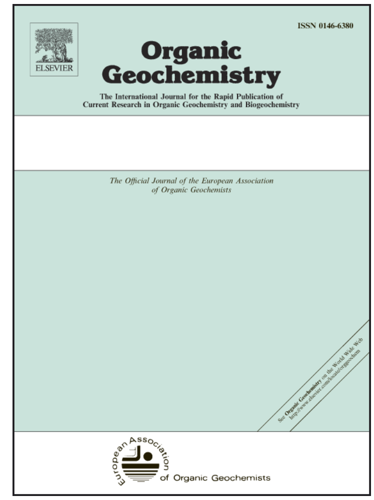
Journal Pre-proofs




# Characterisation of dissolved organic matter in two contrasting arsenic-prone sites in Kandal Province, Cambodia

Oliver C. Moore, Amy D. Holt, Laura Richards, Amy M. McKenna, Robert G.M. Spencer, Dan Lapworth, David A. Polya, Jonathan R. Lloyd, Bart E. van Dongen

PII: S0146-6380(24)00151-7
DOI: https://doi.org/10.1016/j.orggeochem.2024.104886
Reference: OG 104886

To appear in: *Organic Geochemistry*

Received Date: 17 May 2024
Revised Date: 16 October 2024
Accepted Date: 16 October 2024

Please cite this article as: Moore, O.C., Holt, A.D., Richards, L., McKenna, A.M., Spencer, R.G.M., Lapworth, D., Polya, D.A., Lloyd, J.R., van Dongen, B.E., Characterisation of dissolved organic matter in two contrasting arsenic-prone sites in Kandal Province, Cambodia, *Organic Geochemistry* (2024), doi: https://doi.org/10.1016/j.orggeochem.2024.104886

# Characterisation of dissolved organic matter in two contrasting arsenic-prone sites in Kandal Province, Cambodia

Oliver C. Moore[*,1,6], Amy D. Holt[2], Laura Richards[1], Amy M. McKenna[3,4], Robert G.M. Spencer[2], Dan Lapworth[5], David A. Polya[1], Jonathan R. Lloyd[*,1,] and Bart E. van Dongen[1]

[1]Department of Earth and Environmental Sciences and Williamson Research Centre for Molecular Environmental Science, University of Manchester, Manchester, UK

[2] National High Magnetic Field Laboratory Geochemistry Group and Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL 32306, USA.

[3]Ion Cyclotron Resonance Facility, National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32301, USA

[4]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA

[5]British Geological Survey, Maclean Building, Wallingford, UK

*Corresponding authors: oliver.moore@slu.se (Oliver C. Moore), jon.lloyd@manchester.ac.uk (Jonathan R. Lloyd)

## Highlights

- Ultra-high resolution mass spectrometry study of As-prone aquifer dissolved organics
- Optical/molecular analyses indicate overall dominance of terrestrial-derived organics
- Likely microbial processing of organics since percolation into the aquifer
- Analyses support role dissolved organics play during microbial mediated As release

## Abstract

Aquifers throughout Asia are impacted by the release of geogenic arsenic (As) into groundwater by microbial reduction of As-bearing Fe(III) (oxy)hydroxide minerals, severely impacting water quality. Groundwater dissolved organic matter (DOM) is likely key to As release, mainly as electron donor or electron shuttles. This study used optical analyses and ultra-high resolution mass spectrometry to examine the sources and composition of groundwater DOM in the As-prone aquifers of Kandal Province, Cambodia, at boreholes with differing host lithology (clay- and sand-dominated). Groundwater and surface water DOM composition were related to As concentrations, to infer the potential role of DOM in promoting As release. Optical and molecular-level analyses indicated an overall dominance of terrestrial-derived DOM in the groundwater samples, with higher freshness index and relative abundance (RA) of aliphatic compounds in clay compared to sand-dominated lithology. Compared to

[6] Current address: Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Biocentrum, Uppsala, Sweden

surface water, groundwater DOM had relatively lower O/C ratios and nominal oxidation state of carbon (-0.19 to -0.13 compared to 0.04 for ground and surface water, respectively), with a lower %RA of aliphatic compounds and higher %RA of carboxyl-rich alicyclic molecules, suggesting microbial processing of DOM since percolation into the aquifer. Concentrations of As across both sites were negatively correlated with DOM tryptophan:fulvic-like fluorescence and the %RA of aliphatics, potentially indicating microbial degradation of biolabile DOM in connection with As release, which is consistent with its role as an electron donor source. Together these data support DOM composition as an important control on microbial mediated As release.

**Keywords**: Dissolved organic matter, Cambodia, Kandal Province, optical properties, FT-ICR MS, As-cycling.

## 1. Introduction

The contamination of drinking water with arsenic (As) is associated with a range of health problems including cancers, cardiovascular disease, and skin lesions (Hughes *et al.*, 2011). Up to 220 million people worldwide are estimated to be at risk of chronic As exposure (Podgorski & Berg, 2020). The problem is especially severe in south and southeast Asia, where As-contaminated groundwater can be the primary source of drinking water. The problem is so widespread it has been termed *"the largest mass poisoning of a population in history"* (Smith *et al.*, 2000). The As-prone aquifers are found within the Indus, Ganges, Irrawaddy, Mekong and Red River deltas, and the Hetao and Jianghan alluvial plains, all of which drain from the Himalayas. These deltas and alluvial plains consist of sediments containing As-bearing Fe(III) (oxy)hydroxide minerals (Kinniburgh and Smedley, 2001; Fendorf *et al.*, 2010; Mukherjee *et al.*, 2012). Arsenic is thought to be released from these minerals by microbial-mediated reductive-dissolution of the ferric host minerals (Islam *et al.*, 2004; Charlet & Polya, 2006). Key to this process, are the composition and environmental sources of (organic) electron donors. However, organic matter (OM) composition and environmental sources in aquifers depend on a number of site-specific factors (e.g., geomorphological setting and sediment lithology), and the impact of these factors on As release remains poorly understood (Moore *et al.*, 2023). Proposed sources of OM in As-prone aquifers include: 1) draw-down of terrestrial-derived OM (e.g., overlying plant matter) by the hydraulic gradient maintained by pumping, with possible anthropogenic inputs (Harvey *et al.*, 2002; van Dongen *et al.*, 2008; Neumann *et al.*, 2010; Whaley-Martin *et al.*, 2016); 2) OM derived from the sediments, either deposited with the sediments, or transported from upstream (McArthur *et al.*, 2004; van Dongen *et al.*, 2008); and 3) upwelling of thermally mature hydrocarbons (e.g., petroleum) from underlying reservoirs (Rowland *et al.*, 2006, 2009; van Dongen *et al.*, 2008; Al Lawati *et al.*, 2012, 2023; Ghosh *et al.*, 2015; Magnone *et al.*, 2017). Sediment microcosm-based studies have shown that the type and biolability of organic material, not necessarily the total quantity, can influence the prevalence and mechanism of Fe(III) reduction and As release (Rowland *et al.*, 2007, 2009; Glodowska *et al.*, 2020; Qiao *et al.*, 2020; Wang *et al.*, 2021). Therefore, understanding how the composition and environmental sources of aquifer OM relate to As release is of key importance.

To date, research has been predominately conducted on the sedimentary OM (SOM) of As-prone aquifers (McArthur *et al.*, 2004; Rowland *et al.*, 2006, 2007, 2009; van Dongen *et al.*, 2008; Al Lawati *et al.*, 2012, 2013; Neumann *et al.*, 2014; Magnone *et al.*, 2017; Ye *et al.*, 2017; Mao *et al.*, 2018; Pracht *et al.*, 2018). As such, the understanding of the relationship between dissolved organic matter (DOM) composition and As release remains incomplete (Qiao et al., 2021). Biolabile DOM compounds could play a role as electron donors and carbon sources (Ghosh *et al.*, 2015; Qiao *et al.*, 2021) whilst more stable compounds (particularly humic compounds) may act as electron shuttles (Nevin & Lovley, 2002; Rowland *et al*., 2007; Mladenov *et al.*, 2010; Melton *et al.*, 2014; Chen *et al.*, 2017; Kulkarni *et al.*, 2017; Richards *et al.*, 2019b). Furthermore, dependent on composition, DOM can undergo complexation reactions with As and iron (Fe; Mukhopadhyay & Sanyal, 2004; Guo *et al.*, 2011; Fakour & Lin, 2014; Xue *et al.*, 2019), and competitive sorption with As for sites on Fe oxide minerals (Grafe *et al.*, 2002; Ko *et al.*, 2004; Bauer & Blodau, 2006). Given these interactions, it is important to understand how DOM composition and source impacts As release.

The composition of DOM can be characterised by analysis of optical properties, for instance specific ultraviolet absorbance at 254 nm ($SUVA_{254}$) is used as a proxy for DOM aromaticity (i.e., greater $SUVA_{254}$ indicates greater aromaticity; Weishaar *et al.*, 2003). In excitation emission matrix (EEM) analysis, specific excitation and emission wavelengths correspond to the fluorescence properties of DOM (Leenheer & Croué, 2003). These properties can be identified by peak-picking or deconvolution of target EEMs (e.g., for protein-like and humic-like components), accompanied by the calculation of indices (Tye & Lapworth, 2016; Richards *et al.*, 2019b). For example, fluorescence index (FI) and Freshness index, have been related to source and processing of DOM, where higher values indicate greater relative contributions of recently produced, microbially derived OM, over processed, terrestrial OM sources (e.g., soil and vegetation; McKnight *et al*., 2001; Parlanti *et al*., 2000; Wilson & Xenopoulos, 2009; Kulkarni *et al*., 2017). These measurements have become widely used for the characterisation of DOM, including in studies of As-prone aquifers, due to their ease of measurement and cost-effectiveness (Derrien *et al.*, 2019).

Ultra-high resolution Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS), allows for the molecular-level identification and categorisation of thousands of organic compounds, and is considered the most powerful technique for DOM characterisation (Hertkorn *et al.*, 2007; Hockaday *et al.*, 2009; Ohno *et al.*, 2010; Dittmar & Stubbins, 2014; Spencer *et al.*, 2014; Hodgkins *et al.*, 2016; Derrien *et al.*, 2019; McDonough *et al.*, 2022). FT-ICR MS has been applied to the study of DOM in As-prone aquifers in the (Bangladeshi) Bengal Basin (Pracht *et al.*, 2018), the Jianghan Plain (Du *et al.*, 2020; Yu *et al.*, 2020), and Hetao Basin (Qiao *et al.*, 2020, 2021). The combined use of optical and FT-ICR MS techniques has been applied to aquatic environments to understand the source and composition of DOM, as well as relationships between fluorescence properties and molecular composition (e.g., Stubbins *et al.*, 2014; Kellerman *et al.*, 2015, 2018; Wagner *et al.*, 2015). However, few studies have assessed this relationship in As-prone aquifers (Pracht *et al* 2018; Qiao *et al.*, 2020, 2021; Wang et al., 2023; Gao et al., 2023; Li et al., 2024), thus the association between DOM composition and As-cycling remains poorly defined. Previously, incubation studies have used FT-ICR MS to link molecular-level DOM composition to biolability (Spencer *et al.,* 2015; Textor *et al.,* 2018; Kellerman *et al.,* 2018). For instance, D'Andrilli *et al*. (2015) proposed the molecular lability boundary (MLB), whereby organic molecules with H/C ratios ≥1.5 (i.e., aliphatics) were considered most biolabile. Incubation studies of water-extracted sedimentary organic matter (WESOM) from the Hetao Basin, have shown these high H/C compounds (including carbohydrate, and protein-like compounds) are microbially degraded, and associated with Fe(II) and As release from sediments (Qiao *et al.*, 2020). Furthermore, analysis of the optical and molecular properties of the Hetao Basin has shown positive correlations between groundwater As concentrations and more biostable DOM (e.g., humic-like components and lower H/C molecular formulae), suggesting degradation of biolabile DOM coupled with reduction of As-bearing Fe(III) oxyhydroxide minerals and As release, possibly enhanced by the microbial utilisation of stable molecules as electron shuttles (Qiao *et al.*, 2021). Incubation experiments on aquifer organics from Bangladesh have shown similar trends; particularly, the sequential degradation of organic molecules in order of decreasing nominal oxidation state of carbon (NOSC; an index for thermodynamic viability and energy potential of the organic compounds (LaRowe & Cappellen, 2011) in connection with As release (Pracht *et al.*, 2018). Together, these studies highlight that the combined analysis of DOM by optical and molecular techniques in As prone aquifers can provide useful insight into DOM composition, source and processing in relation to As release, highlighting the role of bioavailable DOM in microbial mediated As release.

The aquifers of the Kandal Province, Cambodia, have been characterised extensively in previous studies (Polya *et al.*, 2005; Charlet & Polya, 2006; Rowland *et al.*, 2007, 2008; Tamura *et al.*, 2007; Benner *et al.*, 2008; Kocar *et al.*, 2008, 2014; Polizzotto *et al.*, 2008; van Dongen *et al.*, 2008; Lawson *et al.*, 2013, 2016; Richards *et al.*, 2016, 2017, 2018, 2019a, 2019b; Magnone *et al.*, 2017, 2019a; Uhlemann *et al.*, 2017). High spatial heterogeneity has been found in groundwater As, Fe, and bulk dissolved organic carbon (DOC) concentrations, in the ranges of 11–1095 μg/L, 0–6.9 mg/L, and 0.8–18 mg/L, respectively (Richards *et al.*, 2017). Groundwater abstraction in the region is limited due to relatively low population density and economic development (Richards *et al*., 2018 – and references therein). Studies related to the SOM have shown that the clay-dominated sediments have higher OC content and OM is primarily derived from recently produced terrestrial sources (i.e., plant matter), compared to sand-dominated sediments where a greater proportion of OM may be derived from thermally mature (petroleum derived) hydrocarbons (Rowland *et al*., 2007; van Dongen *et al*., 2008; Magnone *et al*., 2017). Optical characterization of DOM at both sand and clay-dominated sites has shown an overall dominance of terrestrial-derived DOM (i.e., humic-like fluorescence properties; Richards *et al.*, 2019b). Despite this similarity, there are indications of higher biolability of DOM in the clay-dominated area compared to that of the sand-dominated (Richards *et al.,* 2019b), which may result in variability in As release. Comprehensive isotopic-based hydrogeological studies of the area have shown that the inorganic carbon derived from dissolved organic carbon (DOC) degradation (associated with As release and other processes) pre-dates modelled groundwater flow (by hundreds of years) and suggested the importance of in situ oxidation of DOC in As release (Magnone *et al.*, 2019). This indicates that DOM may play an important role in As release at these sites. However, there is currently little understanding of the molecular-level composition of groundwater organics from these sites.

