



University
of Glasgow



Understanding microbial communities through *in situ* omics data synthesis

Dr Umer Zeeshan Ijaz

Reader in Information Engineering, University of Glasgow

<http://userweb.eng.gla.ac.uk/umer.ijaz>

[Water & Environment Group](#) (James Watt School of Engineering, University of Glasgow)

[Bacteria, Immunology, Nutrition, Gastroenterology and OMICS \(BINGO\) Group](#) (School of Medicine, Dentistry & Nursing, University of Glasgow)

[Decentralised Water Technologies](#) (EPSRC, and the Scottish Water)

[A Genetic View into Past Sea Ice Variability in the Arctic \(AGENSI\) Group](#) (Bergen, Norway)

[Glasgow Water Cluster](#) (University of Glasgow)

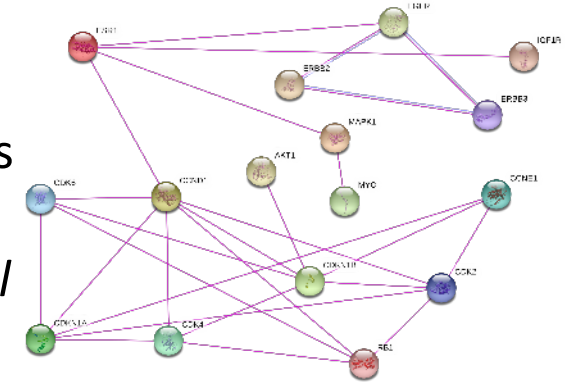
[Blackleg Hub](#) (The James Hutton Institute)

[Bacterial Plant Diseases UK](#) (BBSRC, NERC, Defra, and the Scottish Government)

Information Engineer

Information Engineering

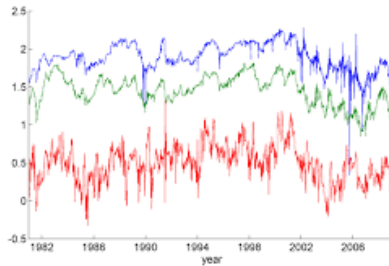
..defined as the generation, distribution, analysis and use of information in systems. This later definition involves the usage of *machine learning*, *data mining* and other *computational methods* to enhance the presentation and understanding of the high-throughput data that is generated by different systems. **An example is bioinformatics applying information engineering to biological data.**



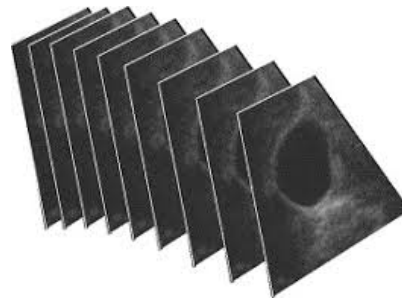
Network inference,
e.g., co-occurrence
networks of taxa



**String
Processing,**
e.g., DNA
sequences



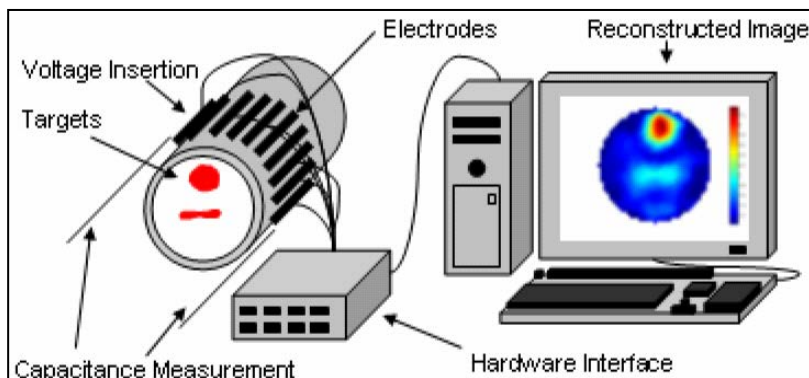
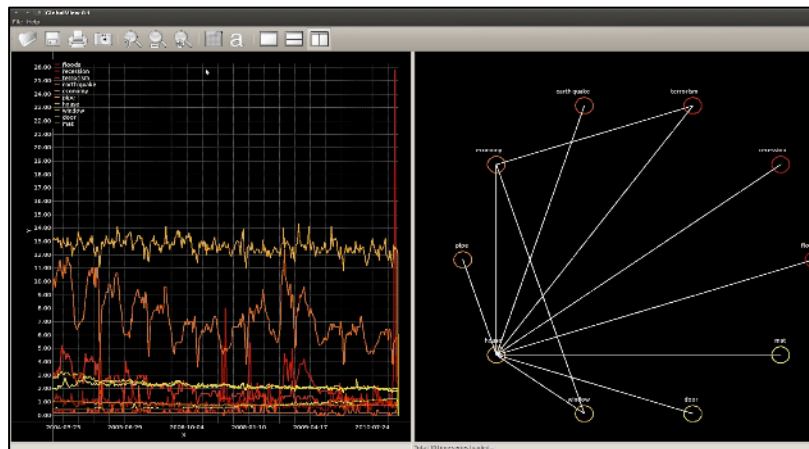
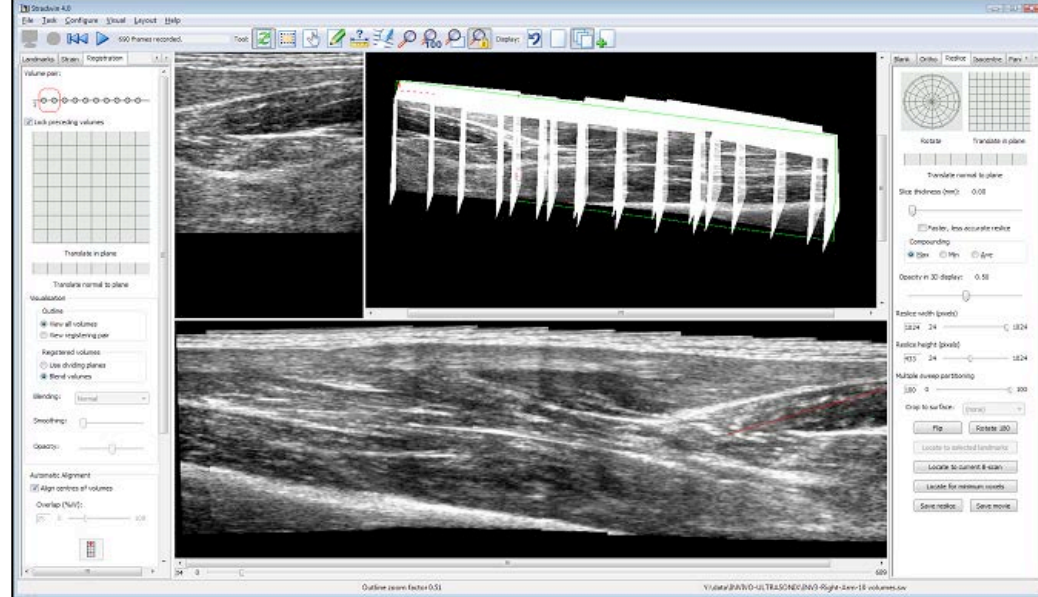
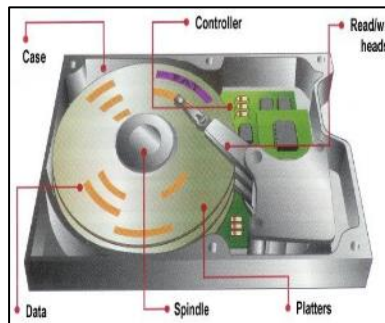
**1-
Dimensional,**
e.g., time
series, Raman
Spectra



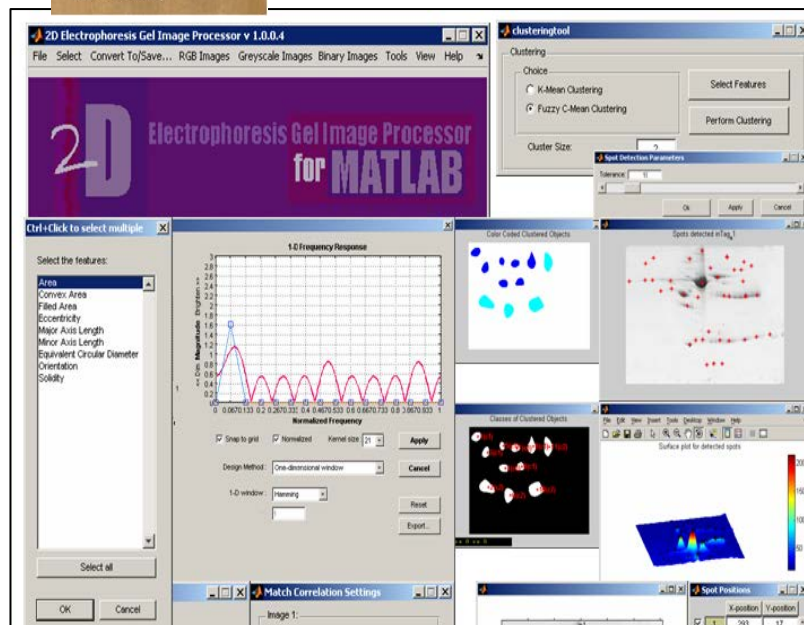
**2-
Dimensional,**
e.g.,
ultrasound
images

| | Columns | | | |
|---|---------|--------|--------|--------|
| | 0 | 1 | 2 | 3 |
| 0 | [0][0] | [0][1] | [0][2] | [0][3] |
| 1 | [1][0] | [1][1] | [1][2] | [1][3] |
| 2 | [2][0] | [2][1] | [2][2] | [2][3] |
| 3 | [3][0] | [3][1] | [3][2] | [3][3] |

**N-
Dimensional,**
e.g., tables
from
meta*omics



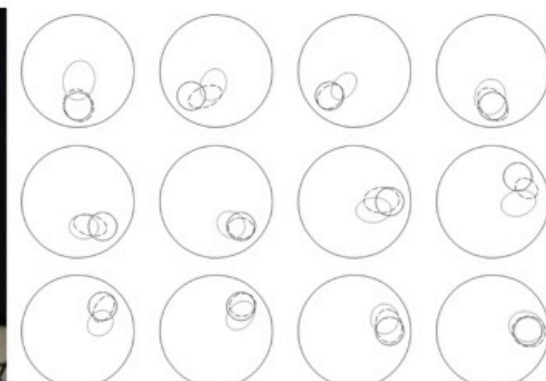
Cambridge/Oxford/
South Korea (2004-2012)



Dynamic Electrical Impedance Tomography Measurement System



Phantom with boundary electrodes where electric current is injected and voltages are recorded



Solid line: True boundary of plastic rod moving in the phantom
Dotted line: Boundary estimated by extended Kalman filter
Dashed line: Boundary estimated by unscented Kalman filter

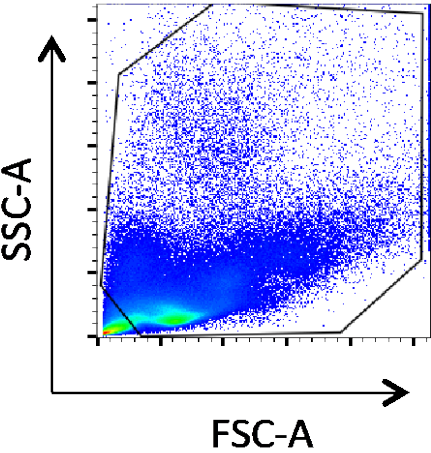
Environmental'Omics Lab (2012-present)

NERC Omics Fellowship: Microbial Ecology



Hardware devices

Whole genome shotgun metagenomics



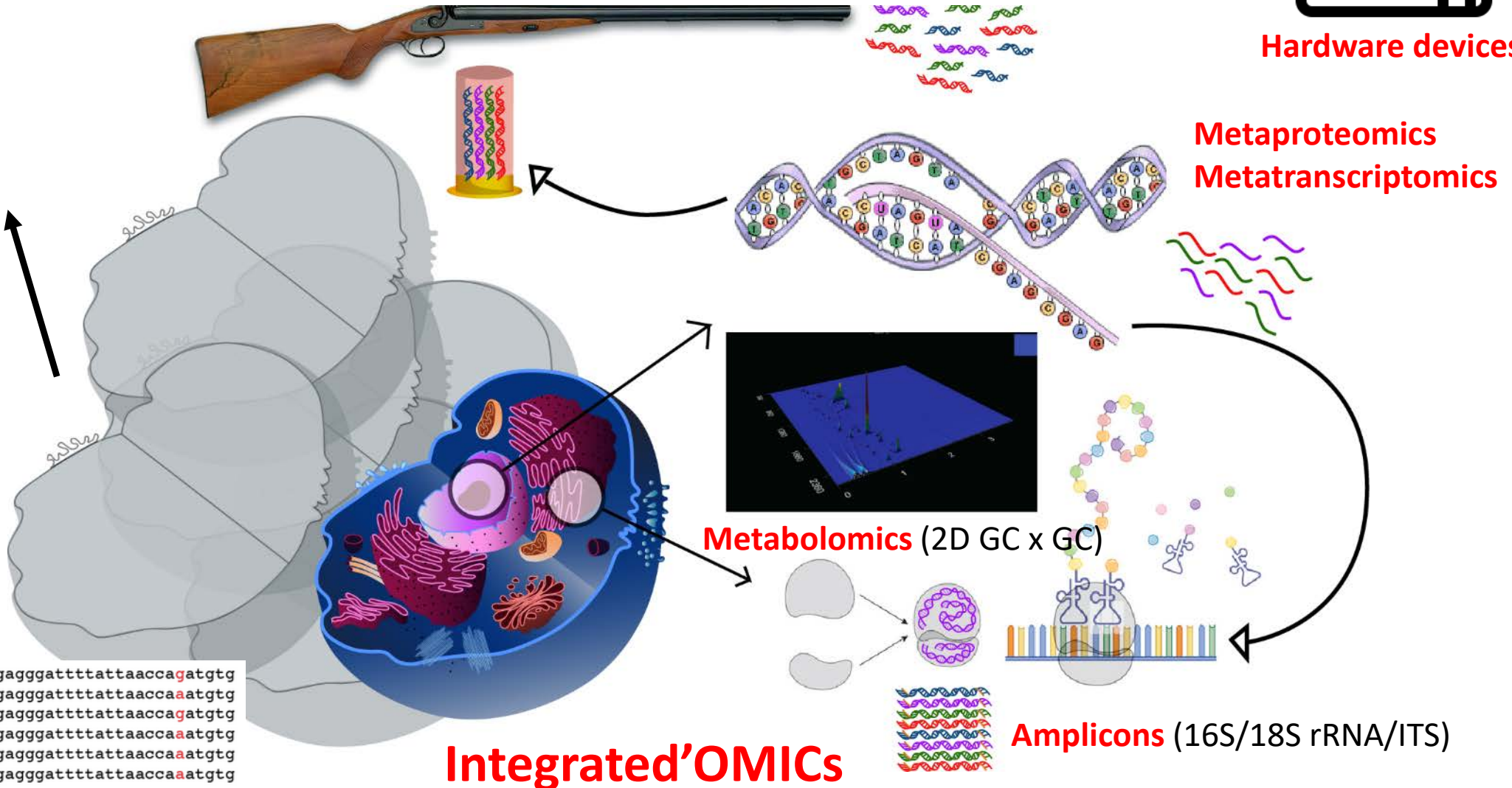
Flow cytometry:
Immunophenotyping

Host associated

```

1: acggttagctacaattattttaaaggaggaggatatttattaaccaatgtg
2: acggttatctacaattattttaaaggaggaggatatttattaaccaatgtg
3: acggttaactacaattattttaaaggaggaggatatttattaaccaatgtg
4: acggttaactacaattattttaaaggaggaggatatttattaaccaatgtg
5: acggttatctacaattattttaaaggaggaggatatttattaaccaatgtg
6: acggttatctacaattattttaaaggaggaggatatttattaaccaatgtg
    
```

Plants: Population genomics (including RADseq)



Integrated'OMICS

Amplicons (16S/18S rRNA/ITS)



| Sample ID | Treatment | Host |
|-----------|-----------|------|
| Sample 1 | treated | A |
| Sample 2 | treated | B |
| Sample 3 | control | A |
| Sample 4 | control | B |
| | | |

