Functional analysis of (pristine) estuarine marine sediments

Rohan Shah
CSIRO
Katie Hillyer
CSIRO
Sarah Stephenson
CSIRO
Joseph Crosswell
CSIRO
Avinash Karpe
CSIRO
Enzo Paolombo
Swinburne University
Oliver Jones
RMIT University https://orcid.org/0000-0002-4541-662X
Daniel Gorman
CSIRO
Levente Bodrossy
CSIRO Marine and Atmospheric Research https://orcid.org/0000-0001-6940-452X
Jodie van de Kamp
CSIRO https://orcid.org/0000-0003-2167-0938
Andrew Bissett
Commonwealth Scientific and Industrial Research Organisation
Andrew Whiteley
CSIRO
Andy Steven
CSIRO
David Beale (david.beale@csiro.au)
CSIRO https://orcid.org/0000-0002-9948-9197

Article

Keywords: Environmental metabolomics, ecometabolomics, systems biology, multi-omics, environmental monitoring, community metabolomics

Posted Date: February 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-141498/v2

Version of Record: A version of this preprint was published on March 18th, 2021. See the published version at https://doi.org/10.1016/j.scitotenv.2021.146526.

Abstract

Traditional environmental monitoring techniques are well suited to resolving acute exposure effects but lack resolution in determining subtle shifts in ecosystem functions resulting from chronic exposure(s). Surveillance with sensitive omics-based technologies could bridge this gap but, to date, most omics-based environmental studies have focused on previously degraded environments, identifying key metabolic differences resulting from anthropogenic perturbations. Here, we apply ‘omic based approaches to pristine environments to establish blueprints of microbial functionality within healthy estuarine sediment communities. We collected surface sediments (n=50) from four pristine estuaries along the Western Cape York Peninsula of Far North Queensland, Australia. Sediment microbiomes were analyzed for 16S rRNA amplicon sequences, central carbon metabolism metabolites and associated secondary metabolites via untargeted metabolic profiling methods. Multivariate statistical analyses indicated heterogeneity amongst the sampled estuaries, however, taxa-function relationships could be established and predicted community metabolism potential. Twenty-four correlated gene-metabolite pathways were identified and used to establish sediment microbial blueprints of carbon metabolism and amino acid biosynthesis. Our results establish a baseline microbial blueprint for the pristine sediment microbiome, one that drives important ecosystem services and to which future ecosurveillance monitoring can be compared.

1. Introduction

Traditional environmental and organism monitoring techniques (e.g., chemical monitoring and bioassays) are highly suited to assessing acute disruptive events [1] but are less able to detect subtle shifts in ecosystem function caused by chronic and/or sub-lethal drivers. In this regard, novel surveillance techniques and monitoring approaches that combine ‘omics-based technologies with traditional methodologies are being developed [2-5]. This has predominantly been driven by research in the health and nutrition domain [6] but has found application within environmental science. For example, omics-based techniques have advanced our understanding of microbial physiology [7], organism-environment interactions and organism function/health [4] based on the premise that metabolomics is effectively one step removed from an organism’s phenotype [8]. This metabolomic-phenotypic link has further expanded our knowledge of specific organismal responses to abiotic stressors [5] and for exploring both physical and anthropogenic contaminants [9-11]. Technologically, we are now entering an era with the potential for developing ecosystem-scale, multi-omic enabled blueprints for sensitive ecosurveillance strategies, strategies that could determine critical drivers of ecosystem functions within pristine and impacted environments. Critically, due to the substantial range of key ecosystem services that microbial communities provide globally, determining system wide impacts upon this component of the biosphere could efficiently provision key monitoring approaches for assessing ecosystem status and the impacts of environmental perturbation.

Microbial communities provide vital goods and services with far reaching consequences for critical ecosystem functions, with the ecological modulators of the rate of delivery of some of these services now

being resolved (e.g., denitrification) [12]. Further, microbial communities are highly dynamic and diverse assemblages, where members can become more or less important for the provision of these services over time [13] and with increasing community diversity thought to build resistance, resilience and functional redundancy into the services [12, 14, 15]. As such, the application of environmental microbial community metabolomics is one specific approach being explored to assess large scale ecosystem function questions [16-19]. However, the consistency and predictability of un-perturbed microbiomes suggest they are under the regulation of the (meta)genetic blueprint [20], and thus, understanding the contributions of genetic factors to measured metabolic phenotypes, how these interact and are subsequently modulated by stressors, is also a key component to resolve. As stated by Bundy *et al.* [4], it is difficult to measure and interpret differing degrees of influence that combine genetics and the impact the environment has on metabolism.

Despite some preliminary studies comparing genetic and site-specific effects on the metabolic profile of environmentally disturbed systems [16, 17, 21, 22], there has yet to be a systematic analysis of how microbial community structure influences the community metabolome of large ecosystems and a variety of metabolic functions. To date, the majority of environmental metabolomics research has focused on model and non-model organisms, typically the response of these organisms to an already perturbed ecosystem, with the detection of subtle shifts in ecosystem-wide functions remaining a challenge [23]. Further, it is yet to be resolved if the community metabolome can be used in routine monitoring programs as a measure of ecosystem function and health. This is typically due to limited metabolite data for integration with community composition (rRNA or metagenome) datasets, a lack of temporal application or, relying on statistical associations and anecdotal linkages to generate ecosystem function assessments. In many cases, 16S rRNA gene sequencing is principally used to profile population diversity and the microbial functional capacity is then inferred as a secondary consideration from 16S rRNA gene sequencing data, with a degree of uncertainty [24]. There are surprisingly few studies that have validated this potential, or perceived, function with detailed metabolism data (i.e., metabolomics), nor overlaid and assessed how metabolomic profiles compare to phylogenetic differences and similarities (i.e., exploring what is different, but also what is the same) in non-perturbed systems [25].

In this study, bacterial 16S rRNA gene sequencing and a combination of targeted central carbon metabolism and untargeted secondary metabolite discovery approaches were integrated. Taxa-function relationships were established to assess the variance and similarities of metabolisms and bacterial community structures in four pristine estuaries during the dry season (winter) of the Western Cape York region, in northern tropical Australia. These integrated datasets enabled a better understanding of the underlying conditions that define the ecosystem health and resilience within these estuaries, through the construction of integrated ‘microbial blueprints’. This information subsequently provisioned the baseline ‘microbial blueprint’ assessment of a highly valued and pristine environment, against which future developments and impacts could be measured.

2. Materials And Methods

2.1 Study area – the Western Cape York Peninsula, Australia

The Western Cape York Peninsula, Far North Queensland, Australia, is considered 'World Heritage Quality' [26]. The peninsula is a remote natural wilderness area in the monsoon tropical zone of northernmost Queensland, Australia (**Figure 1 and Supplementary Figure S1**). The Queensland Government's State of the Environment report card considers the region to be in very good condition in terms of water quality and contaminants [27]. A detailed description of the study location is provided in the supplementary information, as required by the Metabolomics Standard Initiative reporting guidelines [28].

2.2 Sediment sample collection

Sediment samples were collected during the dry season (13^{th} – 26^{th} of July, 2018), with no rainfall occurring during the month prior, and during the collection period [29]. At each site (n=50) a single surficial sediment sample (top 2 cm) was obtained using a clean polycarbonate corer (diameter 10 cm). Samples intended for 16S rRNA gene sequencing and metabolomics were then transferred into DNA-free, sterile 50 mL Greiner tubes and placed on ice immediately upon collection. The tubes were then placed into sterile snap-lock bags and stored at −20°C until transported to the laboratory. Upon arrival at the laboratory, samples were then stored at −80°C. All materials used for the collection and storage of DNA and metabolite samples were soaked for at least 24 h in 5% sodium hypochlorite and rinsed thoroughly (n=5) with Milli-Q water. The core sediment sample was subsampled for nutrients (0.5g), grain size (5.0g), pesticide (0.5g), and metal analysis (0.5g).

2.3 Physicochemical analyses

Sediment samples were analyzed for heavy metals and pesticides and characterized in terms of particle size. Heavy metals were extracted from a 0.5 g sub-sample of sediment, oven-dried, and extracted as previously described [30]. Pesticides were screened as per the Agilent Pesticide Personal Compound Database and Library (PCDL) method (G3878CA, Agilent Technologies, Santa Clara, CA, USA), without modifications, using a sediment-based QuEChERS approach described by Bragança *et al.* [31]. Pesticides were analyzed using an Agilent 6546 liquid chromatography quadrupole time-of-flight mass spectrometer (LC-QToF) with an Agilent Jet Stream source coupled to an Agilent Infinity II UHPLC system (Agilent Technologies, Santa Clara, CA, USA). Particle size analysis of the sediments was undertaken as previously described , with slight modifications, using 5.0 g of sediment and per and AS 4816.1-2002 (R2016) [32] standards. All chemicals were of analytical grade or higher and were purchased from Sigma-Aldrich within Australia unless otherwise stated.

2.4 Bacterial 16S rRNA gene sequencing

2.4.1 DNA sediment extraction

Using 0.25 g of homogenized sediment, DNA was extracted and purified from each sample using the DNeasy PowerSoil© DNA isolation kit (Cat No. 47016, QIAGEN GmbH®, Hilden, Germany) using the

manufacturer's standard protocol. Negative water controls were included in all polymerase chain reaction (PCR) experiments to test for biological contamination during amplification.

2.4.2 16S rRNA gene sequencing and data analysis

16S rRNA amplicon sequencing was performed according to Bissett, Fitzgerald [33], using the Illumina MiSeq platform at the Ramaciotti Centre for Genomics at the University of New South Wales. Amplicons targeting the bacterial 16S rRNA gene (V1-3) were amplified with the primer set 27F-519R [34, 35] (PCR conditions: 10 min at 95°C; 35 cycles of 94°C for 30 s; 55°C for 10 s; 72°C for 45 s with a final extension at 72°C for 10 mins. PCR was performed in 25µl reactions containing 2.5µl Immo Buffer (10x), 0.5µl dNTP Mix (10mM), 1.25µl $MgCl_2$ (50mM), 0.2µl Immolase Polymerase (5U/µl), 1.25µl of each primer and 1µl of template DNA. Negative water controls were included in all polymerase chain reaction (PCR) experiments to test for biological contamination during amplification.

