Supplement of Biogeosciences, 21, 2227–2252, 2024
https://doi.org/10.5194/bg-21-2227-2024-supplement

Biogeosciences Open Access EGU

Supplement of

Environmental controls on the distribution of brGDGTs and brGMGTs across the Seine River basin (NW France): implications for bacterial tetraethers as a proxy for riverine runoff

Zhe-Xuan Zhang et al.

Correspondence to: Arnaud Huguet (arnaud.huguet@sorbonne-universite.fr)

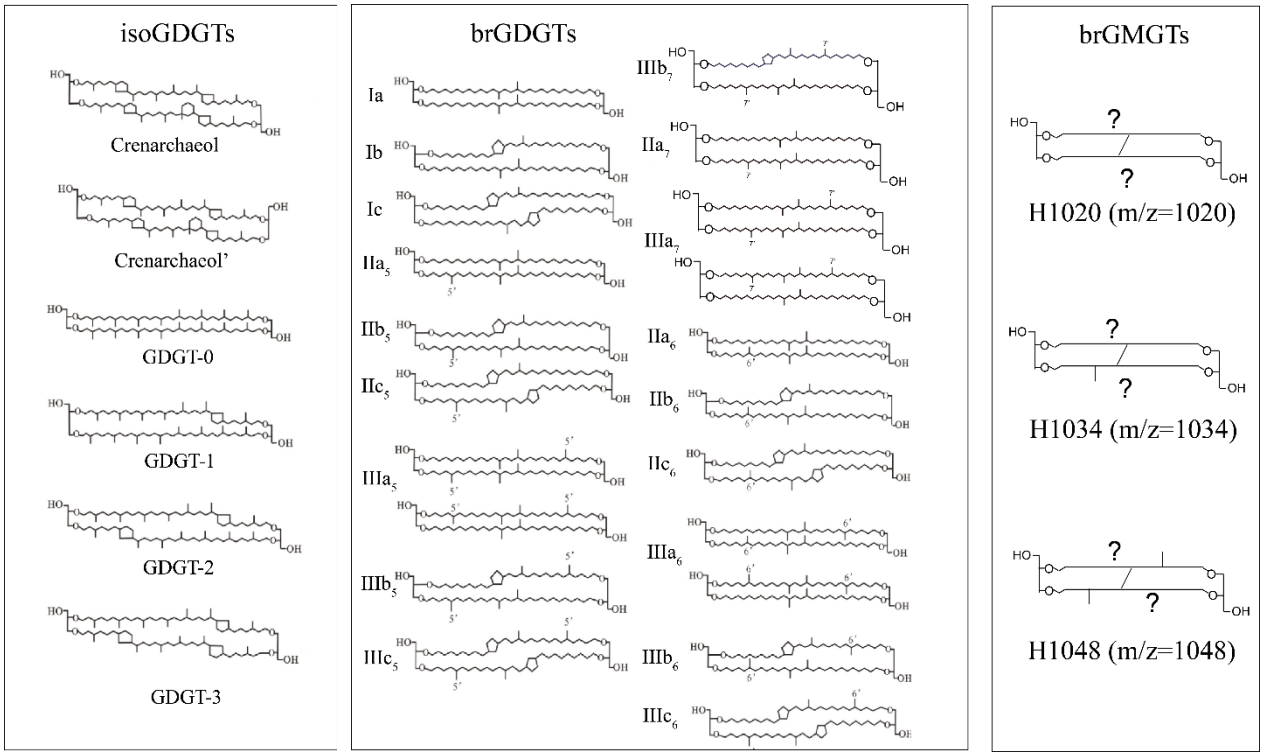

Fig. S1. Structures of isoGDGTs, brGDGTs, and brGMGTs. Note that the structures of brGMGTs and compounds eluting later than 7-Methyl brGDGTs (1050d and 1036d) have not been described.

