

CT-Metabolome Canalization in Coconut

Quantitative evidence for developmental canalization: CT-metabolome integration reveals a structural-metabolic compensation axis in coconut fruit development

Running Title: CT-metabolome canalization in coconut

Ninghuan You^{12†}, Huangsheng Ling^{3†}, Jing Chen³, Wenrao Li¹, Mengxing Huang⁴, Yu Zhang⁵, Hongxing Cao¹, Chengxu Sun¹

¹ Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences, Wenchang, Hainan 571339, China ² School of Life Sciences, Henan University, Kaifeng, Henan 475004, China ³ Haikou Affiliated Hospital of Central South University Xiangya School of Medicine, Haikou, Hainan 570208, China ⁴ College of Information and Communication Engineering, Hainan University, Haikou, Hainan 570228, China ⁵ College of Computer Science and Technology, Hainan University, Haikou, Hainan 570228, China

† Co-first authors | * Corresponding authors: Hongxing Cao (caohx@catas.cn), Chengxu Sun (suncx@catas.cn)

Summary

Coconut is a major oil crop in the tropics, yet the structural-metabolic differences in fruit development across varieties remain poorly understood. Here we integrate CT phenomics (7,686 DICOM slices, five varieties, five developmental stages) with widely targeted LC-MS/MS metabolomics (624 metabolites) to dissect the structural-metabolic divergence between the elite dwarf variety Wenye 5 (W5) and the local tall landrace CK. CT

phenotyping revealed that four tissue-specific CT values (Hounsfield Units) constitute a quantitative developmental clock spanning approximately 1,200 HU. Metabolomic profiling uncovered a reciprocal metabolic pattern: W5 was enriched in flavonoids (26 differential species, all higher in W5, up to 16-fold), whereas CK accumulated amino acids, organic acids and free fatty acids (15 differential species, all higher in CK). This metabolic profile is the inverse of a classical growth–defense trade-off: W5, despite its structural growth investment (thin shell, accelerated endosperm development), maintains elevated chemical defense capacity, suggesting structural-metabolic compensation. Correlation analysis of 18 CT parameters against 624 metabolites identified 3,376 significant associations in W5 (55 pairs with $|r| > 0.99$) versus none in CK. Permutation tests (10,000 shuffles) and robustness checks (Spearman rank correlation, first-differencing Pearson correlation) confirmed that this asymmetry is not attributable to chance or shared developmental trends. We propose that the inter-variety asymmetry in CT-metabolite association strength may reflect differences in developmental canalization associated with breeding-induced genetic homogenization. CT-based germination prediction (AUC = 0.716, $n = 75$) demonstrates the translational potential of this integrative framework for non-destructive seed viability assessment.

Keywords: Coconut | CT phenomics | Metabolomics | Growth–defense compensation | Developmental canalization | Multi-omics integration

Significance Statement

This study establishes coconut—with its unique ~1,200 HU density range spanning four tissue compartments—as an ideal system for CT phenotyping. Integrating 7,686 DICOM slices with 624 metabolites reveals a structural-metabolic compensation axis: the improved dwarf variety W5 invests in structural efficiency (thin shell, early endosperm) but compensates chemically with elevated flavonoid content. The striking asymmetry in CT-metabolite correlation strength (55 pairs with $|r| > 0.99$ in W5 versus none in CK) provides a quantitative statistical signature consistent with developmental canalization, a hypothesis requiring DNA-level validation. This integrative framework provides a quantitative phenotype-genotype connection for assessing developmental uniformity and enables non-destructive germination prediction (AUC = 0.716), with broad applicability to breeding programs.

1 | Introduction

Coconut (*Cocos nucifera* L.) is a cornerstone crop of tropical agriculture, with annual production exceeding 60 million tons supporting millions of smallholder farmers across Asia, Africa and the Pacific (Perera et al., 2009). In Hainan, China, systematic breeding programs have developed the Wenye series of dwarf varieties (Lu and Liu, 2021), selected by breeders over decades from local tall landraces for early maturation, dwarf stature, high yield and thin shell (Li et al., 2020). The structural and biochemical differences in fruit development between these improved varieties and their progenitor landraces have not been systematically characterized.

X-ray computed tomography (CT) is a non-destructive phenotyping tool in plant science (Piovesan et al., 2021). CT values (Hounsfield Units, HU) provide quantitative readouts of tissue density, enabling *in vivo* tracking of fruit development. In horticultural crops, CT has been used to detect internal disorders in pomegranate (Khodaei et al., 2023), evaluate internal quality of durian (Saechua et al., 2021) and assess oil palm fruit maturity (Tarmizi et al., 2021). In coconut, a computational CT phenotyping pipeline has been established covering non-invasive 3D quantitative imaging (Zhang et al., 2023), automated tissue-level segmentation via an improved DeepLabV3+ network (Liu et al., 2023) and clinical CT postprocessing for multi-compartment quantification (Lin et al., 2023). Yu et al. (2022) further integrated Micro-CT with DeepLabV3+ for high-precision segmentation, extracting 78 phenotypic parameters. These foundations have been extended to germination dynamics (Lin et al., 2025), automated sprouting stage classification via few-shot generative models (Mehmood et al., 2026), cross-modal association with lignan metabolites (Sun et al., 2024) and location-specific density variation analysis (Lin et al., 2024). Wang et al. (2024) reviewed the broader application of CT in agricultural non-destructive testing.