FASTQ paired-end reads were merged using FLASH v1.2.1 [36], (min-overlap=30, max-overlap=250). Merged sequences that were deemed too short or too long (minlength=400, maxlength=520), had ambiguous base calls or homopolymer runs >8 were removed using 'screen.seqs' in Mothur v1.41.1 [37] and dereplicated using the fastx_uniques function in USEARCH v11 [38]. Dereplicated sequences were denoised and resolved into amplicon sequence variants (ASV's) using the UNOISE3 algorithm in USEARCH v11 [39] and default arguments. Merged and unfiltered sequences were then mapped to these ASV's at 97% similarity to produce an ASV abundance table using the "otu_tab" function in USEARCH v11. Sequences were classified using the RDP naïve Bayesian classifier [40] using MOTHUR v1.41.1 [37] against the MOTHUR distribution of the Silva132 databases [41] respectively and a probability cut-off of 60%. Sequences not classified as "Bacteria", not classified at Phylum level, classified as "chloroplast" or "Mitochondria" were removed.

ASV data were further clustered at 97% sequence similarity, using USEARCH -cluster_fast on size (abundance) sorted ASVs, for prediction of the metagenomic function using PICRUSt2 [42]. The PICRUSt2 workflow was used with default arguments to make stratified and non-stratified metagenomic predictions.

2.5 Community microbial metabolomics

2.5.1 Metabolite sediment extraction

Dried core sections (0.5 g, freeze-dried) were weighed into 7 mL bead mill tubes (SK 38, Bertin Corp., Thermo Fisher Scientific, Australia). Polar and semi-polar metabolites were extracted via modified QuEChERS (Anastassiades *et al.* 2003). Briefly, 1.0 g of magnesium sulphate and sodium acetate (6:1.5 *w/w*) was added, followed by 2 mL cold MilliQ water (MQ) and 2 mL cold acetonitrile (ACN) (at 4˚C). MQ contained 0.25 µg mL^{-1} of the internal standards (IS) L-phenylalanine (1-^{13}C) and succinic acid (1,4-$^{13}C_2$). Samples were then bead-milled (Precellys Evolution, Thermo Fisher Scientific, Australia) at 6,000 rpm, 3 x 30 s, with a 20 s break, followed by 10 min sonication at room temperature. To separate the ACN fraction,

samples were centrifuged at 4,750 *g* for 30 min at 4˚C. The ACN fraction was then removed to a fresh tube and the extraction repeated with a further 2 mL cold ACN. The extracts were combined and filtered via Captiva EMR-Lipid cartridges to remove excess lipids (3 mL, 300 mg, Agilent, Mulgrave, Australia). Biological blanks were prepared in the same way, without biological material. Pooled biological quality control (PBQC) samples were prepared with 200 µL aliquots from each biological sample. Two 1 ml aliquots of each sample extract was then transferred into 1.5 mL high recovery glass vials and samples concentrated sequentially via rotational vacuum centrifuge (2,000 rpm, 38˚C, 24 h). Dried extracts were then reconstituted in 50 µL methanol/water (20:80, *v/v*) and vortexed for 5 min at room temperature before metabolomics analysis.

2.5.2 Central carbon metabolism metabolites (LC–QqQ-MS)

Central carbon metabolism metabolites were analyzed using an Agilent 6470 liquid chromatography triple quadrupole mass spectrometer (LC-QqQ-MS) coupled with an Agilent Infinity II Flex ultra-high-performance liquid chromatography (UHPLC) system (Agilent Technologies, Santa Clara, CA, USA). The instrument was operated using the Agilent Metabolomics dMRM Database and Method . Collected data were processed using MassHunter Quantitative Analysis (for QQQ) software (Version 10.0, Agilent Technologies, USA), normalized to IS in preparation for downstream analyses.

2.5.3 Discovery polar metabolites (LC–QToF-MS)

Untargeted polar metabolites were analyzed using an Agilent 6546 liquid chromatography time-of-flight mass spectrometer (LC-QToF) with an Agilent Jet Stream source coupled to an Agilent Infinity II UHPLC system (Agilent Technologies, USA). Chromatographic separation was achieved by injection (3 µL) of sample onto an Agilent InfinityLab Proroshell 120 HILIC-Z Peek lined column (2.1 mm x 150 mm, 2.7 µm). Each sample was analyzed in positive and negative ionization mode. The mobile phase for the positive ion mode chromatography was (A) 10 mM ammonium formate in water with 0.1% formic acid and (B) 10 mM ammonium formate in acetonitrile/water (90:10, *v:v*) with 0.1 % formic acid operated for 20 minutes with a nonlinear gradient starting at 98% B. The column temperature was set at 25°C. The mobile phase for the negative ion mode chromatography was (A) 10 mM ammonium acetate in water with 2.5 µM medronic acid and (B) 10 mM ammonium acetate in acetonitrile/water (85:15, *v:v*) with 2.5 µM medronic acid operated for 26 minutes with a nonlinear gradient starting at 96% B. The column temperature was set at 50°C. The detector gas temperature was 225°C with a drying gas rate of 6 and 13 L min^{-1} for the positive and negative ion modes of operation. The sheath gas temperature and flow were 225°C and 10 L min^{-1} for positive ions, and 350°C and 12 L min^{-1} for negative ions. The nebulizer pressure was also 40 psi and 35 psi for positive and negative ions respectively. The acquisition range was 50 to 1600 *m/z*, at 3 spectrum per second. Reference mass ions were 922.009198 m/z (positive ion) and 68.9957 *m/z* and 980.02163 m/z (negative ion). Collected data were processed using MassHunter Profinder software (Version 10.0, Agilent Technologies, USA), normisled to IS and putatively identified against the Agilent METLIN Metabolite PCDL (G6825-90008, Agilent Technologies, Santa Clara, CA, USA) based on MSMS spectra and library threshold score of 0.8.

2.6 Heavy metals in sediments (ICP-MS)

Trace heavy metals were analyzed using an Agilent 7700× quadrupole-type ICP-MS (Agilent Technologies, Mulgrave, Australia), equipped with an Agilent ASX-520 autosampler. Marine sediment samples (0.5 g) from each site were used for the analysis of trace heavy metals following the EPA Method 3051A [30]. Trace heavy metals were extracted, in triplicate, using concentrated nitric acid (9 mL) by microwave-assisted digestion (Multiwave 3000, PerkinElmer Inc., Melbourne, VIC, Australia). The instrument was operated in He-mode. The integration time was 0.3 s per mass, 1 point per mass, 3 replicates, and 100 sweeps per replicate. The Agilent Environmental standard for ICP-MS was used for quantification.

2.7 Statistical analysis and data integration

The trace metal and physicochemical data were used to assess the baseline condition of the sampled estuaries against the current Australian and New Zealand Guidelines for Fresh and Marine Water Quality in sediments. The grain size data were analyzed using GRADISTAT [43]. This was to account for any bias in the subsequent bacterial 16S sequencing and metabolomics analysis arising from differences in sediment physical characteristics. The bacterial 16S sequencing and metabolomics data were subjected to further statistical analysis using multivariate statistics. The data were first imported, matched by sample location identifiers (metadata), and log-transformed to normalize the data using SIMCA 16 (MKS Data Analytics Solutions, Uméa, Sweden). Partial Least Square-Discriminant Analysis (PLS-DA) was performed by finding successive orthogonal components from the estuary-specific datasets with maximum squared covariance and was subsequently used to identify the common relationships among the multiple datasets.

MetaboAnalyst 4.0 (Xia Lab, McGill University, Quebec, Canada) was used for differential metabolite abundances and pathway enrichment analysis and metabolites with Benjamini–Hochberg adjusted p-value of ≤ 0.05 and, Fold Changes (FC) of < 0.5 (downward regulation) or > 2.0 (upward regulation), were considered to be statistically significant [44]. Chemical clusters based on structural similarity were created for metabolic examination using the ChemRICH analysis [45]. The identified metabolites were used to build metabolic pathways using the KEGG Mapper.

Automated taxonomic-to-phenotypic mapping was performed using Burrito [46] and MIMOSA2 [47], two metagenomics systems biology web tools. These tools integrate the predicted metagenomic function and metabolic data with the taxonomic composition of each sample. Alpha-diversity indices such as Shannon-Weiner's H', Simpson's D_1, Simpson's dominance D_2, Simpson's evenness E, and Pielou's evenness J were used to calculate the taxonomic diversity and functional diversity within each community. For diversity between communities, beta-diversity indices such as Jaccard's C_J and Sorensen's C_S were calculated. For multiple group comparisons (≥ 3 groups), significant features were identified using one-way ANOVA statistical analyses ($p < 0.05$) using a posthoc Dunn's significance test. Venn diagrams, presented in the supplementary section, were created using the InteractiVenn online web tool (http://www.interactivenn.net/) [48].

3. Results And Discussion

The sampled estuaries were assessed against the Australian and New Zealand Guidelines for Fresh and Marine Water Quality in sediments [49], using the collected physicochemical data (See supplementary information). Overall, the four estuaries sampled are considered healthy and below the trigger values for concern stated in the guidelines for tropical estuaries in Australia. This supports the most recent aquatic ecosystem report card published by the Queensland Government that stated the region as being “good quality” in terms of water quality, pollution, and introduced flora and fauna species.

It is noted that some water quality and sediment parameters were at, or above, the guideline trigger values, including chlorophyll-a, total nitrogen (TN), nitric oxide (NOx), total phosphorus (TP), and turbidity **(Supplementary Tables S1-S2)**. Furthermore, measured concentrations of iron and aluminium were also high compared to the marine sediment guideline trigger values across the four estuaries (**Supplementary Table S1**). However, this was to be expected based on the geological composition of the surrounding soil/substrate and, for the region, is considered naturally occurring. It should be noted that bauxite mining does occur out of the Embley-Hey estuary. Again, the concentrations of iron and aluminum were observed consistently high in all estuaries irrespective of localized mining activities. Sediment particle size analysis was also consistent for the region (**Supplementary Figure S2 and Table S3**). The distinct wet-dry seasonality of the semi-arid Cape York estuaries has been shown to trap and redistribute sediments over tidal and seasonal cycles, leading to an extreme, but highly variable, turbidity [50, 51]. As such, turbidity above 20 NTU for all estuary sites was expected and may account for elevated chlorophyll-a and primary nutrients. The analyzed sediments were free from any pesticide residues screened as part of the Agilent Pesticide PCDL (data not shown).

