Additional Files 1: Supporting Information

From mechanism to application: decrypting light-regulated denitrifying microbiome through geometric deep learning

Yang Liao[a], Jiyong Bian[a], Jing Zhao[a], Ziwei Zhang[b], Siqi Xu[a], Yijian Qin[b], Xuan Luo[c], Shiyu Miao[a], Rui Li[a], Ruiping Liu[a,*], Meng Zhang[d], Wenwu Zhu[b], Huijuan Liu[a], Jiuhui Qu[a].

[a] Center for Water and Ecology, State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing 100084, China

[b] Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China;

[c] Weiyang College, Tsinghua University, Beijing 100084, China

[d] School of Electronic and Information Engineering, Beihang University, Beijing, 100191, China

*Correspondence: rpliu@tsinghua.edu.cn

This file includes:

Supporting text

Figures S1 to S15

Tables S1 to S5

Legends for Datasets S1 to S7

Supplementary References

Other supporting materials for this manuscript include the following:

Datasets S1 to S7

Supporting Information Text

Supporting Methods

Fluorescence assay. Bacterial Viability Detection Kit (DOJINDO) was used for cell staining. Briefly, before staining, the cell density of bacterial suspension was adjusted to similar level and washed by PBS (0.01 M, pH = 7.4) twice. After that, we used CFDA, a coloring reagent that can fluoresce when binding to intracellular esterase, indicating esterase activity and cellular viability. PI, a coloring reagent that can combine with nucleic acid, was utilized to indicate dead cells. Stains and cell suspension were vortexed to mix well and follow the instruction for further sample pretreatment. The fluorescence images of stained cells were shotted by confocal laser scanning microscopy (CLSM, Zeiss LSM880). The CLSM data were further processed by ZEN 3.1 (blue edition) to obtain 2.5D images to enhance the resolving power of cellular physiology after photo-denitrification under different illumination conditions.

RNA extraction and metatranscriptomic sequencing. After photo-denitrification, cell suspensions were collected for RNA extraction. Total RNAs were extracted with the E.Z.N.A.® Soil RNA Midi Kit (Omega Bio-tek, Norcross, GA, U.S.) following the manufacturer's instruction. The RNA concentration and purity were quantified with NanoDrop2000 (Thermo Fisher Scientific,U.S.). RNA quality assessment was conducter with a RNA6000 Nano chip (total RNA) in an Agilent 2100 Bioanalyzer, and further determined by RNA integrity number (RIN). Ribo-zero Magnetic kit was used for rRNA removal followed the manufacturer's instruction (Epicentre, an Illumina® company). cDNA libraries were constructed using TruSeq™ RNA sample prep kit (Illumina). The barcoded libraries were paired-end sequenced on the Illumina Hiseq 2500 platform at Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China), using HiSeq 4000 PE Cluster Kit and HiSeq 4000 SBS Kits.

Metatranscriptomic data quality control and genes expression quantification. To obtain high-quality reads, SeqPrep (https://github.com/jstjohn/SeqPrep) and Sickle (https://github.com/najoshi/sickle) were utilized to strip 3' and 5' ends and remove low-quality reads (length<50 bp or quality value <20 or having N bases), respectively. rRNA reads were removed using SortMeRNA [1] aligning to the SILVA 128

version database. Clean reads were assembled by Trinity [2] and the longest transcripts were extracted through python scripts as unigenes. Then, open reading frames (ORFs) prediction were performed with TransGeneScan [3] and sequences with 95% identity were clustered as the non-redundant gene catalogues by CD-HIT [4]. The longest sequence was utilized to represent the gene catalogue. After that, reads were mapped to representative genes and gene expression was quantified in the form of Fragments Per Kilobase of transcript per Million fragments mapped (FPKM) that calculated by RSEM [5] to eliminate the disturbance caused by genes length and sequencing depth.

Gene functional annotations. Taxonomic annotations were carried out through aligning non-redundant gene catalogs against NR database with by BLAST [6]. Furthermore, genes were aligned against eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) and KEGG database (Kyoto Encyclopedia of Genes and Genomes) to obtain functional information via Diamond (http://www.diamondsearch.org/index.php, version 0.8.35). The subcellular information, including signal peptide and transmembrane domain, were predicted using Diamond against SignalIP 6.0 database (https://services.healthtech.dtu.dk/service.php?SignalP) and TMHMM/2.0 database (http://www.cbs.dtu.dk/services/TMHMM/), respectively. All the e-value cutoff was set as $1e^{-5}$.

Spatial distribution and enrichment analysis on significant functional pathways. We conducted enrichment and function analysis based on KEGG database on critical clusters and functionally co-expressed genes. We projected the distribution of critical clusters on 2D tSNE space to decipher the overall expression pattern. To identify essential pathways in functional genes panels, we annotated these genes to KEGG level 3 pathways and summarized the expression pattern in response to blue and yellow light. For blue light, pathways with p-value < 0.01, fold change < 0.5 or >2 and expression level < 10 FPKM were selected as significant functional pathways. For yellow light, we adjusted the threshold of p-value and expression level to select about top 10 significant pathways.

Demonstrating the effects of nitrate on superoxide level. Gradient nitrate concentrations (0, 100, 500, 1000 mg N L^{-1}) batches were employed to investigate the effects of nitrate concentration on cellular total ROS production and superoxide level. After activation, microbiota were incubated with and without nitrate for 15~20 hours, at the end of which most of the nitrate was depleted due to bio-denitrification under

yellow and dark conditions. All gradient experiments were performed under dark conditions to control nitrate to be the only variable, thus proving nitrate reduction, superoxide production and some light-sensing genes were in a co-expression genes module. Photo-denitrification with and without nitrate under yellow and blue light conditions was carried out to investigate the "bio-quenching" effects of nitrate on cellular superoxide. The initial nitrate concentration was 500 mg NO_3-N L^{-1}. Detailed implementations were similar to the gradient experiments as mentioned above.

Quenching experiments. To investigate the contribution of superoxide on nitrate removal, we conducted quenching experiments on nitrate under dark, blue and yellow light. The variation of nitrate concentration with different illumination conditions after incubation with quenching agents was used to quantify the quenching ratio of different ROS. Dimethyl sulfoxide (DMSO) was employed as ·OH scavenger, and L-histidine for $^1O_2$, SOD for $\cdot O_{2^-}$, respectively. 1 mM of quenching regents were supplemented to reactors at the start of photo-denitrification. After about 14 hours, nitrate concentrations were measured to calculate quenching ratios. Quenching ratio (QR) was calculated as follow:

$$A = \frac{\frac{c_{N0}-c_{Nt}}{c_{N0}} - \frac{c_{q0}-c_{qit}}{c_{q0}}}{\frac{c_{N0}-c_{Nt}}{c_{N0}}} \times 100\%, \quad (1)$$

where c_{N0} (mg NO_3-N L^{-1}) and c_{qi0} (mg NO_3-N L^{-1}) were the initial nitrate concentration of control and quenching group, c_{Nt} (mg NO_3-N L^{-1}) and c_{qit} (mg NO_3-N L^{-1}) were the nitrate concentration of control and quenching group during nitrate reduction.

Analytics. During photo-denitrification, cell suspensions were periodically sampled by syringe for further assays. After filtered by 0.22 μm micropore filter, nitrate, nitrite and acetate concentration [5,39]were assayed through iron chromatography (Thermo Fisher Scientific, Integrion ICS-5000). All UV-vis spectra and absorbance were obtained through microplate reader (Thermo Fisher Scientific, TENCAN-Spark). Protein concentrations were quantified by the Micro BCA Protein Assay Kit (Thermo Fisher Scientific Inc., USA).

Statistics and visualization. All statistical analysis and bioinformatics visualization were performed through R (4.2.2). The results of experiments are shown as means ± s.d. Unsupervised learning and model evaluation were conducted through Python. Data visualization was performed in R. Some elements and

components for schematic illustrating were downloaded from Servier Medical Art (https://smart.servier.com/), licensed under a Creative Commons Attribution 3.0 Unported License.

Supporting Discussions

Discussion S1. Exploratory analysis on datasets. Light irradiation activated transcriptional activities, increasing transcripts number by 137.4% for blue light and 17.9% for yellow light (**Table S1**). Compared to yellow light, blue light triggered a higher number of differentially expressed genes (DEGs), as well as more substantial and significant transcriptional variance with higher p-value (**Fig. S3a, b**). Among 56,991 genes, blue light induces more differentially expressed genes (DEGs) (36,373 genes) than yellow light (6,293 genes), and more genes were up-regulated after light illumination, indicating that blue light triggered metabolism fluxes redirection to some rare metabolism pathways. This could attribute to the activation of genetic parts or components, such as promoters, ion channels, pumps etc. [7] This was also supported by the volcano plots (**Fig. S3b**) that blue light displayed notable transcriptional changes and most of the top 10 highly-expressed genes were down-regulated. Whereas most of the highly-expressed genes of yellow groups remained minor changes.

Dimension reduction on samples under different light exposure conditions revealed the light-induced phenotype divergence (**Fig. S3c**). The expression profiles induced by blue light clustered far apart from dark and yellow light groups in both 2D and 3D PCA. Yellow light groups exhibited similar PC1 and PC2 with dark, which possessed cumulative 64.6% and 94.3% of variance. Whereas divergence presented in PC3 which accounts for 2.9% variance, implying that yellow light also induced slight differences in the whole gene expression profile. Dimension reduction on genes through T-distributed stochastic neighbor embedding (tSNE) depicted the spatial distribution of valid DEGs and the light-dependent clustering was evident, indicating the light-induced genetic co-expression (**Fig. 2*b***).

Light-induced metabolism changes can be attributed to gene co-expression, i.e. expression levels of photo-responsive genes could impact the expression of genes in the same gene panel through the gene regulatory network [8,9]. Orthologous groups and functional annotations of database are commonly used to define co-expressed gene panels [10]. We annotated the valid genes to eggNOG database and labeled the

eggNOG classification in tSNE space. No obvious gene cluster was formed, indicating that eggNOG failed to capture the light-induced co-expression (**Fig. S4**). Moreover, the lack of annotation, denoted by NA, was another bottleneck in decrypting the optogenetic regulatory mechanism. The above results showed that the prior-knowledge-based methods cannot apply to all situations. Therefore, contextually customed models were needed [11].

We made subcellular information annotations on all genes as described in Material and Methods to supplement the biological knowledge for modeling. There were 1,285 genes predicted to encode secretory protein (**Fig. S5a**). Blue light contributed to a larger proportion of genes (4.93% DEGs) that encoded secretory protein compared to yellow light, whereas yellow light triggered the expression of transmembrane protein (**Fig. S5b**).

Discussion S2. Determination of gene panels and results evaluation. To better define co-expression gene panels in accordance with biological meaning, we need to set appropriate cluster number. Too large cluster number might separate co-expressed functions into sub-clusters, while too small value might bring about multi-functions in one cluster. We compared DGI and K-means with and without subcellular information to determine the number of co-expressed gene panels (**Fig. S7, Fig. 3*c*, *d***). Hierarchical clustering (HC) was not taken into consideration due to its severe bias (**Fig. 3*a*, *b***). We compared cluster number of 7, 10 and 24. It can be observed that DGI model performed better than K-means regardless of cluster number. We set cluster number as 7 for further analysis given the better distinguished capability as regard to both clustering approaches and heterogenous information integration.

