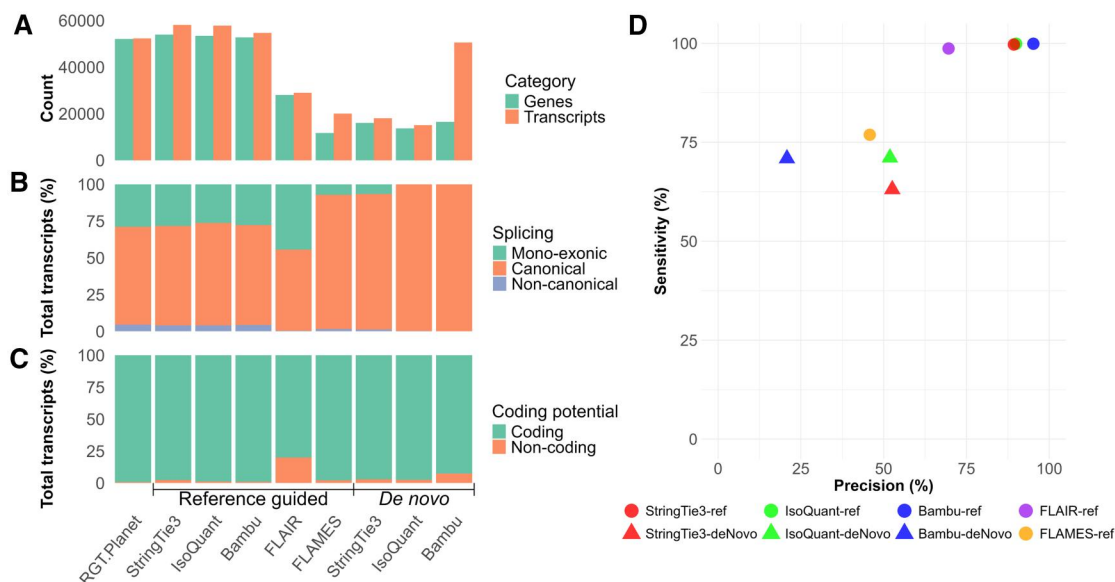


Figure 2 Comparative metrics to benchmark reference-guided and *de novo* methods for genome annotation of barley (cv. RGT Planet) against reference annotations of RGT Planet. (A) Summary of transcript and gene counts generated for annotation tools, (B) splicing patterns, (C) analysis of transcript annotations for coding potential, and (D) transcript-level precision and sensitivity values compared to reference annotations of RGT Planet. One annotation was produced for each tool, and performance values represent complete results, not statistical estimates based on replicates.

IsoQuant, and Bambu, 784 of which were novel to RGT Planet. Most tool-specific transcripts were novel, with FLAIR annotating the highest number (8462) (Fig. 3A). The tools used for *de novo* annotation collectively identified 54 801 transcripts. The largest set, comprising 35 085 transcripts, was generated by Bambu, of which only 1428 were supported by the reference annotation (Fig. 3B). The intersection across *de novo* tools comprised 11 050 transcripts, 3713 of which were unsupported.

3.2 Transcript structure and validation

SQANTI3 classification revealed insights into isoform diversity. Full splice matches (FSMs, known annotations) made up the majority of annotations across all tools and methods, except Bambu *de novo*, which was dominated (53%) by incomplete splice matches (ISMs, transcripts with matching splice junctions to known transcripts but missing exons). Notably, 98% of these Bambu *de novo* ISMs were 3' fragments, meaning 5' exons were

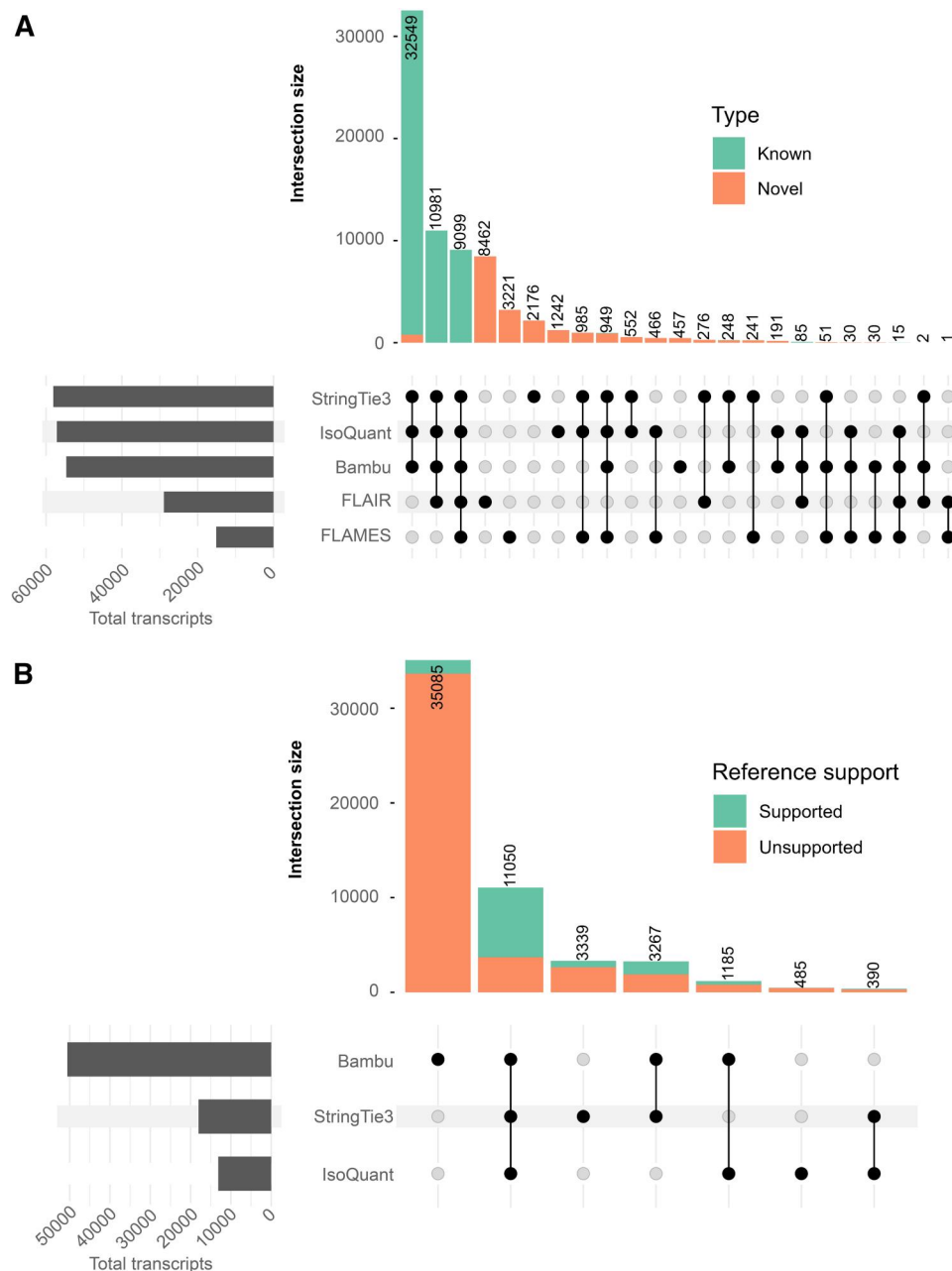


Figure 3 Upset plots showing the shared transcript structure discovery across annotation tools for (A) reference-guided and (B) *de novo* genome annotation of barley (cv. RGT Planet) using direct RNA-seq data. Transcripts with identical intron structures were classified as ‘matching’ using GffCompare. For reference-guided methods, annotations were compared with the RGT Planet reference annotation and classified as known or novel. For *de novo* methods, transcripts were classified as supported or unsupported based on whether corresponding reference models were present. Bar charts display the total number of transcript annotations generated by each tool.

missing. Further analysis into 3' fragment ISMs across all annotations revealed that FLAMES (reference-guided) and Bambu (*de novo*) had the shortest median ISM length of 863 and 899 bp, respectively (Fig. 6, available as [supplementary data](#) at *Bioinformatics Advances* online).

