Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Biotechnology

Continuous Plankton Recorder in the omics era: from marine microbiome to global ocean observations[$]

Luigi Vezzulli[1,4], Jaime Martinez-Urtaza[2,4] and Rowena Stern[3,4]

First routinely deployed in 1931 the Continuous Plankton Recorder (CPR) technology has established the most extensive, marine biological sampling programme in the world. With more than 90 years of sampling, over a total of 8 million nautical miles covered and 500 000 curated samples, the CPR survey provides a gold mine of information available to marine researchers. Such information is likely to exponentially increase thanks to new cutting-edge molecular technologies that are beginning to be applied on CPR samples. In this review we aim to address the exciting developments that the genomic revolution is having on CPR applications from the study of marine microbiome to ocean plankton communities leading to a new 'digital era' of the global ocean CPR observation programme.

Addresses
[1] Department of Earth, Environmental and Life Sciences (DISTAV), University of Genoa, Corso Europa 26, 16132 Genoa, Italy
[2] Department of Genetics and Microbiology, Facultat de Biociéncies, Universitat Autònoma de Barcelona (UAB), Bellaterra, 08193 Barcelona, Spain
[3] The Marine Biological Association the Laboratory, Citadel Hill Plymouth, PL1 2PB Devon, UK

Corresponding author: Vezzulli, Luigi (luigi.vezzulli@unige.it)
[4] These authors contributed equally.

Current Opinion in Biotechnology 2022, 73:61–66

This review comes from a themed issue on Environmental biotechnology

Edited by Luigi Vezzulli and Marco Ventura

For complete overview of the section, please refer to the article collection, "Environmental Biotechnology"

Available online 24th July 2021

https://doi.org/10.1016/j.copbio.2021.07.016

The Continuous Plankton Recorder survey from prewar to modern time

The CPR is a mechanical automatic plankton recording device invented by Sir Alister Hardy to observe patterns of distribution [1,2]. A prototype was deployed in Antarctica in 1925–1926 and a modified version deployed regularly to survey the North Sea from 1939 and then across the Atlantic in 1959, only breaking for World War II [2]. The recorder is towed behind ships of opportunity at 7–10 m depth. Plankton (from bacteria and viruses to copepods and other taxa) enters a small aperture and is captured on a band of silk mesh (270 mm mesh size). The recorder's gear mechanisms advance the silk bands at a constant rate to capture and cover the plankton for protection. Up to 450 nautical miles are collected on a spool and fixed with formalin to a final 4% concentration [1]. The silk is cut into samples representing 18.5 km (10 nautical miles), equivalent to 3 m^3 of water [3]. The samples are numbered, and only alternate, odd-numbered samples are analysed by light microscopy recording up to 700 taxonomic entities in a quasi-logarithmic manner that's remained unchanged from 1958 [1] to compare real changes in taxa. There are now sister surveys forming the Global Alliance of CPR Surveys expanding to other oceans [4•]. Approximately 500 000 samples have been archived since 1958 (URL: https://www.cprsurvey.org/), of which 272 705 have been microscopically analysed for taxonomic and abundance information, a resource than can be exploited to capture the unmeasured planktonic component.

Genomic revolution applied to archived formalin-fixed CPR samples

The 'Genomic revolution' begin in the mid 1970's and culminated in recent years with the explosion of 'omics' sciences mainly in relation to the advent of the Next-Generation Sequencing (NGS) technologies [5]. Because of long term stability of DNA in preserved biological samples, the 'omics revolution' opened a new era in the analysis of the ocean's largest plankton archive.

The principal challenges facing molecular analyses of genetic material extracted from CPR samples is the fragmentation of the DNA and cross-linked protein-DNA complexes, induced by formalin fixation of the samples [6]. The chemical reaction of formalin with nucleotides and proteins also leads to the occasional deamination of cytosine to uracil which results in a C-T conversion when the DNA is replicated for amplification in the laboratory [7,8]. Methods developed to extract DNA from formalin-preserved CPR samples are based on the recovery of plankton from the silk which is initially washed to dilute and remove residual formalin [9–11].

