

Identification of Seed Metabolites and Microbiota Members associated with Germination and Emergence in Common Bean

Louna Colaert-Sentenac

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Élisabeth Planchet

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Cyril Abadie

Ecophysiologie et Génomique Fonctionnelle de la Vigne, University of Bordeaux, Bordeaux Science Agro, INRAE, ISVV

Julie Lalande

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Sherif Hamdy

GEVES, Station Nationale d'Essais de Semences

Coralie Marais

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Audrey Dupont

GEVES, Station Nationale d'Essais de Semences

Laurence Le Corre

GEVES, Station Nationale d'Essais de Semences

Claude-Emmanuel Koutouan

FNAMS, Impasse de verger

Marie-Hélène Wagner

GEVES, Station Nationale d'Essais de Semences

Matthieu Barret

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Guillaume Tcherkez

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Béatrice Teulat

`beatrice.teulat@institut-agro.fr`

Institut Agro, INRAE, IRHS, SFR QUASAV, Université d'Angers

Marie Simonin

Research Article

Keywords: indicator microbial taxa, metabolite biomarkers, germination speed, *Phaseolus vulgaris* L., multidimensional characterisation, seed morphology, seed quality, seedling emergence

Posted Date: August 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10390664/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Seed quality is a complex trait shaped by morphological, biochemical and microbiological properties that are rarely characterised simultaneously, limiting our ability to identify robust predictive indicators of germination speed and seedling emergence across varieties. Here, we performed a multi-factor characterisation of eight common bean (*Phaseolus vulgaris* L.) varieties, combining seed morphometrics, untargeted GC-MS metabolomics on three seed organs, and amplicon sequencing of bacterial and fungal communities, to identify indicators of germination speed and emergence percentage.

Results

The eight varieties showed substantial variation in both traits, used as physiological seed quality proxies. Seed weight and size variation between varieties were correlated with germination speed. The intravariety variance of seed weight was independently correlated with emergence performance. Metabolome composition differed strongly across seed organs, with variety as the dominant driver. Individual-seed metabolomic profiles in the plumule and cotyledon were associated with germination speed but not emergence, yielding 16 plumule and three cotyledon candidate metabolite markers. Fungal community composition was associated with both germination speed and emergence, while bacterial communities were associated with emergence only. Nine fungal and four bacterial taxa were identified as candidate indicators. Inter-kingdom co-occurrence network analysis revealed that fungi with similar germination speed associations tend to cluster in the same modules, suggesting that community-level modules rather than individual taxa may constitute more robust microbial indicators.

Conclusions

These results demonstrate that germination speed and emergence capacity are governed by distinct seed properties, and provide morphological, metabolic and microbial candidate indicators for integration into targeted seed quality assessment frameworks for common bean.

Background

High-quality seeds are essential in agriculture to ensure high germination rate, as well as rapid and uniform emergence and seedling establishment, which affect crop production. Seed quality is a multifactorial trait that integrates physical, genetic, physiological, and sanitary components. Seed quality is evaluated through physical criteria (non damaged seeds, uniform size, no insect infestation, moisture content), variety purity, physiological quality (fast and high-rate germination and emergence, protein, oil,

fatty acid, and amino-acid content), and health-related criteria (cleanliness, absence of disease-causing microorganisms) [1, 2].

Seed physiological quality is best evaluated by germination and emergence assays, conducted with precise parameter settings for official evaluation [3]; however, those tests are time-, resource-, and material-consuming. To evaluate seed physiological quality, morphological proxies are historically the most commonly used indicators, with many studies linking seed weight and seed energy stock quantity, seed color and imbibition time before germination [4]. During the transition from quiescent dry state to active growth, seeds rely on tightly regulated metabolic networks involving carbohydrates, lipids, amino acids, antioxidants, and signaling molecules that support energy supply, stress tolerance, and developmental progression [5, 6]. Consistent with this view, seed metabolic signature could provide valuable insight into seed germination ability, as previously evidenced for maize, tomato, rice, and wheat [7–9]. Indeed, as evidenced for dry seed mRNAs and proteins [10–12], metabolome could constitute a reservoir of ready-to-use molecules when the imbibition starts, thus impacting seed germination. Metabolic pathways involved in seed quality and seed germination are traditionally explored through GWAS, transcriptomic, or proteomic studies, but the dry seed metabolite pool has yet received little attention. The difference between the composition of seed organs could be especially relevant, given their specialized roles in germination [13]. In parallel, the seed microbiota is increasingly viewed as a functional component of seed biology rather than a passive contaminant. Seeds carry microbial communities that can contribute to plant establishment, health, and resilience [14–16], and are now considered a primary inoculum for the next plant generation. Yet routine seed quality evaluation usually focuses on harmful microorganisms and does not account for potentially beneficial seed-associated taxa. As a result, the relationships between seed microbiota, seed metabolome, and germination and emergence performance remain poorly resolved, as well as their contribution to the differences in seed quality between varieties, with only few studies exploring their interactions [17].

Here, we addressed this issue with a comprehensive multi-factor study to explore traits associated with common bean (*Phaseolus vulgaris* L.) seed germination and seedling emergence, taking advantage of a collection of French varieties of commercial significance. Common bean was selected as model legume species due to its importance as staple food, displaying high genetic and phenotypic diversity [18, 19]. As one of the most widely cultivated grain legumes, common bean provides essential protein amino acids, fibers, and micronutrients, making it a critical component of human diets, especially in developing countries [19–21]. In addition, bean varieties exhibit substantial diversity in seed morphology, composition, and germination behaviour, offering an excellent opportunity to explore correlations between these traits and overall seed quality [22], that to the best of our knowledge has not yet been studied. Seeds used for these analyses were obtained from plants cultivated at the same time in the same location, limiting environmental effect on seed development and microbiota assembly.

Germination and emergence assays, GC-MS metabolomic analyses of various seed tissues, and analyses of the seed microbiome across varieties were used to identify indicators of germination and emergence. We examined their correlation with seed properties using multivariate statistical methods.

Our working questions were: (i) how do mature seed morphological, metabolic, and microbiota-related parameters relate to seed germination speed and seedling emergence percentage, and can we therefore identify quality biomarkers? (ii) are such biomarkers valid across different bean varieties?

Methods

All statistical analyses were conducted on R Statistical Software (v.4.5.3; R Core Team 2025). All figures were generated using *ggplot2* v4.0.2 package. A list of package versions used is provided in the code.

Plant material

Eight *Phaseolus vulgaris* L. (common bean) varieties were selected to represent a large panel of cultivated bean varieties in France (Caprice, Contender, Deezer, Facila, Flavert, Linex, Vanilla, and Vezer), including dwarf French string bean, and shelling bean. The seeds for all varieties were produced in 2020 as part of a single field trial (organic farming conditions in a single plot, with one line per variety) at the National Federation of Seed Multipliers experimental station (FNAMS, 43°57'35.654"N 0°23'45.428"E, Condom, Gers, France). Harvested seeds were dried at room temperature, cleaned by filtering soil and non-seed structures on sieves at the National Seed Testing Station (GEVES, France), and stored in paper bags at 4°C in the dark until further use.

Seed germination and seedling emergence indicators

Seed germination assays

Seed germination assays were performed at the National Seed Testing Station (GEVES, France) on wet pleated paper in a Jacobsen germinator set at 20°C, 95% humidity and equipped with four RGB cameras [23]. The assay was performed on 4 replicates of 24 non-damaged seeds for each variety. A seed was considered germinated as soon as the radicle emerged from seed testa. Germination imaging was recorded every 2 h over 6 days (144 h). Germination time-course curves for each variety were used to determine two germination indicators: the final germination percentage (FGP) and its variety average (mean FGP), and the germination speed per replicate (time to 50% of germinated seeds - t50germ) and its variety average - mean t50germ).

Seedling emergence assays

Seedling emergence assays were performed on the Phenotic platform [24] on four biological replicates of 20 seeds without defaults for each variety. Seeds were grown in potting mix [K® TraySUBSTRAT Klasmann-Deilmann France] in seedling trays, with tray position randomised across varieties. Plants were grown in a climatic chamber set at 23°C day and 20°C night, for 16 h day and 8 h night, under 70% humidity and 150 $\mu\text{mol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ LED light. The number of non-germinated seeds, normal seedlings that have emerged, or abnormal seedlings was recorded on the seventh day according to ISTA guidelines for common bean [3]. Mean final emergence percentage (FEP) and its variety average (mean FEP) were used as the indicator of emergence of varieties.

Statistical analyses - Germination and emergence indicators

Germination kinetic curves were fitted per variety using a four-parameter Hill function (*FourPHFit.bulk*, germinationmetrics [25]) to derive the time to 50% of germinated seeds for each replicate (t50germ) [26]. FGP, t50germ and FEP were compared between varieties by one-way ANOVA (R *aov* function) with post-hoc pairwise comparison of adjusted means (multcomp package).

Seed weight and morphological characteristics

Seed measurements

Seed morphological characteristics were measured on 100 individual seeds for each variety at GEVES (France). Their morphological characteristics, including seed area (**Table S1**), were determined from images acquired by 2D X-rays radiography [23]. These images were analysed using Python Programming Language, OpenCV, and AVIZO (Thermo Fisher Scientific). This approach also enabled the identification of seeds presenting mechanical defects (visual inspection for tegument fissures, protruding teguments, insect damage and spread cotyledons). The proportion of seeds presenting defects was quite low (0-16.7% across varieties, 3.8% overall), and those seeds were excluded from further processing. All seeds were individually weighed (mg) on a precision balance, then sown on wet flat paper at 20°C on the automated germinator to record individual imbibition and germination time (hours to radicle emergence).

Statistical analyses

To assess variations in seed weight and size diversity between varieties and their associations with germination speed and emergence percentage, univariate statistics were used.

Variety-level means and standard deviations were computed for seed weight and all morphological variables across the 100 seeds per variety. Thanks to individual measures, the coefficient of variation (CV = SD/mean) was also computed for seed weight and area as an index of intravariety size heterogeneity. Between-variety Spearman rank correlations were computed between variety-level means of all morphological variables and mean t50germ and mean FEP, using the eight variety means as the unit of analysis (n = 8). Significance was assessed at $\alpha = 0.05$ with no correction for multiple testing given the exploratory nature of the analysis.

The relationship between individual seed weight and projected area was visualised using scatter plots with variety-level 95% confidence ellipses. Logarithmic regression ($y \sim \log(x)$) was used to model the relationship between mean seed weight and mean t50germ across varieties, as visual inspection suggested a non-linear relationship. Linear regression was used for all other trait pairs. R^2 values from ordinary least squares regression are reported alongside Spearman rank correlation coefficients to characterise both the strength and the functional form of the relationships.

To test whether the between-variety relationship between seed size and germination speed was recapitulated at the individual seed level, Spearman rank correlations between individual seed weight or

area, and individual germination time were computed separately within each variety, restricted to germinated seeds ($n = 43\text{--}90$ per variety depending on germination rate). This analysis addresses whether heavier or larger individual seeds within a variety germinate faster or slower than lighter or smaller seeds of the same variety, independently of between-variety differences in mean seed size.

Seed metabolome analysis

Sample preparation, extraction, and GC-MS analysis

The metabolomes of different seed organs (plumule, cotyledons, and hypocotyl-radicle - HR-axis) were studied by GC-MS across all varieties. To avoid too much bias in metabolome composition due to seed size, only seeds with a weight within $\pm 50\%$ standard deviation around the mean seed size for each variety were kept [27]. Four mature dry seeds of each variety were dissected to isolate the organs. Each organ was weighted and ground into a fine powder with metallic beads. Extraction of metabolites were carried out on 4 mg dry weight with 200 μL 90%/10%(v/v) methanol - MilliQ water with adonitol (55 $\mu\text{mol L}^{-1}$, Sigma-Aldrich®) as an internal standard. The extraction volume was adjusted proportionally for samples < 4 mg (plumule, HR-axis). Extracts were centrifuged for 10 min at 1,500 g . For each extract, 10 μL of supernatant was transferred into a vial with insert, and 10 μL was used to prepare a quality control (QC), i.e. a mix of all seed samples, which was also prepared in 10 μL per vial. All vials were spin-dried overnight at room temperature, sealed and kept in dark dry conditions until GC-MS analysis.

GC-MS analyses were conducted with a GC-MS-Orbitrap Q Exactive (Thermo Scientific). Samples were automatically derivatized with 20 μL of 20 $\text{mg}\cdot\text{mL}^{-1}$ methoxyamine-pyridine solution for 90 min at 37°C, then with 30 μL of N-methyl-N-(trimethylsilyl)-trifluoroacetamide (MSTFA) for 30 min at 37°C. Five μL of alkane mix (C_9 to C_{36} alkanes at 3 $\mu\text{g}\cdot\mu\text{L}^{-1}$, Connecticut n-Hydrocarbon Mix, Supelco) was added before injection to compute the retention index. One μL sample was injected in splitless mode in a TG-5 SILMS column (30 m x 0.25 mm x 0.25 μm , Thermo Scientific) at 230°C, using a Trace 1300 Series GC (Thermo Scientific). Samples were carried with a constant helium flow (1 $\text{mL}\cdot\text{min}^{-1}$). Temperature in the GC oven was set at 70°C for 1 min, then raised at 15°C $\cdot\text{min}^{-1}$ to 325°C and kept such for 4 min.

MS analyses were done in positive polarity in full MS scan mode with source set at weight scan range: 50–750 m/z , resolution: 60,000, AGC: target 1E6, MS transfer line: 300°C, and filament delay: 4.12 min. Metabolites were ionized by electron impact at 70 eV, 250°C. Blanks were injected at the beginning and at the end of each batch to assess column cleanliness.

QC samples were added to each batch to assess system stability throughout the batches and to condition the analytical platform. Analytes were identified automatically with TraceFinder® (Thermo Scientific) using retention time and m/z ion of major characteristic fragments and confirmation fragment(s) by comparison to a local compound database [28]. Peak integration by TraceFinder® was checked manually for all metabolites. Integral values were normalized using the internal standard.

Statistical analysis of metabolomic data

Relative analyte abundances were imported in R. A total of 252 annotated analytes were identified across organs. Metabolites detected in less than 20% of samples within an organ were excluded, leading to 221 retained metabolites per organ. Data were Z-scaled. Outlier samples were identified by principal component analysis (PCA) and removed when outside the Hotelling's T^2 ellipse at the 95% confidence interval. As a result, 38 cotyledon, 53 HR-axis, and 46 plumule samples across 8 varieties were kept.

