Package: SACAT (via r-universe)

November 15, 2026

Type Package

Title Structural-Absence and Conditional-Abundance Testing

Version 0.99.2

URL https://github.com/Statwright/SACAT

BugReports https://github.com/Statwright/SACAT/issues

Description Provides depth-aware taxon-level inference for microbiome count data. SACAT separates group differences in structural-absence probability from differences in mean log relative abundance among samples in which a taxon is present. It accounts for finite-depth nondetection, centers abundance effects against a target-excluded cross-taxon reference, reports component-specific and omnibus tests, and visualizes statistical evidence, observed detection, estimated zero sources, and abundance distributions.

License GPL (>= 3)

Encoding UTF-8

Depends R (>= 4.5.0)

Imports graphics, grDevices, grid, parallel, Rcpp, SummarizedExperiment, statmod, stats, utils

LinkingTo Rcpp

Suggests BiocStyle, knitr, rmarkdown, testthat (>= 3.2.0)

Config/testthat/edition 3

Roxygen list(markdown = TRUE)

NeedsCompilation yes

Config/roxygen2/version 8.1.0

VignetteBuilder knitr

biocViews Microbiome, Metagenomics, DifferentialExpression, Regression, StatisticalMethod, Visualization

Config/pak/sysreqs zlib1g-dev

Repository https://biocstaging.r-universe.dev

Date/Publication 2026-11-15 01:57:01 UTC

RemoteUrl https://github.com/BiocStaging/SACAT

RemoteRef HEAD

RemoteSha 976eedb31c5fdb2076153d9bbcdb765cf59baae9

Contents

gems_microbiome 2
plot.sacat 3
sacat 5
sacat_se 9

Index **11**

`gems_microbiome` *Gut microbiome profiles from children in the GEMS study*

Description

Stool 16S rRNA gene counts from diarrhea-free control children in the Global Enteric Multicenter Study (GEMS), reported by Pop et al. (2014).

Usage

```
data(gems_microbiome)
```

Format

A SummarizedExperiment::SummarizedExperiment with 41 bacterial genera in rows and 328 children in columns. The `counts` assay contains raw genus-level read counts. Sample metadata in `colData()` contain:

`group` Age group: `Infant` (1–11 months, 191 children) or `Older` (24–58 months, 137 children).

`age_months` Age in months.

`country` Study country: Bangladesh, The Gambia, Kenya or Mali.

`reads` Total reads for the sample in the original count table, before genus aggregation and filtering.

Details

The example includes controls younger than 12 months or at least 24 months old. Counts were summed across features with the same bacterial genus name. Genera detected in at least 20% of the selected children were retained. The preparation script is installed under `scripts/prepare_gems_microbiome.R`.

Source

Counts and sample annotations were adapted from the Bioconductor data package `msd16s` (version 1.32.0) by Joseph N. Paulson, Hector Corrada Bravo, and Mihai Pop (doi:10.18129/B9.bioc.msd16s). The source package is distributed under the Artistic License 2.0.

References

Pop M, Walker AW, Paulson JN, et al. (2014). Diarrhea in young children from low-income countries leads to large-scale alterations in intestinal microbiota composition. Genome Biology, 15, R76. doi:10.1186/gb2014156r76.

Examples

```
data("gems_microbiome")
dim(gems_microbiome)
table(gems_microbiome$group, gems_microbiome$country)
```

`plot.sacat` *Plot a SACAT association profile*

Description

Aligns component evidence, observed detection with estimated zero sources, and observed and fitted abundance for selected taxa.

Usage

```
## S3 method for class 'sacat'
plot(
  x,
  features = NULL,
  selection = c("significant", "top", "all"),
  max_features = 24L,
  alpha = 0.05,
  group_colors = c("#4C78A8", "#E45756"),
  zero_colors = c("#6A9DC0", "#E5A36C", "#E9E9E9"),
  file = NULL,
  width = NULL,
  height = NULL,
  dpi = 300,
  group_labels = NULL,
  show_component_guides = TRUE,
  ...
)
```

