Supplementary Information

Sections:

1: Alternative averaging approach for calculating multifunctionality also indicates bicultures outperform monocultures.

2: Supplementary data for transcriptomic analyses determining differential expression of lipid genes.

3: Supplementary PCA data illustrating genomic distinctions between populations belonging to FAME overyielding versus non-overyielding bicultures.

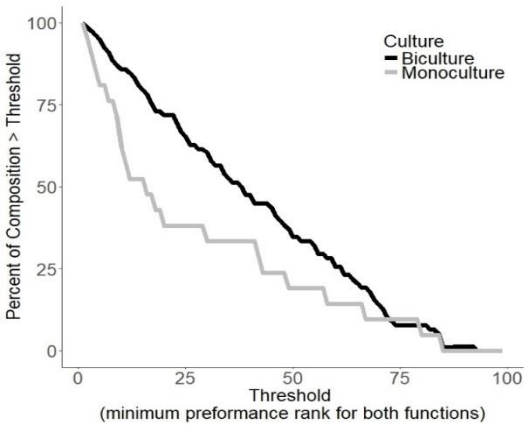
4: Shifts in expression at the gene-level caused by the diversity treatment significantly predict shifts metrics of biomass quality for bio-oil that depend on fatty acid composition, including Cetane Number, and Higher Heating Value.

5: Background experiments used for species and media selection, as well as growth curves of all treatments in the main experiment.

6: Biculture growing treatment frequently altered expression of genes regulating photosynthesis and nutrient assimilation.

<u>Section 1</u>: Alternative averaging approach for calculating multifunctionality also indicates bicultures outperform monocultures.

A)



B)

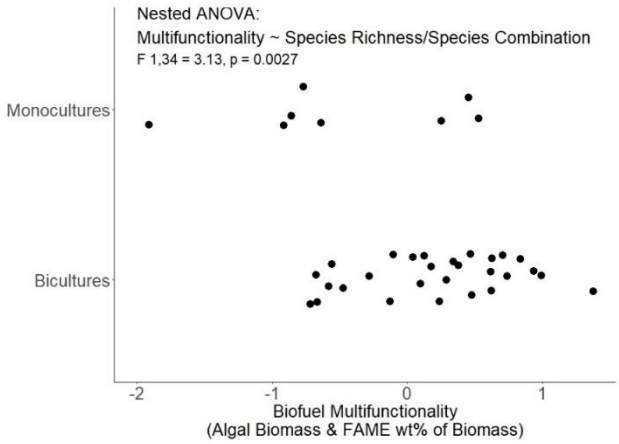


FIG S1 There are several approaches for calculating multifunctionality. (a) Here, we show that a higher percentage of bicultures versus monocultures simultaneously exceed thresholds for biomass yield and quality. (b) Additionally, we find that bicultures outperform monocultures via the averaging multifunctionality metric. We calculated this averaging multifunctionality index where FAME (wt % of samples) and Biomass (i.e., log transformed values of population cell density multiplied by estimates of cell biovolume) values were standardized via conversion to z-scores, and the FAME and Biomass standard scores were then averaged together to yield a multifunctionality value. Averaged multifunctionality index values across biological replicates are shown. Note the nested ANOVA uses all biological replicates with species combination nested within species richness.

Section 2: Supplementary data for transcriptomic analyses determining differential expression of lipid genes.

TABLE S1 Summary of species physiological and transcriptome data, including mean cell biovolumes, mean read abundance per sample for each algal species, transcript number per species before and after filtering by mapped taxon (within the Diaphoretickes), and presence/absence of functional annotation via Gene Ontology.

Species (order)	Species Biovolumes (μm^3)	Mean Library Size	Total Transcripts	Transcripts Mapped to 'Green' Taxa	GO Annotated Transcripts Mapped to 'Green' Taxa	GO Annotated Lipid Transcripts
Chlorella sorokiniana (Chlorellales)	61.46	8,951,025	31,452	9,560	5,892	168
Closteriopsis acicularis (Chlorellales)	511.24	6,880,563	37,466	11,334	6,766	200
Cosmarium turpinii (Desmidiales)	46330	5,324,166	58,333	14,196	10,691	381
Pandorina charkowiensis (Chlamydomonadales)	897.94	2,620,093	41,825	14,610	9,360	290
Scenedesmus acuminatus (Sphaeropleales)	244.99	8,729,722	50,367	12,214	7,584	263
Selenastrum capricornutum (Sphaeropleales)	98.72	3,554,743	32,085	10,696	6,663	223
Staurastrum punctulatum (Desmidiales)	7110.93	4,476,754	61,441	12,629	9,786	383
Tetraedron minimum (Sphaeropleales)	887.98	3,283,709	47,246	10,327	6,405	208

TABLE S2 Our species combination treatment had a reproducible effect on algal lipid gene expression, as described with a summary table of multi-dimension scaling plots that indicate similarity in gene expression via proximity of points. An MDS plots was constructed for each of the eight algal species on all 24 biological replicates across the monoculture and biculture conditions. Reported are pairwise euclidean distances among points (i.e. biological replicates) within the same treatment versus pairwise euclidean distances among points belonging to different treatments. For the majority of species, replicates within treatments were significantly more similar in gene expression than across treatments, as indicated via t-tests where asterisks indicate significance of $p < 0.05$). This same pattern was found when repeating this process for all transcripts, except *Selenastum* showed significantly more similar expression within treatment than across treatments.