Recent Developments

- Omics approaches
 - Amplicons/Metagenomics, Metabolomics (2D GCxGC), Flow Cytometry, Metatranscriptomics, single-cell RNASeq, Population Genomics
- Bioinformatics
 - Benchmarking studies (Library preparations/Protocols/Platforms)
 - Workflow development
 - Text mining strategies
- Multivariate statistical analyses
 - Statistical packages (standalone/run on mobile phones)

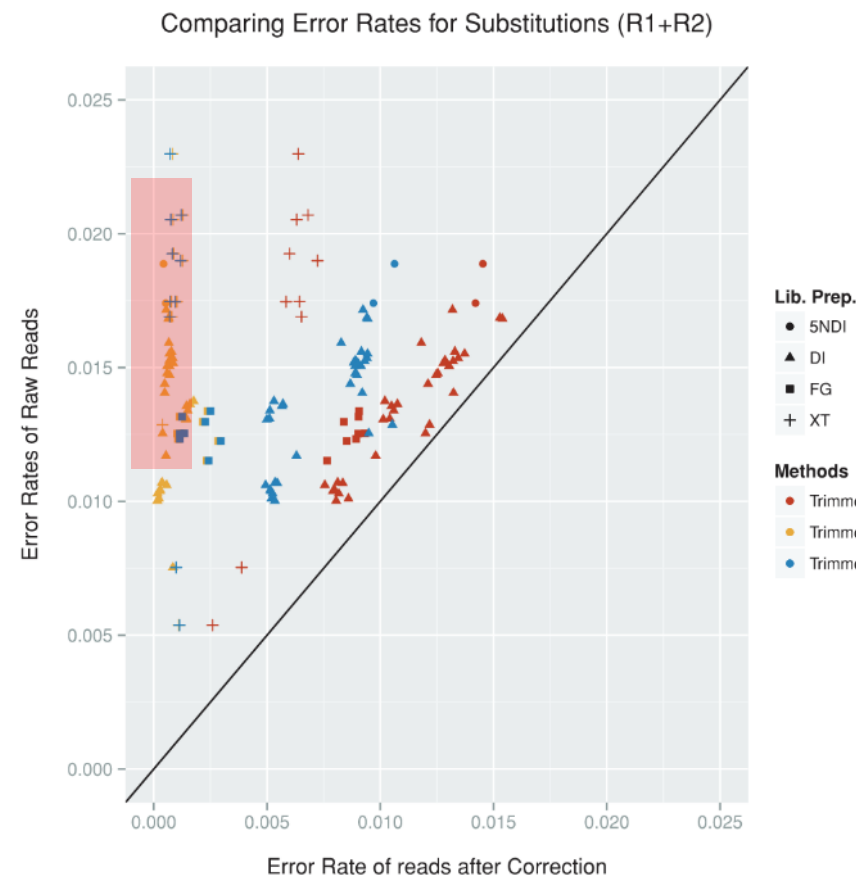
Benchmarking Studies

Published online 13 January 2015

Nucleic Acids Research, 2015, Vol. 43, No. 6 e37
doi: 10.1093/nar/gku1341

Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform

Melanie Schirmer^{1,*}, Umer Z. Ijaz¹, Rosalinda D'Amore², Neil Hall², William T. Sloan¹ and Christopher Quince¹



Software strategies to decrease error rates

SCIENTIFIC REPORTS

(2019) 9:9328 | <https://doi.org/10.1038/s41598-019-44974-x>

DNA extraction and amplicon production strategies deeply influence the outcome of gut mycobiome studies

Alessandra Frau¹, John G. Kenny^{2,7}, Luca Lenzi², Barry J. Campbell¹, Umer Z. Ijaz¹, Carrie A. Duckworth¹, Michael D. Burkitt^{1,4}, Neil Hall², Jim Anson⁶, Alistair C. Darby² & Christopher S. J. Probert¹

Schirmer et al. BMC Bioinformatics (2016) 17:125
DOI 10.1186/s12859-016-0976-y

BMC Bioinformatics

RESEARCH ARTICLE

Open Access

Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data

Melanie Schirmer^{1,2,4,*}, Rosalinda D'Amore³, Umer Z. Ijaz⁴, Neil Hall³ and Christopher Quince⁵

Gerasimidis et al. BMC Res Notes (2016) 9:365
DOI 10.1186/s13104-016-2171-7

BMC Research Notes

RESEARCH ARTICLE

Open Access

The effect of DNA extraction methodology on gut microbiota research applications

Konstantinos Gerasimidis¹, Martin Bertz¹, Christopher Quince², Katja Brunner¹, Alanna Bruce¹, Emilie Combet¹, Szymon Calus³, Nick Loman⁴ and Umer Zeeshan Ijaz^{3*}

D'Amore et al. BMC Genomics (2016) 17:55
DOI 10.1186/s12864-015-2194-9

RESEARCH ARTICLE

Open Access

A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling

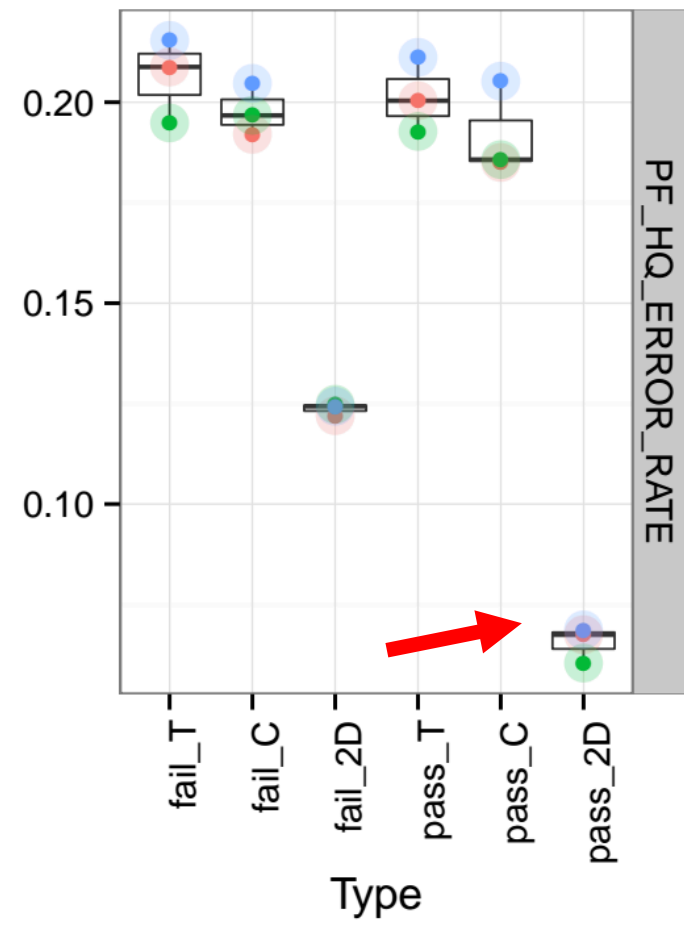
Rosalinda D'Amore^{1,1}, Umer Zeeshan Ijaz^{2,1}, Melanie Schirmer², John G. Kenny¹, Richard Gregory¹, Alistair C. Darby¹, Migun Shakya³, Mircea Podar⁴, Christopher Quince^{5*} and Neil Hall^{1*}

NanoAmpli-Seq: a workflow for amplicon sequencing for mixed microbial communities on the nanopore sequencing platform

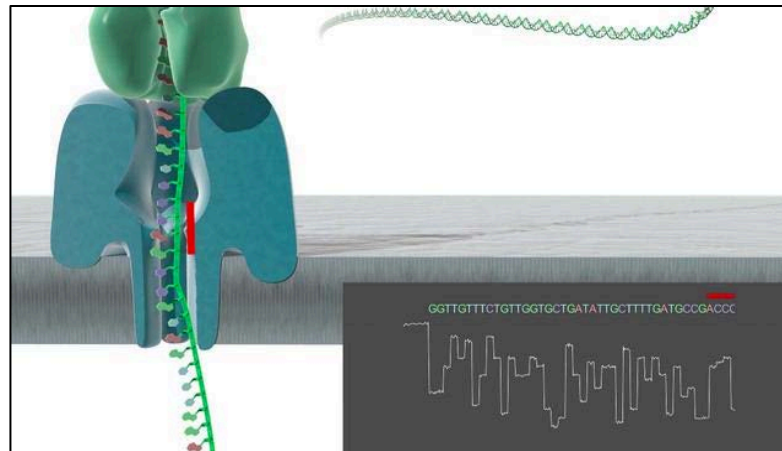
Szymon T Calus, Umer Z Ijaz, Ameet J Pinto

GigaScience, Volume 7, Issue 12, December 2018, giy140, <https://doi.org/10.1093/giga-science/giy140>

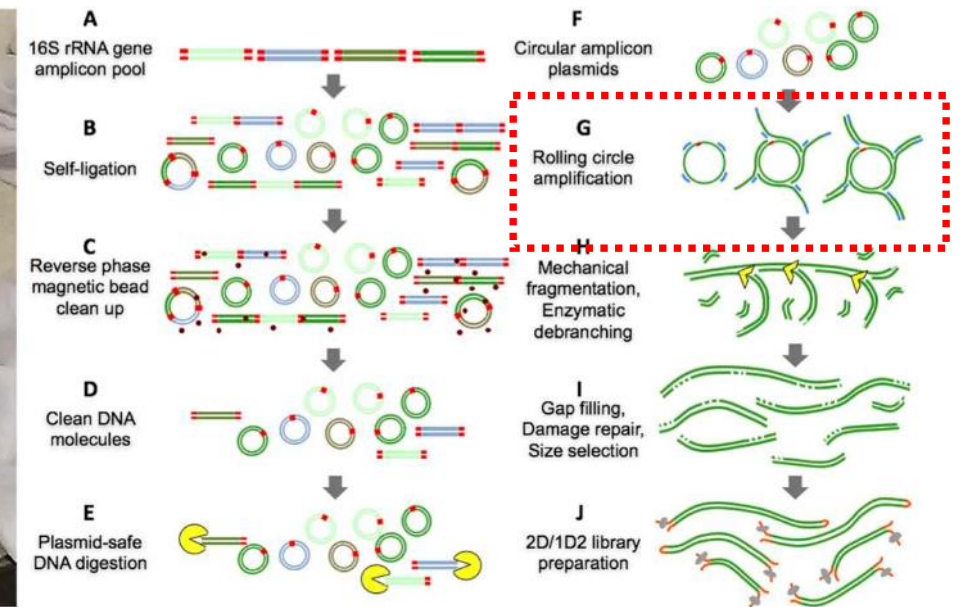
Published: 23 November 2018 Article history



Template (T)
Complement (D)



Overview of the sample preparation protocol for 16S rRNA gene amplicon pool preparation, plasmid-like structure construction, enzymatic debranching and mechanical fragmentation, and 2D and 1D2 library preparation including intermediate cleanup steps.



Solution: Concatamerization

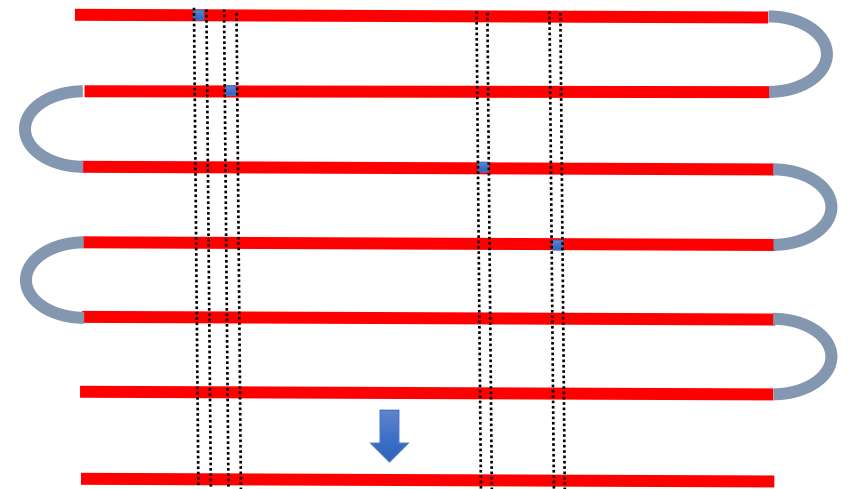


Table 2: Number of OTUs detected and consensus sequence accuracy for all experiments using the nanoClust for OTU clustering and consensus calling approach

| Protocol Experiment | Number of OTUs | | | | Average consensus accuracy (%) | | | |
|------------------------|--------------------|---------------------------------|--------------------|---------------------------------|--------------------------------|---------------------------------|--------------------|---------------------------------|
| | 2D One organism | 1D ² One organism | 2D Ten organism | 1D ² Ten organism | 2D One organism | 1D ² One organism | 2D Ten organism | 1D ² Ten organism |
| Run | 1 | 4 | 2 | 3 | 1 | 4 | 2 | 3 |
| Theoretical | 1 | 1 | 10 | 10 | | | | |
| INC-Seq aligner | Blastn | | | | | | | |
| OTUs detected | 1 | 1 | 11 | 11 | 99.36 | 99.5 | 99.47 | 99.61 |
| Spurious OTUs | 0 | 0 | 1 | 1 | – | – | 99.22 | 99.37 |
| Non-Detect | 0 | 0 | 0 | 0 | – | – | – | – |
| INC-Seq aligner | Graphmap | | | | | | | |
| OTUs detected | 1 | 1 | 11 | 11 | 99.43 | 99.43 | 99.44 | 99.61 |
| Spurious OTUs | 0 | 0 | 1 | 1 | – | – | 99.29 | 99.5 |
| Non-Detect | 0 | 0 | 0 | 0 | – | – | – | – |
| INC-Seq aligner | POA | | | | | | | |
| OTUs detected | 1 | 2 | 10 | 12 | 99.5 | 99.61 | 99.60 | 99.52 |
| Spurious OTUs | 0 | 1 | 0 | 2 | – | 99.57 | – | 98.67 |
| Non-Detect | 0 | 0 | 0 | 0 | – | – | – | – |

1.6KB amplicons with ~99.5% accuracy for mixed communities using nanopore

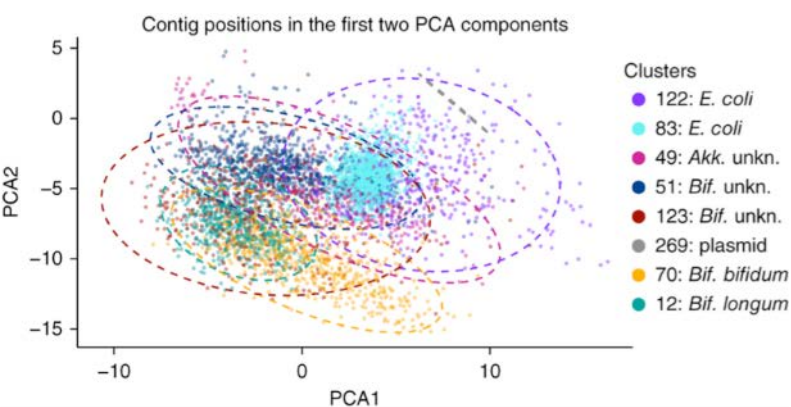
Binning metagenomic contigs by coverage and composition

Johannes Alneberg^{1,8}, Brynjar Smári Bjarnason^{1,8}, Ino de Bruijn^{1,2}, Melanie Schirmer³, Joshua Quick^{4,5}, Umer Z Ijaz³, Leo Lahti^{6,7}, Nicholas J Loman⁴, Anders F Andersson^{1,9} & Christopher Quince^{3,9}



A Java-based statistical framework for integration of shotgun metagenomics with other omics technologies (Orges Koci; PhD student)

<https://github.com/KociOrges/cviewer>



Each cluster represents contigs (assembled from shotgun sequencing reads) belonging to the same species



Schematic demonstrating the software layout and the main features that are supported in the system.