Arguments

x	A sacat object fitted with component = "all" and store_plot_data = TRUE.
features	Optional unique taxon names, displayed in the supplied order.
selection	Select omnibus-significant taxa, the top-ranked taxa, or all eligible taxa. Ignored when features is supplied.
max_features	Maximum number of automatically selected taxa. Ignored for selection = "all" or explicit features.
alpha	Adjusted-p threshold for selection and component markers.
group_colors	Two colors, in reference/comparison order, for observed abundance distributions.
zero_colors	Three colors for detected samples, estimated SA, and estimated ND, respectively.
file	Optional PDF, PNG, or SVG filename. Without it, draw on the current graphics device.
width, height	Output dimensions in inches when file is supplied. Defaults to 10.5 inches wide and a height based on the number of taxa.
dpi	Resolution of PNG output.
group_labels	Optional reference/comparison labels for display.
show_component_guides	Show BH component discovery boundaries when p_adjust_method is "BH" or "fdr". A component with no discoveries has no boundary.
...	Reserved for graphical extensions.

Details

Panel A shows the structural-absence (SA) score statistic and the signed square root of the adjusted relative-abundance (RA) likelihood-ratio statistic. Positive values indicate higher SA or reference-corrected RA in the comparison group. Filled circles mark component adjusted p-values at or below alpha; open circles mark values above it. Dashed lines show component-specific BH discovery boundaries computed from the complete retained families. Taxon superscripts summarize omnibus adjusted p-values.

Panel B shows each group's observed detection fraction and estimated contributions of structural absence and nondetection (ND) to observed zeros. For each zero, the SA contribution is its posterior SA probability from the group-inclusive structural plotting fit; the ND contribution is its complement. Positive counts contribute to detected samples. These quantities are averaged within the observed groups and sum to one. An undivided, unfilled zero segment marked -- denotes an unavailable zero-source fit.

Panel C shows positive observed relative abundances as points and boxplots, with library sizes from the fitted analysis. Boxplots summarize log10 relative abundance: the box spans the quartiles, the line marks the median, and whiskers extend to observations within 1.5 interquartile ranges. Diamonds show the exponentiated mean fitted log relative abundance conditional on presence, standardized over all observed covariate profiles. Both are displayed as percentages on a log scale. The raw abundance contrast is the natural log of the ratio of the comparison-group diamond to the

reference-group diamond. The RA test in Panel A subtracts the estimated cross-taxon reference from this contrast.

Plotting uses data prepared by `sacat()` with `store_plot_data = TRUE`. It does not refit the model or change statistical results.

Value

Invisibly, a list containing the displayed data and graphical settings.

Examples

```
example("sacat", echo = FALSE)
plot(fit, features = c("Haemophilus", "Lactobacillus", "Faecalibacterium"),
     group_labels = c("1-11 mo", "24-58 mo"))
```

`sacat` *Structural-Absence and Conditional-Abundance Testing*

Description

Fits SACAT to microbiome count data and tests a binary group contrast in structural-absence probability, relative abundance conditional on presence, or both. Sequencing depth enters the latent-state count model so that an observed zero can receive different posterior support for structural absence and nondetection across samples.

Usage

```
sacat(
  counts,
  metadata,
  formula,
  group,
  library_size,
  taxa_are_rows = TRUE,
  reference = NULL,
  p_adjust_method = "BH",
  component = c("all", "structural_absence", "relative_abundance"),
  full_output = FALSE,
  min_positive_samples = 3L,
  min_reference_taxa = 4L,
  structural_quadrature_points = 1001L,
  workers = 1L,
  verbose = FALSE,
  abundance_quadrature_points = 41L,
  store_plot_data = FALSE
)
```