The Silhouette Coefficient Index (SCI), an inter-cluster similarity indicator, dropped dramatically when combining expression matrix and subcellular information. For blue light, the integration of subcellular information significantly improved the clustering performance. While yellow light presented no significant improvement, which could be attributed to less pronounced expression variance and smaller dataset size [12].

We defined functional assignment score (FAS) to evaluate the biological function matching level quantitatively. We compared the FAS of pathways closely related to light, including oxidative stress and optogenetic switches. Longevity regulating pathway and peroxisome involve in oxidative stress resilience. Cytochrome P450 (CYP450) is responsive to light-wavelength at 450 nm. Metabolism of xenobiotics by

Cytochrome P450 and peroxisome exhibited more effective enhancement by DGI model and integration of subcellular information. Xenobiotics were chemicals that involve in the inter-species signaling and substance exchanges [13]. Generally, DGI model outperformed K-means and the integration of subcellular information assisted identify biological functions (**Fig. 2*e***). FAS of yellow light groups were higher than blue light, which was attributed to the smaller size of the yellow light dataset so that genes tended to co-express. With the increase of gene counts, pathways would contain more hub gene nodes in metabolism network that involve multi-functions with divergent expression patterns. Nitrogen metabolism was a good case in point [14]. The various expression patterns well supported it (**Fig. S6b**). The divergent performance of blue and yellow light was attributed to the size of datasets that larger datasets contain more hub genes and pathways that involve multi-functions with divergent expression patterns, such as nitrogen metabolism [14]. The various expression patterns well supported it (**Fig. S6b**).

Discussion S3. Detailed functions analysis on highly-expressed pathways of HGPs in enrichment analysis. Aging was shared by the HGPs of both blue and yellow light (**Fig. 3*c*, *d***). Aging in bacteria refers to the process of senescence, where microorganisms undergo a decline in viability and fitness over time. This could be attributed to the accumulation of cellular damage over time, such as DNA damage, oxidative stress, and protein misfolding, as well as regulated by genetic and environmental factors like nutrient availability, stress responses and quorum sensing [15]. The representative pathway in KEGG Brite Aging is the longevity regulating pathway, a set of genes associated with signaling, oxidative stress and protein synthesis that contribute to the collective response to promote cellular fitness and ultimately longevity.

For other pathways in blue light's HGP, signaling transduction, another highly-expressed KEGG Brite, facilitated the signaling process needed for aging. Specifically, FoxO signaling pathway includes genes that involve in the regulatory network of longevity, oxidative stress resistance and cell-cycle control, so it is also relevant to ROS [16]. MAPK signaling pathway is another signaling transduction pathway that was essential in cell signaling, proliferation, growth and division [17]. Besides ROS-related pathways, the expression levels of Inositol phosphate metabolism and streptomycin biosynthesis were also striking. The inositol phosphate signaling network is central to nutrient responses and coordinates cellular responses to nutrient uptake and utilization from growth factor signaling to energy homeostasis [18]. Moreover, the presence of myo-inositol can promote the biosynthesis of streptomycin as an intermediate [19].

Streptomycin is a broad-spectrum antibiotic, mainly produced by Streptomyces griseus to compete with other bacteria [20]. Therefore, those two pathways play pivotal roles in bacterial virulence, proliferation, apoptosis and cross-species competition.

As for other pathways in HGPs, blue light was featured with more pathways that involved in ROS production and cellular antioxidant capability than yellow light. Generally, most of disease-related pathways involve in the generation of ROS, which explained the presence of Chemical carcinogenesis – reactive oxygen species, Lipid and atherosclerosis, Pertussis, Legionellosis and Shigellosis for blue light, Tuberculosis for yellow light. Corresponding to the ROS-related pathways in HGPs of blue light, there were also large number of antioxidant pathways. Peroxisome is the principal pathway in the antioxidant system, containing a set of antioxidant enzymes that are vital in converting ROS into safer molecules. Glutamatergic synapse and GABAergic synapse involved in glutathione metabolism, which was a critical antioxidant compound.

In contrast, the rest pathways of yellow light's HGP were mostly for protein synthesis, such as chaperone and folding catalysts, exosome, messenger RNA biogenesis and secretion system. Exosomes are small extracellular vesicles secreted and played a role in intercellular communication, such as transmitting signals and molecules and remodeling the extracellular matrix [21], capable of carrying a variety of cargos like proteins, lipids and genetic materials [22]. The significant enrichment and expression levels of exosomes and Chaperones and folding catalysts, along with secretion system, suggested that the yellow light triggered highly-active proteins secretion for microbial interactions. The activated glyoxylate and dicarboxylate metabolism corresponded to the activated acetate utilization (**Fig. 1A**).

Discussion S4. Detailed functions analysis on highly-expressed pathways of SGPs in enrichment analysis. Among the SGP of blue light (**Fig. S11a**), Carbapenem biosynthesis, porphyrin metabolism and sulfur relay system were dominant pathways with relatively high expression (**Table S4**). Porphyrin metabolism involves intermediate metabolites in the production of vital molecules such as heme, chlorophyll coenzyme F430, cobalamin (vitamin B12 coenzyme). These substances were crucial cofactors for basic cellular metabolism. Folate (vitamin B9) is also one of them, explaining the upregulation of folate biosynthesis. Correspondingly, sulfur relay system was also upregulated to facilitate the biosynthesis of those cofactors through signaling messengers like Ubiquitin and ubiquitin-like proteins (Ubls).

Additionally, both sulfur relay system and carbapenem biosynthesis play essential roles in protein synthesis, folding and modification. Besides those dominant pathways, other pathways were mostly related to inter-cellular interactions, such as signaling, substances exchanges and aggregations. Terpenoid is a group of active substances with diverse biological functions, whose mostly well-known application is medicine. Galactose metabolism is essential for the formation of biofilm, which contained exopolysaccharides (EPS) and can assist bacteria to be more resistant to antibiotics like streptomycin [23], corresponding to the presence of Exopolysaccharides biosynthesis, streptomycin biosynthesis. This was because blue light caused stress on microbiota and only species capable of secreting EPS for protection and producing active or antibacterial substances to compete with other species could survive [24,25]. Correspondingly, secretion system of blue light was also presented in SGP, playing a role in cellular protection, inter-cellular signaling and substance exchange. To sum up, most of SGP's pathways of blue light were associated with inter-species signaling for survival under environmental stimulation.

As for yellow light, sulfur metabolism exhibited strikingly high enrichment (**Fig. S11b**), contributing to energy metabolism of phototrophic organisms [26]. In contrast, the expression level of yellow light's SGP was significantly lower compared to blue light, and ROS-related pathways took the majority, including Chemical carcinogenesis – reactive oxygen species, Cardiac muscle contraction and ascorbate and aldarate metabolism. Ascorbate and aldarate metabolism is a biochemical pathway involved in the breakdown and synthesis of ascorbic acid (vitamin C) and related compounds. These compounds play a crucial role in antioxidant capability. It is reported that exposure to light can increase the ascorbate level [27]. The activated ROS-related pathways implied that light-induced ROS might serve as signals or facilitate signal transduction to regulate diverse microbial metabolism [28]. Other two pathways with relatively high expression level were insect hormone biosynthesis and Limonene and pinene degradation, which involve signaling related to the metabolism of fatty acids, isoprenoids and cytochrome P450 (CYP15A1) [29,30]. The rest significant pathways with high fold changes and considerable expression levels were mainly related to xenobiotics and amino acid metabolism, i.e. Chloroalkane and chloroalkene degradation, Phenylalanine, tyrosine and tryptophan biosynthesis, etc. Chloroalkane and chloroalkene degradation is the second largest xenobiotic biodegradation pathway [31]. This indicated that yellow light might promote the

synthesis of macromolecules like secretory proteins, corresponding to the up-regulated protein synthesis activities in HGP (**Fig. 4*d***).

Discussion S5. Landmark genes analysis distinguished high-resolution co-expression for regulation strategy development. We defined landmark genes as the top three highly-expressed genes among the modularity class. We searched the modularity classes that contained PDGs, including narG (g_51049, P09152, nitrate reductase), narK1 (g_19066, Q9RA46, nitrate/nitrite transporter) and nirK (g_06597, P25006, nitrite reductase), and conducted subnetwork co-expression analysis as follows.

For blue light, both narK1 and nirK belonged to class 0, whose landmark genes were Myo-inositol-1-phosphate synthase (MIPS), nitric oxide reductase subunit B (NorB) and isocitrate dehydrogenase (IDH) (**Dataset 5, Fig. 5*b***). MIPS is an important enzyme involved in the biosynthesis of phosphatidylinositol and phosphoinositides, and plays important roles in signal transduction, cellular growth, etc [32]. NorB was the enzyme for next step of denitrification after nitrite reductase and co-expressed with narK. IDH plays a crucial role in tricarboxylic acid cycle (TCA cycle), which is essential in energy production and cellular metabolism. Therefore, nitrite reduction and its transport are highly associated with the energy supply and can be induced by some active substances like inositol. Nitrate reductase narG belonged to class 3, characterized by superoxide dismutase (SOD), ribosomal protein and 4-hydroxy-tetrahydrodipicolinate synthase (DapA). DapA involves in lysine biosynthesis, which is also a precursor to many proteins and serves as a building block for glutamate, playing a role in cellular antioxidant capability [33]. Superoxide dismutase was the top landmark gene, as well as vital in ROS scavenging and signaling, indicating superoxide's association with denitrification.

For PDGs of yellow light datasets, only narK1 presented in HGP but without any connection with other nodes (**Fig. S12**). The expression levels of molecular chaperone, the landmark genes of yellow light's HGP, were striking (**Dataset S6**), indicating that yellow light centralized the metabolism fluxes for protein synthesis. LuxR family is a DNA-binding transcriptional regulator that can regulate the expression of plenty of crucial physiological events, such as virulence factor production, biofilm formation, quorum sensing (QS), acetate metabolism, motility, bioluminescence, etc. [34,35] It also presented in the landmark genes of yellow light's HGP (Data S6), inferring its role in sensing light signals. Pilus synthesis was also considerable, indicating it played a role under yellow light illumination. Combining with the overall

profiles of yellow lights' landmark genes of HGP and functional pathways analysis in **Fig. 4*d***, it could be learned that yellow light promoted ROS generation to signal protein synthesis and secretion.

Regarding to SGPs, some landmark genes in SGP of blue light also involved in the synthesis of vital antioxidants and enzymes related to oxidative stress release (**Dataset S5**). A large number of landmark genes involved in the metabolism of antioxidants, such as peroxiredoxin, alkyl hydroperoxide reductase and dye decolorizing peroxidase. These enzymes can catalyze reactions that produce or mediated typical antioxidants, including vitamin C, vitamin E, glutathione (GSH) and flavonoids [36,37]. In contrast, most of landmark genes in yellow light's SGP were related to electron transfer and energy metabolism, such as aldehyde dehydrogenase (NAD+), NADH dehydrogenase and ubiquinol-cytochrome c reductase (Data S6).