Research article | Bioinformatics | Ecology | Ecosystem Science | Environmental Sciences | Microbiology

Extending SEQenv: a taxa-centric approach to environmental annotations of 16S rDNA sequences

Bioinformatics tool Biodiversity Biogeography Bioinformatics Microbiology

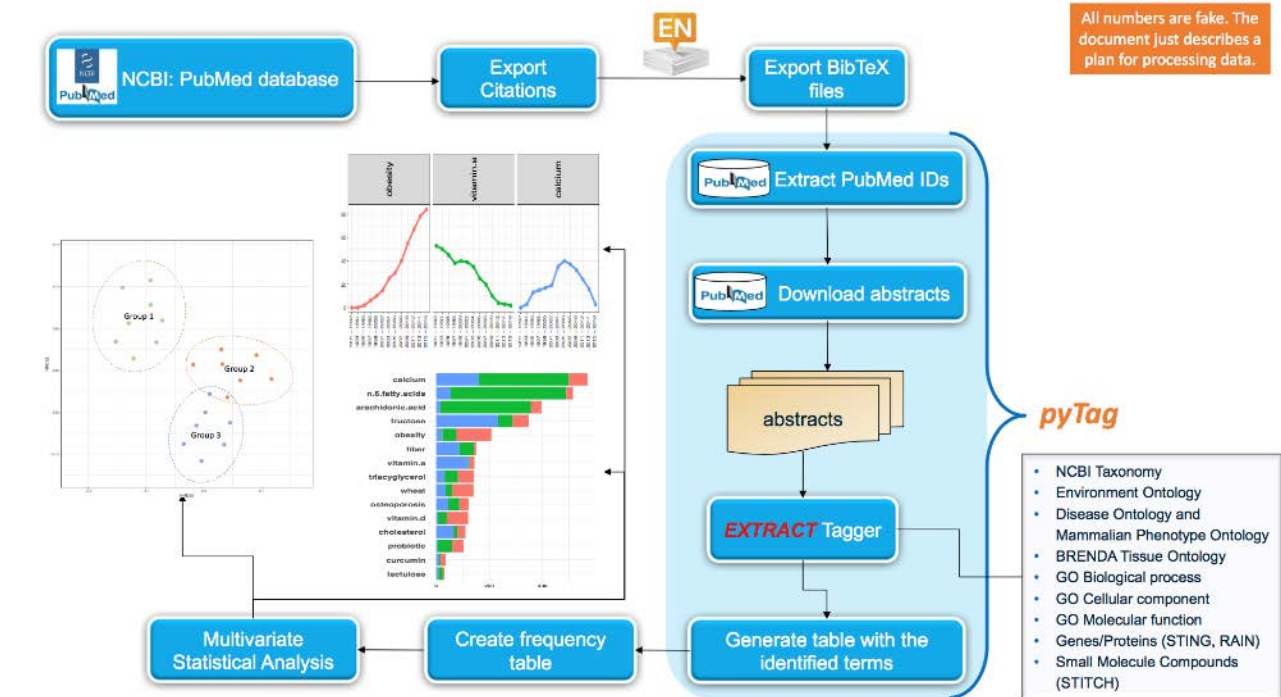
Published October 10, 2017 PubMed 29038749

Research article | Bioinformatics | Algorithms and Analysis of Algorithms | Data Mining and Machine Learning

Software Engineering Visual Analytics

Want to publish review articles? Eh?

How about you sit and relax and let the pipeline do systematic reviews for you?





Optimized R functions for analysis of ecological community data using the R virtual laboratory (RvLab)

Constantinos Varso[†], Theodore Patkos[†], Anastasis Oulas[§], Christina Pavloudi^{§,||}, Alexandros Gougousis[§], Umer Zeeshan Ijaz[†], Irene Filiopoulou[§], Nikolaos Pattakos[§], Edward Vanden Berghe[#], Antonio Fernández-Guerra^o, Sarah Faulwetter[§], Eva Chatzinikolaou[§], Evangelos Pafilis[§], Chrysoula Bekiri[†], Martin Doerr[†], Christos Arvanitidis[§]



Mobile RvLab application web services and screen shots. Overall workflow of the mobRvLab interface.

microbiomeSeq: An R package
for microbial community
analysis in an environmental
context

Introduction to microbiomeSeq

Data normalisation

Alpha diversity with ANOVA

Beta diversity

Canonical Correspondence Analysis

Differential Abundance

Co-occurrence pattern analysis

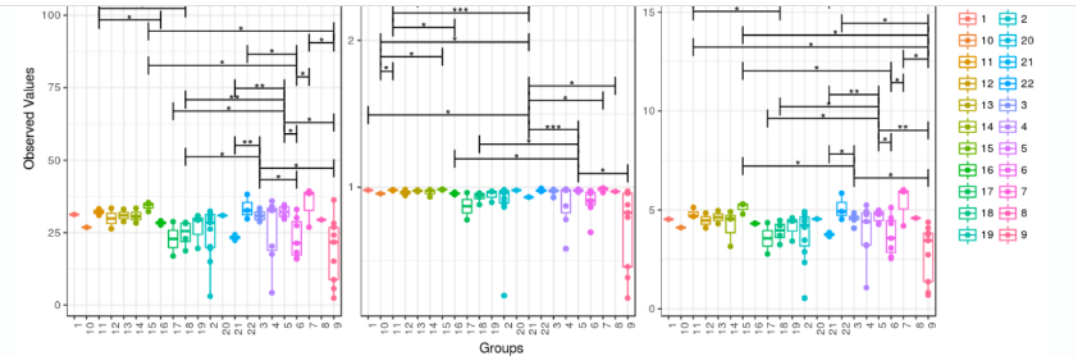
Correlation between taxa abundance
and environmental variables

Dependencies

References

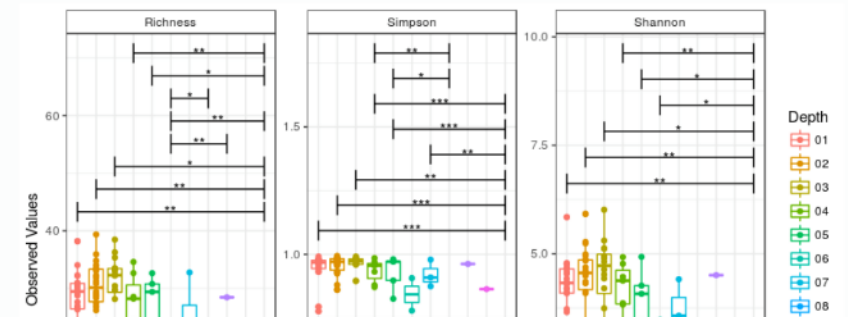
Alfred Ssekagiri, William T. Sloan, * Umer
Zeeshan Ijaz (* Correspondence:
Umer.Ijaz@glasgow.ac.uk)

2018-02-14



Grouping by Depth categorical variable, we obtain the following plot.

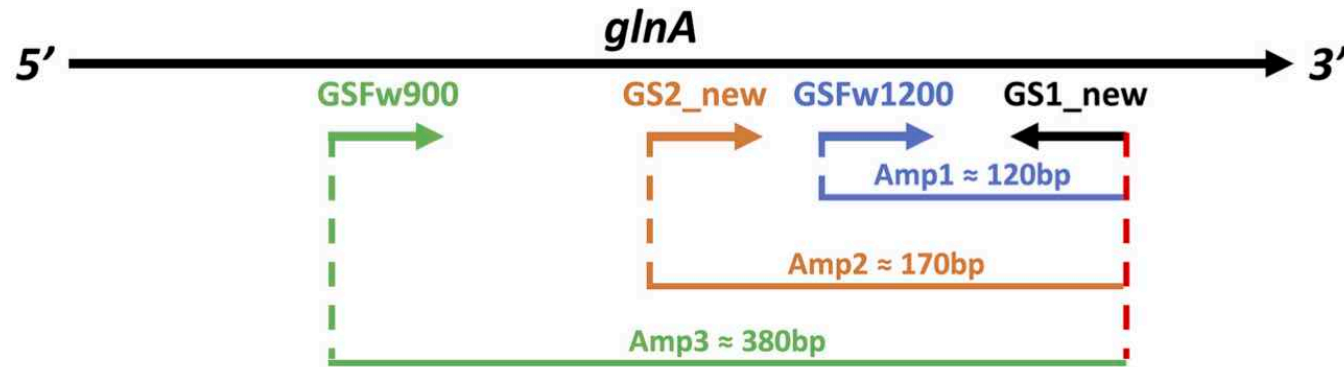
```
p <- plot_anova_diversity(physeq, method = c("richness", "simpson", "shannon"),  
  grouping_column = "Depth", pValueCutoff = 0.05)  
print(p)
```



Differential ratio amplicons (R_{amp}) for the evaluation of RNA integrity extracted from complex environmental samples

Fabien Cholet✉, Umer Z. Ijaz, Cindy J. Smith

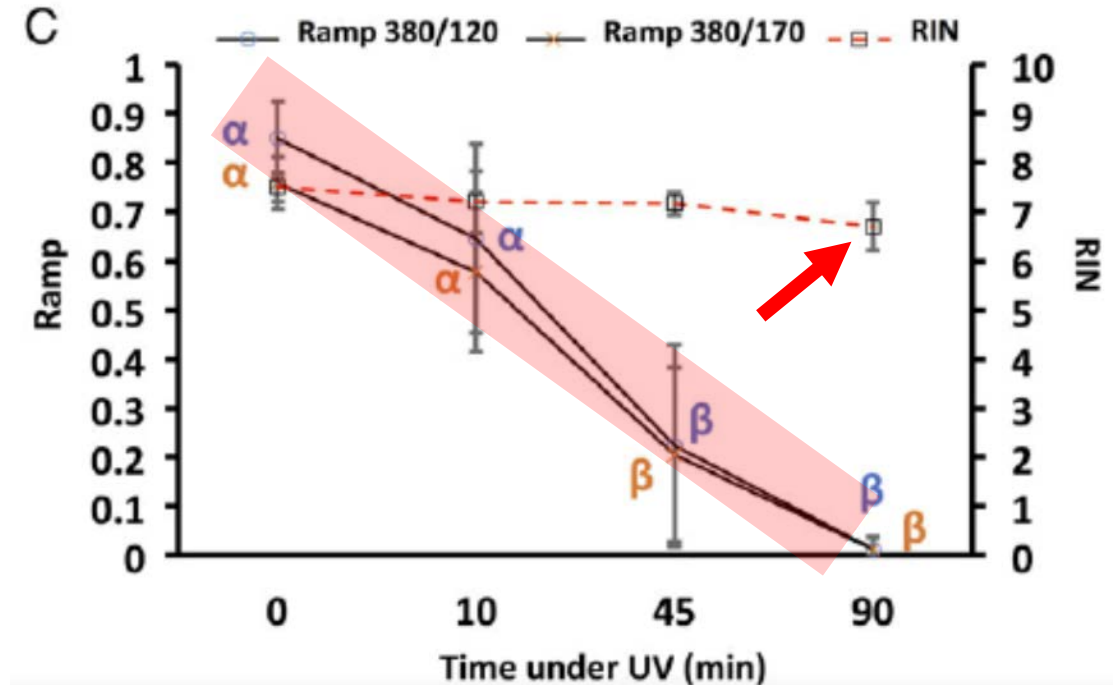
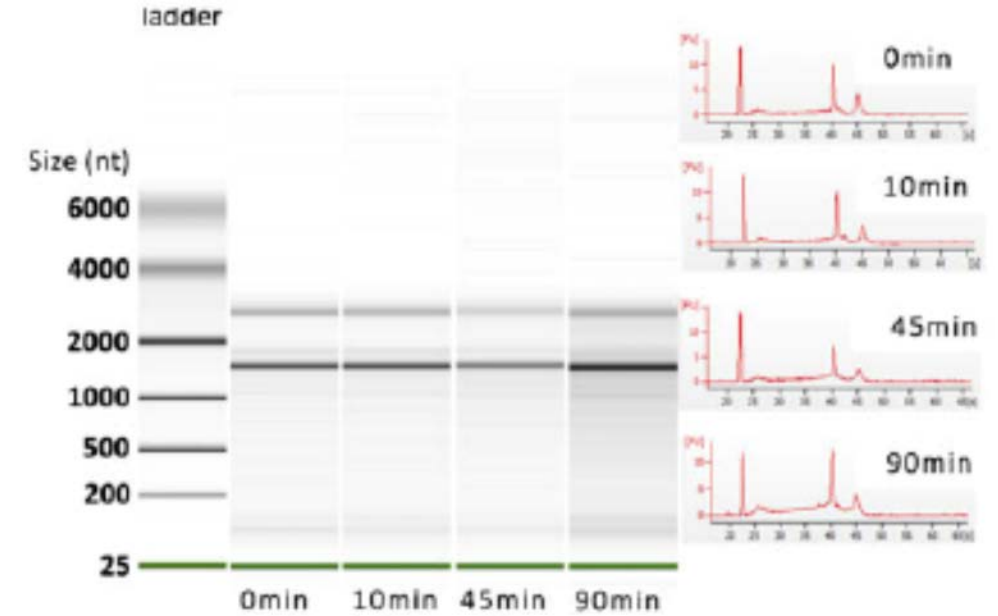
First published: 25 December 2018 | <https://doi.org/10.1111/1462-2920.14516> | Citations: 6



$$R_{amp} \text{ 380/120} = \frac{35 - Ct \text{ Amp 3}}{35 - Ct \text{ Amp 1}}$$

$$R_{amp} \text{ 380/170} = \frac{35 - Ct \text{ Amp 3}}{35 - Ct \text{ Amp 2}}$$

Schematic representation of primer binding sites along the Bacterial *glnA* gene. Primers are represented by arrows pointing to the right (forward primers) or to the left (reverse primer). The amplicons (Amp) generated by the different primer combinations are represented as coloured lines. The formulas used to calculate the two Ramp indexes are detailed under the figure.



Engineering Applications/Products

- Exploiting and harnessing microbial communities in both **Natural** and **Engineered** systems
- Can we **emulate natural processes** through bioreactors? (artificial intestine)?
- Can we improve drinking/waste water treatment systems?
 - Better filter design
 - Harvesting additional heat from households (Decentralised Water Technologies; EPSRC Programme Grant)
- Can we come up with a dietary intervention that can **modulate gut microbiome** against IBDs?
- Can we utilize microfluidics to select for those microbes that play a particular function?

