scientific reports

Check for updates

OPEN

Genome concatenation enables accurate dual RNA-seq mapping for lignocellulolytic fungi in coculture

Bruna Pena Sollero[1]✉, Rutiane Moreira de Jesus Costa[2], Roberto Coiti Togawa[1], Érica de Castro Costa[2] & Robert Neil Gerard Miller[2]

Dual RNA-seq enables simultaneous transcriptome profiling of interacting organisms, yet accurate read assignment remains challenging, particularly in fungal cocultures where phylogenetic proximity and high percentages of homologous sequences can be shared. Here, we evaluate two mapping strategies for dual RNA-seq analysis of *Trichoderma reesei* RUT-C30 and *Pleurotus citrinopileatus* cocultivated on sugarcane bagasse, based on a sequential alignment against individual genomes and a concatenated reference genome approach. Illumina NovaSeq sequencing generated 175 million high-quality reads, with mapping rates exceeding 96% across all strategies. No significant differences were observed between mapping strategies or genome order in the sequential approach ($p = 0.976$ and $p = 1.0$, respectively). Cross-mapping and multimapping rates were uniformly low (< 2% and < 1.7%, respectively), consistent with the substantial evolutionary divergence between the two fungi, and ensuring that read assignment to each genome was not significantly affected. The concatenated strategy reduced computational time by ~ 93% and yielded slightly more balanced genome-specific read assignments and lower multimapping frequencies. Together, these results demonstrate that concatenated genome mapping offers an efficient and accurate framework for resolving mixed fungal transcriptomes, appropriate for investigation of coculture systems and gene expression modulation during the degradation of lignocellulosic biomass.

Keywords Dual RNA-seq, Sequential mapping, Combined mapping, Fungi, Coculture

Illumina RNA-seq is a next-generation sequencing (NGS) approach widely employed for transcriptome analysis owing to its high throughput, sensitivity, and ability to generate both quantitative and qualitative data with single-base resolution. This technology enables accurate and sensitive gene expression quantification, appropriate for detection of both known and novel transcripts. Moreover, it can be applied to virtually any species, including those lacking a reference genome, thereby facilitating investigation of genetic processes underlying specific phenotypes[1].

At the genome-wide level, dual RNA-Seq is a technique that is appropriate for simultaneous quantification of transcripts from multiple interacting organisms[2], with application across different types of interspecies relations such as host-pathogen, mutualistic and commensal interactions[3]. In dual RNA-seq experiment, appropriate strategies for sample collection and RNA isolation for each organism are essential. Whilst gene transcripts from the different organisms are then resolved computationally by mapping sequence reads to their respective reference genomes or transcriptomes[2–4], this simultaneous analysis introduces significant computational challenges when processing reads derived from distinct organisms. A key difficulty lies in read alignments, which must account for both potential multiple-mapping events within a single reference, where reads can align equally well to several loci, and cross-mapping events that can occur between the two reference genome sequences. Such events can critically affect the accuracy of read assignment between the different organisms. At present, there is no clear consensus on whether sequences that align to more than one genome should be retained or discarded[5].

Various mapping strategies have been proposed, each with distinct limitations. Sequential mapping, in which reads are first aligned to one genome and the remaining unmapped reads are subsequently mapped to the second genome, can bias results toward the first genome analysed, potentially inflating its expression

[1]Embrapa Recursos Genéticos e Biotecnologia, Laboratório de Bioinformática, Parque Estação Biológica – PqEB, Avenida W5 Norte, Caixa Postal 02372, Brasília DF 70770-917, Brazil. [2]Instituto de Ciências Biológicas, Departamento de Biologia Celular, Universidade de Brasília, Campus Universitário Darcy Ribeiro, Brasília DF 70910-900, Brazil. ✉email: bruna.sollero@embrapa.br

estimates. Conversely, mapping against a concatenated (composite) reference genome can reduce cross-mapping artifacts and alleviate some of these issues. Nevertheless, this approach requires careful preprocessing to address overlapping or poorly annotated regions and entails higher computational demands[4].

Although the number of dual RNA-seq studies simultaneously analyzing transcriptome data from multiple organisms has been steadily increasing, particularly those focused on the dynamics of host-pathogen interactions[2,6–8], the application of dual RNA-seq to fungal coculture systems remains comparatively limited, despite its considerable biotechnological potential[9,10]. Fungal coculture, defined as the controlled, axenic cultivation of two or more fungal species within a shared growth environment[11], represents a powerful experimental framework for exploring interspecies interactions relevant to biomass bioconversion and the production of industrially valuable enzymes, biofuels, and other high-value bioproducts. Whilst there have been numerous studies highlighting the potential of fungal cocultures for enhancing enzyme activities and accelerating lignocellulose degradation, in particular involving *Trichoderma reesei RUT-C30*, which is one of the most efficient lignocellulose degraders[12–15], there have been comparatively few investigations on the cooperative, competitive or antagonistic interspecies interactions that occur between fungal cocultures at the transcriptome level[16]. While RNA-seq analyses of fungal monocultures have considerably advanced our understanding of the genes and pathways involved in lignocellulose deconstruction[17–20], dual RNA-seq approaches in fungal co-cultures provide a powerful framework to dissect transcriptional dynamics in interacting species. Systematic evaluation of different read-mapping strategies can further resolve the contributions of each organism, offering unprecedented insights into cooperative biomass degradation and enabling the identification of novel gene networks and enzymatic systems optimized through interspecies interactions. While the technical feasibility of dual RNA-seq is now well established[2], the inherent complexity of whole transcriptome analysis across multiple organisms necessitates careful optimization of bioassay design, RNA and cDNA library preparation, and bioinformatics pipelines for downstream analysis of datasets. With no published work yet to evaluate mapping strategies for mixed RNA-seq data originating from different fungal species grown in coculture, here, we describe a transcriptome mapping strategy that is appropriate for accurate analysis of dual RNA-seq data generated from two lignocellulolytic fungi - an Ascomycete, namely *Trichoderma reesei* RUT-C30, and a Basidiomycete, namely *Pleurotus citrinopileatus*, - grown in coculture on sugarcane bagasse as the sole carbon source. We compare the efficiency and robustness of a conventional sequential mapping approach, in which reads are aligned separately to each genome, with an alternative strategy employing a composite reference genome for simultaneous transcriptome mapping.

