Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Candidate alternative splice-junction architecture at a Ganoderma carnitine acetyltransferase-like locus: an accession-traceable RNA-seq reanalysis

Kiarash Sadeghian Esfahani

arturestrong@gmail.com

Independent researcher https://orcid.org/0000-0002-1844-0675

Research Article

Keywords: Ganoderma, alternative splicing, splice junction, accession audit, RNA-seq, computational reanalysis

Posted Date: November 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11714081/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Candidate alternative splice-junction architecture at a Ganoderma carnitine acetyltransferase-like locus: an accession-traceable RNA-seq reanalysis

Kiarash Sadeghian Esfahani

Independent researcher, Isfahan, Iran

Corresponding author: Kiarash Sadeghian Esfahani | Email: arturestrong@gmail.com

Article type: Research Article (focused computational secondary analysis)

Running title: Candidate Ganoderma splice-junction architecture

Abstract

A focused reanalysis of public Ganoderma transcriptomic resources evaluated two candidate splice-junction paths (A and B) at a putative carnitine acetyltransferase-like locus (MSTRG.5584). Forty RNA-seq run labels from five genomic backgrounds were reconciled with accession metadata and examined using targeted splice-spanning read counts; 11 undercovered libraries underwent complete read-1 recounting. Both candidate junction paths were observed in all five backgrounds. In libraries with any focal read support, mean junction-A fractions were descriptively greater in mycelium-labeled than in fruiting-body-labeled groups, with background-level differences ranging from +0.190 to +0.535. However, the held-out four-background exact two-sided sign test was not significant (P = 0.125), and the original discovery screen had an approximate false-discovery rate of q = 0.136. A separate Ganoderma lingzhi GL0102 genome contained a closely matching candidate genomic segment (approximately 99.5% identity over 16.7 kb); an external PacBio-derived run yielded eight A-like and six B-like splice-spanning read identifiers. These observations support the documentation of candidate alternative junction architecture across datasets, not confirmation of tissue-regulated isoform switching. Uneven RNA-seq sampling depth, uncertain biological-specimen independence, and incomplete full-length transcript validation limit further inference.

Keywords: Ganoderma; alternative splicing; splice junction; accession audit; RNA-seq; computational reanalysis

1. Introduction

Alternative pre-mRNA processing can generate different transcript structures at the same fungal locus. Such diversity is particularly relevant to developmentally changing basidiomycetes, yet observed splice-junction reads do not, by themselves, establish distinct stable proteins or regulatory causation. In Ganoderma, alternative splicing has been experimentally characterized at Swi6, thaumatin-like protein and Gllac7 loci (Lian et al., 2021; Liu et al., 2023; Wang et al., 2025). Hence, neither fungal alternative splicing nor development-associated transcript variation constitutes a new phenomenon.

Reference genomic work in Ganoderma (Chen et al., 2012), along with strand-specific transcript profiling (Shao et al., 2017), provides a foundation for comparative transcript reanalysis. Ganoderma comparative genome resources provide an opportunity to assess individual transcript architectures across genetic backgrounds. Yu et al. (2024) assembled and compared multiple Ganoderma genomes and generated transcriptomes from mycelia and fruiting bodies. Their biological focus was pan-genomic variation and spore development rather than the MSTRG.5584 junction alternative investigated here. Independently, GL0102 long-read transcript resources permit a restricted external check of splice-junction architecture (Wang et al., 2024).

A putative carnitine acetyltransferase-like locus represents an interesting target because alternative transcript structure could change predicted open reading frames. However, fungal carnitine acetyltransferases may carry multiple localization determinants, and translation and compartmental targeting must not be inferred from computational transcript sequence alone (Hynes et al., 2011).

The present secondary analysis asks three narrow questions: whether two candidate MSTRG.5584 junction paths are supported by splice-spanning reads across genomic backgrounds; whether their observed

proportions differ consistently between the two source tissue labels under explicit depth and statistical limitations; and whether an independent Ganoderma genome and long-read library contain corresponding junction evidence. The analysis was deliberately designed to separate descriptive conservation, exploratory tissue association and functional inference.

