Spectral algal fingerprinting and long sequencing in synthetic algal-microbial communities

Ayagoz Meirkhanova[1], Sabina Marks[2], Nicole Feja[2], Ivan A. Vorobjev[1,3], Natasha S. Barteneva[*,1,4]

[1]School of Science and Humanities, Nazarbayev University, Astana 010000, Kazakhstan

[2]Faculty of Biology, University of Duisburg-Essen, Campus Essen, 45141 Essen, Germany

[3]National Laboratory Astana, Nazarbayev University, Astana 010000, Kazakhstan

[4]The Environmental Research and Efficiency Cluster, Nazarbayev University, Astana 010000, Kazakhstan

*Corresponding Author:

Professor Natalie (Natasha) S. Barteneva,

Department of Biology, Nazarbayev University,

53 Kabanbay Batyr Avenue, 010000, Astana, Kazakhstan

Phone: +7-7786357336; 1-617-935-5829

e-mail: natalie.barteneva@nu.edu.kz

Abstract

Full-spectrum cytometry offers a novel technique in environmental research, enabling rapid, high-throughput analysis of single algal cells from different taxa by capturing entire emission spectra. In this work we demonstrate label-free identification and enumeration of phytoplankton, overcoming traditional methods limitations. We propose a strategy for monitoring synthetic algal communities based on integration of high-throughput spectral cytometry and long sequencing analysis.

Keywords: synthetic algal-microbial communities; full-spectrum cytometry; spectral flow cytometry; phytoplankton; imaging flow cytometry**;** long sequencing; nanopore-based sequencing

In recent decades, synthetic biology has made significant progress in building artificial microbial and algal communities[1]. However, the difficulty of engineering complex biological systems has been humbling. As biosystems evolve, they change and refine themselves, and the evolutionary process sets them apart from non-biological systems[2]. Community selection experiments are related to fundamental questions in ecology and evolution as well as applied biotechnological problems. Uni-algal systems are not stable, and bacterial contamination is likely if the algal community is maintained heterotrophically on a medium with a fixed source of carbon[3] or dead algae present. The algal-microbial systems development is often overdominated by one or two algae or bacterial part of the community, with some exceptions[4]. It is related to the absence of methods allowing the following dynamics of mixed algal-microbial systems.

Traditional methods for studying microbial and algal communities, such as microscopy and pigment analysis, have throughput, taxonomic resolution, and functional characterization constraints and were primarily designed to differentiate a specific taxonomic group or provide averaged integrated data at a population level[5-7]. Conventional flow cytometry struggles to distinguish cells from other particulate matter in the water matrix, and imaging flow cytometry is capable to differentiate only few fluorescent channels. Full-spectrum cytometry, on the other hand, has recently transformed the field by offering high-throughput, multidimensional analysis of algal cells based on their size, complexity, and spectral fingerprint[8-10].

By capturing the entire emission spectrum from each single cell, full-spectrum cytometry may provide insights into phytoplankton community composition, structure, dynamics[11-12], and responses to environmental stressors that are accompanied by shifts in community structure[13]. The spectral cytometry allows for the identification of highly autofluorescent cellular subpopulations by virtual filtering[11] or with autofluorescence finder[14]. The work presented herein demonstrates the proof-of-concept that contemporary spectral cytometry analysis has advanced to a level that enables good

identification and enumeration of different algae based on their autofluorescent pigment signatures in

parallel with a long sequencing amplicon-based 16S rRNA analysis at the species level providing a

powerful tool for advancing our understanding of phytoplankton ecology, synthetic biology, and

evolution.

The first step in the analysis of complex communities requires construction of a library with unique

spectral signatures for different phytoplankton groups. For this purpose, single phytoplankton cultures

were recorded on ID7000 (Sony Biotechnology), and individual spectral signatures were extracted.

(**Figure 1a**). Clear separation between the two *Microcystis* autofluorescent populations was observed

across all lasers, except for 808 nm, given the optimal set of virtual filters (VFs) (**Figure 1b-d**). The

same workflow was applied for the acquisition of spectral signatures of the remaining phytoplankton

monocultures (**Figure 1e**), resulting in a spectral library used for further analysis.