	GO Annotated Lipid Transcripts		
Species	within	among	
Chlorella	0.82	1.2	*
Closteropsis	1.6	1.8	
Cosmarium	1.0	1.4	*
Pandorina	1.4	1.9	*
Scenedesmus	0.66	1.0	*
Selenastrum	1.0	1.2	
Straurastrum	0.67	1.5	*
Tetraedron	1.2	1.7	*

<u>Section 3</u>: Supplementary PCA data illustrating genomic distinctions between populations belonging to FAME overyielding versus non-overyielding bicultures.

TABLE S3 Variable loadings describe axes of PCA illustrated in Fig. 3. Bolding indicates Gene Ontology groups more heavily weighting PC2 and PC3, which best separated non-overyielding versus FAME overyielding bicultures. Only Gene Ontology terms shared across all eight species were used in PCA.

Gene Ontology Term	Function	PC1	PC2	PC3
GO:0000062	fatty-acyl-CoA binding	-0.012	-0.017	0.185
GO:0003865	**3-oxo-5-alpha-steroid 4-dehydrogenase activity**	-0.077	**-0.447**	**-0.425**
GO:0004149	**dihydrolipoyllysine-residue succinyltransferase activity**	-0.064	**0.130**	-0.136
GO:0004583	dolichyl-phosphate-glucose-glycolipid alpha-glucosyltransferase activity	0.038	-0.083	-0.022
GO:0004806	triglyceride lipase activity	-0.044	-0.065	-0.130
GO:0005543	phospholipid binding	-0.009	-0.163	0.084
GO:0006629	lipid metabolic process	-0.029	-0.143	0.099
GO.0006631	fatty acid metabolic process	-0.001	-0.107	-0.005
GO.0006633	fatty acid biosynthetic process	-0.079	-0.111	0.054
GO.0006635	fatty acid beta-oxidation	-0.010	-0.022	-0.081
GO.0006644	phospholipid metabolic process	-0.040	-0.413	0.255
GO.0006650	**glycerophospholipid metabolic process**	0.022	-0.219	**0.349**
GO.0006694	steroid biosynthetic process	-0.011	-0.063	0.271
GO.0008202	**steroid metabolic process**	-0.081	**-0.453**	**-0.444**
GO.0008289	lipid binding	-0.338	-0.004	0.109
GO.0008610	**lipid biosynthetic process**	-0.911	**0.135**	0.035
GO.0008654	phospholipid biosynthetic process	-0.013	-0.031	0.175
GO.0009107	lipoate biosynthetic process	0.029	-0.085	0.131
GO.0009245	**lipid A biosynthetic process**	0.011	-0.275	**0.321**
GO.0009249	protein lipoylation	0.068	0.058	0.026
GO.0009742	brassinosteroid mediated signaling pathway	0.003	0.089	-0.134
GO.0016042	lipid catabolic process	-0.101	-0.155	0.116
GO.0016125	sterol metabolic process	-0.032	-0.167	0.011
GO.0016992	lipoate synthase activity	0.056	0.013	0.062
GO.0019915	lipid storage	0.014	-0.162	-0.070
GO.0043550	regulation of lipid kinase activity	0.027	-0.134	0.212
GO.0044255	cellular lipid metabolic process	-0.035	-0.147	-0.092
GO.0051861	glycolipid binding	-0.060	-0.204	0.084