Coconut is a favorable system for CT phenotyping. Its fruit comprises four tissue compartments—fiber layer (mesocarp), shell (endocarp), solid endosperm (kernel) and liquid endosperm (coconut water)—spanning a density range from approximately −950 HU (mature fiber) to +256 HU (mineralized shell). This ~1,200 HU dynamic range, comparable to the span between lung tissue and cortical bone in medical CT, allows a single non-destructive scan to track four independent developmental processes simultaneously. No other non-destructive modality—RGB imaging (surface only), near-infrared spectroscopy (3–5 mm penetration), ultrasound (scattered by hard shells) or MRI (weak signal in dry tissues)—can resolve all four compartments.

Parallel advances in metabolomics have provided high-resolution profiles of the chemical dynamics of fruit development. In coconut, Liang et al. (2021) profiled temporal dynamics of fatty acid biosynthesis in the endosperm; Yang et al. (2022) characterized primary and secondary metabolites across eight developmental stages; Zhao et al. (2023) analyzed variety-specific accumulation across eight varieties; and Hou et al. (2024) tracked flavonoid dynamics in coconut water across developmental stages.

The key insight is that CT captures physical outcomes of development (density changes from air-filling, mineralization and lipid deposition), while metabolomics captures the chemical driving forces. The relationship is not coincidental—every HU change results from metabolic activity: programmed cell death and water translocation in the fiber layer, oxalate biosynthesis in the shell, and fatty acid and protein synthesis in the endosperm. By systematically mapping CT-metabolite associations, CT can be transformed from a descriptive tool into a mechanistically interpretable one.

The concept of developmental canalization, first proposed by Waddington (1942), describes the capacity of developmental systems to produce consistent phenotypic outcomes despite genetic or environmental perturbations. This buffering property, often enhanced by genetic homozygosity through directional selection, may leave detectable statistical signatures in population-level phenotypic correlations. While canalization has been extensively studied in model organisms such as *Drosophila* and *Arabidopsis*, its manifestation in fruit tree crops—particularly in the context of breeding-induced genetic homogenization—remains unexplored.

The growth–defense trade-off is a well-established concept in plant biology (Huot et al., 2014), describing resource allocation between biomass accumulation and chemical protection. This has been demonstrated in model species (Züst et al., 2011) and crop domestication (Schauer et al., 2005; Doebley et al., 2006), but how it operates in fruit trees with multi-month developmental timescales remains unclear. Coconut, with its four-layer structure, breeding history and contrasting variety characteristics, provides an opportunity to investigate how artificial selection shapes structural-metabolic relationships.

Seed germination represents another dimension: thin shells selected for ease of kernel extraction may compromise seed viability. CT-based germination prediction has been validated in oil palm (Tarmizi et al., 2021) and reviewed across multiple species (Goodman et al., 2022), but is unexplored in coconut.

Dwarf varieties subjected to generations of directional selection (such as W5) are expected to possess higher genetic homozygosity, which may canalize developmental programs—

buffering trajectories against genetic and environmental variation (Waddington, 1942). This canalization may leave detectable signatures in structural phenotypes and metabolic programs, but its extent has not been quantified in fruit trees.

This study integrates CT phenomics (7,686 DICOM slices, five varieties, five developmental stages) with widely targeted metabolomics (624 metabolites in W5 and CK) to test the hypothesis that breeding-induced genetic homogenization produces detectable statistical signatures in developmental canalization, specifically by comparing CT-metabolite correlation strength between an improved dwarf variety and a genetically heterogeneous landrace. Four specific questions are addressed: (i) Can four tissue-specific CT values constitute a quantitative developmental clock? (ii) Do metabolic differences between W5 and CK exhibit a growth-defense pattern, and if so, how does it map onto structural phenotypes? (iii) Is there asymmetry in CT-metabolite association strength between varieties? (iv) Can CT features be used for non-destructive germination prediction?

2 | Results

2.1 | CT phenotyping reveals variety-specific developmental trajectories

CT phenotyping was performed on 7,686 DICOM slices from five coconut varieties (Wenye 2/3/4/5 and the local tall landrace CK) at five developmental stages (2, 4, 6, 8 and 10 months after pollination). The four tissue-specific CT values exhibited distinct, reproducible developmental trajectories, collectively forming a quantitative 'CT developmental clock' (Figure 1).

Fiber layer CT decreased from -13 HU at 2 months to -795 HU at 10 months across all varieties, following a logistic decay curve ($R^2 > 0.95$). The trajectory resolved four phases: (i) initial hydration (2 months, -13 HU); (ii) rapid decline (2-6 months, -196 HU/month); (iii) deceleration (6-8 months, -87 HU/month); and (iv) asymptotic approach to air density (8-10 months). This ~ 780 HU decline reflects progressive replacement of water by air in mesocarp fiber cells.