3.1 Community metabolomics analysis

In total, 1681 discovery polar metabolite features were detected, of which 727 metabolite features were identified in the sediment samples collected from across four estuaries. Of all the identified metabolites, 53 central carbon metabolism metabolites and 674 discovery polar metabolites were identified against the METLIN metabolites and Agilent PCDL. **Supplementary Figure S3** presents the PCA plots for the metabolomics data and provides a preliminary overview of sample/site clustering. The results indicate no clear discrimination between estuaries. The PCA model was found to be significant, with good linearity R^2X (cum) = 0.639 and predictability Q^2 (cum) = 0.484. Outliers were examined using a distance of observation (DModX) analysis ; no metabolites were considered outliers at the two times DCrit threshold (p = 0.05) (**Supplementary Figure S3)**. To explore potential for metabolic pathway activity levels in each estuarine system, the metabolomic data were subjected to pathway enrichment analysis. Pathway enrichment was used to decipher the metabolic pathways to which the metabolites are associated and resulted in 445 identified metabolites being mapped to the KEGG database. Pathway impact measured the relative significance of all metabolites of a single pathway for the overall spectrum of metabolites. **Supplementary Figure S3** summarises the pathways mapped. We then performed a ChemRICH class annotation of the KEGG identified metabolites and found that 70 chemical clusters were represented.

These chemical groups can be associated with genes and phenotypes. The analysis highlighted saturated and unsaturated fatty acids (FA), glycosides, aminoglycosides, deoxyadenosines, glutamates, indole alkaloids, ketones, monoterpenes, pregnanediones, and purine nucleosides as the most abundant chemical clusters **(Supplementary Figure S4)**. **Figure 2** illustrates the chemical clusters of significance ($p < 0.05$) contributing to metabolic differences between the estuaries **(Supplementary Figure S4)**.

Metabolites that differentiated individual estuaries were harmine (KEGG ID C06538; an indole alkaloid found in fruit and seeds , which was less abundant in the Skardon estuary), deoxyadenosine (C00559, a deoxyadenosines linked to marine microbial natural products [52], which was elevated in Skardon), deoxyguanosine (C00330; a purine nucleosides found in plants, and elevated in Skardon), undecylenic acid (C13910; an unsaturated FA elevated in Skardon that is a known antifungal drug and eukaryotic metabolite produced in plants), and a glycerol 1-phosphate derivative (C04590; a ketone found in marine sediments [53], and less abundant in Skardon). The Wenlock samples were found to have decreased levels of biphenyl (C06588; a deoxyadenosine found in some plants and animal tissues). Other significant key chemicals were fumigaclavine C (C20438; an indole alkaloid derived from marine-based fungus [54]), itaconic acid (C00490; a succinate), and L-serine (C00065; an unsaturated FA), all of which were slightly decreased. Deoxyloganin (C06071; a monoterpene) and oxoglutaric acid (C00026; a glutarate) were lower in the Embley-Hey catchment, likewise 4-hydroxy-L-glutamic acid (C0449; a glutamate) was lower in the Archer-Watson estuary. It should be noted that all the significant metabolites here were identified based on p-values alone and no metabolites were observed with changes greater than 1.5-fold across all the estuary sites. This is indicative of ecosystems that are considered similar in terms of their metabolic function at the time of sampling. The specific pathways that are the same across all the estuaries are discussed in more detail below and form the basis for establishing microbial blueprints of the sampled estuaries.

3.2 Bacterial community analysis

Bacterial communities (with relative abundance ≥ 1%) from the four estuaries were taxonomically diverse, comprising 15 bacterial phyla. A total of 16,066 phylotypes (97% OTUs) across all four estuaries were obtained. The four estuaries shared 3,582 common phylotypes **(Supplementary Figure S5)**.

Taxonomic profiling demonstrated that the most abundant phyla (70.5 – 71.6% relative abundance) were, *Proteobacteria* (mostly classes *g-proteobacteria* and *a-proteobacteria*; Embley-Hey > Skardon > Wenlock > Archer-Watson), *Bacteroidota* (mostly class *Bacteroidia*; Archer-Watson > Wenlock > Skardon > Embley-Hey), *Chloroflexi* (mostly class *Anaerolineae*; Wenlock > Archer-Watson > Skardon > Embley-Hey) and *Desulfobacterota* (mostly classes *Desulfobulbia* and *Desulfobacteria*; Skardon > Embley-Hey > Archer-Watson > Wenlock) **(Supplementary Figure S5)**. Members of *Planctomycetota* (4.5 – 6.1%) and *Acidobacteriota* (3.9 – 4.5%) were also found to have high relative abundances in all four estuaries. While the four estuaries share a largely similar community composition at the phylum level, a distinctive yellow microbial mat on the intertidal area in the upper and mid regions of the Wenlock and the Archer-Watson estuaries was observed. This is most likely due to an increased relative abundance of cyanobacteria in

these estuaries. It is worth noting that elevated turbidity was found in these estuaries **(Supplementary Table S1),** which might offer some competitive advantage to benthic cyanobacteria that are exposed to direct sunlight at low tide [55]. The presence of cyanobacteria, albeit in low relative abundances (2.3% in Wenlock and 3.5% in Archer-Watson), could explain the higher chlorophyll-a content in these samples.

A significant difference in OTU diversity indices (Shannon's diversity (H'), Simpson's diversity (D_1), and Simpson's dominance (D_2); **Supplementary Table S4)** was detected among the four estuaries (Kruskal–Wallis; $P < 0.05$) **(Supplementary Table S4)**. A significant difference in H', D_1 and D_2, were observed between Archer-Watson and Skardon estuaries (Dunn's Test, $P < 0.05$). It should be noted that these estuaries are the most distant geographically and may also be the most different in terms of watershed characteristics and wind/tidal forcing. This was also illustrated by lower beta-diversity indices (C_J and C_s, **Supplementary Table S4**).

3.3 Overall functional gene diversity and structure of microbial communities

The predicted functional profiling of the estuarine microbiome datasets showed the presence of 7,303 KEGG IDs in the four estuaries. The most active pathways are indicated in **Supplementary Figure S6**. We observed a highly similar functional repertoire of the estuarine microbiome present across the four estuaries, despite the differences in their overall taxonomic profile composition. This was also illustrated by the OPLS-DA statistics of the metagenome dataset **(Supplementary Figure S7).**

We did however observe a significant difference in predicted gene diversity indices (H', D_1 and D_2) of microbial communities among the four estuaries (Kruskal–Wallis; $P < 0.05$) **(Supplementary Table S5).** We also observed a significant difference between Archer-Watson and each of the other three estuaries for Shannon's diversity index (H', Dunn's test, $P < 0.05$). A low Jaccard's and Sorensen's index also indicated that there was only a small dis-similarity between these estuaries **(Supplementary Table S5)**.

3.4 Carbon metabolism and biosynthesis of amino acids in estuarine sediments

As a result of the significant similarities observed between these geographically distinct estuaries, the estuaries were further analyzed in terms of characterizing the microbial blueprints important for microbially-mediated processes. As such, all identified metabolites were integrated with 317 functional genes predicted from the 16S rRNA gene sequencing data that are involved in carbon, nitrogen, and sulphur cycles, amino acid metabolism and metal homeostasis and resistance. **Figure 3** maps all the detected and identified metabolites with the predicted functional genes. A high relative abundance of enzymes involved in carbon metabolism was also predicted based on the 16S rRNA gene sequences **(Figure 4)**. Central carbon metabolism includes the Embden-Meyerhof-Parnas (EMP) pathway of glycolysis, the pentose phosphate pathway (PPP), the citric acid cycle (or TCA cycle), six known carbon fixation pathways, and some pathways of methane metabolism [56].

High relative abundances of TCA associated enzymes were predicted in the sediments based on the predicted functional gene data. Iron (Fe), present at elevated concentrations across all four estuaries

(61.4 – 95.2 mg kg-1), is known to modulate the expression of critical enzymes of the TCA cycle including aconitase (ACO), citrate synthase (CS), isocitric dehydrogenase (IDH1), and succinate dehydrogenase (SDHB) [54]. Fe also increases the formation of reducing equivalents such as NADH by the TCA cycle and thus increased mitochondrial O_2 consumption and ATP formation via oxidative phosphorylation; NAD+, ADP, and ATP were all detected in the current study and demonstrate the importance of the TCA cycle here for establishing a microbial blueprint. Several genes such as *sdhB* and *frdB* that encode enzyme iron-sulphur subunits with a role in oxidative phosphorylation were predicted from the gene sequence data. Furthermore, increased Fe concentrations can also cause repression of glycolysis [57]. This supports the negligible relative abundance of genes involved in glycolysis being predicted in the sampled sediments.

Several intermediates of the pentose-phosphate pathway such as fructose-6-phosphate, fructose-1,6-diphosphate, xylulose-5-phosphate, ribose-5-phosphate, ribulose-1,5-diphosphate, sedoheptulose-7-phosphate and glycerate-3-phosphate were detected and identified. Glycerate-3-phosphate is oxidised to produce pyruvate **(Figure 3)**. Prediction of high relative abundance of genes encoding pyruvate oxidation enzymes dihydrolipoamide S-Acetyltransferase (DLAT) and dihydrolipoamide dehydrogenase (DLD) coupled with the identification of acetyl-CoA in the sediments suggests this pathway is indeed active amongst the sediment microbiomes.

A very high abundance of the genes *acnA* (encoding aconitate hydratase), *sdhB* (encoding succinate dehydrogenase) and *frdB* (encoding fumarate reductase) indicate enhanced reductive citric acid cycle (also called the Arnon-Buchanan cycle) activity **(Figure 3)**. The reductive citric acid cycle is found in microaerophiles and anaerobes, such as green sulphur bacteria belonging to the *Bacteroidota* phylum. In the current study, members of the *Bacteroidota,* phylum were suggested to contribute towards carbon fixation in the metagenome of all estuaries **(Supplementary Figure S8)**.

The reductive acetyl-CoA pathway (also called the Wood-Ljungdahl pathway) is found in strictly anaerobic *Proteobacteria* and *Planctomycetes*, some of which are methane-forming. The most important bifunctional enzymes, carbon monoxide dehydrogenase (encoded by *cooS*) and acetyl-CoA synthase (encoded by *cdhE*), catalyse CO_2 1 CO and CO_2 to a methyl group conversion, respectively. In the current study, members of *Proteobacteria* phylum were suggested to contribute towards methane metabolism in the metagenome of all estuaries **(Supplementary Figure S9)**.