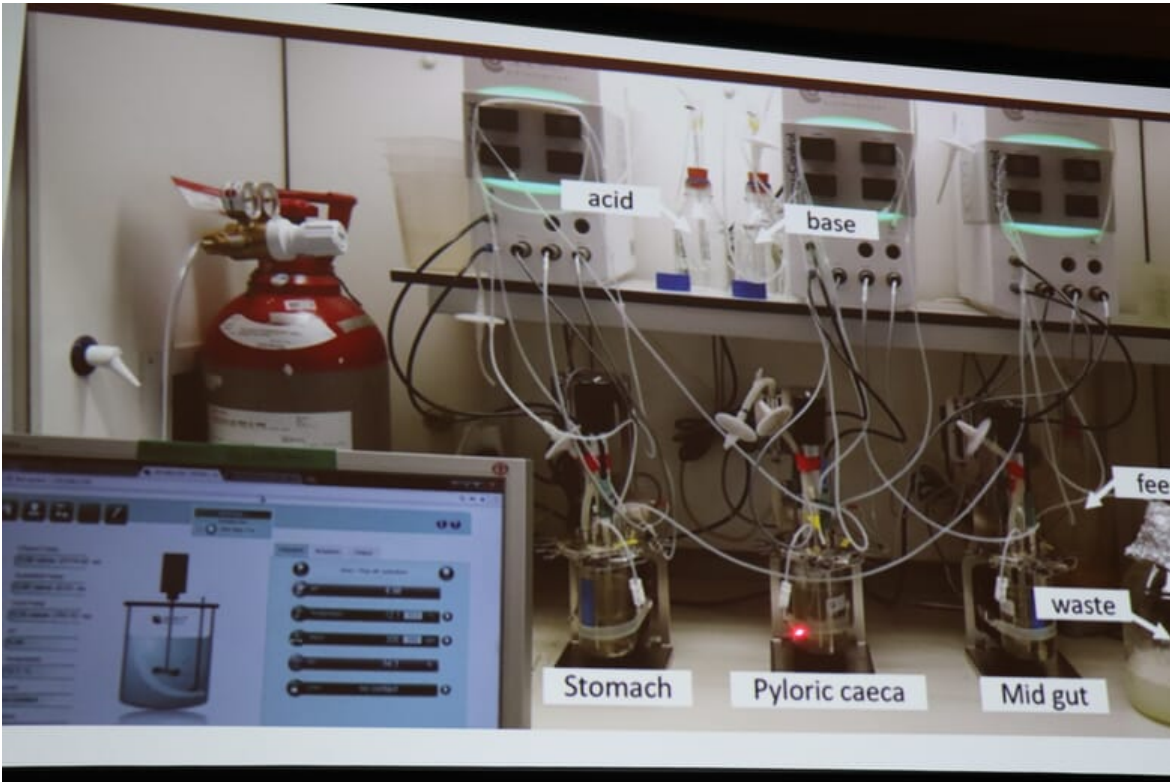
RESEARCH

Open Access



SalmoSim: the development of a three-compartment *in vitro* simulator of the Atlantic salmon GI tract and associated microbial communities

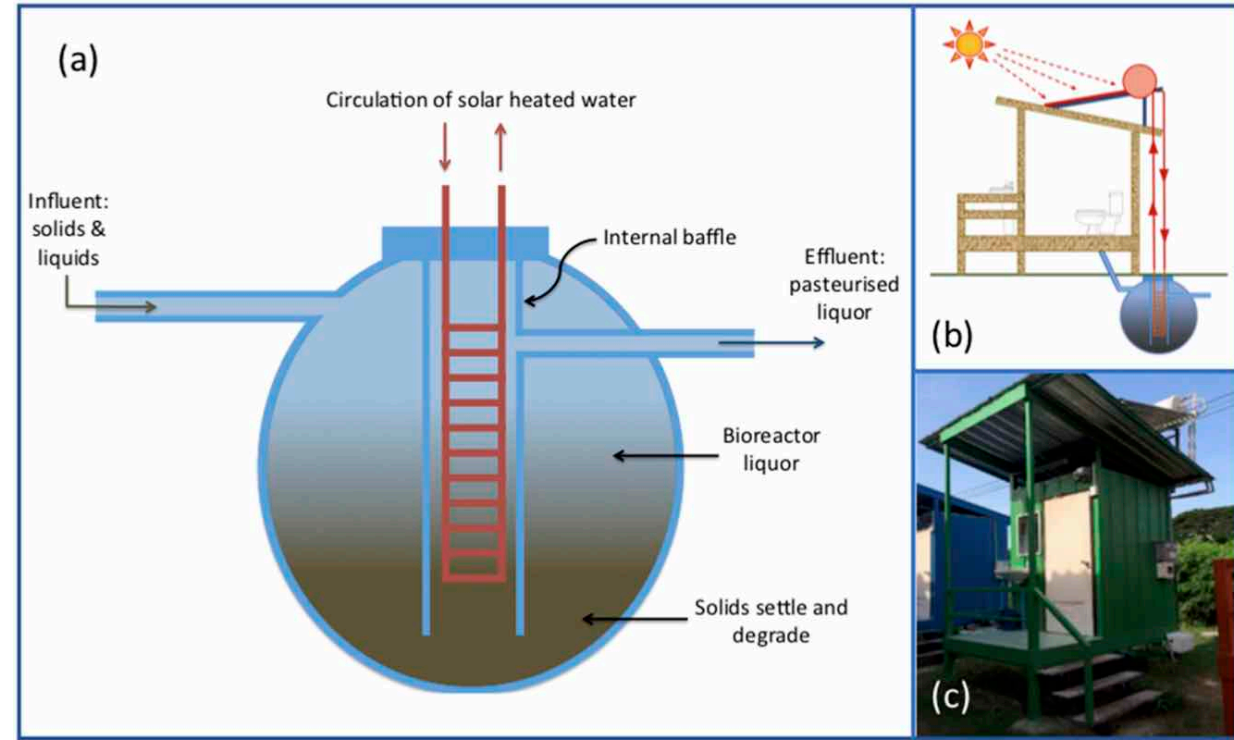
Raminta Kazlauskaitė^{1*}, Bachar Cheaib¹, Chloe Heys¹, Umer Zeeshan Ijaz², Stephanie Connelly², William Sloan², Julie Russel², Laura Rubio³, John Sweetman^{4,5}, Alex Kitts¹, Philip McGinnity^{6,7}, Philip Lyons^{4,5} and Martin Llewellyn¹



Article

Solar Septic Tank: Next Generation Sequencing Reveals Effluent Microbial Community Composition as a Useful Index of System Performance

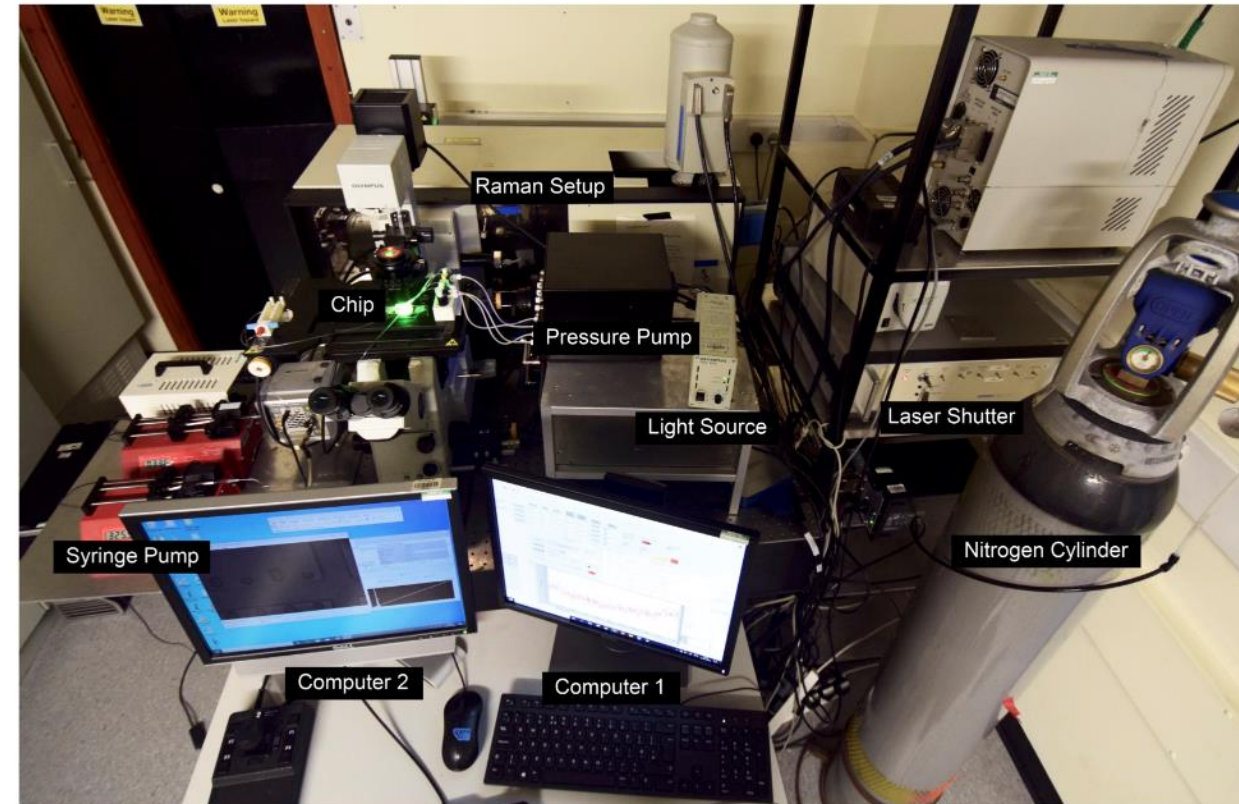
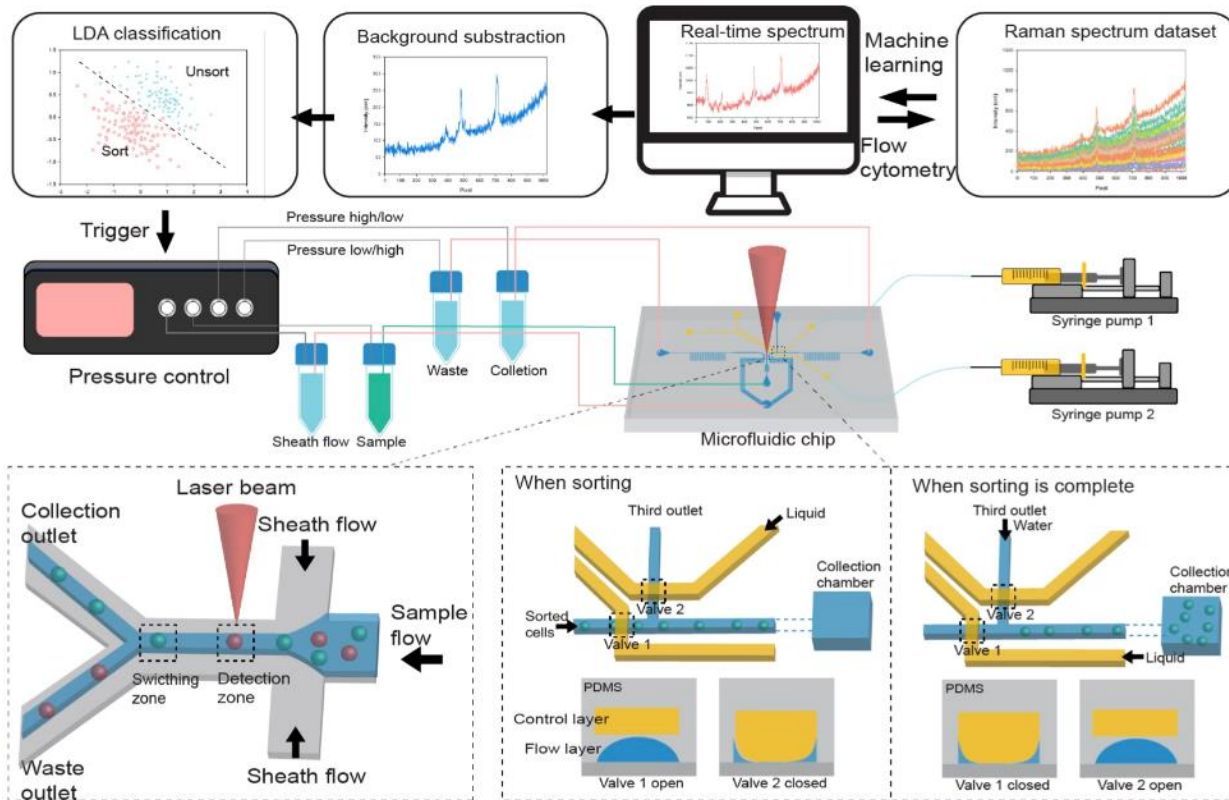
Stephanie Connelly^{1,*,†}, Tatchai Pussayanavin^{2,3,†}, Richard J. Randle-Boggis¹, Araya Wicheansan², Suparat Jampathong², Ciara Keating¹, Umer Z. Ijaz¹, William T. Sloan¹ and Thammarat Koottatep²



Schematic (a) illustrates the principle of the solar heating applied to the SST in contrast to the CT which operates at ambient temperature and without an internal baffle. Schematic (b) illustrates the installation of the SST at the field test site, showing the buried septic tank and the solar collection unit on the roof of the served toilet block seen in the photograph (c).

Raman-Activated Cell Sorting System

(NERC NE/P003826/1)



Overview of Smart Raman Activated Cell Sorting System: The biological specimen is introduced into the chip, hydro-dynamically focused by sheath-flow and flushed uniformly through the laser point. The system collects Raman spectrum every 80 milliseconds and analyzes it real-time. When the spectrum meets the sorting criteria based on our machine learning algorithms, the pressure changes and the bacteria are transferred to the collection direction. The details are given in Yu Chen Fu, MPhil in Research Thesis, University of Glasgow, UK (2021)

<https://theses.gla.ac.uk/82633/>

Treatment of Active Crohn's Disease With an Ordinary Food-based Diet That Replicates Exclusive Enteral Nutrition

Vaios Svolos,¹ Richard Hansen,² Ben Nichols,¹ Christopher Quince,³ Umer Z. Ijaz,⁴ Rodanthi T. Papadopoulou,¹ Christine A. Edwards,¹ David Watson,⁵ Adel Alghamdi,⁵ Asker Brejnrod,⁶ Cecilia Ansalone,⁷ Hazel Duncan,² Lisa Gervais,² Rachel Tayler,² Jonathan Salmond,⁸ Daniele Bolognini,⁹ Robert Klopffleisch,¹⁰ Daniel R. Gaya,¹¹ Simon Milling,⁷ Richard K. Russell,² and Konstantinos Gerasimidis¹



Prepacked CD-TREAT meals (developed by a PhD student, Vaios Svolos), a novel dietary therapy to control recurrence of gut inflammation in children with Crohn's disease in remission.



Contents lists available at ScienceDirect

Bioresource Technology

journal homepage: www.elsevier.com/locate/biortech



First proof of concept for full-scale, direct, low-temperature anaerobic treatment of municipal wastewater

Anna Christine Trego^{a,1}, B. Conall Holohan^{a,b,1}, Ciara Keating^c, Alison Graham^a, Sandra O'Connor^a, Michael Gerardo^d, Dermot Hughes^b, Umer Zeeshan Ijaz^{c,*}, Vincent O'Flaherty^a