Shell CT rose from ~ 0 HU at 2 months to $+202$ HU (W5) and $+256$ HU (other varieties) at 10 months, following a sigmoidal curve with rapid deposition (2-6 months, $+42$ HU/month) and saturation (6-10 months, $+8$ HU/month) phases. This captures progressive shell mineralization through calcium oxalate deposition (Franceschi and Nakata, 2005). W5 exhibited significantly lower terminal shell CT (202 vs 230-256 HU, Tukey HSD $p < 0.01$), consistent with its thin-shell phenotype. Two-way

ANOVA confirmed significant variety $\times$ stage interactions for shell CT ($F_{4,90} = 23.4$, $p < 0.001$) and endosperm CT ($F_{4,90} = 18.7$, $p < 0.001$).

Endosperm CT increased from 0 HU (no solid endosperm at 2 months) to +37 HU at 10 months. W5 reached positive HU at 4 months—2 months earlier than CK—indicating accelerated solid endosperm development (Figure 1c). This event cannot be detected through external observation.

Coconut water CT remained stable at 19–27 HU, with no significant variety or stage effects.

Among the five varieties, W5 maintained the smallest fruit weight (402 to 930 g), CK the largest (1,200 to 2,639 g at 8 months). W5 also showed the thinnest shell, fastest endosperm development and most gradual growth trajectory—a pattern of **structural growth investment**.

2.2 | Metabolomics reveals reciprocal metabolic profiles: W5 accumulates flavonoids, CK accumulates amino acids

Widely targeted metabolomics identified 624 metabolites across 12 major classes: flavonoids (131), phenolic acids (68), amino acids and derivatives (52), lipids (48), nucleotides and derivatives (41), organic acids (39), tannins (28) and terpenoids (14). Volcano plot analysis (Figure 2a) revealed 41 differentially abundant metabolites ($|\log_2FC| > 1$ and $-\log_{10}(FDR) > 1.3$): 26 enriched in W5 and 15 enriched in CK.

W5—flavonoid-dominated chemical defense: We term this pattern **structural-metabolic compensation**: investment in physical protection (shell thickness) and chemical defense (flavonoid accumulation) are inversely coordinated, such that structural efficiency is balanced by elevated chemical defense, rather than both being reduced as in a classical growth–defense trade-off.: All 26 metabolites with higher abundance in W5 were secondary metabolites, predominantly flavonoids (Figure 2b). The most differentially abundant compound was isorhamnetin-3-O-rutinoside (16.0-fold, $\log_2FC = +4.04$, $FDR < 0.001$), a methylated flavonol glycoside with documented antioxidant and antimicrobial activities. Other highly enriched flavonoids included tamarixetin-3-O-glucoside-7-O-rhamnoside (16.2-fold, $\log_2FC = +4.02$), sexangularetin-3-O-glucoside-7-O-rhamnoside (15.5-fold, $\log_2FC = +3.96$) and kaempferol-3-O-rutinoside (13.0-fold, $\log_2FC = +3.70$). Across the flavonoid class, 26 of 131 species (19.8%) showed differential abundance, and all were higher in W5. No flavonoid was higher in CK. This suggests that W5, despite its

structurally growth-oriented phenotype (thin shell, rapid endosperm development), maintains an elevated chemical defense capacity.

CK—amino acid and organic acid predominant: The 15 metabolites higher in CK were dominated by amino acids and organic acids (Figure 2c). L-Proline ($\log_2FC = -3.24$), L-valine ($\log_2FC = -2.27$) and 5-oxoproline ($\log_2FC = -2.88$) were among the most enriched, suggesting active nitrogen metabolism and protein turnover. Organic acids including 3-hydroxybutyric acid ($\log_2FC = -3.11$) and citric acid were also higher in CK. These primary metabolites are associated with growth-related metabolic activity rather than chemical defense.

Class-level enrichment analysis (Figure 2d) confirmed this reciprocal pattern: flavonoids were exclusively enriched in W5 (26 species, 100% of differential flavonoids); amino acids and organic acids were predominantly enriched in CK (87% and 83% of differential species, respectively). Only two metabolite classes (phenolic acids and lipids) showed bidirectional changes, with the majority of their differential species enriched in W5.

This metabolic profile constitutes a **structural-metabolic compensation axis**: W5, the improved dwarf variety with thin shell and accelerated endosperm development (structural growth traits), simultaneously accumulates chemical defense compounds (flavonoids). CK, the local landrace with thick protective shell, shows higher levels of growth-associated primary metabolites (amino acids). Rather than a simple growth-defense trade-off where one variety invests in growth and the other in defense, the pattern suggests structural-metabolic compensation—structural efficiency is balanced by chemical defense within the same variety.

2.3 | Systematic CT-metabolite correlation analysis reveals inter-variety asymmetry

To bridge structural and chemical dimensions, we performed Pearson correlation analysis between 18 CT parameters and 624 metabolites at five developmental time points ($n = 5$ time-point means per variety, $df = 3$), computed separately by variety. The analysis revealed a striking inter-variety asymmetry: W5 yielded 3,376 significant CT-metabolite associations ($FDR\ q < 0.05$), covering 611 metabolites and all 18 CT parameters, of which 55 pairs achieved near-perfect correlation (defined here as $|r| > 0.99$); in stark contrast, CK showed no pair exceeding $|r| > 0.95$ (Figure 4, Figure S1). These 55 pairs represent 0.49% of the 11,232 total tested pairs and 1.63% of the 3,376 significant associations in W5.