Noticeably, high relative levels of genes such as *accABCD,* which encode various subunits of acetyl-CoA carboxylase were predicted in this study **(Figure 3).** These results indicate that the predicted metagenome fixes CO_2 by the 3-hydroxypropionate bicycle pathway commonly found in prokaryotes. This pathway is found in some green non-sulphur bacteria of the phylum *Chloroflexi*. Taxa-functional analysis of the predicted metagenomes also indicated that members of *Chloroflexi* substantially contributed to this pathway **(Supplementary Figure S10)**.

Biomass is generated by fluxes branching out from central carbon metabolism, originating at twelve well-known precursor substances that include glucose-6-phosphate, fructose-6-phosphate, glyceraldehyde-3-phosphate, glycerate-3-phosphate, phosphoenolpyruvate, pyruvate, acetyl-CoA, ribose-5-phosphate, erythrose-4-phosphate, a-ketoglutarate, succinyl-CoA and oxaloacetate, and glycerate-1,3-diphosphate [58]. Most of these precursor molecules were detected in the sediments **(Figure 3)**. Glucose-6-phosphate, fructose-6-phosphate, glyceraldehyde-3-phosphate, and acetyl-CoA produce glycogen, cell wall components, and lipids. Several fatty acids were also identified in the sediments **(Figure 3)**. Glycerate-3-phosphate is an important precursor in the biosynthesis of cysteine, serine, and glycine. Phosphoenolpyruvate produces tyrosine and tryptophan downstream (**Figure 3**).

Pyruvate branches out to form alanine, isoleucine, lysine, leucine, and valine. Isoleucine and leucine were identified in the sediments while several genes encoding enzymes catalysing these conversions were predicted in the sediments **(Figures 3 and 5)**. Leucine is also produced from acetyl-CoA. Several amino acids including arginine, proline, glutamine, and glutamate are produced from a-ketoglutarate and were identified in the sediments. A very high relative proportion of genes involved in the biosynthesis of various amino acids was predicted **(Figure 5).** Members of the phyla *Bacteriodota, Chloroflexi,* and *Proteobacteria* were identified as contributing to the biosynthesis of amino acids in the sediment samples **(Supplementary Figures S8-S10)**.

Community-wide biosynthetic and degradation potential for each metabolite in each sample was predicted through an integrated 16S rRNA gene sequencing-metabolomics approach using the MIMOSA2 webtool. Metabolites whose variation is consistent with expectations based on well-predicted microbial metabolic potential were identified. From these data, a community-wide metabolic model was constructed using MIMOSA2 for each sample and a community metabolic potential (CMP) was calculated, representing the relative capacity of the predicted community enzyme content in that sample to synthesize or degrade each metabolite based on metabolic reference information that links gene predictions to their substrates and products (metabolites) [47]. Twenty-four metabolites with predicted CMPs were identified; however, none of these metabolites were found to be significant using the multi-variate statistical approaches described above.

Figure 6 illustrates the top 5 positively correlated CMP metabolites. Higher positive CMP scores indicate metabolites for which this agreement is statistically "*well-predicted*." Negatively predicted metabolites can be interpreted in several different ways. Possible reasons for a negative correlation between a metabolite's levels and its metabolic potential include incorrectly annotated or missing reactions (i.e., missing metabolites within the identified gene-metabolite pathway). However, contributors identified for negatively correlated metabolites have the potential to represent true relationships but should be interpreted more cautiously than positively correlated metabolites. The top 5 negatively correlated CMP metabolites are presented in **Supplementary Figure S11.**

The CMP successfully predicted alanine à pyruvate conversion through alanine-glyoxylate transaminase activity (K00830). Pyruvate may form 2-oxoisovalerate which produces 2-isopropylmalate downstream.

2-isopropylmalate à 3-isopropylmalate, followed by 3-isopropylmalate à 2-oxoisocaproate is predicted due to the activity of 3-isopropylmalate dehydratase and 3-isopropylmalate dehydratase, respectively. 2-oxoisocaproate may lead to the biosynthesis of leucine. Pyruvate is possibly oxidized to acetyl-CoA by pyruvate dehydrogenase (K00163) or to acetolactate by acetolactate synthase (K01652, K01653, K11258). Acetolactate may further proceed downstream to produce valine or leucine. O-Succinyl-L-homoserine produces succinate (TCA cycle intermediate) and 2-ketobutanoate (the keto acid precursor for isoleucine), by O-Succinyl-L-homoserine succinate-lyase (K01739). Isocitrate was also predicted to form succinate by isocitrate lyase (K01637). Isocitrate, an intermediate of the TCA cycle, may form a-ketoglutarate which branches out to form homocitrate and glutamate, possibly leading to the production of 2-aminoadipate and g-amino butanoate. 4-guanidinobutanoate is obtained from the degradation of g-amino butanoate by glycine amidinotransferase (K00613).

Several amino acids such as methionine, histidine, serine, tryptophan, phenylalanine, and arginine with predicted CMP were indicated in the current study. Arginine degradation also produces 4-guanidinobutanoate. In the current study, arginine à ornithine by arginase (K01476) à citrulline by ornithine carbamoyl transferase (K00611) was predicted. Citrulline is also predicted to be produced from N-acetyl ornithine by acetylornithine deacetylase (K01438). Methionine is synthesized from 4-methylthio-2-oxobutanoic acid by aromatic aminotransferase (K08969, K00832). Degradation of methionine forms 5’-methylthioadenosine via S-adenosyl-L-methionine intermediate by S-adenosyl-L-methionine methylthioadenosine-lyase (K01762).

Histidinal is degraded to histidine by histidinol dehydrogenase (K00013, K14152). Tryptophan synthase (K01694, K01695, K01696, K06001) catalyzes the conversion of serine à tryptophan. Serine à pyruvate is also catalyzed by serine dehydratase (K01754). Tryptophan is further degraded to kynurenine by indoleamine 2,3-dioxygenase (K00463). Another aromatic acid phenylalanine à phenethylamine is predicted by phenylalanine carboxy-lyase (K01593). L-cystine thiocysteine-lyase (K01760) catalyzes cystine à thiocysteine.

3.5 Metal homeostasis and metal resistance in estuarine sediments

Our results indicated very high concentrations of Fe (61.4 – 95.2 mg kg^{-1}) and Al (45.5 – 68.2 mg kg^{-1}) in all estuarine sediments. Some metals, such as Cu, Fe, Ni, and Zn, are essential nutrients and play important roles in the regulation of gene expression, and the activity of biomolecules, including enzymes and cofactors for essential biochemical reactions [59]. Other metals, such as Al, Cd, Pb, As and Hg, have no biological role and are considered nonessential [60]. Unsurprisingly, 125 KEGG genes related to metal homeostasis and metal resistance. These results are summarized in **Supplementary Figure S12** and indicate the presence of a very high proportion of functional genes related to Fe and Zn metabolism, mostly related to membrane transport functions. The role of Fe in transport systems for Al uptake has been previously reported by Auger, Han [61]. Relatively higher proportions of Fe-related genes may reduce the toxic effects of Al found in the sediments. There were no metabolites detected and identified that

related to these specific genes, however, additional metallomics-based approaches could be applied to overcome this analytical limitation (i.e., assessing metal efflux in marine sediment microbiomes) [62].

3.6 Sulphur metabolism in estuarine sediments

High relative proportions of functional genes such as *dsrAB* (encoding dissimilatory sulphite reductase), *aprAB* (encoding adenylylsulphate reductase) and *sat* (encoding sulphate adenylyl transferase) were predicted in the metagenomes from all four estuaries (**Supplementary Figure S13**). SO_4^{2-} à S^{2-} is driven by the oxidation of organic carbon, supplemented by the anaerobic oxidation of methane (CH_4) at the subsurface SO_4^{2-}-CH_4 transition [63]. Most of the S^{2-} is ultimately re-oxidized back to SO_4^{2-}, via diverse sulphur intermediates, by geochemical or microbial reactions [63]. The assimilatory pathway produces reduced sulphur compounds for the biosynthesis of S-containing amino acids such as cysteine.

A high proportion of genes (**Supplementary Figure S13**) involved in sulphur oxidation such as *soxAX* (encoding L-cysteine S-thiosulfotransferase), *soxB* (encoding S-sulfosulfanyl-L-cysteine sulfohydrolase), *soxC* (encoding sulfane dehydrogenase), and *soxYZ* (encoding sulphur-oxidizing protein) were predicted in the current study. The large amounts of Fe in the estuarine sediments could increase the sulphur oxidation. Fe acts as an oxidant for S^{2-} in the deeper sediment layers where it partly binds to S^{2-} to form iron sulphide (FeS) and pyrite (FeS_2).

The sulphur cycle of marine/estuarine sediments is largely microbial-driven, via dissimilatory sulphate (SO_4^{2-}) reduction to sulphide (S^{2-}) [64]. Members of the *Proteobacteria* (mainly classes *Alphaproteobacteria* and *Gammaproteobacteria*) and *Desulfobacterota* (class *Desulfobacteria*) phyla contribute the most to the sulphur metabolism (**Supplementary Figure S13**). Members of *Proteobacteria* have been previously shown to contribute considerably to Fe cycling [65]. While the metabolomics data did not provide any metabolites that directly feed into these pathways, characterizing the relevant proteins that drive these processes may enable greater insight into these functions and curation of microbial blueprints of activity.

3.7 Nitrogen metabolism in estuarine sediments

Several key functional genes (**Supplementary Figure S14**) related to nitrogen metabolism were predicted in all four estuaries, including genes (in decreasing order of relative proportion) involved in dissimilatory nitrate (NO_3^-) reduction to ammonium (NH_4^+) (DNRA), denitrification, assimilatory NO_3^- reduction to NH_4^+ (ANRA), nitrification, and nitrogen fixation. However, the inorganic metabolomics related to these processes is not measurable by the LC-MS metabolomics-based approaches used herein. Nitrogen cycling through the environment is influenced by microbially-driven processes [66]. The denitrification and DNRA processes are mediated by the products of several genes illustrated in **Supplementary Figure S14**. Predicted functional gene analysis indicated the likely presence of these genes in all four estuaries. Taxa-functional analysis indicated that nitrogen metabolism is mediated by a diverse polyphyletic group of

bacteria. The highest contribution to nitrogen cycling genes was attributed to *Proteobacteria*, *Desulfobacterota*, and *Bacteroidota* (**Supplementary Figures S8-S9, S13**).