Low-temperature anaerobic digestion of dilute municipal wastewater

Influent BOD: 0.5-200 mg/L

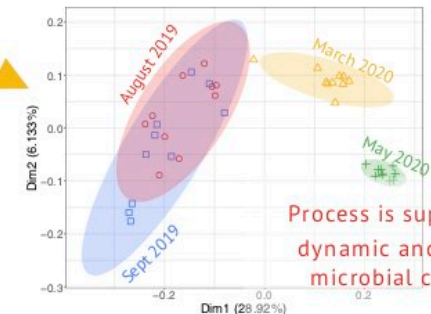
Temperature: 2-18 °C

HRT: 7 h → 3 h

Welsh Water WWTP
Builth Wells, UK

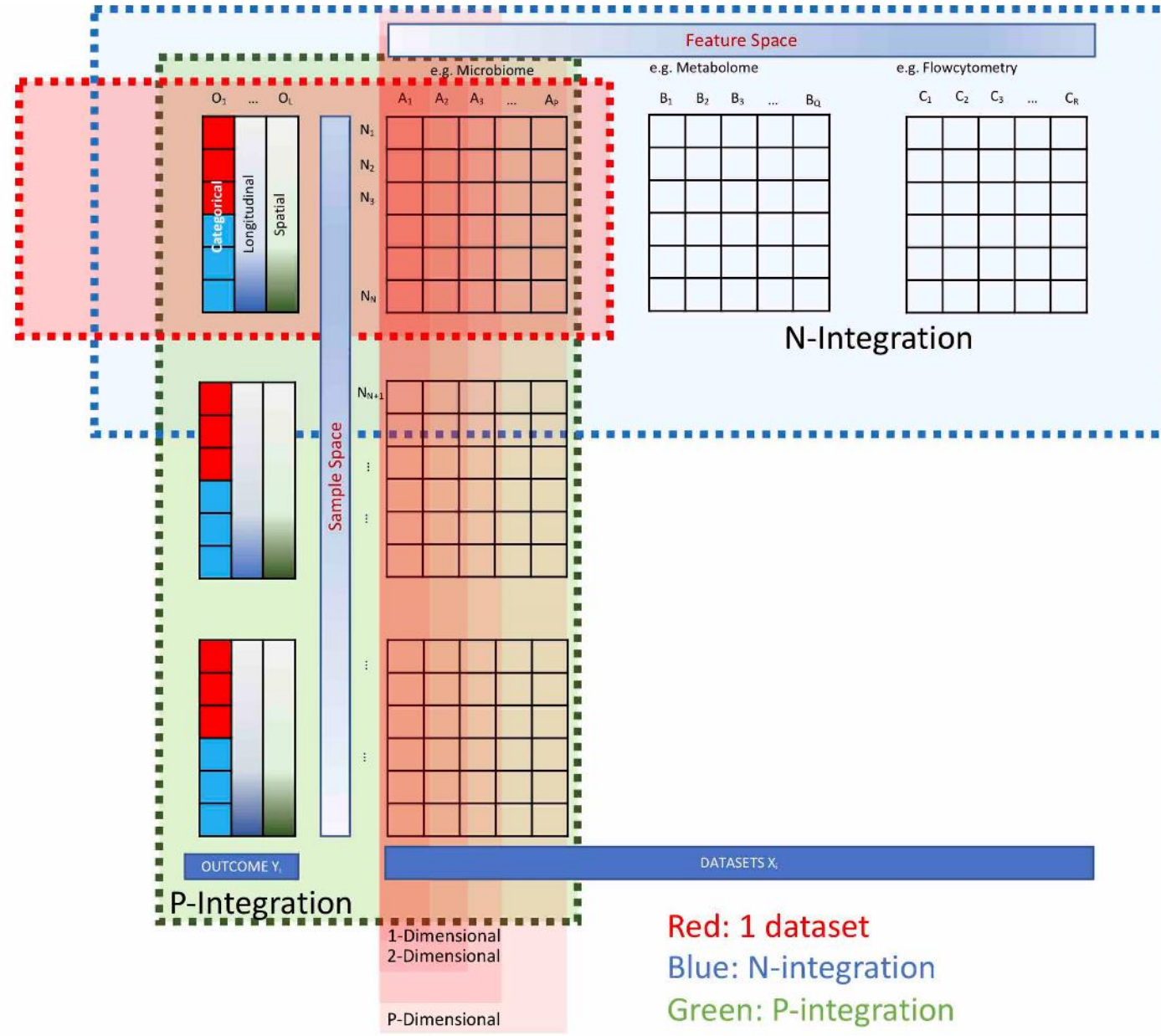


Efficient BOD conversion even at low temperatures and concentrations

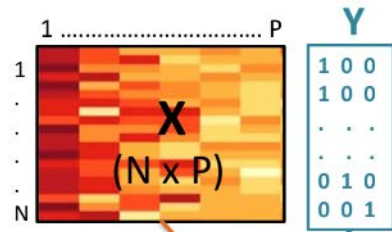


Process is supported by a dynamic and adaptable microbial community

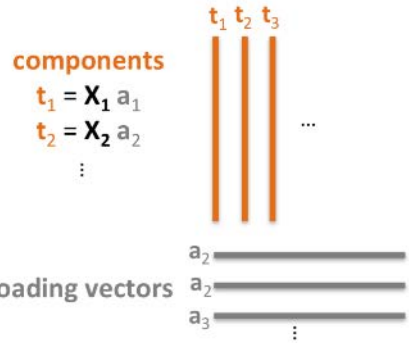
Integrated Omics



sPLS-DA



max $\text{cov}(t, u)$

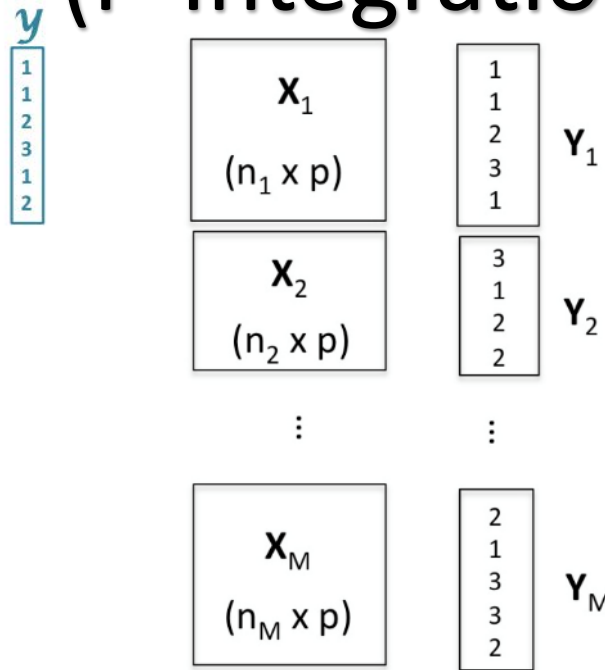


$$\arg \max_{(a_h, b_h)} \text{cov}(X_h a_h, Y_h b_h)$$

$$\text{s.t. } \|a_h\|_2 = \|b_h\|_2 = 1 \text{ and } \|a_h\|_1 \leq \lambda_h$$

Beauty of the method (enforces components of loading vector to go to zero)

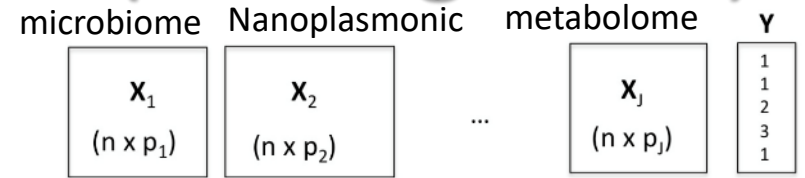
MINT (P-Integration)



$$\arg \max_{a_h, b_h} \sum_{m=1}^M \text{cov}(X_h^{(m)} a_h, Y_h^{(m)} b_h)$$

$$\text{s.t. } \|a_h\|_2 = 1 \text{ and } \|a_h\|_1 \leq \lambda$$

DIABLO (N-Integration)



$$C = \{c_{q,j}\}_{q,j} =$$

Offers tradeoff between Discrimination and correlation

| | X ₁ | X ₂ | X ₃ | Y |
|----------------|----------------|----------------|----------------|---|
| X ₁ | 0 | 0.1 | 0.1 | 1 |
| X ₂ | 0.1 | 0 | 0.1 | 1 |
| X ₃ | 0.1 | 0.1 | 0 | 1 |
| Y | 1 | 1 | 1 | 0 |

$$\arg \max_{a_h^{(1)}, \dots, a_h^{(Q)}} \sum_{q,j=1, q \neq j}^Q c_{q,j} \text{cov}(X_h^{(q)} a_h^{(q)}, X_h^{(j)} a_h^{(j)})$$

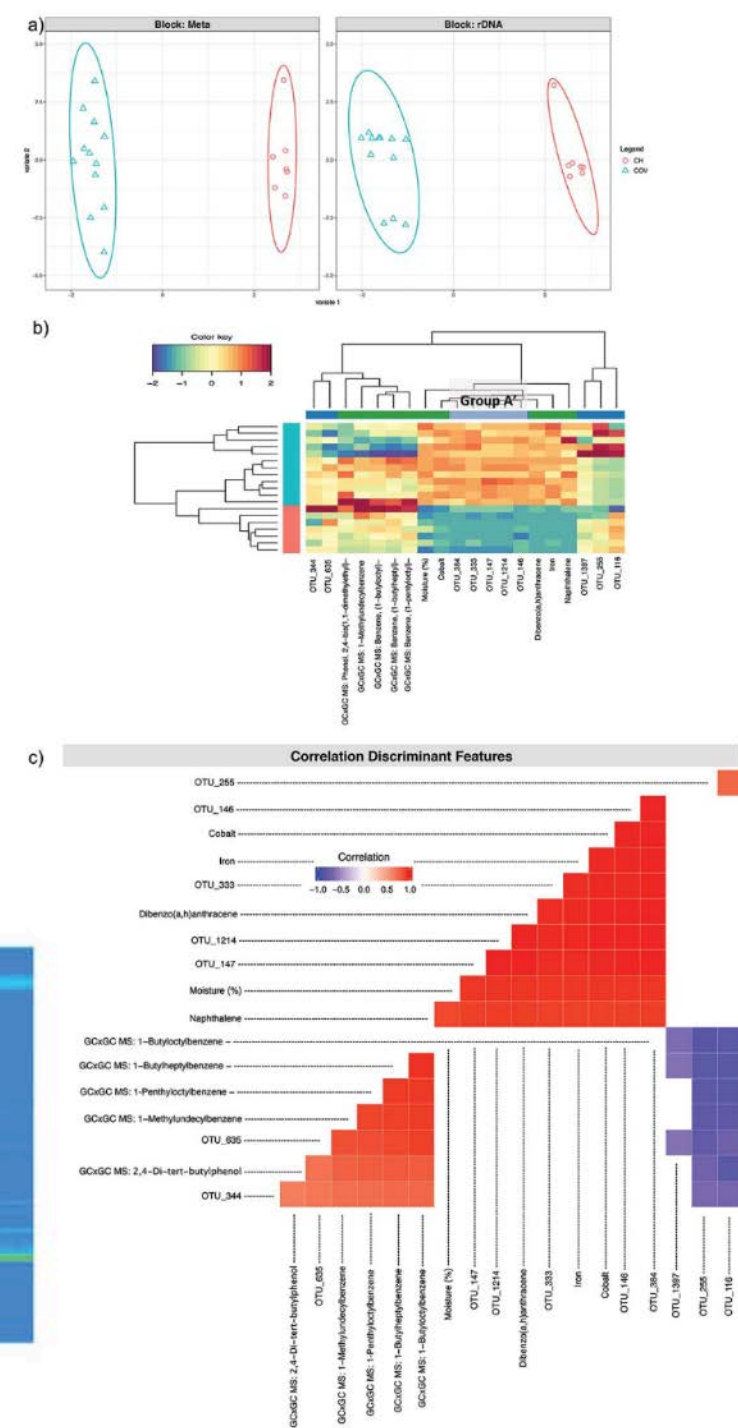
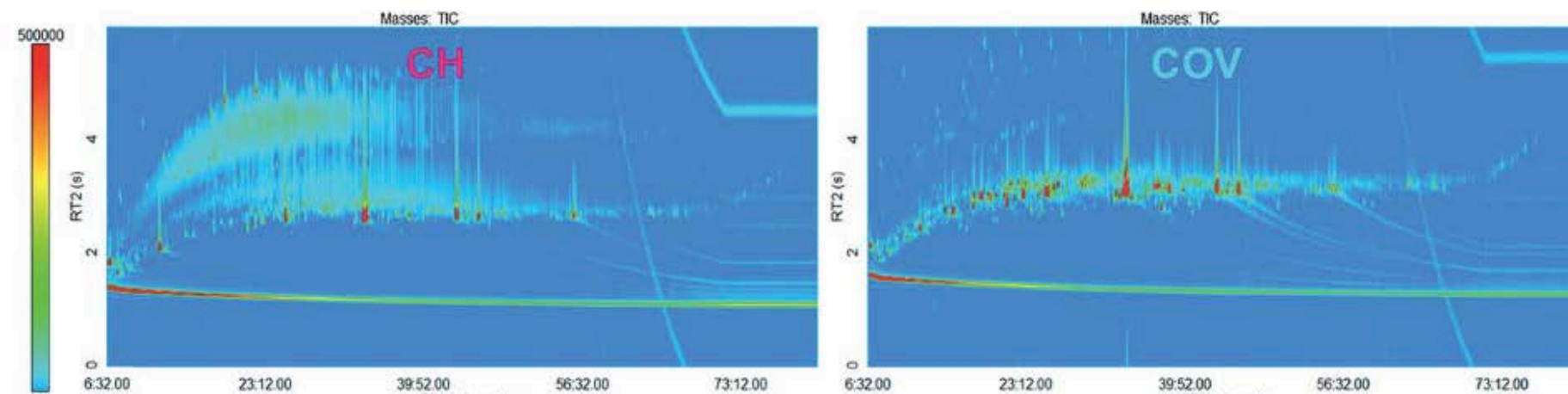
$$\text{s.t. } \|a_h^{(q)}\|_2 = 1 \text{ and } \|a_h^{(q)}\|_1 \leq \lambda^{(q)}$$

PAPER

[View Article Online](#)
[View Journal](#) | [View Issue](#)

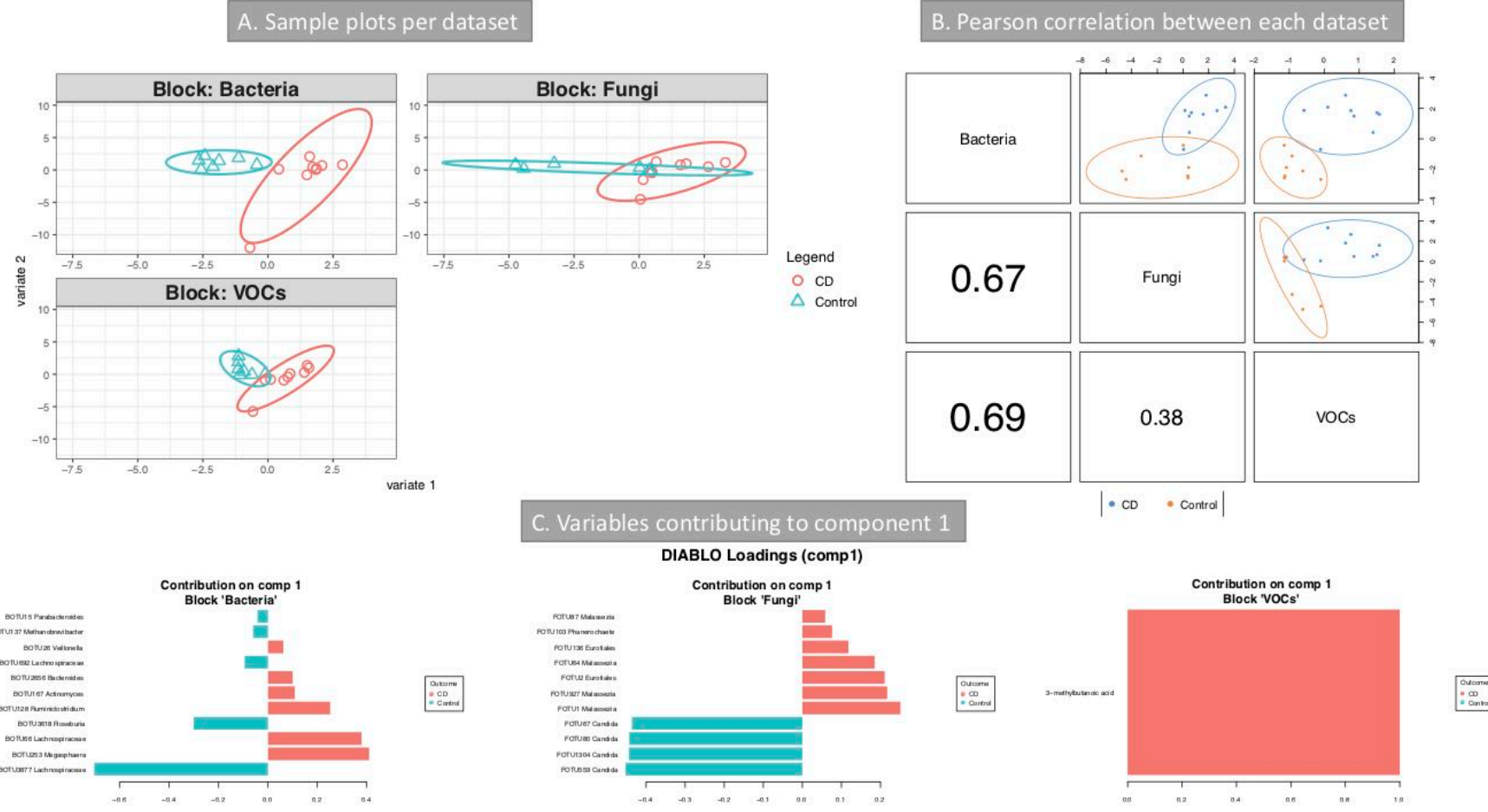
Systems biology approach to elucidation of contaminant biodegradation in complex samples – integration of high-resolution analytical and molecular tools†

Caroline Gauchotte-Lindsay,^{ID}*^a Thomas J. Aspray,[‡]^b Mara Knapp^c and Umer Z. Ijaz^{ID}^a



Inter-kingdom relationships in Crohn's disease explored using a multi-omics approach