Methodological caveat: with $n = 5$ ($df = 3$), confidence intervals for individual coefficients are wide. We do not treat any single r value as mechanistic evidence, focusing instead on inter-variety asymmetry as a global pattern. At the uncorrected threshold ($p < 0.05$), W5 had 5,412 significant pairs versus 3,281 in CK, confirming that the asymmetry does not arise from differential FDR correction stringency. To test whether this asymmetry could arise by chance, we performed a permutation test: time-point labels were shuffled 10,000 times. Under permutations, the median number of $|r| > 0.99$ pairs in W5 was 0 (95% CI 0–1), versus 55 observed ($p < 0.0001$). A leave-one-out sensitivity analysis showed that 48 of the 55 near-perfect pairs ($|r| > 0.99$ threshold) retained $|r| > 0.95$ after removal of any single time point (7 of the 55 pairs dropped below $|r| > 0.95$ but remained > 0.90), indicating that the high correlations are not driven by individual outliers. We further verified that asymmetry is not an artifact of the linearity assumption: Spearman rank correlation and first-differencing Pearson correlation (detrending) both preserved the pattern, with W5 retaining 48 pairs at $|\rho| > 0.95$ (Spearman) and 41 pairs at $|r| > 0.95$ (FD-Pearson), versus 0 and 2 in CK, respectively (Supplementary Table S4).

Three near-perfect associations revealed potential mechanistic links. First, **shell density** correlated negatively with **13-methyltetradecanoic acid** ($r = -0.999$, $p < 0.001$), a branched-chain fatty acid that declined as shell density increased, possibly reflecting down-regulation of membrane lipid metabolism during mineralization. Second, **endosperm density** correlated positively with **dehydroabietic acid** ($r = +0.999$, $p < 0.001$), a diterpene resin acid associated with plant defense that accumulated in parallel with endosperm solidification. Third, **shell thickness** correlated with **citric acid** ($r > +0.99$), which serves as both a TCA cycle intermediate and an oxalate precursor for calcium oxalate deposition, suggesting carbon flux coordination between energy metabolism and structural mineralization.

Class-level analysis revealed systematic patterns: flavonoids correlated negatively with fiber density in W5 (mean $r = -0.83 \pm 0.11$), while amino acids correlated positively with endosperm density (mean $r = +0.88 \pm 0.09$) (Figure 4b). The metabolite co-regulation network further illustrates these class-level interaction patterns (Figure S3).

The most pronounced finding was the asymmetry in correlation strength: CK had no pair exceeding $|r| > 0.95$, contrasting with the 55 pairs observed in W5. This asymmetry raises a question: does CT-metabolite association strength reflect direct effects of breeding selection, or a statistical consequence of genetic homogenization?

2.4 | CT features predict seed germination—the dual role of shell thickness

Using nine DICOM-derived CT features from 75 matched CT-germination samples (45 germinated, 30 non-germinated), a random forest classifier achieved a hold-out AUC of 0.716 (Figure 5a) and 5-fold cross-validation AUC of 0.609 ± 0.089 (Figure S2). The discrepancy between hold-out and CV AUC (0.716 vs 0.609) likely reflects the modest sample size ($n = 75$) and the class imbalance; repeated random train-test splits ($n = 100$) yielded a mean hold-out AUC of 0.678 ± 0.072 , closer to the CV estimate. Six alternative algorithms were compared, confirming random forest as the top performer.

Feature importance analysis (Figure 5b) identified **shell thickness** (8.7% relative importance) and **embryo size** (8.2%) as the most informative predictors. Shell thickness simultaneously served as a central node in the CT-metabolite correlation network ($r > 0.99$ with citric acid) and as the strongest germination predictor. The two most important features together contributed less than 17%, with predictive signal distributed across multiple features.

3 | Discussion

3.1 | The CT developmental clock: a quantitative framework for coconut fruit maturation

The four tissue-specific CT trajectories provide a quantitative, non-destructive framework for in vivo coconut fruit maturation. The fiber layer CT trajectory captures progressive mesocarp air-filling, unique to coconut among the Arecaceae—oil palm mesocarp accumulates oil rather than losing water (Tarmizi et al., 2021). The shell CT trajectory captures a conserved biomineralization program shared among hard-shelled fruits (Schoeman et al., 2021). The endosperm CT trajectory provides a non-destructive measure of solid endosperm formation.

The accelerated endosperm development and lower terminal shell CT of W5 are consistent with its breeding history of early maturation and thin shell. The $\sim 1,200$ HU density span makes coconut a well-suited system for CT phenotyping.