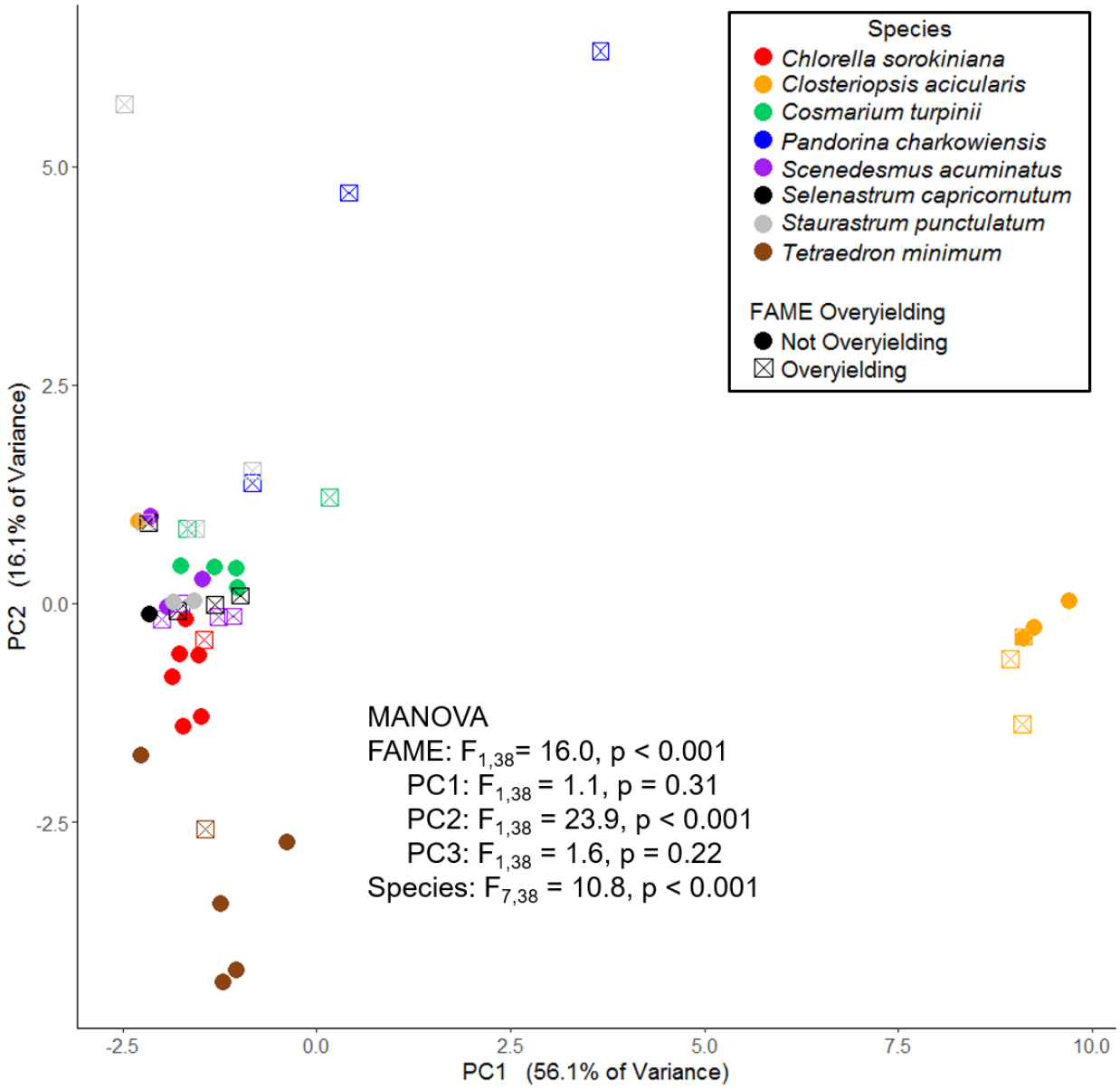


FIG S2 In addition to Fig. 3, here we show PC1 versus PC2 of our principal component analysis that incorporates mean $\log_2$FC values for each algal species of 28 different Gene Ontology groups involved in lipid regulatory functions. Note PC1 is largely differences across species, while PC2 and PC3 shown in Fig. 3 more clearly illustrate that bicultures overyielding versus non-overyielding in FAME production exhibit distinct patterns of gene expression.

<u>Section 4</u>: Shifts in expression at the gene-level caused by the diversity treatment significantly predict shifts in metrics of biomass quality for bio-oil, including FAME, Cetane Number, and Higher Heating Value

