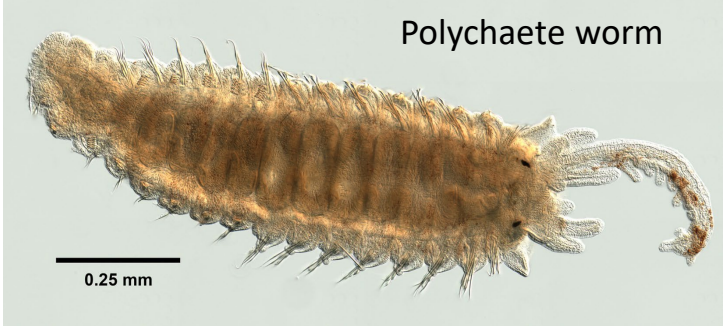
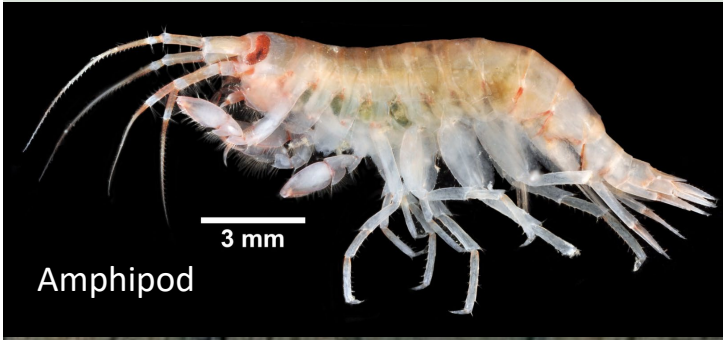
Evaluation of infauna community structure through microscopy and eDNA

Joanna Strzelecki, Sarah Stephenson, Jodie van de Kamp, Lev Bodrossy, Mick Haywoood, John Keesing, Lydianne Matio, Damian Thomson and Melanie Trapon| September 2019

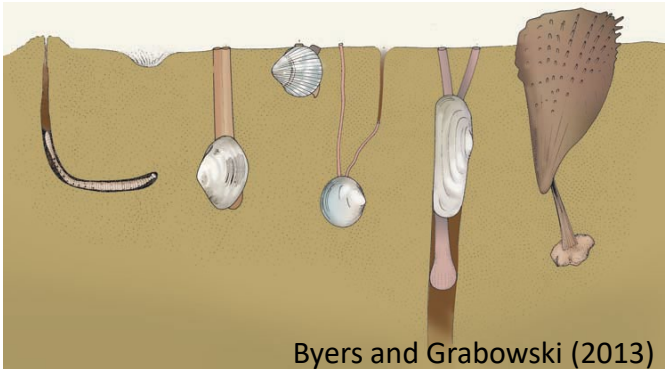




Importance of Infauna

- burrow into bottom sediment
- connected to the water by tubes and tunnels

- Numerous, widely distributed, diverse
- An important link in food web
  - link primary producers with higher trophic levels
  - food for economically important species: crabs, shrimps, fish
- Major role in recycling organic matter
  - ingestion
  - bioturbation
  - bio-irrigation
- Indicator of environmental condition (do not move very far so they cannot avoid pollution)
  - nutrient pollution
  - chemical contaminants
  - ocean acidification



Byers and Grabowski (2013)

Aim

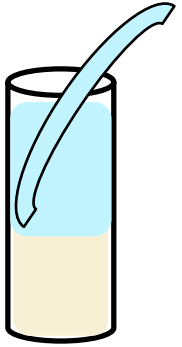
- describe and compare infauna using traditional biodiversity assessments and eDNA

Methods field

Infauna is patchy on a small spatial scale – need to homogenise sample before subsampling for microscopy and DNA analyses (or preserve for morphology &DNA Creer et al 2010)