[$] Given his role as Guest Editor, Luigi Vezzulli had no involvement in the peer review of the article and has no access to information regarding its peer-review. Full responsibility for the editorial process of this article was delegated to Marco Ventura.

Bacterial and eukaryotic DNA is then release by lysozyme (1–2 hours) and proteinase K (4–24 hours) digestion of cell material and a heating step (e.g. 90°C for 1 hours) is applied to partially reverse formalin modification of nucleic acids. After heating Uracil-N-Glycosilase (UNG) is added to specifically remove deaminated cytosine residues [12]. At this stage DNA extraction can be carried out through chemical means, such as phenol-chloroform [10] or spin column-based nucleic acid purification [11]. Using a concentration step by ultrafiltration, variable DNA yields are obtained (generally in the range 10–1000 ng) from CPR samples depending on sampling areas and the amount of plankton captured through the filtering process. Assessment of fragment size is 200–800 bp in the oldest samples (>50 years) which reduces with decreasing sample age, where DNA fragments of >2000 bp are generally detected in recent samples <10 years old [11].

Following the development of effective protocols to extract and purify genetic material with sufficient quality and quantity, a portfolio of methods has been developed and successfully applied for downstream molecular analysis of CPR samples, from amplification based, for example, end-point PCR, quantitative Real Time PCR, Digital Droplet PCR [9,10,13,14,15••] to next generation sequencing, for example, amplicon sequencing, shotgun and targeted metagenomics methods [11,15••,16]. New big data and bioinformatics technologies are now being applied to large scale spatial-temporal analysis of the plankton communities collected by the CPR survey.

Prokaryotic plankton

Molecular analysis of prokaryotic communities in CPR samples has historically been hampered by sample contamination. This may occur during (i) preparation and loading of the silk in the CPR before each tow, (ii) silk manipulation during cutting and microscopic analysis of the sample and (iii) long term storage in the CPR archive. Points (ii) and (iii) are the most critical as the chance of contamination at these stages are exacerbated by the presence of formalin in the samples.

A partial solution to such problems has been circumvented by using untouched plankton samples and by targeting strictly marine species in molecular studies. For example, archived CPR samples have been employed to retrospectively investigate long-term variations of Vibrio populations in the marine environment [11]. Bacteria of the genus 'Vibrio' include several species pathogenic for humans and animals and are only found in aquatic environments mainly in association with plankton organisms especially zooplankton [17]. A molecular index called 'Vibrio relative abundance index' was developed and applied to measure and compare Vibrio concentration in archived CPR samples [11]. Results from these studies provided first experimental evidence on the link between rising Vibrio pathogens in the ocean and their associated human diseases with increase of sea surface temperature over the last half a century [18].

More recently microbiological analysis of CPR samples was employed to investigate aquatic reservoirs of Vibrio cholerae, the causative agent of epidemic cholera, in endemic countries [13]. By combining large scale CPR plankton sampling, microbial culture and ultrasensitive molecular methods, namely Droplet Digital PCR (ddPCR) and targeted genomics, the presence of V. cholerae was investigated in a 96 600 L volume of surface water collected on a 322 nautical mile (596 km) transect in Lake Tanganyika the longest and second largest (by volume) lake in the world. V. cholerae was detected and identified in a large area of the lake albeit toxigenic strains (e.g. V. cholerae O1 or O139) were not detected in plankton samples [19••]. This study represents to date the largest investigation ever conducted for the research of V. cholerae in large tropical lakes. In addition, it provides an important proof of concept on the use of the CPR technology in cholera studies which could be employed in other areas of the world, where the establishment of environmental reservoirs of epidemic V. cholerae strains are suspected to play a role in the origin of cholera outbreaks. For instance, the developed methodology was recently employed to investigate presence of non-O1/O139 V. cholerae in CPR samples collected before, during and after a V. cholerae disease outbreak occurred in March 2018 in Vancouver Island (British Columbia, Canada).

