

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Genetic interrogation of PsbS structure, function, and agronomic potential

Permalink

<https://escholarship.org/uc/item/41c983sb>

Author

Patel-Tupper, Dhruv

Publication Date

2022

Peer reviewed|Thesis/dissertation

Genetic interrogation of PsbS structure, function, and agronomic potential

By

Dhruv P. Patel-Tupper

A dissertation submitted in partial satisfaction of the
requirements for the degree of

Doctor of Philosophy

in

Plant Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Krishna K. Niyogi, Chair

Professor Henrik Scheller

Professor Graham Fleming

Fall 2022

Abstract

Genetic interrogation of PsbS structure, function, and agronomic potential

by

Dhruv P. Patel-Tupper

Doctor of Philosophy in Plant Biology

University of California, Berkeley

Professor Krishna K. Niyogi, Chair

Photosynthesis necessitates photoprotection. In land plants, Photosystem II Subunit S (PsbS) is necessary for a vast majority of non-photochemical quenching (NPQ) capacity *in planta*. But in the 20+ years since its characterization, our understanding of the mechanisms in which PsbS induces NPQ, the structural differences that define its success and efficiency in quenching, and the potential of engineering NPQ as a trait for photosynthetic efficiency remain poorly understood. This work seeks to narrow knowledge gaps in our understanding and application of PsbS-dependent NPQ. In Chapter 1, I introduce light harvesting and briefly discuss persistent questions and contrasting dogmas in the field. In Chapter 2, I describe new *Nicotiana benthamiana* CRISPR/Cas9 NPQ mutants and discuss their potential in interrogating the ultrafast biophysical mechanisms that underlie quenching. In Chapter 3, I show how CRISPR/Cas9 can also be used to alter expression of PsbS and drive beneficial agronomic traits. Finally in Chapter 4, I showcase the potential behind our current era of transcriptomics—mining the Viridiplantae for diverse LHCSR and PsbS orthologs and applying a high-throughput transient expression pipeline to assess their relative ability to induce NPQ *in planta*. In sum, the confluence of phenotypes described here expands our functional knowledge of PsbS-dependent NPQ and provides exciting ways to interrogate the biophysical mechanisms and structural differences that modulate NPQ efficiency *in vivo*.

Acknowledgements

How do you encapsulate a journey as long, tumultuous, transformative, and exhilarating as a Ph.D.? I could never have expected all the changes that would come from the last six years, but I am infinitely grateful to the community that has gotten me here today.

First, I'd like to thank my Ph.D. advisor Krishna Niyogi. Kris, I joined your lab knowing that I could ask exciting questions with the independence to make mistakes along the way. While I absolutely underestimated the number of mistakes and dead ends...I am leaving confident in my ability to think critically, act creatively, and do *really* good science. Thank you for showing me that giddy feeling you get at a new discovery never changes, and for being a role model in scientific curiosity.

There are several other faculty members that have meaningfully impacted my life. I'd like to thank Peggy Lemaux for her tireless advocacy in all things science communication (and sorghum). Your enthusiasm is contagious, and it has been a treat working with you in the lab and in the community. I'd also like to thank Mary Wildermuth for the opportunities she has given me, and her confidence in my ability to shine as an educator. Mary, seeing how thoughtfully you approach teaching, even amidst an actual pandemic, is something I hope to emulate in my future career. I'm grateful to have worked with Michi Taga, whose tireless advocacy for diversity, equity, and inclusion has given me hope for the future of academia. Thank you to my committee members Graham Fleming and Henrik Scheller for their time and effort in guiding my dissertation. And thank you to those PMB faculty members who have supported me in difficult, important dialogues on inclusion over my time here.

I would also like to express my deepest gratitude to the members of the Niyogi lab for the technical and emotional support they have given me. I truly do not know what we would do without the tireless leadership of our lab manager, Marilyn Kobayashi. Marilyn, your understanding of departmental and grant dealings is second only to the boundless empathy you offer. I've also had exceptional guidance from two senior scientists in the lab, Setsuko Wakao and Masakazu Iwai. Setsuko, your genuine, blunt feedback has been such a blessing to my work and development as a scientist. Masa, time and time again you show me how elegant science can be, and to this day your grasp of biochemistry inspires me. I'd like to thank several former graduate students and post-docs in the lab, Chris Gee, Cindy Amstutz, Daniel Westcott, Simon Alamos, and Vincent Boudreau, for your support, advice, and guidance. To the younger graduate students following in our footsteps, be confident in your abilities, diligent in your experimental design, and forgiving to yourself when things undoubtedly fail (sorry!).

One of the greatest privileges I've had is serving as a lab mentor. To my enthusiastic undergraduate mentees Sienna Marley and Sophie Ma, thank you for your faith in me and for taking a gamble on plant biology—I know you both will do amazing things in

science and medicine and I'm proud of what we have achieved together. To the lab technicians I've been lucky to work with: Armen Kelikian, Anna Leipertz, Aviva Lucas, Raquel Ponce, and Jay Yung—I can't wait to see the exciting things you all discover. Armen, I would not have made it this far, this quickly, without your tireless effort, attention to detail, and dedication to our science. Anna, I'm so grateful for the curiosity, enthusiasm, and personality you've brought to our projects. I'm excited to see where you take Team Rice CRISPRies after Armen and I leave!

I had an outstanding network of graduate student peers that kept me going. I am especially grateful to Tim Jeffers, Ilea Chau-Ly, and Alex Styer for all the laughs and tears we have been able to share as a cohort. Tim, I consider myself so incredibly lucky to have gone through this experience with you in the Niyogi lab. Thank you for your boundless optimism and excitement, and for how you transformed my science in ways I could never imagine. Ilea (and Jordan), thanks for your endless support in my forays into baking and weightlifting. I'm never looking back! And Alex, I'm so glad we went from recruitment weekend roommates to lifelong friends. Seeing you all marry the loves of your lives has been one of the greatest gifts of this experience.

I would be remiss to exclude the two most fashionable scientists I've ever had the pleasure to meet. To Thien Crisanto, thanks for bringing that chaotic good energy to our lab and sharing my love for candy and Coke Zero. To Nicholas Karavolias, it's been a privilege seeing how much you've grown from our undergraduate Plant Physiology class, and I'm so grateful to be sharing in this research/FFAR Fellows/science outreach journey with you.

To my fellow (juris) doctors, Jillian Berk and Eleanor Walker, thanks for keeping me grounded in my sporadic scientific spirals. Jillian, who would have thought we would have gone from high school frenemies (just kidding) to adult frenemies (less kidding)—thanks for always being able to make me cry from laughing and for all your emotional support in science and otherwise. Eleanor, thanks for rounding out our pandemic bubble, for always complimenting my baking, and opening your home office to me, my dissertation, & only the most productive amount of tears.

I've had such loving support from my family. To my siblings Sonal and Dhylan, I love how much our relationship has grown during my Ph.D. and I'm so grateful for it. To my parents Ila and Pete, thank you for believing in me and supporting my dreams and life journey wholeheartedly. Thank you to my family in the East Bay for welcoming me back to California with open arms, and for being so forgiving when I fell off the face of the earth for school. Preena Bhakta and Sonya Lal, you two have made this scientific journey and embracing my queerness so much easier. And to Mitul Bhakta, I wish I could share in this moment with you. I love you.

Finally, to Gus Hathaway Patel-Tupper. You changed my life as soon as you became a part of it; you are the best thing to come from my entire UC Berkeley experience. Thank you for your patience and endless support, for dealing with me and my terrifyingly spontaneous sobbing fits, and for celebrating with me so genuinely when I've discovered something new. I may have written a dissertation, but you planned and executed an entire wedding amidst it all. And I'm so grateful that through you, I now also have Gail, Evan, and Leni Tupper in my life too—thank you all for giving me a second home and more love than I could have imagined in joining a new family.

To you all, thank you from the bottom of my heart for making this journey possible!

Table of Contents

CHAPTER 1 : UNDERSTANDING LAND PLANT NON-PHOTOCHEMICAL QUENCHING	1
1.1 LIGHT HARVESTING, A DOUBLE-EDGED SWORD	1
1.2 STRUCTURAL ACCLIMATION TO ATTENUATE LIGHT INTERCEPTION	1
1.3 NPQ FOR SENSITIVE, ULTRAFAST PHOTOPROTECTION	2
1.4 MEASURING NPQ EN MASSE	3
1.5 A LONG-STANDING DEBATE: HOW DOES PSBS INDUCE NPQ?	4
1.6 ADDRESSING KNOWLEDGE GAPS IN PSBS-DEPENDENT NPQ	6
CHAPTER 2 : TOWARDS AN UNDERSTANDING OF ULTRAFAST NON-PHOTOCHEMICAL QUENCHING MECHANISMS IN <i>NICOTIANA BENTHAMIANA</i>	7
2.1 ACKNOWLEDGEMENTS	7
2.2 ABSTRACT	7
2.3 INTRODUCTION	7
2.4 MATERIALS AND METHODS	9
2.4.1 PLANT MATERIAL AND GROWTH CONDITIONS	9
2.4.2 CONSTRUCT CLONING AND GUIDE RNA DESIGN	10
2.4.3 CHLOROPHYLL FLUORESCENCE PHENOTYPING OF NPQ	11
2.4.4 KANAMYCIN SENSITIVITY ASSAY FOR HIGH-THROUGHPUT CAS9 TRANSGENE DETECTION	12
2.4.5 GENOTYPING AND MUTATION ANALYSIS OF T ₀ AND T ₁ LINES	12
2.4.6 HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY	13
2.5 RESULTS	13
2.5.1 CRISPR/CAS9 GENE-EDITING OF <i>NICOTIANA BENTHAMIANA</i> NPQ ORTHOLOGS	13
2.5.2 NbPSBS1 AND NbPSBS2 BOTH CONTRIBUTE TO NPQ IN PLANTA	14
2.5.3 NbVDE1 IS REQUIRED FOR DE-EPOXIDATION OF VIOLAXANTHIN UNDER HIGH LIGHT	15
2.5.4 SINGLE ORTHOLOG NbZEP KNOCKOUTS DECOUPLE ZEAXANTHIN ACCUMULATION AND STOMATAL MISREGULATION PHENOTYPES	17
2.5.5 NbLUT2 KNOCKOUTS LACK LUTEIN AND RETAIN TRIPLE THE VAZ POOL SIZE	19
2.5.6 GENOTYPIC COMPARISONS RESOLVE LIGHT AND TIME-DEPENDENT CONTRIBUTORS TO THE N. BENTHAMIANA NPQ RESPONSE	20
2.6 DISCUSSION	24
CHAPTER 3 : MULTIPLEXED CRISPR/CAS9 MUTAGENESIS OF RICE <i>PSBS1</i> NON-CODING SEQUENCES GENERATES TRANSGENE-FREE OVEREXPRESSION ALLELES	26
3.1 ACKNOWLEDGEMENTS	26
3.2 ABSTRACT	26

3.3 INTRODUCTION	26
3.4 MATERIALS AND METHODS	28
3.4.1 PLANT MATERIAL AND GROWTH CONDITIONS	28
3.4.2 GUIDE RNA DESIGN AND CLONING INTO CAS9 VECTOR PRGEB32	28
3.4.3 IN VITRO gRNA VALIDATION OF UPSTREAM OSPBS1 NON-CODING SEQUENCES	30
3.4.4 INDUCTION OF EMBRYOGENIC CALLI	31
3.4.5 AGROBACTERIUM-MEDIATED TRANSFORMATION	31
3.4.6 CHLOROPHYLL FLUORESCENCE MEASUREMENTS OF NPQ	32
3.4.7 HYGROMYCIN SENSITIVITY ASSAY FOR HIGH-THROUGHPUT CAS9 TRANSGENE DETECTION	32
3.4.8 TOTAL RNA/PROTEIN EXTRACTION OF REPRESENTATIVE LINES WITH VARYING NPQ	33
3.4.9 QUANTITATIVE RT-PCR OF OSPBS1 RELATIVE TO TWO REFERENCE GENES	33
3.4.10 IMMUNOBLOTTING OF WHOLE LEAF PROTEIN EXTRACTS	34
3.4.11 STEADY-STATE NPQ ACCLIMATION TO LOW, MODERATE, AND HIGH LIGHT	34
3.4.12 IN-PARALLEL GAS EXCHANGE AND CHLOROPHYLL FLUORESCENCE ANALYSIS UNDER INCREASING RED LIGHT	35
3.4.13 PCR GENOTYPING OF TRANSGENE-FREE, EDITED LINES	36
3.4.14 PACBIO LONG-READ WHOLE-GENOME SEQUENCING AND ANALYSIS	36
3.5 RESULTS	37
3.5.1 A LIBRARY OF INDEPENDENT EVENTS FOR MULTIPLEXED EDITING UPSTREAM OF OSPBS1	37
3.5.2 HIGH-THROUGHPUT CHLOROPHYLL FLUORESCENCE SCREENING OF EDITED, SEMI-DOMINANT PsBS ALLELES	38
3.5.3 GENE-EDITED, OVEREXPRESSION ALLELES ARE PRESENT, ALBEIT RARE	39
3.5.4 CHLOROPHYLL FLUORESCENCE PHENOTYPES CORRELATE WITH PsBS PROTEIN ABUNDANCE	40
3.5.5 VARYING PsBS ABUNDANCE ALTERS STEADY STATE GAS EXCHANGE PHENOTYPES IN RESPONSE TO RED-LIGHT	43
3.5.6 Q_A REDOX STATE IS AN INCONSISTENT PREDICTOR OF RICE G_{SW}	44
3.5.7 MOST NON-WT PHENOTYPES ARE EXPLAINED BY VARIATION AT THE 5'UTR	46
3.5.8 LONG-READ SEQUENCING RESOLVES COMPLEX STRUCTURAL VARIANTS UNDERLYING OVEREXPRESSION	47
3.6 DISCUSSION	51

CHAPTER 4 : QUANTITATIVE TRANSIENT EXPRESSION OF DIVERSE PSBS PROTEINS REVEALS THE EVOLUTIONARY ORIGIN OF LAND PLANT-LIKE PSBS ACTIVITY **54**

4.1 ACKNOWLEDGEMENTS	54
4.2 ABSTRACT	54
4.3 INTRODUCTION	54
4.4 MATERIALS AND METHODS	56
4.4.1 PLANT MATERIAL AND GROWTH CONDITIONS	56
4.4.2 MINING OF DIVERSE PsBS AND LHCSR	56
4.4.3 EFI-EST CLUSTERING BY SEQUENCE SIMILARITY NETWORKS AND VISUALIZATION VIA CYTOSCAPE	57

4.4.4 MULTIPLE SEQUENCE ALIGNMENT (MSA), PHYLOGENETIC TREE CONSTRUCTION, AND STATISTICAL ANALYSIS	57
4.4.5 TRANSIENT EXPRESSION CONSTRUCT CLONING AND DESIGN	58
4.4.6 GENE SYNTHESIS OF REPRESENTATIVE AND SYNTHETIC LHCSR AND PsbS	59
4.4.7 TRANSIENT EXPRESSION	60
4.4.8 CHLOROPHYLL FLUORESCENCE PHENOTYPING	60
4.4.9 FLUORESCENCE REPORTER QUANTIFICATION	61
4.4.10 ASSESSING NPQ BY FLUORESCENCE PHENOTYPES IN AGGREGATE	61
4.4.11 PLOTTING PREDICTED STRUCTURAL REGIONS AS A REFERENCE FOR MSA	61
4.5 RESULTS	62
4.5.1 IDENTIFICATION OF 269 LHCSR AND 1199 PsbS ORTHOLOGS ACROSS THE VIRIDIPLANTAE	62
4.5.2 SEQUENCE SIMILARITY NETWORKS CONSOLIDATE LHCSR AND PsbS DIVERSITY	63
4.5.3 CLUSTERS MIRROR PHYLOGENETIC DIVERSITY AND FORM BASED ON SEQUENCE QUALITY	64
4.5.4 INTER-CLUSTER COMPARISONS RESOLVE STATISTICALLY SIGNIFICANT DIFFERENCES AS DEMONSTRATED BY "DIVERSITY BAR PLOTS"	66
4.5.5 HIGH-THROUGHPUT CLONING OF DIVERSE LHCSR AND PsbS	68
4.5.6 NPQ BY FLUORESCENCE PLOTS MIRROR STABLE IN PLANTA PHENOTYPES	73
4.5.7 STREPTOPHYTE LHCSR EXHIBIT SLOWLY INDUCIBLE NPQ CAPACITY	75
4.5.8 NICOTIANA BENTHAMIANA PsbS IS NOT THE MOST EFFICIENT QUENCHER IN PLANTA	77
4.5.9 MONOCOT PsbS PERFORM POORLY DESPITE HIGH NATIVE ACTIVITY	79
4.5.10 PAIRWISE COMPARISONS MAY HIGHLIGHT IMPORTANT FUNCTIONAL/STRUCTURAL RESIDUES	81
4.5.11 A RICE "PSEUDOGENE" IS BETTER ADAPTED FOR QUENCHING IN DICOTS	82
4.5.12 LAND-PLANT LIKE PsbS FIRST AROSE IN STREPTOPHYTE ALGAE	83
4.5.13 MULTIPLE SEQUENCE ALIGNMENTS IDENTIFY INTERESTING NPQ PATTERNS THAT REQUIRE FURTHER STATISTICAL INVESTIGATION	85
4.6 DISCUSSION	92
CHAPTER 5 : FUTURE OUTLOOKS	95
CHAPTER 6 : REFERENCES	96

List of Figures

Figure 2-1. Biosynthetic and mechanistic pathways underlying carotenoid production and NPQ in plants.	9
Figure 2-2. Plasmid map of an adapted pCambia2300 backbone for CRISPR/Cas9 mutagenesis of <i>Nicotiana benthamiana</i>	11
Figure 2-3. Gene models and target sites for orthologous <i>N. benthamiana</i> NPQ gene mutagenesis.	14

Figure 2-4. NPQ kinetics of WT, psbs1, and psbs1psbs2 (npq4) tobacco mutants.....	15
Figure 2-5. Changes in xanthophyll pigment profile of NbVDE1 and NbVDE2 mutants in response to high-light.....	16
Figure 2-6. Xanthophyll pigment profile & growth phenotypes of dark-acclimated zep1, zep2, and zep1zep2 mutants.....	18
Figure 2-7. Carotenoid profile of <i>N. benthamiana</i> mutants lacking LUT2-1 and LUT2-2.....	20
Figure 2-8. NPQ kinetics of tobacco mutants under moderate ($750 \mu\text{mol m}^{-2} \text{s}^{-1}$) and high ($1500 \mu\text{mol m}^{-2} \text{s}^{-1}$) actinic light.	22
Figure 3-1. Plasmid map of pRGEB32 after GoldenGate assembly of the synthesized 8 gRNA spacer insert.	30
Figure 3-2. CRISPR/Cas9 target sites and naming scheme for <i>Nipponbare</i> (subsp. <i>japonica</i>) transformants.	37
Figure 3-3. A high-throughput chlorophyll fluorescence resolves quantitative variation and Cas9 transgene segregation of T ₁ progeny.	38
Figure 3-4. Maximum NPQ of 120 putative homozygous alleles spanning 78 T ₀ events.	40
Figure 3-5. Changes in PsbS protein abundance underlie NPQ and ΦPSII phenotypes.	41
Figure 3-6. Chlorophyll fluorescence and gas exchange measurements of Event 2 lines with varying PsbS abundance.	43
Figure 3-7. Correlation of Q _A redox state (1-qL) and g _{sw} as a predictor of iWUE with varying [PsbS].	45
Figure 3-8. Variation near and within the 5'UTR likely drives observed phenotypic variation.	47
Figure 3-9. Long-read sequencing of two overexpression alleles identifies inversions upstream of OsPsbS1.....	51
Figure 4-1. Plasmid map of an adapted pCAMBIA1300 backbone (pUBQ-AtPsbS_cTP-[SpeI_HpaI]-IntF2A-mScarlet+P19) used for quantitative transient expression.....	59
Figure 4-2. Dendrogram and species richness mined for diverse LHCSR and PsbS.....	63
Figure 4-3. Sequence similarity networks of mined LHCSR and PsbS orthologs.....	64
Figure 4-4. Stringency and cluster-dependent differences in sequence length and phylogenetic composition.	65
Figure 4-5. Multiple sequence alignment and Chi-squared analysis of amino acid enrichment within SSN clusters.	67
Figure 4-6. In-parallel cloning for high-throughput quantitative transient expression of LHCSR and PsbS variants.....	69
Figure 4-7. Quantitative transient expression of Arabidopsis PsbS recapitulates phenotypes previously observed in stable Arabidopsis transformants.....	74
Figure 4-8. Aggregate MqE and MqTotal phenotypes for all infiltrated LHCSR and PsbS.....	75
Figure 4-9. Streptophyte LHCSR show a weak ability to induce MqTotal but not MqE.....	76
Figure 4-10. Native <i>N. benthamiana</i> PsbS is not the most efficient PsbS at quenching in planta.....	78
Figure 4-11. Monocot PsbS induce NPQ poorly relative to the entire LHCSR/PsbS library.	80

Figure 4-12. PsbS from the monocot <i>Zea mays</i> is atypical in quenching capacity relative to other monocot PsbS.	81
Figure 4-13. A rice “pseudogene” exhibits higher NPQ capacity per fluorescence than its dominant <i>in vivo</i> ortholog.....	82
Figure 4-14. Cross green algal PsbS analysis identifies robust land plant-like activity in the Zygnematophyceae.....	84
Figure 4-15. A UPGMA dendrogram of all tested PsbS relative to MqE and MqTot slopes.	86
Figure 4-16. UPGMA MSA in dendrogram order of all tested PsbS against MqE.....	87
Figure 4-17. PsbS MSA of all amino acids in descending MqE order, excluding chlorophyte green algae.....	88
Figure 4-18. PsbS MSA of protonatable residues in descending MqE order, excluding chlorophyte green algae	90
Figure 4-19. PsbS MSA of hydropathy in descending MqE order, excluding chlorophyte green algae.....	91
Figure 4-20. Correlation of hydrophobic residue at amino acid 140 with observed MqE	92

List of Tables

Table 2-1. <i>Nicotiana benthamiana</i> orthologs and gRNA spacer sequences used for CRISPR/Cas9-mediated knockout.....	10
Table 2-2. Primers for genomic DNA amplification of <i>Nicotiana benthamiana</i> NPQ gene orthologs.....	12
Table 2-3. <i>A. thaliana</i> NPQ proteins and their respective <i>N. benthamiana</i> orthologs.....	14
Table 2-4. NbPsbS knockout alleles isolated across five independent transformants.....	15
Table 2-5. NbVDE knockout alleles isolated across ten independent transformants.....	17
Table 2-6. Viable NbZEP knockout alleles isolated across eleven independent transformants..	19
Table 2-7. NbLUT2 knockout alleles isolated across eleven independent transformants.	20
Table 3-1. OsPsbS1 non-coding sequence gRNA spacer sequences and positions relative to the start codon.....	29
Table 3-2. Oligomers for <i>in vitro</i> Cas9-gRNA editing of <i>O. sativa</i> (ssp. <i>indica</i>) OsPsbS1 non-coding sequences.....	31
Table 3-3. qRT-PCR primer pairs and observed primer efficiency in WT across a 1:81 dilution series.	34
Table 3-4. Primers used for PCR amplification and genotyping of OsPsbS1 NCS.....	36
Table 4-1. PsbS and LHCSR query proteins used to BLAST against published genomes and transcriptomes.	56
Table 4-2. Overhangs used for Gibson Assembly of synthesized LHCSR, PsbS, and Chlamy LHCB.	60
Table 4-3. Synthesized LHCSR, PsbS, and CrLHCB sequences used in this study, with cTP-cleaved mature protein length.....	70

Chapter 1 : Understanding land plant non-photochemical quenching

1.1 Light harvesting, a double-edged sword

Efficient photosynthesis is essential for productive food crops. Concordantly chlorophyll (Chl), the light-absorbing pigment in light-harvesting complexes (LHCs), is incredibly efficient at absorbing red and blue light, with a long-lived ($\sim$ ns) singlet excited state ($^1\text{Chl}^*$) sufficient for charge separation and formation of electrochemical potential. However, this highly energetic reaction does not come without consequence, providing Chl the opportunity to become a potent photosensitizer.

Natural light is often in excess. Such conditions can leave $^1\text{Chl}^*$ without an adequate electron acceptor (e.g. oxidized plastoquinone) or available quencher (e.g. carotenoids), resulting in the formation of the longer-lived ($\sim$ ms) triplet chlorophyll state ($^3\text{Chl}^*$) by intersystem crossing¹. Interaction of $^3\text{Chl}^*$ with molecular oxygen (O_2) produces arguably the most photoinhibitory of chloroplastic reactive oxygen species (ROS), singlet oxygen ($^1\text{O}_2$)^{2,3}. Carotenoids, namely lutein and zeaxanthin, are present in sufficient proximity to quench $^3\text{Chl}^*$ and $^1\text{O}_2$ in and around LHCII⁴. However, at the reaction center (RC), weak coupling of Chl to the relatively distant carotenoid β -carotene limits the efficiency of $^3\text{Chl}^*$ quenching necessary to efficiently prevent formation of $^1\text{O}_2$ ^{5,6}. Sustained accumulation of $^1\text{O}_2$ induces photooxidative damage to pigments, proteins, and lipids within the thylakoid membrane that limit photosynthetic efficiency⁷. Given that steady-state CO_2 assimilation saturates at light intensities well-below peak daily irradiance⁸, plants must regulate—and largely dissipate—excess light to maintain plant fitness.

1.2 Structural acclimation to attenuate light interception

Plants employ myriad acclimation strategies to regulate total light absorption by Chl in excess light. Adjusting leaf angle⁹, architecture¹⁰, and reflectance¹¹ are several macro-scale strategies used to minimize functional light interception. At the cellular scale, chloroplast stacking can further decrease the functional penetrance of excess light across the three-dimensional cell “canopy”¹². At the sub-cellular level, reductions in total chlorophyll¹³ and functional chlorophyll around reaction centers by modulating antennae size¹⁴ can further decrease excitation pressure on photosystem II (PSII). Additional partitioning of light between photosystem I (PSI) and PSII by LHCII state transitions further fine-tunes the molecular scale at which light is absorbed¹⁵. However, these light avoidance strategies lack the magnitude of impact, timescale of action, and specificity to fully explain photoprotection *in planta*.

Light intensity fluctuates by region, season, and climate, forming the basis of broader sun-shade photosynthetic strategies that allow plants to thrive in different ecosystems¹⁶. However, light intensity also fluctuates by time of day, local shading, and most dramatically, by the diurnal movement of sunflecks in dense, fluid canopies¹⁷. To cope with these dynamic environments, photosynthetic organisms evolved the ability to biophysically dissipate excess energy by quenching $^1\text{Chl}^*$ at the photosystem scale—a suite of mechanisms termed non-photochemical quenching (NPQ)¹⁸.

1.3 NPQ for sensitive, ultrafast photoprotection

NPQ is comprised of several components defined by their relaxation kinetics and molecular players. In plants, the rapidly induced, energy-dependent quenching (qE) requires the PSII subunit S (PsbS), the photoprotective carotenoid zeaxanthin (Zx), and a trans-thylakoid pH gradient (ΔpH), regulating quenching that is induced on a timescale of seconds to minutes^{18,19}. Zeaxanthin (Zx)-dependent quenching (qZ) quenches on the timescale of minutes²⁰ and is contingent on Zx produced via a xanthophyll cycle²¹. A recently discovered form of sustained antennae quenching (qH) has been found to be lipocalin-dependent, and occurs under stress conditions such as cold and high light²². The most slowly relaxing component of NPQ, photoinhibitory quenching (qI), results from photoinactivation of PSII by ROS—relaxing on the timescale of hours to days²³. In concert, NPQ mechanisms provide the flexibility to cope with light fluctuations at varying intensities and timescales.

Core NPQ proteins have co-opted photosynthetic cues to maintain high sensitivity in NPQ induction and relaxation. When linear electron flow is high (i.e. in excess light), stromal inorganic phosphate (P_i) becomes limiting, constraining ATP synthase activity and resulting in lumen acidification^{24,25}. PsbS contains two evolutionarily conserved, protonatable residues (E122, E226) that sense this change and are necessary for qE^{26,27}. Protonation of these residues by the ΔpH results in a conformational change within PsbS²⁸ that is thought to trigger quenching *in vivo*. The luminal enzyme responsible for xanthophyll de-epoxidation of violaxanthin (Vx) to the quenching carotenoid Zx, violaxanthin de-epoxidase (VDE), is also activated by acidic pH²⁹. VDE binds to the thylakoid membrane after a low pH-dependent conformational change that allows binding of its substrate Vx³⁰ to trigger de-epoxidation using ascorbate as a co-substrate³¹. In contrast, the stromal enzyme responsible for relaxation of qZ by conversion of Zx back to Vx, zeaxanthin epoxidase (ZEP), is degraded under high light³² thus sustaining qZ capacity under light stress. A similar negative feedback loop to maintain NPQ occurs in qI; D1 functions as a “suicide” protein that quenches ROS in its oxidative damage, transiently suppressing the abundance of available RCs for light

harvesting while shielding other photosystem components with substantially slower turnover from irreversible damage³³. Post-translational regulation of NPQ by photosynthesis provides rapid and specific control over photoprotection.

However, while NPQ protects the photosynthetic apparatus from excess light, it also competes with light harvesting and can limit photosynthetic efficiency. This effect becomes significantly consequential under fluctuating light, where rapid induction and slower relaxation of NPQ can limit CO₂ assimilation at estimates of up to 30%³⁴. Optimizing the light-dependent and light-independent reactions of photosynthesis under dynamic, field-relevant light conditions is a rapidly expanding field³⁵, and natural variation in light response may provide reservoirs for breeding for higher photosynthetic efficiency^{36,37}. Excitingly, a transgenic approach to increase the relaxation of fast-acting NPQ by overexpression of VDE, PsbS, and ZEP (hereafter VPZ) increased biomass accumulation of the field model crop *Nicotiana tabacum* by ~15%³⁸. This approach was also successfully applied in an elite crop variety for the first time, achieving increases in soybean yield of ~20%³⁹. However, failures to reproduce significant VPZ biomass phenotypes in *Arabidopsis thaliana*⁴⁰, potato⁴¹, and bentgrass⁴² warrant caution. Achieving fast NPQ relaxation at the expense of wild-type (WT) NPQ capacity can bring substantial fitness costs, emphasizing the need to optimize VPZ protein stoichiometry and screen for optimal phenotypes rigorously before assessing biomass and yield.

1.4 Measuring NPQ en masse

¹Chl* can be photochemically (qP) and non-photochemically (NPQ) quenched on a picosecond timescale. However, a third, nanosecond-resolution pathway can also quench excited chlorophyll by red-shifted re-emission, or fluorescence (F). The relatively slow speed makes chlorophyll fluorescence a negligible quencher, but its inverse relationship to qP and NPQ has proved to be a robust and essential tool for non-destructive monitoring of photoprotection *in vivo*^{43,44}. In brief, pulse-amplitude modulated (PAM) fluorometry capitalizes on the dynamics between qP, NPQ, and F to infer changes in quenching in the presence and absence of strong actinic light⁴⁵. Quenching of maximum fluorescence (F_m) in the light (F_m') allows quantification of NPQ as a Stern-Volmer relationship: $([F_m - F_m'] / F_m')$.

Time-correlated single photon counting (TCSPC) of chlorophyll fluorescence lifetime measurements has built on the principles of PAM fluorometry and has increased the sensitivity with which we can monitor changes in quenching⁴⁶. Compared to fluorescence yields by PAM, fluorescence lifetime measurements are not influenced by processes like light scattering, photobleaching, or chloroplast movement that reduce

fluorescence yield but do not quench excited chlorophyll⁴⁷. Lifetime snapshots at various light/dark exposures allow quantification of an analogous fluorescence quenching ratio, NPQ_τ, from a biexponential fit of amplitude-weighted average chlorophyll fluorescence lifetimes. Correlative biochemical analysis, such as photosynthetic pigment profiling by high-performance liquid chromatography (HPLC) to understand dynamic changes in xanthophyll kinetics or immunoblotting to monitor the turnover of D1 and consequences of photoinhibition, has also proven useful in understanding NPQ at the leaf/thylakoid scale. However, changes in physiology are insufficient in fully describing the mechanisms underlying quenching.

1.5 A long-standing debate: How does PsbS induce NPQ?

PsbS is required for NPQ *in planta*. Since its characterization over 20 years ago¹⁹, hundreds of publications have followed to interrogate the function of this ubiquitous protein in plants. Despite this, there is no agreed upon theory regarding the mechanisms of PsbS-dependent NPQ and the site of quenching *in vivo*.

There are several areas of strong consensus. In plants, NPQ is dependent on the presence of PsbS¹⁹, which is able to sense changes in ΔpH via two protonatable residues²⁷ and, in concert with increasing amounts of Zx⁴⁸, induces WT levels of NPQ. Overexpression of PsbS increases NPQ capacity⁴⁹, even in the absence of Zx⁵⁰, though at reduced magnitude. PsbS is found in the grana stacks enriched with PSII-LHCII and exists as a dimer in the dark, monomerizing in response to excess light⁵¹. And in the absence of lutein and zeaxanthin, PsbS lacks the ability for rapidly reversible NPQ⁵². These results, made possible by a rich genetic library of Arabidopsis NPQ mutants, have opened the door to address several fundamental, unanswered questions.

Where is PsbS? Well before its involvement in NPQ was identified, PsbS was characterized as a 22-kDa non-essential oxygen-evolving complex protein that co-purified with RCs, dependent on detergent conditions⁵³. Later co-immunoprecipitation of PsbS in maize thylakoids at relatively weak detergent conditions identified interactions with the minor antennae proteins CP29 (LHCB4) and major LHCII trimers⁵⁴. Two recent *in vivo* cross-linking studies have reinforced high-light-dependent and independent interactions with major and minor LHCII⁵⁵, showing an enrichment in interactions between PsbS and minor LHCII in the presence of Zx⁵⁶. However specific residues necessary for interaction are still unknown, and spatial resolution in PsbS co-migration has been confounded by a large distribution range of PsbS across varying levels of PSII-LHCII complex assembly^{51,57}. Innovative approaches to resolve the elusive

binding site necessary for PsbS-dependent NPQ will be necessary to advance our fundamental understanding of its function.

What does PsbS interact with? Outside of a general co-migration with PSII-LHCII, there are conflicting records regarding the allosteric and specific protein interactions that underlie PsbS activity. Early in the characterization of PsbS, this four-transmembrane LHC-like protein was identified to bind chlorophyll and pigments⁵⁸, though it was also structurally stable in its absence⁵⁹. However, *in vitro* and *in vivo* reconstitution of PsbS showed that it lacks the xanthophyll- and chlorophyll-binding capacity observed in other LHCII proteins⁶⁰, and the two conserved chlorophyll-binding residues appear more structurally essential for salt bridge formation and protein structure than functional binding of chlorophyll⁶¹. Whether monomerization of PsbS and interaction with LHCII creates a novel inter-protein pigment-binding site remains to be proven. Molecular dynamics simulations that reveal a Zx-binding interface between CP29 and monomerized PsbS under low pH, which is also destabilized under neutral pH, warrant *in vivo* investigation⁶². To that aim, site-directed mutagenesis of CP29 chlorophyll-binding sites has shown utility in describing their contribution to NPQ *in vivo*⁶³. Further research placing these structural differences within the larger PSII-LHCII supercomplex will be required for a complete description of how quenching is mediated at these sites.

Can PsbS be the site of quenching in vivo? While the NPQ phenotype of a PsbS knockout mutant (*npq4*) is dramatic, its absence is epistatic to other fast-acting NPQ components and thus complicates genetic dissection by chlorophyll fluorescence. Work by Johnson and Ruban has circumvented this using the cyclic electron flow inducer diaminodurine (DAD) to induce artificially high Δ pH. In doing so they report the ability to induce WT levels of NPQ in thylakoids lacking PsbS⁶⁴ and partially complement WT+DAD-like NPQ in thylakoids lacking lutein and zeaxanthin⁶⁵. The authors posit that this result, in concert with *in vitro* evidence of the role of LHCII aggregation in quenching chlorophyll fluorescence lifetime⁶⁶, suggests that PsbS and zeaxanthin serve as allosteric regulators that sense changes in Δ pH and potentiate LHCII's inherent capacity for quenching rather than act directly as the quencher itself *in vivo*^{67,68}. Two recent studies challenge this thinking.

The first arises from the characterization of NPQ in a no minor antennae (*NoM*) *Arabidopsis* mutant⁶⁹. Dall'Osto et al. reveal the presence of a biphasic NPQ response in the absence of minor antennae that requires PsbS and, in its second phase, Zx. Transient absorption (TA) spectroscopy revealed that this biphasic NPQ is dependent entirely on excitation energy transfer (EET) between the Chl Q_y state and the Zx S_1 state⁷⁰, with the consistency in NPQ phenotype between *NoM* and *NoM lut2* mutants suggesting either

(1) lutein is not involved in quenching LHCII trimers still present in the *NoM* mutant or (2) Vx, Ax, and/or Zx functionally compensate for the loss of lutein *in vivo*. Both conclusions challenge the *in vivo* significance of density- and pH-dependent quenching in LHCII in the absence of PsbS and Zx *in vitro*^{71–74}.

Snapshot TA, an advancement on traditional single-timepoint TA approaches, disentangles ultrafast EET from charge transfer (CT) and the formation of the Zx radical cation *in vivo*⁷⁵. Interestingly, Park and Fischer et al. report the presence of biphasic EET⁷⁶ reminiscent of NoM EET-dependent NPQ⁶⁹. Excited state absorption signatures of CT, associated with rapidly inducible NPQ_r, were inhibited by the PsbS-inhibiting cross-linker DTSSP⁷⁷. Taken together, these two pivotal studies have revealed the significance of PsbS, Zx, and minor LHCII in rapidly inducible NPQ by CT *in vivo*. In contrast EET may be explained by the accumulation and subsequent quenching by Zx in LHCII trimers. Further spectroscopic studies of mutants may provide more direct causality. A more holistic understanding of thylakoid-level changes in high light (e.g. membrane thinning) may help reconcile the aforementioned fast-acting NPQ kinetics on what may be independent signatures of LHCII aggregation⁷⁸. And while follow-up work has challenged the significance of the minor antennae in quenching given the eventual convergence of NoM and WT steady-state NPQ⁷⁹, the impact of NPQ on photosynthesis under fluctuating light is non-trivial. A deeper understanding of these dynamics may drive efforts to optimize photosynthetic efficiency *in vivo*^{38,39}.

1.6 Addressing knowledge gaps in PsbS-dependent NPQ

NPQ is a highly conserved, photoprotective process that sustains plant fitness⁸⁰. However, basic and applied knowledge gaps limit a holistic understanding *in vivo*. I seek to narrow these gaps in our understanding of PsbS-dependent NPQ through (1) the development of genetic tools for disentangling ultrafast NPQ, (2) genetic engineering of PsbS for improved photosynthetic and water use efficiency, and (3) a comprehensive screen of diverse NPQ proteins to further our understanding of the requirements of rapidly inducible qE. In doing so, this work has the capacity to expand our functional understanding of land plant NPQ.

Chapter 2 : Towards an understanding of ultrafast non-photochemical quenching mechanisms in *Nicotiana benthamiana*

2.1 Acknowledgements

Nicholas Karavolias cloned the first-generation of CRISPR/Cas9 constructs. Bailey Van Bockern and Lucero Jimenez, under the guidance of David M. Tricoli, generated the *Nicotiana benthamiana* transformants under contract with the Ralph M. Parsons Foundation Plant Transformation Facility at UC Davis. Aviva Lucas screened mutants for homozygous knockouts of NbLUT2, NbVDE, and NbZEP. Lam Lam, whose spectroscopic data is not reported here, has been a great motivator in pursuing these genetic tools.