3.2 | The structural-metabolic compensation axis: a revised view of the growth-defense relationship

The reciprocal metabolic profile between W5 and CK does not follow the classical growth-defense trade-off pattern. In the classical model (Huot et al., 2014), fast-growing, improved

varieties accumulate fewer secondary metabolites, whereas slow-growing landraces invest more in chemical defense. This has been demonstrated in domesticated tomato (Schauer et al., 2005) where larger fruit size co-occurs with reduced secondary metabolites, and in maize (Doebley et al., 2006) where selection for starch content occurred at the expense of defense pathways.

The coconut pattern is fundamentally different. W5, the improved dwarf variety, exhibits both structural growth traits (thin shell, early endosperm development, smaller fruit) and elevated chemical defense (flavonoid accumulation). CK, the local landrace with thick protective shell, shows higher levels of growth-associated primary metabolites (amino acids, organic acids). Rather than a linear trade-off, this constitutes a **structural-metabolic compensation axis**: structural and chemical defense investments are inversely coordinated across tissue types (shell vs. endosperm), not across varieties.

We propose two non-mutually-exclusive interpretations. **First**, the compensation may reflect a resource reallocation strategy: W5's thin shell reduces structural investment in physical protection, freeing resources for flavonoid biosynthesis in the endosperm. The observation that shell thickness is the strongest germination predictor (Figure 5a) and correlates with citric acid—an oxalate precursor for shell mineralization—supports carbon flux coordination between structural and chemical compartments. **Second**, the flavonoid enrichment in W5 may be a pleiotropic consequence of dwarfing gene selection. Dwarfing genes in cereals (e.g., rice *sd1*, wheat *Rht*) are associated with altered gibberellin signaling, which can modulate secondary metabolism (Spielmeyer et al., 2002). Whether similar mechanisms operate in coconut dwarf varieties remains to be determined.

Regardless of the mechanism, this structural-metabolic compensation suggests that breeding selection for structural traits (thin shell, early maturity) has indirect metabolomic consequences that cannot be predicted from structural phenotypes alone—underscoring the value of multi-omics integration.

3.3 | Inter-variety CT-metabolite association asymmetry: a statistical signal consistent with breeding canalization

The central observation in this study—55 pairs with $|r| > 0.99$ in W5 versus none exceeding $|r| > 0.95$ in CK—requires careful interpretation, as two competing explanations are plausible.

Hypothesis A ('tightened coupling'): Breeding selection that targeted specific phenotypes (shell thickness, fruit weight, maturation timing) may have incidentally tightened structural-

metabolic coordination for the selected traits, producing stronger CT-metabolite correlations in W5. Under this view, the asymmetry is a byproduct of multivariate phenotypic selection.

Hypothesis B (developmental canalization): W5, as a genetically homozygous variety developed through generations of directional selection, exhibits canalized developmental programs—trajectories buffered against environmental fluctuations and genetic background variation (Waddington, 1942). All W5 individuals follow nearly identical trajectories, producing tight statistical correlations at the population level. CK, as a heterogeneous landrace, possesses substantial genetic variation that dilutes correlation strength.

We find Hypothesis B more parsimonious for three reasons. **First**, the asymmetry is global rather than trait-specific: the 55 near-perfect correlations span diverse parameter combinations—from endosperm density to branched-chain fatty acids, from shell thickness to TCA cycle intermediates. If selection targeted only specific traits (shell thickness, fruit size), the global scope of the asymmetry—covering all 18 CT parameters and 611 metabolites—would require the coordinated selection of an implausibly large number of independent pathways. **Second**, metabolic variation is broader in CK than W5 (median CV: 1.13 vs 0.98, $p = 0.017$, Wilcoxon test), ruling out the possibility that CK's weaker correlations result from insufficient metabolic variation. **Third**, the observation that CK maintains denser inter-metabolite correlations (mean $|r| = 0.493$ vs 0.441 in the co-regulation network) suggests that the weaker CT-metabolite coupling in CK reflects genuine developmental decoupling rather than a statistical artifact.

Under this interpretation, CT-metabolite association strength measures not the tightness of coupling at the individual level, but developmental canalization at the population level—a statistical signature of breeding-induced genetic homogenization (Doebley et al., 2006). If validated at the DNA level (by correlating canalization metrics with genomic diversity indices such as nucleotide diversity π , runs of homozygosity or fixation index F_{st}), CT-metabolite association strength could become a practical indicator for assessing developmental uniformity in coconut breeding populations.

3.4 | CT-based germination prediction

The two strongest predictors—shell thickness and embryo size—correspond to known germination biology. Shell thickness is the primary physical barrier for radicle emergence (Tarmizi et al., 2021); the coconut haustorium plays a key role in mobilizing endosperm reserves during germination (Sugimura and Murakami, 1990). Embryo size reflects growth potential and

storage reserves. This method reduces germination assessment from 2–3 months to approximately 15 minutes, providing a practical screening tool for coconut nurseries.

3.5 | Limitations and future directions

The metabolomic analysis was limited to two varieties (W5 and CK). Extension to all five varieties would allow comprehensive characterization of the compensation axis across a broader genetic range. CT-metabolite correlations are associative rather than causal; functional validation experiments (e.g., CT monitoring following targeted inhibition of flavonoid biosynthesis) are needed. Our interpretation of CT-metabolite asymmetry as developmental canalization requires DNA-level validation—specifically, comparing genomic diversity indices across varieties and correlating them with CT-metabolite association strength. Integrating transcriptomics could identify the regulatory genes underlying selection-driven canalization.