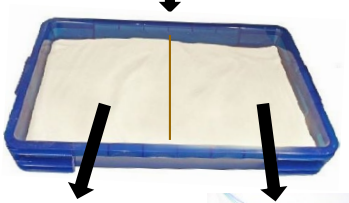
10 cm

10 cm



Water siphoned out

Sediment emptied to plastic bag and homogenised by gentle shaking



tipped into a tray and split in ½

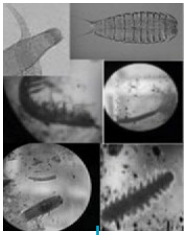
DNA (-20°C)

Microscopy (formalin)

Methods microscopy

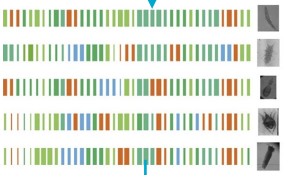
- Samples sieved through 4, 2.8, 2, 1.4, 1 and 0.5 mm sieves to remove preservative and fine sediment
- Sieved samples sorted into basic groups
- Specimens identified to the lowest taxonomic unit and counted

Methods eDNA



- samples were thawed
- shells or rocks removed
- sediment homogenized in a bench-top blender on the highest speed for 2 minutes

DNA extracted in triplicate from a 0.5 g subsample of homogenised sediment using the QIAGEN DNeasy PowerSoil® Kit

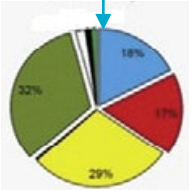


18SV9 rDNA 18SV4 rDNA
mt16S and COI genes

Next generation sequencing of all target genes using the Illumina MiSeq platform

Next generation sequencing

custom pipeline Greenfield Hybrid Amplicon Pipeline v2.1 (GHAPv2.1) (uses USEARCH sequence analysis tools (Edgar, 2013))

