Interplay between high-energy quenching and state transitions in *Chlamydomonas reinhardtii*: a single-cell approach.

Aliénor Lahlou[1, 2*], Marcelo Orlando[3], Sandrine Bujaldon[3], William Gaultier[2], Eliora Israelievitch[3], Peter Hanappe[2], Thomas Le Saux[1], Ludovic Jullien[1], David Colliaux[2*], Benjamin Bailleul[3*]

1. PASTEUR, Department of Chemistry, École Normale Supérieure, PSL University, Sorbonne University, CNRS, Paris, France
2. Paris Research, Sony Computer Science Laboratories, Paris, France
3. UMR7141: Chloroplast Biology and Light Sensing in Microalgae, Institut de Biologie Physico-Chimique, Sorbonne University, CNRS, Paris, France

[*] Corresponding authors alienor.lahlou@sony.com , David.Colliaux@sony.com, bailleul@ibpc.fr

Abstract

Studying cell-to-cell heterogeneity is essential to understand how unicellular organisms respond to stresses. Here, we combine single-cell measurements of chlorophyll *a* fluorescence with machine learning to study light stress responses in the model green alga *Chlamydomonas reinhardtii*. This framework allows us to score the extent of three high-light responses at the single cell level – state transitions, high-energy quenching, and photoinhibition – and reveals significant cell-to-cell response heterogeneity. The unexpectedly strong correlation between state transitions and high-energy quenching also indicates a cellular control over the relative extent of these processes that cannot be detected by bulk measurements. Accordingly, new models were derived for the coordinated functional response to light stress in *Chlamydomonas*. The improved phenotypic characterization allowed by this approach will facilitate the identification of genetic determinants of intercellular heterogeneity through single cell–omics and isolation of cells displaying traits of interest for biotechnology.

Teaser: Scoring of NPQ components in single cells algae unveils the general orchestration of high light photosynthetic responses.

Short title: Scoring of NPQ components in *C. reinhardtii*

Field codes: Biophysics of photosynthesis / Single cell physiology

Abbreviations

ChlF: endogenous chlorophyll *a* fluorescence

PSII: photosystem II

NPQ: Non-Photochemical Quenching of chlorophyll fluorescence

LEF: Linear Electron Flow

CEF: Cyclic Electron Flow

MIN: Tris-minimal culture medium

WT: wild type

STT7: state-transition 7 kinase

LHCII: Light Harvesting Complex II

LHCSR: stress-related light-harvesting complex

PAM: Pulse Amplitude Modulated

SP: Saturating Pulse of light

HL: High Light

F_m': maximal fluorescence yield in light-adapted sample

F_m: maximal fluorescence yield in dark-adapted sample

qT: NPQ component due to state-transition

qE: energy-dependent quenching

qI: photoinhibition-related quenching

Pop-0: population of *stt7-1* without HL-treatment without qI

Pop-qE: population of *stt7-1* with HL-treatment without qI

Pop-qI: population of *stt7-1* without HL-treatment with qI

Pop-qT: population of wild type *wt4a*$^-$ without HL-treatment without qI

LDA: Linear Discriminant Analysis

3D NPQ space: three-dimensional space with axis defined by elementary NPQ components

q̃T: axis of the 3D NPQ space defined by ***Pop-qT***

q̃E: axis of the 3D NPQ space defined by ***Pop-qE***

q̃I: axis of the 3D NPQ space defined by ***Pop-qI***

Introduction (1383 words)

Most of the existing knowledge on photosynthesis has been derived from measurements at the organism or tissue level for plants, and bulk measurements for unicellular organisms (1). Those studies have revealed major insights into the biology of photosynthesis, yet they fail to account for the cell-to-cell heterogeneity, which is observed even within monoclonal populations (2,3). In contrast, accessing the biophysical heterogeneity at the single-cell level can reveal a richer set of information since the individual responses are not smoothed-out by averaging. These data can be used to confront hypothetical models, draw out interactions between biophysical traits, and perform selection of individuals endowed with traits of interest.

Population-level studies typically use reverse genetics (or forward genetic screens) to create knockout mutants of one trait and examine the impact on other traits. However, mutagenesis can lead to secondary mutations and physiological changes, and often fails to capture the complexity of trait interaction, such as nonlinearity or threshold effects. In contrast, single-cell studies offer significant advantages for understanding the interplay between co-occurring traits in constant genetic background. When trait dispersion is sufficiently high, it provides a continuum of natural trait variations in synchronised isogenic cells in the same environment, a situation that is impossible to achieve when comparing two populations (4). The present study serves as a proof of concept for leveraging the often underestimated benefits of single-cell analyses over population-level studies by exploiting cell-to-cell heterogeneity to gain insights into the interplay between traits.

With regards to single-cell tools, there has been a surge in the development of single-cell sequencing technologies probing "omics" data from genomics, proteomics, metabolomics, lipidomics and glycomics (5), all of them requiring cell lysing. A new challenge is to integrate them with non-invasive physiological measurements and further connect changes in gene expression and proteins, sugar and fat synthesis with physiological characteristics. For this purpose, being able to quantify a physiological process using a non-invasive and specific biological observable is highly desirable.

Chlorophyll a fluorescence (ChlF) emitted by photosystem II (PSII) is common to all photosynthetic organisms. It is the preferred observable for non-invasive studies of photosynthesis as it is highly sensitive and benefits from a century-long history of research (6,7). After absorption of light by the pigments of the PSII light-harvesting system, excited chlorophyll can relax through three main de-excitation pathways. Firstly, the absorbed light energy can initiate photosynthetic electron transfer through photochemistry (8). Secondly, light energy can be dissipated through non-radiative de-excitation, releasing heat. Lastly, it can be dissipated by radiative de-excitation, in the form of fluorescence – that can be measured optically. A useful framework for assessing PSII performances is the "lake" model, which postulates that numerous PSII complexes are integrated into a shared light-harvesting antenna and compete for excitons (9). In this model, the relative

efficiencies of photochemistry, heat dissipation and ChlF emission are determined by the values of the respective rate constants associated to each process. While the rate constant governing fluorescence emission remains invariant, the rate constants of the other processes can be modulated by biological factors. Photochemistry is modulated by the saturation level of the photosynthetic chain, while heat dissipation can be influenced by the regulated production of non-photochemical quenchers in the light-harvesting antenna. Because of such kinetic competition between the three de-excitation pathways, variations of ChlF can provide valuable information on changes in photochemistry or heat dissipation. ChlF has been widely used as a diagnostic tool to assess the overall status of photosynthesis and to diagnose stress in plants and microalgae, down to the single-cell level (4,10–21). In this work, we selected the response of photosynthetic organisms to high light (HL) stress as a case study. This topic has become the focus of many studies because better understanding the response to high light stress is presently debated for improving the photosynthetic efficiency and biomass production (22–26). We aim to show how single-cell ChlF data can be used to score the various light-stress responses, and thus study and exploit intercellular variations within these responses, in all their complexity.

In natural environments, the intensity and wavelength distribution of sunlight fluctuates significantly over time scales ranging from seconds to months, caused by factors such as surface wave motion, vertical movement of phytoplankton, overlapping and shifting leaves, and passing clouds (27,28). Photosynthetic organisms have evolved various processes to cope with these fluctuations, enabling them to dissipate excess energy when light intensity peaks, and maximize photon intake during low light conditions (28–30). In particular, several processes mitigate the risk of reactive oxygen species production and PSII photodamage under high light stress, by regulating the photon absorption capacity of the photosystems and/or the efficiency of heat dissipation. These processes, altogether, induce fluorescence changes, and are referred to as non-photochemical quenching (NPQ) (31,32). They have been grouped in three categories reflecting the underlying set of biochemical reactions. In this study, these groups are referred to as NPQ components and associated biological processes (33). The first group corresponds to high energy quenching (qE), involving the rapidly reversible generation of a non-photochemical quencher in the light-harvesting antenna of PSII and is induced by low luminal pH. A second type of non-photochemical quenching is associated with photodamaged PSII (qI), has a slower relaxation and its mechanism is not yet fully understood (34). The last process, state transitions (qT), involves the reversible movement of a fraction of the light-harvesting complexes between the photosystems. Although qT is not a form of non-photochemical quenching *per se,* it changes the absorption cross-section of PSII with resulting in variations of ChlF emission (35,36). Our study aims to answer the following questions: *What are the cell-to-cell variations in the extent of a given NPQ component when it is the only active, or between NPQ components when more than one are active? Can we benefit from cell-to-cell variations to explore the interaction between NPQ components and learn about the biological orchestration of these high light responses?*

To address these questions, it is crucial to untangle the light-stress responses that coexist and often complement each other. This task is challenging, because qE, qT, and qI occur simultaneously over time scales ranging from seconds to hours, each modifying the ChlF kinetics, resulting in complex, intertwined signatures. Established metrics have been empirically developed to quantify NPQ

components from ChlF traces, but they often rely on basic mathematical treatments of specific values such as timestamps, inflection points or local extrema (7,37–40) or more recently on machine learning (41). To better separate contributions from concomitant processes, complementary observables such as fluorescence lifetime or multispectral acquisition can be introduced (42,43) but these strategies comes at the cost of complexifying instruments and protocols. In this report, we present a novel computational method to analyze complex ChlF traces and separate the contributions of co-occurring processes. It makes the *a priori* strong hypothesis that the behavior of a population (expressing multiple processes) can be satisfactorily accounted for by the referential of populations of a training set (expressing one elementary process). We chose to perform this analysis on the model unicellular green alga *Chlamydomonas reinhardtii* (30) because this organism allows for easy isolation and manipulation of elementary processes involved in high light response (high energy quenching, state-transition and photoinhibition–related quenching). By selecting appropriate mutants, choosing specific growth conditions, or using customized treatments, we can build a dataset of reference populations, each exhibiting at most one NPQ component (qE, qT, qI, or none). We trained an algorithm borrowed from the domain of image compression (44) to extract the key ChlF patterns representing elementary kinetics. Then, we introduced an algorithm designed for class-separability problems (45) to project the ChlF traces into a three-dimensional space where each axis represents an NPQ component. This transformation is then applied to wild-type (WT) strains to score the three types of NPQ components (qE, qT, qI). This approach, based on well-characterized responses to light stress, provides clear and unbiased scores for NPQ components in the WT strain. As a result of our experimental design, we separated the analysis of qI from that of qE and qT. In our case study, we quantify the NPQ components in synchronised isogenic cells to investigate the cell-to-cell heterogeneity of qE and qT separately. Then, we exploit the observed cell-to-cell heterogeneity of qE and qT when they are co-expressed to study the interplay between the underlying biological processes.

Results

New instrument and protocol to probe high light responses in single cells

We developed a fully automated microscope capable of applying versatile illumination protocols to capture ChlF traces from microalgae immobilized on agarose. This setup allows to monitor the response of each microalga independently over time (see Methods, Fig. 1A, S1, S6). Our illumination protocols (see Methods, Fig. S8, S10, Table S2) are similar to those used in previous non-photochemical quenching (NPQ) studies on *Chlamydomonas reinhardtii* (37,38): high-intensity actinic light induces the NPQ response, while brief saturating pulses (SP) provide temporary saturation of the photosynthetic electron transport chain. The reference protocol applies high light (HL) for 15 min followed by a 15 min of dark relaxation, with SPs delivered every 20-sec to achieve high temporal resolution of ongoing processes (see Fig. S9 for the evaluation of the actinic effect of the SP). During SPs, when the photochemical yield becomes minimal, ChlF acts as a reliable indicator of competing non-photochemical quenching processes. Although pulse-amplitude modulated (PAM) fluorescence microscopy can image individual microalgal cells, the low photon count is often too low compared to camera noise, making single-cell resolution difficult (10,12,20,21,46). In contrast to traditional PAM methods (47,48), we chose direct SP-induced

fluorescence as the optimal approach to capture all photons emitted from the sample and study the response of *Chlamydomonas reinhardtii* to constant HL at the single-cell level. Since each SP delivers a constant amount of photons, changes in ChlF amplitudes in response to the SP reflect changes in ChlF yield. The SP-induced fluorescence is named F_m before the illumination and F_m' during and after the HL treatment. This approach improves the signal-to-noise ratio, while overcoming limitations often associated with direct fluorescence (see Discussion section).

For each experiment, we submitted hundreds of individual cells (Methods and Table 1) to our illumination protocol. We applied an image segmentation algorithm to each frame to retrieve the ChlF temporal trace of each algae, called F_m' trace hereafter (Fig S6). An example individual F_m' traces of a wildtype (WT) population – we chose the strain *wt4a*$^-$ in this work, otherwise specified– expressing qE and qT is shown in Fig 1A.

Building a reference dataset of elementary qE, qT and qI traces

Quantifying the relative contributions of the various NPQ components in the traces shown in Fig 1A is notoriously challenging with classical methods that depend on identifying local extrema and inflection points. To address this, recent approaches have introduced unsupervised multivariate analysis of NPQ components (41). In this study, we developed a machine learning-based workflow to accommodate the variability in the kinetic properties of each NPQ component. As a first step, we defined three distinct populations – specific strains conditioned under appropriate settings – each expressing a single dominant NPQ component, with minimal expression of the others. These populations provided elementary traces representing a single NPQ component, capturing variability in both amplitudes and time constants. Combined with a fourth population lacking all NPQ component, these traces formed the training dataset. Our approach assumes that the ChlF trace of a cell expressing multiple processes, such as in WT (Fig 1A), can be modelled as a linear combination of elementary traces. Leveraging extensive knowledge and resources on high light stress responses in *C. reinhardtii*, we established strains and conditions where particular NPQ traits are either present or absent (37,42,49).