2.2 Abstract

Snapshot transient absorption (TA) spectroscopy has dramatically improved our ability to temporally resolve signatures of carotenoid-chlorophyll charge transfer and excitation energy transfer *in vivo*. However, attempts to apply this approach in *Arabidopsis thaliana* mutants have been unsuccessful. To circumvent this issue, I have generated CRISPR/Cas9 mutant knockouts in the broad-leafed, high NPQ model plant *Nicotiana benthamiana*. Across four target genes, PsbS, VDE, ZEP, and LUT2, I identified functional ortholog(s) *in planta* and confirmed expected differences in NPQ and pigment phenotypes. Using the generated suite of mutants, I show how adjusting light intensity and temporal resolution reveals interesting discrepancies between the two model plant species. The germplasm generated has potential to improve our knowledge of ultrafast NPQ by Snapshot TA.

2.3 Introduction

Forward genetic screens facilitated by pulse-amplitude modulated (PAM) fluorometry and high-performance liquid chromatography (HPLC) have driven our understanding of land plant non-photochemical quenching (NPQ). From these studies we have learned that WT NPQ capacity requires PsbS (NPQ4) *in vivo*¹⁹, is augmented by zeaxanthin (Zx) produced by VDE (NPQ1)⁴⁸, and is relaxed by recovery of violaxanthin (Vx) by ZEP (ABA1/NPQ2) under light-limiting conditions⁴⁸. A role in NPQ was also attributed to the carotenoid lutein (Lut), as knockouts in *lut2*⁸¹ and *npq1* revealed additive contributions to NPQ^{52,82}. A xanthophyll biosynthetic pathway is summarized below (**Figure 2-1A**).

These genetic screens have also revealed the incredible plasticity inherent to photosynthesis. Lutein is the most abundant xanthophyll in higher plants, present constitutively in LHCII with a role in both light harvesting and quenching of excited

chlorophyll⁸³. In the absence of Lut (e.g. *lut2*), violaxanthin (Vx) occupies lutein-binding sites in LHCII and leads to significantly decreased ³Chl* quenching capacity and increased ¹O₂ yield⁸⁴. Loss of lutein is concomitant with a proportionate increase in VAZ (Vx + Ax + Zx) pool size⁸¹. The contributions of Zx to NPQ *in vivo* appear significantly larger than that of Lut⁸⁵, saturating on a timescale of minutes²⁰. While flexibility in xanthophyll binding has obscured our ability to decouple the contribution of each xanthophyll genetically, higher order mutants like *lut2npq1* have proven essential in describing redundant and synergistic roles in ROS scavenging and NPQ⁵². In bulk, chlorophyll-carotenoid triplet-energy transfer can be up to 95% efficient⁸⁶. Thus xanthophylls are critical in rapidly inducible NPQ of ¹Chl* and scavenging of the longer lived ³Chl* and ¹O₂ photosensitizers (**Figure 2-1B**).

While physiological studies have provided a strong understanding of the essential components of NPQ, we have only recently been able to describe its underlying mechanisms at an ultrafast (fs-ps) timescale. A recent advance, snapshot transient absorption (TA) spectroscopy, has revealed an exciting ability to track the excited state absorption signals indicative of charge transfer (CT) and excitation energy transfer (EET) mechanisms driving NPQ *in vivo*⁷⁵. Coupled with changes in chlorophyll fluorescence lifetime measured by time correlated single photon counting (TCSPC)⁴⁶, Park and Fischer et al. were able to identify the co-existence of CT⁷⁷ and EET⁷⁶ *in vivo* and largely correlate those signatures with the activity of PsbS and Zx by chemical inhibition.

However, the conclusions made were constrained by the efficiency of chemical inhibition and other complicating factors, such as the presence of Zx prior to light exposure. Genetic mutants in a species amenable to Snapshot TA would help more specifically determine the contributions of each missing protein to ultrafast NPQ processes. In this chapter, I describe my efforts to generate relevant germplasm in the allotetraploid *Nicotiana benthamiana*, a broad-leaved model dicot with significantly larger biomass and higher NPQ capacity relative to *Arabidopsis thaliana*. Knockouts in orthologs of four Arabidopsis genes, PsbS, VDE, ZEP, and LUT2, recapitulate expected phenotypes. Interestingly, there are significant divergences on the relative dependence of NPQ on Zx and Lut. Leveraging different testing conditions, I reveal differences in light- and time-dependent NPQ phenotypes that will prove useful in investigating these mutants by Snapshot TA.

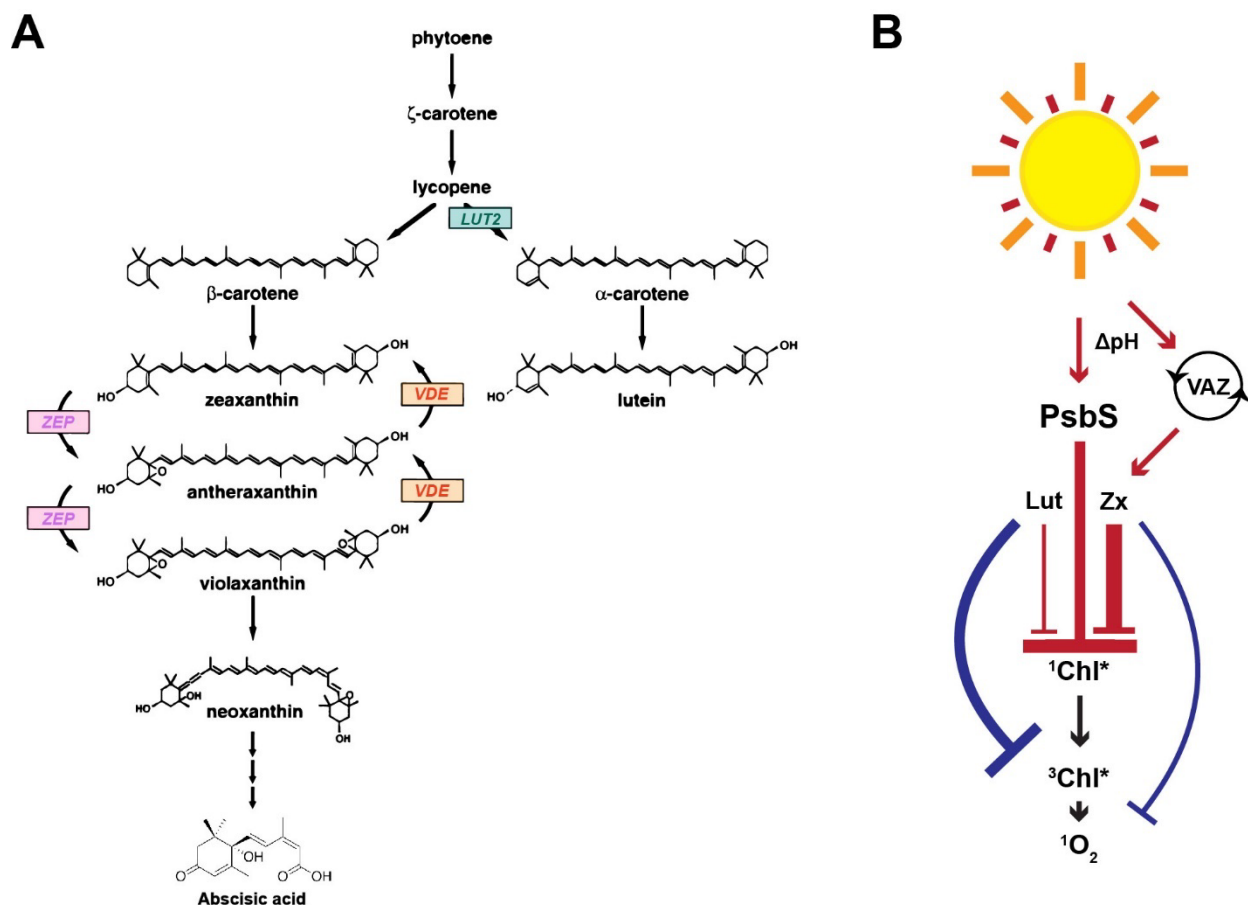


Figure 2-1. Biosynthetic and mechanistic pathways underlying carotenoid production and NPQ in plants.

(A) Biosynthetic pathways adapted from Niyogi et al., 1997⁸⁷. Enzymes of interest are shown in color: LUT2, blue-green, VDE/NPQ1 (orange), and ZEP/ABA1/NPQ2 (purple). **(B)** Schematic describing the roles of PsbS, Lutein (Lut), and Zeaxanthin (Zx) in photoprotection. Pathways implicated in NPQ are shown in red, whereas pathways associated with ROS scavenging are shown in blue, weighted by their relative contribution.

2.4 Materials and Methods

2.4.1 Plant material and growth conditions

Nicotiana benthamiana (accession Nb-1) plants were grown under ambient light conditions at a 10-hour daylength in a south-facing greenhouse. Seeds were germinated directly on a mixture of 4 parts Sunshine Mix #1 (Sungro) and 1 part perlite and coated with a thin layer of sand to stabilize seedlings. Plants were bottom watered for two weeks, the first week under a high humidity dome. Pots were then thinned and moved into flats with drainage to be top watered through maturity. Plants were fertilized with JR Peter's Blue 20-20-20 fertilizer (<https://www.jrpeters.com/>) monthly.

2.4.2 Construct cloning and guide RNA design

Candidate gRNA were identified using CRISPR-P (crispr.hzau.edu.cn)⁸⁸ and two high-scoring candidates for each gene were chosen depending (1) on their ability to target both *N. benthamiana* orthologs and (2) sequence similarity to the orthologous Arabidopsis gene downstream of the chloroplast transit peptide sequence (**Table 2-1**). Each set of 2 gRNAs was synthesized as a gBlock (Integrated DNA Technologies), interspersed with a gRNA scaffold and tRNA linker to allow for polycistronic gRNA expression as previously described⁸⁹. The insert was cloned into a modified pCambia2300 backbone (**Figure 2-2**) containing Cas9 and an Arabidopsis U6-26 promoter driving expression of candidate gRNAs. Each set of 2 gRNA was cloned into pCambia2300_Cas9 via GoldenGate Assembly for *Agrobacterium*-mediated transformation by the Ralph M. Parsons Foundation Plant Transformation Facility at UC Davis (<https://ptf.ucdavis.edu/>).

Table 2-1. *Nicotiana benthamiana* orthologs and gRNA spacer sequences used for CRISPR/Cas9-mediated knockout

Gene Name	Draft Genome Location	Position from ATG	Location	Spacer Sequences (5' -> 3')
PsbS1	Niben101Scf11852g00012.1	+331 : +351 +422 : +442	Exon 2 Exon 2	ACAGGTTGTACCAAAGCCAA; GTTGGCCGTGTTGCTATGAT
PsbS2	Niben101Scf05304g05008.1	+351 : +371 +422 : +442	Exon 2 Exon 2	
VDE1	Niben101Scf07893g00003.1	+1983 : +2003 +2088 : +2108	Exon 4 Exon 4	GGGAAATGGTTCATAACTCG; TGGAGAATACGGACACCTGA
VDE2	Niben101Scf00177g07008.1	+2503 : +2523 +2608 : +2628	Exon 5 Exon 5	
ZEP1	Niben101Scf16082g00007.1	+296 : +316 +474 : +494	Exon 1 Exon 1	GAAAAGGGGATTTGAGGTGT; GGTTTGGTTGATGGTGTTC
ZEP2	Niben101Scf01553g01001.1	+296 : +316 +474 : +494	Exon 1 Exon 1	
LUT2-1	Niben101Scf18343g00013.1	+2150 : +2170 +2298 : +2318	Exon 3 Exon 4	TAGTCGCCATTTACTGCACG; ATCTTAACGAAAGTGGAT
LUT2-2	Niben101Ctg13249g00004.1+ Niben101Ctg15093g00004.1*	+240 : +260 +389 : +409	Exon 2 Exon 3	

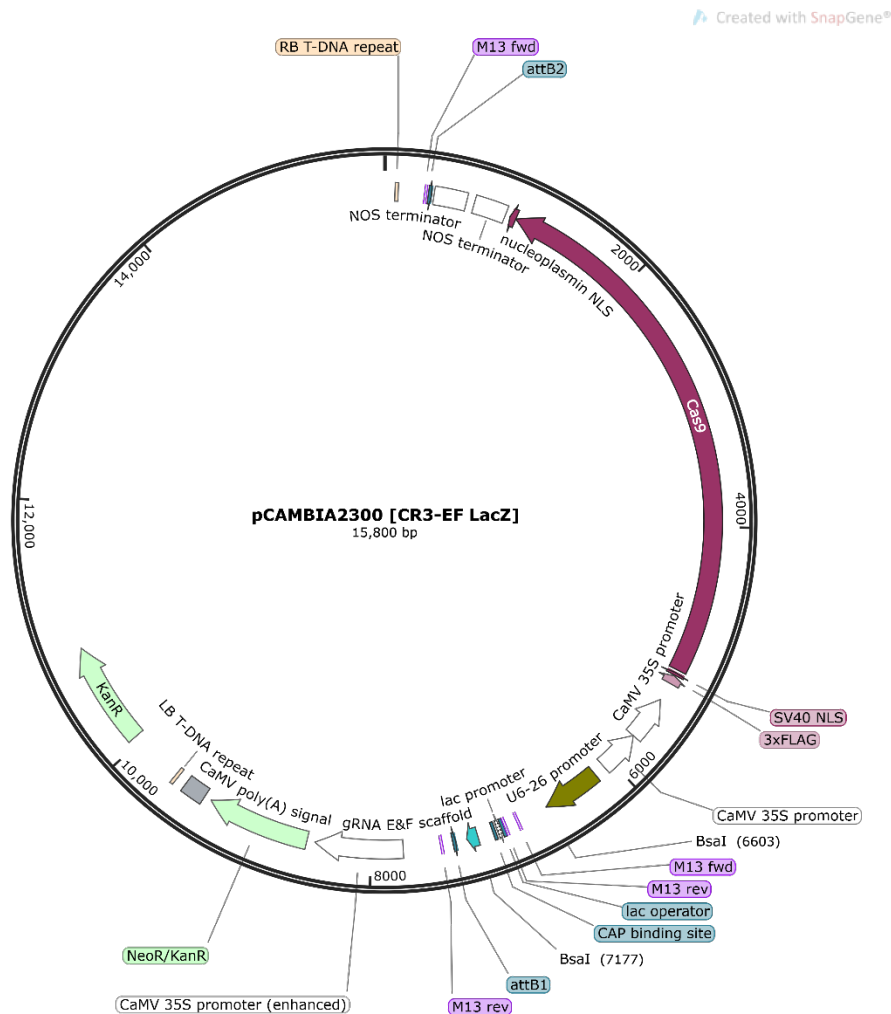


Figure 2-2. Plasmid map of an adapted pCambia2300 backbone for CRISPR/Cas9 mutagenesis of *Nicotiana benthamiana*.

An *Arabidopsis* U6-26 promoter drives expression of the polycistronic gRNA-tRNA cassette. A constitutive 35S promoter drives expression of a nuclear-localized SpCas9 codon-optimized for *Arabidopsis*. A plant selectable marker (Kanamycin^R) was included for *Agrobacterium*-mediated transformation of embryogenic tobacco calli.

2.4.3 Chlorophyll fluorescence phenotyping of NPQ

CRISPR/Cas9 knockouts of target orthologs were identified by whole plant and/or leaf punch phenotyping of chlorophyll fluorescence measurements at room temperature using an Imaging-PAM Maxi (Walz) pulse-amplitude modulation fluorometer *in vivo*. Fluorescence yield measurements in the dark (F_0 , F_m) and after actinic light exposure (F_0' , F_m') were measured using periodic saturating pulses over a 10-minute period of blue actinic light ($1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ unless otherwise noted) and 10 minutes of darkness. Plants were dark acclimated for at least 30 minutes prior to analysis.

Cross-genotype differences in NPQ were quantified using an FMS2+ (Hansatech Instruments Ltd), measuring chlorophyll fluorescence measurements under a period of 15 minutes high light, 5 minutes darkness, 5 minutes high light, 10 minutes darkness at two white light intensities: $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ and $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ light *in vivo*. In both instances, NPQ was calculated by (Equation 1).

$$\text{Equation 1: } NPQ = [(F_m - F_m')/F_m'].$$

2.4.4 Kanamycin sensitivity assay for high-throughput Cas9 transgene detection

Floated leaf punches assayed for NPQ capacity were used to determine kanamycin sensitivity and segregation of the transgene. Kanamycin (50 mg/mL milliQ H₂O) was added to each well to reach a final concentration of 50 µg/mL of antibiotic. Plates were incubated for 3-4 days at 100 µmol m⁻² s⁻¹ fluorescent light with a 14-hour day length (27°C day/25°C night temperature), following which the Imaging-PAM Maxi (Walz) was used to identify differences in the maximum efficiency of Photosystem II (F_v/F_m) after 30 minutes dark acclimation using (Equation 2) in biological duplicate.

$$\text{Equation 2: } F_v/F_m = (F_m - F_o)/(F_m)$$

Sensitive, transgene-free plants had a decline in F_v/F_m of ~0.3-0.4, whereas transgenic plants maintained a WT F_v/F_m of ~0.6-0.8 across at least two replicates per plant screened.

2.4.5 Genotyping and mutation analysis of T₀ and T₁ lines

DNA from putative knockouts was isolated and genotyped by Phire Plant Direct PCR Master Mix (ThermoScientific™, Catalog #F160L) using the supplied dilution buffer. Orthologs were amplified using primers that spanned the two gRNA target sites for each gene of interest but were specific to each of the highly homologous orthologs (**Table 2-2**). PCR products were purified and sequenced by Sanger sequencing. Segregating gene-edited mutations were identified in the T₀ population via SangerTrace analysis (ice.synthego.com)⁹⁰, and promising knockout candidates were analyzed for stable, heritable phenotypes and genotypes in the T₁ generation.

Table 2-2. Primers for genomic DNA amplification of *Nicotiana benthamiana* NPQ gene orthologs.

Oligo ID	Oligo (5' -> 3')	Fragment Size	Tm
oNb1 ZEP_1.F	CCTCATTGGTTCAGCTTC	806 bp	58
oNb2 ZEP_1.R	GACATGCAGTATTAAATGACACC		
oNb3 ZEP_2.F	GGATTCCCCGACAGAGTTG	530 bp	58
oNb4 ZEP_2.R	GCTTGTGTATGATTGTACAGTGC		
oNb5 PsbS_1.F	GGCAGGGAGGCAAATACTAAC	308 bp	58
oNb6 PsbS_1.R	TTTCGTTTACCGCCTTTC		
oNb7 PsbS_2.F	CCAAAGCTCTGCCAAAAAGG	428 bp	58
oNb8 PsbS_2.R	AGTTAAATGGAACGTCCGTGC		
oNb9 LUT2_1.F	CAACAAACAGAACCTCTTGTTTC	500 bp	58
oNb10 LUT2_2.F	GAAACATCTTGTTCTCTGGAGC	500 bp	58
oNb11 LUT2.R	CCAGATGCAACAGTGACAAA		
oNb12 VDE.F	CTAATATGCTGGAGTGATTCTGC		

oNb13 VDE_1.R	GCACACAGAGATACGGAAC	458 bp	58
oNb14 VDE_2.R	CTATCAGCAAGGTTTAATCCAGC	567 bp	58
oNb15 Cas9.F	CGAAGAGGGCATCAAAGAG	410 bp	58
oNb16 Cas9.R	GCTGTCTCTTGATGAAGCC		

2.4.6 High-performance liquid chromatography

Total leaf chlorophylls and carotenoids (neoxanthin, violaxanthin, antheraxanthin, lutein, chlorophyll *b*, zeaxanthin, chlorophyll *a*, and β -carotene) were quantified by high-performance liquid chromatography (1100 HPLC, Agilent) as described³¹ and quantified against a dilution series of standards. Briefly, flash frozen leaf tissue, either dark-acclimated or exposed to high light, was flash frozen in liquid nitrogen for quantification. Leaf tissue was ground in tubes containing Lysing Matrix D beads through a FastPrep-24 5G™ High-Speed Homogenizer (6.0 m/s 1x40s, MP Biomedical). Pigments were extracted twice in 150 μ L 100% acetone until the remaining leaf debris was white in color. Samples were extracted on ice and in the dark to minimize pigment degradation and evaporation of the solvent.

2.5 Results

2.5.1 CRISPR/Cas9 gene-editing of *Nicotiana benthamiana* NPQ orthologs

Mutant libraries of *Nicotiana benthamiana* have lagged considerably behind those of *Arabidopsis thaliana* and other model plants, despite its key role in plant immunity studies, as a chassis for transient expression, and its potential for use with transient absorption spectroscopy. Multiplexed CRISPR/Cas9 mutagenesis was employed to knockout orthologs of core *Arabidopsis* NPQ proteins. Orthologs of candidate NPQ genes (NPQ4/PSBS, NPQ1/VDE, NPQ2/ABA1/ZEP, and LUT2) were identified using the *Nicotiana benthamiana* draft genome sequence v1.0.1 ([Sol Genomics Network](#))⁹¹ and *Arabidopsis* protein sequences as queries. The dual-targeted gRNA spacer sequences used are found in (**Table 2-1**). Gene structure was largely similar across pairs of genes, excluding LUT2-2, which was manually assembled by splicing two draft contigs (**Figure 2-3A, Table 2-3**). Guide RNA (gRNA) spacers were cloned into a modified pCAMBIA2300 backbone for CRISPR/Cas9 mutagenesis (**Figure 2-2**), separated by a tRNA linker and gRNA scaffold to allow for polycistronic expression⁸⁹. Four constructs were designed targeting each pair of genes: PSBS_ko, VDE_ko, ZEP_ko, LUT2_ko (**Figure 2-3B**). Knockout phenotypes were assessed by chlorophyll fluorescence phenotyping of NPQ, and stable alleles lacking the Cas9 transgene were identified by antibiotic sensitivity (**Figure 2-3C**).

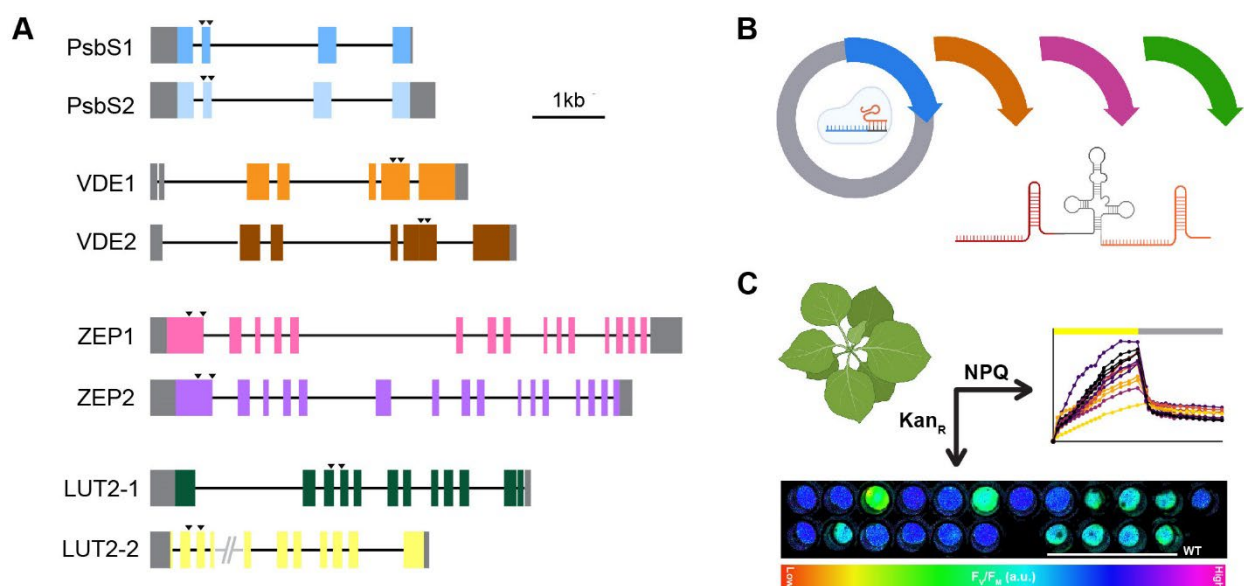


Figure 2-3. Gene models and target sites for orthologous *N. benthamiana* NPQ gene mutagenesis.

(A) Gene models assembled from the SolGenomics draft genome of Nb-1, numbered within each pair by percent homology to their respective Arabidopsis ortholog. Exons are shown by colored boxes, introns by black lines, and untranslated regions in gray. gRNA spacers are marked by black triangles. (B) Schematic for dual gRNA expression of each set of target genes using an intervening tRNA linker. (C) Screen for knockout NPQ phenotypes and sensitivity to the Cas9 selectable marker kanamycin, as shown by a decline in F_v/F_m similar to non-transgenic WT (indicated by the white line).

Table 2-3. *A. thaliana* NPQ proteins and their respective *N. benthamiana* orthologs.

<i>A. thaliana</i> mutant	At Gene Locus	<i>N. benthamiana</i> orthologs	Nb Gene Locus
npq4 / psbs	At1g44575	PsbS1	Niben101Scf11852g00012
		PsbS2	Niben101Scf05304g05008
npq1/ vde	At1g08550	VDE1	Niben101Scf07893g00003
		VDE2	Niben101Scf00177g07008
npq2 / aba1 / zep	At5g67030	ZEP1	Niben101Scf16082g00007
		ZEP2	Niben101Scf01553g01001
lut2	At5g57030	LUT2-1	Niben101Scf18343g00013
		LUT2-2	Niben101Ctg13249Ctg00004*
			Niben101Ctg15093Ctg00004*

* Indicates a manually assembled gene model based on orthology across two draft contigs

2.5.2 NbPsbS1 and NbPsbS2 both contribute to NPQ in planta

Arabidopsis mutants in PsbS lack the ability to rapidly induce and sustain WT NPQ capacity¹⁹. Knockouts of both tobacco orthologs show a similar loss in NPQ capacity. Knockout of tobacco *psbs1* alone resulted in a ~75% reduction in NPQ, suggesting that NbPsbS2 is responsible for ~25% of WT NPQ capacity at the tested light intensity (Figure 2-4). Independent isolated alleles are summarized in (Table 2-4).

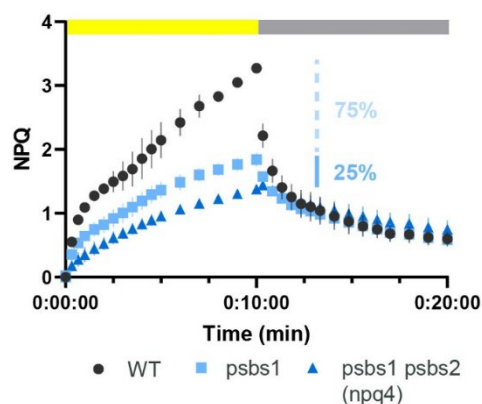


Figure 2-4. NPQ kinetics of WT, *psbs1*, and *psbs1psbs2* (*npq4*) tobacco mutants.

NPQ was induced for 10 minutes at $1000 \mu\text{mol m}^{-2} \text{s}^{-1}$ of high-light and monitored for relaxation for 10 minutes at $0 \mu\text{mol m}^{-2} \text{s}^{-1}$. Data from WT ($n=3$, black circle), *psbs1* ($n=6$, light blue square), and *psbs1psbs2/npq4* ($n=8$, dark blue triangle) are shown ± 1 SEM.

Table 2-4. *NbPsbs* knockout alleles isolated across five independent transformants.

T ₀ Parent	Fixed, Cas9- T ₁ Progeny	PSBS1	PSBS2	Allele ID	Mutant ID in this study
PSBS_ko-1	3	-8 bp	-1 bp	npq4_1-1	psbs1psbs2 (npq4)
PSBS_ko-1	63	-8 bp		psbs1	psbs1
PSBS_ko-3	20, 39, 57	-7 bp	-1 bp	npq4_3-1	
PSBS_ko-3	59	-7 bp	-2 bp	npq4_3-2	
PSBS_ko-4	2, 51, 86	-8 bp	-4 bp	npq4_4-1	
PSBS_ko-4	10	-8 bp	-2 bp	npq4_4-2	
PSBS_ko-4	82	-2 bp	-2 bp	npq4_4-3	
PSBS_ko-5	9, 38, 52, 54	-2 bp	-1 bp/-1 bp	npq4_5-1	

2.5.3 *NbVDE1* is required for de-epoxidation of violaxanthin under high light

VDE is responsible for the de-epoxidation of violaxanthin to zeaxanthin in response to high light⁴⁸. To test the contributions of each *NbVDE* ortholog, single (*vde1* or *vde2*) and double (*vde1vde2*) mutants were exposed to $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ for one hour and assayed for changes in pigment profiles by HPLC (**Figure 2-5**).

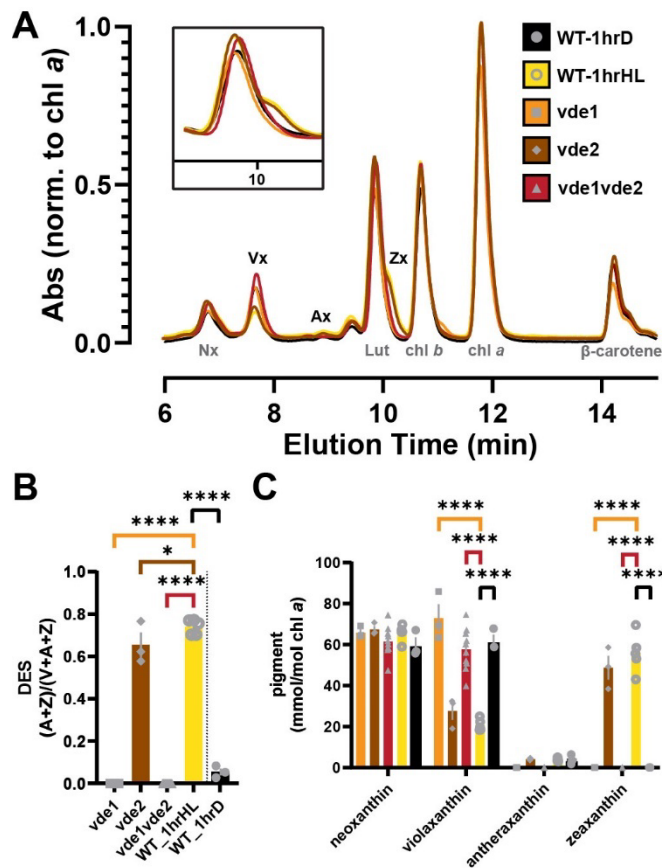


Figure 2-5. Changes in xanthophyll pigment profile of NbVDE1 and NbVDE2 mutants in response to high-light.

(A) Overlay of representative chromatograms normalized to chlorophyll *a* resolving neoxanthin (Nx), violaxanthin (Vx), antheraxanthin (Ax), lutein (Lut), zeaxanthin (Zx), chlorophyll *b* (chl *b*), chlorophyll *a* (chl *a*), and β-carotene after 1 hour at 1500 μmol m⁻² s⁻¹ unless otherwise noted. An inset figure provides increased resolution of lutein (left) and zeaxanthin (right) peaks. **(B)** De-epoxidation state (DES) and **(C)** individual xanthophyll pigment concentration normalized to chlorophyll *a*. Data for a dark-acclimated WT (*n*=3, black, circle) is provided as a baseline against 1-hour high-light acclimated WT (*n*=5, yellow, open circle), *vde1* (*n*=3, orange, square), *vde2* (*n*=3, brown, diamond), and *vde1vde2* (*n*=10, red, triangle) data shown ± 1 SEM. Pairwise significance in (B,C) was determined by ordinary one-way ANOVA (α=0.05) using Dunnett's test for multiple comparisons against WT-1hrHL with significance denoted by asterisks (***p*≤0.01, ****p*≤0.001, *****p*<0.0001).

Chromatogram overlays of representative mutants and WT samples show a robust accumulation of Zx after HL in WT and *vde2* (**Figure 2-5A**). The *vde1* and *vde1vde2* mutants lack any detectable de-epoxidized xanthophyll pigments, illustrated by a significantly lower de-epoxidation state (*p*<0.0001) (**Figure 2-5B**). Assessment of individual xanthophyll pigments substantiated this result (*p*<0.0001), with *vde1* and *vde1vde2* almost fully phenocopying the dark-acclimated WT sample—the exception being a trace amount of antheraxanthin present in WT dark samples that was absent with functional knockout of *vde* (**Figure 2-5C**). There was no significant difference in individual pigments between WT-1hrHL and *vde2* (**Figure 2-5C**) despite a weak observed reduction in de-epoxidation state (*p*=0.024) (**Figure 2-5B**). Together, these data show that NbVDE1 is the only functional ortholog in *N. benthamiana* VAZ cycling. Independent isolated alleles are summarized in (**Table 2-5**).

Table 2-5. *NbVDE knockout alleles isolated across ten independent transformants.*

T₀ Parent	Fixed, Cas9- T₁ Progeny	VDE1	VDE2	Allele ID	Mutant ID in this study
VDE_ko-2	3, 14, 37, 39, 42	-11 bp	-1 bp	npq1_2-1	vde1vde2 (npq1)
VDE_ko-2	23, 24	+1 bp	+1 bp	npq1_2-2	
VDE_ko-3	4, 37, 181	+1 bp		vde1	vde1
VDE_ko-3	123, 124, 146		-1 bp	vde2	vde2
VDE_ko-3	27, 33, 96, 156			WT	
VDE_ko-7	19, 25, 36, 43	-1 bp	+1 bp	npq1_7-1	

2.5.4 Single ortholog NbZEP knockouts decouple zeaxanthin accumulation and stomatal misregulation phenotypes

ZEP is necessary for the production of antheraxanthin, violaxanthin, and neoxanthin⁴⁸, the latter being an important precursor of the stomata-regulating plant hormone abscisic acid⁹². Correspondingly, knockouts of ZEP should constitutively accumulate zeaxanthin even in the dark. To assess the consistency of this phenotype, HPLC on dark-acclimated *zep1*, *zep2*, and *zep1zep2* mutants was performed (**Figure 2-6**).

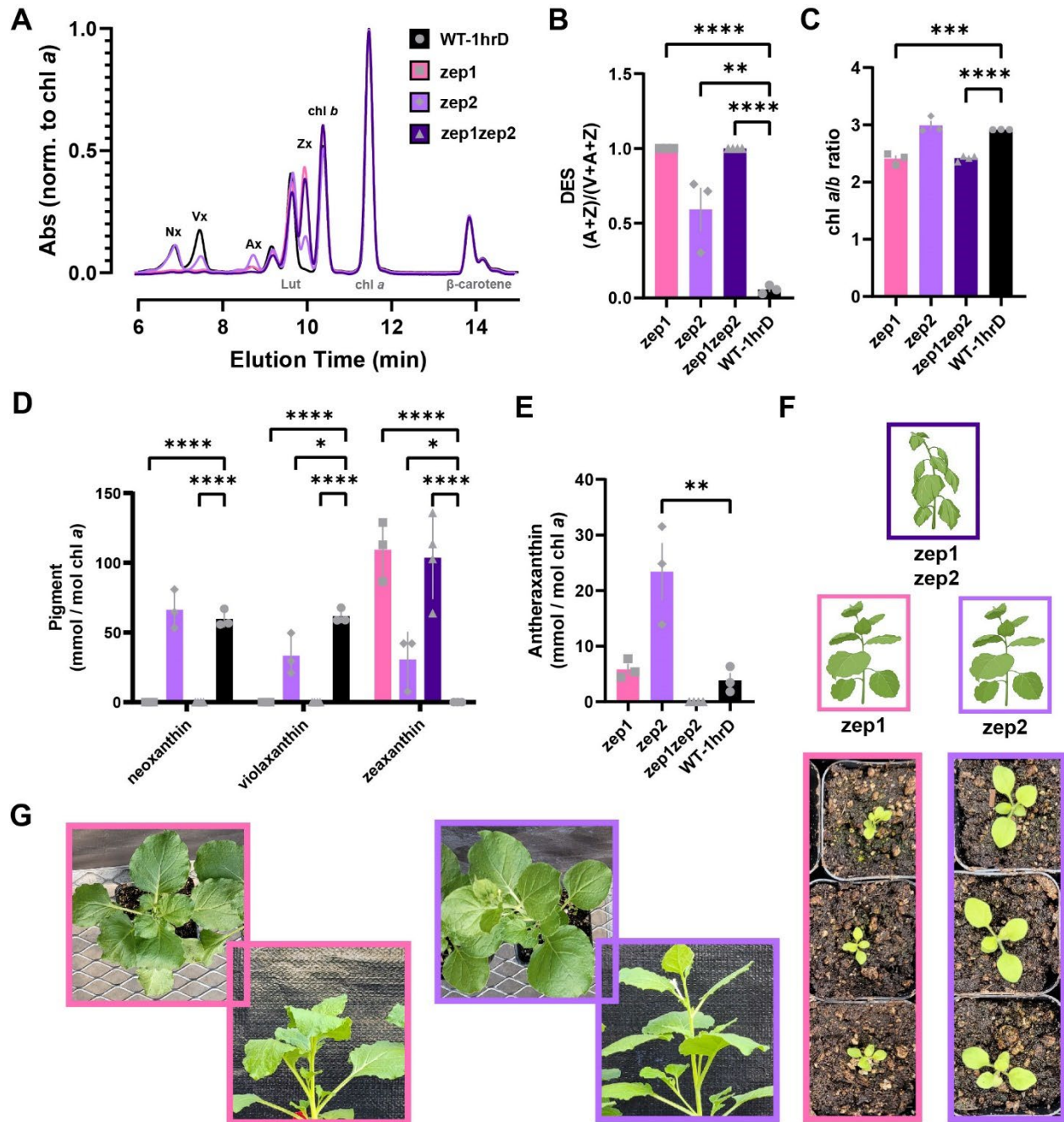


Figure 2-6. Xanthophyll pigment profile & growth phenotypes of dark-acclimated *zep1*, *zep2*, and *zep1zep2* mutants.

(A) Overlay of representative chromatograms normalized to chlorophyll *a* resolving neoxanthin (Nx), violaxanthin (Vx), antheraxanthin (Ax), lutein (Lut), zeaxanthin (Zx), chlorophyll *b* (chl *b*), chlorophyll *a* (chl *a*), and β -carotene after 1-hour at $0 \mu\text{mol m}^{-2} \text{s}^{-1}$. (B) De-epoxidation state (DES) (C) Chl *a/b* ratio, and (D,E) individual xanthophyll pigment concentrations normalized to chlorophyll *a*. Anthraxanthin is plotted on a separate y-axis to increase contrast between genotypes. (F,G) Cartoon and representative two-week and six-week old plants of *zep1* and *zep2*. Data for dark-acclimated WT ($n=3$, black, circle) is provided as a baseline against *zep1* ($n=3$, pink, square), *zep2* ($n=3$, lavender, diamond), and *zep1zep2* ($n=4$, purple, triangle), data shown ± 1 SEM. Cartoon icons were taken from Biorender. Pairwise significance in was determined by ordinary one-way ANOVA (B, C, E; $\alpha=0.05$) or two-way ANOVA (D, $\alpha=0.05$)

using Dunnett's test for multiple comparisons against WT-1hrD with significance denoted by asterisks (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p < 0.0001$).