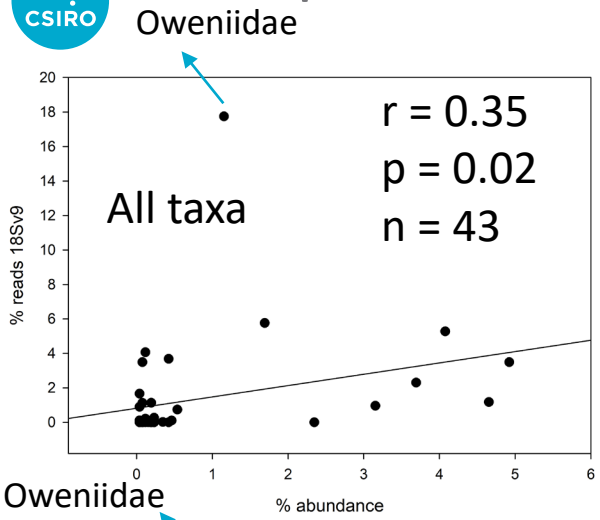


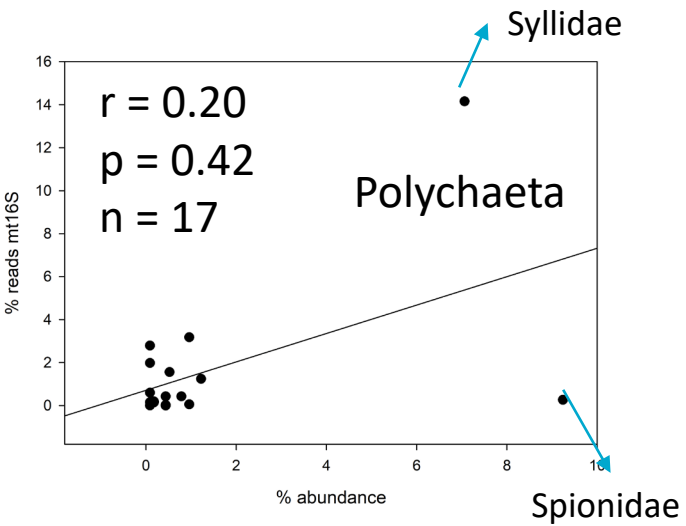
Taxonomic assignment

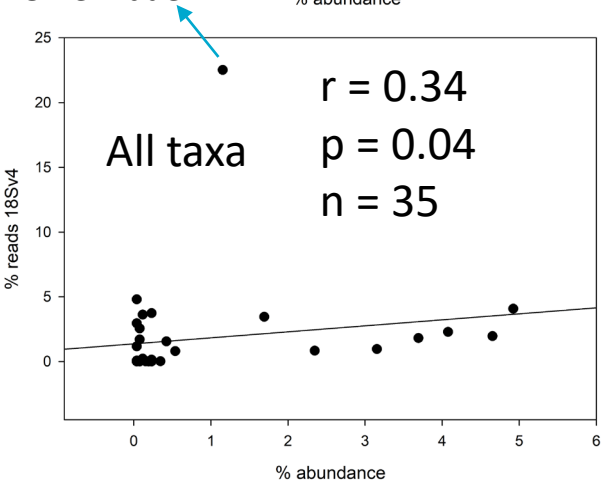
Number of OTUs

taxa	Microscopy	18Sv9 rDNA	18Sv4 rDNA	mt16S (polychaetes)
Total OTUs	123	2043	2042	531
Polychaeta	36	132	59	230
Nematoda	1	367	239	0
Gastropoda	4	28	9	23
Bivalvia	7	36	19	83