2. Materials and methods

2.1. Public accessions and sample provenance

The primary dataset was obtained from the public Ganoderma pan-genome study, BioProject PRJNA992876 (Yu et al., 2024). Forty runs were assigned to five genomic-background comparisons, each with four library labels per tissue. Genomic-background identifiers used here are S1, S2, S3, S4 and S6; these are distinct from source RNA sample-group labels S1–S10. An aggregate design summary is provided in Supplementary Table S1; the complete run-to-group, tissue and BioSample mapping is deposited as a machine-readable accession crosswalk in Zenodo (https://doi.org/10.5281/zenodo.23689684). Original Table S20 filename labels and ENA sample accession metadata were cross-checked, identifying no mismatched SRR assignments in 40 runs. The source publication describes four biological replicates per condition, although one BioSample groups each four-library condition and an extraction-by-extraction specimen record was unavailable. Accordingly, test statistics that require independence are conditional on the original experimental description.

2.2. Candidate transcript and reference structure

The two computational transcript models MSTRG5584_A and MSTRG5584_B were reconstructed and aligned. Model A is 2,932 nucleotides and model B 2,377 nucleotides, with an A-specific 555-nucleotide segment bounded by otherwise identical aligned transcript regions. Open reading frames were evaluated in silico by searching candidate ATG start sites and first in-frame stop codons. This analysis is exploratory: no ribosome profiling, proteomics or organellar localization data were available.

2.3. Observed-read mapping and junction counting

Primary short-read evidence was collected by strain-specific RNA-to-genome alignment with HISAT2 (Kim et al., 2019). Spliced primary alignments were filtered for mapping quality at least 20, unique alignment when the NH tag was present, and at least 15 nucleotides of contiguous aligned sequence flanking each splice junction. Junction A and B counts were calculated per SRR library. An initial bounded approach used nonrandom read-1 FASTQ prefixes, typically up to five million reads; following forensic consistency checks, 11 libraries were updated by recounting complete read-1 files containing approximately 21 million reads each. Original S1 data came from the earlier discovery analysis, for which uniform full-depth processing was not established here. These strata are therefore not uniform in read depth or paired-end treatment.

The junction-A fraction was computed as A/(A+B) for libraries with at least one focal supporting read, retaining zero-support libraries explicitly as nonevaluable. A predefined adequacy criterion required A+B at least 10 for at least three libraries in each tissue for a given genomic background. This is an analytical coverage threshold, not a test of biological significance.

2.4. Internal comparisons and statistical sensitivity analysis

S1 served as the discovery background and S2, S3, S4 and S6 as internal evaluation backgrounds from the same primary study. A background-specific difference was calculated from the mean library-level A fraction in mycelium minus that in fruiting-body libraries. The exploratory held-out statistic averaged this difference across backgrounds, including only libraries meeting the ten-read depth criterion for the permutation analysis. Within-background label permutations were combined with 100,000 Monte Carlo draws using a fixed seed (55842026); an equal-background-weight exact sign test and leave-one-background-out analyses were also performed. The original event had been selected after genome-wide screening, so internal held-out permutation P values are not replacement multiple-testing-adjusted P values. The original approximate Benjamini–Hochberg FDR of 0.136 is preserved.

2.5. Independent long-read genomic and junction assessment

A 16.7-kb S1 reference interval spanning the candidate locus was aligned to Ganoderma lingzhi GL0102 genome assembly GCA_047651895.1 (Wang et al., 2024; BioProject PRJNA1177388) using minimap2 (Li, 2018). The aligned region was used to map the two junction coordinate pairs into the independent genome. One external PacBio-derived library (SRR31124588, labeled GL0102_Pe_isoseq) was tested by splice-aware minimap2 alignment (Li, 2018). A supporting junction was counted only from a primary mapping with MAPQ at least 20 and aligned blocks of at least 15 bases on both sides; agreement within two reference bases was permitted. A strand check was applied when an XS tag was present; absent strand tags were not independent proof of transcript orientation. Different developmental sampling and limited evidence about full-length molecular integrity precluded an independent tissue-difference test.

2.6. Computational reproducibility and AI-assisted scripting

The derived junction counts, accession crosswalk, computational sequence reconstructions, and statistical outputs are detailed in the accompanying Supporting Information and cited at https://doi.org/10.5281/zenodo.23689684. OpenAI ChatGPT assisted with the development and review of some analysis and packaging scripts. The deposited files are intended to allow independent inspection and reproduction of the summarized computations; the current manuscript does not claim that every raw public library was uniformly reprocessed. Biological interpretation and source verification remain the responsibility of the author.

