Uncovering Microbial Drivers of Saponin Production in *Abrus cantoniensis*

Project Summary

Abrus cantoniensis ("Herba Abri") is a medicinal plant known for producing hepatoprotective triterpenoid saponins, yet its metabolite yield varies across environments. This project investigates whether heat-driven shifts in the plant's endophytic microbiome influence biosynthetic gene expression and saponin output. The project will (i) use metagenomic 16S rRNA sequencing to profile microbial diversity and strain-level structure across a natural temperature gradient, (ii) apply metatranscriptomics and LC-MS/MS to link endophyte gene expression to metabolite production, and (iii) integrate multi-omic data using statistical tools to identify heat-stable biosynthetic gene clusters (BGCs). The top microbial strains and BGCs will be prioritised for downstream synthetic biology applications. This approach addresses major challenges in environmental genomics including uncharacterised microbiomes, functional annotation of BGCs, and linking genetic variation to ecological function.

Research Aims

This proposal hypothesizes that differences in hepatoprotective saponin production in *Abrus cantoniensis* are influenced not only by host genetics but also by differences in its endophytic microbiome and its biosynthetic activity under environmental heat stress at the strain level. This project aims to identify which microbial biosynthetic gene clusters are heat-responsive and most closely associated with saponin production by integrating metagenomics, metatranscriptomics, and metabolomics data under natural temperature gradients. By pinpointing functional microbial partners, this research will improve the consistency of herbal medicine quality and support the development of synthetic biology tools for sustainable phytopharmaceutical production.

Objectives

1. Profile bacterial endophyte community composition in root, stem, leaf, and seed tissues of *A. cantoniensis* across a natural coastal heat gradient (27–33 °C) using 16S rRNA gene sequencing and QIIME2 analysis.
2. Quantify microbial biosynthetic gene cluster (BGC) expression and triterpenoid saponin levels in the same tissues using rRNA-depleted RNA-seq and targeted LC-MS/MS.
3. Integrate microbial, transcriptomic, and metabolomic data using redundancy analysis, partial Mantel tests, and random forest regression to identify and prioritise heat-responsive BGCs that are predictive of high saponin yield.

Background & Significance

A. cantoniensis ("Herba Abri") is a traditional medicine in Cantonese and Southeast Asian systems, widely used for its hepatoprotective and anti-inflammatory properties. Liang et al. (2023) demonstrated that host genetic variation affects flavonoid profiles in *Abrus* species, revealing differences in vicenin/schaftoside isomers and immune transcriptomic responses between *A. cantoniensis* and *A. mollis*. However, their study did not explore how environmental factors such as heat may drive chemical diversity in the plant's microbiome. In addition, Wang et al. (2022) identified three active metabolites: β-sitosterol, stigmasterol, and butin which act as neuroprotectants and regulate epilepsy-related pathways through multi-target interactions, including oxidative stress and apoptosis. This highlights the diverse bioactivity of *A. cantoniensis* metabolites and also the potential impact of environmental or microbial factors on their abundance. Xu et al. (2024) demonstrated the hepatoprotective function of *A. cantoniensis* extracts in a CCl_4-induced rat liver injury model, where saponin-rich extracts significantly reduced oxidative damage, inflammation, and cell death. Together, these studies highlight the importance of host genetics and bioactive metabolite diversity in *A. cantoniensis*. My proposal builds directly on this foundation by investigating whether heat-driven shifts in the endophytic microbiome modulate microbial BGC expression and the plant's pool of hepatoprotective triterpenoid saponins.

Recent studies have highlighted the importance of endophytes in shaping the pharmacological properties of medicinal plants. Sharma et al. (2023) reviewed how microbial biosynthetic gene clusters adapt to environmental stress and regulate flavonoid synthesis through host-mirror and host-independent pathways. Duhan et al. (2020) further demonstrated that endophytes isolated from *Tinospora cordifolia* exhibited plant-beneficial properties such as phosphate solubilization and antimicrobial activity, suggesting that they play a role in supporting host health and promoting metabolite biosynthesis. Similarly, Sharma et al. (2024) isolated endophytes from *Momordica charantia* with cytotoxic activity against lung cancer cells and attributed these effects to strain-specific secondary metabolites, highlighting the need to explore the genetic basis of microbial bioactivity through BGCs. Complementing these findings, Rieusset et al. (2020) found that *Pseudomonas* spp overproduced secondary metabolites in biofilm mode, highlighting how environmental and structural factors affect the expression of BGCs. Taken together, these studies support the core hypothesis of this project: thermally driven turnover of the endophytic microbiome of *Abrus cantoniensis* may regulate the expression of microbial BGCs and modulate the production of hepatoprotective triterpenoid saponins.