Discussion S6. Molecular biology mechanism of wavelength-dependent denitrification. We extracted the subcellular location and expression information of photo-denitrification genes (PhNGs) to reconstruct metabolism scheme (**Fig. S15**). Most phototransduction genes up-regulated by blue and yellow light encoded calmodulin (CaM) and G-protein (**Fig. S6a, Table S2**). Light activated G-protein coupled receptor (GPRC) signaled particulate guanylate cyclase (pGC) to transform guanosine triphosphate (GTP) to guanosine monophosphate (cGMP), similarly for cyclic adenosine monophosphate (cAMP) [38]. Both of them were second messengers that transmitted signals through regulatory networks and influenced denitrification ***[39–41]***. Specifically, genes for denitrification were hub genes, involving multiple pathways like two-component system, transporters, and etc. [41] Light signals from receptors of phototransduction were transmitted along the metabolism network to two-component regulatory system, influencing the nitrogen metabolism through nitrogen regulatory kinase NtrB (GlnL) [40]. As shown by the expression levels comparison (**Fig. S6a**), phototransduction was more sensitive to blue light, thus generating more signals to produce diverse metabolites with great potential as active substances [42]. As for denitrification, blue light inhibited the expression of NarK1 but enhanced NarK2 (Data S7), two transmembrane proteins for nitrite and nitrate transportation. Therefore, nitrite was stored in the cytoplasm and thus separated from nitrite reductase in the periplasm, explaining the effective nitrite accumulation under blue light (**Fig. 2*a***).

Discussion S7. Molecular biology mechanism of nitrate-superoxide co-regulation. It was intriguing that microbiota with higher superoxide levels after light exposure enhanced nitrate removal, which was contrary

to the common sense on superoxide cytotoxicity [43]. Our co-expression and topological network models succeeded in discovering the nitrate-superoxide correlation mechanism. This was also related to subcellular location. Peroxisome is essential in ROS detoxification and signaling. We extracted all valid DEGs subjected to peroxisome to figure out the nitrate-superoxide correlation mechanism (**Dataset S7**). SOD was the main enzyme in peroxisome that relieve oxidative stress caused by superoxide. There were several genes encoding SOD. The intracellular SOD (g_41985, Q9RUV2) exhibited the highest expression level. Blue light contributed to higher expression levels of transmembrane SOD (g_29329, P31108), suggesting it scavenged the extracellular superoxide and prevented oxidative stress signaling between cells [28]. Therefore, micro-stimulation of superoxide contributed to inter-cellular signaling to promote substances exchanges and collective cellular metabolism.

Discussion S8. Molecular biology mechanism of wavelength-divergent secretion system. Notably, a large proportion of DEGs of secretion system encode pilin secretion and fimbrial assembly proteins, which were up-regulated by yellow light (**Dataset S7**). This suggested that more pilis were formed in response to illumination for electron transfer and energy metabolism [44]. Blue light significantly up-regulated the expression of Chaperone protein ClpB, one of proteins of type VI secretion system, which is a protein nanomachine that is widespread in Gram-negative bacteria and is used to translocate effector proteins directly into neighboring cells as effective weapons for inter-bacterial competition [45]. The competition was collective responses towards irradiation stress to obtain sufficient resources for survival [46]. Despite unfavorable conditions caused by blue light [25,41], it was still within the threshold. Microbiota tended to secrete more bioactive substances as inter-cellular signals to maintain homeostasis (**Fig. S11a**), which accelerated the proliferation and evolution of the microbial community, corresponding to the strengthened nitrate removal in self-catalysis experiments (**Fig. 4*f***).

Supporting Figures and Tables

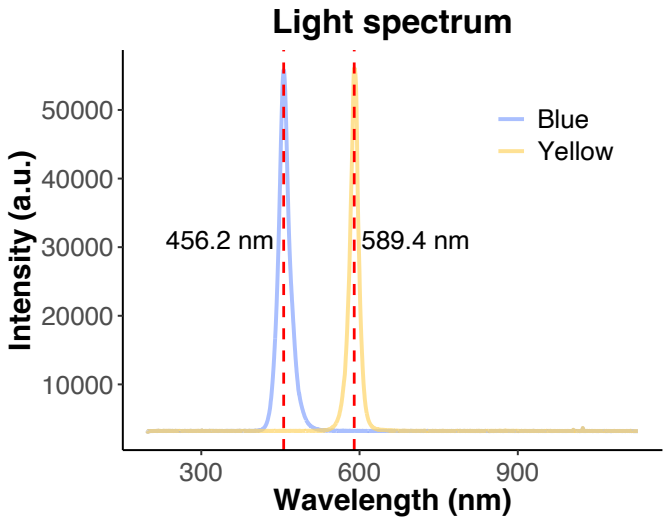

Fig. S1. The Light spectrums of LEDs employed to regulate photo-denitrification.

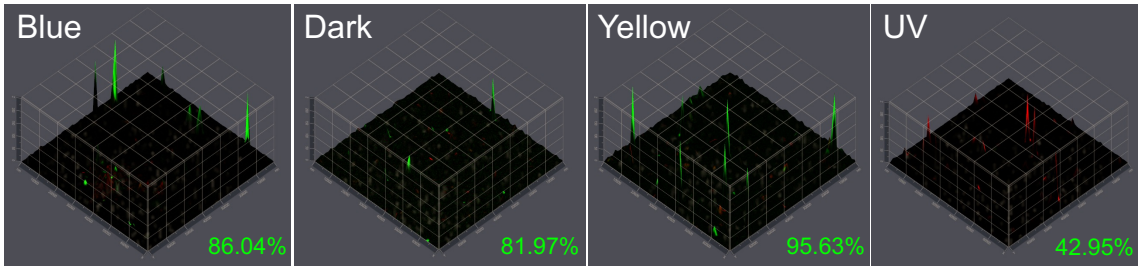

Fig. S2. 3D Fluorescence images that depicted the cellular viability and membrane damage. Green: active cells. Red: dead cells. The peak's height represents fluorescence intensity. The green font percentages represented the ratio of live cells.

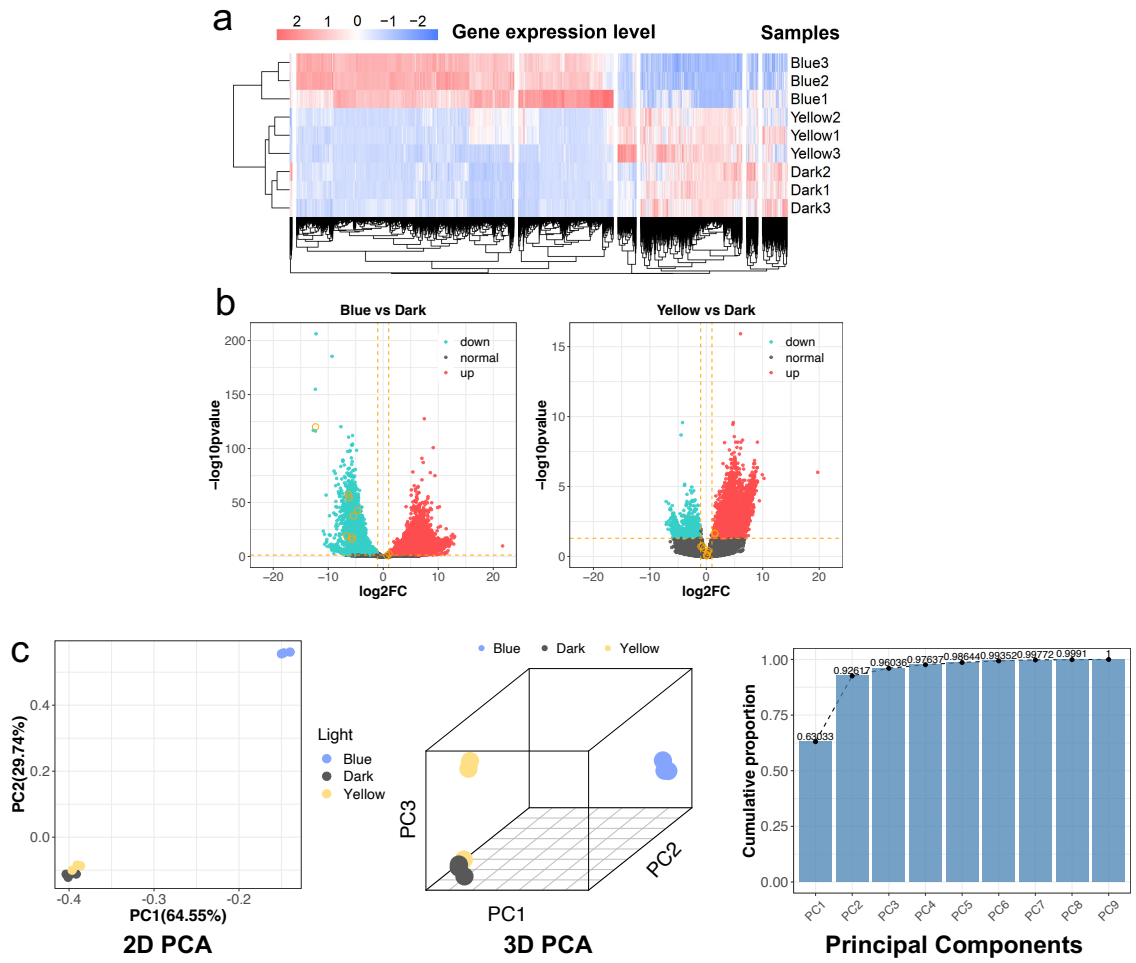

Fig. S3. Gene expression profiles and dimension reduction analysis. (a) Gene expression patterns of valid DEGs. The expression levels were log normalized and scaled based on FPKM. (b) Volcano plots of the DESeq results of blue and yellow lights. The dark condition was utilized as control. The orange cycle highlighted the top 10 highly-expressed genes. (c) Dimension reduction through principal components analysis (PCA).

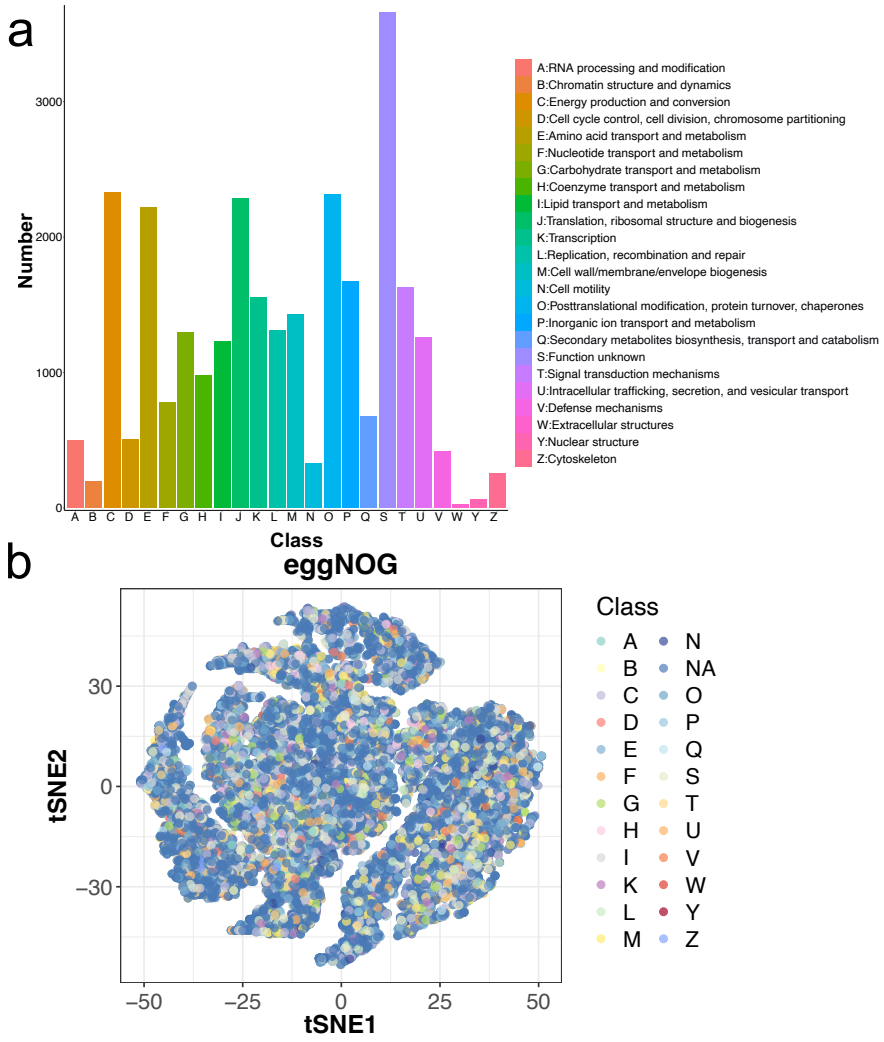

Fig. S4. EggNOG class distribution on all differentially expressed genes (DEGs). (a) Genes counts that assigned to different eggNOG class. Class abbreviation and more information see Supplementary Table 2. (b) Gene expression distribution and corresponding eggnog labels of valid DEGs after dimension reduction through tSNE. NA denotes genes lack of eggNOG annotation.

