

Diversity of foliar endophytic ascomycetes in the endemic Corsican pine forests

Adrien Taudière^{1, 2*}, Jean-Michel Bellanger¹, Christopher Carcaillet^{2, 3}, Laetitia Hugot⁴, Finn Kjellberg¹, Aarón Lecanda⁵, Annick Lesne^{6, 7}, Pierre-Arthur Moreau⁸, Karin Scharmann⁵, Sebastian Leidel⁵, Franck Richard^{1*}

Affiliation and Address:

1/ UMR 5175 CEFE – CNRS - Université de Montpellier - Université Paul Valéry Montpellier 3 - EPHE – IRD - INSERM, Campus CNRS, 1919 Route de Mende, F-34293 Montpellier, France

2/ École Pratique des Hautes Études (EPHE) & Paris Research University, Paris, France

3/ Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés (UMR5023 CNRS), Université Lyon 1, Villeurbanne Cedex, France

4/ Conservatoire Botanique National de Corse. Route d'Aleria. 20250 Corte, France.

5/ Max Planck Research Group for RNA Biology, Max Planck Institute for Molecular Biomedicine, Von-Esmarch-Strasse 54, 48149 Muenster, Germany

6/ Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75252, Paris, France

7/ Institut de Génétique Moléculaire de Montpellier,, University of Montpellier, CNRS, Montpellier, France

8/ Département de Botanique, Faculté des Sciences Pharmaceutiques et Biologiques, Université Lille - Nord de France B. P. 83, 59006 Lille Cedex, France.

* Corresponding authors: adrien.taudiere@zaclys.net; franck.richard@cefe.cnrs.fr

Supplementary materials (Available at https://github.com/adrientaudiere/FEF_paper):

FIG S1 (pdf): Bioinformatics pipeline. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S1.pdf

SM S2 (markdown and pdf): UNIX scripts for bioinformatics pipeline. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S2.Rmd and https://github.com/adrientaudiere/FEF_paper/blob/master/S2.pdf

SM S3 (pdf and associated rnw): Ecological analysis of the Uparse dataset discarding unique sequences in the Bio-informatic pipeline. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S3.pdf

SM S4 (pdf and associated rnw): Ecological analysis of the Uparse dataset discarding unique sequences in the Bio-informatic pipeline but keeping sequences putatively assigned to “Tracheophyta” by ITSx. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S4.pdf

SM S5 (pdf and associated rnw): Ecological analysis of the Uparse dataset allowing unique sequences in the Bio-informatic pipeline. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S5.pdf

SM S6 (pdf and associated rnw): Ecological analysis of the Swarm dataset. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S6.pdf

SM S7 (pdf and associated rnw): Ecological analysis of the Open Ref dataset. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S7.pdf

SM S8 (pdf and associated rnw): Ecological analysis of the Closed Ref dataset. Available at: https://github.com/adrientaudiere/FEF_paper/blob/master/S8.pdf

FIG S9 (html): Comparison of four taxonomic assignment methods (RDP, Blast, Mothur and Uclust). We build this figure using Krona (Ondov et al. 2011). Available at: http://htmlpreview.github.io/?https://github.com/adrientaudiere/FEF_paper/blob/master/S9.html

FIG S10 (html): FastQC result of sequences quality before quality filtering. Available to download: http://htmlpreview.github.io/?https://github.com/adrientaudiere/FEF_paper/blob/master/S10.html

FIG S11 (html): FastQC result of sequences quality after quality filtering. Available to download: http://htmlpreview.github.io/?https://github.com/adrientaudiere/FEF_paper/blob/master/S11.html

FIG S12 (this document): Assignment quality by RDP classifier measured by bootstrap values in relation to the number of sequences by OTUs.

FIG S13 (this document): Number of sequences by OTUs (molecular Abundances) in function of the more precise taxonomic rank assigned with confidence by RDP classifier (bootstrap value > 80%).

FIG S14 (this document): (a) Distribution of molecular abundances (number of sequences) according to the taxonomy of OTUs. (b) Taxonomic distribution of OTUs represented by at least 1 000 sequences.