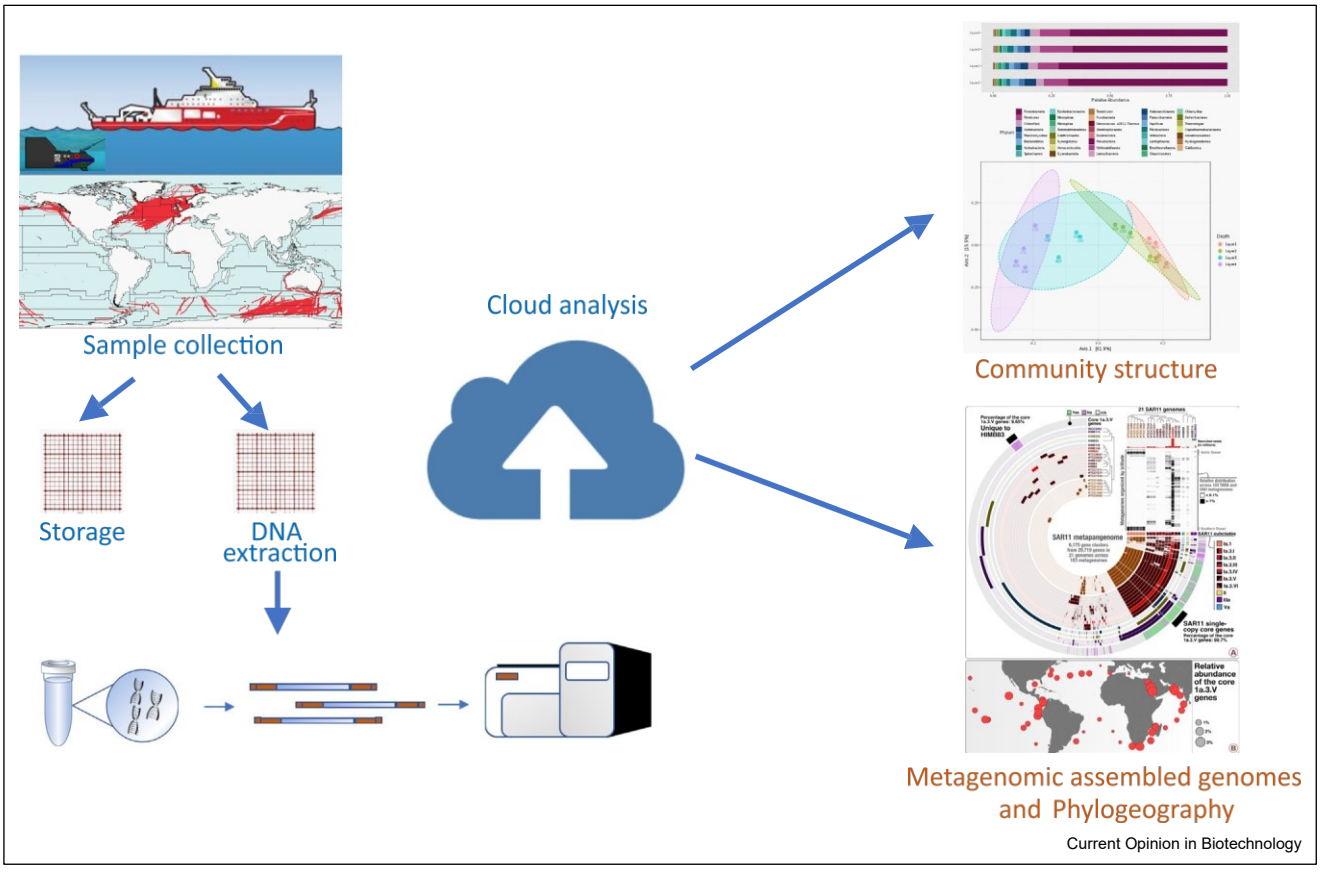
Eukaryotic plankton

With its large spatial and temporal diversity, archived CPR samples can uncover hidden eukaryotic plankton distribution changes over climate-relevant scales [20]. Early proof of concept studies successfully identified echinoderms from Mitochondrial 16S rDNA sequences in CPR samples [21]. A cytochrome c oxidase I (COI) metabarcoding comparison of zooplankton from CPR samples (gCPR) with eDNA sampling (2 L) yielded 1.5 more taxa and rarer types with the former and two-thirds of the species were identified by both methods [22•]. Yet, PCR amplification and sequencing of a coccolith morphology motif (CMM) in Emiliania huxleyi successfully identify variants of these coccolithophores [23] from samples as far back as 1972 [10].

Coccolithophores collected by CPR and Niskin bottles in parallel during a coccolithophore bloom were compared, revealing the same dominant CMM genotype variant in both samples [24] in similar proportions.

The DNA extraction methodology developed by Ripley et al. [10] allowed for the retrospective genetic determination of whether Karenia mikimotoi in the Celtic Sea was a non-native species introduced from Japan using the rbcL gene marker to demarcate allopatric populations. CPR

Figure 1



Next Generation of Genomic-based CPR Analysis.

samples from 1963 determined that this species was native and present before its assumed date of introduction (M Al-Kandari, PhD thesis, University of Plymouth, 2012).

Novel spatial patterns of Pseudo-nitzschia spp. were uncovered in Northeast and central Pacific CPR samples from 2002 to 2008 [15••]. Some Pseudo-nitzschia spp can cause harmful algal blooms therefore there is a real incentive to be able to discriminate the taxa present [15••]. However, this pennate diatom genus cannot be speciated by light microscopy alone. Using a partial