The aims of this study were to: (i) investigate the relationships between molecular-level and optical properties of groundwater DOM in the As-prone aquifers of Kandal Province, Cambodia; (ii) to provide insights into the potential environmental sources of the DOM and its subsequent microbial processing; and (iii) identify connections between DOM composition and As release. To achieve this, groundwater DOM composition was assessed at boreholes (*n*=8) with differing lithology (clay- or sand-dominated) and compared to surface water DOM (*n*=1). The composition of DOM at these sites was characterised using optical and, for the first time, molecular-level analysis from ultrahigh resolution 21 Tesla FT-ICR MS. Multivariate statistics were used to relate bulk and molecular-level DOM composition to DOC concentrations and inorganic geochemistry published elsewhere (from Bassil *et al*., 2024), ultimately providing insight into the role of DOM in As cycling.

## 2. Materials and Methods

### 2.1. Site Description and Field Sampling

The field sites for this study were located between the Mekong and Bassac Rivers, in Kandal Province, Cambodia. The two sites were contrasted by local stratigraphy, with one being sand-dominated and one clay-dominated (Uhlemann *et al.*, 2017). Four 18 m deep boreholes (with 5 m of screening) ~2–3 m apart, were drilled at each site in January 2019 and fully developed in May 2019 (Bassil *et al*., 2024). These boreholes are located near sample sites along hydrogeological transects from previous studies in the area (Richards *et al.*, 2016, 2017, 2018, 2019a, 2019b; Magnone *et al.*, 2017, 2019) , with the boreholes of the clay-dominated site being located near previously-named sites LR12–LR14 and the boreholes of the sand-dominated site near LR01 (Fig. 1). Fieldwork for this study was conducted in January 2020, with the groundwater from each well pumped to the surface using a submersible pump (Proactive Environmental Products Tornado 12-volt). Prior to sampling, each well was flushed until the oxidation-reduction-potential readings were stabilised. Water samples for DOC concentration, inorganic geochemistry (As and Fe) and DOM compositional analyses were taken from all eight boreholes, as well as a surface water sample ($n$=1) from a lake near the boreholes of the clay-dominated site (Fig. 1).

Inorganic geochemical and DOC analyses were performed and the results thereof are discussed separately (Bassil *et al*., 2024). Briefly, groundwater samples were filtered through 0.45 μm cellulose filters, and transported and stored (at 4 °C) to the Manchester Analytical Geochemistry Unit (MAGU). The As and Fe concentrations were determined by inductively coupled plasma mass spectrometry (ICP-MS) after acidification with 2% v/v $HNO_3$, using established methods (Richards *et al*., 2017; Bassil et al., 2024). Bulk DOC concentrations were measured using a Shimadzu TOC-V CPN Analyser, at MAGU using standard protocols (Richards *et al*., 2017; Bassil et al., 2024). Sampling for DOM composition analysis was conducted alongside DOC and inorganic geochemistry. Briefly, samples for EEM analyses were taken by filtration of 20 mL of sample water through 0.45 μm MERK cellulose nitrate filters using a BD-plastic syringe, into acid-washed (10% $HNO_3$ $\geq$ 24 hours) and pre-combusted (430 °C; >3 hours) amber glass vials (Richards *et al.*, 2019b). Samples were stored at 4 °C in the dark. Similarly, for FT-ICR MS analysis, 200 mL of sample water was filtered through pre-combusted 0.7 μm Whatman GF/F filters, using a pre-combusted (430 °C, >3h) and pre-rinsed glass pump filtration system, into 250 mL acid-washed (10% HCl for 72 hours) and pre-rinsed polycarbonate bottles. Samples were stored frozen (-20 °C) in the dark.

### 2.2. Excitation-emission matrix (EEM) analysis

Optical analyses were conducted at the British Geological Survey laboratory (Wallingford, UK) following previously described methodology (Lapworth & Kinniburgh, 2009; Tye & Lapworth, 2016; Richards *et al.*, 2019b). Specific ultraviolet absorbance at 254 nm ($SUVA_{254}$) was calculated by division of the absorbance at 254 nm ($A_{254;}$ $m^{-1}$) by the DOC concentration (mg/L ; Weishaar *et al.*, 2003). The EEMs were blank corrected prior to peak-picking and peaks reported in Raman Units (RU). Peak-picking was performed for the EEM values corresponding to tryptophan-like (TPH-like), tyrosine-like (TYR-like), fulvic acid-like (FA-like), humic acid-like (HA-like) fluorescence. Greater values of single-component variables (TPH-like, TYR-like, FA-like and HA-like fluorescence) indicate a greater relative abundance of the compounds they correspond to. The ratio between TPH-like and FA-like fluorescence (TPH:FA-like), was calculated as a proxy for microbial-derived over terrestrial-derived DOM, which was indicated by greater TPH:FA-like values (Parlanti *et al.*, 2000; Wilson & Xenopoulos, 2009; Richards *et al.*, 2019b). Peak-picking was performed for the calculation of FI (McKnight *et al.*, 2001; Cory *et al.*, 2010), humification index (HIX; Zsolnay *et al.*, 1999), HIX corrected for inner filtration ($HIX_{corr}$; Ohno, 2002), and "freshness index" (the ratio between relatively freshly produced DOM (microbial-derived or produced in situ, "β") and more stable DOM (i.e., terrestrial derived carbon compounds, "α"; β:α; Parlanti *et al.*, 2000; Wilson & Xenopoulos, 2009; Kulkarni *et al.*, 2017; Table S1).

## 2.3. Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS)

The samples were shipped frozen to the National High Magnetic Field Laboratory (Tallahassee, Florida, USA) for preparation and analysis. Filtered samples were acidified to pH 2 (10 M HCl) and solid phase extracted (SPE) onto preconditioned Bond Elut 100 mg PPL columns following standard methods (Agilent technologies; Dittmar *et al.*, 2008). The volume of sample extracted was adjusted dependent to the DOC concentration of the sample to achieve a target loading of ~40 µg C, assuming an extraction efficiency of 60% (Dittmar *et al.*, 2008). Following extraction, PPL columns were dried with a flow of ultrahigh purity nitrogen gas and eluted with 1 mL of HPLC methanol into pre-combusted (550°C, 5h) glass vials and stored at -20 °C until analysis. The eluted samples were analysed using a custom built 21 Tesla FT-ICR MS, using negative-mode electrospray ionization (Hendrickson *et al.*, 2015). Each mass spectrum was formed from 100 scans, conditionally co-added, and subsequently phase corrected (Xian *et al., 2010)*. Spectra were internally calibrated in Predator analysis using the 'walking' calibration and 10-15 abundant homologous series that covered each spectra's molecular weight distribution (Blakney *et al.*, 2011; Savory *et al.*, 2011). Peaks with a signal greater than the root mean square baseline noise plus 6σ were exported to a peak list, and elemental composition assigned between 170–1000 Da using Petroorg© (Corilo, 2014). Assignments of molecular formulae were within the bounds $C_{1\text{-}100}$, $H_{4\text{-}200}$, $N_{0\text{-}4}$, $O_{1\text{-}30}$, and $S_{0\text{-}2}$ (error ± 0.3 ppm). Formulae were classed according to elemental stoichiometry (CHO, CHON, CHOS, and CHONS). Double bond equivalent (DBE; an index for the degree of saturation), modified aromaticity index ($AI_{mod}$; an index for the degree of aromaticity; Koch & Dittmar, 2006; 2016) and NOSC were calculated for assigned formulae, as described elsewhere (LaRowe & Cappellen, 2011). Carboxylic-rich alicyclic molecules (CRAM-like) compounds were identified by ratios between DBE values and the number of carbon, hydrogen, and oxygen atoms in each formulae. Formulae which fell within the bounds: DBE/C, 0.30–0.68; DBE/H, 0.20–0.95 and DBE/O, 0.77–1.75 were classed as CRAM-like (Hertkorn *et al.*, 2006). Compound class assignments were made based on ratios between C, O, and H, and $AI_{mod}$ values, as follows: condensed aromatics ($AI_{mod}$>0.67); polyphenolics ($AI_{mod}$ values of 0.5–0.67); highly unsaturated and phenolic compounds (HUPs; $AI_{mod}$ <0.5, H/C <1.5); aliphatics (H/C ≥ 1.5, O/C ≤ 0.9 and $N = 0$); peptide-like (H/C ≥ 1.5, O/C ≤ 0.9 and $N > 0$); and sugar-like (H/C ≥ 1.5 and O/C > 0.9; Spencer *et al.*, 2014; McDonough *et al.*, 2020). All formulae and compound classes were calculated as percentage relative abundance (%RA; i.e., normalized to intensity of all assigned peaks). The combined concentrations of assigned sugar-like and peptide-like formulae were negligible (<0.3 %RA) and thus not discussed further. Compounds classes were also separated into low O/C (O/C <0.5) and high O/C (O/C ≥0.5) formulae.

### 2.4 Data analysis

The processed EEM and FT-ICR MS datasets, along with the geochemical data (Bassil *et al.*, 2024) were analysed using R (version 3.6.1, included "stats" package). Unpaired Welch *t*-tests were performed, using the "t.test" function, between the groundwater samples taken from the sand- and clay-dominated sites, for all key geochemical, optical and molecular-level DOM metrics that had normal distributions (normality was determined by the Shapiro-Wilk test, where variables that returned *p*-values <0.05 were rejected; Table 1).

To identify covariations between geochemical data (As, Fe and DOC concentrations), and optical and molecular-level DOM properties, principal component analysis (PCA) was performed. The optical and molecular parameters selected for PCA were the indices, ratios and % Ras; all single components with absolute measurements (e.g. $Abs_{254}$ and TPH-mean and max) were excluded. These parameters were centred and scaled and the PCAs were performed using the *rda* (redundancy analysis) function within the *vegan* package. The distribution of the parameters (shown by the arrows) was obtained using the *envfit* function. The significance of each parameter was determined by *p*-values above 0.05, which were also obtained using the *envfit* function. The *stat_elipse* function was used to draw ellipses around the samples grouped by site, at 95% confidence.

Linear regression was performed between Fe and As concentrations, and DOM compositional metrics (e.g., TPH:FA-like fluorescence, $HIX_{cor}$, and %RA of condensed aromatics and aliphatics) and between optical and molecular-level aromaticity metrics (i.e., $SUVA_{254}$ and $AI_{mod}$ , respectively). Pearson's correlations were reported for each regression and calculated using the "cor" and "cor.test" functions. The Welch t-tests and the Pearson's correlations were performed at 95% confidence level, and statistical significance was confirmed when the returned *p*-values were <0.05.

## 3. Results

### 3.1. Geochemical and optical DOM properties

As reported elsewhere, groundwater Fe and As concentrations ranged from 4.0 to 7.4 mg/L and 9.8 to 66 µg/L, respectively (Bassil *et al.*, 2024). Bulk DOC concentrations ranged from 5.5 to 14 mg/L. Higher concentrations of DOC were found in groundwater of the clay-dominated site than the groundwater of the sand-dominated site, with correspondingly lower As concentrations (Table 1 and Fig. 2; Bassil *et al.*, 2024). These values were largely comparable to those of previous studies of the region (e.g., Richards *et al.*, 2019b). Optical metrics showed groundwater DOM was distinct between the clay-dominated and sand-dominated sites, and the surface water sample (Table 1 and Fig. 2). The mean $SUVA_{254}$ values were 4.74 (2.52–6.32), 1.81 (1.52–2.12), and 2.40 L mg $C^{-1}$ $m^{-1}$ in the clay-dominated site groundwater, sand-dominated site groundwater, and surface water sample, respectively (Table 1). The single - component fluorescence parameters and TPH:FA ratios indicate an overall dominance (terrestrial-associated) HA and FA-like fluorescence components over (microbial-associated)

TPH-like components, in the groundwater and surface water. However, some microbial influence in the groundwater is suggested by the increased FI values in groundwater (1.41-1.51) compared to surface water (1.20) - which compare to end-member values of 1.55 and 1.21 for microbial-derived and terrestrial-derived fulvic acids, respectively (Cory *et al.,* 2010). The single component parameters (HA-like, FA-like, and TPH-like fluorescence) were higher and more heterogeneous at the clay-dominated site compared to the sand-dominated site (e.g., 2.82–9.41 RU and 0.77–0.86 RU mean FA-like fluorescence respectively; 1.04–3.53 RU and 0.20–0.24 RU mean TPH-like fluorescence respectively; Table 1 and Fig. 2). Indices that were higher in the groundwater DOM of the clay-dominated site compared to that of the sand-dominated included TPH:FA-like ratios (0.34–0.38 vs 0.25–0.28; $p<0.01$), $SUVA_{254}$ (2.52–6.32 vs 1.52–2.12 L mg $C^{-1}$ $m^{-1}$ ; $p=0.05$), FI (1.44–1.51 vs 1.31–1.41; $p=0.01$), and β:α (0.65–0.70 vs 0.51–0.60; $p<0.01$; Table 1). On average, $HIX_{corr}$ values were higher at the sand-dominated site compared to clay-dominated, though not statically significant (0.80–0.91 and 0.92–0.94 at clay-dominated and sand-dominated sites, respectively; Fig. 2 and Table 1). The TPH:FA and $HIX_{corr}$ values were higher in the surface water than in the groundwater, whether as the FI values were higher in the groundwater (Fig. 2 and Table 1). The surface water values for $SUVA_{254}$ and β:α were between those of the two groundwater sites Fig. 2 and Table 1).

### 3.2. Molecular-level DOM properties

Between 9,445 and 11,405 molecular formulae were assigned in the groundwater of the clay-dominated site, 9,839 to 10,345 molecular formulae in the groundwater of the sand-dominated site, and 8,473 molecular formulae in the surface water (Table 1), which may indicate higher molecular diversity in the groundwaters than in the surface water sample. The groundwater of the sand-dominated site had slightly higher H/C ratios (1.17–1.18) than the groundwater of the clay-dominated site (1.13–1.17; Table 1). The surface water had a lower H/C ratio (1.13) and higher O/C ratio (0.55) than the groundwaters (0.47–0.48; Table 1). Higher $AI_{mod}$ values were found in the groundwater of the clay-dominated site (0.28–0.30) than in the groundwater of the sand-dominated site (0.27–0.28), and the surface value (0.28) was between those of the groundwater sites (Table 1 and Fig.3). Overall NOSC values were negative in the groundwater (-0.19 to -0.13) and slightly positive in the surface water (0.04; Table 1 and Fig.3).