Chromatogram overlays highlight dramatic differences in steady-state xanthophyll pigments (**Figure 2-6A**). The *zep1* and *zep1zep2* mutants maintain an entirely de-epoxidized xanthophyll pool ($p < 0.0001$), with the *zep2* mutant maintaining an intermediate DES relative to WT ($p = 0.001$) (**Figure 2-6B**). The *zep1* mutant also phenocopies *zep1zep2* in its reduced chlorophyll *a/b* ratio (**Figure 2-6C**) and saturated zeaxanthin pool size, at the expense of neoxanthin and violaxanthin (**Figure 2-6D**). The two genotypes differ only in the presence (*zep1*) or absence (*zep1zep2*) of antheraxanthin ($p = 0.264$ one-way ANOVA, $p = 0.009$ unpaired t-test) (**Figure 2-6E**).

In contrast, the *zep2* mutant has a WT-like chlorophyll *a/b* ratio and neoxanthin pool size (**Figure 2-6C,D**). Consistent with differences in DES, the *zep2* mutant shows modest reductions in violaxanthin ($p = 0.05$) and increases in zeaxanthin ($p = 0.032$) and antheraxanthin ($p = 0.0012$) (**Figure 2-6D,E**).

Mutations in single and double mutants also corresponded to significant changes in plant fitness and overall physiology. The *zep1zep2* mutant was only able to grow under the high-humidity conditions used for germination, and was unable to maintain turgor pressure and grow past the seedling stage necessary for seed set (**Figure 2-6F, Table 2-6**). In contrast, both *zep1* and *zep2* mutants were able to grow under ambient greenhouse conditions with more modest differences in growth. While the *zep2* mutant was indistinguishable from WT plants, the *zep1* mutant exhibited somewhat stunted growth, frilled leaf edges, and reduced leaf turgor (**Figure 2-6F,G**). Given the near lethality of the *zep1zep2* double knockout, only one of eleven T₀ transformants produced viable, stable progeny (**Table 2-6**).

Table 2-6. Viable *NbZEP* knockout alleles isolated across eleven independent transformants.

T ₀ Parent	Fixed, Cas9- T ₁ Progeny	ZEP1	ZEP2	Allele ID	Mutant ID in this study
ZEP_ko-4	38, 44, 58	+1 bp	-6 bp	zep1 ZEP2*	zep1
ZEP_ko-4	18, 39, 63	-3 bp	-8 bp	ZEP1* zep2	zep2
ZEP_ko-4	46 [–] , 68 [–] , 72 [–] , 88 [–]	+1 bp	-8 bp	zep1zep2	zep1zep2

* Indicates an in-frame mutation that may maintain protein structure and function

[–] Indicates progeny that did not survive to seed set

2.5.5 *NbLUT2* knockouts lack lutein and retain triple the VAZ pool size

LUT2 acts at the γ -carotene branch point, producing the precursor α -carotene necessary for lutein production⁸¹. Consistent with what was reported in Arabidopsis, tobacco *lut2-1/lut2-2* double mutants lacked lutein (**Figure 2-7A,B**) and retained triple the VAZ pool

size (**Figure 2-7C**). The knockout phenotype was phenocopied in a *LUT2-1*lut2-2* mutant containing an in-frame deletion in the first ortholog.

Initial observation of pigment profile differences between overnight dark-acclimated WT and *lut2-1lut2-2* mutants showed possible presence of a zeaxanthin peak partially obscured by the lutein peak shoulder (**Figure 2-7A**). Increased time resolution confirmed the zeaxanthin peak was true (**Figure 2-7B**), contributing in part to the increased DES observed even after overnight dark acclimation, despite no detectable antheraxanthin or zeaxanthin after overnight WT dark acclimation ($p < 0.0001$) (**Figure 2-7D**). Independent isolated alleles are summarized in (**Table 2-7**).

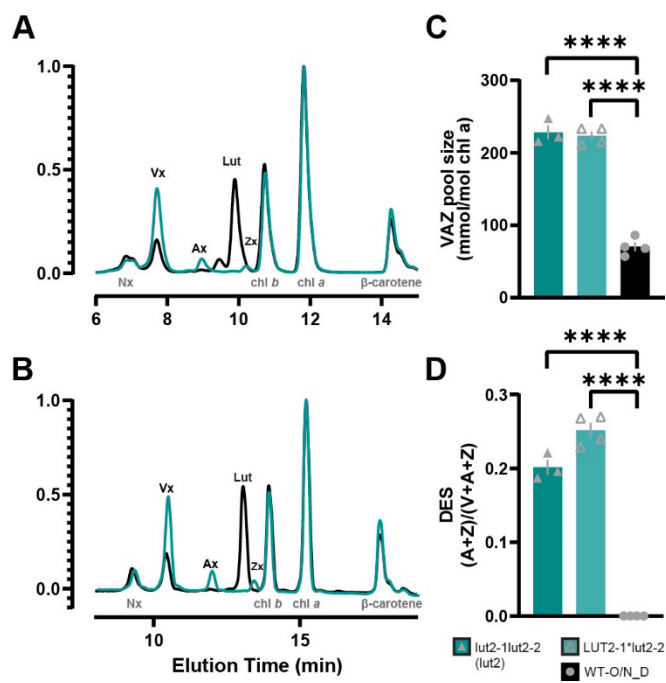


Figure 2-7. Carotenoid profile of *N. benthamiana* mutants lacking *LUT2-1* and *LUT2-2*.

Overlay of representative chromatograms normalized to chlorophyll *a* resolving neoxanthin (Nx), violaxanthin (Vx), antheraxanthin (Ax), lutein (Lut), zeaxanthin (Zx), chlorophyll *b* (chl *b*), chlorophyll *a* (chl *a*), and β -carotene after an overnight dark acclimation with a (**A**) 20-minute and (**B**) 30-minute total run time. (**C**) Vx + Ax + Zx (VAZ) pool size and (**D**) De-epoxidation state (DES) normalized to chlorophyll *a* analyzed from samples run with the 30-minute run-time (B). Data for overnight dark-acclimated WT ($n=4$, black, circle) is provided as a baseline against *lut2-1lut2-2* (*lut2*) ($n=3$, teal, filled triangle) and *LUT2-1*lut2-2* ($n=4$, blue-green, open triangle), data shown ± 1 SEM. Pairwise significance in (C,D) was determined by ordinary one-way ANOVA ($\alpha=0.05$) using Dunnett's test for multiple comparisons against WT-O/N_D with significance denoted by asterisks (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p < 0.0001$).

Table 2-7. *NbLUT2* knockout alleles isolated across eleven independent transformants.

To Parent	Fixed, Cas9- T ₁ Progeny	LUT2-1	LUT2-2	Allele ID	Mutant ID in this study
LUT2_ko-4	1, 6, 25, 35	-3 bp	+1 bp	LUT2-1* lut2-2	LUT2-1* lut2-2
LUT2_ko-6	51, 65, 69	-1 bp	+1 bp/+1 bp	lut2_6-1	lut2-1lut2-2 (lut2_6-1)
LUT2_ko-7	20	+1 bp	+ 1bp	lut2_7-1	lut2-1lut2-2 (lut2_7-1)

* Indicates an in-frame mutation that may maintain protein structure and function

2.5.6 Genotypic comparisons resolve light and time-dependent contributors to the *N. benthamiana* NPQ response

Screening for T₁ knockouts confirmed loss-of-function phenotypes observed in orthologous *Arabidopsis* NPQ mutants and identified dominant orthologs in carotenoid production and qE activity. To test the contributions of the altered pigment profiles on

NPQ, T₂ progeny of homozygous, loss-of-function parents were grown and assessed for their NPQ capacity under fluctuating light. Additionally, given the observed differences in DES of WT with 1-hour (**Figure 2-5, Figure 2-6**) and overnight (**Figure 2-7**) dark acclimation, extra care was taken to also assess the contributions of WT DES on NPQ.

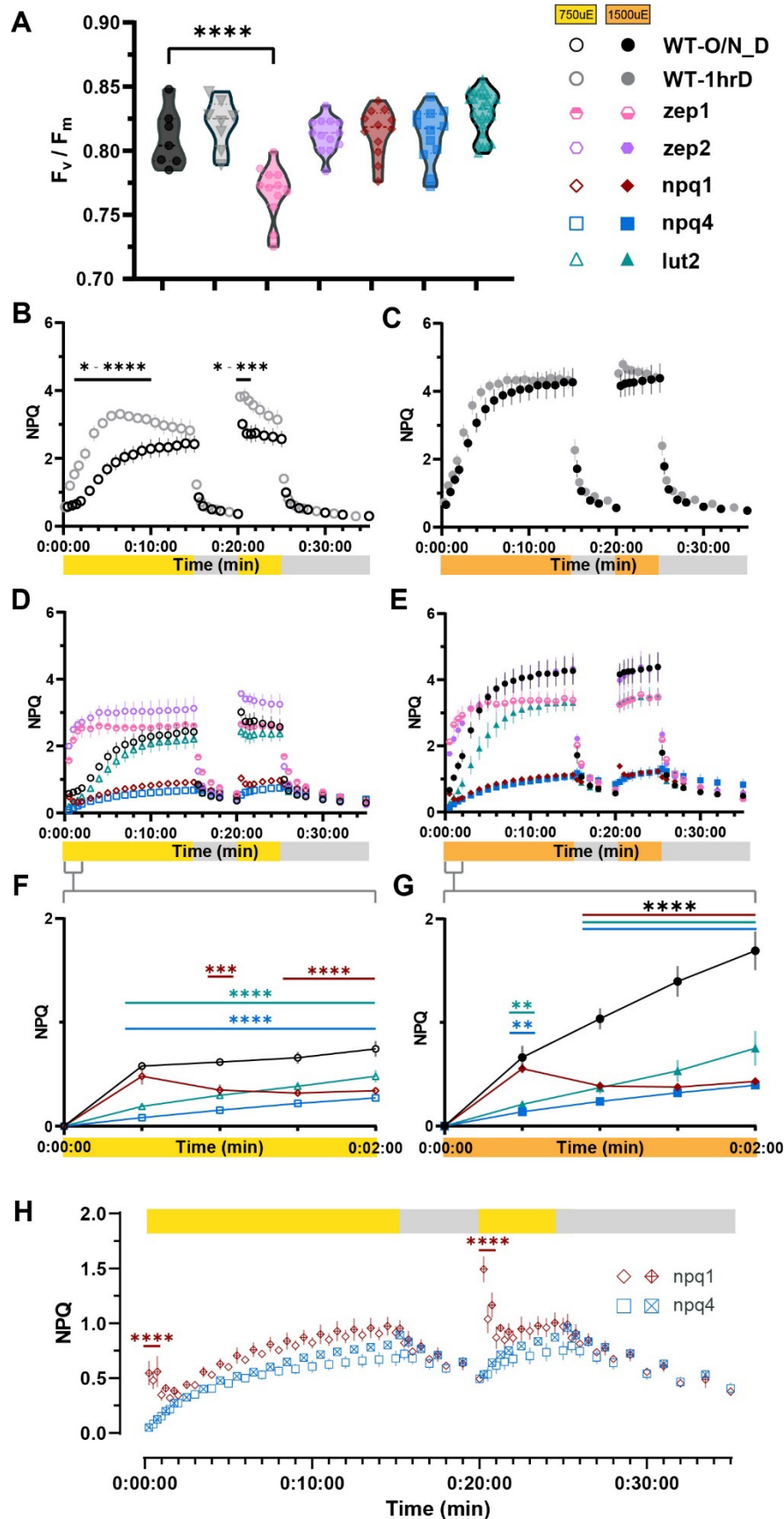


Figure 2-8. NPQ kinetics of tobacco mutants under moderate ($750 \mu\text{mol m}^{-2} \text{s}^{-1}$) and high ($1500 \mu\text{mol m}^{-2} \text{s}^{-1}$) actinic light.

(A) Truncated violin plot of F_v/F_m for all assayed plants (n=12-25 replicates). **(B)** NPQ kinetics of WT dark-acclimated for 1 hour (gray) and overnight (black) at $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ (yellow bar, open circle) and **(C)** $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ (orange bar, closed circle) actinic light. **(D)** NPQ kinetics of genotypes listed in panel (A) at $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ (yellow bar) and **(E)** $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ (orange bar) actinic light after overnight dark acclimation. **(F,G)** NPQ in the first two minutes of (D,E) excluding *zep1* and *zep2*. **(H)** NPQ kinetics of *npq1* and *npq4* at two measurement frequencies. For NPQ graphs, n=3-4 replicates each. Markers described as follows: *zep1* (pink half-hex), *zep2* (purple hexagon), *npq1* (red diamond), *npq4* (blue square), *lut2* (blue-green, triangle). Open markers are samples assayed at $750 \mu\text{mol m}^{-2} \text{s}^{-1}$. Closed markers are samples assayed at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$. Pairwise significance was determined by ordinary one-way (A, $\alpha=0.05$) or two-way ANOVA (B,C,F-H, $\alpha=0.05$) using Dunnett's test for multiple comparisons against WT-O/N_D with significance denoted (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p < 0.0001$).

Despite similar F_v/F_m between WT dark-acclimation conditions (**Figure 2-8A**), the kinetics of NPQ induction significantly differed under $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($0.024 < p < 0.0001$) (**Figure 2-8B**) but not $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ (**Figure 2-8C**) of actinic light. After overnight dark acclimation (WT-O/N_D) NPQ induction at $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ was slow to rise and convex in shape. In contrast, 1-hour dark acclimation (WT-1hrD) or assessment of NPQ at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ resulted in a concave induction curve reminiscent of phenotypes observed in *Arabidopsis*. To reduce variation in phenotype due to starting DES, all genotypes and controls were dark acclimated overnight for the remaining NPQ measurements.

Of the remaining genotypes only *zep1*, a mutant with constitutive zeaxanthin at the expense of violaxanthin and neoxanthin (**Figure 2-6**), showed a significant decline in quantum efficiency of PSII ($p < 0.0001$) (**Figure 2-8A**). This may partially explain the lower NPQ observed relative to *zep2*, though both showed the rapid early rise in NPQ indicative of a saturated zeaxanthin pool (**Figure 2-8D,E**). Notably, the difference in phenotype between *zep1*, *zep2*, and WT was smaller at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$, with WT phenocopying *zep2* at all timepoints after 4 minutes.

Interestingly, the *lut2* mutant was indistinguishable from WT at $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ but showed modest reductions in total NPQ at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($p > 0.009$) (**Figure 2-8D,E**). This contrasts with the *npq1* mutant which was almost indistinguishable from the *psbs* knockout *npq4* at both light intensities. Closer examination in 1) the first two minutes of NPQ induction and 2) at increased measurement frequencies revealed subtle differences in quenching.

At both light intensities, *lut2* and *npq4* mutants had significantly lower NPQ in the first two-minutes of actinic light ($0.001 < p < 0.0001$). Interestingly, the *npq1* mutant was indistinguishable from WT in the first 30 seconds of actinic light before declining to *npq4*-like levels of quenching for subsequent timepoints ($0.0003 < p < 0.0001$) (**Figure 2-8F,G**). Phenotyping of an independent cohort of plants at a staggered measurement frequency reveal that the NPQ capacity of *npq1* significantly differed from *npq4* only in the first 45 seconds of each $750 \mu\text{mol m}^{-2} \text{s}^{-1}$ actinic light window ($p < 0.0001$) (**Figure 2-8F**). Also noteworthy, the induction of NPQ in the first 45 seconds showed a higher magnitude of change in the second period of high-light (NPQ increase of ~ 1) compared to the first incidence of high-light (NPQ increase of ~ 0.5). While large macro-scale differences and similarities distinguished dominant NPQ phenotypes, adjusting light intensity and measuring frequency revealed $< 1\text{min}$ differences in NPQ induction kinetics across genotypes.

2.6 Discussion

Our understanding of the underlying mechanisms of NPQ has been limited to few plant species and even narrower genetic diversity. In this work, I broaden our cross-species understanding of NPQ through CRISPR/Cas9 knockouts of NPQ genes in the model plant *Nicotiana benthamiana*. The resulting germplasm has great potential to advance our understanding of the ultrafast kinetics driving NPQ, and the observed NPQ phenotypes prompt new discussion points in understanding quenching *in vivo*.

Unsurprisingly, knockout of these highly conserved orthologs behaved more or less as expected. Mutations in *NbPsbS1/2* (*npq4*) abolished a vast majority of NPQ capacity. However, the presence of a stable “knockdown” *psbs1* only mutant provides an interesting opportunity to test PsbS dosage against CT and EET signatures. Knockouts of *NbVDE* (*npq1*), *NbZEP* (*npq2*), and *NbLUT2* (*lut2*) showed consistent differences in pigment profile relative to published Arabidopsis mutants. In contrast to *NbVDE* and *NbLUT2*, which appear to only have a single functional copy of each trait (*NbVDE1* and *NbLUT2-1*), we see a similar dual dependence of orthologs in *NbZEP1/2* as *NbPsbS1/2*. This unexpected windfall provided two viable alleles: *Nbzep1*, a mutant line with an entirely de-epoxidized xanthophyll pool lacking violaxanthin and neoxanthin entirely, and *Nbzep2*, a weaker mutant allele physiologically indistinguishable from WT except in its NPQ capacity. These results laid the basis for specific, highly controlled assessment of NPQ across the four target genes.

First, I sought to reconcile differences in NPQ due to starting de-epoxidation state after 1-hour (no detectable Zx, minimal Ax) and overnight (no Zx or Ax) dark acclimation. Unexpectedly, these modest differences in DES corresponded to significant differences in the induction of NPQ under moderate ($750 \mu\text{mol m}^{-2} \text{s}^{-1}$) light. A DES of ~5% was sufficient to drive a rapid rise in NPQ, whereas a DES of 0% lead to a functional plateau in NPQ for the first 2-minutes of light exposure (**Figure 2-8F**). This result is also interesting in the context of available spinach CT⁷⁷ data, which found that a similarly small de-epoxidized pool was sufficient to drive CT in WT and +DTT thylakoids, the latter intended to inhibit Zx accumulation. The presence of rapidly induced NPQ under both DES at high ($1500 \mu\text{mol m}^{-2} \text{s}^{-1}$) light may correlate with an increase in VDE activity that rapidly compensates for the absence of starting Ax. This compensation of VDE at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ is well corroborated by the tobacco *zep* mutant phenotypes which showed significantly higher NPQ in the 2nd light phase at moderate but not high light. In fact at high light, WT rapidly converged in NPQ capacity with the *zep2* knockdown line, showcasing a robust capacity for Zx-dependent NPQ in this species.

Perhaps the most significant deviations from Arabidopsis-like NPQ arose from studying the tobacco *vde* and *lut2* mutants. Knockout of *npq1* abolished almost all NPQ capacity and led to *npq4*-like NPQ at steady state. This qE-less phenotype, observed in Arabidopsis only in the *lut2npq1* double mutant⁵², suggests that Zx is the primary carotenoid involved in longer-term ¹Chl* quenching in tobacco. Consistent with this phenotype, knockout of tobacco *lut2* did not compromise NPQ at moderate light and only mildly ($0.03 < p < 0.05$) reduced NPQ capacity at high light.

Additionally, unique cross-genotype differences were observed early in light exposure. The *lut2* mutant had NPQ that was indistinguishable from *npq4* for the first minute of both moderate and high light. This was in contrast to *npq1*, which was statistically indistinguishable from WT in the first 30 seconds of light exposure before dropping down to *npq4*-like levels. From this converse relationship, it is enticing to think about a physiological mechanism where lutein contributes transiently to NPQ before being structurally displaced (e.g. by a conformational change in protein) and/or functionally replaced (e.g. by zeaxanthin at quenching sites). The rapid rise in NPQ of *npq1* at 15 seconds, nearly doubling in magnitude between the first and second actinic light periods, may suggest a role in which PsbS responds to increases in ΔpH to trigger but is then unable to sustain quenching (e.g. due to the absence of an appropriate quencher). While we are still optimizing our ability to resolve CT and EET in tobacco thylakoids, detailed dissection of temporal and light intensity dependent factors on NPQ reveals interesting physiological differences to investigate.

Chapter 3 : Multiplexed CRISPR/Cas9 mutagenesis of rice *PsbS1* non-coding sequences generates transgene-free overexpression alleles

3.1 Acknowledgements

Michelle Tjahjadi, under the guidance of Myeong-Je Cho, generated the rice transformants at the Innovative Genomics Institute Plant Transformation Facility. Armen Kelikian managed our expansive rice mutant library, screened thousands of NPQ and physiological phenotypes, and led our greenhouse yield trials. Anna Leipertz was responsible for RNA and protein extraction of our representative expression lines, and cis-regulatory variant sequencing. The *Sorghum bicolor* PsbS antibody used for OsPsbS1 immunoblots was generously provided by Steven J. Burgess (University of Illinois). Nina Maryn analyzed the Pacbio long-read whole genome sequencing data, visualized identified structural variants, and contributed relevant information to “Materials and Methods”.

3.2 Abstract

Understanding CRISPR/Cas9’s capacity to generate native overexpression (OX) alleles would accelerate agronomic gains achievable by gene editing. To generate OX alleles with increased RNA and protein abundance, we used multiplexed CRISPR/Cas9 mutagenesis of conserved, non-coding DNA sequences located upstream of the rice *PsbS1* gene. I describe a high-throughput phenotyping pipeline that resolved 120 transgene-free, gene-edited alleles with varying NPQ capacity *in vivo*—ranging from complete knockout to overexpression. Overexpression of OsPsbS1 increased protein abundance 2-4-fold—similar to fold changes seen by transgenic overexpression. Increased PsbS abundance enhanced NPQ capacity, tolerance to photoinhibition, and improved water-use efficiency under red light, though the latter’s dependence on Q_A redox state is unclear. Across our resolved genetic variation, we identify the role of 5’UTR indels and inversions in driving knockout/knockdown and overexpression phenotypes, respectively. Our results may inform future gene-editing strategies for overexpression and have opened the door to the pursuit of gene-edited, non-transgenic VPZ rice plants.

3.3 Introduction

Optimizing photosynthetic efficiency remains one of the largest gaps in engineering sustainable and productive crop varieties^{93,94}. Excitingly, several recent breakthroughs have showcased the ability to bring informed photosynthetic design principles to

improve agronomic traits in the field. A photorespiratory bypass, made possible in part by cross-phyla overexpression of green algal glycolate dehydrogenase, increased field-grown biomass of *Nicotiana tabacum* by 20%⁹⁵, and enhanced relative fitness over WT when grown under elevated temperatures⁹⁶. Overexpression of putative photosynthetic transcriptional regulators OsDREB1C⁹⁷ and zmm28⁹⁸ have also shown significant agronomic yield gains in model rice and elite maize varieties respectively, though the causal effects on photosynthesis are still unclear. Noteworthy to this work, transgenic overexpression of three highly conserved NPQ genes from *Arabidopsis thaliana* (*AtVDE*, *AtPsbS*, and *AtZEP*, hereafter VPZ) increased relaxation of but did not compromise NPQ under fluctuating light conditions, resulting in increases in field-grown tobacco biomass by 15%³⁸ and field-grown soybean yield up to 20%³⁹. The prospects for rationally designing photosynthesis in agronomic environments are bright.

However, these approaches relied on expression of foreign DNA, or transgenes, which incur regulatory complications in their use and can be susceptible to gene silencing across generations⁹⁹. NPQ genes are found in all plants, and NPQ proteins are highly conserved in their function. We reasoned that altering endogenous gene expression could achieve similar agronomic gains without the need for stable transgenes. As a proof-of-concept we targeted the rice Photosystem II Subunit S (*OsPsbS1*) gene, a core factor in high-light and fluctuating-light tolerance, to generate mutants with increased *OsPsbS1* expression and improved NPQ capacity.

Quantitative variation in rice *OsPsbS1* has already been observed between *japonica* (high NPQ) and *indica* (low NPQ) subspecies¹⁰⁰, potentially highlighting NPQ as a trait that has recently undergone selection. Transgenic overexpression of PsbS in rice has been shown to increase radiation use efficiency in greenhouse-grown plants¹⁰¹, and overexpression in tobacco has been implicated in increasing intrinsic water use efficiency (iWUE) under red light¹⁰². As a diploid model C₃ monocot with single-copy orthologs of VDE, ZEP, and a single, functional PsbS ortholog *OsPsbS1*¹⁰³, and as a food crop that accounts for over 20% of all global calories consumed, rice is an impactful and genetically tractable target for improvement.

CRISPR/Cas9 has dramatically expanded our capacity to produce targeted loss-of-function mutations. Recently, CRISPR/Cas9 editing of cis-regulatory elements has also demonstrated the utility of this approach to decrease rather than abolish gene expression through partial loss-of-function alleles in tomato¹⁰⁴ and maize¹⁰⁵. Similar multiplexed mutagenesis has shown utility in identifying and editing putative negative regulators in optimizing rice architecture¹⁰⁶. However, our ability to increase gene expression without the use of persistent transgenes is still lacking, and finer genome

engineering by prime editing is bottlenecked by low editing, transformation, and regeneration efficiencies¹⁰⁷. To circumvent these limitations, we describe a high-throughput pipeline for screening novel NPQ alleles generated by CRISPR/Cas9 non-coding sequence mutagenesis upstream of *OsPsbS1*. Our findings reveal the ability to generate a large phenotypic range of activity, including overexpression, with significant effects on NPQ and iWUE. Finally, we identify themes in cis-regulation across our library that may inform gene editing for overexpression within other traits.

3.4 Materials and Methods

3.4.1 Plant material and growth conditions

Rice cultivar Nipponbare (*Oryza sativa sp. japonica*) seeds were germinated on Whatman filter paper for five days at 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$ fluorescent light with a 14-hour day length (27°C day/25°C night temperature). Seedlings were transferred to soil composed of equal parts Turface and Sunshine Mix #4 (Sungro) and grown under seasonal day length (10-14 hours) in a south-facing greenhouse that fluctuated in temperature (38°C High/16°C Low) and relative humidity (45-60%). Plants were fertilized with a 0.1% Sprint 330 iron supplement after transplanting at 2 weeks post-germination and at the onset of grain filling at 10 weeks post germination, and JR Peter's Blue 20-20-20 fertilizer monthly. Flats were kept full of water to mimic flooded growth conditions. Genotypes were randomized across flats and throughout the greenhouse to minimize positional effects. At the V4-5 leaf stage, T₁ progeny and WT controls were assayed for differences in NPQ capacity and sensitivity to the selectable marker hygromycin.

To assess differences in photosynthetic efficiency and yield, gene-edited T₂ plants were grown in a larger greenhouse with more homogenous light exposure and field-relevant growth conditions (40°C High/27°C Low, 30-50% relative humidity) without supplemental light. Plants were grown and yield data were collected concurrent to the California rice growing season (April – October 2022), after which the plants were dried down for 2 weeks and above-ground dry biomass, yield, 1000 seed weight was quantified by destructive harvest.

3.4.2 Guide RNA design and cloning into Cas9 vector pRGEB32

Eight gRNA target sites were identified upstream of the functional *PsbS* ortholog in rice, *PsbS1* (LOC_Os01g64960), using CRISPR-P (crispr.hzau.edu.cn)⁸⁸ and a 1.5-kb region upstream of the *PsbS1* start codon in a draft genome of *Oryza sativa sp. indica* cultivar IR64 (<http://schatzlab.cshl.edu/data/rice/>)¹⁰⁸ (**Table 3-1**). The eight gRNA spacers were assembled into a DNA cassette interspersed with scaffolds and tRNA linkers for

polycistronic gRNA expression as previously described⁸⁹, and synthesized (Genscript). The insert was cloned into the pRGE32 rice *Agrobacterium*-mediated transformation vector (Addgene Plasmid #63142) via GoldenGate Assembly (**Figure 3-1**).

Table 3-1. *OsPsbS1* non-coding sequence gRNA spacer sequences and positions relative to the start codon.

Insert	Orientation (rel. to ORF)	Position from ATG in <i>O. sativa ssp. indica</i> cultivar IR64	Position from ATG in <i>O. sativa ssp. japonica</i> cultivar Nipponbare	Spacer Sequence (5' -> 3')
gRNA1	R	-1163 : -1183	-3837 : -3857	GCGAGACACTAAAATACATT
gRNA2	F	-973 : -953	-3647 : -3627	TCTTGTCCTGGATGTAATT
gRNA3	R	-888 : -908	-3562 : -3582	AGATTCAGGAGTAACAAAAA
gRNA4	F	-769 : -749	-3443 : -3423	TGTTGCATGTGGTCCGTCGA
gRNA5	R	-585 : -605	-3259 : -3279	CACAAAAAAGTACGGGAATG
gRNA6	F	-374 : -394	-374 : -394	TACCAACCACCTGCTCTTCT
gRNA7	R	-263 : -283	-263 : -283	GTCTCCCCGAATCCCTTCTA
gRNA8	F	-166 : -146	-166 : -146	CTACGCCTCCCACCCGCCAC
gRNA scaffold	GTTT TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTGAAAAAGTGGCACCGAGTCGG TGC			
tRNA linker	AACAAAGCACCAAGTGGTCTAGTGGTAGAATAGTACCCTGCCACGGTACAGACCCGGGTTGATTCCCGGCTGG TGCA			

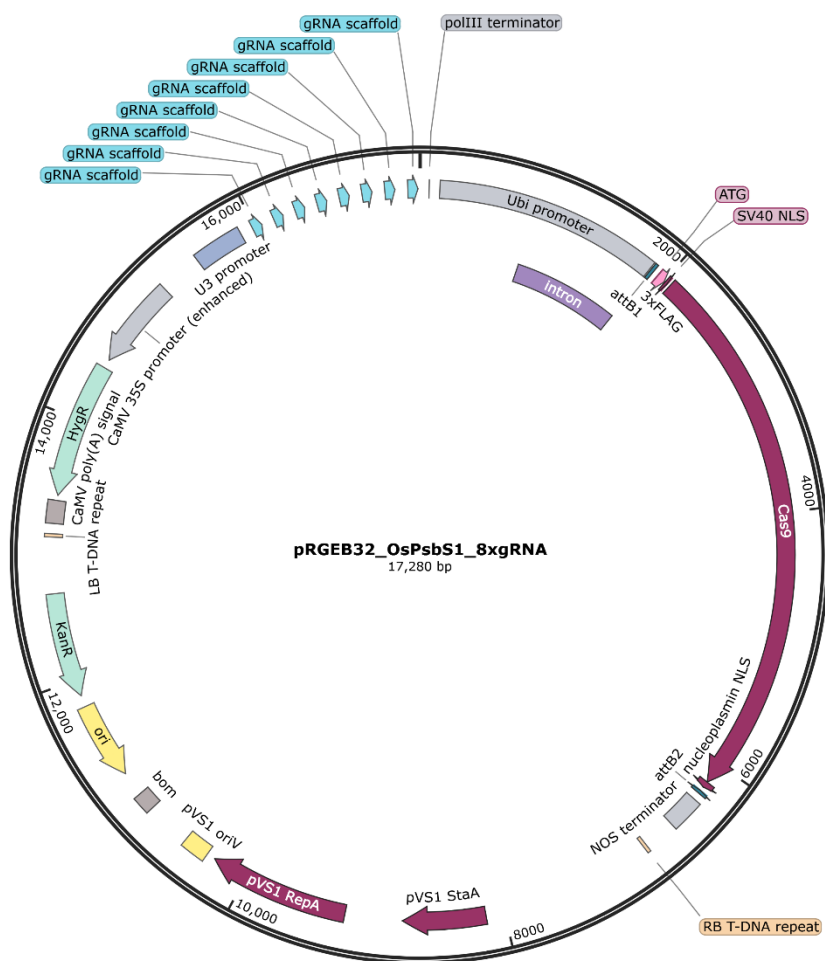


Figure 3-1. Plasmid map of pRGEb32 after GoldenGate assembly of the synthesized 8 gRNA spacer insert.

An *OsU3* promoter drives expression of the polycistronic gRNA-tRNA cassette with scaffolds (blue arrows). A *ZmUbi* promoter drives expression of a dual nuclear-localized SpCas9 codon-optimized for rice. A hygromycin resistance gene (HygR, green) was used for *Agrobacterium*-mediated transformation of embryogenic rice calli.

3.4.3 In vitro gRNA validation of upstream *OsPsbS1* non-coding sequences

Each candidate gRNA spacer was synthesized as a ssDNA oligomer with additional 5' and 3' overhangs necessary for T7 RNA Polymerase transcription and binding of the gRNA scaffold primer respectively (Table 3-2). Oligomers were synthesized into dsDNA via Phusion™ High-Fidelity PCR (NEB) and transcribed into RNA using the HiScribe RNA synthesis kit (NEB). Residual dsDNA was digested via a DNaseI treatment, and RNA was purified using the RNeasy mini-kit (Qiagen). Cas9 was individually complexed with each gRNA spacer in a buffer of 2 mM HEPES, 10 mM NaCl, 0.5 mM MgCl₂, 10 μM EDTA, pH 6.5) for 20 minutes at 37°C before co-incubating the complexed ribonucleoprotein with 100 ng of the PCR-amplified, 2-kb region upstream of *OsPsbS1* (*ssp. indica*) overnight at 37°C to verify activity of all eight guides (Table 3-2).

Table 3-2. Oligomers for *in vitro* Cas9-gRNA editing of *O. sativa* (*ssp. indica*) *OsPsbS1* non-coding sequences

5' overhang	TAATACGACTCACTATAGGG
3' overhang	GTTTAAGAGCTATGCTGGAA
gRNA scaffolding primer	AAAAGCACC GACTCGGTGCCACTTTTCAAGTTGATAACGGACTAGCCTTATTTAACTTGCTATGCTGTTCCAGCATAGCTCTTAAAC
EXAMPLE: OsPsbS1 gRNA1	TAATACGACTCACTATAGGG-[GCGAGACACTAAAATACATT]-GTTTAAGAGCTATGCTGGAA
oDP211 OsPsbS1 NCS_F	CTACCTATGCAACATGTGACCC
oDP214 OsPsbS1 NCS_R	AGACAGAGGTATGTCAATGTGTTATTG

3.4.4 Induction of embryogenic calli

Mature seeds of rice (*Oryza sativa ssp. japonica* cv. Nipponbare) were de-hulled, and surface-sterilized for 20 min in 20% (v/v) commercial bleach (5.25% sodium hypochlorite) plus a drop of Tween 20. Three washes in sterile water were used to remove residual bleach from seeds. De-hulled seeds were placed on callus induction medium (CIM) medium (N6 salts and vitamins¹⁰⁹, 30 g/L maltose, 0.1 g/L myo-inositol, 0.3 g/L casein enzymatic hydrolysate, 0.5 g/L L-proline, 0.5 g/L L-glutamine, 2.5 mg/L 2,4-D, 0.2 mg/L BAP, 5 mM CuSO₄, 3.5 g/L Phytigel, pH 5.8) and incubated in the dark at 28 °C to initiate callus induction. Six- to eight-week-old embryogenic calli were used as targets for transformation.

3.4.5 *Agrobacterium*-mediated transformation

Embryogenic calli were dried for 30 min prior to incubation with an *Agrobacterium tumefaciens* EHA105 suspension (OD_{600nm} = 0.1) carrying the cloned binary vector, pRGE32_OsPsbS1_8xgRNA. After a 30 min incubation, the *Agrobacterium* suspension was removed. Calli were then placed on sterile filter paper, transferred to co-cultivation medium (N6 salts and vitamins, 30 g/L maltose, 10 g/L glucose, 0.1 g/L myo-inositol, 0.3 g/L casein enzymatic hydrolysate, 0.5 g/L L-proline, 0.5 g/L L-glutamine, 2 mg/L 2,4-D, 0.5 mg/L thiamine, 100 mM acetosyringone, 3.5 g/L Phytigel, pH 5.2) and incubated in the dark at 21°C for 3 days. After co-cultivation, calli were transferred to resting medium (N6 salts and vitamins, 30 g/L maltose, 0.1 g/L myo-inositol, 0.3 g/L casein enzymatic hydrolysate, 0.5 g/L L-proline, 0.5 g/L L-glutamine, 2 mg/L 2,4-D, 0.5 mg/L thiamine, 100 mg/L timentin, 3.5 g/L Phytigel, pH 5.8) and incubated in the dark at 28°C for 7 days. Calli were then transferred to selection medium (CIM plus 250 mg/L cefotaxime and 50 mg/L hygromycin B) and allowed to proliferate in the dark at 28°C for 14 days. Well-proliferating tissues were transferred to CIM containing 75 mg/l hygromycin B. The

remaining tissues were subcultured at 3- to 4- week intervals on fresh selection medium. When a sufficient amount (about 1.5 cm in diameter) of the putatively transformed tissues was obtained, they were transferred to regeneration medium (MS salts and vitamins ¹¹⁰, 30 g/L sucrose, 30 g/L sorbitol, 0.5 mg/L NAA, 1 mg/L BAP, 150 mg/L cefotaxime) containing 40 mg/L hygromycin B and incubated at 26 °C, 16-hr light, 90 $\mu\text{mol m}^{-2} \text{s}^{-1}$. When regenerated plantlets reached at least 1 cm in height, they were transferred to rooting medium (MS salts and vitamins, 20 g/L sucrose, 1 g/L myo-inositol, 150 mg/L cefotaxime) containing 20 mg/L hygromycin B and incubated at 26 °C under conditions of 16-hr light (150 $\mu\text{mol m}^{-2} \text{s}^{-1}$) and 8-h dark until roots were established and leaves touched the Phytatrays lid. When possible, multiple regenerants per calli were recovered. Plantlets were then transferred to soil.

3.4.6 Chlorophyll fluorescence measurements of NPQ

Leaf punches were sampled from mature, fully developed leaves at leaf stage V3-5 and floated on 270 μL of water in a 96-well plate. Plates were dark acclimated for at least 30 minutes prior to analysis. *In vivo* chlorophyll fluorescence measurements were determined at room temperature using an Imaging-PAM Maxi (Walz) pulse-amplitude modulation fluorometer. Fluorescence levels after dark acclimation (F_o , F_m) and during light acclimation (F_o' , F_m') were monitored in two ways:

To resolve phenotypically segregating *OsPsbS1* NCS gene-edited alleles from a single parent, a single leaf punch at the V3 stage was exposed to a 4-minute period of high-intensity actinic light (1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$) using periodic saturated pulses. Putative homozygous lines were identified as those progeny in the top and bottom 10% of total NPQ. The leaf punches were then used to determine hygromycin sensitivity and Cas9 transgene segregation as described (3.3.7).

Candidates above lacking Cas9 were re-sampled at the V5 leaf stage, phenotyping two leaf punches from two mature leaves per plant to compare NPQ relative to Nipponbare WT plants. NPQ was quantified during a 10-minute period of high-intensity actinic light (1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$) and 10 minutes dark relaxation (0 $\mu\text{mol m}^{-2} \text{s}^{-1}$) using periodic saturating pulses. NPQ in both cases was calculated (Equation 1).

$$\text{Equation 1: } NPQ = [(F_m - F_m') / F_m'].$$

3.4.7 Hygromycin sensitivity assay for high-throughput Cas9 transgene detection

Floated leaf punches assayed for NPQ capacity were used to determine hygromycin sensitivity and segregation of the transgene. Hygromycin B (50 mg/mL 1x PBS) was added to each well to a final concentration of 20 $\mu\text{g/mL}$ of antibiotic. Plates were

incubated under rice germination chamber conditions for three days, following which the Imaging-PAM Maxi (Walz) was used to identify differences in the maximum efficiency of Photosystem II (F_v/F_m) after 30 minutes dark acclimation using (Equation 2).

$$\text{Equation 2: } F_v/F_m = (F_m - F_o)/(F_m)$$

Sensitive, transgene-free plants had a decline in F_v/F_m of >0.3-0.4, whereas transgenic plants maintained a WT F_v/F_m of ~0.7-0.8.