Most annotation sets, except for FLAMES, included intergenic transcripts (transcripts from novel genes). All except FLAIR included a notable number of novel not in catalogue (NNC) transcripts, denoting annotations with at least one novel splice junction within a known gene. Other structural categories, such as genic (overlapping known exons/introns), and novel in catalogue (NIC, novel combinations of known splice sites) represented a smaller proportion (Fig. 4A). Intron retention, the most common alternative splicing event in plants (Chaudhary *et al.* 2019), was highest in FLAMES (415 transcripts), followed by IsoQuant (339), StringTie3 (170), FLAIR (68), and Bambu (9), in reference-guided annotations.

Validation of novel junctions with Illumina short reads showed support across most methods and tools. IsoQuant and StringTie3 had the highest support among reference-guided methods (95% and 93%, respectively), followed by Bambu (89%) and FLAMES (85%). Despite over 20% of FLAIR transcripts being

intergenic, only 15 novel junctions were identified (47% Illumina support). All novel FLAIR junctions were within known genes, indicating that all intergenic annotations were mono-exonic. Novel junctions annotated by *de novo* methods showed high Illumina support, led by IsoQuant (94%) and closely followed by Bambu (91%) and StringTie3 (90%) (Fig. 4B).

3.3 Visualization reveals tool-specific annotation patterns

Annotations and read alignments were visualized for qualitative evaluation of transcript models, revealing several tool-specific patterns. Some tools, particularly Bambu *de novo*, annotated transcripts with identical 3' intron/exon structure to known transcripts which differed in the length/count of 5' exons, typically with low 5' read coverage (Fig. 5A). IsoQuant sometimes annotated additional genes and transcripts at loci with existing annotations (Fig. 5B).

Reference-guided annotations were visualized for four novel and five disease resistance-associated genes. StringTie3 and IsoQuant captured all novel genes, though both missed 5' exons

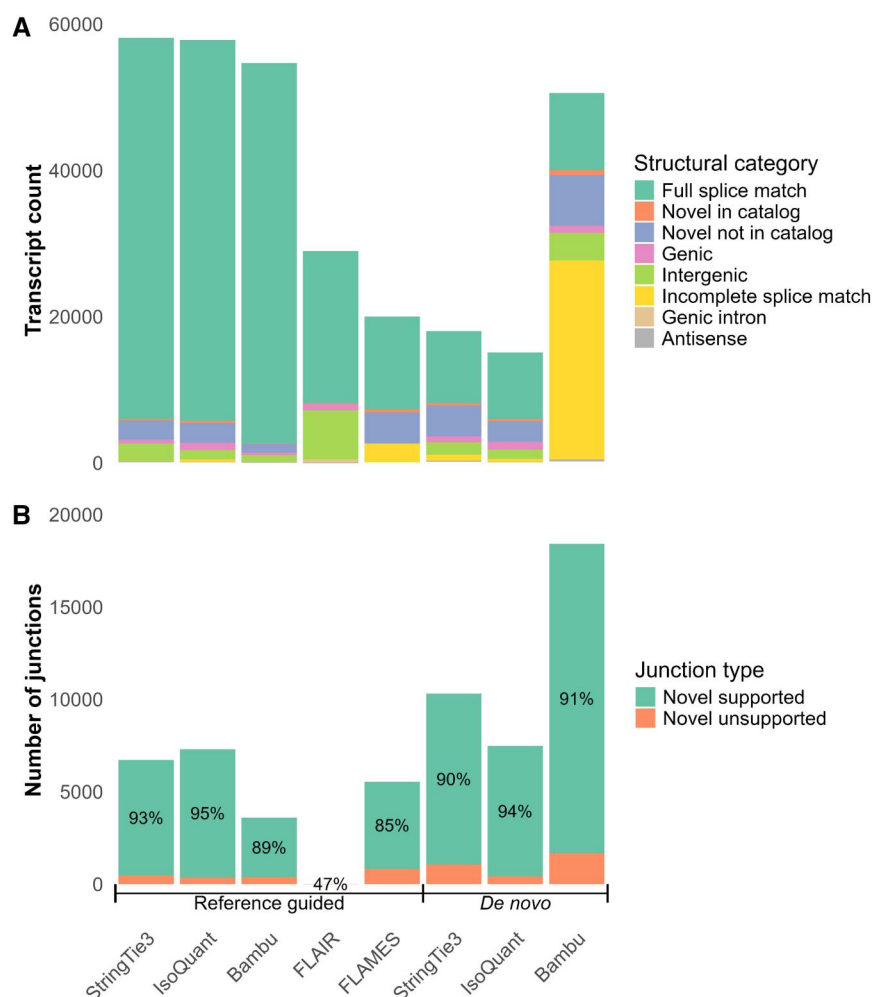


Figure 4 SQANTI3 analysis of annotations generated for barley (cv. RGT Planet) with reference-guided and *de novo* methods. (A) Structural category of transcripts compared to the RGT Planet reference annotation, and (B) Illumina short-read support for novel splice junctions, the percentage of novel junctions supported is shown on the bars.