Results

Novaseq 6000 sequencing of the coculture cDNA libraries yielded a total of 193,029,095 reads, which, after quality trimming, resulted in 175,497,498 reads (90%) for downstream analysis. Using the program STAR, over 96% of transcripts were successfully mapped (Table 1), irrespective of the strategy employed or genome order in the sequential approach.

To assess whether there were significant differences in the number of mapped reads between mapping strategies (Sequential vs. Concatenated), between genomes (T vs. P) and between genome order within the sequential strategy (TP vs. PT), we performed ANOVA models. No significant differences were detected in the number of uniquely mapped reads between mapping strategies ($p = 0.976$) or between genome orders ($p = 1.0$) (Table 1). However, the difference between the numbers of uniquely mapped reads to each genome was highly significant ($p = 8.95\text{E-}08$).

The average number of uniquely mapped reads for *Pleurotus* was 11,898,579 in the PT mapping order and 11,898,565 in the TP order, whereas for *Trichoderma* the averages were 22,488,766 and 22,488,752, respectively (Fig. 1). On average, *Trichoderma* samples exhibited nearly twice as many uniquely mapped reads as those from *Pleurotus* (Table 1), suggesting either intrinsic differences in genome mappability or unequal representation of both organisms in the coculture transcriptome, as also reported in previous multi-genomes transcriptome studies[9].

The difference in the number of uniquely mapped reads between *Trichoderma* and *Pleurotus* (Diff. T-P) was not significant ($p = 0.999$) but consistently lower across all samples when using the concatenated mapping strategy compared with the sequential strategy. This trend indicates that concatenated reference genomes may promote a more balanced alignment of mapped reads between the two fungal genomes in coculture conditions.

In addition to this potential advantage, the concatenated approach was markedly more efficient, reducing runtime to 3 h 40 min, compared to 65 h 02 min and 46 h 54 min for the sequential mapping strategies (Trichoderma→Pleurotus and Pleurotus→Trichoderma, respectively). Based on the mean runtime, the concatenated strategy required only approximately 6–7% of the time needed for sequential mapping, corresponding to an approximate 93% reduction in computational time.

Table 2 summarizes the analysis of cross-mapping ambiguity metrics, revealing discrepancies in read assignments between mapping strategies, as well as mapping agreements. Cross-mapping metrics generally indicate potential mapping errors; however, the proportion of such ambiguous reads was consistently low - below 2% across all samples. The number of reads classified as "Cross-mapped (seqA vs concB)" (0.67%–2.01%) exceeded those in "Cross-mapped (seqB vs concA)" (0.01%), indicating a higher frequency of mismatches or misassignments when *Pleurotus* served as genome A in the sequential mapping strategy (Table 2).

Overall, "Concordant reads" ("FP seq→ TP through conc (B)"), i.e., false positives [FP] in the sequential approach that were converted to true positives [TP] in the concatenated approach, ranged from 26% to 71% across the mapping orders and were consistently higher in the PT sequential order compared with the TP sequential order (Table 2). Given this effect of genome order in the sequential approach, data indicates that the concatenated mapping strategy provided greater mapping support.

Uniquely Mapped Reads							
Mapping Strategy 1: Sequential - Genome order Trichoderma (T), Pleurotus (*P*)							
Samples	**Total n Reads Mapped**	**n Reads (T)**	**1 st Mapping (T)**	**n Reads (P)**	**2nd Mapping (P)**	**Total % mapped**	**Diff. T-P**
CO_BAG_R1	33,635,116	17,684,623	52.48%	15,161,229	45.08%	97.56%	2,523,394
CO_BAG_R2	34,927,296	21,686,874	61.97%	12,479,662	35.73%	97.70%	9,207,212
CO_BAG_R3	34,559,356	22,962,928	66.31%	10,735,121	31.06%	97.37%	12,227,807
CO_BAG_R4	34,692,815	23,162,681	66.61%	10,946,190	31.55%	98.16%	12,216,491
CO_BAG_R5	37,682,915	26,971,883	71.44%	10,145,394	26.92%	98.36%	16,826,489
Mapping Strategy 1: Sequential - Genome order Pleurotus (*P*), Trichoderma (T)							
Samples	**Total *n* Reads Mapped**	***n* Reads (*P*)**	**1 st Mapping (*P*)**	***n* Reads (T)**	**2nd Mapping (T)**	**Total % mapped**	**Diff. T-*P***
CO_BAG_R1	33,635,116	15,171,110	44.42%	17,674,749	52.55%	96.97%	2,503,639
CO_BAG_R2	34,927,296	12,490,177	35.17%	21,676,481	62.06%	97.23%	9,186,304
CO_BAG_R3	34,559,356	10,745,067	30.57%	22,952,977	66.42%	96.99%	12,207,910
CO_BAG_R4	34,692,815	10,956,320	31.09%	23,152,453	66.74%	97.83%	12,196,133
CO_BAG_R5	37,682,915	10,155,446	26.54%	26,961,944	71.55%	98.09%	16,806,498
Mapping Strategy 2: Concatenated							
Samples	**Total *n* Reads Mapped**	***n* Reads (*P*)**	**Mapping (*P*)**	***n* Reads (T)**	**Mapping (T)**	**Total % mapped**	**Diff. T-*P***
CO_BAG_R1	33,635,116	15,241,108	44.86%	17,678,880	52.04%	96,91%	2,437,772
CO_BAG_R2	34,927,296	12,564,664	35.64%	21,676,069	61.48%	97,13%	9,111,405
CO_BAG_R3	34,559,356	10,799,033	30.99%	22,957,250	65.88%	96,88%	12,158,217
CO_BAG_R4	34,692,815	11,008,691	31.48%	23,155,212	66.22%	97,71%	12,146,521
CO_BAG_R5	37,682,915	10,197,846	26.88%	26,967,066	71.09%	97,98%	16,769,220