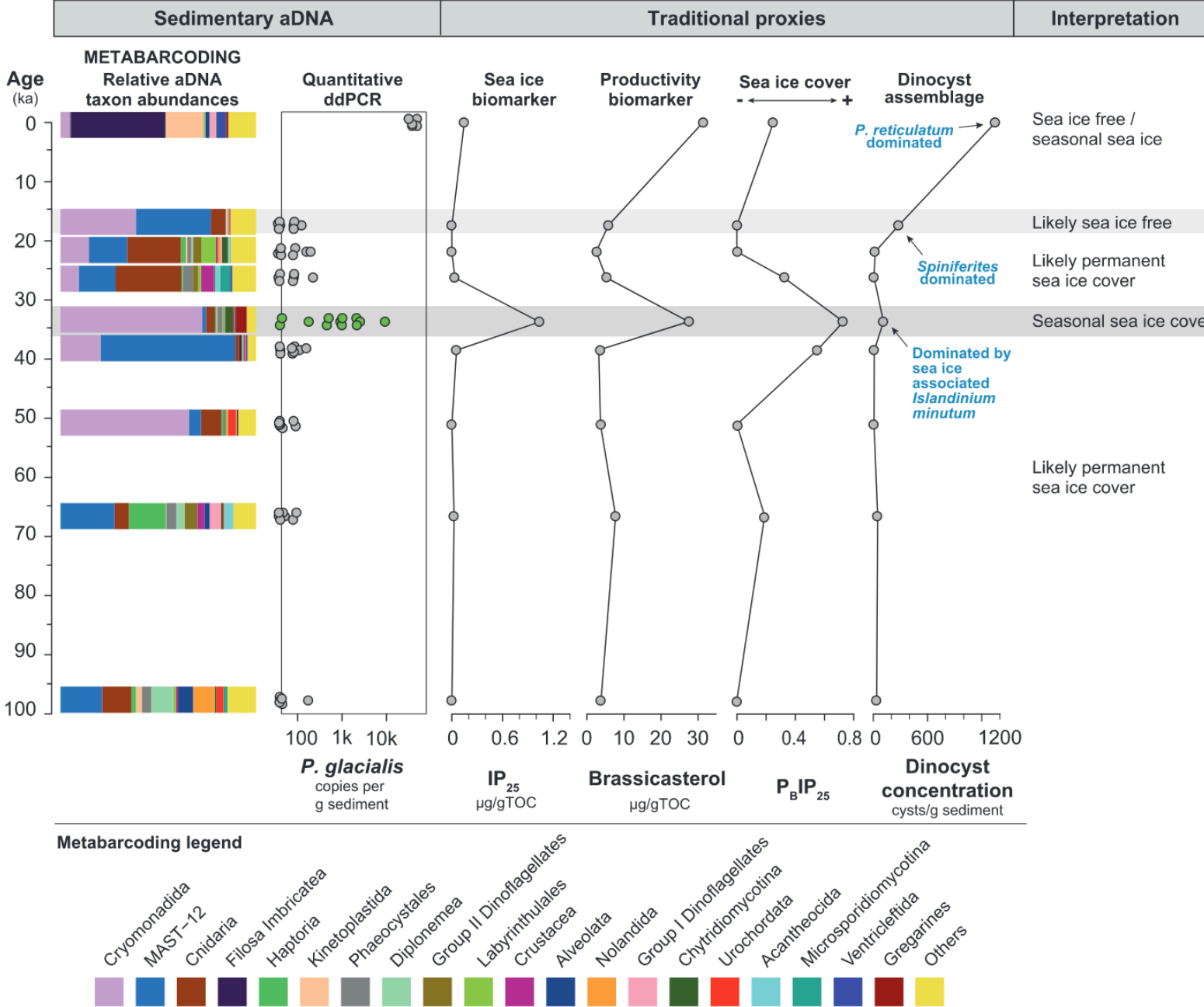
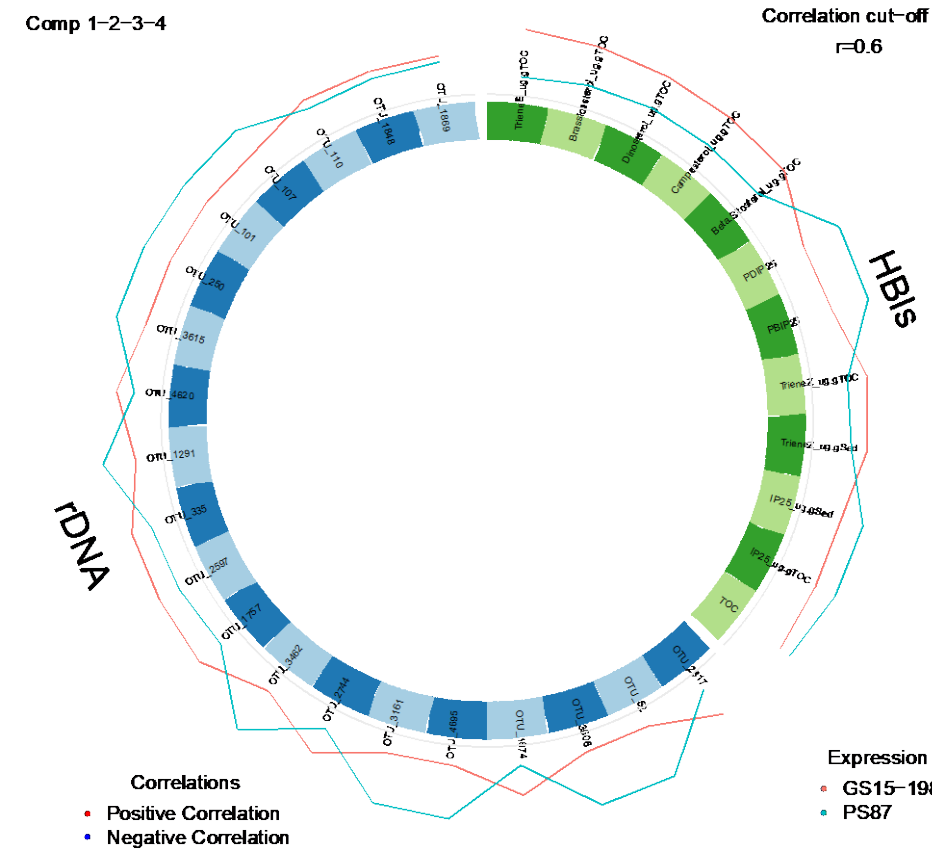
Alessandra Frau^a, Umer Z. Ijaz^b, Rachael Slater^a, Daisy Jonkers^c, John Penders^c, Barry J. Campbell^d, John G. Kenny^e, Neil Hall^{f,g}, Luca Lenzi^h, Michael D. Burkittⁱ, Marieke Pierik^e, Alistair C. Darby^d, and Christopher S. J. Probert^a



Integration of metabolomics (volatile organic compounds, VOCs) and metagenomics (bacterial and fungi) data (Crohn's disease (CD) vs Controls). 16S and 18S rRNA datasets were from transverse colon biopsy samples, and VOCs from stool samples. Samples groups were CD (n=9) vs Controls (n=7). **A** three omics sample plots are shown separately. **B** Pearson's correlation of the three data set for component 1. **C** variables contributing to component 1.

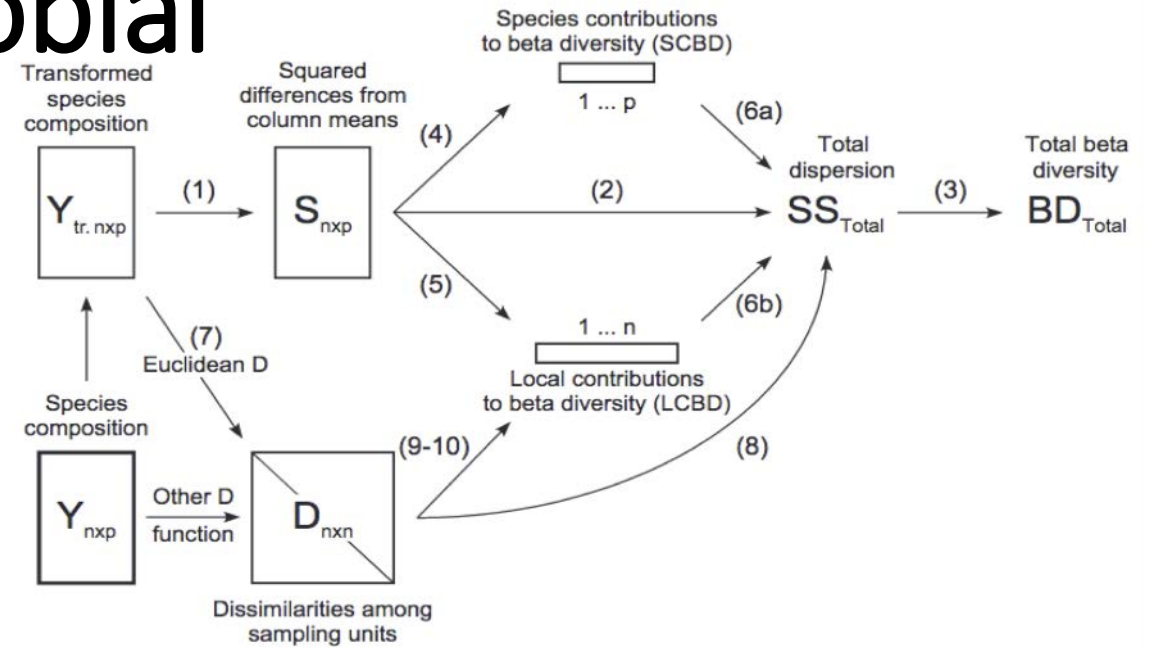
The potential of sedimentary ancient DNA for reconstructing past sea ice evolution

Stijn De Schepper¹ · Jessica L. Ray² · Katrine Sandnes Skaar² · Henrik Sadatzki^{3,7} · Umer Z. Ijaz⁴ · Ruediger Stein^{5,6} · Aud Larsen²

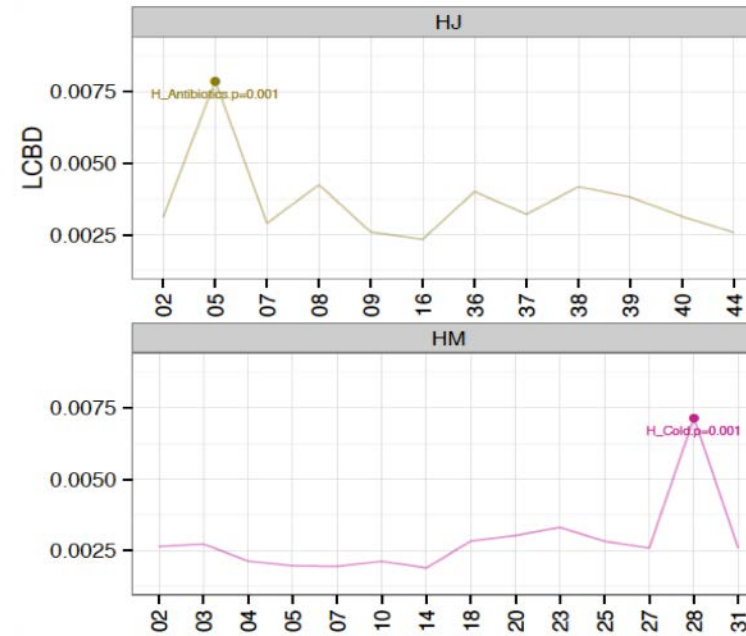


Stability Analysis in Microbial Communities

- Particularly important in view of Waste Water treatment systems
- Can we come up with simple proxies to judge stability in both **abundance** and **function** of microbial communities
- Local Contribution to Betadiversity (LCBD) as a measure of '**dysbiosis**'

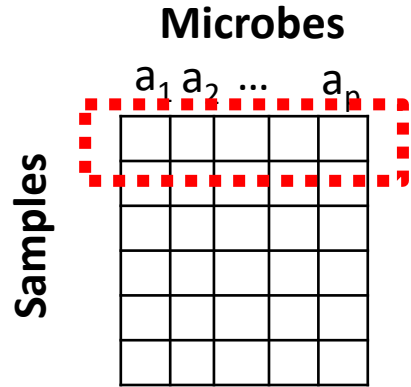


Temporal Profile →



With Asha Ram (PhD student)
Pharyngeal Microbiome

Functional Robustness/Stability Parameters: attenuation/buffering



Given an original community and perturbed community with N unique pathways with average copy number $a_j \forall j \in 1, 2, \dots, N$ and $b_j \forall j \in 1, 2, \dots, N$, respectively, the **cosine dissimilarity** between two functional profiles is

$$1 - \frac{\sum_j^N a_j b_j}{\sqrt{\sum_j^N a_j^2} \sqrt{\sum_j^N b_j^2}}$$

Create perturbations

$$p_i = a_i \times m_i^{d_i}$$

Per sample basis, create say
100 perturbations

Weighted Unifrac Distance

Fit a function $f = \frac{1}{e^a} t^b$ where a is **attenuation coefficient**, and b is the **buffering coefficient**, t denotes the **magnitude of the taxonomic perturbation**.

$m_i \forall i \in 1, 2, \dots, N$, are sampled from a uniform distribution over the interval $(0, M]$, M being the **maximum perturbation magnitude**

perturbation directions $d_i \in \{-1, 1\} \forall i \in 1, 2, \dots, N$, with an equal chance of either directions

Attenuation a describes the expected rate at which increases in the taxonomic perturbation magnitude are expected to increase functional shifts

Buffering b indicates how large a perturbation must be before a functional profile shift becomes noticeable and approaches the expect shift magnitude defined by attenuation

RESEARCH

Open Access

Response and oil degradation activities of a northeast Atlantic bacterial community to biogenic and synthetic surfactants

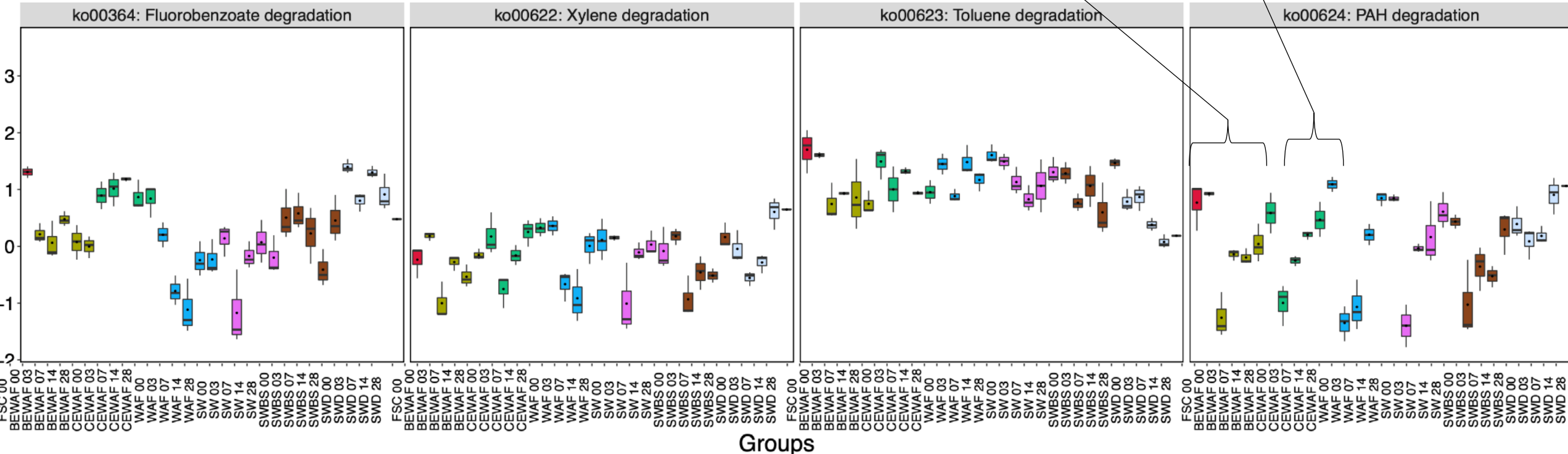
Christina N. Nikolova¹, Umer Zeeshan Ijaz^{2*}, Clayton Magill³, Sara Kleindienst⁴, Samantha B. Joye⁵ and Tony Gutierrez^{1*}



Biogenic Surfactant Treatment

Synthetic Surfactant Treatment

Attenuation Values



Microbiome Community Assembly Mechanisms

- Finding **ecological phenomena** underpinning microbial community assembly
- Main question: Stochasticity vs Determinism or Neutral vs Niche
- Procedure: **Null modeling**
 - Record an observed original metric and then apply randomization(999) (abundance table/phylogenetic tree) whilst preserving a certain property (typically alpha diversity)
 - Deviation of observed original metric from the average of null models has the power to predict ecological processes
- e.g., NRI/NTI \leftarrow alpha diversity measure to differentiate between “**environmental filtering**” and “**competitive exclusion**”
- e.g., QPE framework \leftarrow beta diversity measure to differentiate between “**homogeneous selection**”, “**heterogeneous selection**”, “**drift limitation**”, and “**homogeneous drift**”

Exploration of marine bacterioplankton community assembly mechanisms during chemical dispersant and surfactant-assisted oil biodegradation