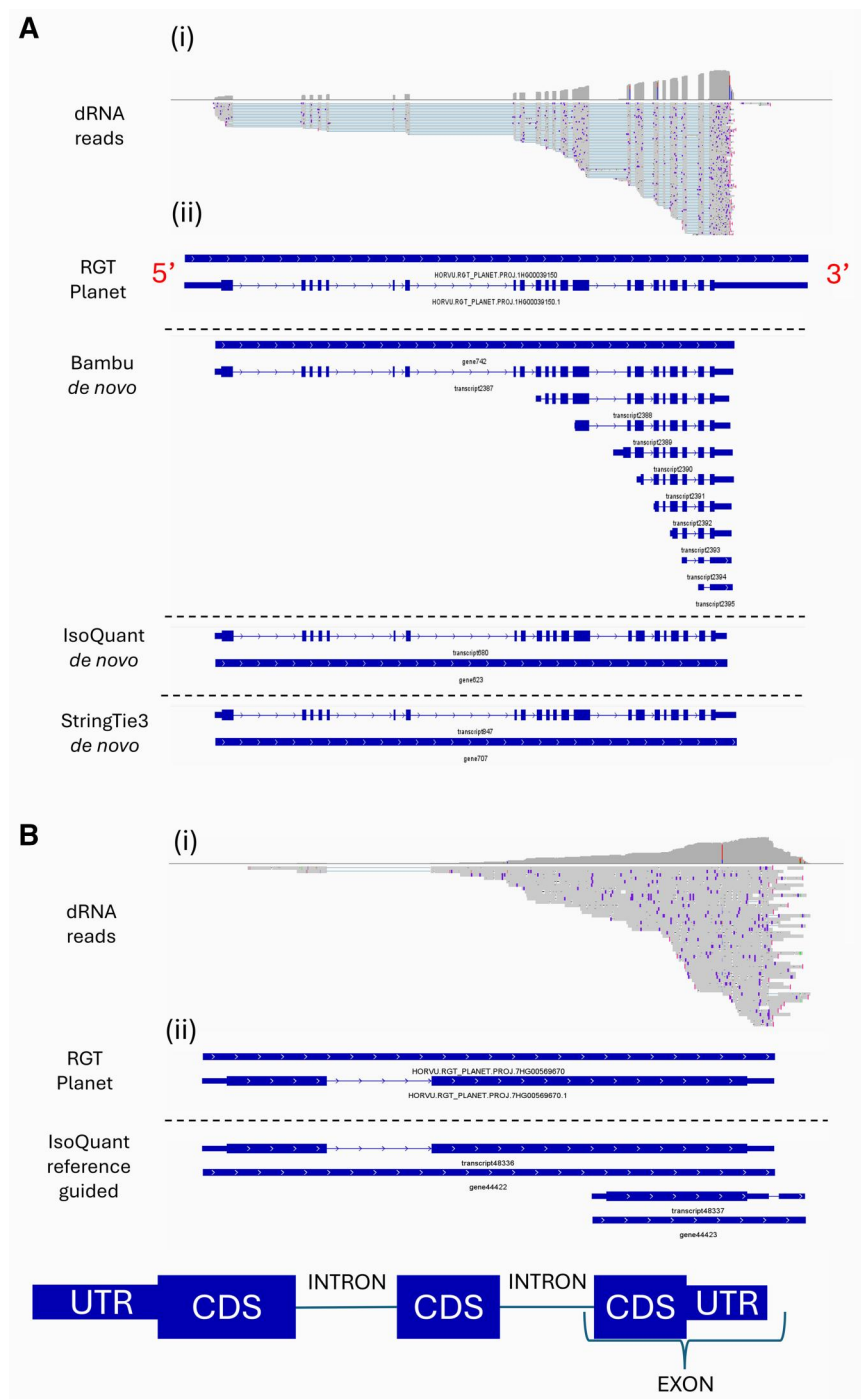


Figure 5 Representative examples of two patterns observed in the visual investigation of annotations produced using reference-guided and *de novo* tools informed by Nanopore direct RNA-seq data. (A) Abundance of annotations with identical splicing at the 3' end, differing in the length of their 5' exon for Bambu *de novo* and not supported by reference annotations or StringTie3 and IsoQuant *de novo*, and (B) Duplicated transcript models at loci with existing annotations generated by IsoQuant reference-guided. (i) Read alignments with corresponding coverage profiles, (ii) annotations. In transcript annotations, thinner lines represent introns, and thicker lines represent exons, split into coding sequences (CDS) and untranslated regions (UTR). Arrows point from the 5' to 3' direction.

at one locus. IsoQuant also introduced an extra gene and transcript at one locus. Bambu annotated three novel genes (one missing a 5' exon), FLAIR annotated two (both missing 5' exons), and FLAMES did not annotate any novel genes (Fig. 7A, available as [supplementary data](#) at *Bioinformatics Advances* online).

For resistance-associated genes, every tool except FLAMES generated transcripts at all five loci. StringTie3 and Bambu matched the reference, while FLAIR and IsoQuant annotated additional 3' fragments and genes, respectively. FLAMES annotated transcripts at three loci, all with extra 5' and/or 3' truncated

transcripts (Fig. 7B, available as [supplementary data](#) at *Bioinformatics Advances* online).

For *de novo* annotations, 10 resistance-associated genes expressed in the NFNB-infected sample were visualized. StringTie3 successfully annotated full-length transcripts at eight loci, while IsoQuant annotated transcripts at seven, including one with extra gene annotations and another lacking complete resolution. Bambu also annotated eight loci but generated additional 5' truncated transcripts at four loci and failed to fully resolve two transcripts (Fig. 8, available as [supplementary data](#) at *Bioinformatics Advances* online).

3.4 Novel transcript discovery and functional characterization

Based on strong performance, the reference-guided IsoQuant and StringTie3 annotations were selected for further analysis.

Comparison with RGT Planet and PanBaRT20 identified 994 (73 with predicted GO terms) and 774 (62 with predicted GO terms) previously unannotated transcripts for StringTie3 and IsoQuant, respectively. REVIGO summarization of GO terms revealed substantial overlap between StringTie3 and IsoQuant, sharing terms including 'cell surface receptor signalling pathway', 'nucleosome assembly', and 'protein phosphorylation'. Some tool-specific terms were observed, including 'defence response' for StringTie3 and 'biosynthetic process' for IsoQuant (Fig. 6). Visual inspection indicated that 74% and 76% of novel, functionally annotated transcripts were supported by ≥ 3 reads for the StringTie3 and IsoQuant annotations, respectively. Approximately 10% of the visualized StringTie3 transcripts and 5% of the IsoQuant transcripts were classified as false as they appeared significantly shorter than read alignments. Short-read RNA-seq analysis identified 7533 significantly differentially expressed (DE) transcripts during disease using the StringTie3 annotation (284 previously