Global metabolome variation across organs was explored by PCA on data Z-scaled across all samples jointly. Metabolic differences between organs were assessed by PERMANOVA (adonis2, vegan package v2.6, Euclidean distance, 999 permutations). Homogeneity of multivariate dispersion across organs was tested with betadisper followed by permutest (999 permutations, pairwise comparisons). Metabolites with significantly higher median abundance in a given organ relative to at least another organ were identified by Kruskal-Wallis tests with Benjamini-Hochberg (BH) correction (adjusted $p < 0.05$), followed by Dunn post-hoc pairwise comparisons. A heatmap was generated from these filtered, Z-scaled data using *ComplexHeatmap* v.2.28.0 [29].

For each organ, the association between metabolome and germination speed and emergence percentage was assessed via supervised multivariate statistics with orthogonal partial least squares regression - OPLS, *rop/s* v.1.44.0 [30] - using metabolites as X predictive variables and per-replicate t50germ or FEP as Y continuous response variables. Since individual seeds used for metabolome profiling were not the same as that used for germination and emergence assays, a random within-variety permutations was performed to assess the robustness of the OPLS analysis. To do so, 100 independent permutations were performed for each dataset. This number of permutations was determined by a preliminary stability analysis confirming convergence of metabolite significance frequencies and sign of loading. The quality of each OPLS statistical model was assessed by the cumulative R and Q^2 (cross-validated R^2), averaged across 100 permutations. Metabolite significance was evaluated based on frequency of significance across permutations. That is, a metabolite was kept as a candidate if it met simultaneously: (i) large loading value in the OPLS, i.e. $|pq1| > 0.05$, and its variable projection VIP was larger than unity in $\geq 70\%$ of permutations; (ii) univariate significance i.e. Spearman correlation with t50germ or FEP with Benjamini-Hochberg-adjusted $p < 0.05$ in $\geq 70\%$ of permutations; and (iii) the sign of the OPLS loading $pq1$ was consistent across $\geq 80\%$ of permutations. Results were visualized as volcano plots (mean $pq1$ vs. $-\log_{10}$ mean BH-adjusted p-value). Candidate metabolite biomarkers were then filtered by coefficient of variation of OPLS loading $pq1 < 0.5$ and BH-corrected p-value < 0.5 (reproducibility of the effect size and statistical significance). This analysis identified an initial set of significant metabolites. Subsequently, a further analysis was carried out to identify the most robust metabolites through three complementary, independent approaches :

1. Pearson or Spearman correlations ($n = 8$ varieties, test selected based on Shapiro-Wilk normality assessment) between variety-median scaled metabolite abundances and variety mean t50germ or mean FEP. Due to the limited number of varieties ($n = 8$), results are reported as raw correlation coefficients and p-values. This approach works at the variety level rather than the individual seed

level, and is therefore independent of the within-variety phenotype permutation used in the OPLS pipeline.

2. Linear mixed models - *lme4* v.2.0.1, [31] - metabolite ~ t50germ + (1|Variety)) testing the within-variety phenotype association while explicitly controlling for variety as a random effect.
3. Elastic net regression - *glmnet* v.5.0, [32], α optimised by cross-validation- performing independent sparse variable selection from the full metabolome, testing whether OPLS-identified candidates are also among the most parsimonious predictors of t50germ when all 221 metabolites compete simultaneously for inclusion in a penalised regression model.

Seed microbiota analysis

Seed preparation, DNA extraction, library construction and sequencing

Seed microbiota analyses were conducted by metabarcoding on three replicates of macerate from batches of 30 seeds each, for each variety.

Seed batches were rinsed three times with sterile water to remove dust or debris on the surface of the seeds before being soaked in 2 mL per gram of seeds of Phosphate Buffer Saline sterile solution (Sigma-Aldrich, Saint-Louis, MO, USA) supplemented with 0.05% [v/v] of Tween 20 (Sigma-Aldrich), under agitation (220 rpm) at 6°C for 16 h. Macerates were centrifuged in 15 mL tubes (4,000 g, 15 min, 20°C), pellet was resuspended in 1 mL supernatant and transferred in a 1.5 mL tube, and stored at -20°C until extraction. DNA was extracted from the macerate using NucleoSpin® 96 Food kit (Macherey-Nagel, Düren, Germany), following the manufacturer's instructions. Bacterial and fungal DNA were amplified through PCRs with a high-fidelity Taq DNA polymerase (AccuPrime Taq DNA Polymerase System, Invitrogen, Carlsbad, CA, USA), with primers for a portion of the bacterial *gyrB* gene (*gyrB_aF64/gyrB_aR353*) [33] and primers for region 1 of the fungal internal transcribed spacer of nuclear DNA (ITS1F/ITS2) [34], respectively. The first PCR was performed with 5 µL of 10X buffer, 1 µL of forward and reverse primers (100 µM for *gyrB* and 10 µM for ITS), 0.2 µL of Taq and 5 µL of DNA. Cycling conditions were such : initial denaturation step at 94°C (3 min), then 35 cycles of amplification at 94°C (30 s), 55°C for *gyrB* and 50°C for ITS (45 s), and 68°C (90 s), then a final elongation step at 68°C (10 min). Amplicons were purified with magnetic beads (Sera-Mag™, Merck, Kenilworth, New Jersey, USA). A second PCR was performed to incorporate Illumina adapters and barcodes. Cycling conditions were such: initial denaturation step at 94°C (1 min), then 12 cycles of amplification at 94°C (1 min), 55°C (1 min), and 68°C (1 min), then a final elongation step at 68°C (10 min). Amplicons were again purified with magnetic beads, quantified with QuantIT PicoGreen dsDNA Assay Kit (Invitrogen) and pooled in equimolar concentration. Concentration of the pool was measured via quantitative PCR (KAPA Library Quantification Kit, Roche, Basel, Switzerland). Amplicon libraries were mixed with 5% PhiX and sequenced with MiSeq reagent kits v3 600 cycles (Illumina, San Diego, California, USA) for 2 x 250 bp on the Illumina MiSeq platform (ANAN platform, SFR QuaSav, Angers, France). A blank extraction kit control, a PCR-negative control, and a PCR-positive control (*Lactococcus piscium* DSM6634, a fish pathogen that is not plant-associated) were included in each PCR plate. The raw amplicon sequencing data are

available on the European Nucleotide Archive (ENA) with the accession number PRJEB59579 (files named V[variety number]S1, with variety number from 1 to 8: Deezer, Caprice, Linex, Vanilla, Facila, Vezer, Contender, Flavert).

Sequencing data processing

Sequenced read integrity was verified for file integrity with gzip, then processed using QIIME2 v2019.7 [35]. Primer trimming and read demultiplexing were performed with *cutadapt* 2.7 plugin. Denoising (including merging for *gyrB* marker), chimera filtering and Amplicon Sequence Variants (ASVs) inference were conducted using the DADA2 (v1.10.0) [36] plugin within QIIME2 with max-ee = 2. For the *gyrB* gene, truncation parameters were set to 180 bp (forward) and 160 bp (reverse). For the ITS marker, the single-end mode using the forward reads only was used without truncation nor trimming as recommended by [34]. ASV taxonomy was assigned using a feature-classifier based on Naïve Bayes algorithm [37, 38] based on the UNITE database for ITS1 (UNITEv10 (tgz), version 19.02.2025) [39]; and on our local *gyrB* database (http://.fa.gz/train_set_gyrB_v5 [40]. For marker *gyrB*, ASVs assigned to "parE" (a *gyrB* paralog) or unclassified at the phylum level were excluded from further analysis. For marker ITS, ASVs not classified as Fungi (e.g., Viridiplantae, Protista) or unassigned were excluded. The resulting feature tables were filtered to retain ASVs present in at least two samples with a minimum number of 20 reads. Samples with fewer than 1000 total reads were excluded to mitigate bias due to low sequencing depth. Unrarefied and rarefied ASV tables, representative sequences, and associated taxonomic classifications were exported to R for downstream ecological analysis.

Statistical analyses of microbial community composition

Microbiota analyses were conducted with the Phyloseq package v1.54.2 [41]. Alpha rarefaction curves were generated to assess sequencing depth sufficiency, and the datasets were rarefied for alpha diversity metric calculations to the minimal sampling depth (1069 reads per sample for *gyrB* marker, 7435 reads per sample for ITS marker). α -diversity (richness, Chao1, Shannon) was estimated on rarefied data and differences between varieties tested with Kruskal-Wallis tests followed by Dunn's post-hoc comparisons with Benjamini-Hochberg correction. β -diversity was assessed on unrarefied data filtered at 5% prevalence and transformed with robust centered-log ratio, using PCoA on Euclidean distances for visualisation and PERMANOVA (999 permutations, `vegan::adonis2`) for statistical testing, with homogeneity of dispersions verified beforehand using `vegan::betadisper`. Taxonomic composition was visualised at genus level from non-rarefied counts filtered at 5% prevalence, with genera below 1% maximum relative abundance grouped as 'Other'. Those classical microbial analyses, along with their results, are detailed in **Note S1**.

Associations between variety-mean germination and emergence (mean t50germ, meanFEP) and seed microbiota composition were assessed for bacterial (*gyrB*) and fungal (ITS) communities separately, on the non-rarefied read counts filtered by prevalence $\geq 15\%$ to avoid taxa present only in replicates of one variety. The effect of community composition on t50germ, and on FEP, was tested with PERMANOVA on the rCLR-transformed Euclidean distance matrix with 999 permutations. As germination speed and

emergence percentage were nearly uncorrelated (Spearman $r = -0.05$) and showed no collinearity ($VIF < 2$ for both), the relation between microbiota and each variable was tested in separate models to maximize statistical power given the limited sample size ($n = 21$). Associations were visualized using distance-based redundancy analysis (db-RDA, *vegan::dbRDA*), with species scores assigned from the rCLR matrix (*vegan::sppscores*). The ten ASVs contributing most to the constrained axes were identified by their Euclidean distance from the origin in ordination space and labeled using the best available taxonomic assignment.

To identify individual ASVs whose relative abundance was significantly associated with seed germination and emergence performance, five complementary statistical approaches were applied to non-rarefied read counts filtered by prevalence ($\geq 15\%$ of samples):

1. Variety-level Spearman correlation (SpearmanVar) with BH correction, on per-variety means of microbial relative TSS-normalized abundances, and mean t50germ or mean FEP values, tested if the correlation held at the variety level,
2. Within-variety permutation Spearman correlation (SpearmanPerm) with BH correction, on 100 random within-variety permutations of microbiota data and t50germ or FEP values, tested if samples with higher relative abundance of this taxon tend to have lower t50germ and higher FEP. A taxon was retained as a candidate if it met simultaneously: BH-adjusted $q < 0.25$ in $\geq 70\%$ of permutations, and a consistent sign of association in $\geq 80\%$ of permutations. This approach mirrors the permutation strategy applied in the metabolomics analysis.
3. Linear model-based association analysis (MaAsLin2, *Maaslin2* R package, TSS normalization, log transformation, Benjamini-Hochberg correction, same within-variety permutation design), tested if there was a linear association between the log-transformed relative abundance of a taxon and t50germ or FEP, after accounting for compositionality through normalisation,
4. Compositional correlation analysis (ALDEx2 v1.44.0, Monte Carlo Dirichlet sampling, 128 instances, Spearman correlation test) tested if the association between a taxon and germination performance was robust to the stochastic uncertainty inherent in compositional count data, when abundance is expressed relative to the geometric mean of all taxa.
5. Finally, bias-corrected compositional differential abundance analysis (ANCOMBC2, *ANCOMBC* R package) tested if the association between a taxon and variety-mean t50germ or FEP was robust to differences in total microbial load between samples.

A relaxed significance threshold of $q \leq 0.25$ was applied across all methods given the limited effective sample size ($n = 8$ varieties), and results are interpreted as exploratory. To complement initial taxa affiliation, ASVs of candidate indicator taxa that were not resolved below phylum level by the SILVA/UNITE-DADA2 taxonomy pipeline were subjected either: for bacteria, to sequence-based identification by BLASTn against the NCBI non-redundant nucleotide database (bacteria, taxid:2) with identity thresholds of $\geq 97\%$, $\geq 95\%$, and $\geq 90\%$ used for species-, genus-, and family-level assignments, respectively; for fungi, to sequence-based re-identification using the DADA2 assignTaxonomy function against the UNITE general FASTA release for Fungi v.10.0, with the minimum bootstrap threshold set to

30 and reverse complement matching enabled (tryRC = TRUE). Bootstrap confidence values (0-100) reflecting the consistency of k-mer based assignment across 100 resamples of each query sequence were used to assign taxonomy at each rank, with thresholds of ≥ 80 , ≥ 60 , and ≥ 40 applied for species-, genus-, and family-level assignments, respectively. Assignments below the corresponding threshold were treated as unreliable.

Cross-method consistency was assessed by building a comparison table recording, for each taxon-variable combination, which methods identified a significant association and the corresponding effect sizes (Spearman rho from SpearmanVar and SpearmanPerm, MaAsLin2 log fold-change, ALDEx2 Spearman rho, ANCOM-BC2 log fold-change). Taxa confirmed by at least two independent methods were considered robust candidates and retained for further visualization.

To better understand the microbiota structure, a co-occurrence network approach was used. Analyses were performed on non-rarefied read counts filtered by prevalence ($\geq 15\%$ of samples) for both bacterial (*gyrB*, $n = 96$ ASVs) and fungal (ITS, $n = 93$ ASVs) communities.

The co-occurrence of bacterial and fungal taxa was inferred with a sparse co-abundance network, using the SPIEC-EASI algorithm - *SpiecEasi* R package, v1.99.0 [42] - that estimates sparse inverse covariance matrices from compositional microbiome data using Meinshausen-Bühlmann neighbourhood selection method fitting sparse regression models for each node against all others. For multi-kingdom analysis, we used the *multi.spiec.easi* function, which handles compositionality within each domain separately while estimating cross-domain associations jointly. Network sparsity was optimized using the StARS (Stability Approach to Regularization Selection) criterion on 20 values of λ (minimum ration: 10^{-2}), with 50 subsampling repetitions, a stability threshold of 0.05, and a subsampling ratio of 0.8. The optimal regularization parameter was selected at the lambda value first exceeding the StARS stability threshold, resulting in a conservative sparse network appropriate for the limited sample size ($n = 21$). The optimal adjacency matrix was used to construct an undirected weighted network using igraph, where nodes represent ASVs and edges represent robust covariance between ASVs. Edges positive and negative weight values indicated respectively co-exclusion and co-occurrence relationships. Unconnected nodes were removed before further analysis. Modules (clusters of co-varying taxa) were identified using the Louvain community detection algorithm with *cluster_louvain* [43], which optimizes network modularity. Network layout was computed using the Fruchterman-Reingold algorithm with absolute edge weights.