Ostracoda	8	38	27	0
Copepoda	1	134	70	0
Echinodermata	4	7	5	2
Decapoda	6	9	6	29
Amphipoda	26	0	0	1
Isopoda	5	0	0	1

Comparison of relative abundances







- Spearman rank correlation coefficient
- Correlations limited to taxa uncovered by both methods

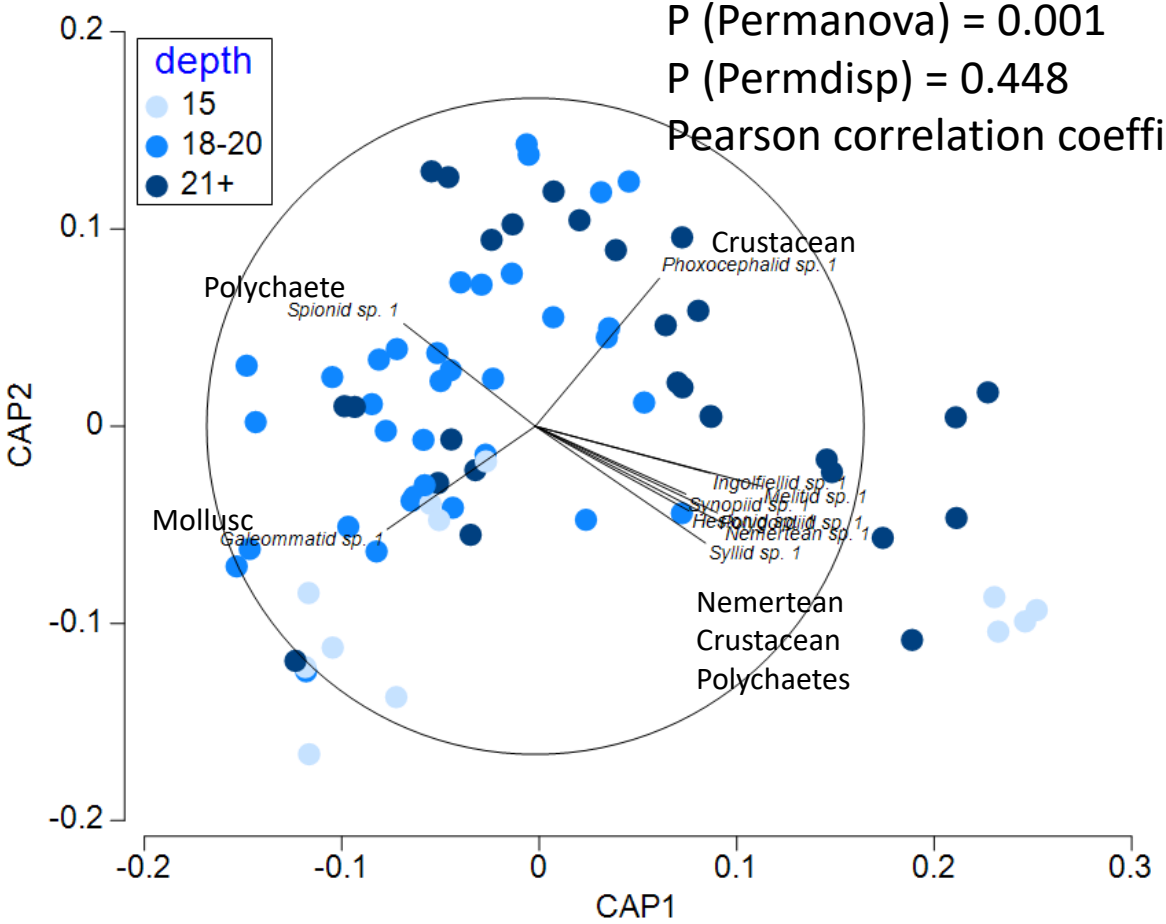
Microscopy: Environmental variables structuring infauna

