Tutorial: Taxon-level microbiome mediation analysis using CAMRA

Yunfei Peng

2026-02-27

This is a practical tutorial on the use of the miMediation package for conducting taxon-level microbiome mediation analysis via CAMRA (Causal Absolute-abundance Mediation from Relative-Abundance data). The methodology is described in detail in the Wang, Li, Peng, and Tang (2026).

Brief Introduction of CAMRA

Identifying microbial taxa that mediate the effect of exposures on health outcomes is central to understanding causal pathways. However, microbiome sequencing primarily measures relative abundance (RA). Conducting regression analyses directly on RA can exhibit distorted null behavior due to compositionality (the unit-sum constraint), leading to inflated false positives in mediation inference. The following figure illustrates how compositionality can induce apparent exposure-taxon and taxon-outcome associations on the RA scale even when AA-scale mediation is sparse and localized.
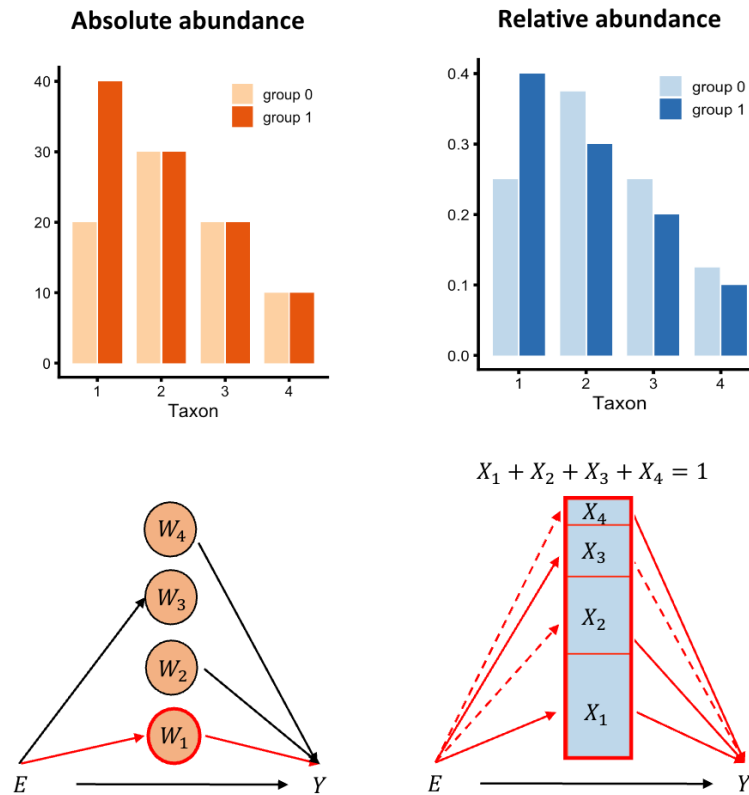
The biologically relevant target is typically the absolute abundance (AA) scale. CAMRA is proposed as a principled remedy: it takes standard RA/count data as input but targets mediation effects defined on the AA scale, recovering the AA-level evidence from RA observations to ensure rigorous False Discovery Rate (FDR) control. If the latent AA variable W_{ik} for subject i and taxon k were observed, the two standard causal mediation models would be:

AA-based taxa model:
$$\mathbb{E}(W_{ik}|E_i) = \exp(\alpha_{k0} + E_i\alpha_k)$$

AA-based outcome model:

$$g(\mathbb{E}(Y_i)|E_i, \log(W_i)) = \beta_0 + E_i\beta_E + \sum_{k=1}^{p} \log(W_{ik})\beta_k$$

The mediation null hypothesis for taxon k is

$$H_{0k} : \alpha_k\beta_k = 0.$$

Because AA is unobserved, CAMRA addresses this through three modular steps:

Step 1. Exposure → Microbiome Association:

CAMRA estimates exposure effects using PALM, which links the RA-based exposure coefficient to its AA-level counterpart, using a robust median-based correction to yield AA-calibrated estimates and p-values p_{α_k}.

Step 2. Microbiome → Outcome Association:

CAMRA utilizes the PALAR transformation, mapping RA profiles to transformed predictors whose coefficients coincide with the AA-based model. A sparse de-biased Lasso procedure is applied to obtain p-values p_{β_k}.

Step 3. Composite-null Mediation Testing:

CAMRA applies the HDMT procedure to the path-specific p-values (p_{α_k} and p_{β_k}) to estimate the proportions of the three sub-nulls, yielding calibrated taxon-level mediation q-values for FDR control.

Application: Gut microbiome mediation of between-country differences in BMI

To illustrate CAMRA, we analyze whether specific gut microbial taxa mediate the association between country of origin (United States vs. China) and body mass index (BMI).

First, we load the package and the dataset. (`data.bmi` is the same dataset we used in our paper).

```
library(miMediation)
data("data.bmi")
Trt <- data.bmi$treatment
table(Trt) # 0: China, 1: United States
```

```
## Trt
##   0   1
## 446 397
```

```
M <- data.bmi$mediators
dim(M)
```

```
## [1] 843 186
```

```
Y <- data.bmi$outcome
summary(Y)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.67   20.65   23.43   24.23   26.13   57.46
```

Then we use `CAMRA` function in the function to detect the significant mediators. The function can also return the q-value of all taxa, and the path-specific p-values p_{α_k} and p_{β_k}.

```
CAMRA_res <- miMediation::CAMRA(
  mediators = data.bmi$mediators,
  treatment = data.bmi$treatment,
  outcome = data.bmi$outcome,
  confounders = NULL,
  fdr.alpha = 0.05,
  seed = 123
)

# View significant microbial mediators
print(CAMRA_res$sig.mediators)
```

```
## [1] "species:Acidaminococcus intestini"
```

References

Wang Q, Li Y, Peng Y, Tang, ZZ (2026). Error control in microbiome mediator discovery: benchmark and remedy. Submitted.

Wei Z, Hong Q, Chen G, Hartert TV, Rosas-Salazar C, Das SR, Shilts MH, Levin AM, Tang ZZ (2026). Fast and reliable association discovery in large-scale microbiome studies and meta-analyses using PALM. Genome Biology, accepted.

Li Y, Wang Q, Feng Z, Wang X, Tang ZZ (2025). PALAR: Estimation of absolute abundance effects in regression with relative abundance predictors. Journal of the American Statistical Association (JASA), DOI: 10.1080/01621459.2025.2596250.

Dai JY, Stanford JL, LeBlanc M (2020). A multiple-testing procedure for high-dimensional mediation hypotheses. Journal of the American Statistical Association (JASA), DOI: 10.1080/01621459.2020.1765785.