To evaluate the specificities of candidate indicator taxa in the network, those taxa were first divided into five categories: associated with faster, or slower, germination speed, or higher, or lower, emergence percentage, or non-indicator. To assess whether taxa from the same association category tend to co-occur, the assortativity of taxa nodes (i.e., the more-than-random occurrence between taxa of the same association category) was computed on each category. Assortativity is a metric bound by $[-1, +1]$, where $+1$ means that every edge connects two nodes of the same category, 0 that edges connect same-category and different-category nodes at random, and -1 that every edge connects nodes of different categories. Assortativity was computed per category on edge subsets (global: all edges, within-kingdom

(bacteria-bacteria and fungi-fungi) or inter-kingdom (bacteria-fungi) edges) using *assortativity_nominal* from igraph v.2.3.1 [44].

Results

Seed germination and seedling emergence vary between bean varieties

Seed germination time courses were monitored for eight common bean varieties on pleated paper in a germination bench (Fig. 1A-B). Germination time curves (Fig. 1A) were used to compute final germination percentage (FGP, **Figure S1A**) and fitted to derive the time to 50% of germinated seeds for each replicate ($t_{50\text{germ}}$) and its variety average (mean $t_{50\text{germ}}$) (Fig. 1B, **Figure S1B**). The varieties are separated into two groups. On the one hand, Vezer, Flavert, Deezer, and Facila showed the fast germination, with short lag phases before first germination (22–38 h), lower mean $t_{50\text{germ}}$ (32–50 h), and high germination uniformity (17–30h time window between 10% and 90% germination). On the other hand, Vanilla, Caprice, Linex, and Contender showed longer lag phases (37–48 h), higher mean $t_{50\text{germ}}$ (54–67 h) and lower uniformity (30–50 h). Mean final germination percentage (mean FGP) also differed between varieties: the four faster-germinating varieties had a mean FGP value ranging from 85.4% to 100%, while amongst slow varieties Contender was notable in reaching a high mean FGP despite its slow germination kinetics.

Seedling emergence was followed during seven days in potting soil (Fig. 1C). Mean final emergence percentage (mean FEP) differed significantly between varieties, ranging from 54% (Vanilla) to 96% (Contender) (Fig. 1D). Mean $t_{50\text{germ}}$ and mean FGP were not correlated across varieties (**Figure S1C**) whereas mean FGP was highly positively correlated with mean FEP ($R^2 = 0.89$, **Figure S1D**). This suggests that the final germination percentage, but not the germination speed, is a determining factor of seedling emergence of varieties under standard conditions. Based on these results, the mean $t_{50\text{germ}}$ and the mean FEP were selected as response variables to which seed metabolic and microbiological characteristics may be associated.

Variety colors reflect FEP ranking and are consistent across figures. Different letters indicate significant differences ($P < 0.05$).

Between-variety seed size influences germination speed; intra-variety seed weight variance reflects seedling emergence

Seed weight and size (i.e. projected area) were characterized for individual seeds before imbibition, and their influence on variety seed germination and seedling emergence variability was explored. The eight varieties showed substantial diversity for both traits (Fig. 2, **Figure S2A-B**). Variety means ranged from 94 to 522 mg for weight and from 38 to 97 mm² for area. Caprice, Contender and Linex presented the largest and heaviest seeds and the largest intravariety variance for both traits. Conversely, Vanilla and Vezer had the smallest and lightest seeds, with a small variance.

Between-variety but not within-variety seed size predicts germination speed

When varieties were compared (inter-variety analysis), seed weight and projected area were positively correlated with germination speed ($R^2 = 0.94$ and $R^2 = 0.62$ between mean t50germ and mean seed weight or seed area, respectively; **Figure S2D-E**). This indicated that larger and heavier seeds tended to germinate more slowly than small and light seeds. However, Spearman correlations between individual seed weight and germination time within each variety showed a significant positive correlation only in Facila ($\rho = +0.37$, $p = 0.001$) and Vanilla ($\rho = +0.31$, $p = 0.019$) (**Table S2**). No significant correlation was observed in the other six varieties, even for those presenting the largest seeds (Caprice, Linex, and Contender) which drive the relationship between seed size and mean germination speed when comparing varieties (**Figure S2C**).

Intra-variety variance of seed weight reflects final emergence percentage

Unlike germination speed, neither seed weight nor projected area were significantly correlated with emergence (mean FEP) across varieties (**Figure S2F-G**). Given the high disparity in seed weight distribution between varieties, the coefficient of variation of individual seed weight within each variety (referred to as cv_weight) was explored further. Cv_weight did not correlate with mean seed weight itself ($\rho = +0.24$, $p = 0.570$), nor with mean germination speed ($\rho = +0.21$, $p = 0.610$, **Figure S2H**), but was significantly positively correlated with mean FEP across varieties ($\rho = +0.83$, $p = 0.010$, **Figure S2I**, **Table S3**). In other words, varieties with more heterogeneous seed weight distributions, such as Contender and Flavert, tended to achieve higher final emergence. Conversely, Vanilla, the variety with the most homogeneous seed weights (cv_weight = 0.17), showed the lowest emergence percentage (53.75%). Collectively, these results suggest that mean-variety seed weight and size are related to germination speed but unrelated to seedling emergence capacity. Rather, seedling emergence capacity appeared to be linked to intra-variety seed weight heterogeneity.

Metabolome composition reflects germination potential

Metabolome composition differs strongly across seed organs

Untargeted GC-MS metabolomics was performed on individual seeds dissected into three organs: cotyledon, hypocotyl-radicle axis (HR-axis), and plumule. After quality filtering (removal of samples with null or aberrant internal standard abundance) and normalization by internal standard, 252 annotated metabolites were retained prior to prevalence filtering, for 1 to 12 replicates per organ per variety (**Table S4**).

Principal component analysis (PCA) showed clear separation of organ type (cotyledon on the one hand, other organs on the other hand) along the first two principal components (PC1: 24.0%, PC2: 6.3% of total variance; Fig. 3A). Organ-level clustering was consistent across varieties, indicating that organ-specific metabolic signature was independent of the genotype. PERMANOVA on Euclidean distances confirmed that organ identity explained 19.4% of total metabolomic variance ($F = 16.2$, $p = 0.001$), with this

separation concentrated in the major axes of variation (PERMANOVA on PC1-PC2: $R^2 = 0.58$, $F = 92.2$, $p = 0.001$). The multivariate dispersion differed significantly across organs (betadisper: $F = 13.6$, $p = 0.001$), with cotyledon samples being substantially more homogeneous than HR-axis or plumule samples (average distance to group median: 8.98, 13.93, and 14.98 respectively; pairwise tests: Cotyledon vs HR-axis $p = 0.001$, Cotyledon vs Plumule $p = 0.001$, HR-axis vs Plumule $p = 0.424$). That is, there were differences on average between organs along with greater within-organ metabolic variability in the HR axis and plumule, relative to cotyledons.

Metabolites significantly enriched in an organ compared to at least one other organ were identified by Kruskal-Wallis tests and Dunn pairwise tests with BH correction. Twelve analytes were enriched in **cotyledons**: central carbon metabolites including sugar phosphates (malic acid, ribose-5-phosphate, ribulose-5-phosphate, pyruvic acid, saccharic acid, erythrose-4-phosphate, fructose-6-phosphate), aromatic compounds (shikimic acid, 2,6-dimethylaniline), and *N*-acetyl-L-alanine (**Table S5**). Six metabolites were enriched in the **HR-axis**, including the allantoin (purine catabolism product), methionine, sinapinic acid, and D-ribose-5-phosphate. Twelve metabolites were enriched in the **plumule**, including compounds of nitrogen metabolism (2,3-diaminopropanoic acid, 4-aminobutanoic acid, cysteine, citrulline, methionine sulfone), adenine, sucrose, and metabolites of various pathways (heptadecanoic acid, urocanic acid, citramalic acid, maleamate, allocholic acid).

Variety drives metabolome composition within each organ

Separate PCA for each organ, colored by variety, showed clear clustering between varieties (Fig. 3B). Within-organ PERMANOVA (**Table S6**) revealed that variety was a main driver of metabolome composition of the three organs, explaining 51.8% of metabolomic variance in cotyledons ($F = 4.61$, $p = 0.001$), 41.3% in plumules ($F = 3.81$, $p = 0.001$) and 27.5% in the HR-axis ($F = 2.44$, $p = 0.001$), with homogeneous dispersion across varieties in all organs (betadisper: Cotyledon $F = 1.96$, $p = 0.094$; HR-axis $F = 0.91$, $p = 0.493$; Plumule $F = 2.07$, $p = 0.075$). Since varieties also differ in their germination and emergence performance, we tested whether these traits were directly associated with metabolome composition, despite the PCA not displaying a clear gradient in mean t50germ or mean FEP along the major axes of variation. PERMANOVA analyses revealed that germination speed (mean t50germ) was significantly associated with metabolome composition of the three organs (mainly in cotyledons: $R^2 = 0.080$, $F = 3.13$, $p = 0.001$; plumule: $R^2 = 0.056$, $F = 2.59$, $p = 0.001$ and then HR-axis: $R^2 = 0.039$, $F = 2.07$, $p = 0.020$). As for mean FEP, it was only significantly associated with the metabolome composition of plumule ($R^2 = 0.067$, $F = 3.15$, $p = 0.001$) and HR-axis ($R^2 = 0.033$, $F = 1.75$, $p = 0.045$).

Individual seed metabolome associates consistently with germination speed but not with emergence. Individual metabolites significantly associated with germination speed (t50germ) and emergence (FEP) were investigated through orthogonal partial least squares regression (OPLS). Metabolite significance was evaluated by concurrent multivariate and univariate analysis. OPLS regression models showed good to excellent predictive ability for t50germ across all organs (Plumule: $R^2Y = 0.958 \pm 0.018$, $Q^2Y = 0.782 \pm$

0.035; Cotyledon: $R^2Y = 0.911 \pm 0.054$, $Q^2Y = 0.695 \pm 0.045$; HR-axis: $R^2Y = 0.846 \pm 0.023$, $Q^2Y = 0.570 \pm 0.041$). Models for FEP showed moderate predictive ability (**Table S7**).

Many metabolites showed associations with FEP in the mean volcano plots (51 in cotyledon, 11 in HR-axis, 20 in plumule, **Table S8**). Amongst them, we kept only those with a consistent loading across permutations in the dataset (see Material and methods). We thereby obtained three metabolites in cotyledons and 16 in plumules (listed in Fig. 4B; **Table S9**). In cotyledons, p-coumaric acid and arachidic acid were negatively associated with t50germ, indicating higher abundance in faster-germinating seeds, while cystine was positively associated with t50germ, indicating higher abundance in slower-germinating seeds. In plumules, eleven metabolites were positively related to t50germ (higher abundance in slower-germinating seeds), including fatty acids (petroselinic, heptadecanoic, stearic, and palmitic acids), allose, sinapinic acid, and β -gentiobiose. Five metabolites were negatively associated with t50germ (higher abundance in faster-germinating seeds), including thymine, cytosine, and putrescine. Robustness verification (via Spearman correlations, linear mixed models, and elastic net regression) confirmed p-coumaric acid as being the most robust marker of t50germ in cotyledons, and deoxycholic acid, stigmasterol, putrescine, and heptadecanoic acid in plumules.

Several metabolites showed associations with FEP in the mean volcano plots (four in cotyledon, one in HR axis, 18 in plumule, Fig. 4A, **Table S8**), but none reached the permutation frequency threshold required to ensure reliability, indicating that metabolite-FEP associations are present but unsteady across the whole dataset.

Seed microbiota reflects variety's identity and emergence performance

Seed microbiota is dominated by a few bacterial and fungal genera

Illumina sequencing yielded an average of 45,050 raw reads of *gyrB* gene sequences and 29,445 of ITS gene sequences, respectively. The sequences were filtered and merged to an average of 8,552 final bacterial and 16,374 final fungal reads per sample (1,069 and 7,435 after rarefaction), from which 1,579 bacterial and 290 fungal ASVs were identified (884 and 167 after rarefaction).

The most prevalent genera were *Pantoea*, *Pseudomonas*, *Kosakonia*, and *Erwinia*, comprising in average more than 54% of the relative abundance in the bacterial communities, and *Alternaria*, *Cladosporium*, and *Stemphylium*, comprising in average more than 90% of the relative abundance in the fungal communities (**Figure S4.A-D**).

Fungal, but not bacterial, community composition is variety-dependent

α -diversity varied significantly between varieties for both bacterial and fungal communities (Kruskal-Wallis, $p = 0.034$ and $p = 0.025$, respectively). β -diversity analysis revealed contrasting patterns (Fig. 5A-B): variety had no significant effect on bacterial community composition (PERMANOVA: $R^2 = 0.37$, $p = 0.4$) but strongly structured fungal communities ($R^2 = 0.80$, $p = 0.001$), though the low number of

replicates per variety (n = 2–3) prevented identification of individually significant variety pairs. For further details on general microbiota composition, see **Note S1**.

Fungal communities associate with both germination speed and emergence

Seed bacterial community composition (*gyrB*) was not significantly associated with mean t50germ (PERMANOVA: $F = 1.50$, $R^2 = 0.073$, $p = 0.238$) but was significantly associated with mean FEP (PERMANOVA: $F = 3.18$, $R^2 = 0.143$, $p = 0.022$). In contrast, fungal community composition (ITS) was significantly associated with both traits (PERMANOVA: $F = 3.01$, $R^2 = 0.137$, $p = 0.049$; and $F = 3.58$, $R^2 = 0.159$, $p = 0.042$, respectively). db-RDA biplots were used to visualize these associations, with the ten ASVs most strongly driving the constrained axes identified and labeled by their lowest available taxonomic assignment (Fig. 5C-F, **Table S10**).

These results suggest that seed-associated fungal communities are more broadly linked to germination speed and emergence performance than bacterial communities, which appear specifically associated to FEP rather than to t50germ.

Individual fungal taxa associate with emergence rather than germination speed

Individual microbial taxa associated with germination and emergence traits were identified in an exploratory analysis with five independent statistical methods (Spearman rank correlation on replicates or on variety means, linear model-based association analysis, compositional correlation analysis, and bias-corrected compositional differential abundance analysis). Across the five statistical methods applied, a total of 27 taxa reached the exploratory significance threshold ($q \leq 0.25$) (**Table S11**). Of these, four bacterial taxa were confirmed by at least two independent methods and four fungal taxa were confirmed by at least three independent methods, and retained as robust candidates (Fig. 6). For fungal communities (ITS1), a *Stemphylium trifolii*, one unassigned fungi and an unassigned *Saccharomycetales* ASVs showed strong positive associations with FEP ($\rho = 0.73$, 0.72 and 0.65), one *Acremonium* ASV showed a negative association with FEP ($\rho = -0.60$). No fungal ASV was detected as associated with low t50germ. For bacterial communities (*gyrB*), an unassignable *Actinomycetes* ASV negatively associated with mean t50germ ($\rho = -0.57$, confirmed by Spearman and ALDEx2), suggesting faster germination in seeds harbouring this taxon. *Kosakonia cowanii*, one *Erwinia* ASV, and a *Reyranella soli* were positively associated with mean FEP, as confirmed by Spearman correlation and ANCOM-BC2 or mean Spearman correlation ($\rho = 0.68$, 0.66 , and 0.84 respectively). As such, seeds harbouring higher abundances of these bacteria tended to germinate faster.