At a descriptive level, this study provides a systematic CT phenomic record of multi-variety coconut fruit development. At a mechanistic level, it reveals a structural-metabolic compensation axis through cross-modal correlation. At a hypothesis-generating level, it proposes that CT-metabolite association strength may serve as a candidate quantitative phenotype for developmental uniformity, with testable predictions for genetic validation.

Beyond validation, future research should explore whether CT-metabolite association strength can serve as a predictive indicator for breeding population uniformity. If validated across diverse genetic backgrounds, this metric could be integrated into breeding programs as a non-destructive, high-throughput screening tool for developmental consistency. Additionally, extending this integrative framework to other fruit tree crops with hard shells (e.g., walnut, macadamia) could reveal whether the compensation axis and canalization signatures are conserved features of breeding selection.

4 | Experimental Procedures

4.1 | Plant Materials and Experimental Design

Five dwarf coconut varieties: Wenye 2 (W2), Wenye 3 (W3), Wenye 4 (W4), Wenye 5 (W5)—improved varieties bred by the Coconut Research Institute, Chinese Academy of Tropical Agricultural Sciences—and CK (Hainan Lvga, local tall landrace). Fruits were collected from the Wenchang Experimental Station, Hainan (19°32'N, 110°47'E), harvested at 2, 4, 6, 8 and 10 months after pollination, with 3–5 biological replicates per variety × month

(total $n \approx 100$ fruits). Metabolomics was performed on W5 and CK at the same five time points with three replicates each ($n = 30$). For germination prediction, 100 fruits were tracked for 6 months, of which 75 had matched CT records.

4.2 | CT Scanning and Computational Phenotyping

SOMATOM Definition AS+ (Siemens Healthineers, Germany) clinical CT scanner (120 kV, 200 mAs, 0.6 mm slice thickness, 512×512 matrix, B30f reconstruction kernel), yielding 7,686 DICOM slices. The pipeline included automated equatorial slice selection, tissue-level density profiling and DICOM-derived feature extraction, implemented in Python 3.11 (pydicom, OpenCV, scikit-learn). See Supplementary Methods for details.

4.3 | Metabolomics Analysis

Freeze-dried endosperm samples (100 mg) were extracted with 70% methanol/water and analyzed by UHPLC-MS/MS (Q-Exactive Orbitrap, Thermo Fisher Scientific) in MRM mode. Metabolites were identified against an in-house standard library, mzCloud and HMDB databases, yielding 624 metabolites. Metabolite identification confidence followed MSI guidelines (Sumner et al., 2007): 178 metabolites (28.5%) were confirmed at Level 1 (authentic standard match on retention time and MS/MS), and the remaining 446 (71.5%) at Level 2 (database spectral match).

4.4 | Statistical Analysis

Developmental trajectories: two-way ANOVA (variety $\times$ month) with Tukey HSD post-hoc tests. Differential metabolites: two-tailed Welch t-test with Benjamini-Hochberg FDR correction (Benjamini and Hochberg, 1995; $q < 0.05$). CT-metabolite correlations: Pearson correlation was computed separately by variety using $n = 5$ time-point means ($df = 3$). To control for multiple testing, p-values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995) across all 11,232 pairs (18 CT parameters $\times$ 624 metabolites) per variety, with significance threshold set at $q < 0.05$. Permutation test: time-point labels shuffled 10,000 times. Robustness checks: Spearman rank correlation and first-differencing Pearson correlation (detrending). Coefficient of variation analysis: two-sided Wilcoxon rank-sum test. Metabolite co-regulation networks (194,376 pairs) constructed from 15 individual samples per variety. Random forest classifier: nine DICOM-derived CT features, $n_estimators = 200$, $max_depth = 5$, $class_weight = 'balanced'$, 75:25 train-test split, 5-fold cross-validation, AUC evaluation. Hyperparameters were selected based on a preliminary grid search ($n_estimators$: 100/200/500; max_depth : 3/5/7/10), with the combination maximizing mean CV

AUC. Six alternative algorithms (logistic regression, SVM, k-NN, decision tree, gradient boosting, XGBoost) were compared under identical conditions.

4.5 | Data and Code Availability

All data supporting this study are publicly available. Metabolomics raw data and CT DICOM archives (7,686 slices) are available from the corresponding authors without restriction. Processed abundance tables, CT feature matrices and analysis scripts are accessible at palm.suncx.top (DOI pending upon Zenodo deposition). Interactive tools (CT viewer, metabolic network browser, germination predictor) are hosted at the same URL for community use.