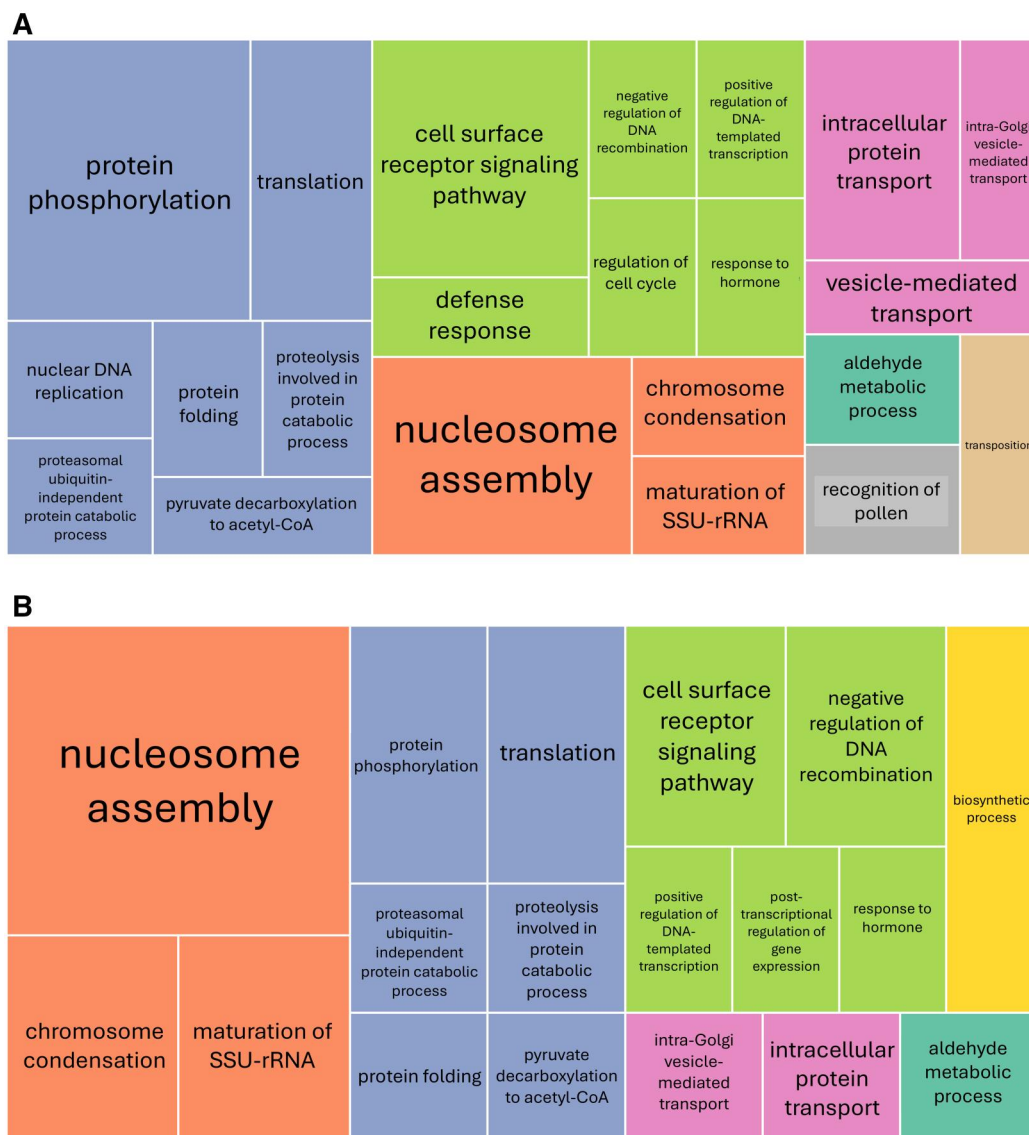


Figure 6 Treemaps showing predicted gene ontology (GO) terms for novel transcripts identified by (A) StringTie3 and (B) IsoQuant using barley direct RNA data. Functional annotation of novel transcripts was performed using InterProScan, and GO term summarization and clustering were conducted with REVIGO to create a treemap. Box sizes represent GO term frequency, and colours indicate groups of semantically similar terms as defined by REVIGO.

unannotated) and 7065 using the IsoQuant annotation (152 previously unannotated).

4 Discussion

This study benchmarked five bioinformatic tools for barley genome annotation using direct RNA (dRNA) sequencing, applying both reference-guided and *de novo* approaches. While previous benchmarking efforts have focused on model organisms, this work provides a crop plant-specific perspective. Comparisons to LRGASP (Pardo-Palacios *et al.* 2024b) and two other benchmarking studies (Sagniez *et al.* 2024, Su *et al.* 2024), provided valuable reference points for interpreting tool performance (Table 4, available as [supplementary data](#) at *Bioinformatics Advances* online). Unlike these earlier studies, which evaluated StringTie2, the present work incorporates the updated StringTie3 release (Shinder *et al.* 2025).

4.1 5' Truncation is a major limitation of dRNA data

A consistent challenge across tools was apparent 5' transcript truncation. Bambu *de novo* frequently annotated transcripts with identical 3' ends to known transcripts with missing 5' exons, supported by SQANTI3's detection of a high proportion of ISMs. Visualization of novel, functionally annotated StringTie3 reference-guided transcript predictions also showed that 10% lacked 5' exons despite read support. This reflects a known limitation of ONT dRNA sequencing, which initiates sequencing of the RNA-complementary DNA (cDNA) hybrid molecule from the 3' poly-A tail, leading to stronger 3' coverage and tapering at the 5' end. As read coverage declined towards 5' ends, annotation tools likely misclassified truncated reads as independent isoforms. The challenge of ONT 5' truncation has been discussed in previous studies (Calvo-Roitberg *et al.* 2024), and was further noted in LRGASP, which reported that tools generally demonstrated greater transcript 3' resolution. The limitation was also noted by Su *et al.* (2024), who simulated ONT datasets with complete reads and datasets with 5' and/or 3' read truncation. The study reported overall decreases in annotation precision and sensitivity for most tools using truncated datasets, supporting the negative impact of incomplete reads on annotation quality.

Several strategies could be trialled in future work to mitigate the impact of 5' truncation. Firstly, the use of alternative reverse transcriptases may optimize the process to ensure RNA-cDNA hybrids are representative of full-length transcripts. Additionally, an increase in flow cell output may improve full-length transcript detection. Given the immense size of the barley transcriptome (Coulter *et al.* 2022), deeper sequencing is likely needed to achieve sufficient transcript coverage, which may be facilitated by continued improvements in ONT technology. The use of 5' adapters could also be explored to improve transcript resolution. This technology replaces 5' caps with an identifiable oligomer, allowing for full-length read detection, and has previously been applied to plant, viral and human transcriptomes (Parker *et al.* 2020, Mulroney *et al.* 2022, Ugolini *et al.* 2022). More stringent filtering for minimum read length during preprocessing may also be beneficial, although this introduces the

risk of excluding true short transcripts. Addressing 3' bias remains a key priority for improving dRNA-informed annotations.

4.2 StringTie3 is a top performing reference-guided tool

Analysis of reference-guided approaches revealed significant differences between tools. Overall, FLAMES and FLAIR performed less reliably. FLAIR, for example, annotated a large proportion of intergenic, mono-exonic transcripts, a finding supported by a 2025 study that identified similar issues using FLAIR with ONT cDNA data from human brain samples (Santucci *et al.* 2025). FLAIR also exhibited poor reliability in other analyses, annotating few genes and transcripts, with a high percentage of non-coding predictions. Sagniez *et al.* (2024) also concluded that FLAIR was not as successful as other tools. FLAMES, on the other hand, did not annotate any intergenic transcripts or novel genes. This was consistent with Su *et al.* (2024), who tested FLAMES on 25 experimental long-read datasets and found a minimal number of intergenic transcripts in only 14. Furthermore, FLAMES was less sensitive and precise, with a higher percentage of ISM models. FLAIR and FLAMES showed the highest detection of tool-specific transcripts, suggesting that 'novel' annotations detected exclusively by one tool may be low-quality predictions and should be critically evaluated. Su *et al.* (2024) noted that the quality of reference annotations had a greater impact on the performance of FLAIR and FLAMES than other tools. Our results further support this finding and demonstrate the unsuitability of FLAIR and FLAMES for the dRNA-informed annotation of complex crop genomes, such as barley, which lack reference resources for comprehensive conditions.

In contrast, StringTie3, IsoQuant, and Bambu performed relatively well in reference-guided mode, showing high precision and near-perfect sensitivity. High sensitivity was expected for IsoQuant and Bambu, which automatically output reference annotations regardless of expression. StringTie3 was run to merge the reference annotation with the assembled transcripts, achieving similarly high sensitivity. All three tools produced annotations with coding and splicing profiles similar to the reference, and thus, these metrics were not a viable basis for comparison. IsoQuant revealed limitations in handling read truncation and Bambu, while strong overall, fell behind StringTie3 in gene annotation, splice junction support, and performance in visualizations. These findings highlight StringTie3 as a top performer for reference-guided dRNA annotation in barley.

4.3 StringTie3 and IsoQuant are top *de novo* annotation tools

De novo annotation comparisons revealed greater variation. Bambu annotated a significant proportion of 3' fragment ISMs, likely due to the high novel discovery rate (NDR=1), which retains all read classes and therefore identifies many 'novel' transcripts arising from truncated reads. This greatly increased the total transcript count and led to a high proportion of tool-specific annotations unsupported by reference transcripts, thereby inflating the number of 'false positives' in sensitivity and precision calculations. Consequently, Bambu-generated

transcripts had low precision, despite comparable sensitivity values to other *de novo* tools. These findings suggest that Bambu's model for *de novo* annotation, trained on human data (Chen *et al.* 2023), is unsuitable for barley. Bambu enables users to train models on species-specific data, which may improve performance, however, this was beyond the scope of the present study. StringTie3 and IsoQuant *de novo* performed more consistently, with similar precision and sensitivity, and SQANTI3 transcript classification profiles. IsoQuant omitted mono-exonic transcripts due to default ONT settings that suppress unspliced transcripts to avoid false positives; this setting can be changed by users. IsoQuant slightly outperformed StringTie3 in short-read junction support, while StringTie3 performed better in visual comparisons, highlighting strengths for each tool.

