This manuscript is a non-peer reviewed preprint submitted to EarthArXiv

Environmental controls on the brGDGT and brGMGT distributions across the Seine River basin (NW France): Implications for bacterial tetraethers as a proxy for riverine runoff

Zhe-Xuan Zhang[1,2], Edith Parlanti[2], Christelle Anquetil[1], Jérôme Morelle[3], Anniet M. Laverman[4], Alexandre Thibault[5], Elisa Bou[6], Arnaud Huguet[1]*

1. Sorbonne Université, CNRS, EPHE, PSL, UMR METIS, Paris, 75005, France
2. Univ. Bordeaux, CNRS, Bordeaux INP, EPOC, UMR 5805, F-33600 Pessac, France
3. Department of Biology and CESAM – Centre for Environmental and Marine Studies, University of Aveiro, Campus de Santiago, Aveiro, 3810-193, Portugal
4. Univ. Rennes 1, CNRS, ECOBIO-UMR 6553, Rennes, 35000, France
5. Antea Group, Innovation Hub, 803 boulevard Duhamel du Monceau, Olivet, 45160, France
6. Université de Toulouse, CNRS, Toulouse INP, Université Toulouse 3 - Paul Sabatier (UPS), Laboratoire Ecologie Fonctionnelle et Environnement, Route de Narbonne 118, Toulouse, 31062, France

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

Abstract. Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are bacterial lipids that have been largely used as environmental proxies in continental paleorecords. Another group of related lipids, branched glycerol monoalkyl glycerol tetraethers (brGMGTs), has recently been proposed as a potential paleotemperature proxy. Nevertheless, the sources and environmental dependencies of both brGDGTs and brGMGTs along the river-sea continuum are still poorly understood, complicating their application as paleoenvironmental proxies in some aquatic settings. In this study, the sources of brGDGTs and brGMGTs and the potential factors controlling their distributions are explored across the Seine River basin (NW France), which encompasses the freshwater to seawater continuum. BrGDGTs and brGMGTs were analyzed in soils, Suspended Particulate Matter (SPM), and sediments (*n*=237) collected along the land-sea continuum of the Seine basin.~~To this aim, brGDGTs and brGMGTs were analyzed in soils, Suspended Particulate Matter (SPM) and sediments (*n*=237) collected all along this basin, from land to sea.~~ Both types of compounds (i.e. brGDGTs and brGMGTs) are shown to be produced *in situ*, in ~~both~~ freshwater and saltwater, based on their high concentrations and distinct distributions in ~~in~~ aquatic settings (SPM and sediments) *vs*. soils. ~~Both types of compounds are shown to be produced *in situ*, in freshwater as well as saltwater.~~ Redundancy analysis further shows that both salinity and nitrogen ~~loadings~~ dominantly control the brGDGT distributions. Furthermore, the relative abundance of 6-methyl *vs*. 5-methyl brGDGTs (IR_{6Me} ratio), Total Nitrogen (TN), $\delta^{15}N$ and chlorophyll *a* concentration co-vary in a specific geographical zone with low salinity, suggesting that 6-methyl brGDGTs are preferentially produced under low-salinity and high-productivity conditions. In contrast with brGDGTs, brGMGT distribution appears to be primarily regulated by salinity, with a distinct influence on the individual homologues. Salinity is positively correlated with homologues H1020a and H1020b, and negatively correlated with compounds H1020c and H1034b in SPM. This suggests that bacteria ~~thriving~~ living in freshwater preferentially produce compounds H1020c and H1034b, whereas bacteria primarily growing in saltwater appear to be predominantly responsible for the production of homologues H1020a and H1020b. Based on the

abundance ratio of the freshwater-derived compounds (H1020c and H1034b) *vs.* saltwater-derived homologues (H1020a and H1020b), a novel proxy, Riverine IndeX (RIX) is proposed to trace riverine organic matter inputs, with high values (>0.5) indicating higher riverine contribution. We successfully applied RIX ~~RIX was then applied~~ to the Godavari River basin (India) and a paleorecord across the upper Paleocene and lower Eocene from the Arctic Coring Expedition at Lomonosov Ridge, showing its potential applicability in both modern samples and in paleorecords.

1 Introduction

Branched glycerol dialkyl glycerol tetraethers (brGDGTs) are membrane lipids produced by ~~unknown~~ bacteria, some of them belonging to the ~~, although some of them were attributed to the~~ phylum *Acidobacteria* (Sinninghe Damsté et al., 2011; Chen et al., 2022; Halamka et al., 2023). These compounds were observed to occur ubiquitously in a wide range of terrestrial and aquatic environments (Schouten et al., 2013; Raberg et al., 2022). The distribution of brGDGTs (number of cyclopentane moieties and methyl groups; cf. structures in Fig. S1) ~~was empirically~~has been linked with pH and Mean Annual Air Temperature (MAAT) in soils (Weijers et al., 2007; De Jonge et al., 2014; Véquaud et al., 2022), peats (Naafs et al., 2017; Véquaud et al., 2022) and lake sediments (Tierney et al., 2010; Russell et al., 2018; Martínez-Sosa et al., 2021). The brGDGT-based proxies (i.e. MBT'$_{5ME}$ and CBT') have been largely applied to reconstruct paleoclimate~~MAAT and pH~~ from sedimentary archives (Coffinet et al., 2018; Harning et al., 2020; Wang et al., 2020).

In aquatic settings, brGDGTs were initially suggested to be transported by erosion to the sediments~~predominantly derived from watershed soils and transported by erosion in the sediments~~ (Hopmans et al., 2004). Based on this assumption, the Branched and Isoprenoid Tetraethers (BIT) index was defined as the abundance ratio of the major brGDGTs to crenarchaeol (isoprenoid GDGT mainly produced by marine *Nitrososphaerota*). The BIT index ranges between 0 and 1, with high BIT values (around 1) reflecting a higher contribution of terrestrial organic matter compared to marine organic matter~~It is comprised between 0 and 1, with high BIT values (around 1) reflecting higher contribution of terrestrial organic matter compared to marine organic matter~~ (Hopmans et al., 2004).

Over the last years, the BIT index has been broadly used to~~for~~ quantify~~ing~~ the relative contribution of terrestrial organic matter in aquatic systems (Xu et al., 2020; Yedema et al., 2023) and to evaluat~~ing~~e the reliability of the TEX_{86} palaeothermometer (Cramwinckel et al., 2018). However, several studies have shown that brGDGTs can also be produced *in situ* in aquatic settings, ~~either in~~including rivers (e.g. De Jonge et al., 2015; Freymond et al., 2017; Kim et al., 2015; Zell et al., 2014, 2013), lakes (Tierney and Russell, 2009), ~~or~~ and marine settings (Dearing Crampton-Flood et al., 2019; Zeng et al., 2023). This ~~,~~ add~~sing complication~~ complexity ~~for~~ to the identification of brGDGT sources in aquatic ecosystems and to ~~for~~ the application of ~~the~~ brGDGTs as (paleo)environmental proxies, including the BIT index.

The BIT values have all the more to be carefully interpreted, especially considering the~~ir~~ potential influence ~~from~~of the selective degradation~~as they could also be influenced~~ ~~by the selective degradation~~ of branched *vs*. isoprenoid GDGTs (Smith et al., 2012). Thus, complementary molecular proxies for quantifying the input of terrestrial organic matter to aquatic settings are still needed. These proxies~~, which~~ may cross-validate other available terrestrial proxies, such as the $\delta^{13}C$ of organic carbon (Lamb et al., 2006), heterocyst glycolipids (Kang et al., 2023), and long-chain diols (Lattaud et al., 2017). Recently, a machine-learning approach (BIGMaC model) was proposed to infer the origin of environmental samples (e.g. soil, peat, marine and lake settings) based on their GDGT distribution (Martínez-Sosa et al., 2023). ~~Recently, a machine learning approach was proposed to classify GDGT sources in various environments, including soil, peat, marine, and lake (Martínez-Sosa et al., 2023).~~ While such an approach shows potential for differentiating distinct sources of GDGTs, its application to aquatic systems has not yet been extensively explored.~~While this approach shows potential for differentiating distinct sources of GDGTs, its application in aquatic systems has not yet been extensively explored.~~

The improvement of chromatographic methods allowed the separation and quantification of 5-, 6- and 7-methyl brGDGTs (methyl groups at the fifth, sixth, and seventh positions; Fig. S1) that previously co-eluted ~~The improvement of chromatographyanalytical methods allowed the separation and quantification of 5-, 6- and 7-methyl brGDGTs (methyl groups at the fifth, sixth, and seventh positions; Fig. S1), that in previous chromatographic protocols co-eluted~~ (De Jonge et al., 2013, 2014; Ding et al., 2016). This led to the development of new brGDGT-based proxies based on these specific brGDGT isomers~~This results in more precise quantifications of the isomers of brGDGTs, thereby significantly influencing the brGDGT-based proxies~~ (De Jonge et al., 2014). Compounds eluting later than 7-methyl brGDGTs are tentatively designated 1050d and 1036d, as their exact chemical structures are currently unknown (Wang et al., 2021). The fractional abundance of the individual brGDGT isomers was shown to be influenced by distinct environmental factors. For example, the relative abundance of 5-methyl brGDGTs was correlated with temperature, whereas that of 6-methyl brGDGTs was correlated with pH (De Jonge et al., 2014). In addition~~ to temperature and pH, other environmental factors may influence brGDGT distributions in terrestrial and aquatic settings and hence the application and interpretation of brGDGT-derived proxies. For example,~~ , recent studies in lakes observed an influence of salinity on the relative abundance of 6-methyl, 7-methyl brGDGTs and their late-eluting compounds (Wang et al., 2021; Kou et al., 2022). This suggests that salinity could also control the distribution of these compounds in other systems like river-sea continuums but this assumption has not yet been studied.

Compared with brGDGTs, the branched glycerol monoalkyl glycerol tetraethers (brGMGTs, also referred as H-brGDGTs) are a much less studied group of lipids.~~Compared with brGDGTs, the branched glycerol monoalkyl glycerol tetraethers (brGMGTs or H-brGDGTs in some previous studies) are a much less studied group of lipids.~~ Recent studies have revealed their presence in diverse environments, including peatlands (Naafs et al., 2018; Tang et al., 2021), marine settings (Liu et al., 2012; Xie et al., 2014), rivers (Kirkels et al., 2022a), soils (Baxter et al., 2021; Kirkels et al., 2022a) and lakes (Baxter et al., 2019, 2021).

BrGMGTs are labelled as H1020, H1034, and H1048 respectively (cf. in Fig. S1), with isomers suggested by a suffix letter (a-c) following the order in which they elute according to Baxter et al. (2019). These compounds are structurally similar to brGDGTs, but possess an additional covalent carbon–carbon bond between the alkyl chains, leading to "H-shaped" structure. The bridge of brGMGTs was considered to be a primary adaptation to heat stress (Naafs et al., 2018; Baxter et al., 2019). Their presumed membrane stability under high temperature conditions was inferred from the behaviour of isoprenoid glycerol monoalkyl glycerol tetraethers (isoGMGTs), which were identified in a hyperthermophilic methanogen (Morii et al., 1998) and deep-sea hydrothermal vents (Schouten et al., 2008). Although a rigorous chemical characterization of brGMGTs is lacking and the source organisms of brGMGTs are unknown, correlations between the relative abundances of brGMGTs and MAAT were observed in peat soils (Naafs et al., 2018) and lakes (Baxter et al., 2019), showing their potential as temperature proxies. In addition to temperature, anoxic conditions may also trigger brGMGT production in the anoxic zone of peats (Naafs et al., 2018; Tang et al., 2021), anoxic part of the water column and/or sediments in lakes (Baxter et al., 2021), in regularly inundated soils (Kirkels et al., 2022a), as well as in the oxygen minimum zone in the marine environments (Xie et al., 2014). Furthermore, shifts in microbial community composition in response to other unknown environmental factors seem to control the relative abundances of brGMGTs in peats and lignites (Elling et al., 2023). Henceforth, in order to use the brGMGTs as environmental proxies in sedimentary records, it is still necessary to understand which factors control their distributions in riverine and marine water columns and sediments. This ~~, which~~ remains to date poorly understood (Sluijs et al., 2020; Bijl et al., 2021).

Based on previous studies of brGDGTs and brGMGTs in terrestrial and marine settings (Dearing Crampton-Flood et al., 2019; Wang et al., 2021; Kirkels et al., 2022a, 2022b; Kou et al., 2022), we hypothesize (1) that both brGDGTs and brGMGTs can be produced *in situ* in aquatic systems and (2) that brGDGT and brGMGT distribution are influenced by surrounding environmental factors and vary spatially along the land-sea continuum. These compounds have a potential to be used as proxies of riverine organic matter inputs along estuaries. These hypotheses were tested by examining and comparing the distribution of brGDGTs and brGMGTs in soils, suspended particulate matter (SPM) and sediments ($n = 237$) collected all along the Seine River basin (NW France), covering its riverine and estuarine parts. The aim of the present study was (1) to investigate the sources of brGDGTs and brGMGTs along the Seine land-sea continuum, (2) to determine the predominant environmental controls affecting the distribution of these molecules and (3) to assess the potential of brGMGTs as a riverine runoff proxy.

2 Material and methods

2.1 Study area

The Seine River basin (Seine River and its estuary; Fig. 1a) is more than 760km long ~~and is characterized by high population density~~, draining through the greater Paris region (over 12 million inhabitants) to the English Channel (Flipo et al., 2021). The Seine Estuary is a macrotidal ~~(large tidal range)~~ estuary according to its ~~high~~ large tidal range, small depth and morphology.

The maximum flows are generally observed in winter (over 700 m^3/s; Fig. 1b), whereas the minimum flows are observed in summer (below 250 m^3/s; Fig. 1b). The tide influences the estuary up to the city of Poses (site 5, KP 202 in Fig. 1a; KP represents kilometric point and is defined as the distance in kilometers from the city of Paris), where a dam constitutes the boundary between the river and the estuary. Based on spatiotemporal variations of salinity, the estuary can be divided into two major parts. The upstream estuary corresponds to the freshwater tidal sector (KP 202 to KP 298, from site 5 to site 12; Fig. 1a and Table 1) and the downstream estuary is influenced by a salinity gradient (starting at KP 298, from site 12 to the coastal area; Fig. 1a and Table 1) (Romero et al., 2016; Druine et al., 2018). ~~The tide influences the estuary up to the city of Poses (site 5, KP 202 in Fig. 1a; KP represents kilometric point and is defined as the distance in kilometers from the city of Paris), where a dam constitutes the boundary between the river and the estuary. The estuary can be divided into two major parts: the upper section mainly influenced by freshwater (KP 202 to KP 298, from site 5 to site 12; Fig. 1a and Table 1) and the lower section (starting at KP 298, from site 12 to the coastal area; Fig. 1a and Table 1).~~

Table 1. Location of the sampling sites along the Seine Basin, with the type of samples collected

Site	Name	Longitude (°)	Latitude (°)	KP	Zone	Date	Type
1	Marnay sur Seine	3.56	48.51	-200	River	2020-11	Sub-surface SPM (*n*=1)
2	Bougival	2.13	48.87	40	River	2020-11	Sub-surface SPM (*n*=1)
3	Triel sur Seine	2.00	48.98	80	River	2020-11	Sub-surface SPM (*n*=1)
4	Les Andelys	1.40	49.24	175	River	2019-6; 2019-7; 2020-9	Sub-surface and bottom SPM (*n*=6)
5	Poses	1.24	49.31	202	Upstream estuary	2016-4; 2020-11	Sub-surface SPM (*n*=2)
6	Oissel	1.10	49.34	229.4	Upstream estuary	2019-6; 2019-7; 2020-9	Sub-surface and bottom SPM (*n*=18)

7	Rouen	1.03	49.43	243	Upstream estuary	2016-4	Sub-surface SPM (*n*=1); Sediments (*n*=10)
8	Petit Couronne	1.01	49.38	251.3	Upstream estuary	2020-9; 2021-2; 2021-3	Sub-surface SPM (*n*=3)
9	Haulot Sur Seine	0.98	49.36	255.6	Upstream estuary	2019-6	Sub-surface SPM (*n*=1)
10	Val des Leux	0.92	49.40	265.55	Upstream estuary	2019-6; 2019-7; 2020-9	Sub-surface and bottom SPM (*n*=18)
11	Duclair	0.87	49.48	278	Upstream estuary	2020-9; 2021-2; 2021-3	Sub-surface SPM (*n*=3)
12	Heurtauville	0.82	49.45	297.65	Downstream estuary	2019-6	Sub-surface SPM (*n*=1)
13	Caudebec	0.75	49.52	310.5	Downstream estuary	2015-4; 2015-9; 2016-4; 2019-6; 2019-7; 2020-9; 2021-2; 2021-3	Sub-surface and bottom SPM (*n*=24)
14	Vatteville-La-Rue	0.67	49.47	318	Downstream estuary	2019-6	Sub-surface SPM (*n*=1)
15	Tancarville	0.47	49.47	337	Downstream estuary	2015-1; 2015-4; 2015-9; 2019-6; 2019-7; 2020-9; 2021-2; 2021-3	Sub-surface and bottom SPM (*n*=24); Sediments (*n*=20)
16	Berville-Sur-Mer	0.37	49.44	346	Downstream estuary	2019-6	Sub-surface SPM (*n*=1)
17	Fatouville	0.32	49.44	350	Downstream estuary	2015-4; 2015-7; 2015-9; 2016-4	Sub-surface SPM (*n*=4); Sediments (*n*=28)
18	Honfleur	0.23	49.43	355.8	Downstream estuary	2015-4; 2015-9; 2019-6; 2020-9; 2021-2; 2021-3	Sub-surface SPM (*n*=6)

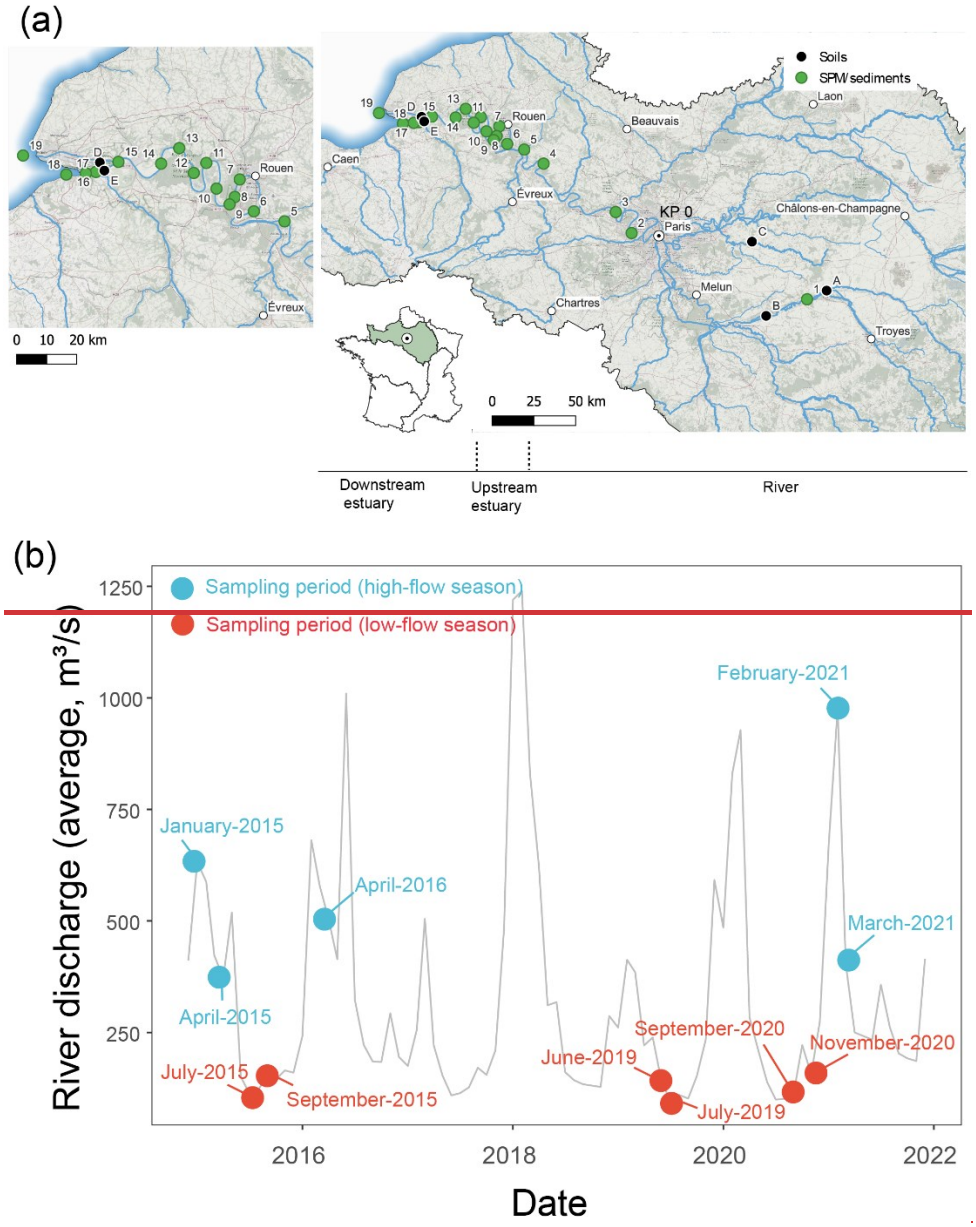
19	La Carosse	0.03	49.48	370	Downstream estuary	2015-7; 2016-4; 2016-4	Sub-surface SPM (*n*=2); Sediments (*n*=10)
A	n.a.	3.72	48.56	n.a.	Soil (around the river)	2021-9	Soil (*n*=1)
B	n.a.	3.23 3.26	48.43 48.42	n.a.	Soil (around the river)	2021-9	Soil (*n*=5)
C	n.a.	3.11 3.13	48.83 48.85	n.a.	Soil (around the river)	2021-10	Soil (*n*=3)
D	n.a.	0.38 0.38 0.38	49.47 49.46 49.45	n.a.	Soil (around the downstream estuary)	2021-3; 2021-9	Soil (*n*=8)
E	n.a.	0.41 0.41	49.44 49.45	n.a.	Soil (around the downstream estuary)	2018-2; 2018-6; 2018-8; 2018-10; 2020-9; 2021-3	Soil (*n*=34)

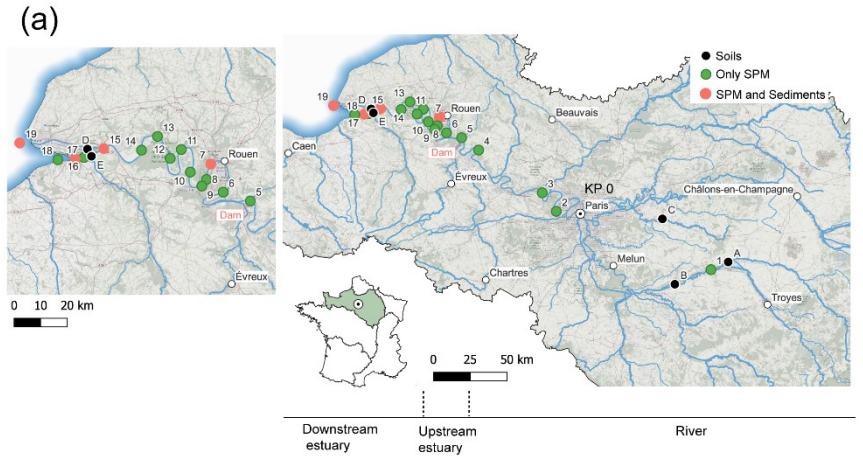
2.2 Sampling

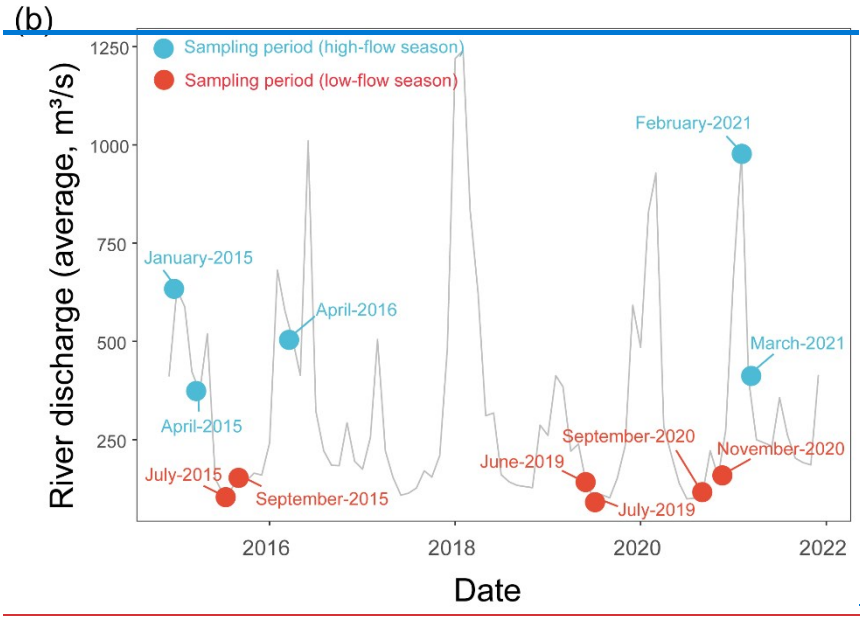
From June 2019 to March 2021, water samples (*n*=102) were collected across the Seine River (Fig. 1a). Sub-surface water (ca. 1m depth) samples were collected in high-flow (over 250 m^3/s) and low-flow (below 250 m^3/s) periods from the three zones (river, upstream estuary and downstream estuary) of the Seine River basin (Table 1). At 5 sites (sites 4, 6, 10, 13, and 15, Fig. 1a and Table 1), both sub-surface and bottom water (2.2-16 m depth) samples were retrieved using a pump into precleaned 20L FLPE Nalgene carboys. Estuarine water samples (sites 6, 10, 13, and 15; Fig. 1a and Table 1) were collected at three tide periods (high tide, low tide and mid tide). For these sites, 0.25-43 L of water were immediately filtered using pre-combusted Whatman GF/F 0.7 µm glass fiber filters. After filtration, filters were freeze-dried, scratched and stored frozen at -20°C prior to analysis.

Additional SPM samples (*n*=16; Table 1) used in this study for brGDGT and brGMGT analysis were collected from the upstream and downstream estuary (site 5, 7, 13, 15, 17, 18, and 19; Fig. 1a and Table 1) in 2015 and 2016, as detailed by Thibault et al. (2019). Sediments (*n*=68) from ~~8~~ 7 cores (10-cm depth~~10 cm depth~~) were collected in the river channel at the

same sites as these SPM samples in 2015 and 2016 using a UWITEC corer as described by Thibault et al. (2019) (Table 1). These sediments were further sliced (1-cm thickness) and freeze-dried. For each core, ten samples were analyzed for brGDGTs and brGMGTs, except for the one collected at site 17 in April 2016, where no lipids were detected between 4-5 and 5-6 cm depth.~~Eight samples are available for biomarker analysis in the core collected from site 17 in April 2016, while each of the other six cores has ten samples available.~~ Surficial soils (*n*=9) were collected in the lateral area of the upstream section of the ~~Seine river~~Seine River in 2021 (site A, B, and C, Fig. 1a and Table 1) and freeze-dried. Additional wetland soils and mudflat sediments (*n*=42) were collected in the downstream estuary in 2018, 2020, and 2021 (site D and E, Fig. 1a and Table 1), representing allochthonous material transported into the estuary by tidal effect. These samples were collected at low tide using a plexiglass® core (4.5 cm depth), and back to the laboratory, homogenized, freeze-dried, and ground using a ball mill (model MM400, Retsch®).


(a)
Soils
SPM/sediments
Rouen
Évreux
Caen
Beauvais
Laon
KP 0
Paris
Châlons-en-Champagne
Melun
Chartres
Troyes
0 10 20 km
0 25 50 km
Downstream estuary
Upstream estuary
River
(b)
Sampling period (high-flow season)
Sampling period (low-flow season)
River discharge (average, m³/s)
Date
January-2015
April-2015
July-2015
September-2015
April-2016
June-2019
July-2019
September-2020
November-2020
February-2021
March-2021


(a)
Soils
Only SPM
SPM and Sediments
Rouen
Dam
Évreux
Caen
Beauvais
KP 0
Paris
Châlons-en-Champagne
Melun
Chartres
Troyes
0 10 20 km
0 25 50 km
Downstream estuary
Upstream estuary
River


(b)
Sampling period (high-flow season)
Sampling period (low-flow season)
River discharge (average, m³/s)
Date
January-2015
April-2015
July-2015
September-2015
April-2016
June-2019
July-2019
September-2020
November-2020
February-2021
March-2021

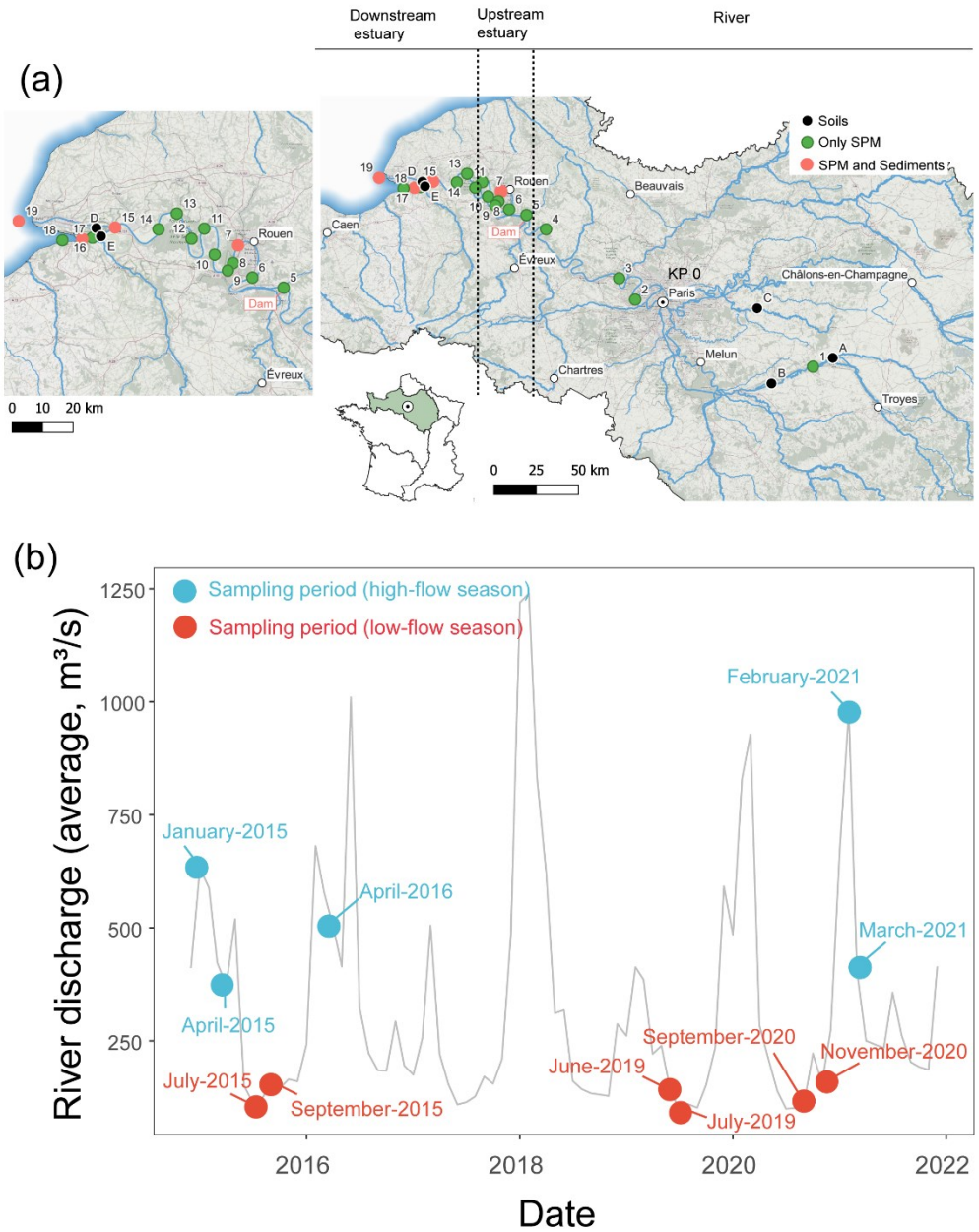


Figure 1: (a) Geographical locations of sampling sites in the Seine River Basin (KP: kilometric point, the distance in kilometers from the city of Paris (KP 0)). The sampling sites from upstream estuary and downstream estuary are shown in the zoom-in figure. Sub-surface SPM was collected for all sites from site 1 to site 18, while both sub-surface and bottom SPM were collected at sites 4, 6, 10, 13, and 15. (b) Mean monthly water discharge for the Seine River at the Paris Austerlitz station from 2015 to 2021 (data from https://www.hydro.eaufrance.fr/). Bullets represent the sampling period in high-flow (>250 m³/s - blue) and low-flow (<250 m³/s - red) seasons.