FIG S15 (this document): Distribution of OTUs (left panel) and sequences (right panel) across putative ecology

Table S16 (this document): Result of PerManova including tree individual

FIG S17 (this document): Molecular abundance (number of sequences) distribution of OTUs showing significant differences across sites (DESeq2 R package).

Table S18 (this document): Effect of sites, host age and elevation on FEF diversity (ANOVA).

Figure S12: Assignment quality of OTUs by RDP classifier measured by bootstrap values in relation to the number of sequences for each OTU. Smoothed curves were obtained using the non-parametric loess function in R 3.2.4 package (R Core Team 2016), with span argument value of 0.35. Dots represent OTUs and the level of their final assignment, from the phylum (in green) to the species (in orange). Envelopes indicate confidence intervals (0.95 by default).

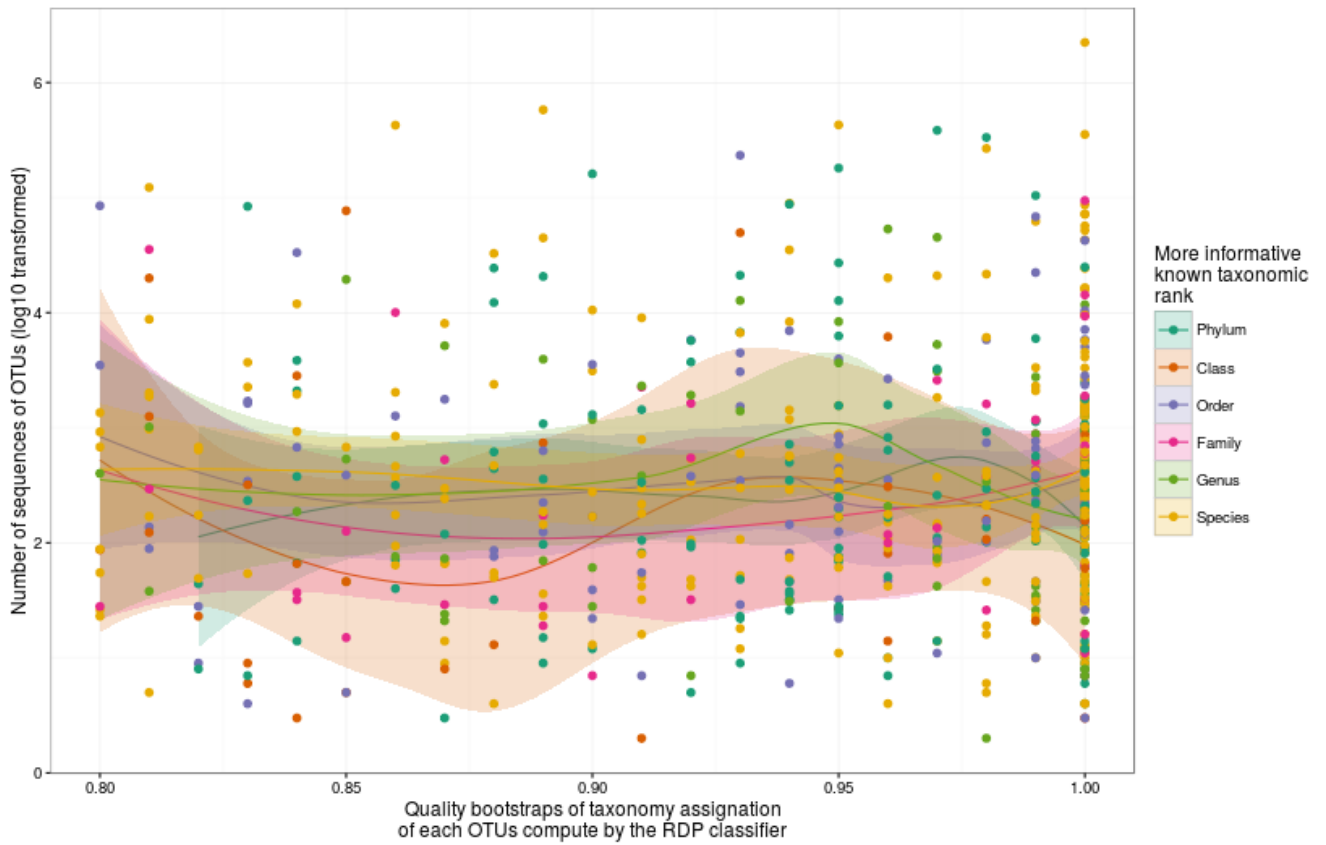
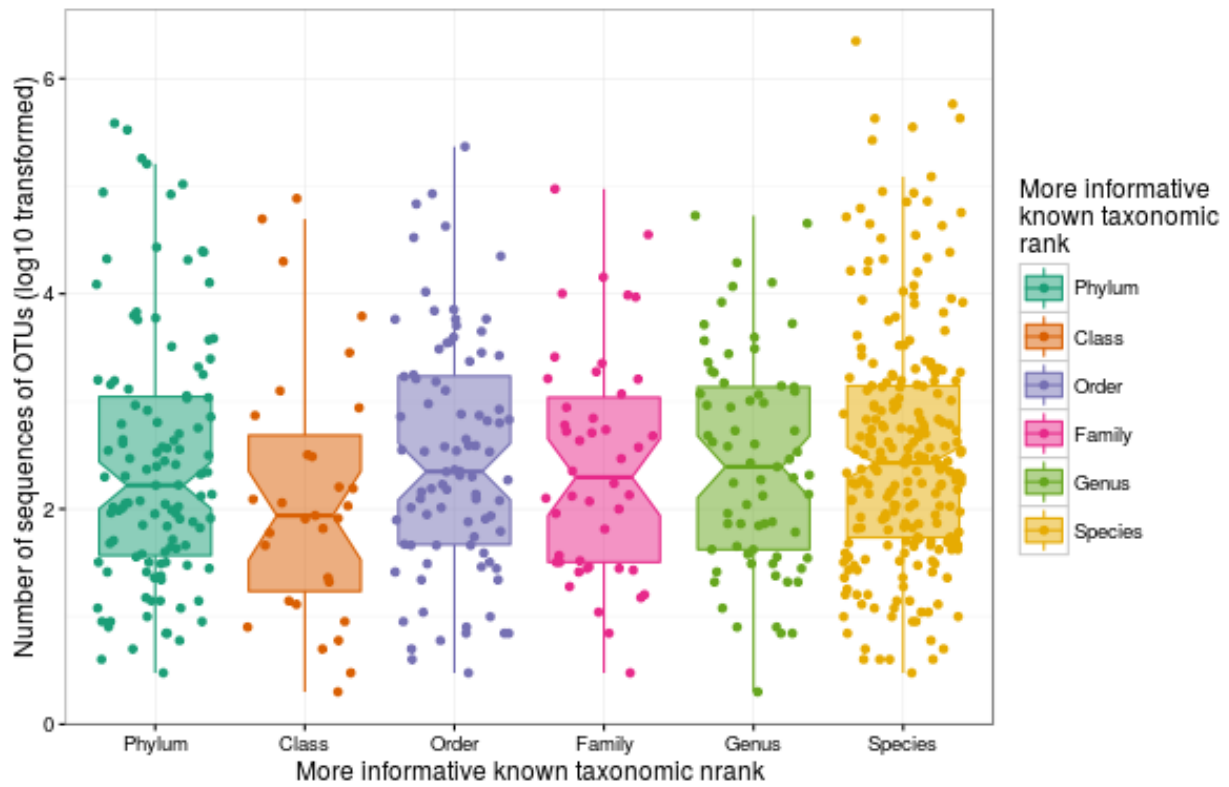


Figure S13: Number of sequences by OTU (molecular Abundances) in function of the more precise taxonomic rank assigned with confidence by RDP classifier (bootstrap value > 80%).



