

Hyperspectral time-series clustering of microalgal growth kinetics for strain-aware harvest timing

Rajalakshmi Elumalai^a, Karthikeyan Meenatchi Sundaram^b, Sunil Chinnadurai^b, Ankita Juneja^c, Karthik Rajendran^a,

*

^aThe NET ZERO Lab, Department of Environmental Science and Engineering, SRM University-AP, Amaravati, 522 240, Andhra Pradesh, India

^bDepartment of Electronics and Communication Engineering, SRM University-AP, Amaravati, 522 240, Andhra Pradesh, India

^cDepartment of Chemical Engineering, State University of New York College of Environmental Science and Forestry, Syracuse, NY, 13210, USA

Corresponding author: rajendran.k@srmap.edu.in

Table S1: Phylum-to-species taxonomic classification of the twenty strains

Sl. No.	Phylum	Class	Order	Species / Genus
1	Chlorophyta	Chlorodendrophyceae	Chlorodendrales	<i>Tetraselmis chuii</i>
2	Ochrophyta	Eustigmatophyceae	Eustigmatales	<i>Nannochloropsis oculata</i>
3	Haptophyta	Pavlovophyceae	Pavloales	<i>Pavlova lutheri</i>
4	Chlorophyta	Chlorophyceae	Sphaeropleales	<i>Scenedesmus</i> sp.
5	Chlorophyta	Chlorophyceae	Chlamydomonadales	<i>Dunaliella</i> sp.
6	Cyanobacteria	Cyanophyceae	Chroococcales	<i>Chroococcus minor</i>
7	Chlorophyta	Trebouxiophyceae	Chlorellales	<i>Chlorella vulgaris</i>
8	Cyanobacteria	Cyanophyceae	Chroococcales	<i>Aphanocapsa montana</i>
9	Cyanobacteria	Cyanophyceae	Chroococcales	<i>Gloeocapsa</i> sp.
10	Chlorophyta	Trebouxiophyceae	Chlorellales	<i>Picochlorum</i> sp.
11	Chlorophyta	Chlorodendrophyceae	Chlorodendrales	<i>Platymonas</i> sp.
12	Cyanobacteria	Cyanophyceae	Nostocales	<i>Anabaena</i> sp.
13	Chlorophyta	Chlorophyceae	Chlamydomonadales	<i>Chlamydomodium</i> sp.
14	Chlorophyta	Chlorophyceae	Sphaeropleales	<i>Coelastrella</i> sp.
15	Chlorophyta	Chlorophyceae	Sphaeropleales	<i>Desmodesmus</i> sp.
16	Chlorophyta	Chlorophyceae	Sphaeropleales	<i>Messastrum</i> sp.
17	Chlorophyta	Chlorophyceae	Sphaeropleales	<i>Monoraphidium</i> sp.
18	Cyanobacteria	Cyanophyceae	Nostocales	<i>Nostoc</i> sp.
19	Cyanobacteria	Cyanophyceae	Oscillatoriales	<i>Spirulina</i> sp.
20	Chlorophyta	Chlorophyceae	Sphaeropleales	<i>Tetradasmus</i> sp.