Environmental temperature has been identified as a key factor influencing the composition and function of microbial communities in different ecosystems, providing ecological support for the proposed thermal gradient approach. Liu et al. (2022) demonstrated that high temperature and waterlogging significantly altered the rice rhizosphere microbiome, enriching nitrogen-transforming taxa while reducing functional complexity. This reveals how abiotic stresses reprogram the role of microorganisms in plant health. Similarly, Podar et al. (2020) revealed that thermal gradients in Iceland hot springs led to a sharp decline in microbial diversity and the emergence of thermophilic taxa with unique biosynthetic properties, emphasizing that temperature is a selective force driving microbial turnover and functional specialization. Complementing these patterns, Ma et al. (2022) showed that fine-scale microclimate changes along an elevation gradient in a subtropical forest strongly affected bacterial and fungal diversity, which correlated with changes in soil properties and plant traits. Taken together, these studies highlight that increasing temperatures can drive changes in microbial community assembly and biosynthetic gene expression. Using the example of *Abrus cantoniensis*, this strengthens the theoretical basis for studying how heat-driven endophyte turnover affects the expression of microbial biosynthetic gene clusters and the production of hepatoprotective triterpenoid saponins in plants.

Environmental gradients profoundly influence the composition, structure, and function of endophyte communities, making endophytes both indicators and mediators of plant stress responses. Zhang et al. (2025) demonstrated that the microbial community associated with the endosperm of *Pinus armandi* changes predictably with climate change, with specific taxa enriched at higher altitudes contributing to host resilience. Similarly, Giauque and Hawkes (2013) showed that endophytic fungi from hotter, drier environments conferred greater drought tolerance to *Panicum hallii*, highlighting the adaptive value of climate-selected microbial communities. Dastogeer et al. (2022) further reinforced this framework by demonstrating how endophytes mediate thermotolerance by upregulating antioxidant activity, modulating phytohormones, and altering secondary metabolites. Taken together, these studies demonstrate the ecological rationale for combining thermal gradient sampling with multi-omics to reveal how thermoresponsive endophytes influence the medicinal chemistry of *A. cantoniensis*.

Objective 1 – Genome census along the heat gradient

Experiment 1 aims to investigate whether rising temperatures will affect the composition of the endophytic microbial community in *Abrus cantoniensis*. By sampling across a well-defined coastal heat gradient (27-33°C), this experiment will determine whether specific microbes are more adapted to high temperatures and may be involved in regulating the biosynthesis of the plant's medicinal chemical constituents, particularly the hepatoprotective triterpenoid saponins.

The samples will be collected from roots, stems, leaves and seeds of *A. cantoniensis* grown in three locations spanning a natural temperature gradient with mean annual temperatures of 27°C, 30°C and 33°C. Three plant samples will be collected from each location and all tissues will be surface sterilized for endophyte isolation. DNA will be extracted from each sample and sequenced using 16S rRNA gene sequencing to analyze bacterial diversity. This method targets conserved and variable regions of the 16S ribosomal RNA gene, allowing for genus-level taxonomic identification. This method is based on the work of Ingrey et al. (2021), who used 16S rRNA sequencing to identify endophytes in Australian bush medicines. This study will use the same method to track how temperature affects the microbial diversity and potential functions of *A. cantoniensis* on a quantitative thermal gradient.

The resulting sequencing data will be processed using the QIIME2 bioinformatics pipeline. Following the methods of Dumigan and Deyholos (2024), sequences will be quality filtered, denoised, and classified using a classifier trained on a reference 16S rRNA gene database to assign taxonomy to unknown microbial sequences. Community composition will be compared under different temperature conditions using the Shannon diversity index and non-metric multidimensional scaling (NMDS) plots based on Bray-Curtis dissimilarity. Shannon index values help determine whether microbial diversity increases or decreases with temperature. The higher the Shannon index, the more diverse the microbial community. Cullen et al. (2023) demonstrated the effectiveness of this metric in detecting exercise-driven microbial changes in animal gut studies and highlighted its ecological sensitivity. NMDS plots used by Hou et al. (2020) can visualize how microbial communities cluster with temperature conditions, revealing broader compositional shifts. Taken together, these analyses will reveal how heat affects endophyte diversity and help identify microbes that may be involved in regulating the medicinal properties of plants.