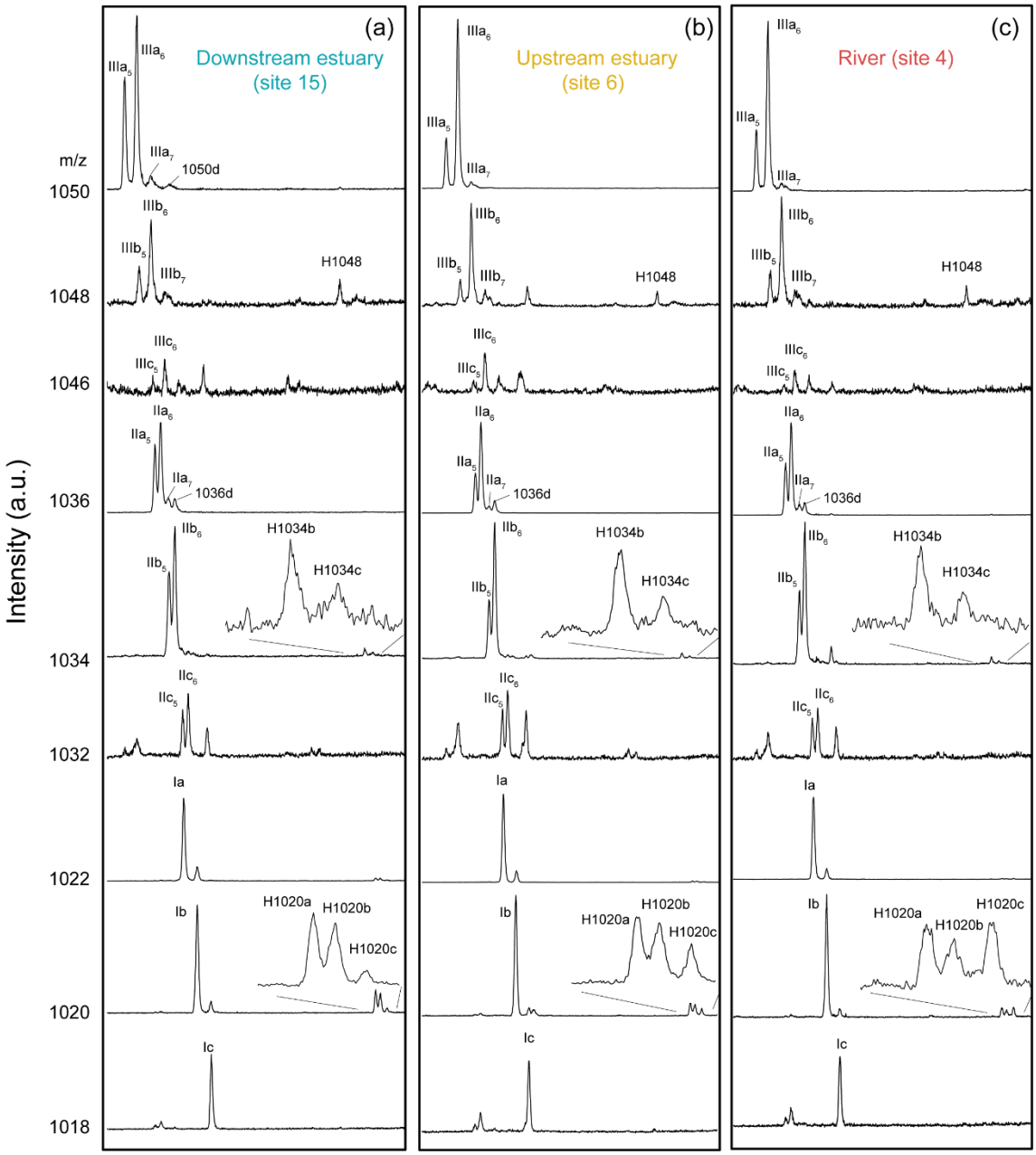

Fig. S2. Extracted chromatograms of brGDGTs and brGMGTs for the SPM samples collected in (a) site 15 (Tancarville, September 2020), (b) site 6 (Oissel, July 2019) and (c) site 4 (Les Andelys, July 2019). The nomenclature for the penta- and hexamethylated brGDGTs: 5-methyl brGDGTs (IIIa$_5$, IIIb$_5$, IIIc$_5$, IIa$_5$, IIb$_5$, and IIc$_5$); 6-methyl brGDGTs (IIIa$_6$, IIIb$_6$, IIIc$_6$, IIa$_6$, IIb$_6$, and IIc$_6$); 7-methyl brGDGTs (IIIa$_7$, IIIb$_7$, and IIa$_7$). 1050d and 1036d represent compounds eluting later than IIIa$_7$ and IIa$_7$, respectively.

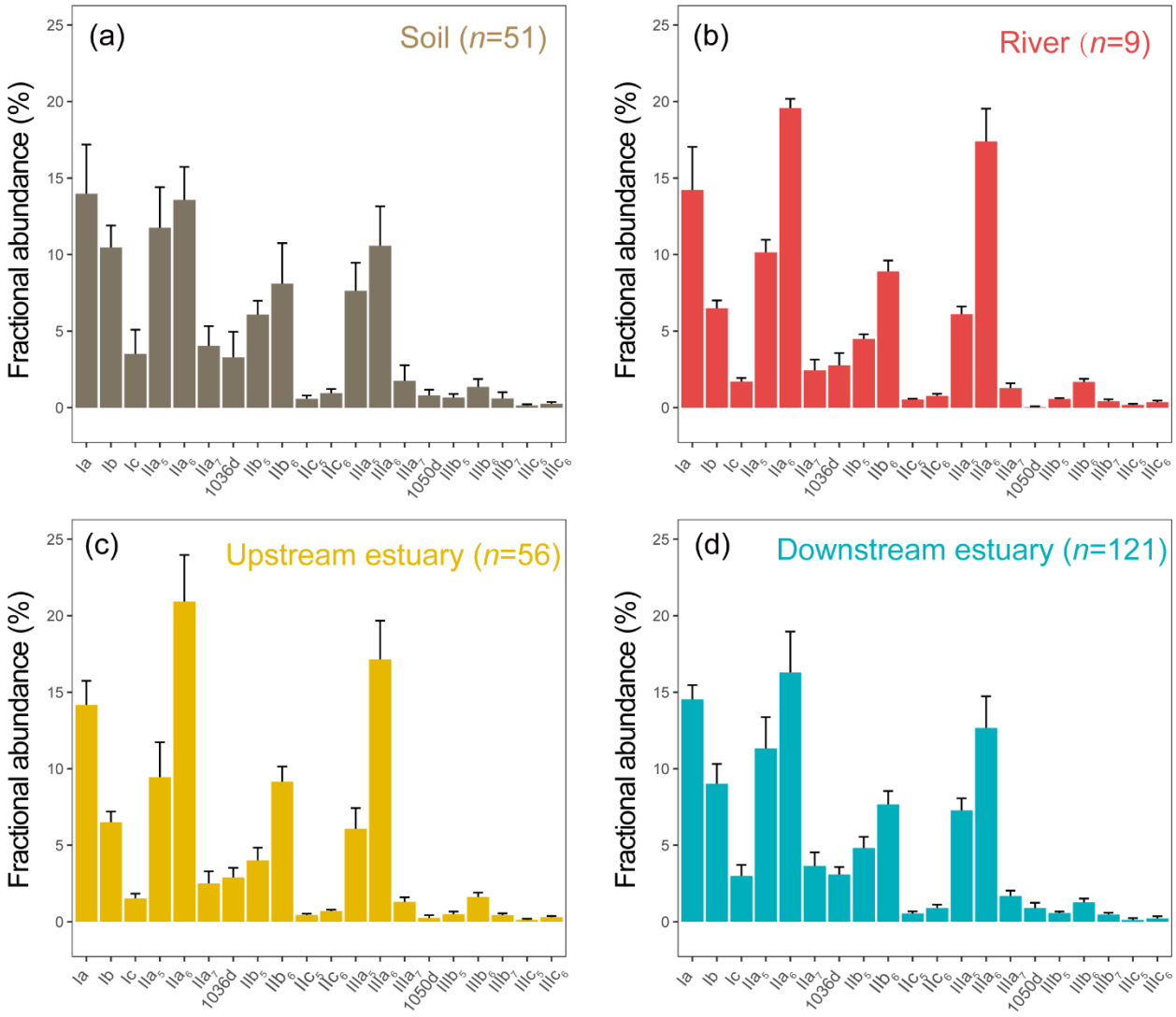

Fig. S3. Distribution of brGDGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (*n*=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin.