The most abundant formulae class was CHO (61.2–70.7 %RA in the groundwater, 63.6 %RA in the surface water), followed by CHON, CHOS and CHONS (Fig. 4A). The groundwater of the clay-dominated site and the surface water contained higher %RA of sulfur-containing formulae (CHOS, CHONS) than the groundwater of the sand-dominated site (10.4-12.1 %RA in surface water and clay-dominated site groundwater compared to 5.0-5.8 %RA in sand-dominated site groundwater; Table 1 and Fig. 4A). HUP compounds dominated the molecular-level composition of all samples (90.1–93.6 %RA in the groundwater, 86.1 %RA in the surface water), polyphenolics were the second most abundant compound class (3.3–6.9 %RA in the groundwater, 7.8 %RA in the surface water), followed by aliphatics (2.7–3.6 %RA in the groundwater, 4.6 %RA in the surface water) and condensed aromatics (<0.3–0.4 %RA in the groundwater, 1.2 %RA in the surface water; Table 1 and Fig. 4B). HUP and polyphenolic compounds in groundwater samples were dominated by low O/C ratios (<0.5, e.g. 51.5-56.6%RA low oxygen HUPs; Table 1), whilst those in the surface water sample predominantly had high O/C ratios (0.5 or higher; e.g. 59.4%RA HUPs; Table 1 and Fig. 4B). The groundwater also had significantly higher CRAM-like formulae (70.5–74.9%RA) compared to the surface water (52.9%RA; Fig. 3 and Table 1).

### 3.3. Covariations between geochemical, optical and molecular characteristics

Strong positive correlations were found between $SUVA_{254}$ and $AI_{mod}$ ($r = 0.97$; Fig. 5) as well as strong negative correlations between the $SUVA_{254}$ and CRAM-like relative abundances, though only for the groundwater samples ($p<0.01$; Fig. 5).

Overall, optical and molecular analysis showed a dominance of terrestrial-derived DOM in both clay- and sand-dominated sites (Fig. 2–4), although comparison of FI values indicate that the fulvic acids in the groundwater had a substantial microbial influence (Fig. 2). However, DOM composition differed between groundwaters, particularly clay-dominated groundwater had a higher freshness index and %RA of aliphatic compounds compared to sand-dominated groundwater. In order to understand what drives this variability, and identify relationships between DOM composition and As release, a principal component analysis was conducted (PCA). Two PCAs were performed given the substantial differences in composition between the surface water and groundwater (Fig. 2 and 3). The first PCA was conducted on all the samples collected (i.e., the groundwater and surface water samples; Fig. 6A) to identify the covariations in DOM composition and As concentrations between surface and groundwaters. The second PCA was performed on the groundwater samples only (Fig. 6B), in order to identify the differences between (and co-variations within) the clay-dominated and sand-dominated sites.

When the surface water sample is included, the clay-dominated site groundwater and surface water was differentiated from the sand-dominated site groundwater by PC1 (48.9% of total variance), whilst the surface water is differentiated from the clay-dominated site groundwater by PC2 (31.6% of total variance; Fig. 6A). The surface water sample is characterised by higher NOSC values, and an increased %RA of condensed aromatics and aliphatics, as well as a high O/C polyphenolics and high O/C HUP compounds (Fig. 6A). Fe and As concentrations were associated with positive values on PC1, and thus DOM with relatively higher low O/C HUP and CHO-only compounds, when all samples are included (Fig. 6A).

The PCA performed solely on the groundwater samples also showed differentiation by site along PC1 (56.5% of total variance), with those of the clay-dominated site on the negative side, and those of the sand-dominated site on the positive side (Fig. 6B). Furthermore, the groundwater-only PCA showed separation by TPH:FA-like fluorescence, β:α, FI, as well as %RA of sulfur-containing formulae (CHOS, CHONS), polyphenolics, O/C HUPs, aliphatics, and NOSC values (weighted by RA) on the negative side of PC1; and higher Fe concentrations, As concentrations, %RA of CRAMs, low O/C HUPs, CHO formulae on the positive side of PC1. PC2 (23.9% of total variance), which appeared to be driven by variables related to aromaticity (e.g., $SUVA_{254}$ and polyphenolics on the negative side of PC2, and low oxygen HUPs on the positive side of PC2; Fig. 6B). There was no clear site-separation along PC2, however there was wider variation in the clay-dominated site samples compared to the sand-dominated site samples along PC2 (Fig. 6B). Along PC1, higher Fe and As concentrations are weakly associated with the sand-dominated site groundwater (i.e., Welch t-tests show As concentrations are statistically similar between the groundwaters of the clay-dominated and sand-dominated sites; Table 1). Higher DOC concentrations appeared to be associated with the clay-dominated site groundwater (Fig. 6B), although this association is not statistically significant according to the Welch t-tests between the sites ($p=0.07$; Table 1). Furthermore, in both PCAs, high O/C HUPs, condensed aromatics, and $HIX_{corr}$ had *envfit p*-values of >0.05 (Fig. 6 and Box S1), indicating that these variables did not significantly contribute to overall variation, though may still be important with respect to the sites they are distributed with.

## 4. Discussion

### 4.1. DOM sources and microbial processing

The $SUVA_{254}$ values were comparable to waters of As-prone aquifers elsewhere, such as in the Hetao Basin (1.43 to 2.09 L mg $C^{-1}$ $m^{-1}$ surface water, 1.33 to 7.35 L mg $C^{-1}$ $m^{-1}$ in groundwater; Qiao *et al*., 2021) the Bengal Basin (2.48 to 3.23 L mg $C^{-1}$ $m^{-1}$ in surface water, 0.40 to 4.65 L mg $C^{-1}$ $m^{-1}$ in groundwater; Kulkarni *et al*., 2017), Datong Basin (0.72 L mg $C^{-1}$ $m^{-1}$ in surface water, 0.18 to 0.62 L mg $C^{-1}$ $m^{-1}$ in groundwater; Pi *et al*., 2015) and the Indus Delta (0 to 8.1 L mg $C^{-1}$ $m^{-1}$ in groundwater; Malik *et al*., 2020). Although dissolved Fe is known to increase absorbance at 254 nm (Poulin *et al*., 2014) and thus influence $SUVA_{254}$ values, a previous study on groundwaters from these Cambodian field sites compared shaken to unshaken samples to assess the potential impact of Fe oxyhydroxide particulates in the samples (brought into suspension by shaking) and found no significant difference in $Abs_{254}$ readings (Richards *et al.,* 2019b). In addition, in this study, no correlation was found between Fe concentrations and $SUVA_{254}$ values in the groundwater samples (Supplementary Fig. S2). Furthermore, the strong positive correlations between groundwater $SUVA_{254}$ values with independent molecular-level metrics of aromaticity $AI_{mod}$ ($r = 0.97$; Fig. 5), is in agreement with the use of $SUVA_{254}$ as a proxy for aromaticity (Weishaar *et al.*, 2003).

The higher heterogeneity of (optical and molecular) DOM signatures in the clay-dominated sites, compared to that of the sand-dominated site, is likely due to the lower recharge rates and lower connectivity between the wells in clay-dominated lithology (Uhlemann *et al.,* 2017; Richards *et al.,* 2017 and 2019b). The overall optical characteristics with respect to the groundwaters of the clay-dominated and sand-dominated sites (Fig. 2), were also very similar to those of previous studies in the Kandal Province, and the domination of humic and fulvic acid-like fluorescence over that of tryptophan-derived (associated with microbially-derived DOM) were consistent with the importance of surface-derived terrestrial organics on groundwater hydro-geochemistry (Richards *et al*., 2019b). At the molecular-level, the presence of polyphenolics in the groundwater was also in line with terrestrial, surface derived organics from plant-material into the groundwater (Hättenschwiler & Vitousek, 2000). The higher abundance of polyphenolics in the clay-dominated site groundwater compared to the sand-dominated site groundwater (Fig. 4) is also consistent with associations between clay-dominated aquifer sediments and plant-derived material, observed in previous studies of aquifers in Cambodia (Rowland *et al*., 2007; van Dongen *et al*., 2008; Magnone *et al*., 2017) and northern Vietnam (Al Lawati *et al*., 2012; Glodowska *et al*., 2020). This may indicate build-up of surface/plant-derived compounds in the clay-dominated aquifers and/or the preservation of plant-derived material in the clay-dominated sediments or pores, and mobilisation into the groundwater.

The similarity in % RA of sulfur-containing formulae (CHON and CHOS) between the surface sample and the clay-dominated site groundwater (Fig. 4A) possibly reflects the close proximity of the lake (where the surface water sample was taken) to the boreholes of the clay-dominated site. The lower O/C ratios, NOSC values, and higher H/C ratios and %RA of CRAM-like molecules in groundwater (from both sites) compared to surface water (Fig. 3 and 4), are likely explained by biodegradation and are broadly consistent with that of unconfined aquifers <41 m depth, in a recently proposed framework for DOM cycling in groundwater (McDonough *et al.*, 2022). Under this framework, aromatic DOM from the surface percolates into aquifers and biodegradation leads to a removal of compounds with a higher NOSC. This results in a decrease in O/C ratios and an increase in H/C ratios, as well as an increased %RA of CRAM-like molecules down flow path (McDonough *et al.*, 2022). The removal of these higher NOSC compounds during microbial degradation in the subsurface is consistent with OM-degradation studies of As-prone aquifer sediments from Bangladesh and the Hetao Basin, which have shown mean %RA weighted NOSC decrease during bioincubation, as compounds with a higher NOSC preferentially metabolized (Pracht *et al.*, 2018; Qiao *et al.*, 2020). A study in the As-prone Jianghan Plain, suggested that organic compounds with CRAM-like molecular formulae may also be reused in biogeochemical processes (including As release; Yu *et al.*, 2020). However, there is no clear evidence for the reuse of CRAM-like molecular formulae as electron donors in this study, although this cannot be completely excluded given the high %RA of CRAM-like molecular formulae in the groundwater compared to the surface (Fig. 3). Therefore, the differences in DOM molecular composition between the surface and groundwater observed in this study, particularly NOSC values, and in the relative abundance of high O/C formulae, CRAM-like molecular formulae (Fig. 6A), may reflect the microbial processing of DOM that has occurred since percolation from the surface into the aquifer. This microbial processing may also explain the microbial influence in groundwater DOM suggested by the FI values observed (Fig. 2; Cory *et al.,* 2010). Previous studies in the Kandal Province have shown groundwater DOC age at 15 m (i.e., the depth of sampling in this study) to be ~5960 years BP in the clay-dominated area (LR12–LR14 in Fig. 1; near the clay-dominated site boreholes in this study), and ~1240 years BP near the sand-dominated area (LR01 in Fig. 1; near the sand-dominated site boreholes in this study; Magnone *et al.*, 2019). Radiocarbon dating of DOC from previous studies at these sites combined with the molecular-level composition of DOM observed here indicate that the microbial processing of DOM may occur during these time-periods, even at the relatively shallow depth of 15 m. However, the differences in residence time between the clay-dominated site and sand-dominated site groundwaters (Magnone *et al.*, 2019) did not appear to have exerted a significant effect on the %RA of aliphatics and heteroatom (N and S) containing formulae (Table 1), which are thought to accumulate in anoxic groundwater (such as the groundwaters in this study), due to the lack of degradation (McDonough *et al.*, 2022).

Additionally, SOM may be an important DOM source and thus may influence the compositions observed here. Previous studies in the area have shown that inorganic carbon derived from oxidation of DOC predated modelled groundwater flow by hundreds of years, and thus SOM derived compounds may drive As release, opposed to surface-derived sources (Magnone *et al.*, 2017, 2019). The importance of SOM has further been highlighted in previous microcosm studies, which indicated microbial Fe(III)/As(V) reduction and As release where SOM was used as electron donor, including petroleum-derived compounds (Rowland *et al.*, 2007, 2009). The detection of petroleum-derived compounds in FT-ICR MS data is possible, but challenging to identify when DOM composition has multiple sources. It is possible that some organic compounds (derived from co-deposition or upwelling) could have been sorbed onto the sediment grains (Bauer & Blodau, 2006). It is also possible that some (higher NOSC) terrestrial-derived DOM compounds could have been sorbed onto aquifer sediments down the flow path, which has been suggested to occur in shallow aquifers (McDonough *et al.*, 2022). In the event of any sorption of organic compounds onto the sediments, they would not appear in the DOM pool, and it is possible that any degradation/oxidation occurred whilst bound to the sediments.

Sulfur-containing formulae (CHOS and CHONS) were the lowest in %RA of all the formula classes (e.g., CHOS formulae 5.0 to 12.1%; Table 1). This contrasts with that previously measured in the recharge waters and sediment porewaters of Bangladesh (47.3 to 58.9 %RA; Pracht *et al*., 2018) and is also marginally lower than values previously measured in aquifer waters of the Hetao Basin (7.0 to 16.0 %RA in groundwaters, 16.6 to 23.3 %RA in surface waters; calculated from Qiao *et al*., 2021). The relatively lower abundance of sulfur-containing formulae noted in this study is likely a reflection of the lower amounts of sulfur in the Kandal Province groundwaters (Richards *et al*., 2017), highlighting that sulfur species may not play a major role in the DOM cycling at these sites.

Nonetheless, the association between sulfur-containing formulae and tryptophan-like fluorescence found here may indicate biodegradation and the accumulation of S-containing organics, further indicating the importance of microbial processes in forming the DOM compositions observed at these sites (Fig. 6). This association has also been previously noted in the Hetao Basin groundwaters (Qiao *et al*., 2021), where microbial sulfate reduction and the production of sulfides allows for abiotic sorption or incorporation of S- into DOM (Qiao *et al*., 2021 and references therein). Furthermore, Pracht *et al.* (2018) showed an increase in organic sulfur compounds during incubation of aquifer material from Bangladesh. At these sites in the Kandal Province, sulfate-reducing conditions have previously been noted from modelling of mineral saturation, for instances supersaturation of pyrite, orpiment and realgar were observed at 6 and 9 m depth at near the sand-dominated site of this study (LR01), and 21 m depth near the clay-dominated site of this study (LR14; Fig. 1) but not at 15 m depth (at which the groundwater samples in this study were taken; Richards *et al*., 2017). Sulfur-containing formulae may suggest microbial processing of DOM (Spencer *et al*., 2019; Behnke *et al.,* 2021; Vaughn *et al.,* 2021). Furthermore, in the McDonough *et al* (2022) framework, heteroatom-containing formulae (i.e., N or S containing species) are thought to accumulate during biodegradation under suboxic or anoxic conditions and may also increase in abundance due to hydrogenation on unsaturated formulae and sulfurization (aided by sulfate reduction). As such, the higher %RA of CHOS and CHONS formulae at the clay-dominated site groundwater, than that of the sand-dominated groundwater (Table 1) might therefore suggest greater microbial cycling than that at the sand-dominated site, either through sulfate reduction or microbial cycling of the organic matter (Wagner *et al* 2015; Yuan *et al*., 2017). However, high abundances of S-containing formulae were also found in the (oxic) surface water (Table 1) which might suggest microbial production of S-containing compounds (McDonough *et al*., 2022), especially since %RA of aliphatics were higher in the surface water compared to that of the groundwater (4.6 %RA in surface water, compared to 2.7 to 2.6 %RA in the groundwater; Table 1). Incubations studies would be needed to further understand the impact of microbial degradation on the DOM at these aquifers and the relation to Fe, As and sulfur cycles thereof.