Notably, *Kosakonia cowanii* was identified both as a robust indicator taxa associated with mean FEP from two methods and as one of the 10 taxa most strongly driving the db-RDA axis associated with mean t50germ and mean FEP. A one-method candidate for FEP *Cladosporium herbarum* (ALDEx2 only, $\rho = -0.69$ negative) also appears to drive both fungal dbRDA ordination.

Fungi with similar associations to germination or emergence preferentially co-occur

To evaluate the tendency of taxa to connect preferentially to other taxa with similar association with germination or emergence properties, a multi-kingdom co-occurrence network was inferred (stability = 0.0345, optimal lambda = 0.606). After removal of isolated nodes, 168 taxa remained in the connected network with 194 edges (93 bacteria, 75 fungi, Fig. 7A). Edge weights ranged from - 0.227 to 0.375, with positive weights indicating positive covariance and negative weights indicating exclusion between taxa (**Table S12**).

Taxa correlating with FEP or t50germ were distributed across multiple modules. Fungal indicator taxa of the same association category (associated with faster or slower t50germ, or higher or lower FEP) tended to appear in the same modules (dark red M6 on top of Fig. 7A, light-blue M23 on left side, Fig. 7B, **Table S12**). Also, bacterial taxa associated with both t50germ and FEP, or with either but with different signs, tended to appear in the same modules (turquoise M3 on the bottom-right and M5, orange M8, % on the bottom). The mixed modules (orange/red : containing both mean t50germ and mean FEP-related taxa) suggest that some community members had a role in both germination and emergence traits. Note that one candidate taxa associated with higher FEP, *Kosakonia cowanii*, was removed from the network as it was not connected to any other taxa.

Assortativity coefficients (Fig. 7C), all positive, revealed that indicator taxa co-occur more with other indicator taxa of the same category than expected by chance, especially for high FEP-associated fungi (assortativity = 0.383). Taxa associated with fast germination co-occurred more strongly by kingdom (4 bacteria, assortativity 0.202; 2 fungi, assortativity 0.370) than inter kingdom (0.071). Assortativity coefficients were similar across categories.

Discussion

Germination speed and seedling emergence capacity are independent in common bean

The distinction between germination and emergence is well established, with extensive experimental and conceptual work showing that these two phases rely on partly distinct underlying processes and determinants in crop and model species [45–48], and that rapid germination does not guarantee successful seedling establishment [49, 50]. Contender, the slowest-germinating variety achieving the highest emergence percentage (95.8%), illustrates this functional decoupling and highlights the risk of using germination speed as the sole proxy for seed quality in field contexts. Building on this framework, we investigated quantitative indicators in common bean associated with early germination processes (metabolic activation, radicle protrusion) on the one hand, and traits associated with seedling emergence (hypocotyl elongation, ability to overcome soil resistance, and early seedling establishment) on the other hand. The aim of this study was to identify practical, physiologically grounded indicators that could be used to predict each phase, rather than relying on a single measure of seed quality that

aggregates germination and emergence. To do so, we used accurate, individual-seed weight and morphological measurement, a fine individual-seed per-organ metabolome characterisation, and seed batch microbiota analysis for eight varieties of common bean (*Phaseolus vulgaris* L.). We took the precaution to study seeds cultivated in the same assay. The eight varieties, representative of the morphological and genetic diversity of cultivated varieties in France, display a wide range of germination speeds and emergence capacities we used to identify proxies of seed quality. Given the agronomical importance of such proxies, their strength and reliability of those proxies should be further evaluated in and compared across various conditions, in germination and emergence assays, to emulate field conditions.

Seed size predicts germination speed between varieties, while intra-variety size variance predicts emergence

When varieties were compared, we found that seed weight and size were strong predictors of germination time (e.g. weight vs mean t50germ: $R^2 = 0.94$), which is consistent with reports in other legume species where large seed mass leads to slower water uptake kinetics and longer imbibition times [51–54]. We nevertheless recognise that it is not a simple correlation and some legume species are associated with an opposite, or more complex, relationship [47]. Also, the relationship between seed weight and t50germ was not found anymore when individual seeds, rather than variety averages, were considered. Thus, it seems that in common bean, the mass-germination time relationship was a statistical property at the variety scale, rather than a reflection of intrinsic mechanical effects of seed mass on, e.g., imbibition. In other words, differentiating varieties encompasses not only morphological differences (such as average seed weight), but also physiological differences that dictate germination timing in this species.

Quite surprisingly, a positive correlation was found between intra-variety seed weight variation and mean FEP ($\rho = +0.83$, $p = 0.010$), independent of mean seed weight itself ($\rho = +0.24$, $p = 0.570$) (**Figure S2I**). From a physiological perspective, it might reflect a decoupling during seed development, between reserve accumulation and cotyledon development on the one hand, and the biosynthesis of cellular actors involved in germination processes (such as rRNA accumulation, hormone content, etc.) on the other hand. That is, varieties associated with high FEP give priority to the development of the molecular machinery for germination, regardless of natural seed biomass variations, due to, e.g., variable phloem transport rate, pod position on the plant, or seed position within the pod [27]. From an ecological point of view, it suggests that varieties producing more heterogeneous seed populations are better adapted to unpredictable soil conditions during emergence. This is in line with the concept of “bet-hedging” in seed biology, whereby variability in morphology and developmental stage within a seed population ensures a higher success rate of emergence under variable environmental conditions [55, 56].

Metabolomes of seed organs associate with germination speed but not emergence capacity

As for proteins and transcripts, metabolites accumulated during seed maturation and preserved during desiccation can play a role in germination [57]. Seed organs have different physiological functions, reflected in their composition [58–60]. Here, the three dissected seed organs exhibited distinct metabolomic profiles that were consistent with their developmental functions. The cotyledons were enriched in metabolites of primary carbon metabolism including sugar phosphates (reflecting sugar interconversions and starch accumulation), as well as pyruvate and malate (glycolysis and Krebs cycle) and *N*-acetyl-alanine (stabilisation of accumulated proteins) [61]. The hypocotyl-radicle axis was enriched in allantoin, which is a nitrogen transport metabolite in French bean [62], and methionine, a sulphur transport molecule resynthesised from *S*-methylmethionine in legume developing seeds [63]. The plumule was enriched in amino acids, adenine, polyamine (citrulline) and sucrose, reflecting metabolism sustaining leaf cell division and growth [13, 64].

Within each organ, variety was a dominant driver of the metabolome ($R^2 = 0.28\text{--}0.52$), confirming that seed metabolomic profiles were primarily dictated by the genotype, as found in tomato [7, 65] but not in *Medicago* [27]. Also, individual-seed metabolomic profiles in plumule and cotyledon were significantly related with germination speed but not emergence performance, suggesting that these two traits have different metabolic mechanisms. That is, germination speed probably relies more on compounds inherited from the dry seed while emergence depends more on post-germination metabolism, including metabolic pathways during reserve remobilisation. Interestingly, putrescine was identified as the most robust plumule biomarker of fast germination regardless of the statistical method used. It has been shown that putrescine plays a role in stress mitigation [66], seed germination [67–70] and growth of young soybean seedlings [71]. Thymine and cytosine were also biomarkers of fast germination alongside putrescine, and probably reflect cell division and thus nucleotide synthesis capability. Conversely, free saturated fatty acids (heptadecanoic, palmitic, and stearic acids) and stigmasterol were more abundant in plumules of slower-germinating seeds, perhaps reflecting increased cell membrane rigidity, improper fatty acid composition balancing, or altered carbon partitioning towards lipid at the expense of carbohydrates and proteins. A link with membrane properties would be consistent with the documented role of membrane fluidity in controlling turgor-driven cell expansion required for radicle emergence [72] and the key role of changes in fatty acid composition during imbibition, as found in *Medicago* [73]. Altogether, these results suggest that plumule metabolome can be a source of germination speed indicators, and highlight the distinction between germination speed and emergence capacities. The last could depend on factors beyond the seed chemistry, including seed microbiota composition.

Fungal community composition is variety-specific and broadly linked to germination performance

Seed microbiota role in germination and seedling emergence is a recent field: a growing number of studies focus on seed microbiota characterisation, while few link that diversity to plant development [74–76]. In the present study, fungal community composition was strongly variety-dependent (PERMANOVA $R^2 = 0.80$), whereas bacterial community composition showed no significant variety effect ($R^2 = 0.37$, $p = 0.4$). The higher impact of host variety, or species, on fungal communities has already been described [77–79]. This contrast may reflect differences in transmission mode and community assembly, although these mechanisms are yet poorly understood. The dominance of *Pantoea*, *Pseudomonas*, *Kosakonia* and *Erwinia* in the bacterial communities, and of *Alternaria*, *Cladosporium* and *Stemphylium* in the fungal communities, is consistent with the core seed microbiota composition reported across crop species in a meta-analysis [79].

The composition of the fungal community was significantly associated with both germination speed and emergence (PERMANOVA $R^2 = 0.137$ and 0.159), while bacterial communities were associated with emergence only ($R^2 = 0.143$, $p = 0.022$). This functional dissociation suggests that fungi may be more tightly coupled to the biochemical activation of germination, possibly through interactions with seed coat chemistry or plumule metabolism, while bacteria could influence the later stages of seedling establishment and soil emergence, consistent with their documented roles in phytohormone production, phosphate solubilisation and induced systemic resistance [80–82]. Such fungal-tegumen interaction could be further studied in relation to the seed coat metabolome, which we were not able to analyse in this study. The positive association of *Saccharomycetales/Geotrichum* ASVs with final emergence ($p = +0.73$ and $+0.72$) could incite further research on the growth-promoting potential of seed yeasts [83, 84]. One *Kosakonia cowanii* ASV was found positively associated with emergence ($p = +0.68$, two independent methods) and identified as a top driver of the db-RDA community axes, providing evidence from community-level and taxon-level analyses. This is coherent with the plant-growth promoting potential reported for this species [85]. Similarly, a *Cladosporium herbarum* ASV, associated with low FEP ($p = -0.69$, 1 method), was in the top db-RDA taxa. This species has been shown to negatively impact *Phaseolus vulgaris* development [86, 87]. Those associations should be further investigated, for example through seed inoculation [88].

Co-occurrence networks are useful to study microbial ecology and inter-individual correlation or covariance from abundance data. Although they do not reflect direct, biotic interaction, the edges that link taxa nodes allude to facilitation, mutualism, and niche sharing for positive co-occurrence or to the indirect consequences of competition, predation, and environmental filters for negative co-occurrence [89–92]. In our case, the cross-domain co-occurrence network revealed that the seed microbiota is structured with respect to germination performance: fungi with similar emergence capacity associations preferentially co-occur. Such association of several candidate indicator taxa in the same modules weights their value as individual indicator, but also of the whole module as a strong proxy to evaluate the seeds emergence capacities. Indeed, identification of microbial modules associated with ecosystemic functions is a promising research avenue to both find indicators and better understand microbiota contribution to those functions [93–95]. Such modules' role in seedling emergence can then be validated

with inoculation approaches, using microbial consortia (previously “synthetic communities”) inoculation on seeds to infer causal relationships in plant-microbiota interactions [96, 97].

Conclusion

Taken together, our results demonstrate that germination speed and emergence capacity in common bean are governed by distinct and largely independent seed properties, each with characteristic morphological, metabolic and microbial signatures. Seed size and plumule metabolome composition were informative indicators of germination speed, while intravariety seed weight heterogeneity and fungal community composition were found informative for emergence. This multi-factor picture supports the development of complementary rather than single-proxy seed quality assessment frameworks, potentially integrating morphometric, metabolomic and microbiota measurements on dry seeds prior to sowing [98, 99]. Future research involving larger variety panels, individual seed phenotyping designs, and spermosphere analysis at the seed-to-soil interface [100, 101] will be needed to validate the candidate indicators and assess their applicability across environments and agronomic contexts.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Funding

This research was conducted in the frame of the 3rd Programme for Future Investments (France 2030), operated by the SUCSEED project (ANR- 20- PCPA-0009) and funded by the ‘Growing and Protecting crops Differently’ French Priority Research Program (PPR-CPA), part of the national investment plan operated by the French National Research Agency (ANR).

Author Contribution

C.-E. K. directed the seed harvest. B.T., E.P., M.S., G.T., and M.B. designed the experiments. C.M., A.D., and M.-A.W. performed the germination assay. L.L.C., and S.H. performed the X-ray analyses. L.C.-S.

performed the emergence assay. E.P., C.A. dissected the dry seeds and prepared the metabolomic experiment, C.A. and J.L. ran the metabolomic experiment, and L.C.-S. processed the raw metabolomic data, with guidance from C.A. and G.T.. M.S. and C.M. performed the metabarcoding assay. L.C.-S. analyzed all experimental data. L.C.-S, G.T., B.T., E.P. and M.S. were involved in the interpretation of the results. L.C.-S wrote the article with B.T., E.P., G.T., and M.S. All authors read, corrected, and approved the final manuscript.

Acknowledgement

This research was conducted in the frame of the 3rd Programme for Future Investments (France 2030), operated by the SUCSEED project (ANR- 20- PCPA-0009) and funded by the 'Growing and Protecting crops Differently' French Priority Research Program (PPR-CPA), part of the national investment plan operated by the French National Research Agency (ANR). We thank Elise Vendeuvre for the gift of Linex, Deezer, Vezer, and Caprice commercial seeds used as mother-plants. We thank Fernand Roques, Elodie Gauvin, and Emmanuelle Laurent for the seeds they grew and provided, and Jérôme Lemaire for cleaning and sampling seed lots. We thank Didier Demilly and the Phenotic Angers Seed & Plant Phenotyping Facility (INRAE-IRHS, Angers University, Institut Agro, GEVES) for experiments on seed germination and seedling emergence.