Depth

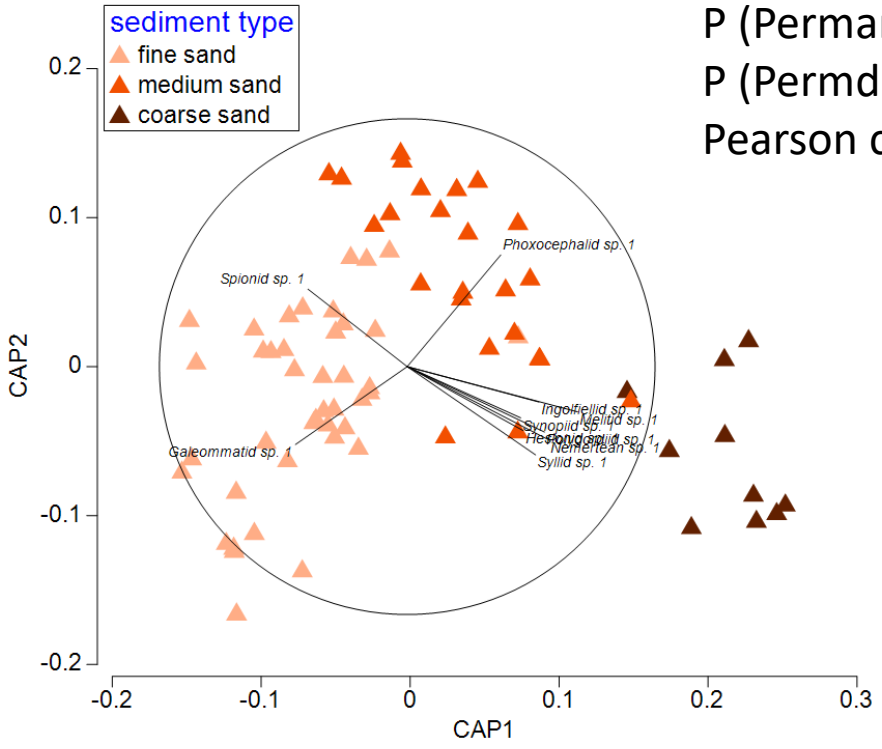
P (Permanova) = 0.001

P (Permdisp) = 0.448

Pearson correlation coefficient >0.5



Microscopy: environmental variables structuring infauna



Sediment type

P (Permanova) = 0.001

P (Permdisp) = 0.379

Pearson correlation coefficient >0.5

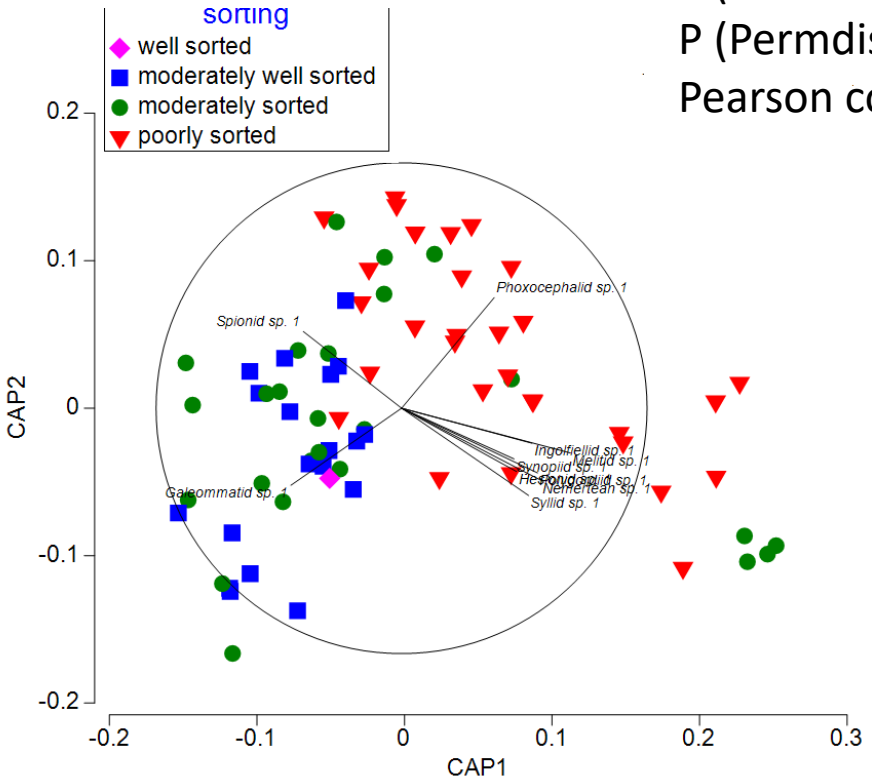
Microscopy: environmental variables structuring infauna

Sediment sorting (uniformity of grain size)