LSU marker amplification approach from McDonald et al. [25], species-level identification was successful using the now defunct 454 pyrosequencing technology and with standard sanger sequencing. Both methods produced similar species profiles and two main species. Pseudo-nitzschia fraudulenta was the only species found by electron microscopy (EM) and sequencing, indicating many are broken up, or an introduced PCR amplification bias in taxa detected.

A highly unusual application of CPR survey was reported where live oceanic microbes were recovered from a survey in the Tasman Sea when the Australian CPR survey captured an extreme and large dust storm event in 2009 [26]. The samples were black and covered in tiny 5 mm fungal spores. Incredibly after rinsing off the formalin, the fungi could be cultured and sequenced. The ribosomal Internal Transcribed Spacer (ITS), 28S large ribosomal subunit gene and b-tubulin gene marker sequences identified the fungus Aspergillus syndowii in these samples. This species is associated with corals and could be a good indicator of coral health.

Perspective for a global ocean CPR molecular observation programme

The rapid progress of sequencing technologies and computational analysis have revolutionised the way that oceans are investigated and introduced unprecedent levels of resolution in the study of the seas [27••,28]. It is possible today to capture a snapshot of prokaryotic and eukaryotic communities in the oceans simply with the access to the DNA extracted from samples [29]. The application of genomic science to explore the marine biosphere with initiatives like the Sorcerer II Global Ocean Sampling (GOS) [30] and the most recent Tara

Oceans [27••] has unlocked a hidden world of microbes in the oceans, expanding our understanding of biodiversity and ecology and reshaping the vision of the global ocean microbiome [31].