### 4.2. DOM composition and association to As release

Positive Pearson's correlations were found between groundwater As concentrations and indicators of terrestrial-derived OM and some indicators of relatively biostable DOM (e.g., $HIX_{corr}$ values; $r$=0.73; $p$=0.039; Supplementary Fig. S1). This is in line with DOM analysis of relatively shallow (<40 m depth) groundwater in the Hetao Basin that indicated positive correlations between As concentrations and %RA of biostable compounds (e.g., humics; Qiao *et al*., 2020, 2021; Wang et al., 2023). These observations are further supported by negative correlations between groundwater As concentrations and optical properties, such as TPH:FA-like fluorescence ($r$ = -0.78; $p$ = 0.023; Richards *et al*., 2019b; Qiao *et al*., 2020, 2021), as well as %RA of aliphatics ($r$= -0.82; $p$ = 0.013). Here, similar associations were observed between groundwater Fe concentrations and the aforementioned metrics (e.g., TPH:FA ratio and %RA of aliphatics), although these were not statistically significant (p >0.05; Supplementary Fig. S1). However, statistically significant positive Pearson's correlations did occur between groundwater Fe concentrations and abundances of condensed aromatics ($r$=0.74; $p$=0.038), but not As concentrations ($r$=0.49; $p$=0.213; Supplementary Fig. S1).

The negative associations between TPH:FA ratios and %RA of aliphatics with As and Fe concentrations (Fig. 6) may suggest the degradation of biolabile DOM in connection of Fe(III)/As(V) reduction and As release, and possibly the utilisation of stable compounds (e.g., humics) as electron shuttles in Fe(III)/As(V) reduction (Qiao *et al.*, 2020, 2021). The differences in Fe and As distributions in relation to organic variations (Supplementary Fig. S1 and Table S2) could be related to the different sorption behaviours of these elements to organic compounds and mineral surfaces. Although no data on the SOM characteristics and biodegradability is available from the boreholes used in this study, comparative molecular studies between (water-extractable) SOM and groundwater DOM in other regions indicated a relative higher biolability for the SOM (Qiao *et al.*, 2020). It is therefore likely that SOM could have acted as electron donors in Fe(III) and/or As(V) reduction and As release with the more stable end-products released into the groundwater DOM. No statistically significant correlations occurred between either Fe or As concentrations and ß:α (vs Fe, $p$=0.373; vs As, $p$=0.064) and NOSC (vs Fe, $p$=0.628; vs As, $p$=0.876; Supplementary Table S2), although typically higher NOSC and ß:α values were associated with lower As concentrations (Fig. 6). This likely reflects the complexity in groundwater flow and the respective sources of Fe, As and organic compounds present in the Kandal Province aquifers (Lawson *et al.*, 2013, 2016; Richards *et al.*, 2017, 2018; Magnone *et al.*, 2019), highlighting that more spatial and temporal (e.g., pre and post monsoon) analyses would be needed to substantiate the relationship between DOM composition and As and Fe release. Furthermore, the trends identified in this study and studies of shallow groundwaters (<40 m) of the Hetao Basin (Qiao et al., 2020, 2021; Wang et al., 2023) somewhat contrast with studies of deeper groundwaters (>40 m) in the Songen and Hetao Basins (>40 m), which have identified positive relationships between DOM biodegradability and As concentrations in some studied aquifers (Gao et al., 2023; Li et al., 2024). These results highlight the need to include multiple depths of sampling as well as spatial variation in future studies.

Overall, the present study indicates that the combined use of FT-ICR MS and optical analyses can yield valuable information about the composition of groundwater DOM in As-prone aquifers, and the relation thereof with As-release from sediments. However, larger sample sizes taken across wider transects (combined with models of groundwater flow) and incubation experiments on the samples, would be needed to confirm and decipher any potential trends between groundwater As and the molecular characteristics of groundwater DOM, along hydrogeological flow paths. Future studies would therefore benefit from application of FT-ICR MS alongside optical characterisation on As-prone aquifers, along groundwater flow transects at multiple depths with contrasting hydrogeochemical characteristics (e.g., As concentrations and methane release), to further understand the relationship between As release and carbon cycling in aquifers. High spatial and temporal resolution is possible through optical measurements due to low costs and sample volumes required (Kulkarni *et al.*, 2017; Richards *et al.*, 2019b), whilst FT-ICR MS provides unparalleled resolution (Spencer *et al.,* 2014, 2015; Derrien *et al.,* 2019; McDonough *et al.,* 2022) – the combined use of both methods therefore can provide "*the best of both worlds"*. Future studies should also include incubation experiments to directly measure groundwater DOM biolability in conjunction with molecular composition to robustly connect DOM composition and biolability to Fe, As (and sulfur) cycling in the environment. Such experiments could allow cross-referencing of changes in biogeochemical processes (i.e. DOM composition and As release) with microbial communities and functions. A detailed understanding of these processes at molecular-level would significantly aid the prediction of As cycling in vulnerable aquifers and therefore help improve groundwater management practices.

## 5. Conclusion

This study aimed to provide insights into the environmental sources and processing of groundwater DOM, and its role in As release in the aquifers of Kandal Province, Cambodia, by analysis and comparison between novel molecular and optical properties with geochemical data. Fluorescence indices suggest overall dominance of terrestrial-derived DOM in groundwaters in both the clay-dominated and sand-dominated aquifer sites (although a degree of microbial processing is suggested by the FI values), and potentially higher biodegradability in the groundwaters of the clay-dominated site. Fluorescence indices for aromaticity were validated by comparison to the molecular data and fluorescence indices for microbial-sources (e.g., high FI and high TPH:FA ratio) covaried with sulfur-containing formulae suggesting microbial processing of DOM. The combined molecular and fluorescence analyses shows that groundwater DOM in the shallow aquifers is influenced by a variety of sources (including terrestrial/surface-derived compounds, e.g., polyphenolics), which in turn may be influenced by subsurface microbial processing (suggested by high CRAM-like abundances and low NOSC values in groundwaters compared to surface). Negative correlations between dissolved As and Fe with TPH:FA and %RA of aliphatic compounds likely indicate degradation of biolabile compounds and accumulation of stable compounds, in connection with As release. This is consistent with the role of biolabile organic compounds as electron donors in As release, by reductive dissolution as well as possible electron shuttles at both sites. Together, these findings highlight the role DOM composition plays in microbially mediated As release.

## Author contributions

Oliver Moore: lead researcher, conducted fieldwork, analysed peak-picked and post-processed data, interpreted results, and wrote manuscript. Laura Richards: supported fieldwork, assisted in interpretation of EEM data, contextualisation, and manuscript review. Amy Holt: performed SPE-extractions on groundwater samples for FT-ICR MS, processed and post-processed data in PetroOrg, manuscript editing. Amy Mckenna: performed FT-ICR MS analysis at NHMFL facility. Robert Spencer: manuscript editing. Dan Lapworth: EEM analysis at BGS facility, processed and QA'd EEM data, manuscript review. Naji Bassil: supported fieldwork. David Polya: supported fieldwork, contextualisation, and manuscript review. Bart van Dongen: conceptual guidance, contextualisation, assisted in data interpretation, manuscript review. Jonathan Lloyd: contextualisation, manuscript review, funding acquisition (GOAM).

## Funding

This work was funded by NERC through a PhD studentship of the Manchester-Liverpool Earth Atmosphere and Ocean Doctoral Training Program (EAO DTP; NE/L002469/1) and supported by the NERC-funded GOAM project (NE/P01304X/1). Dan Lapworth was supported by the BGS International NC programme 'Geoscience to tackle Global Environmental Challenges'. LAR acknowledges a Dame Kathleen Ollerenshaw Fellowship and a NERC Exploring the Frontiers Award (NE/X010813/1). The molecular characterisation portion of this work was performed at the National High Magnetic Field Laboratory ICR User Facility, which is supported by the National Science Foundation Division of Chemistry and Division of Materials Research through DMR-2128556 and the State of Florida.

## Acknowledgements

We would like to thank Chivuth Kong, Pheaktra Sovann, all staff at the Royal University of Phnom Penh and the Resource Development International Cambodia (RDIC), as well as all drivers and borehole drillers, who helped facilitate the fieldwork. We would also like to thank Daren Gooddy and Peter Williams for their assistance in facilitating the EEM analysis at BGS, Wallingford, UK. Dan Lapworth publishes with the permission of the Executive Director of the British Geological Survey. We thank the anonymous reviewers for their feedback, which has contributed to the improvement of this manuscript.

## Conflicts of interest

No conflict of interest to declare.

## Figure captions

**Fig. 1**. Location of the boreholes in Kandal Province, Cambodia, drilled in January 2019 and sampled in January 2020 for this study and parallel studies (Bassil *et al*., 2024). Yellow star marks the boreholes of the sand-dominated site (*n*=4), orange star marks the boreholes of the clay-dominated (*n=4)* site. LR marks transects of previous hydrogeochemical surveys in the area (from 2014 to 2015), including EEM data (Magnone *et al*., 2017, 2019; Richards *et al*., 2017, 2018, 2019b) and previous electrical resistivity surveys (Uhlemann *et al*., 2017). Adapted from Richards *et al*. (2017). The surface water sample was collected from a lake near the boreholes of the clay-dominated site. Surface water was additionally collected from a seasonal pond/wetland located in very close proximity to the clay site and the sample location falls within the same orange star marked for that site.

**Fig. 2.** Key physiochemical and optical properties of the groundwater dissolved organic matter (DOM) sampled from the clay-dominated (Clay) and sand-dominated (Sand) aquifer sites in Kandal Province, Cambodia – sampled in January 2020. Total As and dissolved organic carbon (DOC) were calculated from Bassil *et al* (2024). Specific ultraviolet absorbance ($SUVA_{254}$; L mg $C^{-1}m^{-1}$) was calculated from the absorbance at 254 nm divided by the DOC, and is a proxy for aromaticity (Weishaar *et al*., 2003). FA-like = fulvic acid-like fluorescence. TPH-like = tryptophan-like fluorescence. ß:α = "freshness index", the ratio between recently derived DOM ("β") and more decomposed DOM ("α") (Parlanti *et al*., 2000; Wilson & Xenopoulos, 2009;

Fellman *et al*., 2010; Kulkarni *et al*., 2017). $HIX_{corr}$ = humification index with inner filtration correction (Ohno, 2002). FI = fluorescence index, with threshold values for microbial and terrestrial-derived DOM indicated on figure (McKnight *et al*., 2001). Boxplots show groundwater sites, separated by stratigraphy, and the surface water (SW) sample value is marked by the broken line. The outcome of Welch t-test (95% confidence) between the groundwaters of the clay-dominated and sand-dominated sites is indicated by the *p*-values; significance is indicated when $p<0.05$ (this could not be performed on the $HIX_{corr}$ due to non-normal distribution).

**Fig. 3.** Modified aromaticity index ($AI_{mod}$), nominal oxidation state of carbon (NOSC), and the relative abundances (%RA) of carboxyl-rich acyclic molecules (CRAM-like), in the aquifer groundwaters of the clay-dominated (Clay) and sand-dominated (Sand) sites ($n$=4 at each site) and a single surface water sample (SW; $n$=1) in the Kandal Province, Cambodia. The outcome of Welch t-test (95% confidence) between the groundwaters of the clay-dominated and sand-dominated sites is indicated by the *p*-values, significance is indicated when $p<0.05$ (this could not be determined on NOSC values due to non-normal distributions).

**Fig. 4.** Heteroatom classes (A) and compound classes (B) by relative abundance (RA) within the groundwaters of the clay-dominated (Clay; $n$=4) and sand-dominated (Sand; $n$=4) boreholes and the surface water sample ($n$=1). HUPs = highly unsaturated and phenolic compounds. Low O/C = <0.5; high O/C ≥0.5.

**Fig. 5.** Pearson's correlations between specific ultraviolet absorbance at 254 nm ($SUVA_{254}$), and modified aromaticity index ($AI_{mod}$; top) and relative abundance (%RA) of carboxyl-rich alicyclic-like molecules (CRAM-like; bottom). Line of best fit and r values are for groundwater samples only. Clay GW = groundwater samples from the clay-dominated site ($n$=4). Sand GW = groundwater samples from the sand-dominated site ($n$=4).

**Fig. 6.** Principal component analysis (PCA) of Fe, As and dissolved organic carbon (DOC) concentrations, along with optical and FT-ICR MS metrics. (A) Groundwater and surface water samples. (B) Groundwater samples only. Red circles show close parameter co-variations; blue circles show the respective plots locations of the clay-dominated site (negative side of PC1 in both PCAs) and sand-dominated site (positive side of PC1 in both PCAs) groundwater samples, and the surface water sample (negative side of PC1 and PC2 in first PCA). $SUVA_{254}$ = specific ultraviolet absorbance at 254 nm ($Abs_{254}$/DOC). TPH:FA-like = the ratio between tryptophan-like and fulvic acid-like fluorescence. $HIX_{corr}$ = humification index with inner filtration correction (Ohno, 2002). ß:α = "freshness index", the ratio between recently derived DOM ("β") and more decomposed DOM ("α") (Parlanti *et al*., 2000; Wilson & Xenopoulos, 2009; Fellman *et al*., 2010; Kulkarni *et al*., 2017). HUPs = highly unsaturated and phenolic compounds. $AI_{mod}$= modified aromaticity index (Koch & Dittmar, 2006, 2016). NOSC = nominal oxidation state of carbon (LaRowe & Cappellen, 2011). Clay GW = groundwater samples from the clay-dominated site ($n$=4). Sand GW = groundwater samples from the sand-dominated site ($n$=4).