In *Chlamydomonas reinhardtii*, the high-energy quenching (qE) is the main heat dissipation mechanism preventing PSII photoinhibition (50). qE occurs and relaxes within seconds and is regulated by conformational changes in the Light Harvesting Complexes Stress Related 1 and 3 (LHCSR1 and LHCSR3) induced by low lumenal pH (29,51). For our experiments, we used cells grown under low light in acetate-containing medium to create qE-free conditions, as these conditions suppress LHCSR gene expression (see western blot in Fig. 1B and (37,38,52)). In contrast, to induce qE, we subjected cells to a 1h20-4h HL-treatment under the microscope after transferring them to acetate-free medium (see Methods), which triggers LHCSR production (Fig. 1B and (37,38,52)).

State transitions (qT) occur and relax within minutes, regulated by LHCII phosphorylation via a kinase/phosphatase system (35,36,53,54). The regulation of the activity of state-transition 7 kinase (STT7) by the redox state of the PQ pool is essential for qT. In state I, when all mobile LHCIIs are bound to PSII, the absorption cross-section is large, resulting in high fluorescence

levels. In state II, fluorescence decreases when LHCII binds to PSI. The STT7 mutant is unable to perform qT and remains blocked in state I. For our experiments, we use the WT strain (*wt4a*$^-$) to study conditions with qT and the STT7 knockout mutant (stt7-1 #a6, (53)), generated in the *wt4a*$^-$ genetic background, to study conditions without qT (see STT7 quantification by western blots in Fig 1B).

Based on this, the untreated STT7 knockout mutant was expected to lack both qE and qT. However, we observed a significant fluorescence decrease in this population during the 15 min HL period, which did not relax during the subsequent dark period (see Fig. S10). Several slowly relaxing NPQ components have been described, including qI quenching (34), xanthophyll cycle-dependent qZ quenching (40,55) and sustained qH quenching (56). However, key proteins involved in qH quenching are absent in the *Chlamydomonas reinhardtii* genome (56,57). This slowly induced and slowly relaxing quenching was exacerbated by the chloroplastic translation inhibitor lincomycin (Fig S4), consistent with qI quenching (34). PSII photodamage and qI occur when excessive light disrupts the balance of ongoing degradation and repair processes of reaction centers, particularly the D1 protein encoded in the chloroplast genome (58,59). Therefore, we attribute the slowly reversible quenching observed in our dataset to qI, altough a contribution from qZ cannot be strictly excluded (40). Manipulating qI is more challenging than manipulating other NPQ components, as there are no mutant providing qI-free ChlF traces. Although the primary goal of this study was to investigate the heterogeneity and interaction of qE and qT, the inclusion of elementary traces for qI in the training dataset was essential for the machine learning-based approach to work effectively. We found that repeated exposure to our reference protocol progressively reduced the extent of qI, likely reaching a steady-state were PSII repair balanced PSII damage (59). Therefore, for our training dataset, we used the third or fourth iteration of the reference protocol to create conditions without qI contribution and the first iteration to represent conditions with qI.

We used the above considerations to generate the required four populations expressing only qE, only qT, only qI or none of those (Table 1). In practice, we used the 4th repeat of our reference protocol on the untreated WT population to obtain elementary traces of qT (called ***Pop-qT*** hereafter), as it expresses the STT7 kinase but not LHCSR3 proteins. The 4th repeat of our reference protocol on the HL-treated *stt7-1* mutant provided elementary traces of qE (***Pop-qE***) since this mutant then expresses LHCSR3 proteins but lacks STT7 kinase. We considered the untreated *stt7-1* population as our control for the absence of both qE and qT. In this population, the 1st repeat of our protocol was used for the elementary traces of qI (***Pop-qI***), whereas the 3rd repeat was used as the reference population displaying none of the NPQ components (***Pop-0***). We used the 3rd repeat instead of the 4th as the best compromise to minimize qE and qI because the mutant *stt7-1* is known to be more prone to express LHCSR than the WT when exposed to HL (see Fig 1B, Fig S10b and (37,60,61)).

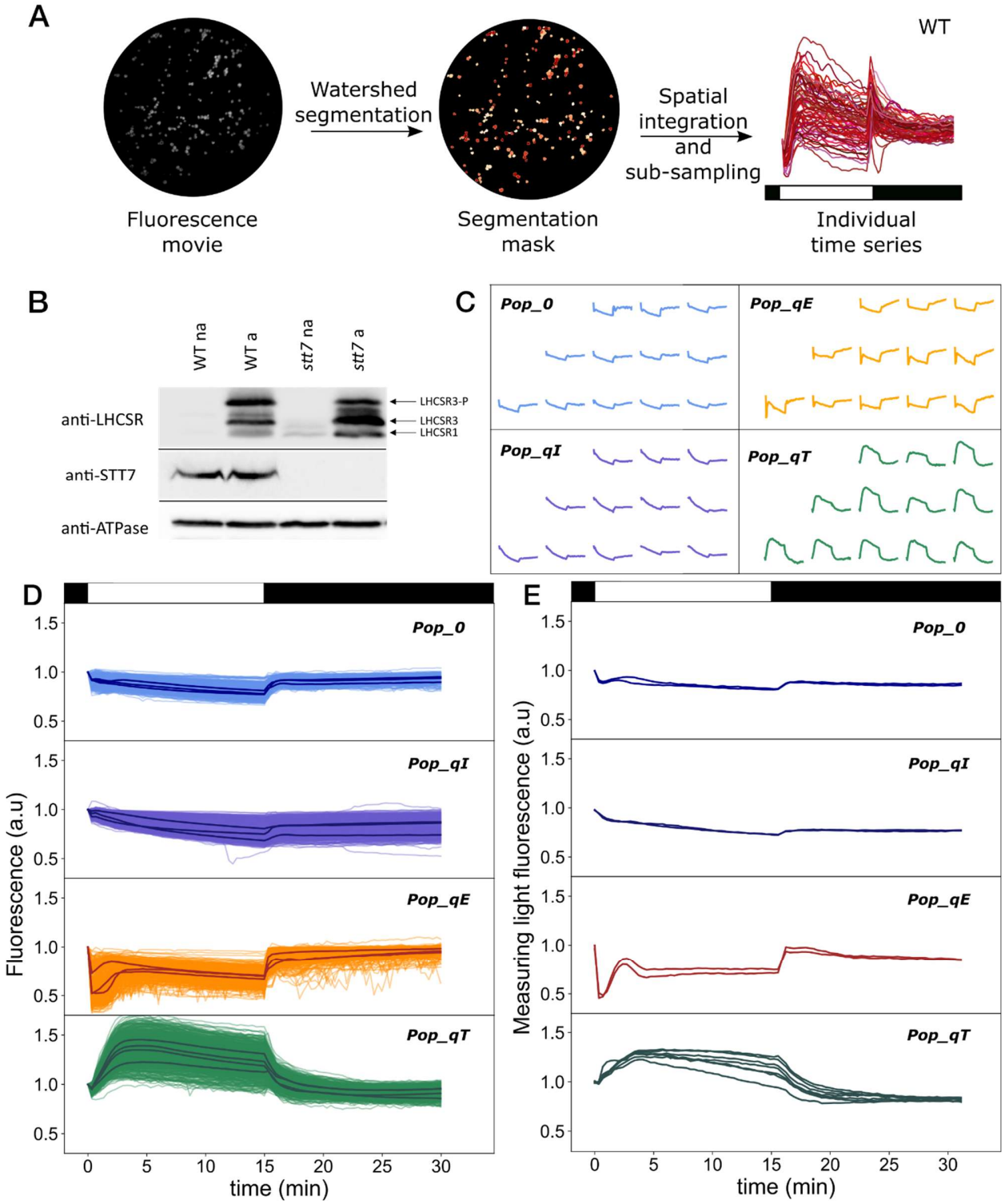
A
Fluorescence movie
Watershed segmentation
Segmentation mask
Spatial integration and sub-sampling
WT
Individual time series
B
WT na
WT a
stt7 na
stt7 a
anti-LHCSR
anti-STT7
anti-ATPase
LHCSR3-P
LHCSR3
LHCSR1
C
Pop_0
Pop_qE
Pop_qI
Pop_qT
D
Fluorescence (a.u)
time (min)
E
Measuring light fluorescence (a.u)
time (min)

Figure 1: *Acquisition of the reference dataset representing elementary ChlF traces for each NPQ component*. **A**: *Extraction of the F_m' traces from single cells of the WT imaged with the fluorescence microscope.* Sample displaying qE and qT: population of the HL-treated WT strain *wt4a*$^-$. **B**: immuno-quantification of LHCSR (LHCSR1, LHCSR3 and phosphorylated LHCSR3) and STT7 proteins, as well as AtpB (beta-ATP synthase subunit) as a loading control. Wt4a$^-$ and STT7 knock-out mutant were sampled as described in Methods, before (na) and after (a) HL-treatment. **C:** *Randomly sampled F_m' traces of the four populations used to build the reference dataset*. Samples: populations of ***Pop-0*** (untreated *stt7-1* mutant, 3rd repeat), ***Pop-qI*** (untreated *stt7-1* mutant, 1st repeat), ***Pop-qE*** (HL-treated *stt7-1* mutant, 4th repeat) and (untreated *wt4a*$^-$, 4th repeat); **D**: *F_m' single cell traces of the four populations used to build the reference dataset*. The panels represent the superimposed single cell traces of > 400 cells from at least 3 independent biological replicates (see Table 1 for details) and the average trace is shown as black lines; **E**: *Validation of the microscope setup*. *F_m'* traces measured on the same four populations as in **D** using a reference macro fluorescence imaging set-up (see Methods) on two independent biological replicates for *stt7*-1 (5 for the *wt4a*$^-$). In **D** and **E**, the *F_m'* data are normalized to the initial F_m, in the dark-adapted cell before the start of the actinic light period. The illumination protocol (HL: white boxes; dark: black boxes) are indicated on the top of the **A**, **D** and **E** panels.

Fig. 1C displays several randomly sampled traces for the four populations, demonstrating the diversity of the ChlF responses in the training set. The whole dataset is plotted in Fig. 1D for ***Pop-qE***, ***Pop-qT***, ***Pop-qI*** and ***Pop-0***. These *F_m'* traces were collected from movies displaying 100 to 300 algae cells in at least three independent biological replicates (see Table 1, Methods and Fig. S5 for sample preparation details). The algae populations were synchronised but not isogenic (Methods). Mean responses calculated from individual traces of each biological replicate are shown as a black line in Fig. 1D. The data convincingly reproduce the classical trends of the NPQ components (37,38,42,49):

- ***Pop-qE*** traces show a steep decay at the onset of light and rapid relaxation of the fluorescence decline at the end of light consistent with the contribution of qE. Both phases are too fast to be resolved with the time resolution used. In most cells, qE reached a transitory maximum before partially relaxing during the HL period. Such transitory qE has been reported in *Chlamydomonas* (42,49,60) or in plants, where the relaxation of qE in the light was attributed to the concomitant activation of carbon fixation in the Calvin-Benson-Bassham cycle (62).
- ***Pop-qT*** traces show a gradual increase of fluorescence at the onset of light and a gradual relaxation during the dark period, both phases lasting a few minutes, indicating qT and the movement of LHCII from PSI (state II) to PSII (state I) in the HL period, and back to PSI in the dark. It might seem counter intuitive that LHCIIs switch from PSI to PSII under high light conditions. Indeed, PSII is already under excessive light pressure and this should favor PQ pool reduction, STT7 activation and state II. However, this behavior has been documented previously (37) and attributed to kinase inactivation by negative redox control by the ferredoxin/thioredoxin system (63) or to steric hindrance preventing LHCII kinase phosphorylation due to HL-induced conformational changes in PSII (64).
- ***Pop-qI*** traces show a continuous decline of *F_m'* during the HL period, which does not fully recover during 15 min of darkness, characteristic of the contribution of qI, with an overall decay between the initial and final *F_m'*.
- ***Pop-0*** is not completely flat as would be expected if all NPQ components were absent. However, the slight decrease of *F_m'* at the light onset, representative of qE, and the slight decay of

fluorescence between the beginning and the end of the experiment, representative of qI, are both negligible compared to their equivalent in the ***Pop-qE*** and ***Pop-qI***, respectively.

Moreover, these results compare well with the fluorescence traces measured at the population level using a conventional imaging fluorometer (macro set-up, see Methods) by following the same light and dark protocols, including the actinic and saturating pulse (SP) sequences (Fig. 1E) (37,65). The difference with the single cell microscope is the detection of fluorescence induced by short pulses of low light intensity set not to perturbate the physiological state of photosynthesis rather than direct fluorescence from SP. This agreement validates our instrument and acquisition protocol using direct fluorescence.

Name	n° of experiments	n° of algae	strain	HL-treatment	reference protocol repeat	qT	qE	qI
Pop-0	3	433	*stt7-1*	∅	3rd			
Pop-qI	3	453	*stt7-1*	∅	1st			✓
Pop-qE	3	480	*stt7-1*	✓	4th		✓	
Pop-qT	4	936	*wt4a⁻*	∅	4th	✓		

Table 1: Description of the dataset (strains, conditioning, and acquisition). ***Pop-0***: untreated *stt7-1* (third repeat of the reference protocol); ***Pop-qI***: *untreated stt7-1* (first repeat of the reference protocol); ***Pop-qE***: *HL-treated stt7-1* (fourth repeat of the reference protocol); ***Pop-qT***: *untreated wt4a⁻* (fourth repeat of the reference protocol). Shaded cells: NPQ component absent, checkmark: NPQ component present.