Table 1. STAR-derived numbers and percentages of uniquely mapped reads for replicate *Trichoderma reesei* RUT-C30 and *Pleurotus citrinopileatus* coculture samples, considering a sequential mapping strategy (strategy 1), initiating with *Trichoderma* (TP) or *Pleurotus* (PT), and a concatenated mapping strategy (strategy 2). Differences in read numbers between *Trichoderma* (T) and *Pleurotus* (P) are also presented (Diff. T-P).

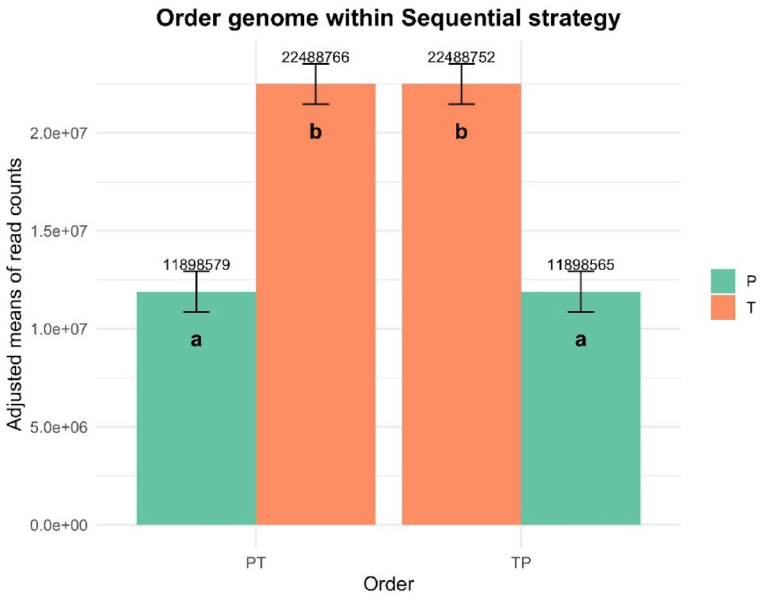

Fig. 1. Bar chart of adjusted mean read counts uniquely mapped to each fungal genome (*Pleurotus* and *Trichoderma*) under the sequential mapping strategy. The effect of mapping order is presented, with different letters indicating significance at $p < 0.001$.

The number of "Rescued reads" was at most 0.01%. Although the number of such observations was very low, the final metric of Table 2, which identifies reads mapped exclusively in the concatenated strategy (and absent in the sequential strategy), further highlights the advantage of the concatenated approach.

In general, True and False Positives (TP and FP) and True and False Negatives (TN and FN) metrics showed that the concatenated model resolved a greater number of cases correctly.

Samples/Total *n* Reads Mapped	Metrics	*n* Reads (PT)	% Reads	*n* Reads (TP)	% Reads
CO_BAG_R1 33,635,116	Cross-mapped (seqA vs. concB)	5,465	1.62%	4,350	1.29%
	Cross-mapped (seqB vs. concA)	23	0.01%	19	0.01%
	FP seq→ TP through conc. (B)	17,671,488	52.54%	15,152,059	45.05%
	FN seq. rescued by conc. (A)	3,541	0.01%	220	0.00%
	FN seq. rescued by conc. (B)	220	0.00%	3,541	0.01%
CO_BAG_R2 34,927,296	Cross-mapped (seqA vs. concB)	6,491	1.86%	3,669	1.05%
	Cross-mapped (seqB vs. concA)	38	0.01%	31	0.01%
	FP seq→ TP through conc. (B)	21,667,533	62.04%	12,472,793	35.71%
	FN seq. rescued by conc. (A)	3,203	0.01%	282	0.00%
	FN seq. rescued by conc. (B)	270	0.00%	3,205	0.01%
CO_BAG_R3 34,559,356	Cross-mapped (seqA vs. concB)	6,943	2.01%	2,852	0.83%
	Cross-mapped (seqB vs. concA)	24	0.01%	28	0.01%
	FP seq→ TP through conc. (B)	22,948,321	66.40%	10,728,077	31.04%
	FN seq. rescued by conc. (A)	4,357	0.01%	296	0.00%
	FN seq. rescued by conc. (B)	334	0.00%	4,357	0.01%
CO_BAG_R4 34,692,815	Cross-mapped (seqA vs. concB)	6,707	1.93%	3,243	0.93%
	Cross-mapped (seqB vs. concA)	24	0.01%	43	0.01%
	FP seq→ TP through conc (B)	23,146,746	66.72%	10,940,628	31.54%
	FN seq. rescued by conc (A)	1,634	0.00%	331	0.00%
	FN seq. rescued by conc (B)	340	0.00%	1,634	0.00%
CO_BAG_R5 37,682,915	Cross-mapped (seqA vs. concB)	7,388	1.96%	2,508	0.67%
	Cross-mapped (seqB vs. concA)	22	0.01%	48	0.01%
	FP seq→ TP through conc (B)	26,955,943	71.53%	10,140,352	26.91%
	FN seq. rescued by conc (A)	2,157	0.01%	2,031	0.01%
	FN seq. rescued by conc (B)	2,026	0.01%	2,157	0.01%