2.3 Elemental and isotopic analyses

Elemental and isotopic analyses of the soils (surficial soils and mudflat sediments, *n*=51) and SPM (*n*=102) collected from 2018 to 2021 were performed following the method described in Thibault et al. (2019). The samples were split, and one aliquot was decarbonated. Briefly, 40 mg of SPM and 1 g of soils/sediments samples were firstly decarbonated by adding 10 mL of 3 M HCl for 2 h with magnetic stirring at room temperature. Subsequently, these samples were rinsed using ultrapure water and centrifuged until reaching neutral pH. The obtained decarbonated samples were stored at −20 °C and freeze dried. Both decarbonated and non-decarbonated samples (~6 mg for SPM and ~20 mg for soils) were enclosed in a tin capsule. Total Organic Carbon content (TOC) and stable carbon isotopic composition ($\delta^{13}C$) were measured in decarbonated samples using an elemental analyzer coupled with an isotope ratio mass spectrometer (Thermo Fisher Scientific Delta V Advantage) at the ALYSES platform (Sorbonne University / IRD, Bondy, France). Total Nitrogen (TN) and nitrogen isotope ($\delta^{15}N$) were measured in non-decarbonated samples as acidification could impact the N contents (Ryba and Burgess, 2002). The isotopic composition ($\delta^{13}C$ or $\delta^{15}N$) was expressed as relative difference between isotopic ratios in samples and in standards (Vienna Pee Dee Belemnite for carbon ~~or~~ and atmospheric N_2 for nitrogen). Additional elemental and isotopic ~~analyses~~ data ~~for~~based on~~of~~ SPM and sediments collected in 2015 and 2016 (*n*=84) were obtained ~~carried out as described in~~from Thibault et al. (2019).

2.4 Lipid extraction and analyses

The lipids from surficial soils and mudflat sediments (4~~–~~–20g, *n*=51) and from SPM samples (~150mg, *n*=102) were extracted ultrasonically (3×) with 20 to 40 mL of dichloromethane (DCM): methanol (MeOH) (5/1, v/v) per extraction. Lipids from the SPM and sediments samples (*n*=84) collected in 2015 and 2016 were previously extracted by Thibault (2018) following the same method. The total lipid extracts were then separated into fractions of increasing polarity on an activated silica gel column (ca. 10 mg), using (i) 30 mL of heptane, (ii) 30 mL of heptane:DCM (1/4, v/v), and (iii) 30 mL of DCM/MeOH (1/1, v/v) as eluents. An aliquot (30%) of the third (polar) fraction containing GDGTs and GMGTs was dried, re-dissolved in heptane, and passed through a 0.2µm polytetrafluoroethylene (PTFE) filter (Ultrafree-MC; Merck). C_{46} Glycerol Trialkyl Glycerol Tetraether (GTGT) was used as an internal standard (Huguet et al., 2006). 5 µl of this standard (0.01025 mg/mL) was typically added to 45 µl of sample.

GDGTs and GMGTs were analyzed using a Shimadzu LCMS 2020 high pressure liquid chromatography coupled with mass spectrometry with an atmospheric pressure chemical ionization source (HPLC-APCI-MS) in selected ion monitoring mode, modified from Hopmans et al. (2016) and Huguet et al. (2019). Tetraether lipids were separated with two silica columns in tandem (BEH HILIC columns, 2.1 × 150 mm, 1.7 µm; Waters) thermostated at 30°C. Injection volume was 30 µL. The flow rate was set at 0.2 mL/min. GDGTs and GMGTs were eluted isocratically for 25 min with 82% A/18% B (A= hexane, B=hexane/isopropanol 9/1, v/v), followed by a linear gradient to 65% A/35% B in 25 min, then a linear gradient to 100% B in

30 min, and back to 82% A/18% B in 4 min, maintained for 50 min. Identification of the different brGMGT isomers was achieved by comparison of peak retention time with that of known brGMGTs in Baxter et al. (2019) and Kirkels et al. (2022a). Semi-quantification of ~~br~~GDGTs and brGMGTs was performed by comparing the integrated signal of the respective compound with the signal of a C_{46} synthesized internal standard (Huguet et al., 2006) assuming their response factors to be identical.

The detection limit was set at a signal-to-noise ratio (SNR) of 3. Peaks with lower SNR (<3) are not distinguishable from the background noise and are considered below the limit of quantification.

2.5 Calculation of GDGT proxies

The IR_{6Me} index represents the proportion of 6-methyl brGDGTs *vs*. 5-methyl brGDGTs and was calculated according to De Jonge et al. (2015; Eq. 1) with Roman numbers referring to the structures in annex (Fig. S1):

$$IR_{6Me} = \frac{II_{a_6} + II_{b_6} + II_{c_6} + III_{a_6} + III_{b_6} + III_{c_6}}{II_{a_5} + II_{b_5} + II_{c_5} + II_{a_6} + II_{b_6} + II_{c_6} + III_{a_5} + III_{b_5} + III_{c_5} + III_{a_6} + III_{b_6} + III_{c_6}} \quad (1)$$

The IR_{7Me} index represents the proportion of 7-methyl brGDGTs and late-eluting isomers following Wang et al. (2021; Eq. 2):

$$IR_{7Me} = \frac{III_{a_7} + 1050d + II_{a_7} + 1036d}{III_{a_5} + III_{a_6} + III_{a_7} + 1050d + II_{a_5} + II_{a_6} + II_{a_7} + 1036d} \quad (2)$$

The IR_{6+7Me} index represents the average value between IR_{6Me} and IR_{7Me} according to Wang et al. (2021; Eq. 3):

$$IR_{6+7Me} = \frac{IR_{6Me} + IR_{7Me}}{2} \quad (3)$$

The ACE index was calculated as follows (Turich and Freeman, 2011; Eq. 4):

$$ACE = \frac{Archaeol}{Archaeol + GDGT{-}0} \times 100 \quad (4)$$

The BIT index including the 6-methyl brGDGTs was calculated following De Jonge et al. (2015; Eq. ~~2~~5):

$$BIT = \frac{I_a + II_{a_5} + II_{a_6} + III_{a_5} + III_{a_6}}{I_a + II_{a_5} + II_{a_6} + III_{a_5} + III_{a_5} + crenarchaeol}$$

(~~2~~5)

Based on replicate injections of 3 different samples, the averaged standard deviations were 0.004 for IR_{6Me}, 0.005 for IR_{7Me}, 0.003 for IR_{6+7Me}, 8.5364 for ACE, and 0.032 for BIT.~~Based on replicate injections (n=6), the average analytical error was 0.005 for IR_{6Me} and 0.06 for BIT.~~

2.6 Water quality measurements

Water turbidity was measured by a CTD Probe Sea-bird®. Water temperature, dissolved oxygen, salinity, and pH were measured using an automated YSI 6000 multi-parameter probe (YSI inc., Yellow springs, OH, USA). Chlorophyll *a* (Chl *a*) concentrations were measured on water samples after filtration on Whatman GF/F 0.7 µm glass fiber filters, which were stored frozen (-20° C) before analysis. Chl *a* was extracted from filters with incubation in 10 ml of 90% acetone for 12 hours in the dark at 4°C. After two centrifugations (1700 g, 5 min), Chl *a* concentrations were measured using a Turner Designs Fluorometer according to the method of Strickland and Parsons (1972) as described in the reference protocol of SNO SOMLIT (Service d'observation du Milieu Littoral). Water quality measurements were performed at the Laboratoire Ecologie Fonctionnelle et Environnement (Université de Toulouse) as well as at UMR BOREA (Université de Caen Normandie).

2.7 Statistical analyses

All statistical analyses were performed using the R software (version 4.2.1). The non-parametric statistical tests were used due to the non-normal distribution of the dataset (tested by Shapiro–Wilk normality test; p-values < 0.05). Specifically, the Spearman's correlation was used to investigate potential correlations among different features (environmental parameters, fractional abundances of brGDGTs and brGMGTs, and proxies derived from these compounds), and the unpaired two-samples Wilcoxon test (also known as Mann-Whitney test or Wilcoxon rank sum test) was used for two independent group comparisons. Significance level is indicated by asterisks: *p-value < 0.05; **p-value < 0.01; ***p-value < 0.001; ****p-value < 0.0001; ns (not significant), p-value > 0.05.

A Principal Component Analysis (PCA) was performed on the fractional abundances of brGDGTs and brGMGTs, using the R packages factoextra and FactoMineR. The different groups of samples were highlighted by adding 95% concentration ellipses. The proportion of variance in brGDGT and brGMGT compositions that can be explained by different groups was evaluated by permutational multivariate analysis of variance using distance matrices (adonis) in the adonis2 function of the R package v~~V~~egan, using the Bray-Curtis distances and 999 permutations.

A Redundancy analysis (RDA) was performed using the R package vegan to investigate the relationship between environmental parameters and brGDGT or brGMGT distributions in SPM. Angles between brGDGTs or brGMGTs and environmental factors were used to identify the potential ~~correlations~~relationships. Right angles (90°) reflect a lack of linear correlations, whereas small or straight angles (close to 0° or 180°, respectively) imply positive or negative linear correlations.

The compounds that are close to each other were assumed to be strongly linked, representing similar distribution patterns and comparable responses to the environmental conditions. To evaluate the relative importance of each explanatory variable (environmental parameters) on brGDGT or brGMGT distributions, a hierarchical partitioning method implemented in the R package rdacca.hp was used. Briefly, this approach suggests that shared variance can be decomposed into equal components based on the number of involved predictors (environmental factors), allowing for the estimation of the relative importance of each predictor by adding its partial R^2 to the sum of all allocated average shared R^2. While most selection procedures, such as forward selection, use predictor ordering to assess variable importance, hierarchical partitioning calculates individual importance (the sum of unique and total average shared effects) from all subset models, generating an unordered assessment of variable importance ~~This method calculated the individual importance (sum of the unique and total average shared effects) from all subset models, generating an unordered assessment of variable importance~~ (Lai et al., 2022).

Spatio-temporal variations of environmental factors and proxies derived from brGDGTs and brGMGTs were assessed after applying a locally estimated scatterplot smoothing (LOESS) method. This method ~~allows the identification of~~identifies nonlinear data patterns and buffers the effect of aberrant data and outliers. LOESS was implemented by the geom_smooth function of the R package ggplot2.

2.8 Machine learning

The BigMac model, developed by Martínez-Sosa et al. (2023) based on brGDGT and isoGDGT distribution, was applied. Subsequently, using the same algorithm (random forest), we developed our own model based on either brGDGTs or brGMGTs.

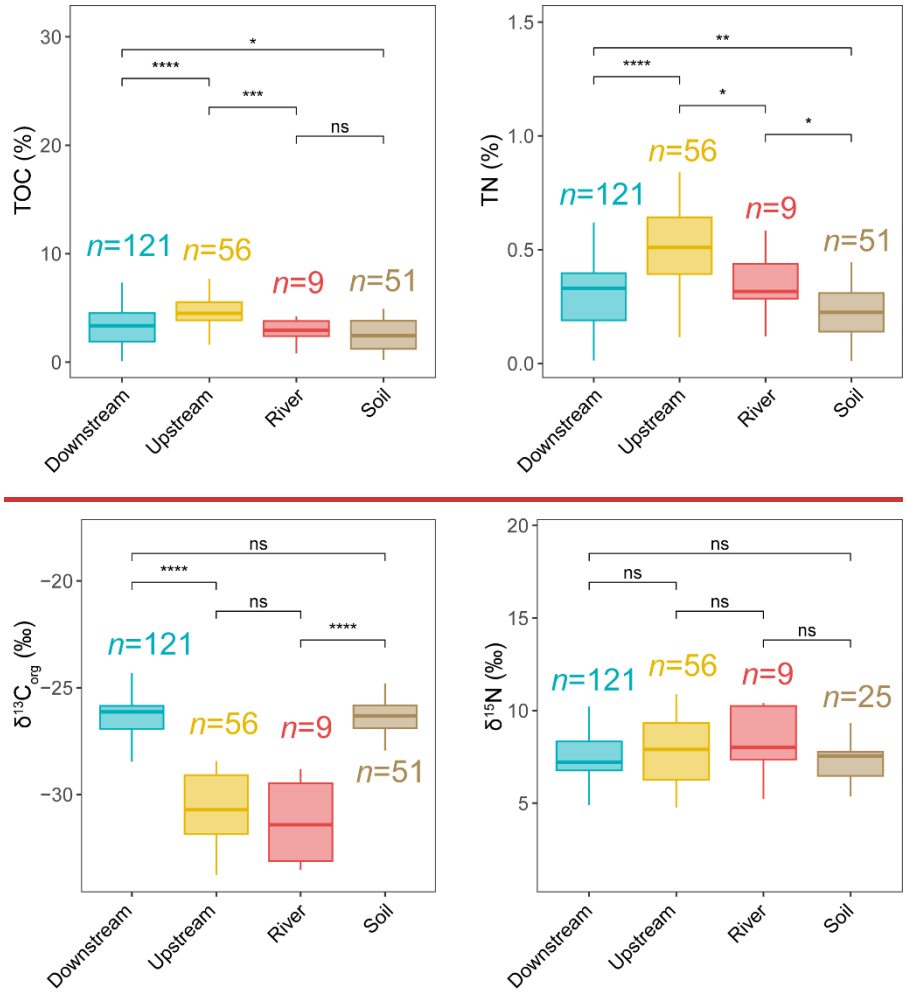
For independent models, ~~O~~our lipid dataset was split into a training set (75%) and a test set (25%). We then used a supervised machine-learning algorithm (random forest) to train models. This algorithm was applied to classify the downstream estuary and soil samples based on brGDGTs or brGMGTs as input, implemented using the scikit-learn library (https://github.com/scikit-learn/) (Pedregosa et al., 2011) in Python (version 3.10.12). Hyperparameter tuning was conducted using a randomized search approach implemented through the RandomizedSearchCV function in scikit-learn.

SHapley Additive exPlanations (SHAP) is a game-theoretical method used to interpret machine learning models (Lundberg et al., 2020). SHAP analysis was applied to identify which compounds were important for the classifications, implemented by the SHAP library in Python. A higher SHAP value indicates a more substantial contribution of the feature (brGDGTs or brGMGTs) to the predicted outcome (downstream estuary or soils).

~~(Pedregosa et al., 2011)(Lundberg et al., 2020)~~

3 Results

3.1 Distribution of bulk parameters from land to sea


TOC (%)
30
20
10
0
*
****
***
ns
n=121
n=56
n=9
n=51
Downstream
Upstream
River
Soil
TN (%)
1.5
1.0
0.5
0.0
**
****
*
*
n=121
n=56
n=9
n=51
Downstream
Upstream
River
Soil
δ13Corg (‰)
−20
−25
−30
ns
****
ns
****
n=121
n=56
n=9
n=51
Downstream
Upstream
River
Soil
δ15N (‰)
20
15
10
5
ns
ns
ns
ns
n=121
n=56
n=9
n=25
Downstream
Upstream
River
Soil

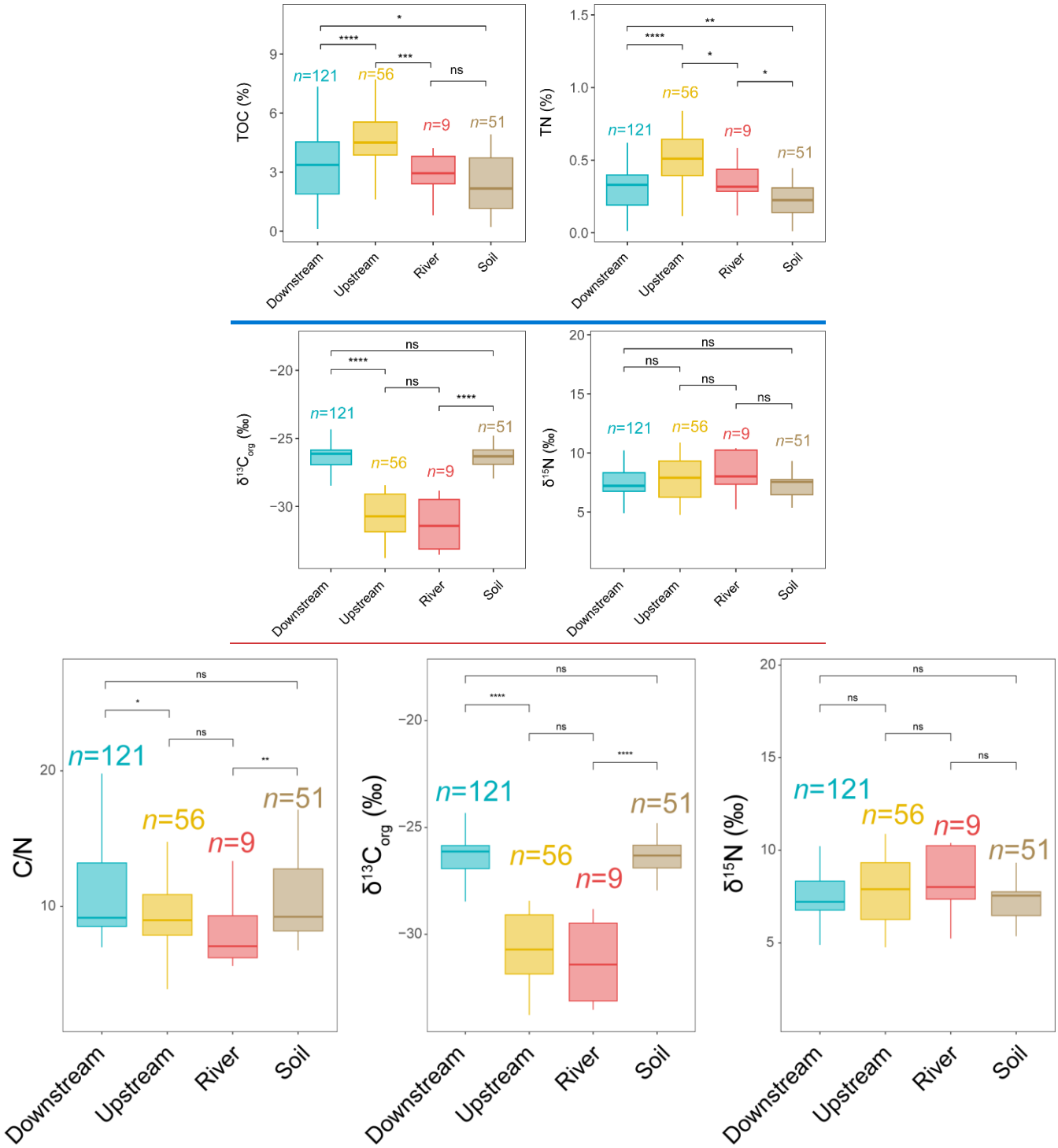


Figure 2: Distribution of bulk parameters (~~TOC, TN~~C/N, $\delta^{13}C_{org}$ and $\delta^{15}N$) from soils (surficial soils and mudflat sediments) as well as river, upstream estuary and downstream estuary samples across the Seine River basin. Box plots of upstream and downstream

estuary samples are based on SPM and 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, $p > 0.05$).

The total organic carbon (TOC) content was significantly higher in the upstream estuary (4.64±1.42 %, based on SPM and river channel sediments) than in downstream estuary (3.30±1.69 %, based on SPM and sediments), soils (3.~~02~~03±3.49 %, based on surficial soils and mudflat sediments) and river (2.88±1.14 %, based on SPM) (Table S1) ~~(Fig. 2)~~. The total nitrogen (TN) content was higher in the upstream estuary (0.51±0.17 %, based on SPM and sediments) than in the river (0.37±0.15 %, based on SPM), downstream estuary (0.31±0.14 %, based on SPM and river channel sediments), and soils (0.24±0.17 %, based on surficial soils and mudflat sediments) (Table S1) ~~(Fig. 2)~~. Lower C/N values were observed in the river (8.04±4.31, based on SPM) and upstream estuary (9.42±3.67, based on SPM and sediments) compared to the downstream estuary (10.73±3.59, based on SPM and sediments) and soils (11.59±4.79, based on surficial soils and mudflat sediments) (Fig. 2). Much lower values of $\delta^{13}C_{org}$ were observed in river (-31.30±1.91 ‰, based on SPM) and upstream estuary (-30.62±1.66 ‰, based on SPM and sediments) than in the downstream estuary (-26.45±1.34 ‰, based on SPM and sediments) and soils (-26.55±1.13 ‰, based on surficial soils and mudflat sediments) (Fig. 2). In addition, no significant differences in δ^{15}~~3C~~N_{org} were observed along the river basin (Fig. 2).

3.2 Distribution of brGDGTs from land to sea

The different 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) were detected in all studied samples. The brGDGT chromatograms from upstream samples (SPM and river channel sediments) differed markedly from downstream estuarine samples (SPM and sediments). For example, 6-methyl brGDGTs were much more abundant than 5-methyl brGDGTs in the river (SPM) and upstream estuary (SPM), whereas the strong predominance of 6-methyl *vs.* 5-methyl brGDGTs decreased in the downstream estuarine SPM samples (Fig. S2~~3~~). Furthermore, the peaks of the recently described 7-methyl brGDGTs and their late-eluting isomers (i.e. 1050d) were more pronounced in the downstream estuary than in the rest of the Seine basin (Fig. S2~~3~~).

The relative abundances of the brGDGTs were determined all along the Seine River basin (Fig. 3~~4~~ and Figs. S3~~2~~, S4~~3~~). The 6-methyl brGDGTs ($IIIa_6$, IIa_6, $IIIb_6$, IIb_6, and $IIIc_6$) were significantly higher in river (SPM) and upstream estuary (SPM and river channel sediments) than in soils (surficial soils and mudflat sediments) and downstream estuary (SPM and river channel sediments). In addition, the relative abundances of 7-methyl brGDGTs ($IIIa_7$ and IIa_7) and their late-eluting compound (1050d) were significantly higher in downstream estuary (SPM and river channel sediments) and soils (surficial soils and mudflat sediments) than in river (SPM) and the upstream estuary (SPM and river channel sediments).

The concentration of total brGDGTs also showed differences along the land to sea continuum (Fig. S54a). The total brGDGTs concentration decreased from river (10.51 ± 5.91 μg/g organic carbon (C_{org}), based on SPM samples) to upstream estuary (7.52 ± 5.09 μg/g C_{org}, based on SPM and sediments) and downstream estuary (4.95 ± 4.09 μg/g C_{org}, based SPM and sediments). In soils from all the Seine basin, the concentration in total brGDGTs (1.55 ± 1.61 μg/g C_{org}, based on surficial soils and mudflat sediments) was significantly lower than that in SPM and sediments (Fig. S54a).

A Principal Component Analysis (PCA) was performed to statistically compare the fractional abundances of brGDGTs from different location (river, upstream and downstream estuary, based on SPM and sediments collected in the river channel), which explained 40.954.1% of the variance in the first two dimensions (Fig. 4a). The first axis (PC1) explained 40.9% of the variance, with negative loadings for most of the 6-methyl brGDGTs and positive loadings for the remaining brGDGTs (Fig. 45a). Samples from the downstream estuary clustered well apart from those from the river and upstream parts. Specifically, the brGDGT distribution was dominated by 6-methyl brGDGTs ($IIIa_6$, $IIIb_6$, $IIIc_6$, IIa_6, and IIb_6) in river and upstream estuarine samples, whereas in downstream estuary, it was driven by 5-methyl brGDGTs (III_5, IIa_5, IIc_5, IIb_5 and $IIIb_5$), tetramethylated brGDGTs (Ia, Ib, and Ic), 7-methyl brGDGTs ($IIIa_7$, IIa_7, and IIb_7), and their late-eluting compounds (1050d and 1036d). The brGDGT distributions of soils (surficial soils and mudflat sediments) were included in the PCA biplot performed on SPM and river channel sediments. This revealed that the brGDGT distribution in soils mostly overlap with the one in downstream estuarine SPM and river channel sediments (Fig. 45a).

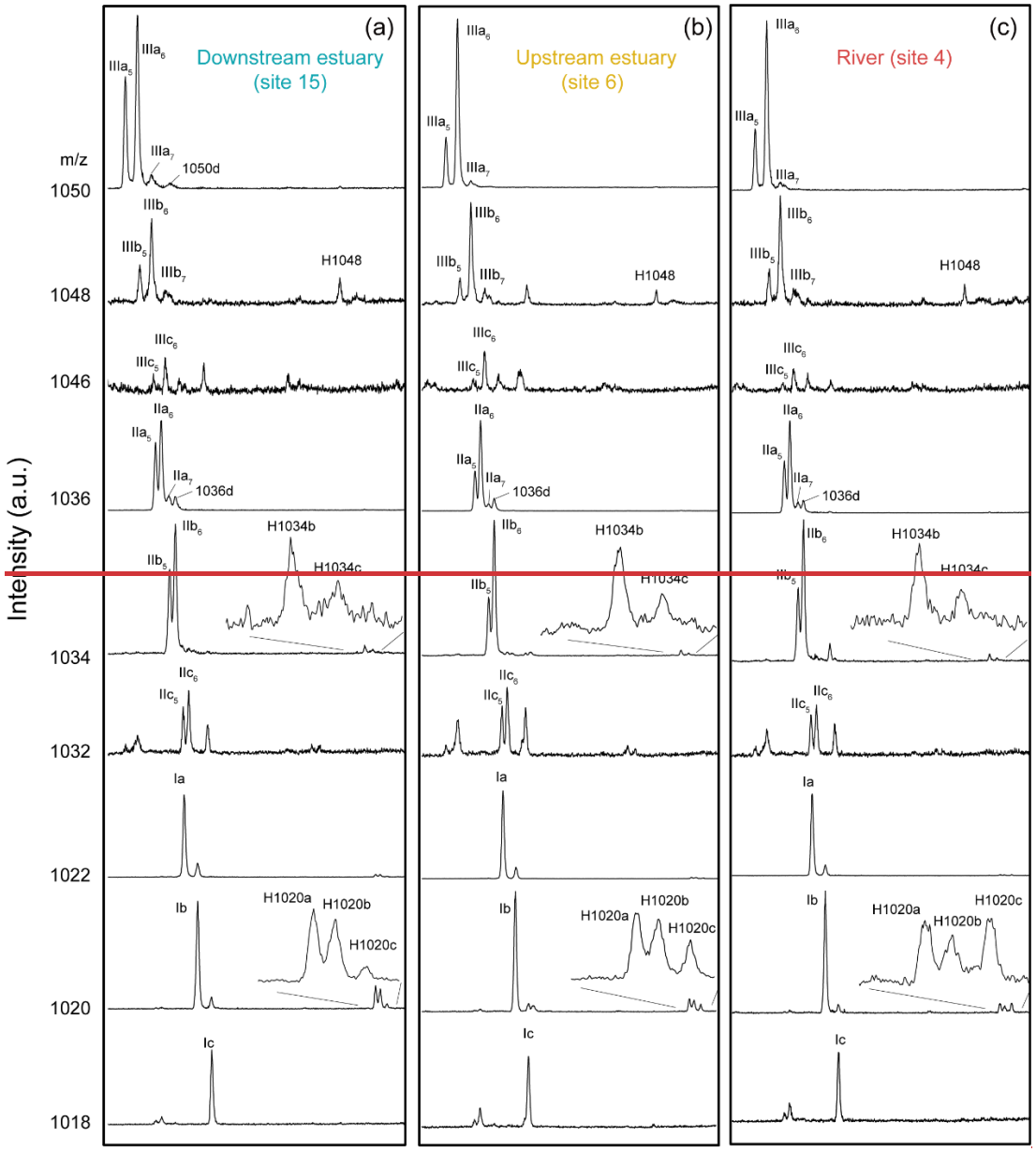


Figure 3: Extracted chromatograms of brGDGTs and brGMGTs for the SPM samples collected in (a) site 15 (Tancarville, September 2020), (b) site 6 (Oissel, July 2019) and (c) site 4 (Les Andelys, July 2019). The nomenclature for the penta- and hexamethylated brGDGTs: 5-methyl brGDGTs (IIIa$_5$, IIIb$_5$, 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$_7$).