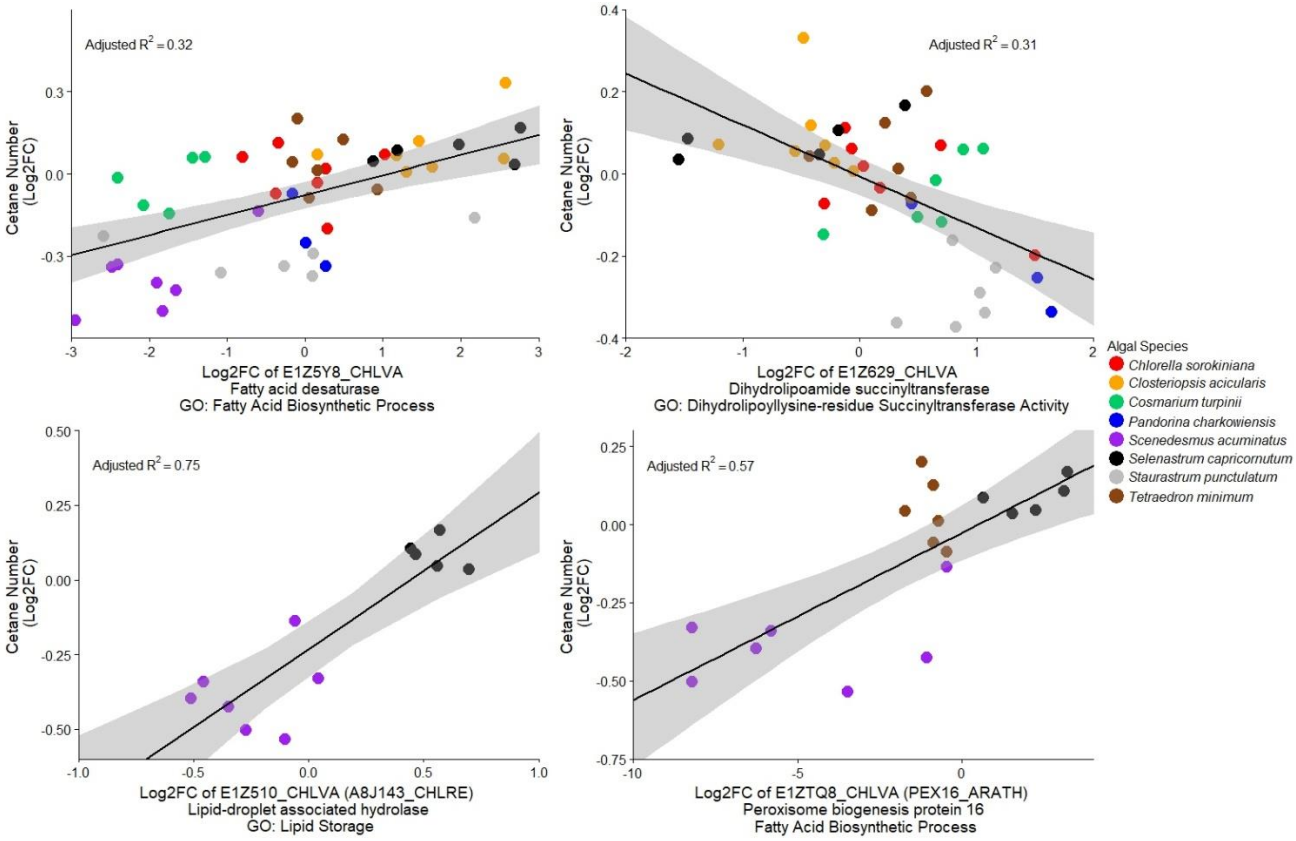


FIG S3 Across multiple species of algae, we found that the relative change in gene expression between monoculture and biculture growing conditions significantly predicts the relative change between the monoculture and biculture levels of Cetane Number, which was measured from the extracted fatty acid methyl esters. Correlation between $\log_2$ fold change of gene expression