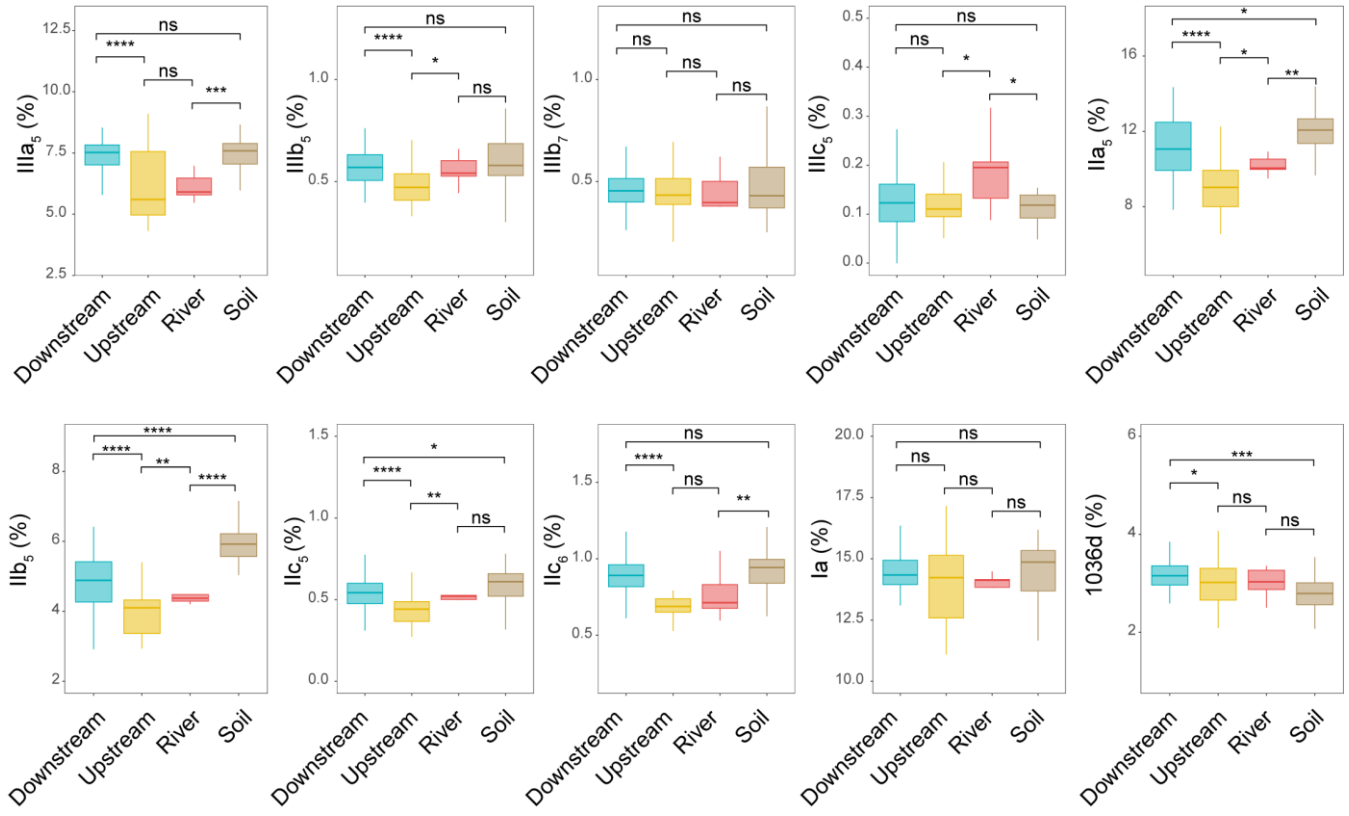

Fig. S4. Relative abundance of individual brGDGTs over 20 brGDGTs (IIIa$_5$, IIIb$_5$, IIIc$_5$, IIa$_5$, IIb$_5$, IIc$_5$, IIIa$_6$, IIIb$_6$, IIIc$_6$, IIa$_6$, IIb$_6$, IIc$_6$, IIIa$_7$, IIIb$_7$, IIa$_7$, Ia, Ib, Ic, 1050d, and 1036d) across the Seine River basin. Box plots of upstream and downstream estuary are composed of SPM and river channel sediments, whereas those of river are composed of SPM. Boxes show the upper and lower quartiles of the data, and whiskers show the range of the data, which are color-coded based on the sample type (river in red, upstream estuary in yellow, and downstream estuary in blue). The center-line in the boxes indicates the median value of the dataset. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

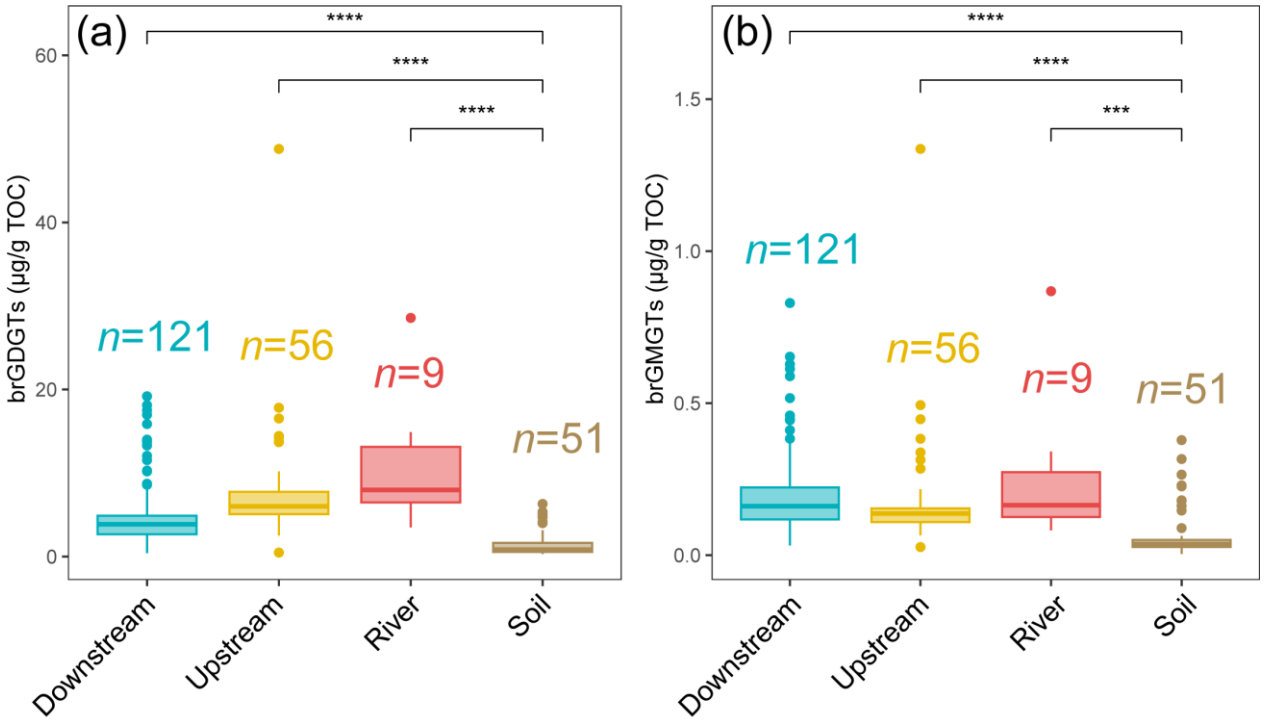

Fig. S5. Concentrations (normalized to total organic carbon) of (a) total brGDGTs and (b) total brGMGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin. Box plots of upstream and downstream estuary samples are based on SPM and river channel sediments, whereas those of river samples are based only on SPM. Boxes show the upper and lower quartiles of the data, and whiskers show the range of the data, which are color-coded based on the sample type (river in red, upstream estuary in yellow, and downstream estuary in blue). The center-line in the boxes indicates the median value of the dataset. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

Table S1. Description of available environmental parameters

	River	Upstream estuary	Downstream estuary	Soil
Min temperature (°C)	20	8.49	6.4	n.a.
Max temperature (°C)	23.41	24.4	23.38	n.a.
Mean temperature (°C)	21.51	20.09	18.27	n.a.
Number of samples	6	44	62	n.a.
Min salinity	0	0	0.1	n.a.
Max salinity	0.3	0.32	32.3	n.a.
Mean salinity	0.2	0.22	3.77	n.a.
Number of samples	6	43	60	n.a.
Min discharge (m^3/s)	99	99	99	n.a.
Max discharge (m^3/s)	156	978	978	n.a.
Mean discharge (m^3/s)	129.78	183.62	218.85	n.a.
Number of samples	9	48	62	n.a.
Min TOC (%)	0.82	0.75	0.11	0.22
Max TOC (%)	4.22	7.71	7.35	22.28
Mean TOC (%)	2.88	4.64	3.3	3.03
Number of samples	9	57	120	51
Min TN (%)	0.12	0.12	0.01	0.01
Max TN (%)	0.58	0.84	0.619	1.07
Mean TN (%)	0.37	0.51	0.31	0.24
Number of samples	9	57	120	51