Denitrification is one of the major nitrogen loss pathways that uses multiple electron donors such as hydrogen gas, hydrogen sulphide, or other organic compounds. Microorganisms can grow using DNRA, the key reaction of which is NO_3^- à NH_4^+, by coupling it to the oxidation of electron donors, such as organic matter, methane, hydrogen, sulphur compounds, and iron. The environmental importance of DNRA is not well established ; however, DNRA appears to be favoured over denitrification in marine and lake sediments when there is an excess of electron donors relative to NO_3^- [67]. The results of this study agree with this observation (**Supplementary Figure S14**). High DNRA in estuarine samples indicates that nitrogen is retained in the system. This is evident from amino acid production discussed earlier.

The competition for nitrogen in estuaries is mostly for the inorganic forms of ions, commonly NH_4^+ or NO_3^-. NH_4^+ is the reduced form of nitrogen and, therefore, requires less energy to build amino acids. As such, NH_4^+ is preferred over NO_3^- and makes the competition for NH_4^+ fierce in estuaries.

The assimilation of NH_4^+ in prokaryotes occurs mainly via ammonium transporter proteins. Primary products of assimilated ammonia are glutamine and glutamate, which constitute the central reservoir of nitrogen for many biosynthetic pathways. The first pathway involves a high-affinity ammonium transporter, encoded by *amtB*, that moves NH_4^+ into the cell, where NH_4^+ + glutamate à glutamine by glutamine synthetase [68]. The functional gene *amtB* was predicted in all the four estuaries indicating an assimilatory process in the studied environments. The other mechanism involves a low-affinity transport system in which NH_3 is passively (diffusion) transported into the cell and NH4+ + α-ketoglutarate à glutamate by glutamate dehydrogenase [68]. Glutamate, α-ketoglutarate, and glutamine were identified in the estuarine samples. Glutamate is the principal source of nitrogen for the production of N-amine and is involved in transamination reactions at the core of amino acid metabolism.

4. Conclusion

The use of omics-based surveillance techniques (particularly metabolomics) for environmental science has grown substantially in the last twenty to thirty years. Indeed, environmental toxicology was one of the major applications of the very earliest metabolomics studies [69]. Today scientists are moving away from individual ‘omic approaches towards a more holistic, systems-based approach for environmental assessment.

The major limitations of research in environmental metabolomics, and indeed much of environmental science to date, has been the major focus on already perturbed systems and looking at what is different – not what is the same. Indeed, it is very hard to find an unperturbed system to develop a baseline against which impacts can be measured. In this study, the approaches normally used to look at changes in systems were used, for the first time, to establish a microbial blueprint for a “pristine” system during

the dry season which was combined with a range of metadata. These data not only inform the story of these estuaries today, but also establishes a baseline against which future monitoring can be assessed. This is especially important for the studied estuaries herein, given the planned development of Northern Australia.

The systems-based approach also allowed us to tease out important relationships, for example, the increase in the 125 KEGG genes related to metal homeostasis and metal resistance observed in the estuaries studied**.** There were no metabolites detected and identified that related to these specific genes so they may have been missed with a metabolomics-based approach. Conversely only using genomics would have omitted metabolites that were positively correlated with community metabolic potential and function (including 2-oxisocapraote, tryptophan, histidine citrulline and succinic acid).

This study provides an important first step towards the use of a multi-omics microbial blueprint approach, which could be used to monitor more estuaries and sediment communities around the world. Further expansion to include planktonic and sessile bacterial populations, in addition to capturing seasonal variance, would enable a truly dynamic and integrative microbial blueprint to be established. Such an estuarine and sediment microbial blueprint would be relevant to ecological investigations measuring the impact of a range of factors, such as climate change, contaminants, disease, food restriction, infection, parasite load, and biogeochemical cycles.

5. Acknowledgement

The 1290 Infinity II Flex pump and 6470 LC-QqQ-MS instruments used in this study were provided by Agilent Technologies to CSIRO. All images and figures were created or modified using BioRender.com using data outputs from Shiny Burrito and Shiny MIMOSA2.

References

1. Maloney, E.M., *How do we take the pulse of an aquatic ecosystem? Current and historical approaches to measuring ecosystem integrity.* Environmental Toxicology and Chemistry, 2019. **38**(2): p. 289-301.
2. Lankadurai, B.P., E.G. Nagato, and M.J. Simpson, *Environmental metabolomics: an emerging approach to study organism responses to environmental stressors.* Environmental Reviews, 2013. **21**(3): p. 180-205.
3. Ekman, D.R., et al., *Metabolite Profiling of Fish Skin Mucus: A Novel Approach for Minimally-Invasive Environmental Exposure Monitoring and Surveillance.* Environmental Science & Technology, 2015. **49**(5): p. 3091-3100.
4. Bundy, J.G., M.P. Davey, and M.R. Viant, *Environmental metabolomics: a critical review and future perspectives.* Metabolomics, 2008. **5**(1): p. 3.

5. Álvarez-Muñoz, D. and M. Farré, *Chapter 12 - Future trends in environmental metabolomics analysis*, in *Environmental Metabolomics*, D. Álvarez-Muñoz and M. Farré, Editors. 2020, Elsevier. p. 339-341.
6. Shah, R.M., et al., *An Integrated Multi-Disciplinary Perspective for Addressing Challenges of the Human Gut Microbiome.* Metabolites, 2020. **10**(3): p. 94.
7. Pinu, F.R., et al., *Systems Biology and Multi-Omics Integration: Viewpoints from the Metabolomics Research Community.* Metabolites, 2019. **9**(4): p. 76.
8. Beale, D.J., et al., *Review of recent developments in GC–MS approaches to metabolomics-based research.* Metabolomics, 2018. **14**(11): p. 152.
9. Long, S.M., et al., *Metabolomics Provide Sensitive Insights into the Impacts of Low Level Environmental Contamination on Fish Health—A Pilot Study.* Metabolites, 2020. **10**(1): p. 24.
10. Jones, O.A.H., et al., *Metabolomics and its use in ecology.* Austral Ecology, 2013. **38**(6): p. 713-720.
11. Hillyer, K.E., et al., *13C metabolomics reveals widespread change in carbon fate during coral bleaching.* Metabolomics, 2017. **14**(1): p. 12.
12. Bell, T., et al., *The contribution of species richness and composition to bacterial services.* Nature, 2005. **436**(7054): p. 1157-1160.
13. Bertilsson, S., et al., *Links between bacterial production, amino-acid utilization and community composition in productive lakes.* The ISME Journal, 2007. **1**(6): p. 532-544.
14. Comte, J., L. Fauteux, and P. del Giorgio, *Links between metabolic plasticity and functional redundancy in freshwater bacterioplankton communities.* Frontiers in Microbiology, 2013. **4**(112).
15. Allison, S.D. and J.B.H. Martiny, *Resistance, resilience, and redundancy in microbial communities.* Proceedings of the National Academy of Sciences, 2008. **105**(Supplement 1): p. 11512.
16. Beale, D.J., et al., *A multi-omics based ecological analysis of coastal marine sediments from Gladstone, in Australia's Central Queensland, and Heron Island, a nearby fringing platform reef.* Sci Total Environ, 2017. **609**: p. 842-853.
17. Beale, D.J., et al., *Seasonal metabolic analysis of marine sediments collected from Moreton Bay in South East Queensland, Australia, using a multi-omics-based approach.* Science of The Total Environment, 2018. **631-632**: p. 1328-1341.
18. Shah, R.M., et al., *Influence of Human Activities on Broad-Scale Estuarine-Marine Habitats Using Omics-Based Approaches Applied to Marine Sediments.* Microorganisms, 2019. **7**(10): p. 419.
19. Jones, O.A.H., et al., *Community Metabolomics in Environmental Microbiology*, in *Microbial Metabolomics: Applications in Clinical, Environmental, and Industrial Microbiology*, D.J. Beale, K.A. Kouremenos, and E.A. Palombo, Editors. 2016, Springer International Publishing: Cham. p. 199-224.
20. Haney, C.H. and F.M. Ausubel, *Plant microbiome blueprints.* Science, 2015. **349**(6250): p. 788.
21. Jones, O.A.H., et al., *Metabolomic analysis of soil communities can be used for pollution assessment.* Environmental Toxicology and Chemistry, 2014. **33**(1): p. 61-64.
22. Kimes, N.E., et al., *Metagenomic analysis and metabolite profiling of deep–sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill.* Frontiers in Microbiology, 2013. **4**: p. 50.

23. Sardans, J., J. Peñuelas, and A. Rivas-Ubach, *Ecological metabolomics: overview of current developments and future challenges.* Chemoecology, 2011. **21**(4): p. 191-225.
24. Janda, J.M. and S.L. Abbott, *16S rRNA Gene Sequencing for Bacterial Identification in the Diagnostic Laboratory: Pluses, Perils, and Pitfalls.* Journal of Clinical Microbiology, 2007. **45**(9): p. 2761.
25. Mohit, V., et al., *Phylogenetic differences in attached and free-living bacterial communities in a temperate coastal lagoon during summer, revealed via high-throughput 16S rRNA gene sequencing.* Applied and environmental microbiology, 2014. **80**(7): p. 2071-2083.
26. IUCN, *The World's Greatest Natural Areas: An Indicative Inventory of Natural Sites of World Heritage Quality.* International Union for Conservation of Nature, 1982.
27. Queensland Governemnt. *State of the Environment: Condition of aquatic ecosystem health* 2014 [cited 2020; Available from: https://www.stateoftheenvironment.des.qld.gov.au/pollution/water-quality/condition-of-aquatic-ecosystem-health.
28. Morrison, N., et al., *Standard reporting requirements for biological samples in metabolomics experiments: environmental context.* Metabolomics, 2007. **3**(3): p. 203-210.
29. Bureau of Meteorology. *Queensland in July 2018: Warm days overall; dry in the west and south*. [cited 2020; Available from: http://www.bom.gov.au/climate/current/month/qld/archive/201807.summary.shtml.
30. EPA Method 3051A, *Method 3051, Microwave assisted acid digestion of sediments, sludges, soils and oils*. 2007, US EPA, Washington DC, USA.
31. Bragança, I., et al., *QuEChERS: A new sample preparation approach for the determination of ibuprofen and its metabolites in soils.* Science of The Total Environment, 2012. **433**: p. 281-289.
32. AS 4816.1-2002 (R2016), *Determination of particle size distribution by gravitational liquid sedimentation methods Part 1: General principles and guidelines*. 2002, Standards Australia.
33. Bissett, A., et al., *Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database.* Gigascience, 2016. **5**(1): p. s13742-016-0126-5.
34. Lane, D., *16S/23S rRNA sequencing*. Nucleic acid techniques in bacterial systematics, ed. E. Stackebrandt and M. Goodfellow. 1991. 115-175.
35. Lane, D.J., et al., *Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses.* Proceedings of the National Academy of Sciences, 1985. **82**(20): p. 6955-6959.
36. Magoč, T. and S.L. Salzberg, *FLASH: fast length adjustment of short reads to improve genome assemblies.* Bioinformatics, 2011. **27**(21): p. 2957-2963.
37. Schloss, P.D., et al., *Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities.* Applied and Environmental Microbiology, 2009. **75**(23): p. 7537-7541.
38. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-2461.