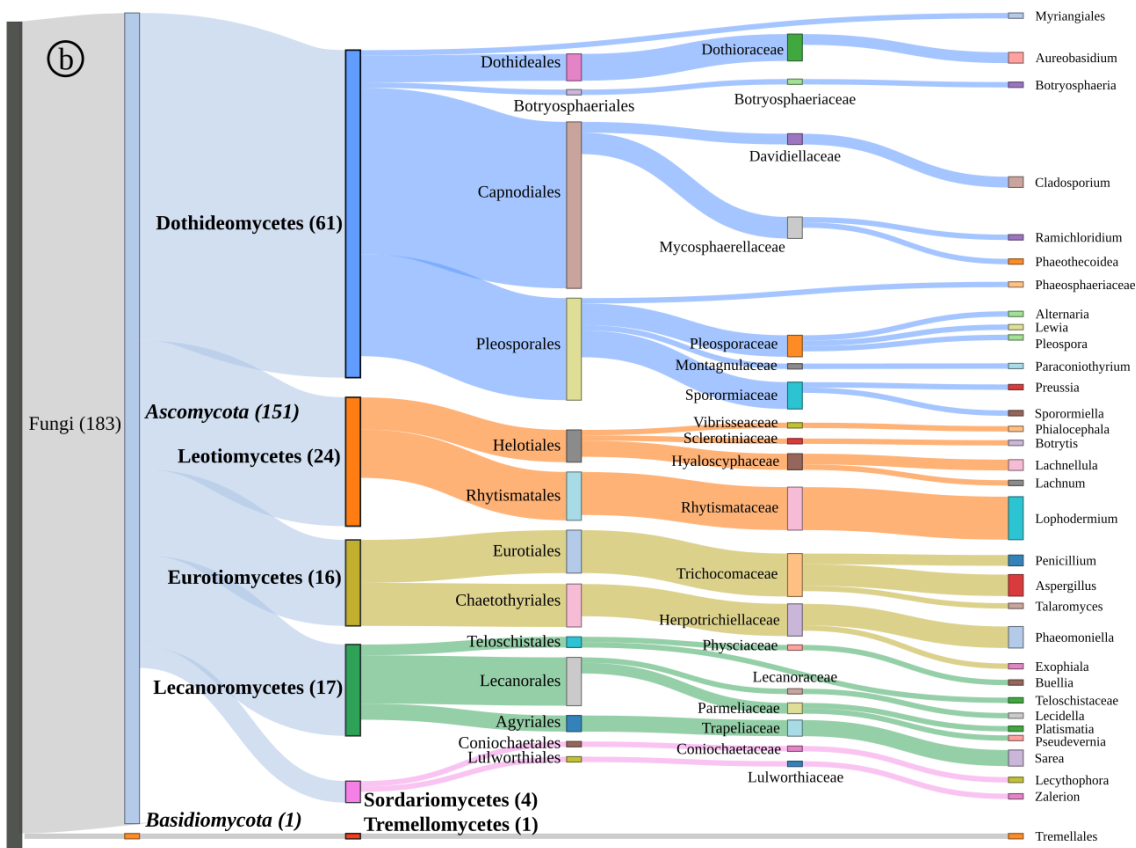
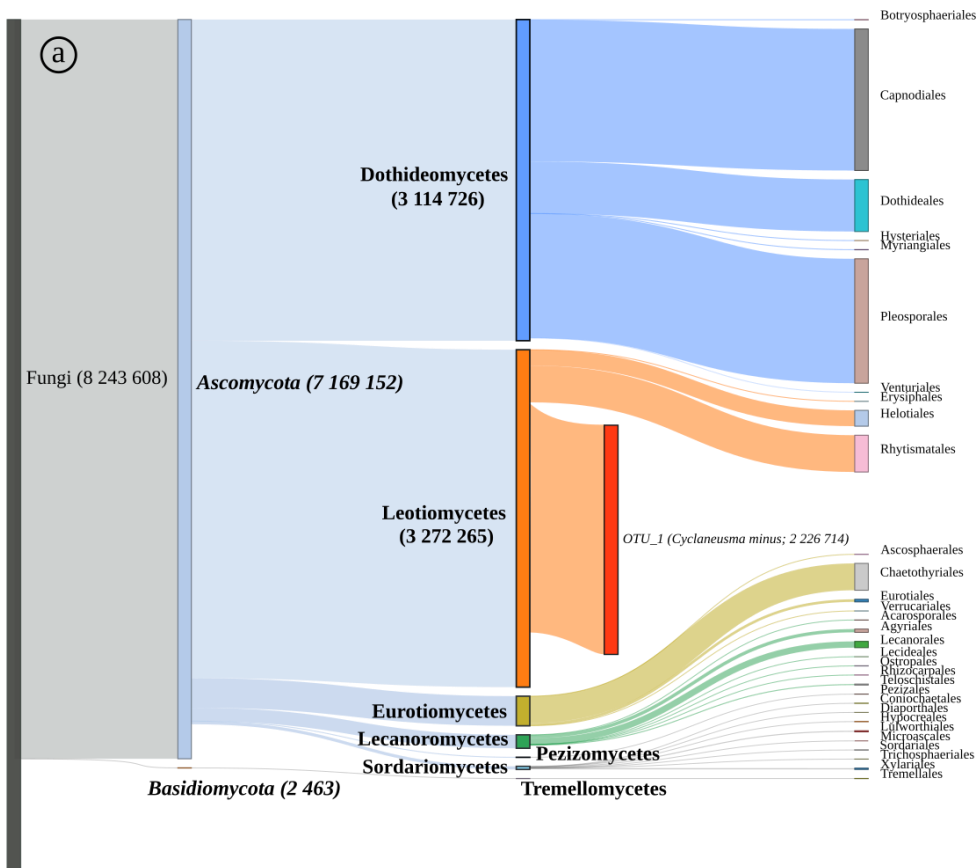