Initial microbial community studies were primarily based on the application of 16S metabarcoding, which provide a partial picture of microbial communities present in marine ecosystem [32]. The widespread use of shotgun metagenomics has enabled a deeper scrutiny of the microbial world, overcoming the limits imposed by the use of a target-based strategy and expanding the analysis to a genome-wide level [33]. Sequencing technologies today provide more extensive and better-quality data, and it is possible to reconstruct almost complete genomes from metagenomic sequences with an incredible level of completeness and accuracy [34•]. These genomes can be analysed within an ecological or evolutionary contest, or used to identify the biogeographical patterns of species distribution on a global scale (Figure 1) [35]. The advent of the third generation of sequencing technologies has made even simpler and more affordable the implementation of genome-based studies. The access to portable and cheap instruments, such as Oxford Nanopore, using simple protocols for an end-to-end metagenomics solution can today be applied to sequence samples in the field potentially within the payload of the CPR device or in parallel to it. Another innovation of third generation sequencing devices is the generation of long reads (longer than 100 Kb) which is critical to improve the metagenomics sequence assembly to generate complete genomes [36]. An additional benefit of Oxford Nanopore technology is adaptive sampling: their software can select which DNA fragments are sequenced [37•] than can effectively enrich low abundant and rare species in metagenomic samples, that can otherwise be overridden by the signal from a dominant species.

Despite the clear benefits introduced by the last generation of sequencing technologies, it is still necessary to improve some critical aspects and overcome some problems limiting the application of these technologies to CPR samples. The improvement of the nucleic acid extraction/collection procedures is probably today one of the major challenges to expand the use of new sequencing approaches to CPR samples as recently highlighted by POGO (Partnership for Observation of the Global Ocean) scientists during a CPR Molecular Workshop held in Hobart Tasmania in November 2019 (https://pogo-ocean.org/innovation-in-ocean-observing/activities/dna-sampling-tools-for-the-global-cpr-survey/).

As mentioned previously, DNA is degraded and fragmented during its preservation in formalin, which limits the application of deep and long-read sequencing and most of its benefits. Extraction protocols need to be improved to yield more abundant, longer and better-quality DNA to expand the use of new sequencing technologies in formalin-preserved samples. Introducing less aggressive methodologies avoiding denaturation and dehydration, such as isotachophoresis [38], which works solely based on the separation by electrophoretic mobility, can be an alternative to obtain a higher yield of pure and less fragmented nucleic acid free of contamination. These strategies should consider approaches to repair and restore old DNA in the case of historical formalin-fixed samples [39].

An alternative to dismiss the damage caused by formalin preservation is to follow the initiatives introduced in many of the major biological collections in the world to create biorepositories based on collections of nucleic acid extracts [40]. Performing DNA extraction at the reception of samples to generate a DNA bank of CPR samples could be an option to preserve high quality DNA for future investigations making available genomic resources to researchers everywhere. This goes in parallel with implementing procedures to minimise contamination of the samples during manipulation and storage, for example, by preparing and preserving samples in sterile conditions.

Conclusions

The Continuous Plankton Recorder (CPR) Survey is a story of success. Running since 1931 the survey has collected marine plankton globally (coupled with ocean physical, chemical and biological observations) with the resulting data used by marine biologists, resource managers, scientific institutes, and governmental bodies across the world. Recently, DNA-based tools and next generation sequencing have started a new era of the CPR programme that is paving the way for large spatial and temporal scale studies of ocean communities with a level of resolution never achieved before. In this timeframe, the establishment of a DNA biorepository together with this traditional sample archive is a necessary step to fully realise the potential of the CPR survey in the post genomic era. Minimising sample contamination, improving current nucleic acids extraction protocols and applying new technologies such as target metagenomics and long-read sequencing will greatly contribute to this goal (Figure 1). In addition, updated and cheaper CPR instruments such as the Plankton Indicator type 2017 (a small plankton Indicator that can be towed with greater ease for short distance) coupled with molecular analysis of the samples might find application in coastal monitoring programmes such as pathogen detection and surveillance in human and aquaculture settings worldwide.

Overall, the genomic revolution is expected to have a profound impact on the CPR survey in the future, including social-ecological applications embracing human health (e.g. V. cholerae and microbial pathogens studies, identifying harmful algal bloom species) as well as changes in species distributions which could have

consequences for marine resources. Thanks to the implementation of DNA-based molecular tools decades long times series could be generated on organisms that could not previously be identified using a microscope. Big data from omics studies together with the fact that samples already exist from many of the world ocean basins is expected to lead to a new 'digital era' of the global ocean CPR observation programme.