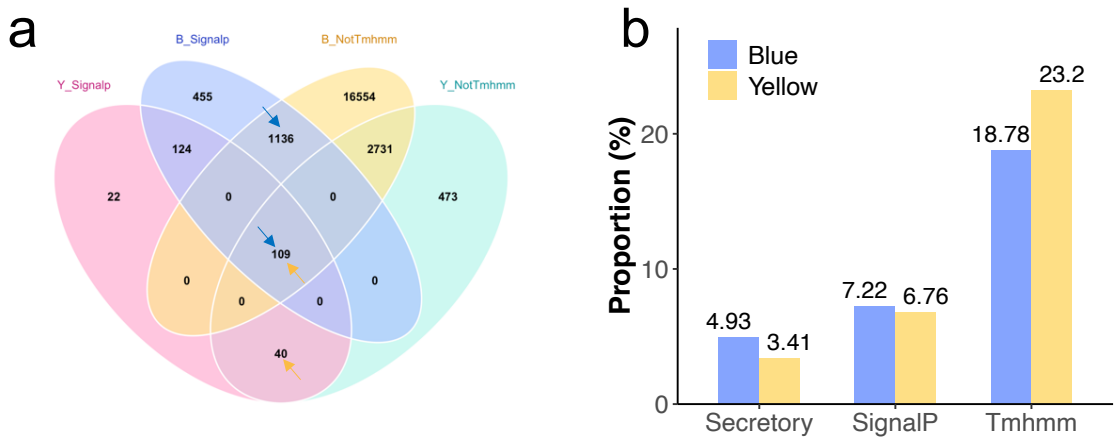

Fig. S5. Subcellular location profiles of blue light's and yellow light's valid DEGs. (a) Blue and yellow errors denoted secretory protein responsive to blue and yellow light. B_Signalp and Y_Signalp represent DEGs that encoded signal peptides of blue and yellow light. B_NotTmhmm and Y_NotTmhmm represent DEGs that encode protein without transmembrane domains of blue and yellow light. The overlaps of Signalp and NotTmhmm were secretory proteins. (b) Proportion of secreted protein, signal peptide and transmembrane protein in DEGs of blue and yellow light.

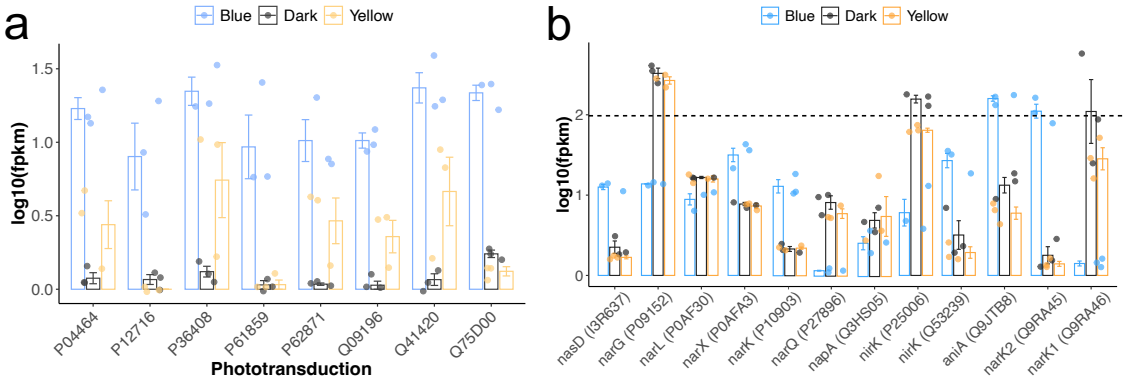

Fig. S6. Expression patterns of major DEGs that related to phototransduction and nitrate conversion. (a) Expression levels of DEGs that annotated to phototransduction. Genes were denoted by Swissprot ID. Functional information can be found in Table S2. (b) Expression levels of nitrate- and nitrite-related genes, namely partial denitrification genes (PDGs). Genes with expression above the dash line were regarded as highly-expressed genes. See Table S3 for details.

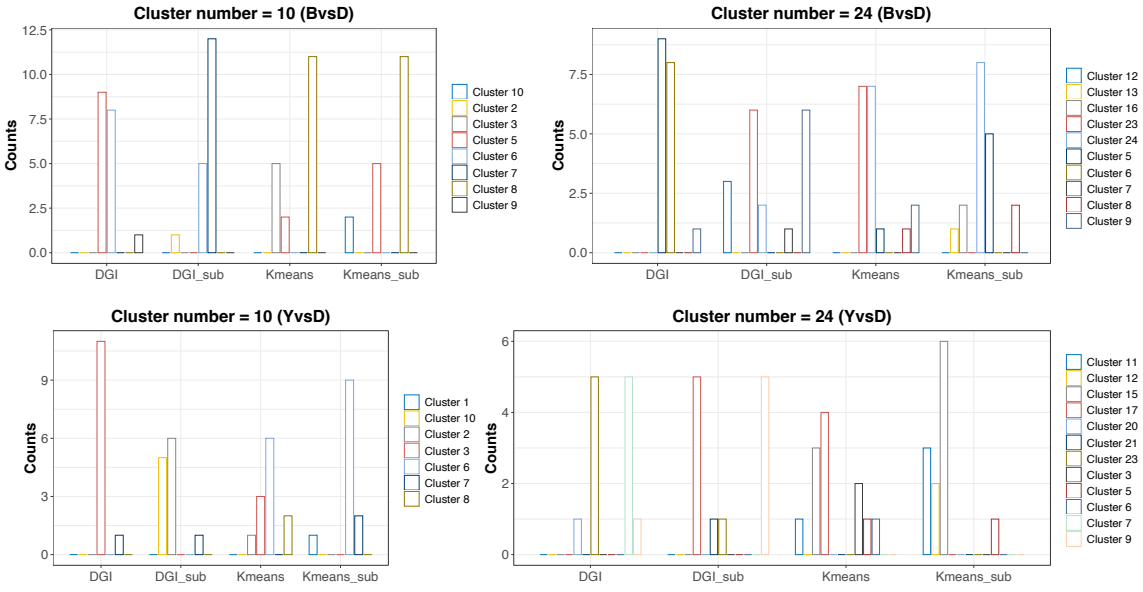

Fig. S7. Cluster assignment of phototransduction genes under different cluster number. BvsD: blue vs dark, indicating DEGs of blue light group. YvsD: yellow vs dark, indicating DEGs of yellow light group.

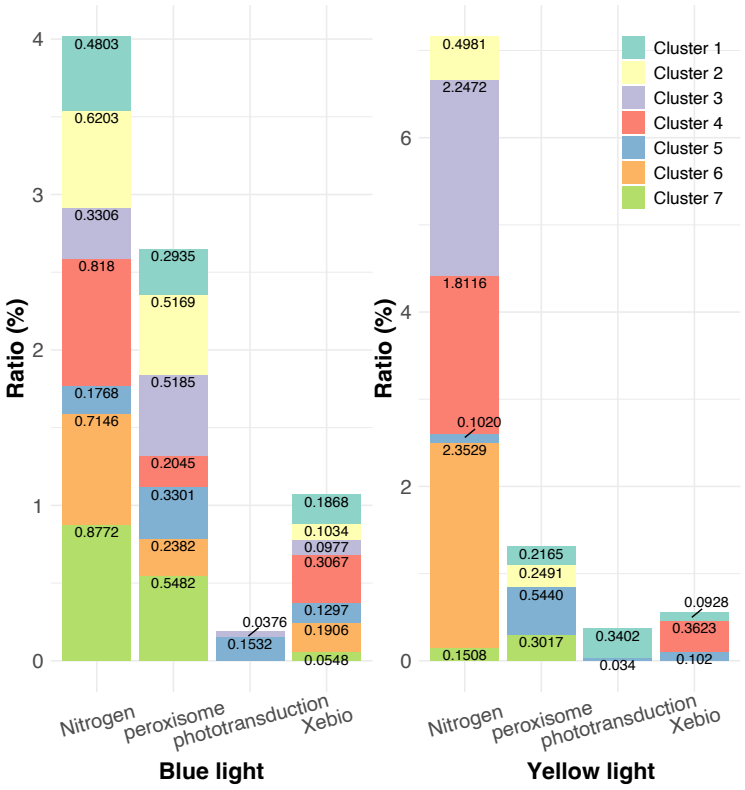

Fig. S8. Cluster distribution of functional pathways. Nitrogen: nitrogen metabolism. Xebio: metabolism of xenobiotics by cytochrome P450.

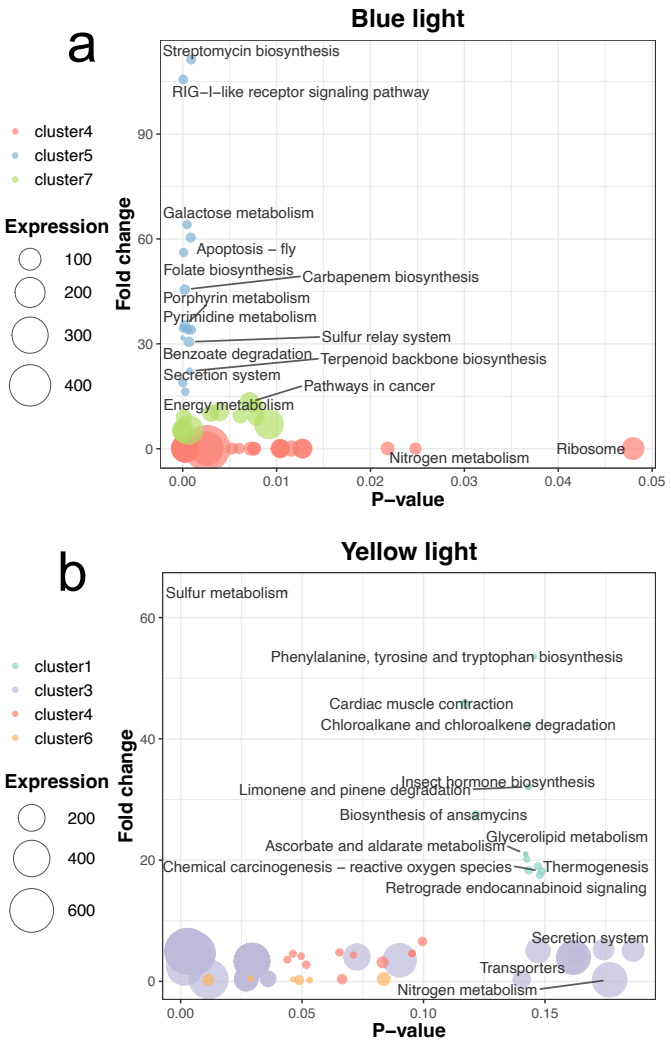

Fig. S9. Pathway enrichment analysis on significant photo-denitrification pathways of blue (a) and yellow light (b). Fold changes were calculated with the dark group as control. Expression was the mean gene expression under blue or yellow light.