Figure 1. Schematic overview of the proposed experimental workflow with MinION Mk1C-based sequencing platform (Oxford Nanopore) and ID7000. It includes: **(a)** recording and isolation of single spectral signatures with "Autofluorescence Finder" tool; **(b)** a set of optimal virtual filters for discrimination of autofluorescent populations with a set of optimal virtual filters (**AF-A** - *Microcystis* sp. and **AF-B** – debris); **(c)** respective emission spectra (red line – **AF-A**, blue line – **AF-B**); **(d)** *Microcystis*; **(e)** spectral signatures library corresponding to respective phytoplankton species, is then constructed, and utilized during subsequent analysis, including unsupervised clustering techniques (e.g., t-SNE) in parallel with species-level bacterial classification.

To discriminate multiple components in a suspension, a mixture of picoplanktonic cyanobacteria was used and recorded for further analysis (**Supplementary Figure 1**). In addition, components of the mix represented variable taxonomic groups at the Order level (*Microcystis* sp., belonging to Chroococcales), as well as members of the same Order (*Synechococcus* sp., *Synechocystis* sp., belonging to Synechococcales. The gating strategy is presented in **Suppl. Figure 1c**, where a combination of [VF-647]_A vs [VF-488]_A reveals two subpopulations, corresponding to *Synechocystis* sp. and population "A". Due to the similarity in the spectra of *Microcystis* sp., and *Synechococcus* sp. subpopulation "A" was then differentiated based on SSC values. Identified populations could not be discriminated based on FSC and SSC only (**Suppl. Figure 1d**). At this point, the maximum number of individual phytoplankton groups that could be discriminated reached 14 (**Supplementary Figure 2**).

Next, applicability of full-spectrum cytometry in the temporal monitoring of synthetic phytoplankton communities was assessed. Setting up an artificial mix experiment with 3 representatives of major phytoplankton groups – Cyanobacteria, Chlorophyta, and Cryptophyta (**Supplementary Table 1, Figure 3**), growth of which was monitored for the period of 9 days.

Triplicates were also analyzed in parallel on BD FACSAria (BD Biosciences, USA), ImageStreamX

MarkII (Cytek, USA), and microbial community composition was assessed using Nanopore-based

full-length 16S rRNA gene sequencing. Firstly, a library containing 9 individual spectral signatures

covering each phytoplankton species within the mix was constructed (**Figure 2a).** As seen from the

graph, regions unique for each of the groups are present, specifically between channels 10-15 (566-

646 nm) of the spectra. This information was further utilized to define a set of optimal VF pairs to

discriminate these populations in a complex mix. VFs were defined in a way that clustered

subpopulations within the same group, while maintaining the difference between the groups. A set of

filters [VF-320]_A vs [VF-561]_A (**Figure 2b**) was chosen as an optimal pair for the first step within

the gating strategy (**Figure 2c**), where three groups are identified and discriminated from the rest of

the unwanted events (background noise/cellular debris).

The next steps involved pairwise combinations of defined filters to discriminate the components of

each of the subpopulations. As mentioned previously, due to the high similarity of spectra between

Microcystis sp. and *Synechococcus* sp., SSC was used to differentiate these populations. A comparison

of the data obtained on 3 instruments revealed the limitations of some of the instruments

(**Supplementary Figure 4**). Opt-SNE dimensionality reduction algorithm was utilized to minimize

"gating bias" and group data into clusters with similar characteristics in low-dimensional space. 3

datasets were compared: synthetic mix data obtained from the ID7000 instrument using a set of pre-

defined VFs and a set of VFs matching the filters used on FACSAria, and lastly, data obtained from