Despite the limitations associated with dRNA data, both StringTie3 and IsoQuant produced relatively robust *de novo* annotations. The annotations were comparable to those of reference-guided methods in terms of predicted transcript coding potential and relatively high junction support and were mostly comprised of full splice matches. However, dRNA annotation guided by high-quality reference resources remains preferable to *de novo* approaches. Published reference annotations are typically curated using diverse biological evidence and computational methods, whereas *de novo* annotations rely on sequencing datasets that may be affected by technical constraints. Nonetheless, comprehensive reference annotations are limited for many non-model plant species and their diverse genotypes. For example, although the 2024 barley pangenome includes curated annotations for 76 genotypes (Jayakodi *et al.* 2024), hundreds of thousands of barley accessions exist globally (Vioni *et al.* 2023), representing extensive species diversity and numerous uncharacterized genotypes. This gap is significant, given that genotype-specific reference resources for barley have been shown to improve transcript quantification and differential expression analyses compared with using resources from another genotype (Guo *et al.* 2022). Reference resources are even more limited across plant species more broadly, with genome assemblies available for only 0.26%–0.29% of extant green plant species (Bernal-Gallardo and de Folter 2024). Continued expansion of plant genomic resources is therefore essential to support accurate transcriptomic analyses, an effort well-positioned to benefit from long-read sequencing technologies, including dRNA. Overall, in the absence of curated reference annotations, results from this study suggest that dRNA-informed *de novo* annotation with StringTie3 or IsoQuant provides a viable preliminary strategy for complex, non-model plants such as barley.

4.4 StringTie3 and IsoQuant enable the discovery of novel, disease-related transcripts

Further analysis of StringTie3 and IsoQuant reference-guided annotations revealed novel transcripts, several of which had relevant predicted biological functions. One of the most prevalent GO terms identified was 'nucleosome assembly', a cellular process that organizes DNA and histones into nucleosomes. A genome-wide gene expression study in NFNB-susceptible barley previously reported significant suppression of nucleosome

assembly genes during infection (Moolhuijzen *et al.* 2021), likely reflecting a shift in host resource allocation from routine cellular maintenance to defence. Additional predicted GO terms directly associated with disease response included 'cell surface receptor signalling', which mediates recognition of pathogen-associated molecular patterns (Dodds *et al.* 2024), and 'protein phosphorylation', a core driver of immune signalling cascades such as mitogen-activated protein kinase (MAPK) pathways (Park *et al.* 2012). These observations highlight the value of condition-specific long-read data for annotation enrichment, allowing the discovery of transcripts relevant to pathogen defence. However, visual inspection revealed that both tools also produced transcript models considerably shorter than the aligned reads, indicating truncated isoforms arising from dRNA limitations. Thus, although StringTie3 and IsoQuant successfully enriched disease-relevant annotations, improving 5' read completeness will be critical for maximizing annotation accuracy.

4.5 Limitations

While this study provides novel insights into dRNA-informed barley annotation, several limitations should be noted. Firstly, this study utilized default or recommended tool parameters and minimal read filtering (Phred ≥ 10 ; length ≥ 100 bp). More stringent filtering or customized tuning of tools may improve annotation performance, although given the relatively low accuracy of dRNA sequencing (Zhang *et al.* 2024), excessive filtering risks discarding reads that could still annotate valid transcript models when aligned using tolerant algorithms such as minimap2. Secondly, unlike other benchmarking efforts, spike-in controls were not included to evaluate isoform detection accuracy (Hardwick *et al.* 2016). Thirdly, although several metrics were used to assess the quality of novel transcripts, confirmation requires experimental validation, raising broader questions regarding what constitutes sufficient evidence for novel transcript annotation in the era of expansive transcript discovery enabled by long-read sequencing. Finally, computational tools alone cannot guarantee comprehensive or accurate annotations and careful curation is needed to ensure that transcript models reflect biological reality, rather than artefacts (Bagherian *et al.* 2025). Integrating short- and long-read data, using multiple annotation tools, filtering mono-exonic transcripts lacking conserved protein domains, and routine manual visualization are recommended steps to improve annotation reliability (Vuruputoor *et al.* 2023).

5 Conclusion

This study provides new insights into dRNA-based genome annotation in barley and demonstrates the utility of StringTie3 and IsoQuant for discovering biologically meaningful transcripts under disease pressure. These findings could guide further improvements to barley reference annotations using dRNA data generated under diverse conditions, including varied biotic and abiotic stresses. The identification of hundreds of novel transcripts in barley, despite its relatively well-curated existing resources, also suggests strong potential for applying this workflow in plant species with more complex or incomplete

reference resources. Future applications may be supported by the Nextflow-based analysis pipeline provided with this study, which streamlines dRNA pre-processing, annotation and evaluation. Continued refinement of long-read sequencing technology and annotation tools will be essential for building complete, functionally accurate plant transcriptome resources that underpin crop improvement and agrigenomic research.

Acknowledgements

This work was supported by the high-performance computing resources of the Pawsey Supercomputing Research Centre's Setonix Supercomputer (<https://doi.org/10.48569/18sb-8s43>). We thank Drs. Ashley Jones, Simon Dunbar, Alka Saxena, and Johannes Debler for valuable discussions about the use of Oxford Nanopore Technologies and RNA sequencing.

Supplementary material

Supplementary material is available at *Bioinformatics Advances* online.

Conflicts of interest

None declared.

Funding

This work was supported by a co-investment between the Grains Research and Development Corporation and Curtin University [CUR1403-002BLX]; and the Analytics for the Australian Grains Industry [CUR2210-005OPX]. Research activities were undertaken by the Centre for Crop and Disease Management.

Data availability

The data underlying this article are available in the National Centre for Biotechnology Information Sequence Read Archive at <https://www.ncbi.nlm.nih.gov/>, and can be accessed with the BioProject ID PRJNA1282305. The code underlying this article is available on GitHub at https://github.com/jadedavis5/benchmarking_paper (10.5281/zenodo.15636533). The commands include embedded use of a Nextflow-based pipeline to streamline processing, available at <https://github.com/jadedavis5/NanoFLORA> (10.5281/zenodo.15636556).

References

Alhabsi A, Ling Y, Crespi M *et al.* Alternative splicing dynamics in plant adaptive responses to stress. *Annu Rev Plant Biol* 2025; **76**:687–717. <https://doi.org/10.1146/annurev-arplant-083123-090055>

Amarasinghe SL, Su S, Dong X *et al.* Opportunities and challenges in long-read sequencing data analysis. *Genome Biol* 2020; **21**:30. <https://doi.org/10.1186/s13059-020-1935-5>

Amin US, Biswas S, Elias SM *et al.* Enhanced salt tolerance conferred by the complete 2.3 kb cDNA of the rice vacuolar Na (+)/H(+) antiporter gene compared to 1.9 kb coding region with 5' UTR in transgenic lines of rice. *Front Plant Sci* 2016; **7**:14. <https://doi.org/10.3389/fpls.2016.00014>

Andrews S. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> (29 October 2024, date last accessed).