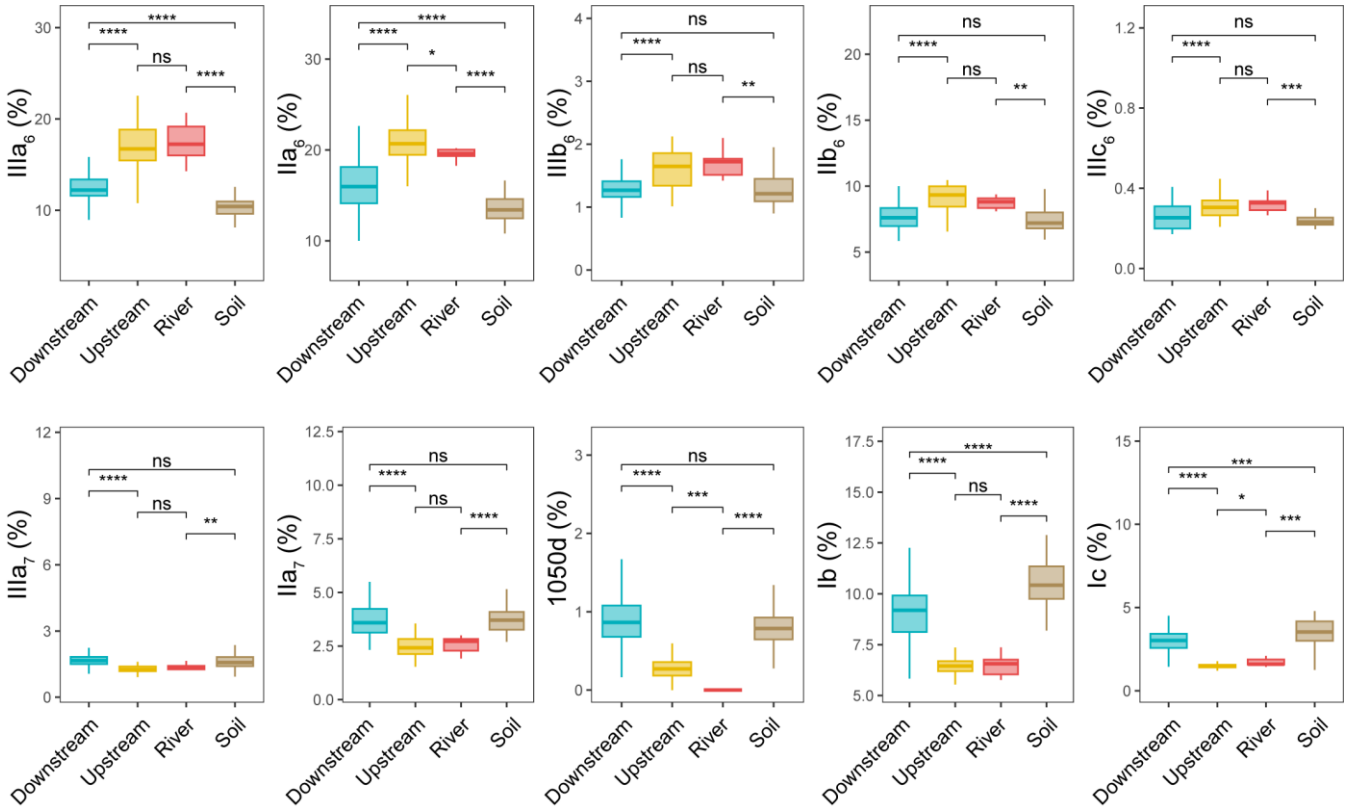


Figure 34: Relative abundances of selected individual brGDGTs over 20 brGDGTs (IIIa$_5$, IIIb$_5$, IIIc$_5$, IIa$_5$, IIb$_5$, IIc$_5$, IIIa$_6$, IIIb$_6$, IIIc$_6$, IIa$_6$, IIb$_6$, IIc$_6$, IIIa$_7$, IIIb$_7$, IIa$_7$, Ia, Ib, Ic, 1050d, and 1036d) from soils (surficial soils and mudflat soils/sediments, *n*=51), river (*n*=9), upstream estuary (*n*=56), and downstream estuary (*n*=121) samples across the Seine River basin: cyclopentane-containing tetramethylated brGDGTs (Ib and Ic), 6-methyl brGDGTs (IIa$_6$, IIIa$_6$, IIb$_6$, IIIb$_6$, and IIIc$_6$), 7-methyl brGDGTs (IIa$_7$ and IIIa$_7$) and brGDGTs 1050d. Box plots of upstream and downstream estuary samples are based on SPM and 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. Boxes are color-coded based on the sample type (soil in brown, river in red, upstream estuary in yellow, and downstream estuary in blue). 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).**

A Redundancy analysis (RDA) was performed to investigate the influence of the environmental factors (TOC, TN, temperature, water discharge and salinity) on the brGDGT distributions in SPM samples (Fig. 56a and Table 2). It allowed to explain 39.79% of the variability through two dimensions. The RDA triplot (Fig. 56) showed how these factors correlate to the distributions of individual brGDGTs. The first axis of the RDA explained 33.16% of the variability and was primarily correlated with salinity and TN, whereas the second axis explained 6.63% of the variability and was associated with temperature, water discharge and TOC (Fig. 56a and Table 2). Based on hierarchical partitioning, salinity and TN were the two most important variables in explaining the brGDGT variations (individual importance of 14.97 % for salinity and 13.47 % for TN; Fig. 56b

and Table 2). Compared with the salinity and TN, other available parameters have much lower individual importance (3.68 % for water discharge, 3.6 % for temperature and 2.12 % for TOC; Fig. 5~~6~~b and Table 2).

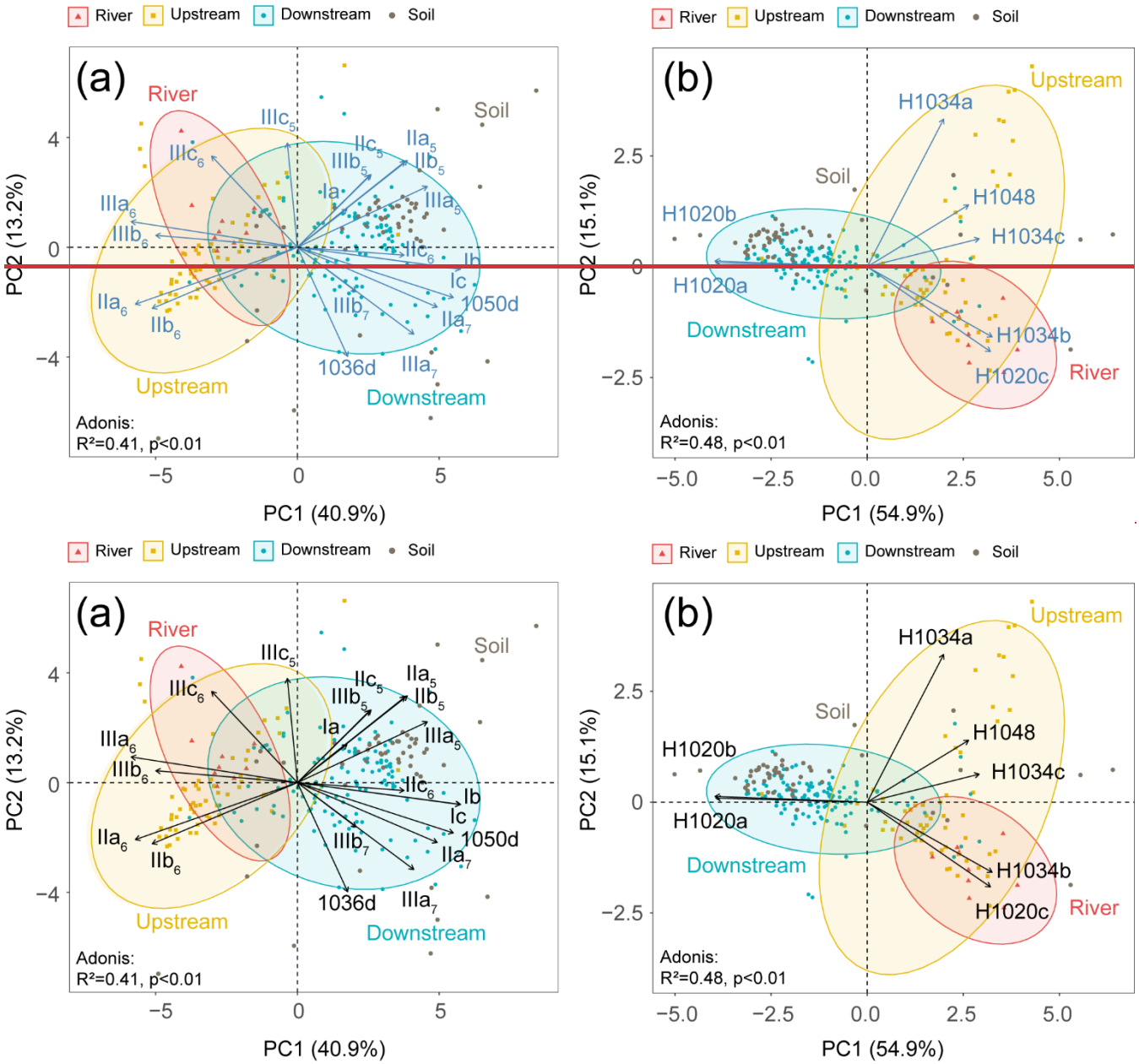


Figure 4~~5~~: PCA analysis of fractional abundances of (a) brGDGTs and (b) brGMGTs. The coordinates of passive individuals (soils) are added passively as an overlay. They~~, which~~ are predicted based on the information provided by the existing PCA performed on SPM and sediments (active individuals). ~~PCA scores of soils are added passively as an overlay. Their coordinates are predicted based on the information provided by the PCA performed on SPM and sediments (active individuals).~~ Adonis analysis was used to evaluate how variation can be explained by the variables (999 permutations).

Table 2. RDA results

	variables	RDA scores		Individual Importance (%)
		Axis 1	Axis 2	
brGDGTs	TOC	0.17	-0.49	2.12 *
	TN	0.72	-0.23	13.47 ***
	Temperature	0.41	-0.82	3.6 **
	Salinity	-0.73	-0.10	14.97 ***
	Water discharge	-0.30	0.93	3.68 **
brGMGTs	TOC	0.08	0.68	3.5 **
	TN	-0.34	0.09	4.18 **
	Temperature	0.04	0.64	1.17 ns
	Salinity	0.83	-0.50	17.45 ***
	Water discharge	-0.17	-0.71	2.16 *

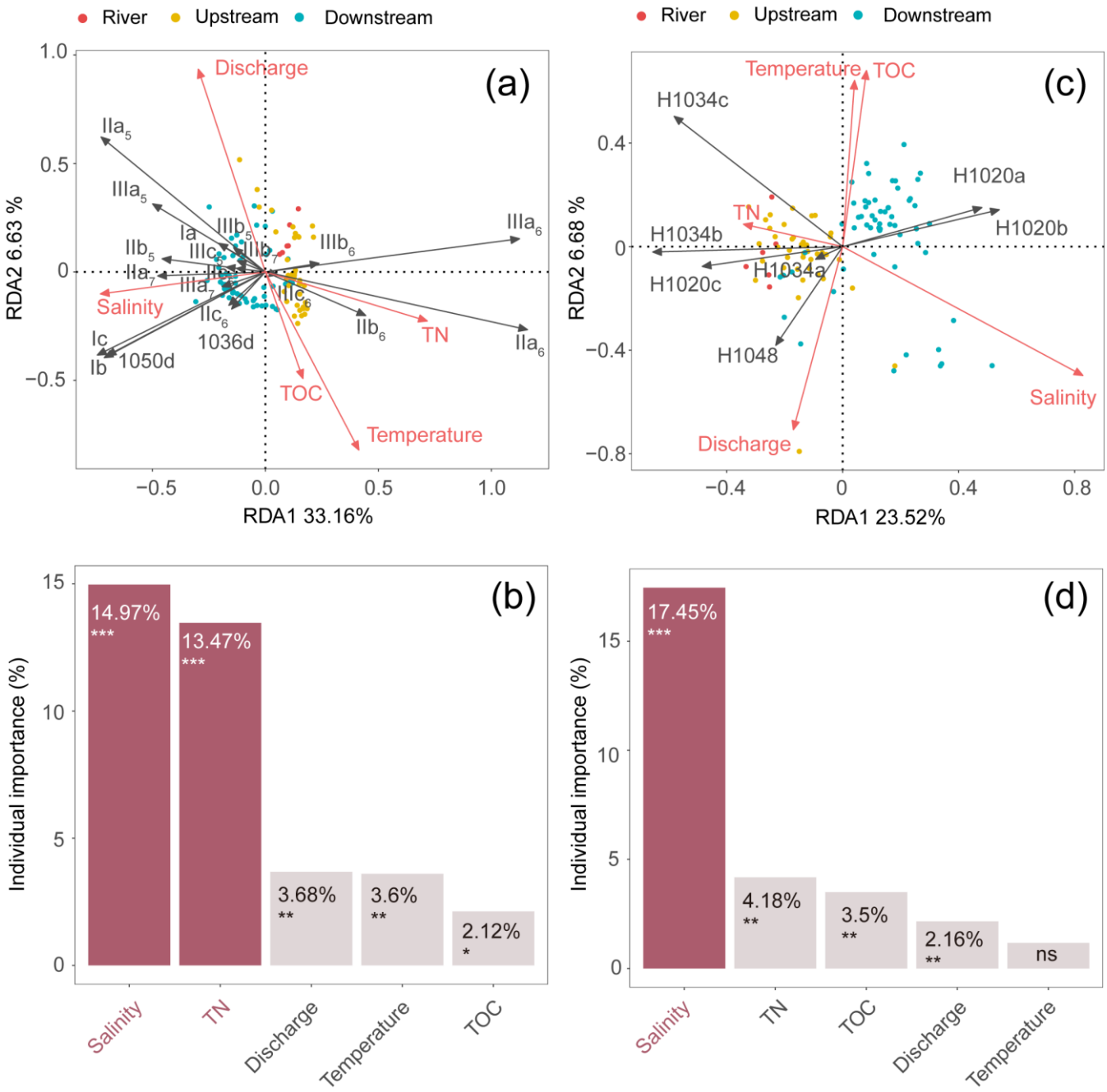


Figure ~~5~~6: RDA analysis showing relationships between environmental factors (TN, TOC, salinity, temperature, discharge, red arrows) and fractional abundances of (a) brGDGTs and (c) brGMGTs. The individual importance of the environmental factors (TN, TOC, salinity, temperature, and discharge) explaining the variation in (b) brGDGT and (d) brGMGT distributions was determined by hierarchical partitioning analysis. The dataset used for RDA analysis is composed of SPM from river (*n*=6; red), upstream estuary (*n*=42; yellow) and downstream estuary (*n*=59; blue). Significance level is indicated by asterisks: **p* < 0.05; ***p* < 0.01; ****p* < 0.001; ns, not significant, *p* > 0.05. *p*-values are derived from permutation tests (999 randomizations).

3.3 Distribution of brGMGTs from land to sea

~~The brGMGTs identified in (2019)previous studies were detected in the samples collected across the Seine River basin (Fig. 3).~~ The brGMGTs (H1020a, H1020b, H1020c, H1034a, H1034b, H1034c, and H1048) identified ~~in~~by Baxter et al. (2019) were detected in the samples collected across the Seine River basin. H1034a is the least abundant isomer and is below detection limit for most of the SPM and sediment samples in the Seine River basin (Fig. S6 and S7~~5~~). The chromatograms revealed distinct distributions in brGMGTs in the different parts of the basin (SPM and sediments), with e.g. a higher intensity for the homologue H1020c in the river samples (SPM) than in those from the upstream (SPM) and downstream estuary (SPM) (Fig. S2~~3~~). These spatial variations were apparent when calculating the fractional abundances of the individual brGMGTs (Fig. ~~7~~6 and Figs. S5, S6). From ~~upstream~~ river to downstream estuary, the relative abundances in H1020a and H1020b increased, whereas those in 1020c and H1034b decreased (Fig. ~~7~~6). In SPM and river channel sediments, the total brGMGT concentration was observed to be slightly (but not significantly) higher in the riverine part (0.26 ± 0.24 μg/g C_{org}) than in downstream estuary (0.20 ± 0.13 μg/g C_{org}) and upstream estuary samples (0.17 ± 0.18 μg/g C_{org}; Fig. S5~~4~~b). The total brGMGT concentrations were the lowest in soils (surficial soils and mudflat sediments) all over the basin (0.07 ± 0.09 μg/g C_{org}; Fig. S5~~4~~b).

The PCA analysis based on the brGMGT relative abundances (Fig. 4~~5~~b) explained 70 % of the variance in the first two dimensions, which separates samples from different parts of the basin. ~~allows to observe that samples from the different parts of the basin clustered well apart from each other.~~ The first axis explained 54.9 % of the variance, separating downstream estuarine samples from riverine and upstream estuarine samples, with negative loadings for two brGMGTs (H1020a and H1020b), and positive loadings for the remaining brGMGTs (H1020c, H1034a, H1034b, H1034c, and H1048). The second axis explained 15.1% of the variance and mainly separated the riverine and upstream estuarine samples, with higher relative abundances of compounds H1020c and H1034b in riverine samples (Fig. 4~~5~~b). The soil brGMGT distributions were passively added to the PCA biplot based on SPM and sediments, revealing that the soils largely overlap with the SPM and sediments collected in the downstream estuary (Fig. 4~~5~~b).

The RDA was performed to investigate the factors that could explain the variability of brGMGT distributions in SPM samples (Fig. 5~~6~~c and Table 2) and ~~allows to~~ explains 30.2 % of the variance in the first two axes. The RDA triplot showed that the first axis, accounting for 23.52 % of the variability, was mainly associated with salinity and to a lesser extent TN, while the second axis (6.68 %) was mainly driven by temperature, TOC and water discharge (Fig. 5c and Table 2). Based on hierarchical partitioning, salinity had the highest individual importance (17.45 %) in explaining the variability of brGMGT distribution followed by TN (4.18 %), TOC (3.5 %), and water discharge (2.16 %) (Fig. 5~~6~~d and Table 2).

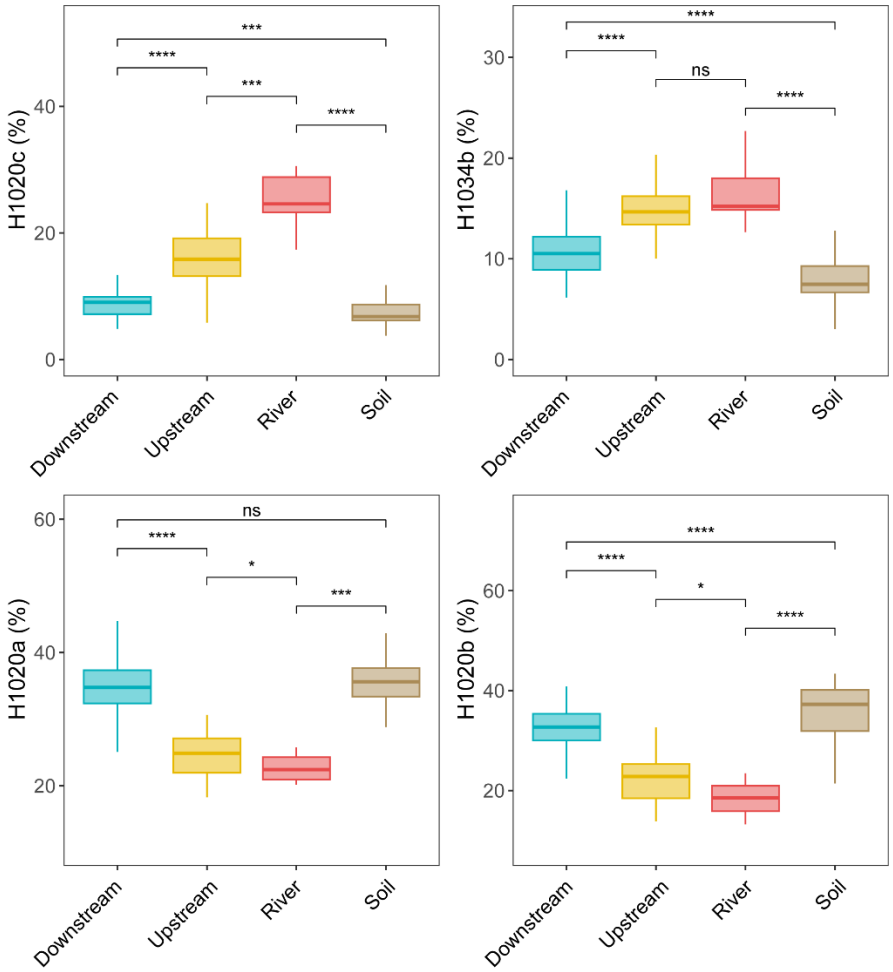


Figure 6~~7~~: Relative abundance of selected individual brGMGTs over 7 brGMGTs (H1020a, H1020b, H1020c, H1034a, H1034b, H1034c, and H1048) from soils (surficial soils and mudflat soils/sediments, *n*=51), river (*n*=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) across the Seine River 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. ~~Boxes are color-coded based on the sample type (soil in brown, river in red, upstream estuary in yellow, and downstream estuary in blue).~~ Statistical testing was performed by a Wilcoxon test (*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$; ****$p < 0.0001$; ns, not significant, $p > 0.05$).

4 Discussion

4.1 Sources of brGDGTs and environmental controls on their distribution

4.1.1 Sources of brGDGTs

In order to determine the predominant origin of brGDGTs in the Seine River basin, the overall brGDGT concentrations and distributions in SPM and river channel sediments (n=186) were compared with those in soils (surficial soils and mudflat sediments, n=51). The brGDGT concentrations (normalized to C_{org}) and relative abundances of several brGDGTs (i.e. $IIIa_6$, IIa_6, $IIIb_6$, IIb_6, and $IIIc_6$) in the SPM and sediments were significantly higher than those in soils (p<0.05, Wilcoxon test; Fig. S54a and Fig. 34). Such differences in brGDGT concentrations and relative abundances between soils and aquatic settings (SPM and sediments) imply that at least part of the brGDGTs in the water column and sediments of the Seine River basin is produced *in situ*. This is in agreement with previous findings which suggested an *in situ* aquatic contribution to the brGDGT pool (Peterse et al., 2009; De Jonge et al., 2015; Crampton-Flood et al., 2021; Kirkels et al., 2022b).

More specifically, the fractional abundances of the two major 6-methyl brGDGTs (IIa_6 and $IIIa_6$) are significantly higher in the Seine River and upstream estuary than in soils (Fig. 34). This confirms that these brGDGTs are mostly produced within the river, adding to the growing body of evidence supporting riverine 6-methyl brGDGT production in water column and/or sediment (De Jonge et al., 2015; Bertassoli et al., 2022; Kirkels et al., 2022b). A subsequent shift in the brGDGT distributions in the downstream estuary compared to the upstream areas is observed in the Seine River basin. The PCA analysis shows a separation of downstream estuarine samples (influenced by seawater intrusion) from riverine and upstream estuary ones (without significant seawater intrusion) (Fig. 45a). This difference is predominantly driven by the higher abundances of 6-methyl brGDGTs in riverine and upstream estuarine samples *vs.* higher abundances of 5- and 7-methyl brGDGTs as well as compounds Ib, Ic, and late eluting brGDGTs 1050d, 1036d in downstream estuarine samples (Figs. 34, 5a 4a and Fig. S43). The decrease in the fractional abundance of 6-methyl brGDGTs from the upstream estuary to the downstream estuary cannot be explained by the dam located at Poses (Fig. 1a). This dam separates the riverine part of the Seine from the upstream estuarine section. Even during the low-flow season (Fig. 1b), at least part of the water from the Seine River upstream of Poses flows into the estuary (Romero et al., 2019). Thus, This indicates that the dam should not prevent (part of) the riverine brGDGTs associated with SPM from reaching the estuary. However, Iit cannot be excluded that part of the riverine sediments is trapped by this dam. In additionNevertheless, all our estuarine samples were collected downstream of the dam, implying that the observed changes in brGDGT abundance and distribution within the estuary are a result ofintrinsic to -the biogeochemical functioning of the Seine estuary and cannot be attributed to the dam, not the dam.

Instead, a shift in brGDGTs along the land-sea continuumIt may reflect the fact that riverine 6-methyl brGDGTs are more easily degraded than soil-derived homologues and only partially transferred downstream. In addition to that, the riverine brGDGT signal may be diluted by brGDGTs from other sources during downstream transport. The firstThis hypothesis is

based in a previous study, which showed a shift in brGDGT distribution from the Yenisei River to the Kara Sea (De Jonge et al., 2015). They interpreted this to be a preferential degradation of labile (riverine) 6-methyl brGDGTs and the enrichment in less labile (soil-derived) 5-methyl brGDGTs during transport (De Jonge et al., 2015). This suggests that only limited amounts of riverine 6-methyl brGDGTs are transferred to the ocean, as also shown in other recent studies (Cao et al., 2022; Kirkels et al., 2022b). Such preferential degradation of 6-methyl brGDGTs over other brGDGTs could be attributed to variations in how these molecules are attached to soil particles (Huguet et al., 2008). Indeed, the higher degradation of 6-methyl brGDGTs upstream could be attributed to their different attachment to particles compared to downstream. The median diameter of the SPM was monitored between February 2015 and June 2016 in both the upstream (sites 7 and 10) and downstream (sites 15 and 17) parts of the Seine Estuary (Druine, 2018). The particle size showed only slight dispersion (80-110 µm) under various hydrological conditions in the upstream estuarine section. The homogeneity in particle size in the upstream estuary likely reflects its predominantly continental origin (i.e. from the Seine River before the dam at Poses). In contrast, a large variability in the size of SPM particles was observed in the downstream estuary (15-20 µm to 80-90 µm), attributed to the complex flocculation and defragmentation processes of particles in this part of the estuary (Druine, 2018). Hence, the variability in the size of SPM particles from upstream to downstream could influence the distribution of brGDGTs in the Seine estuary.

In addition to this hypothesis, a shift in brGDGT distribution during downstream transport could be explained by mixing with autochthonous (i.e. estuarine-produced) brGDGTs (Crampton-Flood et al., 2021). The relative abundance of several brGDGTs (i.e. Ib, Ic, $IIIa_7$, IIa_7 and 1050d) in the downstream part of the Seine River basin is indeed significantly higher than the one in the upstream part (p<0.05, Wilcoxon test; Fig. 3~~4~~), suggesting *in situ* brGDGT production in saltwater. Such a saltwater contribution can be visualized by the PCA based on brGDGT distribution, showing the positive score of the aforementioned compounds with the first axis (Fig. 4~~5~~a). This axis is dominated by downstream samples influenced by seawater intrusion in the Seine Estuary (Fig. 4~~5~~a).

It should be noted that brGDGT distributions in soils were roughly similar to those observed in downstream estuarine samples (SPM and river channel sediments~~; Fig. 45a~~) based on the PCA (Fig. 4a). Additionally, no significant differences were observed in the fractional abundances of several brGDGTs ($IIIb_6$, IIb_6, $IIIc_6$, $IIIa_7$, IIa_7, 1050d, $IIIa_5$, $IIIb_5$, $IIIb_7$, $IIIc_5$, IIc_6, and Ia) between soils and downstream samples (Fig. 3 and Fig. S4). This~~e~~ similarity in brGDGT distributions ~~between soils and downstream estuarine samples~~ may be due to the influx of brGDGTs from the downstream soils into the downstream estuary~~the overrepresentation of downstream soil samples~~, as 82% of the soils were collected downstream (Fig. 1a and Table 1). Hence, it cannot be excluded that brGDGTs detected in downstream estuarine samples are at least partly derived from soils of the watershed. Nevertheless, the soil-derived brGDGT contribution to the downstream estuarine samples is expected to be much lower than the autochthonous one, as the average brGDGT concentration in soils was ca. 3 times lower than the one in downstream estuarine (i.e. SPM and river channel sediment) samples (Fig. S5~~4~~a).

In order to further assess whether downstream estuarine samples could be distinguished from soils, w~~W~~e ~~then~~ applied the machine learning model (BigMac)~~,~~ developed by Martínez-Sosa et al. (2023)~~,~~ to our dataset with isoGDGT and brGDGT data as input. ~~This might allow us to assess whether downstream estuarine samples could be distinguished from soils. The output of the BigMac model indicates~~ M~~that m~~ost of our samples (SPM, sediments, and soils) ~~are~~were predicted as lake-type, with only one soil sample (soil6) collected at site B predicted as soil-type. This ~~sugg~~model suggests ~~ests~~ that, when considered altogether, the isoGDGT and brGDGT ~~signatures~~distributions ~~in our aquatic samples differ from those in the soils investigated by Martínez-Sosa et al. (2023), but they~~ are similar ~~to~~ in aquatic and soil samples from the Seine estuary and differ from the soil-type samples described by Martínez-Sosa et al. (2023) ~~the soils from the Seine River basin and to lake-type samples as described in Martínez-Sosa et al. (2023). This similarity indicates a comparable data pattern for both isoGDGTs and brGDGTs but does not necessarily imply similarity in brGDGT alone. Additionally, s~~Since the BigMac model does not include a river-type or estuary-type category (Martínez-Sosa et al., 2023), further inclusion of both isoGDGT and brGDGT data from global riverine/estuarine samples in the BigMac model may help to enhance predictions for river-type or estuary-type SPM and sediment samples.