Using the reference dataset to build a 3-dimensional (qE, qT, qI) space

We used the reference dataset with a machine learning framework to derive a quantitative estimate of the three NPQ components from any ChlF trace. This process involves constructing a 3D space where the three axes represent the elementary biological processes associated with the NPQ components qE, qT, and qI, respectively (3D NPQ space hereafter). The framework performs two consecutive dimension reduction steps, each optimizing a different metric, to transform the F_m' traces into a biologically interpretable 3D vector. Once trained, the framework can be used to analyse new data. A new F_m' trace is first reduced to a vector of coefficients that describe the linear combination of basic waveforms reconstructing the F_m' trace. This vector is then projected onto a 3D space with axes denoted q̃T, q̃E, and q̃I (Fig. 2A), attributing a score to each NPQ component. It is to note that the scores attributed to the different NPQ components, along the three axes, cannot be simply related to the NPQ metric from the literature.

The first step of dimension reduction aims at learning a minimal set of ChlF waveforms from the training set to reconstruct generic input ChlF traces with minimal reconstruction error. It enables the interpretation of new data as a combination of traces learned during training, if our

working hypothesis is valid. Indeed, we postulate that data comprising the three NPQ components can be considered as a linear combination of the features described by the training set. At the opposite, failure to reconstruct the traces within the training error margin either invalidates this assumption, or indicates that a supplementary NPQ component is present, that is not covered by the training dataset. We implemented a dictionary learning method (see Methods) (44), where the convergence of the algorithm depends on hyper-parameters such as the sparsity level, number of atoms, and the size of the training set (Fig. S11). After evaluating their influence on the reconstruction error (see Fig. S12), we selected a reconstruction error threshold (2.10^{-4}) corresponding to 10 dictionary atoms (Fig. 2A).

The second step is a supervised dimension reduction aimed at maximizing the separability of the annotated training populations. This step learns a matrix to project the output of the first step onto a 3D space using a method based on Linear Discriminant Analysis (LDA, see Methods) (45). The LDA metric used for training maximizes separability by minimizing the datapoint distances within a single training population (***Pop-0, Pop-qE, Pop-qT, Pop-qI***) while maximizing the separation between different populations. This supervised machine learning approach requires labels for the training populations. LDA learns a projection matrix from the 10-dimensional dictionary output space to a 3D space by identifying three hyperplanes that effectively separate the four training populations. This dimensionality constraint explains the need to select four classes to construct the reference dataset. Once the 3D projection was obtained, we defined the final 3D basis of the NPQ elementary components by assigning the x, y and z axes to the directions along which the training populations ***Pop-qE, Pop-qT*** and ***Pop-qI*** are spread (principal components) and redefining the origin thanks to the population ***Pop-0***. Therefore, the point cloud of each training population extends along a separate axis and, since each population represents an elementary HL stress-response, the axes were named q̃T, q̃E or q̃I according to the corresponding NPQ components (see Methods).

The 3D projection results of our machine learning framework applied to the training dataset are shown in Fig. 2B,C,D with each panel representing a 2D projection for clarity. The q̃T, and q̃E axes are well defined, validating our approach to control the molecular processes through the STT7 mutation for qT and HL-treatment for qE. Firstly, the scatterplots of populations expressing the qT or qE component extend only along the q̃T or q̃E axis, respectively. Secondly, the scatterplots of those not expressing the component overlap at the origin of the corresponding axis. In the plane (q̃T, q̃E), the populations ***Pop-qI*** and ***Pop-0*** overlap since they exhibit neither qE nor qT (Fig. 2B). The scatterplot of the population ***Pop-qI*** along the q̃I axis is also consistent, eventhough it is less separated from the other populations compared to the other axis cases. We comment in more details the qI axis in the supplementary materials (Fig. S16).

The two consecutive steps applied to the training set allow the qualitative separation of point clouds corresponding to populations exhibiting distinct NPQ components. Beyond class separation, this representation also allows a quantitative scoring of the NPQ components. Fig. 2E visually demonstrates that the amplitude of the different NPQ components in the fluorescence response increases as the corresponding data point moves away from the origin. For example, the ***Pop-qT*** population extends along the q̃T axis, with points further from the origin showing a more significant increase in fluorescence at light onset and relaxation to the basal level in the dark. Similarly, the ***Pop-qE*** population exhibits a gradual increase in fluorescence drop at light onset

along the q̃E axis. We present the traces of the ***Pop-qI*** and ***Pop-0*** training populations for comparison and provide the evolution of the fluorescence trace along the q̃I axis in Fig. S16.

We then compared the results of the machine learning framework with those based on *ad hoc* metrics used on bulk populations in (37) to quantitatively measure the extent of NPQ components. These *ad hoc* metrics are based on F_m' variation over specific time phases, identified as representative of elementary NPQ contributions. We observed a quantitative correlation between the NPQ metrics and our scores (see Fig. S13).

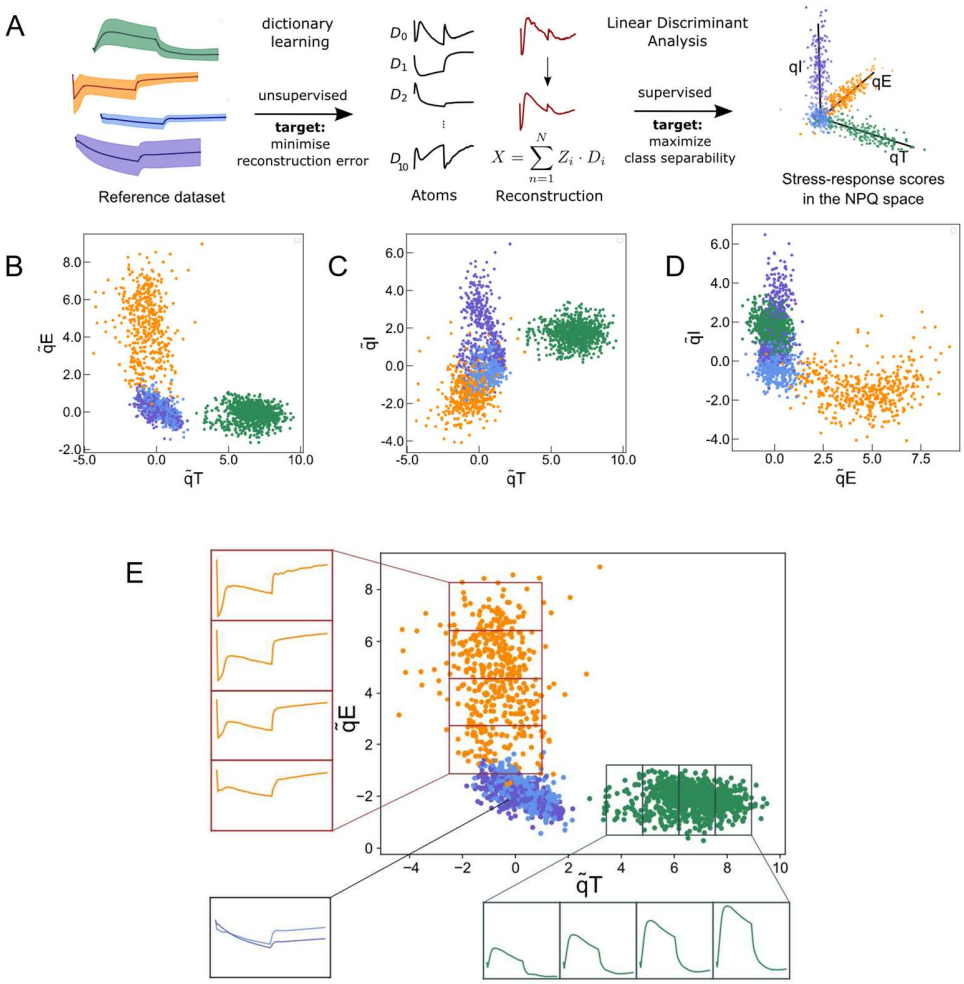

Figure 2: *From the reference dataset to a base of elementary NPQ components in a 3D space (dataset: see Table 1).* **A**: Flow of the process. A sub-set of the reference dataset traces is first used to perform unsupervised training of an algorithm of dictionary learning which learns to reconstruct the traces with a minimal set of elementary traces (atoms). The resulting reconstructed traces are then combined with the labels of the four training populations to perform a Linear Discriminant Analysis (LDA). It results in a vector in a 3D space, where each axis corresponds to a population of the reference dataset expressing an elementary stress. The axis is named according to this corresponding population: q̃T, q̃E or q̃I; **B, C, D**: 2D projections of the reference dataset onto the 3D NPQ space after performing Linear Discriminant Analysis on the atom code (dimension 10) from applying dictionary learning. **B**: (q̃T, q̃E), **C**: (q̃T, q̃I), **D**: (q̃E, q̃I). **E** Qualitative analysis of the evolution of the fluorescence traces along the q̃E and q̃T axis. The orange population corresponds to the population ***Pop-qE***. The span of (98% of) the population along the q̃E axis is split into 4 equal parts, and the average fluorescence trace within each quarter are computed using the fluorescence traces (normalized to the first point) corresponding to the data points falling inside the quarter. The same is performed for the ***Pop-qT*** population (green point cloud) along the q̃T axis. For comparison, the average of the fluorescence traces (normalized to the first point) of the ***Pop-qI*** and ***Pop-0*** populations are also plotted. For all plots, the span of the y-axis is equal to 0.7, with the first point falling at position 1.

Cell-to-cell variations in qE and qT

We used the 3D projection of traces to assess the cell-to-cell dispersion of ***Pop-qT*** and ***Pop-qE*** populations along the q̃T and q̃E axes, respectively. The dispersions of qE and qT in these two populations appear significant. This observation prompted us to perform a series of additional experiments to distinguish between the contribution of biological cell-to-cell variations and experimental noise such as fluorescence noise, instrumental noise, variations of incident light in the field of view, sample-related heterogeneity or imprecise dictionary reconstruction. To assess the level of experimental noise, we exploited our microscope setup, which allows us to apply the illumination protocol to the same cell maintained in the same position. We estimated that the distance in 3D between projections of the same cell under two consecutive repeats would provide a good estimate of the experimental noise, assuming the HL stress response remains consistent between the two repeats. Therefore, we compared the distribution of pairwise distance between two consecutive repeats on the same cell ($D_{i,i}$) to that between all the other cells in the repeat ($D_{i,j}$). If the distance distributions of $D_{i,i}$ and $D_{i,j}$ are similar, this suggests that experimental noise governs the scatter of the points cloud. Conversely, if the distribution of $D_{i,j}$ exceeds that of $D_{i,i}$ this would indicate that the scatter of the point clouds is mainly due to cell-to-cell biological variations. While the training set consisted in synchronised non-monoclonal populations, there we performed new experiments with synchronized monoclonal populations of untreated *wt4a*$^-$ (***Pop-qT***) and HL-treated *stt7-1* mutant (***Pop-qE***) to minimize genetic diversity and the contribution of cell cycle to intercellular variations (see Methods). We collected F_m' traces by applying the same illumination protocol used on the training populations (Fig. 1D), projected them into the 3D space described above, and retrieved the NPQ scores.

First, we observed that the distributions of NPQ scores after two consecutive repeats were statistically similar (Fig. 3A,D), indicating that the average response to HL remained identical (the distribution of the remaining scores and the tests of statistical equality are presented in Fig. S15 and Table S3). We investigated the influence of illumination heterogeneity in the microscope field of view and found no significant correlation with the value along each NPQ-axis (Fig. S14A,B,C).

We then demonstrated that the distance distribution was twice as narrow for $D_{i,i}$ as for $D_{i,j}$ (Fig. 3B,E; a statistical test validated that the distribution variances are different, see Table S3). $D_{i,i}$ is an upper estimate of the experimental noise, as it includes a slight contribution from "real" biological variations in the HL response of each cell between the two repeats. Thus, we conclude that the scatter of the cloud points is mainly dominated by intercellular biological variations.

The distributions displayed in Fig. 3A,D indicate that the variation coefficients (standard deviation relative to the mean) are ~40% and ~20% for q̃E in ***Pop-qE*** and q̃T in ***Pop-qT***, respectively. Interestingly, the variation coefficients are not statistically different from those of the non-monoclonal synchronized algae for the reference dataset (Fig. 3C,F), which supports that the distribution is not governed by a genetic diversity. We further showed that the cell surface measured in the field of view does not alone explain the variance either, and that there was no visible heterogeneity over the sample that could be attributed to the sample design (Fig. S14G,H,I). The results obtained from populations expressing only one NPQ component confirmed that the intercellular dispersions of NPQ scores are significant and dominated by biological variations.