Table S2: Supporting calculations used in preprocessing and validation

No.	Name	Formula	Variables
1.	Relative reflectance	$R = \frac{I - I_d}{I_w - I_d}$	R : relative reflectance; I : raw sample image; I_d : dark-reference image; I_w : white-reference image
2.	Standard Normal Variate (SNV)	$x_{\text{SNV}} = \frac{x - \bar{x}}{s(x)}$	x : input spectrum; $\bar{x}$: mean of x ; $s(x)$: standard deviation of x
3.	Band-average reflectance	$B = \frac{1}{N} \sum_{i=1}^N R_i$	B : band value; R_i : reflectance at wavelength i ; N : number of wavelength points in the band window
4.	Silhouette score	$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))}$	$a(i)$: mean distance from strain i to strains in its own cluster; $b(i)$: mean distance from strain i to strains in the nearest other cluster
5.	Kruskal–Wallis statistic	$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$	N : total sample size; k : number of groups; R_j : rank sum of group j ; n_j : size of group j
6.	Area under the curve (trapezoidal rule)	$\text{AUC} = \sum_{t=1}^{T-1} \frac{y_t + y_{t+1}}{2} (d_{t+1} - d_t)$	y_t : band signal on day d_t ; T : number of measured days
7.	Early/late growth-phase slope	$\text{slope} = \frac{y_{\text{end}} - y_{\text{start}}}{d_{\text{end}} - d_{\text{start}}}$	$y_{\text{start}}, y_{\text{end}}$: band signal at the start and end of the phase; $d_{\text{start}}, d_{\text{end}}$: corresponding days

Table S3: Composition of the three-culture media (BG-11, BB, ASN-III) used to culture the strains

Component	BG-11 (g/L)	BB Medium (g/L)	ASN-III (g/L)
NaNO ₃	1.5	0.25	0.75
K ₂ HPO ₄	0.04	0.075	0.02
KH ₂ PO ₄	—	0.175	—
MgSO ₄ ·7H ₂ O	0.075	0.075	3.5
CaCl ₂ ·2H ₂ O	0.036	0.025	0.5
NaCl	—	0.025	25.0
KCl	—	—	0.5
MgCl ₂ ·6H ₂ O	—	—	2.0
Na ₂ CO ₃	0.02	—	0.02
Citric acid	0.006	—	0.003
Ferric ammonium citrate	0.006	—	0.003
EDTA (disodium)	0.001	0.05	0.0005
KOH	—	0.031	—
FeSO ₄ ·7H ₂ O	—	0.005	—
H ₃ BO ₃	0.00286	0.01142	0.00286
MnCl ₂ ·4H ₂ O	0.00181	0.00144	0.00181
ZnSO ₄ ·7H ₂ O	0.000222	0.00882	0.000222
Na ₂ MoO ₄ ·2H ₂ O	0.000390	—	0.000390
MoO ₃	—	0.00071	—
CuSO ₄ ·5H ₂ O	0.0000790	0.00157	0.0000790
Co(NO ₃) ₂ ·6H ₂ O	0.0000494	0.000494	0.0000494
Distilled water	1000 mL	1000 mL	1000 mL
pH	7.1	6.6–6.8	7.5