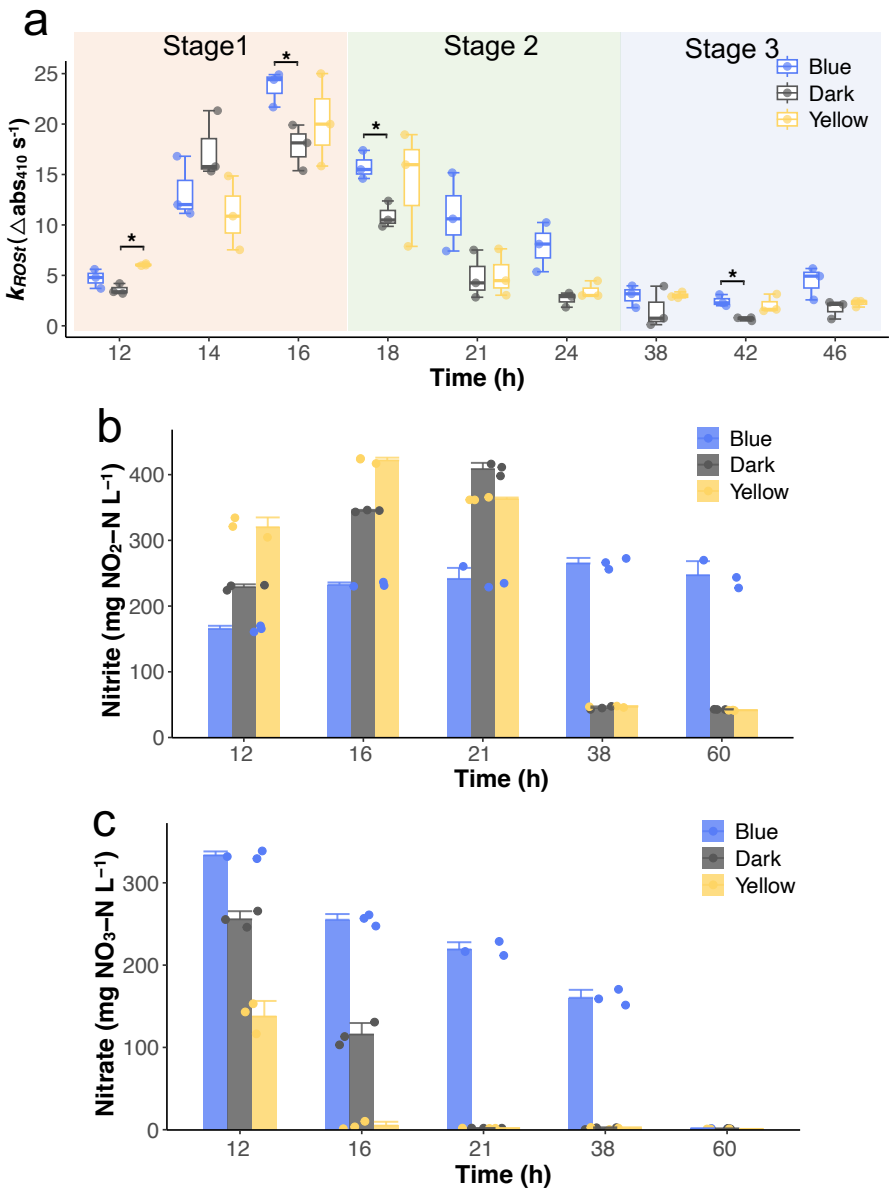

Fig. S10. Wet-lab validations on co-expression between total reactive oxygen species (ROS) levels and photo-denitrification (a) Total ROS variation in the time course. Photo-denitrification was divided into three stages based on main nitrogen metabolism. Stage1: nitrate reduction. Stage 2: nitrite reduction. Stage 3: Inorganic nitrogen depletion. Nitrite (b) and nitrate (c) concentrations variation during photo-denitrification. It can be observed that total ROS shared similar trends with nitrite concentrations.

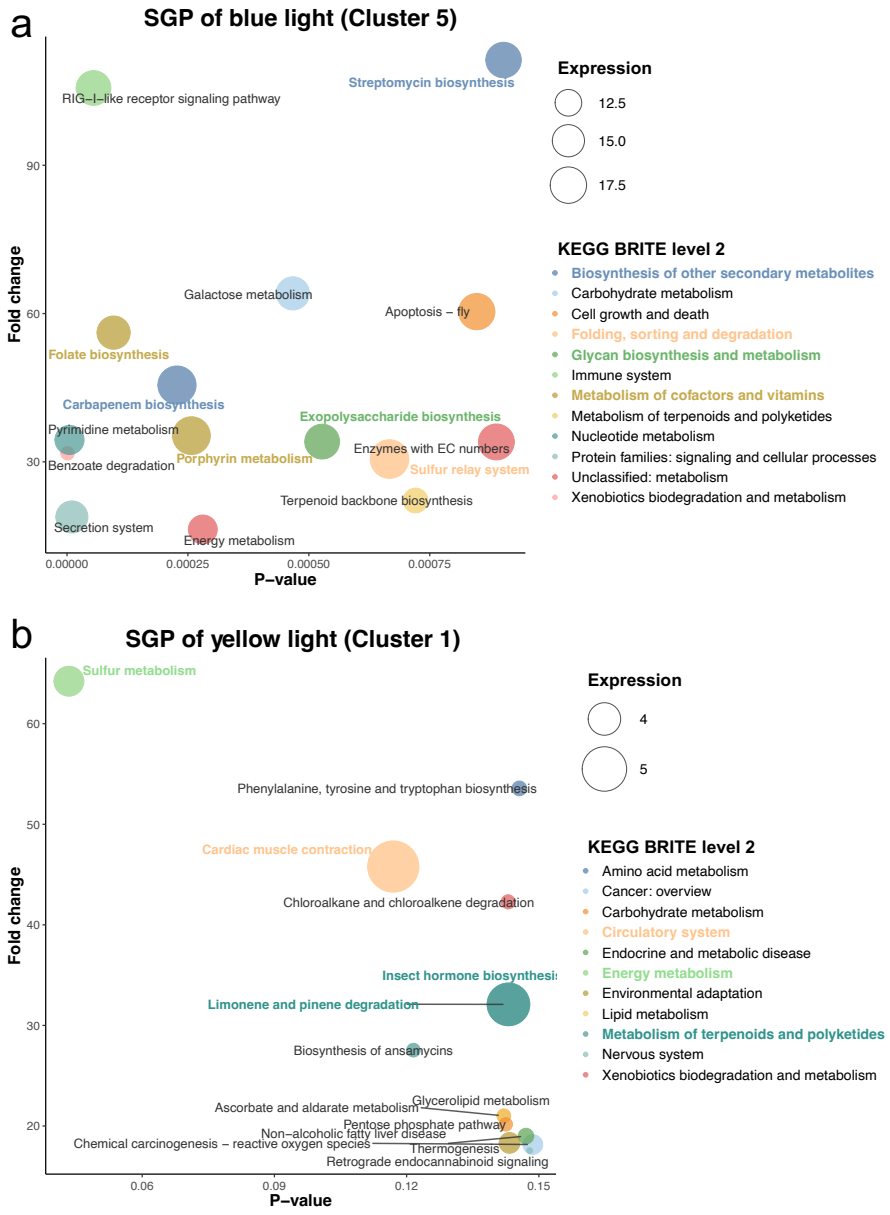

Fig. S11. Pathways enrichment analysis of signaling gene panels (SGPs) of blue (a) and yellow (b) light. The most highly expressed pathways (Table S4) and their KEGG Brite was highlighted by corresponding color and bold font. Bubble size denoted mean expression levels (FPKM) under blue or yellow light, respectively. Fold changes were calculated with the dark group as control.

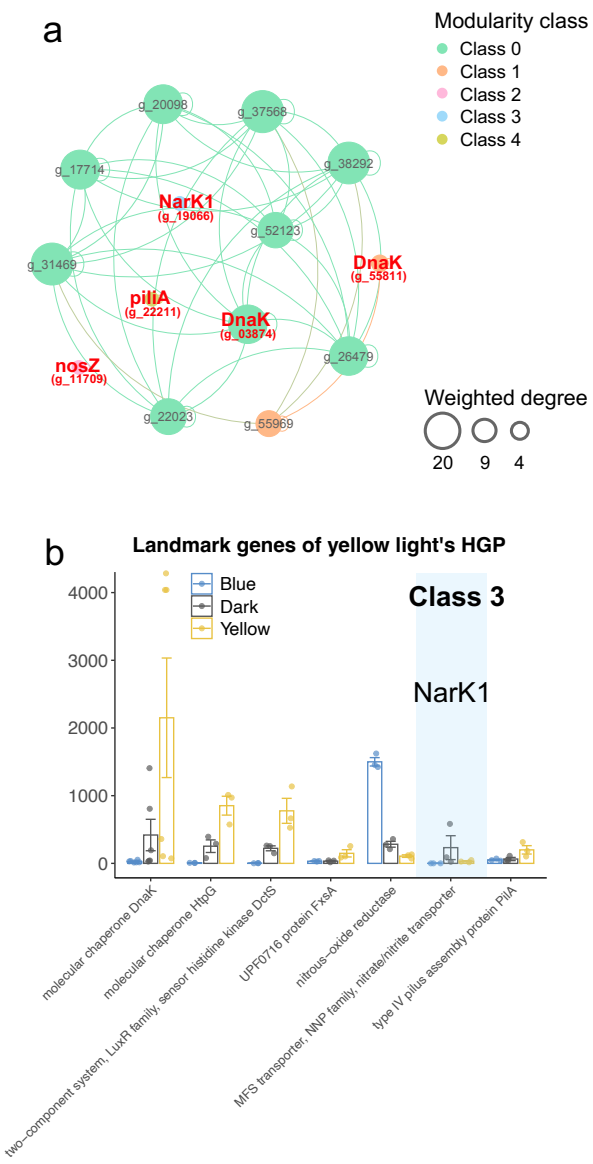

Fig. S12. The topological network model and corresponding landmark genes of yellow light's HGP. (a) The gene topological network. Details on the topological information of gene nodes and landmark genes were summarized in Data S2, S6. The bold red font highlighted the landmark genes with the highest expression. The bold black font highlighted the principal denitrification genes. (b) Expression levels of landmark genes. Background highlighted the modularity class PDGs subjected to.