Author Contributions

Ninghuan You: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft. Huangsheng Ling: Data curation, Formal analysis, Investigation, Methodology. Jing Chen: CT technical guidance, Data fitting, Resources. Wenrao Li: Investigation, Resources. Mengxing Huang: Methodology, Supervision, Writing – review & editing. Yu Zhang: Methodology, Software, Validation. Hongxing Cao: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. Chengxu Sun: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

This work was supported by the Key Research and Development Project of Hainan Provincial Department of Science and Technology (ZDYF2026XDNY143), the Central Finance Forestry Science and Technology Promotion Demonstration Fund Project of Hainan Province (Qiong [2024] TG07), and the International Science and Technology Cooperation Research and Development Project of Hainan Provincial Department of Science and Technology (GHYF2025027).

Conflicts of Interest

The authors declare no conflicts of interest.

References

Agati G and Tattini M (2010) Multiple functional roles of flavonoids in photoprotection. *New Phytologist* 186, 786–793.

Benjamini Y and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B* 57, 289–300.

Doebley J, Gaut BS and Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127, 1309–1321.

Falcone Ferreyra ML, Rius SP and Casati P (2012) Flavonoids: biosynthesis, biological functions, and biotechnological applications. *Frontiers in Plant Science* 3, 222.

Franceschi VR and Nakata PA (2005) Calcium oxalate in plants: formation and function. *Annual Review of Plant Biology* 56, 41–71.

Goodman RC and Orozco J (2022) X-ray computed tomography for seed quality assessment. *Seed Science Research* 32, 1–15.

Huot B, Yao J, Montgomery BL and He SY (2014) Growth-defense tradeoffs in plants: a balancing act to optimize fitness. *Molecular Plant* 7, 1267–1287.

Khodaei M, Khojastehpour M, Golzarian MR and Mohebbi M (2023) Detection of internal disorders in pomegranate fruit using X-ray computed tomography. *Postharvest Biology and Technology* 198, 112240.

Li J, Cao HX, Sun CX, Zhang J and Fan HK (2020) Breeding and selection of Wenye dwarf coconut varieties [In Chinese]. *Chinese Journal of Tropical Crops* 41, 2031–2038.

Liang Y, Yuan Y, Li B and Zheng Y (2021) Metabolomic analysis of coconut endosperm reveals temporal dynamics of fatty acid biosynthesis. *Journal of Agricultural and Food Chemistry* 69, 13128–13139.

Lin S, Zhang Y, Luo L, Huang M, Cao H, Hu J, Sun C and Chen J (2023) Visualization and quantification of coconut using advanced computed tomography postprocessing technology. *PLOS ONE* 18(2), e0282182.

Lin S, Sun C, Luo L, Huang M, John Martin JJ, Cao H, Hu J, Bai Z, He Z, Zhang Y and Chen J (2024) Exploring the density and morphology of coconut structures at two locations: a time-based analysis using computer tomography. *PeerJ* 12, e18206.

Lin S, Sun C, Luo L, Huang M, Cao H, Hu J, Bai Z, Zhang Y and Chen J (2025) The observation of internal structure changes and survival prediction modeling of mature coconut during germination based on computed tomography imaging. *Industrial Crops & Products* 233, 121396.

- Liu Q, Zhang Y, Chen J, Sun C, Huang M, Che M, Li C and Lin S (2023) An improved DeepLab V3+ network based coconut CT image segmentation method. *Frontiers in Plant Science* 14, 1139666.
- Lu L and Liu R (2021) Research progress on coconut germplasm resources, cultivation and utilization [In Chinese]. *Chinese Journal of Tropical Crops* 42, 1795–1803.
- Mehmood A, Bai X, Zhang Y, Bai X, Sun C, Bhatti UA, Huang M, Lin S, Cao H and Chen J (2026) Phenotyping coconut germination stage classification based on few-shot image generation using a modified FastGAN for non-destructive processing. *Journal of Food Composition and Analysis* 153, 109068 (advance online publication).
- Perera L, Perera SACN, Bandaranayake CK and Harries HC (2009) Coconut. In: Vollmann J and Rajcan I (eds) *Oil Crops*. New York: Springer, pp. 369–396.
- Piovesan A, Vancauwenberghe V, Van De Looverbosch T, Verboven P and Nicolai B (2021) X-ray computed tomography for 3D plant imaging. *Trends in Plant Science* 26, 1171–1185.
- Saechua W, Pinitnanthakorn T, Narkrugsu W and Teerachaichayut S (2021) Non-destructive evaluation of internal quality of durian fruit using X-ray computed tomography. *Journal of Food Engineering* 307, 110648.
- Schauer N, Zamir D and Fernie AR (2005) Metabolic profiling of leaves and fruit of wild species tomato: a survey of the *Solanum lycopersicum* complex. *Journal of Experimental Botany* 56, 297–307.
- Schoeman L, Williams PJ, du Plessis A and Manley M (2021) X-ray micro-computed tomography for non-destructive characterization of food microstructure. *Trends in Food Science and Technology* 109, 154–167.
- Spielmeyer W, Ellis MH and Chandler PM (2002) Semidwarf (sd-1), “green revolution” rice, contains a defective gibberellin 20-oxidase gene. *Proceedings of the National Academy of Sciences* 99, 9043–9048.
- Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TWM, Fiehn O, Goodacre R, Griffin JL, Hankemeier T, Hardy N, Harnly J, Higashi R, Kopka J, Lane AN, Lindon JC, Marriott P, Nicholls AW, Reilly MD, Thaden JJ and Viant MR (2007) Proposed minimum reporting standards for chemical analysis. *Metabolomics* 3, 211–221.