3.4.8 Total RNA/protein extraction of representative lines with varying NPQ

2 cm of leaf tissue from the youngest fully developed leaf was collected and flash-frozen in RNase-free, DNase-free tubes containing Lysing Matrix D (FastPrep-24™) at midday. Leaf tissue was ground on dry ice using a FastPrep-24 5G™ High-Speed Homogenizer (6.0 m/s for 2 x 40 s, MP Biomedical). Protein and mRNA were extracted from the same leaf sample (NucleoSpin RNA/Protein kit, REF740933, Macherey-Nagel GmbH & Co., Düren, Germany).

3.4.9 Quantitative RT-PCR of *OsPsbS1* relative to two reference genes

Extracted mRNA was treated with DNase (ThermoFisher Scientific) and transcribed to cDNA using Omniscript Reverse Transcriptase (Qiagen) and a 1:1 mixture of random hexamers and oligo dT as recommended by the manufacturer. Quantitative reverse transcription PCR was used to quantify *OsPsbS1* transcripts relative to *OsUBQ* and *OsUBQ5* transcripts, in biological triplicate with technical duplicates using published methods to normalize qRT-PCR expression to multiple reference genes^{111,112}. Samples were run on a 7500 Fast Real-Time PCR system (Applied Biosystems, Gent, Belgium) in a total volume of 20 µL using 4 µL of 1:10 diluted cDNA. Primers were empirically validated on a 5-step dilution series of WT cDNA (1:1 to 1:81). All final primer pairs had an amplification efficiency between 90-105% and linear amplification within the dynamic range tested. A single peak in melt-curve analysis was observed for each gene of interest, verifying specificity of the amplicon. Primers and primer efficiencies are shown (**Table 3-3**).

Table 3-3. *qRT-PCR primer pairs and observed primer efficiency in WT across a 1:81 dilution series.*

Gene ID	Primer Name	Sequence	Primer Efficiency
UBQ ¹¹³ LOC_Os03g13170	oDP578q UBQ-F_Fu2020	TGCACCCTAGGGCTGTCAAC	95.3%
	oDP579q UBQ-R_Fu2020	GGCGAGTGACGCTCTAGTTCTT	
UBQ5 ¹¹⁴ LOC_Os01g22490	oDP580q UBQ5-F_Jain2006	ACCACTTCGACCGCCACTACT	105.1%
	oDP581q UBQ5-R_Jain2006	ACGCCTAAGCCTGCTGGTT	
PsbS ¹¹³ LOC_Os01g64960	oDP584q PsbS1F-Fu2020	CTGTTCCGGCAGGTCCAAG	95.9%
	oDP586q PsbS1R2-DP	CAAACCCGAGCATGGCGA	

3.4.10 Immunoblotting of whole leaf protein extracts

Precipitated protein was resuspended in the supplied protein solubilization buffer (PSB-TCEP) and quantified using a TCA colorimetric assay¹¹⁵. Samples containing 5 µg total protein were resolved using pre-cast SDS-PAGE Any KD™ gels (BIO-RAD), transferred to a polyvinylidene difluoride membrane (Immobilon-FL 0.45 µm, Millipore) via wet transfer, and blocked with 3% nonfat dry milk for immunodetection. Membranes were cut and incubated with the following antibodies. A rabbit polyclonal antibody raised against sorghum PsbS (SbPsbS) was generously shared by Steven J. Burgess (University of Illinois) and used at a 1:2,500 dilution. A rabbit polyclonal antibody raised against a synthetic peptide of the β-subunit of ATP synthase (Atpβ) was obtained from Agrisera (catalogue no. AS05 085) and used at 1:10,000 dilution. After incubation with an HRP-conjugated, anti-rabbit secondary antibody from GE Healthcare (1:10,000 dilution), bands were detected by chemiluminescence using SuperSignal West Femto Maximum Sensitivity Substrate (Thermo Scientific). Protein bands were quantified by densitometry with ImageQuant TL software (version 7.0 GE Healthcare Life Sciences, Pittsburgh, PA, USA). PsbS abundance was quantified relative to WT based on a dilution series of WT PsbS protein.

3.4.11 Steady-state NPQ acclimation to low, moderate, and high light

To assess steady-state differences in NPQ capacity across representative lines, genotypes varying in PsbS abundance were grown and the youngest fully expanded leaf of 8-week-old vegetative rice plants was phenotyped. Whole plants were dark acclimated for 1 hour and phenotyped at the following regime, quantifying chlorophyll fluorescence by a multiphase flash routine (LI6800, LI-COR, Lincoln, NE, USA) every 3 minutes: 15 min 0 µmol m⁻² s⁻¹ – 45 min 100 µmol m⁻² s⁻¹ – 45 min 800 µmol m⁻² s⁻¹ – 45

min 1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ – 15min 0 $\mu\text{mol m}^{-2} \text{s}^{-1}$, using increasing red-light intensity and a total of 10 $\mu\text{mol m}^{-2} \text{s}^{-1}$ blue-light for all light-dependent measurements. In addition to Equations 1+2, operating efficiency of PSII (ΦPSII) was also monitored (Equation 3).

$$\text{Equation 3: } \Phi\text{PSII} = (F_m' - F_s') / F_m'$$

3.4.12 In-parallel gas exchange and chlorophyll fluorescence analysis under increasing red light

Photosynthetic gas exchange dynamics were measured on the youngest, fully expanded flag leaf of 12-week-old flowering rice plants. Gas exchange measurements were performed using an open gas exchange system (LI6800, LI-COR, Lincoln, NE, USA) equipped with a 2-cm² leaf chamber and integrated modulated fluorometer. Whole plants were low light acclimated for 1-2 hours to mitigate afternoon depression of photosynthesis and dark acclimated for at least 30 minutes to allow for concurrent phenotyping of gas exchange (e.g. A_n , g_{sw}) and chlorophyll fluorescence parameters (e.g. F_v/F_m , NPQ). For all measurements the chamber conditions were set to: 400 ppm chamber [CO₂], 27°C chamber temperature, 1.4 kPa vapor pressure deficit of the leaf, 500 $\mu\text{mol s}^{-1}$ flow rate, and a fan speed of 10,000 rpm. Samples were assayed within the boundaries of ambient daylength (8 am – 5 pm).

Steady state photosynthesis and stomatal conductance was monitored in response to changes in red light intensity (100% red LED's, λ_{peak} 630 nm). Light intensity was varied from 0, 50, 80, 110, 140, 170, 200, 300, 400, 600, 800, 1000, 1500, and 2000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ with 10-20 minutes of acclimation per light step. Steady state was reached when the stomatal conductance, g_{sw} , maintained a slope less than 0.005 +/- 0.00025 SD over a 40 second period and when net assimilation rate, A_n , showed variation less than 0.5 +/- 0.25 SD over a 20 s period. Net assimilation rate, stomatal conductance, and intracellular [CO₂] was logged. A saturating pulse was then applied to collect all relevant chlorophyll fluorescence parameters using a multiphase flash routine. In addition to NPQ and F_v/F_m , we assessed intrinsic water use efficiency (iWUE), ΦPSII , and Q_A redox state (1-qL) (Equations 3-5). The derivation of 1-qL assumes a "lake" model for photosynthetic antenna complexes.

$$\text{Equation 4: } iWUE = A_n/g_{sw}$$

$$\text{Equation 5: } 1-qL = 1- (F_q'/F_v')/(F_o'/F_s')$$

3.4.13 PCR genotyping of transgene-free, edited lines

50 mg of leaf tissue was ground via bead beating (Lysing Matrix D, FastPrep-24™) and genomic DNA was extracted in 2xCTAB buffer at 65°C for 15 minutes. DNA was separated via chloroform phase separation and precipitated using isopropanol. The pellet was washed briefly in 70% ethanol before being dried and resuspended in 1xTE buffer.

Two overlapping regions spanning the 8 gRNA target sites were PCR-amplified via Phusion™ High-Fidelity PCR using 5x GC-rich buffer (**Table 3-4**). PCR products were amplified from at least two putative homozygotes per line and sequenced by Sanger Sequencing. Contigs of the ~4.3-kb upstream region were assembled using Snapgene.

Table 3-4. Primers used for PCR amplification and genotyping of *OsPsbS1* NCS.

Amplicon	Forward Primer (5' -> 3')	Reverse Primer (5' -> 3')	Amplicon Size	Add. Sequencing Primer
Distal end (gRNA 1-5)	oDP214: AGACAGAGGTATGTCAA TGTGTTATTG	oDP498: AAAGAGCAAATGGCC TACCA	2384bp	oDP500: CGTGTCTCCACGTC TTCTT
Proximal end (gRNA 6-8)	oDP497: TGGTAGGCCATTTGCTC TTT	oDP496: CATCCATCCAAATTCC AACCT	2167bp	oDP499: GAGCAAACACTCAG GCACAA

3.4.14 PacBio long-read whole-genome sequencing and analysis

Leaf tissue used for high-molecular weight genomic DNA (HMW gDNA) extraction was dark starved for 4 days before being sampled and flash frozen in liquid nitrogen. DNA was isolated using the NucleoBond HMW DNA kit (TakaraBio, Catalog #740160.2) with the following modifications: Plant leaves were ground by pestle and mortar under liquid nitrogen, where 1 g of ground leaf tissue was resuspended in 2.5 times the amount of recommended lysis buffer and incubated in a 50°C water bath for 4 hours. The amount of Binding Buffer H2 was also proportionately increased. HMW gDNA was resuspended by gentle pipetting in water and assessed for quality by Femto Pulse Analysis (median fragment length 13-15 kb).

HMW gDNA was pooled and sequenced using a PacBio Sequel II (QB3 Genomics, UC Berkeley, Berkeley, CA, RRID:SCR_022170) by HiFi circular consensus sequencing (CCS) on a single 8M SMRTcell. Untrimmed reads for overexpression lines N2-4_OX and N19-

1_OX were first quality checked using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were mapped to the *Oryza sativa* v7.0 reference genome¹¹⁶ using pbmm2, a minimap2 wrapper by Pacific Biosciences specifically for HiFi reads and assembled using pbIPA, Pacific Biosciences phased assembler (<https://github.com/PacificBiosciences/pbbioconda>). Sniffles (<https://github.com/fritzsedlazeck/Sniffles>) and pbsv (<https://github.com/PacificBiosciences/pbbioconda>) were used to identify structural variants from mapped reads. D-Genies¹¹⁷ was used to align the de novo assembly to the reference genome and generate dot plots. Integrative Genomics Viewer¹¹⁸ was used to visualize mapped reads, aligned assemblies, and sniffles structural variation output.

3.5 Results

3.5.1 A library of independent events for multiplexed editing upstream of *OsPsbS1*

To generate novel cis-regulatory variation upstream of *OsPsbS1*, eight specific and evolutionarily conserved gRNA target sites (**Table 3-1**) were introduced into Nipponbare (subsp. *japonica*) rice calli via *Agrobacterium*-mediated transformation (**Figure 3-2**). The design avoided targeting a putative QTL for NPQ activity in rice upstream of *OsPsbS1*^{100,103} (**Figure 3-2A**). 23 fertile, independent transformants were generated, and multiple sister lines were recovered from the same transformed calli when possible, yielding 78 T₀ plants (**Figure 3-2B,C**).

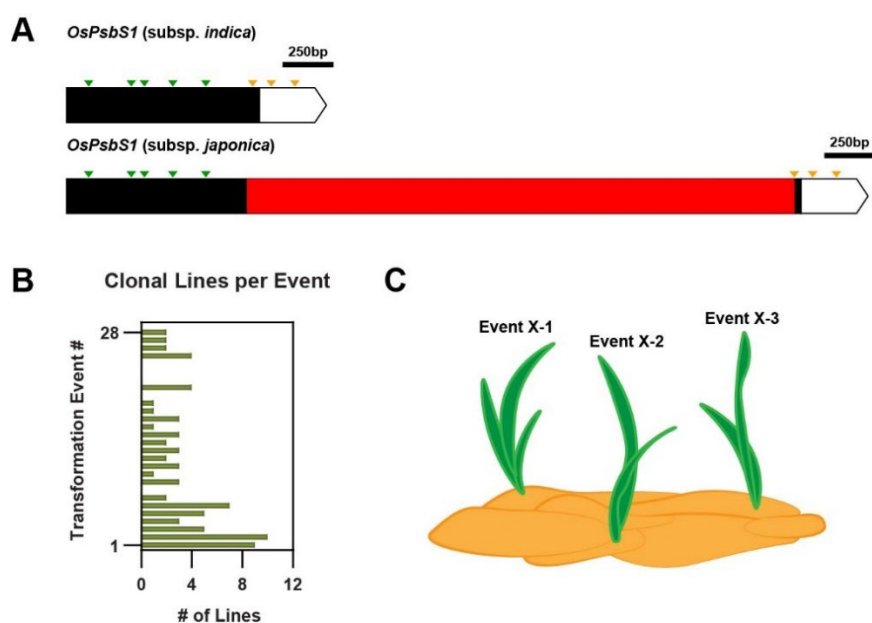


Figure 3-2. CRISPR/Cas9 target sites and naming scheme for Nipponbare (subsp. *japonica*) transformants.

(A) Target gRNA sites (triangles) distributed in distal (green) and proximal (yellow) regions upstream of *OsPsbS1*. An insertion specific to subsp. *japonica* varieties is marked in red. **(B)** Number of lines recovered per event for twenty-three independent transformants. **(C)** Multiple sister lines were regenerated from the same transformed callus when possible, differentiated by hyphens.

3.5.2 High-throughput chlorophyll fluorescence screening of edited, semi-dominant *PsbS* alleles

In vivo chlorophyll fluorescence screening was used to resolve unique phenotypes across the 78 diploid T_0 transformants, yielding up to 156 gene-edited alleles. However, phenotypes in the T_0 generation may be somatic and competition between alleles may mask dynamic changes in gene expression and activity. To circumvent this, each transformant was screened in the T_1 generation where alleles were fixed to homozygosity and the hemizygous Cas9 transgene could be segregated away (**Figure 3-3A**).

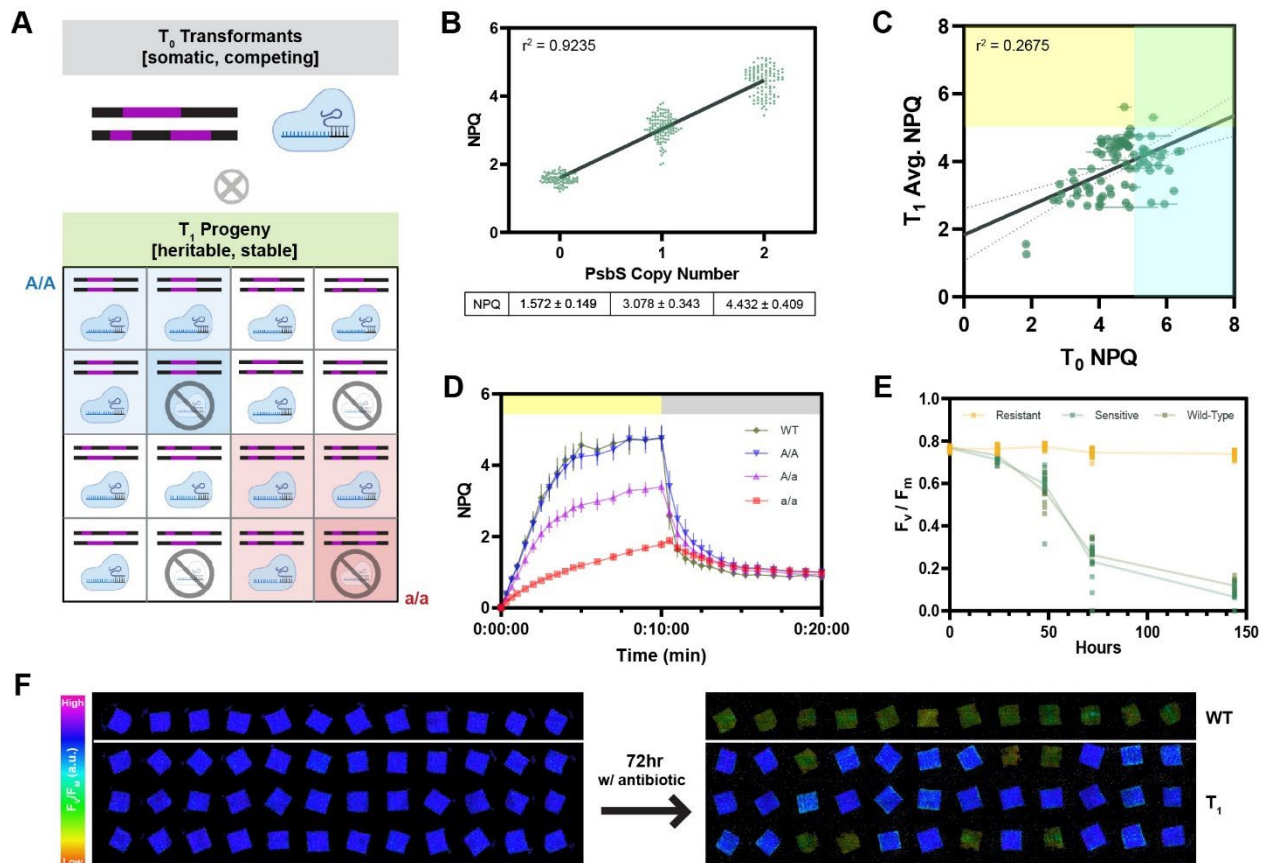


Figure 3-3. A high-throughput chlorophyll fluorescence resolves quantitative variation and Cas9 transgene segregation of T_1 progeny.

(A) Mendelian segregation of cis-regulatory *OsPsbS1* edits (purple) and inheritance of the hemizygous Cas9 transgene. **(B)** Linear regression of maximum NPQ capacity after 10 minutes of $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ of blue light in *PsbS* knockout, heterozygous, and WT lines ($n = 95, 117$, and 101 biological replicates respectively). **(C)** T_0 max NPQ weakly correlates with average max NPQ of its corresponding T_1 alleles ($n = 78$ T_0 plants, 2-4 biological replicates per T_1 population). Individuals with NPQ exceeding $+2$ SD of WT in the T_0 generation, T_1 generation, or both generations are shown in blue, yellow, and green respectively. **(D)** Chlorophyll fluorescence phenotyping of NPQ resolves homozygous alleles (A/A, blue triangle; a/a, red square) in 36 progeny of a single segregating T_0 parent relative to WT ($n=4$, green diamond). Data shown are means $\pm$ SEM. **(E,F)** Addition of the plant selection antibiotic hygromycin to

leaf punches phenotyped in (D) identifies sensitive individuals with inhibited F_v/F_m that lack the Cas9 transgene. Figure (A) was made using a Cas9 icon from Biorender.

PsbS is a semi-dominant trait that shows a strong linear correlation between copy number and NPQ capacity ($r^2 = 0.9235$) (**Figure 3-3B**). Despite this, T_0 max NPQ was a poor predictor of the average of T_1 fitness ($r^2 = 0.2675$). T_0 phenotypes overestimated the number of putative T_1 overexpressors (blue) and failed to identify a candidate stable overexpressor (yellow) (**Figure 3-3C**). To resolve stable, heritable alleles, populations of T_1 progeny were screened at high throughput to determine differences in NPQ capacity and Cas9 transgene segregation. Putative homozygous alleles were identified phenotypically (A/A and a/a, **Figure 3-3D**), with more subtle differences in population phenotypes resolved by binning the upper and lower 10% of NPQ phenotypes. Leaf punches used to assess NPQ capacity in (D) were treated with the plant selection antibiotic hygromycin and monitored for a decline in the quantum efficiency of photosystem II (F_v/F_m), a common indicator of plant stress and photodamage (**Figure 3-3E,F**). Obvious differences could be observed after only 72 hours of incubation, with a decrease in F_v/F_m of over 75%. Using this approach, homozygous, Cas9-free plants could be identified after <2 hours of work, allowing us to rapidly identify fixed germplasm to move into the T_2 generation.

3.5.3 Gene-edited, overexpression alleles are present, albeit rare

Putative homozygous alleles were identified via pairwise comparison between WT plants and progeny from a single T_0 parent, as depicted in (**Figure 3-3D**). To assess the variation in phenotypes across all putative stable alleles, maximum NPQ for all 120 phenotypically resolved alleles was plotted (**Figure 3-4A**).

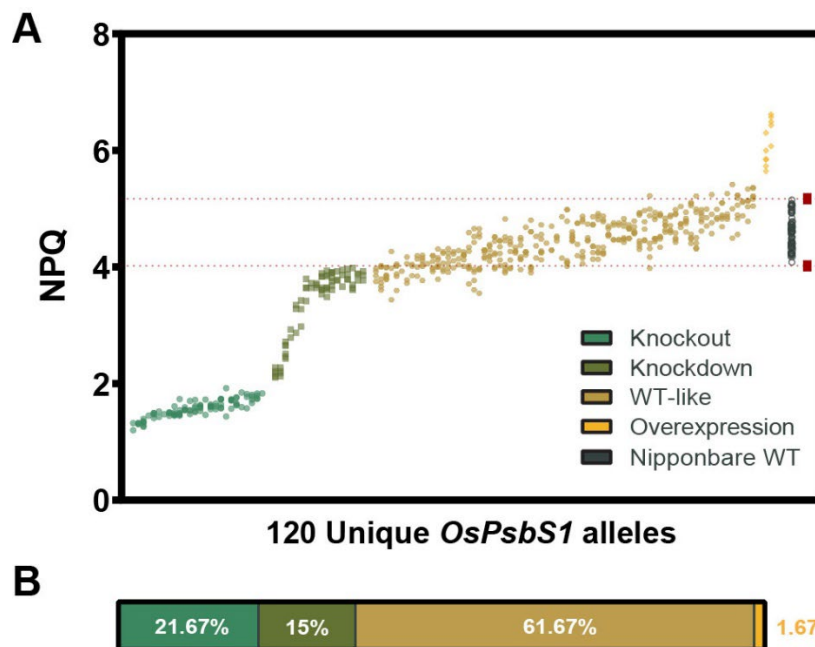


Figure 3-4. Maximum NPQ of 120 putative homozygous alleles spanning 78 T_0 events.

(A) Max NPQ ($n=1-4$ tech. replicates each) after 10 minutes at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ light plotted for each unique allele, sorted from lowest to highest average NPQ capacity. WT NPQ of 64 biological replicates spanning all phenotyping experiments is shown (gray, open circles) with boundaries demarcating ± 2 STD (red, dashed lines). Lines with all biological replicates below/above the WT boundary were binned by phenotype: knockout (teal, circle), knockdown (dark green, square), WT-like (brown, hexagon), and overexpressor (yellow, diamond). **(B)** Proportions of each phenotype observed.

Almost two-thirds of the 120 stable alleles isolated were WT-like in NPQ capacity (61.67%), with the second and third largest groups being knockout (21.67%) and knockdown (15%) phenotypes, respectively. Two independent overexpression alleles were isolated, comprising 1.67% of the total phenotypic variation (**Figure 3-4B**).

3.5.4 Chlorophyll fluorescence phenotypes correlate with PsbS protein abundance

To better understand the phenotypic variation generated, representative lines from the two events that yielded overexpression alleles, Event2 and Event19, were assessed for their ability to acclimate to low, moderate, and high light (**Figure 3-5A**). Assayed individuals showed no differences in initial F_v/F_m (**Figure 3-5B**) or NPQ at $100 \mu\text{mol m}^{-2} \text{s}^{-1}$ (**Figure 3-5C**). At $800 \mu\text{mol m}^{-2} \text{s}^{-1}$ of light, the Event2-5 knockout line (2-5_KO) and Event2-1 strong knockdown (2-1_SKD) line showed significantly reduced NPQ capacity relative to the Event2 WT control ($p < 0.0001$ and $p = 0.0041$ respectively). In contrast, the Event19-1 overexpression line (19-1_OX) showed significantly higher NPQ capacity than the Event19 WT control under the same condition ($p = 0.0001$). At $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$, significant differences between all non-WT-like alleles were observed ($p < 0.0003$) (**Figure 3-5C**).

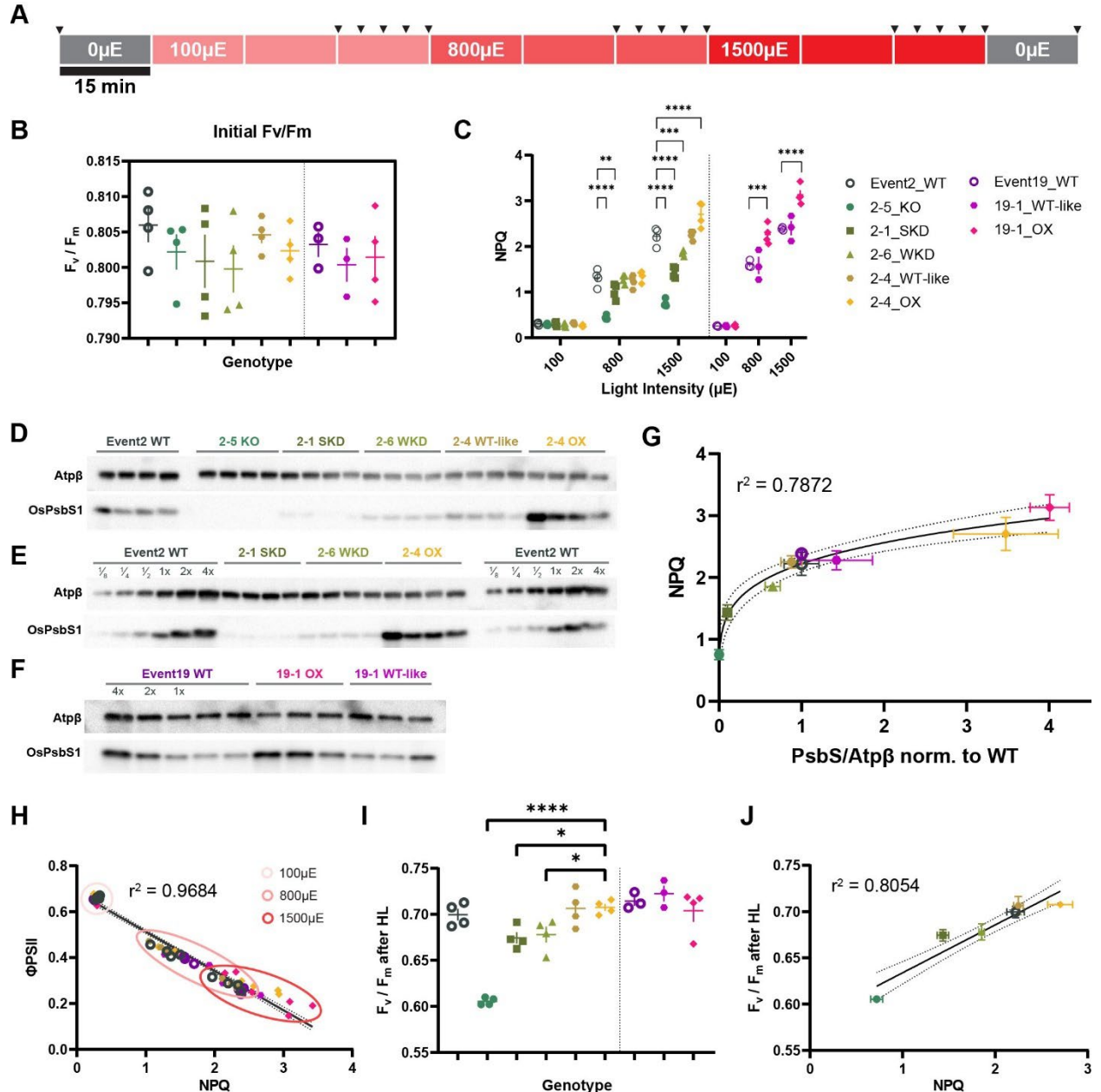


Figure 3-5. Changes in PsbS protein abundance underlie NPQ and Φ_{PSII} phenotypes.

(A) Light intensity acclimation scheme to assess steady state chlorophyll fluorescence traits. For each light intensity, steady-state chlorophyll fluorescence values of the last 5 measurements (spanning 15 minutes) were averaged for data visualization. Measurements of initial F_v/F_m and F_v/F_m after HL are noted by single black triangles at the beginning and end of the schematic ($n=3-4$ biological replicates plotted on top of mean $\pm$ 1 SEM). **(B)** Initial F_v/F_m of representative lines in Event2 and Event19: Event2 Nipponbare WT control (black, open circle, Event2_WT), Event2 knockout (teal, circle, 2-5_KO), Event2 strong knockdown (dark green, square, 2-1_SKD), Event2 weak knockdown (light green, triangle, 2-6_WKD), Event2 WT-like (brown, hexagon, 2-4_WT-like), Event2 overexpression (yellow, diamond, 2-4_OX), Event19 Nipponbare WT control (purple, open circle, Event19_WT), Event19-1 WT-like (magenta, hexagon, 19-1_WT-like), and Event19-1 overexpression (pink, diamond, 19-1_OX). **(C)** Differences in steady-state NPQ capacity at 100, 800, and 1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$. **(D-F)** Immunoblot of OsPsbS1 of representative Event2 and Event19 lines with 5 μg total protein loaded per well unless otherwise noted. Representative blots of two technical replicates are

shown. **(G)** NPQ capacity at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ plotted against OsPsbS1 band density normalized to 1) Atp β band intensity and 2) WT PsbS band intensity on the same blot (PsbS/Atp β norm. to WT = 1). Data were fit to a logarithmic curve using each individual data point (mean $\pm$ 1 SEM shown), with dashed lines constraining the 95% confidence interval. **(H)** Correlation between NPQ and ΦPSII of lines with WT or higher NPQ capacity at all three light intensities denoted by colored circles (light pink, $100 \mu\text{mol m}^{-2} \text{s}^{-1}$; pink, $800 \mu\text{mol m}^{-2} \text{s}^{-1}$; dark pink, $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ and a fitted linear regression with dashed lines constraining the 95% confidence interval. **(I)** Residual F_v/F_m 15 minutes following the end of the increasing light intensity regime (n=3-4 biological replicates plotted on top of mean $\pm$ 1SEM). **(J)** A linear regression of residual F_v/F_m against NPQ capacity at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$. For all pairwise comparisons, significance was determined by ordinary one-way ANOVA ($\alpha=0.05$) using Dunnett's test for multiple comparisons against Nipponbare WT (B,C) or the respective Event2 or Event19 overexpression datasets (I) and is denoted by asterisks (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p < 0.0001$).

Immunoblot analysis of the OsPsbS1 protein highlights the dynamic range of absolute protein abundance (**Figure 3-5D**) and abundance relative to a Nipponbare WT control dilution series (**Figure 3-5E,F**). In Event2, the 2-1_SKD allele and 2-6 weak knockdown allele (2-6_WKD) accumulated 1/8 and 1/4 of the WT OsPsbS1 protein level, respectively. In contrast, the 2-4 overexpression allele (2-4_OX) accumulated 2-4 times the WT OsPsbS1 protein, similar to 19-1_OX. 2-5_KO is below the threshold of PsbS detection. When normalized against the loading control, chloroplastic Atp β , as well as the Nipponbare WT controls of each membrane to allow comparison across experiments, it was possible to fit a logarithmic relationship between OsPsbS1 protein abundance and NPQ capacity at $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ for all genotypes ($r^2=0.7872$), consistent with PsbS dosage-dependent activity observed previously in *A. thaliana*¹¹⁹ (**Figure 3-5G**).

NPQ is photoprotective but can compete with light harvesting. To assess the extent to which photosynthesis was constrained in the isolated overexpression alleles, ΦPSII , the operating efficiency of PSII in the light, was plotted against NPQ at all three light conditions for WT and OX lines (**Figure 3-5H**). A strong negative correlation was observed ($r^2 = 0.9684$), with individuals with the highest NPQ (i.e. 19-1_OX) showing a lower ΦPSII at both $800 \mu\text{mol m}^{-2} \text{s}^{-1}$ and $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ relative to other genotypes at that light intensity. To determine the extent to which NPQ was photoprotective, F_v/F_m was measured 15 minutes after the final light step of the assayed light regime. Relative to 2-4_OX, the 2-5_KO, 2-1_SKD, and 2-6_WKD alleles showed greater amounts of photoinhibition as indicated by declines in F_v/F_m ($p < 0.0001$, $p = 0.011$, $p = 0.024$ respectively) (**Figure 3-5I**). There was no difference in phenotype between the overexpression and WT-like alleles for each event. However, plotting F_v/F_m after the light regime against NPQ capacity at the last $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ last step revealed a strong correlation between NPQ capacity and the maintenance of the quantum efficiency of PSII (**Figure 3-5J**). Thus steady-state acclimation of rice lines at varying PsbS abundances and light intensities revealed possible tradeoffs between light harvesting and photoprotection at varying PsbS levels.

3.5.5 Varying *PsbS* abundance alters steady state gas exchange phenotypes in response to red-light

Previous work has shown that transgenic overexpression of *PsbS* does not compromise steady-state CO₂ assimilation in rice^{101,120}, but it does increase intrinsic water-use efficiency (iWUE) under red-light in tobacco¹⁰². To determine the extent to which these phenotypes are consistent in rice genotypes with varying endogenous *PsbS* abundance, six genotypes were assessed for steady-state chlorophyll fluorescence and gas exchange parameters under increasing red light (**Figure 3-6**).

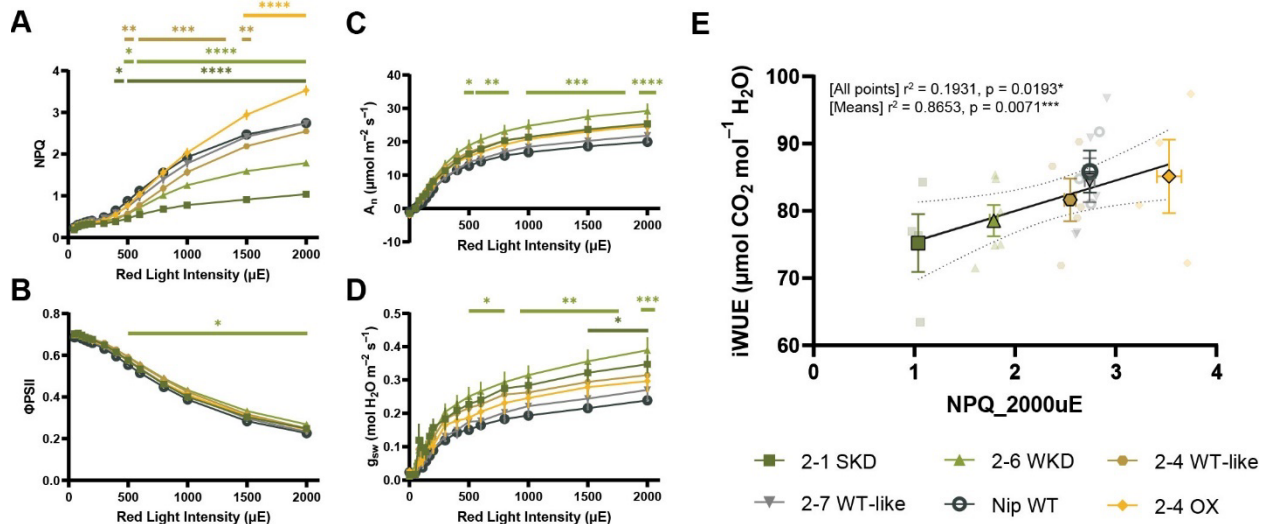


Figure 3-6. Chlorophyll fluorescence and gas exchange measurements of Event 2 lines with varying *PsbS* abundance.

(**A**) NPQ (**B**) Operating efficiency of PSII (Φ PSII) (**C**) CO₂ assimilation (A_n), and (**D**) Stomatal conductance (g_{sw}) as a function of incident red light on mature flag leaves ($n=4$ biological replicates, data shown $\pm$ 1SEM). Significance was determined by two-way ANOVA ($\alpha=0.05$) using Dunnett's test for multiple comparisons against the Nipponbare WT control, denoted by asterisks (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p < 0.0001$). (**E**) Linear regression of iWUE (A_n/g_{sw}) as a function of maximum NPQ capacity for each genotype ($n=4$ biological replicates, individual data points and means $\pm$ 1SEM shown). Genotypes are marked as followed: Nipponbare WT control (black, open circle, Nip_WT), Event2 strong knockdown (dark green, square, 2-1_SKD), Event2 weak knockdown (light green, triangle, 2-6_WKD), Event2 WT-like1 (brown, hexagon, 2-4_WT-like), Event2 WT-like2 (gray, inverted triangle, 2-7_WT-like), and Event2 overexpression (yellow, diamond, 2-4_OX).

Consistent with the high-throughput screening of NPQ (**Figure 3-4**) and acclimation to low, moderate, and high light intensities (**Figure 3-5**), knockdown lines 2-1_SKD and 2-6_WKD showed significantly lower NPQ at light intensities greater than 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ ($p < 0.0001$) relative to the Nipponbare WT control. In contrast, 2-4_OX showed significantly higher NPQ only at 1500 and 2000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ ($p < 0.0001$) (**Figure 3-6A**). The 2-4_WT-like line showed modest reductions in NPQ between 500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ and 1500 $\mu\text{mol m}^{-2} \text{s}^{-1}$ ($0.002 < p < 0.0004$), though a second gene-edited WT-like control (2-7_WT-like) showed no significant differences from Nipponbare WT across all

phenotypes measured (**Figure 3-6A-E**). Mild knockdown in 2-6_WKD increased ΦPSII at light intensities over $500 \mu\text{mol m}^{-2} \text{s}^{-1}$ ($0.041 < p < 0.006$) (**Figure 3-6B**).

Consistent with an increased ΦPSII , weak knockdown of PsbS also increased steady-state CO_2 assimilation at light intensities over $500 \mu\text{mol m}^{-2} \text{s}^{-1}$ with increasing statistical significance ($p=0.016$; $p<0.001$) (**Figure 3-6C**). Stomatal conductance was also significantly higher in the 2-6_WKD line ($p=0.043$; $p=0.0007$) with a more modest increase observed in the 2-1_SKD line ($p=0.043$ - 0.048) relative to Nipponbare WT (**Figure 3-6D**).

To assess differences in iWUE due to varying PsbS activity, high-light ($>600 \mu\text{mol m}^{-2} \text{s}^{-1}$) iWUE was averaged per replicate and plotted against maximum NPQ. A significant, positive correlation ($p=0.019$ for all data points, $p=0.007$ for means) of increased iWUE with higher maximum NPQ capacity was observed (**Figure 3-6E**).

3.5.6 Q_A redox state is an inconsistent predictor of rice g_{sw}

The increased iWUE phenotype observed in transgenic tobacco was predicted to be mediated by the Q_A redox state, approximated by $1-qL^{102}$. Surprisingly, $1-qL$ did not significantly differ across lines with varying OsPsbS relative to the Nipponbare WT control (**Figure 3-7A**).