n.a.= not applicable

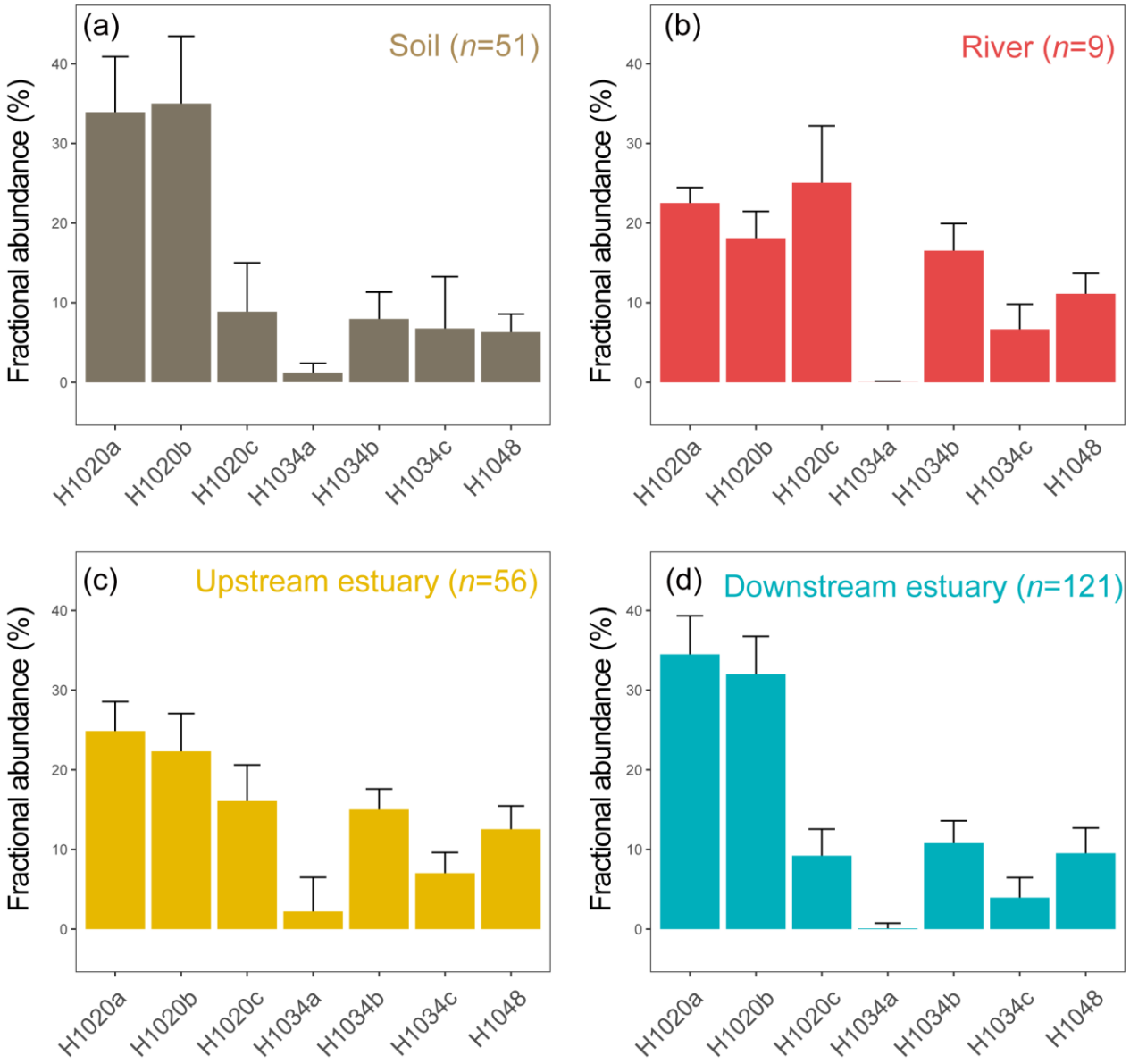

Fig. S6. Distribution of brGMGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin.

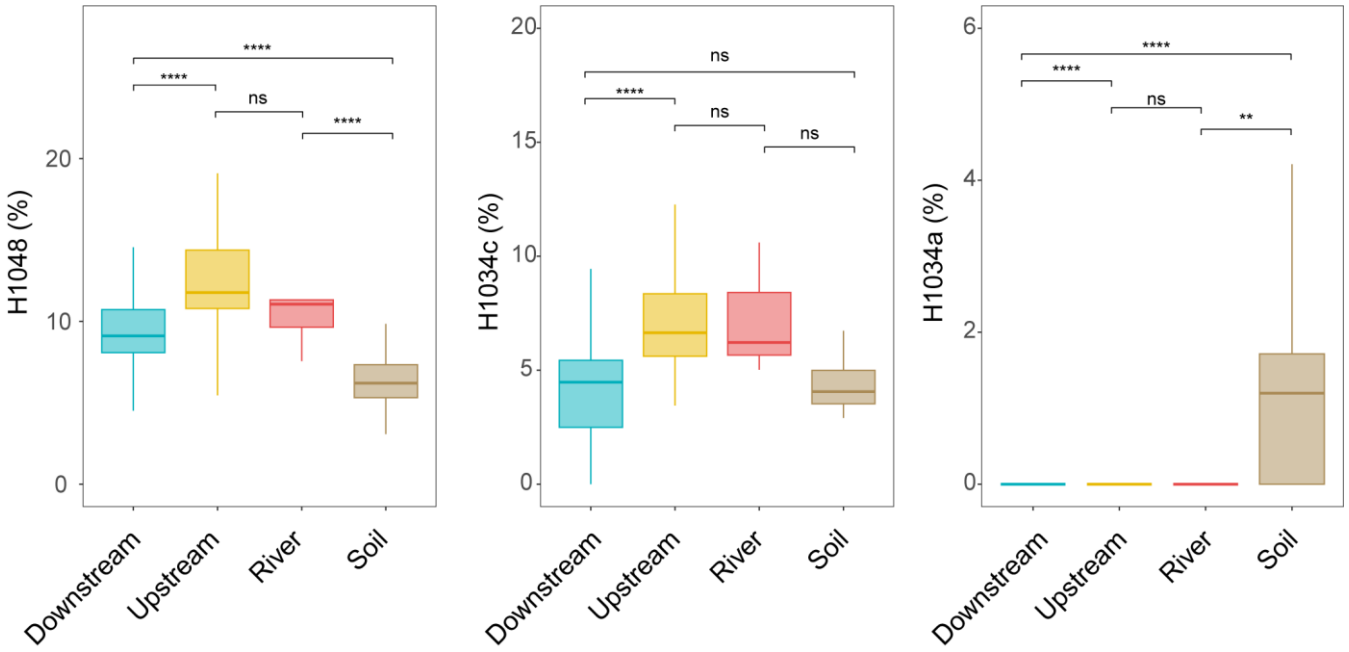

Fig. S7. Relative abundance of brGMGTs over 7 brGMGTs (H1020a, H1020b, H1020c, H1034a, H1034b, H1034c, and H1048) across the Seine River basin. Box plots of upstream and downstream estuary are composed of SPM and sediments, whereas those of river are composed of SPM. Boxes show the upper and lower quartiles of the data, and whiskers show the range of the data, which are color-coded based on the sample type (river in red, upstream estuary in yellow, and downstream estuary in blue). The center-line in the boxes indicates the median value of the dataset. Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