Arguments

counts	Raw non-negative integer counts in a matrix or data frame.
metadata	Sample metadata in a data frame. Row names must contain all sample names in counts. Variables used by formula, group, or a metadata-based library_size must be complete. Prepare a complete analysis set before calling sacat() while keeping counts, metadata, and any named depth vector aligned.
formula	A one-sided formula with an intercept, an additive main effect for group, and any adjustment terms, for example ~ disease + age + sex. Interactions involving group, offsets, and random-effect terms are not supported.
group	Name of the binary metadata variable to test.
library_size	Either the name of a numeric column in metadata or a numeric vector named by sample containing the original sequencing depth for each sample. Depths must be positive integers at least as large as the sample totals in counts.
taxa_are_rows	Logical. If TRUE, taxa are rows and samples are columns in counts. If FALSE, samples are rows.
reference	Optional reference value for group. A factor uses its first level by default; logical and numeric 0/1 groups use FALSE or 0. Character groups and other numeric codings require an explicit reference.
p_adjust_method	Method passed to stats::p.adjust() separately for each requested component and omnibus family.
component	Analysis to run. "all" fits both components and reports the omnibus analyses. "structural_absence" fits the structural-absence component. "relative_abundance" fits the relative-abundance component.
full_output	Logical. If TRUE, the returned object includes detailed component fits. Non-default quadrature rules and fitted latent-scale standard deviations above two trigger fixed-fit higher-order quadrature comparisons. The detailed abundance table records the comparison order and the conditional-log-likelihood and effect discrepancies.
min_positive_samples	Minimum number of samples with a positive count required to retain a taxon before fitting. The default is 3L. This preliminary filter does not replace the positive-mark design-rank and information checks. Because it changes the retained multiplicity and reference families, it should be chosen before inspecting results.
min_reference_taxa	Minimum number of eligible target-excluded taxa required to form an abundance reference. The default is 4L; values below 3L are not supported.
structural_quadrature_points	Number of Gauss-Hermite nodes used by the structural arm to refine the conditional-present fit and evaluate the structural-absence model. The default is 1001L. Smaller values reduce computation at the cost of numerical accuracy; full_output = TRUE provides a higher-order comparison.

`workers` Number of parallel worker processes. The default `1L` runs sequentially. Values above one use an ordered, cross-platform PSOCK cluster and do not change the taxon order.

`verbose` Logical. If `TRUE`, report arm-level start and completion messages.

`abundance_quadrature_points`
Number of Gauss-Hermite nodes used by the shared five-start conditional-present search and by subsequent relative-abundance refinement and inference. When the structural component is requested, this search supplies the starting point for refinement at `structural_quadrature_points`. The default is `41L`. Nondefault values use the same log-domain rule as the structural arm. With `full_output = TRUE`, nondefault rules and fitted latent-scale standard deviations above two trigger a fixed-fit higher-order comparison.

`store_plot_data`
Logical. If `TRUE`, retain observed positive abundances, groupwise detection and estimated zero-source shares, and standardized fitted abundance summaries used by `plot.sacat()`. This option requires `component = "all"`. Structural companion summaries reuse the configured worker pool. Higher-order quadrature comparisons remain controlled by `full_output`. This setting only controls the stored plotting data; the default `FALSE` keeps the fitted object compact.

Details

The structural-absence component estimates the conditional-present count distribution from positive counts, estimates structural-absence nuisance parameters from the detection indicators, and evaluates a nuisance-orthogonalized estimating-function statistic for the group effect with a matching empirical sandwich variance. Positive `z_structural_absence` values indicate greater structural absence in the comparison group.

The relative-abundance component estimates a covariate-standardized comparison-minus-reference difference in mean log relative abundance conditional on taxon presence. Every taxon is fitted with the same zero-truncated conditional-mark likelihood. Conditioning on a positive count removes the structural-presence probability from this likelihood. Zero counts have zero mark score, and all samples contribute their covariate values to sample standardization.

Taxon-specific effects are centered against a target-excluded cross-taxon reference. The standard error uses centered, sample-aligned influence contributions. A sandwich-adjusted profile likelihood-ratio test compares the fitted likelihood with a constrained fit whose raw abundance contrast equals the estimated reference. The likelihood-ratio statistic is multiplied by the raw contrast's inverse-information variance divided by the reference-corrected sandwich variance and compared with a chi-squared distribution with one degree of freedom. `z_relative_abundance` is its signed square root, with the sign of the reference-corrected effect. A positive `estimate_relative_abundance` indicates that the target taxon's present-conditional contrast exceeds the shared compositional background. With `full_output = TRUE`, the detailed abundance table records the raw taxon estimate and the reference quantities used to construct the reported contrast.

Taxa with positive counts in fewer than `min_positive_samples` samples are excluded before component fitting and omitted from the multiple-testing families. Every retained taxon remains in each requested family. If a component cannot be formed, an operational value of one retains the taxon in that family; the formation indicator and reason identify values assigned to unavailable component tests. Abundance-component availability is the proportion of retained taxa with

formed_relative_abundance = TRUE. The diagnostics record positive-count support and the reason for each unavailable test.