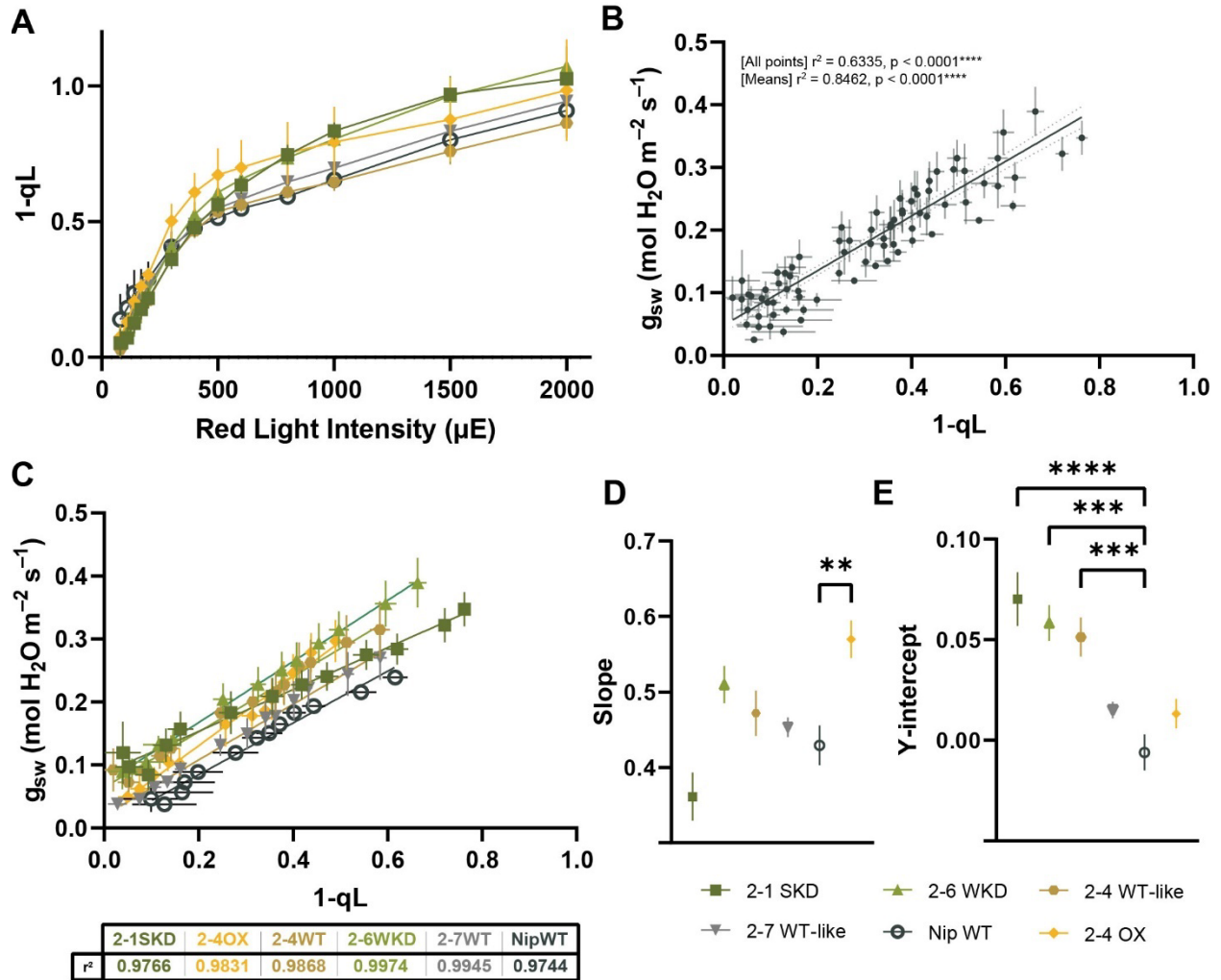


Figure 3-7. Correlation of Q_A redox state (1-qL) and g_{sw} as a predictor of iWUE with varying [PsbS].

(A) Q_A redox state (1-qL) as a function of incident red light on mature flag leaves (n=4 biological replicates, data shown $\pm$ 1SEM). Significance was determined by two-way ANOVA ($\alpha=0.05$) using Dunnett's test for multiple comparisons against the Nipponbare WT control. (B) Linear regression of 1-qL as a function of g_{sw} for all genotypes (n=24 biological replicates) and (C) individuals genotypes (n=4 biological replicates each) at all light steps exceeding $50 \mu\text{mol m}^{-2} \text{s}^{-1}$ (data shown $\pm$ 1SEM). (D) Slopes and (E) y-intercepts of linear regression lines by genotype (data shown $\pm$ 1SEM). Genotypes are marked as followed: Nipponbare WT control (black, open circle, Nip_WT), Event2 strong knockdown (dark green, square, 2-1_SKD), Event2 weak knockdown (light green, triangle, 2-6_WKD), Event2 WT-like1 (brown, hexagon, 2-4_WT-like), Event2 WT-like2 (gray, inverted triangle, 2-7_WT-like), and Event2 overexpression (yellow, diamond, 2-4_OX). Pairwise significance in (D,E) was determined by ordinary one-way ANOVA ($\alpha=0.05$) using Dunnett's test for multiple comparisons against Nipponbare WT and is denoted by asterisks (** $p \leq 0.01$, *** $p \leq 0.001$, **** $p < 0.0001$).

To further interrogate the putative correlation, 1-qL was plotted against stomatal conductance for all genotypes and light intensities. While a significantly non-linear ($p < 0.0001$) and relatively strong goodness of fit ($r^2 = 0.634$ for all points, $r^2 = 0.846$ for means) was observed (Figure 3-7B), the correlation was weaker than what has been reported in tobacco ($r^2 = 0.98$, $p < 0.001$)¹⁰². Leveraging the greater dynamic range in NPQ

capacity uncovered in this study, the correlation was assessed individually across all six tested genotypes (**Figure 3-7C**). While the goodness of fit significantly improved ($r^2 > 0.974$), the slope of the linear regression significantly differed between Nipponbare WT and 2-4_OX ($p = 0.007$), and the y-intercepts of the remaining non-WT lines all differed significantly ($p < 0.001$) (**Figure 3-7D,E**). In summary, all linear regressions of Q_A redox state and stomatal conductance, excluding that of the 2-7_WT-like control, significantly differed from Nipponbare WT as PsbS level was varied.

3.5.7 Most non-WT phenotypes are explained by variation at the 5'UTR

Differences in NPQ capacity, protein abundance, and iWUE substantiated observed phenotypic variation generated by non-coding sequence mutagenesis. To assess the causal mutations underlying these phenotypes, the ~4.3-kb upstream cis-regulatory region was sequenced by Sanger sequencing. 107 of the 120 unique alleles (**Figure 3-4**) could be amplified by PCR (89.2%).

Three alleles containing large deletions of the five distal gRNA target sites were observed (**Figure 3-4A**). Interestingly, all 3 lines had NPQ that was indistinguishable from WT (**Figure 3-4B**). As a result, it is likely that most of the observed phenotypic variation can be explained by proximal gRNA variants within and near the 5'UTR. Cis-regulatory analysis of Event 2 lines with varying NPQ (**Figure 3-5** thru **Figure 3-7**) revealed varying 5'UTR deletions. The size and relative location of the deletions corresponded with knockout (KO) and knockdown (KD) phenotypes, with most variation being driven by the second proximal gRNA (**Figure 3-4C**). As shown by the 24-5_18 WT-like allele, the transcription start site (TSS) is not essential for OsPsbS1 expression.

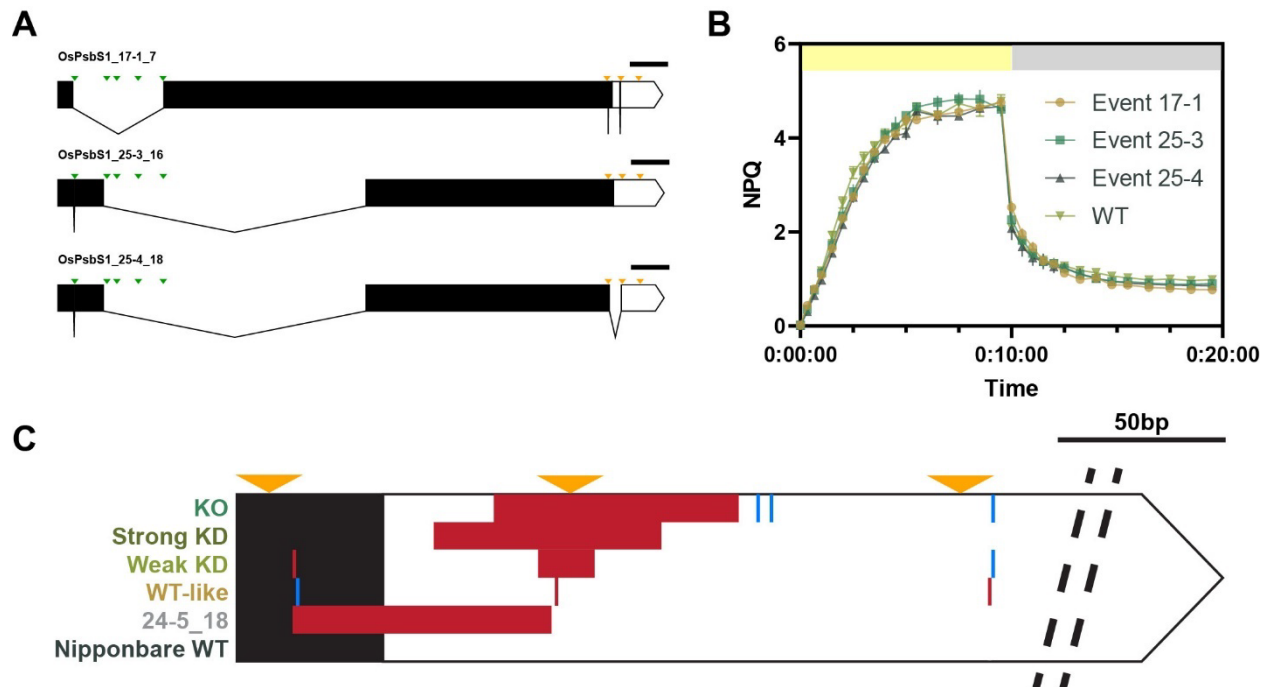


Figure 3-8. Variation near and within the 5'UTR likely drives observed phenotypic variation.

(A) Three unique cis-regulatory mutants with large deletions (black junctions) at distal gRNA sites (green triangles) mapped onto the Nipponbare (*sp. japonica*) promoter. Scale bar is 250 bp. **(B)** NPQ kinetics of large distal deletion lines in (A) as follows: Event17-1_7 (brown circle), Event 25-3_16 (teal square), Event 25-4_18 (gray triangle), Nipponbare WT (green inverted triangle). Data shown $\pm$ 1 SEM. **(C)** PCR-genotyped Event 2 mutations at proximal gRNA (yellow triangles) of lines with varying NPQ capacity used in this study (Figures 3-5, 3-6, 3-7). Event 2-4_OX could not be genotyped by PCR. Mutations in the cis-regulatory promoter (black box) and 5'UTR (white box) are shown upstream of the *OsPsbS1* coding sequence. Dashed lines indicate 146 bp of WT sequence not shown before the start codon. Deletions are shown in red, and insertions are shown in blue. Scale bar is 50 bp.

3.5.8 Long-read sequencing resolves complex structural variants underlying overexpression

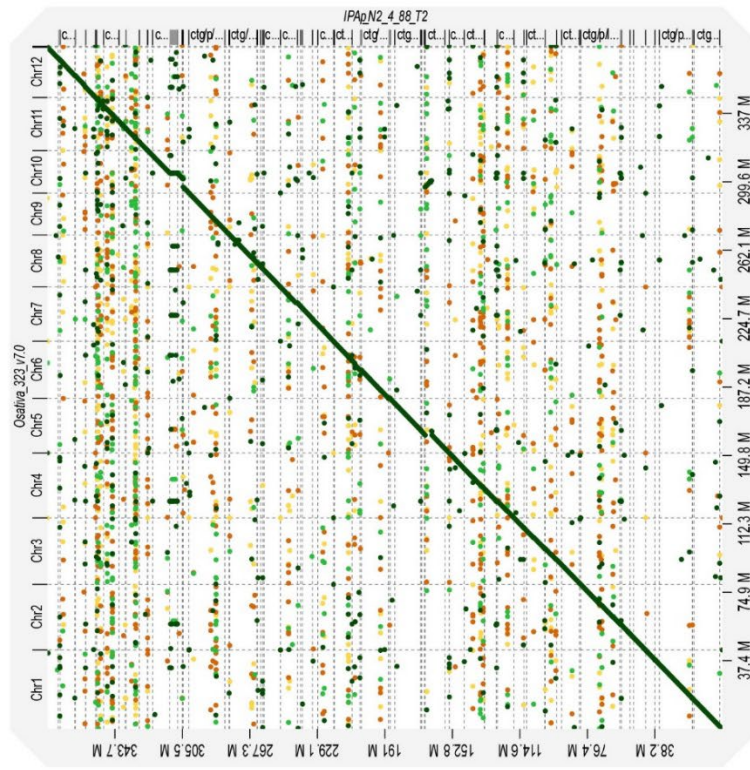
Unexpectedly, neither overexpression allele (Event2-4OX or Event19-1_OX) could be analyzed by PCR genotyping. To determine whether complex, structural variants were underlying these phenotypes, a homozygous 2-4_OX T₂ line and segregating 19-1_OX T₁ line were sequenced by long-read HiFi circular consensus sequencing (Pacific Biosciences). The 19-1 allele was sequenced in the T₁ generation due to unexpected sterility, also found in 5 of the originally 28 independent transformants (**Figure 3-2B**). Reads were assembled *de novo* and mapped onto the Nipponbare *O. sativa* v7.0 reference genome¹¹⁶ using bioinformatic tools developed by Pacific Biosciences (See Materials and Methods, 3.3.14).

First, all sequence and structural variants across the 12 rice chromosomes were visualized by dot plots (**Figure 3-9A**). Sequences were largely identical to the reference

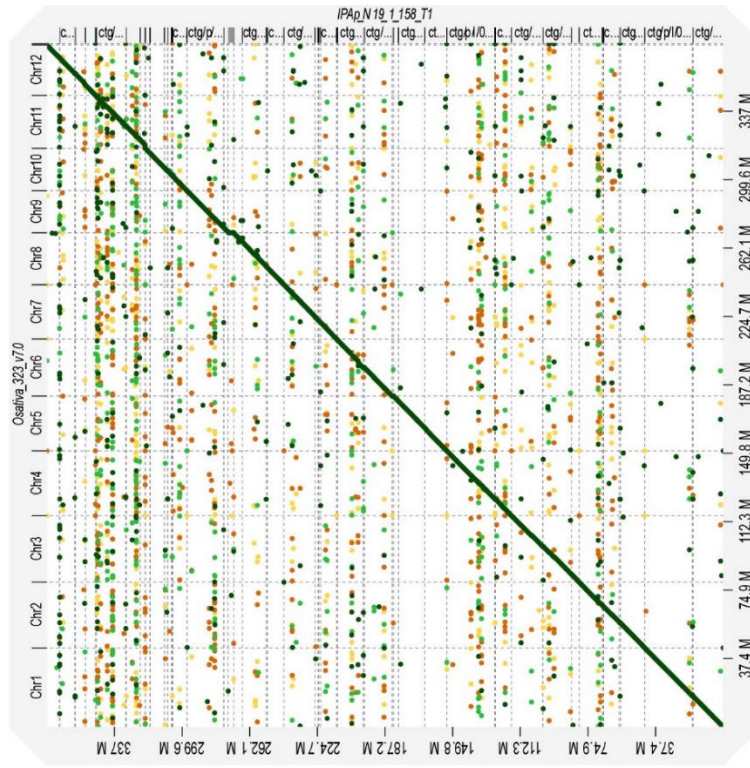
genome, as indicated by the contiguous green intersecting line. Small chromosomal variants (e.g. indels, SNP's) are visualized as scattered dots, with a highly consistent pattern between 2-4_OX and 19-1_OX, likely representing inherent sequence differences between the reference genome and lab strain used. In contrast, the two sequenced strains did show modest differences in chromosome-level structure, indicated by gaps in continuity that indicate larger insertions and deletions.

Increased resolution of the *OsPsbS1* Chromosome 1 locus highlighted large differences in structural variation not observable at the whole-genome scale (**Figure 3-9B**). The 2-4_OX line showed robust signatures of a ~252-kb inversion via dot plot, which was further substantiated at the sequence level by the identification of looped, head-to-head single long-read sequences visualized by Integrative Genome Viewer (IGV) (**Figure 3-9C**). In contrast, the 19-1_OX line showed no appreciable differences by dot plot. However, local investigation revealed the presence of a smaller ~4-kb inversion with a Sniffles variant called allele frequency of 0.361 (**Figure 3-9D**), the low frequency potentially explained by sequencing in the T₁ generation where homozygosity could not be confirmed. Thus, inversions of varying sizes account for observed PsbS overexpression phenotypes.

A



B



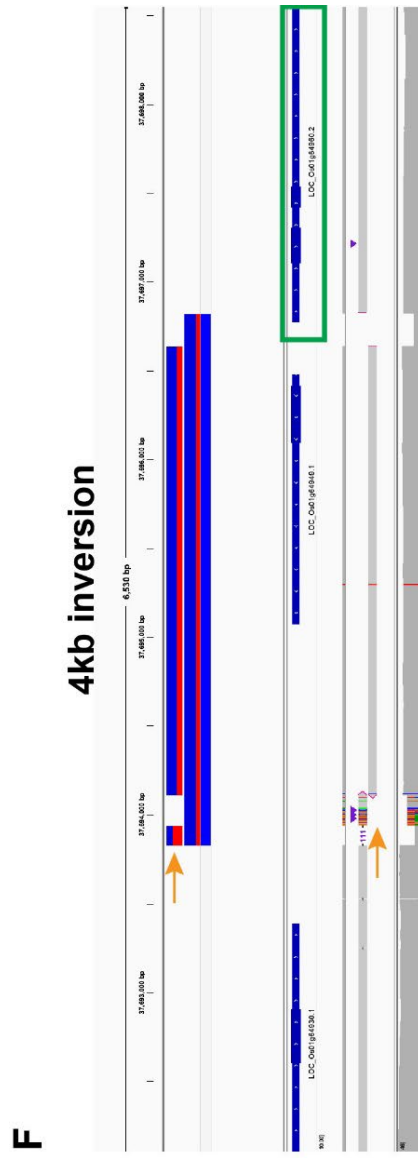
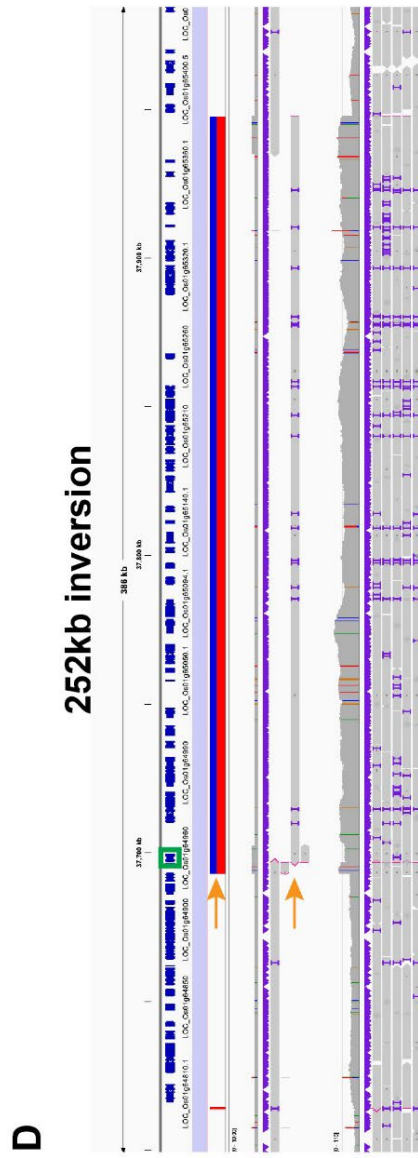
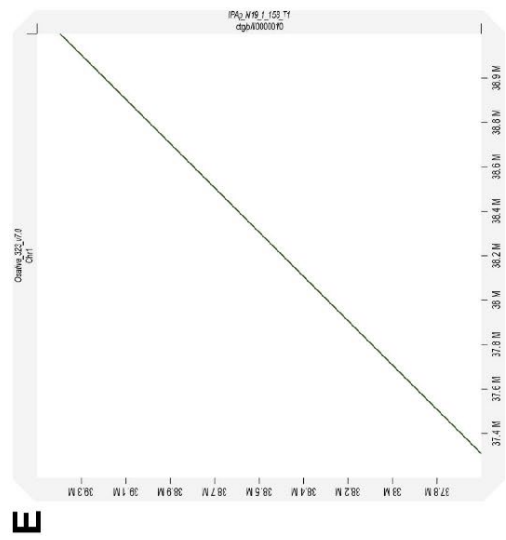
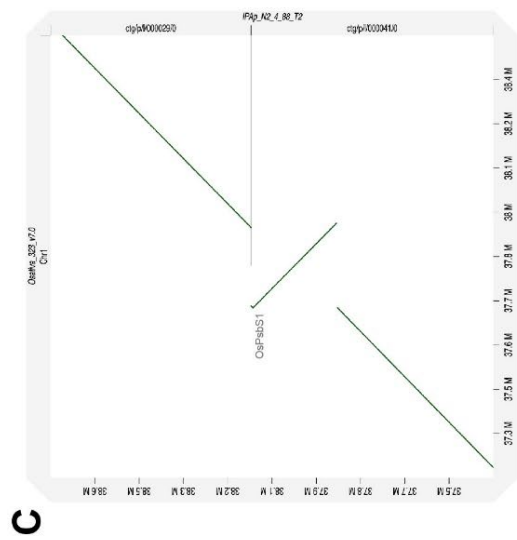


Figure 3-9. Long-read sequencing of two overexpression alleles identifies inversions upstream of *OsPsbS1*.

(A) Dot plot of 2-4_OX and **(B)** 19-1_OX, plotting sequenced OX variants (x-axis) against the reference genome (y-axis). Colors indicate the strength of matched sequences, with yellow being the lowest confidence (0-25%) and dark green being the highest (75-100%). Continuity on the intersecting axis (diagonal green) indicates high similarity between both genomes, with chromosome-level insertions and deletions shown by gaps in continuity. **(C+D)** Increased resolution of the Chr 1 locus (~37.85 Mbp) harboring *PsbS* for 2-4_OX and **(E+F)** 19-1_OX via dot plot (left) and Integrative Genomics Viewer (IGV, right). Green boxes denote the *OsPsbS1* gene (LOC_Os01g64960). Strength of the called structural variant was quantified by Sniffles (topmost orange arrow, quantified in red) and substantiated by head-to-head sequenced fragments (bottommost orange arrow).

3.6 Discussion

As our understanding of basic biology grows, so does our need for specific and efficient gene-editing tools to modify biology for application in agriculture and medicine. The CRISPR/Cas9 toolkit has the potential to meet that need if the design principles to modulate gene expression are well understood. Here we show the ability to use CRISPR/Cas9 mutagenesis of non-coding sequences to achieve native overexpression of genes at levels that compete with transgenic overexpression. Facilitated by a high-throughput screening pipeline for rapid detection of homozygous, Cas9-free progeny (**3.4.2**), which revealed a diverse array of quantitative phenotypes (**3.4.3**), these results have the potential to inform novel gene-editing strategies to optimize crop phenotypes.

Consistent with a large body of literature, we find that overexpression of *PsbS* increases NPQ capacity^{49,101,102}. Overexpression of *OsPsbS1* in the 2-4_OX allele only increased steady-state NPQ capacity at very high light intensities ($1500 + 2000 \mu\text{mol m}^{-2} \text{s}^{-1}$). Regardless, a strong correlation was observed between NPQ capacity and iWUE at light intensities as low as $600 \mu\text{mol m}^{-2} \text{s}^{-1}$ (**3.4.5**). However, this phenotype is poorly explained by Q_A redox state (**3.4.6**), which did not significantly differ and varied in its correlation with stomatal conductance across genotypes. The greater phenotypic range of NPQ values achieved here relative to transgenic tobacco may explain the increased resolution and variance within this trait, though whether red-light-mediated stomatal response is a more complex trait regulated in part by Q_A redox state and/or *PsbS* remains to be understood. Robust yield measurements in the field may help to reconcile differences in photoinhibition, photoprotection, and iWUE relative to plant fitness.

While the phenotypes observed reinforce observed differences in agronomic traits dependent on *PsbS* activity, the most important gains from this study come from the structural variants underlying changes in gene expression (**3.4.7, 3.4.8**). Due to the distribution of gRNA sites in our multiplex experimental design (**Figure 3-2A**), we found that distal regions of the *OsPsbS1* promoter, which are conserved between *indica* and *japonica* subspecies, were dispensable for expression of *OsPsbS1* (**Figure 3-8B**), and

much of the phenotypic variation observed is due to indels near and within the 5'UTR. Whether these differences are due to the presence/absence of cis-regulatory element motifs or due to changes in post-transcriptional mRNA processing or stability remains to be understood. However, the reduced genetic complexity (e.g. need for fewer guides, smaller editing region) present 5'UTR mutagenesis as an attractive approach for achieving controlled knockdown to bypass negative epistasis¹²¹. The potential of 5'UTR mutagenesis for overexpression, such as in the editing of competitive upstream open reading frames (ORFs) that may inhibit translation¹²², warrants further investigation.

Surprisingly, the two overexpression alleles that increased NPQ capacity were driven by inversions upstream of *OsPsbS1* (**Figure 3-9**). Allele 19-1_OX presents an interesting example as it carries a clean inversion between the distal and proximal gRNAs that may readily be replicated or introgressed into other varieties. While 19-1 displayed unexpected event-specific sterility, likely arising from off-target editing or somaclonal variation related to tissue culture due to tolerance to a substantially larger inversion in 2-4OX, the shared pattern of inversions driving overexpression in both alleles presents an interesting area for further investigation.

A revolution in long-read sequencing has revealed the pervasiveness of genomic structural variants. Complex structural variants (CSVs) including translocations, insertions, and inversions are persistent at both the population and pan-genome level as reported in tomato¹²³, rapeseed¹²⁴, maize¹²⁵, grapevine¹²⁶, and rice¹²⁷, affecting upwards of 20% of all genes. In fact, much of the current understanding of structural variants is driven by the prevalence of these changes driving various human cancers and diseases¹²⁸⁻¹³⁰. Unlike humans however, plants exhibit a much greater tolerance to (and thus abundance of) structural variants that is likely driven *in vivo* by transposable elements and recombination—both documented as key drivers of crop domestication^{131,132}.

In fact, chromosomal inversions have also been implicated in gene overexpression during the domestication of peach¹³³ and the complex rearrangements that underlie multiple myeloma¹³⁴. Recently, Lu et. al. showed that CRISPR/Cas9 could be used to drive native overexpression via promoter swapping¹³⁵, generating inversions in ~3% of transformed calli that increased gene expression of *OsPPO1* at varying frequencies. However, these inversions came at the cost of expression of the opposite promoter, knocking out gene expression of the target Calvin-Benson cycle protein 12 (*OsCP12*) gene (LOC_Os01g19740). In the instance of our 19-1_OX allele, the other inversion breakpoint occurred downstream of the 3'UTR of the neighboring gene (LOC_Os01g64930) and likely does not interfere with its expression. In the 2-4_OX allele,

evidence of a possible inversion-duplication due to differences in read coverage may further expand the types of variants that can be used to drive native overexpression, though the duplication is not observed within the dot plot. The work presented here, unbiased in its target design, reinforces and expands the genome-engineering potential of inversions for native gene overexpression.

In our dataset, overexpression alleles harboring an inversion account for 2 of 13 unique alleles that could not be PCR amplified and sequenced (3.4.7). Of the remaining 11 alleles, 8 were phenotypic knockouts, 1 was a phenotypic knockdown, and 2 were WT-like in NPQ activity. It is possible that CSVs also underlie these remaining alleles, increasing the potential frequency of CSV alleles from 1.7% to 10.8% by CRISPR/Cas9. Efforts to increase efficiency and frequency of inversions, such as by the use of paired prime-editing strategies may be promising¹³⁶. Additionally, guided design of inversions to minimize transcriptomic perturbations, such as inversions into transcriptionally active epigenetic marks¹³⁷ rather than promoter swaps, may prove fruitful in future plant engineering efforts. Further research into the potential behind editing other non-coding sequence types (e.g. introns and 3'UTRs) and identifying promising overexpression targets in other NPQ genes, *OsVDE* and *OsZEP*, will bring us closer to optimizing native photosynthetic efficiency in crop plants.

Chapter 4 : Quantitative transient expression of diverse PsbS proteins reveals the evolutionary origin of land plant-like PsbS activity

4.1 Acknowledgements

Sienna Marley identified and verified putative LHCSR and PsbS orthologs by BLAST, visualized sequence similarity networks using Cytoscape, and assessed early sequence differences by phylogenetic trees. Tim Jeffers was invaluable in processing the large volume and diversity of data collected in this project; he envisioned and executed the chi-squared analysis to identify significant amino acid variants between and within clusters, visualized and fit the aggregate NPQ data across experiments, and generated the UPGMA dendrograms of the tested PsbS sequences. Tim contributed written descriptions of his contributions in "Materials and Methods".

4.2 Abstract

The presence of pH-sensing orthologs implicated in non-photochemical quenching are ubiquitous across the Viridiplantae. However, our understanding of these proteins, Light-Harvesting Complex Stress-Related proteins (LHCSR) in green algae and mosses and Photosystem II Subunit S proteins (PsbS) in land plants, has been restricted to mutant studies in few model species. To understand the functional sequence diversity of LHCSR and PsbS, 1468 orthologs were mined across 1400 species spanning green algae to angiosperms. Structural diversity was consolidated via sequence similarity networks with statistically significant differences in amino acid sequence verified between clusters. A subset of representative LHCSR and PsbS were synthesized and tested by quantitative transient expression in a *psbs* knockout mutant of *Nicotiana benthamiana*. Here, we show that a PsbS capable of land plant-like NPQ first arose in streptophyte algae and not mosses. We identify functional diversity across PsbS orthologs that show differing abilities to induce and sustain NPQ. We also observe unexpected differences within angiosperm orthologs, suggesting the possibility of co-evolution of PsbS sequence across species. This work is an exciting foundational point to assess structure-function NPQ relationships at the amino acid sequence-level, with the potential to one day exert synthetic control over NPQ.

4.3 Introduction

The Viridiplantae (or as translated, "green plants") arose from an ancient endosymbiosis event ~1.5 billion years ago¹³⁸, giving rise to a severely underestimated 500,000+ species that span from green algae to land plants¹³⁹. Across these species, the

fundamental mechanics of photosynthesis are extremely well conserved. However, the structures that support efficient light harvesting differ dramatically in adaptation to various ecological niches. Concomitantly, photoprotective non-photochemical quenching (NPQ) mechanisms have co-evolved to maintain fitness¹⁴⁰.

Our early understanding of NPQ was dominated by land plants, with genetic screens revealing the importance of Photosystem II Subunit S (PsbS)¹⁹ and zeaxanthin (Zx)⁴⁸ in quenching excited singlet state chlorophyll (¹Chl*) under excess light. PsbS is an atypical light-harvesting complex-like protein whose oddity arises from its four-transmembrane structure²⁷ and its inability to bind chlorophyll or pigments relative to other minor LHCII proteins⁶⁰. PsbS harbors two protonatable residues that are essential for rapidly inducible NPQ (qE)²⁷ and are conserved in all PsbS orthologs studied to date.

In contrast, green algae require Light-Harvesting Complex Stress-Related (LHCSR) proteins to rapidly induce NPQ *in vivo*. This protein's functional role in green algal NPQ was not revealed until almost 10 years after PsbS, in a forward genetic screen of *Chlamydomonas reinhardtii*¹⁴¹. Similar to PsbS, the dominant *C. reinhardtii* ortholog LHCSR3 senses changes in Δ pH through 3 lumen-exposed, protonatable residues that account for 70% of qE capacity¹⁴². Attempts to fully abolish qE by site-directed mutagenesis have failed, evidenced by the inability of this protein to tolerate additional acidic residue substitutions in its C-terminal tail¹⁴³. Interestingly LHCSR3 is able to bind xanthophylls¹⁴⁴ and exhibits both Δ pH and carotenoid-dependent quenching capacity¹⁴⁵.

In terms of NPQ, the moss *Physcomitrium patens* represents an interesting evolutionary intermediate between *Arabidopsis* and *Chlamydomonas*—constitutively expressing both LHCSR and PsbS, which contribute additively to NPQ^{146,147}. In contrast to the lutein-dependent activity of *Chlamydomonas* LHCSR3 (CrLHCSR3)¹⁴⁴, *Physcomitrium* requires Zx for over 90% of its *in vivo* quenching capacity¹⁴⁸. Both PpLHCSR1¹⁴⁹ and CrLHCSR3¹⁵⁰ have been shown to also quench at Photosystem I (PSI), unlike the grana-localized PsbS protein. There is some evidence of green algal PsbS contributing to high light tolerance, but pleiotropic effects on LHCSR abundance and lack of genetic mutants in other green algal species have made its direct contributions to NPQ unclear^{151–153}.

Comparisons across these model organisms have revealed functional conservation in NPQ, but what drove the transition between LHCSR and PsbS is still unclear. A better understanding of how PsbS became the dominant quencher *in vivo* may help to identify important protein-protein interactions (e.g. with light-harvesting complexes, LHCs) that co-evolved during the transition to land. Additionally, a better understanding of extant

functional diversity may provide new insights into essential and non-essential components of PsbS structure in modulating NPQ. Finally, analysis at the amino acid sequence level may help inform future efforts surrounding the engineering of photosynthetic efficiency and NPQ. In this work, I describe my efforts to consolidate structural diversity across over 1400 LHCSR+PsbS orthologs, testing representative sequences in a unified experimental system to provide new insights into the sequences and structures that sustain NPQ *in planta*.

4.4 *Materials and Methods*

4.4.1 *Plant material and growth conditions*

Nicotiana benthamiana (accession Nb-1) plants were grown under ambient light conditions at a 10-hour daylength in a south-facing greenhouse. WT and stable knockout *psbs1 psbs2 (npq4)* seeds were germinated directly on a mixture of 4 parts Sunshine Mix #1 (Sungro) and 1 part perlite and coated with a thin layer of sand to stabilize seedlings. Plants were bottom watered for two weeks, the first week under a high humidity dome. Pots were then thinned and moved into flats with drainage to be top-watered through maturity. Plants were fertilized with JR Peter's Blue 20-20-20 fertilizer (<https://www.jrpeters.com/>) monthly.

4.4.2 *Mining of diverse PsbS and LHCSR*

LHCSR and PsbS orthologs were found in publicly available, high-quality reference genomes: land plants from Phytozome¹⁵⁴; green algae from Phycocosm¹⁵⁵; and other species across the Viridiplantae via NCBI¹⁵⁶. Protein sequences in (Table 4-1) were used as queries, after omitting the predicted chloroplast transit peptide as determined by PredAlgo¹⁵⁷ or TargetP¹⁵⁸ for LHCSR and PsbS orthologs, respectively. In addition, we expanded our representative species diversity by mining orthologs sequenced by the 1000 Plant Transcriptomes Project¹⁵⁹. Putative orthologs greater than 190 amino acids in length with an e-value cutoff of $<e^{-13}$ were then reciprocal BLAST'd against the three representative genomes to rule out closely related, but distinct light-harvesting complex genes. In total, we identified 1199 PsbS and 269 LHCSR orthologs across all methods.

Table 4-1. *PsbS and LHCSR query proteins used to BLAST against published genomes and transcriptomes.*

Clade	Species	Ortholog	GeneID	Length of cTP
Green Algae	<i>Chlamydomonas reinhardtii</i>	CrLHCSR3	Cre08.g367400 Cre08.g367500	32 aa
Green Algae	<i>Chlamydomonas reinhardtii</i>	CrPsbS	Cre01.g016600 Cre01.g016750	19 aa
Moss	<i>Physcomitrium patens</i>	PpLHCSR1	Pp3c15_11070V3	25 aa

Moss	<i>Physcomitrium patens</i>	PpPsbS	Pp3c20_23430	63 aa
Angiosperm	<i>Arabidopsis thaliana</i>	AtPsbS	At1g44575	54 aa

4.4.3 EFI-EST clustering by sequence similarity networks and visualization via Cytoscape

Using (<https://efi.igb.illinois.edu/efi-est/>)¹⁶⁰, 1199 PsbS and 269 LHCSR orthologs were clustered separately using an e-value cutoff of -50 for SSN edge calculation between sequences. Alignment stringencies were first determined using EFI-EST output scores to determine at what stringency threshold alignment scores became monotonic and refined empirically to determine at what stringencies cohesive clusters formed. For PsbS, all sequences were run first at a relatively high stringency of 95. The remaining unclustered sequences were then run through the same pipeline at a lower stringency of 55. The same was done for orthologs of LHCSR, at stringencies of 115 (high) and 75 (low).

Sequence similarity networks were visualized using Cytoscape¹⁶¹. To identify distinct sub-clusters, the sequence similarity network was further distinguished by gLay Community Structure Analysis¹⁶², scoring unique peptides by their relative neighborhood connectivity to other sequences in the same cluster. Clusters were named by the stringency they were formed at (High or Low) and ordered numerically by size of cluster (e.g. the High_1 cluster contained more orthologs than High_2, and so on).

4.4.4 Multiple sequence alignment (MSA), phylogenetic tree construction, and statistical analysis

Multiple sequence alignment (MSA) and phylogenetic tree construction were conducted in R using in-suite phylogenetic analysis tools, which will be released on Github upon future publication. Briefly, sequence tables were converted to fasta format using seqinr (v4.2-16). Fasta files were read into R using biostrings (v2.60.2), and multiple sequence alignment was conducted through "msa" function(v1.24.0), using the ClustalW algorithm. The MSA file was used to generate a distant alignment identity matrix through seqinr ("dist.alignment(matrix= "identity")") which was processed with the phangorn package (version 2.10.0) into a rooted tree by UPGMA. The trees were visualized with the ggtree (3.0.4) function in R, with the ggtree output object being saved to append cluster or phenotypic information to the tree. To visualize evolution of amino acid positions, MSA objects were converted into long-form data frames to visualize with ggplot graphics, which allowed us to append cluster ID, structure, and NPQ phenotype information to MSA. The above steps were repeated often to remove erroneous (e.g. rare insertion) or outlier sequences that became apparent in the MSA visualization. The Chi-squared test

was used to test the significance (i.e. p -value) of amino acids being non-randomly distributed across clusters at each aligned amino acid position for both LHCSR and PsbS, with residuals extracted per amino acid position and cluster to quantify its relative intra-cluster uniqueness. Frequency of amino acids at each position was calculated and ordered. Altogether, the “Diversity Bar” plots capture the most frequent and most unique amino acid at each position when compared across clusters (see 4.4.4).

4.4.5 Transient expression construct cloning and design

An empty pCAMBIA1300 backbone was used to clone a quantitative transient expression vector library. In brief, constitutive expression of the gene of interest is driven by the Arabidopsis UBQ10 (At4g05320) promoter, 5'UTR, and first intron (673bp total). Two unique restriction enzyme sites (HpaI and SpeI) were used to linearize the backbone prior to Gibson Assembly. 30-bp adapters were added either by primer overhang or gene synthesis for scarless, in-frame cloning with the constitutive promoter and downstream elements. Vectors differed in their transit peptide (No TP, AtPsbS chloroplast TP [54 aa], AtVDE luminal transit peptide [113 aa]) and in their fluorescence reporter (mScarlet, sfGFP, and mTagBFP2) for a total of 10 constructs including the no TP, no reporter control. The gene of interest was assembled in-frame with an Intein-F2A (IntF2A) linker which is post-translationally cleaved to promote stoichiometric expression of the target gene and reporter at near 1:1 frequencies^{163,164}. To enhance transient expression efficiency, a viral suppressor of RNA silencing, P19, was cloned from the pEAQ vector series into a neutral site between the left and right T-DNA borders¹⁶⁵. A table of overhangs used for cloning is included below, alongside a schematic of the backbone used in this study: pUBQ-AtPsbS_cTP-[SpeI_HpaI]-IntF2A-mScarlet+P19.



Figure 4-1. Plasmid map of an adapted pCambia1300 backbone (pUBQ-AtPsbS_cTP-[SpeI_HpaI]-IntF2A-mScarlet+P19) used for quantitative transient expression.

Representative map shown for a 10-vector library varying in transit peptide (no transit peptide, chloroplast, lumen) and reporter (mScarlet, sfGFP, mTagBFP2). Genes of interest were cloned in by Gibson/HiFi assembly in-frame with the IntF2A linker and AtUBQ10 promoter.