Data Availability

The datasets generated and analyzed during current study are included as Supplementary material, or can be found with codes used for this manuscript analysis, on GitHub at this address:

<https://github.com/LounaCS/Multidimensional-Drivers-of-Seed-Germination-and-Emergence.git>

References

1. Corbineau F. Markers of seed quality: from present to future. *Seed Sci Res.* 2012;22:S61–8. <https://doi.org/10.1017/S0960258511000419>.
2. Rahman A, Cho B-K. Assessment of seed quality using non-destructive measurement techniques: a review | *Seed Science Research | Cambridge Core.* *Seed Sci Res.* 2016;26:285–305. <https://doi.org/10.1017/S0960258516000234>.
3. Don R, Ducournau S, Committee IG. *ISTA Handbook on Seedling Evaluation, Edition F.* 2018. 2018. <https://doi.org/10.7939/83911>
4. Mecha E, Erny GL, Guerreiro ACL, Feliciano RP, Barbosa I, Bento da Silva A, et al. Metabolomics profile responses to changing environments in a common bean (*Phaseolus vulgaris* L.) germplasm collection. *Food Chem.* 2022;370:131003. <https://doi.org/10.1016/j.foodchem.2021.131003>.
5. El-Maarouf-Bouteau H. The Seed and the Metabolism Regulation. *Biology.* 2022;11:168. <https://doi.org/https://doi.org/10.3390/biology11020168>.

6. Gonçalves J, de Gasparini P, Picoli K, de Costa EA, Araujo MD-BL, Zsögön WL. Metabolic control of seed germination in legumes. *J Plant Physiol.* 2024;295:154206. <https://doi.org/10.1016/j.jplph.2024.154206>.
7. Kazmi RH, Willems LAJ, Joosen RVL, Khan N, Ligterink W, Hilhorst HWM. Metabolomic analysis of tomato seed germination. *Metabolomics.* 2017;13:145. <https://doi.org/10.1007/s11306-017-1284-x>.
8. Domergue J-B, Abadie C, Limami A, Way D, Tcherkez G. Seed quality and carbon primary metabolism. *Plant Cell Environ.* 2019;42:2776–88. <https://doi.org/10.1111/pce.13618>.
9. Guo H, Lyv Y, Zheng W, Yang C, Li Y, Wang X, et al. Comparative Metabolomics Reveals Two Metabolic Modules Affecting Seed Germination in Rice (*Oryza sativa*). *Metabolites.* 2021;11:880. <https://doi.org/10.3390/metabo11120880>.
10. Rajjou L, Gallardo K, Debeaujon I, Vandekerckhove J, Job C, Job D. The Effect of α -Amanitin on the Arabidopsis Seed Proteome Highlights the Distinct Roles of Stored and Neosynthesized mRNAs during Germination. *Plant Physiol.* 2004;134:1598–613. <https://doi.org/10.1104/pp.103.036293>.
11. Weitbrecht K, Müller K, Leubner-Metzger G. First off the mark: early seed germination. *J Exp Bot.* 2011;62:3289–309. <https://doi.org/10.1093/jxb/err030>.
12. Sano N, Rajjou L, North HM. Lost in Translation: Physiological Roles of Stored mRNAs in Seed Germination. *Plants.* 2020;9:347. <https://doi.org/10.3390/plants9030347>.
13. Galland M, He D, Lounifi I, Arc E, Clément G, Balzergue S, et al. An Integrated Multi-Omics Comparison of Embryo and Endosperm Tissue-Specific Features and Their Impact on Rice Seed Quality. *Front Plant Sci.* 2017;8. <https://doi.org/10.3389/fpls.2017.01984>.
14. War AF, Bashir I, Reshi ZA, Kardol P, Rashid I. Insights into the seed microbiome and its ecological significance in plant life. *Microbiol Res.* 2023;269:127318. <https://doi.org/10.1016/j.micres.2023.127318>.
15. Romão IR, do Carmo Gomes J, Silva D, Vilchez JI. The seed microbiota from an application perspective: an underexplored frontier in plant–microbe interactions. *Crop Health.* 2025;3:12. <https://doi.org/10.1007/s44297-025-00051-6>.
16. Xu X-Y, Wang C-L, Xu J-Y, Dong C-J-H, Tan C, He Y-X, et al. Seed–microbiome interactions: Mechanistic insights and utilization toward seed performance for sustainable agriculture. *Plant Commun.* 2026;7. <https://doi.org/10.1016/j.xplc.2026.101716>.
17. Ahmed S, Singh P, Middleton J, Merritt D, Jenkins S, Nichols P. Impact of Seed Maturation on the Morphology, Nutrition, Microbiome Composition and Germinability of Subterranean Clover (*Trifolium subterraneum*) Seeds. *Grass Forage Sci.* 2025;80:e12725. <https://doi.org/10.1111/gfs.12725>.
18. Vidak M, Lazarević B, Javornik T, Šatović Z, Carović-Stanko K. Seed Water Absorption, Germination, Emergence and Seedling Phenotypic Characterization of the Common Bean Landraces Differing in Seed Size and Color. *Seeds.* 2022;1:324–39. <https://doi.org/10.3390/seeds1040027>.
19. Uebersax MA, Cichy KA, Gomez FE, Porch TG, Heitholt J, Osorno JM, et al. Dry beans (*Phaseolus vulgaris* L.) as a vital component of sustainable agriculture and food security—A review. *Legume Sci.* 2023;5:e155. <https://doi.org/10.1002/leg3.155>.

20. Boye J, Zare F, Pletch A. Pulse proteins: Processing, characterization, functional properties and applications in food and feed. *Food Res Int.* 2010;43:414–31. <https://doi.org/10.1016/j.foodres.2009.09.003>.
21. McDermott J, Wyatt AJ. The role of pulses in sustainable and healthy food systems. *Ann N Y Acad Sci.* 2017;1392:30–42. <https://doi.org/10.1111/nyas.13319>.
22. Ribeiro ND, dos Santos GG, Maziero SM, dos Santos GG. Genetic diversity and selection of bean landraces and cultivars based on technological and nutritional traits. *J Food Compos Anal.* 2021;96:103721. <https://doi.org/10.1016/j.jfca.2020.103721>.
23. Demilly D, Ducournau S, Wagner M-H. Digital imaging of seed germination. *Plant Image Analysis.* CRC; 2014.
24. Angers Plant Phenotyping Facility (APPF). PHENOTIC. 2023. <https://doi.org/10.17180/ykbz-2v85>.
25. Aravind J, Vimala Devi S, Radhamani J, Jacob SR, Srinivasan K, germinationmetrics. Seed Germination Indices and Curve Fitting. 2023. <https://doi.org/10.5281/zenodo.8267035>
26. Soltani E, Ghaderi-Far F, Baskin CC, Baskin JM. Problems with using mean germination time to calculate rate of seed germination. *Aust J Bot.* 2015;63:631–5. <https://doi.org/10.1071/BT15133>.
27. Domergue J-B, Lalande J, Beucher D, Satour P, Abadie C, Limami AM, et al. Experimental Evidence for Seed Metabolic Allometry in Barrel Medic (*Medicago truncatula* Gaertn). *Int J Mol Sci.* 2022;23:8484. <https://doi.org/10.3390/ijms23158484>.
28. Abadie C, Lalande J, Tcherkez G. Exact mass GC-MS analysis: Protocol, database, advantages and application to plant metabolic profiling. *Plant Cell Environ.* 2022;45:3171–83. <https://doi.org/10.1111/pce.14407>.
29. Gu Z. Complex heatmap visualization. *iMeta.* 2022;1:e43. <https://doi.org/10.1002/imt2.43>.
30. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. *J Proteome Res.* 2015;14:3322–35. <https://doi.org/10.1021/acs.jproteome.5b00354>.
31. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw.* 2015;67:1–48. <https://doi.org/10.18637/jss.v067.i01>.
32. Tay JK, Narasimhan B, Hastie T. Elastic Net Regularization Paths for All Generalized Linear Models. *J Stat Softw.* 2023;106:1–31. <https://doi.org/10.18637/jss.v106.i01>.
33. Barret M, Briand M, Bonneau S, Préveaux A, Valière S, Bouchez O, et al. Emergence Shapes the Structure of the Seed Microbiota. *Appl Environ Microbiol.* 2015;81:1257–66. <https://doi.org/10.1128/AEM.03722-14>.
34. Pauvert C, Buée M, Laval V, Edel-Hermann V, Fauchery L, Gautier A, et al. Bioinformatics matters: The accuracy of plant and soil fungal community data is highly dependent on the metabarcoding pipeline. *Fungal Ecol.* 2019;41:23–33. <https://doi.org/10.1016/j.funeco.2019.03.005>.

35. Bolyen E, Rideout JR, Dillon MR, Bokulich NA, Abnet CC, Al-Ghalith GA, et al. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol.* 2019;37:852–7. <https://doi.org/10.1038/s41587-019-0209-9>.
36. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods.* 2016;13:581–3. <https://doi.org/10.1038/nmeth.3869>.
37. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve Bayesian Classifier for Rapid Assignment of rRNA Sequences into the New Bacterial Taxonomy. *Appl Environ Microbiol.* 2007;73:5261–7. <https://doi.org/10.1128/AEM.00062-07>.
38. Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, et al. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome.* 2018;6:90. <https://doi.org/10.1186/s40168-018-0470-z>.
39. Abarenkov K, Nilsson RH, Larsson K-H, Taylor AFS, May TW, Frøslev TG, et al. The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: sequences, taxa and classifications reconsidered. *Nucleic Acids Res.* 2024;52:D791–7. <https://doi.org/10.1093/nar/gkad1039>.
40. Briand M, Rué O, Barret M. gyrB database for taxonomic assignment formatted for DADA2. 2025. <https://doi.org/10.57745/DD7RZ8>
41. McMurdie PJ, Holmes S. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE.* 2013;8:e61217. <https://doi.org/10.1371/journal.pone.0061217>.
42. Kurtz Z, Mueller C, Miraldi E, Bonneau R, SpieckEasi. Sparse Inverse Covariance for Ecological Statistical Inference. 2025.
43. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theory Exp.* 2008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>. 2008:P10008.
44. Csárdi G, Nepusz T, Müller K, Horvát S, Traag V, Zanini F et al. igraph for R: R interface of the igraph library for graph theory and network analysis. 2026. <https://doi.org/10.5281/zenodo.20103238>
45. Finch-Savage WE, Leubner-Metzger G. Seed dormancy and the control of germination. *New Phytol.* 2006;171:501–23. <https://doi.org/10.1111/j.1469-8137.2006.01787.x>.
46. Dias PMB, Brunel-Muguet S, Dürr C, Huguet T, Demilly D, Wagner M-H, et al. QTL analysis of seed germination and pre-emergence growth at extreme temperatures in *Medicago truncatula*. *Theor Appl Genet.* 2011;122:429–44. <https://doi.org/10.1007/s00122-010-1458-7>.
47. Gardarin A, Coste F, Wagner M-H, Dürr C. How do seed and seedling traits influence germination and emergence parameters in crop species? A comparative analysis. *Seed Sci Res.* 2016;26:317–31. <https://doi.org/10.1017/S0960258516000210>.
48. Onofri A, Mesgaran MB, Ritz C. A unified framework for the analysis of germination, emergence, and other time-to-event data in weed science. *Weed Sci.* 2022;70:259–71.

49. Larson JE, Anacker BL, Wanous S, Funk JL. Ecological strategies begin at germination: Traits, plasticity and survival in the first 4 days of plant life. *Funct Ecol.* 2020;34:968–79. <https://doi.org/10.1111/1365-2435.13543>.
50. Reed RC, Bradford KJ, Khanday I. Seed germination and vigor: ensuring crop sustainability in a changing climate. *Heredity.* 2022;128:450–9. <https://doi.org/10.1038/s41437-022-00497-2>.
51. Norden N, Daws MI, Antoine C, Gonzalez MA, Garwood NC, Chave J. The relationship between seed mass and mean time to germination for 1037 tree species across five tropical forests. *Funct Ecol.* 2009;23:203–10. <https://doi.org/10.1111/j.1365-2435.2008.01477.x>.
52. Ambika S, Manonmani V, Somasundaram G. Review on Effect of Seed Size on Seedling Vigour and Seed Yield. *Res J Seed Sci.* 2014;7:31–8. <https://doi.org/10.3923/rjss.2014.31.38>.
53. Vidak M, Lazarević B, Javornik T, Šatović Z, Carović-Stanko K. Seed Water Absorption, Germination, Emergence and Seedling Phenotypic Characterization of the Common Bean Landraces Differing in Seed Size and Color. *Seeds.* 2022;1:324–39. <https://doi.org/10.3390/seeds1040027>.
54. Upreti P, Bandara MS, Tanino KK. The Role of Seed Characteristics on Water Uptake Preceding Germination. *Seeds.* 2024;3:559–74. <https://doi.org/10.3390/seeds3040038>.
55. Sharma A, Choudhary P, Chakdar H, Shukla P. Molecular insights and omics-based understanding of plant–microbe interactions under drought stress. *World J Microbiol Biotechnol.* 2023;40:42. <https://doi.org/10.1007/s11274-023-03837-4>.
56. Safdar LB, Fisk ID, Bhosale RA, Griffiths S, Hawkesford MJ, Boden SA, et al. Heterogeneity in seed traits: an overlooked breeding opportunity. *Trends Plant Sci.* 2026;31:379–82. <https://doi.org/10.1016/j.tplants.2025.12.009>.
57. Rosental L, Nonogaki H, Fait A. Activation and regulation of primary metabolism during seed germination. *Seed Sci Res.* 2014;24:1–15. <https://doi.org/10.1017/S0960258513000391>.
58. Verma G, Mishra S, Sangwan N, Sharma S. Reactive oxygen species mediate axis-cotyledon signaling to induce reserve mobilization during germination and seedling establishment in *Vigna radiata*. *J Plant Physiol.* 2015;184:79–88. <https://doi.org/10.1016/j.jplph.2015.07.001>.
59. Xu H-H, Liu S-J, Song S-H, Wang W-Q, Møller IM, Song S-Q. Proteome changes associated with dormancy release of Dongxiang wild rice seeds. *J Plant Physiol.* 2016;206:68–86. <https://doi.org/10.1016/j.jplph.2016.08.016>.
60. Xue X, Yu Y-C, Wu Y, Xue H, Chen L-Q. Locally restricted glucose availability in the embryonic hypocotyl determines seed germination under abscisic acid treatment. *New Phytol.* 2021;231:1832–44. <https://doi.org/10.1111/nph.17513>.
61. Domergue J-B, Abadie C, Limami A, Way D, Tcherkez G. Seed quality and carbon primary metabolism - Domergue – 2019 - Plant, Cell & Environment - Wiley Online Library. 2019. <https://onlinelibrary.wiley.com/doi/10.1111/pce.13618>. Accessed 19 May 2026.
62. Quiles FA, Raso MJ, Pineda M, Piedras P. Ureide metabolism during seedling development in French bean (*Phaseolus vulgaris*). *Physiol Plant.* 2009;135:19–28. <https://doi.org/10.1111/j.1399-3054.2008.01173.x>.