FACSAria. Both sets of filters utilized on the ID7000 resulted in a clear separation of all 9

subpopulations within the mix, while a single population of *Scherffelia* sp. was not possible to

discriminate from the rest on the FACSAria instrument. These results demonstrate the higher

flexibility a virtual filter system offers in the analysis of multi-component mixes. In addition to more

precise population separation obtained with the ID7000 instrument, it was possible to track the

abundance of nearly all mix components, as demonstrated in **Supplementary Figure 5**. Overlay of

temporal spectral signatures recorded for the period of 9 days for each mix member was used to assess

the extent of signal intensity dispersion. As revealed through the coefficient of variation, regions most

susceptible to differences in intensity generally included channels 1-13 (413.6-612.6 nm) and 30-34

(785.4-835.1 nm). High variation within these regions was also observed when comparing single

spectral signatures between excitation lasers (**Supplementary Figure 6**).

Figure 2. Tracking temporal changes in synthetic community composition. **(a)** overlay plot of

individual spectral signatures for each phytoplankton group representative (Cyanobacteria,

Chlorophyta, and Cryptophyta); red regions indicate regions of interest used for setting an optimal

pair of VFs; **(b)** "Autofluorescence Finder" tool with VFs set to regions of interest, demonstrating the

best separation of subpopulations on [VF-320_A] VS [VF-561_A] plot; **(c)** The gating strategy,

composed of sequential sub-gating steps, used to identify phytoplankton subpopulations within the

synthetic community; **(d)** Microbial community composition in experimental triplicates at phylum

and species level revealed through full-length 16S rRNA gene-based Nanopore sequencing.

In conclusion, by constructing spectral libraries for various phytoplankton groups and differentiating

between algal taxa and debris based on autofluorescent pigment signatures and light scattering,

spectral cytometry overcomes the limitations of traditional methods and provides detailed insights

into community structure, dynamics, and responses to environmental stressors. The potential for

parallel sequencing and associated microbial analysis at the multi-omics level is needed to fully

elucidate the microbial functions occurring in artificial algal communities. This technology provides

possibilities for a wide range of research applications, from addressing fundamental ecological and

evolutionary questions to applied biology, thereby advancing the field of environmental research.

173 **Methods**

174 *Sample preparation and instrument setup*

Freshwater algal cultures (**Supplementary Table 1**) were obtained from The Culture Collection of

Algae at the University of Göttingen (Germany) and Central Collection of Algal Cultures at the

University of Duisburg-Essen (Germany). Cultures were maintained in Basal Medium (BM) at 21°C,

under 12/12 L/D cycle in algae growth chamber AL-30L2 (Percival, USA). Samples were recorded

in parallel on 3 instruments: 7 lasers ID7000 Spectral Cell Analyzer (Sony Biotechnology, USA), 5

lasers BD FACSAria (BD Biosciences, USA) and ImageStreamX MarkII (Amnis - Cytek, USA)

equipped with 40X and 60X objectives. Samples prepared for cytometric analysis included mono- and

mixed cultures and were recorded at spectral cytometer using a standard 24-tube rack. Each sample

was placed in round-bottom 5mL tubes, with minimum volume of 250uL, and arranged in specified

positions on the rack. Prior to data acquisition, instrument settings were adjusted and set in accordance

with each sample (FSC and SSC gains depending on algal culture, threshold value – 11%, PMT

voltage ranging from 40% to 70%). Samples were recorded at minimum flow rate (=1) and at least

50,000 events was set for data acquisition. Scatter plots were used for initial assessment of instrument

settings in the "Preview" mode prior to recording.

Autofluorescence Finder Tool and Virtual Filters setup

The ID7000 software's Autofluorescence Finder Tool was used to analyze the algae samples,

treating autofluorescence as an independent fluorescent parameter. This feature is crucial for the

taxonomic identification of different algae. By incorporating excitation lasers ranging from 320nm to

808 nm, we included unique emission regions in the algal fluorescent signature, distinct from the Chl

peak and highly variable among different algae. This setup allowed for the detection of individual

autofluorescent signals for each of the recorded monoculture. Following the identification of the best

separation between autofluorescent populations, subpopulations of interest were gated and examined

for their spectral signatures. Each population demonstrated a unique autofluorescence spectrum.