The BigMac model distinguishes the type of samples using IIa_6 and crenarchaeol as the two most important ~~characteristics~~predicting variables. When accounting for both isoGDGTs and brGDGTs in the Seine River basin, the fractional abundance of crenarchaeol *vs.* total GDGTs (i.e. isoGDGTs + brGDGTs) varies significantly, whereas the one of IIa_6 does not differ significantly between the downstream estuary and soils (Fig. S8). Hence, the inclusion of isoGDGTs in the model ~~the Seine River basin~~ may ~~dilute~~highly reduce the differences between sample types, as we observe significant differences in the fractional abundance of IIa_6 when ~~considering brGDGTs alone~~calculated vs. total brGDGTs only (Fig. 3). As the BigMac model relies on ~~uses~~ both isoGDGT and brGDGT distribution, with no option of using brGDGTs alone, we chose to perform an independent analysis to assess the similarity ~~of brGDGT~~in brGDGT relative abundance between ~~in~~ downstream SPM and sediment samples ~~to~~on the one hand and ~~-~~soils from~~in~~ the Seine River basin on the other hand. This model was developed using the same algorithm (random forest) as Martínez-Sosa et al. (2023). Binary classification (downstream estuary *vs*. soils) was performed based on fractional abundances of brGDGTs. The trained model (Fig. S9) indicated distinguishable brGDGT distributions between downstream estuary (SPM and sediments) and soil samples, supporting the *in* ~~-~~*situ* production of brGDGTs in the downstream estuary. Although most of our soil samples were collected from the downstream estuary and showed similarity with the downstream SPM and sediment samples through PCA and comparison of fractional abundances, we were ~~still~~ able to distinguish their brGDGT compositions using machine learning.
~~This is further supported by a machine learning approach aimed at distinguishing brGDGT types in distinct environments. Martínez-Sosa et al. (2023) showed the efficiency of random forest algorithm to distinguish the origin of GDGTs from a global dataset comprised of soil, peat, marine and lake samples. Consequently, we chose to use this algorithm to train and classify brGDGTs from downstream estuary and soil samples in the Seine River basin. Binary classification (downstream estuary *vs.* soils) was performed using a random forest model based on fractional abundances of brGDGTs. This trained model (Fig. S8)~~

~~indicated distinguishable brGDGT distributions between downstream estuary (SPM and sediments) and soil samples, supporting in situ production of brGDGTs in the downstream estuary.~~

~~(2023)~~

4.1.2 Environmental controls on the brGDGT distribution

As several individual brGDGTs are suggested to be preferentially produced either in the riverine or estuarine parts of the Seine basin, their distribution might be related to ambient environmental factors. The RDA (performed on SPM samples) highlights the relationships between the available environmental variables (salinity, TN, TOC, water discharge and temperature) and the relative abundances of brGDGTs. Hierarchical partitioning indicates that salinity is the most important factor influencing the brGDGT distribution (14.97 %) in the Seine River basin (Fig. 5~~6~~b and Table 2). Salinity is related to the relative abundances of compounds Ib, Ic, 7-methyl brGDGTs and the late-eluting homologs 1050d and 1036d that scored negatively on the first axis of the RDA (Fig. 5~~6~~a). This is in line with the positive significant correlations between salinity and the relative abundances of these compounds (Fig. S10~~7~~). This trend~~s~~ also supports the assumption ~~made~~ about the aquatic production of ring-containing tetramethylated brGDGTs (Ib and Ic) in Svalbard fjords which was thought to be linked to a salinity change (Dearing Crampton-Flood et al., 2019). The 7-methyl brGDGTs and their late-eluting isomers were also shown to be much more abundant in hypersaline lakes than those of lower salinity (Wang et al., 2021). Such a salinity-dependent brGDGT composition has previously been interpreted by membrane adaptation to salinity changes or by a shift in bacterial community composition (Dearing Crampton-Flood et al., 2019; Wang et al., 2021). Hence, the significant positive correlations between salinity and these compounds in the Seine River basin suggest that brGDGT-producing bacteria have similar physiological mechanisms (i.e., membrane adaptation) to those reported in other aquatic settings (lakes and fjords) and/or that the diversity of these bacteria is changing along the river-sea continuum. ~~However, t~~The salinity proxy (IR_{6+7me}) proposed by Wang et al. (2021) does not show significant correlations with salinity in this study (p>0.05, Wilcoxon test; Fig. S10~~9~~). This suggests that the IR_{6+7me} index is relatively insensitive in the Seine Estuary, potentially due to the preferential production of ~~certain~~ 6-methyl brGDGTs in specific estuarine regions (i.e. KP 255.6-337).

~~The relative abundances of several 6-methyl brGDGTs (i.e. IIa_6, $IIIa_6$, and IIb_6) in the Seine River basin reveal significant negative correlations with salinity (p<0.05, Wilcoxon test; Fig. S7), which is in contrast with the positive relationships previously found in lakes (Wang et al., 2021). The distinct behavior of 6-methyl brGDGTs between lakes and the Seine river-sea continuum might be due to the lower salinity range in the Seine River basin (0-32 psu) *vs.* the lakes (0-376 psu) investigated by Wang et al. (2021). This suggests that the limited range of salinity variation in the Seine River basin might be insufficient to trigger significant 6-methyl brGDGT production, as observed in hypersaline lakes.~~

~~Alternatively~~Indeed, the significant negative correlations between the salinity and the relative abundance of 6-methyl brGDGTs is observed in the Seine basin (Fig. S10~~9~~), which suggests that the bacteria producing 6-methyl brGDGTs are preferentially present in the low salinity area of the estuary. To explore this ~~hypothesis~~further, we investigated the spatio-temporal variations of the 6-methyl *vs*. 5-methyl brGDGTs ratio: IR_{6Me} (Fig. 7~~8~~). High IR_{6Me} values (0.69 ± 0.10) are associated with enhanced *in situ* production of 6-methyl brGDGTs within the Yenisei river (De Jonge et al., 2015). In the Seine River basin, seasonal variation in IR_{6me} is observed. Specifically, much higher IR_{6Me} values are observed in a specific zone of the estuary (260 < KP < 340) with a low salinity range (1.18 ± 2.71 ~~psu~~) during low-flow season (Fig. 7~~8~~), suggesting that 6-methyl brGDGTs are preferentially produced in this zone when water discharge is low. Similarly, preferential production of 6-methyl brGDGT at low discharges was previously observed in other river systems, including the Amazon River basin (Kirkels et al., 2020; Crampton-Flood et al., 2021; Bertassoli et al., 2022) as well as Black and White Rivers (Dai et al., 2019). It was suggested that the enhanced 6-methyl brGDGT production at low flows was due to slow flow velocity and reduced soil mobilization. Although these hypotheses could account for the temporal variation in IR_{6Me} in the Seine River basin, they are unlikely to explain the substantially high IR_{6Me} values in this specific zone. Other environmental variables such as dissolved oxygen contents (Wu et al., 2021) and pH (De Jonge et al., 2014, 2015) were previously suggested to have a potential influence on 6-methyl brGDGT distributions. Nevertheless, these two environmental factors do not co-vary with IR_{6Me} in the present study and can be ruled out as causes of variation in 6-methyl brGDGT distribution along the Seine river-sea continuum (Fig. 7~~8~~). Hence, the production of 6-methyl brGDGTs in this zone of the Seine Estuary has to be triggered by other factors, such as the nutrient concentration.

High nutrient levels were shown to favor the production of 6-methyl versus 5-methyl brGDGTs in the water column of mesocosm experiments (Martínez-Sosa and Tierney, 2019). As the nutrient concentration is higher in the upstream part of the Seine estuary (Wei et al., 2022), ~~-~~and this zone is characterized by high proportions of agricultural land use (Flipo et al., 2021), the substantial production of 6-methyl brGDGT observed in the aforementioned zone (260 < KP < 340; Fig. 7) during low flows could be attributed to elevated nutrient levels, particularly nitrogen, resulting from intense agricultural activities.~~the substantial 6-methyl brGDGT production observed in the aforementioned zone (260 < KP < 340, Fig. 8) at low flows could be due to the high amount of nutrients, especially nitrogen.~~ This is supported by the RDA triplot showing strong correlation of TN with the brGDGT distribution in the Seine basin, with the major 6-methyl brGDGTs (i.e. IIa_6 and $IIIa_6$) plotting close to TN in the RDA triplot (Fig. 5~~6~~a). In addition, TN and $\delta^{15}N$ are observed to co-vary with IR_{6Me} and to peak in the same zone (260 < KP < 340; Fig. 7~~8~~) during the low-flow season. Nitrate from sewage effluents and manure are generally enriched in ^{15}N compared to other sources, leading to much elevated $\delta^{15}N$ values (10–25‰) (Leavitt et al., 2006; Andrisoa et al., 2019). Nutrients, in the form of nitrogen, can be concentrated at low discharges, thus triggering phytoplankton blooms (Romero et al., 2019). Hence, the elevated TN and $\delta^{15}N$ signals in a specific zone of the estuary (260 < KP < 340) could be attributed to the increase of nitrogen loadings and ^{15}N-enriched nitrate uptake by phytoplankton developing intensively during the low-flow season. The much higher chlorophyll *a* ~~concentrations~~concentration in this zone under low discharge conditions supports the

hypothesis of phytoplankton blooms (Fig. 7~~8~~). This high phytoplankton biomass might consequently create an environment that accelerates the growth and production of heterotrophic bacteria, which can in turn transform phytoplankton-derived organic matter (Buchan et al., 2014). As the brGDGT-producers were suggested to have a heterotrophic lifestyle (Weijers et al., 2010; Huguet et al., 2017; Blewett et al., 2022), they may transform phytoplankton-derived organic matter and thus participate in N-cycling during blooms. Hence, the co-variations of all the parameters (IR_{6Me}, TN, $\delta^{15}N$, and Chl *a* concentration) peaking in the low salinity area during low-flow season suggest that low salinity range and high phytoplankton productivity represent favorable conditions for 6-methyl brGDGT production.

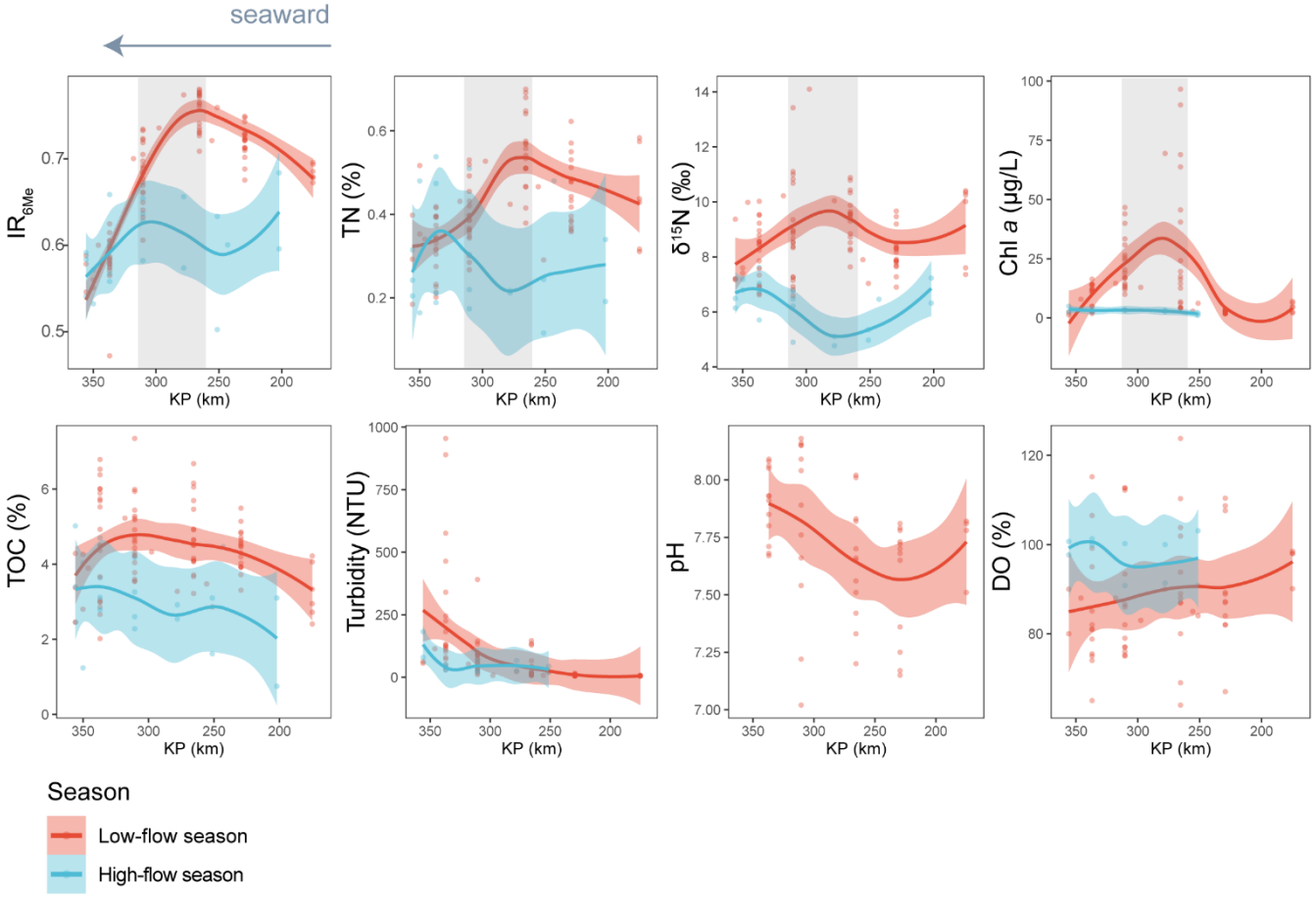


Figure 7~~8~~: Spatio-temporal variations of IR_{6Me} and several environmental factors, including TN (%), $\delta^{15}N$ (‰), Chla (µg/L), TOC (%), turbidity (NTU), pH, and dissolved oxygen saturation (DO, %). The trends showing variations were based on locally estimated scatterplot smoothing (LOESS) method with 95% confidence intervals. KP (kilometric point) represents the distance in kilometers from the city of Paris (KP 0). Dataset is composed of SPM. The shaded area highlights a zone (KP 255.6-337) where IR_{6Me} and several environmental parameters co-vary.

4.2 Sources of brGMGTs and environmental controls on their distribution

4.2.1 Sources of brGMGTs

Similarly to the brGDGTs, the brGMGTs can also be produced *in situ* within the water column and/or sediments (Baxter et al., 2021; Kirkels et al., 2022a). In previous studies, brGMGTs were detected only in part of the soils surrounding the Godavari River basin (India; Kirkels et al., 2022a) and Lake Chala (East Africa; Baxter et al., 2021), suggesting a limited brGMGT production in soils in comparison to aquatic settings. Consistently, in the Seine River basin, concentrations of brGMGTs in SPM and sediment samples are significantly higher than those in soils ($p<0.05$, Wilcoxon test; Fig. S5~~4~~), pointing out their predominant aquatic source.

A notable compositional shift in brGMGT distribution is observed along the Seine River basin, as revealed by the separation of riverine, upstream and downstream estuarine samples in the PCA (Fig. 4~~5~~b). The relative abundance of 2 brGMGTs (H1020c and H1034b) gradually decreases across the basin (Fig. 6~~7~~) and is significantly correlated with those of 6-methyl brGDGTs (Fig. S11~~08~~). As 6-methyl brGDGTs are mainly produced in freshwaters in the Seine basin, this suggests that brGMGTs H1020c and H1034b and 6-methyl brGDGTs have a common freshwater origin and that the mixture of fresh and marine waters along the estuary leads to the dilution of these compounds during downstream transport. H1020c is the dominant brGMGT homologue in SPM from the riverine zone of the Seine and one of the most abundant brGMGT in the upstream part of the estuary (Fig. 6~~7~~). Such a trend was also observed in SPM and riverbed sediments from the upper part of the Godavari River basin, which was attributed to *in situ* riverine brGMGT production of this compound (Kirkels et al., 2022a).

The fractional abundance of H1020a and H1020b homologues gradually increases along the Seine River basin. This is consistent with the higher abundances of H1020a and H1020b previously reported in marine sediments from the Bay of Bengal (Kirkels et al., 2022a). The predominance of these compounds in such samples was attributed to their *in situ* production in the marine realm. In line with this hypothesis, the relative abundances of brGMGTs H1020a and H1020b positively correlate with brGDGTs Ib, Ic, $IIIa_7$, IIa_7 and 1050d (Fig. S11~~07~~) in the Seine Estuary, suggesting a similar marine origin.

4.2.2 Environmental controls on the distribution of brGMGTs

The current knowledge on the parameters controlling the brGMGT distributions in the terrestrial and marine realm is still limited, as there is little literature available (Kirkels et al., 2022a).~~The current knowledge on the parameters controlling the brGMGT distributions in the terrestrial and marine realm is still limited.~~ The correlations between the brGDGT and brGMGT relative abundances in the Seine River basin (Fig. S11~~08~~) suggest that both types of compounds might be derived from overlapping source microorganisms, with common environmental factors controlling their membrane lipid composition. In the Seine River basin, salinity is shown to be the main environmental parameter influencing the brGMGT distribution, as also observed for brGDGTs (Fig. 5~~6~~). This is reflected in the significant ($p<0.05$) increase in the relative abundances of homologues

H1020a and H1020b with salinity and a concomitant significant negative correlation between this parameter and the relative abundances of homologues H1020c and H1034b (p<0.05, Wilcoxon test; Fig. 8~~9~~, a-d). Nevertheless, the individual effect of TN on brGMGT relative abundances is observed to be much lower compared to that observed for brGDGTs (Fig. 5~~6~~ and Table 2). This implies that, while having common controlling factors such as the salinity, they are also influenced by distinct parameters (i.e. TN), likely indicating distinct sources. This is consistent with a recent study showing that brGDGTs and brGMGTs likely originate from overlapping, but not identical origins (Elling et al., 2023).

The shift in brGMGT distribution observed across the Seine River basin (Fig. 4) could be due to a change in the diversity of brGMGT-producing bacteria and/or to an adaptation of these microorganisms to environmental changes occurring from upstream to downstream. The latter hypothesis seems unlikely, as a physiological adaptation of a given bacterial community would make it difficult to explain why the relative abundance of three isomers of compound H1020, which share a similar structure, varies differently in response to salinity changes. Hence, a shift in brGMGT-producing bacterial communities across the basin is more likely. Compounds H1020c and H1034b could predominantly be produced by bacteria preferentially growing in freshwater, and homologues H1020a and H1020b by bacteria preferentially living in brackish or saltwater.

4.3 Potential implications for brGMGTs as a proxy for riverine runoff in modern systems

The distinct brGMGT distributions in freshwater and saltwater could be used to trace the Organic Matter (OM) produced upstream all along the Seine basin. To trace such a riverine runoff signal, we propose a new proxy, the Riverine IndeX (RIX), based on the fractional abundances of brGMGTs H1020c and H1034b versus H1020a and H1020b (Eq. 6):

$$\mathrm{RIX} = \frac{H1020c+H1034b}{H1020c+H1034b+H1020a+H1020b} \qquad (6)$$

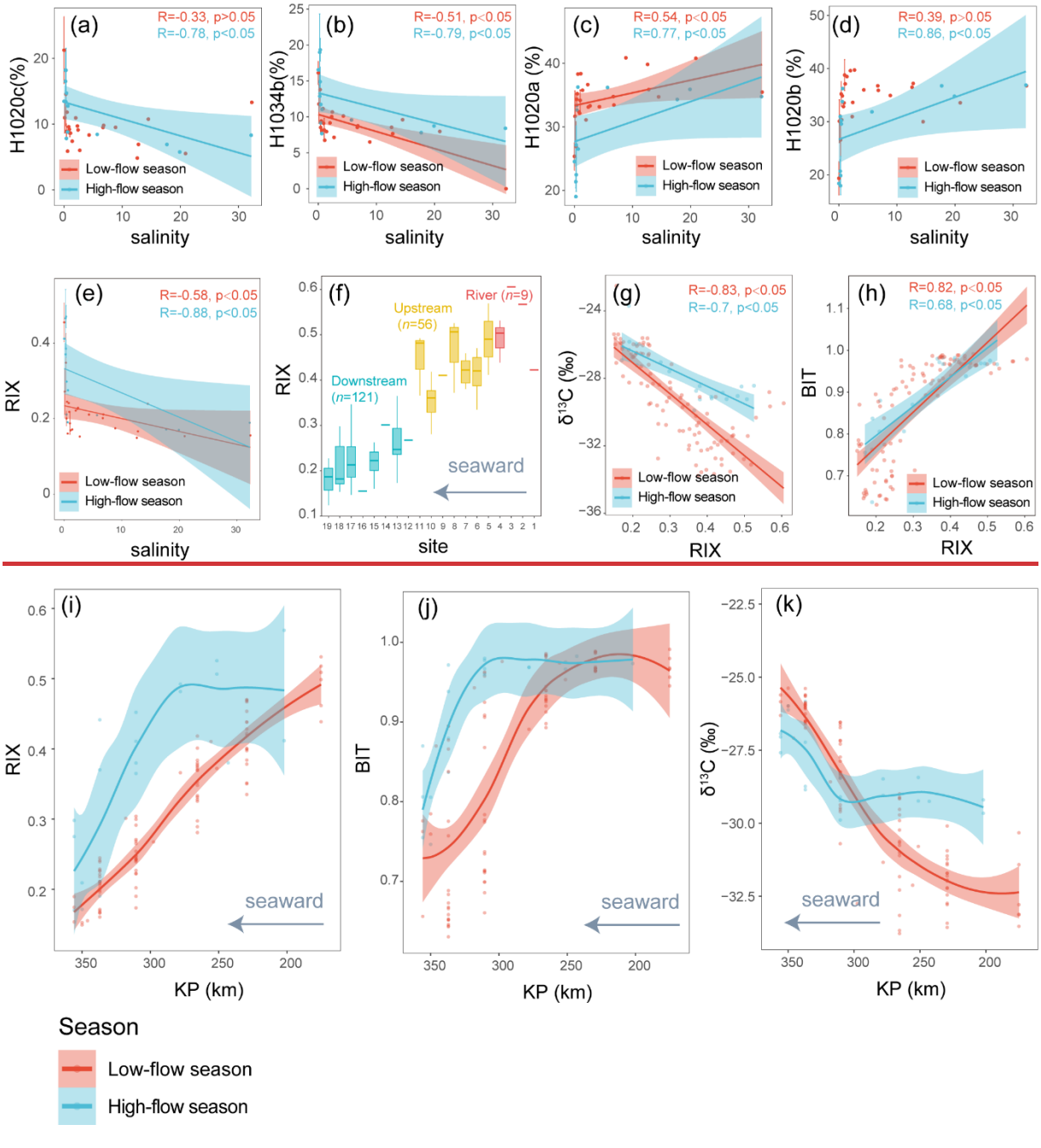
The rationale for RIX as a riverine runoff signal is that in freshwater environments, the pool of brGMGTs is dominated by H1020c and H1034b, whereas H1020a and H1020b prevail in saltwater environments. This is further supported by the significant negative correlation between RIX and salinity (Fig. 8e).

Since the other salinity proxies (i.e. ACE and IR_{6+7Me}) have shown positive correlations with salinity in previous studies (Turich and Freeman, 2011; Wang et al., 2021), ~~we~~they were expected ~~them~~ to be positively correlated with salinity and negatively correlated with RIX in the Seine River basin. However, the ACE index (Turich and Freeman, 2011) and IR_{6+7Me} (Wang et al., 2021) do not show significant correlations with salinity in the Seine River basin (p>0.05, Wilcoxon test; Fig. S10) and ~~, and these indices show significant positive correlations (with low correlation coefficients~~show weak but significant relations ~~)~~ with the RIX ~~in the Seine River basin~~ (Fig. S13). This could be attributed to the influence of other factors than salinity on these indices (i.e. ACE and IR_{6+7Me})~~ being influenced by factors other than salinity~~. ~~Compared with RIX, other salinity proxies,~~

~~including the ACE index (Turich and Freeman, 2011) and IR$_{6+7Me}$ (Wang et al., 2021), do not show significant correlations with salinity in the Seine River basin (p>0.05, Wilcoxon test; Fig. S9). (Turich and Freeman, 2011; Wang et al., 2021)~~Indeed, ~~A~~while~~although~~ ACE has been successfully applied in hypersaline systems (Turich and Freeman, 2011), it performs less effectively in certain saline settings due to the complex sources of archaeol and GDGTs (Huguet et al., 2015) and/or distinct ionization efficiencies between these compounds (He et al., 2020; Wang et al., 2021). Similarly, ~~I~~the IR$_{6+7Me}$ may be influenced by the preferential production of 6-methyl brGDGTs related with nitrogen nutrient loadings in a specific region of the estuary (KP 255.6-337), as discussed in 4.1.2. Consequently, only the -RIX successfully tracks salinity variations in this basin, while ACE and IR$_{6+7Me}$ show relative insensitivity. However, quantitatively reconstructing salinity with RIX is an important step forward that warrants further investigation. This ~~, which~~ requires comparing brGMGT distributions~~s~~ from various aquatic environments (e.g. estuaries and lakes) across salinity gradients~~ for a salinity reconstruction~~.

~~Moreover, t~~The RIX ~~is~~was calculated for the SPM and sediment samples from the Seine River basin, showing an obvious decreasing trend from upstream to downstream (Fig. 8f). The RIX in river (0.51±0.06, SPM) and upstream estuarine (0.40±0.07, SPM and sediments) samples is significantly higher than for downstream estuarine (0.23±0.06, SPM and sediments) samples. RIX values around 0.50 could therefore be considered reflecting the riverine endmember, while those below 0.30 could represent the saltwater endmember.

A significant overlap between brGMGT distribution soils and downstream samples was observed (Fig. 4b). This suggests that a portion of the brGMGT signal in the water masses of the Seine may be partially derived from surrounding soils. Hence, RIX was also calculated for the soil samples. The RIX values of the soil samples were 0.21±0.13, close to those of the downstream estuarine samples. A large variance in the soil brGMGT concentration was observed (Fig. S5b), suggesting that further investigation is needed to better understand the environmental controls on the brGMGT production in soils. As with brGDGTs, we applied a random forest algorithm to distinguish brGMGT distributions between downstream estuary and soil samples. This trained model accurately distinguishes soils from downstream estuarine samples (Fig. S12~~1~~), indicating in situ production of brGMGTs in the downstream estuary. Given the significantly low brGMGT concentrations in soils (p<0.05, Wilcoxon test; Fig. S5b) and the distinct distributions between brGMGT in soils and aquatic settings identified through PCA (Fig. 4) and machine learning (Fig. S12~~1~~), it can be assumed that the impact of soil-derived brGMGTs on the observed RIX signal in the water column of the Seine basin is low.


(a)
R=-0.33, p>0.05
R=-0.78, p<0.05
H1020c(%)
salinity
(b)
R=-0.51, p<0.05
R=-0.79, p<0.05
H1034b(%)
salinity
(c)
R=0.54, p<0.05
R=0.77, p<0.05
H1020a (%)
salinity
(d)
R=0.39, p>0.05
R=0.86, p<0.05
H1020b (%)
salinity
Low-flow season
High-flow season
(e)
R=-0.58, p<0.05
R=-0.88, p<0.05
RIX
salinity
(f)
Upstream (n=56)
River (n=9)
Downstream (n=121)
seaward
RIX
site
(g)
R=-0.83, p<0.05
R=-0.7, p<0.05
δ13C (‰)
RIX
(h)
R=0.82, p<0.05
R=0.68, p<0.05
BIT
RIX
(i)
RIX
KP (km)
(j)
BIT
KP (km)
(k)
δ13C (‰)
KP (km)
seaward
Season
Low-flow season
High-flow season

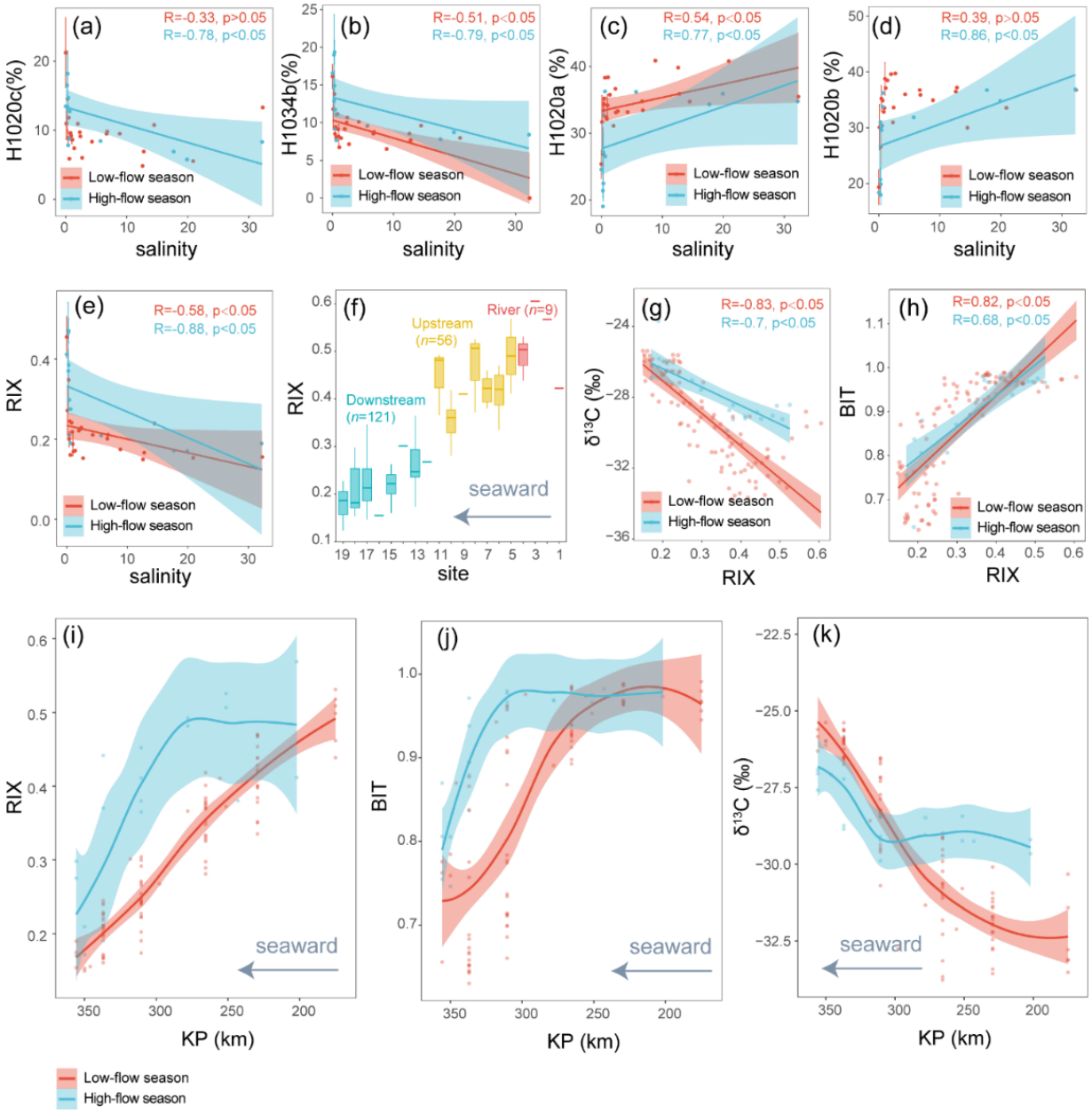


Figure ~~8~~9: (a-e) The correlation between salinity and the relative abundance of brGMGTs and RIX, analyzed through linear regression.~~Salinity plotted versus relative abundance of brGMGTs and RIX.~~ Shaded area ~~represent~~represents 95% confidence intervals. Vertical error bars indicate mean ± s.d for samples with the same salinity. Dataset is composed of SPM. (f) Distribution of RIX across the Seine River basin. ~~Boxes are color-coded based on the sample type (river in red, upstream estuary in yellow, and downstream estuary in blue).~~ 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. Dataset is composed of SPM and sediments. (g-h) RIX plotted versus $\delta^{13}C$ and BIT through the linear regression. Shaded area ~~represent~~represents 95% confidence intervals. (i-k) Spatio-temporal variations of RIX and several other terrestrial proxies, including BIT and $\delta^{13}C$ (‰). The trends showing spatio-temporal variations were based on locally estimated scatterplot smoothing (LOESS) method with

95% confidence intervals. KP (kilometric point) represents the distance in kilometers from the city of Paris (KP 0). Dataset is composed of SPM.