If microbial community composition changes significantly along the temperature gradient, as reflected by changes in Shannon diversity and distinct clustering features on the NMDS plot, it supports the hypothesis that heat affects the structure of the endophytic bacterial community in

Abrus cantoniensis. Specifically, increases or decreases in diversity or enrichment of specific taxa in warmer regions may indicate the presence of heat-tolerant microorganisms that are involved in regulating the medicinal chemical properties of *Abrus cantoniensis*. These patterns are useful for further investigation into whether these microorganisms carry biosynthetic gene clusters (BGCs) associated with saponin production, thereby linking environmental temperature to the functional microorganism's function.

If the microbial composition remains stable across the temperature gradient, this may indicate that temperature alone is not the only factor driving endophyte turnover. In this case, functional changes (e.g., differential gene expression) rather than structural shifts could explain the variability in saponins, supporting the need for metatranscriptome analysis in Objective 2.

Objective 2 – Metatranscriptomic Profiling of Matched Tissues

Experiment 2 will determine whether elevated temperatures affect the functional activity of endophytes by measuring expression levels of microbial BGCs and saponin production in the same tissues. This experiment tests whether changes in gene expression, rather than community composition alone, can explain high-temperature-induced changes in medicinal compound production and plant stress responses.

Using root, stem, leaf, and seed tissues collected in Experiment 1, the experirment will extract total RNA from *Abrus cantoniensis* samples using the SDS-LiCl method developed by Vennapusa et al. (2020). This protocol is optimized for high-protein plant tissues and provides consistent yields of high-integrity RNA suitable for transcriptomic applications. RNA will be DNase treated to remove contaminating DNA, quality checked and enriched for messenger RNA using the Ribo-Zero Meta Ribosomal RNA Removal Kit. Then, construct strand-specific RNA sequencing libraries and sequence them on the Illumina NovaSeq platform, targeting approximately 60 million paired-end reads per sample. A similar workflow was used by Roy et al. (2024) in Hibiscus hibiscus, where rRNA-depleted RNA sequencing enabled detailed analysis of the diversity of viral transcripts, highlighting the applicability of this approach to complex plant microbiome systems.

The resulting reads will be aligned to publicly available bacterial reference genomes and representative endophyte sequences using Bowtie2, as demonstrated by Azizpour et al. (2022) to accurately distinguish host and microbial reads in plant microbiome studies. Transcript abundance will be quantified as transcripts per million (TPM) using featureCounts. Expression values will

focus on BGCs, especially those related to triterpene and saponin biosynthesis. In parallel, triterpene saponins will be extracted from the same tissues using 85% methanol and analyzed by liquid chromatography tandem mass spectrometry (LC-MS/MS), measuring their relative concentrations under a heat gradient.
This approach follows the protocol of Khakimov et al. (2016) and allows for precise identification of saponins based on aglycone mass and sugar composition. These paired transcriptomic and metabolomic datasets will form the basis for Objective 3, which is to correlate microbial gene expression patterns with saponin production to identify candidate functional groups that are heat sensitive.

If microbial BGCs involved in triterpenoid and saponin biosynthesis have increased expression (TPM) in samples from warmer regions and these samples also exhibit higher saponin concentrations, it would support the hypothesis that heat affects microbial activity and metabolite production in *Abrus cantoniensis*. The strong correlation between BGC expression and saponin levels suggests that endophytic microorganisms contribute functionally to the medicinal chemistry of plants under heat stress. This suggests that temperature-driven microbiome activity, rather than composition alone, plays a key role in regulating the production of bioactive compounds.

If no clear correlation is found between BGC expression and saponin levels, it may indicate that saponin biosynthesis is primarily plant-driven or post-transcriptionally regulated. Alternatively, important microbial contributors may be low in abundance or missed in the reference genome map. Follow-up studies could include untargeted metabolomics or plant transcriptome analysis.

Objective 3 – Integration of Microbiome, Expression, and Chemistry Data

Experiment 3 will identify which microbial biosynthetic gene clusters are most closely associated with the production of saponins in *Abrus cantoniensis*. By integrating microbial and metatranscriptome RNA sequencing data and LC-MS/MS data, this experiment will test whether specific endophytic functions can explain the changes in the levels of bioactive compounds under high temperature stress and ultimately identify candidate BGCs for probiotic or synthetic biology applications.