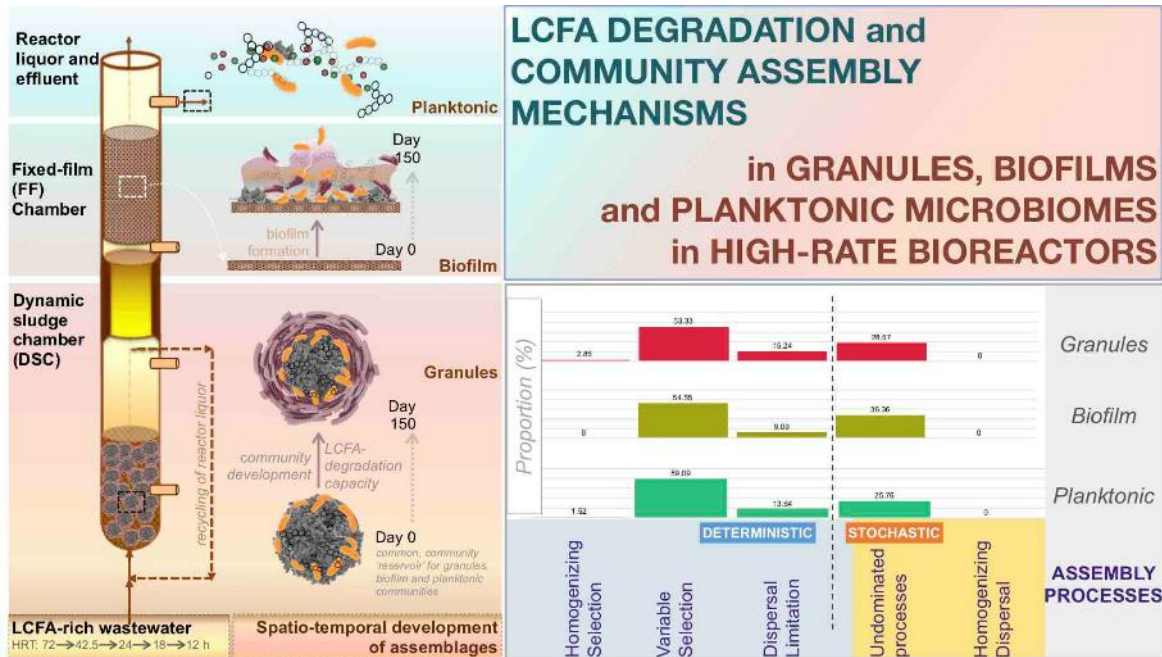
Christina Nikolova¹ | Umer Zeeshan Ijaz² | Tony Gutierrez¹

| Null model | Description | Input | Reference |
|--------------------------------------|--|--|---|
| Normalised stochasticity ratio (NST) | A method that allows any type of beta-diversity measure, whether it is incidence-based (presence-absence) or abundance-based and gives a percentage of stochasticity for samples belonging to a single category. | Any beta-diversity metric as long as it is normalised to have a range between 0 and 1. Abundance-based Ružička and incidence-based Jaccard metrics were deemed to have a superior performance in view of the simulations done by the author. | Ning et al. (2019) |
| Hill numbers | The effect of deterministic and stochastic factors on explaining the differences between multiple categories. | Jaccard and Bray-Curtis dissimilarity metrics. | Modin et al. (2020) |
| β-null deviation | Differentiation between niche and neutral community assemblage processes. | Abundance-based (Bray-Curtis) β-null deviation metric. Abundance-based phylogenetic (UniFrac) measure. | Tucker et al. (2016) Lee et al. (2017) |
| Quantitative Process Estimate (QPE) | Determination of the dominant assembly process between two given communities. | Abundance table and phylogenetic tree on which βNTI and βRC are calculated for all pairs of samples. | Stegen et al. (2013) |
| Lottery model | Determination of clade-based community assembly where a single taxonomic group solely occupies a given niche due to a competitive advantage over other species (i.e., lottery winner). | Proportionally normalised abundance matrix. | Verster & Borenstein (2018) |
| Phylogenetic dispersion | Characterisation of the species distribution across treatments in relation to priority effects. | Abundance table and phylogeny for samples that have temporal behaviour which is captured as probability of recruitment. This is calculated as difference in phylogenetic diversity fitted with a logistic regression. | Darcy et al. (2020) |



Microbial community assembly and dynamics in Granular, Fixed-Biofilm and planktonic microbiomes valorizing Long-Chain fatty acids at 20 °C

Suniti Singh^{a,b,c}, Johanna M. Rinta-Kanto^a, Piet N.L. Lens^{a,b}, Marika Kokko^a, Jukka Rintala^a, Vincent O'Flaherty^c, Umer Zeeshan Ijaz^{d,*}, Gavin Collins^{c,d}

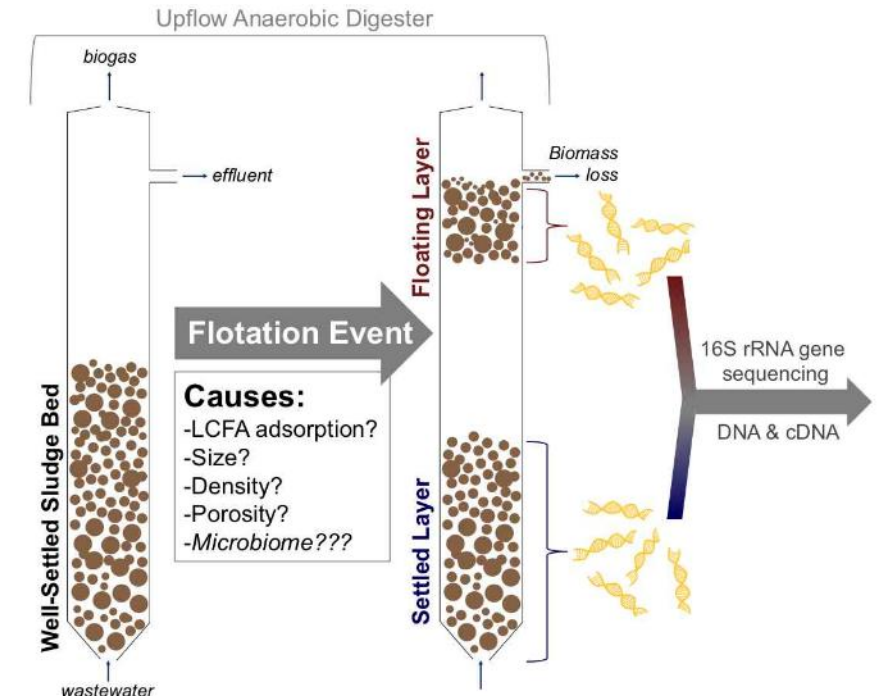


Combined Stochastic and Deterministic Processes Drive Community Assembly of Anaerobic Microbiomes During Granule Flotation

Anna Christine Trego^{1†}, Paul G. McAteer^{1,2†}, Corine Nzetue¹, Therese Mahony¹, Florence Abram², Umer Zeeshan Ijaz^{3*} and Vincent O'Flaherty^{1*}



a)



b)

Schematic illustrating: a) Photograph of the reactor set-up with floating and settled sludge visible, and b) Research methodology.

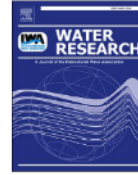
Meta-analysis (Amplicons)

Water Research 212 (2022) 118106

Contents lists available at ScienceDirect

Water Research

journal homepage: www.elsevier.com/locate/watres



Microbiomes in drinking water treatment and distribution: A meta-analysis from source to tap

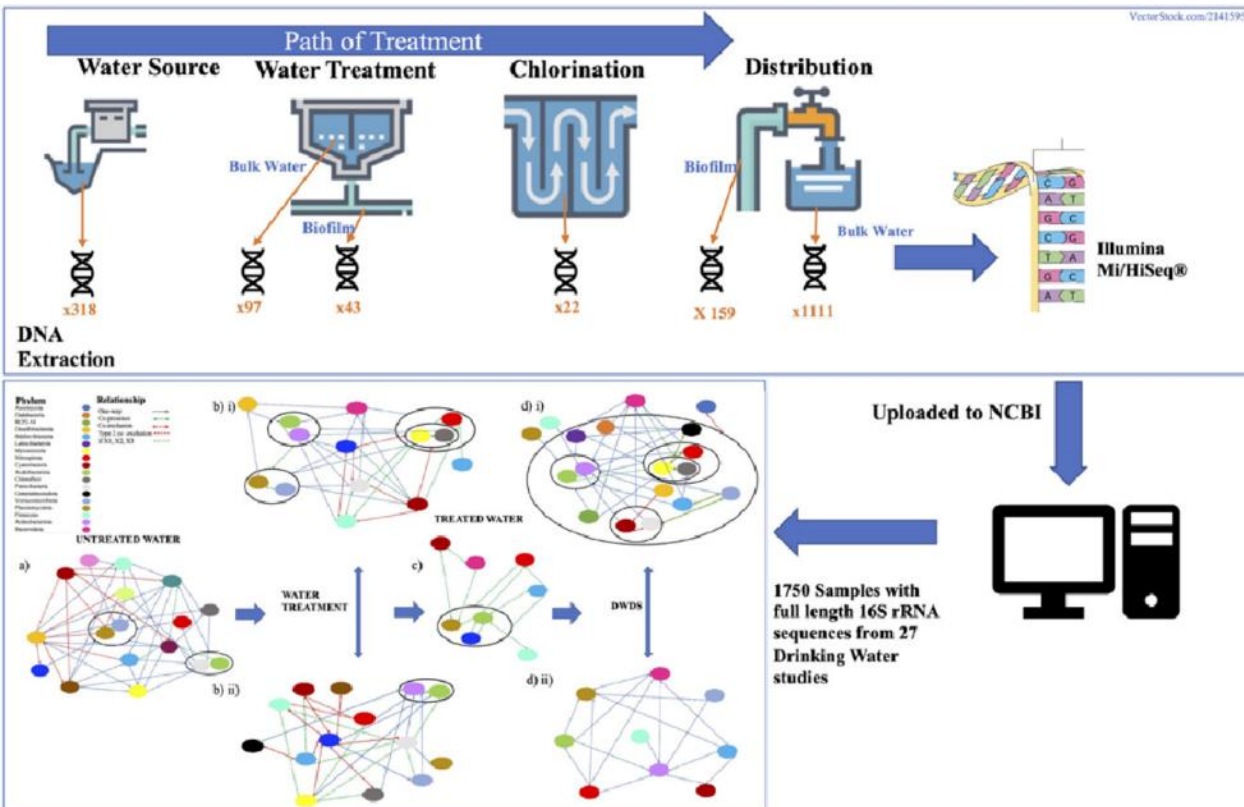
Claire Thom^{a,b,*}, Cindy J Smith^a, Graeme Moore^b, Paul Weir^b, Umer Z Ijaz^a

^a School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

^b School of Civil and Environmental Engineering, Nanyang Technological University, Singapore

* Corresponding author. E-mail: claire.thom@ntu.edu.sg

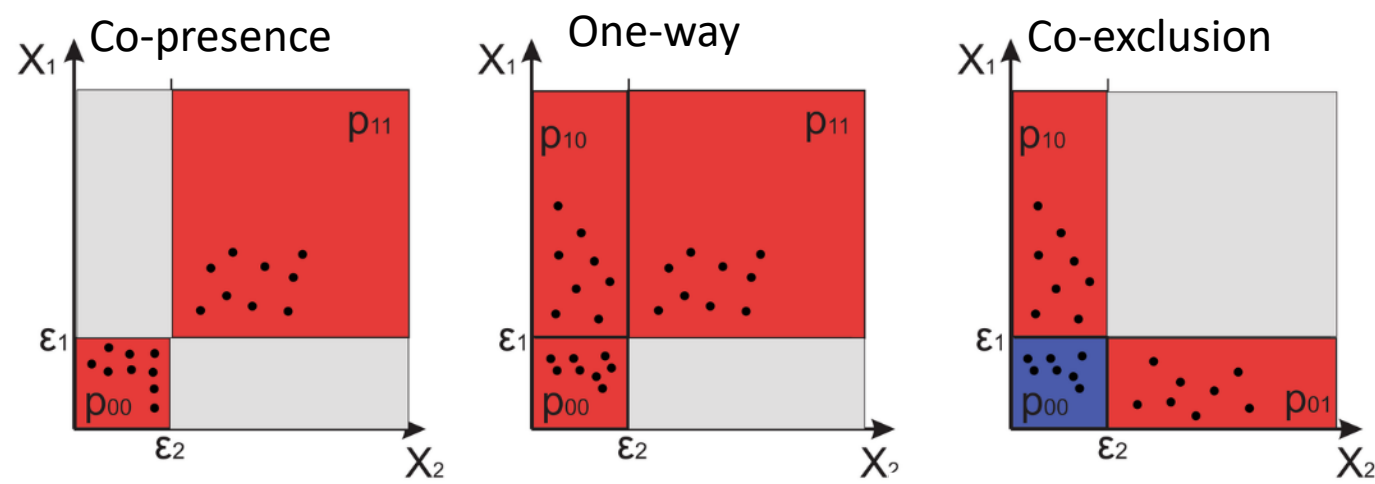
1



- Collation Strategy (How to combine several V-regions together
 - Read loss (86% of abundance lies in 3% of the ASVs)
- Taxonomic Assignment Strategy
 - Naïve Bayesian Classifier (QIIME2)
 - Bayesian Common Ancestor Algorithm
- Phylogenetic tree generation

Network Inference

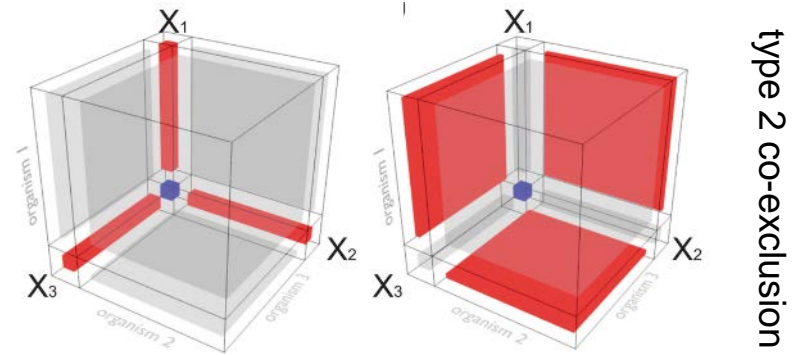
- Constructing networks of relationships
 - Causality vs Association
 - Static vs Dynamic
- Subcommunity detection (Modularity)
- Network-wide statistical measures
 - Keystone nodes/modules
- Co-occurrence networks (Correlation; microbiomeSeq)
- Sparse-inverse covariance estimation (spiecEasi)
- **Boolean Networks**



type 1 co-exclusion

$$S_{3D \text{ co-exclusion type 1}} = \max_{\epsilon_1, \epsilon_2, \epsilon_3} (p_{000} + p_{100} + p_{010} + p_{001}),$$

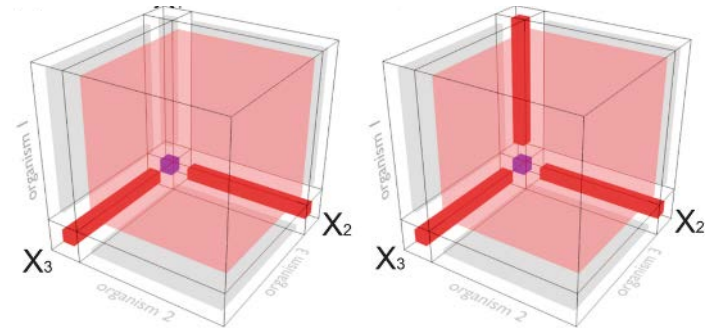
where $p_{100} > m$; $p_{010} > m$; $p_{001} > m$;



$$S_{3D \text{ co-exclusion type 2}} = \max_{\epsilon_1, \epsilon_2, \epsilon_3} (p_{000} + p_{110} + p_{101} + p_{011}),$$

where $p_{110} > m$; $p_{101} > m$; $p_{011} > m$.