Conflict of interest statement

Nothing declared.

Acknowledgements

We wish to thank all the past and present members and supporters of the Continuous Plankton Recorder survey, whose efforts have allowed the long-term establishment and maintenance of the CPR dataset and archive. This paper was supported by funding from Natural Environment Research Council (NERC), omics themed research funding (grant NE/S013431/1).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Reid PC, Colebrook JM, Matthews JBL, Aiken J, Barnard R, Batten SD, Beaugrand G, Buckland C, Edwards M, Finlayson J et al.: The Continuous Plankton Recorder: concepts and history, from Plankton Indicator to undulating recorders. Prog Oceanogr 2003, 57:117-173.

2. Richardson AJ, Walne AW, John AWG, Jonas TD, Lindley JA, Sims DW, Stevens D, Witt M: Using continuous plankton recorder data. Prog Oceanogr 2006, 68:27-74.

3. John EH, Batten SD, Stevens D, Walne AW, Jonas T, Hays GC: Continuous plankton records stand the test of time: evaluation of flow rates, clogging and the continuity of the CPR time-series. J Plankton Res 2002, 24:941-946.

4. • Batten SD, Abu-Alhaija R, Chiba S, Edwards M, Graham G, Jyothibabu R, Kitchener JA, Koubbi P, McQuatters-Gollop A, Muxagata E et al.: A global Plankton diversity monitoring program. Front Mar Sci 2019, 6

This article describes concept, methodology and use of the Continuous Plankton Recorder (CPR) survey for global plankton diversity monitoring.

5. Goodwin S, McPherson JD, McCombie WR: Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016, 17:333-351.

6. Einaga N, Yoshida A, Noda H, Suemitsu M, Nakayama Y, Sakurada A, Kawaji Y, Yamaguchi H, Sasaki Y, Tokino T et al.: Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. PLoS One 2017, 12.

7. Chen G, Mosier S, Gocke CD, Lin M-T, Eshleman JR: Cytosine deamination is a major cause of baseline noise in next-generation sequencing. Mol Diagn Ther 2014, 18:587-593.

8. Mcdonough SJ, Bhagwate A, Sun Z, Wang C, Zschunke M, Gorman JA, Kopp KJ, Cunninghamid JM: Use of FFPE-derived DNA in next generation sequencing: DNA extraction methods. PLoS One 2019, 14.

9. Kirby RR, Reid PC: PCR from the CPR offers a historical perspective on marine population ecology. J Mar Biol Assoc UK 2001, 81:539-540.

10. Ripley SJ, Baker AC, Miller PI, Walne AW, Schroeder DC: Development and validation of a molecular technique for the analysis of archived formalin-preserved phytoplankton samples permits retrospective assessment of Emiliania huxleyi communities. J Microbiol Meth 2008, 73:118-124.

11. Vezzulli L, Brettar I, Pezzati E, Reid PC, Colwell RR, Höfle MG, Pruzzo C: Long-term effects of ocean warming on the prokaryotic community: evidence from the vibrios. ISME J 2012, 6:21-30.

12. Serizawa M, Yokota T, Hosokawa A, Kusafuka K, Sugiyama T, Tsubosa Y, Yasui H, Nakajima T, Koh Y: The efficacy of uracil DNA glycosylase pretreatment in amplicon-based massively parallel sequencing with DNA extracted from archived formalin-fixed paraffin-embedded esophageal cancer tissues. Cancer Genet 2015, 208:415-427.

13. Vezzulli L, Stauder M, Grande C, Pezzati E, Verheye HM, Owens NJP, Pruzzo C: GbpA as a novel qPCR target for the species-specificdetection of Vibrio cholerae O1, O139, Non- O1/Non-O139 in environmental, stool, and historical continuous plankton recorder samples. PLoS One 2015, 10.