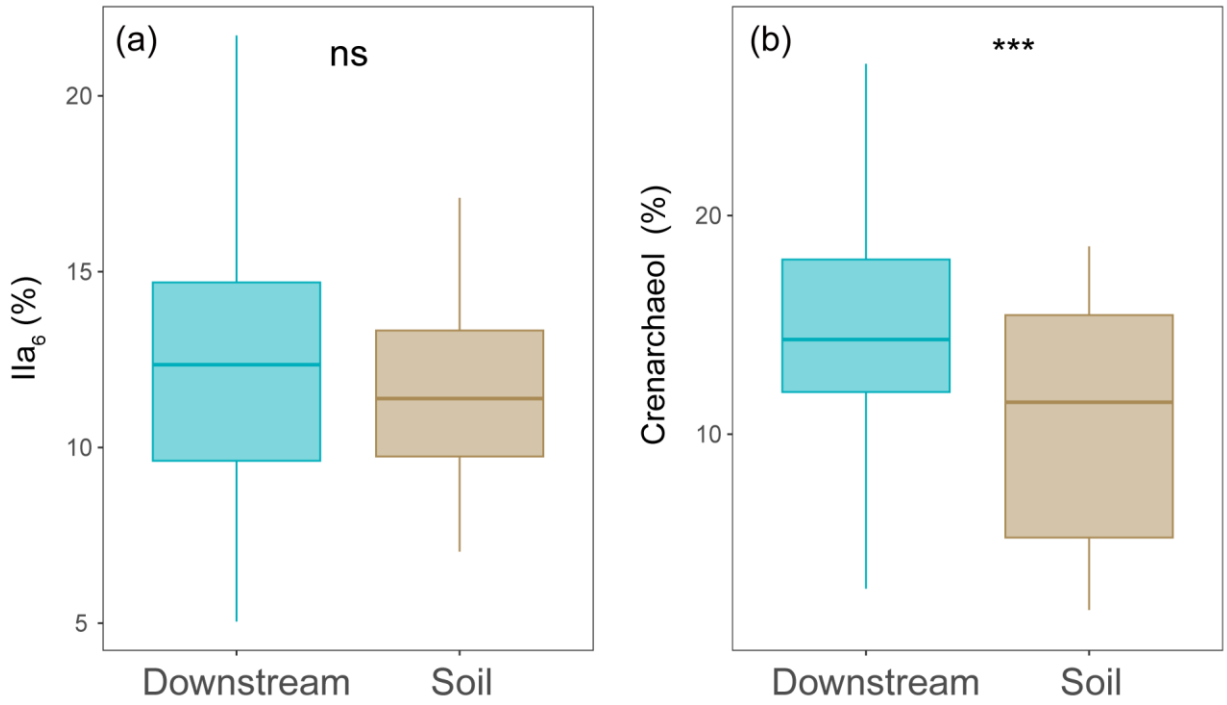

Fig. S8. Relative abundance of IIa$_6$ (a) and crenarchaeol (b) over 19 GDGTs (GDGT-0, GDGT-1, GDGT-2, GDGT-3, Crenarchaeol, Crenarchaeol', IIIa$_5$, IIIa$_6$, IIIb$_5$, IIIb$_6$, IIa$_5$, IIa$_6$, IIb$_5$, IIb$_6$, IIc$_5$, IIc$_6$, Ia, Ib, and Ic) used in the BigMac model. Boxes show the upper and lower quartiles of the data, and whiskers show the range of the data, which are color-coded based on the sample type (downstream estuary in blue and soil in brown). The center-line in the boxes indicates the median value of the dataset. Statistical testing was performed by a Wilcoxon test (*$p < 0.001$; ns, not significant, $p > 0.05$).**