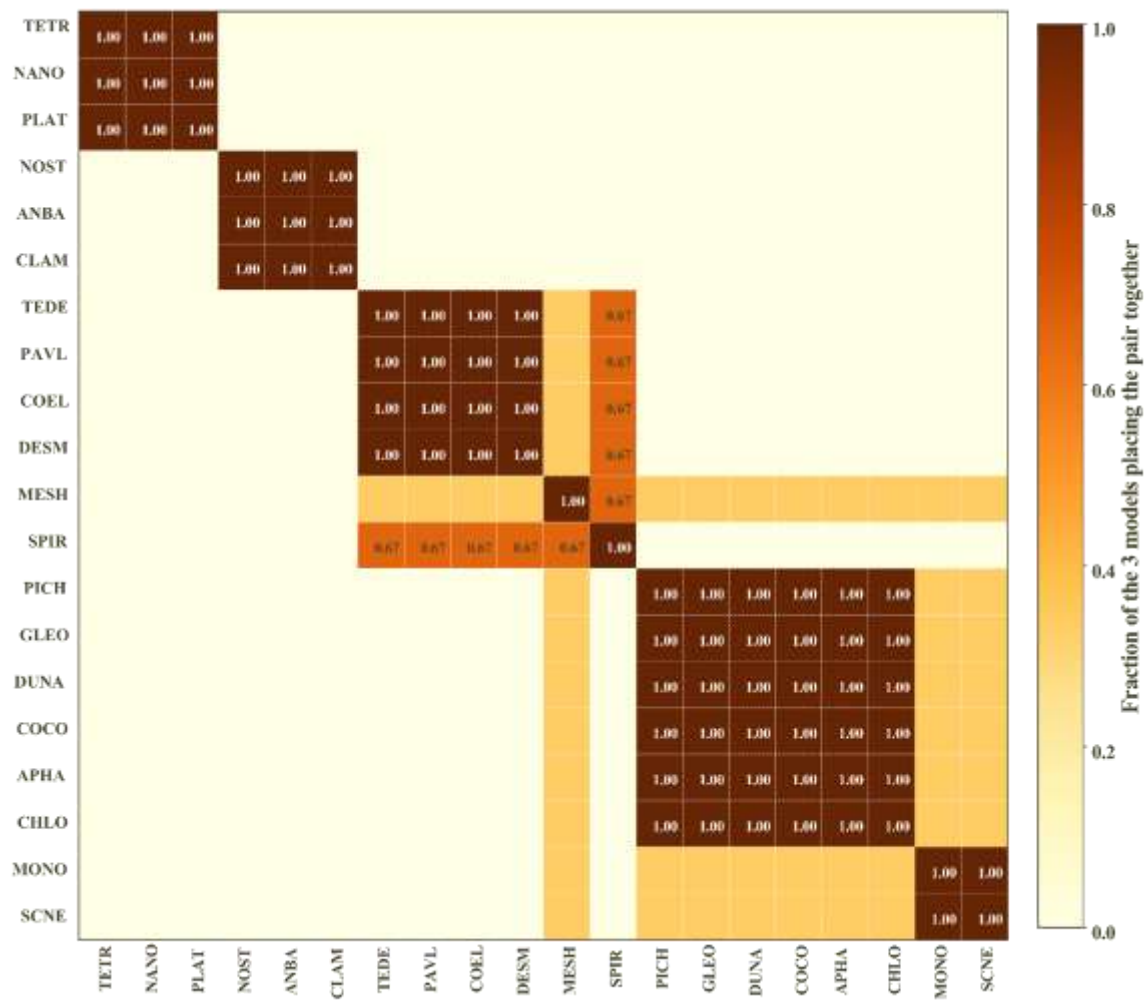


Fig S1: Species-by-species co-association matrix across the three clustering models, strains sorted by consensus cluster. Cell shade shows the fraction of models (0-1) placing each strain pair together; the block structure along the diagonal marks the 28 unanimous pairs, and lighter off-diagonal cells mark the 28 contested pairs (Section 3.2).

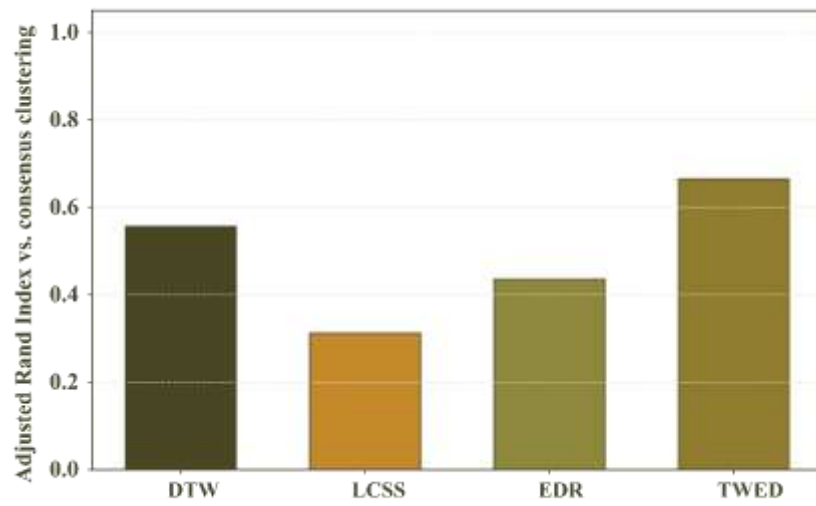


Fig S2. Adjusted Rand Index between the consensus partition and clusterings built from four elastic distance metrics (DTW, LCSS, EDR, TWED), reported as a direct empirical check on the DTW-vs-Euclidean caveat