4.4.6 Gene synthesis of representative and synthetic LHCSR and PsbS

Representative LHCSR and PsbS sequences were selected across the defined clusters by (1) lab and agronomic relevance, (2) intra-cluster neighborhood connectivity as quantified by gLay¹⁶², (3) origin species, and (4) sequence quality. In addition, a synthetic consensus sequence was synthesized for each High stringency cluster, using a previously described pipeline¹⁶⁶. In short, sequences of individual clusters were aligned via MAFFT and assessed for the highest frequency amino acid at each position along the multiple sequence alignment. A gap was only incorporated when it exceeded 50% of all diversity at that position. Of the 18 synthetic LHCSR and PsbS consensus sequences generated, 17 still retained a strong predicted chloroplast target score by TargetP¹⁵⁸, suggesting robust structural conservation.

In total 63 PsbS (13 synthetic), 22 LHCSR (4 synthetic), and 11 Chlamy LHCSR (LHCB) proteins were selected, and their native cTP was predicted by TargetP 2.0¹⁵⁸. The predicted cTP was removed at the predicted site, and when predicted at low confidence, was manually aligned to other sequences in that cluster with strong transit peptide

scores to determine the probable transit peptide cut site by homology. For the Chlamy LHCb proteins, transit peptides were determined using PredAlgo¹⁵⁷ and evidence from published proteomic datasets. Sequences were codon optimized for *Nicotiana tabacum* and synthesized by Twist Biosciences with 30-bp overhangs for Gibson Assembly into the target construct (AtPsbS cTP).

Table 4-2. Overhangs used for Gibson Assembly of synthesized LHCSR, PsbS, and Chlamy LHCb.

Target Vector	Overhangs for Gibson Assembly [listed 5' -> 3']
No TP 5' overhang	TCGATTAATCTGAGTTTTCTGATTAACAG
AtPsbS cTP 5' overhang	AGACAATCATTCGTTCCCTCTCGCTCTCTTC
AtVDE luTP 5' overhang	GCGTTCCTTATTGTTCCATCTGCAGATGCA
IntF2A 3' overhang	TGTTTGCCTTCGGAAGTCTGAGATACTTACA

4.4.7 Transient expression

Transient expression was performed largely as described¹⁷. Briefly, *A. tumefaciens* strains were grown overnight in 1xLB media and resuspended in 1 mL induction media (10 mM MgCl₂, 10 mM MES pH 5.6, 150 μ M acetosyringone) and incubated for 2-3 h at 28 °C. Constructs were injected into *N. benthamiana npq4* leaves and demarcated using a black sharpie. Plants were placed under low light growth chamber conditions (50 μ mol m⁻² s⁻¹ 12/12 day-length) for 2-3 days prior to leaf sampling and NPQ and fluorescence quantification. A construct in the same backbone encoding a FLAG-tagged GUS without a fluorescence reporter was used as a negative control.

4.4.8 Chlorophyll fluorescence phenotyping

Leaf punches were sampled from mature, infiltrated tobacco plants and floated on 300 μ L of water in a 96-well plate. Plates were dark acclimated for at least 30 minutes prior to analysis. *In vivo* chlorophyll fluorescence measurements were determined at room temperature using an Imaging-PAM Maxi (Walz) pulse-amplitude modulation fluorometer. Fluorescence levels after dark acclimation (F_o , F_m) and during light acclimation (F_o' , F_m') were monitored through a 15-minute period of high-intensity actinic light (1500 μ mol photons m⁻² s⁻¹), a 5-minute period of dark relaxation (0 μ mol photons m⁻² s⁻¹), another 5-minute period of high-intensity actinic light (1500 μ mol photons m⁻² s⁻¹), and a final 10-minute period of dark relaxation (0 μ mol photons m⁻² s⁻¹). Leaf punches were re-phenotyped on consecutive days, with a cutoff of $F_v/F_m > 0.72$ to ensure representative NPQ capacity. NPQ was calculated using the formula below.

$$\text{Equation 6: } NPQ = [(F_m - F_m')/F_m'].$$

4.4.9 Fluorescence reporter quantification

Activity of the mScarlet reporter protein was quantified using a Tecan Infinite M1000 Pro (Tecan, USA) on leaf punches phenotyped for NPQ. Fluorescence was measured at $\lambda_{\text{Ex}} = 570 \text{ nm}$ (bandwidth 8 nm) and $\lambda_{\text{Em}} = 605 \text{ nm}$ (bandwidth 8 nm) at a Gain of 85, 95, and 105 to ensure fluorescence was within a quantifiable dynamic range. 9 technical replicates were taken from each leaf in a 3x3 format with a boundary of 600 μm per well.

4.4.10 Assessing NPQ by fluorescence phenotypes in aggregate

Single datapoint differences in NPQ were determined between the infiltrated leaf disc and the GUS negative control (n=8 per replicate) to allow for correlative analysis by fluorescence. All F_m and F_m' trace values were assembled into a longform dataframe to be amenable to data wrangling and descriptive functions by R tidyverse (v.1.3.2) suite of packages. NPQ was calculated at each time point after actinic light induction as $\text{NPQ} = (F_m - F_m') / F_m'$. The parameters MqE (1:20 – 2:00min; start of first actinic light phase) and MqTotal (24:00 – 25:00min; end of second actinic light phase) were used. A value for MqE or MqTot was considered significantly different if the difference between the GOI and GUS exceeded 1 standard deviation of the corresponding plate GUS control.

Equation 7:

$$\text{MqE} = \text{Avg}[\text{GOI_NPQ}_{1:20} + \text{GOI_NPQ}_{1:40} + \text{GOI_NPQ}_{2:00}] - \text{Avg}[\text{GOI_NPQ}_{1:20} + \text{GOI_NPQ}_{1:40} + \text{GOI_NPQ}_{2:00}]$$

Equation 8:

$$\text{MqTotal} = \text{Avg}[\text{GOI_NPQ}_{24:00} + \text{GOI_NPQ}_{24:30} + \text{GOI_NPQ}_{25:00}] - \text{Avg}[\text{GOI_NPQ}_{24:00} + \text{GOI_NPQ}_{24:30} + \text{GOI_NPQ}_{25:00}]$$

Significant MqE and MqTot values were plotted against log-normalized mScarlet fluorescence (a.u.) at Gain 95. Values that exceeded the detection threshold (>65000 a.u.) were extrapolated from normalized mScarlet values at Gain 85. To find a regression line for each construct relative to protein abundance, the base R linear model function ("lm") was used to find the values of control MqE or MqTot as a function of (linear) mScarlet arbitrary fluorescence values. Slopes of graphs were only generated on graphs with replicates exceeding the significance threshold. For constructs where no replicates had significant MqE or MqTot, the NPQ parameters were assigned a value of 0. Parameters were plotted as a heatmap against an MSA of all tested sequences.

4.4.11 Plotting predicted structural regions as a reference for MSA

To create a structural reference for multiple sequence alignments of diverse PsbS protein, the amino acid coordinates of *Arabidopsis* PsbS secondary structures were manually picked from public AlphaFold structures

(<https://alphafold.ebi.ac.uk/entry/Q9XF91>). This *in silico* model recapitulates the four transmembrane helices captured in the spinach PsbS crystal structure⁶¹ and predicted topological models for *Arabidopsis* PsbS²⁷, including sites of lumen-exposed glutamates occurring between transmembrane helices 1-2, and 3-4 (labeled from the N-terminus). Additionally, non-canonical helices were predicted by AlphaFold with high confidence, one of which contains the lumen-exposed glutamate between transmembrane helices 3-4. These smaller length helices were labeled as "Little folds". The same structural designation was used to model *Chlamydomonas reinhardtii* LHCSR1 (<https://alphafold.ebi.ac.uk/entry/P93664>) as a reference for LHCSRs, where the three transmembrane domains were also recapitulated by AlphaFold¹⁴⁰. To visualize the structures in reference to aligned MSA sequences, a gap-filling R function was to fill in alignment gaps ("-") in the reference sequence. Alignments to the reference guide revealed erroneous sequences (e.g. incomplete transcripts, incorrectly spliced sequences) rooted by highly conserved transmembrane helices. Transmembrane helices were modeled in green, little folds in light blue, and predicted N-terminal transit peptides (when uncleaved) in pink.

4.5 Results

4.5.1 Identification of 269 LHCSR and 1199 PsbS orthologs across the Viridiplantae

The One Thousand Plant Transcriptomes Initiative¹⁵⁹ has expanded the capacity for cross-phylogenetic sequence analysis. Using query proteins from *Chlamydomonas reinhardtii*, *Physcomitrium patens*, and *Arabidopsis thaliana* (**Table 4-1**), LHCSR and PsbS orthologs were mined across 12 broad phylogenetic clades (**Figure 4-2A**) spanning 1400 species across the Viridiplantae (**Figure 4-2B**). Species diversity was dominated by eudicots (740 species in total), with sizeable representation from green algae (207 species), monocots (129 species), bryophytes (89 species), and several other basal land plant clades. Consistent with our current understanding, LHCSR orthologs were found almost exclusively within the Green Algae and Bryophyte clades. PsbS was found ubiquitously across all clades, but the proportion of LHCSR:PsbS per species was on average weighted towards LHCSR in Bryophytes (1.5 LHCSR per PsbS ortholog) and Green Algae (1.23 LHCSR per PsbS ortholog) (**Figure 4-2C**).

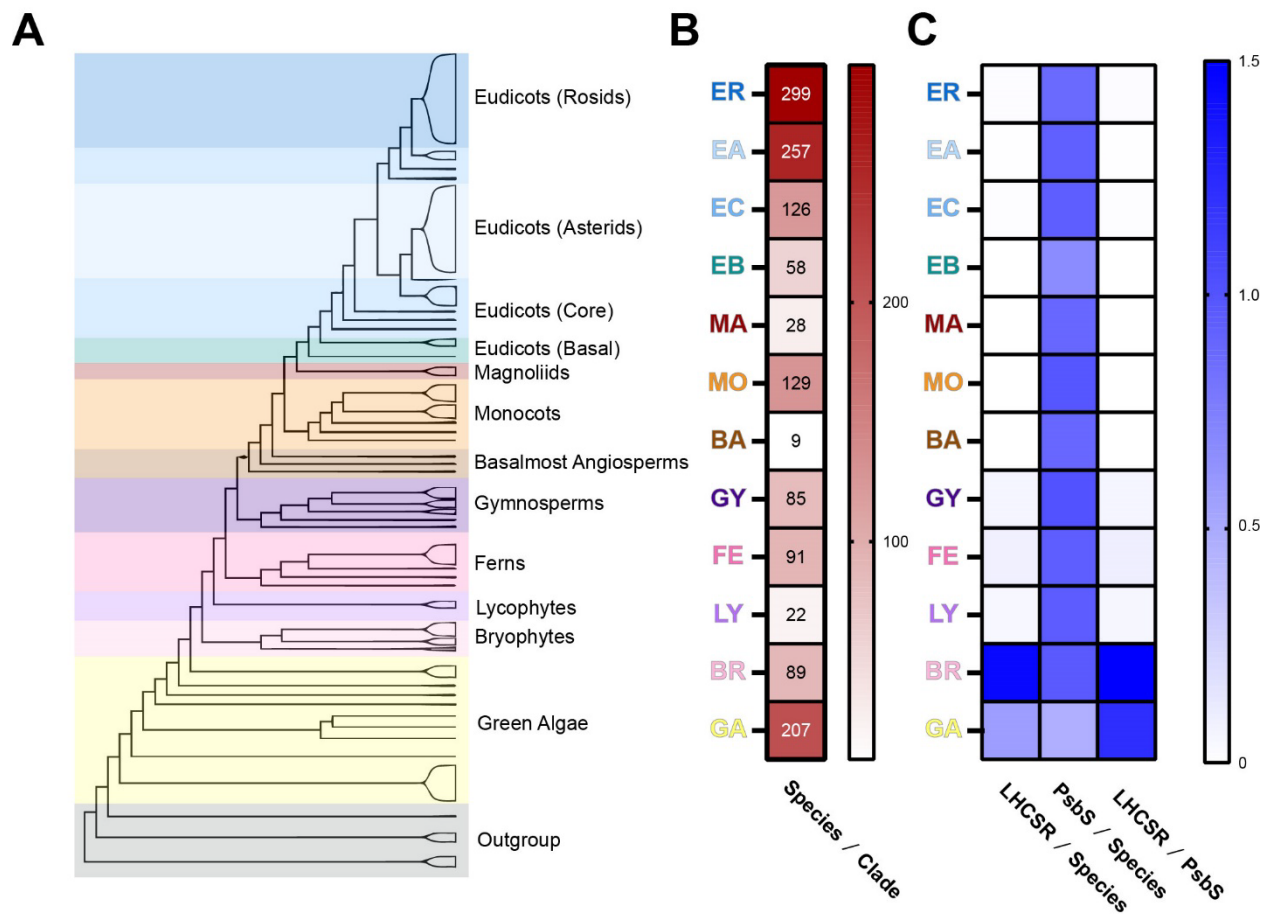


Figure 4-2. Dendrogram and species richness mined for diverse LHCSR and PsbS

(A) Dendrogram adapted from the One Thousand Plant Transcriptomes Initiative¹⁵⁹. Representative clades across the Viridiplantae are color-coded: Eudicots (Rosids), [ER] dark blue; Eudicots (Asterids), [EA] light blue; Eudicots (Core), [EC] blue; Eudicots (Basal), [EB] teal; Magnoliids, [MA] red; Monocots, [MO] orange; Basalmost Angiosperms, [BA] brown; Gymnosperms, [GY] purple; Ferns, [FE] pink; Lycophytes, [LY] lavender; Bryophytes, [BR] light pink; Green Algae, [GA] yellow; Red/brown algal outgroup, gray. **(B)** Heatmap of the number of species per clade used in this study. **(C)** Heatmap of the proportion of LHCSR, PsbS, and the proportion of LHCSR:PsbS found per species within each clade.

4.5.2 Sequence similarity networks consolidate LHCSR and PsbS diversity

To consolidate functional diversity, we utilized the Enzyme Function Initiative – Enzyme Similarity Tool (<https://efi.igb.illinois.edu/efi-est/>)¹⁶⁰, a sequence similarity network (SSN) platform that clusters protein sequences based on similarity via all-by-all BLAST. Sequences were first clustered at an empirically determined “High” stringency, before re-clustering the ungrouped sequences at a “Low” stringency, as described in Materials and Methods. The resulting sequence similarity networks (SSNs) reduced phylogenetic diversity into manageable clusters of LHCSR or PsbS orthologs (**Figure 4-3**).

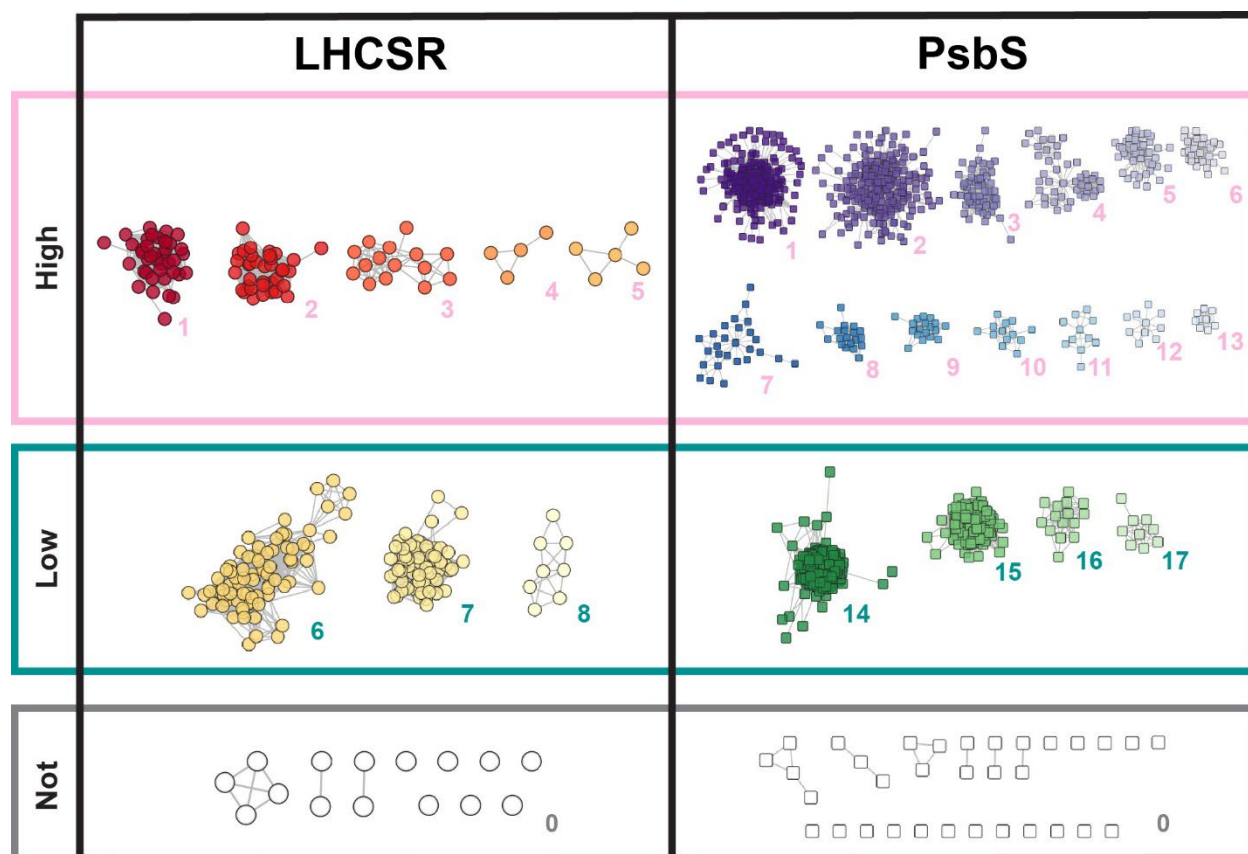


Figure 4-3. Sequence similarity networks of mined LHCSR and PsbS orthologs.

269 LHCSR orthologs grouped into 9 clusters (5 at High stringency, 3 at Low stringency, 1 unclustered). 1199 PsbS formed 17 clusters (13 at High stringency, 4 at Low stringency, 1 unclustered). Clusters are ordered within each sub-category by size.

4.5.3 Clusters mirror phylogenetic diversity and form based on sequence quality

To assess the extent SSN's captured meaningful phylogenetic differences, clusters were further analyzed for their sequence quality and composition. For both LHCSR (**Figure 4-4A**) and PsbS (**Figure 4-4B**), the sequences within the High stringency clusters showed an enrichment at the expected sequence length (LHCSR: 240-250 aa; PsbS 250-280 aa). In contrast, the Low and Not stringency clusters for both groups often peaked at lower sequence lengths. Manual alignments of a portion of these low-sequence length sequences revealed they were largely truncated sequences, likely the result of transcriptome quality and/or assembly. The appearance of full-length sequences in Low and Not clusters suggest the presence of divergent sequences that may be substantially more atypical than those of High clusters.

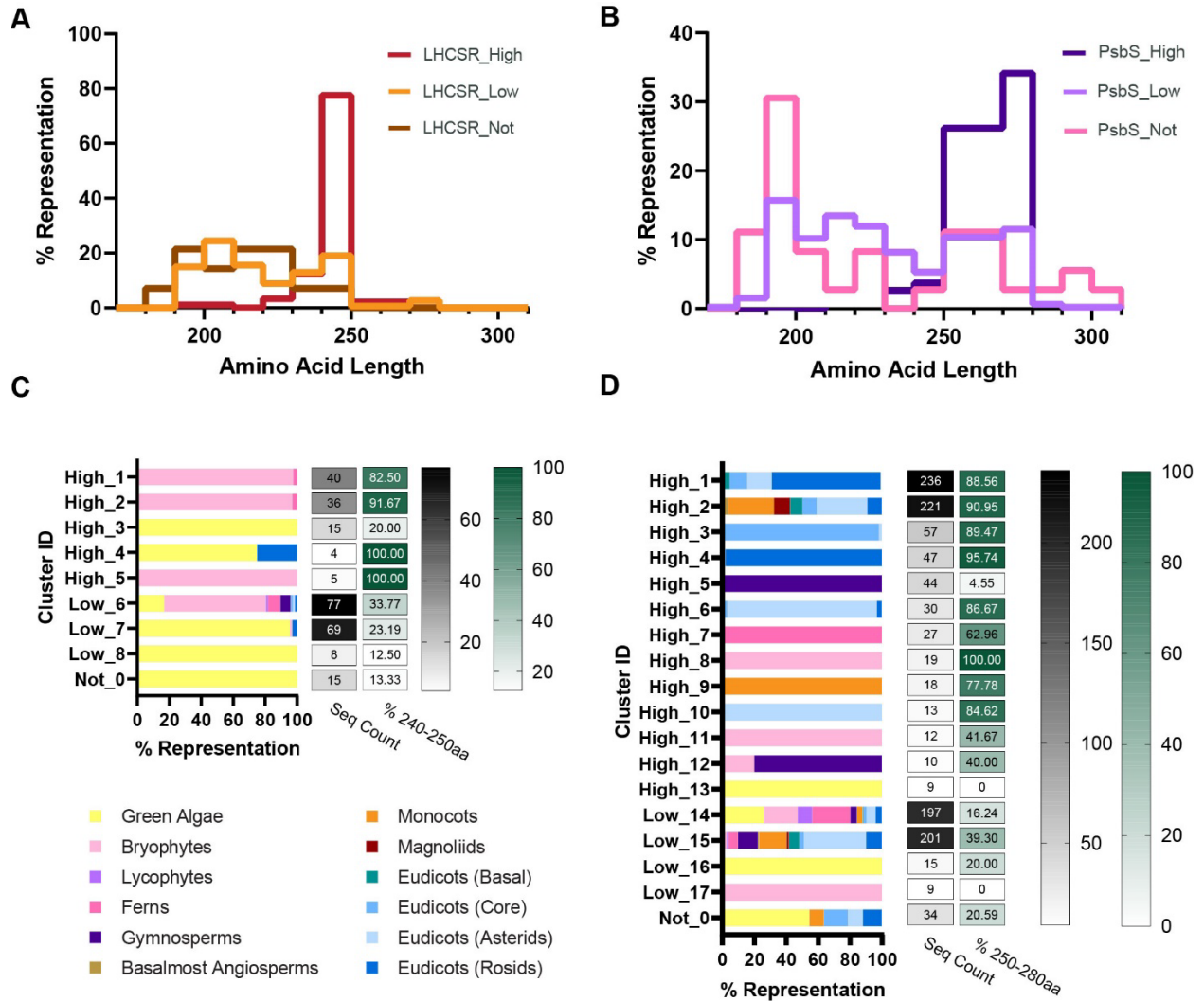


Figure 4-4. Stringency and cluster-dependent differences in sequence length and phylogenetic composition.

(A) Histogram of the number of mined peptides by amino acid sequence length for LHCSR (LHCSR_High, red; LHCSR_Low, orange; LHCSR_Not, brown) and (B) PsbS (PsbS_High, purple; PsbS_Low, lavender; PsbS_Not, pink) at High, Low, and Not Clustered stringencies. (C) Clade composition of each LHCSR and (D) PsbS cluster. The heatmap in black describes the number of sequences within each cluster. The heatmap in green describes the percentage of those sequences found within the peak amino acid sequence length range determined by their respective “High” clusters (i.e. 240-250aa for LHCSR, 250-280aa for PsbS). Clades are color-coded: Eudicots (Rosids), [ER] dark blue; Eudicots (Asterids), [EA] light blue; Eudicots (Core), [EC] blue; Eudicots (Basal), [EB] teal; Magnoliids, [MA] red; Monocots, [MO] orange; Basalmost Angiosperms, [BA] brown; Gymnosperms, [GY] purple; Ferns, [FE] pink; Lycopphytes, [LY] lavender; Bryophytes, [BR] light pink; Green Algae, [GA] yellow; Red/brown algal outgroup, gray.

The species composition of each LHCSR (**Figure 4-4C**) and PsbS (**Figure 4-4D**) cluster was assessed next. Consistent with (**Figure 4-2C**), LHCSR clusters were dominated by orthologs from Green Algae and Bryophytes. Surprisingly, LHCSR orthologs were detected in a few Eudicot (Rosids), Ferns, and Gymnosperm species. However, ground truthing of those largely truncated sequences in closely related species via NCBI-BLAST

failed to resolve a sufficiently orthologous hit and may be due to non-target algal contamination in wild samples.

In contrast, PsbS orthologs were found across the Viridiplantae. A vast majority of PsbS clusters were dominated by a single clade or highly related clades. The most notable exception was PsbS_Low14 and PsbS_Low15, which contained representative PsbS from almost all 12 phylogenetic clades. However, the low frequency of full-length sequences (16.24% and 39.3% respectively) may suggest this diversity is an artifact from clustering of fragmented sequences.

Consistent with **(Figure 4-4A)** and **(Figure 4-4B)**, High stringency clusters often had a greater proportion of “full-length” amino acids. The notable exceptions LHCSR_High_3 (Avg. length 253 ± 4.5 aa, 95% confidence interval) and PsbS_High_5 (Avg. length 243.2 ± 6.6 aa, 95% confidence interval) were on average just above and below threshold cutoffs respectively, potentially due to differences in transit peptide length. Altogether, these data suggest that clustering by SSN was a meaningful and intuitive way to consolidate phylogenetic diversity.

4.5.4 Inter-cluster comparisons resolve statistically significant differences as demonstrated by “diversity bar plots”

The fundamental algorithms underlying clustering via all-by-all BLAST differ from ClustalW-based multiple sequence alignments (MSA's), with the latter being more computationally intensive but also more sensitive to conserved sequence motifs. To assess the extent to which all-by-all BLAST generated meaningful and significantly different clusters, a subset of up to 32 random sequences per cluster, including those that were ultimately selected for phenotyping, were analyzed by MSA. In total, 387 PsbS sequences (31.9% of all diversity) and 64.8% of all 177 LHCSR sequences (64.8% of all diversity) were tested.

The results of the LHCSR and PsbS MSAs were further analyzed to confirm that candidate representative sequences for *in planta* testing represented most of the distinguishing sequence features unique to each cluster. To find amino acid positions that were informative of protein cluster divergence, the Chi-squared test was used to test the significance (i.e. *p*-value) of amino acids being non-randomly distributed across clusters at each aligned amino acid position. The resulting Chi-squared values were extracted and used to quantify residuals of the amino acids that significantly differed from other clusters at the same position, excluding rare outliers (red and black boxed amino acids). Taken together, the “diversity bar plot” highlights the conservation of

amino acids at each position, and depending on the size of the residual, the uniqueness of that amino acid relative to other clusters at the same position.

A representative diversity plot is shown for LHCSR and PsbS (**Figure 4-5**). Notably, the existence of low (**Figure 4-5A**) and high (**Figure 4-5B**) species diversity within a cluster does not result in large-scale differences in intra-sequence similarity (i.e. there are no obviously significant differences in color patterning between the representative sequences [top] and total cluster diversity [bottom]). While candidate sequences carrying atypical amino acids do exist (e.g. PsbS_O.sativa), on average, the candidate representative sequences captured nearly all divergent amino acids that were statistically distinguished within the cluster.

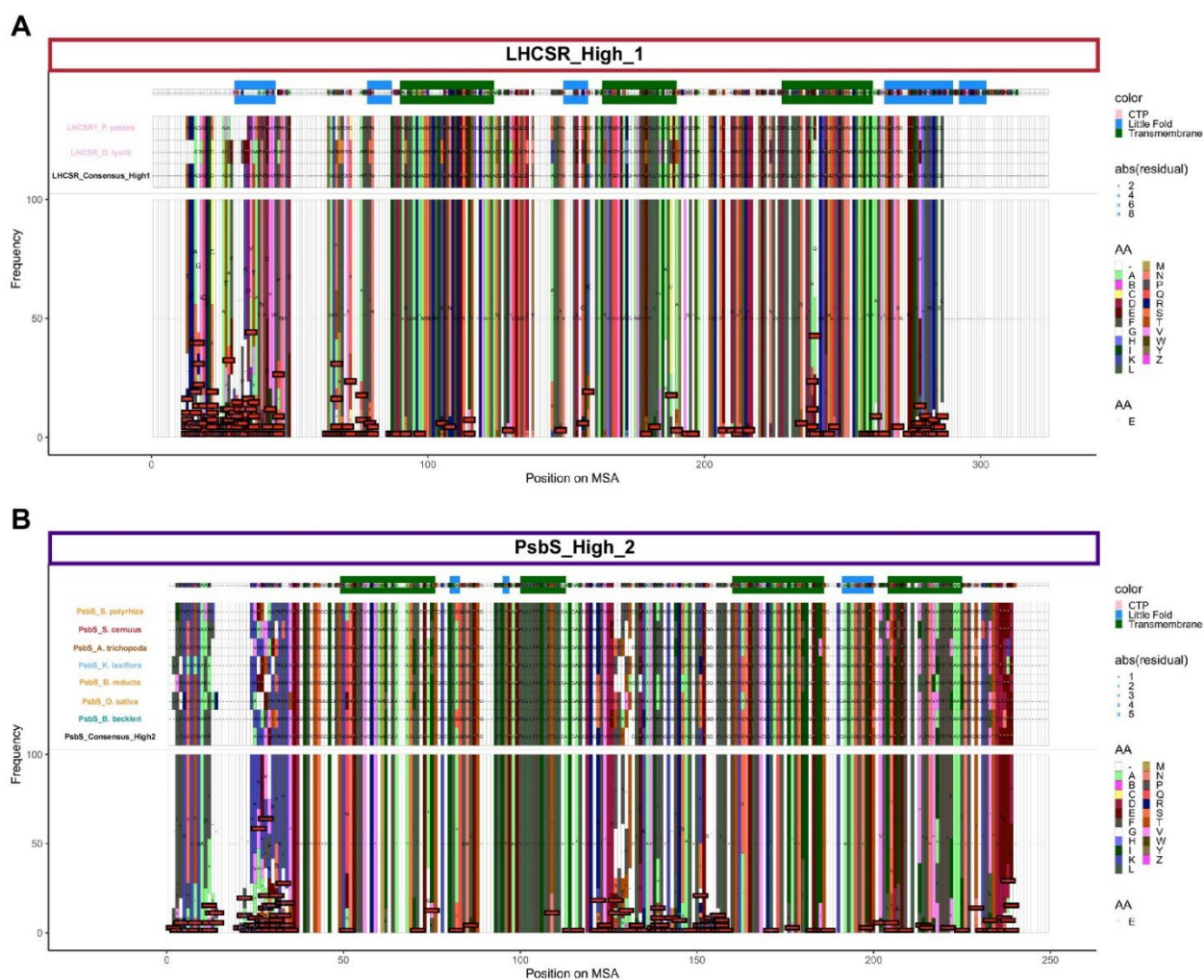


Figure 4-5. Multiple sequence alignment and Chi-squared analysis of amino acid enrichment within SSN clusters.

Representative outputs for **(A)** a cluster with low phylogenetic diversity (LHCSR_High_1) and **(B)** a cluster with high phylogenetic diversity (PsbS_High_2) mapped against AlphaFold models of LHCSR and PsbS, respectively. Frequency

of amino acids at each position for a given cluster were calculated, ordered by frequency, and color-coded. The values of Chi-squared test residuals were extracted from the larger all sequence MSA to quantify the level of enrichment of that amino acid (i.e. the larger the letter the greater the “uniqueness” of the cluster’s amino acid relative to other clusters). Highlighted in red boxes are atypical amino acids excluded from analysis. Species (left) are color-coded by clade as described in (**Figure 4-4**).

4.5.5 High-throughput cloning of diverse LHCSR and PsbS

Sequence similarity networks have thus far proven their ability to (1) consolidate significantly large datasets not amenable to MSA (**Figure 4-3**), (2) maintain phylogenetic diversity without bias (**Figure 4-4**), and (3) retain unique amino acids that distinguish clusters from one another (**Figure 4-5**). To capitalize on this library of genetic diversity, 22 LHCSR and 63 PsbS were identified and synthesized without their predicted chloroplast transit peptide (cTP) (**Table 2-1**) for Gibson assembly into a self-designed quantitative transient expression vector (**Figure 4-6A**).

In short, synthesized gene inserts were introduced into a linearized transient expression vector by Gibson/HiFi assembly in a 96-well plate. Competent *E. coli* cells were added and heat shocked in the same 96-well array, with transformants spotted at varying densities until single colony resolution was achieved. Assembly efficiency exceeded 97% by colony PCR. Thus, individual colonies were immediately picked from agar plates and miniprepmed to confirm in-frame assembly. Sequence-verified plasmids were transformed into *Agrobacterium* (GV3101) for transient expression and frozen in a 96-well array of glycerol stocks (**Figure 4-6B**).

Transient expression of *Nicotiana benthamiana npq4* plants deficient in PsbS revealed a robust capacity to induce NPQ across infiltrated constructs relative to WT positive controls (n=4) and GUS+FLAG infiltration negative controls (n=8). Phenotyping on multiple, consecutive days showed increases in both NPQ (**Figure 4-6B**) and mScarlet fluorescence, increasing the expression range available across experiments.

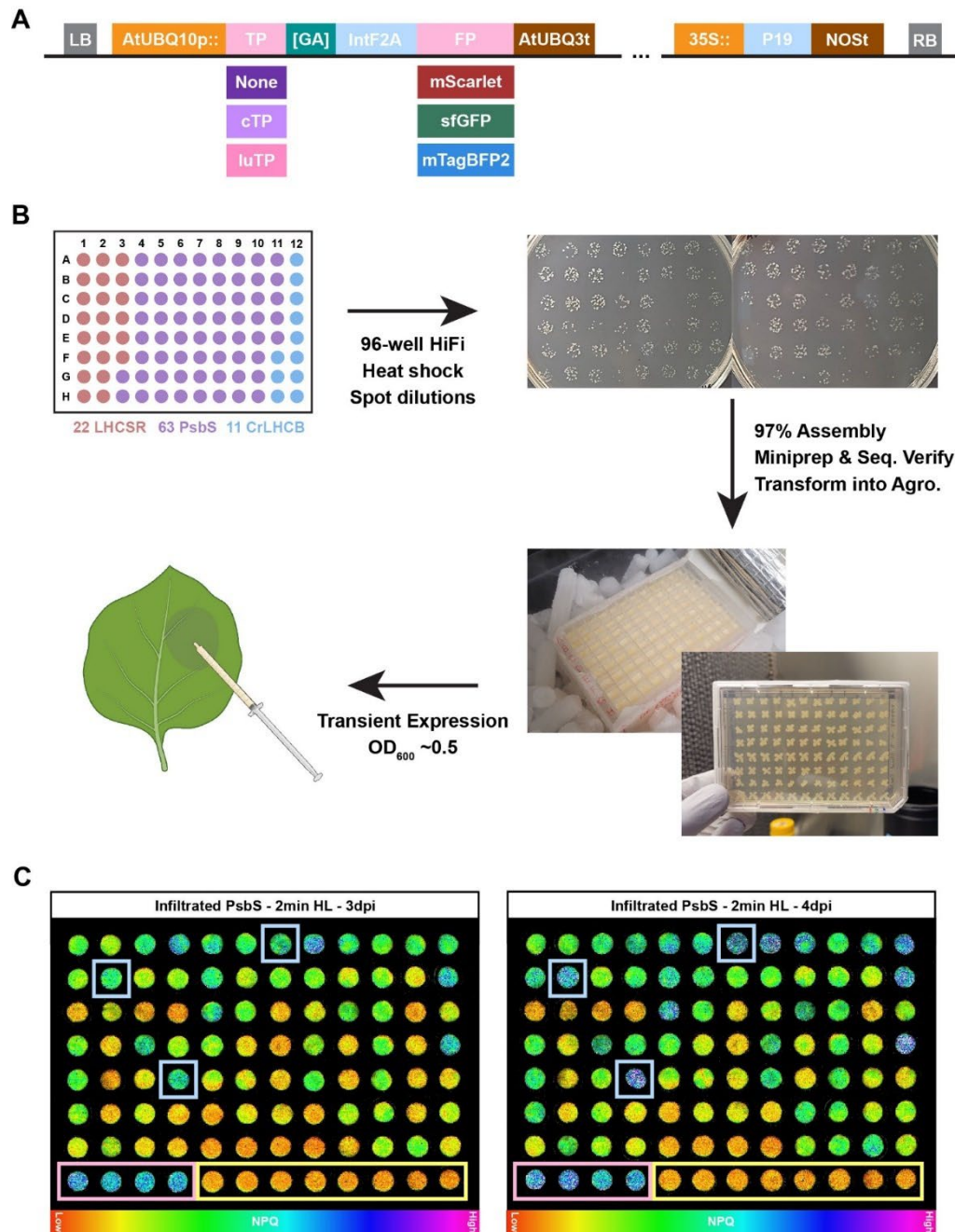


Figure 4-6. In-parallel cloning for high-throughput quantitative transient expression of LHCSR and PsbS variants

(A) Schematic of quantitative transient expression 10-vector backbone library varying in transit peptide (TP) and fluorescent reporter protein (FP). Expression of LHCSR/PsbS variants is driven by the Arabidopsis *UBQ10* promoter and 5'UTR, in-frame with a cleaved Intein-F2A (IntF2A) linker for stoichiometric expression of a cytosolic fluorescence protein. In this study, the chloroplast transit peptide (cTP) of Arabidopsis PsbS (First 54 aa, At1g44575) was used for co-expression with mScarlet. **(B)** 23 representative LHCSR, 63 representative

PsbS, and all 11 Chlamy LHCB genes were synthesized without their predicted cTP by Twist Biosciences. Inserts were assembled into the digested backbone in 2 μ L reactions in 96-well plates, heat shocked into XL-1 competent *E. coli*, and spotted in dilutions to single-colony resolution. Assembly efficiency exceeded 97% in a representative 72 assessed colonies. Constructs were sequence verified and transformed by electroporation into competent *Agrobacterium tumefaciens* strain GV3101 for transient expression. **(C)** Strains were infiltrated into the *N. benthamiana npq4* mutant generated in Chapter 3 and sampled as leaf punches to be phenotyped in 96-grid arrays by an Imaging-PAM (Walz). Leaf discs meeting the F_v/F_m threshold ($F_v/F_m > 0.72$) were phenotyped on multiple days post-infiltration (dpi) to take advantage of the accumulation of PsbS and mScarlet over time (examples of increased NPQ capacity between days in blue boxes) relative to the WT (pink) and npq4+GUS (yellow) controls.

Table 4-3. Synthesized LHCSR, PsbS, and CrLHCB sequences used in this study, with mature protein length.