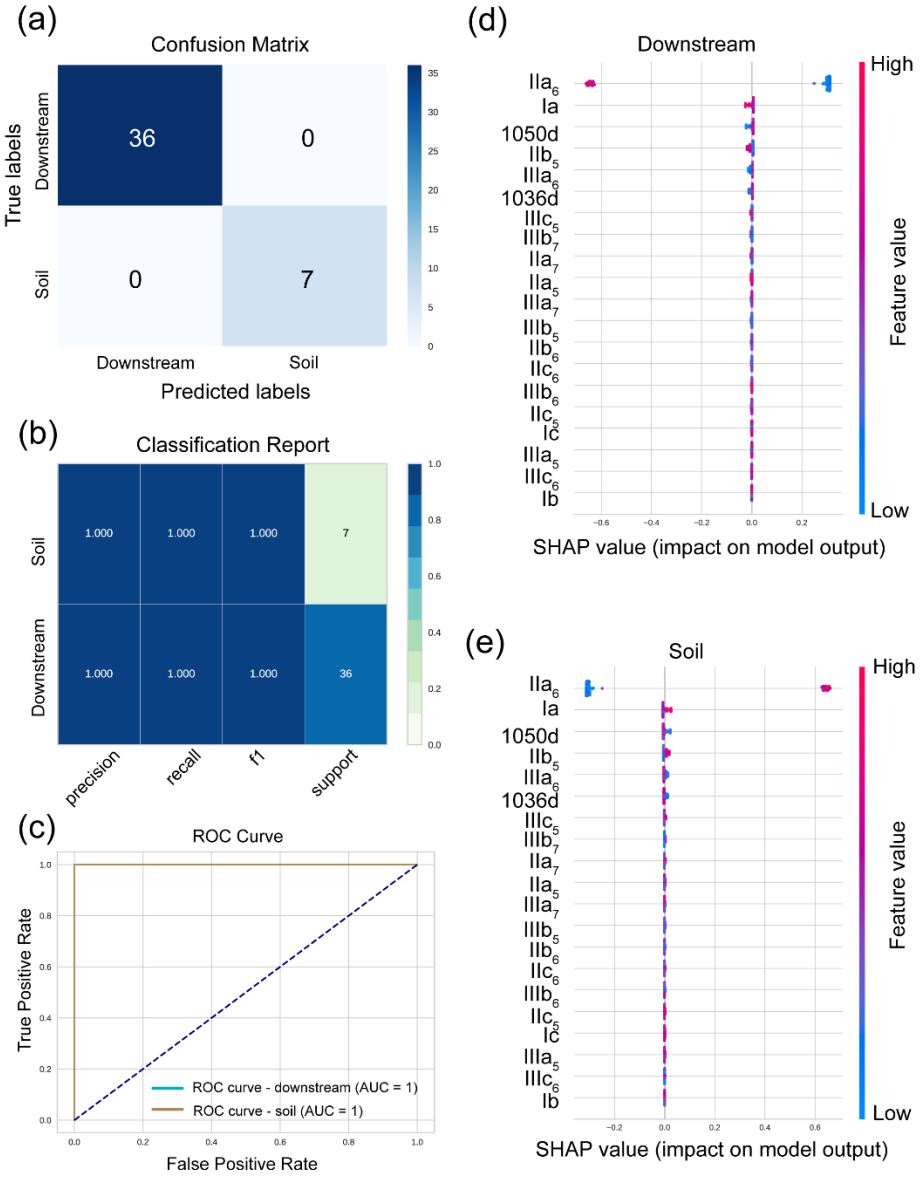

Fig. S9. Evaluation of the random forest model based on brGDGTs through the confusion matrix (a), classification report (b), and receiver operating characteristic (ROC) curve (c). SHAP summary plots (d-e) show the feature importance obtained from the random forest algorithm and the SHAP library. Each bullet in the plot represents a single sample in the training set, with the color indicating the feature value (fractional abundance of the brGDGTs) from low (blue) to high (pink). The bullets positioned on the right side of the SHAP summary plot correspond to positive SHAP values, indicating a positive effect on the model output (downstream estuary or soils). The bullets on the left side of the plot indicate negative SHAP values, suggesting a negative effect on the model output. The variables (brGDGTs) with higher impact on the model performance are shown at higher positions. Training sets include downstream SPM and sediment samples (d) and soils (e).

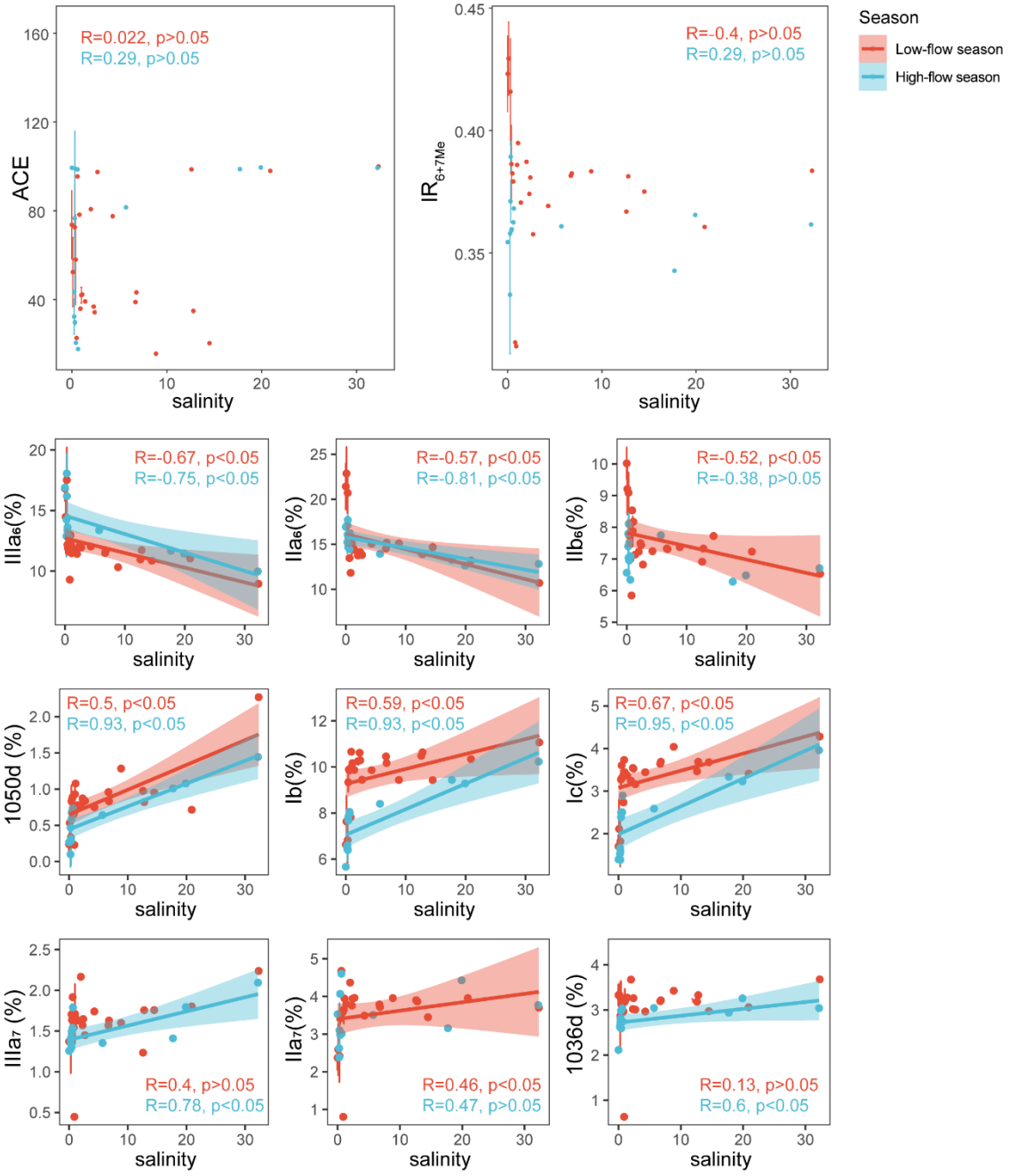

Fig. S10. Salinity plotted versus ACE, IR_{6+7Me}, relative abundance of 6-methyl and 7-methyl brGDGTs ($IIIa_6$, IIa_6, IIb_6, $IIIa_7$ and IIa_7) as well as compounds 1050d, 1036d, Ib, and Ic through the linear regression. Shaded area represents 95% confidence intervals. Vertical error bars indicate mean ± s.d for samples with the same salinity. Dataset is composed of SPM.