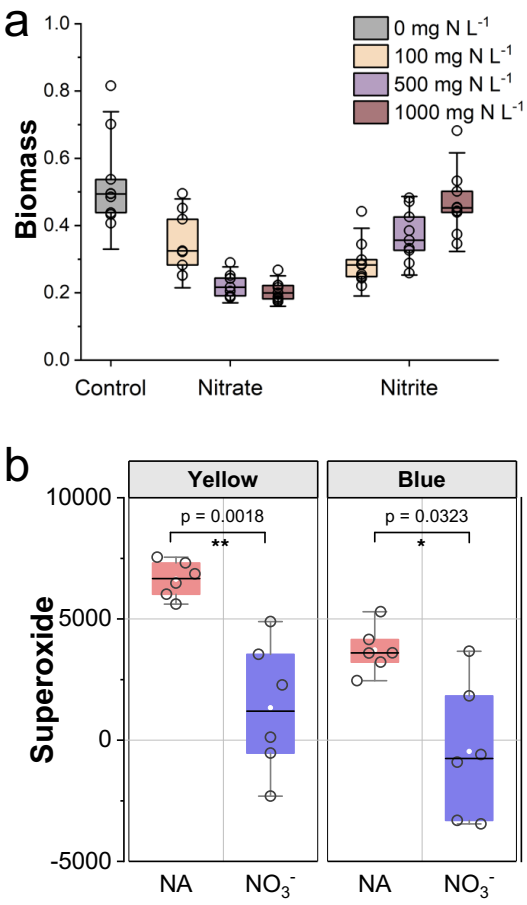

Fig. S13. Co-expression of nitrate metabolism and superoxide production. (a) Biomass synthesis under different nitrate and nitrite concentrations. (b) Influence of nitrate absence on superoxide level under yellow and blue light.

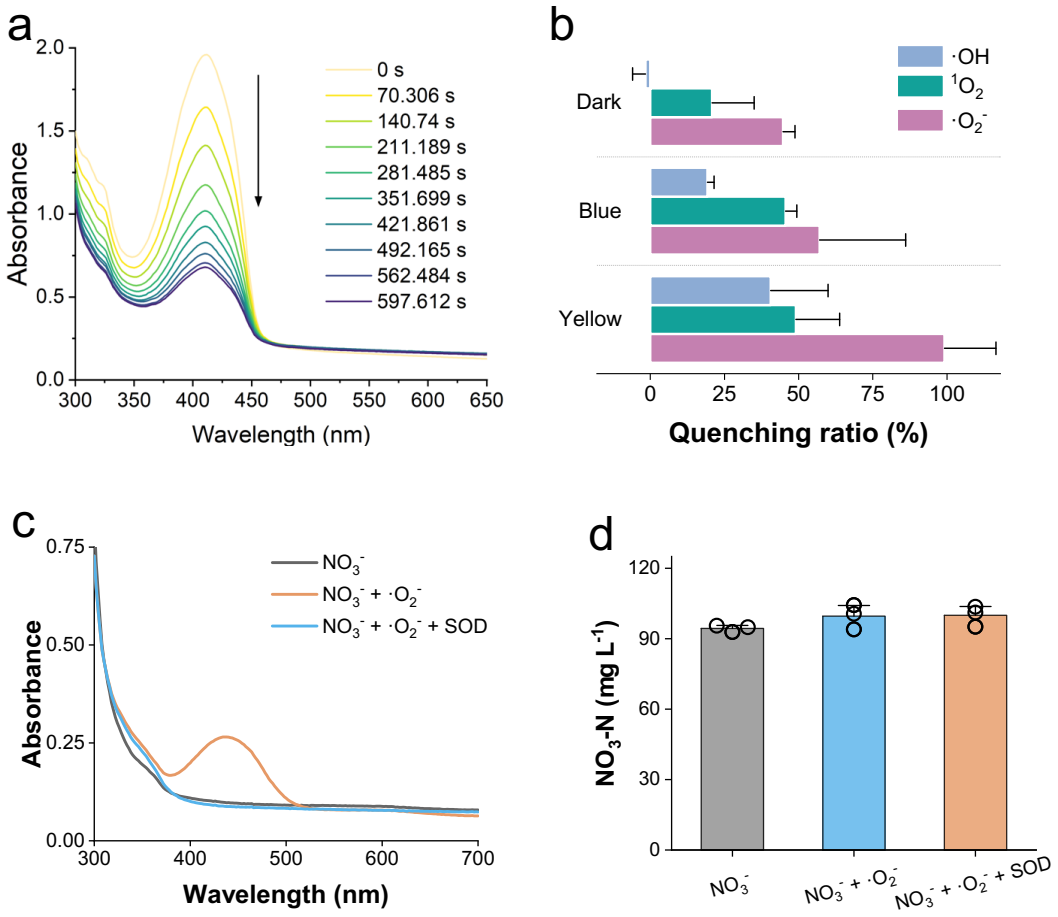

Fig. S14. ROS assay and addition of superoxide. (a) Dynamic total ROS production under dark condition. (b) Quenching experiments under different illumination conditions. Quenching ratio was calculated based on nitrate removal efficiency. (c) Detection of superoxide added by biological methods. (d) Non-biological reactions between nitrate, superoxide and SOD.

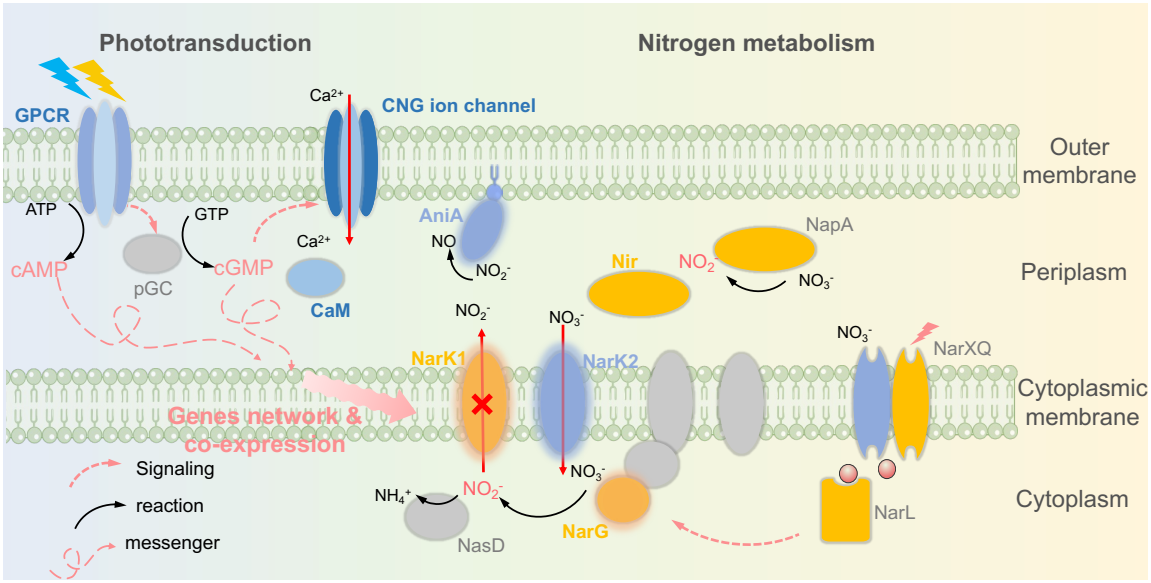

Fig. S15. Mechanistic scheme of light-regulated denitrification. It was reconstructed based on the subcellular location of critical enzymes involved in phototransduction and nitrogen metabolism. CaM: Calmodulin. GPCR: G-protein coupled receptor. CNG ion channel: cyclic nucleotide-gated ion channel. For details on genes and protein abbreviations see Table S2, S3.

Table S1. Statistic summary of unigenes obtained through clean reads assembly.

1. Transcripts Information

Samples	Number of transcripts	Average number of transcripts	Total length (bp)	Average total length	Min length	Max length	Average length	Average length
Blue1	134,599		85,919,295		197	85,629	638.3	
Blue2	135,862	132,951	75,682,473	77,591,722	84	57,433	557.1	583
Blue3	128,391		71,173,398		197	112,849	554.3	
Dark1	52,490		33,818,001		94	117,282	644.3	
Dark2	57,157	56,014	35,918,000	35,877,126	197	112,284	628.4	641
Dark3	58,395		37,895,376		197	126,032	648.9	
Yellow1	69,613		41,305,106		99	330,546	593.4	
Yellow2	68,999	66,051	40,442,350	39,587,255	197	273,774	586.1	600
Yellow3	59,541		37,014,309		133	115,134	621.7	

2. Changing ratio

Light	-	Average number of transcripts	-	Average total length	-	-	-	Average length
Blue	-	137.4%	-	116.3%	-	-	-	-8.9%
Yellow	-	17.9%	-	10.3%	-	-	-	-6.3%

Table S2. Highly expressed genes that encode phototransduction of blue and yellow light datasets. Groups of blue and yellow light were manually annotated and merged. SignalP and Tmhmm represented signal peptide and transmembrane protein. 1 and 0 were utilized to denote whether or not this gene encodes signal peptide and transmembrane protein.

Gene id	SwissProt	Genes	SwissProt Description	SignalP	Tmhmm	Notes
g_34466	Q75D00	ACT1	RecName: Full=Actin	0	0	Actins are highly conserved proteins that involve in various types of cell motility Catalytic activity: ATP + H2O = ADP + H+ + phosphate
g_01051	P12716	Actin, cytoplasmic	RecName: Full=Actin, cytoplasmic; Contains: RecName: Full=Actin, cytoplasmic, intermediate form; Flags: Precursor	0	0	
g_09501	P04464	Calmodulin	RecName: Full=Calmodulin; Short=CaM	0	0	**Calmodulin** mediates the control of a large number of enzymes, ion channels and other proteins by Ca2+. Among the enzymes to be stimulated by the calmodulin-Ca2+ complex are a number of protein
g_49798	P04464	Calmodulin	RecName: Full=Calmodulin; Short=CaM	0	0	
g_09982	P61859	cmd-1	RecName: Full=Calmodulin; Short=CaM >P61860.2 RecName: Full=Calmodulin; Short=CaM >P61861.2 RecName: Full=Calmodulin; Short=CaM >Q9UWF0.4 RecName: Full=Calmodulin; Short=CaM	0	0	

Gene id	SwissProt	Genes	SwissProt Description	SignalP	Tmhmm	Notes
						kinases and phosphatases.
g_31770	Q09196	cdc4	RecName: Full=Myosin regulatory light chain cdc4	0	1	Involved in cytokinesis. Required for the formation and function of the contractile ring.
g_24609	P62871	GNB1	RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P62872.3 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P62873.3 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P62874.3 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin	0	0	**Guanine nucleotide-binding proteins (G proteins)** are involved as a modulator or transducer in various transmembrane signaling systems. The beta and gamma chains are required for the GTPase activity, for replacement of GDP by GTP, and For **G protein-effector** interaction (By

Gene id	SwissProt	Genes	SwissProt Description	SignalP	Tmhmm	Notes
			beta chain 1 >P54311.4 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1			similarity). Required for normal chemotaxis in response to cAMP and for
g_48714	P62871	GNB1	RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P62872.3 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P62873.3 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P62874.3 RecName: Full=Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1 >P54311.4 RecName: Full=Guanine	0	0	aggregation during scorocarp development

Gene id	SwissProt	Genes	SwissProt Description	SignalP	Tmhmm	Notes
			nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-1; AltName: Full=Transducin beta chain 1			
g_33597	P36408	gpbA	RecName: Full=Guanine nucleotide-binding protein subunit beta	0	0	
g_08757	Q41420	PCM3	PUTATIVE PSEUDOGENE: RecName: Full=Putative calmodulin-3; Short=CaM-3	0	0	Calmodulin mediates the control of a large number of enzymes, ion channels and other proteins by Ca2+. Among the enzymes to be stimulated by the calmodulin-Ca2+ complex are a number of protein kinases and phosphatases.