39. Edgar, R.C., *UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing.* BioRxiv, 2016: p. 081257.
40. Wang, Q., et al., *Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy.* Applied and Environmental Microbiology, 2007. **73**(16): p. 5261-5267.
41. Quast, C., et al., *The SILVA ribosomal RNA gene database project: improved data processing and web-based tools.* Nucleic acids research, 2012. **41**(D1): p. D590-D596.
42. Douglas, G.M., et al., *PICRUSt2 for prediction of metagenome functions.* Nature Biotechnology, 2020: p. 1-5.
43. Blott, S.J. and K. Pye, *GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments.* Earth surface processes and Landforms, 2001. **26**(11): p. 1237-1248.
44. Karpe, A.V., et al., *Untargeted metabolic profiling of Vitis vinifera during fungal degradation.* FEMS Microbiology Letters, 2015. **362**(10).
45. Barupal, D.K. and O. Fiehn, *Chemical Similarity Enrichment Analysis (ChemRICH) as alternative to biochemical pathway mapping for metabolomic datasets.* Scientific Reports, 2017. **7**(1): p. 14567.
46. McNally, C.P., et al., *BURRITO: An interactive multi-omic tool for visualizing taxa–function relationships in microbiome data.* Frontiers in microbiology, 2018. **9**: p. 365.
47. Noecker, C., et al., *Metabolic Model-Based Integration of Microbiome Taxonomic and Metabolomic Profiles Elucidates Mechanistic Links between Ecological and Metabolic Variation.* mSystems, 2016. **1**(1): p. e00013-15.
48. Heberle, H., et al., *InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams.* BMC Bioinformatics, 2015. **16**(1): p. 169.
49. NWQMS, *Australian and New Zealand Guidelines for Fresh and Marine Water Quality*, N.W.Q.M. Strategy, Editor. 2000, Australian and New Zealand Environment and Conservation Council and Agriculture and Resource Management Council of Australia and New Zealand, Canberra, ACT. p. 1-103.
50. Crosswell, J.R., G. Carlin, and A. Steven, *Controls on carbon, nutrient, and sediment cycling in a large, semiarid estuarine system; princess Charlotte Bay, Australia.* Journal of Geophysical Research: Biogeosciences, 2020. **125**(1): p. e2019JG005049.
51. Wolanski, E., *An evaporation-driven salinity maximum zone in Australian tropical estuaries.* Estuarine, Coastal and Shelf Science, 1986. **22**(4): p. 415-424.
52. Xiong, Z.-Q., et al., *Recent advances in the discovery and development of marine microbial natural products.* Marine drugs, 2013. **11**(3): p. 700-717.
53. Liu, X.-L., et al., *Predominance of parallel glycerol arrangement in archaeal tetraethers from marine sediments: Structural features revealed from degradation products.* Organic Geochemistry, 2018. **115**: p. 12-23.
54. Li, Y.-X., et al. *Fumigaclavine C from a marine-derived fungus Aspergillus fumigatus induces apoptosis in MCF-7 breast cancer cells*. Marine drugs, 2013. **11**, 5063-5086 DOI:

10.3390/md11125063.

55. Jia, Z., et al., *Light as a possible regulator of MIB-producing Planktothrix in source water reservoir, mechanism and in-situ verification.* Harmful Algae, 2019. **88**: p. 101658.
56. Sudarsan, S., et al., *The Functional Structure of Central Carbon Metabolism in Pseudomonas putida KT2440.* Applied and Environmental Microbiology, 2014. **80**(17): p. 5292-5303.
57. Oexle, H., E. Gnaiger, and G. Weiss, *Iron-dependent changes in cellular energy metabolism: influence on citric acid cycle and oxidative phosphorylation.* Biochimica et Biophysica Acta (BBA) - Bioenergetics, 1999. **1413**(3): p. 99-107.
58. Noor, E., et al., *Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy.* Molecular Cell, 2010. **39**(5): p. 809-820.
59. Prabhakaran, P., M.A. Ashraf, and W.S. Aqma, *Microbial stress response to heavy metals in the environment.* Rsc Advances, 2016. **6**(111): p. 109862-109877.
60. Jones, O.A.H., et al., *The use of metabolomics in the study of metals in biological systems.* Metallomics, 2015. **7**(1): p. 29-38.
61. Auger, C., et al., *Metabolic reengineering invoked by microbial systems to decontaminate aluminum: Implications for bioremediation technologies.* Biotechnology Advances, 2013. **31**(2): p. 266-273.
62. Pi, H. and J.D. Helmann, *Ferrous iron efflux systems in bacteria.* Metallomics, 2017. **9**(7): p. 840-851.
63. Jørgensen, B.B., A.J. Findlay, and A. Pellerin, *The Biogeochemical Sulfur Cycle of Marine Sediments.* Frontiers in Microbiology, 2019. **10**(849).
64. Wasmund, K., M. Mußmann, and A. Loy, *The life sulfuric: microbial ecology of sulfur cycling in marine sediments.* Environmental Microbiology Reports, 2017. **9**(4): p. 323-344.
65. Baker, B.J., et al., *Genomic resolution of linkages in carbon, nitrogen, and sulfur cycling among widespread estuary sediment bacteria.* Microbiome, 2015. **3**(1): p. 14.
66. Kuypers, M.M., H.K. Marchant, and B. Kartal, *The microbial nitrogen-cycling network.* Nature Reviews Microbiology, 2018. **16**(5): p. 263.
67. Kraft, B., et al., *The environmental controls that govern the end product of bacterial nitrate respiration.* Science, 2014. **345**(6197): p. 676-679.
68. Damashek, J. and C.A. Francis, *Microbial nitrogen cycling in estuaries: from genes to ecosystem processes.* Estuaries and Coasts, 2018. **41**(3): p. 626-660.
69. Martyniuk, C.J. and D.B. Simmons, *Spotlight on environmental omics and toxicology: a long way in a short time.* Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 2016. **19**: p. 97-101.

Figures

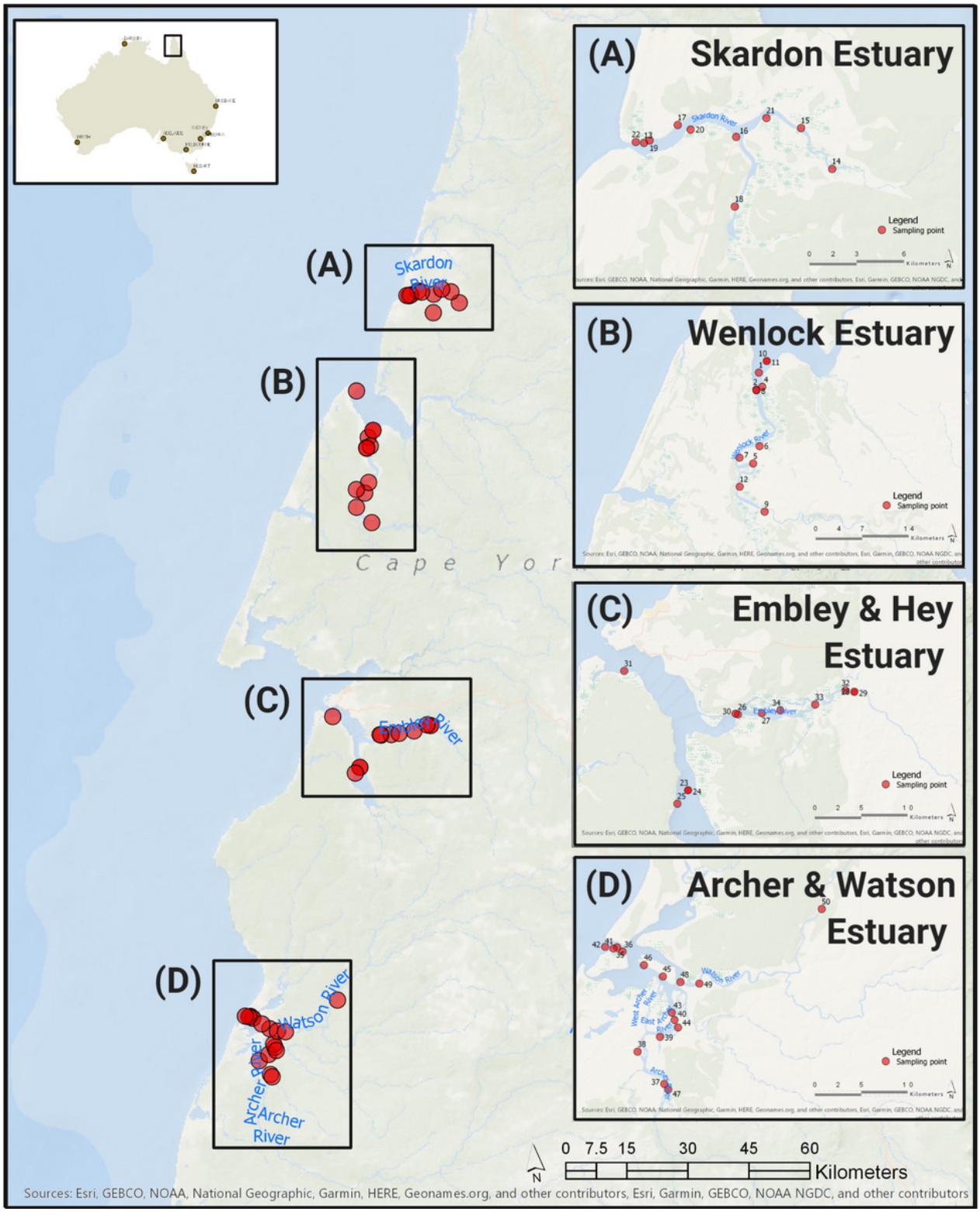

Figure 1

Map of the Western Cape York Peninsula (Far North Queensland, Australia), with the four sampled estuary systems identified. The sediment sampling sites (n=50) are annotated by the red circles.