63. Tegeder M. Transporters involved in source to sink partitioning of amino acids and ureides: opportunities for crop improvement. *J Exp Bot.* 2014;65:1865–78. <https://doi.org/10.1093/jxb/eru012>.
64. Bewley JD, Bradford KJ, Hilhorst HWM, Nonogaki H, Seeds. *Physiology of Development, Germination and Dormancy*, 3rd Edition. New York, NY: Springer; 2013. <https://doi.org/10.1007/978-1-4614-4693-4>
65. Rosental L, Perelman A, Nevo N, Toubiana D, Samani T, Batushansky A, et al. Environmental and genetic effects on tomato seed metabolic balance and its association with germination vigor. *BMC Genomics.* 2016;17:1047. <https://doi.org/10.1186/s12864-016-3376-9>.
66. Yang XF, Yao CJ, Li YJ, Zhang SJ. Effect of Exogenous Putrescine on Wheat Seed Germination and Physiological Characteristics Under Salt Stress. *Appl Ecol Environ Res.* 2025;23:9363–74. https://doi.org/10.15666/aeer/2305_93639374.
67. Farooq M, Basra SMA, Rehman H, Hussain M. Seed Priming with Polyamines Improves the Germination and Early Seedling Growth in Fine Rice. *J New Seeds.* 2008;9:145–55. <https://doi.org/10.1080/15228860802087297>.
68. Xu S, Hu J, Li Y, Ma W, Zheng Y, Zhu S. Chilling tolerance in *Nicotiana tabacum* induced by seed priming with putrescine. *Plant Growth Regul.* 2011;63:279–90. <https://doi.org/10.1007/s10725-010-9528-z>.
69. Blázquez MA, Polyamines. Their Role in Plant Development and Stress. *Annu Rev Plant Biol.* 2024;75 Volume 75, 2024:95–117. <https://doi.org/10.1146/annurev-arplant-070623-110056>
70. Kibar H, Kibar B. Enhancing Cowpea Resilience: Role of Putrescine and Salicylic Acid in Mitigating Salt Stress Effects on Germination, Early Growth and Biochemical Content. *Biol Bull.* 2025;52:69. <https://doi.org/10.1134/S1062359024612631>.
71. Lin PPC. Polyamine Anabolism in Germinating *Glycine max* (L.) Seeds 1 | *Plant Physiology* | Oxford Academic. *Plant Physiol.* 1984;76:372–80. <https://doi.org/https://doi.org/10.1104/pp.76.2.372>.
72. Pagnussat L, Burbach C, Baluška F, de la Canal L. Rapid endocytosis is triggered upon imbibition in *Arabidopsis* seeds. *Plant Signal Behav.* 2012;7:416–21. <https://doi.org/10.4161/psb.19669>.
73. Doria E, Pagano A, Ferreri C, Larocca AV, Macovei A, de Araújo S. How Does the Seed Pre-Germinative Metabolism Fight Against Imbibition Damage? Emerging Roles of Fatty Acid Cohort and Antioxidant Defence. *Front Plant Sci.* 2019;10. <https://doi.org/10.3389/fpls.2019.01505>.
74. Nelson EB. The seed microbiome: Origins, interactions, and impacts. *Plant Soil.* 2018;422:7–34. <https://doi.org/10.1007/s11104-017-3289-7>.
75. Simonin M, Préveaux A, Marais C, Garin T, Arnault G, Sarniguet A, et al. Transmission of synthetic seed bacterial communities to radish seedlings: impact on microbiota assembly and plant phenotype. *Peer Community J.* 2023;3. <https://doi.org/10.24072/pcjournal.329>.
76. Romão IR, do Carmo Gomes J, Silva D, Vilchez JI. The seed microbiota from an application perspective: an underexplored frontier in plant–microbe interactions. *Crop Health.* 2025;3:12. <https://doi.org/10.1007/s44297-025-00051-6>.

77. Özkurt E, Hassani MA, Sesiz U, Künzel S, Dagan T, Özkan H, et al. Seed-Derived Microbial Colonization of Wild Emmer and Domesticated Bread Wheat (*Triticum dicoccoides* and *T. aestivum*) Seedlings Shows Pronounced Differences in Overall Diversity and Composition. *mBio*. 2020;11. 10.1128/mbio.02637-20.
78. Rochefort A, Simonin M, Marais C, Guillerme-Erckelboudt A-Y, Barret M, Sarniguet A. Transmission of Seed and Soil Microbiota to Seedling. *mSystems*. 2021;6:e00446–21. <https://doi.org/10.1128/mSystems.00446-21>.
79. Simonin M, Briand M, Chesneau G, Rochefort A, Marais C, Sarniguet A, et al. Seed microbiota revealed by a large-scale meta-analysis including 50 plant species. *New Phytol*. 2022;234:1448–63. <https://doi.org/10.1111/nph.18037>.
80. Glick BR. Bacteria with ACC deaminase can promote plant growth and help to feed the world. *Microbiol Res*. 2014;169:30–9. <https://doi.org/10.1016/j.micres.2013.09.009>.
81. Pieterse CMJ, Zamioudis C, Berendsen RL, Weller DM, Wees SCMV, Bakker PAHM. Induced Systemic Resistance by Beneficial Microbes. *Annu Rev Phytopathol*. 2014;52:2014:347–75. <https://doi.org/10.1146/annurev-phyto-082712-102340>.
82. Compant S, Samad A, Faist H, Sessitsch A. A review on the plant microbiome: Ecology, functions, and emerging trends in microbial application. *J Adv Res*. 2019;19:29–37. <https://doi.org/10.1016/j.jare.2019.03.004>.
83. Marchi M, Bosc-Bierne A, Labourgade L, Lerenard T, Corff JL, Aligon S et al. Unexplored Yeast diversity in Seed Microbiota. 2026;:2024.11.27.625647. <https://doi.org/10.1101/2024.11.27.625647>
84. Sharma S, Nagrale DT, Chaurasia A, Ranjan R, Jha GK, Kumar S. The emerging roles of yeasts as plant growth promoters: mechanisms, benefits, and agricultural applications. *World J Microbiol Biotechnol*. 2026;42:224. <https://doi.org/10.1007/s11274-026-04810-7>.
85. Hernández Flores JL, Martínez YJ, Ramos López MÁ, Saldaña Gutierrez C, Reyes AA, Armendariz Rosales MM, et al. Volatile Organic Compounds Produced by *Kosakonia cowanii* Cp1 Isolated from the Seeds of *Capsicum pubescens* R & P Possess Antifungal Activity. *Microorganisms*. 2023;11:2491. <https://doi.org/10.3390/microorganisms11102491>.
86. Guimarães GR, Pereira FS, Matos FS, Mello SCM, Carvalho DDC. Suppression of seed borne *Cladosporium herbarum* on common bean seed by *Trichoderma harzianum* and promotion of seedling development. *Trop Plant Pathol*. 2014;39:401–6. <https://doi.org/https://doi.org/10.1590/S1982-56762014000500007>.
87. El-Dawy EGAEM, Gherbawy YA, Hussein MA. Morphological, molecular characterization, plant pathogenicity and biocontrol of *Cladosporium* complex groups associated with faba beans. *Sci Rep*. 2021;11:14183. <https://doi.org/10.1038/s41598-021-93123-w>.
88. Arnault G, Marais C, Prévieux A, Briand M, Poisson A-S, Sarniguet A, et al. Seedling microbiota engineering using bacterial synthetic community inoculation on seeds. *FEMS Microbiol Ecol*. 2024;100:fiae027. <https://doi.org/10.1093/femsec/fiae027>.

89. Faust K, Raes J. Microbial interactions: from networks to models. *Nat Rev Microbiol.* 2012;10:538–50. <https://doi.org/10.1038/nrmicro2832>.
90. Poudel R, Jumpponen A, Schlatter DC, Paulitz TC, Gardener BBM, Kinkel LL, et al. Microbiome Networks: A Systems Framework for Identifying Candidate Microbial Assemblages for Disease Management. *Phytopathology®.* 2016;106:1083–96. <https://doi.org/10.1094/PHYTO-02-16-0058-FI>.
91. Faust K. Open challenges for microbial network construction and analysis. *ISME J.* 2021;15:3111–8. <https://doi.org/10.1038/s41396-021-01027-4>.
92. Lee KK, Kim H, Lee Y-H. Cross-kingdom co-occurrence networks in the plant microbiome: Importance and ecological interpretations. *Front Microbiol.* 2022;13. <https://doi.org/10.3389/fmicb.2022.953300>.
93. Banerjee S, Schlaeppi K, van der Heijden MGA. Keystone taxa as drivers of microbiome structure and functioning. *Nat Rev Microbiol.* 2018;16:567–76. <https://doi.org/10.1038/s41579-018-0024-1>.
94. Babalola OO, Fadiji AE, Enagbonma BJ, Alori ET, Ayilara MS, Ayangbenro AS. The Nexus Between Plant and Plant Microbiome: Revelation of the Networking Strategies. *Front Microbiol.* 2020;11.
95. Dundore-Arias JP, Michalska-Smith M, Millican M, Kinkel LL. More Than the Sum of Its Parts: Unlocking the Power of Network Structure for Understanding Organization and Function in Microbiomes. *Annu Rev Phytopathol.* 2023;61 Volume 61, 2023:403–23. <https://doi.org/10.1146/annurev-phyto-021021-041457>
96. Vorholt JA, Vogel C, Carlström CI, Müller DB. Establishing Causality: Opportunities of Synthetic Communities for Plant Microbiome Research. *Cell Host Microbe.* 2017;22:142–55. <https://doi.org/10.1016/j.chom.2017.07.004>.
97. Mehlferber EC, Arnault G, Joshi B, Partida-Martinez LP, Patras KA, Simonin M, et al. A cross-systems primer for synthetic microbial communities. *Nat Microbiol.* 2024;9:2765–73. <https://doi.org/10.1038/s41564-024-01827-2>.
98. Rajjou L, Duval M, Gallardo K, Catusse J, Bally J, Job C, et al. Seed Germination and Vigor. *Annu Rev Plant Biol.* 2012;63:507–33. <https://doi.org/10.1146/annurev-arplant-042811-105550>.
99. Finch-Savage WE, Bassel GW. Seed vigour and crop establishment: Extending performance beyond adaptation. *J Exp Bot.* 2016;67:567–91. <https://doi.org/10.1093/jxb/erv490>.
100. Lemanceau P, Barret M, Mazurier S, Mondy S, Pivato B, Fort T, et al. Plant Communication With Associated Microbiota in the Spherosphere, Rhizosphere and Phyllosphere. *Advances in Botanical Research.* Academic; 2017. pp. 101–33. <https://doi.org/10.1016/bs.abr.2016.10.007>.
101. Saccaram C, Simonin M, Boutet S, Brosse C, Peng S, François T, et al. Elucidating the interplay between metabolites and microorganisms in the spherosphere of common bean (*Phaseolus vulgaris* L.) seeds. *mSystems.* 2025;10:e00707–25. <https://doi.org/10.1128/msystems.00707-25>.

Figures

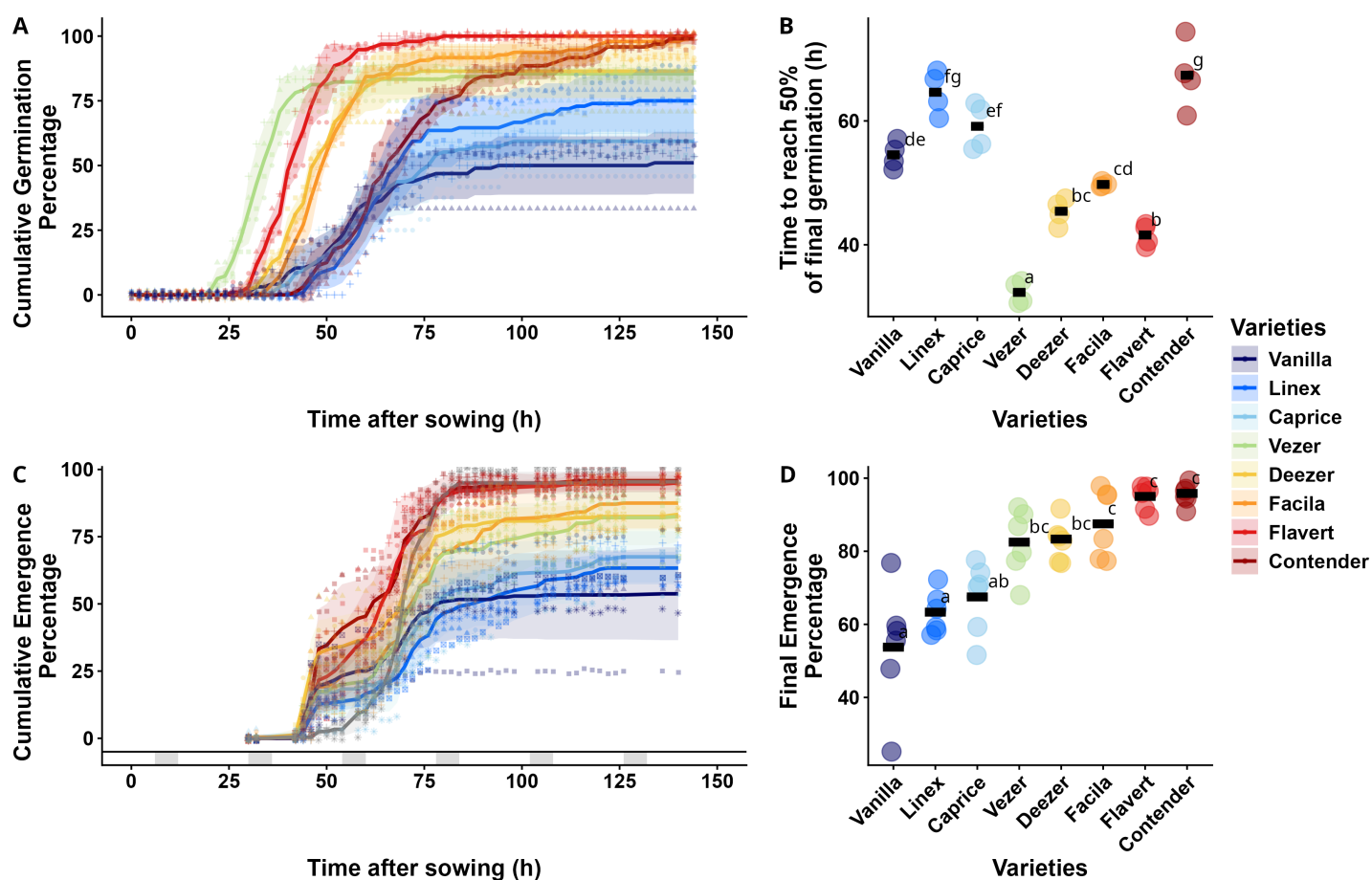


Figure 1

Seed germination and seedling emergence traits of 8 common bean varieties

A: Cumulated germination percentage over 5 days(pleated paper on Jacobsen germinators, 20°C, 144 h, 4*24 seeds). Replicates are shown as different shapes, varieties as colors. Bold lines: variety mean; shaded zones: standard deviation interval.