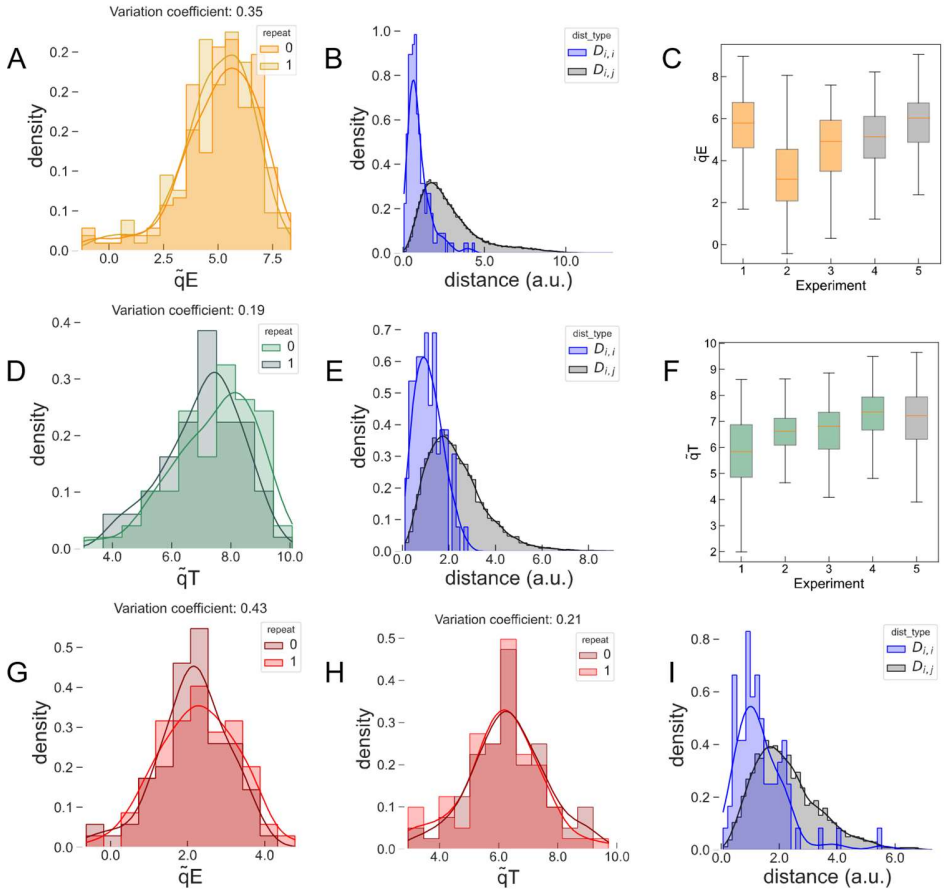

Figure 3: *Intercellular variations in qE and qT in monoclonal synchronised populations*. **A**, **D**: Distribution of the q̃E (**A,** 176 algae) and q̃T (**D,** 77 algae) values of two consecutive repeats of the same experiment on isogenic and

synchronised cells of the HL-treated *stt7-1* (**A**) and untreated WT *wt4a*$^-$ (**D**) without qI. We evaluate the variation coefficient at 19% for qT, 35% for qE; **B**, **E**: Distributions of the cell-to-cell ($D_{i,i}$ – blue, SD: 0.7 (**B**), 0.3 (**E**)) *vs* cell-to-population ($D_{i,j}$ – black, SD: 1.8 (**B**), 1.5 (**E**)) distances in the 3D NPQ space retrieved from **B**: two isogenic HL-treated *stt7-1* populations (2 consecutive repeats; 176 algae); **E**: HL-treated *wt4a*$^-$ population without qI (**E**; 2 consecutive repeats; 77 algae). **C**: Dispersion of the q̃E values collected from the training set (orange) and experiments with isogenic synchronized populations (grey); **F**: Dispersion of the q̃T values collected from the training set (green) and experiments with isogenic synchronized populations (grey); **G**, **H**: Distribution of the q̃E (**G**) and q̃T (**H**) values of two repeats of the same experiment on isogenic and synchronised cells of the HL-treated *wt4a*$^-$ population without qI. We evaluate the variation coefficient at 21% for qT, 43% for qE; **I**: Distributions of the cell-to-cell ($D_{i,i}$ – blue, SD: 0.7) vs cell-to-population ($D_{i,j}$ – black, SD: 1.3) distances in the 3D NPQ space retrieved from two isogenic *wt4a*$^-$ populations (2 consecutive repeats; 77 algae) See Table S3 for statistical tests related to the figure.

Projection of wild type traces in the 3D NPQ space

We then applied our approach to populations expressing both qE and qT and exploited the natural intercellular variations of these traits to study their interaction. Using the same method, we analysed the projection of single cell F_m' traces from the HL-treated WT *wt4a*$^-$ in 3D space. First, the monoclonal population *wt4a*$^-$ was HL-treated for 4 h to induce the expression of qE, in addition to constitutively expressed qT. We verified that the average reconstruction error with the dictionary atoms was close to the selected threshold ($2.5 \cdot 10^{-4}$ for 74 cells), validating the hypothesis that the WT response can be accounted for by a combination of the training data. Then, we verified that experimental noise contributed minimally to the dispersion of the points cloud. Fig. 3G,H indicate that successive repeats of the HL protocol on the same cell showed similar mean scores for qE and qT. Moreover, the $D_{i,i}$ distribution was narrower than $D_{i,j}$ as previously observed (Fig. 3I), confirming once again that the dispersion is dominated by intercellular biological variations.

The mean qE score was lower than that of the untreated *stt7-1* population, consistent with the lower expression of qE in the wild-type compared to *stt7-1* mutant as reported in the literature (53,60,66). However, the population exhibited similar coefficients of variation of qE and qT to those observed in training populations expressing only one trait (43% for qE, 21% for qT in Fig. 3G,3H). Even though it suggests qE and qT dispersions are largely independent, we looked into their correlation at the single-cell level to investigate possible interactions. This was done both by exploiting the natural variability of qE and qT within a given *wt4a*$^-$ population, and by comparing different *wt4a*$^-$ populations exposed to different durations of the HL-treatment and expressing different averaged levels of qE.

We analyzed how intercellular variations evolve over the qE activation process by applying three sequential HL-treatments of 80 min each, with measurements of F_m' traces after each step, leading to a total of four point clouds in the 3D space. The time evolution of the point clouds in the (q̃T, q̃E) plane is presented in Fig. 4A. Fig. 4B demonstrates that our machine learning framework successively provides quantitative estimates of qE and qT from the complex F_m' traces depicted in Fig. 1A, and we show in Fig. S12 that the traces are properly reconstructed by the dictionary method. As the q̃E score increases, the characteristic drop at the onset of light increases. Similarly, as the q̃T score decreases, the amplitude of the gradual increase in fluorescence at the onset of light also decreases, reflecting the transition from state II to state I. The progressive increase of q̃E scores with HL exposure time (Fig. 4A,C) is in line with reported results showing that few hours of exposure to HL are needed to reach a steady qE level (37). Surprisingly, we also observed a progressive decrease in q̃T scores (Fig. 4A,C). The alignment of the respective affine fits in Fig. 4A demonstrated that the two phenomena, qE and qT, are strongly correlated, both at the centroids

of the scatter plot and among individual cells of the three populations treated with HL for 80, 160 or 240 min (Fig. 4A), which we confirmed by statistical tests (Table S3). As suggested in the literature (30,42,49), this observation supports the synergistic action of qE and qT in response to HL. We argue below for mechanical coupling between the two processes, rather than a coupling at the level of the gene expression or protein biogenesis.

We performed the same experiments with another wild-type strain, *cc124*. Similarly to the wild type strain *wt4a⁻*, the time course of the scatter plot with the duration of the HL treatment also shows a gradual increase in the q̃E score and a continual decrease in the q̃T score (Fig. 4D,F). The time course of the distribution of cell coordinates for each strain, and the means of the distributions show similar trends in both strains. However, unlike *wt4a⁻*, the correlation observed at the population level (point cloud centroids) is not reflected among individual cells of the same population (Fig. 4D). Specifically, the affine fitting curve of the point cloud centroids is not aligned with the respective affine fit within each population (Fig. 4D). When the traces are displayed in a grid (Fig. 4E), we observe phenotypes where both qE and qT are significant, a situation not found in *wt4a⁻*, yet we note that it was properly reconstructed by the dictionary method (Fig. S12). The different behaviour in *wt4a⁻* and *cc124* indicates that the correlation between traits observed at the population level, when all cells are averaged (Fig. 4C,F), may mask different interactions or synergies between qE and qT at the individual cell level (Fig. 4A,D). This finding highlights the interest of single-cell analysis when exploring the orchestration of the various physiological responses of cells to a given stress.

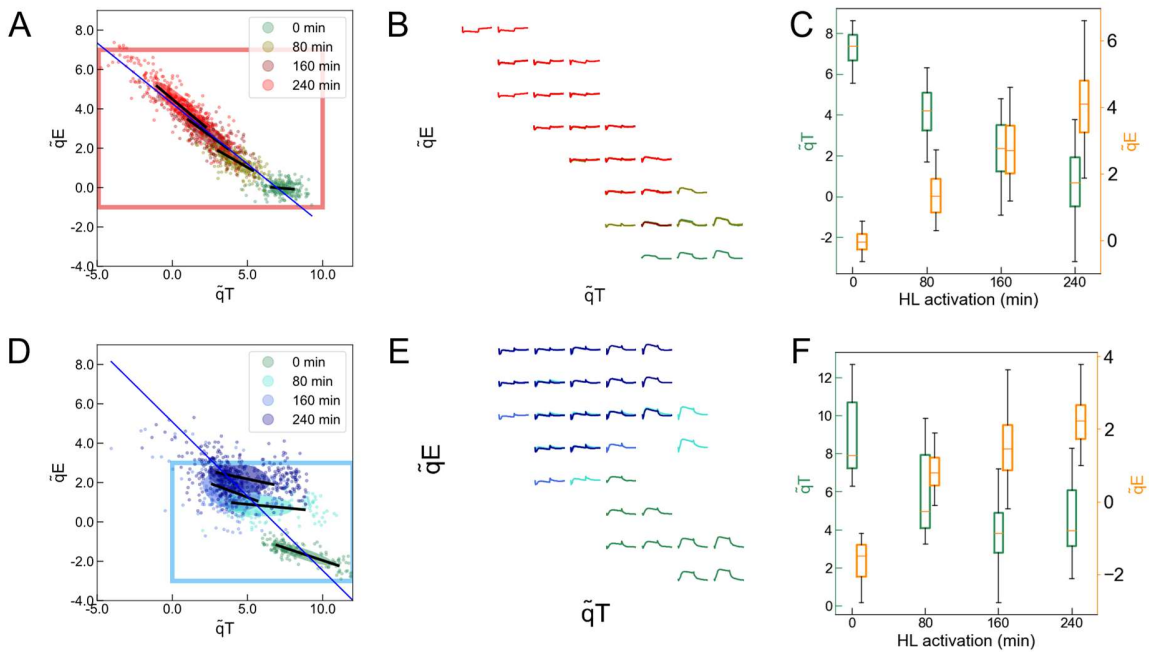

Figure 4: *Evolution of the distribution of wild type populations throughout HL-treatment*. **A**: Wild type *wt4a⁻* population taken from the growth light not HL-treated and displaying minimal qI component (untreated *wt4a⁻* equivalent to ***Pop-qT***, 4th repeat, 196 cells) exposed to three consecutive HL-treatment protocols of 80 min to promote

the expression of LHCSR3 proteins and induce qE. The blue line indicates the direction of the affine fitting curve of the point cloud centroids while the black line indicates the affine curve fit of each point cloud separately; **B**: Grid projection of the F_m' traces from **A**; **C**: Amplitude of the trait along the q̃T axis (left, green) and q̃E axis (right, orange) throughout the HL-treatment of wild type *wt4a⁻*. **D**: Wild type *cc124* population taken from the growth light not HL-treated and displaying minimal qI component (4th repeat, 164 cells) exposed to three consecutive HL-treatment protocols of 80 min to promote the expression of LHCSR3 proteins and induce qE. The blue line indicates the direction of the affine fitting curve of the point cloud centroids while the black line indicates the affine curve fit of each point cloud separately; **E**: Grid projection of the F'_m traces from **D**; **F**: Amplitude of the trait along the q̃T axis (left, green) and q̃E axis (right, orange) throughout the HL-treatment of wild type *cc124*. In **B** and **D**, the axis limits have been adjusted to the rectangles (blue, A – red, B) and a trace is represented in the grid only if two or more datapoints fall within the box.

Discussion

In this report, we developed a fluorescence microscope capable of capturing direct fluorescence of single cells of the green alga *Chlamydomonas reinhardtii*. We also established an acquisition protocol to record maximal fluorescence (F_m') traces during alternating periods of darkness and bright light exposure, and introduced a novel machine learning framework to analyse these traces. Using a well-defined training set, we created a three-dimensional NPQ space, with each axis representing one of three NPQ components (qE, qT and qI). By projecting any F_m' trace obtained under the same reference light protocol on this 3D space, we successfully separated and quantified the contributions of co-occurring NPQ components. We thus effectively deconvoluted the three NPQ contributions in an unbiased manner. This work raises several questions regarding the fundamental biological responses to light stress in *Chlamydomonas reinhardtii*, the validity of our approach compared to established protocols, and its potential applications in the broader field of phenomics.

Quality check of the single cell F_m' traces

In this work, we placed great emphasis on validating the results of our analysis framework to ensure that biases and noise in single-cell F_m' traces were sufficiently low, thus not affecting our interpretation of intercellular variability or trait correlations. We consider that this has been fully validated via three tasks which we see as key quality checks:

- *Comparison with the reference instrument*. Before analyzing the F_m' traces at the single-cell level, we confirmed that the average F_m' trace of our data matched the F_m' trace obtained from the entire population using a reference instrument. This close agreement (Fig. 1D,E) indicated that there was no systematic bias in the acquisition protocol of our custom instrument. Note that SP were applied at a higher rate than usual (37,38) to allow the monitoring of fast NPQ kinetics. We evaluated their negligible actinic contribution compared to the high-light exposure (see Fig. S9).
- *Assessment of light heterogeneity.* To rule out the possibility of bias due to light heterogeneity, which could disappear after averaging over cells, we examined the

correlation between NPQ scores and the light intensity profile over the field of view of the microscope measured by actinometry (67). The lack of statistically significant correlation between the NPQ score and light intensity among cells confirmed the absence of light heterogeneity bias (Fig. S14).

- *Assessment of the noise impact.* We assessed the impact of noise, which could potentially affect our conclusions on the heterogeneity of intercellular NPQ components. By conducting two successive experiments with each cell assumed to have consistent high light response in the two repeats, we analyzed the noise distributions and found that the variation in light response among a population is explained by biological variations and not experimental noise (Fig. 3B,E,I). It is here important to note that achieving identical behavior across all cells in repeated experiments is an unattainable goal. Therefore, our protocol likely overestimates the noise level and underestimates the signal-to-noise ratio.