~~The shift in brGMGT distribution observed across the Seine River basin (Fig. 5) could be due to a change in the diversity of brGMGT-producing bacteria and/or to an adaptation of these microorganisms to environmental changes occurring from upstream to downstream. The latter hypothesis seems unlikely, as a physiological adaptation of a given bacterial community would make it difficult to explain why the relative abundance of three isomers of compound H1020, which share a similar structure, varies differently in response to salinity changes. Hence, a shift in brGMGT-producing bacterial communities across the basin is more likely. Compounds H1020c and H1034b could predominantly be produced by bacteria preferentially growing in freshwater, and homologues H1020a and H1020b by bacteria preferentially living in brackish or saltwater.~~

~~4.3 Potential implications for brGMGTs as a proxy for riverine runoff in modern systems~~

~~The distinct brGMGT distributions in freshwater and saltwater could be used to trace the Organic Matter (OM) produced upstream all along the Seine basin. To trace such a riverine runoff signal, we propose a new proxy, the Riverine IndeX (RIX), based on the fractional abundances of brGMGTs H1020c and H1034b versus H1020a and H1020b (Eq. 3):~~

$$\text{RIX} = \frac{H1020c + H1034b}{H1020c + H1034b + H1020a + H1020b} \quad (3)$$

~~The rationale for RIX as a riverine runoff signal is that in freshwater environments, the pool of brGMGTs is dominated by H1020c and H1034b, whereas H1020a and H1020b prevail in saltwater environments. This is further supported by the significant negative correlation between RIX and salinity (Fig. 9e). (Turich and Freeman, 2011)(Wang et al., 2021)(Huguet et al., 2015)(He et al., 2020; Wang et al., 2021)The RIX is calculated for the SPM and sediment samples from the Seine River basin, showing an obvious decreasing trend from upstream to downstream (Fig. 9f). The RIX in river (0.51±0.06, SPM) and upstream estuarine (0.40±0.07, SPM and sediments) samples is significantly higher than for downstream estuarine (0.23±0.06, SPM and sediments) samples. RIX values around 0.50 could therefore be considered reflecting the riverine endmember, while those below 0.30 could represent the saltwater endmember.~~

~~A significant overlap between brGMGT distribution soils and downstream samples was observed (Fig. 45b). This suggests that part a portion of the brGMGT signal in the water masses of the Seine may be partially derived from surrounding soils. Hence, RIX was also calculated for the soil samples. The RIX values of the soil samples were 0.21±0.13, close to those of the downstream estuarine samples. Notably, aA large variance in the soil brGMGT concentration was observed (Fig. S5b), which indicatessuggesting that further investigation is needed to better understand the environmental controls on the brGMGT distributionsproduction in soils. As with brGDGTs, we applied a random forest algorithm to~~

~~distinguish brGMGT distributions between downstream estuary and soil samples. This trained model accurately distinguishes soils from downstream estuarine samples (Fig. S11), indicating in situ production of brGMGTs in the downstream estuary. Given the significantly low brGMGT concentrations in soils (p<0.05, Wilcoxon test; Fig. S5b) and the distinct distributions between brGMGT in soils and aquatic settings identified through PCA (Fig. 4) and machine learning (Fig. S11), it can be assumed that the impact of soil-derived brGMGTs on the observed RIX signal in the water column of the Seine basin is low.~~~~However, the average concentrations of brGMGTs are an order of magnitude lower in the soils than in the river channel sediments and SPM samples of the Seine basin (Fig. S4b). Therefore, it can be assumed that the impact of soil-derived brGMGTs on the observed RIX signal in the water column of the Seine basin is low.~~

In order to test the general applicability of the RIX, it was then applied to soils, riverine and marine samples (SPM and sediments) collected in the Godavari River basin and Bay of Bengal (Kirkels et al., 2022a). This site represents the only other river-sea continuum besides the Seine basin for which brGMGT data are presently available. Significant differences in RIX between the soils, SPM and sediment samples from the Godavari River basin are observed (p<0.05, Wilcoxon test; Fig. 9~~10~~). RIX values in soils (0.49±0.16) around the Godavari River basin are significantly lower ~~compared to~~than those the river samples (p<0.05, Wilcoxon test; Fig. 9). Therefore, the potential soil contribution~~s~~ would dilute the riverine brGMGT signal, further decreasing the RIX in marine sediments. This is consistent with the observations in the Seine River basin. However, given the distinct distributions between soil and aquatic samples and the lower brGMGT concentration in soils (Kirkels et al., 2022a), the influence of soil-derived brGMGTs on riverine RIX value~~s~~ may be limited. In addition, 96% of the RIX values in riverine SPM and riverbed sediments from the Godavari basin exceed 0.5, whereas all of the RIX values observed in marine sediments from the Bay of Bengal are below 0.3. This suggests that the RIX cutoff values defined using the samples from the Seine basin may be broadly applicable and valid across other river-sea continuums. This deserves further studies.

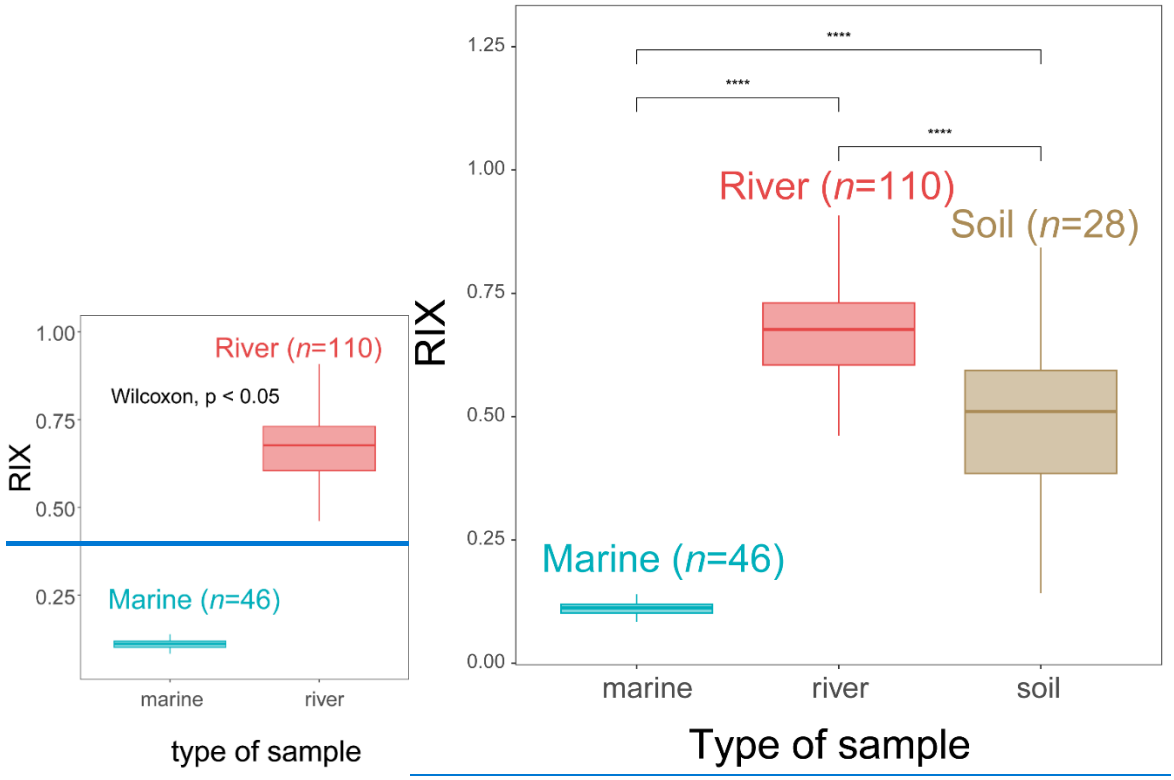


Figure 9~~10~~: RIX in the soils, SPM and sediment samples from Godavari River basin (India) and Bay of Bengal sediments (data from Kirkels et al. (2022a)). Statistical testing was performed by a Wilcoxon test. Boxes show the upper and lower quartiles of the data, and whiskers show the range of the data, which are color-coded based on the sample type (river in red~~ and~~, marine in blue, and soil in brown). The center-line in the boxes indicates the median value of the dataset.

Further confirmation of the RIX potential as a tracer of riverine OM comes from the significant correlations observed between this index and other commonly used proxies for tracing OM sources, i.e. the BIT and $\delta^{13}C_{org}$ (p<0.05, Wilcoxon test; Fig. 8~~9~~, g-h). These proxies show roughly similar spatial and temporal variations in the Seine River basin. It is worth noting that another terrestrial proxy (C/N) was not included because it may be ineffective in tracing terrestrial OM in this anthropogenic estuary. ~~The C/N ratio is commonly used as a bulk terrestrial indicator, with higher values indicating a greater terrestrial contribution than marine sources (Bianchi and Canuel, 2011). However, other parameters such as decomposition processes, remineralization, and distinct sources could complicate its applications (Lamb et al., 2006). In the Seine River basin, C/N values were significantly lower in the river and upstream estuary than in downstream samples (Fig. 2). Given that anthropogenic OM contains more nitrogen than natural OM, an increase in anthropogenic sources would result in a decrease in C/N values (Van Den Hende et al., 2011; Liu et al., 2020). As a result, the observed decrease in C/N values in river and upstream estuarine samples could be attributed to a higher contribution of nitrogen from anthropogenic sources, possibly due~~

~~to intense agricultural activities as discussed in 4.1.2. (Liu et al., 2020)As BIT, RIX, and $\delta^{13}C_{org}$ provide similar information, they may be more reliable terrestrial tracers compared to C/N in the Seine River basin.fu~~

In the low-flow season, RIX and BIT gradually decrease while $\delta^{13}C_{org}$ increase across the basin (Fig. 8~~9~~, i-k). Such trends during the low discharge periods likely reflects the continuous dilution process of riverine OM caused by the mixing of fresh and marine water masses (Thibault et al., 2019). The gradual dilution of the riverine OM signal along the Seine River basin could be due to the increase of seawater intrusion, and thus marine-derived OM, at low discharges (Ralston and Geyer, 2019; Kolb et al., 2022). In contrast, during the high-flow season, no such gradual dilution trend is observed. Instead, at high discharges, the RIX, BIT and $\delta^{13}C_{org}$ remain roughly stable from KP 202 to 310.5, before, steeply decreasing for BIT index and RIX, and increasing for $\delta^{13}C_{org}$. This trend can be explained by the fact that at high flow rates, the limit of saltwater intrusion in the estuary shifts seawards rather than landwards, allowing the riverine OM to be flushed further downstream than under low discharge conditions. After this region, the riverine OM is diluted because of the mixing with marine water masses, as observed during the low-flow season. The trends observed in the Seine Estuary are consistent with previous studies in other regions, showing that terrestrial OM was only effectively transported downstream at high flow rates (Kirkels et al., 2020, 2022b).

Although the BIT is successfully used in the Seine River basin as well as in previous studies to trace riverine (terrestrial) OM inputs (Hopmans et al., 2004; Xu et al., 2020), this index can be biased by *in situ* production of brGDGTs in the water column and/or sediments (Sinninghe Damsté, 2016; Dearing Crampton-Flood et al., 2019) and selective degradation of crenarchaeol *vs.* brGDGTs (Smith et al., 2012). Hence, high BIT values do not necessarily indicate higher contribution of terrestrial OM in some settings (Smith et al., 2012). Unlike the BIT index, based on two different families of compounds (isoGDGTs and brGDGTs), the RIX is based on 4 compounds from the same family (brGMGTs) that likely have similar degradation rates and therefore not influenced by selective degradation. Furthermore, the RIX is based on the relative abundances of abundant brGMGTs which are all predominantly produced in aquatic settings, two of them (H1020c and H1034b) being mainly produced in freshwater and two of them (H1020a and H1020b) mainly in saltwater. Therefore, the RIX is based on compounds which are more specifically produced in the two endmembers (freshwater or saltwater), which could avoid the biases encountered with the BIT. Overall, our work shows that, in addition to the BIT and $\delta^{13}C_{org}$, the RIX successfully captures the spatio-temporal dynamics of riverine OM in the Seine River basin, making this proxy a promising and complementary one tracing riverine runoff in modern samples.

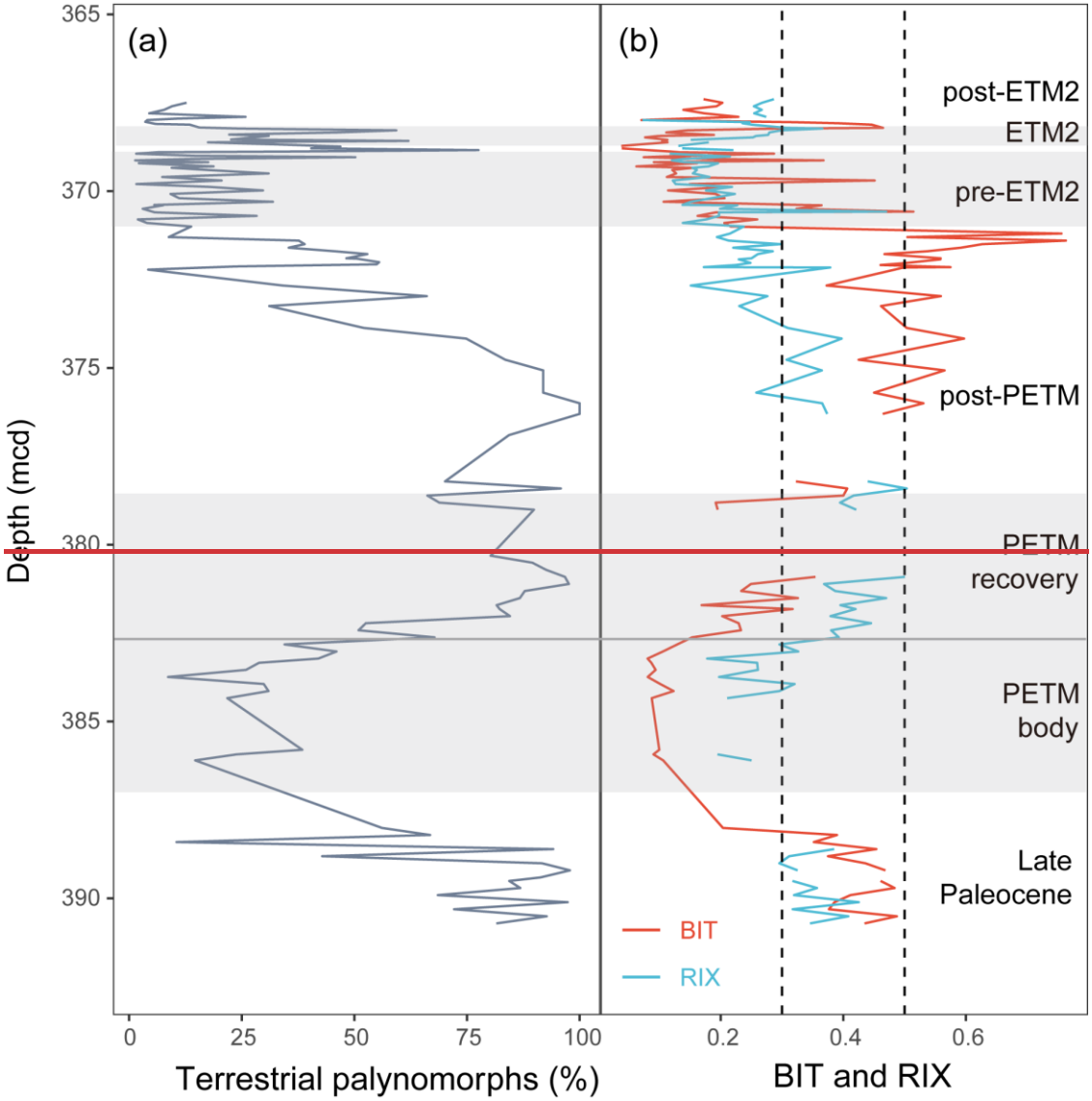
4.4 Application of the RIX to a paleorecord across the upper Paleocene and lower Eocene

To further test the applicability of the RIX as a riverine proxy in paleorecords, we calculated RIX and compared this proxy with BIT using the published brGMGT and GDGT data (Sluijs et al., 2020) from IODP Expedition 302 Hole 4A (cf. location in Fig. S14~~3~~) across the upper Paleocene and lower Eocene. This core is considered to record significant changes in terrestrial

inputs (i.e. continental spores and pollen) due to sea level changes over time ~~This core is considered proximal to the coast and has considerable changes in terrestrial inputs (i.e. continental spores and pollen) over time~~ (Sluijs et al., 2009, 2006), making it a suitable paleorecord for testing runoff proxies.

In the late Palaeocene, the relative abundance of terrestrial palynomorphs (spores and pollen) remains at high levels (Fig. 10~~1~~a), indicating enhanced terrestrial inputs during this period. This is consistent with the presence of abundant amorphous organic matter in this interval that is presumed to have originated from terrestrial sources (Sluijs et al., 2006). Enhanced freshwater contribution is also successfully captured by RIX and BIT (Fig. 10~~1~~b), with values exceeding 0.3 for both proxies.

Palynomorph assemblages from the body of Paleocene-Eocene Thermal Maximum (PETM) are characterized by substantially lower terrestrial palynomorphs (Fig. 10~~1~~a). This indicates a relative decrease of riverine runoff, which is also evidenced by a drop of both RIX and BIT indices (Fig. 10~~1~~b). Such decreased runoff during the PETM body was previously attributed to a rise in sea level (Sluijs et al., 2006), which has been recorded in many other sites worldwide (Speijer and Morsi, 2002; Harding et al., 2011; Sluijs et al., 2014).~~Such decreased runoff during the PETM body was previously attributed to a local sea level rise, resulting in a greater distance from the shoreline (Sluijs et al., 2006; Speijer and Morsi, 2002).~~ During the PETM recovery, increased runoff is reflected by the gradual increase in relative abundance of terrestrial palynomorphs (Fig. 10~~1~~b), which was interpreted as a consequence of a drop in sea level (Sluijs et al., 2006). Higher freshwater influx during this period is also indicated by increases in RIX and BIT (Fig. 10~~1~~b).


(a)
(b)
Depth (mcd)
365
370
375
380
385
390
post-ETM2
ETM2
pre-ETM2
post-PETM
PETM
recovery
PETM
body
Late
Paleocene
BIT
RIX
0
25
50
75
100
0.2
0.4
0.6
Terrestrial palynomorphs (%)
BIT and RIX

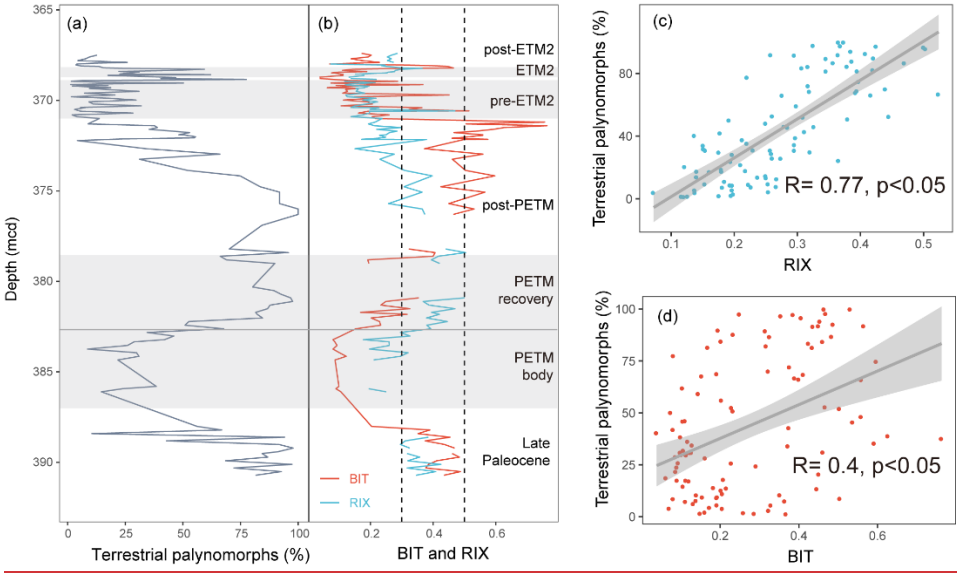


Figure 10~~1~~: Comparison between (a) terrestrial palynomorphs (%) and (b) BIT and RIX across the upper Paleocene and lower Eocene between 391 and 367 meters composite depth below sea floor (mcd) of IODP Expedition 302 Hole 4A. Terrestrial palynomorphs data are from Sluijs et al. (2006) and Sluijs et al. (2009). RIX and BIT were calculated using data from Sluijs et al. (2020). Grey shading represents Eocene Thermal Maximum 2 (ETM2), pre-ETM2 interval, and Paleocene-Eocene Thermal Maximum (PETM). Dotted line represents cutoff values of RIX (below 0.3 for marine contribution and above 0.5 for riverine contribution). Linear regression of the RIX (c) and BIT (d) ~~to~~against the terrestrial palynomorphs. Shaded area represents 95% confidence intervals.

The relative abundance of terrestrial palynomorphs decreases after the termination of the PETM (Fig. ~~11a~~10a). Similar decreasing trends are also evident for RIX during post-PETM (Fig. ~~11b~~10b). However, BIT levels remain high until the pre-Eocene Thermal Maximum 2 (ETM2), showing distinct patterns compared to terrestrial palynomorphs and RIX. One possible explanation for this distinction could be the sedimentary *in situ* production of brGDGTs (Peterse et al., 2009). Indeed, several studies have indicated that elevated BIT values may not accurately represent high levels of terrestrial inputs in marine environments (Smith et al., 2012). In contrast, the similar trends observed between RIX and terrestrial palynomorphs highlight the reliability of RIX as a valuable complementary runoff proxy (Fig. ~~11b~~10b).

At the onset of the pre-ETM2, there is a significant decrease in terrestrial palynomorphs (Fig. ~~11a~~10a). Meanwhile, there is a sharp increase in the proportions of normal marine dinocysts, which was interpreted as a transgressive signal (Sluijs et al., 2008; Willard et al., 2019; Sluijs et al., 2009). Throughout the pre-ETM2 interval, the relative abundance of terrestrial palynomorphs remains consistently low (Fig. 10~~1~~a). Additionally, the dinocyst assemblages suggest normal marine conditions for this period (Sluijs et al., 2009). These normal marine conditions are also well documented by RIX (Fig. 10~~1~~b), as most of the samples demonstrate low values (below 0.3). In contrast, BIT values exhibit some fluctuation, with several samples

displaying high values (Fig. 10~~1~~b). This variability in BIT values could also be attributed to the sedimentary *in situ* production of brGDGTs.

During the ETM2 interval, increase of terrestrial palynomorphs suggests an increased runoff from the continent to the site (Fig. 10~~1~~a). Enhanced runoff signal during ETM2 is also supported by the dominance of low-salinity dinocyst taxa and the presence of massive amorphous organic matter (Sluijs et al., 2009; Willard et al., 2019). Both RIX and BIT show increasing trends for this interval (Fig. 10~~1~~b), indicating that both indices are reflecting a runoff signal during this period.

Following the ETM2, there is a decline in the relative abundance of terrestrial palynomorphs (Fig. 10~~1~~a), which indicates a shift toward normal marine conditions. This shift is also supported by the dominance of normal marine dinocysts and low concentrations of massive amorphous organic matter (Sluijs et al., 2009; Willard et al., 2019). Additionally, this shift towards normal marine conditions is in line with the lower values (below 0.3) observed for both RIX and BIT (Fig. 10~~1~~b).

Overall, RIX and BIT exhibit similar trends with terrestrial palynomorphs in the late Palaeocene, PETM, ETM2, and post-ETM2 periods. Both lipid proxies are likely reliable indicators of riverine runoff for these intervals. However, distinctions between RIX and BIT become more apparent especially in the post-PETM and pre-ETM2 periods when normal marine conditions prevail, and sedimentary *in situ* production of brGDGTs may occur, resulting in high BIT values. RIX proves to be particularly valuable for these intervals, avoiding the possible biases associated with BIT. This indicates that RIX performs better in this core compared with BIT, which is further supported by a higher correlation coefficient observed between RIX and terrestrial palynomorphs (0.77; Fig. 10c) compared with BIT and terrestrial palynomorphs (0.4; Fig. 10d).

5 Conclusions

In this study, the brGDGT and brGMGT concentrations and distributions in soils, SPM and sediments (n=237) across the Seine River basin were investigated. Higher concentrations and distinct distributions of brGDGTs and brGMGTs in SPM and sediments compared with soils imply that both types of compounds can be produced *in situ* in aquatic settings. The distribution of both brGDGTs and brGMGTs are largely related to salinity, but only brGDGT distributions are significantly influenced by nitrogen nutrient loadings. In addition, covariations of IR_{6Me}, TN, $\delta^{15}N$, and Chl *a* concentration within the low salinity region suggest that riverine (6-methyl) brGDGT production is favored by low-salinity and high-productivity conditions.

In the Seine River basin, salinity correlates positively with H1020a and H1020b, and negatively with H1020c and H1034b. This indicates that compounds H1020c and H1034b could be produced by bacteria that preferentially grow in freshwater, while homologues H1020a and H1020b could be produced by bacteria that mainly live in saltwater. Based on this, a novel proxy, the Riverine IndeX (RIX) is proposed to trace riverine OM input. The average value of RIX for the riverine samples is 0.51,

which is much higher than that in downstream estuarine (0.23 on average) samples. This suggests that RIX values over 0.5 imply considerable riverine contributions, whereas RIX values below 0.3 indicate higher marine contributions. This cutoff value defined in the Seine River basin also works in the Godavari River basin (India), which implies that this novel proxy based on brGMGTs may be broadly applicable in modern samples. Finally, RIX was applied to a paleorecord across the upper Paleocene and lower Eocene and showed similar trends with terrestrial palynomorphs. Altogether, this showed the potential of the RIX as a riverine runoff proxy in modern and deep times. Additional studies are needed to test its general applicability in modern samples and paleorecords.

Supplement

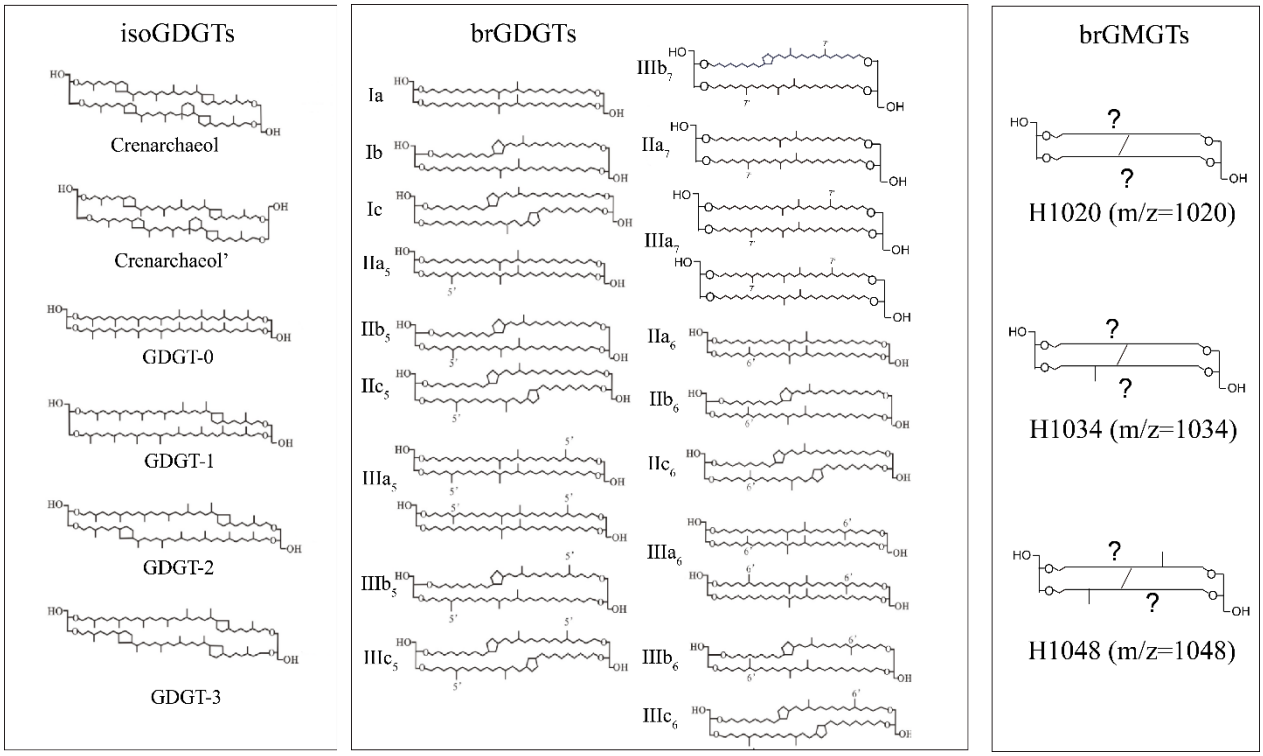


Fig. S1. Structures of isoGDGTs, brGDGTs, and brGMGTs. Note that the structures of brGMGTs and compounds eluting later than 7-Methyl brGDGTs (1050d and 1036d) have not been described.