3. Results

3.1. The junction pair and proposed coding consequence

The reconstructed candidate models differ by a single 555-nt A-specific sequence, and both retained an otherwise matching transcript context. Model B contained a candidate ORF of 642 amino acids from its annotated start. From the shared upstream start in model A, the first predicted in-frame stop follows 183 amino acids, whereas a downstream initiation candidate produces a 623-amino-acid predicted ORF. These are alternative computational start/stop scenarios, not measured translation products.

3.2. Tissue-labeled junction differences were same-direction across backgrounds

Both candidate junction paths were detected in all five genomic-background comparisons. Among libraries with at least one focal supporting read, the mean proportion of A was descriptively greater in mycelium-labeled groups for each background (Table 1; Figure 1). Individual run counts, including zero-support observations and heterogeneous sequencing depths, are available in the accession-resolved deposit. After targeted complete-read recounting, all four internal evaluation backgrounds met the predefined minimum of three depth-adequate libraries per tissue. This coverage gate supports descriptive comparison; it does not establish standardized sequencing depth, biological replication, or normalized transcript expression.

Table 1. Descriptive mean junction-A fractions by genomic background (libraries with A + B > 0); the N ≥ 10 columns indicate the number of libraries passing the coverage gate in each tissue.

Background	MY mean A	FR mean A	Difference	MY ≥10	FR ≥10
S1	0.765	0.325	+0.440	4	4
S2	0.821	0.366	+0.455	3	3
S3	0.640	0.421	+0.219	3	3
S4	0.752	0.562	+0.190	3	3
S6	0.686	0.151	+0.535	3	3

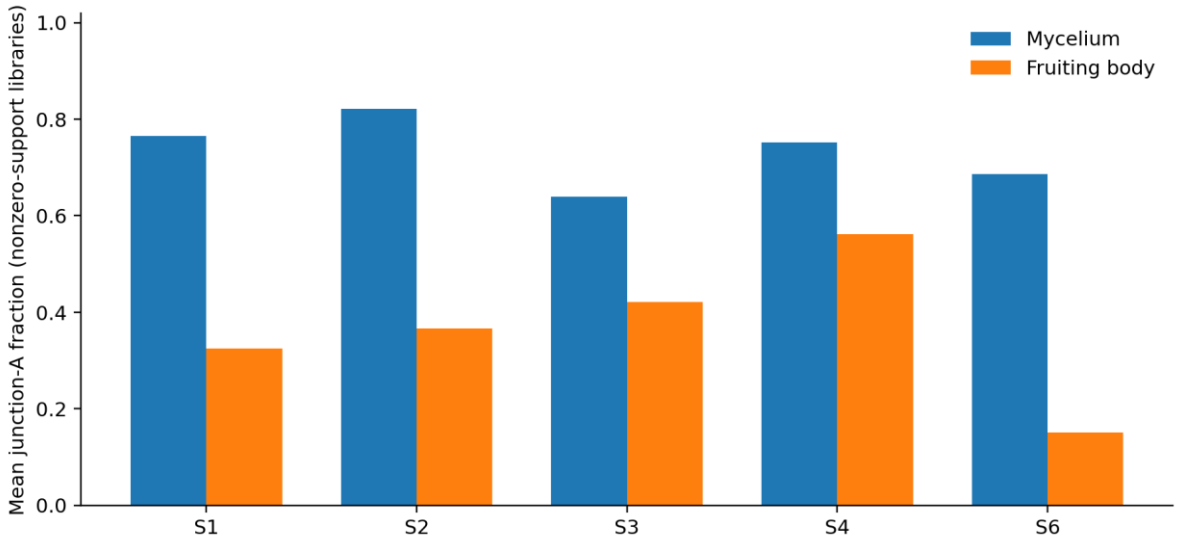


Figure 1. Descriptive mean library-level junction-A fraction by background and source tissue label. Mixed sampling depths and unresolved specimen independence preclude confirmatory inference.