Table 2. Metrics for mapping strategies based on comparison of numbers and percentages of reads between the sequential mapping orders (PT- *Pleurotus* followed by *Trichoderma* and TP- *Trichoderma* followed by *Pleurotus)*.

Samples	Metrics	*n* Cross-mapping (seq-pAtB vs. seq-tApB)	% Reads
CO_BAG_R1	crossAA	9,960	0.03%
	crossBB	8	0.00%
CO_BAG_R2	crossAA	10,513	0.03%
	crossBB	13	0.00%
CO_BAG_R3	crossAA	10,003	0.03%
	crossBB	11	0.00%
CO_BAG_R4	crossAA	10,175	0.03%
	crossBB	9	0.00%
CO_BAG_R5	crossAA	10,099	0.03%
	crossBB	11	0.00%

Table 3. Numbers and percentages of reads cross-mapping to the alternate genome when mapping order was reversed. *CrossAA* denotes the number of reads assigned to *Pleurotus* when it was mapped first compared with those assigned to *Trichoderma* when it was mapped first. *CrossBB* represents the reciprocal comparison (*Trichoderma* mapped second vs. *Pleurotus* mapped second).

Analysis of cross-mapping ambiguity within strategies is summarized in Table 3.

Cross-mapping within the same approach (sequential), i.e. testing reads mapped to the opposite genome when the mapping order was inverted (crossAA or crossBB= tApB vs. pAtB), was extremely low (< 0.03%) (Table 3). This finding reinforces that, although the sequential approach is more time-consuming, it introduces no significant ambiguity regarding read origin. Because the two genomes are highly divergent and share virtually no homologous regions, read assignment becomes highly specific, and mapping ambiguity is expected to be minimal. Genomically distant species inherently reduce the chance of reads aligning to both references with similar scores. The Mash program detected no shared k-mers between the two genomic sketches (MinHash-based summaries of the k-mer content of each genome), yielding a Mash distance of 1 (corresponding to

approximately 0% identity using the 1 – *dist approximation*), with a p-value of 1, confirming the absence of statistically meaningful similarity. Consistently, NUCMER/DNADIFF identified only four small 1:1 alignments (100–150 bp) with an average identity (AvgIdentity) of 88.9% within these small regions. Both the reference and query aligned fractions equal to 0.0002, resulting in an estimated overall genomic identity of 0.000%.

Compared with previous results[21], the percentage of ambiguous reads and the number of multimapped reads between strategies represent key criteria for evaluating mapping performance. For this reason, a multimapping analysis was also conducted (Table 4).

Although the differences were not statistically significant between mapping strategies ($p = 0.722$) nor between mapping order within the sequential approach ($p = 0.934$) (Table 4), the concatenated strategy consistently showed a lower proportion of reads aligning to multiple genomic locations, an outcome associated with reduced mapping ambiguity and improved precision.

Discussion

Transcriptomic analysis provides a powerful framework for investigating fungal coculture systems, enabling quantification of host gene expression changes that underpin biotechnological development in areas that include biofuel production, pharmaceuticals and biopesticide development[11]. Interactions within fungal cocultures, whether cooperative, competitive, or antagonistic, can profoundly influence transcriptional profiles, particularly of genes involved in secondary metabolism, enzymatic degradation, and bioactive compound synthesis. Importantly, many of these products are not necessarily expressed under monoculture conditions, underscoring the relevance of coculture transcriptomics for revealing latent biosynthetic potential for biotechnological applications. Several studies have demonstrated that fungal cocultivation can enhance the production of industrially important enzymes. For instance, Gao and collaborators[10] analysed the transcriptomes of *Penicillium* and *Trichoderma* in coculture, identifying differentially expressed genes involved in lignocellulolytic enzyme systems and secondary metabolism, including antibiotic biosynthesis. Similarly, Liu and collaborators[22] reported substantial upregulation of genes encoding cellulases, hemicellulases, and xylanases in *Penicillium–Trichoderma* coculture systems, reflecting a synergistic enhanced capacity for lignocellulose degradation. These findings emphasize the value of coculture transcriptomics for elucidating interspecies regulation of metabolism and enzyme secretion.

Given the biotechnological potential of coculture transcriptomics, together with the complexity of mixed transcriptomes, there is a need for accurate analytical frameworks to resolve the transcriptional contributions of each species within a mixed system. In this context, dual RNA-seq has emerged as an effective approach for studying coculture. With regard to the parsing of mixed sequence reads from the different organisms, mapping strategies differ among analytical pipelines, with some employing concatenated reference genomes[3,5,7,9], and others aligning reads to multiple genomes in parallel[6,8,23,24]. To our knowledge, this is the first study to apply dual RNA-seq to two phylogenetically distinct fungal species - an ascomycete (*Trichoderma*) and a basidiomycete (*Pleurotus*) - comparing a one-step concatenated genome mapping strategy with a conventional two-step sequential approach based on individual genome alignments of reads from cocultured samples.