Anjanappa RB, GUISSEM W. Current progress and challenges in crop genetic transformation. *J Plant Physiol* 2021; **261**:153411. <https://doi.org/10.1016/j.jplph.2021.153411>

Bagherian M, Harris G, Sathishkumar P *et al.* Start right to end right: authentic open reading frame selection matters for nonsense-mediated decay target identification. *Genes (Basel)* 2025; **16**:1297. <https://www.mdpi.com/2073-4425/16/11/1297>

Bedre R, Irigoyen S, Schaker PDC *et al.* Genome-wide alternative splicing landscapes modulated by biotrophic sugarcane smut pathogen. *Scientific Reports*, 2019; **9**:8876. <https://doi.org/10.1038/s41598-019-45184-1>

Bernal-Gallardo JJ, de Folter S. Plant genome information facilitates plant functional genomics. *Planta* 2024; **259**:117. <https://doi.org/10.1007/s00425-024-04397-z>

Blum M, Andreeva A, Florentino LC *et al.* InterPro: the protein sequence classification resource in 2025. *Nucleic Acids Res* 2025; **53**:D444–56. <https://doi.org/10.1093/nar/gkae1082>

Calixto CPG, Guo W, James AB *et al.* Rapid and dynamic alternative splicing impacts the *arabidopsis* cold response transcriptome. *Plant Cell* 2018; **30**:1424–44. <https://doi.org/10.1105/tpc.18.00177>

Calvo-Roitberg E, Daniels RF, Pai AA. Challenges in identifying mRNA transcript starts and ends from long-read sequencing data. *Genome Res* 2024; **34**:1719–34. <https://doi.org/10.1101/gr.279559.124>

Camacho C, Coulouris G, Avagyan V *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* 2009; **10**:421. <https://doi.org/10.1186/1471-2105-10-421>

Camargo AP, Sourkov V, Pereira GAG *et al.* RNAsamba: neural network-based assessment of the protein-coding potential of RNA sequences. *NAR Genom Bioinform* 2020; **2**:lqz024. <https://doi.org/10.1093/nargab/lqz024>

Chaudhary S, Khokhar W, Jabre I *et al.* Alternative splicing and protein diversity: plants versus animals. *Front Plant Sci* 2019; **10**:708. <https://doi.org/10.3389/fpls.2019.00708>

Chen Y, Sim A, Wan YK *et al.* Context-aware transcript quantification from long-read RNA-seq data with bambu. *Nat Methods* 2023; **20**:1187–95. <https://doi.org/10.1038/s41592-023-01908-w>

Chenna S, Ivanov M, Nielsen TK *et al.* A data-driven genome annotation approach for cassava. *Plant J* 2024; **119**:1596–612. <https://doi.org/10.1111/tpj.16856>

Chong GL, Foo MH, Lin W-D *et al.* Highly ABA-induced 1 (HA1)-interacting protein HIN1 and drought acclimation-enhanced splicing efficiency at intron retention sites. *Proc Natl Acad Sci USA* 2019; **116**:22376–85. <https://doi.org/10.1073/pnas.1906244116>

Claros MG, Bautista R, Guerrero-Fernández D *et al.* Why assembling plant genome sequences is so challenging. *Biology (Basel)* 2012; **1**:439–59. <https://doi.org/10.3390/biology1020439>

- Coulter M, Entizne JC, Guo W *et al.* BaRTv2: a highly resolved barley reference transcriptome for accurate transcript-specific RNA-seq quantification. *Plant J* 2022;**111**:1183–202. <https://doi.org/10.1111/tbj.15871>
- Dainat J. 2025. AGAT: Another Gff Analysis Toolkit to Handle Annotations in any GTF/GFF Format. Zenodo. [10.5281/zenodo.3552717](https://doi.org/10.5281/zenodo.3552717) (12 January 2025, date last accessed).
- De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* 2023;**39**:btad311. <https://doi.org/10.1093/bioinformatics/btad311>
- Di Tommaso P, Chatzou M, Floden EW *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* 2017;**35**:316–9. <https://doi.org/10.1038/nbt.3820>
- Dinesh-Kumar SP, Baker BJ. Alternatively spliced *N* resistance gene transcripts: their possible role in tobacco mosaic virus resistance. *Proc Natl Acad Sci USA* 2000;**97**:1908–13. <https://doi.org/10.1073/pnas.020367497>
- Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21. <https://doi.org/10.1093/bioinformatics/bts635>
- Dodds PN, Chen J, Outram MA. Pathogen perception and signalling in plant immunity. *Plant Cell* 2024;**36**:1465–81. <https://doi.org/10.1093/plcell/koae020>
- Dominguez Del Angel V, Hjerde E, Sterck L *et al.* Ten steps to get started in genome assembly and annotation. *F1000Res* 2018;**7**:ELIXIR-148. <https://doi.org/10.12688/f1000research.13598.1>
- Dong X, Du MRM, Gouil Q *et al.* Benchmarking long-read RNA-seq analysis tools using *in silico* mixtures. *Nat Methods* 2023;**20**:1810–21. <https://doi.org/10.1038/s41592-023-02026-3>
- Ewels P, Magnusson M, Lundin S *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;**32**:3047–8. <https://doi.org/10.1093/bioinformatics/btw354>
- Gleeson J, Leger A, Prawer YDJ *et al.* Accurate expression quantification from nanopore direct RNA sequencing with NanoCount. *Nucleic Acids Res* 2022;**50**:e19. <https://doi.org/10.1093/nar/gkab1129>
- Gremme G, Steinbiss S, Kurtz S. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform* 2013;**10**:645–56. <https://doi.org/10.1109/tcbb.2013.68>
- Guo W, Coulter M, Waugh R *et al.* The value of genotype-specific reference for transcriptome analyses in barley. *Life Sci Alliance* 2022;**5**:e202101255. <https://doi.org/10.26508/lsa.202101255>
- Guo W, Schreiber M, Marosi VB *et al.* A barley pan-transcriptome reveals layers of genotype-dependent transcriptional complexity. *Nat Genet* 2025;**57**:441–50. <https://doi.org/10.1038/s41588-024-02069-y>
- Gupta N, Verma V. Next-generation sequencing and its application: empowering in public health beyond reality. In: Arora PK (ed.), *Microbial Technology for the Welfare of Society*. Springer Singapore, 2019, 313–41. https://doi.org/10.1007/978-981-13-8844-6_15
- Hardwick SA, Chen WY, Wong T *et al.* Spliced synthetic genes as internal controls in RNA sequencing experiments. *Nat Methods* 2016;**13**:792–8. <https://doi.org/10.1038/nmeth.3958>
- Harrison PW, Amode MR, Austine-Orimoloye O *et al.* Ensembl 2024. *Nucleic Acids Res* 2024;**52**:D891–9. <https://doi.org/10.1093/nar/gkad1049>
- Huang J, Lu X, Wu H *et al.* *Phytophthora* effectors modulate genome-wide alternative splicing of host mRNAs to reprogram plant immunity. *Mol Plant* 2020;**13**:1470–84. <https://doi.org/10.1016/j.molp.2020.07.007>
- Jayakodi M, Lu Q, Pidon H *et al.* Structural variation in the pan-genome of wild and domesticated barley. *Nature* 2024;**636**:654–62. <https://doi.org/10.1038/s41586-024-08187-1>
- Jayakodi M, Padmarasu S, Haberer G *et al.* The barley pan-genome reveals the hidden legacy of mutation breeding. *Nature* 2020;**588**:284–9. <https://doi.org/10.1038/s41586-020-2947-8>
- Jiang C, Kan J, Gao G *et al.* Barley2035: a decadal vision for barley research and breeding. *Molecular Plant*, 2025;**18**:195–218. <https://doi.org/10.1016/j.molp.2024.12.009>
- Jones P, Binns D, Chang HY *et al.* InterProScan 5: genome-scale protein function classification. *Bioinformatics* 2014;**30**:1236–40. <https://doi.org/10.1093/bioinformatics/btu031>
- Kovaka S, Zimin AV, Pertea GM *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol*, 2019;**20**:278. <https://doi.org/10.1186/s13059-019-1910-1>
- Krassowski M, Arts M, Lager C. *Max. Krassowski/Complex-Upset (Version v1.3.5)*. Zenodo. 2022. [10.5281/zenodo.7314197](https://doi.org/10.5281/zenodo.7314197) (28 October 2025, date last accessed).
- Kress WJ, Soltis DE, Kersey PJ *et al.* Green plant genomes: what we know in an era of rapidly expanding opportunities. *Proc Natl Acad Sci USA* 2022;**119**:e2115640118. <https://doi.org/10.1073/pnas.2115640118>
- Kurtzer GM, Sochat V, Bauer MW. Singularity: scientific containers for mobility of compute. *PLoS One* 2017;**12**:e0177459. <https://doi.org/10.1371/journal.pone.0177459>
- Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 2011;**12**:323. <https://doi.org/10.1186/1471-2105-12-323>
- Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;**34**:3094–100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li H, Handsaker B, Wysoker A, 1000 Genome Project Data Processing Subgroup *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9. <https://doi.org/10.1093/bioinformatics/btp352>
- Li H, Li A, Shen W *et al.* Global survey of alternative splicing in rice by direct RNA sequencing during reproductive development: landscape and genetic regulation. *Rice (NY)* 2021;**14**:75. <https://doi.org/10.1186/s12284-021-00516-6>
- Li T, Wang H, Zhang Y *et al.* Comprehensive profiling of epigenetic modifications in fast-growing moso bamboo shoots. *Plant Physiol* 2023;**191**:1017–35. <https://doi.org/10.1093/plphys/kiac525>
- Liu XX, Guo QH, Xu WB *et al.* Rapid regulation of alternative splicing in response to environmental stresses. *Front Plant Sci* 2022;**13**:832177. <https://doi.org/10.3389/fpls.2022.832177>
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. <https://doi.org/10.1186/s13059-014-0550-8>
- Mascher M, Wicker T, Jenkins J *et al.* Long-read sequence assembly: a technical evaluation in barley. *Plant Cell* 2021;**33**:1888–906. <https://doi.org/10.1093/plcell/koab077>