3.3. Exploratory statistics do not establish post-selection significance

For S2, S3, S4 and S6, the mean background-level MY-minus-FR junction-A fractions after imposing the depth threshold were 0.444, 0.148, 0.161 and 0.485, respectively. The equal-background mean difference was 0.309. The two-sided library-label Monte Carlo permutation estimate was P=0.00083, conditional on the source study's biological-replicate design. In contrast, the four-background exact two-sided sign test gave P=0.125, and the original genome-wide discovery screen had approximate FDR q=0.136. The results therefore do not support a confirmatory claim of differential splice regulation. The small number of distinct backgrounds and single-study origin of all short-read observations remain central limitations.

3.4. External genomic match and candidate PacBio junction support

The S1 candidate interval aligned to a 16,752-nt region of the GL0102 assembly with 16,668 matching aligned positions (approximate 99.5% identity; alignment coverage of the queried interval reported as 100%). Two mapped junction paths were located in the corresponding region. In GL0102 primordium-associated PacBio-derived run SRR31124588, the available computational recount reported eight A-like and six B-like splice-spanning read identifiers under its mapping and anchor filters. The external specimen was not a matched mature-fruiting-body comparison, and read orientation and full-length molecule status were not independently proven. This is orthogonal candidate junction support rather than independent reproduction of tissue-associated regulation.

Table 2. Evidence hierarchy and scope of inference.

Evidence	Observation	Permissible inference
Internal RNA-seq	40 runs; 4/4 held-out depth gates	Junction paths observed; exploratory group contrast
Source design	40 accession matches; 10 BioSamples	Published replicate labels supported, not specimen identity
External GL0102 genome	16.7-kb mapping, ~99.5% identity	Candidate orthologous region
External PacBio	8 A-like; 6 B-like read identifiers	Candidate junction corroboration, not expression replication
Statistics	Held-out sign P=0.125; original FDR q~0.136	Confirmatory tissue-regulation claim not established

4. Discussion

The principal observation is repeated read-level detection of two candidate splice-junction paths across five Ganoderma genomic backgrounds. A candidate matching genomic interval and splice-spanning long-read identifiers in the separately generated GL0102 resource offer an additional sequence-based check. The

same-direction tissue-label differences remain exploratory: they are not an independently replicated demonstration of developmental regulation, and the statistical conclusions change when genomic background rather than individual run label is taken as the inferential unit.

Functional interpretations of the predicted transcript differences must remain constrained. An A-specific 555-nucleotide segment alters an in silico initiation and stop-codon scenario, but there is no evidence here for translation from either alternative initiation site, transcript stability, organelle targeting, enzymatic activity or phenotype. Prior experimental reports of splice-sensitive functions in Ganoderma demonstrate why this distinction matters (Lian et al., 2021; Liu et al., 2023; Wang et al., 2025). Future work using targeted RT-PCR, independent sample extraction, full-length transcript validation and localized protein assays could distinguish a genuine regulated isoform switch from alignment or expression-context effects.

This work also illustrates inferential difficulties in targeted reanalysis of public fungal RNA-seq. Four nominal biological replicates per condition are reported by the primary publication, and run-level accession and label checks were consistent. Nevertheless, each four-run group shares one BioSample, so this study cannot independently verify physical specimen independence. Most libraries were assayed as nonrandom prefixes of read 1, whereas a selected subset was rescued at full R1 depth to reach an event-level coverage threshold. The cross-background contrasts are therefore not a harmonized genome-wide differential splicing experiment. Positive results of exploratory label-shuffling cannot correct initial candidate-selection bias, and the relatively large cluster-level sign-test P value demonstrates that the inference changes with statistical unit.

The biological reach of this result is deliberately narrower than that of experimental Ganoderma splicing papers. Its added value lies in the accession-level reconstruction and audit trail across five backgrounds and the opportunity for independent sequence-based confirmation; the interpretation remains hypothesis-generating. Further technical reprocessing without independent well-characterized biological specimens would not, by itself, solve the limitations exposed by the present analysis.

5. Conclusions

A targeted public-data reanalysis documents two candidate alternative splice-junction paths at the putative Ganoderma MSTRG.5584 carnitine acetyltransferase-like locus, together with a closely matching external genomic interval and putative long-read junction corroboration. These findings supply accession-traceable evidence for a testable transcript-architecture hypothesis. They do not establish regulated tissue-specific splicing, stable protein products, altered subcellular localization, enzymatic activity, or phenotype. Independent specimen-resolved and full-length-transcript experiments would be needed to test those possibilities.