Previous studies have adopted similar strategies in other biological systems. Westermann and collaborators[6], for example, quality-filtered RNA-seq reads were aligned in parallel against respective host and pathogen reference genomes, with cross-mapped reads, i.e. those aligning equally well to both reference genomes, excluded from downstream analyses. Similarly, in the present study, the STAR aligner enabled the systematic identification and filtering of multi-mapped and ambiguous reads. While such reads were still observed following preprocessing and alignment, their proportion was quantified and accounted for, minimizing potential biases in genome-specific read assignment. According to O'keeffe and Jones[4], the STAR program, which employs a seed and anchor algorithm, based on the Maximal Mappable Prefix, has been shown to outperform other mapping programs in handling non-continuous reads and mismatches.

Daly and collaborators[9] demonstrated the value of composite (concatenated) reference genomes for mapping multi-species dual RNA-seq, revealing differences in gene content and enzymatic activities between three fungal species. Similarly, literature[7,3]] highlighted the advantages of concatenated references for reducing false positives, minimizing computational demand, and improving mapping accuracy by allowing the software to determine the most appropriate genome assignment for each read. Consistent with these reports, the concatenated mapping strategy in the present study yielded results comparable in accuracy with the sequential approach, while requiring only ~ 14% of its computational time. Although statistical precision did not differ significantly, the concatenated approach consistently produced fewer multi-mapped reads, thereby increasing mapping specificity and reducing ambiguity. These features make it advantageous for large-scale analyses in mixed fungal transcriptomes, where multiple genomes from closely related species can be involved.

Samples	Total reads	SEQ_TP	% Mult. Reads	SEQ_PT	% Mult. Reads	CONC	% Mult. Reads
COBAGR1	33,635,116	565,005	1.68	567,171	1.69	545,233	1.62
COBAGR2	34,927,296	551,513	1.58	559,513	1.6	538,845	1.54
COBAGR3	34,559,356	536,436	1.55	540,459	1.56	520,918	1.51
COBAGR4	34,692,815	394,124	1.14	399,465	1.15	375,026	1.08
COBAGR5	37,682,915	370,923	0.98	376,136	1	357,591	0.95

Table 4. Number and percentage of multimapped reads identified in the sequential strategies (genome orders: SEQ_TP and SEQ_PT) and the concatenated strategy (CONC).

No significant statistical differences were observed between mapping orders in the sequential approach, although variations in read distribution (TP - True Positives vs. FP - False Positives) suggest that mapping order may introduce a potential methodological bias, particularly when the two examined genomes differ in relative transcript abundance or genome assembly quality. Such bias implies that analyses relying solely on sequential mapping may lead to skewed interpretations of differential expression. In this study, *Trichoderma* consistently received approximately twice as many uniquely mapped reads as *Pleurotus*, which may reflect differences in relative biomass abundance, superior genome assembly quality, or genuine biological variation in global transcriptional activity. This asymmetry has methodological implications when interpreting coculture species dominance and interaction dynamics, as noted by O'Keeffe and Jones[4], who observed that host-pathogen sequence read ratios can significantly affect the accuracy and biological interpretation of dual RNA-seq data. Aprianto and collaborators[7] observed comparable findings when mapping reads to a chimeric genome containing the concatenated genomes of *H. sapiens* and *S. pneumoniae*, with one-step mapping chosen to minimize false negative rates. High alignment rates were achieved (>99%), with reads distributed between epithelial (42%) and pneumococcal (58%) origins. As in the present study, the high proportion of usable reads from each library reflected the quality and suitability of the concatenated mapping approach for dual RNA-seq. Baddal and collaborators[25] analyzed NTHi colonization in differentiated ciliated bronchial epithelium using dual RNA-seq in a sequential approach, and found no reciprocal cross-mapping between host and pathogen. Unlike the current study, however, their system involved the interaction between an eukaryotic host and a prokaryotic pathogen, organisms which are genomically far more distinct than the two fungal species investigated here. Espindula and others[5] also concluded that combined mapping approaches resulted in lower numbers of cross-mapped reads and improved consistency in dual RNA-seq studies.

Together, these results demonstrate that the genomes are sufficiently divergent to allow unambiguous read assignment between *Pleurotus* and *Trichoderma*. Moreover, when the sequential mapping began with *Trichoderma*, the proportion of cross-mapped reads was even lower and close to zero. Despite the high genomic divergence between the two fungal species, some orthologous genes are expected to retain substantial sequence similarity, potentially leading to ambiguous read assignments, particularly among multimapped reads. Therefore, when applying the sequential mapping approach to the two fungal genomes, results must be interpreted with caution, as genome order can introduce minor variations in read distribution that may influence downstream analyses. Mapping bias remains a key consideration for accurate read assignment between related genomes, reinforcing that the concatenated reference genome strategy provides a more robust framework for quantifying genome-specific expression or abundance.

The evaluation of FP (False Positive)/TP (True Positive)/FN (False Negative) metrics confirmed that the concatenated mapping approach minimized information loss and order-dependent bias, consistent with results reported for *Glycine max–Bradyrhizobium* and *Zea mays–Fusarium* in dual RNA-seq systems[5].