Ortholog	Cluster	Gene_ID	ID in this study	Phylogenetic Clade	CTP_trim length
LHCSR	High_1	CMEQ_scaffold_2002752 Orthotrichum lyellii	LHCSR Orthotrichum lyellii	Bryophytes	212
LHCSR	High_1	Pp3c9_3440V3.1.p Physcomitrium patens	LHCSR1 Physcomitrium patens	Bryophytes	208
LHCSR	High_1	LHCSR Consensus High_1	LHCSR Consensus High_1	*	211
LHCSR	High_2	HRWG_scaffold_2002881 Buxbaumia aphylla	LHCSR Buxbaumia aphylla	Bryophytes	208
LHCSR	High_2	Pp3c15_11070V3.1.p Physcomitrium patens	LHCSR2 Physcomitrium patens	Bryophytes	208
LHCSR	High_2	LHCSR Consensus High_2	LHCSR Consensus High_2	*	209
LHCSR	High_3	Cre08.g365900 Chlamydomonas reinhardtii	LHCSR1 Chlamydomonas reinhardtii	Green Algae	224
LHCSR	High_3	Cre08.g367400 Chlamydomonas reinhardtii	LHCSR3 Chlamydomonas reinhardtii	Green Algae	227
LHCSR	High_3	LHCSR Consensus High_3	LHCSR Consensus High_3	*	222
LHCSR	High_4	FPCO_scaffold_2024786 Interfilum paradoxum	LHCSR Interfilum paradoxum	Green Algae	207
LHCSR	High_4	GAQ89003.1 Klebsormidium nitens	LHCSR Klebsormidium nitens	Green Algae	208
LHCSR	High_4	HAOX_scaffold_2024389 Spirogyra sp.	LHCSR Spirogyra sp.	Green Algae	207
LHCSR	High_5	IRBN_scaffold_2012024 Scapania nemorosa	LHCSR Scapania nemorosa	Bryophytes	208
LHCSR	High_5	YBQN_scaffold_2127488 Odontoschisma prostratum	LHCSR Odontoschisma prostratum	Bryophytes	208
LHCSR	High_5	LGOW_scaffold_2103234 Schistochila sp.	LHCSR Schistochila sp.	Bryophytes	208
LHCSR	Low_6	Cz16g02040.t1 Chromochloris zofingiensis	LHCSR Chromochloris zofingiensis	Green Algae	215
LHCSR	Low_6	WEEQ_scaffold_2087311 Phaeoceros carolinianus	LHCSR Phaeoceros carolinianus	Bryophytes	205
LHCSR	Low_7	IGS.gm_8_00476 Chlorella variabilis	LHCSR Chlorella variabilis	Green Algae	201
LHCSR	Low_8	50637 Micromonas pusilla	LHCSR Micromonas pusilla	Green Algae	220
LHCSR	Low_8	XP_003079276.1 Ostreococcus tauri	LHCSR Ostreococcus tauri	Green Algae	208
LHCSR	Not 0	AER58182.1 Ulva linza	LHCSR Ulva linza	Green Algae	185
LHCSR	Not 0	ABD37894.1 Mesostigma viride	LHCSR Mesostigma viride	Green Algae	188
PsbS	High_1	VIT_218s0001g02740.1 Vitis vinifera	PsbS Vitis vinifera	Eudicots (Rosids)	214
PsbS	High_1	Gorai.002G222200.1 Gossypium raimondii	PsbS Gossypium raimondii	Eudicots (Rosids)	214
PsbS	High_1	PsbS Consensus High_1	PsbS Consensus High_1	*	212
PsbS	High_2	LOC_Os01g64960.1 Oryza sativa	PsbS1 Oryza sativa	Monocots	218
PsbS	High_2	Kalax.1050s0003.1.p Kalanchoe laxiflora	PsbS Kalanchoe laxiflora	Eudicots (Core)	212
PsbS	High_2	URDJ_scaffold_2000285 Amborella trichopoda	PsbS Amborella trichopoda	Basalmost Angiosperms	213
PsbS	High_2	OPDF_scaffold_2005802 Saururus cernuus	PsbS Saururus cernuus	Magnoliids	213
PsbS	High_2	PsbS Consensus High_2	PsbS Consensus High_2	*	213
PsbS	High_3	EGOS_scaffold_2083970 Allionia incarnata	PsbS Allionia incarnata	Eudicots (Core)	213
PsbS	High_3	Spov3_chr5.00161 Spinacia oleracea	PsbS Spinacia oleracea	Eudicots (Core)	212
PsbS	High_3	PsbS Consensus High_3	PsbS Consensus High_3	*	212
PsbS	High_4	Lus10025676 Linum usitatissimum	PsbS Linum usitatissimum	Eudicots (Rosids)	212

PsbS	High_4	AT1G44575 Arabidopsis thaliana Araport	PsbS_Arabidopsis thaliana	Eudicots (Rosids)	213
PsbS	High_4	PsbS_Consensus_High_4	PsbS_Consensus_High_4	*	212
PsbS	High_5	UEVI_scaffold_2060015 Fokienia hodginsii	PsbS_Fokienia hodginsii	Gymnosperms	223
PsbS	High_5	YLPM_scaffold_2081596 Pseudotaxus chienii	PsbS_Pseudotaxus chienii	Gymnosperms	221
PsbS	High_5	PsbS_Consensus_High_5	PsbS_Consensus_High_5	*	224
PsbS	High_6	EJBY_scaffold_2042976 Anticharis glandulosa	PsbS_Anticharis glandulosa	Eudicots (Asterids)	213
PsbS	High_6	Lsat_1_v5_gn_6_35480.1 Lactuca sativa	PsbS_Lactuca sativa	Eudicots (Asterids)	212
PsbS	High_6	PsbS_Consensus_High_6	PsbS_Consensus_High_6	*	212
PsbS	High_7	DN57662_c0_g1_i1.p1 Astrolepis windhamii	PsbS_Astrolepis windhamii	Ferns	218
PsbS	High_7	YQEC_scaffold_2074641 Woodsia ilvensis	PsbS_Woodsia ilvensis	Ferns	217
PsbS	High_7	DN36247_c0_g1_i7.p2 Pleopeltis polypoides	PsbS_Pleopeltis polypoides	Ferns	220
PsbS	High_7	BMJR_scaffold_2008197 Adiantum raddianum	PsbS_Adiantum raddianum	Ferns	217
PsbS	High_7	PsbS_Consensus_High_7	PsbS_Consensus_High_7	*	218
PsbS	High_8	DHWX_scaffold_2073610 Fontinalis antipyretica	PsbS_Fontinalis antipyretica	Bryophytes	228
PsbS	High_8	JMXW_scaffold_2041479 Bryum argenteum	PsbS_Bryum argenteum	Bryophytes	215
PsbS	High_8	PsbS_Consensus_High_8	PsbS_Consensus_High_8	*	216
PsbS	High_9	Sobic.003G370000.1.p Sorghum bicolor	PsbS_Sorghum bicolor	Monocots	217
PsbS	High_9	NP_001105228.2 Zea mays	PsbS_Zea mays	Monocots	217
PsbS	High_9	PsbS_Consensus_High_9	PsbS_Consensus_High_9	*	205
PsbS	High_10	GNPX_scaffold_2007565 Oxera neriifolia	PsbS_Oxera neriifolia	Eudicots (Asterids)	216
PsbS	High_10	BAHE_scaffold_2024222 Solenostemon scutellarioides	PsbS_Solenostemon scutellarioides	Eudicots (Asterids)	211
PsbS	High_10	PsbS_Consensus_High_10	PsbS_Consensus_High_10	*	214
PsbS	High_11	ZQRI_scaffold_2066360 Timmia austriaca	PsbS_Timmia austriaca	Bryophytes	217
PsbS	High_11	Pp3c20_23430V3.1.p Physcomitrium patens	PsbS_Physcomitrium patens	Bryophytes	216
PsbS	High_11	PsbS_Consensus_High_11	PsbS_Consensus_High_11	*	217
PsbS	High_12	AWQB_scaffold_2001714 Picea engelmannii	PsbS_Picea engelmannii	Gymnosperms	220
PsbS	High_12	IOVS_scaffold_2005335 Pseudotsuga wilsoniana	PsbS_Pseudotsuga wilsoniana	Gymnosperms	221
PsbS	High_12	PsbS_Consensus_High_12	PsbS_Consensus_High_12	*	217
PsbS	High_13	Cre01.g016750 Chlamydomonas reinhardtii	PsbS_Chlamydomonas reinhardtii	Green Algae	219
PsbS	High_13	Vocar.0007s0069.1 Volvox carteri	PsbS_Volvox carteri	Green Algae	211
PsbS	High_13	PsbS_Consensus_High_13	PsbS_Consensus_High_13	*	208
PsbS	Low_14	GAQ81861.1 Klebsormidium nitens	PsbS_Klebsormidium nitens	Green Algae	206
PsbS	Low_14	WSJO_scaffold_2045470 Mesotaenium braunii	PsbS_Mesotaenium braunii	Green Algae	208
PsbS	Low_14	HAOX_scaffold_2024336 Spirogyra sp.	PsbS_Spirogyra sp.	Green Algae	208
PsbS	Low_14	STKJ_scaffold_2009269 Cosmarium ochthodes	PsbS_Cosmarium ochthodes	Green Algae	200
PsbS	Low_14	Traes_3B_D88E04B94.1 Triticum aestivum	PsbS_Triticum aestivum	Monocots	216
PsbS	Low_14	JPYU_scaffold_2004029 Marchantia polymorpha	PsbS_Marchantia polymorpha	Bryophytes	216
PsbS	Low_14	ZYCD_scaffold_2041681 Selaginella acanthonota	PsbS_Selaginella acanthonota	Lycophytes	214

PsbS	Low_15	XP_027942273.1 Vigna_unguiculata	PsbS_Vigna_unguiculata	Eudicots (Rosids)	213
PsbS	Low_15	RXRQ_scaffold_2135870 Phaeoceros_carolinianus	PsbS_Phaeoceros_carolinianus	Bryophytes	207
PsbS	Low_15	LOC_Os04g59440.1 Oryza_sativa	PsbS2_Oryza_sativa	Monocots	214
PsbS	Low_15	ABC59515.1 Nicotiana_benthamiana	PsbS1_Nicotiana_benthamiana	Eudicots (Asterids)	215
PsbS	Low_15	ABC59516.1 Nicotiana_benthamiana	PsbS2_Nicotiana_benthamiana	Eudicots (Asterids)	215
PsbS	Low_16	AER58182.1 Ulva_linza AFW18037.1 Ulva_prolifera	PsbS_Ulva_linza	Green Algae	203
PsbS	Low_16	140514 Chlorella_variabilis	PsbS_Chlorella_variabilis	Green Algae	211
PsbS	Low_16	Dusal.0048s00033.1.p Dunaliella_salina	PsbS_Dunaliella_salina	Green Algae	216
PsbS	Not_0	Cz03g01250.t1 Chromochloris_zofingiensis	PsbS1_Chromochloris_zofingiensis	Green Algae	220
PsbS	Not_0	OUS48235.1 Ostreococcus_tauri	PsbS_Ostreococcus_tauri	Green Algae	195
PsbS	Not_0	Pavir.5KG661100.1.p Panicum_virgatum	PsbS_Panicum_virgatum	Monocots	249
PsbS	Not_0	Cz04g24050.t1 Chromochloris_zofingiensis	PsbS2_Chromochloris_zofingiensis	Green Algae	253
PsbS	Not_0	GDKF01008623.1_frame+3_#1 Auxenochlorella_protothecoides	PsbS_Auxenochlorella_protothecoides	Green Algae	204
LHCB	*	Cre16.g673650 CrLHCB5_CP26	CrCP26	Green Algae	280
LHCB	*	Cre17.g720250 CrLHCB4_CP29	CrCP29	Green Algae	280
LHCB	*	Cre01.g066917 CrLHCBM1	CrLHCBM1	Green Algae	232
LHCB	*	Cre12.g548400 CrLHCBM2	CrLHCBM2	Green Algae	222
LHCB	*	Cre04.g232104 CrLHCBM3	CrLHCBM3	Green Algae	228
LHCB	*	Cre06.g283950 CrLHCBM4	CrLHCBM4	Green Algae	239
LHCB	*	Cre03.g156900 CrLHCBM5	CrLHCBM5	Green Algae	239
LHCB	*	Cre06.g285250 CrLHCBM6	CrLHCBM6	Green Algae	239
LHCB	*	Cre12.g548950 CrLHCBM7	CrLHCBM7	Green Algae	222
LHCB	*	Cre06.g284250 CrLHCBM8	CrLHCBM8	Green Algae	227
LHCB	*	Cre06.g284200 CrLHCBM9	CrLHCBM9	Green Algae	239

4.5.6 NPQ by fluorescence plots mirror stable in planta phenotypes

The ability to resolve significant NPQ phenotypes is dependent on sufficient throughput and quantitative resolution. To test the ability of this system to generate reproducible and correlative phenotypes, *Arabidopsis* and *Chlamydomonas* PsbS were first tested for their relative capacity to induce NPQ *in planta*.

Infiltration in the *npq4* mutant provides a relatively noiseless background to quantify increases in qE relative to WT biological variation (**Figure 4-7A**). From the kinetics of WT and *npq4*, two infiltrated LHCSR/PsbS-dependent parameters were selected: the average gain in magnitude of qE (MqE) within the first 2 minutes of high-light exposure, and the average gain in magnitude of total NPQ (MqTot) after stabilization of NPQ capacity in the second high-light phase, both relative to the *npq4* background mutant. These two parameters were selected to capture the potential contributions of rapidly inducible Δ pH-dependent NPQ and additive, longer-acting Zx-dependent NPQ, respectively. However, it was necessary to have a robust baseline *in vivo* to confirm the consistency in the observed transient NPQ phenotypes.

Conveniently Niyogi et al., 2004¹¹⁹ reported NPQ data, analogous to MqTot in this study, of hemizygous *Arabidopsis npq4* mutants complemented with PsbS across a broad phenotypic and expression range (**Figure 4-7B**). Interestingly, they note a logarithmic shape in PsbS dosage by immunoblot with NPQ capacity. This trend closely mirrors *Arabidopsis* PsbS transient expression data for MqE (**Figure 4-7C**) and MqTot (**Figure 4-7D**) despite use of a stoichiometric fluorescence reporter rather than quantification of PsbS directly. Interestingly, the capacity for *Chlamydomonas* PsbS to induce NPQ is significantly lower for both parameters, showcasing the utility and resolution of this approach. As shown by the intersection of infiltrated datapoints (yellow) and the phenotypic range of WT leaves (green), transient expression could reach and exceed native levels of PsbS MqE and MqTot activity. Altogether, this validation work encouraged extensive screening of our LHCSR and PsbS library.

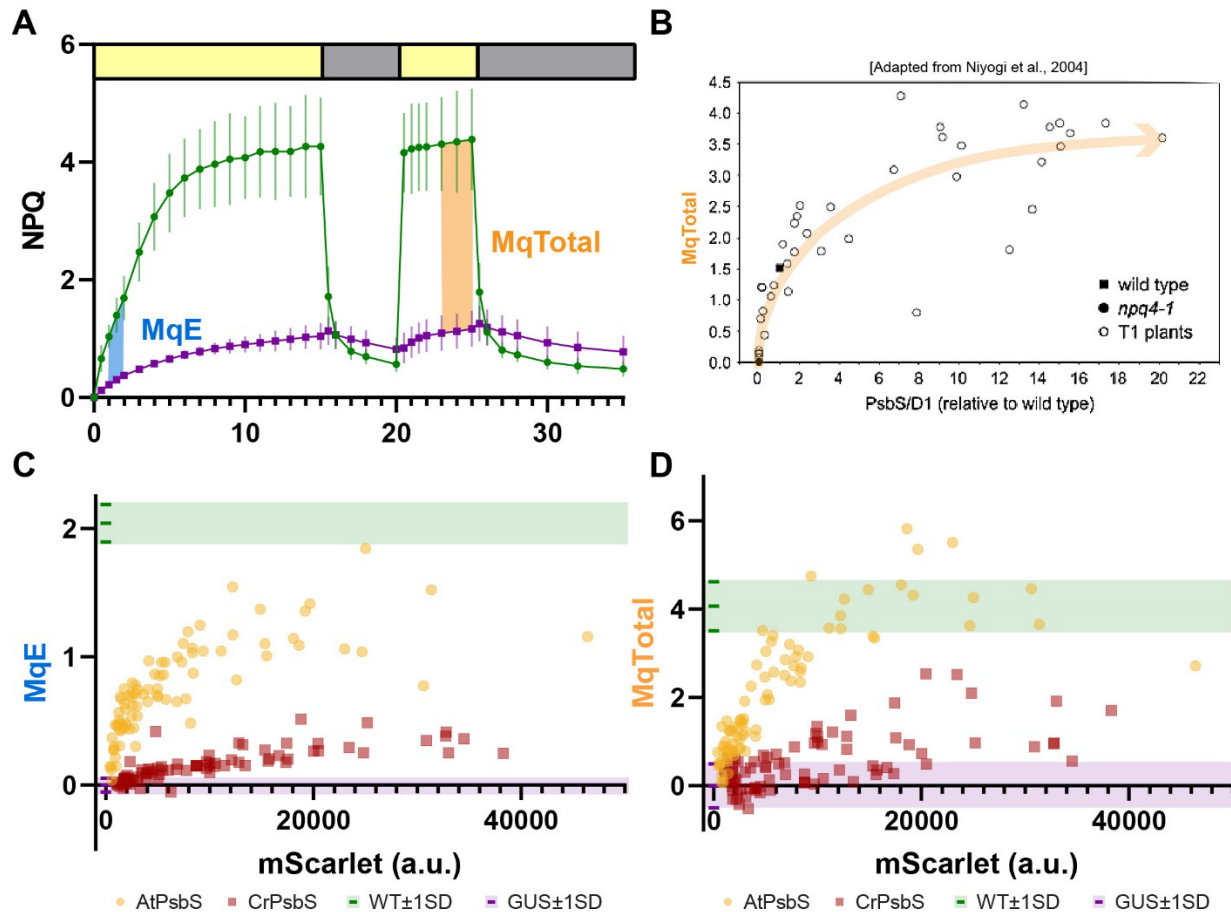


Figure 4-7. Quantitative transient expression of *Arabidopsis* *PsbS* recapitulates phenotypes previously observed in stable *Arabidopsis* transformants.

(A) Schematic of fluctuating light regime used to phenotype transiently expressed leaf samples (15 min $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ – 5 min $0 \mu\text{mol m}^{-2} \text{s}^{-1}$ – 5 min $1500 \mu\text{mol m}^{-2} \text{s}^{-1}$ – 10 min $0 \mu\text{mol m}^{-2} \text{s}^{-1}$), with periods of actinic light marked in yellow and periods of darkness marked in gray for WT (green) and *npq4* (purple) tobacco plants, data shown ± 1 SEM. The two parameters used in this study are marked, MqE (blue, difference in rapidly-inducible NPQ capacity) and MqTotal (orange, difference in total NPQ capacity) induced by the transiently expressed constructs. **(B)** Figure adapted from Niyogi et al., 2004¹¹⁹ describing the difference in MqTotal in T₁ plants with varying *PsbS*/D1 as determined by immunoblot. **(C)** MqE and **(D)** MqTotal as a factor of mScarlet fluorescence (570/605) for *Arabidopsis* *PsbS* (AtPsbS, yellow circle) and *Chlamydomonas* *PsbS* (CrPsbS, red square). Average ± 1 SD values for WT (green) and the *npq4*+GUS negative control (purple) are shown as colored bars.

All 22 LHCSR and 63 *PsbS* were infiltrated in biological replicates ($n=4-6$ each) across 5 independent experiments, with each leaf disc phenotyped on 2-3 consecutive days to capture a greater dynamic range in expression and phenotypes post-infiltration. In total, NPQ traces for over 5000 leaf spots were collected and plotted against mScarlet on a log scale (**Figure 4-8**). NPQ was marked as significantly different (red) if its value exceeded $+1$ SD of the internal GUS negative control (gray). Almost all the observed phenotypic variation in MqE and MqTot arose from constructs expressing various *PsbS*.

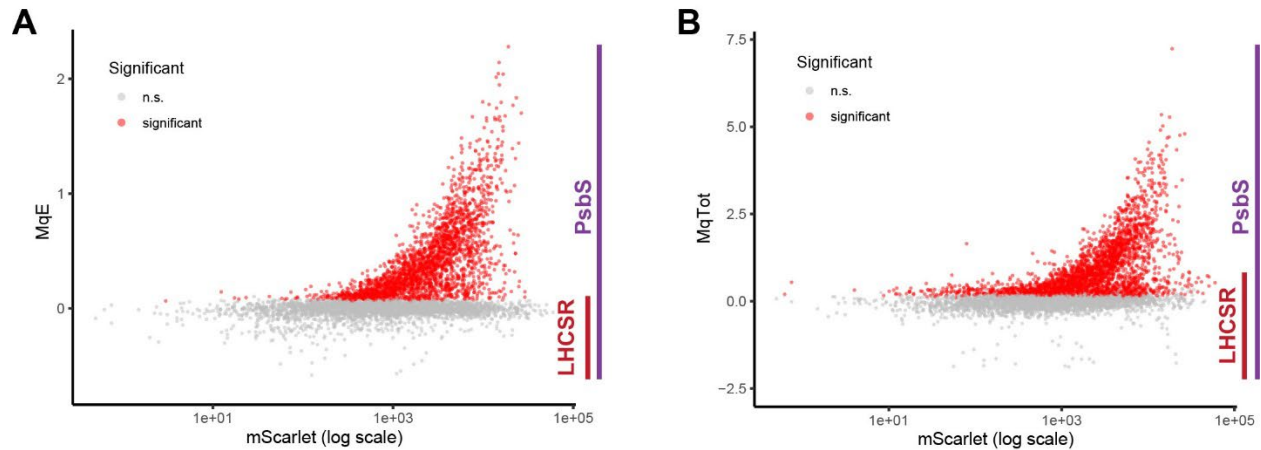


Figure 4-8. Aggregate *MqE* and *MqTotal* phenotypes for all infiltrated *LHCSR* and *PsbS*.

(A) *MqE* and **(B)** *MqTotal* (*MqTot*) on a log scale of *mScarlet* fluorescence (570/605). Infiltrated leaf discs with phenotypes within $\pm 1SD$ of all *npq4*+*GUS* negative control replicates ($n=240$) are shown in gray. Significant NPQ phenotypes are plotted on a linear access and marked in red. Purple (*PsbS*) and Red (*LHCSR*) delineate the boundaries of NPQ phenotypes observed. A total of 5760 infiltrated leaf spots were collected across the 85 *LHCSR* and *PsbS* constructs (data shown $n > 5000$ after filtering by $F_v/F_m > 0.72$ threshold).

The visualization of this data allows for relatively intuitive comparisons of *MqE* and *MqTot* relative to the quenching capacity of all infiltrated leaf discs. In the next several subsections, I briefly emphasize interesting phenotypes before putting them into the context of our larger sequence diversity. Aside from one interesting case study within the *LHCSR* dataset, the bulk of the phenotypes described will focus on the role of diverse *PsbS* in modulating NPQ *in planta*.

4.5.7 Streptophyte *LHCSR* exhibit slowly inducible NPQ capacity

Streptophyte green algae form a subclade that are more sister to bryophytes and other land plants than chlorophyte green algae like *Chlamydomonas reinhardtii*. Given their relative evolutionary proximity, we wondered if *LHCSR* from these green algal species would be competent for NPQ *in planta*. *LHCSR* from three representative streptophytes showed no ability to induce *MqE*. However, all 3 showed a very modest but statistically significant ability to induce *MqTot* at high levels of *mScarlet* (**Figure 4-9**).

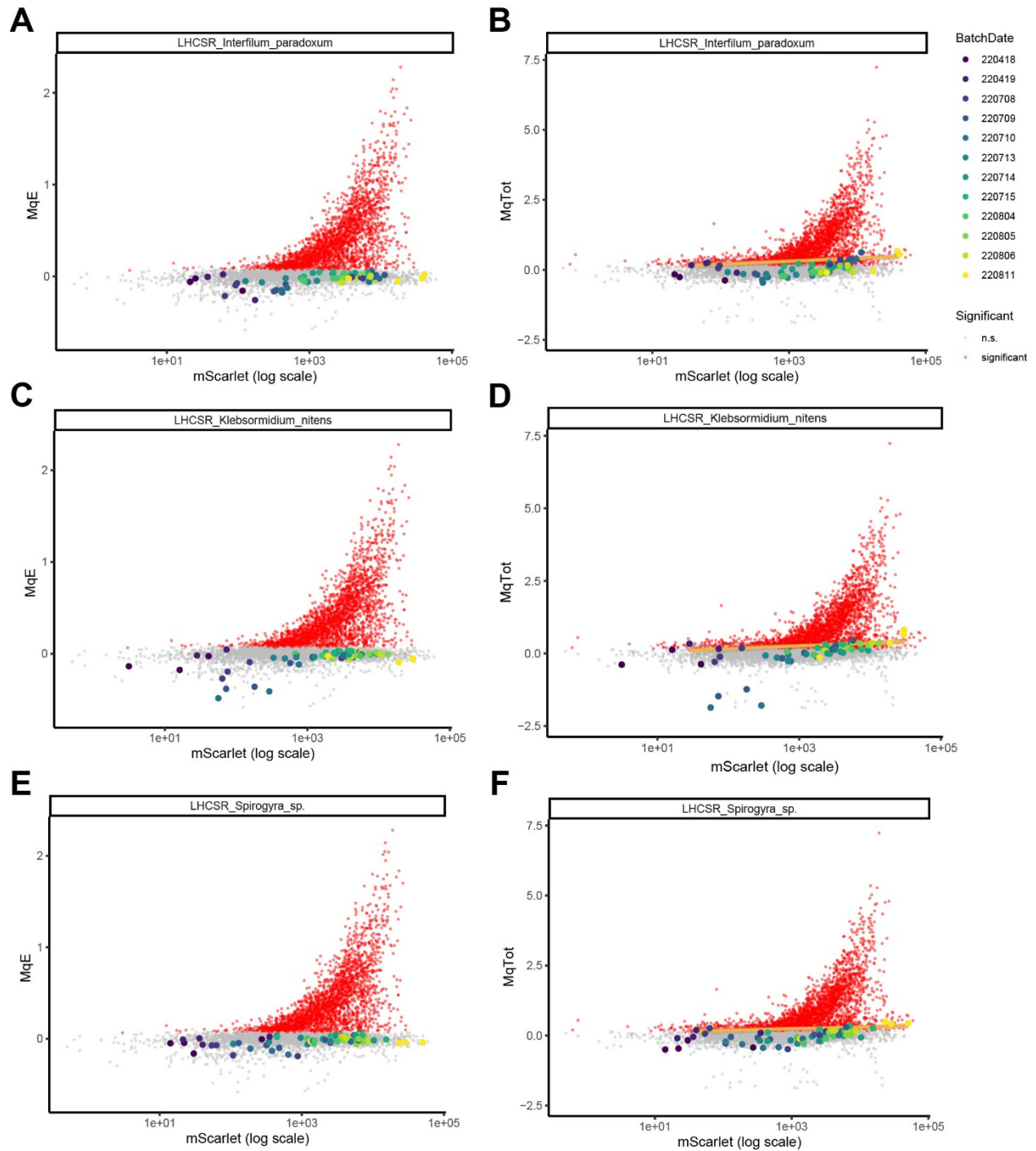


Figure 4-9. Streptophyte LHCSR show a weak ability to induce MqTotal but not MqE.

(A, C, E) MqE and (B, D, F) MqTotal of LHCSR from 3 streptophyte algae from the Klebsormidiaceae (*Interfilum paradoxum*, *Klebsormidium nitens*) and Zygnematophyceae (*Spirogyra sp.*). Data points are color-coded by experiment batch date and overlaid on top of all aggregate LHCSR and PsbS expression data. A linear regression was drawn for all data points over the significance threshold for MqE (left, blue) or MqTotal (right, orange) (> 1SD over npq4+GUS negative control).

4.5.8 Nicotiana benthamiana PsbS is not the most efficient quencher in planta

It is not unreasonable to assume that the native PsbS (NbPsbS1 or NbPsbS2) is best adapted for quenching in our transient expression system. Interestingly, both NbPsbS1 and NbPsbS2 show moderate capacity to induce MqE and MqTot relative to the total phenotypic variation, suggesting those sequences do not hold an inherent property for high native NPQ. In fact, PsbS from the liverwort *Marchantia polymorpha* showed the highest ability to rapidly induce MqE, sharing a high but less competitive capacity for MqTot (**Figure 4-10**). Robust phenotypes observed across multiple experimental batch dates further reinforced the resolution and sensitivity possible in this quantitative transient expression system.

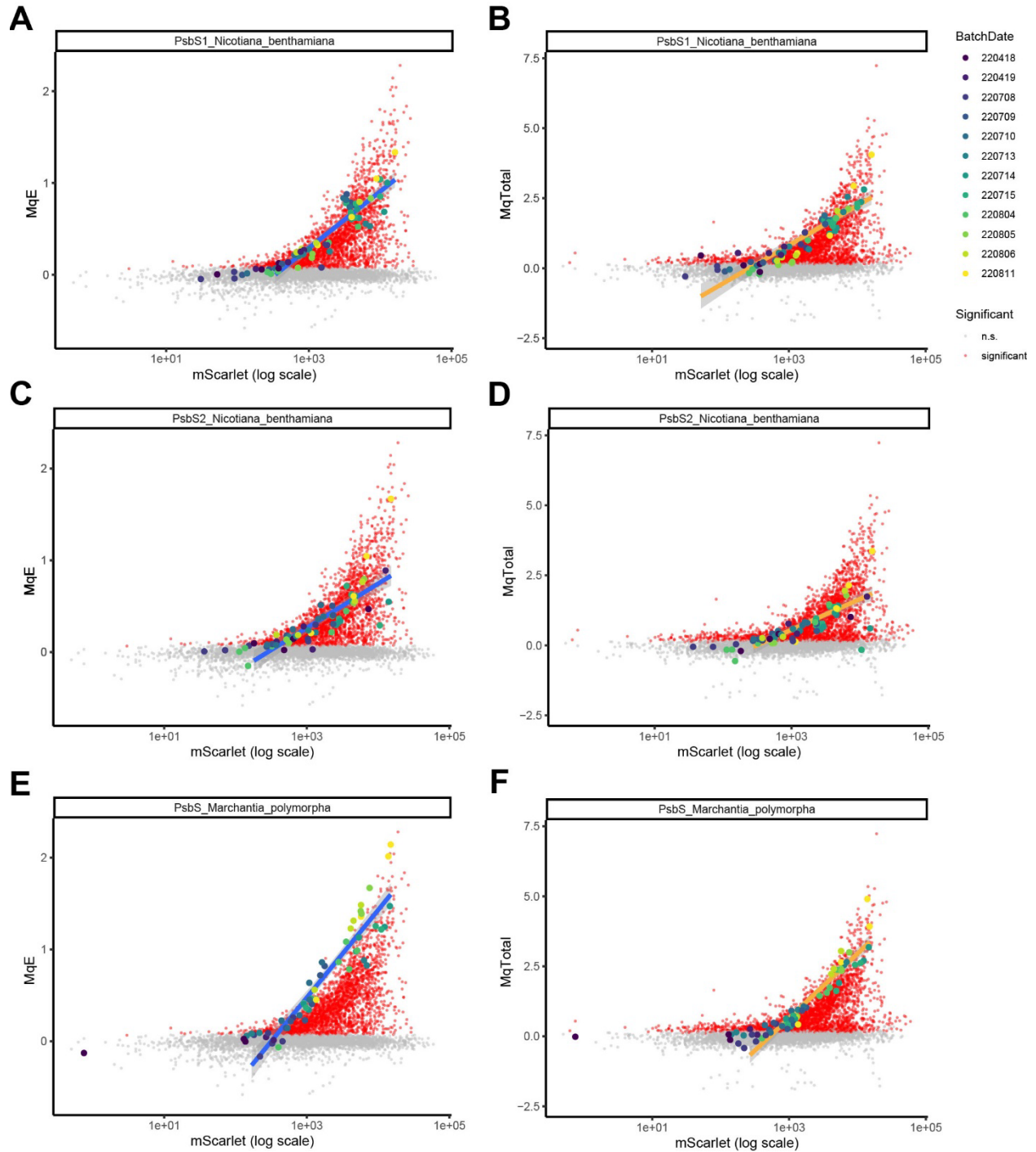


Figure 4-10. Native *N. benthamiana* PsbS is not the most efficient PsbS at quenching in planta

(A, C, E) MqE and (B, D, F) MqTotal of *N. benthamiana* PsbS1, *N. benthamiana* PsbS2, and the highest magnitude MqE PsbS from the liverwort *Marchantia polymorpha*. Data points are color-coded by experiment batch date and overlaid on top of all aggregate LHCSR and PsbS expression data. A linear regression was drawn for all data points over the significance threshold for MqE (left, blue) or MqTotal (right, orange) ($> 1SD$ over npq4+GUS negative control).

4.5.9 Monocot PsbS perform poorly despite high native activity

Monocot species like *Sorghum bicolor* are well adapted to growth in excess light conditions. Additionally, (**Chapter 3**) described robust OsPsbS-dependent capacity for photoprotection in rice. However despite this high *in vivo* capacity, monocot PsbS from rice (**Figure 4-11A+B**), sorghum (**Figure 4-11C+D**), and the synthetic “monocot” PsbS from cluster High_9 (**Figure 4-11E+F**) all have substantially lower MqE and MqTot than a majority of the infiltrated population.

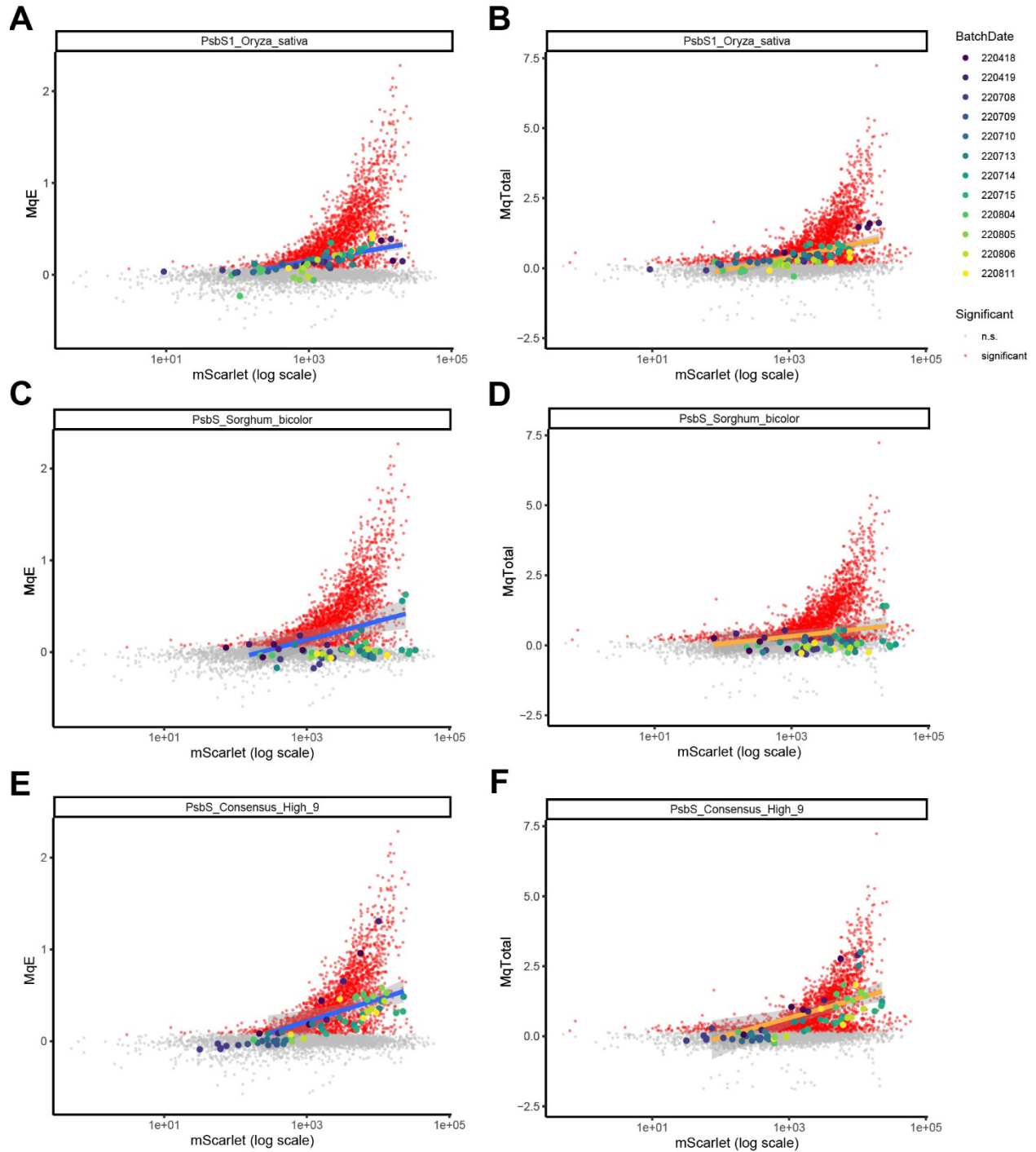


Figure 4-11. Monocot *PsbS* induce NPQ poorly relative to the entire LHCSR/*PsbS* library.

(A, C, E) MqE and (B, D, F) MqTotal of three representative monocots: *O. sativa* PsbS1, *S. bicolor* PsbS, and a synthetic PsbS consensus sequence generated from the monocot PsbS cluster (PsbS_High9). Data points are color-coded by experiment batch date and overlayed on top of all aggregate LHCSR and PsbS expression data. A linear regression was drawn for all data points over the significance threshold for MqE (left, blue) or MqTotal (right, orange) (> 1SD over npq4+GUS negative control).

4.5.10 Pairwise comparisons may highlight important functional/structural residues

The last remaining monocot PsbS from Cluster High_9 was isolated from *Zea mays*. In stark contrast with the other described monocot PsbS, we see a robust capacity for MqE and MqTot (**Figure 4-12A+B**) that rivals that of the native NbPsbS1 and NbPsbS2 proteins (**Figure 4-10**). Superimposed on this graph are the fitted slopes of *S. bicolor* MqE and MqTot (dashed lines), which surprisingly only differ in 4 amino acids (**Figure 4-12C**). Similar pairwise comparisons may be additionally fruitful in identifying sites necessary for quenching in tobacco.

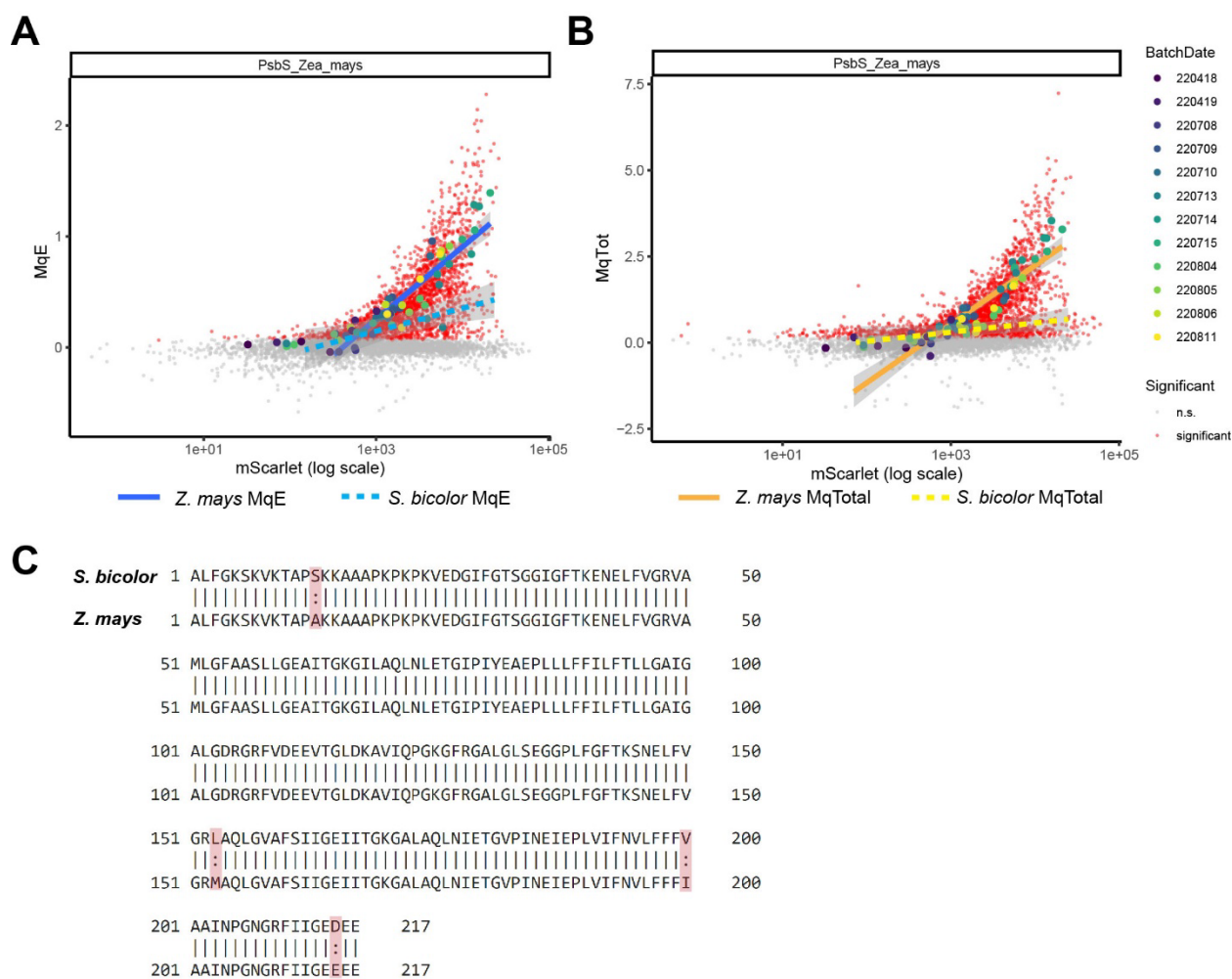


Figure 4-12. PsbS from the monocot *Zea mays* is atypical in quenching capacity relative to other monocot PsbS.