Comparison to established protocols and quantification of NPQ components

After ensuring the quality of the single cell F_m' traces, it was crucial to implement an appropriate framework for data processing and a precise evaluation of the three NPQ contributions. To achieve this, we used an innovative approach. Rather than relying on traditional *ad hoc* metrics which deconvolute the NPQ components (qE, qT and qI) according to their relaxation times or their local extrema in the F_m' traces, we have opted for a machine learning approach. This choice was motivated by the desire to take full advantage of the in-depth knowledge on high light responses and rich kinetic information and intercellular variability present in single-cell ChlF traces. The use of *Chlamydomonas* as a model organism has allowed us to create clear scenarios where one of the three components has mainly influenced the F_m' traces, which is essential for a supervised machine learning process. The approach is based on a key hypothesis –that can be *a posteriori* tested: that a random F_m' trace, containing the three NPQ components, can be represented as a linear combination of elementary traces found in the training dataset. This implies that for every F_m' trace, the kinetics of each NPQ contribution (for example qE) will be reflected somewhere in the collection of single-cell F_m' traces contained in the population representing this NPQ contribution (for example ***Pop-qE*** for qE) in the training dataset.

We confirmed that the NPQ scores generated by our machine learning framework were both qualitatively and quantitatively consistent with the *ad hoc* measurements commonly used in the literature to assess the amplitude of the NPQ components on the training dataset (Fig. 2,S13,S16). More specifically, we observed a strong, but not linear, monotonous correlation between the qE score and the rapid decrease of NPQ at the light onset. In addition, our qT score has shown an excellent agreement with the slow evolution of the F_m' traces at light onset, right after the rapid decrease due to qE, which is traditionally used as an *ad hoc* estimation of qT (Fig. S13). We also validated that the progression of the qE score during HL-treatment (Fig. 4) reflected the well-documented progressive genetic expression of LHCSR3 proteins (37,52). Although we have not studied it in detail, the qualitative behavior associated with the qI component is also consistent with the literature (59), with a continuous decrease of fluorescence in the light and no significant relaxation over the 15 minutes in darkness (Fig. S4,S16). The selection of the reference dataset is critical for the first step of our machine learning framework and it has a significant impact on the

final scoring of the NPQ components. Instead of using mutants, we decided to exploit the property that LHCSR proteins are not expressed in a constitutive manner to create conditions without qE. This choice proved correct since the typical signatures of qE (rapid decrease and increase of F_m' at light onset and offset, respectively) were absent in the ***Pop-qI*** and ***Pop-qT*** populations, which were used to define the qI and qT axis respectively and were not expected to display qE (Fig. 1). Conversely, the qE signature was clearly observed in the *npq4* strain (Fig. S3), a characteristic previously associated with the presence of the second LHCSR protein, LHCSR1 (52,61). We adopted the untreated *stt7-1* population, ***Pop-0***, without HL treatment as the population intended to display none of the NPQ components. However, qE slightly increased during the repeats of 15-min illumination needed to reduce the contribution of qI. This population is thus the weak point of our reference dataset.

Eventually, one may methodologically interrogate our strong working hypothesis: that the kinetic responses of a variety of strains can be scored in a framework of reference built from the kinetic responses of a few strains. Indeed, without an appropriate choice of the reference set, the kinetics of an NPQ component could be influenced by another component. Here, the introduction of an intermediate reconstruction step with Dictionary Learning assesses the quality of the reconstruction of the F_m' traces by an average mean-squared error test. We could harness the results to validate or reject the hypothesis. Hence, we showed that the F_m' trace of strains exhibiting several NPQ components is effectively captured by a framework built exclusively on proper populations expressing a single NPQ component (Fig. S12). As shown in Fig. S12, we have confirmed that this kind of hypothesis is applicable to a wide range of strains that were not included in the training set and do not share the same genetic background.

Extent and origin of the intercellular variability in qE and qT

One of the aims of our single-cell study was to assess the intercellular variability of various responses to high light stress, which we found to be greater than the variability introduced by instrumental noise or data analysis as mentioned above (Fig. 3). Given the reported causal relationships between cell size and photosynthetic activity (68), we investigated whether the NPQ score was influenced by cell size. As shown in Fig. S14D,E,F, no significant correlation was observed between cell size and NPQ score. Another potential source of variability is related to the cell cycle (69). Genome-wide investigations of diurnal gene expression in synchronized *Chlamydomonas reinhardtii* cells revealed that most of the transcriptome is expressed with strong periodicity, forming distinct clusters (70,71). Specifically, genes associated with complex structures and processes involved in photosynthesis exhibit unique co-expression profiles, which are highly predictive of function and temporal order (70–72). To minimize this source of variability, we used monoclonal cultures of cells synchronized on a 12h/12h light/dark cycle (see Methods and Fig. S5,S7). Therefore, we did not identify any significant sources of heterogeneity in cell states. We also discarded the hypothesis that this heterogeneity could come from the position in the pad (Fig. S14G,H,I).

We conclude that the observed variability in qE (coefficient of variation ~0.4) and qT (coefficient of variation ~0.2) likely reflects stochastic gene expression (73,74), although we cannot entirely exclude the possibility that some of the heterogeneity was introduced during cell preparation prior to measurement.

Molecular origins of the strong correlation between qE and qT

The substantial intercellular variability of qE and qT allowed to identify a strong correlation between qE and qT in synchronized monoclonal cultures of *wt4a*$^-$ (Fig. 4A). This result is particularly significant because it emerges within the same genotype and cellular context, among cells with common histories. The correlation can be summarized as follows: the higher the qE under HL, the less pronounced the transition from state II to state I. Remarkably, this correlation extends from "pure qT" (the highest qT score without qE, q̃E = 0, q̃T = 7) to "pure qE" (the highest qE score without qT, q̃E = 5, q̃T = 0) , across four populations subjected to 0, 80, 160 or 240 min of HL treatment. To contextualize these scores, a q̃E score of 5 corresponds to a rapid initial decrease in F_m' of ~40 % which translates to a NPQ value of approximately 2/3, while a q̃T score of 7 corresponds to a change of approximately 40% in F_m' due to state transitions. Although significant, this change is smaller than the up to 80% of LHCII reported in *Chlamydomonas reinhardtii* when ensuring that cells undergo a transition from pure state I to pure state II (75). Notably, positive q̃T scores indicate state transitions from state II (in the dark) to state I (at the end of the HL period).

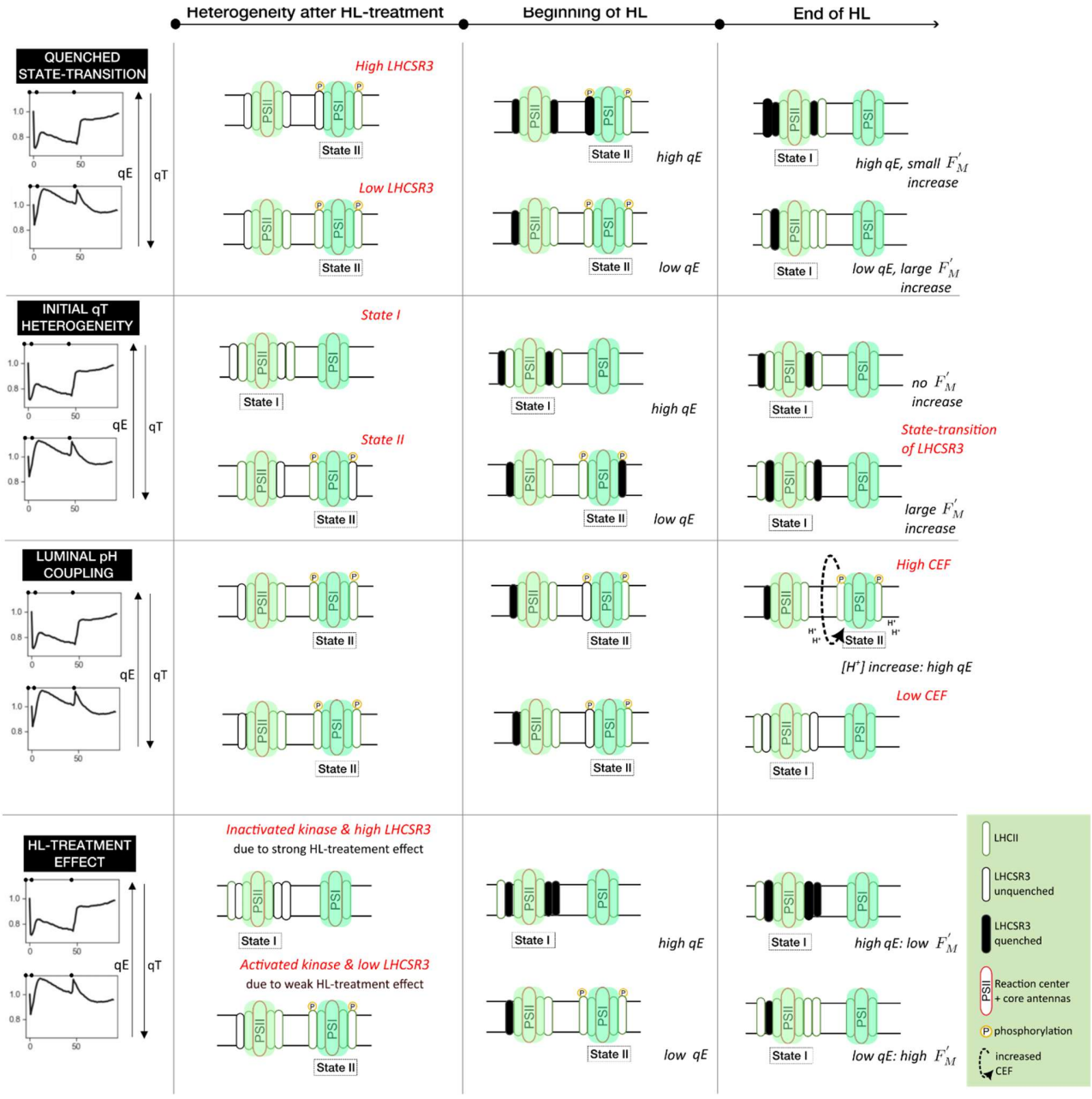

Figure 5: *Four tentative hypotheses accounting for the observed correlation between qE and qT.* The "heterogeneity after HL treatment" column represents the initial state of the dark-adapted sample (first F_m' at the beginning of the light protocol). The "beginning of HL" column represents the situation after less

than 1 min of HL (first F_m' after the onset of HL). The "End of HL" column represents the situation at the end of the HL period (last F_m' before offset of light).

We grouped the different potential links between qE and qT in four groups, which allow to identify tentative hypotheses, that are not mutually exclusive and can together explain the strong correlation between qE and qT in the HL-treated *wt4a*$^-$ (Fig. 5):

- **"Quenched State Transitions"** scenario**.** This scenario suggests an initial heterogeneity in qE capacity (LHCSR3 amount) between cells. During transitions from state II to state I, the visible increase in fluorescence is not only proportional to the amount of LHCII moving from PSI to PSII. It also depends on the level of quenching qE occurring in PSII at the time of the switch. Higher quenching in PSII results in a smaller increase in PSII fluorescence for each additional LHCII that attaches. The "Quenched State Transitions" concept predicts a correlation between qE and qT, consistent with our observations, but can only explain a minor part of the qE/qT correlation. Indeed, in this scenario, the amplitude of qT would be expected to attenuate by approximately 40%, rather than disappear completely as seen in Fig 4, for the highest qE (q̃E = 5). This consideration highlights the need for alternative explanations.

- **"Initial qT Heterogeneity"** scenario. This scenario suggests an initial heterogeneity in LHCII distribution between cells, with some cells starting in state I and others in state II. It builds on previous reports suggesting that LHCSR proteins (LHCSR3 and potentially LHCSR1) are not exclusively associated with PSII but can migrate between the two photosystems, similar to LHCII, likely facilitated by STT7-mediated phosphorylation (37,60). In support of this hypothesis, we detect phosphorylated LHCSR3 in the *wt4a*$^-$ before the HL-treatment (Fig 1B). In this context, cells in state I would have LHCSR3 bound to PSII and would exhibit high qE but no additional qT capacity. Conversely, cells in state II would transition to state I (q̃T positive) but would not exhibit a rapid decrease of fluorescence (qE) since LHCSR proteins are initially bound to PSI.

- **"Mechanistic coupling through luminal pH"** scenario. One might look for a mechanistic link between the two phenomena, such as how qT might influence luminal pH (trigger of qE) or how qE might affect the redox state of the PQ pool (trigger of qT). If the redox state of the PQ pool were the connecting factor, a higher qE at the onset of light should result in a more oxidized PQ pool (as qE reduces the reducing pressure on PSII), thus favoring state I over state II. However, our results show the opposite behavior. Let's consider the second possibility of a coupling through luminal pH, with all cells starting from state II. One could hypothesize that cells remaining in state II, characterized by higher PSI light absorption

than PSII, would favor Cyclic Electron Flow (CEF) around PSI over LEF (76). This would lead to increased proton pumping, lower luminal pH and higher qE (77). In such a scenario, cells exhibiting no qT would display a high qE, whereas cells undergoing a transition to state I would switch to a LEF mode and display smaller qE.