**Tables**

**Table 1** Physicochemical and dissolved organic matter (DOM) characteristics of aquifer groundwater and surface water sampled in the Kandal Province, Cambodia, in January 2020. Values are reported as means with ranges given in parentheses.

| | Clay-dominated site groundwater mean ($n$=4) | Sand-dominated site groundwater mean ($n$=4) | Surface water ($n$=1) | $p$ value (Welch t-test at 95% confidence)[b] | Richards *et al.*, (2019b) at 15 m[c] |
|---|---|---|---|---|---|
| | | **Water geochemistry [a]** | | | |
| As (µg/L) | 33 (9.8–59) | 61 (55–66) | 4 | >0.05 (0.12) | 15–552 |
| Fe (mg/L) | 5.7 (4.0–7.4) | 6.6 (5.7–7.3) | 0 | >0.05 (0.33) | 0.6–27 |
| DOC (mg/L)[d] | 8.4 (5.5–14) | 2.9 (2.3–3.5) | 11 | >0.05 (0.07) | 0.9–12 |
| | | **Fluorescence properties (EEM)** | | | |
| | | **Absorbance** | | | |
| $Abs_{254}$ ($m^{-1}$)[e] | 39.0 (14.1–57.4) | 5.24 (4.96–5.50) | 26.9 | <0.05 (0.04) | |
| $SUVA_{254}$ (L mg $C^{-1}$ $m^{-1}$)[f] | 4.74 (2.52–6.32) | 1.18 (1.52–2.12) | 2.40 | <0.05 (0.05) | ~0.0–12.8 |
| TPH-like mean (RU)[g] | 2.53 (1.04–3.53) | 0.22 (0.20–0.24) | 0.93 | <0.05 (0.03) | 0.25–3.62 |
| TPH-like max (RU)[g] | 3.06 (1.26–4.28) | 0.28 (0.26–0.29) | 1.09 | <0.05 (0.03) | |
| FA-like mean (RU)[h] | 7.06 (2.82–9.41) | 0.82 (0.77–0.86) | 2.14 | <0.05 (0.03) | 0.85–7.92 |
| FA-like max (RU)[h] | 7.47 (3.01–9.99) | 0.89 (0.84–0.96) | 2.33 | <0.05 (0.03) | |
| HA-like mean (RU)[i] | 4.43 (1.67–6.09) | 0.58 (0.55–0.63) | 1.41 | <0.05 (0.03) | |
| HA-like max (RU)[i] | 4.77 (1.79–6.65) | 0.63 (0.58–0.67) | 1.49 | <0.05 (0.03) | |
| | | **Fluorescence indices** | | | |

| | | | | | |
|---|---|---|---|---|---|
| FI[j] | 1.47 (1.44–1.51) | 1.31–1.41 (1.37) | 1.20 | <0.05 (0.01) | 1.23–1.55 |
| TPH:FA[k] | 0.36 (0.34–0.38) | 0.27 (0.25–0.28) | 0.43 | <0.05 (0.00) | 0.27–0.46 |
| HIX[l] | 8.12 (3.96–10.5) | 13.2 (11.8–14.8) | 11.92 | <0.05 (0.03) | |
| HIX (corr)[m] | 0.88 (0.80–0.91) | 0.93 (0.92–0.94) | 0.92 | - | 0.89–0.95 |
| ß/α[n] | 0.67 (0.65–0.70) | 0.56 (0.51–0.60) | 0.59 | <0.05 (0.00) | 0.54–0.74 |
| **<u>Molecular properties (FT-ICR MS)</u>** | | | | | |
| Number of assigned formulae | 10670 (9445–11405) | 10142 (9839–10345) | 8473 | >0.05 (0.31) | |
| Average mass (weighted by RA)[o] | 554 (544–560) | 556 (550–559) | 487 | >0.05 (0.76) | |
| **Ratios with carbon (weighted by RA)** | | | | | |
| O/C | 0.47 (0.47–0.48) | 0.47 (0.47–0.48) | 0.55 | >0.05 (0.75) | |
| N/C | 0.01 (0.01–0.01) | 0.02 (0.01–0.02) | 0.02 | <0.05 (0.00) | |
| H/C | 1.15 (1.13–1.17) | 1.18 (1.17–1.18) | 1.13 | - | |
| **Indices (weighted by RA)[o]** | | | | | |
| $AI_{mod}$[s] | 0.29 (0.28–0.30) | 0.28 (0.27–0.28) | 0.28 | <0.05 (0.04) | |
| NOSC[t] | -0.15 (-0.19–-0.13) | -0.18 (-0.19–-0.17) | 0.04 | - | |
| **Formula classes (%RA)[o]** | | | | | |
| CHOS | 11.2 (10.4–12.1) | 5.5 (5.0–5.8) | 11.5 | <0.05 (0.00) | |

| | | | | |
|---|---|---|---|---|
| CHONS | 1.9 (1.4–2.2) | 0.4 (0.3–0.5) | 1.9 | <0.05 (0.00) |
| CHON | 23.1 (21.9–24.6) | 23.4 (22.7–23.8) | 23.0 | - |
| CHO | 63.8 (62.1–65.5) | 70.7 (70.5–71.0) | 63.6 | <0.05 (0.02) |
| **Compound classes (%RA)[o]** | | | | |
| Aliphatics | 3.2 (3.0–3.6) | 2.7 (2.7–2.8) | 4.6 | - |
| PPs (low oxygen)[p] | 3.9 (3.1–4.8) | 2.0 (1.8–2.2) | 2.20 | <0.05 (0.02) |
| PPs (high oxygen)[p] | 1.8 (1.2–2.2) | 1.5 (1.3–1.7) | 5.63 | >0.05 (0.21) |
| HUPs (low oxygen)[q] | 53.0 (51.5–56.7) | 55.8 (54.5–56.6) | 26.7 | - |
| HUPs (high oxygen)[q] | 37.9 (35.2–40.2) | 37.5 (36.8–38.5) | 59.4 | >0.05 (0.74) |
| Condensed aromatics | <0.3 | <0.3–0.4 | 1.2 | >0.05 (0.29) |
| CRAM-like [o,r] | 72.1 (70.5–74.3) | 74.4 (73.6–74.9) | 52.2 | >0.05 (0.06) |

[a] Geochemical values averaged from the technical duplicates of each sample from Bassil *et al* (2024); [b] *p*-values calculated by the Welch t-test, to compare the differences between the distributions of the groundwater between the clay- and sand-dominated sites, *p*-values <0.05 indicate statistically significant differences at 95% confidence level, variables marked with a hyphen (-) indicate where samples were not normally distributed and thus Welch t-test could be performed; [c] Groundwater EEM values from samples taken at 15 m depth in a previous study of the Kandal Province (Richards *et al.*, 2019b) included for comparison ($SUVA_{254}$ values are averaged across all samples from the previous study); [d] DOC = total dissolved organic carbon; [e] Abs254 = absorbance at 254 nm; [f] $SUVA_{254}$ = specific ultraviolet aborbance ($Abs_{254}$/DOC); [g] TPH-like = tryptophan-like fluorescence; [h] FA-like = fulvic acid-like fluorescence; [I] HA-like = humic acid-like fluorescence; [j] FI = fluorescence index (McKnight *et al.*, 2001); [k] TPH:FA-like = the ratio between TPH-like and FA-like fluorescence; [l] HIX = humification index (Zsolnay *et al.*, 1999). [m] HIX (corr) = humification index with inner filtration correction (Ohno, 2002); [n]. β:α= “freshness index”, the ratio between freshly produced DOM (microbial-derived or produced in situ, “β”) and more processed, biostable DOM (i.e., terrestrial derived carbon compounds, “α”; Parlanti *et al.*, 2000; Wilson & Xenopoulos, 2009; Kulkarni *et al.*, 2017).

%RA = percentage relative abundance; [p]. PPs = polyphenolics; [q] HUPs = highly unsaturated and phenolic; [r]CRAM-like = carboxyl-rich acyclic-like molecules (Hertkorn *et al.*, 2006); [s].$AI_{mod}$= modified aromaticity index (Koch & Dittmar, 2006, 2016); [t]. NOSC = nominal oxidation state of carbon (LaRowe & Cappellen, 2011).

**Data availability**

The datasets from the high resolution 21 Tesla FT-ICR MS analysis are available in an OSF database (https://osf.io/j8yns/). The fluorescence data is available on request.

**References**

Agilent Technologies. Agilent Bond Elut PPL. https://www.agilent.com/cs/library/usermanuals/public/bond-elut-ppl-5994-4162en-agilent.pdf, accessed: 17/05/2024.

Bassil, N. M., Richards, L. A., Kong, C., Moore, O., van Dongen, B. E., Polya, D. A., Lloyd, J. R. 2024. Groundwater geochemical data from two sites in Cambodia, southeast of Phnom Penh, 2019 - 2020. NERC EDS National Geoscience Data Centre. (Dataset). https://doi.org/10.5285/6bb55fd0-c4ef-4044-8acc-d6641c19e00c

Bauer, M., Blodau, C., 2006. Mobilization of arsenic by dissolved organic matter from iron oxides, soils and sediments. Science of the Total Environment 354, 179-90.

Behnke, M.I., McClelland, J.W., Tank, S.E., Kellerman, A.M., Holmes, R.M., Haghipour, N., Eglinton, T.I., Raymond, P.A., Suslova, A., Zhulidov, A.V. and Gurtovaya, T., 2021. Pan-arctic riverine dissolved organic matter: Synchronous molecular stability, shifting sources and subsidies. Global Biogeochemical Cycles 35, e2020GB006871.

Benner, S.G., Polizzotto, M.L., Kocar, B.D., Ganguly, S., Phan, K., Ouch, K., Sampson, M., Fendorf, S., 2008. Groundwater flow in an arsenic-contaminated aquifer, Mekong Delta, Cambodia. Applied Geochemistry 23, 3072–3087.

Blakney, G.T., Hendrickson, C.L., Marshall, A.G., 2011. Predator data station: A fast data acquisition system for advanced FT-ICR MS experiments. International Journal of Mass Spectrometry 306, 246–252.

Charlet, L., Polya, D.A., 2006. Arsenic in Shallow, Reducing Groundwaters in Southern Asia: An Environmental Health Disaster. Elements 2, 91–96.

Chen, W., Westerhoff, P., Leenheer, J.A., Booksh, K., 2003. Fluorescence Excitation-Emission Matrix Regional Integration to Quantify Spectra for Dissolved Organic Matter. Environmental Science & Technology 37, 5701–5710.

Chen, Z., Wang, Y., Jiang, X., Fu, D., Xia, D., Wang, H., Dong, G., Li, Q., 2017. Dual roles of AQDS as electron shuttles for microbes and dissolved organic matter involved in arsenic and iron mobilization in the arsenic-rich sediment. Science of the Total Environment 574, 1684–1694.

Coble, P., 1996. Characterization in marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. Marine Geochemistry 51 (4), 325-346.

Corilo, Y., 2014. PetroOrg Software. https://www.research.fsu.edu/research-offices/oc/technologies/petroorg-software, accessed: 10/02/2024.

Cory, R.M., Miller, M.P., McKnight, D.M., Guerard, J.J. and Miller, P.L., 2010. Effect of instrument-specific response on the analysis of fulvic acid fluorescence spectra. Limnology and Oceanography: Methods 8, 67-78.

D'Andrilli, J., Cooper, W.T., Foreman, C.M., Marshall, A.G. 2015. An ultrahigh-resolution mass spectrometry index to estimate natural organic matter lability. Rapid Communications in Mass Spectrometry 29, 2385–2401.

Derrien, M., Brogi, S.R., Gonçalves-Araujo, R., 2019. Characterization of aquatic organic matter: Assessment, perspectives and research priorities. Water Research 163, 114908.

Dittmar, T., Koch, B., Hertkorn, N., Kattner, G. 2008. A simple and efficient method for the solid-phase extraction of dissolved organic matter (SPE-DOM) from seawater. Limnology and Oceanography: Methods 6, 230–235.

Dittmar, T., Stubbins, A., 2014. Dissolved Organic Matter in Aquatic Systems. In: Turekian, K.K., Holland, H.D. (Ed.), Treatise on Geochemistry 2nd Edition. Elsevier Science, pp. 125-156.

van Dongen, B.E., Rowland, H., Gault, A.G., Polya, D., Bryant, C., Pancost, R.D., 2008. Hopane, sterane and n-alkane distributions in shallow sediments hosting high arsenic groundwaters in Cambodia. Applied Geochemistry 23, 3047–3058.

Du, Y., Deng, Y., Ma, T., Xu, Y., Tao, Y., Huang, Y., Liu, R., Wang, Y. 2020. Enrichment of Geogenic Ammonium in Quaternary Alluvial–Lacustrine Aquifer Systems: Evidence from Carbon Isotopes and DOM Characteristics. Environmental Science & Technology 54, 6104–6114.

Ebdon, D., 1985. Statistics in geography. In: Statistics in geography, 2nd ed., r. Blackwell, Oxford.

Fakour, H., Lin, T.F., 2014. Experimental determination and modeling of arsenic complexation with humic and fulvic acids. Journal of Hazardous Materials 279, 569–578.

Fendorf, S., Michael, H., van Geen, A., 2010. Spatial and temporal variations of groundwater arsenic in South and Southeast Asia. Science 328, 1123–1127.

Fellman, J.B., Hood, E., Spencer, R.G., 2010. Fluorescence spectroscopy opens new windows into dissolved organic matter dynamics in freshwater ecosystems: A review. Limnology and Oceanography 55, 2452-2462.

Gao, Z., Guo, H., Qiao, W., He, C., Shi, Q., Ke, T., Cao, Y. Dissolved Organic Matter Sources in High Arsenic Groundwater From a Sand-Gravel Confined Aquifer. Journal of Geophysical Research: Biogeosciences, 128, e2022JG007178.

Ghosh, D., Routh, J., Bhadury, P., 2015. Characterization and microbial utilization of dissolved lipid organic fraction in arsenic impacted aquifers (India). Journal of Hydrology 527, 221–233.

Glodowska, M., Stopelli, E., Schneider, M., Lightfoot, A., Rathi, B., Straub, D., Patzner, M., Duyen, V.T., Berg, M., Kleindienst, S., Kappler, A., 2020. Role of in Situ Natural Organic Matter in Mobilizing As during Microbial Reduction of $Fe^{III}$ -Mineral-Bearing Aquifer Sediments from Hanoi (Vietnam). Environmental Science & Technology 54, 4149–4159.

Guo, H., Zhang, B., Li, Y., Berner, Z., Tang, X., Norra, S., Stüben, D. 2011. Hydrogeological and biogeochemical constrains of arsenic mobilization in shallow aquifers from the Hetao basin, Inner Mongolia. Environmental pollution 159, 876–883.