Table S3. Annotations of nitrate- and nitrite-related genes, i.e. partial denitrification genes (PDGs). Red font denoted highly expressed DEGs in Fig. S6b. Genes with expression levels < 3 FPKM were filtered. SignalP and Tmhmm represented signal peptide and transmembrane protein. 1 and 0 were utilized to denote whether or not this gene encodes signal peptide and transmembrane protein.

Gene id	SwissProt ID	Genes	SwissProt_Description	SignalP	Tmhmm
g_06597	P25006	nirK	RecName: Full=Copper-containing nitrite reductase; AltName: Full=Cu-NIR; Flags: Precursor	0	0
g_08093	I3R637	nasD	RecName: Full=Assimilatory ferredoxin-dependent nitrite reductase; Short=NiR; AltName: Full=Ferredoxin:nitrite reductase	0	0
g_13189	P0AFA3	narX	RecName: Full=Nitrate/nitrite sensor protein NarX >P0AFA2.1 RecName: Full=Nitrate/nitrite sensor protein NarX >P0AFA4.1 RecName: Full=Nitrate/nitrite sensor protein NarX	0	1
g_19066	Q9RA46	narK1	RecName: Full=Probable nitrate/nitrite antiporter NarK1	0	1
g_25425	P0AF30	narL	RecName: Full=Nitrate/nitrite response regulator protein NarL >P0AF29.1 RecName: Full=Nitrate/nitrite response regulator protein NarL >P0AF28.1 RecName: Full=Nitrate/nitrite response regulator protein NarL >P0AF31.1 RecName: Full=Nitrate/nitrite response regulator protein NarL	0	0
g_26855	Q53239	nirK	RecName: Full=Copper-containing nitrite reductase; AltName: Full=Cu-NIR; Flags: Precursor	0	1
g_30989	I3R637	nasD	RecName: Full=Assimilatory ferredoxin-dependent nitrite reductase; Short=NiR; AltName: Full=Ferredoxin:nitrite reductase	0	0
g_32124	Q9RA45	narK2	RecName: Full=Probable nitrate/nitrite antiporter NarK2	0	1
g_32633	P10903	narK	RecName: Full=Nitrate/nitrite transporter NarK; AltName:	0	1

			Full=Nitrite extrusion protein 1; AltName: Full=Nitrite facilitator 1		
g_37842	P27896	narQ	RecName: Full=Nitrate/nitrite sensor protein NarQ	0	0
g_37935	P0AF30	narL	RecName: Full=Nitrate/nitrite response regulator protein NarL >P0AF29.1 RecName: Full=Nitrate/nitrite response regulator protein NarL >P0AF28.1 RecName: Full=Nitrate/nitrite response regulator protein NarL >P0AF31.1 RecName: Full=Nitrate/nitrite response regulator protein NarL	0	0
g_41999	P10903	narK	RecName: Full=Nitrate/nitrite transporter NarK; AltName: Full=Nitrite extrusion protein 1; AltName: Full=Nitrite facilitator 1	0	1
g_48246	P09152	narG	RecName: Full=Respiratory nitrate reductase 1 alpha chain; AltName: Full=Nitrate reductase A subunit alpha; AltName: Full=Quinol-nitrate oxidoreductase subunit alpha	0	0
g_51049	P09152	narG	RecName: Full=Respiratory nitrate reductase 1 alpha chain; AltName: Full=Nitrate reductase A subunit alpha; AltName: Full=Quinol-nitrate oxidoreductase subunit alpha	0	0
g_55724	Q9JTB8	aniA	RecName: Full=Copper-containing nitrite reductase; Flags: Precursor	1	0
g_56773	Q3HS05	napA	RecName: Full=Periplasmic nitrate reductase; Flags: Precursor >A4VJ11.1 RecName: Full=Periplasmic nitrate reductase; Flags: Precursor	0	0

Table S4. Pathways enrichment analysis on the HGPs and SGPs of blue and yellow light.

Table S4.1 Hub gene panel of blue light

level3_pathway_name	KEGGBrite2	pvalue	Fold change	Expression
Inositol phosphate metabolism	Carbohydrate metabolism	0.00260	0.148	490.854
Streptomycin biosynthesis	Biosynthesis of other secondary metabolites	0.00257	0.145	246.667
FoxO signaling pathway	Signal transduction	0.00027	0.085	159.365
Longevity regulating pathway	Aging	0.00027	0.085	159.365
Longevity regulating pathway - multiple species	Aging	0.00027	0.085	159.365
MAPK signaling pathway - fly	Signal transduction	0.00027	0.085	159.365
Ribosome	Protein families: genetic information processing	0.04797	0.075	99.955
Peroxisome	Transport and catabolism	0.00001	0.078	91.166
Huntington disease	Neurodegenerative disease	0.00009	0.075	85.180
Lipid and atherosclerosis	Cardiovascular disease	0.00005	0.078	83.126
Legionellosis	Infectious disease: bacterial	0.01276	0.055	70.600
Shigellosis	Infectious disease: bacterial	0.01276	0.055	70.600
Plant-pathogen interaction	Environmental adaptation	0.01036	0.074	67.760
NOD-like receptor signaling pathway	Immune system	0.01043	0.054	59.981
Salmonella infection	Infectious disease: bacterial	0.01043	0.054	59.981
Longevity regulating pathway - worm	Aging	0.01157	0.064	45.326
Chemical carcinogenesis - reactive oxygen species	Cancer: overview	0.00001	0.073	44.445

Flagellar assembly	Cell motility	0.00722	0.053	33.015
Mitochondrial biogenesis	Protein families: genetic information processing	0.02183	0.093	32.210
Monobactam biosynthesis	Biosynthesis of other secondary metabolites	0.00061	0.039	31.504
GABAergic synapse	Nervous system	0.00760	0.037	30.935
Glutamatergic synapse	Nervous system	0.00760	0.037	30.935
Nitrogen metabolism	Energy metabolism	0.02480	0.055	25.354
Bacterial motility proteins	Protein families: signaling and cellular processes	0.00524	0.054	24.734
Exosome	Protein families: signaling and cellular processes	0.00601	0.075	21.584
Lysine biosynthesis	Amino acid metabolism	0.00006	0.040	21.406
mean	-	0.00783	0.071	92.492

Table S4.2 Hub gene panel of yellow light

level3_pathway_name	KEGGBrite2	pvalue	Fold change	Expression
Longevity regulating pathway - worm	Aging	0.00287	4.987	1600.498
Messenger RNA biogenesis	Protein families: genetic information processing	0.00287	4.987	1600.498
RNA degradation	Folding, sorting and degradation	0.00287	4.987	1600.498
Tuberculosis	Infectious disease: bacterial	0.00287	4.987	1600.498
Exosome	Protein families: signaling and cellular processes	0.00538	4.647	1413.447

Table S4.2 Hub gene panel of yellow light

Mitochondrial biogenesis	Protein families: genetic information processing	0.00538	4.647	1413.447
Chaperones and folding catalysts	Protein families: genetic information processing	0.00422	4.671	1183.239
Antigen processing and presentation	Immune system	0.02937	3.358	852.291
Chemical carcinogenesis - receptor activation	Cancer: overview	0.02937	3.358	852.291
Estrogen signaling pathway	Endocrine system	0.02937	3.358	852.291
Fluid shear stress and atherosclerosis	Cardiovascular disease	0.02937	3.358	852.291
IL-17 signaling pathway	Immune system	0.02937	3.358	852.291
Lipid and atherosclerosis	Cardiovascular disease	0.02937	3.358	852.291
Membrane trafficking	Protein families: genetic information processing	0.02937	3.358	852.291
Necroptosis	Cell growth and death	0.02937	3.358	852.291
NOD-like receptor signaling pathway	Immune system	0.02937	3.358	852.291
Pathways in cancer	Cancer: overview	0.02937	3.358	852.291
PI3K-Akt signaling pathway	Signal transduction	0.02937	3.358	852.291
Plant-pathogen interaction	Environmental adaptation	0.02937	3.358	852.291
Progesterone-mediated oocyte maturation	Endocrine system	0.02937	3.358	852.291
Prostate cancer	Cancer: specific types	0.02937	3.358	852.291
Proteasome	Protein families: genetic information processing	0.02937	3.358	852.291
Protein phosphatases and	Protein families:	0.02937	3.358	852.291

Table S4.2 Hub gene panel of yellow light

associated proteins	metabolism			
Protein processing in endoplasmic reticulum	Folding, sorting and degradation	0.02937	3.358	852.291
Salmonella infection	Infectious disease: bacterial	0.02937	3.358	852.291
Th17 cell differentiation	Immune system	0.02937	3.358	852.291
Peptidases and inhibitors	Protein families: metabolism	0.00176	2.366	840.141
DNA repair and recombination proteins	Protein families: genetic information processing	0.16173	3.946	797.835
Mismatch repair	Replication and repair	0.16173	3.946	797.835
Protein kinases	Protein families: metabolism	0.09018	3.491	776.097
Two-component system	Signal transduction	0.07259	4.038	452.943
Glyoxylate and dicarboxylate metabolism	Carbohydrate metabolism	0.14717	5.080	393.962
Ribosome biogenesis	Protein families: genetic information processing	0.01149	0.364	374.515
Secretion system	Protein families: signaling and cellular processes	0.18633	5.103	291.366
Longevity regulating pathway - multiple species	Aging	0.17416	5.256	262.407
General function prediction only	Poorly characterized	0.16301	4.414	148.270
Chromosome and associated proteins	Protein families: genetic information processing	0.02699	0.371	117.693
Cytoskeleton proteins	Protein families: signaling and cellular processes	0.02699	0.371	117.693

Table S4.2 Hub gene panel of yellow light				
Prokaryotic defense system	Protein families: signaling and cellular processes	0.02699	0.371	117.693
Nitrogen metabolism	Energy metabolism	0.17664	0.268	68.868
Transporters	Protein families: signaling and cellular processes	0.14045	0.339	54.838
Bacterial chemotaxis	Cell motility	0.03599	0.435	47.834
Flagellar assembly	Cell motility	0.03599	0.435	47.834
mean	-	0.04938	3.214	810.347

Table S4.3 Signaling gene panel of blue light				
level3_pathway_name	KEGGBrite2	pvalue	Fold change	Expression
Sulfur relay system	Folding, sorting and degradation	0.00067	30.562	19.014
Carbapenem biosynthesis	Biosynthesis of other secondary metabolites	0.00023	45.531	18.698
Porphyrin metabolism	Metabolism of cofactors and vitamins	0.00026	35.295	18.430
Apoptosis - fly	Cell growth and death	0.00085	60.387	17.438
Enzymes with EC numbers	Unclassified: metabolism	0.00089	34.027	17.180
Streptomycin biosynthesis	Biosynthesis of other secondary metabolites	0.00090	111.291	16.904
Exopolysaccharide biosynthesis	Glycan biosynthesis and metabolism	0.00053	34.094	16.483
RIG-I-like receptor signaling pathway	Immune system	0.00005	105.636	16.443
Folate biosynthesis	Metabolism of cofactors and	0.00010	56.142	15.579

Table S4.3 Signaling gene panel of blue light				
	vitamins			
Galactose metabolism	Carbohydrate metabolism	0.00047	64.080	15.535
Secretion system	Protein families: signaling and cellular processes	0.00001	18.851	14.948
Pyrimidine metabolism	Nucleotide metabolism	0.00001	34.440	13.524
Energy metabolism	Unclassified: metabolism	0.00028	16.318	13.399
Terpenoid backbone biosynthesis	Metabolism of terpenoids and polyketides	0.00072	22.186	11.840
Benzoate degradation	Xenobiotics biodegradation and metabolism	0.00000	31.734	10.032
mean	-	0.00040	46.705	15.696