- **"LHCSR3/STT7 accumulation"** scenario. It does not imply a direct mechanistic coupling at the level of photosynthetic function, but rather reflects a correlation between the relative accumulation of active STT7 and LHCSR3. It is well established that LHCSR3 levels increase with exposure to high light (37,52). Furthermore, STT7 is inactivated in high light (64), and in land plants the amount of the kinase STN7 itself decreases under high light (63). Therefore, the strong correlation between qE and qT may arise from a common external trigger –HL treatment- linking LHCSR3 induction and STT7 degradation or inactivation. In this setting, some cells would respond more rapidly to HL, expressing higher levels of LHCSR3 and inactivating more STT7 (therefore starting closer to state I). These cells would display high qE and low qT. Conversely, slower-responding cells would exhibit lower qE but higher qT.

We also monitored the evolution of qE and qT during HL-treatment in another wild-type strain, *cc124*. In both wild-type strains, population mean qE and qT values showed a correlation during HL-treatment, suggesting similar mechanisms at the bulk level. However, the consistent correlation observed both between and within populations in *wt4a*$^-$ was not present in *cc124*, indicating that different mechanisms of interaction may be at play between these two traits. This finding in *cc124* conveys two important insights. First, it highlights the value of single cell studies, as two populations may orchestrate their responses to high light differently, while still displaying similar behaviour at the macro (population) level. Second, it highlights the potential influence of the Yule-Simpson paradox in photosynthesis research, where a correlation observed at the population level may not accurately reflect a causal relationship at the level of individuals within that population (78). We have proposed several hypotheses that are not mutually exclusive, so it is possible that the observed differences between these two wild types arise from varying contributions of the phenomena described in these hypotheses.

Perspectives for single-cell photosynthesis measurements

Our machine learning framework is solely based on signal dynamics, without requiring any prior knowledge of temporal phases. Therefore, since it is independent of the specific biological question, it should be adaptable to other single-cell phenotyping scenarios involving kinetic responses of ChlF. Hence, it could be applied to other species or stresses, such as nutrient deficiency or extreme temperatures, provided that these stresses trigger detectable fluorescence changes. However, some limitations may also intervene.

The first concern is the use of direct fluorescence detection. This method is very effective for studying non-photochemical quenching under high light conditions (i.e. photoprotection

performance), due to its excellent sensitivity and constant light intensity. However, for examining other stress responses under low or moderate light, particularly phenotyping based on PSII yield (i.e. photosynthesis performance), alternative method may be more appropriate. This is because direct fluorescence inherently suffers from reduced signal-to-noise as light intensity decreases. Additionally, parameters such as the maximum PSII quantum yield (Fv/Fm), which require fluorescence measurements at very low and very high light intensities, can provide precious additional information but are limited by the dynamic range of the camera (79). A viable alternative, feasible with current setup, is the use of weak "measuring light" (e.g. PAM or similar techniques), combined with a more sensitive camera, to determine fluorescence yields. This approach overcomes the limitation of dynamic range and ensures that the signal-to-noise ratio for the fluorescence yield remains unchanged by the applied actinic light.

The second limitation is that our machine learning framework requires an initial training phase using a representative population for each phenotype to be assessed. This step restricts our approach to a limited number of species for which the necessary resources and expertise are available to build such an informed training dataset. In fact, we experienced this drawback when including the contribution of qI in our analysis, for which the underlying biological processes are not as well controlled and informed as for qE and qT. We see however in (41) that unsupervised approach can have difficulties to extract generic signatures across wild-types and mutants.

Third, our methodology assumes that the wild-type response can be estimated by a combination of elementary responses. Although verified here, this hypothesis might not systematically be valid in the study of other processes influencing ChlF traces. To anticipate this limitation, we implemented a strategy to evaluate this hypothesis, based on a validity criterion associated to the reconstruction capacity of the Dictionary Learning algorithm exposed to new and more complex data.

A promising direction for future work is to integrate this non-invasive approach relying on ChlF measurements with flow cytometry/microfluidics and/or omics methods, which are rapidly advancing and improving our ability to assess cellular states. This integration could allow correlations between cell morphology and various photosynthetic traits, facilitating applications such as varietal selection, directed evolution, screening of single cells after mutagenesis or isolation of species of interest from natural samples. Combined with single-cell omics, as a final step since this analysis is destructive, this approach would allow to identify in more detail the genetic determinants underlying the heterogeneity observed in non-invasive methods, thus advancing our understanding of cellular responses.

Materiel and Methods

Experimental Design: optical calibration

Microscope set-up: We developed a fully automated Python-based microscope capable of applying versatile illumination protocols to photosynthetic organisms (leaves or microalgae – see Fig. S1 and Table S1). An Arduino board controls the 470±10 nm and 405±7 nm LEDs at the millisecond precision. The blue LED is used for the high actinic light, the HL-treatment and the

relaxation, and the purple LED is used for the saturating pulses. We chose these wavelengths because it is possible to calibrate precisely their intensity with an actinometer-based protocol (67). A Python algorithm (72,73) segments with a watershed algorithm each cell from the first frame of the movie (Fig. 1A) to extract the mean fluorescence from the pixels corresponding to each segmented alga over the time frames and create a database of F_m' traces for statistical analysis. A thorough description of the set-up and segmentation method is provided in the Supplementary Information (Fig. S1,S6, Table S1).

Calibration of the light intensity: The blue and purple LEDs were calibrated using the point detector (MPPC). The relaxation time associated to the conversion between ON and OFF state of the Reversibly Photoswitchable Fluorescent Protein Dronpa-2 was monitored, according to the protocol described in (67). Dronpa-2 was extracted and prepared as described in (67): we used 10 μM aqueous solution in Tris buffer pH 7.4 (50 mM Tris, 150 mM NaCl). We repeated the protocol for different LED current levels. We used a 525/30 filter to collect Dronpa-2 fluorescence.

Calibration of the light intensity in the field of view: To evaluate the distribution of light intensity in the field of view, we performed the light calibration of the 470 and 405 LEDs with Dronpa-2 and used a camera instead of a single-point detector to measure the fluorescence's response dynamics (67). The intensity maps and histograms demonstrate the homogeneity of the illumination, respectively above 85% and 70% for the 470 and 405 LEDs (see Fig. S2).

Macro set-up: Fluorescence emissions were measured using a Speedzen fluorescence imaging setup (JBeamBio, France) at room temperature with actinic light at 620 nm (740 µmol(photons).m^{-2}.s^{-1}), saturating pulses at 620 nm (5800 µmol(photons).m^{-2}.s^{-1}, 250 ms) and pulses for the "measuring" light (470 nm, 100 µmol(photons).m^{-2}.s^{-1}, 280 µs).

Experimental Design: Biological samples

Strains: The *Chlamydomonas reinhardtii* strains wild type 4A *mt-* (CC-4603), CC-124 wild type *mt-* and *npq4 mt-* (CC-4615) were obtained from the Chlamydomonas Resource Center and *stt7-1 (#a6)* was from the ChlamyStation (80) collection. *stt7-1* is in the same genetic background than *wt4a*$^-$. *stt7-1 (#a6)* is a clone allelic to *stt7-1* (81) and non-leaky compared to *stt7-9* (53,82).

The strains were synchronized in a 12/12 light cycle (50 µmol(photons).m^{-2}.s^{-1} – 23 ^{o}C) and grown in mixotrophic conditions (Tris-Acetate-Phosphate medium (83)). The synchronized strains grown on Petri dishes were inoculated in liquid medium on Fridays and left to grow under agitation over the weekend. The culture was diluted with fresh Tris-Acetate-Phosphate medium on Mondays and every following days. The experiments were always performed in the exponential phase, between Tuesdays and Fridays. For each experiment, the whole illumination protocol started 2 h after the beginning of the light phase.

Sample preparation for microscope experiments: In all experiments, samples are collected from cultures in exponential phase (from 5×10^5 to 5×10^6 cells/mL), centrifuged (960 g for 8 min)

and transferred to Tris-minimal (MIN) medium at approximately 10^7 cells/mL. The algae are deposited on agarose pads for observations based on protocol from (84). A 1% suspension of agarose in MIN medium was heated to dissolve the agarose. The agarose pads were prepared by dropping 175 µL of melted agarose onto a glass slide holding a spacer *(AB-0578 ThermoFisher)*. To make its surface flat, the pad was covered with a thick microscope glass slide and left for 5 min at 4 ^{o}C for hardening. The cover slide was removed and the preparation was left for 10 min to allow thermal adaptation and evaporation of excess humidity. About 5 µL of algae suspension was dropped onto the agarose and left uncovered for 10 min at room temperature, to remove excess water. Finally, a coverslip was gently placed over the agarose pad for observation. A side opening in the spacer ensured gas exchanges during the experiments (see Fig. S5A).

Sample preparation for macro set-up experiments: Cells are harvested in the same conditions as for microscope experiments and go through the same light protocols. They are centrifuged and resuspended in MIN (same protocol as above) at a final concentration of 2×10^7 cells/mL. After at least 30 min of relaxation under the growth conditions they are either used directly or high light pre-treated as explained below. For observations under the SpeedZen, the cell suspensions are placed in a metal plate inside open cavities of 85 µL.

Monoclonal cultures: The synchronized monoculture of *stt7-1* and *wt4a*$^-$ were obtained by the streak plate protocol (85). The populations were subcloned once. The mother *wt4a*$^-$ population had not been sub-cloned for more than a year, while the mother *stt7-1* population had been obtained by crossing less than one year before.

Analysis of polypeptides: Cells were grown, harvested and conditioned according to the conditions described in Methods for macro set-up experiments. The population of *wt4a*$^-$ and *stt7-1* were harvested (2.10^7 cells) before (na) and after (a) HL-treatment of four hours. Total cellular proteins were resuspended in 200 mM dithiothreitol (DTT) and 200 mM Na_2CO_3 and solubilized in the presence of 2% SDS at 100 °C for 60 s. Cell extracts were loaded on equal chlorophyll basis, separated by SDS-PAGE with 8 M urea (86), and electro-transferred onto nitrocellulose membranes (Amersham Protran 0.1 µm NC) by semi-dry method. Immunodetection was performed using antibodies against LHCSR (60), STT7 (53) and AtpB (beta-ATP synthase subunit) (87). Signals were visualized by enhanced chemiluminescence (ECL - Clarity Biorad) and detected with a luminescent image analyzer ChemidocTM (Biorad).

Experimental Design: Light protocols

Reference protocol: The algae are exposed to 15 min of 400 µmol(photons).m^{-2}.s^{-1}, 470±10 nm, and 15 min of dark. Throughout the whole experiment, saturating pulses (SP) are shone every 20 s (1400 µmol(photons).m^{-2}.s^{-1}, 405±7 nm, 200 ms). The whole fluorescence is collected with a camera,

but only the response specific to saturating pulses, F_m', is analysed. The actinic light is turned off when the SP is applied. The light protocols are illustrated in details in Fig. S8, S10 and Table S2. The four repeats of the reference protocol and the HL-treatment are performed under the microscope without displacement of the sample. The last two repeats are considered qI-free because the F_m level at the beginning of the exposure is very close to the F_m' level at the end of the exposure (see Fig. S4, S10, S16).

HL-treatment: The algae deposited on a pad are exposed to 400 µmol(photons).m^{-2}.s^{-1}, 470±10 nm for 1h20–4 h under the microscope. The HL-treatment is followed by 45 min of 40 µmol(photons).m^{-2}.s^{-1}, 470±10 nm. For macro set-up experiments, the algae suspensions are subjected to 550 µmol(photons).m^{-2}.s^{-1} of white light for 4 h under strong agitation. The HL-treatment is followed by 45 min of relaxation under low light (15 µmol(photons).m^{-2}.s^{-1}, white LED panel).

Monoclonal experiments: The cells were dark-adapted for 15 min and exposed to the reference protocol four consecutive times. Then followed 4 h of HL-treatment, 45 min of low light, 15 min dark and another set of four consecutive exposures to the reference protocol. The last two repeats were used for building Fig. 3A-C. For the synchronized monoclonal *wt4a*$^-$ population, the cells were dark-adapted for 15 min and exposed to the reference protocol four consecutive times. The last two repeats were used in Fig. 3D-F. Then followed 4 h of HL-treatment, 45 min of low light to partly relax the photoinhibition, 15 min of dark and another set of four consecutive exposures. The last two repeats were used for Fig. 3G-I.

Data analysis

Prior to the machine learning protocol, the outliers were removed from the training dataset. The algae with a surface smaller than 5 pixels were discarded. Then, a KD-Tree (88) was built to represent the Euclidean distance relationships between the inputs. We manually set a threshold D=0.01 to remove the points with all neighbours further than D. In total, 66 fluorescence traces out of 2302 were removed from the training set.

Dictionary Learning (Fig. 2A, Fig. S11,S12) (44) is used to decompose any complex series of data into pre-learned arbitrary waveforms that capture its most important features. The training phase is unsupervised: it exploits all the F_m' traces from the dataset without any label. It learns to identify characteristic patterns in the ChlF traces that are stored in a dictionary of basic waveforms (called atoms). Its metric is the accuracy of reconstruction of a ChlF trace as a linear combination of the atoms. We used the sparse coding method to decompose the data into a linear combination of atoms from a dictionary (44) implemented in *scikit-learn* (89). The coefficients weighting the atoms form a feature vector. The dimension of this feature vector is lower than the dimension of the original F_m' trace (91) and is equal to N_D, the number of atoms in the dictionary. To build the dictionary, the randomly-sampled equilibrated dataset containing $n_{samples}$ F_m' traces

from ***Pop-qE, Pop-qT, Pop-qI*** and ***Pop-0*** was used as a training set and fitted with least angle regression following (44). The metric optimized is the reconstruction fidelity with a penalty on the sparsity level, by minimizing the following function:

$$\min_{D,X} \frac{1}{2} \|Y - DX\|_2^2 + \lambda \|X\|_1 \quad (1)$$

Where:

- **Y** is the input data matrix of dimension $91 \times n_{samples}$,
- **D** is the dictionary matrix with N_D atoms of dimension 91 (normal matrix),
- **X** is the coefficient matrix of dimension N_D,
- $\|Y - DX\|_2^2$ is the squared Euclidean norm of the reconstruction error,
- $||\mathbf{X}||_1$ is the L1 norm (sum of absolute values) of the coefficients,
- λ is the regularization parameter that controls the sparsity level.