Data from Experiments 1 and 2, including microbial composition (16S rRNA), biosynthetic gene cluster expression, and saponin concentrations, will be integrated through multivariate analysis. First, redundancy analysis (RDA) will assess how much of the variation in saponin levels across the temperature gradient can be explained by microbial BGC expression. Hou et al. (2022) effectively used RDA to link fatty acid traits to steroidal saponin accumulation in *Dioscorea zingiberensis*, demonstrating its utility in revealing gene-metabolite relationships. Similarly, this study will apply RDA to determine which microbial BGCs are most closely associated with saponin biosynthetic functions in *Abrus cantoniensis* under heat stress.

Next, partial Mantel tests will be used to determine whether microbial community composition (quantified by Bray-Curtis dissimilarities) is significantly associated with saponin profiles while controlling for geographic location. This helps to distinguish functional associations from spatial confounding factors. Bray-Curtis dissimilarities quantify differences in community composition based on the presence and abundance of taxa. Both Sun et al. (2013) and Yang et al. (2019) applied partial Mantel tests to distinguish between the roles of environmental and geographic factors in shaping microbial communities. Their studies support this approach to test whether community-level variation is associated with saponin variation in *A. cantoniensis*.

Finally, random forest regression will be used to rank individual BGCs by their predictive power for saponin concentrations. The top-ranked BGCs are those that are consistently associated with elevated saponin levels in heat-stressed tissues and will be cross-referenced with taxonomic and functional annotations. Xia et al. (2024) investigated the effects of soil microbial communities on the accumulation of saponins by comparing Sanqi ginseng grown in soils from geographic origin and non-origin soils. Using random forest regression analysis, the authors identified key bacterial taxa whose abundance was strongly associated with elevated saponin levels. This study directly supports the planned random forest regression approach to rank microbial BGCs based on their predictive ability for saponin production in *Abrus cantoniensis*.. 3-4 heat-responsive BGCs will be

prioritized for downstream validation because of their consistent correlation with increased saponin production in heat-stressed tissues.

If specific microbial BGCs are strongly associated with elevated saponin levels identified by redundancy analysis and ranked by random forest regression, it would support the hypothesis that endophytic gene expression changes driven by high temperature affect the production of medicinal compounds in *Abrus cantoniensis*. Consistent associations between certain BGCs and high saponin production in multiple tissues and under high temperature conditions would suggest a role for functional microorganisms. If partial Mantel tests confirm that these associations remain significant after controlling for geographic location, it would strengthen the conclusion that microbial activity, not just plant genetics or location, is driving variation in bioactive metabolites.

If no strong associations were found, saponin biosynthesis may be primarily regulated by the plant or influenced by unmeasured abiotic factors. Future studies could incorporate untargeted metabolomics or host transcriptome analyses to capture more drivers. Furthermore, this framework could be extended to test microbiome-metabolite relationships under salt stress, which would require more sophisticated ecophysiological designs and provide a natural extension for translating heat-stable microbial functions into broader climate resilience applications.

Azizpour, N. *et al.* (2022) ‘Identification of endophytic microbiota of Phytoplasma-infected Russian olive trees “Elaeagnus angustifolia l.” in the northwest of Iran’, *Forests*, 13(10), p. 1684.

Cullen, J.M., Shahzad, S. and Dhillon, J. (2023) ‘A systematic review on the effects of exercise on gut microbial diversity, taxonomic composition, and microbial metabolites: Identifying research gaps and future directions’, *Frontiers in Physiology*, 14.

Dastogeer, K.M. *et al.* (2022) ‘Microbe-mediated thermotolerance in plants and pertinent mechanisms- a meta-analysis and Review’, *Frontiers in Microbiology*, 13.

Duhan, P., Bansal, P. and Rani, S. (2020) ‘Isolation, identification and characterization of endophytic bacteria from medicinal plant Tinospora cordifolia’, *South African Journal of Botany*, 134, pp. 43–49.

Dumigan, C.R. and Deyholos, M.K. (2024) ‘Soil and seed both influence bacterial diversity in the microbiome of the Cannabis sativa seedling endosphere’, *Frontiers in Plant Science*, 15.