Declarations

Author contributions: Kiarash Sadeghian Esfahani — conceptualization, computational investigation, data curation, analysis, interpretation, and manuscript approval. The sole author is responsible for verification of reported analyses.

Funding: No specific funding was reported for this independent computational reanalysis.

Potential competing interests: No competing interests to declare.

Ethics statement: This secondary analysis used publicly available fungal genomic and transcriptomic sequence data; no human participants or animal experiments were involved.

Data availability: The author-derived junction-count table, accession crosswalk, computational transcript reconstructions, statistical summaries and associated reproducibility materials are publicly available in Zenodo, version 1: https://doi.org/10.5281/zenodo.23689684. The underlying original sequence reads and genomes are maintained by their depositors under BioProjects PRJNA992876 and PRJNA1177388; original FASTQ files were not republished in the Zenodo record.

Code and computational materials: Derived processing files, scripts, accession crosswalks and candidate sequence reconstructions are described in the linked Zenodo data record. The original read datasets remain with their source repositories.

Consent for publication: Not applicable; no identifiable human participant data were used.

Generative AI and AI-assisted technologies disclosure

OpenAI ChatGPT was used to assist manuscript organization and language editing, and to support development, review, and packaging of some computational analysis scripts. Its output was not treated as experimental evidence. The sole human author is responsible for the accuracy and integrity of the analysis, claims, reference checks, and submitted content.

References

Chen, S., Xu, J., Liu, C., Zhu, Y., Nelson, D. R., Zhou, S., et al. (2012). Genome sequence of the model medicinal mushroom Ganoderma lucidum. Nature Communications, 3, 913. https://doi.org/10.1038/ncomms1923

Hynes, M. J., Murray, S. L., Andrianopoulos, A., & Davis, M. A. (2011). Role of carnitine acetyltransferases in acetyl coenzyme A metabolism in Aspergillus nidulans. Eukaryotic Cell, 10(4), 547–555. https://doi.org/10.1128/EC.00295-10

Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnology, 37(8), 907–915. https://doi.org/10.1038/s41587-019-0201-4

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Lian, L., Zhang, G., Zhu, J., Wang, Y., Wang, L., Liu, R., Shi, L., Ren, A., & Zhao, M. (2021). Swi6B, an alternative splicing isoform of Swi6, mediates the cell wall integrity of Ganoderma lucidum. Environmental Microbiology, 23(8), 4405–4417. https://doi.org/10.1111/1462-2920.15627

Liu, R., Yang, Z., Yang, T., Wang, Z., Chen, X., Zhu, J., Ren, A., Shi, L., Yu, H., & Zhao, M. (2023). PRMT5 regulates the polysaccharide content by controlling the splicing of thaumatin-like protein in Ganoderma lucidum. Microbiology Spectrum, 11(6), e02906-23. https://doi.org/10.1128/spectrum.02906-23

Shao, J., Chen, H., Yang, D., Jiang, M., Zhang, H., Wu, B., Li, J., Yuan, L., & Liu, C. (2017). Genome-wide identification and characterization of natural antisense transcripts by strand-specific RNA sequencing in Ganoderma lucidum. Scientific Reports, 7, 5711. https://doi.org/10.1038/s41598-017-04303-6

Wang, L., Shi, P., Ping, Z., Huang, Q., Jiang, L., Ma, N., Wang, Q., Xu, J., Zou, Y., & Huang, Z. (2024). The golden genome annotation of Ganoderma lingzhi reveals a more complex scenario of eukaryotic gene structure and transcription activity. BMC Biology, 22, 271. https://doi.org/10.1186/s12915-024-02073-y

Wang, L., Li, Z., Ping, Z., Dai, L., Wang, Q., & Zou, Y. (2025). Alternative splicing of Gllac7 regulate lignin degradation in Ganoderma lucidum. International Journal of Biological Macromolecules, 309, 143086. https://doi.org/10.1016/j.ijbiomac.2025.143086

Yu, H., Wang, S., Wang, L., Wu, W., Xu, W., Wu, S., Li, X., Xu, W., Huang, Z., Lin, Y., & Wang, H. (2024). Pan-genomic characterization and structural variant analysis reveal insights into spore development and species diversity in Ganoderma. Microbial Genomics, 10(11), 001328. https://doi.org/10.1099/mgen.0.001328

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- GanodermaSupplementaryInformation.docx