We explored several parameters detailed in the Supplementary Information, including the normalization of the F_m' traces prior to the training. For the analysis described in the Main Text, we selected $N_D = 10$, $n_{samples}$=300, $\lambda = 10^{-6}$, and normalized by the sum of the F_m' to obtain a probability distribution.

Linear Discriminant Analysis The dimension reduction algorithm, multiclass Linear Discriminant Analysis (45) (LDA) finds the best linear combinations of input features that maximizes the separability among differently-labelled data. The metric optimized is derived from two matrices, S_B and S_W which represent respectively the separation between (B) the means of the classes and the dispersion of the datapoints within (W) each class. We obtain a projection from a space of dimension 10 to a space of dimension $c - 1$ where c is the number of classes (in our case $c = 4$, ***Pop-qE, Pop-qT, Pop-qI*** and ***Pop-0***). The LDA is solved by Singular Value Decomposition (scikit-learn Python library) (89). Once the projection $\mathbf{T} \in 10 \times 3$ is obtained, we operate a transformation to align the significant axis with the canonical space (x, y, z) for easier analysis. The first step is to identify the principal directions of each class in the dataset and associate it to an axis by collecting the corresponding eigenvector, using Principal Component Analysis (90). They form a transfer matrix **R** which allows to change the base of the space. Then, the origin is identified by using biological knowledge. The average of the point clouds representing ***Pop-qE***, ***Pop-qI*** and ***Pop-0*** is expected to show no qT as the strain *stt7-1* is an STT7 knockout mutant: it is the origin of the q̃T axis. The average of the point cloud ***Pop-qI*** is expected to show no qE because it has only been exposed to growth light. It is set as the origin of the axis q̃E. The average of the point clouds ***Pop-qE, Pop-qT*** and ***Pop-0*** are supposed to show no qI since the last F_m' equals the first F_m and are therefore used as the zero reference for the q̃I axis.

Evaluation of the cell-to-cell *vs* cell-to-population distances: (Fig. 3B,E,I) As the experiments are performed on the same sample with N algae, we can derive the couple positions in the dictionary learning NPQ space $((x_{3,i},y_{3,i},z_{3,i}),(x_{4,i},y_{4,i},z_{4,i}))$, $i \in [1,N]$ corresponding to algae i for the third and fourth repeats. We introduce the distance $D_{i,j}$ as the distance of the point corresponding to algae i in the third repeat to the point corresponding to algae j in the fourth repeat:

$$D_{i,j} = \sqrt{(x_{4,j} - x_{3,i})^2 + (y_{4,j} - y_{3,i})^2 + (z_{4,j} - z_{3,i})^2} \text{ for } (i,j) \in [1,N] \quad (2)$$

Grid representation of the F_M traces in the NPQ space: To build Fig. 4B,E we grouped the data points in a 8×8 grid and plotted the average F_m' trace if there were more than three data points in a grid element. The traces from different point clouds are overlaid with the corresponding colours indicated in Fig. 4A,D.

SI references: (37,38,40,49,52,58,59,61,67,84,91–97)

References

1. Cole B, Bergmann D, Blaby-Haas CE, Blaby IK, Bouchard KE, Brady SM, et al. Plant single-cell solutions for energy and the environment. Communications biology. 2021;4(1):1-12.
2. Brehm-Stecher BF, Johnson EA. Single-cell microbiology: Tools, technologies, and applications. Microbiology and molecular biology reviews. 2004;68(3):538-59.
3. Lidstrom ME, Konopka MC. The role of physiological heterogeneity in microbial population behavior. Nat Chem Biol. oct 2010;6(10):705-12.
4. Damodaran SP, Eberhard S, Boitard L, Rodriguez JG, Wang Y, Bremond N, et al. A millifluidic study of cell-to-cell heterogeneity in growth-rate and cell-division capability in populations of isogenic cells of Chlamydomonas reinhardtii. PloS one. 2015;10(3):e0118987.
5. Miao Z, Humphreys BD, McMahon AP, Kim J. Multi-omics integration in the age of million single-cell data. Nature Reviews Nephrology. nov 2021;17(11):710-24.
6. Krause GH, Weis E. Chlorophyll fluorescence as a tool in plant physiology: II. Interpretation of fluorescence signals. Photosynthesis Research. 1984;5(2):139-57.
7. Papageorgiou GC, Govindjee, Govindjee, éditeurs. Chlorophyll a Fluorescence: A Signature of Photosynthesis. Dordrecht: Springer Netherlands; 2004. (Advances in Photosynthesis and Respiration; vol. 19).

8. Redding KE, Santabarbara S. Chapter 16 - Photosystems I and II. In: Grossman AR, Wollman FA, éditeurs. The Chlamydomonas Sourcebook (Third Edition) [Internet]. London: Academic Press; 2023 [cité 6 août 2024]. p. 525-60. Disponible sur: https://www.sciencedirect.com/science/article/pii/B9780128214305000225
9. Lavergne J, Trissl HW. Theory of fluorescence induction in photosystem II: derivation of analytical expressions in a model including exciton-radical-pair equilibrium and restricted energy transfer between photosynthetic units. Biophysical Journal. 1 juin 1995;68(6):2474-92.
10. Snel JFH, Dassen HHA. Measurement of Light and pH Dependence of Single-Cell Photosynthesis by Fluorescence Microscopy. Journal of Fluorescence. sept 2000;10(3):269-269.
11. Konert G, Steinbach G, Canonico M, Kaňa R. Protein arrangement factor: A new photosynthetic parameter characterizing the organization of thylakoid membrane proteins. Physiologia plantarum. 2019;166(1):264-77.
12. Westerwalbesloh C, Brehl C, Weber S, Probst C, Widzgowski J, Grünberger A, et al. A microfluidic photobioreactor for simultaneous observation and cultivation of single microalgal cells or cell aggregates. PLOS ONE. avr 2019;14(4):e0216093.
13. Behrendt L, Salek MM, Trampe EL, Fernandez VI, Lee KS, Kühl M, et al. PhenoChip: A single-cell phenomic platform for high-throughput photophysiological analyses of microalgae. Science advances. 2020;6(36):eabb2754.
14. Andersson M, Johansson S, Bergman H, Xiao L, Behrendt L, Tenje M. A microscopy-compatible temperature regulation system for single-cell phenotype Analysis–Demonstrated by thermoresponse mapping of microalgae. Lab on a Chip. 2021;21(9):1694-705.
15. Herdean A, Sutherland DL, Ralph PJ. Phenoplate: An innovative method for assessing interacting effects of temperature and light on non-photochemical quenching in microalgae under chemical stress. New Biotechnology. 2022;66:89-96.
16. Küpper H, Šetlík I, Trtílek M, Nedbal L. A microscope for two-dimensional measurements of in vivo chlorophyll fluorescence kinetics using pulsed measuring radiation, continuous actinic radiation, and saturating flashes. Photosynthetica. 2000;38(4):553-70.
17. Šetlíková E, Šetlík I, Küpper H, Kasalickỳ V, Prášil O. The photosynthesis of individual algal cells during the cell cycle of Scenedesmus quadricauda studied by chlorophyll fluorescence kinetic microscopy. Photosynthesis research. 2005;84(1):113-20.
18. Trampe E, Kolbowski J, Schreiber U, Kühl M. Rapid assessment of different oxygenic phototrophs and single-cell photosynthesis with multicolour variable chlorophyll fluorescence imaging. Marine Biology. juill 2011;158:1667-75.
19. Mohr W, Vagner T, Kuypers MM, Ackermann M, LaRoche J. Resolution of conflicting signals at the single-cell level in the regulation of cyanobacterial photosynthesis and nitrogen fixation. PLOS one. 2013;8(6):e66060.
20. Gachon CMM, Küpper H, Küpper FC, Šetlík I. Single-cell chlorophyll fluorescence kinetic microscopy of Pylaiella littoralis (Phaeophyceae) infected by Chytridium polysiphoniae (Chytridiomycota). European Journal of Phycology. nov 2006;41(4):395-403.
21. Széles E, Nagy K, Ábrahám Á, Kovács S, Podmaniczki A, Nagy V, et al. Microfluidic Platforms Designed for Morphological and Photosynthetic Investigations of Chlamydomonas reinhardtii on a Single-Cell Level. Cells. janv 2022;11(2):285.
22. Kromdijk J, Głowacka K, Leonelli L, Gabilly ST, Iwai M, Niyogi KK, et al. Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. Science (New York, NY). 2016;354(6314):857-61.
23. Kaiser E, Morales A, Harbinson J. Fluctuating light takes crop photosynthesis on a rollercoaster ride. Plant Physiology. 2018;176(2):977-89.
24. Ghosh D, Mohapatra S, Dogra V. Improving photosynthetic efficiency by modulating non-photochemical quenching. Trends in Plant Science. 2022;
25. De Souza AP, Burgess SJ, Doran L, Hansen J, Manukyan L, Maryn N, et al. Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection. Science. 19 août 2022;377(6608):851-4.
26. Sinclair T, Specht J, Cassman K, Purcell L, Rufty T. Comment on “Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection”. Science. 24 févr 2023;379(6634):eade8506.
27. Pearcy RW. Sunflecks and photosynthesis in plant canopies. Annual review of plant biology. 1990;41(1):421-53.

28. Murchie EH, Kefauver S, Araus JL, Muller O, Rascher U, Flood PJ, et al. Measuring the dynamic photosynthome. Annals of botany. 2018;122(2):207-20.
29. Erickson E, Wakao S, Niyogi KK. Light stress and photoprotection in Chlamydomonas reinhardtii. The Plant Journal. 2015;82(3):449-65.
30. Grossmann AR, Wollman FA, éditeurs. The chlamydomonas sourcebook. Volume 2: Organellar and metabolic processes. Third edition. London, UK San Diego, CA Cambridge, MA Oxford, UK: Academic Press, An imprint of Elsevier; 2023.
31. Minagawa J, Tokutsu R. Dynamic regulation of photosynthesis in Chlamydomonas reinhardtii. The Plant Journal. 2015;82(3):413-28.
32. Ruban AV. Nonphotochemical chlorophyll fluorescence quenching: Mechanism and effectiveness in protecting plants from photodamage. Plant physiology. 2016;170(4):1903-16.
33. Goss R, Lepetit B. Biodiversity of NPQ. Journal of Plant Physiology. 1 janv 2015;172:13-32.
34. Nawrocki WJ, Liu X, Raber B, Hu C, de Vitry C, Bennett DIG, et al. Molecular origins of induction and loss of photoinhibition-related energy dissipation qI. Sci Adv. 24 déc 2021;7(52):eabj0055.
35. Wollman FA. State transitions reveal the dynamics and fexibility of the photosynthetic apparatus. The EMBO Journal. juill 2001;20(14):3623-30.
36. Goldschmidt-Clermont M. Chapter 24 - State transitions. In: Grossman AR, Wollman FA, éditeurs. The Chlamydomonas Sourcebook (Third Edition) [Internet]. London: Academic Press; 2023 [cité 6 août 2024]. p. 787-805. Disponible sur: https://www.sciencedirect.com/science/article/pii/B9780128214305000055
37. Allorent G, Tokutsu R, Roach T, Peers G, Cardol P, Girard-Bascou J, et al. A Dual Strategy to Cope with High Light in Chlamydomonas reinhardtii. The Plant Cell. févr 2013;25(2):545-57.
38. Ruiz-Sola MÁ, Petroutsos D. A Toolkit for the Characterization of the Photoprotective Capacity of Green Algae. In: Maréchal E, éditeur. Plastids. New York, NY: Springer US; 2018. p. 315-23.
39. Tietz S, Hall CC, Cruz JA, Kramer DM. NPQ(T): A chlorophyll fluorescence parameter for rapid estimation and imaging of non-photochemical quenching of excitons in photosystem-II-associated antenna complexes. Plant, Cell & Environment. 2017;40(8):1243-55.
40. Nilkens M, Kress E, Lambrev P, Miloslavina Y, Müller M, Holzwarth AR, et al. Identification of a slowly inducible zeaxanthin-dependent component of non-photochemical quenching of chlorophyll fluorescence generated under steady-state conditions in Arabidopsis. Biochimica et Biophysica Acta (BBA) - Bioenergetics. avr 2010;1797(4):466-75.
41. Ramakers LAI, Harbinson J, Wientjes E, van Amerongen H. Unravelling the different components of nonphotochemical quenching using a novel analytical pipeline. New Phytologist [Internet]. [cité 3 déc 2024];n/a(n/a). Disponible sur: https://onlinelibrary.wiley.com/doi/abs/10.1111/nph.20271
42. Steen CJ, Burlacot A, Short AH, Niyogi KK, Fleming GR. Interplay between LHCSR proteins and state transitions governs the NPQ response in Chlamydomonas during light fluctuations. Plant, Cell & Environment. 2022;45(8):2428-45.
43. Harris PD, Eliezer NB, Keren N, Lerner E. Multiparameter-based photosynthetic state transitions of single phytoplankton cells [Internet]. bioRxiv; 2024 [cité 3 déc 2024]. p. 2023.12.31.573751. Disponible sur: https://www.biorxiv.org/content/10.1101/2023.12.31.573751v2
44. Mairal J, Bach F, Ponce J, Sapiro G. Online dictionary learning for sparse coding. In: Proceedings of the 26th Annual International Conference on Machine Learning. Montreal Quebec Canada: ACM; 2009. p. 689-96.
45. Fisher RA. The Use of Multiple Measurements in Taxonomic Problems. Annals of Eugenics. 1936;7(2):179-88.
46. Endo R, Omasa K. Chlorophyll Fluorescence Imaging of Individual Algal Cells: Effects of Herbicide on Spirogyra distenta at Different Growth Stages. Environmental science and technology. 38(15):4165-8.
47. Schreiber U, Schliwa U, Bilger W. Continuous recording of photochemical and non-photochemical chlorophyll fluorescence quenching with a new type of modulation fluorometer. Photosynthesis research. 1986;10(1):51-62.
48. Duysens LN. Transfer and trapping of excitation energy in photosystem II. Ciba Found Symp. 7 févr 1978;(61):323-40.
49. Roach T, Na CS. LHCSR3 affects de-coupling and re-coupling of LHCII to PSII during state transitions in Chlamydomonas reinhardtii. Scientific Reports. mars 2017;7(1):43145.
50. Roach T. LHCSR3-Type NPQ Prevents Photoinhibition and Slowed Growth under Fluctuating Light in Chlamydomonas reinhardtii. Plants. nov 2020;9(11):1604.