and Cetane Number in biculture relative to monoculture were evaluated via simple linear regression models with FDR significance corrections. Orthologs in *Chlorella variabilis* are used for gene annotation when possible. Occasionally, we note orthologs from other model organisms from which we needed to infer gene annotation.

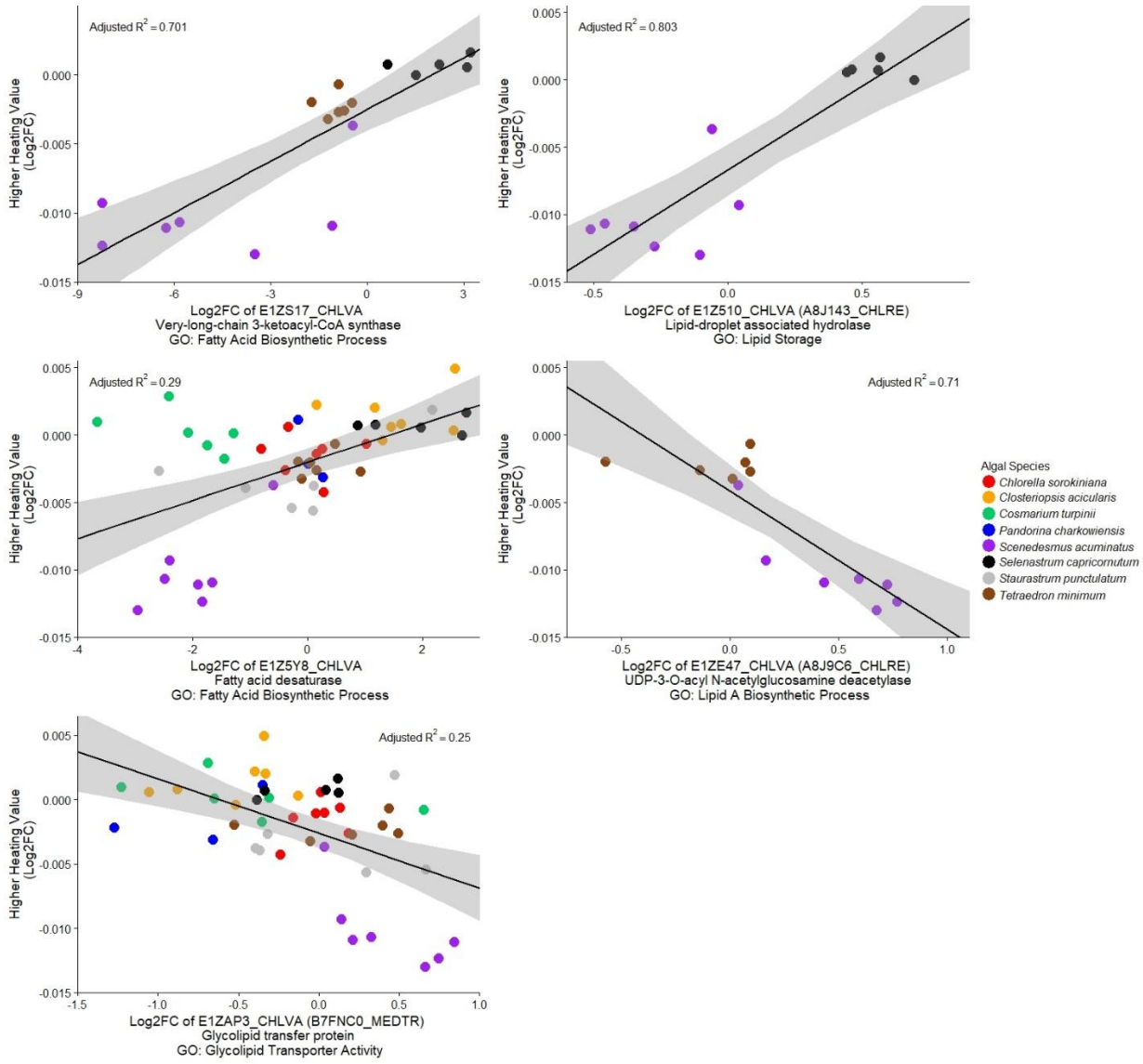
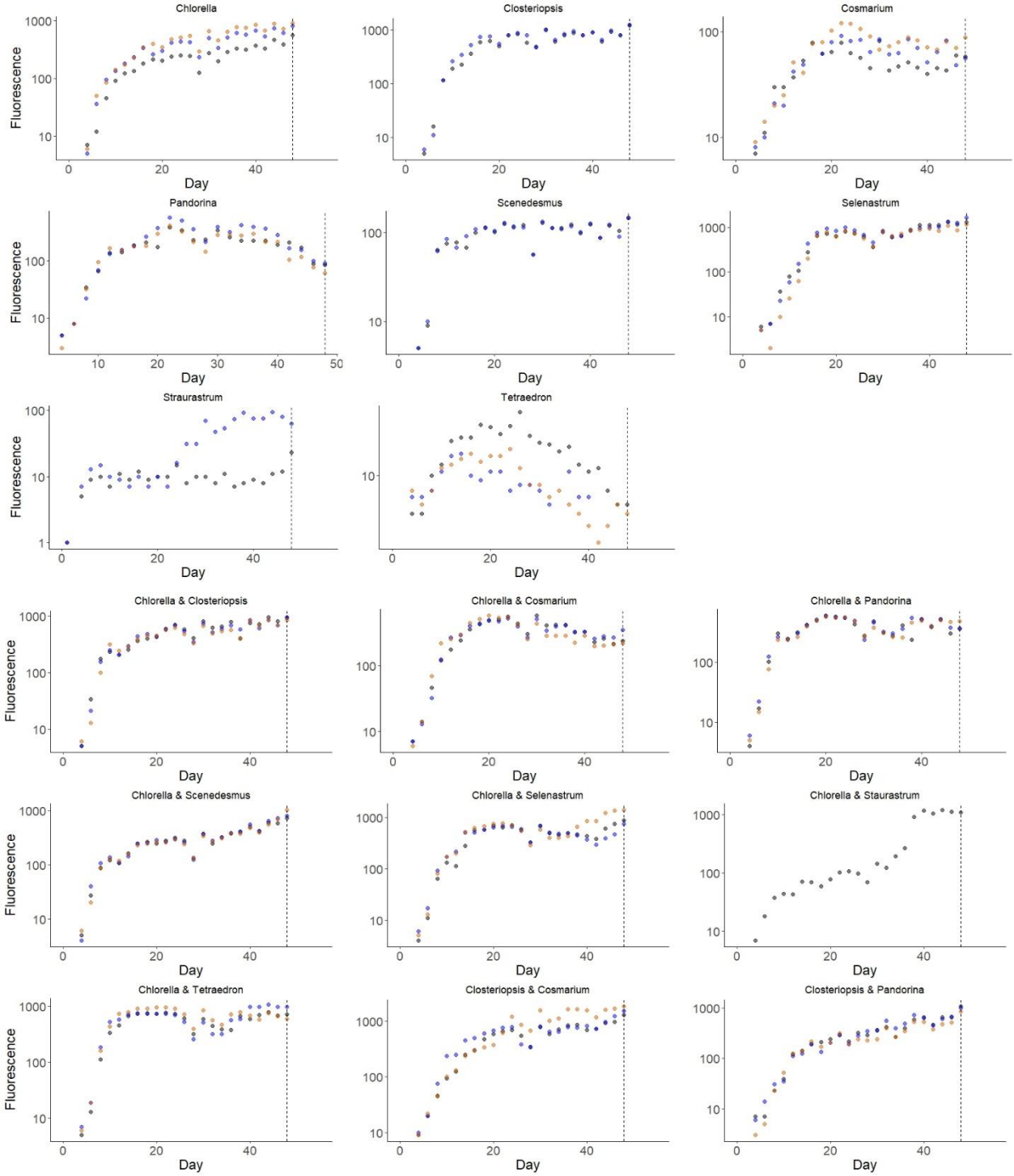