Two structural outcomes are nonregular. A taxon with no observed zeros has no variation in its detection indicator (no_observed_zeros). An intercept-only structural nuisance fit at zero absence probability is recorded as structural_absence_boundary_at_zero. Both nonregular statuses use an operational p-value of one in the primary family and are excluded from the Cauchy sensitivity combination; their signed statistic is undefined.

structural_absence_nonoptimal_nuisance_fit identifies a finite nuisance fit whose detection objective exceeds the exact zero-limit objective. An otherwise supported all-positive taxon can still enter the abundance conditional-mark fit because that fit has no structural nuisance parameter. The component-use columns record the components entering each omnibus calculation.

Requested components return lightweight warning codes even when full_output = FALSE. Structural analyses include warning_structural_absence and nonregular_structural_absence; abundance analyses include warning_relative_abundance. Setting full_output = TRUE additionally retains detailed fitted objects and reference diagnostics.

Value

An object of class sacat with elements:

results A taxon-level table for the requested analyses. Adjusted p-values use the p_adj_ prefix and the method recorded in settings$p_adjust_method.

diagnostics Taxon retention, support counts, component-formation status and reason codes, and numerical warnings.

settings The fitted contrast and analysis settings.

call The matched function call.

plot_data Observed abundances and fitted plotting summaries when store_plot_data = TRUE.

fits Detailed component fits when full_output = TRUE.

Examples

```
data("gems_microbiome")
fit <- sacat(
    counts = SummarizedExperiment::assay(gems_microbiome, "counts"),
    metadata = as.data.frame(SummarizedExperiment::colData(gems_microbiome)),
    formula = ~ country + group,
    group = "group", reference = "Infant", library_size = "reads",
    store_plot_data = TRUE
)
genera <- c("Haemophilus", "Lactobacillus", "Faecalibacterium")
fit$results[match(genera, fit$results$taxon),
            c("taxon", "p_adj_structural_absence", "p_adj_relative_abundance",
              "p_adj_omnibus")]
```

sacat_se *Run SACAT on a SummarizedExperiment*

Description

Extracts a count assay and sample metadata from a SummarizedExperiment::SummarizedExperiment and passes them to `sacat()`. Rows represent taxa and columns represent samples. The statistical model, input validation, and analysis options are those of `sacat()`.

Usage

```
sacat_se(object, formula, group, library_size, assay = "counts", ...)
```

Arguments

`object`	A SummarizedExperiment::SummarizedExperiment or a subclass, with raw non-negative integer counts in the selected assay. Taxon row names and sample column names must be unique and non-empty.
`formula`	A one-sided formula using columns of `colData(object)`. Include an intercept and an additive main effect for `group`.
`group`	Name of the binary variable in `colData(object)` to test.
`library_size`	Name of a numeric column in `colData(object)`, or a numeric vector named by sample, giving the original sequencing depths. Each depth must be a positive integer at least as large as the corresponding count total in the assay.
`assay`	Name or positive integer index of the count assay. Defaults to `"counts"`.
`...`	Additional arguments passed unchanged to `sacat()`, such as `reference`, `component`, `workers`, and `store_plot_data`. Do not supply `counts`, `metadata`, or `taxa_are_rows`: these are determined by `object`.

Details

Original sequencing depths must be supplied explicitly. They are never calculated from assay column sums, which may omit taxa removed before the analysis. The selected assay is converted to a dense matrix by `sacat()`; allow sufficient memory when using a sparse or delayed assay.

Value

A `sacat` object, as returned by `sacat()`.

See Also

`sacat()`, SummarizedExperiment::SummarizedExperiment

Examples

```
data("gems_microbiome")
genera <- c("Haemophilus", "Lactobacillus", "Faecalibacterium")
fit <- sacat_se(
    gems_microbiome[genera, ], ~ country + group,
    group = "group", reference = "Infant", library_size = "reads",
    component = "structural_absence"
)
fit$results[, c("taxon", "p_structural_absence")]
```

Index

∗ **datasets**
gems_microbiome, 2

gems_microbiome, 2

plot.sacat, 3

sacat, 5
sacat(), *5*, *9*
sacat_se, 9
stats::p.adjust(), *6*
SummarizedExperiment::SummarizedExperiment, *2*, *9*