P (Permanova) = 0.001

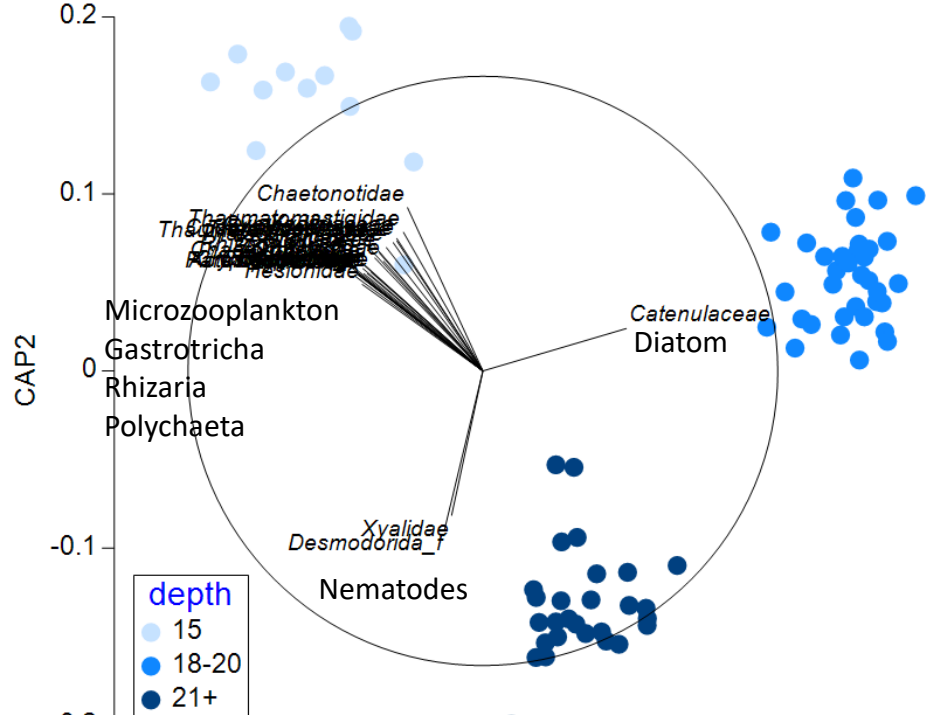
P (Permdisp) = 0.4

Pearson correlation coefficient >0.5



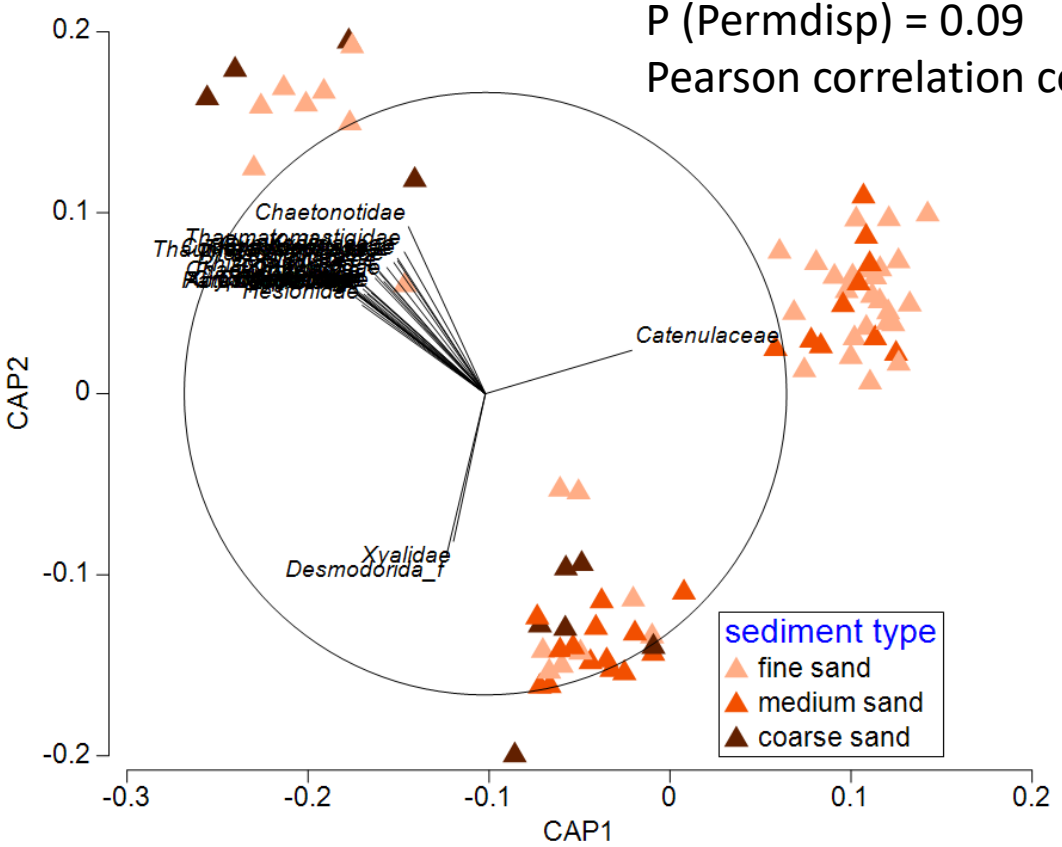
18Sv9 rDNA: environmental variables structuring infauna

Depth
P (Permanova) = 0.001
P (Permdisp) = 0.969
Pearson correlation coefficient >0.5