B: Time (h) to 50% germination (t50germ) estimated by a Four Parameters Hill Function model (data from A, model in **Figure S1B**). Black rectangles: variety mean t50germ.

C: Cumulated emergence percentage over 7 days(potting mix, 140h, 6*40 seeds). Replicates are shown as different shapes, varieties as colors. Bold lines: variety mean; shaded zones: standard deviation interval. Grey vertical zones above the x-axis indicate night periods.

D: Final emergence percentage at 140 h after sowing (6*40 seeds). Black rectangles: variety mean FEP.

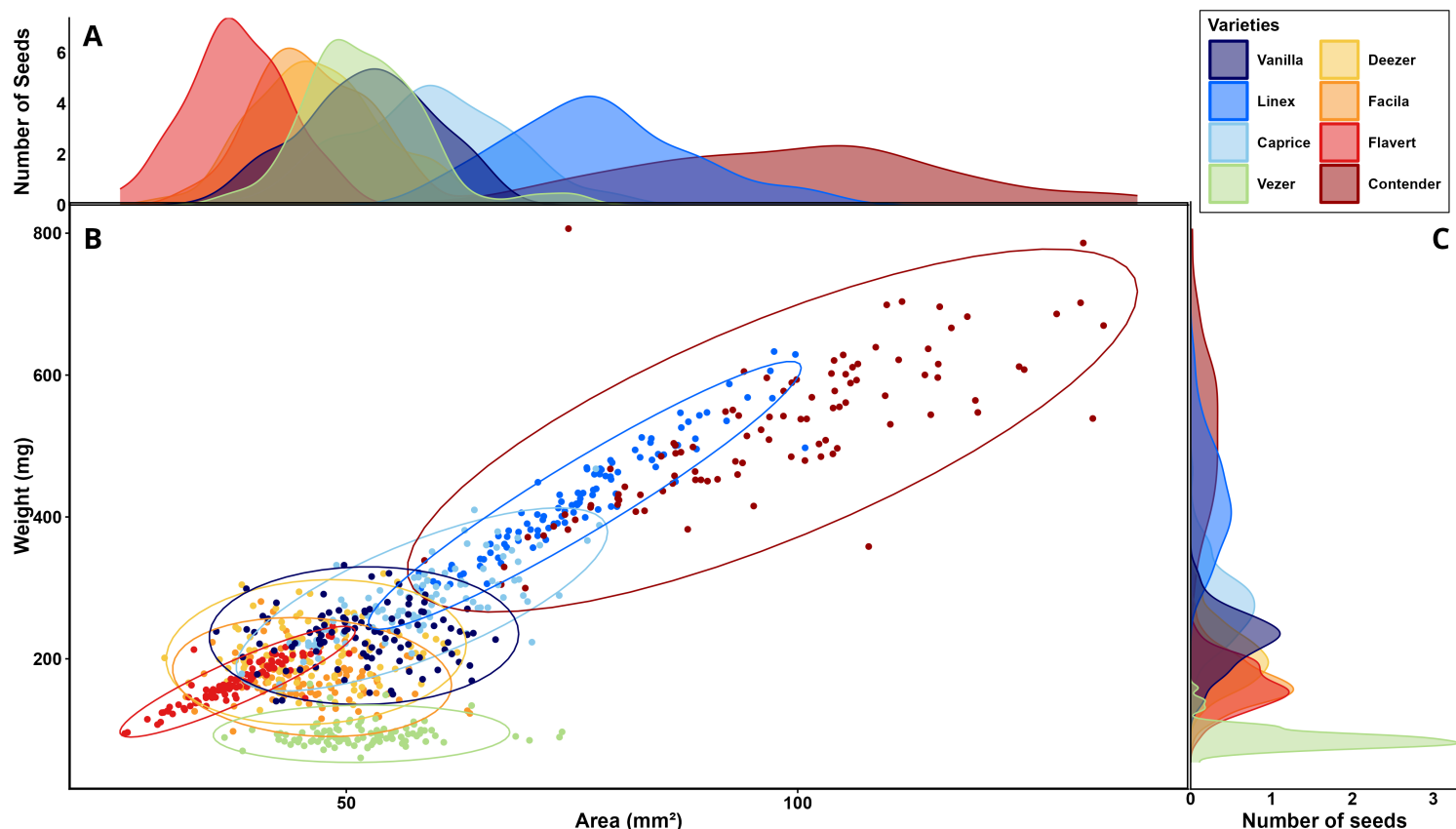


Figure 2

Weight-area distribution of individual seeds (100 seeds for each variety)

A: Distribution of seed area per variety.

B: Weight-area dot plot. Each dot represents one seed. One color is attributed to each variety according to FEP gradient and is consistent across panels. Ellipses represent 95% confidence intervals based on a multivariate t-distribution.

C: Distribution of seed weight per variety.

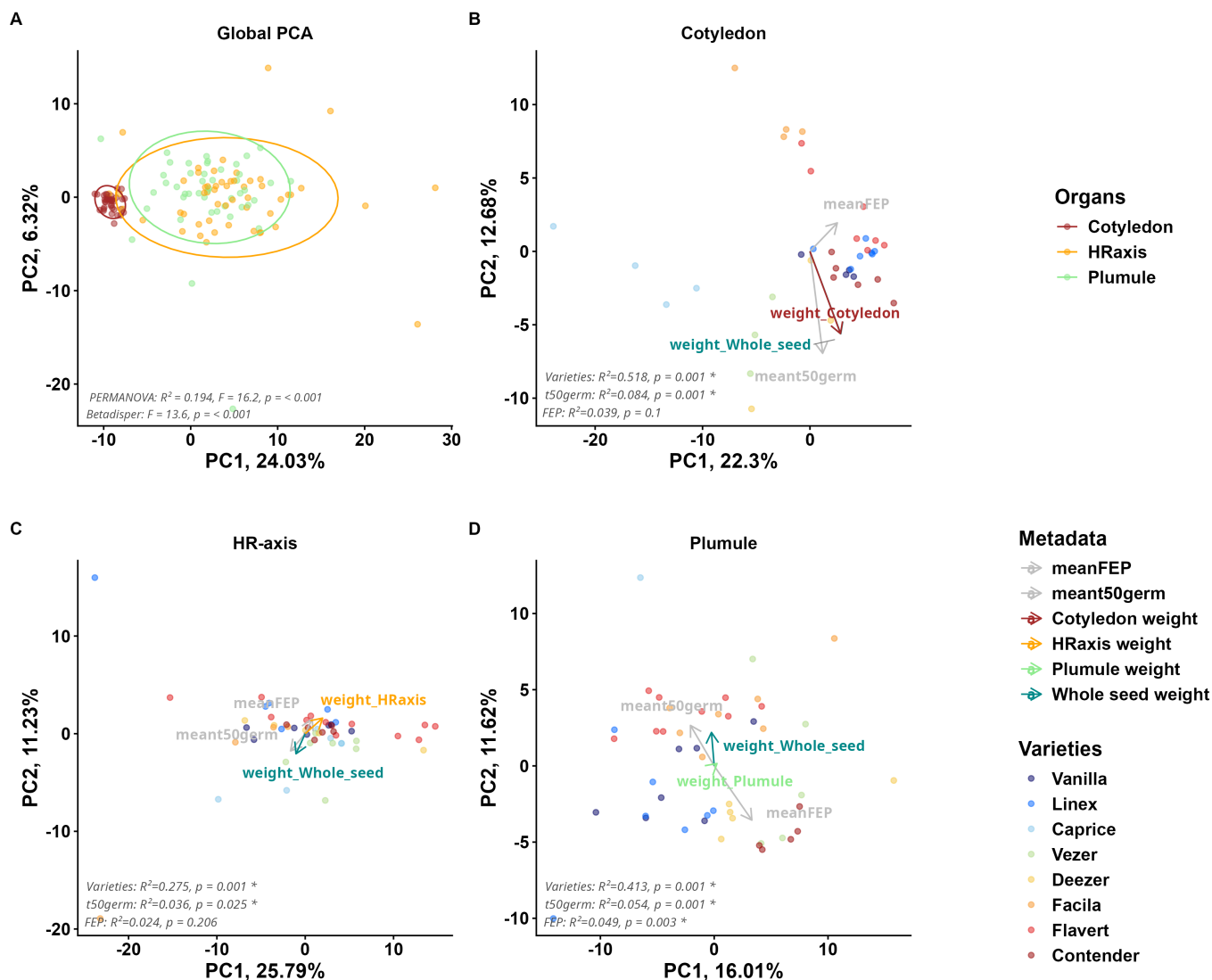


Figure 3

Seed organ metabolome composition

A: Global principal component analysis (PCA) of centered-scaled GC-MS metabolite profiles across three seed organs ($n = 137$ seeds, 221 metabolites after prevalence filtering). Each point represents one individual seed, colored by organ (brown: cotyledon; orange: HR-axis; green: plumule). Ellipses: 95% confidence intervals. PERMANOVA and betadisper statistics are shown in the lower left corner.

B-D: Per-organ PCA on per-organ Z-scaled metabolome profiles. **B:** cotyledon ($n = 38$), **C:** HR-axis ($n = 53$), **D:** plumule ($n = 46$). Each point represents one individual seed, colored by variety. Arrows indicate the direction and strength of association of continuous variables (seed weight, organ weight, mean t50germ, mean FEP) with the principal component axes. PERMANOVA statistics are shown in the lower left corner.

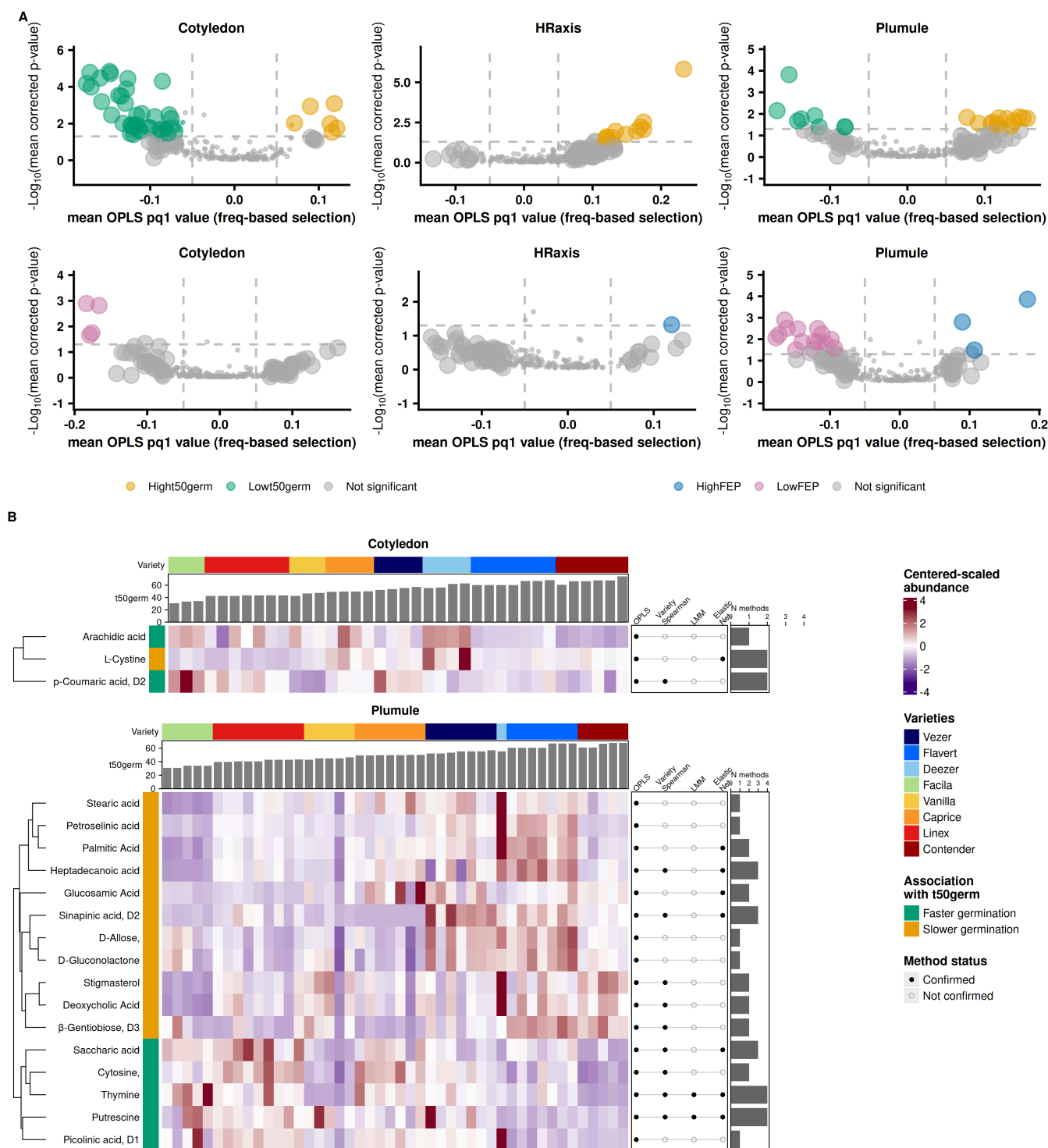


Figure 4

Candidate metabolite biomarkers of germination speed (mean t50germ) or emergence percentage (mean FEP)

A: Mean volcano plots of metabolite associations with germination speed across the three organs. Point size: VIP > 1. Colored points indicate metabolites passing both pq1 ($|pq1| > 0.05$), and significance (BH-

corrected $p < 0.05$) threshold on average across permutations. Dashed lines: significance thresholds. Most depicted metabolites did not reach the following stability criterion.

B: Heatmap of the 19 candidate metabolic markers associated with germination speed (t50germ) identified by the OPLS permutation pipeline and filtered by stability criterion, in the cotyledon (top, 3 metabolites) and plumule (bottom, 16 metabolites). Rows: metabolites (hierarchical clustering, Euclidean distance); columns: individual seeds ordered by variety (color bar, top) then by individual t50germ within variety (grey barplot, top; higher bars = slower germination). Metabolite abundance is shown as per-organ centered-scaled values (purple: below organ mean; white: at organ mean; red: above organ mean). Left strip: sign of association with t50germ (green: faster germination; orange: slower germination). Right panel: confirmation across four methods (OPLS permutation frequency, variety-level Spearman correlation, linear mixed model, and elastic net regression); filled circles: confirmation. Barplot: number of confirming methods (0-4).