Overall, the results demonstrate that the concatenated reference genome strategy achieved comparable mapping accuracy to the sequential strategy and provides a balanced, computationally efficient, biologically robust framework for dual RNA-seq analyses of interacting fungal species. This approach can minimize mapping ambiguity, rescue unassigned reads, and substantially reduces computational processing time. While sequential mapping remains a valid method, its susceptibility to order-dependent bias and higher rates of ambiguous read assignment may compromise the accuracy of differential expression analyses. The concatenated strategy therefore represents a practical, precise and scalable solution for transcriptomic investigations fungal co-cultures, microbial consortia, and other complex multi-organism interactions of biotechnological relevance, where accurate read assignment and quantification are essential for interpreting interspecies dynamics and metabolic cooperation.

We acknowledge that different levels of genetic relatedness between organisms may influence ambiguity and cross-mapping behavior. However, the statistical criteria and evaluation framework presented here are organism-independent and reproducible, providing a practical basis for benchmarking mapping performance in diverse coculture systems. Further taxonomically stratified benchmarking studies on expanded datasets are warranted in this context.

Methods

Study system experimental design

Fungal strains of the ascomycete *T. reesei RUTC-30* and the basidiomycete *Pleurotus citrinopileatus* were initially activated on Potato Dextrose Agar medium (ThermoFisher Scientific, Paisley, UK), then mycelial plugs from each 7-day old culture transferred to a liquid minimal medium (7 g KH_2PO_4, 2 g K_2HPO_4, 0.5 g $MgSO_4$, 1.6 g de $(NH_4)_2$ SO_4, pH 7.0, per L of distilled water), supplemented with non-pretreated sugarcane bagasse (BAG), (obtained from a local source) as sole carbon source (at 1% w/v). Liquid cocutures were grown in 100 mL of the growth media in 125mL Erlenmeyer flasks at 28 °C and 150 rpm over a 6-day period. Culture arrangement followed a randomized block design, with five replicates for each culture.

Sample preparation and collection

Mycelia from cocultures were harvested via filtration after the 6-day growth period and flash frozen in liquid nitrogen. Total RNA isolation from the two organisms was isolated with PureLink™ Plant RNA Reagent (Invitrogen – ThermoFisher Scientific, Waltham, MA, USA), followed by purification using the Direct-zol™ RNA Miniprep Plus Kit (Zymo Research, Irvine, CA, USA) following the manufacturer's instructions. RNA integrity was verified by agarose gel electrophoresis and quantification via NanoDrop™ One Microvolume spectrophotometry (ThermoFisher Technologies, Waltham, MA, USA), employing a quality cut-off value of 1.8 for A260:280 ratios. RNA integrity was also measured on an Agilent 2100 Bioanalyzer system (Agilent Technologies, Palo Alto, CA, USA).

RNA sequencing

RNA samples from the coculture replicates were sequenced as paired-end reads (2 × 150 bp) using NovaSeq 6000 sequencing technology. This approach was chosen to minimize mapping performance biases related to sequencing technology platform. Raw paired-end reads were processed with FastP[26] to remove adaptors and low-quality bases, applying a minimum cutoff of 20 bp after trimming. Read quality was subsequently evaluated with FASTQC (threshold > 30; https://www.bioinformatics.babraham.ac.uk/projects/fastqc).

Trimmed coculture reads from all replicate samples were aligned to the reference genomes *Pleurotus citrinopileatus* (GCA_017312325.1) and *Trichoderma reesei* RUT C-30 (GCA_000513815.1) using STAR (version 2.7.10a)[27]. Two mapping strategies were employed[5,7,9,21]: (i) sequential mapping, where reads were mapped independently to each fungal reference genome, and (ii) composite mapping, in which both reference genomes were concatenated into a single FASTA file for alignment.

Genome mapping strategy (i) - sequential

In dual RNA-seq experiments, transcripts from interacting species are separated in silico.

Here, reads were aligned to both reference fungal genomes using a sequential mapping strategy implemented with STAR. In each mapping round, one genome was designated as the primary reference (A), while unmapped reads were subsequently aligned to the secondary genome (B). Specifically, reads were first mapped to genome A using default parameters with the additional option --outReadsUnmapped Fastx, which outputs unmapped reads in FASTQ format. These unmapped reads were then used as input for a second STAR run, this time mapping them to genome B. To control for potential biases introduced by genome order, the procedure was repeated with the genome reference order inverted. This approach enabled a comprehensive evaluation of mapping efficiency and potential read assignment biases between the two fungal genomes.

Genome mapping strategy (ii) - concatenated

In addition to the sequential approach, a single-step mapping strategy was also performed using a concatenated reference genome. For this, the FASTA files of genome A and genome B were merged with the Linux 'cat' command, and a combined reference index was generated using the STAR genome generation step. Coculture reads were then aligned to this concatenated reference in a single run with STAR, applying standard parameters. In this strategy, both genomes directly compete for read alignment, with STAR assigning each read to its best-matching locus across the combined genomic space.

Comparison criteria

To evaluate the efficiency of the sequential and concatenated mapping strategies, a conceptual context was first provided to support metric interpretation (a, b, c) followed by the establishment of a set of standardized comparison metrics (d), as described below.

a) Sequential strategy: reference assumption for comparison.

For comparative purposes only, the sequential approach was interpreted under the assumption that the first genome in the mapping order (genome A) represented the "expected" reference. Under this framework, reads mapped to genome A were classified as True Positives (TP), while reads mapped to genome B were considered False Positives (FP). This assumption does not imply biological certainty, since in coculture experiments reads may genuinely originate from either species; rather, it provides a consistent baseline for methodological comparison.

b) Within-strategy comparison (sequential order effect).