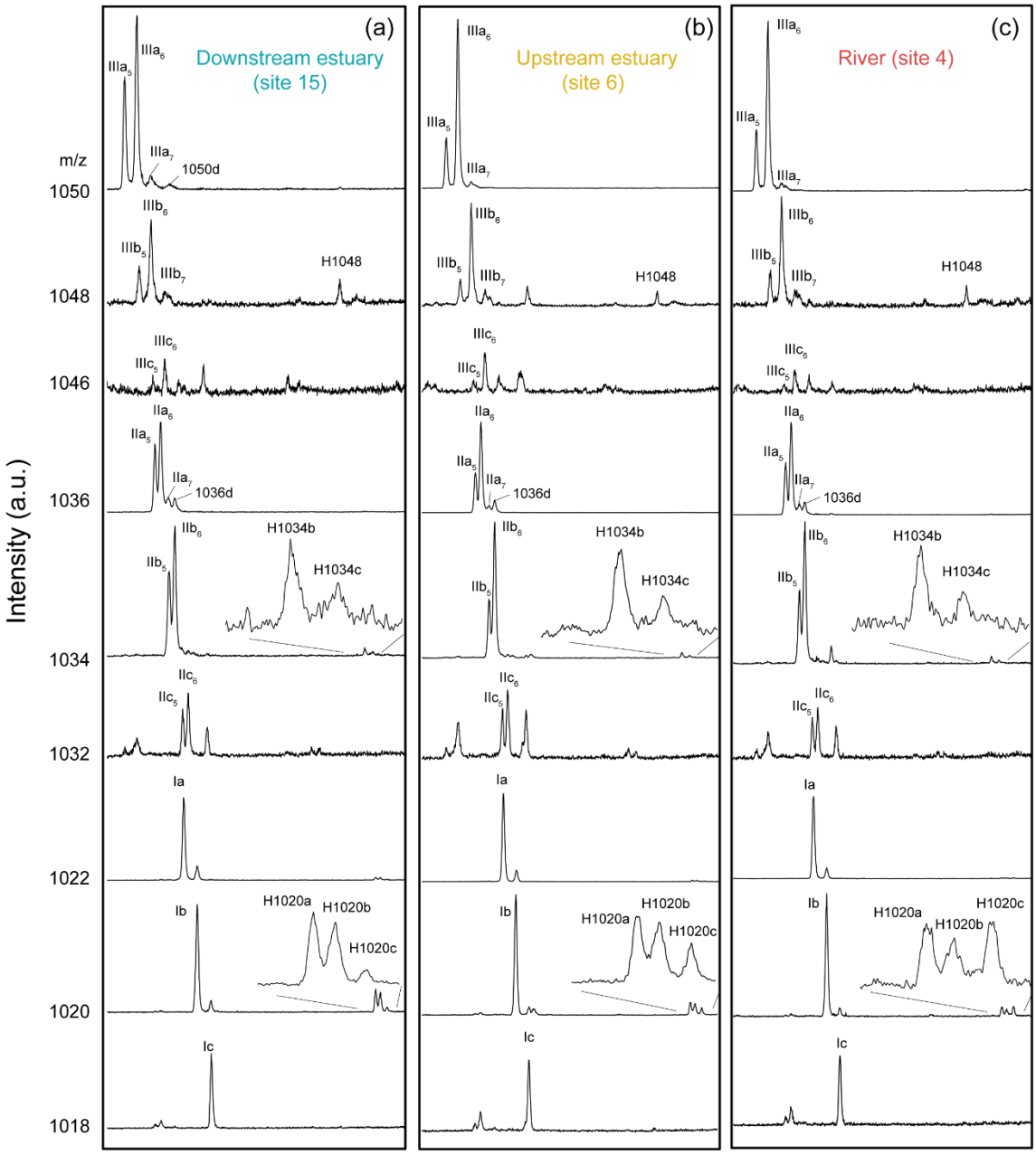


Fig.~~ure~~ S2~~3~~.~~:~~ Extracted chromatograms of brGDGTs and brGMGTs for the SPM samples collected in (a) site 15 (Tancarville, September 2020), (b) site 6 (Oissel, July 2019) and (c) site 4 (Les Andelys, July 2019). The nomenclature for the penta- and hexamethylated brGDGTs: 5-methyl brGDGTs (IIIa$_5$, 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.

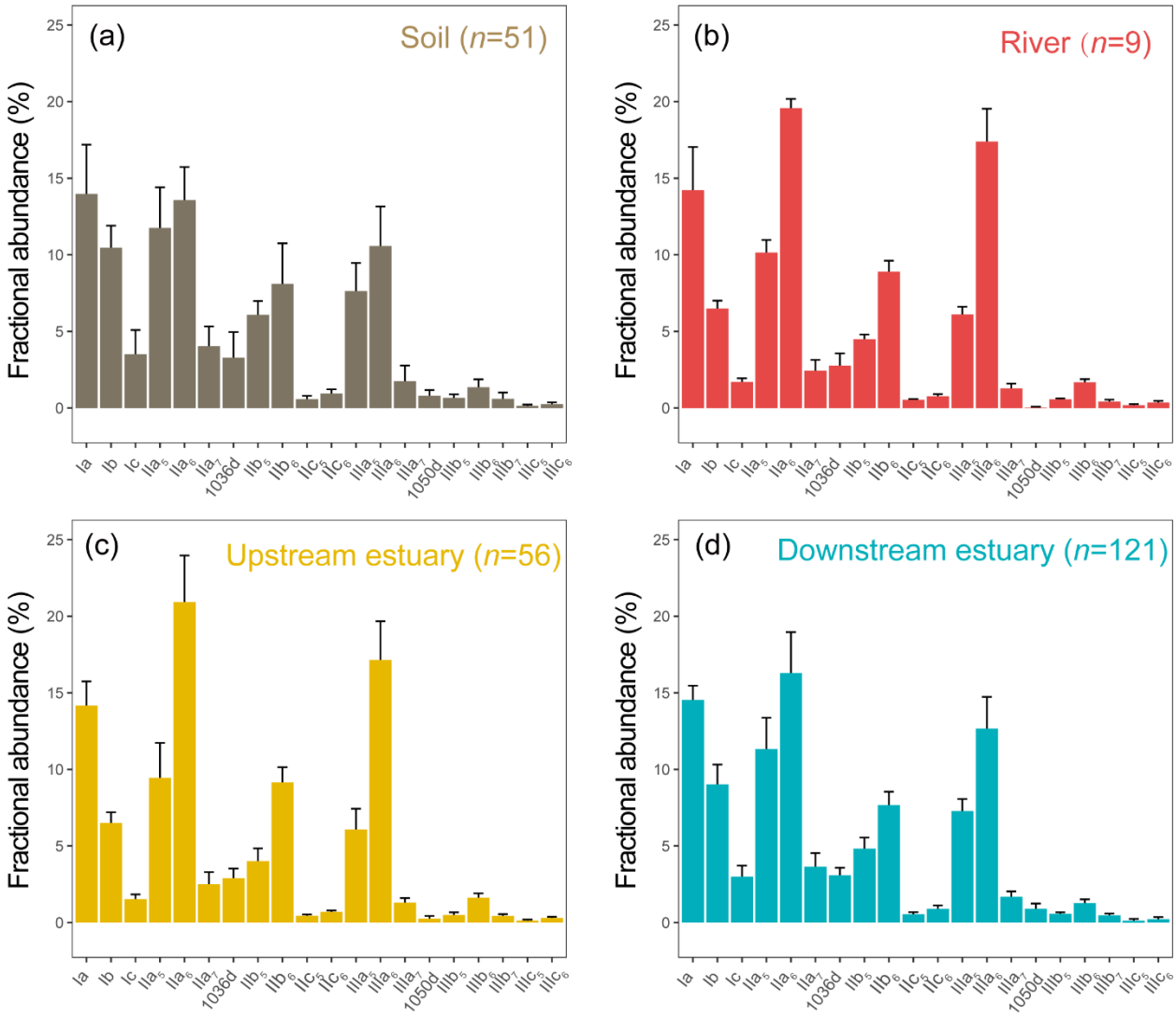
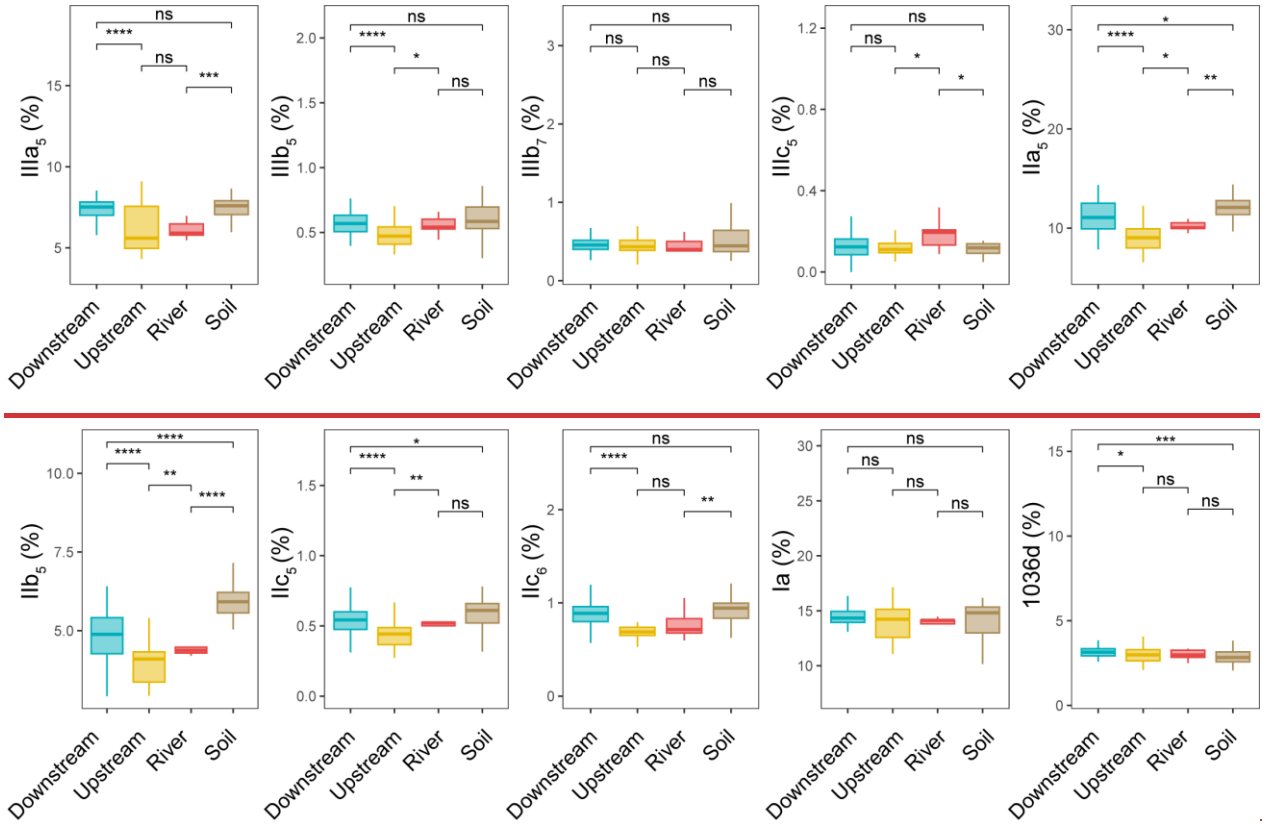


Fig. S32. Distribution of brGDGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (*n*=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin.


IIIa5 (%)
IIIb5 (%)
IIIb7 (%)
IIIc5 (%)
IIa5 (%)
IIb5 (%)
IIc5 (%)
IIc6 (%)
Ia (%)
1036d (%)
Downstream
Upstream
River
Soil
ns
*
**
***
****

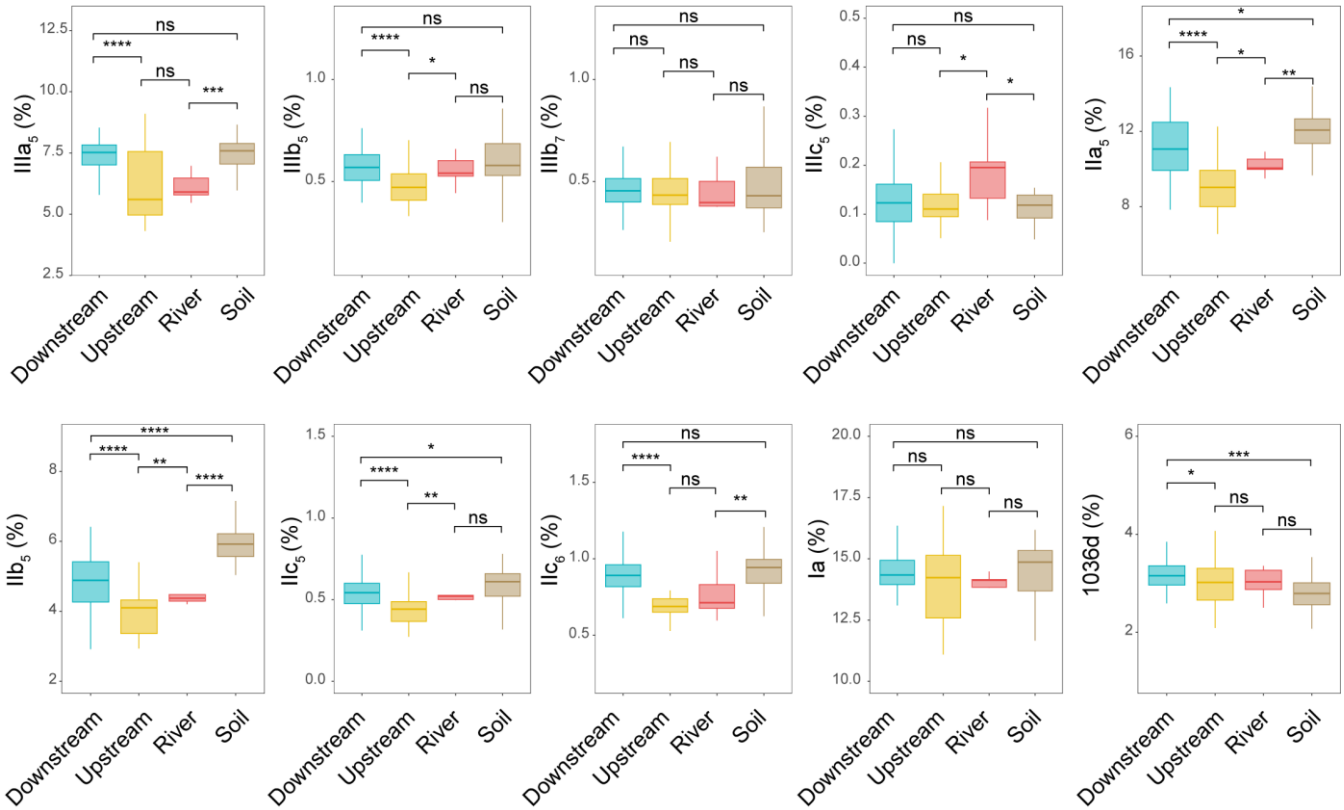


Fig. S~~4~~3. Relative abundance of individual brGDGTs over 20 brGDGTs (IIIa$_5$, IIIb$_5$, IIIc$_5$, IIa$_5$, IIb$_5$, IIc$_5$, IIIa$_6$, IIIb$_6$, IIIc$_6$, IIa$_6$, IIb$_6$, IIc$_6$, IIIa$_7$, IIIb$_7$, IIa$_7$, Ia, Ib, Ic, 1050d, and 1036d) across the Seine River basin. Box plots of upstream and 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, $p > 0.05$).

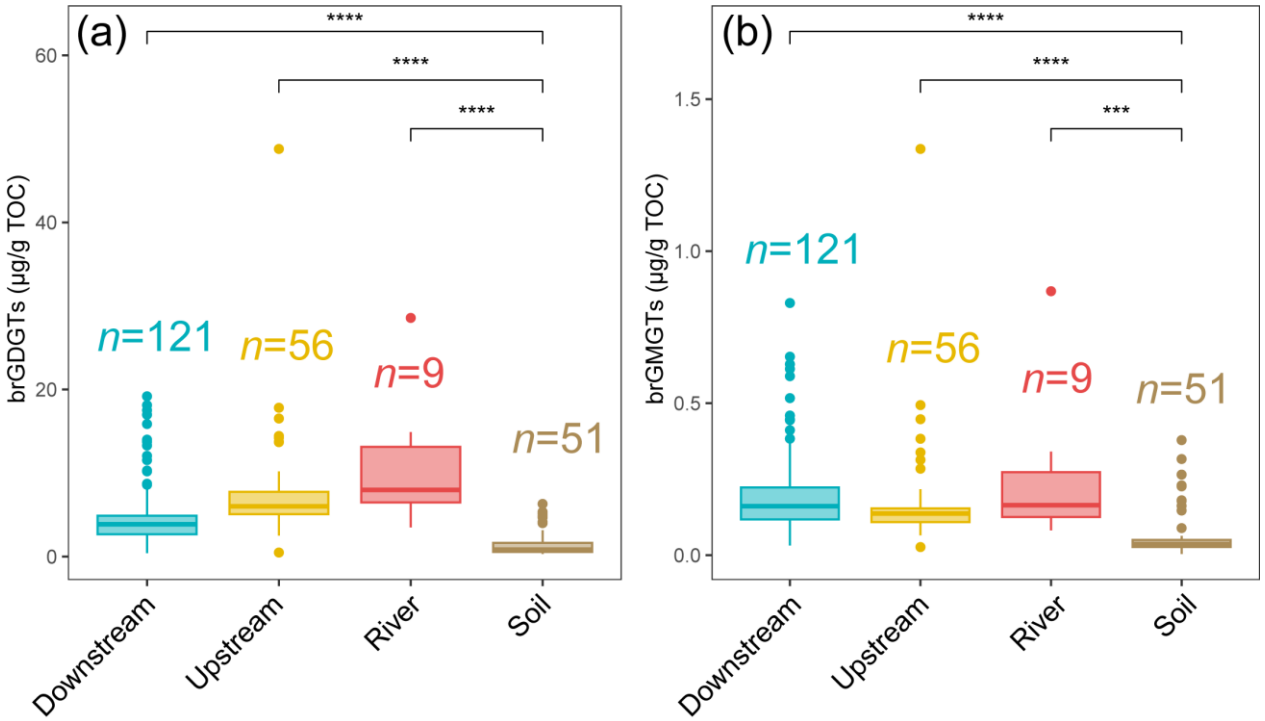


Fig. S54. Concentrations (normalized to total organic carbon) of (a) total brGDGTs and (b) total brGMGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin. Box plots of upstream 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, $p > 0.05$).

975

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

980

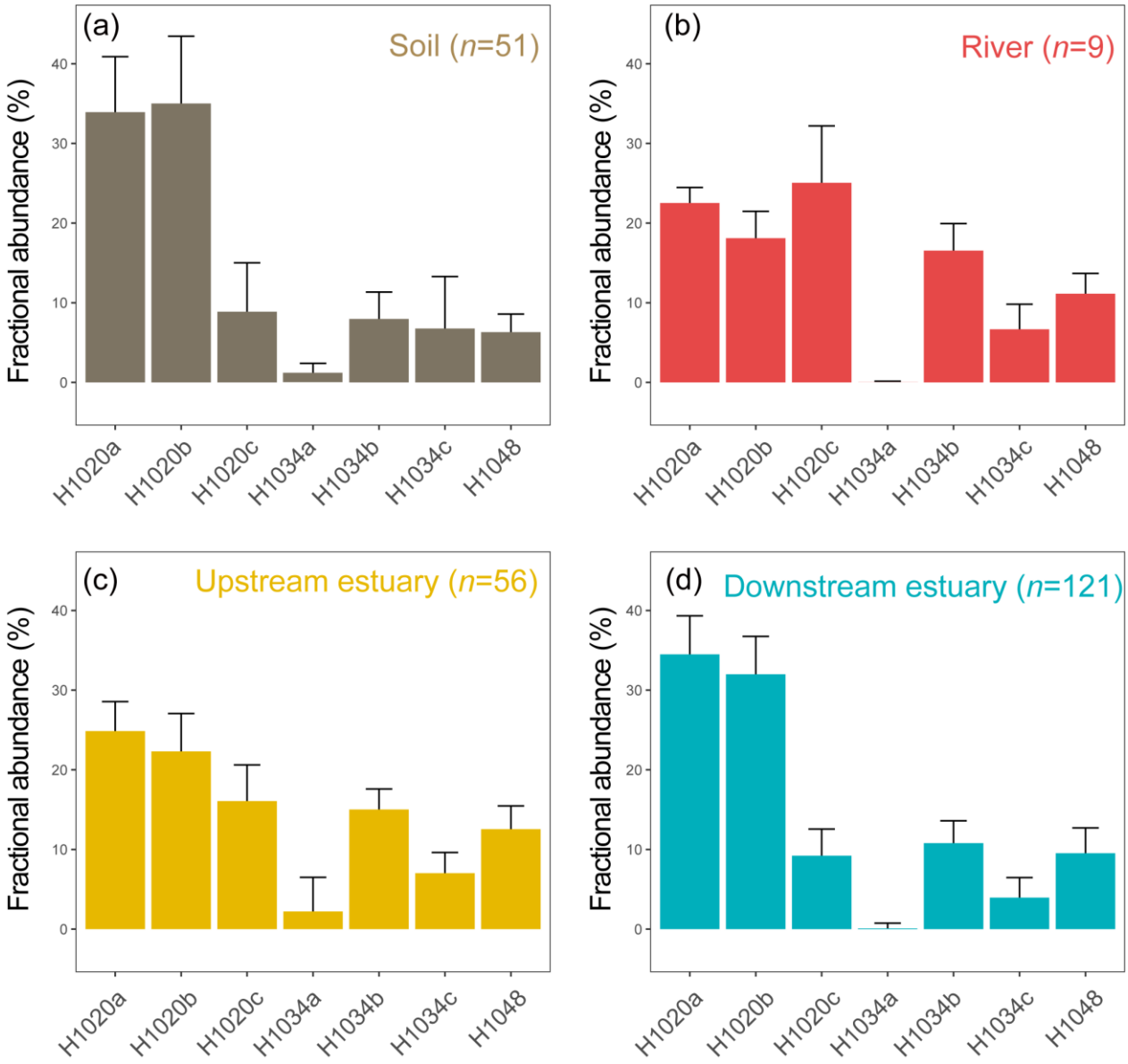


Fig. S65. Distribution of brGMGTs from soils (surficial soils and mudflat sediments, *n*=51) as well as river (n=9), upstream estuary (*n*=56) and downstream estuary (*n*=121) samples across the Seine River basin.

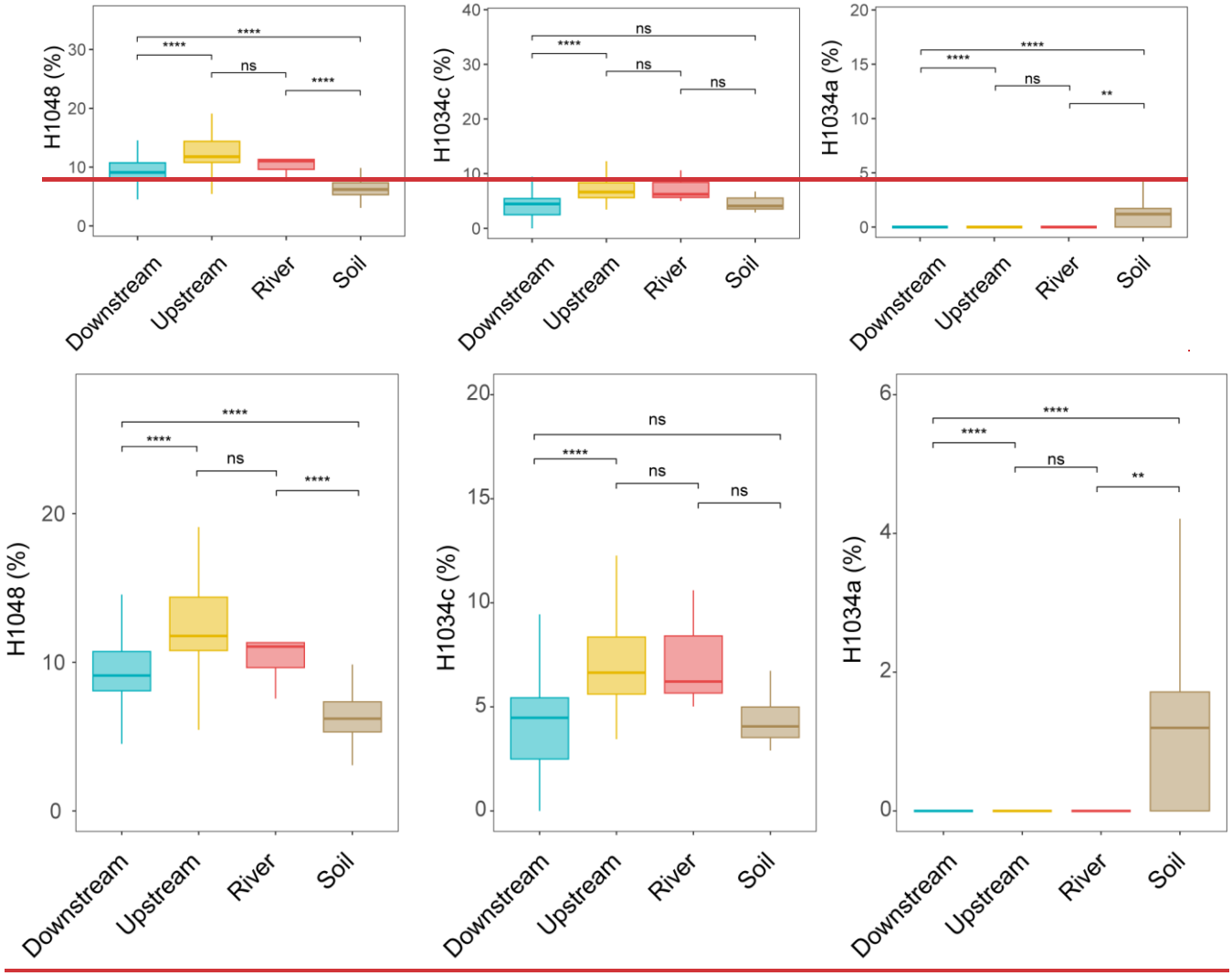


Fig. S76. Relative abundance of brGMGTs over 7 brGMGTs (H1020a, H1020b, H1020c, H1034a, H1034b, H1034c, and H1048) across the Seine River basin. Box plots of upstream and downstream estuary are composed of SPM and sediments, whereas those of river are composed of SPM. Boxes show the upper and lower quartiles 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, $p > 0.05$).

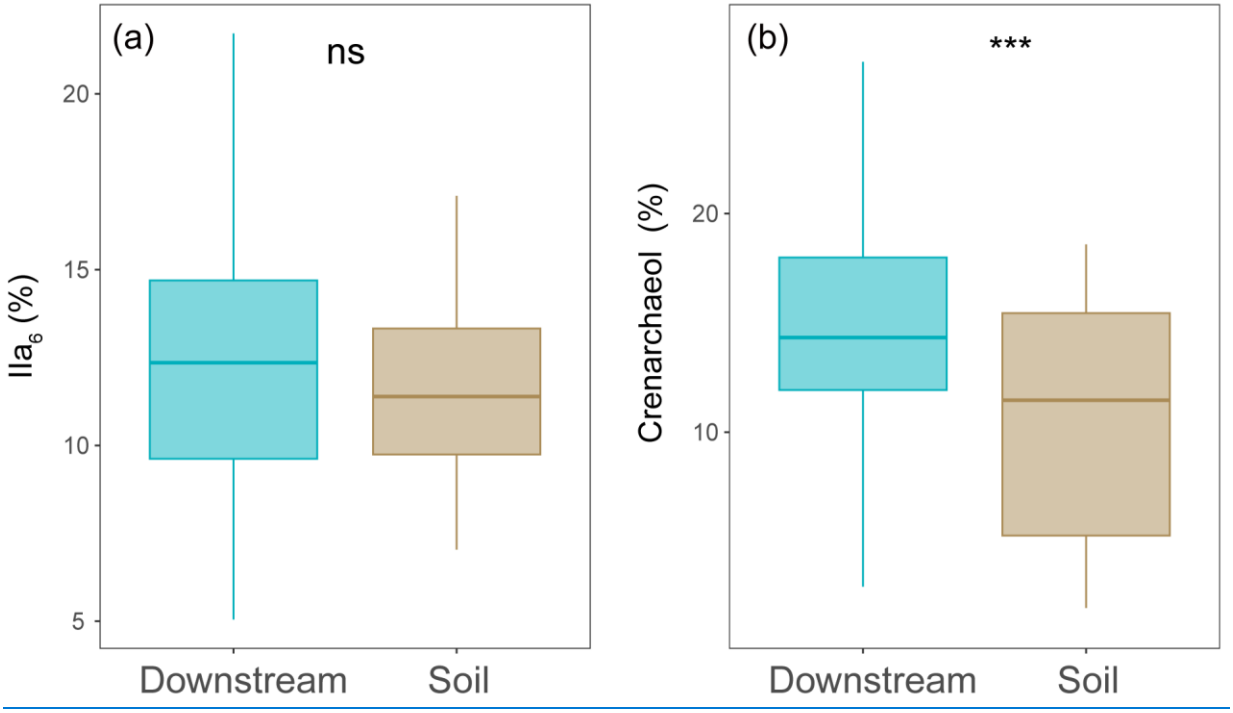


Fig. S8. Relative abundance of IIa$_6$ (a) and crenarchaeol (b) over 19 GDGTs (GDGT-0, GDGT-1, GDGT-2, GDGT-3, Crenarchaeol, Crenarchaeol', IIIa$_5$, IIIa$_6$, IIIb$_5$, IIIb$_6$, IIa$_5$, IIa$_6$, IIb$_5$, IIb$_6$, IIc$_5$, IIc$_6$, Ia, Ib, and Ic) used in the BigMac model. Boxes show the upper and 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, $p > 0.05$).**