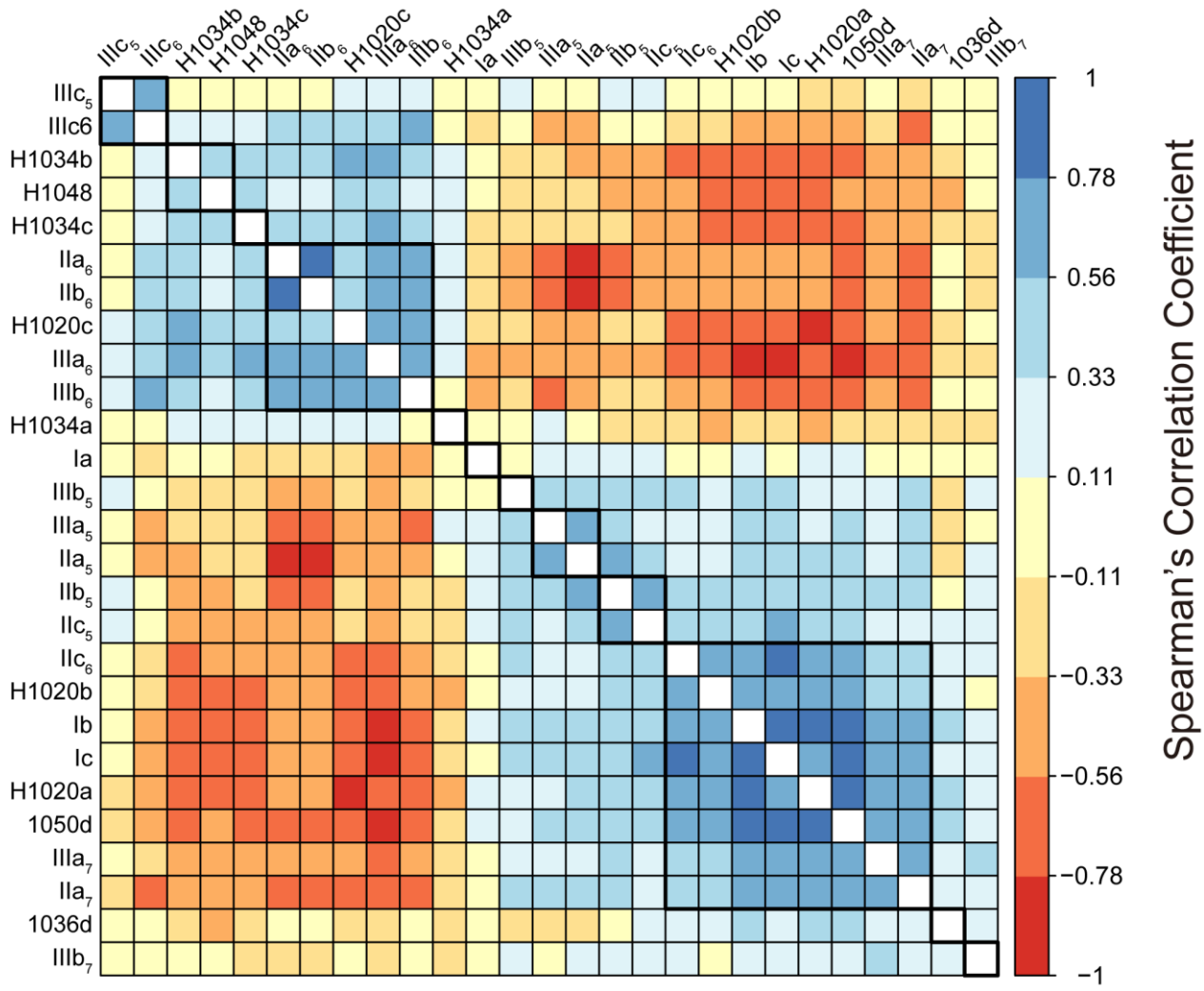

Fig. S11. (a) Correlation plot between fractional abundance of brGDGTs (relative to all brGDGTs) and brGMGTs (relative to all brGMGTs).

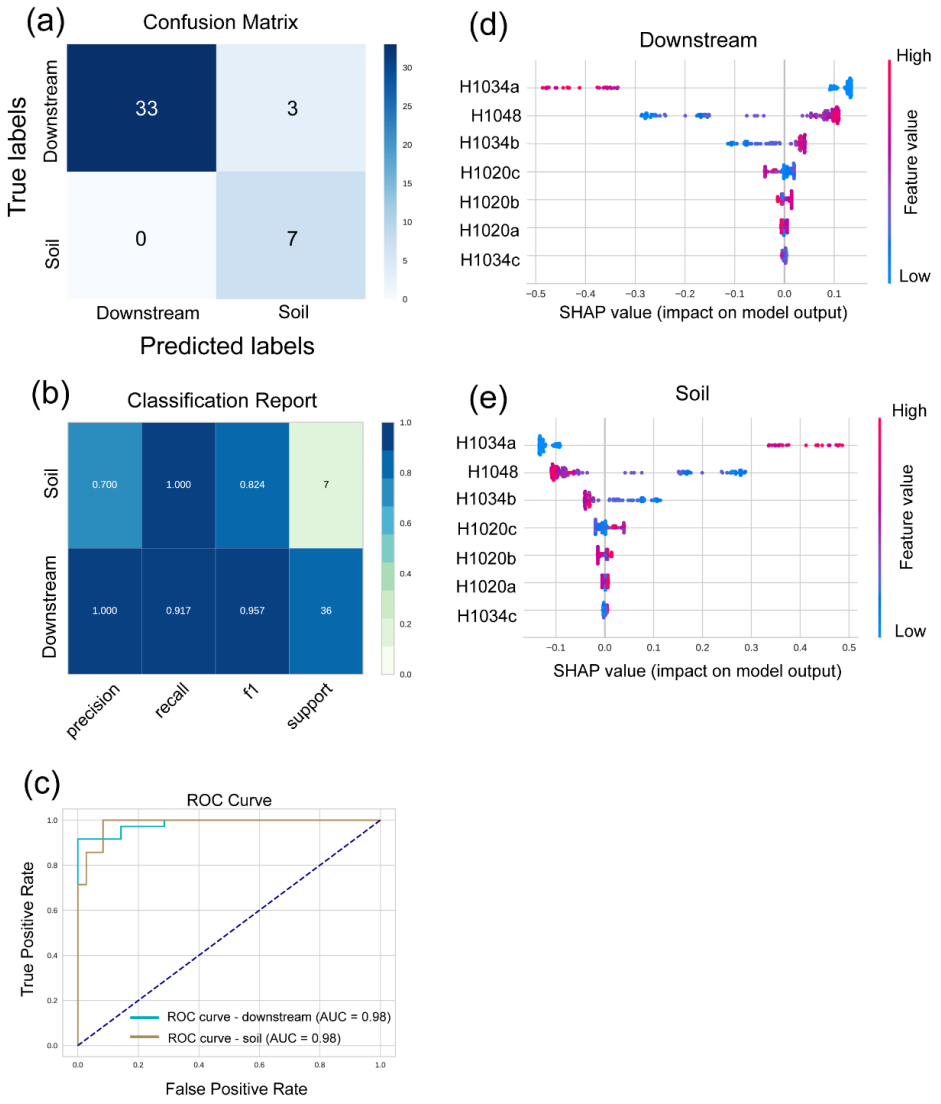

Fig. S12. Evaluation of the random forest model based on brGMGTs through the confusion matrix (a), classification report (b), and receiver operating characteristic (ROC) curve (c). SHAP summary plots (d-e) show the feature importance obtained from the random forest algorithm and the SHAP library. Each bullet represents a single sample within the training set, with the color representing the feature value (fractional abundance of the brGMGTs) ranging from low (blue) to high (pink). The bullets positioned on the right side of the SHAP summary plot correspond to positive SHAP values, indicating a positive effect on the model output (downstream estuary or soils). The bullets on the left side of the plot indicate negative SHAP values, suggesting a negative effect on the model output. The variables (brGDGTs) with higher impact on the model performance are shown at higher positions. The training sets include downstream SPM and sediment samples (d) as well as soils (e).