To evaluate whether the order of genomes influenced mapping outcomes, we identified cross-mapped reads, defined as reads that mapped to genome A when this genome was aligned first but mapped to genome B when the order was inverted. This metric allowed us to detect potential misassignments caused solely by the mapping order.

c) Between-strategy comparison (sequential vs. concatenated).

Comparisons between strategies focused on identifying:

- Reads unmapped in the sequential strategy but successfully aligned in the concatenated strategy, and
- Reads assigned to genome A in one strategy but reassigned to genome B in the other.

d) Defined comparison metrics (Fig. 2).

To quantify accuracy and efficiency, the following metrics were computed:

1. Rescued False Negatives (FN):
2. Concordant Reads:
3. Mismatched Reads (cross-mapping):
4. Multimapped Reads:

All metrics were computed using in-house Bash scripts applied to *.fastq* or *.bam* files generated by each strategy. Although the concatenated approach produced a single alignment file (since both genomes were mapped in one run), reads originating from each species were later identified and separated to enable direct comparison with the sequential strategy. For both approaches, read identifiers were partitioned into genome A (first genome) and genome B (second genome) prior to downstream calculations.

All analyses were performed on a Linux server running Ubuntu 24.04, equipped with an Intel Xeon X7560 @2.27 GHz processor (64 cores total), 256GB RAM, and NFS storage. Read alignment was performed using

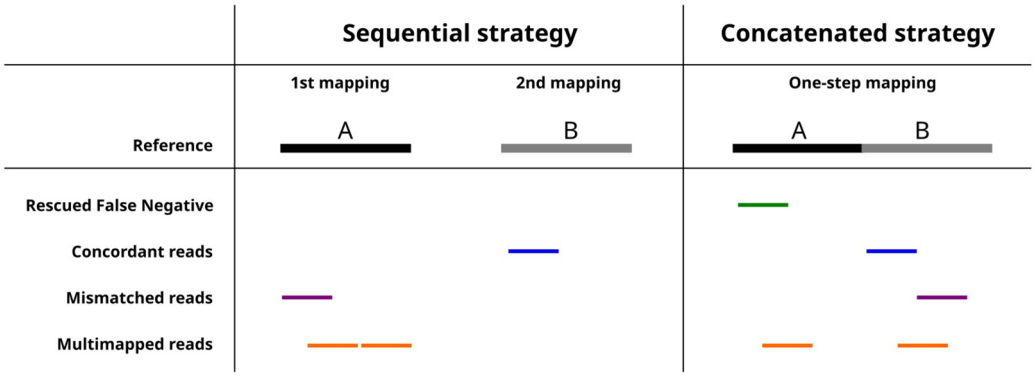

Fig. 2. Schematic representation of criteria/metrics used to assess mapping accuracy, agreement and discrepancies between the sequential and/or concatenated strategies.

STAR v2.7.10a with 16 threads. Runtime was measured using the Unix time command, representing wall-clock time for complete execution.

Statistical model

For comparisons between strategies and across genome order (*Pleurotus-Trichoderma* [PT] vs. *Trichoderma-Pleurotus* [TP]), an additive analysis of variance (ANOVA) model was fitted to assess the main effects of the two factors: (i) group (mapping order: PT vs. TP) and (ii) treatment (sequential vs. concatenated mapping strategy) on the response variable (number of mapped reads). The model considers both effects independently, given that the interaction between the factors was not significant, according to the following equation: $Yijk=\mu+\alpha i+\beta j+\varepsilon ijk$ (1), where Yijk represents the observed response variable, μ is the overall mean, αi is the effect of the i-th level of the group factor, βj is the effect of the j-th level of the treatment factor, and $\varepsilon ijk \sim N(0,\sigma 2)$ is the random error term, assumed to follow a normal distribution with constant variance (homoscedasticity). The model (1) was implemented in R (R Core Team, 2025) using the *aov()* function.

To further contextualize mapping performance, we estimated genomic distances between the two fungi using Mash[28], and performed whole-genome alignments for fine-scale genome comparative analysis using NUCmer[29].

Data availability

The sequencing data supporting the findings of this study have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject accession number **PRJNA1390794** and are publicly available.

Received: 22 December 2025; Accepted: 5 March 2026
Published online: 11 April 2026