14. Di Cesare A, Petrin S, Fontaneto D, Losasso C, Eckert EM, Tassistro G, Borello A, Ricci A, Wilson WH, Pruzzo C et al.: ddPCR applied on archived Continuous Plankton Recorder samples reveals long-term occurrence of class 1 integrons and a sulphonamide resistance gene in marine plankton communities. Env Microbiol Rep 2018, 10:458-464.

15. •• Stern R, Moore SK, Trainer VL, Bill BD, Fischer A, Batten S: Spatial and temporal patterns of Pseudo-nitzschia genetic diversity in the North Pacific Ocean from the Continuous Plankton Recorder survey. Mar Ecol Prog Ser 2018, 606:7-28

This study represents the first application of high-throughput genetic methods to uncover patterns of Pseudo-nitzschia genetic diversity from archival CPR samples, demonstrating the value of using CPR for plankton community analysis in rarely sampled regions of the oceans.

16. Vezzulli L, Grande C, Tassistro G, Brettar I, Höfle MG, Pereira RPA, Mushi D, Pallavicini A, Vassallo P, Pruzzo C: Whole-Genome Enrichment Provides deep insights into Vibrio cholerae metagenome from an African river. Microb Ecol 2017, 73:734-738.

17. Erken M, Lutz C, McDougald D: Interactions of Vibrio spp. with zooplankton. Microbiol Spectr 2015, 3.

18. Vezzulli L, Grande C, Reid PC, Hélaouët P, Edwards M, Höfle MG, Brettar I, Colwell RR, Pruzzo C: Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic. P Natl Acad Sci U S A 2016, 113:E5062-E5071.

19. •• Vezzulli L, Oliveri C, Borello A, Gregory L, Kimirei I, Brunetta M et al.: Aquatic reservoir of Vibrio cholerae in an African Great Lake assessed by large scale plankton sampling and ultrasensitive molecular methods. ISME Commun 2021, 1:20

In this study a Continuous Plankton Recorder was employed to investigate the role of tropical lakes as reservoirs of the cholera-causing bacteria, Vibrio cholerae. This work opens a new dimension on the use of CPR coupled with modern genetic methods to investigate microbial waterborne pathogens and their associated human diseases.

20. McGinty N, Barton AD, Record NR, Finkel ZV, Johns DG, Stock CA, Irwin AJ: Anthropogenic climate change impacts on copepod trait biogeography. Glob Change Biol 2021, 27:1431-1442.

21. Kirby RR, Lindley JA: Molecular analysis of Continuous Plankton Recorder samples, an examination of echinoderm larvae in the North Sea. J Mar Biol Assoc UK 2005, 85:451-459.

22. • Suter L, Polanowski AM, Clarke LJ, Kitchener JA, Deagle BE: Capturing open ocean biodiversity: comparing environmental DNA metabarcoding to the Continuous Plankton Recorder. Mol Ecol 2021, 30:3140-3157 http://dx.doi.org/10.1111/mec.15587

This study compared eDNA metabarcoding using the cytochrome c oxidase I (COI) marker with conventional morphological CPR sampling and bulk DNA metabarcoding of CPR collected plankton. Results showed that COI metabarcoding of bulk DNA samples mostly avoided nonmetazoan amplifications and recovered more zooplankton species than conventional morphological CPR sampling and eDNA metabarcoding.

23. Godrijan J, Drapeau D, Balch WM: Mixotrophic uptake of organic compounds by coccolithophores. Limnol Oceanogr 2020, 65:1410-1421.

24. Highfield A, Evans C, Walne A, Miller PI, Schroeder DC: How many coccolithovirus genotypes does it take to terminate an Emiliania huxleyi bloom? Virology 2014, 466–467:138-145.