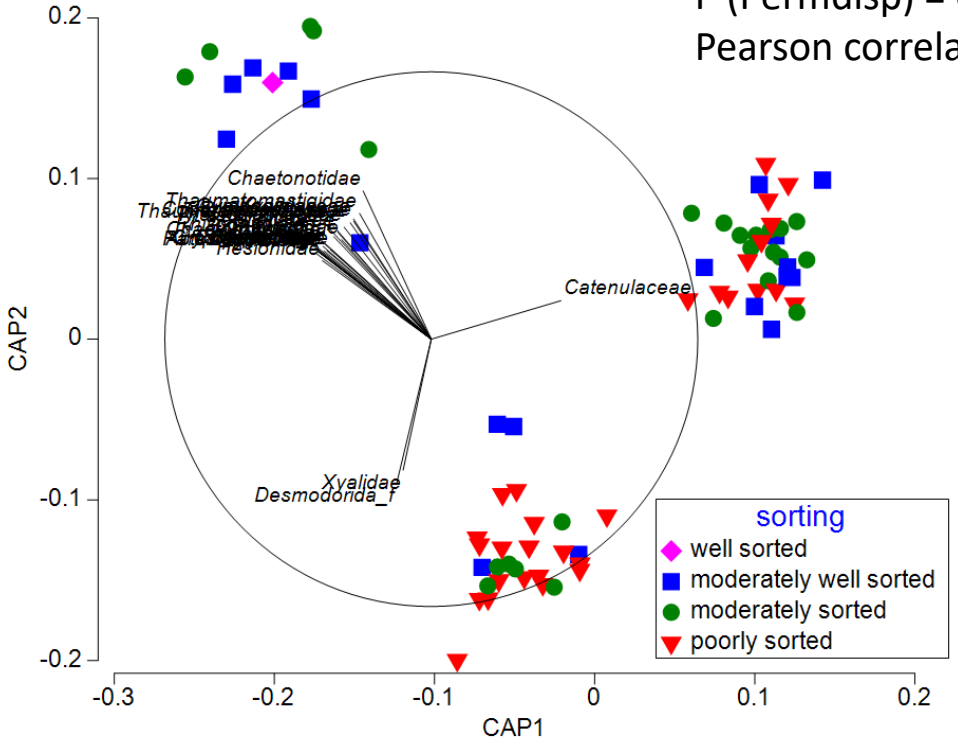
18Sv9 rDNA: environmental variables structuring infauna

Sediment type
P (Permanova) = 0.001
P (Permdisp) = 0.09
Pearson correlation coefficient >0.5



18Sv9 rDNA: environmental variables structuring infauna

Sediment sorting
P (Permanova) = 0.02
P (Permdisp) = 0.084
Pearson correlation coefficient >0.5



Amphipods: false negative

- Low abundance - no
- Incomplete extraction, removal of inhibitors – will try DNeasy Blood & Tissue Kit (QIAGEN)
- Primer mismatch
- Secondary DNA structures
- Too stringent filtering, OTU clustering – no

- Affects
  - relative abundances
  - Diversity α β
  - Relationship with environmental factors

Advantages and challenges

	Microscopy	DNA
Well understood	✓ More reported standard protocols, more comparability, reproducibility	Biases: amplification, sequencing errors, markers, organisms on the surface, in the guts of target taxa, persistence in sediment
Time	Slow (small samples no)	Fast (large no of samples)
Historical data/time series/museum collections	✓	gene sample repositories needed
Soft bodies organisms	Not well preserved	✓
Cryptic species	Morphological similarity	Genetically distinct
Size	Size biased	No size limit
Life stages (different appearance)	Mainly adults	All stages (eggs)
Quantitative	✓	Qualitative, relative no but problematic for eukaryotes
Misidentification	Taxonomic expertise	Further analysis, suboptimal taxonomic resolution false negative, reference database

Thank you

E-DNA methods

- Three reference samples containing crocodile (*Crocodylus porosus), Phaeodactylum* diatom *(Phaeodactylum tricornutum)* and the marine mussel (*Mytillus edulis*) were also processed in three replicates as positive controls
- To identify potential laboratory contamination within DNA samples, three ultra-pure water (nuclease free water) controls were analysed at the same time as the environmental sediment samples
- three contamination controls (DNA free water placed into the blender and treated as a sample) were sequenced to identify potential contamination associated with the use of the blender for homogenisation.