Giauque, H. and Hawkes, C.V. (2013) ‘Climate affects symbiotic fungal endophyte diversity and performance’, *American Journal of Botany*, 100(7), pp. 1435–1444.

Hou, L. *et al.* (2022) ‘Effect of exogenous jasmonic acid on physiology and steroidal saponin accumulation in dioscorea zingiberensis’, *Plant Physiology and Biochemistry*, 186, pp. 1–10.

Hou, W. *et al.* (2020) ‘Epichloë gansuensis endophyte-infection alters soil enzymes activity and soil nutrients at different growth stages of Achnatherum Inebrians’, *Plant and Soil*, 455(1–2), pp. 227–240.

Ingrey, S.D. *et al.* (2021) ‘Australian bush medicines harbour diverse microbial endophytes with broad-spectrum antibacterial activity’, *Journal of Applied Microbiology*, 131(5), pp. 2244–2256.

Khakimov, B. *et al.* (2016) *Screening for triterpenoid saponins in plants using hyphenated analytical platforms* [Preprint].

Liang, Z. *et al.* (2023) ‘A comparative study between *abrus cantoniensis* and *abrus mollis*: Flavonoids profiles by uplc-q-tof-ms and anti-inflammatory mechanism by Transcriptomics’, *Chemistry & Biodiversity*, 20(6).

Liu, J. *et al.* (2022) ‘Impact of waterlogging and heat stress on Rice Rhizosphere microbiome assembly and potential function in carbon and nitrogen transformation’, *Archives of Agronomy and Soil Science*, 69(10), pp. 1920–1932.

Ma, L. *et al.* (2022) ‘When microclimates meet soil microbes: Temperature controls soil microbial diversity along an elevational gradient in subtropical forests’, *Soil Biology and Biochemistry*, 166, p. 108566.

Podar, P.T. *et al.* (2020) ‘Comparative analysis of microbial diversity across temperature gradients in hot springs from Yellowstone and Iceland’, *Frontiers in Microbiology*, 11.

Rieusset, L. *et al.* (2020) ‘Secondary metabolites from plant-associated *pseudomonas* are overproduced in biofilm’, *Microbial Biotechnology*, 13(5), pp. 1562–1580.

Roy, A. *et al.* (2024) ‘Meta-transcriptomic analysis uncovers the presence of four novel viruses and multiple known virus genera in a single hibiscus rosa-sinensis plant in Colombia’, *Viruses*, 16(2), p. 267..

Sharma, G. *et al.* (2023) ‘Therapeutic compounds from medicinal plant endophytes: Molecular and metabolic adaptations’, *Journal of Applied Microbiology*, 134(4).

Sharma, G. *et al.* (2024) ‘Host-specific endophytes of Momordica Charantia: A promising source for affordable lung cancer therapeutics’, *South African Journal of Botany*, 170, pp. 181–193.

Sun, X., Gao, C. and Guo, L. (2013) ‘Changes in soil microbial community and enzyme activity along an exotic plant eupatorium adenophorum invasion in a chinese secondary forest’, *Chinese Science Bulletin*, 58(33), pp. 4101–4108.

Vennapusa, A.R. *et al.* (2020) ‘A universal method for high-quality RNA extraction from plant tissues rich in starch, proteins and fiber’, *Scientific Reports*, 10(1).

Wang, Y. *et al.* (2022) ‘Antiepileptic therapy of abrus cantoniensis: Evidence from Network Pharmacology’, *Evidence-Based Complementary and Alternative Medicine*, 2022, pp. 1–12.

Xia, Q. *et al.* (2024) ‘Deciphering the key soil microbial taxa that contribute to saponin accumulation in a geo-authentic sanqi ginseng production area: Evidence from four different varieties’, *Applied Soil Ecology*, 201, p. 105470.

Xu, C. *et al.* (2024) ‘Protective effects of abrus cantoniensis hance against liver injury through modulation of intestinal microbiota and liver metabolites’, *Ecotoxicology and Environmental Safety*, 279, p. 116495.

Yang, J. *et al.* (2019) ‘A comprehensive census of Lake Microbial Diversity on a global scale’, *Science China Life Sciences*, 62(10), pp. 1320–1331.

Zhang, W.-D. *et al.* (2025) ‘Decoding endosperm endophytes in pinus armandi: A crucial indicator for host response to climate change’, *BMC Microbiology*, 25(1).