- Mehrotra S, Goyal V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics* 2014;**12**:164–71. <https://doi.org/10.1016/j.gpb.2014.07.003>
- Milne L, Bayer M, Rapazote-Flores P *et al.* EORNA, a barley gene and transcript abundance database. *Sci Data* 2021;**8**:90. <https://doi.org/10.1038/s41597-021-00872-4>
- Moolhuijzen P, Lawrence JA, Ellwood SR. Potentiators of disease during barley infection by *Pyrenophora teres f. teres* in a susceptible interaction. *Mol Plant Microbe Interact* 2021;**34**: 779–92. <https://doi.org/10.1094/mpmi-10-20-0297-r>
- Moolhuijzen P, Sanglard LMVP, Paterson DJ *et al.* Spatiotemporal patterns of wheat response to *Pyrenophora tritici-repentis* in asymptomatic regions revealed by transcriptomic and X-ray fluorescence microscopy analyses. *J Exp Bot* 2023;**74**:4707–20. <https://doi.org/10.1093/jxb/erad183>
- Mulroney L, Wulf MG, Schildkraut I *et al.* Identification of high-confidence human poly(A) RNA isoform scaffolds using nanopore sequencing. *RNA* 2022;**28**:162–76. <https://doi.org/10.1261/rna.078703.121>
- Naim F, Khambatta K, Sanglard LMVP *et al.* Synchrotron X-ray fluorescence microscopy-enabled elemental mapping illuminates the ‘battle for nutrients’ between plant and pathogen. *J Exp Bot* 2021;**72**:2757–68. <https://doi.org/10.1093/jxb/erab005>
- Neuwirth E. 2022. Package ‘RColorBrewer’. <https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf> (16 April 2025, date last accessed).
- Oxford Nanopore Technologies. 2024. *Dorado*. <https://github.com/nanoporetech/dorado> (1 December 2024, date last accessed).
- Pardo-Palacios F, Arzalluz-Luque A, Kondratova L *et al.* SQANTI3: curation of long-read transcriptomes for accurate identification of known and novel isoforms. *Nat Methods* 2024a;**21**: 793–7. <https://doi.org/10.1038/s41592-024-02229-2>
- Pardo-Palacios F, Wang D, Reese F *et al.* Systematic assessment of long-read RNA-seq methods for transcript identification and quantification. *Nat Methods* 2024b;**21**:1349–63. <https://doi.org/10.1038/s41592-024-02298-3>
- Park CJ, Caddell DF, Ronald PC. Protein phosphorylation in plant immunity: insights into the regulation of pattern recognition receptor-mediated signaling. *Front Plant Sci* 2012;**3**:177. <https://doi.org/10.3389/fpls.2012.00177>
- Parker MT, Knop K, Sherwood AV *et al.* Nanopore direct RNA sequencing maps the complexity of *Arabidopsis* mRNA processing and m6A modification. *Elife* 2020;**9**:e49658. <https://doi.org/10.7554/eLife.49658>
- Patel H, Ewels P, Manning J *et al.* 2024. *nf-core/rnaseq*: *nf-core/rnaseq* v3.18.0 - *Lithium Lynx*. <https://doi.org/10.5281/zenodo.1400710> (10 March 2025, date last accessed).
- Pertea G, Pertea M. GFF utilities: gffRead and GffCompare. *F1000Res* 2020;**9**:ISCB Comm J-304. <https://doi.org/10.12688/f1000research.23297.2>
- Prjibelski AD, Mikheenko A, Joglekar A *et al.* Accurate isoform discovery with IsoQuant using long reads. *Nat Biotechnol* 2023;**41**:915–8. <https://doi.org/10.1038/s41587-022-01565-y>
- Ranawaka B, An J, Lorenc MT *et al.* A multi-omic *Nicotiana benthamiana* resource for fundamental research and biotechnology. *Nature Plants* 2023;**9**:1558–71. <https://doi.org/10.1038/s41477-023-01489-8>
- Reiser L, Bakker E, Subramaniam S *et al.* The *Arabidopsis* information resource in 2024. *Genetics* 2024;**227**:iyae027. <https://doi.org/10.1093/genetics/iyae027>
- Robinson JT, Thorvaldsdóttir H, Winckler W *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011;**29**:24–6. <https://doi.org/10.1038/nbt.1754>
- RStudio Team. 2022. *RStudio: Integrated Development for R*. <http://www.rstudio.com/> (10 January 2025, date last accessed).
- Sagniez M, Budhraj A, Paré B *et al.* 2024. Assembly arena: benchmarking RNA isoform reconstruction algorithms for nanopore sequencing. *bioRxiv*, <https://doi.org/2003.2021.586080>, 2024, preprint: not peer reviewed. <https://doi.org/10.1101/2024.03.21.586080>
- Sanglard L, Cuellar SM, Gifford C *et al.* 2025. *Isolation of Pyrenophora teres f. teres from Infected Barley Leaves and Re-inoculation of Barley Leaves in a Glasshouse*. <https://doi.org/dx.doi.org/10.17504/protocols.io.3byl462qrgo5/v1> (15 July 2025, date last accessed).
- Santucci K, Cheng Y, Xu SM *et al.* Discovery of novel protein-coding and long non-coding transcripts in distinct regions of the human brain. *J Mol Neurosci* 2025;**75**:30. <https://doi.org/10.1007/s12031-025-02316-9>
- Sato K. History and future perspectives of barley genomics. *DNA Res* 2020;**27**:dsaa023. <https://doi.org/10.1093/dnares/dsaa023>
- Shen W, Le S, Li Y *et al.* SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 2016;**11**: e0163962. <https://doi.org/10.1371/journal.pone.0163962>
- Shinder I, Pertea G, Hu R *et al.* StringTie3 improves total RNA-seq assembly by resolving nascent and mature transcripts. *bioRxiv*, <https://doi.org/10.1101/2025.05.21.655404>, 2025, preprint: not peer reviewed.
- Soneson C, Yao Y, Bratus-Neuenschwander A *et al.* A comprehensive examination of nanopore native RNA sequencing for characterization of complex transcriptomes. *Nat Commun* 2019;**10**:3359. <https://doi.org/10.1038/s41467-019-11272-z>
- Steijger T, Abril JF, Engström PG *et al.*; RGASP Consortium. Assessment of transcript reconstruction methods for RNA-seq. *Nat Methods* 2013;**10**:1177–84. <https://doi.org/10.1038/nmeth.2714>
- Su Y, Yu Z, Jin S *et al.* Comprehensive assessment of mRNA isoform detection methods for long-read sequencing data. *Nat Commun* 2024;**15**:3972. <https://doi.org/10.1038/s41467-024-48117-3>
- Sun K, Li J, Chen C *et al.* Advances in nanopore direct RNA sequencing and its impact on biological research. *Biotechnol Adv* 2025;**85**:108710. <https://doi.org/10.1016/j.biotechadv.2025.108710>
- Sun L, Xu Y, Langhorst B *et al.* *Induro™*, a Novel Reverse Transcriptase for Nanopore Direct RNA Sequencing with Significantly Improved RNA 5' Coverage, 2020. <https://www.neb.com/en/-/media/posters/document/induro-a-novel-reverse-transcriptase-for-nanopore-direct-rna-sequencing.pdf> 2020 (10 November 2025, date last accessed).
- Supek F, Bošnjak M, Škunca N *et al.* REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 2011;**6**: e21800. <https://doi.org/10.1371/journal.pone.0021800>
- Syme RA, Martin A, Wyatt NA *et al.* Transposable element genomic fissuring in *Pyrenophora teres* is associated with genome expansion and dynamics of host-pathogen genetic