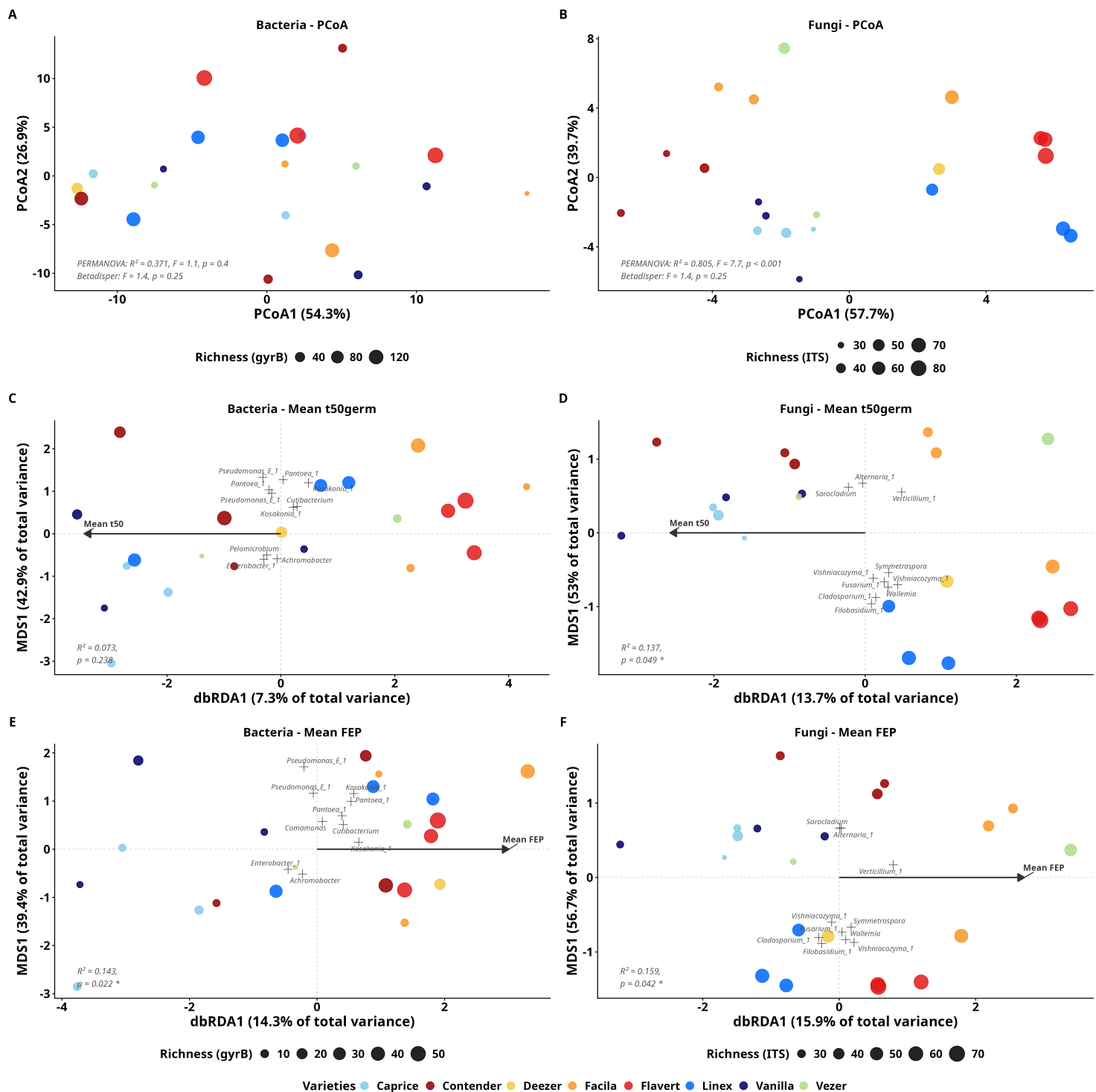


Figure 5

Native seed microbiota community composition and association with germination and emergence traits

A-B: **Principal Coordinates Analysis of bacterial(A) or fungal (B) community profiles** (Aitchison distance, rCLR transformation). Points: individual samples, colored by variety; size: ASV richness.

C-F: **Distance-based redundancy analysis (db-RDA)** of bacterial (C, E) and fungal (D, F) communities constrained by mean germination time (C-D) or mean final emergence percentage (E-F). Arrows: constraint direction and strength. Axis labels: percentage of total inertia explained. Grey axis labels: non-significant axes. Species scores were projected post-hoc onto the ordination space.

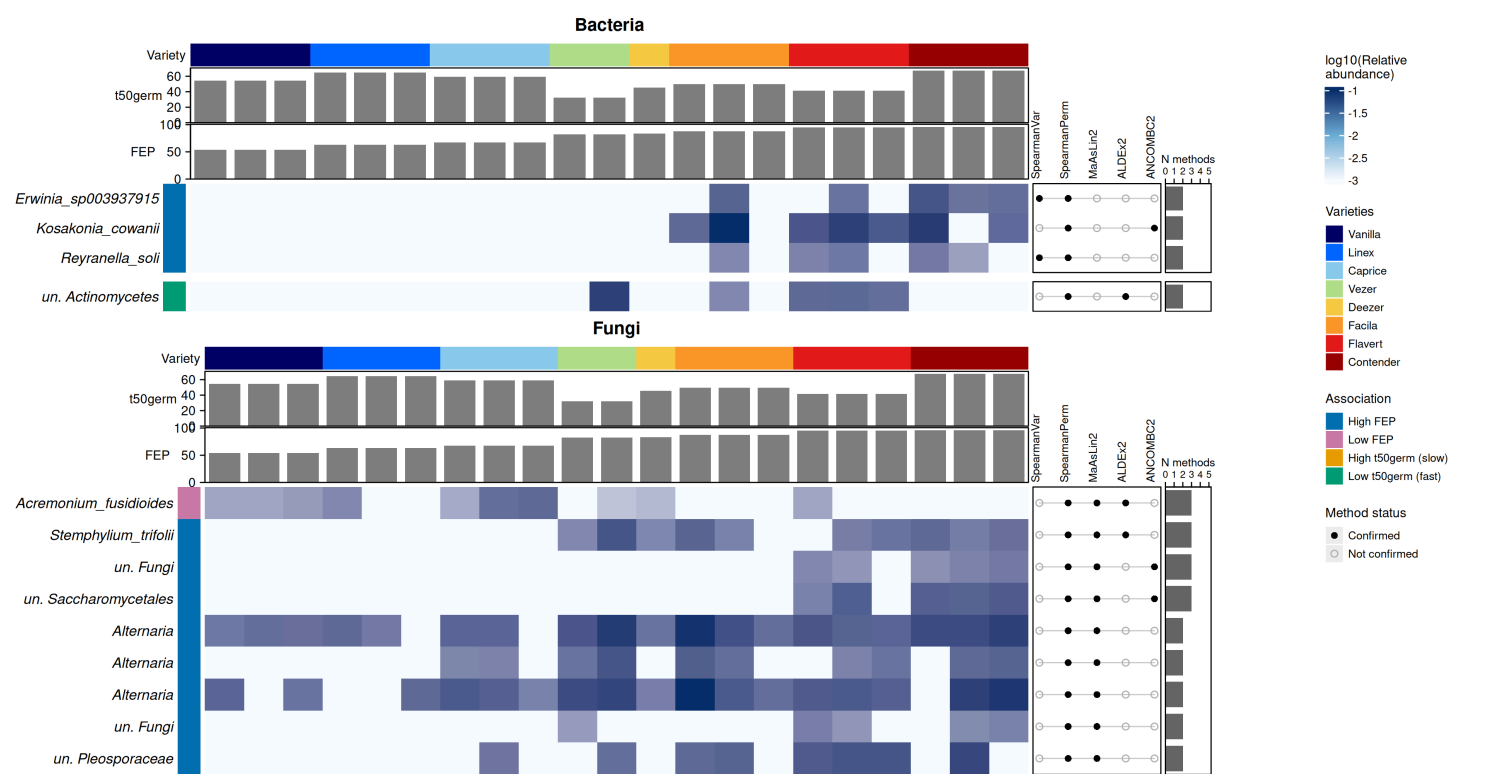


Figure 6

Heatmap of the 13 taxa associated with germination speed (t50germ) or emergence percentage (FEP) identified by at least 2 methods. Rows: taxa; columns: individual seed batch microbiota ordered by variety (color bar, top) then by replicate FEP within variety (lower barplot). Microbial taxa abundance TSS-transformed is shown on a log₁₀ scale. Left color strip: sign of association with t50germ or FEP (green: faster germination; blue: lower FEP, pink: higher FEP). Right panel: confirmation across five methods (Spearman correlation (rho) on variety-mean values and on per-replicate permutations, MaAsLin2 (log fold-change), ALDEx2 (rho), and ANCOM-BC2 (log fold-change)); filled circles: confirmed. Barplot: number of confirming methods. All associations are exploratory (q ≤ 0.25).

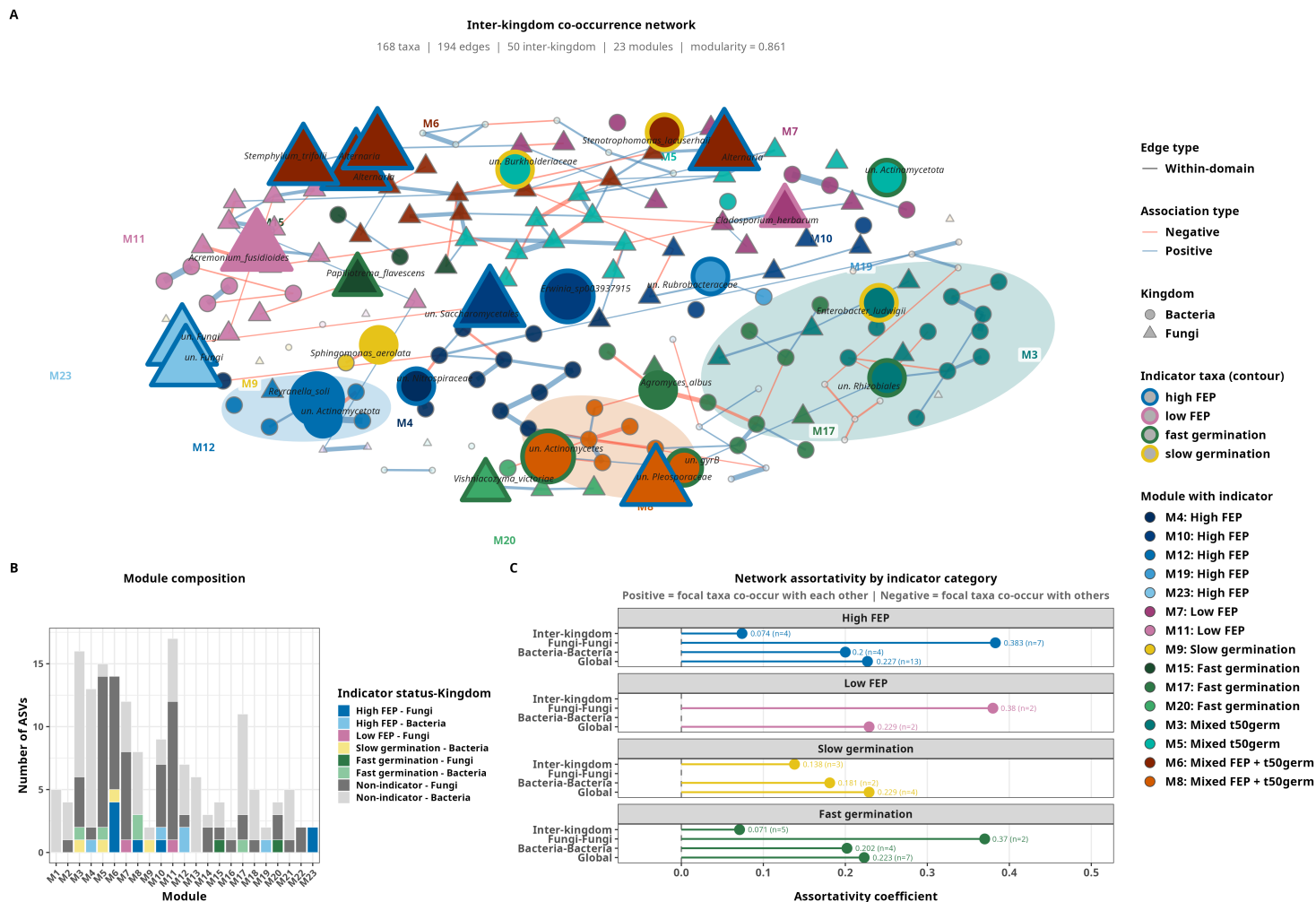


Figure 7

Seed microbiota co-occurrence network and module composition

A: Inter-kingdom co-occurrence network inferred by cross-domain SPIEC-EASI (ASVs with prevalence > 15%). Nodes: ASVs; size: indicator status (large with contour: candidate indicator taxa identified by several (biggest) or one (medium) method, small: taxa in modules containing indicator taxa, tiny grey: others). Candidate indicator taxa contours are colored by association (high FEP: blue; low FEP: pink; high t50germ - slow germination: orange; low t50germ - fast germination: green). Nodes fill color: Louvain module (bright colors: modules containing indicator taxa: high FEP: blue shades; low FEP: pink; high t50germ: yellow; low t50germ: green; FEP and t50germ mixed: orange/red shades; high and low t50germ: cyan; high and low FEP: purple). Edge color and style: sign of the partial correlation between taxa (blue: positive, red: negative, solid: within-domain, dotted: inter-kingdom). Isolated nodes are not shown. *un.* : *unassigned*.

B: Network module composition by kingdom, evidencing candidate indicator taxa associated with each category (high or low t50germ, or high or low FEP).

C: Network assortativity with assortativity coefficient for each trait (All traits, t50germ, FEP) computed across different edge subset (Global: all edges; Bacteria-Bacteria and Fungi-Fungi: within-kingdom edges; Inter-kingdom: edges between bacterial and fungal taxa) for each association sign. Positive

values: taxa from the same association category tend to co-occur more than expected by chance. Absence of value corresponds to non-occurring associations. Node and edge counts per subset are indicated in grey.

Supplementary Files

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

- [ColaertSentenacetalSupplementaryFigures.pdf](#)
- [ColaertSentenacetalSupplementaryNote.docx](#)
- [ColaertSentenacetalSupplementaryTables.xlsx](#)