References

1. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57–63 (2009).
2. Westermann, A. J., Gorski, S. A. & Vogel, J. Dual RNA-seq of pathogen and host. *Nat. Rev. Microbiol.* **10**, 618–630 (2012).
3. Wolf, T., Kämmer, P., Brunke, S. & Linde, J. Two's company: studying interspecies relationships with dual RNA-seq. *Curr. Opin. Microbiol.* **42**, 7–12 (2018).
4. O'Keeffe, K. R. & Jones, C. D. Challenges and solutions for analysing dual RNA-seq data for non-model host–pathogen systems. *Methods Ecol. Evol.* **10**, 401–414 (2019).
5. Espindula, E., Sperb, E. R., Bach, E. & Passaglia, L. M. P. The combined analysis as the best strategy for dual RNA-seq mapping. *Genetics Mol. Biology* **42**, e20190215 (2019).
6. Westermann, A. J., Barquist, L. & Vogel, J. Resolving host–pathogen interactions by dual RNA-seq. *PLoS Pathogens* **13**, e100603 (2017).
7. Aprianto, R., Slager, J., Holsappel, S. & Veening, J. W. Time-resolved dual RNA-seq reveals extensive rewiring of lung epithelial and pneumococcal transcriptomes during early infection. *Genome Biol* **17**, 198 (2016).
8. Vela, S., Wolf, E. S. A., Rollins, J. A., Cuevas, H. E. & Vermerris, W. Dual-RNA-sequencing to elucidate the interactions between sorghum and Colletotrichum sublineola. *Front Fungal Biol* **5**, 1437344 (2024).
9. Daly, P. et al. Transcriptomic responses of mixed cultures of ascomycete fungi to lignocellulose using dual RNA-seq reveal inter-species antagonism and limited beneficial effects on CAZyme expression. *Fungal Genet. Biol.* **102**, 4–21 (2017).
10. Gao, L. et al. Combinatorial Engineering of Transcriptional Activators in Penicillium oxalicum for Improved Production of Corn-Fiber- Degrading Enzymes. (2021). https://doi.org/10.1021/acs.jafc.0c07659
11. Vieira, R. I. M. et al. Fungal Coculture: Unlocking the Potential for Efficient Bioconversion of Lignocellulosic Biomass. 1–36 (2025).
12. Duff, S. J. B., Cooper, D. G. & Fuller, O. M. Effect of media composition and growth conditions on production of cellulase and β-glucosidase by a mixed fungal fermentation. *Enzyme Microb. Technol.* **9**, 47–52 (1987).
13. Ahamed, A. & Vermette, P. Culture-based strategies to enhance cellulase enzyme production from Trichoderma reesei RUT-C30 in bioreactor culture conditions. *Biochem. Eng. J.* **40**, 399–407 (2008).
14. Kolasa, M., Ahring, B. K., Lübeck, P. S. & Lübeck, M. Co-cultivation of Trichoderma reesei RutC30 with three black Aspergillus strains facilitates efficient hydrolysis of pretreated wheat straw and shows promises for on-site enzyme production. *Bioresour Technol.* **169**, 143–148 (2014).
15. Sperandio, G. B. et al. Exploring the Synergistic Secretome: Insights from Co-Cultivation of Aspergillus brasiliensis and Trichoderma reesei RUT-C30. *J. Fungi.* **10**, 677 (2024).

16. Borin, G. P. et al. Comparative transcriptome analysis reveals different strategies for degradation of steam-exploded sugarcane bagasse by Aspergillus niger and Trichoderma reesei. 1–21 (2017). https://doi.org/10.1186/s12864-017-3857-5
17. Midorikawa, G. E. O. et al. Analysis of the Transcriptome in Aspergillus tamarii During Enzymatic Degradation of Sugarcane Bagasse. *Front Bioeng. Biotechnol* **6**, 123 (2018).
18. Corrêa, C. L. et al. Transcriptome Profiling-Based Analysis of Carbohydrate-Active Enzymes in Aspergillus terreus Involved in Plant Biomass Degradation. *Front Bioeng. Biotechnol* **8**, 563415 (2020).
19. Umezawa, K., Niikura, M., Kojima, Y., Goodell, B. & Yoshida, M. Transcriptome analysis of the brown rot fungus Gloeophyllum trabeum during lignocellulose degradation. *PLoS One*. **15**, e0243984 (2020).
20. Ries, L. et al. Genome-wide transcriptional response of Trichoderma reesei to lignocellulose using RNA sequencing and comparison with Aspergillus niger. *BMC Genom.* **14**, 541 (2013).
21. Fruggiero, C., Aufiero, G., D'Angelo, D. & Pasolli, E. & D'Agostino, N. Refining dual RNA-seq mapping: sequential and combined approaches in host-parasitic plant dynamics. *Front Plant. Sci* **15**, 1378894 (2024).
22. Liu, Q. et al. CLR-4, a novel conserved transcription factor for cellulase gene expression in ascomycete fungi. *Mol. Microbiol.* **111**, 373–394 (2019).
23. Kovalchuk, A. et al. Dual RNA-seq analysis provides new insights into interactions between Norway spruce and necrotrophic pathogen Heterobasidion annosum s.l. *BMC Plant. Biol.* **19**, 1–17 (2019).
24. Zamora-Ballesteros, C. et al. Dual rna-sequencing analysis of resistant (Pinus pinea) and susceptible (pinus radiata) hosts during fusarium circinatum challenge. *Int J. Mol. Sci* **22**, 5236 (2021).
25. Baddal, B. et al. Dual RNA-seq of nontypeable haemophilus influenzae and host cell transcriptomes reveals novel insights into host-pathogen cross talk. *MBio* **6**, (2015).
26. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
27. Dobin, A. et al. Ultrafast universal RNA-seq aligner. *Bioinf. 29*. **STAR**, 15–21 (2013).
28. Ondov, B. D., Bergman, N. H. & Phillippy, A. M. Interactive metagenomic visualization in a Web browser. *BMC Bioinform.* **12**, 385 (2011).
29. Marçais, G. et al. MUMmer4: A fast and versatile genome alignment system. *PLOS Comput. Biol.* **14**, e1005944 (2018).

Author contributions
B.P.S. conceived and designed the study. R.M.J.C and E.C.C. conducted the biotic assays. B.P.S., R.C.T., R.M.J.C and E.C.C. conducted the bioinformatic analysis. R.N.G.M. provided intellectual input. B.P.S. and R.N.G.M. wrote and revised the manuscript. All authors have read and approved the final manuscript.

Funding
This work was funded by Brazilian funding bodies FAPDF (Grant 00193–00002388/2023-21) and CAPES (Finance Code 001). RNGM was supported by a fellowship from CNPq (Grant 308165/2021-7).

Declarations

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to B.P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.