Additionally, the Autofluorescence Finder Tool facilitated the distinction between dead and living

algae based on differences in autofluorescence.

Data analysis

Initial data analysis and population identification was performed using the ID7000 software.

Identification of subpopulations within artificial mix involved consecutive gating steps using a

combination of 2D plots with specified virtual filters for best separation. Furthermore, t-Distributed

Stochastic Neighbor Embedding, otherwise known as t-SNE, was done using FCS Express (De Novo

Software). A modified version of t-SNE - opt-SNE, a tool allowing to map high-dimensional

cytometry data onto a two-dimension plot while conserving the original high-dimensional structure,

was used to independently identify subpopulations within the artificial mix, minimizing gating bias.

Obtained clusters were then used to compare data acquired on different cytometers.

Sequencing library preparation and bioinformatic analysis

Samples for Nanopore-based 16S rRNA sequencing were collected during the last day of the

experiment and subsequent procedure consists of the following steps. Firstly, DNA was extracted

using the DNEasy Power Water Kit (Qiagen, Germany) according to the manufacturer's instructions,

with an additional heating lysis step. The full-length 16S gene was then amplified using the universal

16S primer pair 27F (5′-AGAGTTTGATCCTGGCTCAG-3′) and 1492R (5′-

TACGGYTACCTTGTTACGACTT-3′). Library preparation followed the manufacturer's protocol

(ONT, UK). The PCR reaction mixture included nuclease-free water, input DNA, LongAmp Hot Start

Taq 2X Master Mix (New England Biolabs, USA), and the respective 16S barcode primer. PCR

products were next purified using AMPure XP beads (Beckman Coulter, USA). In the final library

preparation step, sequencing adapters were added to the barcoded samples. The sequencing run was

performed on the MinION Mk1C device (Oxford Nanotechnologies, UK) for 40-46 hours. Raw signal

data were stored in FAST5 files and underwent basecalling with the Guppy neural network-based

basecaller integrated into the MinION Mk1C device. The resulting FASTQ files were processed using

Python commands, beginning with read assessment and quality evaluation, followed by quality and

read length filtering. Next steps involved adapter sequence removal and demultiplexing of reads into

their respective barcodes. The demultiplexed reads were then classified using the Emu taxonomic abundance estimator up to the species level (https://gitlab.com/treangenlab/emu)[15-16]. Lastly, the data were rarefied prior to further analysis.

Acknowledgements

We would like to acknowledge contribution of Aigul Kussanova during the initial stages of this work and Polina Len during sequence classification. We would also like to thank Dmitry Malashenkov and Aiym Duisen for helpful discussion and input, and Rudolf Bichele and Steffen Boechner from Sony Biotechnology for their instrumental help and support.

Authors contributions

A.M. designed the study, performed the experiments, analyzed the data. S.M. and N.F. contributed algal cells, in algal cells characterization. I.A.V. and N.S.B. initiated and supervised the project, and N.S.B. funded the project. A.M. and N.S.B. wrote the manuscript with contributions from all authors.

Data availability

Raw sequences obtained during the analysis were deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the BioProject ID PRJNA1131605. Spectral and cytometric data obtained during this study are available from the corresponding author on reasonable request.

Funding

The project was funded by MHES Kazakhstan grant # AP14872028 to N.S.B.

Competing interests

The authors declare no competing financial interests.