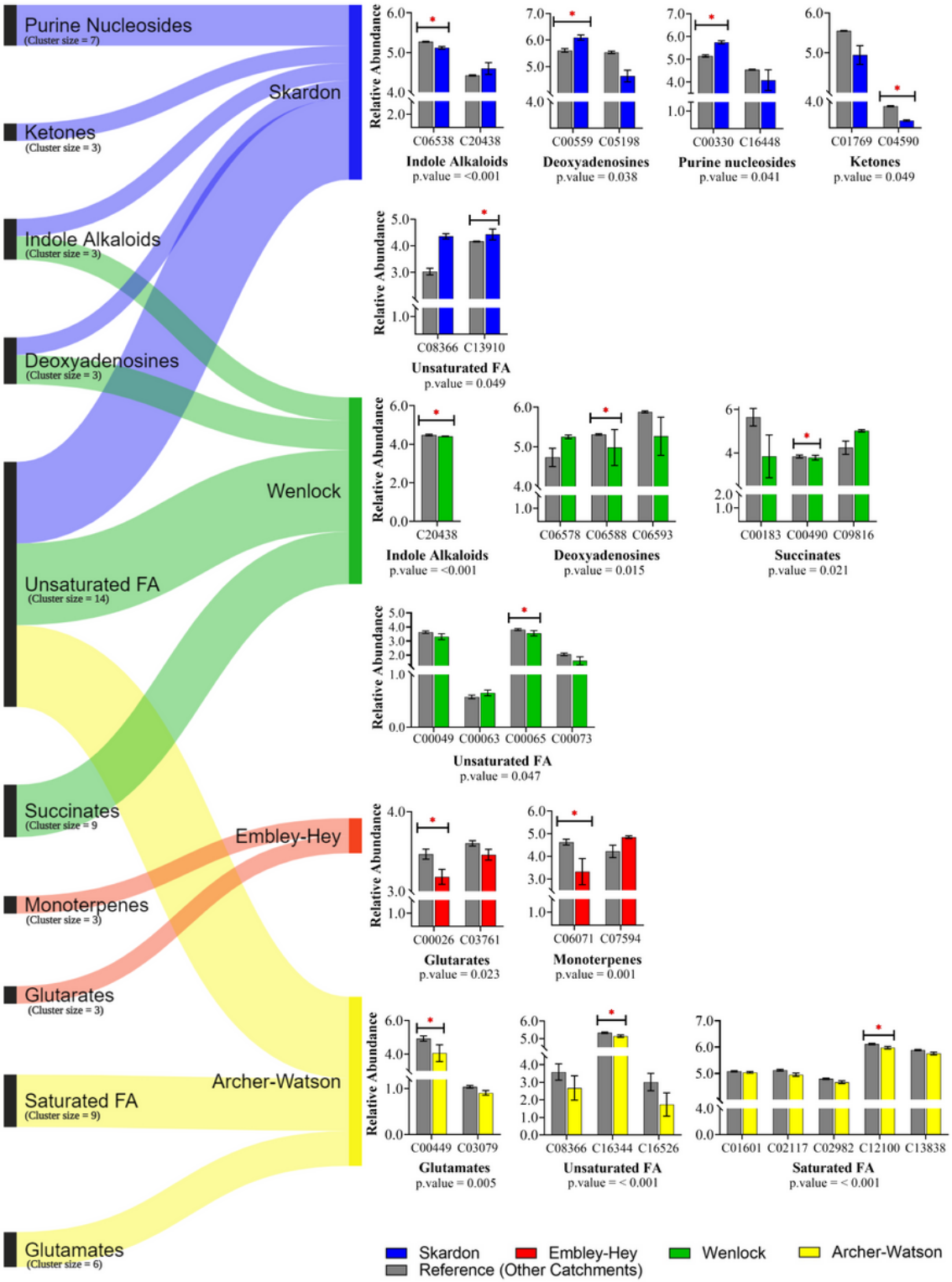

Figure 2

Identified significant metabolite clusters that differentiate the four sampled Western Cape York Peninsula estuaries using ChemRICH. The key compounds identified from each significant cluster are annotated with a red asterisk, in addition to stating the calculated p-value per cluster under the x-axis label. The cluster size is displayed in the parenthesis under each cluster label in the Sankey plot. The reference

catchment (i.e., each grey column) represents the other catchments estuaries to which the estuary of interest is compared against.

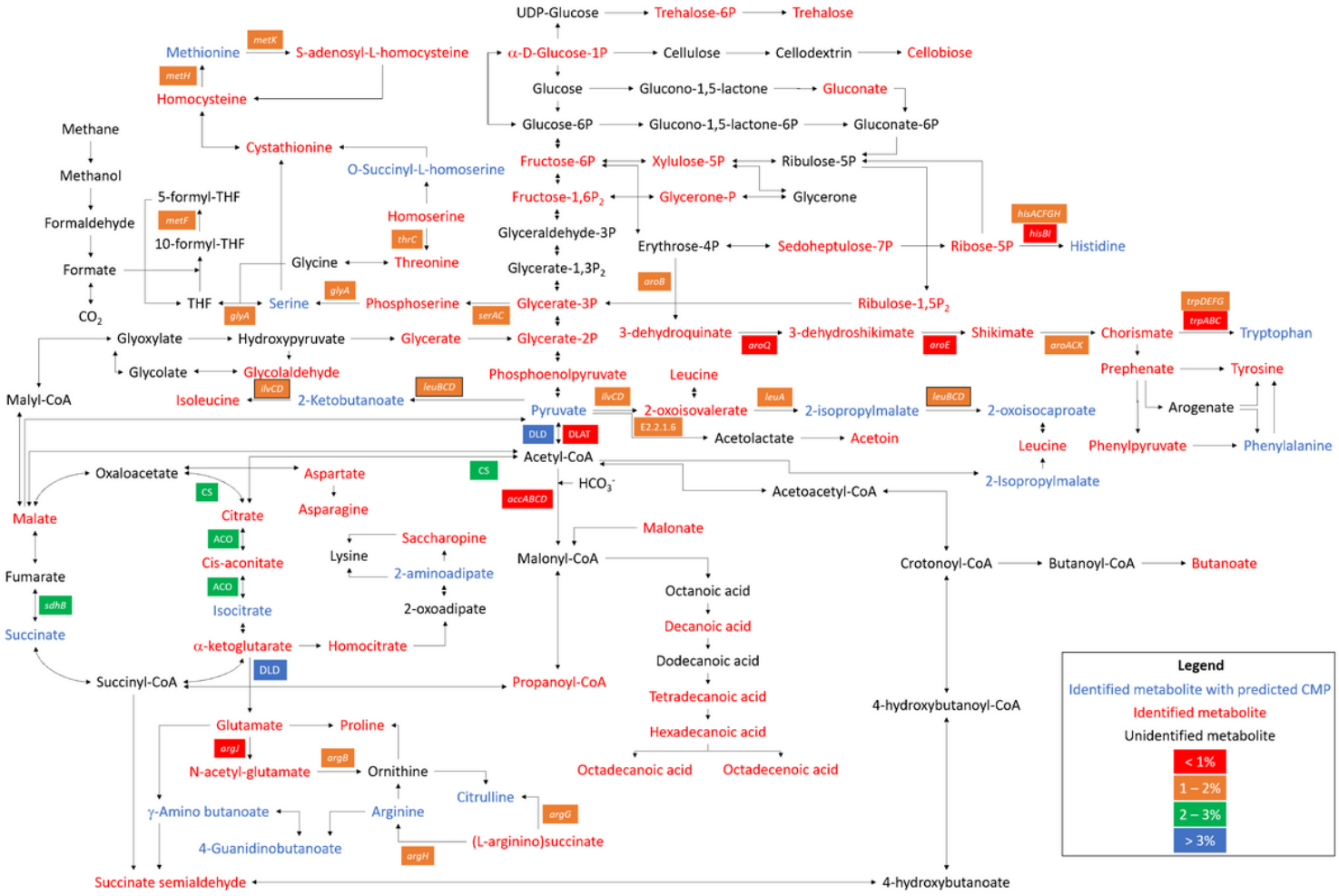

Figure 3

Mapping of metabolites and functional genes involved in carbon metabolism and biosynthesis of amino acids in estuarine sediments from the Western Cape York Peninsula. Identified metabolites with a predicted CMP score are displayed in blue text, identified metabolites (with no CMP score) are displayed in red text, and unidentified metabolites but important in terms of blueprint curation are displayed in black text. Genes are color-coded according to their relative abundance.