25. McDonald SM, Sarno D, Zingone A: Identifying Pseudo-nitzschia species in natural samples using genus-specific PCR primers and clone libraries. Harmful Algae 2007, 6:849-860.

26. Hallegraeff G, Coman F, Davies C, Hayashi A, McLeod D, Slotwinski A, Whittock L, Richardson AJ: Australian dust storm associated with extensive Aspergillus sydowii fungal "Bloom" in coastal waters. Appl Environ Microbiol 2014, 80:3315-3320.

27. •• Sunagawa S, Acinas SG, Bork P, Bowler C, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G et al.: Tara Oceans: towards global ocean ecosystems biology. Nat Rev Microbiol 2020, 18:428-445

This paper reviews the work of Tara Oceans, an international, multi-disciplinary project to assess the complexity of ocean life across comprehensive taxonomic and spatial scales. Tara Oceans is the most extensive sampling of plankton ever made.

28. Ustick LJ, Larkin AA, Garcia CA, Garcia NS, Brock ML, Lee JA, Wiseman NA, Keith Moore J, Martiny AC: Metagenomic analysis reveals global-scale patterns of ocean nutrient limitation. Science 2021, 372:287-291.

29. Cindy B, Simon Neil J, Tiffany S, Haylea M, Todd S, John Kenneth K, Oliver B: Passive eDNA collection enhances aquatic biodiversity analysis. Commun Biol 2021, 4.

30. Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S, Wu D, Eisen JA, Hoffman JM, Remington K et al.: The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol 2007, 5:0398-0431.

31. Yooseph S, Sutton G, Rusch DB, Halpern AL, Williamson SJ, Remington K, Eisen JA, Heidelberg KB, Manning G, Li W et al.: The sorcerer II global ocean sampling expedition: expanding the universe of protein families. PLoS Biol 2007, 5:0432-0466.

32. Giovannoni SJ, Rappé MS, Vergin KL, Adair NL: 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. P Natl Acad Sci U S A 1996, 93:7979-7984.

33. Wei Z-F, Li W-L, Huang J-M, Wang Y: Metagenomic studies of SAR202 bacteria at the full-ocean depth in the Mariana Trench. Deep-Sea Res Pt I 2020, 165.

34. • Cabello-Yeves PJ, Callieri C, Picazo A, Mehrshad M, Haro-Moreno JM, Roda-Garcia JJ, Dzhembekova N, Slabakova V, Slabakova N, Moncheva S et al.: The microbiome of the Black Sea water column analysed by shotgun and genome centric metagenomics. Environ Microb 2021, 16

This study is a nice example on how recent developments in shotgun and genome-resolved metagenomics methods have revolutionised the analysis of marine microbial communities.

35. Tully BJ, Graham ED, Heidelberg JF: The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Sci Data 2018, 5.

36. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, Tyson JR, Beggs AD, Dilthey AT, Fiddes IT et al.: Nanopore sequencing and assembly of a human genome with ultra-long reads. Nat Biotechnol 2018, 36:338-345.

37. • Kovaka S, Fan Y, Ni B, Timp W, Schatz MC: Targeted nanopore sequencing by real-time mapping of raw electrical signal with UNCALLED. Nat Biotech 2021, 39:431-441

In this paper a real-time mapping of raw electrical signal is used for targeted nanopore sequencing of bacterial genomes. Adaptive sampling is a method of software-controlled enrichment unique to Nanopore long-reads sequencing platforms.

38. Persat A, Marshall LA, Santiago JG: Purification of nucleic acids from whole blood using isotachophoresis. Anal Chem 2009, 81:9507-9511.

39. Do H, Dobrovic A: Sequence artifacts in DNA from formalin-fixed tissues: causes and strategies for minimization. Clin Chem 2015, 61:64-71.

40. Yuille M, Dixon K, Platt A, Pullum S, Lewis D, Hall A, Ollier W: The UK DNA banking network: a "fair access" biobank. Cell Tissue Bank 2010, 11:241-251.