Figure S14: (a) Distribution of molecular abundances (number of sequences) according to the taxonomy of OTUs. (b) Taxonomic distribution of OTUs represented by at least 1 000 sequences.

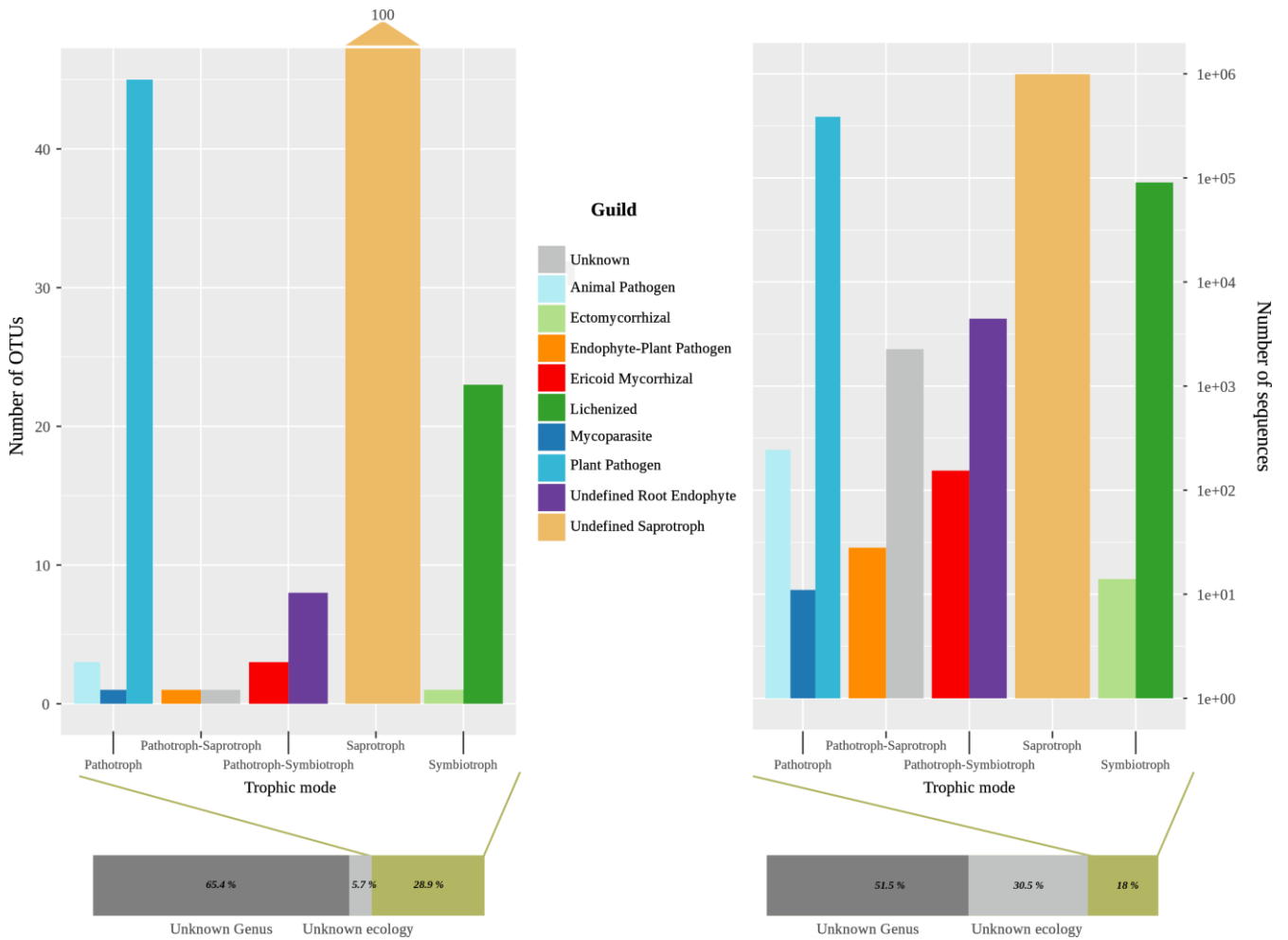


Figure S15: Distribution of OTUs (left) and sequences (right) according to putative functional guilds based on FUNGUILD.

Table S16. Effect of site, host age and tree individual on FEF composition (PERMANOVA). Results with a significant p-value are in bold.

	df	Sums of squares	Mean squares	Pseudo F-statistics	R2	Pr(>F)
Sites	2	1.9928	0.99639	6.8811	0.13599	0.0001
Age	2	0.5816	0.29082	2.0084	0.03969	0.0026
Sites:Age	4	1.3344	0.3336	2.3039	0.09106	0.0001
Tree	18	4.2294	0.23497	1.6227	0.28861	0.0001
Residuals	45	6.516	0.1448		0.44465	
Total	71	14.6543			1	



Figure S17: Molecular abundance (number of sequences) distribution of OTUs showing significant differences across sites. Columns indicate samples. Lines correspond to OTUs detected by the DESq2 method to be significantly different in abundance among sites. OTUs are ranked in function of their similarity of distribution among samples.