Harvey, C.F., Swartz, C.H., Badruzzaman, A.B.M., Keon-Blute, N., Yu, W., Ali, M.A., Jay, J., Beckie, R., Niedan, V., Brabander, D., Oates, P.M., Ashfaque, K.N., Islam, S., Hemond, H.F., Ahmed, M.F., 2002. Arsenic mobility and groundwater extraction in Bangladesh. Science 298, 1602–1606.

Hättenschwiler, S., Vitousek, P.M., 2000. The role of polyphenols in terrestrial ecosystem nutrient cycling. Trends in Ecology & Evolution 15, 238–243.

Hendrickson, C.L., Quinn, J.P., Kaiser, N.K., Smith, D.F., Blakney, G.T., Chen, T., Marshall, A.G., Weisbrod, C.R., Beu, S.C., 2015. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. Journal of the American Society for Mass Spectrometry 26, 1626–1632.

Hertkorn, N., Benner, R., Frommberger, M., Schmitt-Kopplin, P., Witt, M., Kaiser, K., Kettrup, A., Hedges, J.I., 2006. Characterization of a major refractory component of marine dissolved organic matter. Geochimica et Cosmochimica Acta 70, 2990–3010.

Hertkorn, N., Ruecker, C., Meringer, M., Gugisch, R., Frommberger, M., Perdue, E.M., Witt, M., Schmitt-Kopplin, P., 2007. High-precision frequency measurements: indispensable tools at the core of the molecular-level analysis of complex systems. Analytical and Bioanalytical Chemistry 389, 1311–1327.

Hockaday, W.C., Purcell, J.M., Marshall, A.G., Baldock, J.A., Hatcher, P.G. 2009. Electrospray and photoionization mass spectrometry for the characterization of organic matter in natural waters: a qualitative assessment. Limnology and Oceanography: Methods 7, 81–95.

Hodgkins, S.B., Tfaily, M.M., Podgorski, D.C., McCalley, C.K., Saleska, S.R., Crill, P.M., Rich, V.I., Chanton, J.P., Cooper, W.T. 2016. Elemental composition and optical properties reveal changes in dissolved organic matter along a permafrost thaw chronosequence in a subarctic peatland. Geochimica et Cosmochimica Acta 187, 123–140.

Hughes, M.F., Beck, B.D., Chen, Y., Lewis, A.S., Thomas, D.J., 2011. Arsenic exposure and toxicology: A historical perspective. Toxicological Sciences 123, 305–332.

Islam, F.S., Gault, A.G., Boothman, C., Polya, D.A., Chamok, J.M., Chatterjee, D., Lloyd, J.R., 2004. Role of metal-reducing bacteria in arsenic release from Bengal delta sediments. Nature 430, 68–71.

Kellerman, A.M., Kothawala, D.N., Dittmar, T., Tranvik, L.J., 2015. Persistence of dissolved organic matter in lakes related to its molecular characteristics. Nature Geoscience 8, 454–457.

Kellerman, A.M., Guillemette, F., Podgorski, D.C., Aiken, G.R., Butler, K.D., Spencer, R.G.M., 2018. Unifying Concepts Linking Dissolved Organic Matter Composition to Persistence in Aquatic Ecosystems. Environmental Science & Technology 52, 2538–2548.

Kinniburgh, D., Smedley, P., 2001. Arsenic contamination of groundwater in Bangladesh. British Geological Survey Technical Report WC/00/19.Kocar, B.D., Polizzotto, M.L., Benner, S.G., Ying, S.C., Ung, M., Ouch, K., Samreth, S., Suy, B., Phan, K., Sampson, M., Fendorf, S. 2008. Integrated biogeochemical and hydrologic processes driving arsenic release from shallow sediments to groundwaters of the Mekong delta. Applied Geochemistry 23, 3059–3071.

Kocar, B.D., Benner, S.G., Fendorf, S., 2014. Deciphering and predicting spatial and temporal concentrations of arsenic within the Mekong Delta aquifer. Environmental Chemistry 11, 579–594.

Koch, B.P., Dittmar, T., 2006. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Communications in Mass Spectrometry 20, 926–932.

Koch, B.P., Dittmar, T., 2016. From mass to structure: an aromaticity index for high-resolution mass data of natural organic matter. Rapid Communications in Mass Spectrometry 30, 250.

Kulkarni, H.V., Mladenov, N., Johannesson, K.H., Datta, S., 2017. Contrasting dissolved organic matter quality in groundwater in Holocene and Pleistocene aquifers and implications for influencing arsenic mobility. Applied Geochemistry 77, 194–205.

Lapworth, D.J., Kinniburgh, D.G., 2009. An R script for visualising and analysing fluorescence excitation–emission matrices (EEMs). Computers & Geosciences 35, 2160–2163.

LaRowe, D.E., van Cappellen, P., 2011. Degradation of natural organic matter: A thermodynamic analysis. Geochimica et Cosmochimica Acta 75, 2030–2042.

Al Lawati, W.M., Rizoulis, A., Eiche, E., Boothman, C., Polya, D.A., Lloyd, J.R., Berg, M., Vasquez-Aguilar, P., van Dongen, B.E.,2012. Characterisation of organic matter and microbial communities in contrasting arsenic-rich Holocene and arsenic-poor Pleistocene aquifers, Red River Delta, Vietnam. Applied Geochemistry 27, 315–325.

Al Lawati., W.M., Jean, J.S., Kulp, T.R., Lee, M.K., Polya, D.A., Liu, C.C., van Dongen, B.E., 2013. Characterisation of organic matter associated with groundwater arsenic in reducing aquifers of southwestern Taiwan. Journal of Hazardous Materials 262, 970–979.

Lawson, M., Polya, D.A., Boyce, A.J., Bryant, C., Mondal, D., Shantz, A., Ballentine, C.J., 2013. Pond-derived organic carbon driving changes in arsenic hazard found in asian groundwaters. Environmental Science & Technology 47, 7085–7094.

Lawson, M., Polya, D.A., Boyce, A.J., Bryant, C., Ballentine, C.J. 2016. Tracing organic matter composition and distribution and its role on arsenic release in shallow Cambodian groundwaters. Geochimica et Cosmochimica Acta 178, 160–177.

Li, Y., Guo, H., Zhao, B., Gao, Z., Yu, C., Zhang, C., Wu, X., 2024. High biodegradability of microbially-derived dissolved organic matter facilitates arsenic enrichment in groundwater: Evidence from molecular compositions and structures. Journal of Hazardous Materials 470, 134133.

Magnone, D., Richards, L.A., Polya, D.A., Bryant, C., Jones, M., van Dongen, B.E.2017. Biomarker-indicated extent of oxidation of plant-derived organic carbon (OC) in relation to geomorphology in an arsenic contaminated Holocene aquifer, Cambodia. Scientific Reports 7, 13093.

Magnone, D., Richards, L.A., van Dongen, B.E., Bryant, C., Evans, J.A., Polya, D.A., 2019. Calculating 14C mean residence times of inorganic carbon derived from oxidation of organic carbon in groundwater using the principles of 87Sr/86Sr and cation ratio mixing. Geochimica et Cosmochimica Acta 267, 322–340.

Malik, A., Parvaiz, A., Mushtaq, N., Hussain, I., Javed, T., Rehman, H.U., Farooqi, A., 2020. Characterization and role of derived dissolved organic matter on arsenic mobilization in alluvial aquifers of Punjab, Pakistan. Chemosphere 251, 126374, 1-15.

Mao, R., Guo, H., Xiu, W., Yang, Y., Huang, X., Zhou, Y., Li, X., Jin, J., 2018. Characteristics and compound-specific carbon isotope compositions of sedimentary lipids in high arsenic aquifers in the Hetao basin, Inner Mongolia. Environmental pollution 241, 85–95.

McArthur, J.M., Banerjee, D.M., Hudson-Edwards, K.A., Mishra, R., Purohit, R., Ravenscroft, P., Cronin, A., Howarth, R.J., Chatterjee, A., Talukder, T., Lowry, D., Houghton, S., Chadha, D.K., 2004. Natural organic matter in sedimentary basins and its relation to arsenic in anoxic ground water: The example of West Bengal and its worldwide implications. Applied Geochemistry 19, 1255–1293.

McDonough, L.K., O'Carroll, D.M., Meredith, K., Andersen, M.S., Brügger, C., Huang, H., Rutlidge, H., Behnke, M.I., Spencer, R.G.M., McKenna, A., Marjo, C.E., Oudone, P., Baker, A., 2020. Changes in groundwater dissolved organic matter character in a coastal sand aquifer due to rainfall recharge. Water Research 169, 115201.

McDonough, L.K., Andersen, M.S., Behnke, M.I., Rutlidge, H., Oudone, P., Meredith, K., O'Carroll, D.M., Santos, I.R., Marjo, C.E., Spencer, R.G.M., McKenna, A.M., Baker, A. 2022. A new conceptual framework for the transformation of groundwater dissolved organic matter. Nature Communications 13, 1–11.

McKnight, D.M., Boyer, E.W., Westerhoff, P.K., Doran, P.T., Kulbe, T., Andersen, D.T. 2001. Spectrofluorometric characterization of dissolved organic matter for indication of precursor organic material and aromaticity. Limnology and Oceanography 46, 38–48.

Melton, E.D., Swanner, E.D., Behrens, S., Schmidt, C., Kappler, A. 2014. The interplay of microbially mediated and abiotic reactions in the biogeochemical Fe cycle. Nature Reviews Microbiology 12, 797–808.

Mladenov, N., Zheng, Y., Miller, M.P., Nemergut, D.R., Legg, T., Simone, B., Hageman, C., Rahman, M.M., Ahmed, K.M., Mcknight, D.M. 2010. Dissolved organic matter sources and consequences for iron and arsenic mobilization in Bangladesh aquifers. Environmental Science & Technology 44, 123–128.

Moore, O.C., Xiu, W., Guo, H., Polya, D.A., van Dongen, B.E., Lloyd, J.R. 2023. The role of electron donors in arsenic-release by redox-transformation of iron oxide minerals – A review. Chemical Geology 619, 121322.

Mukherjee, A., Scanlon, B.R., Fryar, A.E., Saha, D., Ghosh, A., Chowdhuri, S., Mishra, R., 2012. Solute chemistry and arsenic fate in aquifers between the Himalayan foothills and Indian craton (including central Gangetic plain): Influence of geology and geomorphology. Geochimica et Cosmochimica Acta 90, 283–302.

Mukhopadhyay, D., Sanyal, S.K., 2004. Complexation and release isotherm of arsenic in arsenic-humic/fulvic equilibrium study. Soil Research 42, 815.

Neumann, R.B., Ashfaque, K.N., Badruzzaman, A.B.M., Ali, M.A., Shoemaker, J.K., Harvey, C.F., 2010. Anthropogenic influences on groundwater arsenic concentrations in Bangladesh. Nature Geoscience 3, 46–52.

Neumann, R.B., Pracht, L.E., Polizzotto, M.L., Borhan, M.A., Badruzzaman, M., Ashraf Ali, M., 2014. Biodegradable Organic Carbon in Sediments of an Arsenic-Contaminated Aquifer in Bangladesh. Environmental Science &Technology Letters 1, 221–225.

Nevin, K.P., Lovley, D.R., 2002. Mechanisms for Fe(III) oxide reduction in sedimentary environments. Geomicrobiology Journal 19, 141–159.

Ohno, T., 2002. Fluorescence Inner-Filtering Correction for Determining the Humification Index of Dissolved Organic Matter. Environmental Science & Technology 36, 742–746.

Ohno, T., He, Z., Sleighter, R.L., Honeycutt, C.W., Hatcher, P.G., 2010. Ultrahigh Resolution Mass Spectrometry and Indicator Species Analysis to Identify Marker Components of Soil- and Plant Biomass-Derived Organic Matter Fractions. Environmental Science & Technology 44, 8594–8600.

Omoregie, E.O., Couture, R.M., van Cappellen, P., Corkhill, C.L., Charnock, J.M., Polya, D.A., Vaughan, D., Vanbroekhoven, K., Lloyd, J.R., 2013. Arsenic bioremediation by biogenic iron oxides and sulfides. Applied and Environmental Microbiology 79, 4325–4335.

Parlanti, E., Wörz, K., Geoffroy, L., Lamotte, M., 2000. Dissolved organic matter fluorescence spectroscopy as a tool to estimate biological activity in a coastal zone submitted to anthropogenic inputs. Organic Geochemistry 31, 1765–1781.

Pi, K., Wang, Y., Xie, X., Huang, S., Yu, Q., Yu, M. 2015. Geochemical effects of dissolved organic matter biodegradation on arsenic transport in groundwater systems. Journal of Geochemical Exploration 149, 8–21.

Pi, K., Wang, Y., Xie, X., Ma, T., Su, C., Liu, Y., 2017. Role of sulfur redox cycling on arsenic mobilization in aquifers of Datong Basin, northern China. Applied Geochemistry 77, 31–43.

Podgorski, J., Berg, M., 2020. Global threat of arsenic in groundwater. Science 368, 845–850.

Polizzotto, M.L., Kocar, B.D., Benner, S.G., Sampson, M., Fendorf, S., 2008. Near-surface wetland sediments as a source of arsenic release to ground water in Asia. Nature 454, 505–508.

Polya, D.A., Gault, A.G., Diebe, N., Feldman, P., Rosenboom, J.W., Gilligan, E., Fredericks, D., Milton, A.H., Sampson, M., Rowland, H.A.L., Lythgoe, P.R., Jones, J.C., Middleton, C., Cooke, D.A., 2005. Arsenic hazard in shallow Cambodian groundwaters. Mineralogical Magazine 69, 807–823.

Poulin., B.A., Ryan, J.N., Aiken, G.R., 2014. Effects of iron on optical properties of dissolved organic matter. Environmental science & technology 48, 10098-10106.

Pracht, L.E., Tfaily, M.M., Ardissono, R.J., Neumann, R.B., 2018. Molecular characterization of organic matter mobilized from Bangladeshi aquifer sediment: tracking carbon compositional change during microbial utilization. Biogeosciences 15, 1733–1747.

Qiao, W., Guo, H., He, C., Shi, Q., Xiu, W., Zhao, B., 2020. Molecular Evidence of Arsenic Mobility Linked to Biodegradable Organic Matter. Environmental Science & Technology 54, 7280–7290.

Qiao, W., Guo, H., He, C., Shi, Q., Zhao, B., 2021. Unraveling roles of dissolved organic matter in high arsenic groundwater based on molecular and optical signatures. Journal of Hazardous Materials 406, 124702.