Sun C, Ma X, John Martin JJ, Cao H, Zhang Y, Gao Y, Xing C and Hou M (2024) Preliminary study on the association between lignan metabolites and CT non-destructive testing of coconut fruit at different developmental stages. *PeerJ* 12, e18049.

Sugimura Y and Murakami T (1990) Structure and function of the haustorium in germinating coconut seeds. *Japan Agricultural Research Quarterly* 24, 55–61.

Tarmizi AHA, Shafie SM, Ismail R and Ahmad D (2021) X-ray computed tomography for quality assessment of oil palm fruits. *Scientia Horticulturae* 285, 110182.

Waddington CH (1942) Canalization of development and the inheritance of acquired characters. *Nature* 150, 563–565.

Wang Q, Xue H, John Martin JJ, Hou M, Cao H, Dong Z, Li J and Sun C (2024) Trends and applications of computed tomography in agricultural non-destructive testing. *Agriculture* 14, 2329.

Hou M, John Martin JJ, Song Y, Wang Q, Cao H, Li W and Sun C (2024) Dynamics of flavonoid metabolites in coconut water based on metabolomics perspective. *Frontiers in Plant Science* 15, 1468858.

Yang T, Li H, Zhang R and Chen J (2022) Metabolite profiling of coconut endosperm across eight developmental stages. *Scientia Horticulturae* 304, 111290.

Yu L, Liu L, Yang W, Wu D, Wang J, He Q et al. (2022) A non-destructive coconut fruit and seed traits extraction method based on Micro-CT and DeepLabV3+ model. *Frontiers in Plant Science* 13, 1069849.

Zhang Y, Liu Q, Chen J, Sun C, Lin S, Cao H, Xiao Z and Huang M (2023) Developing non-invasive 3D quantificational imaging for intelligent coconut analysis system with X-ray. *Plant Methods* 19, 24.

Zhao Y, Liu S, Wang H and Li X (2023) Variety-specific metabolite accumulation in coconut water across eight cultivars. *Journal of Food Composition and Analysis* 118, 105192.

Züst T, Joseph B, Shimizu KK, Kliebenstein DJ and Turnbull LA (2011) Using knockout mutants to reveal the growth-defense trade-off in *Arabidopsis*. *Proceedings of the National Academy of Sciences* 108, 11531–11536.

Figure Legends

Figure 1. CT developmental trajectories. (a) Representative equatorial CT slices of W2–W5 with tissue compartments annotated (F = fiber, S = shell, E = endosperm, W = water cavity); (b) fruit weight trajectories; (c) endosperm CT values—W5 reaches positive HU at 4 months, 2 months earlier than CK; (d) shell CT values—reflecting progressive mineralization. Error bars = SD from 3–5 biological replicates.

Figure 2. Metabolomic comparison. (a) Volcano plot of 624 metabolites ($n = 30$: W5 $n = 15$, CK $n = 15$)—blue (W5 $\uparrow$, 26 metabolites, flavonoid-dominated, positive $\log_2FC$), red (CK $\uparrow$, 15 metabolites, amino acid/organic acid-dominated, negative $\log_2FC$); (b) top 10 enriched in W5—flavonoid-dominated chemical defense; (c) top 10 enriched in CK—growth-associated primary metabolites; (d) metabolite class distribution by up/down regulation. The pattern reveals that W5, despite its structurally growth-oriented phenotype (thin shell, accelerated endosperm), accumulates chemical defense compounds—a structural-metabolic compensation axis.

Figure 3. Metabolite class heatmap ($n = 30$ metabolomic samples: W5 $n = 15$, CK $n = 15$). Flavonoids enriched in W5 (defense-oriented chemical profile); amino acids and organic acids enriched in CK (growth-oriented primary metabolism). Z-score normalized across all samples.

Figure 4. CT-metabolite association framework. (a) Network of 3,376 significant associations, with 55 pairs at $|r| > 0.99$ (W5); (b) class-level patterns—flavonoids negatively correlated with fiber density in W5, amino acids positively correlated with endosperm density; (c) asymmetric association strength between W5 and CK.

Figure 5. Germination prediction. (a) ROC curve (AUC = 0.716; $n = 75$ samples: 45 germinated, 30 non-germinated); (b) feature importance—shell thickness and embryo size are the top two predictors.

Figure S1. Complete list of 3,376 significant CT-metabolite associations (FDR $q < 0.05$) in W5. Heatmap of Pearson correlation coefficients between 18 CT parameters and 624 metabolites across five developmental stages.

Figure S2. Five-fold cross-validation AUC distribution for the random forest germination classifier. Box plot showing individual fold AUC values and mean $\pm$ SD.

Figure S3. Metabolite co-regulation network based on 194,376 Pearson correlation pairs across 624 metabolites in W5 samples. Node color indicates metabolite class; edge thickness reflects correlation strength. Only $|r| > 0.95$ pairs are shown for visual clarity.