eDNA methods

- DNA was sent to the Ramaciotti Centre for Genomics Genomics (UNSW Sydney, Australia) for amplicon generation and sequencing
- Different protocols were required for the amplification and library generation of each target gene
- Next generation sequencing of all target genes was carried out using the Illumina MiSeq platform
- Broad eukaryote target genes: Mitochondrial COI (313 bp fragment) of the mitochondrial COI gene
- 18S V4 rDNA - DNA was sent to the Ramaciotti Centre for Genomics where the 18S V4 rDNA region was amplified using gene primers TAReuk454FWD1 and TAReuk-Rev3 and sequenced with 250 bp paired reads
-

Primer sets to target taxa

- PCR amplifications were carried using the AmpliTaq
- Sequenced data were processed using the custom pipeline Greenfield Hybrid Amplicon Pipeline v2.1 (GHAPv2.1) which utilises USEARCH sequence analysis tools (Edgar, 2013).

Target taxa	Amplicon	Length (bp)	Primer forward	Primer reverse	Annealing temp	Reference
Foraminifera	v4 18S rDNA	300	GAAGGGCACCACAAGAACGC	CCACCTATCACAYAATCATG	56°C	Pochon et al., 2016
Polychaetes	mt16S	90	CCGGTYTGAACTCAGMTCA	TGGCACCTCGATGTTGGCT	58°C	Taberlet et al., 2018
Diatoms	v4 18S rDNA	350	GCGGTAATTCCAGCTCCAATAG	CTCTGACAATGGAATACGAATA	58°C	Zimmerman et al., 2011
Crustaceans	v9 18S rDNA	180	TGGTGCATGGCCGTTCTTAGT	CATCTAAGGGCATCACAGACC	56°C	Hardy et al., 2010

eDNA methods

- Samples were thawed just prior to DNA extraction
- Large fragments (shells or rocks) removed from the sample
- Sediment was homogenized using a bench-top blender on the highest speed setting for 2minutes
- DNA was extracted in triplicate from a 0.5 g subsample of homogenised sediment using the QIAGEN DNeasy PowerSoil® Kit
- 18S rDNA and COI used
- 18S rDNA was used to assess the overall eukaryotic community composition, including the micro-, meio- and macrofauna and specific primer pairs were applied to target specific taxonomic groups of interest (polychaetes, diatoms, foraminifera and crustaceans)
- The primers used in COI were designed to be biased against the microbial eukaryotes, thus the results are the macro- and meiofauna

Sequence classification

- 18S OTU sequences were classified by BLASTing them against a curated reference set derived from the SILVA non-bacterial sequences (V128)
- mt16S a custom made mt16S database from blast cleaned up and curated by Paul Greenfield

18S

- The amplicon sequence data was processed using GHAP, an in-house amplicon clustering and classification pipeline built around tools from USearch combined with locally-written tools for demultiplexing (diving sequence reads into seprate files for each index tag/sample, trimming and generating OTU tables
- This pipeline, available at at https://doi.org/10.4225/08/59f98560eba25 takes files of reads and produces tables of classified OTUs and their associated reads counts across all samples
- The amplicon reads are demultiplexed and trimmed, and the read pairs are then merged (using *fastq_mergepairs*) and de-replicated (using *fastx_uniques*) The merged reads are then trimmed again and clustered at 97% similarity (using *cluster_otus*) to generate OTUs (Operational Taxonomic Units)
- Representative sequences from each OTU are then classified by using *ublast* to find the closest match in a set of reference sequences. Eukaryote ribosomal SSU (18S) reads are matched against curated sequences derived from the SILVA v128 SSU reference set, and other amplicons, such as polychaetes and mitochrondial COI, are matched against custom-made reference sets
- The pipeline then maps the merged reads back onto the OTU sequences (using *usearch_global*) to get accurate read counts for each OTU/sample pair, and generates OTU tables in both text and .biom (v1) formats, complete with taxonomic classifications and species assignments
- The OTU tables are then summarised over all taxonomic levels, combining the counts for identified taxa across all OTUs

- Edgar, R.C. (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads, *Nature Methods.* doi: 10.1038/nmeth.2604.
- Quas,t C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.41 (D1): D590-D596.*