References

1. Niu, B. et al. PNAS 114, E2450-9 (2017).

2. Hartwell, L.H. et al. Nature 402, C47-52 (1999).

3. Andersen, R.A. Algal Culturing Techniques. Elsevier, Burlington, USA (2005).

4. Segev, E. et al. elife. 5, e17473 (2016).

5. Poryvkina, L. et al. J. Plankton Res. 16, 1315-27 (1994).

6. Liu, J.Y. et al. Appl. Spectroscopy Rev. 56, 171-92 (2021).

7. Swan, C.M. et al. Deep Sea Res. Part I: Oceanographic Res. Papers. 109, 137-56 (2016).

8. Futamura, K. et al. Cytometry Part A 87, 830-42 (2015).

9. Robinson, J.P. Biotechniques 72, 159-69 (2022).

10. Dott, T. et al. Cytometry Part A 105, 124-38 (2024).

11. Barteneva, N.S. et al. BioRxiv. 2019 Jan 10:516146 (2019).

12. Barteneva, N.S. et al. Using virtual filtering approach to discriminate microalgae by spectral flow cytometer. In: Spectral and Imaging Cytometry: Methods and Protocols, pp. 23-40. NY Springer (2023).

13. Jeppesen, E. et al. Freshwater Biol. 50, 1747-71 (2005).

14. Wanner, N. et al. Front. Bioengineer. 10, 827987 (2022).

15. Curry, K.D. et al. Nature Methods, 19, 845-53 (2022).

16. Len, P. et al. bioRxiv. 2023, 2023-12 (2023).

Supplementary material to "Spectral algal fingerprinting and long sequencing in synthetic algal-microbial communities"

Ayagoz Meirkhanova[1], Sabina Marks[2], Nicole Feja[2], Ivan A. Vorobjev[1,4], Natasha S. Barteneva[*,1,5]

[1]School of Science and Humanities, Nazarbayev University, Astana 010000, Kazakhstan

[2]Faculty of Biology, University of Duisburg-Essen, Campus Essen, 45141 Essen, Germany

[4]National Laboratory Astana, Nazarbayev University, Astana 010000, Kazakhstan

[5]The Environmental Research and Efficiency Cluster, Nazarbayev University, Astana 010000, Kazakhstan

Supplementary Table 1. List of algal cultures used for recordings during artificial mix experiment and for library construction using ID7000 spectral cell analyzer

Phylum	Order	Species	Habitat	Strain number
Cyanobacteria	Nostocales	*Dolichospermum urugayensis*	Freshwater	2498
		Nodularia sphaerocarpa	Freshwater	50.79
	Gloeobacter	*Gloeobacter violaceus*	Freshwater	35.87
	Chroococcales	*Microcystis* sp.	Freshwater	CCAC 7146B
	Synechococcales	*Synechococcus* sp.	Freshwater	CCAC 2944B
		Synechocystis sp.	Freshwater	CCAC 4059B
Chlorophyta	Chlorodendrales	*Scherffelia dubia*	Freshwater	CCAC 1398 B
	Chlamydomonadales	*Chlamydomonas* sp.	Freshwater	CCAC 4527 B
		Haematococcus pluvialis	Freshwater	CCAC 2072B
	Chlorellalels	*Chlorella vulgaris*	Freshwater	211-12
	Sphaeropleales	*Acutodesmus obliquus*	Freshwater	22.81
Cryptophyta	Pyrenomonadales	*Chroomonas* sp.	Freshwater	CCAC 2614B
		Rhodomonas sp.	Freshwater	CCAC 1479B
	Cryptomonadales	*Cryptomonas* sp.	Freshwater	CCAC 2345B
Rhodophyta	Stylonematales	*Chroothece richteriana*	Marine	104.79
	Porphyridiales	*Porphyridium cruentum cf.*	Marine	CCAC 3771 B
Dinophyta	Peridiniales	*Peridinium* sp.	Freshwater	CCAC 1612 B
Euglenophyta	Euglenida	*Euglena sanguinea*	Freshwater	CCAC 3518 B
Bacillariophyta	Fragilariales	*Fragilaria* sp.	Freshwater	CCAC 5509 B