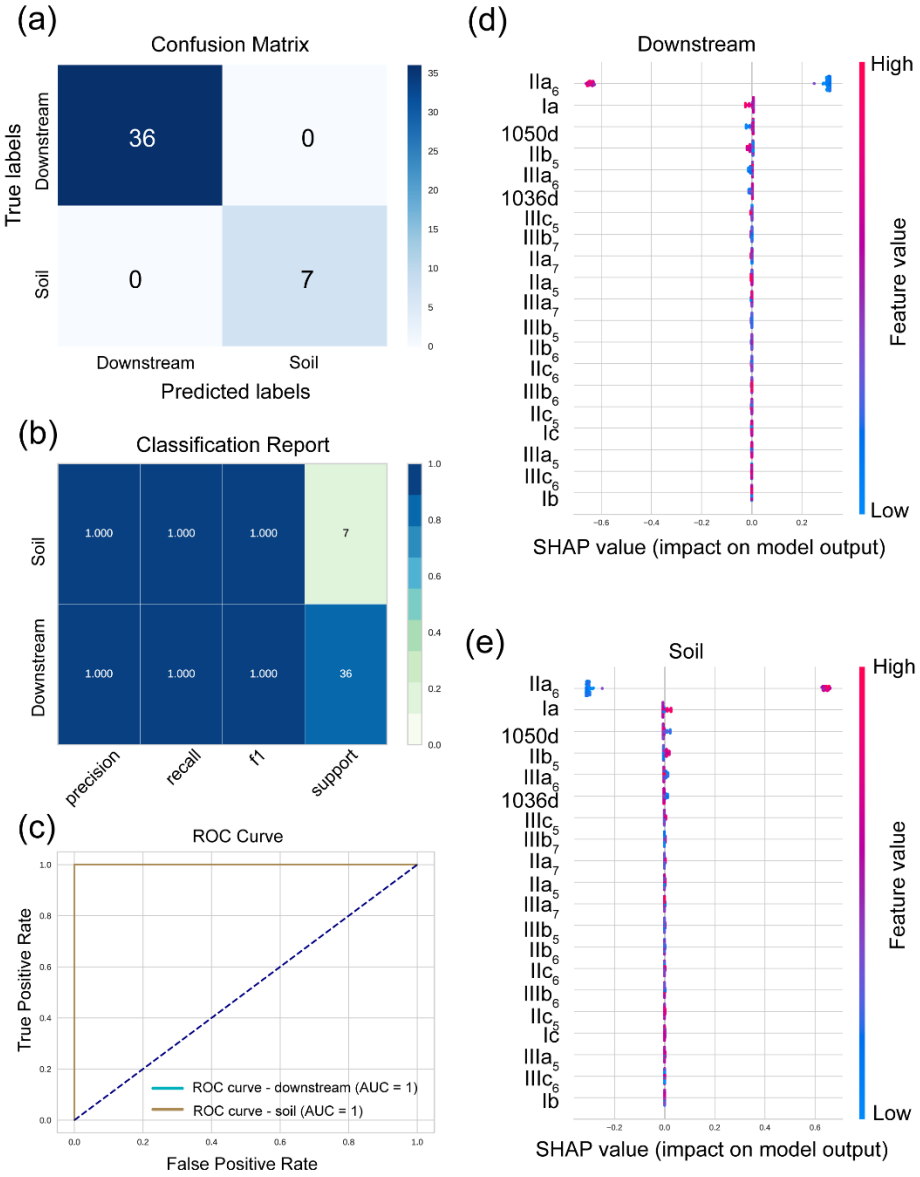
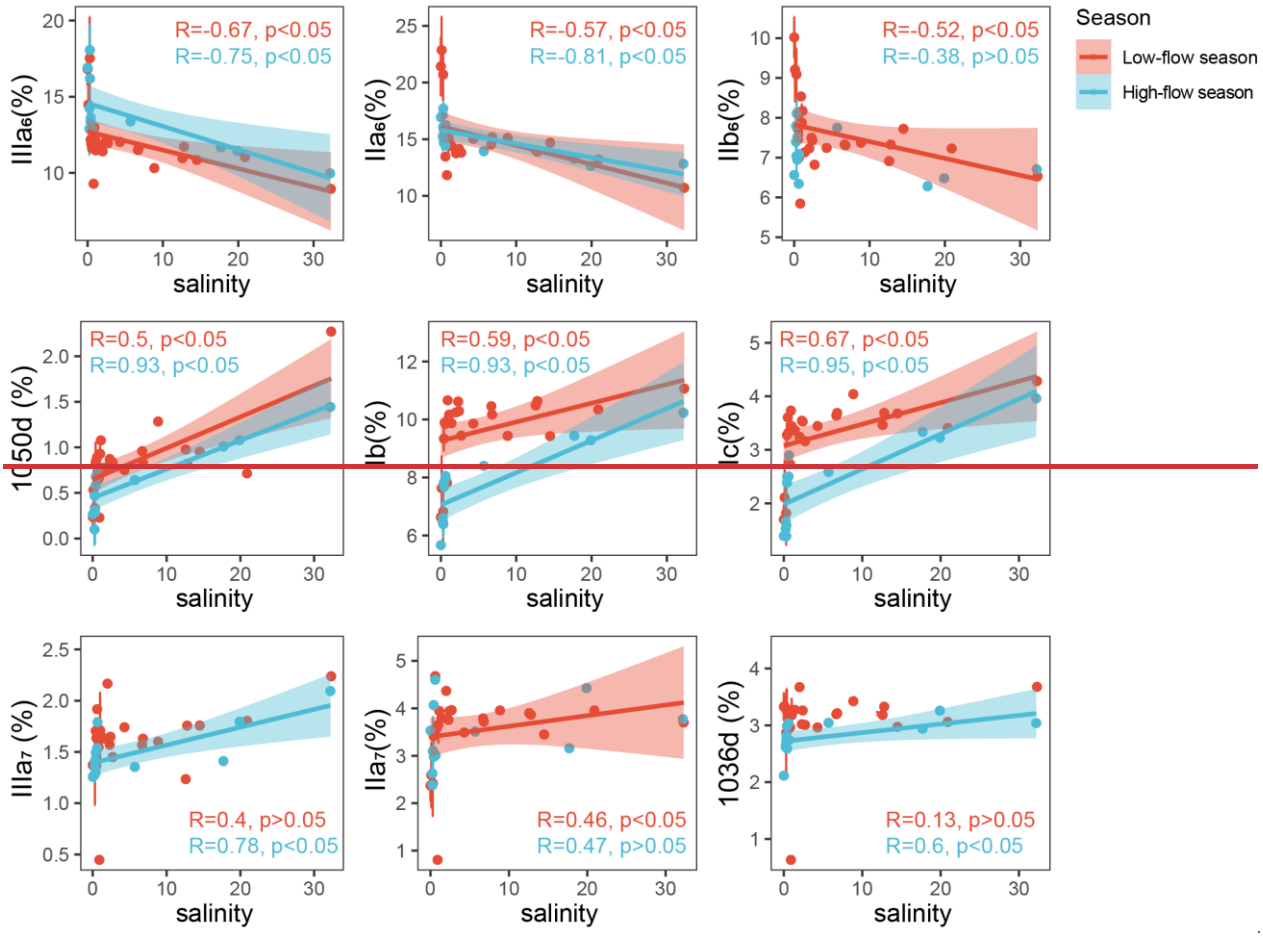


Fig. S98. Evaluation of the random forest model based on brGDGTs through the confusion matrix (a), classification report (b), and receiver operating characteristic (ROC) curve (c). SHAP summary plots (d-e) show the feature importance obtained from the random forest algorithm and the SHAP library. Each 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).


Season
Low-flow season
High-flow season
R=-0.67, p<0.05
R=-0.75, p<0.05
IIIa6(%)
R=-0.57, p<0.05
R=-0.81, p<0.05
IIa6(%)
R=-0.52, p<0.05
R=-0.38, p>0.05
IIb6(%)
R=0.5, p<0.05
R=0.93, p<0.05
1050d (%)
R=0.59, p<0.05
R=0.93, p<0.05
Ib(%)
R=0.67, p<0.05
R=0.95, p<0.05
Ic(%)
R=0.4, p>0.05
R=0.78, p<0.05
IIIa7 (%)
R=0.46, p<0.05
R=0.47, p>0.05
IIa7(%)
R=0.13, p>0.05
R=0.6, p<0.05
1036d (%)
salinity

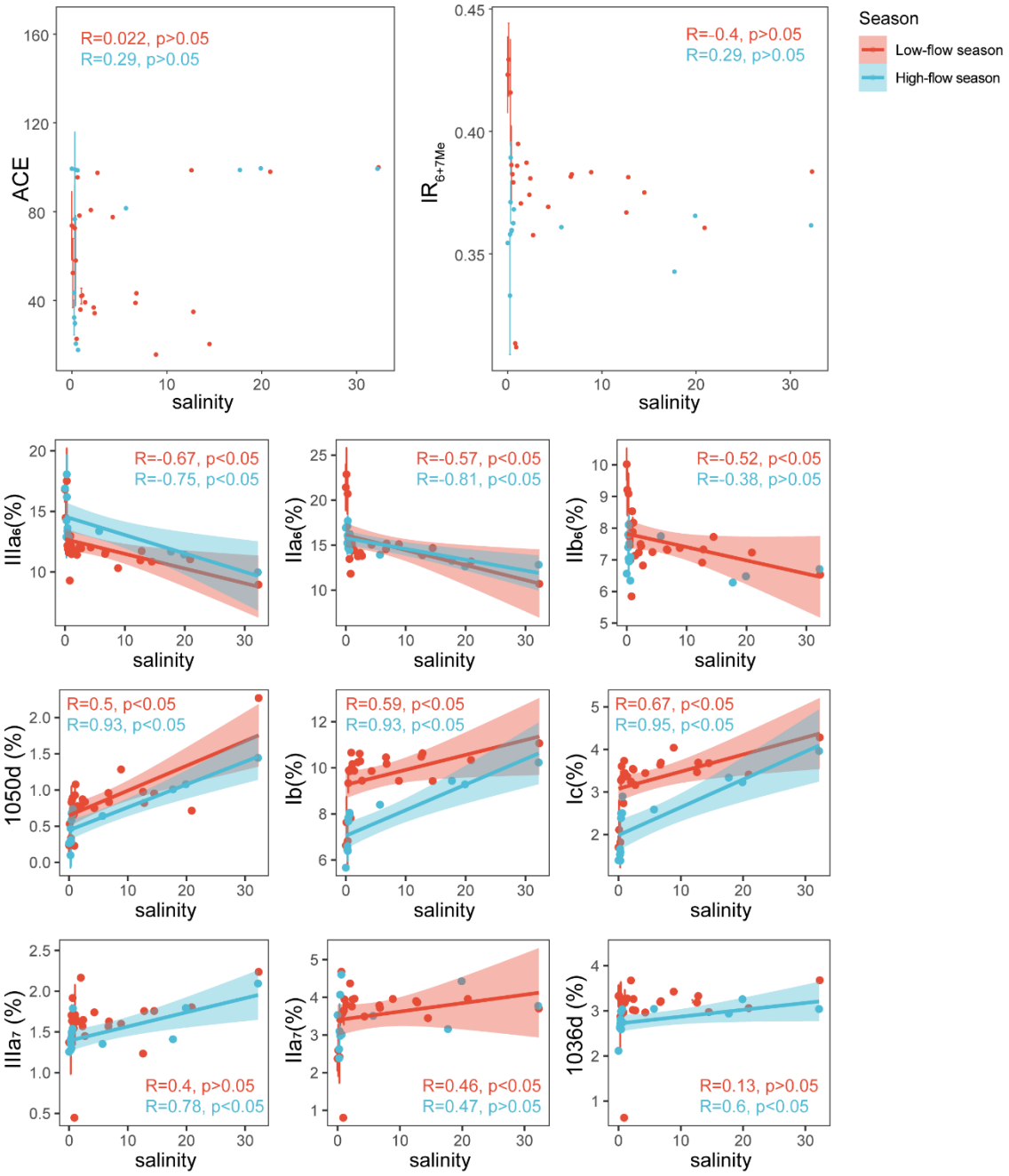


Fig. S109~~7~~. Salinity plotted versus ACE, IR_{6+7Me}, -relative abundance of 6-methyl and 7-methyl brGDGTs ($IIIa_6$, IIa_6, IIb_6, $IIIa_7$ and IIa_7) as well as compounds 1050d, 1036d, Ib, and Ic through the linear regression. Shaded area represents 95% confidence intervals. Vertical error bars indicate mean ± s.d for samples with the same salinity. Dataset is composed of SPM.

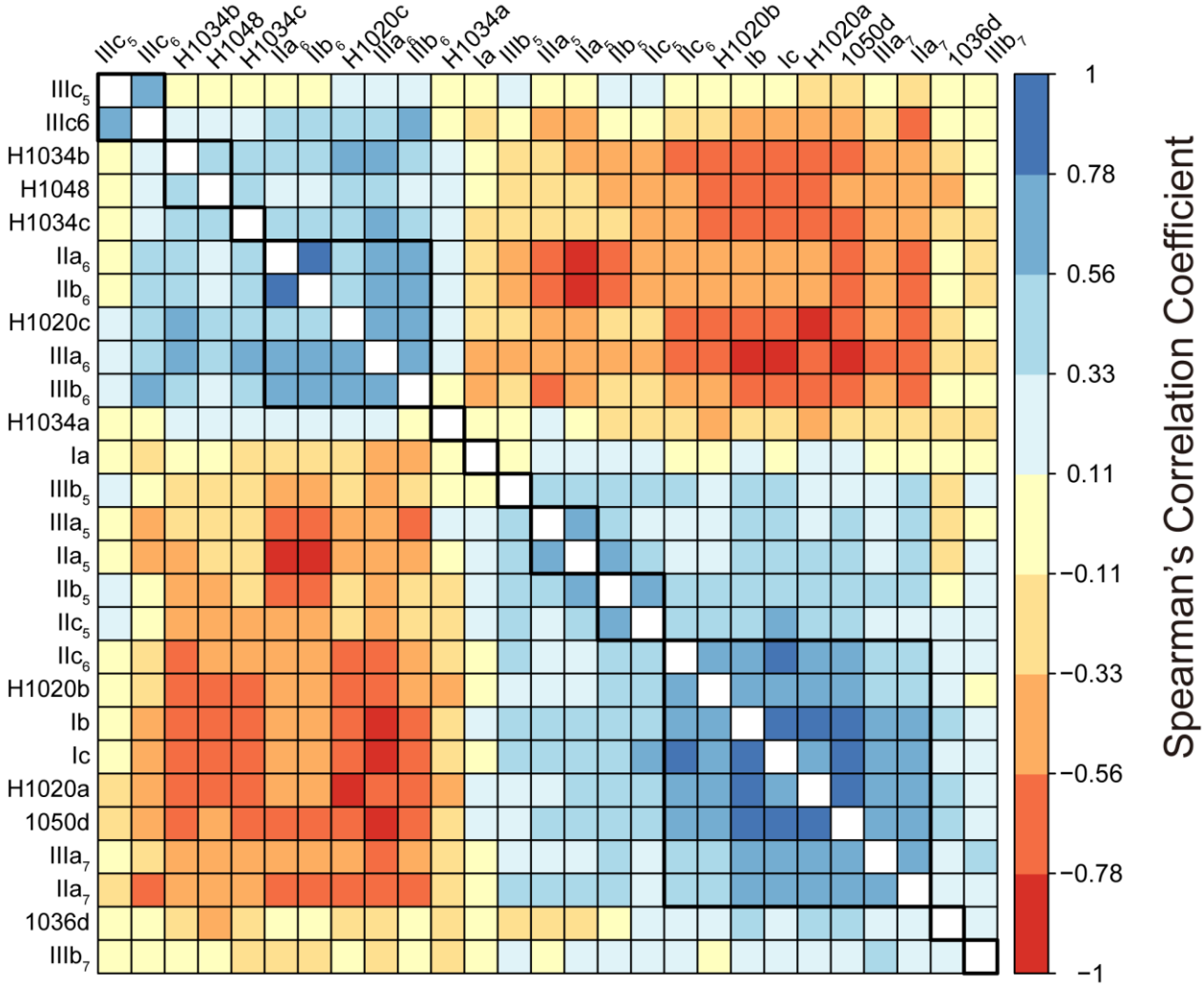


Fig. S1108. (a) Correlation plot between fractional abundance of brGDGTs (relative to all brGDGTs) and brGMGTs (relative to all brGMGTs).

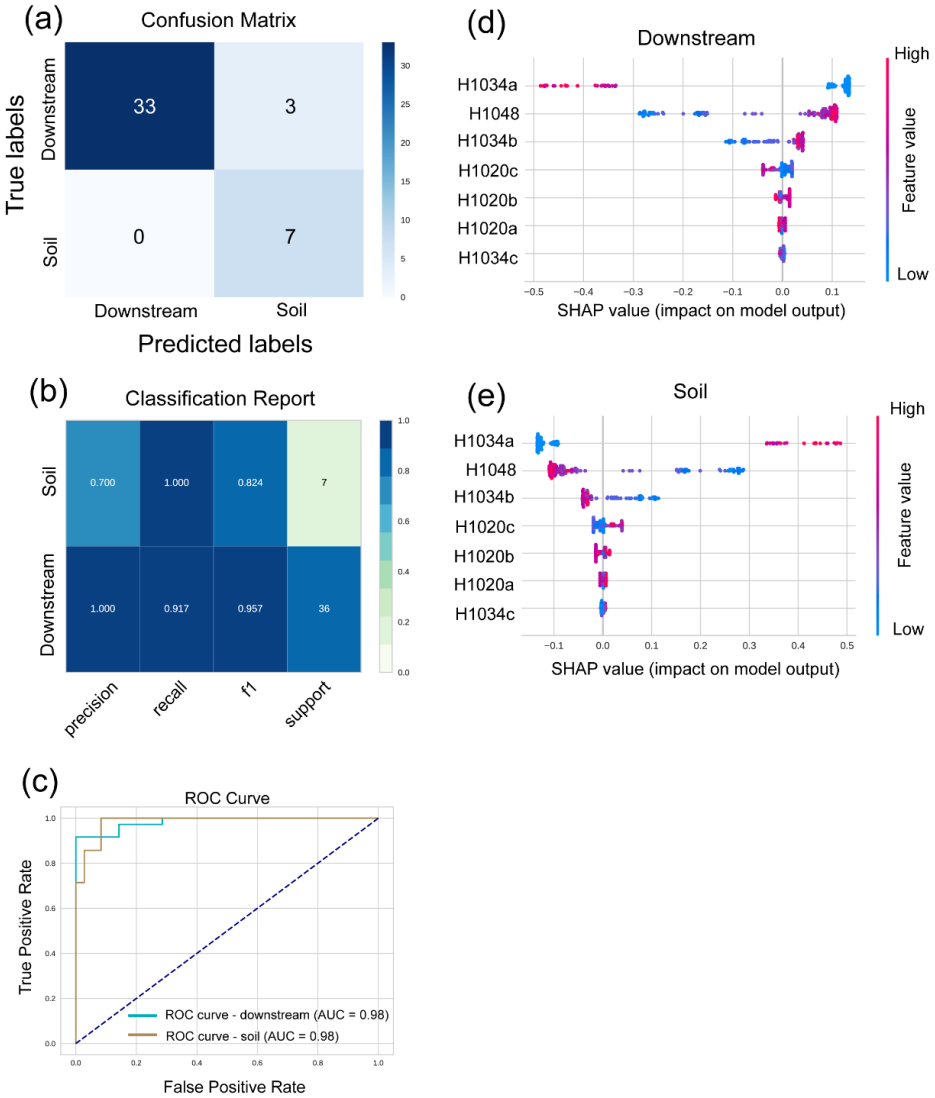


Fig. S12~~1~~. Evaluation of the random forest model based on brGMGTs through the confusion matrix (a), classification report (b), and receiver operating characteristic (ROC) curve (c). SHAP summary plots (d-e) show the feature importance obtained from the random forest algorithm and the SHAP library. 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).

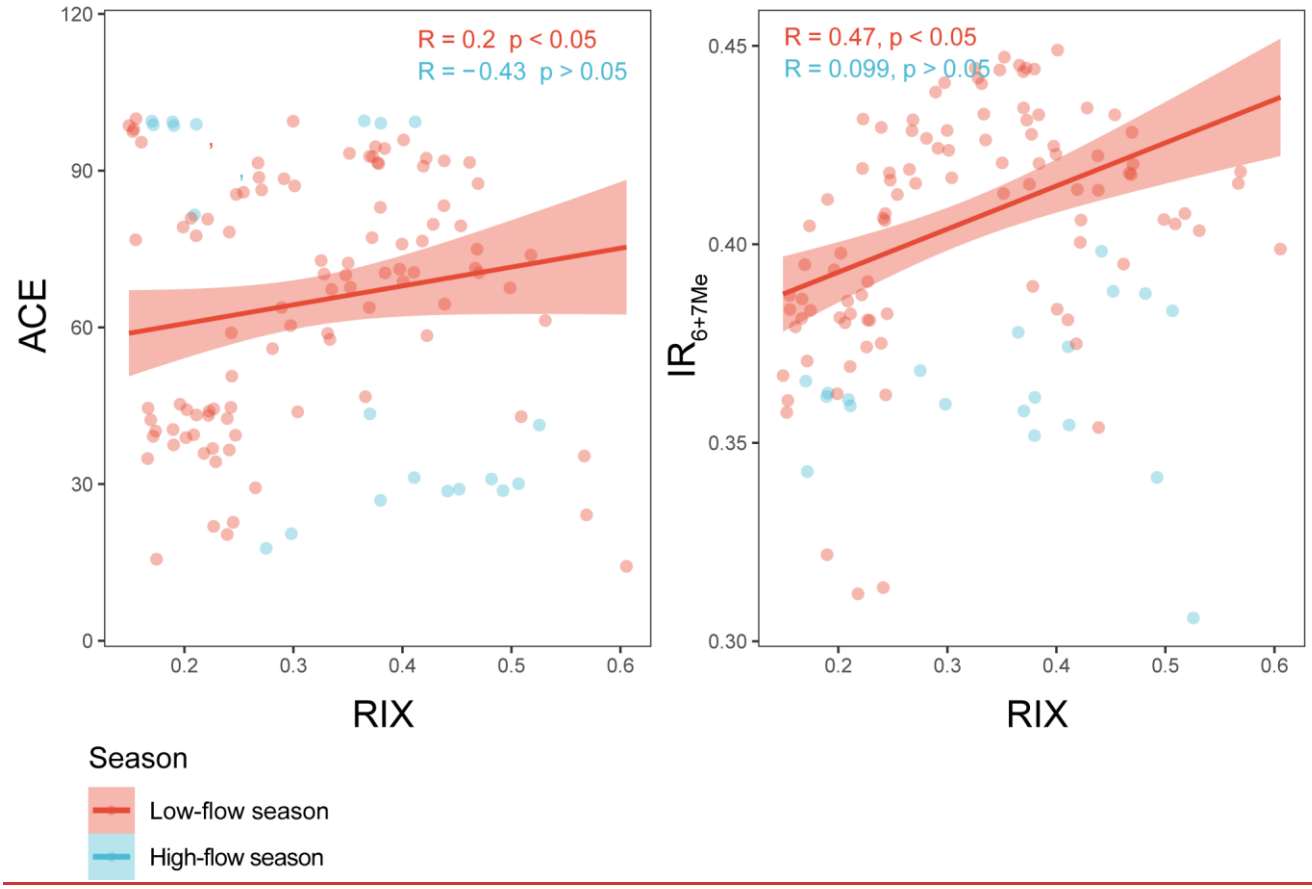


Fig. S132. RIX plotted versus ACE and IR_{6+7Me} through the linear regression. Shaded area represents 95% confidence intervals. Dataset is composed of SPM.

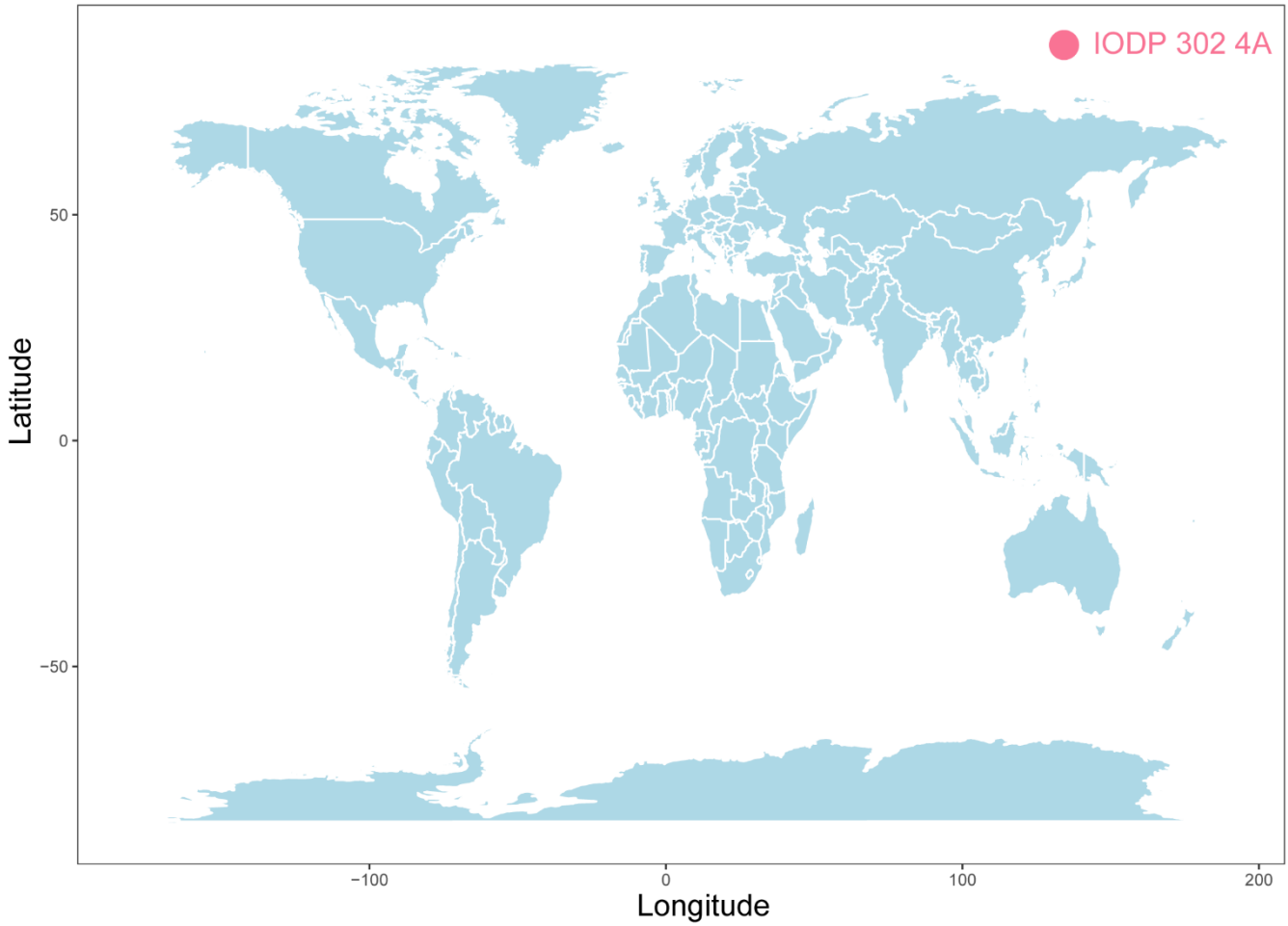


Fig. S143. Location of the IODP Expedition 302 Hole 4A.

Data availability

The brGDGT and brGMGT data will be archived in PANGAEA by the time of publication.

Author contribution

ZXZ, EP, and AH conceptualized this study. AC conducted LC-MS analyses. JM and AML collected and analyzed wetland soils and mudflat sediments. ZXZ, AH, EP, AT, and EB participated in the fieldwork. ZXZ performed the laboratory work, data analyses and wrote the article in consultation with all the other authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgments

We would like to thank members of the SARTRE project for their assistance during the field campaigns. We appreciate the logistical support from Robert Lafite, Frédéric Azémar, Anaëlle Bernard and Mahaut Sourzac. We are grateful to Pascal Claquin and Léo Chasselin for their support in water quality measurements and Micky Tackx for the scientific coordination of the SARTRE project. This study was funded by GIP Seine-Aval (MOSAIC, PHARESEE, FEREE and SARTRE projects) and CNRS/INSU and OFB (EC2CO RUNTIME project). Zhe-Xuan Zhang's PhD scholarship is funded by China Scholarship Council (No. 202004910406) and ASDB from Sorbonne Université. This work was supported by CESAM (UIDP/50017/2020 + UIDB/50017/2020 + LA/P/0094/2020). We would like to thank Vincent Grossi for helpful discussion. We also thank Appy Sluijs for providing the palynological data for the IODP Expedition 302 Hole 4A.

References

Andrisoa, A., Stieglitz, T. C., Rodellas, V., and Raimbault, P.: Primary production in coastal lagoons supported by groundwater discharge and porewater fluxes inferred from nitrogen and carbon isotope signatures, Marine Chemistry, 210, 48–60, https://doi.org/10.1016/j.marchem.2019.03.003, 2019.

Baxter, A. J., Hopmans, E. C., Russell, J. M., and Sinninghe Damsté, J. S.: Bacterial GMGTs in East African lake sediments: Their potential as palaeotemperature indicators, Geochimica et Cosmochimica Acta, 259, 155–169, https://doi.org/10.1016/j.gca.2019.05.039, 2019.

Baxter, A. J., Peterse, F., Verschuren, D., and Sinninghe Damsté, J. S.: Anoxic in situ production of bacterial GMGTs in the water column and surficial bottom sediments of a meromictic tropical crater lake: Implications for lake paleothermometry, Geochimica et Cosmochimica Acta, 306, 171–188, https://doi.org/10.1016/j.gca.2021.05.015, 2021.

Bertassoli, D. J., Häggi, C., Chiessi, C. M., Schefuß, E., Hefter, J., Akabane, T. K., and Sawakuchi, A. O.: Controls on the distributions of GDGTs and n-alkane isotopic compositions in sediments of the Amazon River Basin, Chemical Geology, 594, 120777, https://doi.org/10.1016/j.chemgeo.2022.120777, 2022.

Bianchi, T. S. and Canuel, E. A.: Chemical Biomarkers in Aquatic Ecosystems, Princeton University Press, 417 pp., 2011.

Bijl, P. K., Frieling, J., Cramwinckel, M. J., Boschman, C., Sluijs, A., and Peterse, F.: Maastrichtian–Rupelian paleoclimates in the southwest Pacific – a critical re-evaluation of biomarker paleothermometry and dinoflagellate cyst paleoecology at Ocean Drilling Program Site 1172, Climate of the Past, 17, 2393–2425, https://doi.org/10.5194/cp-17-2393-2021, 2021.

Blewett, J., Elling, F. J., Naafs, B. D. A., Kattein, L., Evans, T. W., Lauretano, V., Gallego-Sala, A. V., Pancost, R. D., and Pearson, A.: Metabolic and ecological controls on the stable carbon isotopic composition of archaeal (isoGDGT and BDGT) and bacterial (brGDGT) lipids in wetlands and lignites, Geochimica et Cosmochimica Acta, 320, 1–25, https://doi.org/10.1016/j.gca.2021.12.023, 2022.

Buchan, A., LeCleir, G. R., Gulvik, C. A., and González, J. M.: Master recyclers: features and functions of bacteria associated with phytoplankton blooms, Nat Rev Microbiol, 12, 686–698, https://doi.org/10.1038/nrmicro3326, 2014.

Cao, J., Lian, E., Yang, S., Ge, H., Jin, X., He, J., and Jia, G.: The distribution of intact polar lipid-derived branched tetraethers along a freshwater-seawater pH gradient in coastal East China Sea, Chemical Geology, 596, 120808, https://doi.org/10.1016/j.chemgeo.2022.120808, 2022.

Chen, Y., Zheng, F., Yang, H., Yang, W., Wu, R., Liu, X., Liang, H., Chen, H., Pei, H., Zhang, C., Pancost, R. D., and Zeng, Z.: The production of diverse brGDGTs by an Acidobacterium providing a physiological basis for paleoclimate proxies, Geochimica et Cosmochimica Acta, https://doi.org/10.1016/j.gca.2022.08.033, 2022.