FIG S4 Across multiple species of algae, we found that the relative change in gene expression between monoculture and biculture growing conditions significantly predicts the relative change between the monoculture and biculture levels of Higher Heating Value, which was measured from the extracted fatty acid methyl esters. Correlation between $\log_2$ fold change of gene expression and Higher Heating Value in biculture relative to monoculture were evaluated via simple linear regression models with FDR significance corrections. Orthologs in *Chlorella variabilis* are used for gene annotation when possible. Occasionally, we note orthologs from other model organisms from which we needed to infer gene annotation.

1

Section 5: Background experiments used for species and media selection, as well as growth curves for all treatments in the main experiment.



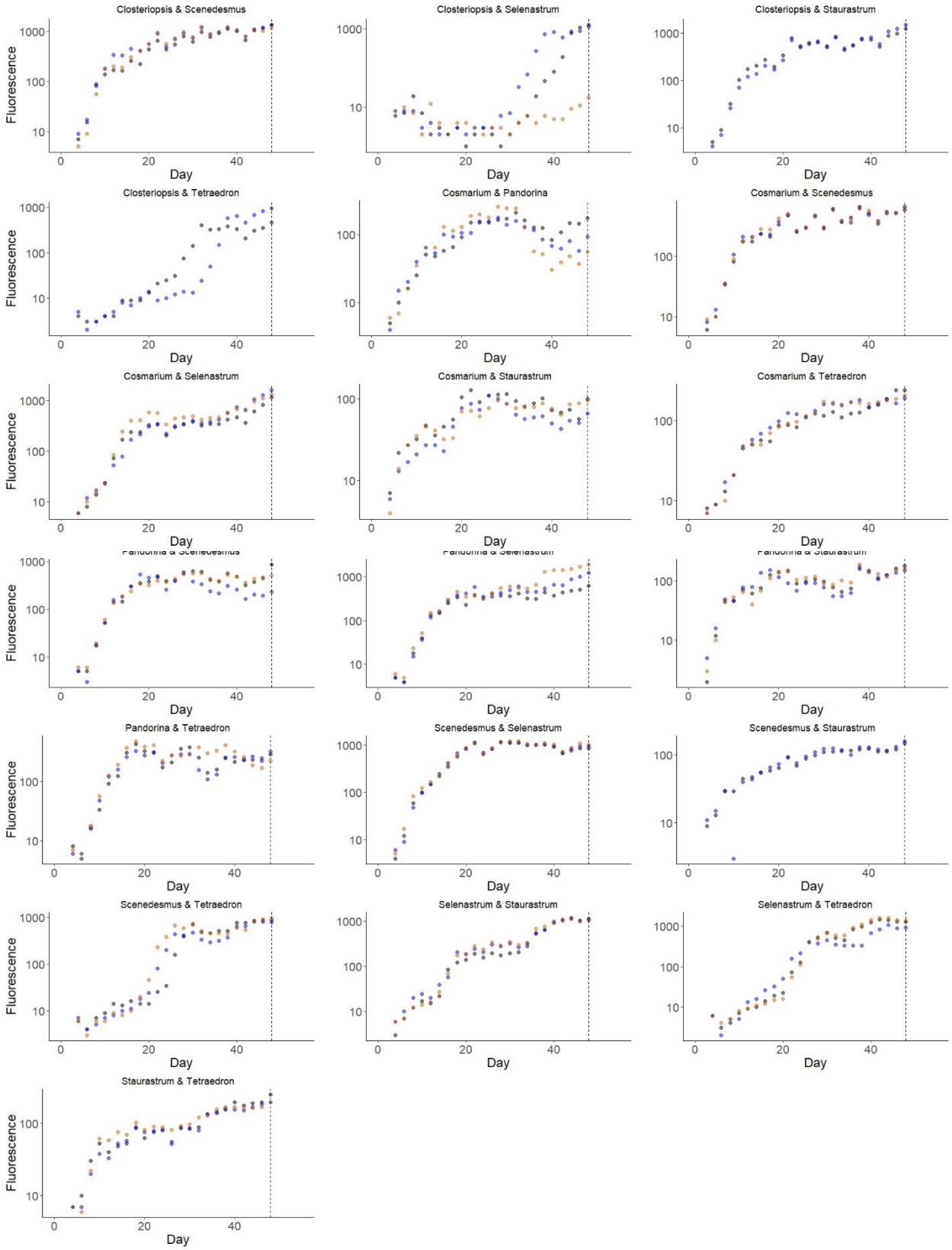

Closteriopsis & Scenedesmus
Closteriopsis & Selenastrum
Closteriopsis & Staurastrum
Closteriopsis & Tetraedron
Cosmarium & Pandorina
Cosmarium & Scenedesmus
Cosmarium & Selenastrum
Cosmarium & Staurastrum
Cosmarium & Tetraedron
Pandorina & Tetraedron
Scenedesmus & Selenastrum
Scenedesmus & Staurastrum
Scenedesmus & Tetraedron
Selenastrum & Staurastrum
Selenastrum & Tetraedron
Staurastrum & Tetraedron
Fluorescence
Day

FIG S5 Growth curves of all monocultures and bicultures using Chlorophyll-a fluorescence. Dashed vertical line at Day 48 indicates when algal biomass was collected for RNA sequencing. Color indicates replicate.

2

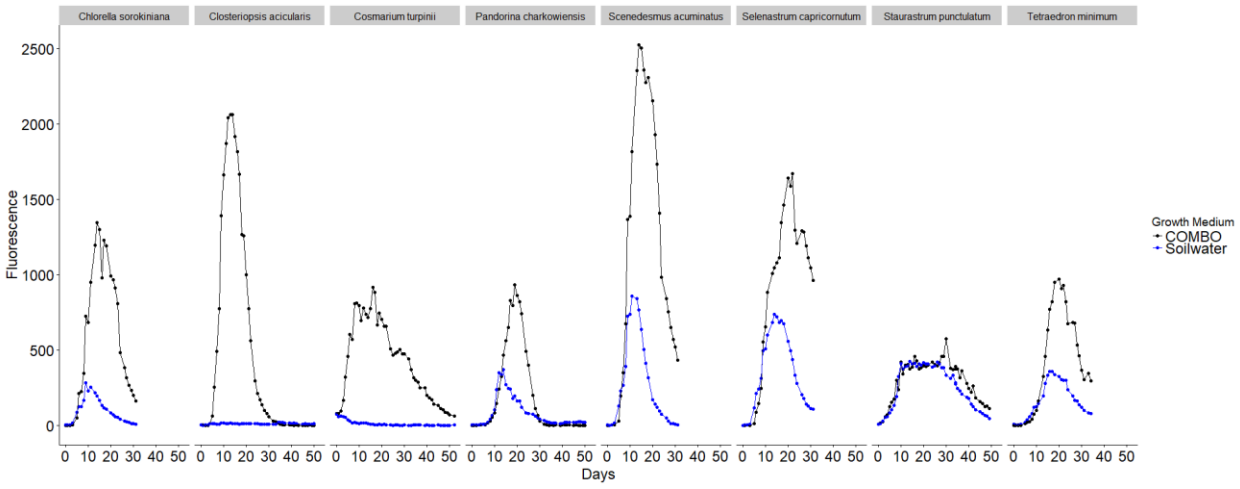


FIG S6 Growth curves of each study species in COMBO versus Soil Water Medium illustrates positive growth of each species in COMBO. Each curve is the average of 8 replicates inoculated into medium at 1000 cells/mL in 48-well plates incubated in 16:8h light chamber.

3

<u>Section 6</u>: Biculture growing treatment frequently altered expression of genes regulating photosynthesis and nutrient assimilation.