Supplementary Figure 1. Discrimination of Picocyanobacteria in a synthetic mix based on autofluorescence. **(a)** Ribbon plot demonstrating combined spectral signatures of 3 picocyanobacterial representatives across 7 excitation lasers: *Synechococcus* sp., *Synechocystis* sp., and *Microcystis* sp. **(b)** Overlay plot of individual spectral signatures for each picocyanobacterial representative (*Synechococcus* sp. - blue, *Synechocystis* sp. - violet, and *Microcystis* sp. - red); red regions indicate regions of interest, used for setting an optimal pair of VFs. **(c)** Dot plot of **[VF-637]_A** against **[VF-488]_A** with 2 autofluorescent populations identified: *Synechocystis* sp., population **A** was then plotted with **[VF-561]_A** against **SSC**_A to resolve two subpopulations *of Microcystis* sp. and *Synechococcus* sp. **(d)** Dot plot demonstrating relative position of each of the identified picocyanobacterial species on a FSC_A against SSC_A plot.

Supplementary Figure 2. Gating strategy for discrimination of 14 groups of phytoplankton species based on autofluorescence in a complex synthetic mix. **(a)** Ribbon plot demonstrating combined spectral signatures of 16 phytoplankton species across 7 excitation lasers. **(b)** Overlay plot of individual spectral signatures for each phytoplankton representative. (c) Gating strategy for discrimination of each subpopulation within the synthetic mix: 1 – Chlorophyta (*Chlorella vulgaris* and *Acutodesmus obliquus*), 2 – *Euglena sanguinea*, 3 – *Peridinium* sp., 4 – *Cryptomonas* sp., 5 – *Fragilaria* sp., 6 – *Haematococcus pluvialis*, 7 – *Chroothece richteriana*, 8 – *Porphyridium cruentum* cf., 9 – filamentous Cyanobacteria (*Dolichospermum urugayensis* and *Nodularia sphaerocarpa*), 10 – *Chroomonas* sp., 11 – *Synechocystis* sp., 12 – *Gloeobacter violaceus*, 13 – *Microcystis* sp., 14 – *Synechococcus* sp.

329
330

331 **Supplementary Figure 3.** The gallery of algal images taken with ImageStream MKII (Amnis-Cytek,

332 USA) of phytoplankton species recorded during the artificial mix experiment.

1	*Microcystis* sp.	4	*Haematococcus* sp.	7	*Cryptomonas* sp.
2	*Synechocystis* sp.	5	*Chlamydomonas* sp.	8	*Rhodomonas* sp.
3	*Synechococcus* sp.	6	*Scherffelia* sp.	9	*Chroomonas* sp.

Supplementary Figure 4. Identification of artificial mix subpopulations compared between SONY

ID700 Spectral Cell Analyzer and BD FACSAria II. The first column represents data recorded on

SONY ID7000 using a set of VFs best optimized for artificial mix discrimination. Opt-SNE

demonstrates successful discrimination of the 9 subpopulations within the mix based on VF data. The

second column represents data recorded on SONY ID7000 using a set of VFs matching optical filters

used on BD FACSAria. Similarly, all 9 subpopulations were discriminated. Last column represents

data recorded on BD FACSAria. Based on the optical configuration of the instrument, only 8

subpopulations were identified using opt-SNE.

Supplementary Figure 5. Temporal overlay of spectral signatures recorded throughout the experiment. Individual spectral signatures for each member of the synthetic mix ((**a**) – *Microcystis* sp., (**b**) - *Synechococcus* sp., (**c**) – *Synechocystis* sp., (**d**) – *Chlamydomonas* sp., (**e**) - *Cryptomonas* sp., (**f**) – *Rhodomonas* sp., (**g**) – *Chroomonas* sp.) recorded over the period of 9 days overlayed to demonstrate temporal changes in signal intensity. To display signal dispersion, coefficient of variation (CV) was calculated and plotted across all detection channels for each of the excitation lasers. A threshold line at 30% was plotted for additional reference.

354 **Supplementary Figure 6.** Overlay plots for each phytoplankton species within a synthetic mix across

355 7 excitation lasers to demonstrate the variability of emission signal for each excitation source. The

356 first row demonstrates overlay plots for Cyanobacterial representatives, second – Chlorophyta, and

357 third – Cryptophyta. The black line denotes cutoff at channel 15 (624 – 635.3 nm).