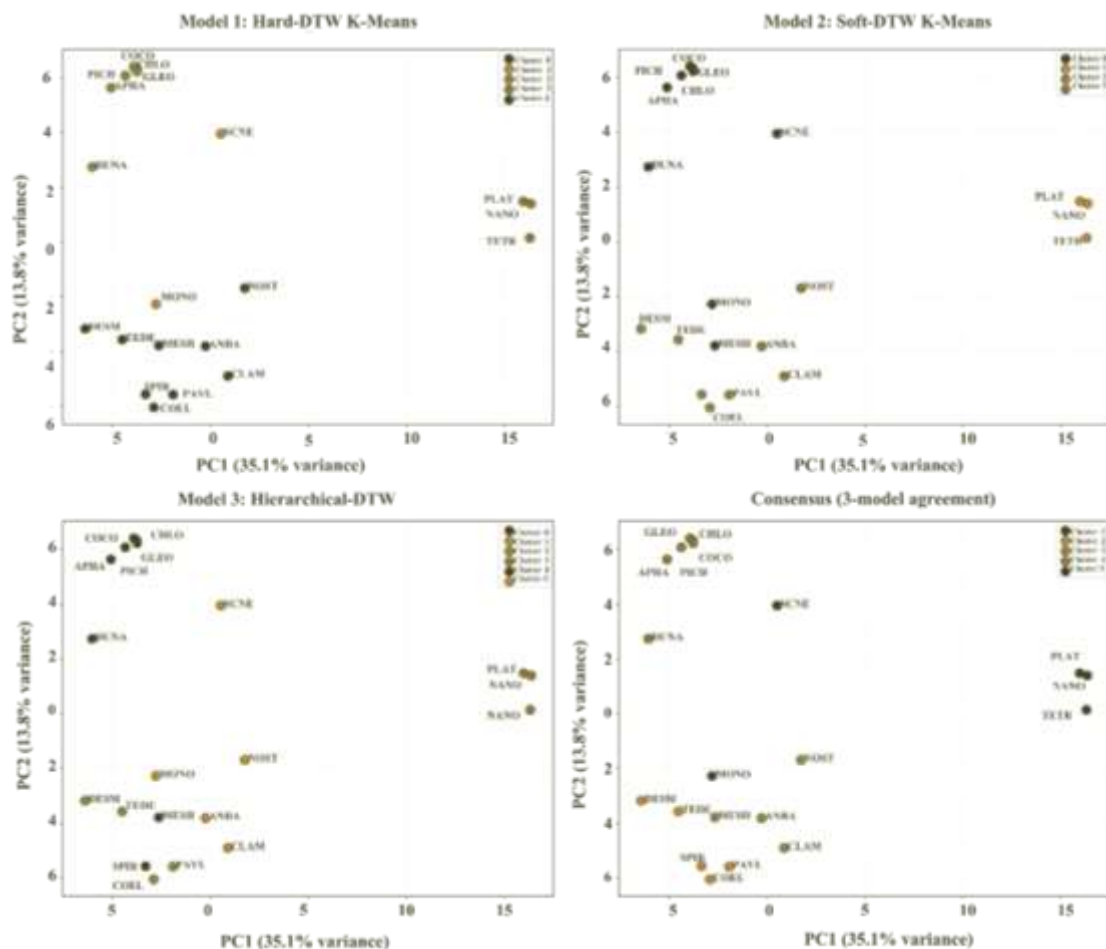


Fig S3. Two-dimensional PCA projection of all twenty strains' standardized trajectories (percent variance explained on each axis), with points colored by cluster assignment under (a) Hard-DTW, (b) Soft-DTW, (c) Hierarchical-DTW, and (d) the consensus partition, complementing the silhouette and Davies-Bouldin values in Table 3 (Section 3.1).