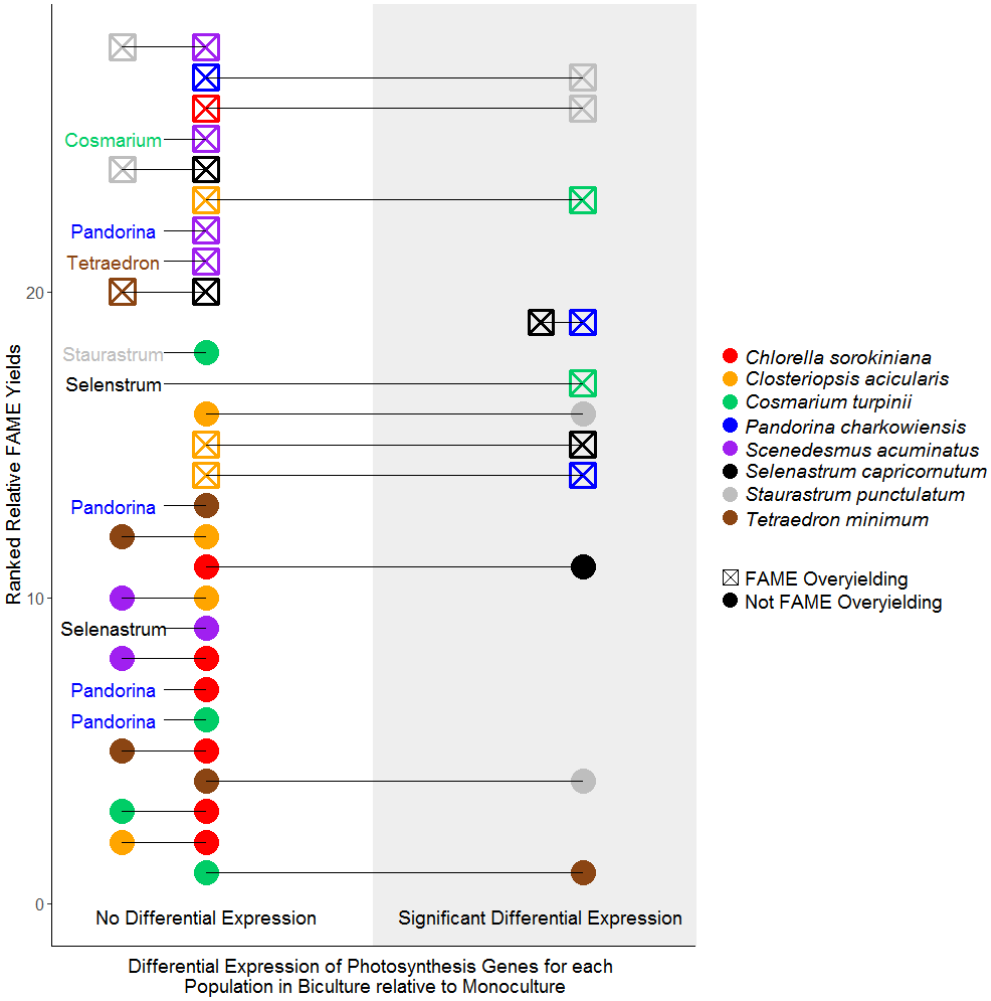


FIG S7 While algae frequently differentially expressed genes regulating photosynthesis in biculture compared to monoculture growing conditions, there was no systematic association of this differential expression with those bicultures that significantly overyielded in FAMEs. Solid lines connect populations growing together in biculture, symbol color indicates species identities in each biculture, and symbol fill indicates whether the biculture is overyielding in biomass quantity (i.e., wt% FAMEs). Location on the x-axis indicates whether those populations significantly differentially express their photosynthesis genes via Kolmogrov-Smirnov tests on the distribution of $\log_2$ fold change values of photosynthesis versus all other genes. Genes included in our photosynthesis group included those annotated as carbonic anhydrases, glutamate semialdehyde aminotransferases, and light harvesting chlorophyll a-b complexes.