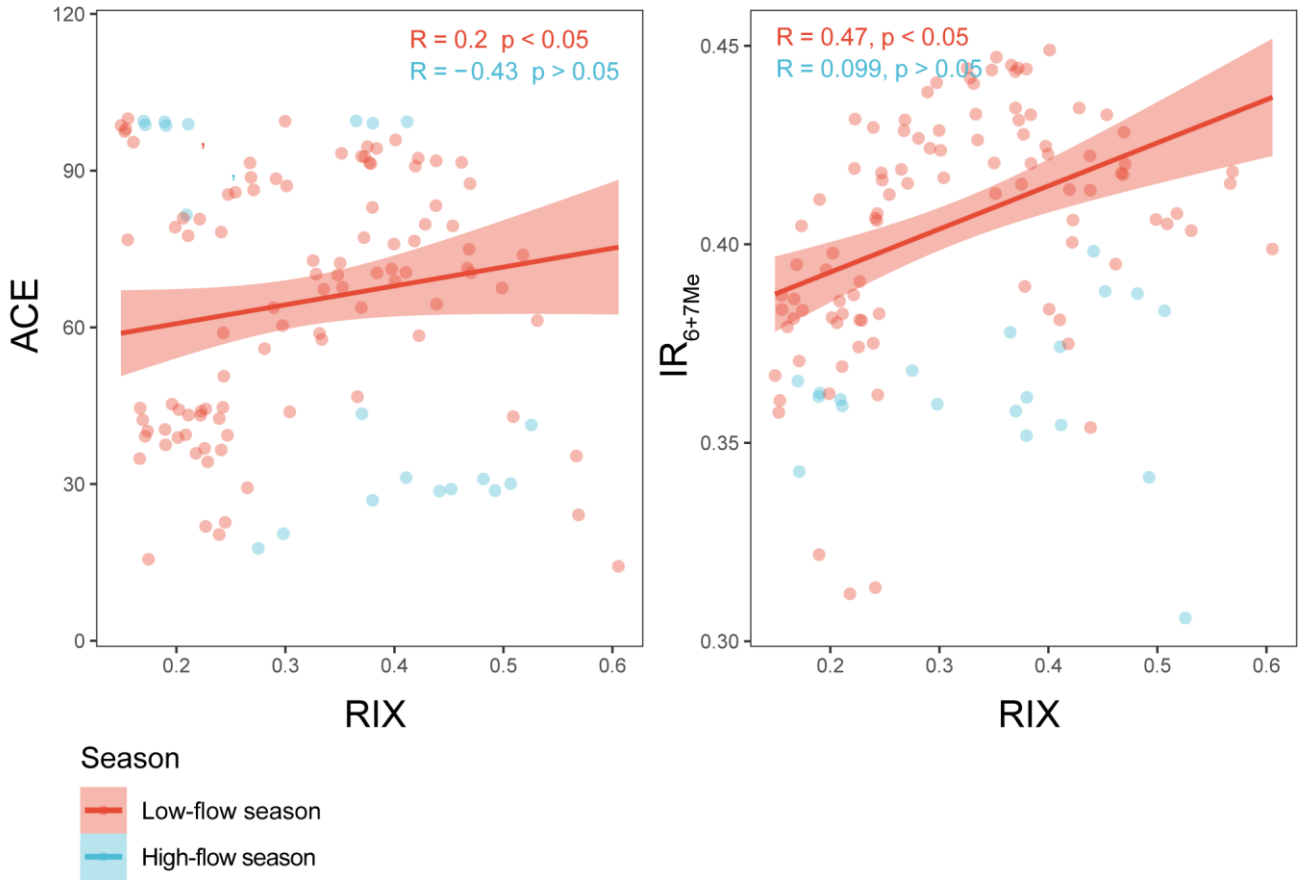

Fig. S13. RIX plotted versus ACE and IR_{6+7Me} through the linear regression. Shaded area represents 95% confidence intervals. Dataset is composed of SPM.

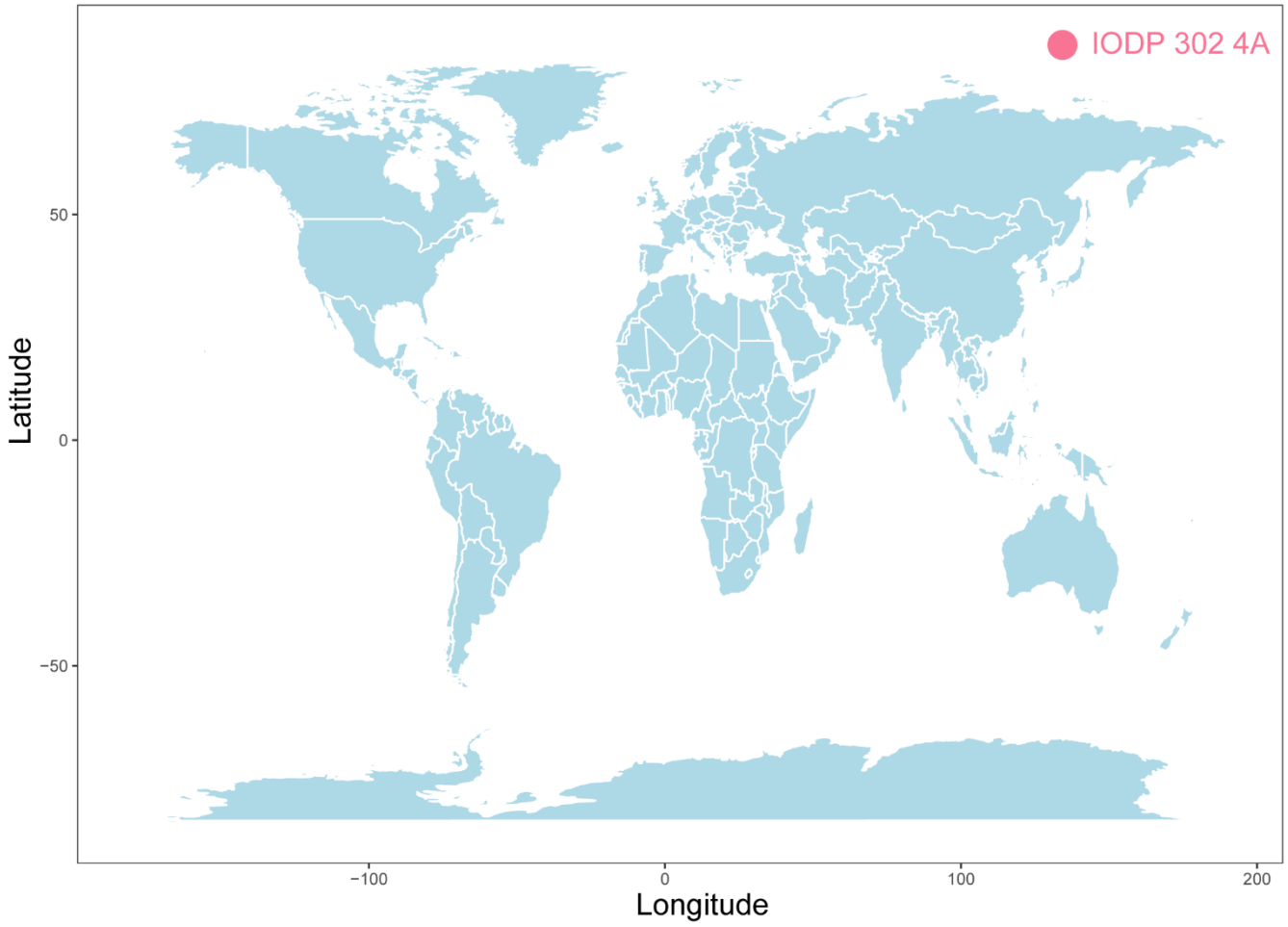

Fig. S14. Location of the IODP Expedition 302 Hole 4A.