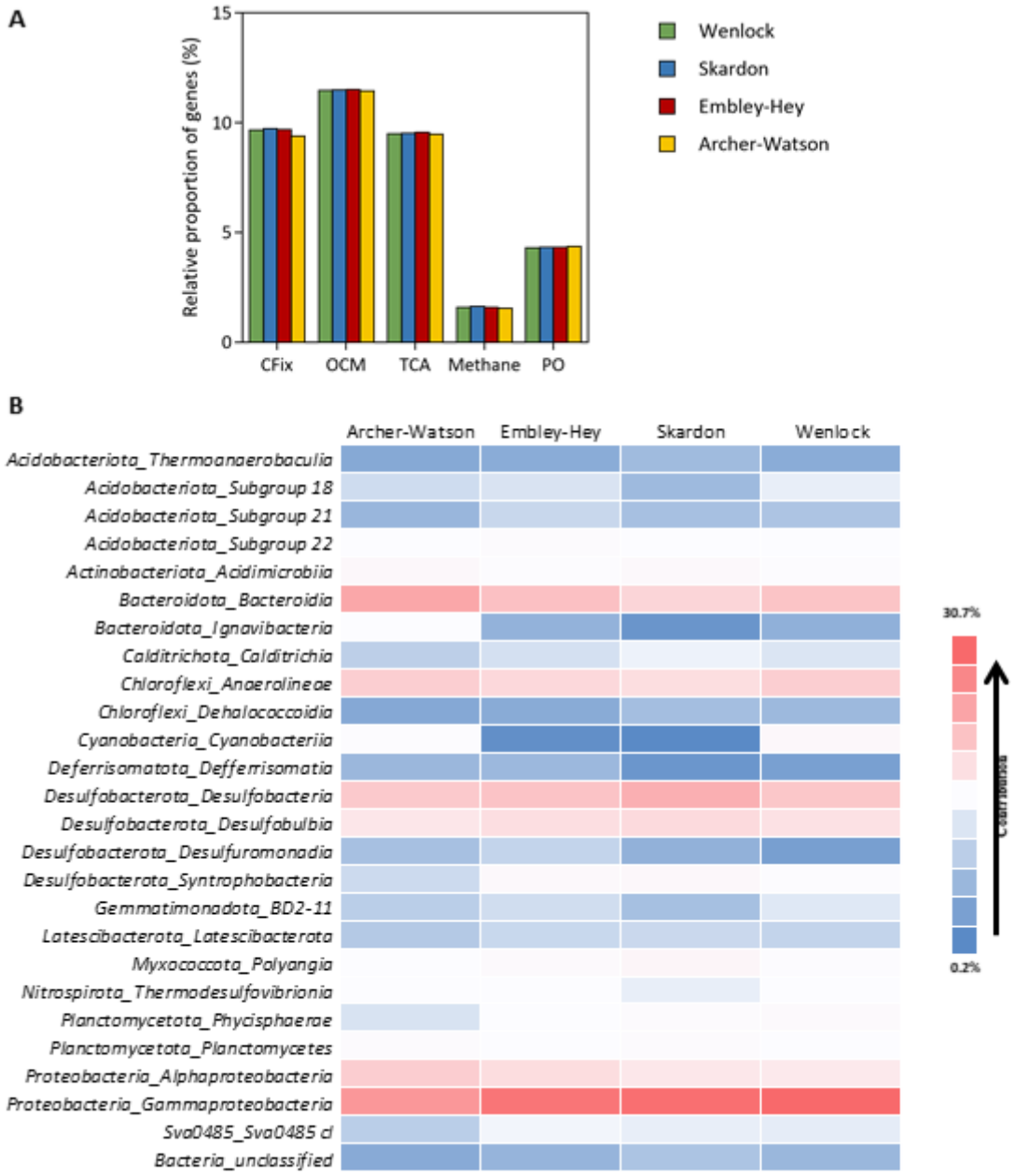

Figure 4

Central carbon metabolism in estuarine sediments. (A) Relative proportion of genes contributing to various pathways within central carbon metabolism, and (B) Taxonomic contribution to carbon metabolism.

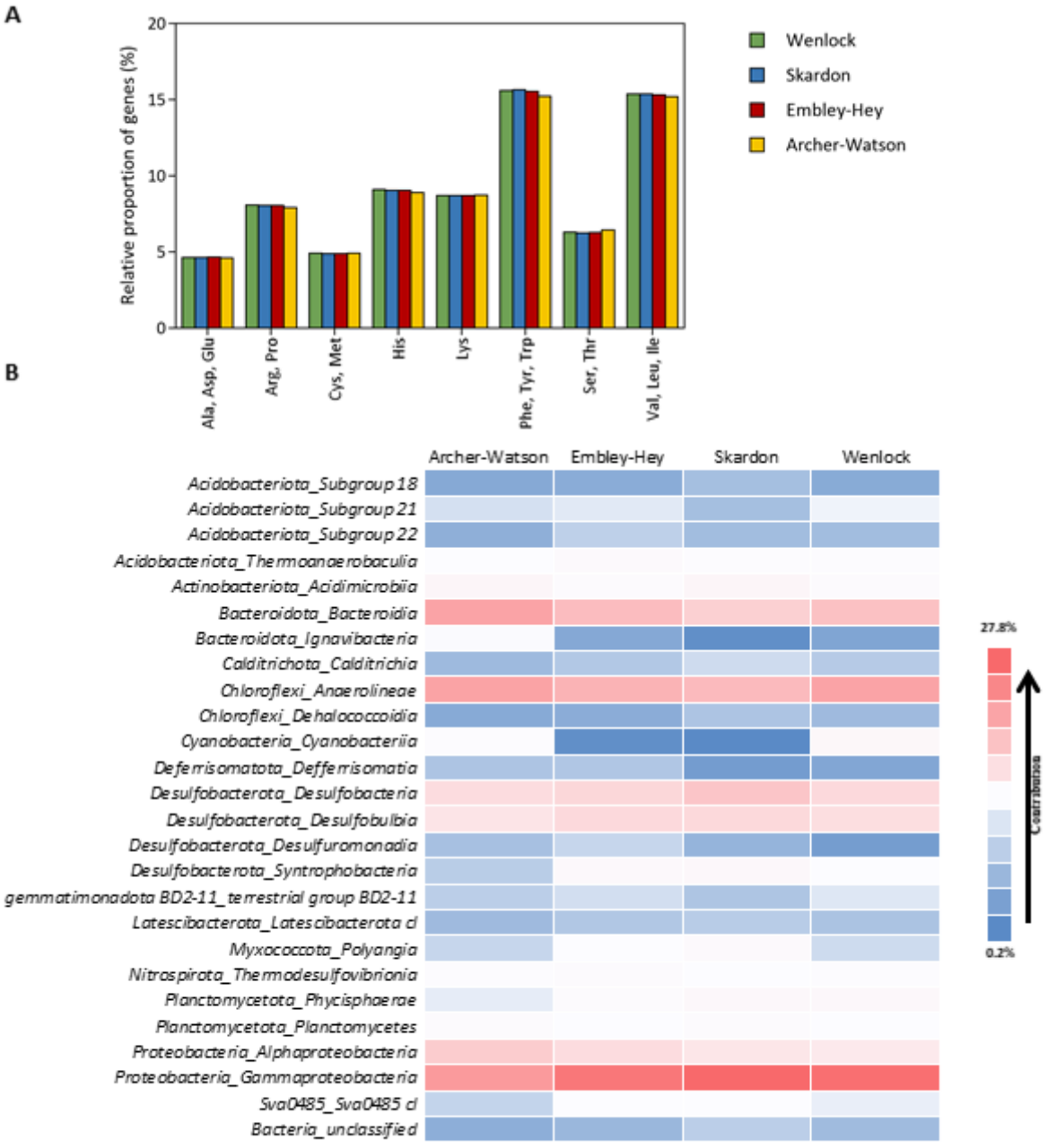

Figure 5

Biosynthesis of amino acids in estuarine sediments. (A) Relative proportion of genes contributing to various pathways related to biosynthesis of amino acids, and (B) Taxonomic contribution to biosynthesis of amino acids.

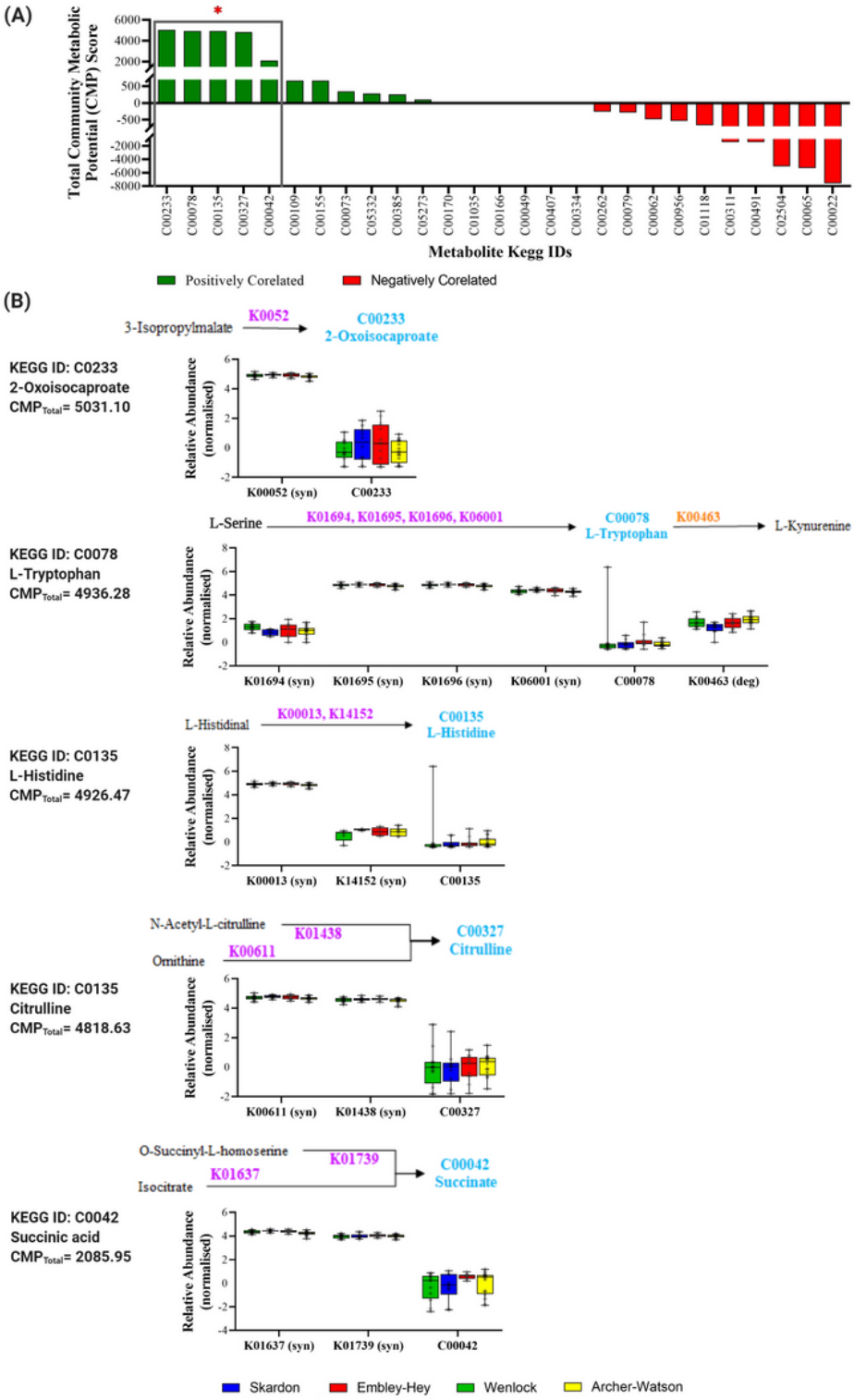

Figure 6

Positively correlated community metabolic potentials (CMP). (A) Top 5 identified metabolites with a positively correlated CMP score, as derived using the MIMOSA2; (B) Identified metabolic pathways (metabolites and genes). Note, metabolites with predicted CMP scores are annotated in blue text, genes contributing to synthesis are marked as (syn) and shown in a purple text in the reaction pathway, genes contributing to degradation are marked with (deg) and shown in orange text.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- ShahetalNatureFNQomicsSupplementary.docx