Coffinet, S., Huguet, A., Bergonzini, L., Pedentchouk, N., Williamson, D., Anquetil, C., Gałka, M., Kołaczek, P., Karpińska-Kołaczek, M., Majule, A., Laggoun-Défarge, F., Wagner, T., and Derenne, S.: Impact of climate change on the ecology of the Kyambangunguru crater marsh in southwestern Tanzania during the Late Holocene, Quaternary Science Reviews, 196, 100–117, https://doi.org/10.1016/j.quascirev.2018.07.038, 2018.

Crampton-Flood, E. D., van der Weijst, C. M., van der Molen, G., Bouquet, M., Yedema, Y., Donders, T. H., Sangiorgi, F., Sluijs, A., Damsté, J. S. S., and Peterse, F.: Identifying marine and freshwater overprints on soil-derived branched GDGT temperature signals in Pliocene Mississippi and Amazon River fan sediments, Organic Geochemistry, 154, 104200, 2021.

Cramwinckel, M. J., Huber, M., Kocken, I. J., Agnini, C., Bijl, P. K., Bohaty, S. M., Frieling, J., Goldner, A., Hilgen, F. J., Kip, E. L., Peterse, F., van der Ploeg, R., Röhl, U., Schouten, S., and Sluijs, A.: Synchronous tropical and polar temperature evolution in the Eocene, Nature, 559, 382–386, https://doi.org/10.1038/s41586-018-0272-2, 2018.

Dai, G., Zhu, E., Liu, Z., Wang, Y., Zhu, S., Wang, S., Ma, T., Jia, J., Wang, X., Hou, S., Fu, P., Peterse, F., and Feng, X.: Compositional Characteristics of Fluvial Particulate Organic Matter Exported From the World's Largest Alpine Wetland, Journal of Geophysical Research: Biogeosciences, 124, 2709–2727, https://doi.org/10.1029/2019JG005231, 2019.

De Jonge, C., Hopmans, E. C., Stadnitskaia, A., Rijpstra, W. I. C., Hofland, R., Tegelaar, E., and Sinninghe Damsté, J. S.: Identification of novel penta- and hexamethylated branched glycerol dialkyl glycerol tetraethers in peat using HPLC–MS2, GC–MS and GC–SMB-MS, Organic Geochemistry, 54, 78–82, https://doi.org/10.1016/j.orggeochem.2012.10.004, 2013.

De Jonge, C., Hopmans, E. C., Zell, C. I., Kim, J.-H., Schouten, S., and Sinninghe Damsté, J. S.: Occurrence and abundance of 6-methyl branched glycerol dialkyl glycerol tetraethers in soils: Implications for palaeoclimate reconstruction, Geochimica et Cosmochimica Acta, 141, 97–112, https://doi.org/10.1016/j.gca.2014.06.013, 2014.

De Jonge, C., Stadnitskaia, A., Hopmans, E. C., Cherkashov, G., Fedotov, A., Streletskaya, I. D., Vasiliev, A. A., and Sinninghe Damsté, J. S.: Drastic changes in the distribution of branched tetraether lipids in suspended matter and sediments from the Yenisei River and Kara Sea (Siberia): Implications for the use of brGDGT-based proxies in coastal marine sediments, Geochimica et Cosmochimica Acta, 165, 200–225, https://doi.org/10.1016/j.gca.2015.05.044, 2015.

Dearing Crampton-Flood, E., Peterse, F., and Sinninghe Damsté, J. S.: Production of branched tetraethers in the marine realm: Svalbard fjord sediments revisited, Organic Geochemistry, 138, 103907, https://doi.org/10.1016/j.orggeochem.2019.103907, 2019.

Ding, S., Schwab, V. F., Ueberschaar, N., Roth, V.-N., Lange, M., Xu, Y., Gleixner, G., and Pohnert, G.: Identification of novel 7-methyl and cyclopentanyl branched glycerol dialkyl glycerol tetraethers in lake sediments, Organic Geochemistry, 102, 52–58, https://doi.org/10.1016/j.orggeochem.2016.09.009, 2016.

Druine, F.: Flux sédimentaire en estuaire de Seine : quantification et variabilité multi-échelle sur la base de mesure de turbidité (réseau SYNAPSES), phdthesis, Normandie Université ; Institut français de recherche pour l'exploitation de la mer, 2018.

Druine, F., Verney, R., Deloffre, J., Lemoine, J.-P., Chapalain, M., Landemaine, V., and Lafite, R.: In situ high frequency long term measurements of suspended sediment concentration in turbid estuarine system (Seine Estuary, France): Optical turbidity sensors response to suspended sediment characteristics, Marine Geology, 400, 24–37, https://doi.org/10.1016/j.margeo.2018.03.003, 2018.

Elling, F. J., Kattein, L., David A. Naafs, B., Lauretano, V., and Pearson, A.: Heterotrophic origin and diverse sources of branched glycerol monoalkyl glycerol tetraethers (brGMGTs) in peats and lignites, Organic Geochemistry, 104558, https://doi.org/10.1016/j.orggeochem.2023.104558, 2023.

Flipo, N., Lestel, L., Labadie, P., Meybeck, M., and Garnier, J.: Trajectories of the Seine River Basin, in: The Seine River Basin, edited by: Flipo, N., Labadie, P., and Lestel, L., Springer International Publishing, Cham, 1–28, https://doi.org/10.1007/698_2019_437, 2021.

Freymond, C. V., Peterse, F., Fischer, L. V., Filip, F., Giosan, L., and Eglinton, T. I.: Branched GDGT signals in fluvial sediments of the Danube River basin: Method comparison and longitudinal evolution, Organic Geochemistry, 103, 88–96, https://doi.org/10.1016/j.orggeochem.2016.11.002, 2017.

Halamka, T. A., Raberg, J. H., McFarlin, J. M., Younkin, A. D., Mulligan, C., Liu, X.-L., and Kopf, S. H.: Production of diverse brGDGTs by Acidobacterium Solibacter usitatus in response to temperature, pH, and O2 provides a culturing perspective on brGDGT proxies and biosynthesis, Geobiology, n/a, https://doi.org/10.1111/gbi.12525, 2022.

Harding, I. C., Charles, A. J., Marshall, J. E. A., Pälike, H., Roberts, A. P., Wilson, P. A., Jarvis, E., Thorne, R., Morris, E., Moremon, R., Pearce, R. B., and Akbari, S.: Sea-level and salinity fluctuations during the Paleocene–Eocene thermal maximum in Arctic Spitsbergen, Earth and Planetary Science Letters, 303, 97–107, https://doi.org/10.1016/j.epsl.2010.12.043, 2011.

Harning, D. J., Curtin, L., Geirsdóttir, Á., D'Andrea, W. J., Miller, G. H., and Sepúlveda, J.: Lipid Biomarkers Quantify Holocene Summer Temperature and Ice Cap Sensitivity in Icelandic Lakes, Geophysical Research Letters, 47, e2019GL085728, https://doi.org/10.1029/2019GL085728, 2020.

He, Y., Wang, H., Meng, B., Liu, H., Zhou, A., Song, M., Kolpakova, M., Krivonogov, S., Liu, W., and Liu, Z.: Appraisal of alkenone- and archaeal ether-based salinity indicators in mid-latitude Asian lakes, Earth and Planetary Science Letters, 538, 116236, https://doi.org/10.1016/j.epsl.2020.116236, 2020.

Hopmans, E. C., Weijers, J. W. H., Schefuß, E., Herfort, L., Sinninghe Damsté, J. S., and Schouten, S.: A novel proxy for terrestrial organic matter in sediments based on branched and isoprenoid tetraether lipids, Earth and Planetary Science Letters, 224, 107–116, https://doi.org/10.1016/j.epsl.2004.05.012, 2004.

Hopmans, E. C., Schouten, S., and Sinninghe Damsté, J. S.: The effect of improved chromatography on GDGT-based palaeoproxies, Organic Geochemistry, 93, 1–6, https://doi.org/10.1016/j.orggeochem.2015.12.006, 2016.

Huguet, A., Grossi, V., Belmahdi, I., Fosse, C., and Derenne, S.: Archaeal and bacterial tetraether lipids in tropical ponds with contrasting salinity (Guadeloupe, French West Indies): Implications for tetraether-based environmental proxies, Organic Geochemistry, 83–84, 158–169, https://doi.org/10.1016/j.orggeochem.2015.02.010, 2015.

Huguet, A., Meador, T. B., Laggoun-Défarge, F., Könneke, M., Wu, W., Derenne, S., and Hinrichs, K.-U.: Production rates of bacterial tetraether lipids and fatty acids in peatland under varying oxygen concentrations, Geochimica et Cosmochimica Acta, 203, 103–116, https://doi.org/10.1016/j.gca.2017.01.012, 2017.

Huguet, A., Coffinet, S., Roussel, A., Gayraud, F., Anquetil, C., Bergonzini, L., Bonanomi, G., Williamson, D., Majule, A., and Derenne, S.: Evaluation of 3-hydroxy fatty acids as a pH and temperature proxy in soils from temperate and tropical altitudinal gradients, Organic Geochemistry, 129, 1–13, https://doi.org/10.1016/j.orggeochem.2019.01.002, 2019.

Huguet, C., Hopmans, E. C., Febo-Ayala, W., Thompson, D. H., Sinninghe Damsté, J. S., and Schouten, S.: An improved method to determine the absolute abundance of glycerol dibiphytanyl glycerol tetraether lipids, Organic Geochemistry, 37, 1036–1041, https://doi.org/10.1016/j.orggeochem.2006.05.008, 2006.

Huguet, C., de Lange, G. J., Gustafsson, Ö., Middelburg, J. J., Sinninghe Damsté, J. S., and Schouten, S.: Selective preservation of soil organic matter in oxidized marine sediments (Madeira Abyssal Plain), Geochimica et Cosmochimica Acta, 72, 6061–6068, https://doi.org/10.1016/j.gca.2008.09.021, 2008.

Kang, M., He, J., and Jia, G.: Evaluation of heterocyte glycolipids with a hexose sugar moiety for tracing terrestrial organic matter in the South China Sea, Chemical Geology, 635, 121604, https://doi.org/10.1016/j.chemgeo.2023.121604, 2023.

Kim, J.-H., Ludwig, W., Buscail, R., Dorhout, D., and Sinninghe Damsté, J. S.: Tracing tetraether lipids from source to sink in the Rhône River system (NW Mediterranean), Frontiers in Earth Science, 3, 2015.

Kirkels, F. M. S. A., Ponton, C., Galy, V., West, A. J., Feakins, S. J., and Peterse, F.: From Andes to Amazon: Assessing Branched Tetraether Lipids as Tracers for Soil Organic Carbon in the Madre de Dios River System, Journal of Geophysical Research: Biogeosciences, 125, e2019JG005270, https://doi.org/10.1029/2019JG005270, 2020.

Kirkels, F. M. S. A., Usman, M. O., and Peterse, F.: Distinct sources of bacterial branched GMGTs in the Godavari River basin (India) and Bay of Bengal sediments, Organic Geochemistry, 167, 104405, https://doi.org/10.1016/j.orggeochem.2022.104405, 2022a.

Kirkels, F. M. S. A., Zwart, H. M., Usman, M. O., Hou, S., Ponton, C., Giosan, L., Eglinton, T. I., and Peterse, F.: From soil to sea: sources and transport of organic carbon traced by tetraether lipids in the monsoonal Godavari River, India, Biogeosciences, 19, 3979–4010, https://doi.org/10.5194/bg-19-3979-2022, 2022b.

Kolb, P., Zorndt, A., Burchard, H., Gräwe, U., and Kösters, F.: Modelling the impact of anthropogenic measures on saltwater intrusion in the Weser estuary, Ocean Science, 18, 1725–1739, https://doi.org/10.5194/os-18-1725-2022, 2022.

Kou, Q., Zhu, L., Ju, J., Wang, J., Xu, T., Li, C., and Ma, Q.: Influence of salinity on glycerol dialkyl glycerol tetraether-based indicators in Tibetan Plateau lakes: Implications for paleotemperature and paleosalinity reconstructions, Palaeogeography, Palaeoclimatology, Palaeoecology, 601, 111127, https://doi.org/10.1016/j.palaeo.2022.111127, 2022.

Lai, J., Zou, Y., Zhang, J., and Peres-Neto, P. R.: Generalizing hierarchical and variation partitioning in multiple regression and canonical analyses using the rdacca.hp R package, Methods in Ecology and Evolution, 13, 782–788, https://doi.org/10.1111/2041-210X.13800, 2022.

Lamb, A. L., Wilson, G. P., and Leng, M. J.: A review of coastal palaeoclimate and relative sea-level reconstructions using δ13C and C/N ratios in organic material, Earth-Science Reviews, 75, 29–57, https://doi.org/10.1016/j.earscirev.2005.10.003, 2006.

Lattaud, J., Kim, J.-H., De Jonge, C., Zell, C., Sinninghe Damsté, J. S., and Schouten, S.: The C32 alkane-1,15-diol as a tracer for riverine input in coastal seas, Geochimica et Cosmochimica Acta, 202, 146–158, https://doi.org/10.1016/j.gca.2016.12.030, 2017.

Leavitt, P. R., Brock, C. S., Ebel, C., and Patoine, A.: Landscape-scale effects of urban nitrogen on a chain of freshwater lakes in central North America, Limnology and Oceanography, 51, 2262–2277, https://doi.org/10.4319/lo.2006.51.5.2262, 2006.

Liu, Q., Liang, Y., Cai, W.-J., Wang, K., Wang, J., and Yin, K.: Changing riverine organic C:N ratios along the Pearl River: Implications for estuarine and coastal carbon cycles, Science of The Total Environment, 709, 136052, https://doi.org/10.1016/j.scitotenv.2019.136052, 2020.

Liu, X.-L., Summons, R. E., and Hinrichs, K.-U.: Extending the known range of glycerol ether lipids in the environment: structural assignments based on tandem mass spectral fragmentation patterns, Rapid Communications in Mass Spectrometry, 26, 2295–2302, https://doi.org/10.1002/rcm.6355, 2012.

Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., and Lee, S.-I.: From local explanations to global understanding with explainable AI for trees, Nat Mach Intell, 2, 56–67, https://doi.org/10.1038/s42256-019-0138-9, 2020.

Martínez-Sosa, P. and Tierney, J. E.: Lacustrine brGDGT response to microcosm and mesocosm incubations, Organic Geochemistry, 127, 12–22, https://doi.org/10.1016/j.orggeochem.2018.10.011, 2019.

Martínez-Sosa, P., Tierney, J. E., Stefanescu, I. C., Dearing Crampton-Flood, E., Shuman, B. N., and Routson, C.: A global Bayesian temperature calibration for lacustrine brGDGTs, Geochimica et Cosmochimica Acta, 305, 87–105, https://doi.org/10.1016/j.gca.2021.04.038, 2021.

Martínez-Sosa, P., Tierney, J. E., Pérez-Angel, L. C., Stefanescu, I. C., Guo, J., Kirkels, F., Sepúlveda, J., Peterse, F., Shuman, B. N., and Reyes, A. V.: Development and Application of the Branched and Isoprenoid GDGT Machine Learning Classification Algorithm (BIGMaC) for Paleoenvironmental Reconstruction, Paleoceanography and Paleoclimatology, 38, e2023PA004611, https://doi.org/10.1029/2023PA004611, 2023.

Morii, H., Eguchi, T., Nishihara, M., Kakinuma, K., König, H., and Koga, Y.: A novel ether core lipid with H-shaped C80-isoprenoid hydrocarbon chain from the hyperthermophilic methanogen Methanothermus fervidus, Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism, 1390, 339–345, https://doi.org/10.1016/S0005-2760(97)00183-5, 1998.

Naafs, B. D. A., Inglis, G. N., Zheng, Y., Amesbury, M. J., Biester, H., Bindler, R., Blewett, J., Burrows, M. A., del Castillo Torres, D., Chambers, F. M., Cohen, A. D., Evershed, R. P., Feakins, S. J., Gałka, M., Gallego-Sala, A., Gandois, L., Gray, D. M., Hatcher, P. G., Honorio Coronado, E. N., Hughes, P. D. M., Huguet, A., Könönen, M., Laggoun-Défarge, F., Lähteenoja, O., Lamentowicz, M., Marchant, R., McClymont, E., Pontevedra-Pombal, X., Ponton, C., Pourmand, A., Rizzuti, A. M., Rochefort, L., Schellekens, J., De Vleeschouwer, F., and Pancost, R. D.: Introducing global peat-specific temperature and pH calibrations based on brGDGT bacterial lipids, Geochimica et Cosmochimica Acta, 208, 285–301, https://doi.org/10.1016/j.gca.2017.01.038, 2017.

Naafs, B. D. A., McCormick, D., Inglis, G. N., and Pancost, R. D.: Archaeal and bacterial H-GDGTs are abundant in peat and their relative abundance is positively correlated with temperature, Geochimica et Cosmochimica Acta, 227, 156–170, https://doi.org/10.1016/j.gca.2018.02.025, 2018.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Peterse, F., Kim, J.-H., Schouten, S., Kristensen, D. K., Koç, N., and Sinninghe Damsté, J. S.: Constraints on the application of the MBT/CBT palaeothermometer at high latitude environments (Svalbard, Norway), Organic Geochemistry, 40, 692–699, https://doi.org/10.1016/j.orggeochem.2009.03.004, 2009.

Raberg, J. H., Miller, G. H., Geirsdóttir, Á., and Sepúlveda, J.: Near-universal trends in brGDGT lipid distributions in nature, Science Advances, 8, eabm7625, https://doi.org/10.1126/sciadv.abm7625, 2022.

Ralston, D. K. and Geyer, W. R.: Response to Channel Deepening of the Salinity Intrusion, Estuarine Circulation, and Stratification in an Urbanized Estuary, Journal of Geophysical Research: Oceans, 124, 4784–4802, https://doi.org/10.1029/2019JC015006, 2019.

Romero, E., Le Gendre, R., Garnier, J., Billen, G., Fisson, C., Silvestre, M., and Riou, P.: Long-term water quality in the lower Seine: Lessons learned over 4 decades of monitoring, Environmental Science & Policy, 58, 141–154, https://doi.org/10.1016/j.envsci.2016.01.016, 2016.

Romero, E., Garnier, J., Billen, G., Ramarson, A., Riou, P., and Le Gendre, R.: Modeling the biogeochemical functioning of the Seine estuary and its coastal zone: Export, retention, and transformations, Limnology and Oceanography, 64, 895–912, https://doi.org/10.1002/lno.11082, 2019.

Russell, J. M., Hopmans, E. C., Loomis, S. E., Liang, J., and Sinninghe Damsté, J. S.: Distributions of 5- and 6-methyl branched glycerol dialkyl glycerol tetraethers (brGDGTs) in East African lake sediment: Effects of temperature, pH, and new lacustrine paleotemperature calibrations, Organic Geochemistry, 117, 56–69, https://doi.org/10.1016/j.orggeochem.2017.12.003, 2018.

Ryba, S. A. and Burgess, R. M.: Effects of sample preparation on the measurement of organic carbon, hydrogen, nitrogen, sulfur, and oxygen concentrations in marine sediments, Chemosphere, 48, 139–147, https://doi.org/10.1016/S0045-6535(02)00027-9, 2002.

Schouten, S., Baas, M., Hopmans, E. C., Reysenbach, A.-L., and Damsté, J. S. S.: Tetraether membrane lipids of Candidatus "Aciduliprofundum boonei", a cultivated obligate thermoacidophilic euryarchaeote from deep-sea hydrothermal vents, Extremophiles, 12, 119–124, 2008.

Schouten, S., Hopmans, E. C., and Sinninghe Damsté, J. S.: The organic geochemistry of glycerol dialkyl glycerol tetraether lipids: A review, Organic Geochemistry, 54, 19–61, https://doi.org/10.1016/j.orggeochem.2012.09.006, 2013.

Sinninghe Damsté, J. S.: Spatial heterogeneity of sources of branched tetraethers in shelf systems: The geochemistry of tetraethers in the Berau River delta (Kalimantan, Indonesia), Geochimica et Cosmochimica Acta, 186, 13–31, https://doi.org/10.1016/j.gca.2016.04.033, 2016.

Sinninghe Damsté, J. S., Rijpstra, W. I. C., Hopmans, E. C., Weijers, J. W. H., Foesel, B. U., Overmann, J., and Dedysh, S. N.: 13,16-Dimethyl octacosanedioic acid (iso-diabolic acid), a common membrane-spanning lipid of Acidobacteria subdivisions 1 and 3, Appl Environ Microbiol, 77, 4147–4154, https://doi.org/10.1128/AEM.00466-11, 2011.

Sluijs, A., Schouten, S., Pagani, M., Woltering, M., Brinkhuis, H., Damsté, J. S. S., Dickens, G. R., Huber, M., Reichart, G.-J., Stein, R., Matthiessen, J., Lourens, L. J., Pedentchouk, N., Backman, J., and Moran, K.: Subtropical Arctic Ocean temperatures during the Palaeocene/Eocene thermal maximum, Nature, 441, 610–613, https://doi.org/10.1038/nature04668, 2006.

Sluijs, A., Brinkhuis, H., Crouch, E. M., John, C. M., Handley, L., Munsterman, D., Bohaty, S. M., Zachos, J. C., Reichart, G.-J., Schouten, S., Pancost, R. D., Damsté, J. S. S., Welters, N. L. D., Lotter, A. F., and Dickens, G. R.: Eustatic variations during the Paleocene-Eocene greenhouse world, Paleoceanography, 23, https://doi.org/10.1029/2008PA001615, 2008.

Sluijs, A., Schouten, S., Donders, T. H., Schoon, P. L., Röhl, U., Reichart, G.-J., Sangiorgi, F., Kim, J.-H., Sinninghe Damsté, J. S., and Brinkhuis, H.: Warm and wet conditions in the Arctic region during Eocene Thermal Maximum 2, Nature Geosci, 2, 777–780, https://doi.org/10.1038/ngeo668, 2009.

Sluijs, A., van Roij, L., Harrington, G. J., Schouten, S., Sessa, J. A., LeVay, L. J., Reichart, G.-J., and Slomp, C. P.: Warming, euxinia and sea level rise during the Paleocene–Eocene Thermal Maximum on the Gulf Coastal Plain: implications for ocean oxygenation and nutrient cycling, Climate of the Past, 10, 1421–1439, https://doi.org/10.5194/cp-10-1421-2014, 2014.

Sluijs, A., Frieling, J., Inglis, G. N., Nierop, K. G. J., Peterse, F., Sangiorgi, F., and Schouten, S.: Late Paleocene–early Eocene Arctic Ocean sea surface temperatures: reassessing biomarker paleothermometry at Lomonosov Ridge, Climate of the Past, 16, 2381–2400, https://doi.org/10.5194/cp-16-2381-2020, 2020.

Smith, R. W., Bianchi, T. S., and Li, X.: A re-evaluation of the use of branched GDGTs as terrestrial biomarkers: Implications for the BIT Index, Geochimica et Cosmochimica Acta, 80, 14–29, https://doi.org/10.1016/j.gca.2011.11.025, 2012.

Speijer, R. P. and Morsi, A.-M. M.: Ostracode turnover and sea-level changes associated with the Paleocene-Eocene thermal maximum, Geology, 30, 23–26, https://doi.org/10.1130/0091-7613(2002)030<0023:OTASLC>2.0.CO;2, 2002.

Strickland, J. D. H. and Parsons, T. R.: A Practical Handbook of Seawater Analysis, 2nd edition., https://doi.org/10.25607/OBP-1791, 1972.

Tang, X., Naafs, B. D. A., Pancost, R. D., Liu, Z., Fan, T., and Zheng, Y.: Exploring the Influences of Temperature on "H-Shaped" Glycerol Dialkyl Glycerol Tetraethers in a Stratigraphic Context: Evidence From Two Peat Cores Across the Late Quaternary, Frontiers in Earth Science, 8, 2021.

Thibault, A.: Dynamique de la matière organique dans la Seine : approche globale et moléculaire, phdthesis, Sorbonne Université, 2018.

Thibault, A., Derenne, S., Parlanti, E., Anquetil, C., Sourzac, M., Budzinski, H., Fuster, L., Laverman, A., Roose-Amsaleg, C., Viollier, E., and Huguet, A.: Dynamics of organic matter in the Seine Estuary (France): Bulk and structural approaches, Marine Chemistry, 212, 108–119, https://doi.org/10.1016/j.marchem.2019.04.007, 2019.

Tierney, J. E. and Russell, J. M.: Distributions of branched GDGTs in a tropical lake system: Implications for lacustrine application of the MBT/CBT paleoproxy, Organic Geochemistry, 40, 1032–1036, https://doi.org/10.1016/j.orggeochem.2009.04.014, 2009.

Tierney, J. E., Russell, J. M., Eggermont, H., Hopmans, E. C., Verschuren, D., and Sinninghe Damsté, J. S.: Environmental controls on branched tetraether lipid distributions in tropical East African lake sediments, Geochimica et Cosmochimica Acta, 74, 4902–4918, https://doi.org/10.1016/j.gca.2010.06.002, 2010.

Turich, C. and Freeman, K. H.: Archaeal lipids record paleosalinity in hypersaline systems, Organic Geochemistry, 42, 1147–1157, https://doi.org/10.1016/j.orggeochem.2011.06.002, 2011.

Van Den Hende, S., Vervaeren, H., Desmet, S., and Boon, N.: Bioflocculation of microalgae and bacteria combined with flue gas to improve sewage treatment, New biotechnology, 29, 23–31, 2011.

Véquaud, P., Thibault, A., Derenne, S., Anquetil, C., Collin, S., Contreras, S., Nottingham, A. T., Sabatier, P., Werne, J. P., and Huguet, A.: FROG: A global machine-learning temperature calibration for branched GDGTs in soils and peats, Geochimica et Cosmochimica Acta, 318, 468–494, https://doi.org/10.1016/j.gca.2021.12.007, 2022.

Wang, H., An, Z., Lu, H., Zhao, Z., and Liu, W.: Calibrating bacterial tetraether distributions towards in situ soil temperature and application to a loess-paleosol sequence, Quaternary Science Reviews, 231, 106172, https://doi.org/10.1016/j.quascirev.2020.106172, 2020.

Wang, H., Liu, W., He, Y., Zhou, A., Zhao, H., Liu, H., Cao, Y., Hu, J., Meng, B., Jiang, J., Kolpakova, M., Krivonogov, S., and Liu, Z.: Salinity-controlled isomerization of lacustrine brGDGTs impacts the associated MBT5ME' terrestrial temperature index, Geochimica et Cosmochimica Acta, 305, 33–48, https://doi.org/10.1016/j.gca.2021.05.004, 2021.

Wei, X., Garnier, J., Thieu, V., Passy, P., Le Gendre, R., Billen, G., Akopian, M., and Laruelle, G. G.: Nutrient transport and transformation in macrotidal estuaries of the French Atlantic coast: a modeling approach using the Carbon-Generic Estuarine Model, Biogeosciences, 19, 931–955, https://doi.org/10.5194/bg-19-931-2022, 2022.

Weijers, J. W. H., Schouten, S., van den Donker, J. C., Hopmans, E. C., and Sinninghe Damsté, J. S.: Environmental controls on bacterial tetraether membrane lipid distribution in soils, Geochimica et Cosmochimica Acta, 71, 703–713, https://doi.org/10.1016/j.gca.2006.10.003, 2007.

Weijers, J. W. H., Wiesenberg, G. L. B., Bol, R., Hopmans, E. C., and Pancost, R. D.: Carbon isotopic composition of branched tetraether membrane lipids in soils suggest a rapid turnover and a heterotrophic life style of their source organism(s), Biogeosciences, 7, 2959–2973, https://doi.org/10.5194/bg-7-2959-2010, 2010.

Willard, D. A., Donders, T. H., Reichgelt, T., Greenwood, D. R., Sangiorgi, F., Peterse, F., Nierop, K. G. J., Frieling, J., Schouten, S., and Sluijs, A.: Arctic vegetation, temperature, and hydrology during Early Eocene transient global warming events, Global and Planetary Change, 178, 139–152, https://doi.org/10.1016/j.gloplacha.2019.04.012, 2019.

Wu, J., Yang, H., Pancost, R. D., Naafs, B. D. A., Qian, S., Dang, X., Sun, H., Pei, H., Wang, R., Zhao, S., and Xie, S.: Variations in dissolved O2 in a Chinese lake drive changes in microbial communities and impact sedimentary GDGT distributions, Chemical Geology, 579, 120348, https://doi.org/10.1016/j.chemgeo.2021.120348, 2021.

Xie, S., Liu, X.-L., Schubotz, F., Wakeham, S. G., and Hinrichs, K.-U.: Distribution of glycerol ether lipids in the oxygen minimum zone of the Eastern Tropical North Pacific Ocean, Organic Geochemistry, 71, 60–71, https://doi.org/10.1016/j.orggeochem.2014.04.006, 2014.

Xu, S., Zhang, Z., Jia, G., Yu, K., Lei, F., and Zhu, X.: Controlling factors and environmental significance of BIT and δ13C of sedimentary GDGTs from the Pearl River Estuary, China over recent decades, Estuarine, Coastal and Shelf Science, 233, 106534, https://doi.org/10.1016/j.ecss.2019.106534, 2020.

Yedema, Y. W., Sangiorgi, F., Sluijs, A., Sinninghe Damsté, J. S., and Peterse, F.: The dispersal of fluvially discharged and marine, shelf-produced particulate organic matter in the northern Gulf of Mexico, Biogeosciences, 20, 663–686, https://doi.org/10.5194/bg-20-663-2023, 2023.

Zell, C., Kim, J.-H., Moreira-Turcq, P., Abril, G., Hopmans, E. C., Bonnet, M.-P., Sobrinho, R. L., and Damsté, J. S. S.: Disentangling the origins of branched tetraether lipids and crenarchaeol in the lower Amazon River: Implications for GDGT-based proxies, Limnology and Oceanography, 58, 343–353, https://doi.org/10.4319/lo.2013.58.1.0343, 2013.

Zell, C., Kim, J.-H., Balsinha, M., Dorhout, D., Fernandes, C., Baas, M., and Sinninghe Damsté, J. S.: Transport of branched tetraether lipids from the Tagus River basin to the coastal ocean of the Portuguese margin: consequences for the interpretation of the MBT'/CBT paleothermometer, Biogeosciences, 11, 5637–5655, https://doi.org/10.5194/bg-11-5637-2014, 2014.

Zeng, Z., Xiao, W., Zheng, F., Chen, Y., Zhu, Y., Tian, J., and Zhang, C.: Enhanced production of highly methylated brGDGTs linked to anaerobic bacteria from sediments of the Mariana Trench, Frontiers in Marine Science, 10, 2023.