51. Müller P, Li XP, Niyogi KK. Non-Photochemical Quenching. A Response to Excess Light Energy1. Plant Physiology. avr 2001;125(4):1558-66.
52. Peers G, Truong TB, Ostendorf E, Busch A, Elrad D, Grossman AR, et al. An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. Nature. nov 2009;462(7272):518-21.
53. Bergner SV, Scholz M, Trompelt K, Barth J, Gäbelein P, Steinbeck J, et al. STT7-Dependent Phosphorylation Is Modulated by Changing Environmental Conditions, and Its Absence Triggers Remodeling of Photosynthetic Protein Complexes. Plant Physiology. juin 2015;168(2):615-34.
54. Cariti F, Chazaux M, Lefebvre-Legendre L, Longoni P, Ghysels B, Johnson X, et al. Regulation of Light Harvesting in Chlamydomonas reinhardtii Two Protein Phosphatases Are Involved in State Transitions. Plant Physiology. 1 août 2020;183(4):1749-64.
55. Kress E, Jahns P. The Dynamics of Energy Dissipation and Xanthophyll Conversion in Arabidopsis Indicate an Indirect Photoprotective Role of Zeaxanthin in Slowly Inducible and Relaxing Components of Non-photochemical Quenching of Excitation Energy. Front Plant Sci [Internet]. 8 déc 2017 [cité 6 août 2024];8. Disponible sur: https://www.frontiersin.org/journals/plant-science/articles/10.3389/fpls.2017.02094/full
56. Malnoë A. Photoinhibition or photoprotection of photosynthesis? Update on the (newly termed) sustained quenching component qH. Environmental and Experimental Botany. 1 oct 2018;154:123-33.
57. Brooks MD. A Suppressor of Quenching Regulates Photosynthetic Light Harvesting [Internet]. UC Berkeley; 2012 [cité 22 août 2023]. Disponible sur: https://escholarship.org/uc/item/0th8p5zq
58. Tyystjärvi E, Aro EM. The Rate Constant of Photoinhibition, Measured in Lincomycin-Treated Leaves, is Directly Proportional to Light Intensity. Proceedings of the National Academy of Sciences of the United States of America. avr 1996;93:2213-8.
59. Tyystjärvi E. Chapter Seven - Photoinhibition of Photosystem II. In: Jeon KW, éditeur. International Review of Cell and Molecular Biology. Academic Press; 2013. p. 243-303. (International Review of Cell and Molecular Biology; vol. 300).
60. Bonente G, Ballottari M, Truong TB, Morosinotto T, Ahn TK, Fleming GR, et al. Analysis of LhcSR3, a Protein Essential for Feedback De-Excitation in the Green Alga Chlamydomonas reinhardtii. PLOS Biology. janv 2011;9(1):e1000577.
61. Girolomoni L, Cazzaniga S, Pinnola A, Perozeni F, Ballottari M, Bassi R. LHCSR3 is a nonphotochemical quencher of both photosystems in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences. mars 2019;116(10):4212-7.
62. Finazzi G, Johnson GN, Dallosto L, Joliot P, Wollman FA, Bassi R. A zeaxanthin-independent nonphotochemical quenching mechanism localized in the photosystem II core complex. Proceedings of the National Academy of Sciences. 17 août 2004;101(33):12375-80.
63. Rintamäki E, Martinsuo P, Pursiheimo S, Aro EM. Cooperative regulation of light-harvesting complex II phosphorylation via the plastoquinol and ferredoxin-thioredoxin system in chloroplasts. Proceedings of the National Academy of Sciences. 10 oct 2000;97(21):11644-9.
64. Vink M, Zer H, Alumot N, Gaathon A, Niyogi K, Herrmann RG, et al. Light-Modulated Exposure of the Light-Harvesting Complex II (LHCII) to Protein Kinase(s) and State Transition in *Chlamydomonas reinhardtii* Xanthophyll Mutants. Biochemistry. 1 juin 2004;43(24):7824-33.
65. Johnson X, Vandystadt G, Bujaldon S, Wollman FA, Dubois R, Roussel P, et al. A new setup for in vivo fluorescence imaging of photosynthetic activity. Photosynthesis Research. oct 2009;102(1):85-93.
66. Nawrocki WJ, Liu X, Croce R. Chlamydomonas reinhardtii Exhibits De Facto Constitutive NPQ Capacity in Physiologically Relevant Conditions. Plant Physiology. janv 2020;182(1):472-9.
67. Lahlou A, Tehrani HS, Coghill I, Shpinov Y, Mandal M, Plamont MA, et al. Fluorescence to measure light intensity. Nat Methods. déc 2023;20(12):1930-8.
68. Malerba ME, Palacios MM, Palacios Delgado YM, Beardall J, Marshall DJ. Cell size, photosynthesis and the package effect: An artificial selection approach. New Phytologist. 2018;219(1):449-61.
69. Cross FR, Umen JG. The chlamydomonas cell cycle. The Plant Journal. 2015;82(3):370-92.
70. Zones JM, Blaby IK, Merchant SS, Umen JG. High-resolution profiling of a synchronized diurnal transcriptome from Chlamydomonas reinhardtii reveals continuous cell and metabolic differentiation. The Plant Cell. 2015;27(10):2743-69.
71. Ma F, Salomé PA, Merchant SS, Pellegrini M. Single-cell RNA sequencing of batch Chlamydomonas cultures reveals heterogeneity in their diurnal cycle phase. The Plant Cell. 2021;33(4):1042-57.

72. Mettler T, Mühlhaus T, Hemme D, Schöttler MA, Rupprecht J, Idoine A, et al. Systems analysis of the response of photosynthesis, metabolism, and growth to an increase in irradiance in the photosynthetic model organism Chlamydomonas reinhardtii. The Plant Cell. 2014;26(6):2310-50.
73. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. Proceedings of the National Academy of Sciences. févr 1997;94(3):814-9.
74. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic Gene Expression in a Single Cell. Science. août 2002;297(5584):1183-6.
75. Delosme R, Olive J, Wollman FA. Changes in light energy distribution upon state transitions: an in vivo photoacoustic study of the wild type and photosynthesis mutants from Chlamydomonas reinhardtii. Biochimica et Biophysica Acta (BBA) - Bioenergetics. 15 févr 1996;1273(2):150-8.
76. Minagawa J. State transitions—The molecular remodeling of photosynthetic supercomplexes that controls energy flow in the chloroplast. Biochimica et Biophysica Acta (BBA) - Bioenergetics. 1 août 2011;1807(8):897-905.
77. Miyake C, Shinzaki Y, Miyata M, Tomizawa K ichi. Enhancement of Cyclic Electron Flow Around PSI at High Light and its Contribution to the Induction of Non-Photochemical Quenching of Chl Fluorescence in Intact Leaves of Tobacco Plants. Plant and Cell Physiology. 15 oct 2004;45(10):1426-33.
78. Simpson EH. The Interpretation of Interaction in Contingency Tables. Journal of the Royal Statistical Society: Series B (Methodological). 1951;13(2):238-41.
79. Oxborough K. Imaging of chlorophyll a fluorescence: Theoretical and practical aspects of an emerging technique for the monitoring of photosynthetic performance. Journal of Experimental Botany. mai 2004;55(400):1195-205.
80. Accueil ChlamyStation database [Internet]. [cité 5 juin 2023]. Disponible sur: http://chlamystation.free.fr/
81. Depège N, Bellafiore S, Rochaix JD. Role of Chloroplast Protein Kinase Stt7 in LHCII Phosphorylation and State Transition in Chlamydomonas. Science. mars 2003;299(5612):1572-5.
82. Bujaldon S. Les antennes photosynthétiques chez Chlamydomonas reinhardtii : biogénèse, fonction et régulation [Internet] [These de doctorat]. Sorbonne université; 2023 [cité 4 déc 2024]. Disponible sur: https://theses.fr/2023SORUS346
83. Gorman DS, Levine RP. Cytochrome f and plastocyanin: Their sequence in the photosynthetic electron transport chain of Chlamydomonas reinhardi. Proceedings of the National Academy of Sciences. déc 1965;54(6):1665-9.
84. Chouket R, Pellissier-Tanon A, Lahlou A, Zhang R, Kim D, Plamont MA, et al. Extra kinetic dimensions for label discrimination. Nature Communications. mars 2022;13(1):1482.
85. Katz DS. The Streak Plate Protocol. Microbe Library [Internet]. 2008 [cité 14 juill 2023]; Disponible sur: https://asm.org:443/Protocols/The-Streak-Plate-Protocol
86. Piccioni RG, Chua N, Bennoun P. A Nuclear Mutant of *Chlamydomonas reinhardtii* Defective in Photosynthetic Photophosphorylation: Characterization of the Algal Coupling Factor ATPase. European Journal of Biochemistry. juin 1981;117(1):93-102.
87. Atteia A, de Vitry C, Pierre Y, Popot JL. Identification of mitochondrial proteins in membrane preparations from Chlamydomonas reinhardtii. Journal of Biological Chemistry. 5 janv 1992;267(1):226-34.
88. Bentley JL. Multidimensional binary search trees used for associative searching. Commun ACM. 1 sept 1975;18(9):509-17.
89. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research. 2011;12(85):2825-30.
90. Duda RO, Hart PE, Stork DG. Part 1: Pattern Classification and Scene Analysis. New York etc. : J. Wiley&sons. Vol. 1. 1973.
91. Zhang R, Chouket R, Plamont MA, Kelemen Z, Espagne A, Tebo AG, et al. Macroscale fluorescence imaging against autofluorescence under ambient light. Light: Science & Applications. nov 2018;7(1):97.
92. Pech-Pacheco JL, Cristobal G, Chamorro-Martinez J, Fernandez-Valdivia J. Diatom autofocusing in brightfield microscopy: A comparative study. In: Proceedings 15th International Conference on Pattern Recognition ICPR-2000. 2000. p. 314-7 vol.3.
93. Kosuge K, Tokutsu R, Kim E, Akimoto S, Yokono M, Ueno Y, et al. LHCSR1-dependent fluorescence quenching is mediated by excitation energy transfer from LHCII to photosystem I in Chlamydomonas reinhardtii. Proceedings of the National Academy of Sciences. avr 2018;115(14):3722-7.

94. Cantrell M, Ware MA, Peers G. Characterizing compensatory mechanisms in the absence of photoprotective qE in Chlamydomonas reinhardtii. Photosynth Res. 1 oct 2023;158(1):23-39.
95. Alienor134. Alienor134/Image_segmentation: 2021.
96. Schreiber U. Pulse-Amplitude-Modulation (PAM) Fluorometry and Saturation Pulse Method: An Overview. In: Papageorgiou GC, Govindjee, éditeurs. Chlorophyll a Fluorescence: A Signature of Photosynthesis. Dordrecht: Springer Netherlands; 2004. p. 279-319. (Advances in Photosynthesis and Respiration).
97. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods. mars 2020;17(3):261-72.

Acknowledgments

The authors would like to thank Francis-Andre Wollman, Olivier Vallon, Setphan Eberhard, Julien Sellés, Pierre Neveu, Vincent Croquette and David Bensimon for helpful discussions, Marie-Aude Plamont for preparing Dronpa-2 extracts, Douglas Boari for participating in the Arduino script development, and Sébastien Marino for illustrations.

Funding:

European Innovation Council Pathfinder Open DREAM (grant no. 101046451)
This material is based upon work supported by the ANRT (Association nationale de la recherche et de la technologie) with a CIFRE fellowship granted to AL.

Author contributions:

Conceptualization: AL, DC, LJ, TLS, BB
Methodology: AL, MO, SB, PH, DC, LJ, TLS, BB
Software: AL, WG, PH, DC
Validation: AL, MO, SB, DC, BB
Formal analysis: AL, MO, SB, DC, BB
Investigation: AL, MO, SB, EI, PH, DC, LJ, TLS, BB
Resources: AL, MO, SB, EI, PH, DC, LJ, TLS, BB
Data curation: AL, MO, WG, DC
Writing – original draft: AL, LJ, DC, BB
Writing – Review and editing: AL, MO, SB, WG, EI, PH, DC, LJ, TLS, BB
Visualisation: AL, MO, SB, WG, DC, BB
Supervision: AL, PH, DC, LJ, TLS, BB
Project Administration: LJ, TLS, PH, DC, BB
Funding acquisition: LJ, TLS, PH, DC, BB

Competing interests:

All other authors declare they have no competing interests.

Data and materials availability:

All data are available in the main text, the supplementary materials or the Github repository https://github.com/DreamRepo/NPQScore-data. DOI: https://doi.org/10.5281/zenodo.14606011