Table S18: Effect of site, host age and light exposure on FEF diversity by ANOVA.

		⁰ H		¹ H		² H	
		Estimate ± sd	P-value	Estimate ± sd	P-value	Estimate ± sd	P-value
	<i>(Intercept)</i>	51.53 ± 22.98	0.0285*	12.22 ± 3.39	0.0006	7.04 ± 1.78	0.0002
	<i>sqrt(readNb)</i>	0.18 ± 0.06	0.0029	0.02 ± 0.01	0.0569	0.01 ± 0.004	0.1459
Sites	<i>Bavella</i>	8.1 ± 10.31	0.4348	-9.56 ± 1.52	<0.0001	-5.39 ± 0.8	<0.0001
	<i>Verghello</i>	-10.86 ± 10.34	0.2976	-3.55 ± 1.53	0.0233	-2.23 ± 0.8	0.0071
Age	<i>Old</i>	-9.72 ± 10.18	0.3432	0.87 ± 1.5	0.5636	0.65 ± 0.79	0.4132
	<i>Young</i>	-21.92 ± 10.48	0.0404	-0.48 ± 1.55	0.7556	0.46 ± 0.81	0.5706
Elevation	<i>Middle</i>	13.4 ± 10.39	0.202	1.88 ± 1.53	0.2251	0.6 ± 0.81	0.4606
	<i>Top</i>	5.74 ± 10.26	0.578	1.34 ± 1.51	0.3794	0.57 ± 0.8	0.4771

⁰H: Species richness

¹H: Exponential of Shannon's entropy index

²H: Inverse of Simpson's concentration index

* Significant p-value (*i.e.* $p < 0.05$) are in bold.

