

easy16S : une application Shiny pour explorer ses données métagénomiques

Cédric Midoux 1*, 2†, 3‡, § Mahendra Mariadassou 2, 3 ¶

Résumé

L'analyse des données du microbiome est devenue un atout majeur pour l'étude de la diversité et de la dynamique microbienne dans divers domaines de la biologie. Après une étape bioinformatique permettant le passage des données de séquences (FASTQ) à une matrice d'abondance des espèces microbiennes (OTU : Operational Taxonomic Unit) et une table d'annotation taxonomique, une étape d'analyse statistique est nécessaire pour répondre aux questions biologiques (*étude de la diversité, comparaison d'abondance en fonction des conditions expérimentales, évolution d'un écosystème, ...*). Il existe une demande croissante d'outils interactifs conviviaux permettant aux chercheurs d'analyser et de visualiser leurs données de manière autonome, sans dépendre d'un biostatisticien ou sans avoir besoin d'acquérir des compétences en programmation R.

Nous présentons ici **easy16S**, un package R et une application Shiny visant à faciliter l'analyse des données du microbiome. Cette application s'appuie sur un objet `phyloseq` (McMurdie and Holmes (2013)) constitué d'une matrice d'abondance des OTU, d'un `data.frame` de métadonnées des échantillons et d'une matrice d'affiliation taxonomique des OTU. Cette application est intuitive et est orientée pour la visualisation des variables d'intérêt sur la structuration des communautés microbiennes.

Après le chargement de ses données brutes, l'utilisateur peut facilement les prétraiter. Cela inclut des options telles que le filtrage des échantillons et des taxons, la modification du tableau d'affiliations, la raréfaction ou la transformation de la matrice de comptage.

L'utilisateur peut ensuite réaliser diverses analyses, telles que :

- Tableaux constituant l'objet `phyloseq`
- Visualisation des métadonnées grâce à `esquisse` (Meyer and Perrier (2024))
- Barplot de composition
- Courbes de raréfaction
- Heatmap d'abondance
- Richesse au sein d'un échantillon
- Dissimilarité entre les échantillons
- *MultiDimensional Scaling*
- Analyse d'abondance différentielle

L'application s'adresse à des utilisateurs débutants souhaitant mener leurs analyses sans compétence technique, à des utilisateurs experts souhaitant visualiser rapidement les patterns et tendances présents au sein de ses données, ainsi qu'aux apprenants lors de sessions de formation.

Mots-clefs : Bioinformatique - Métagénomique - Visualisation - Shiny

Développement

Le package est versionné sur la ForgeMIA d'INRAE (<https://forgemia.inra.fr/migale/easy16s>). Le framework `golem` (Fay et al. (2023)) a été utilisé pour faciliter le développement. Une documentation est disponible

*Université Paris-Saclay, INRAE, PROSE, 92761, Antony, France

†Université Paris-Saclay, INRAE, MaIAGE, 78350, Jouy-en-Josas, France

‡Université Paris-Saclay, INRAE, BioinfOmics, MIGALE bioinformatics facility, 78350, Jouy-en-Josas, France

§cedric.midoux@inrae.fr

¶mahendra.mariadassou@inrae.fr

grâce à `pkgdown` et via intégration continue : <https://easy16s.migale.inrae.fr/>. De plus, une image docker de l'application est déployée grâce à l'intégration continue de GitLab. Enfin, une instance de l'application est disponible en accès libre et est hébergée grâce à shinyproxy sur <https://shiny.migale.inrae.fr/app/easy16S>.

L'application a été soumise à *Journal of Open Source Software* (<https://joss.theoj.org/papers/e03bb9530fd2e1c0621e35352b71691e>).

Lors de cette courte présentation, seront présentés succinctement les besoins ayant motivé le développement de cette application, les méthodologies de développement mises en place, y compris la rédaction de la documentation et l'utilisation en sessions de formation, ainsi que les stratégies de déploiement et de valorisation adoptées.

Références

- Fay, Colin, Vincent Guyader, Sébastien Rochette, and Cervan Girard. 2023. *Golem: A Framework for Robust Shiny Applications*. <https://CRAN.R-project.org/package=golem>.
- McMurdie, Paul J., and Susan Holmes. 2013. "Phyloseq: An r Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data." *PLoS ONE* 8 (4): e61217. <https://doi.org/10.1371/journal.pone.0061217>.
- Meyer, Fanny, and Victor Perrier. 2024. *Esquisse: Explore and Visualize Your Data Interactively*. <https://CRAN.R-project.org/package=esquisse>.