Table S4.4 Signaling gene panel of yellow light				
level3_pathway_name	KEGGBrite2	pvalue	Fold change	Expression
Cardiac muscle contraction	Circulatory system	0.11697	45.798	5.728
Insect hormone biosynthesis	Metabolism of terpenoids and polyketides	0.14310	32.091	4.822
Limonene and pinene degradation	Metabolism of terpenoids and polyketides	0.14310	32.091	4.822
Sulfur metabolism	Energy metabolism	0.04337	64.226	3.833
Thermogenesis	Environmental adaptation	0.14335	18.318	3.373
Chemical carcinogenesis - reactive oxygen species	Cancer: overview	0.14863	18.157	3.336
Non-alcoholic fatty liver disease	Endocrine and metabolic disease	0.14708	18.995	3.212

Phenylalanine, tyrosine and tryptophan biosynthesis	Amino acid metabolism	0.14557	53.568	3.189
Chloroalkane and chloroalkene degradation	Xenobiotics biodegradation and metabolism	0.14302	42.285	3.177
Biosynthesis of ansamycins	Metabolism of terpenoids and polyketides	0.12154	27.538	3.167
Ascorbate and aldarate metabolism	Carbohydrate metabolism	0.14204	21.008	3.165
Glycerolipid metabolism	Lipid metabolism	0.14204	21.008	3.165
Pentose phosphate pathway	Carbohydrate metabolism	0.14250	20.160	3.162
Retrograde endocannabinoid signaling	Nervous system	0.14791	17.547	3.089
mean	-	0.13359	30.914	3.660

Table S5. Topological properties of HGPs and SGPs.

Light	Gene panel	Graph attributes				Nodes attributes			Edge attributes
		Node counts	Edge counts	Network Diameter	Graph density	Avg. degree	Avg. weighted degree	Avg. clustering coefficient	Avg. path length
Blue light	HGP	64	441	5	0.219	13.781	24.045	0.578	2.254
	SGP	216	2371	10	0.102	21.954	39.336	0.623	2.893
Yellow light	HGP	18	53	3	0.346	7.571	12.528	0.762	1.4
	SGP	63	1300	4	0.666	41.27	76.481	0.853	1.374

HGP: hub gene panel.

SGP: signaling gene panel.

Avg.: average

The topological properties correspond to the interaction networks in Fig. 4A, B, Fig. S12 and Fig. S13.

Legends for Datasets

Dataset S1. Topological properties of blue light's hub gene panels (HGP).

Dataset S2. Topological properties of yellow light's hub gene panels (HGP).

Dataset S3. Topological properties of blue light's signaling gene panel (SGP).

Dataset S4. Topological properties of yellow light's signaling gene panel (SGP).

Dataset S5. Landmark genes of blue light.

Dataset S6. Landmark genes of yellow light.

Dataset S7. Subcellular information and functional annotations of critical light-responsive genes.

Supplementary References

1. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–7.

2. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol. 2011;29:644–52.

3. Ismail WM, Ye Y, Tang H. Gene finding in metatranscriptomic sequences. BMC Bioinformatics. 2014;15:S8.

4. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–2.

5. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.

6. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 1997;25:3389–402.

7. Lalwani M, Ip S, Carrasco-López C, Day C, Zhao E, Kawabe H, et al. Optogenetic control of the lac operon for bacterial chemical and protein production. Nature Chemical Biology. 2021;17:1–9.

8. Ribeiro DM, Rubinacci S, Ramisch A, Hofmeister RJ, Dermitzakis ET, Delaneau O. The molecular basis, genetic control and pleiotropic effects of local gene co-expression. Nat Commun. 2021;12:4842.

9. Cao Z-J, Gao G. Multi-omics single-cell data integration and regulatory inference with graph-linked embedding. Nat Biotechnol [Internet]. 2022 [cited 2022 Jun 24]; Available from: https://www.nature.com/articles/s41587-022-01284-4

10. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund S, Cook H, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic acids research. 2018;47.

11. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171:1437-1452.e17.

12. Gelman S, Fahlberg SA, Heinzelman P, Romero PA, Gitter A. Neural networks to learn protein sequence–function relationships from deep mutational scanning data. Proceedings of the National Academy of Sciences. 2021;118:e2104878118.

13. Koppel N, Rekdal V, Balskus E. Chemical transformation of xenobiotics by the human gut microbiota. Science. 2017;356:1246–57.

14. Kuypers MMM, Marchant HK, Kartal B. The microbial nitrogen-cycling network. Nat Rev Microbiol. 2018;16:263–76.

15. Vogel L. Health workers demand health coverage for migrants. CMAJ. 2019;191:E270.

16. Salas-Pérez F, Ramos-Lopez O, Mansego ML, Milagro FI, Santos JL, Riezu-Boj JI, et al. DNA methylation in genes of longevity-regulating pathways: association with obesity and metabolic complications. Aging (Albany NY). 2019;11:1874–99.

17. Egan AJF, Errington J, Vollmer W. Regulation of peptidoglycan synthesis and remodelling. Nat Rev Microbiol. 2020;18:446–60.

18. Tu-Sekine B, Kim SF. The Inositol Phosphate System-A Coordinator of Metabolic Adaptability. Int J Mol Sci. 2022;23:6747.

19. Majumdar SK, Kutzner HJ. Myo-inositol in the Biosynthesis of Streptomycin by Streptomyces griseus. Science. 1962;135:734–734.

20. Singhal S. Digest: Structuring interactions in Streptomyces*. Evolution. 2020;74:207–9.

21. Pegtel DM, Gould SJ. Exosomes. Annu Rev Biochem. 2019;88:487–514.

22. Zhang Y, Liu Y, Liu H, Tang WH. Exosomes: biogenesis, biologic function and clinical potential. Cell Biosci. 2019;9:19.

23. Chai Y, Beauregard PB, Vlamakis H, Losick R, Kolter R. Galactose metabolism plays a crucial role in biofilm formation by Bacillus subtilis. mBio. 2012;3:e00184-00112.

24. Schmid J, Sieber V, Rehm B. Bacterial exopolysaccharides: biosynthesis pathways and engineering strategies. Front Microbiol. 2015;6:496.

25. Bumah VV, Morrow BN, Cortez PM, Bowman CR, Rojas P, Masson-Meyers DS, et al. The importance of porphyrins in blue light suppression of Streptococcus agalactiae. Journal of Photochemistry and Photobiology B: Biology. 2020;212:111996.

26. Dahl C, Hell R, Leustek T, Knaff D. Introduction to Sulfur Metabolism in Phototrophic Organisms. In: Hell R, Dahl C, Knaff D, Leustek T, editors. Sulfur Metabolism in Phototrophic Organisms [Internet]. Dordrecht: Springer Netherlands; 2008 [cited 2023 Mar 19]. p. 1–14. Available from: https://doi.org/10.1007/978-1-4020-6863-8_1

27. Zha L, Liu W, Yang Q, Zhang Y, Zhou C, Shao M. Regulation of Ascorbate Accumulation and Metabolism in Lettuce by the Red:Blue Ratio of Continuous Light Using LEDs. Frontiers in Plant Science. 2020;11.

28. Apel K, Hirt H. Reactive Oxygen Species: Metabolism, Oxidative Stress, and Signal Transduction. Annual review of plant biology. 2004;55:373–99.

29. Helvig C, Koener JF, Unnithan GC, Feyereisen R. CYP15A1, the cytochrome P450 that catalyzes epoxidation of methyl farnesoate to juvenile hormone III in cockroach corpora allata. Proceedings of the National Academy of Sciences. 2004;101:4024–9.

30. Jurenka R. Insect pheromone biosynthesis. Top Curr Chem. 2004;239:97–132.

31. Ni J, Yang H, Chen L, Xu J, Zheng L, Xie G, et al. Metagenomic analysis of microbial community structure and function in a improved biofilter with odorous gases. Sci Rep. 2022;12:1731.

32. Chen C, Chen K, Su T, Zhang B, Li G, Pan J, et al. Myo-inositol-1-phosphate synthase (Ino-1) functions as a protection mechanism in Corynebacterium glutamicum under oxidative stress. MicrobiologyOpen. 2019;8:e00721.

33. Schmitz RA, Dietl A, Müller M, Berben T, Op den Camp HJM, Barends TRM. Structure of the 4-hydroxy-tetrahydrodipicolinate synthase from the thermoacidophilic methanotroph Methylacidiphilum fumariolicum SolV and the phylogeny of the aminotransferase pathway. Acta Crystallogr F Struct Biol Commun. 2020;76:199–208.

34. Zeng L-R, Xie J-P. Molecular basis underlying LuxR family transcription factors and function diversity and implications for novel antibiotic drug targets. Journal of Cellular Biochemistry. 2011;112:3079–84.

35. Zhong X, Lu R, Liu F, Ye J, Zhao J, Wang F, et al. Identification of LuxR Family Regulators That Integrate Into Quorum Sensing Circuit in Vibrio parahaemolyticus. Frontiers in Microbiology [Internet]. 2021 [cited 2023 May 10];12. Available from: https://www.frontiersin.org/articles/10.3389/fmicb.2021.691842

36. Chen YT, Zheng RL, Jia ZJ, Ju Y. Flavonoids as superoxide scavengers and antioxidants. Free Radic Biol Med. 1990;9:19–21.

37. Balasaheb Nimse S, Pal D. Free radicals, natural antioxidants, and their reaction mechanisms. RSC Advances. 2015;5:27986–8006.

38. Johnson J-LF, Leroux MR. cAMP and cGMP signaling: sensory systems with prokaryotic roots adopted by eukaryotic cilia. Trends in Cell Biology. 2010;20:435–44.

39. Emiliani V, Entcheva E, Hedrich R, Hegemann P, Konrad KR, Lüscher C, et al. Optogenetics for light control of biological systems. Nat Rev Methods Primers. 2022;2:1–25.

40. Sanders DA, Gillece-Castro BL, Burlingame AL, Koshland DE. Phosphorylation site of NtrC, a protein phosphatase whose covalent intermediate activates transcription. J Bacteriol. 1992;174:5117–22.

41. Swartz TE, Tseng T-S, Frederickson MA, Paris G, Comerci DJ, Rajashekara G, et al. Blue-Light-Activated Histidine Kinases: Two-Component Sensors in Bacteria. Science. 2007;317:1090–3.

42. Liao Y, Bian J, Miao S, Xu S, Li R, Liu R, et al. Regulation of denitrification performance and microbial topology by lights: Insight into wavelength effects towards microbiota. Water Research. 2022;119434.

43. Georgiou CD, Sun HJ, McKay CP, Grintzalis K, Papapostolou I, Zisimopoulos D, et al. Evidence for photochemical production of reactive oxygen species in desert soils. Nat Commun. 2015;6:7100.

44. Lovley D. Syntrophy Goes Electric: Direct Interspecies Electron Transfer. Annual Review of Microbiology. 2017;71.

45. Hernandez RE, Gallegos-Monterrosa R, Coulthurst SJ. Type VI secretion system effector proteins: Effective weapons for bacterial competitiveness. Cellular Microbiology. 2020;22:e13241.

46. Lee S, Sowa ME, Watanabe Y, Sigler PB, Chiu W, Yoshida M, et al. The structure of ClpB: a molecular chaperone that rescues proteins from an aggregated state. Cell. 2003;115:229–40.