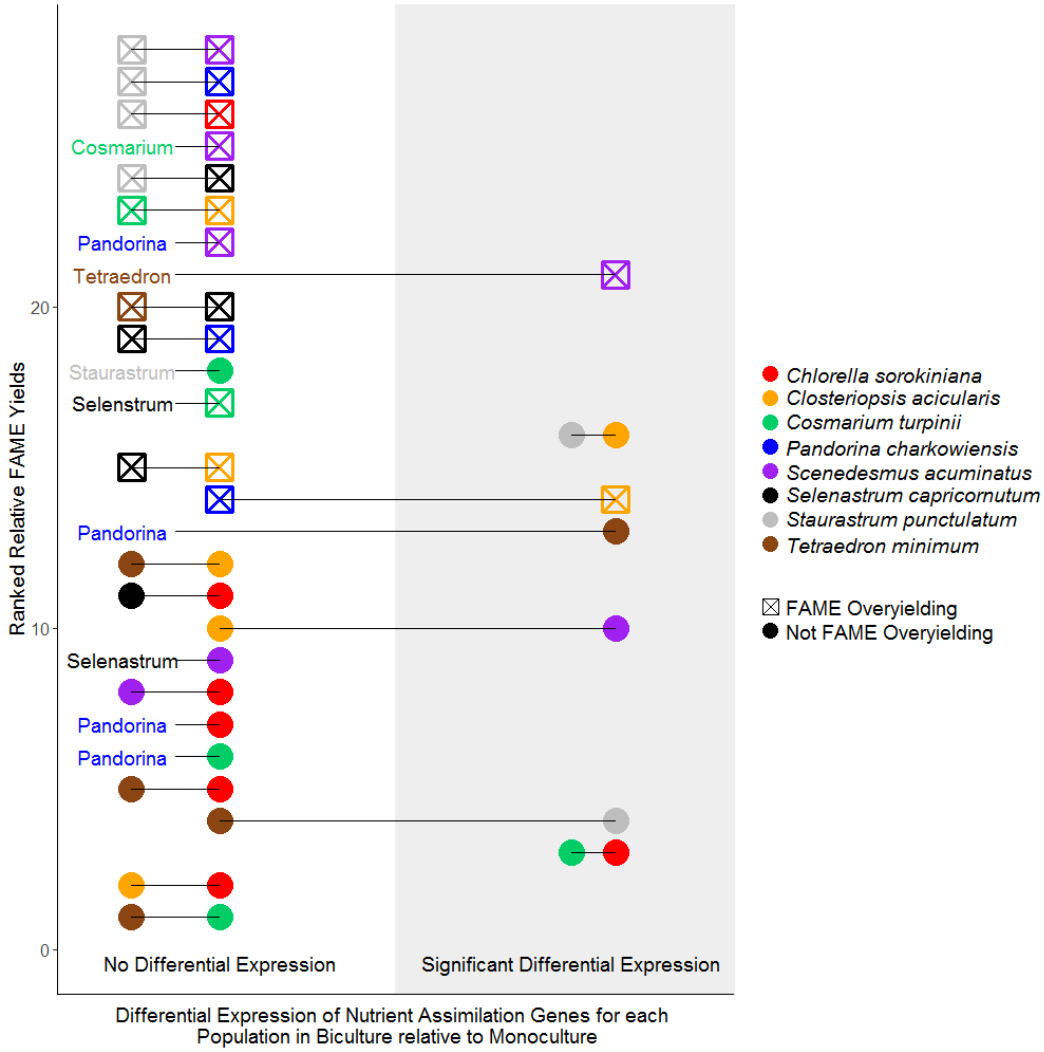


FIG S8 While algae frequently differentially expressed genes regulating nutrient acquisition in biculture compared to monoculture growing conditions, there was no systematic association of this differential expression with those bicultures that significantly overyielded in FAMEs. Solid lines connect populations growing together in biculture, symbol color indicates species identities in each biculture, and symbol fill indicates whether the biculture is overyielding in biomass quantity (i.e., wt% FAMEs). Location on the x-axis indicates whether those populations significantly differentially express their photosynthesis genes via Kolmogrov-Smirnov tests on the distribution of $\log_2$ fold change values of photosynthesis versus all other genes. Genes included in our nutrient assimilation group included those annotated as iron permeases, nitrite transporters, nitrite reducers, nitrate transporters, nitrate reducers, and phosphate transporters.

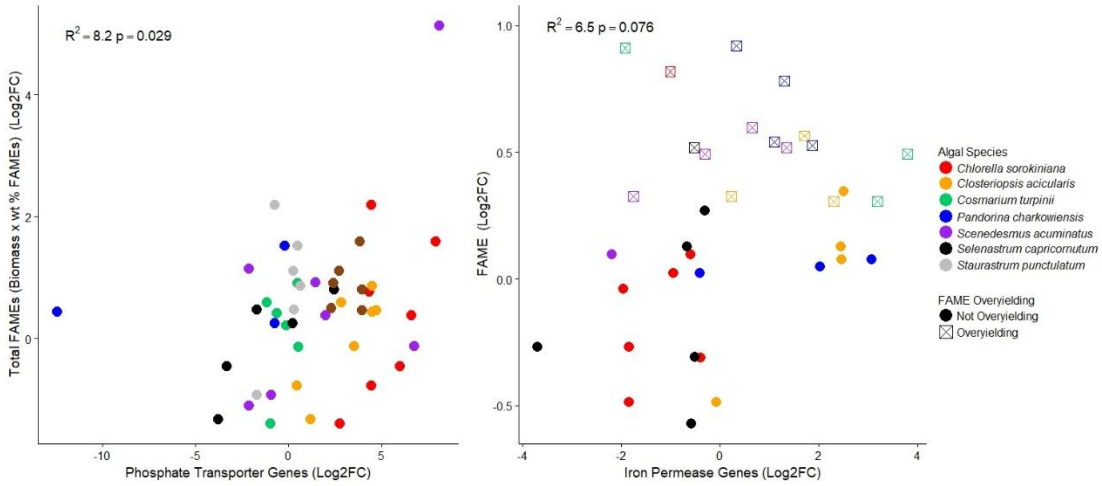


FIG S9 The magnitude of differential expression of genes regulating nutrients assimilation and photosynthesis in biculture relative to monoculture did not significantly predict the magnitude of either FAME overyielding or total production of FAMEs (i.e., biomass multiplied by the weight percentage of FAMEs). Genes tested include: carbonic anhydrases, glutamate semialdehyde aminotransferases, light-harvesting chlorophyll a/b complexes, iron permeases, nitrite transporters, nitrite reducers, nitrate transporters, nitrate reducers, and phosphate transporters. When multiple transcripts within a species were annotated for a single function, i.e. phosphate transport, we use only the transcript with the greatest absolute value log2FC for each treatment. Note we illustrate results from two gene groups with the strongest results, but neither trend remains significant after false discovery rate correction.