Richards, L.A., Sültenfuß, J., Magnone, D., Boyce, A,. Sovann, C., Casanueva-Marenco, M., Ballentine, C., van Dongen, B.E., Polya, D.A., 2016 Age and provenance of groundwater in a shallow arsenic-affected aquifer in the lower Mekong Basin, Kandal Province, Cambodia. Arsenic Research and Global Sustainability - Proceedings of the 6th International Congress on Arsenic in the Environment, AS 2016. pp. 74–75.

Richards, L.A., Magnone, D., Sovann, C., Kong, C., Uhlemann, S., Kuras, O., van Dongen, B.E., Ballentine, C.J., Polya, D.A., 2017. High resolution profile of inorganic aqueous geochemistry and key redox zones in an arsenic bearing aquifer in Cambodia. Science of The Total Environment 590–591, 540–553.

Richards, L.A., Magnone, D., Boyce, A.J., Casanueva-Marenco, M.J., van Dongen, B.E., Ballentine, C.J., Polya, D.A., 2018. Delineating sources of groundwater recharge in an arsenic-affected Holocene aquifer in Cambodia using stable isotope-based mixing models. Journal of Hydrology 557, 321–334.

Richards, L.A., Casanueva-Marenco, M.J., Magnone, D., Sovann, C., van Dongen, B.E., Polya, D.A., 2019a. Contrasting sorption behaviours affecting groundwater arsenic concentration in Kandal Province, Cambodia. Geoscience Frontiers 10, 1701–1713.

Richards, L.A., Lapworth, D.J., Magnone, D., Gooddy, D.C., Chambers, L., Williams, P.J., van Dongen, B.E., Polya, D.A., 2019b. Dissolved organic matter tracers reveal contrasting characteristics across high arsenic aquifers in Cambodia: A fluorescence spectroscopy study. Geoscience Frontiers 10, 1653–1667.

Rowland, H.A.L., Polya, D.A., Lloyd, J.R., Pancost, R.D. 2006. Characterisation of organic matter in a shallow, reducing, arsenic-rich aquifer, West Bengal. Organic Geochemistry 37, 1101–1114.

Rowland, H.A.L., Pederick, R.L., Polya, D.A., Pancost, R.D., van Dongen, B.E., Gault, A.G., Vaughan, D.J., Bryant, C., Anderson, B., Lloyd, J.R., 2007. The control of organic matter on microbially mediated iron reduction and arsenic release in shallow alluvial aquifers, Cambodia. Geobiology 5, 281–292.

Rowland, H.A.L., Gault, A.G., Lythgoe, P., Polya, D.A., 2008. Geochemistry of aquifer sediments and arsenic-rich groundwaters from Kandal Province, Cambodia. Applied Geochemistry 23, 3029–3046.

Rowland, H.A.L., Boothman, C., Pancost, R., Gault, A.G., Polya, D.A., Lloyd, J.R., 2009. The Role of Indigenous Microorganisms in the Biodegradation of Naturally Occurring Petroleum, the Reduction of Iron, and the Mobilization of Arsenite from West Bengal Aquifer Sediments. Journal of Environmental Quality 38, 1598–1607.

Savory, J.J., Kaiser, N.K., McKenna, A.M., Xian, F., Blakney, G.T., Rodgers, R.P., Hendrickson, C.L., Marshall, A.G., 2011. Parts-Per-Billion Fourier Transform Ion Cyclotron Resonance Mass Measurement Accuracy with a “Walking” Calibration Equation. Analytical Chemistry 83, 1732–1736.

Smith, A.H., Lingas, E.O., Rahman, M., 2000. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. Bulletin of the World Health Organization 78, 1093–1103.

Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Marine Chemistry 82 (3-4), 239-254.

Spencer, R.G.M., Guo, W., Raymond, P.A., Dittmar, T., Hood, E., Fellman, J., Stubbins, A., 2014. Source and biolability of ancient dissolved organic matter in glacier and lake ecosystems on the Tibetan Plateau. Geochimica et Cosmochimica Acta 142, 64–74.

Spencer, R.G., Mann, P.J., Dittmar, T., Eglinton, T.I., McIntyre, C., Holmes, R.M.,Zimov, N., Stubbins, A., 2015. Detecting the signature of permafrost thaw in Arctic rivers. Geophysical Research Letters, 42, 2830-2835.

Spencer, R.G., Kellerman, A.M., Podgorski, D.C., Macedo, M.N., Jankowski, K., Nunes, D. and Neill, C., 2019. Identifying the molecular signatures of agricultural expansion in Amazonian headwater streams. Journal of Geophysical Research: Biogeosciences, 124, 1637-1650.

Stubbins, A., Lapierre, J.F., Berggren, M., Prairie, Y.T., Dittmar, T., Del Giorgio, P.A., 2014. What's in an EEM? Molecular signatures associated with dissolved organic fluorescence in boreal Canada. Environmental Science & Technology 48, 10598–10606.

Tamura, T., Saito, Y., Sieng, S., Ben, B., Kong, M., Choup, S., Tsukawaki, S., 2007. Depositional facies and radiocarbon ages of a drill core from the Mekong River lowland near Phnom Penh, Cambodia: Evidence for tidal sedimentation at the time of Holocene maximum flooding. Journal of Asian Earth Sciences 29, 585–592.

Textor, S.R., Guillemette, F., Zito, P.A., Spencer, R.G.M., 2018. An Assessment of Dissolved Organic Carbon Biodegradability and Priming in Blackwater Systems. Journal of Geophysical Research: Biogeosciences, 123, 2998-3015.

Tye, A.M., Lapworth, D.J., 2016. Characterising changes in fluorescence properties of dissolved organic matter and links to N cycling in agricultural floodplain. Agriculture, Ecosystems and Environment 221, 245–257.

Uhlemann, S., Kuras, O., Richards, L.A., Naden, E., Polya, D.A., 2017. Electrical resistivity tomography determines the spatial distribution of clay layer thickness and aquifer vulnerability, Kandal Province, Cambodia. Journal of Asian Earth Sciences 147, 402–414.

Vaughn, D.R., Kellerman, A.M., Wickland, K.P., Striegl, R.G., Podgorski, D.C., Hawkings, J.R., Nienhuis, J.H., Dornblaser, M.M., Stets, E.G. and Spencer, R.G., 2021. Anthropogenic landcover impacts fluvial dissolved organic matter composition in the Upper Mississippi River Basin. Biogeochemistry, 1-25.

Wagner, S., Jaffé, R., Cawley, K., Dittmar, T., Stubbins, A., 2015a. Associations between the molecular and optical properties of dissolved organic matter in the Florida Everglades, a model coastal wetland system. Frontiers in Chemistry 3, 1–14.

Wagner, S., Riedel, T., Niggemann, J., Vahatalo, A.V., Dittmar, T. and Jaffé, R., 2015b. Linking the molecular signature of heteroatomic dissolved organic matter to watershed characteristics in world rivers. Environmental science & technology 49, 13798-13806.

Wang, Y., Tian, X., Song, T., Jiang, Z., Zhang, G., He, C., Li, P., 2023. Linking DOM characteristics to microbial community: The potential role of DOM mineralization for arsenic release in shallow groundwater. Journal of Hazardous Materials 454, 131566.

Weishaar, J.L., Aiken, G.R., Bergamaschi, B.A., Fram, M.S., Fujii, R., Mopper, K., 2003. Evaluation of specific ultraviolet absorbance as an indicator of the chemical composition and reactivity of dissolved organic carbon. Environmental Science & Technology 37, 4702–4708.

Whaley-Martin, K.J., Mailloux, B.J., van Geen, A., Bostick, B.C., Silvern, R.F., Kim, C., Ahmed, K.M., Choudhury, I., Slater, G.F. 2016. Stimulation of Microbially Mediated Arsenic Release in Bangladesh Aquifers by Young Carbon Indicated by Radiocarbon Analysis of Sedimentary Bacterial Lipids. Environmental Science & Technology 50, 7353–7363.

Wilson, H.F., Xenopoulos, M.A. 2009. Effects of agricultural land use on the composition of fluvial dissolved organic matter. Nature Geoscience 2, 37–41.

Xian, F., Hendrickson, C. L., Blakney, G. T., Beu, S. C. & Marshall, A. G., 2010. Automated broadband phase correction of Fourier transform ion cyclotron resonance mass spectra. Analytical chemistry 82, 8807-8812.

Xue, Q., Ran, Y., Tan, Y., Peacock, C.L., Du, H., 2019. Arsenite and arsenate binding to ferrihydrite organo-mineral coprecipitate: Implications for arsenic mobility and fate in natural environments. Chemosphere 224, 103–110.

Ye, H., Yang, Z., Wu, X., Wang, J., Du, D., Cai, J., Lv, K., Chen, H., Mei, J., Chen, M., Du, H., 2017. Sediment biomarker, bacterial community characterization of high arsenic aquifers in Jianghan Plain, China. Scientific Reports 7, 1–11.

Yu, K., Duan, Y., Gan, Y., Zhang, Y., Zhao, K., 2020. Anthropogenic influences on dissolved organic matter transport in high arsenic groundwater: Insights from stable carbon isotope analysis and electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry. Science of the Total Environment 708, 135162.

Yuan, Z., He, C., Shi, Q., Xu, C., Li, Z., Wang, C., Zhao, H. and Ni, J., 2017. Molecular insights into the transformation of dissolved organic matter in landfill leachate concentrate during biodegradation and coagulation processes using ESI FT-ICR MS. Environmental Science & Technology, 51, 8110-8118.

Zsolnay, A., Baigar, E., Jimenez, M., Steinweg, B., Saccomandi, F., 1999. Differentiating with fluorescence spectroscopy the sources of dissolved organic matter in soils subjected to drying. Chemosphere 38, 45–50.

# 6 Characterisation of dissolved organic matter in two contrasting arsenic-prone sites in Kandal Province, Cambodia

Oliver C. Moore[*,1,2], Amy D. Holt[2], Laura Richards[1], Amy M. McKenna[3,4], Robert G.M. Spencer[2], Dan Lapworth[5], David A. Polya[1], Jonathan R. Lloyd[*,1,] and Bart E. van Dongen[1]

[1]Department of Earth and Environmental Sciences and Williamson Research Centre for Molecular Environmental Science, University of Manchester, Manchester, UK

[2] National High Magnetic Field Laboratory Geochemistry Group and Department of Earth, Ocean, and Atmospheric Science, Florida State University, Tallahassee, FL 32306, USA.

[3]Ion Cyclotron Resonance Facility, National High Magnetic Field Laboratory, Florida State University, Tallahassee, FL 32301, USA

[4]Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO, USA

[5]British Geological Survey, Maclean Building, Wallingford, UK

*Corresponding authors: oliver.moore@slu.se (Oliver C. Moore), jon.lloyd@manchester.ac.uk (Jonathan R. Lloyd)

## 6.1 Highlights

- Ultra-high resolution mass spectrometry study of As-prone aquifer dissolved organics
- Optical/molecular analyses indicate overall dominance of terrestrial-derived organics
- Likely microbial processing of organics since percolation into the aquifer
- Analyses support role dissolved organics play during microbial mediated As release

**Declaration of interests**

☐ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Prof Bart van Dongen, a co-author on this paper, is also on the editorial board for Organic Geochemistry. We have previously discussed this with Prof Elizabeth Minor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[2] Current address: Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Biocentrum, Uppsala, Sweden

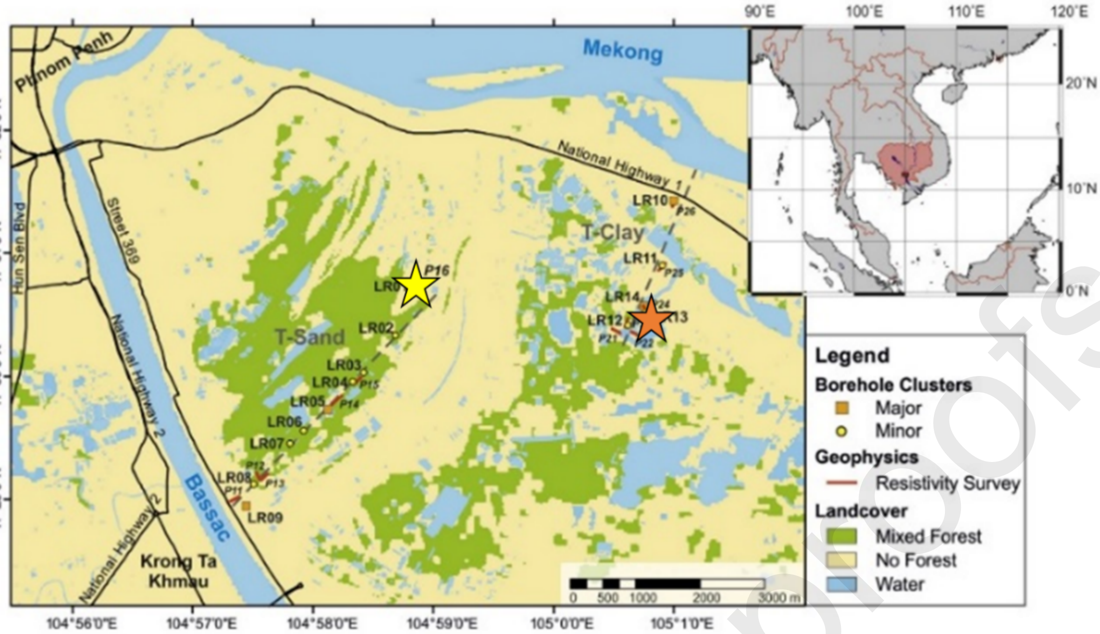



**Fig. 1**. Location of the boreholes in Kandal Province, Cambodia, drilled in January 2019 and sampled in January 2020 for this study and parallel studies (Bassil et al., In Submission). Yellow star marks the boreholes of the sand-dominated site (*n*=4), orange star marks the boreholes of the clay-dominated (*n=4)* site. LR marks transects of previous hydrogeochemical surveys in the area (from 2014 to 2015), including EEM data (Magnone et al., 2017, 2019; Richards et al., 2017, 2018, 2019b) and previous electrical resistivity surveys (Uhlemann et al., 2017). Adapted from Richards et al. (2017). Surface water was additionally collected from a seasonal pond/wetland located in very close proximity to the clay-dominated site and the sample location falls within the same orange star marked for that site.