pattern when the presence of organism X_1 changed patterns between X_2 and X_3 from co-presence to co-exclusion



case where three organisms can be present only all together on one-by-one

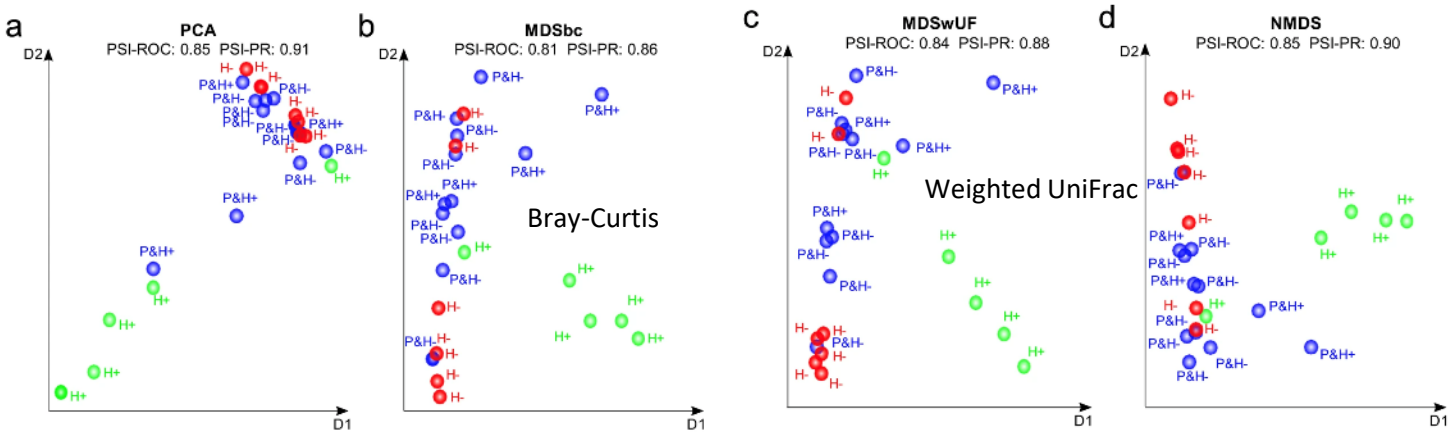
Golovko, G et al. *Microbiome* **8**, 131 (2020).
<https://doi.org/10.1186/s40168-020-00853-6>

Nonlinear machine learning pattern recognition and bacteria-metabolite multilayer network analysis of perturbed gastric microbiome

Claudio Durán, Sara Ciucci, Alessandra Palladini, Umer Z. Ijaz, Antonio G. Zippo, Francesco Paroni Sterbini, Luca Masucci, Giovanni Cammarota, Gianluca Ianiro, Pirjo Spuul, Michael Schroeder, Stephan W. Grill, Bryony N. Parsons, D. Mark Pritchard, Brunella Posteraro, Maurizio Sanguinetti, Giovanni Gasbarrini, Antonio Gasbarrini & Carlo Vittorio Cannistraci

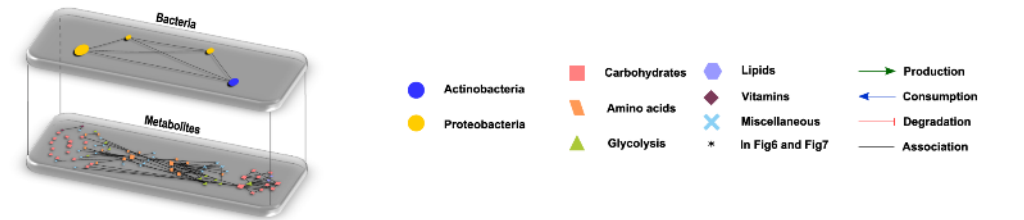
Nature Communications 12, Article number: 1926 (2021) | Cite this article

Traditional approaches to capture beta diversity

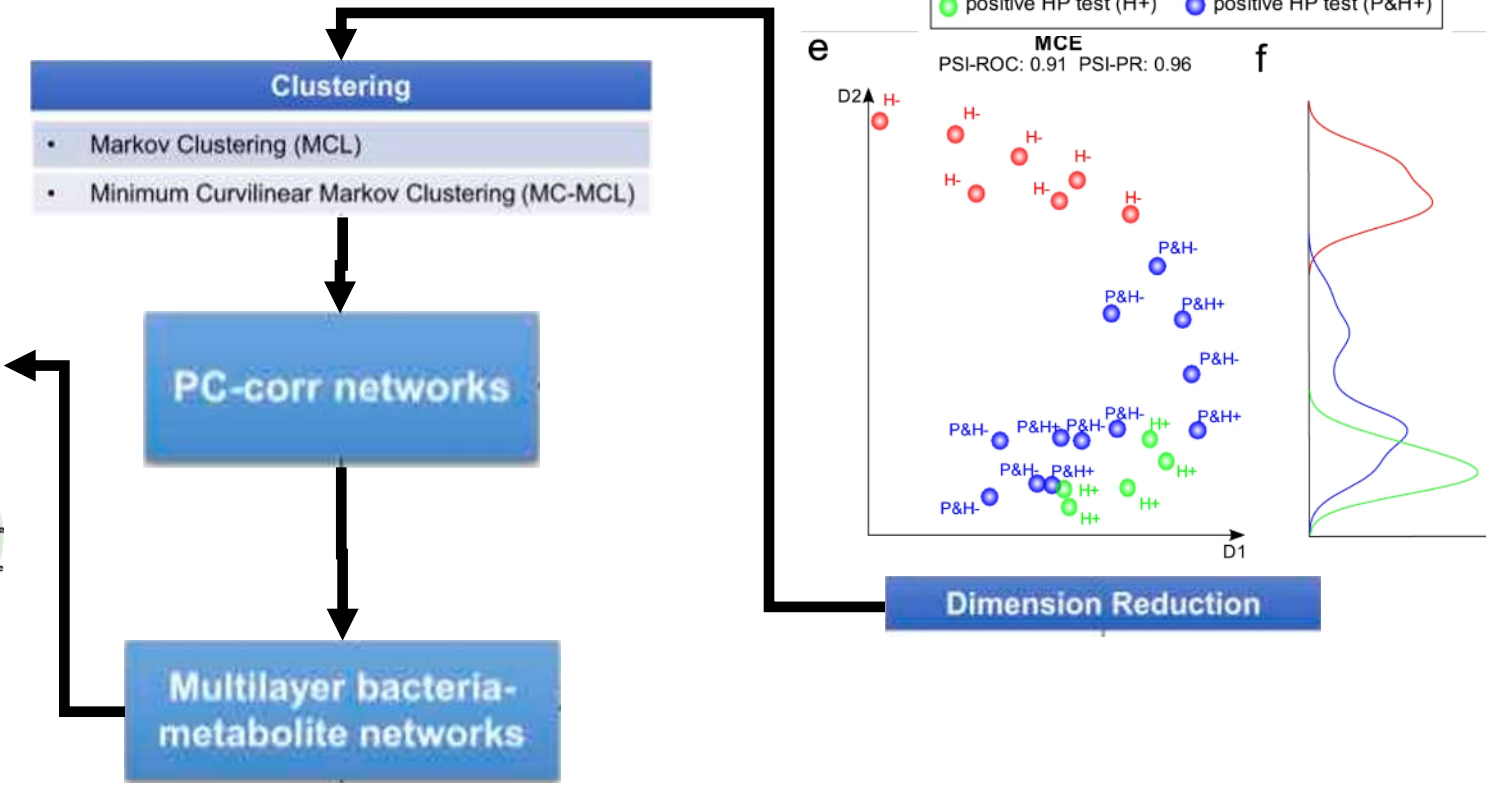
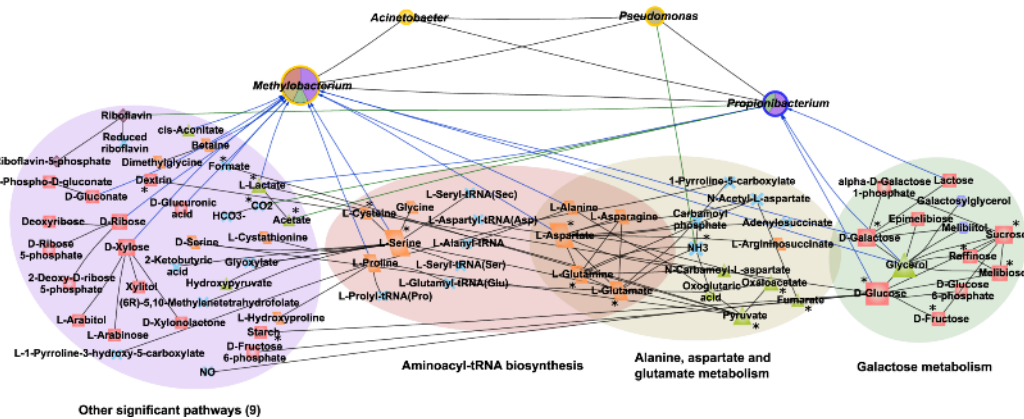


Our Proposal: Minimum Curvilinearity

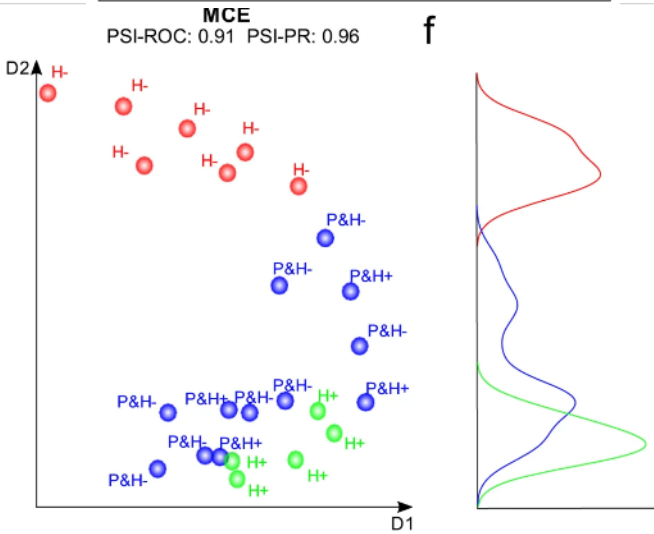
a



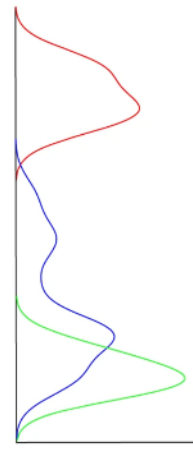
b



e

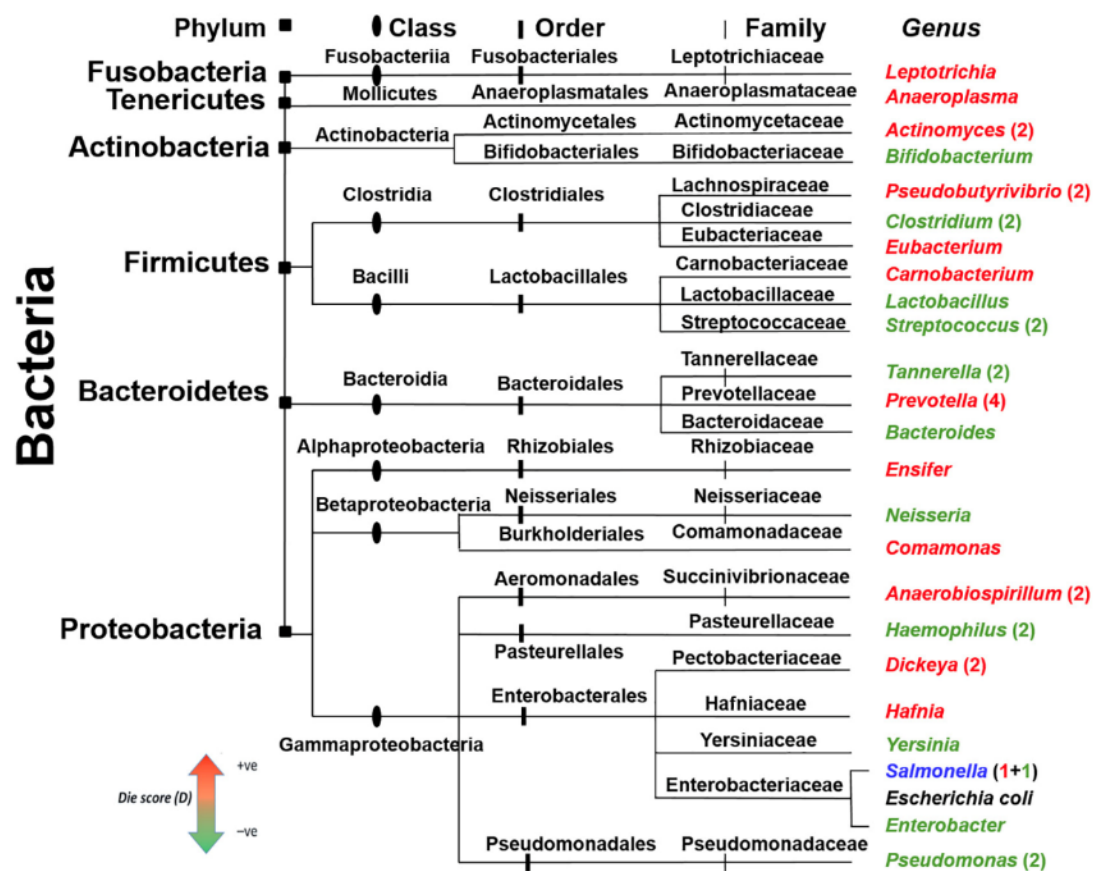


f



Putative Antimicrobial Peptides Within Bacterial Proteomes Affect Bacterial Predominance: A Network Analysis Perspective

Anastasis Oulas^{1,2*}, Margarita Zachariou^{1,2}, Christos T. Chasapis³, Marios Tomazou^{1,2}, Umer Z. Ijaz⁴, Georges Pierre Schmartz⁵, George M. Spyrou^{1,2} and Alexios Vlamis-Gardikas^{6*}



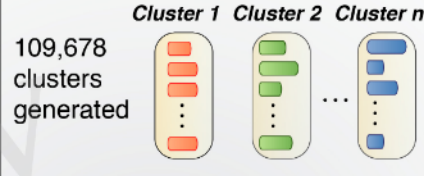
1. Download bacterial & human proteomes from Uniprot



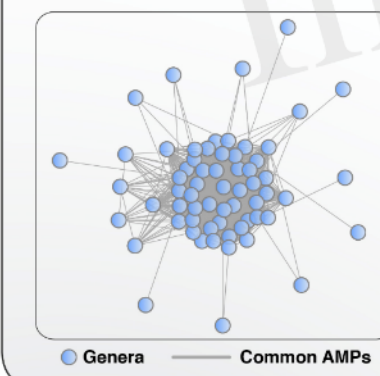
2. Predict AMPs using AMPA software



3. Cluster the predicted AMPs using CdHit (> 80 % sequence similarity)



6. Visualize BW networks for selected bacterial genera



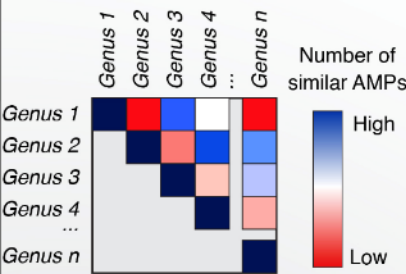
5. Bacterial Wars (BW) database constructed from the edgelist generated from the adjacency matrix

365,294 edges at strain level

Edge weight based on the number of shared AMPs



4. Calculate the adjacency matrix of the number of similar AMPs (same cluster) among all genus pairs



For a node i in a given graph BW , we calculated the *Die* score (D_i) with equation (1):

$$\text{Die Score } D_i = \frac{N_i - S_i}{N_p} \quad (1)$$

Where: S_i represents the strength of the node (i.e., the weighted degree of the node i) and sums up the number of neighboring bacteria that node i is linked to via an edge (e.g., shares common peptides), as well as the edge weight (e.g., number of common peptides). N_i represents the number of pairs with no edges (e.g., no common peptides). It is calculated using the expression $N_i = N_p - \text{deg}_i$, where deg_i represents the degree of the node i . $N_p = (n^2 - n)/2$ represents the number of all possible pairs in the network.

Conclusions

- Understanding microbial communities is somewhat convoluted and hard
- Several challenges (computational/technical) to capture the diversity accurately
- Applying methods from Machine Learning and Numerical Ecology literature to home down to patterns of interests that may have biological relevance.
- Building devices that harness and exploit microbes, and make it easier to understand such complex systems.
- Wider applications in Water/Waste Water Treatment, Agriculture, Aquaculture, Environment and Climate Change, and Health and Diet