- interactions. *Front Genet* 2018;**9**:130. <https://doi.org/10.3389/fgene.2018.00130>
- Tang AD, Felton C, Hrabeta-Robinson E *et al.* Detecting haplotype-specific transcript variation in long reads with FLAIR2. *Genome Biol* 2024;**25**:173. <https://doi.org/10.1186/s13059-024-03301-y>
- Abueg LAL, Afgan E, Allart O *et al.*; The Galaxy Community. The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res* 2024;**52**: W83–94. <https://doi.org/10.1093/nar/gkac410>
- Tian L, Jabbari JS, Thijssen R *et al.* Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biol* 2021;**22**:310. <https://doi.org/10.1186/s13059-021-02525-6>
- Ugolini C, Mulrone L, Leger A *et al.* Nanopore ReCappable sequencing maps SARS-CoV-2 5' capping sites and provides new insights into the structure of sgRNAs. *Nucleic Acids Res* 2022;**50**:3475–89. <https://doi.org/10.1093/nar/gkac144>
- Verstegen H, Köneke O, Korzun V *et al.* The world importance of barley and challenges to further improvements. In: Kumlehn J, Stein N (eds.), *Biotechnological Approaches to Barley Improvement*. Berlin, Heidelberg: Springer, 2014, 3–19. https://doi.org/10.1007/978-3-662-44406-1_1
- Visioni A, Basile B, Amri A *et al.* Advancing the conservation and utilization of barley genetic resources: insights into germplasm management and breeding for sustainable agriculture. *Plants (Basel)* 2023;**12**:3186. <https://doi.org/10.3390/plants12183186>
- Vuruputoor VS, Monyak D, Fetter KC *et al.* Welcome to the big leaves: best practices for improving genome annotation in non-model plant genomes. *Appl Plant Sci* 2023;**11**:e11533. <https://doi.org/10.1002/aps3.11533>
- Wang B, Tseng E, Regulski M *et al.* Unveiling the complexity of the maize transcriptome by single-molecule long-read sequencing. *Nat Commun* 2016;**7**:11708. <https://doi.org/10.1038/ncomms11708>
- Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* 2012;**28**:2184–5. <https://doi.org/10.1093/bioinformatics/bts356>
- Wang X, Zhang H, Nyamesorto B *et al.* A new mode of NPR1 action via an NB-ARC–NPR1 fusion protein negatively regulates the defence response in wheat to stem rust pathogen. *New Phytol* 2020;**228**:959–72. <https://doi.org/10.1111/nph.16748>
- Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag, 2016. <https://ggplot2.tidyverse.org>
- Wickham H, François R, Henry L *et al.* 2023. *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org> (26 March 2025, date last accessed).
- Winkler TS, Vollmer SK, Dyballa-Rukes N *et al.* Isoform-resolved genome annotation enables mapping of tissue-specific beta-lain regulation in amaranth. *New Phytol* 2024;**243**:1082–100. <https://doi.org/10.1111/nph.19736>
- Wright CJ, Smith CWJ, Jiggins CD. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* 2022;**23**: 697–710. <https://doi.org/10.1038/s41576-022-00514-4>
- Xu T, Yang X, Jia Y *et al.* A global survey of the transcriptome of the opium poppy (*papaver somniferum*) based on single-molecule long-read isoform sequencing. *Plant J* 2022;**110**: 607–20. <https://doi.org/10.1111/tpj.15689>
- Yang Y, Zhang L, Tang Q *et al.* Improved genome annotation of *brassica oleracea* highlights the importance of alternative splicing. *Hortic Plant J* 2024;**10**:961–70. <https://doi.org/10.1016/j.hpj.2024.03.005>
- Zhang D, Lv A, Yang T *et al.* Protective functions of alternative splicing transcripts (*CdDHN4-L* and *CdDHN4-S*) of *CdDHN4* from bermudagrass under multiple abiotic stresses. *Gene* 2020;**763s**:100033. <https://doi.org/10.1016/j.gene.2020.100033>
- Zhang J, Liang Y, Zhang S *et al.* Global landscape of alternative splicing in maize response to low temperature. *J Agric Food Chem* 2022;**70**:15715–25. <https://doi.org/10.1021/acs.jafc.2c05969>
- Zhang S, Li R, Zhang L *et al.* New insights into *arabidopsis* transcriptome complexity revealed by direct sequencing of native RNAs. *Nucleic Acids Res* 2020;**48**:7700–11. <https://doi.org/10.1093/nar/gkaa588>
- Zhang T, Li H, Jiang M *et al.* Nanopore sequencing: flourishing in its teenage years. *J Genet Genomics* 2024;**51**:1361–74. <https://doi.org/10.1016/j.jgg.2024.09.007>
- Zhu X-T, Sanz-Jimenez P, Ning X-T *et al.* Direct RNA sequencing in plants: practical applications and future perspectives. *Plant Commun* 2024;**5**:101064. <https://doi.org/10.1016/j.xplc.2024.101064>
- Zmienko A, Wojciechowski P, Figlerowicz M. Editorial: resolving the complexity of plant genomes and transcriptomes with long reads. *Front Plant Sci* 2021;**12**:832257. <https://doi.org/10.3389/fpls.2021.832257>