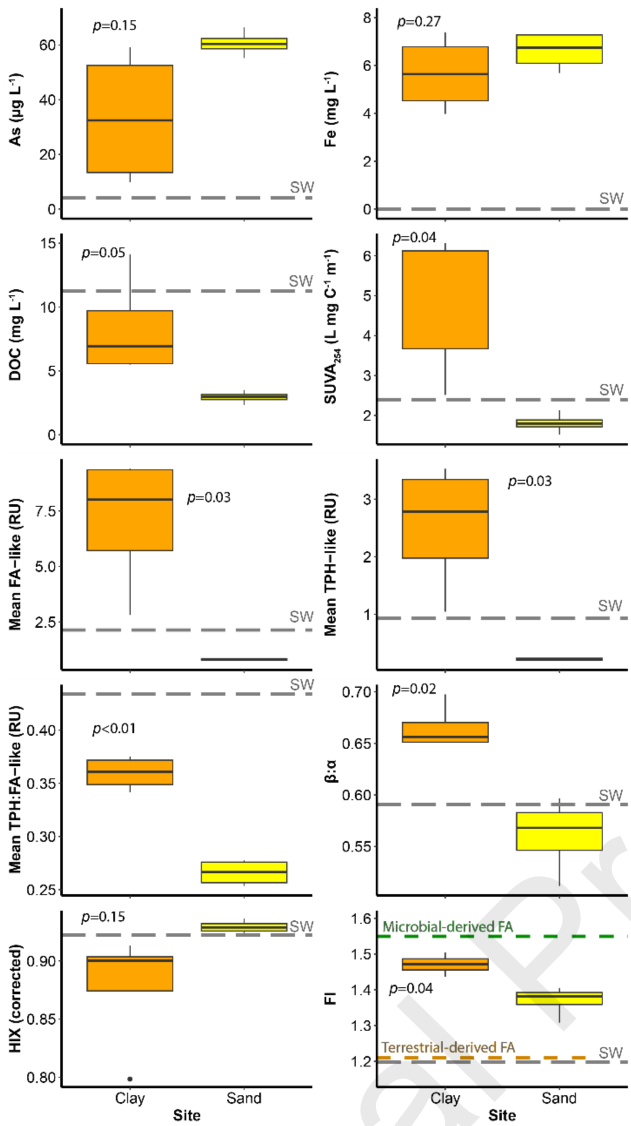



**Fig. 2.** Key physiochemical and optical properties of the groundwater dissolved organic matter (DOM) sampled from the clay-dominated (Clay) and sand-dominated (Sand) aquifer sites in Kandal Province, Cambodia – sampled in January 2020. Total As and dissolved organic carbon (DOC) were calculated from Bassil et al (In Submission). Specific ultraviolet absorbance ($SUVA_{254}$; L mg $C^{-1}m^{-1}$) was calculated from the absorbance at 254 nm divided by the DOC, and is a proxy for aromaticity (Weishaar et al., 2003). FA-like = fulvic acid-like fluorescence. TPH-like = tryptophan-like fluorescence. ß:α = "freshness index", the ratio between recently derived DOM ("β") and more decomposed DOM ("α") (Parlanti et al., 2000; Wilson & Xenopoulos, 2009; Fellman et al., 2010; Kulkarni et al., 2017). $HIX_{corr}$ = humification index with inner filtration correction (Ohno, 2002). FI = fluorescence index, with end-member values for microbial and terrestrial-derived DOM indicated on figure (Cory *et al.,* 2010). Boxplots show groundwater sites, separated by stratigraphy, and the surface water (SW) sample value is marked by the broken line. The outcome of Welch t-test (95% confidence) between the groundwaters of the clay-dominated and sand-dominated sites is indicated by the *p*-values; significance is indicated when $p<0.05$ (this could not be performed on the $HIX_{corr}$ due to non-normal distribution).

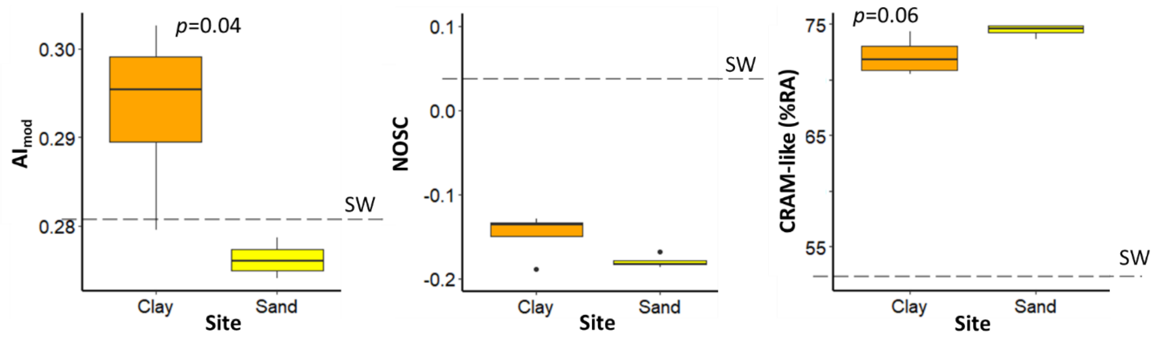



**Fig. 3.** Modified aromaticity index ($AI_{mod}$), nominal oxidation state of carbon (NOSC), and the relative abundances (%RA) of carboxyl-rich acyclic molecules (CRAM-like), in the aquifer groundwaters of the clay-dominated (Clay) and sand-dominated (Sand) sites ($n$=4 at each site) and a single surface water sample (SW; $n$=1) in the Kandal Province, Cambodia. The outcome of Welch t-test (95% confidence) between the groundwaters of the clay-dominated and sand-dominated sites is indicated by the $p$-values, significance is indicated when $p<0.05$ (this could not be determined on NOSC values due to non-normal distributions).

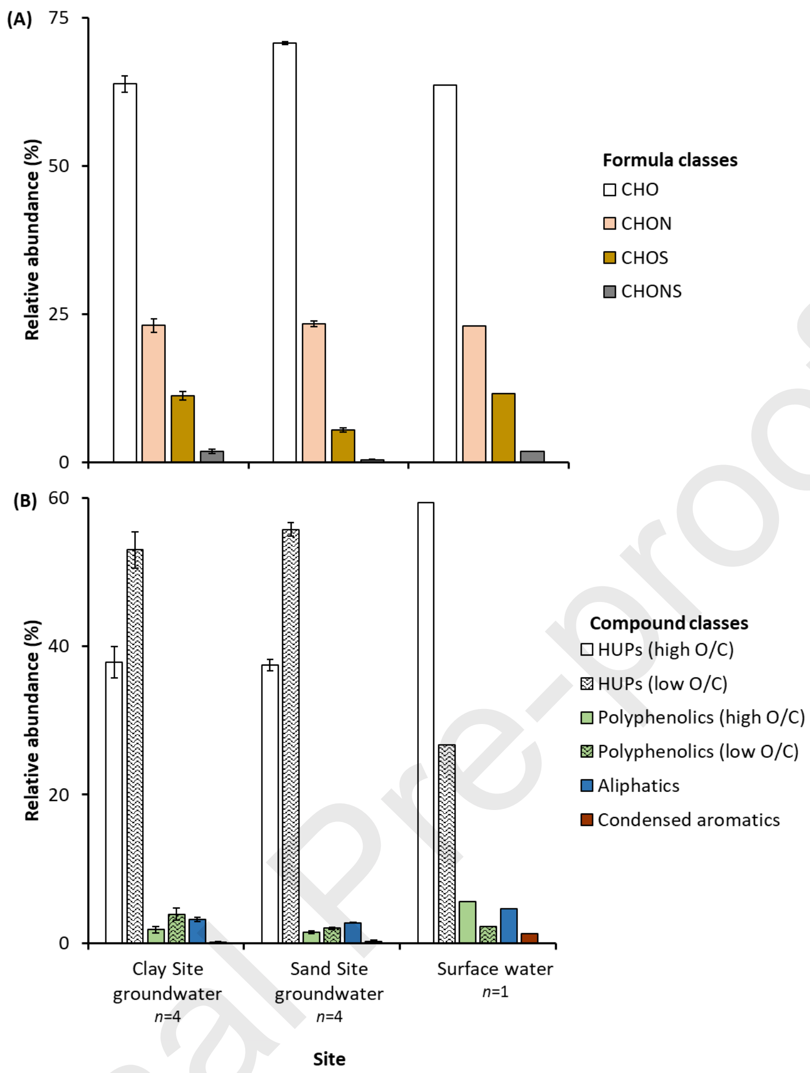



**Fig. 4.** Heteroatom classes (A) and compound classes (B) by relative abundance (RA) within the groundwaters of the clay-dominated (Clay; *n*=4) and sand-dominated (Sand; *n*=4) boreholes and the surface water sample (*n*=1). HUPs = highly unsaturated and phenolic compounds. Low O/C = <0.5; high O/C ≥0.5.

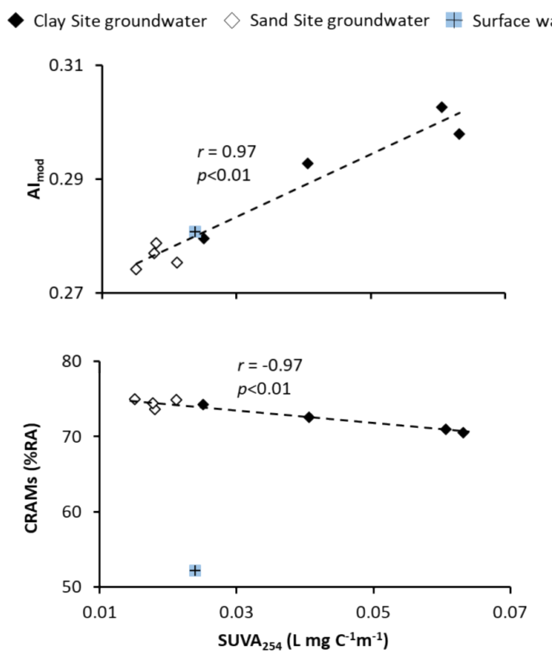



**Fig. 5.** Pearson's correlations between specific ultraviolet absorbance at 254 nm ($SUVA_{254}$), and modified aromaticity index ($AI_{mod}$; top) and relative abundance (%RA) of carboxyl-rich alicyclic-like molecules (CRAM-like; bottom). Line of best fit and r values are for groundwater samples only. Clay GW = groundwater samples from the clay-dominated site ($n$=4). Sand GW = groundwater samples from the sand-dominated site ($n$=4).

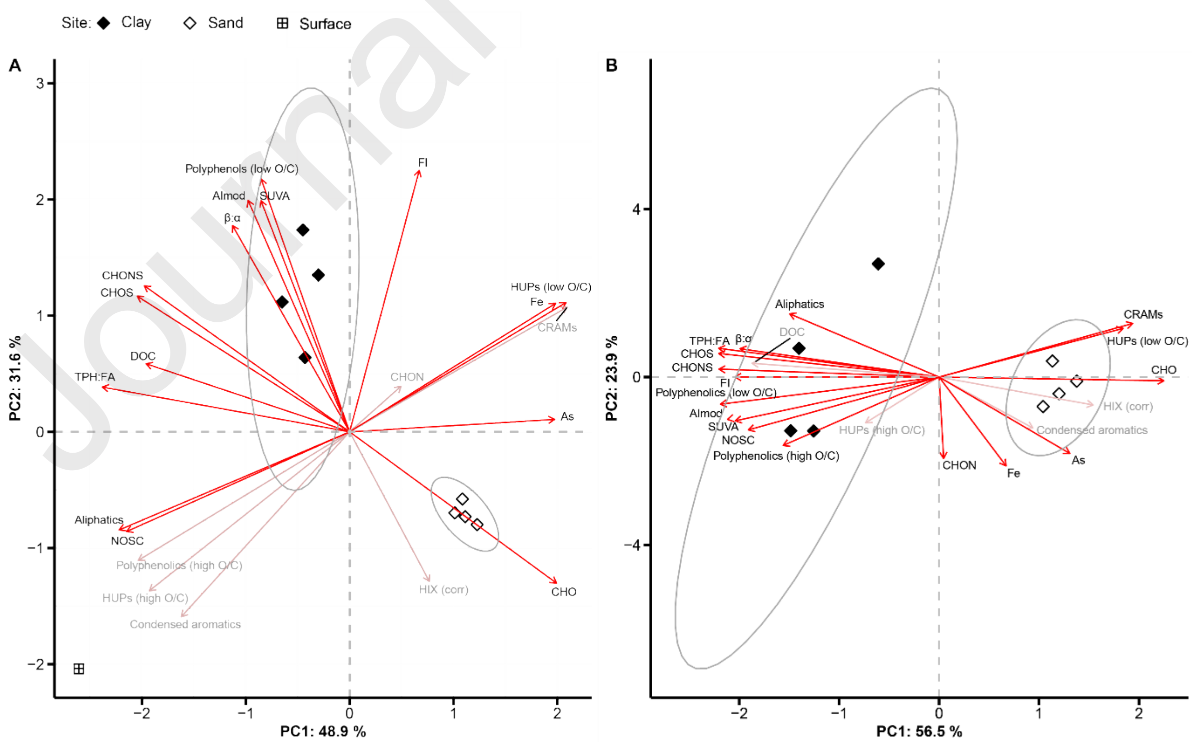


**Fig. 6.** Principal component analysis (PCA) of Fe, As and dissolved organic carbon (DOC) concentrations, along with optical and FT-ICR MS metrics. (A) PCA performed on both groundwater and surface water samples. (B) PCA performed on groundwater samples only. Grey ellipses show the grouping of samples according to site (clay-dominated and sand-dominated), at 95% confidence. Where environmental metrics and their respective arrows are shown in grey and faint colours, this indicates that these factors were not significant to the overall variance according to the envfit *p*-values (Box S1), though may still be important to the respective sites they are distributed with. $SUVA_{254}$ = specific ultraviolet absorbance at 254 nm ($Abs_{254}$/DOC). TPH:FA-like = the ratio between tryptophan-like and fulvic acid-like fluorescence. $HIX_{corr}$ = humification index with inner filtration correction (Ohno, 2002). ß:α = "freshness index", the ratio between recently derived DOM ("β") and more decomposed DOM ("α") (Parlanti et al., 2000; Wilson & Xenopoulos, 2009; Fellman et al., 2010; Kulkarni et al., 2017). HUPs = highly unsaturated and phenolic compounds. $AI_{mod}$= modified aromaticity index (Koch & Dittmar, 2006, 2016). NOSC = nominal oxidation state of carbon (LaRowe & Cappellen, 2011). Clay = groundwater samples from the clay-dominated site (*n*=4). Sand = groundwater samples from the sand-dominated site (*n*=4). Surface = the surface water sample (*n*=1). The raw results of the PCAs can be found in the supplementary material (Box S1).