(**A**) MqE and (**B**) MqTotal of *Zea mays* PsbS. Superimposed on this graph are the linear regressions of MqE (light blue, dashed) and MqTotal (yellow, dashed) from the highly homologous *S. bicolor* PsbS (raw data shown in **Figure 4-11**). Data points for ZmPsbS are color-coded by experiment batch date and overlayed on top of all aggregate LHCSR and PsbS expression data. A linear regression was drawn for all data points over the significance threshold for MqE (left,

blue) or MqTotal (right, orange) ($> 1SD$ over npq4+GUS negative control). **(C)** Sequence alignment of the cTP-trimmed SbPsbS and ZmPsbS sequences, with the 4 diverging amino acids highlighted in red.

4.5.11 A rice “pseudogene” is better adapted for quenching in dicots

Another surprising finding arose from the functional activity of rice PsbS2, a validated pseudogene that does not contribute to rice NPQ *in vivo*¹⁰³. In this experimental system, OsPsbS2 had significantly higher MqE and MqTot than the dominant *in vivo* ortholog OsPsbS1 (**Figure 4-13A+B**). While pairwise comparisons are less clear-cut than those that differentiate sorghum and maize PsbS (**Figure 4-13C**), the presence of both OsPsbS2 and NbPsbS1 in the same cluster (Low_15) may indicate shared sequences that allow for functional NPQ capacity in tobacco. Publicly available gene expression datasets show the *OsPsbS2* “pseudogene” is expressed, though at lower levels relative to *OsPsbS1*, in rice leaves (**Figure 4-13D**). Understanding whether OsPsbS2 protein functionally accumulates *in vivo* in either or both systems would provide interesting insight into the structural requirements necessary for native and transient NPQ activity.

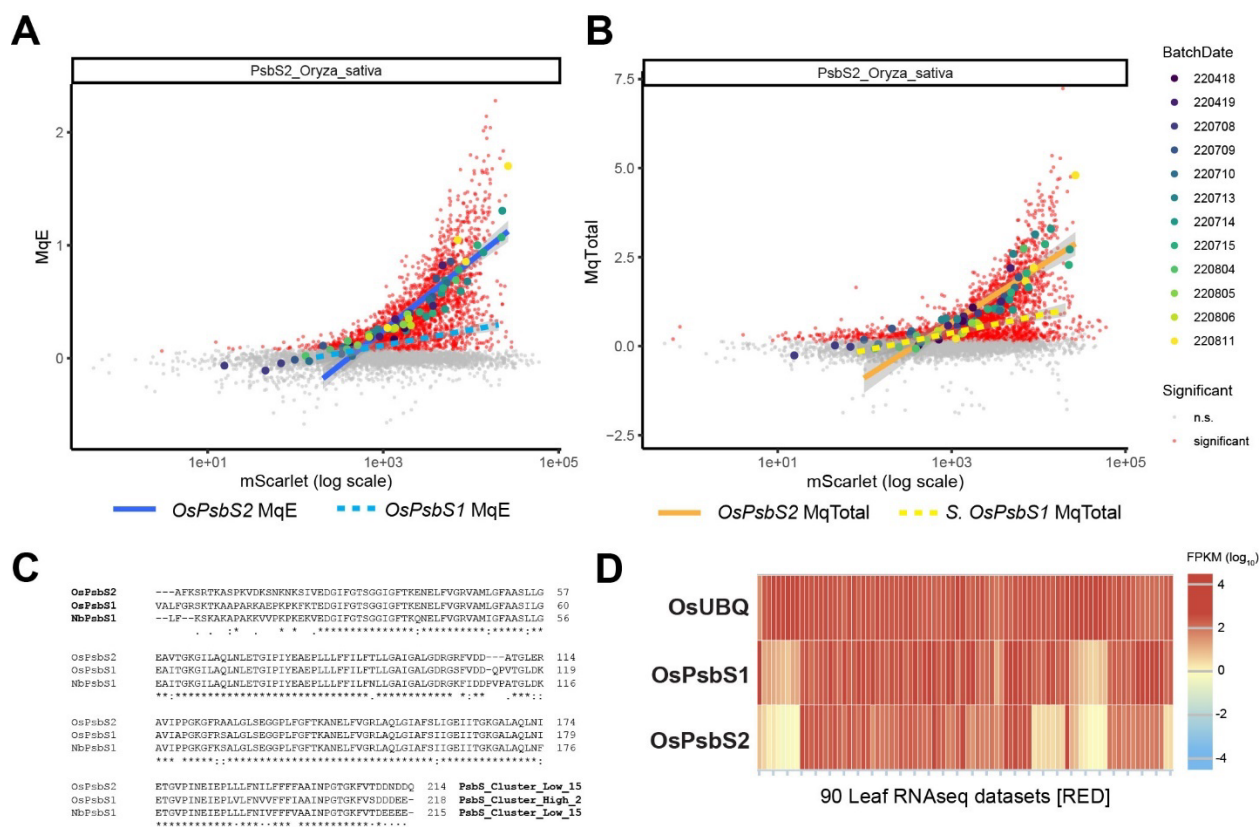


Figure 4-13. A rice “pseudogene” exhibits higher NPQ capacity per fluorescence than its dominant *in vivo* ortholog.

(A) MqE and **(B)** MqTotal of *Oryza sativa* PsbS2. Superimposed on this graph are the linear regressions of MqE (light blue, dashed) and MqTotal (yellow, dashed) from the dominant *in vivo* rice ortholog OsPsbS1 (raw data shown in **Figure 4-11**). A linear regression was drawn for all data points over the significance threshold for MqE (left, blue) or

MqTotal (right, orange) (> 1SD over npq4+GUS negative control). **(C)** Sequence alignment of the cTP-trimmed OsPsbS2, OsPsbS1, and NbPsbS1 found in PsbS clusters Low_15, High_2, and Low_15, respectively. **(D)** Endogenous Log₁₀ FPKM data of *OsUBQ* (LOC_Os03g13170) relative to *OsPsbS1* (LOC_Os01g64960) and *OsPsbS2* (LOC_Os04g59440) across 90 independent RNAseq datasets from the Rice Expression Database¹⁶⁷.

4.5.12 Land-plant like PsbS first arose in streptophyte algae

With similar motivation from our assessment of streptophyte LHCSR-dependent quenching *in planta* (**Figure 4-14**), we sought to interrogate the effect on NPQ of expressing streptophyte PsbS in a land plant. An updated guide tree (**Figure 4-14A**) positions several screened streptophyte species, including *Mesostigma* which lacks PsbS entirely and the Bryophytes which, in *Physcomitrium patens*, have been shown to have both PsbS and LHCSR-dependent NPQ. As species become evolutionarily closer to land plants, we see marked increases in MqE capacity within chlorophyte and streptophyte algae (**Figure 4-14B-E**). In fact, PsbS from the Zygnematophyte *Spirogyra* (the streptophyte lineage most closely related to land plants¹⁵⁹) reaches even higher MqE levels than tobacco (**Figure 4-10**).

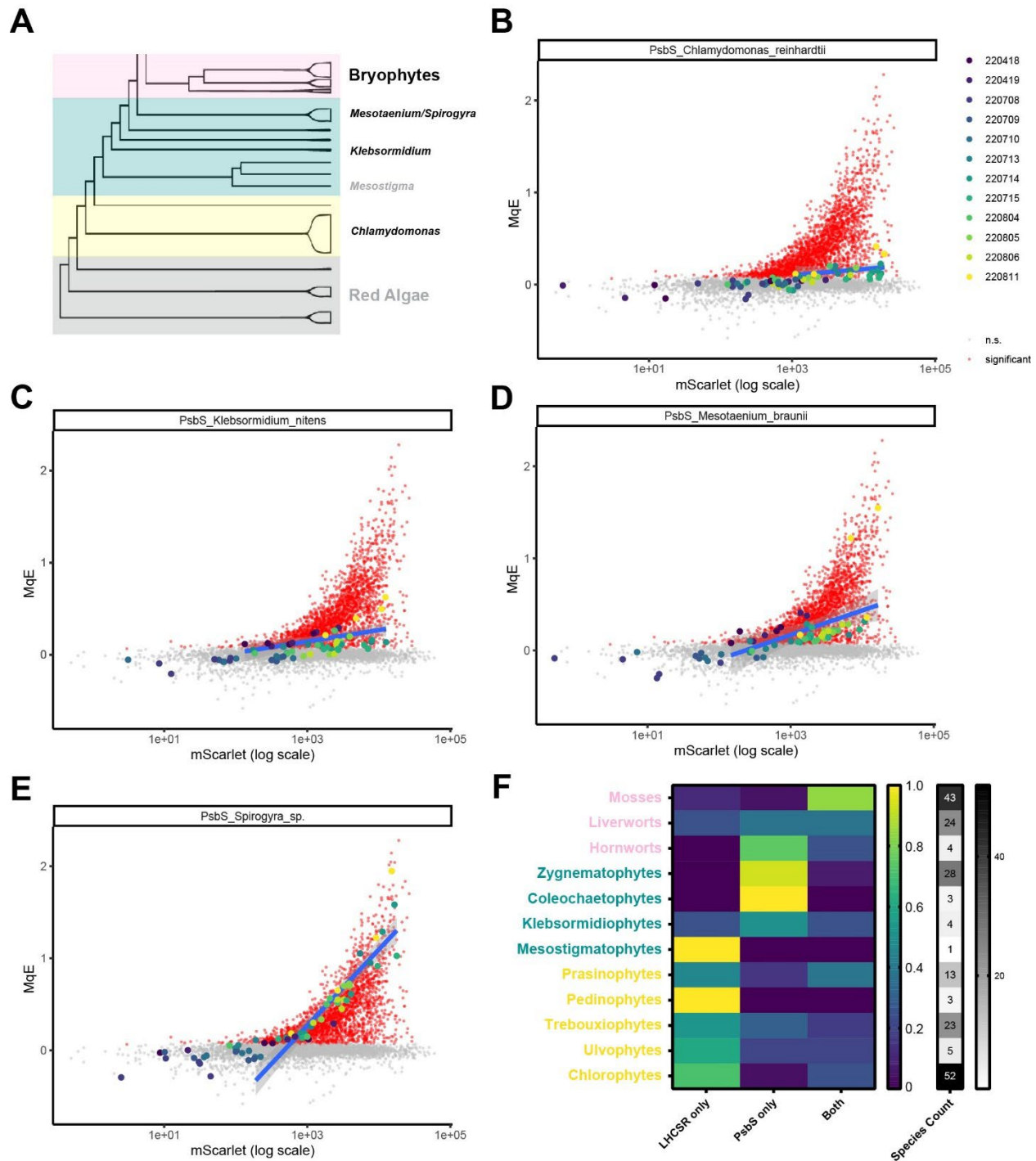


Figure 4-14. Cross green algal *PsbS* analysis identifies robust land plant-like activity in the *Zygnematophyceae*.

(A) Dendrogram of green algal intra-diversity separated into streptophyte (teal) and chlorophyte (yellow) green algae. Branching determined by ASTRAL analysis of 410 single-copy nuclear gene families as described¹⁵⁹. A red algal outgroup and streptophyte algae lacking LHCPS (*Mesostigma viride*) are shown for context. **(B-E)** Representative MqE graphs ordered from most ancestral to most sister to Bryophytes: **(B)** *Chlamydomonas reinhardtii* **(C)** *Klebsormidium nitens* **(D)** *Mesotaenium braunii* **(E)** *Spirogyra sp.* **(F)** Heatmap showing the proportion of species screened with a high-confidence BLAST hit to an ortholog of LHCPS, PsbS, or both within a single class (left). Classes are color-coded

by origin: chlorophyte green algae (yellow), streptophyte green algae (teal), and bryophytes (pink). Number of species per clade is shown as a grayscale heatmap (right).

Cross-phyla analysis of LHCSR and PsbS orthologs across 200 mined species reveals an interesting trend: LHCSR's are ubiquitous in extant chlorophyte green algae (yellow). However, the dominant quencher clearly shifts to PsbS across late-diverging streptophyte green algae (teal). Mosses like *Physcomitrium patens*, which has been evidenced as a necessary LHCSR-to-PsbS "transition state", may be atypical in its co-dependence on both proteins *in vivo* (**Figure 4-14F**).

4.5.13 Multiple sequence alignments identify interesting NPQ patterns that require further statistical investigation

Pairwise comparisons, while satisfying, only provide a sliver of the observed functional diversity. To assess NPQ phenotypes in parallel, all cTP-trimmed sequences were aligned into a UPGMA dendrogram and plotted against their respective slopes of MqE and MqTot (**Figure 4-15**).

Across the tested sequences, we identify groups of PsbS largely segregated by phylogenetic clade. Non-streptophyte algal PsbS are especially divergent in terms of branch length, expectedly sharing a most recent common ancestor with streptophyte green algae. Of the streptophyte algae, *Spirogyra* exhibited the highest capacity for MqE and MqTot (**Figure 4-14**). Consistent with its local branch phenotypes, *Spirogyra* PsbS shares sequence and functional diversity with sister Ferns and Lycophytes (Selaginella).

In terms of NPQ, we see similar MqE and MqTot phenotypes between sequences separated by relatively short branch lengths (e.g. *Nicotiana benthamiana* PsbS1 and PsbS2) and PsbS in highly related phylogenetic clades (e.g. Eudicots (Asterids); Ferns). Contrasting phenotypes, like the differences between *Z. mays* and *S. bicolor*, or the differences between synthetic PsbS Consensus High 1 and High 2 may reveal individual amino acid substitutions with significant impacts on NPQ or protein stability.

Surprisingly, the correlation between MqE and MqTot does not appear to be linear. While *Marchantia* was revealed to have the highest MqE capacity tested (**Figure 4-10**), its capacity for MqTot is more intermediate against the remainder of the tree. This is in contrast to *Arabidopsis* PsbS which showed relatively low MqE capacity but had one of the highest MqTot slopes. This phenotype was also observed in (**Figure 4-7C+D**) with FLAG-tagged *Arabidopsis* PsbS—where AtPsbS-FLAG was unable to induce WT levels of MqE, but it exhibited far above WT levels of MqTot. Divergences between MqE and MqTot may prove interesting in understanding the functional differences underlying NPQ phenotypes between tested orthologs.

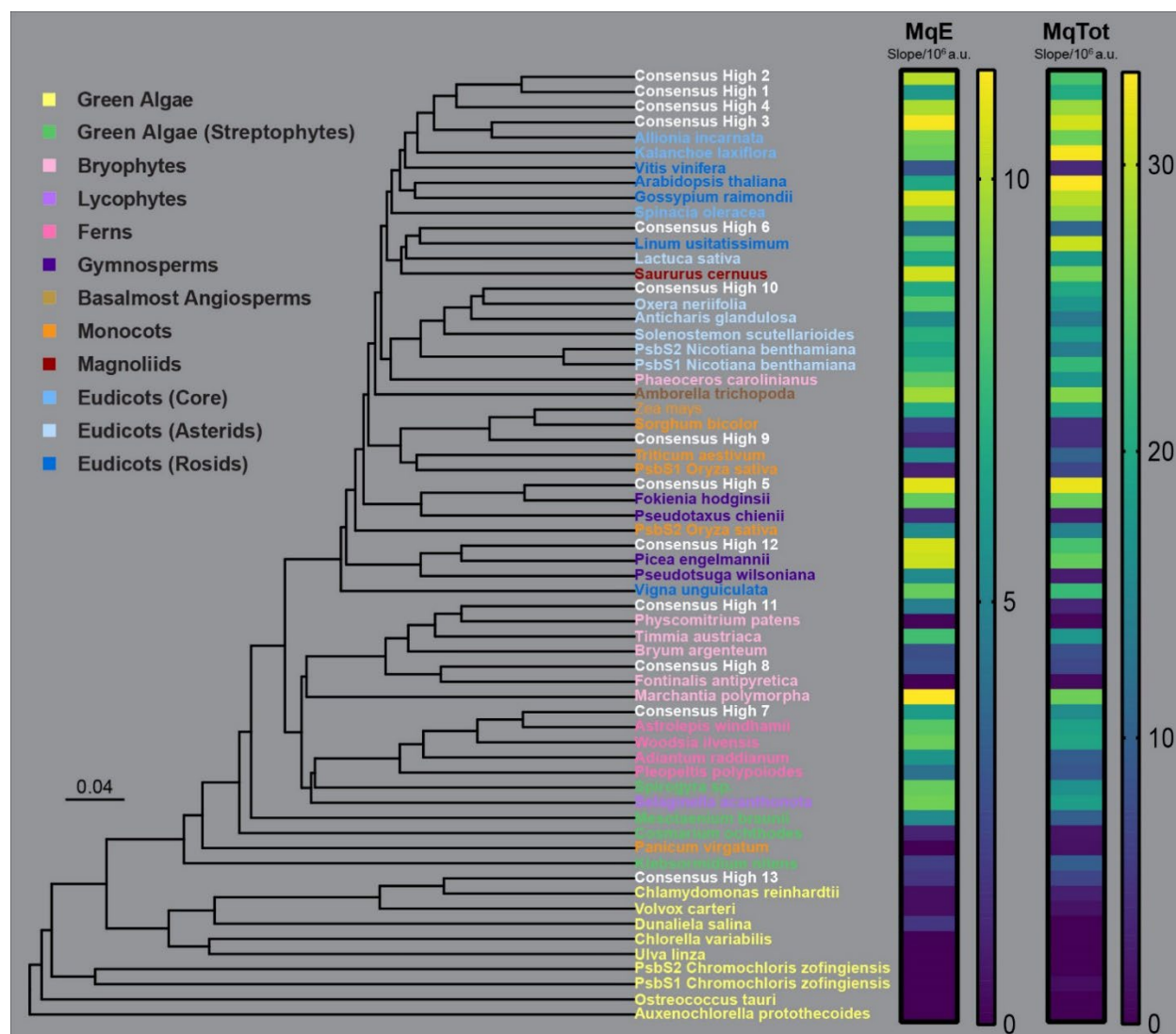


Figure 4-15. A UPGMA dendrogram of all tested PsbS relative to MqE and MqTot slopes.

A dendrogram of all tested PsbS is shown, rooted by the most divergent PsbS variant from *Auxenochlorella protothecoides*. Species are color-coded based on phylogenetic clade and are placed next to their relative MqE and MqTot capacities (Slope/10⁶ a.u. mScarlet), normalized to the highest value for each parameter. Synthetic consensus sequences are included in white.

Increased MSA resolution at the amino acid sequence level emphasized broader structural differences underlying phenotypes (**Figure 4-16**). There is an obvious step change in PsbS structural conservation between chlorophyte green algae (see cutoff at Consensus_High_13) and streptophyte green algae (see *Klebsormium*, *Cosmarium*, *Mesotaenium*). In fact, these 3 extant PsbS may represent sequence transition states, with a scattering of amino acid substitutions at conserved PsbS residues that are fixed following upon reaching the *Marchantia* PsbS branch.

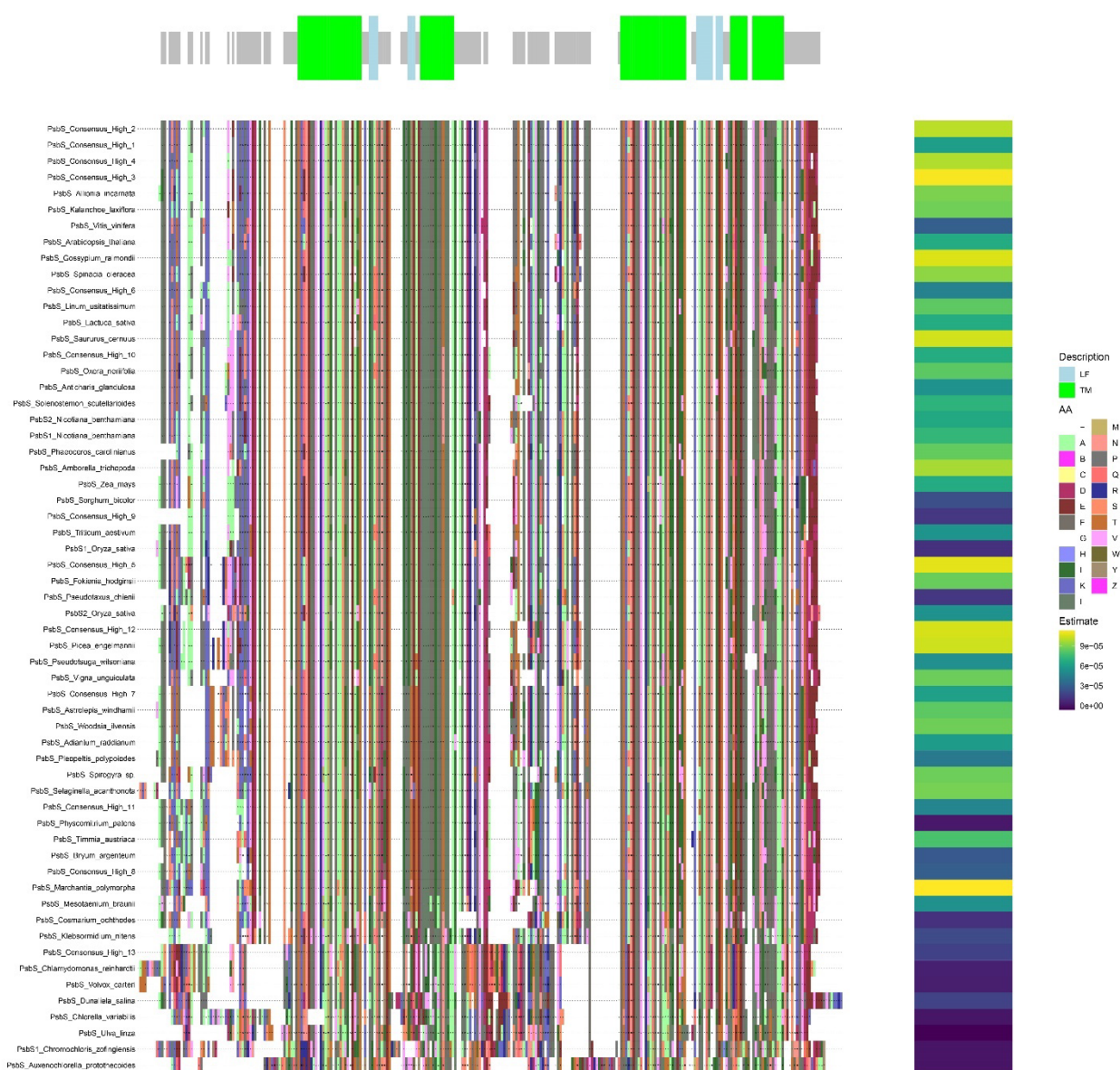


Figure 4-16. UPGMA MSA in dendrogram order of all tested PsbS against MqE.

Sequences are listed in dendrogram order (Figure 4-15), color-coded by amino acid sequence identity, and plotted next to the slope of MqE. Position of substitutions is shown relative to an AlphaFold predicted model of AtPsbS, with transmembrane helices in green, predicted "little folds" in blue, and AtPsbS amino acid sequences in gray.

To improve resolution and help identify candidate residues affecting quenching capacity *in vivo*, the largely non-functional chlorophyte green algae were removed (eliminating gaps from the MSA) and sequences were sorted from highest to lowest MqE (**Figure 4-17**). In doing so, we identify amino acid positions that appear somewhat flexible (i.e. only 2-3 different amino acids ever occupy that site). In contrast, areas with substantial

sequence diversity (e.g. the stomal loop between transmembrane domains 2+3) may be reservoirs of genetic and functional diversity.

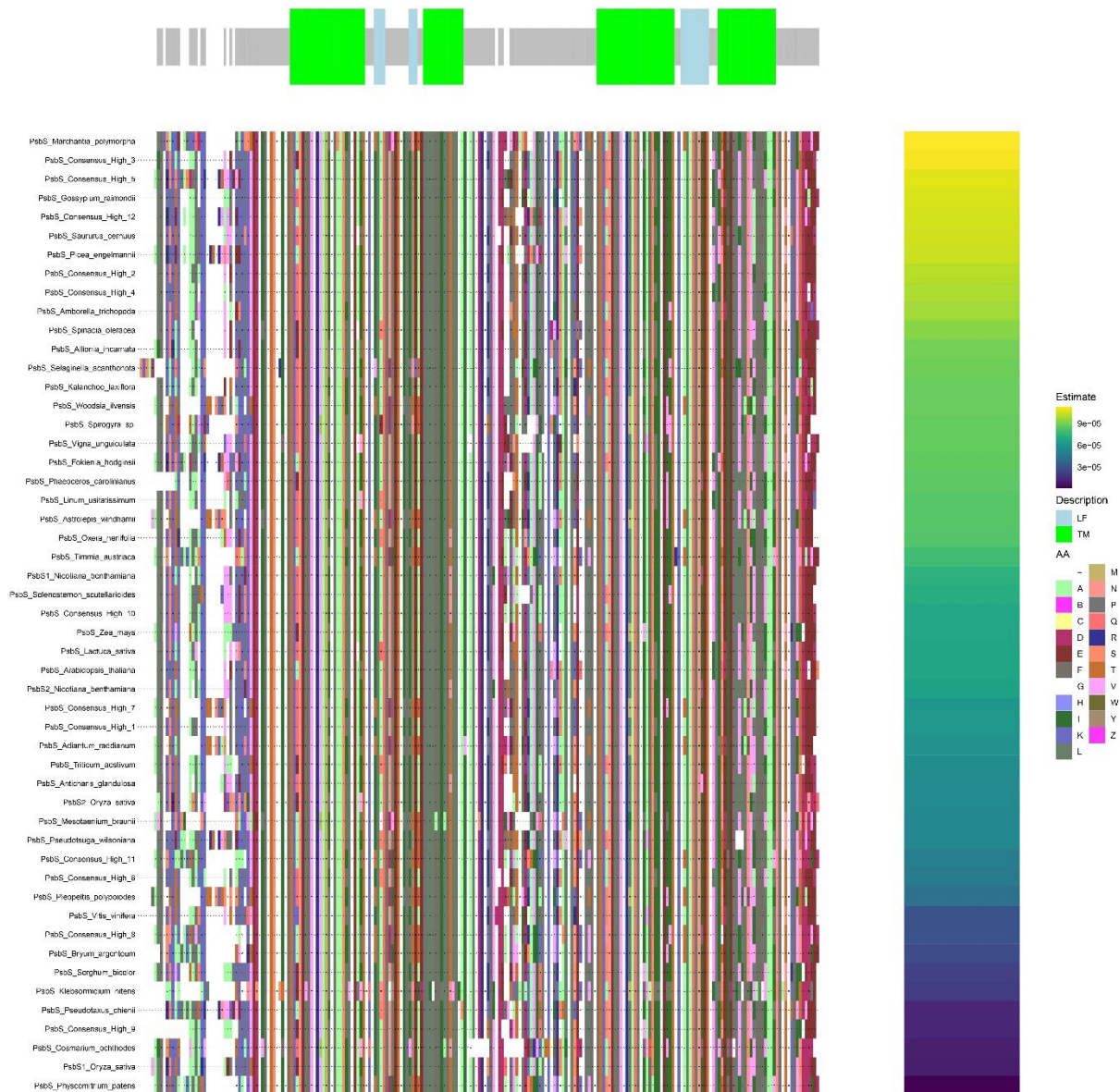


Figure 4-17. *PsbS* MSA of all amino acids in descending *MqE* order, excluding chlorophyte green algae

Sequences are color-coded by amino acid sequence identity and plotted next to the slope of *MqE*. Position of substitutions is shown relative to an AlphaFold predicted model of AtPsbS, with transmembrane helices in green, predicted "little folds" in blue and AtPsbS amino acid sequences in gray.

Reducing genetic complexity by plotting only protonatable residues (**Figure 4-18**) or reducing each amino acid to its characteristic hydrophathy (**Figure 4-19**) may help

identify variants of interest. For example, the former shows a remarkable conservation of the two protonatable residues (between TM1+2, and TM3+4) necessary for PsbS-dependent NPQ in *Arabidopsis*²⁷. However some PsbS are able to tolerate an additional protonatable residue, and in the case of *Spirogyra*, two protonatable residues between TM1+TM2. The implications of those changes on the kinetics of NPQ induction and relaxation, or ratio of MqE:MqTot warrant further investigation.

In the case of hydropathy, it is also possible to identify sites of local diversity. For example, at position 140 we are able to identify a residue that, when hydrophobic, correlates with increased MqE (**Figure 4-20**). More rigorous statistical testing and analysis of sites of adaptive evolution across gene libraries, may facilitate identification of variable and correlative sites that impact quenching capacity *in vivo*.

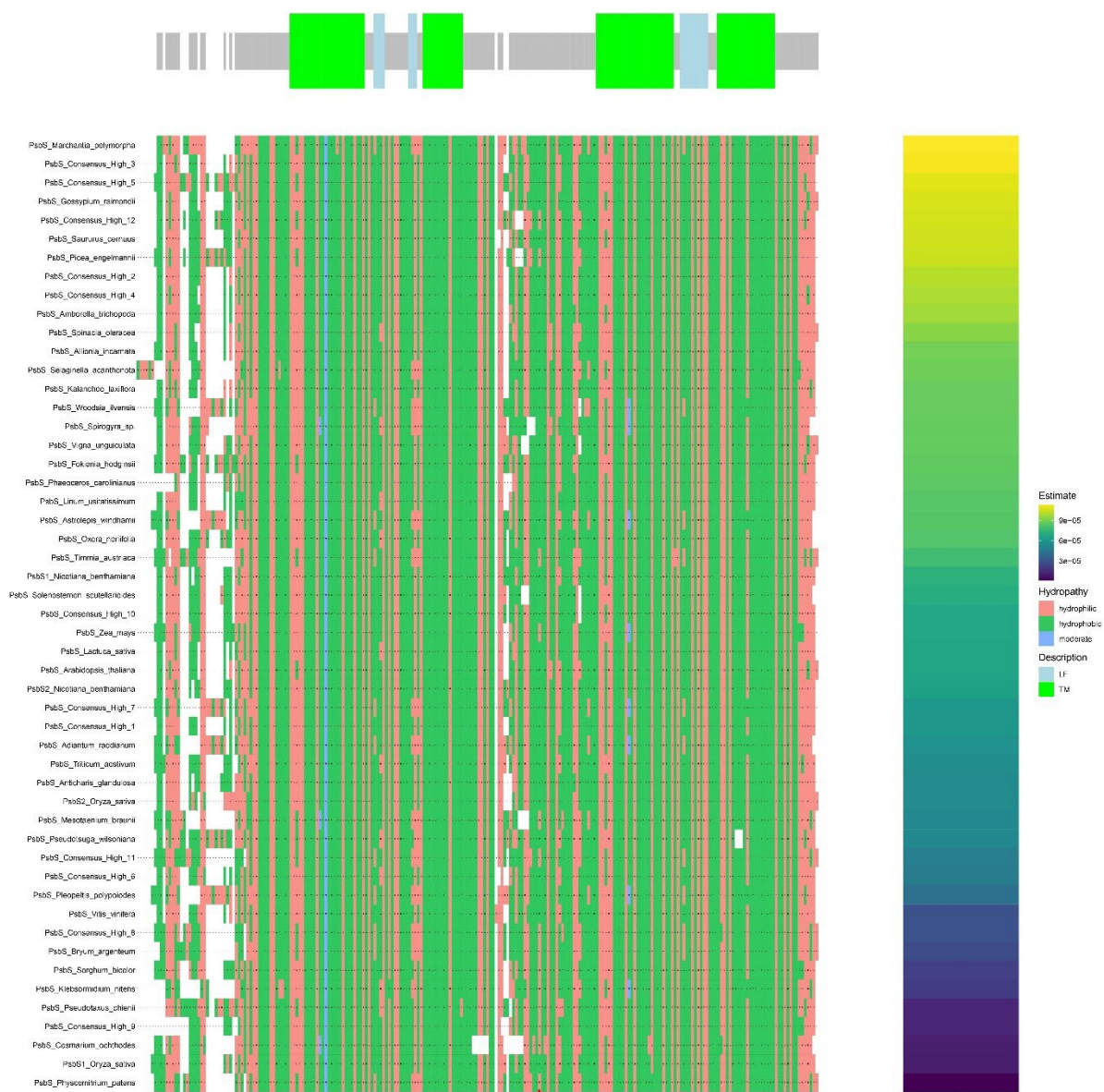


Figure 4-19. *PsbS* MSA of hydropathy in descending *MqE* order, excluding chlorophyte green algae

Sequences are color-coded by hydropathy (orange hydrophilic; green hydrophobic; blue neutral) and plotted next to the slope of *MqE*. Position of substitutions is shown relative to an AlphaFold predicted model of AtPsbS, with transmembrane helices in green, predicted “little folds” in blue and AtPsbS amino acid sequences in gray. An amino acid of interest at position 140 is denoted by a red asterisk (bottom).

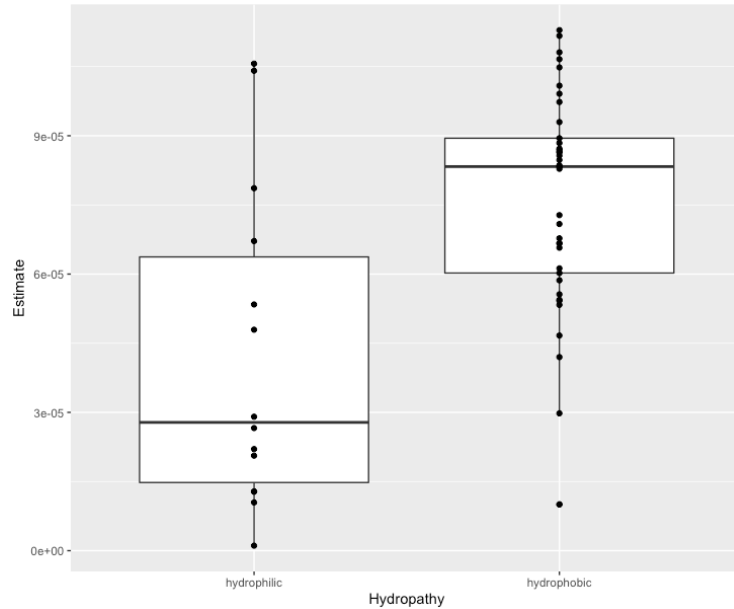


Figure 4-20. Correlation of hydrophobic residue at amino acid 140 with observed *MqE*

MqE slopes of all tested PsbS plotted depending on the identity of the amino acid at position 140 in the PsbS stromal loop. Individual data points are shown, with distribution summarized by box-and-whisker plot.

In sum, the evolutionary context provided by these MSA's will be useful in identifying critical amino acids positions that affect NPQ induction and magnitude. Evolutionary models that monitor sites of selection as well as co-evolution between amino acids may refine the structure-function relationships that remain to be understood.

4.6 Discussion

Advances in sequencing have catapulted our resolution of phylogenetic diversity. Leveraging the One Thousand Plant Transcriptomes Initiative¹⁵⁹, we mined over 1400 LHCSR and PsbS orthologs spanning green algae to land plants. Through sequence similarity networks we identify and statistically validate clusters of sequences that broadly but do not always mirror phylogenetic diversity. I also describe a high-throughput pipeline to clone, test, and screen these variants for differences in NPQ induction kinetics through the parameters *MqE* and *MqTot*. By testing these sequences at a wide range of expression levels in a relatively “noiseless” *psbs* mutant background, we can quantitatively assess the ability of LHCSR and PsbS orthologs to function in NPQ *in planta*. A number of interesting phenotypes were observed, laying the foundation for more specific dissection of sequence by function NPQ relationships.

One of the most significant results of this work is the identification of a land-plant-like, green algal PsbS (*Spirogyra*, **Figure 4-14E**). Sequence-level resolution revealed an obvious jump in sequence similarity between chlorophyte and streptophyte green algae (**Figure 4-16**). A genetic bottleneck (e.g. land plant-like PsbS in the common ancestor of late-branching streptophytes) may have mediated this transition, though quantification

of adaptive evolution in our library is needed. Additionally, we observe a presumably xanthophyll-dependent NPQ (MqTot) by streptophyte LHCSR (**Figure 4-9**), thus identifying Zygnematophytes as the most ancestral clade exhibiting co-dependence on both LHCSR and PsbS. Further analysis of streptophyte NPQ kinetics, such as those reported in *Klebsormium*^{168,169} and *Zygnema*¹⁷⁰, and the relative contributions of LHCSR and PsbS could expand our functional knowledge of these proteins in green algal LHCs.

Importantly, the mined orthologs used in this study are taken from extant species, suggesting selective pressure that maintained either or both LHCSR and PsbS in mosses and streptophyte algae. One explanation posited by Pinnola¹⁷¹ points to the relative lack of lower energy, red-shifted chlorophylls in the PSI of mosses^{172,173} and green algae¹⁷⁴ relative to land plants¹⁷⁵. Would a more energetically active PSI necessitate the need for a quencher like PpLHCSR1¹⁴⁹ and CrLHCSR3¹⁵⁰ that can quench at both photosystems? This PSI safety valve may be more essential in low-light, aquatic environments where microalgae have substantially blue-shifted their absorption of light by pigments to maximize light-harvesting efficiency¹⁷⁶. Cross-comparisons of species lacking and retaining LHCSR/PsbS, their ecological niches, and associated light-harvesting structures may further our understanding of this evolutionary transition.

Another exciting outcome of this study arose in its quantitative sensitivity. MqE and MqTot were used as proxies for rapidly induced Δ pH-dependent NPQ (MqE) and the additive contribution of longer-acting xanthophyll-dependent NPQ (MqTot, or the contributions of both qE and qZ), respectively. These parameters seem somewhat well justified by phenotypes, as demonstrated by the ability to resolve the qZ-like streptophyte LHCSR phenotypes analogous to what has been reported for quenching by PpLHCSR1 in an *Arabidopsis npq4* background¹⁷⁷. If the role of PsbS was to function simply as a trigger, we would expect MqE to linearly correlate with MqTot—given that both parameters are taken from the same sample, in a background with the same WT VDE capacity. While consistencies do exist, there are also significant outliers that dominate in either their MqE (*Marchantia*) or MqTot (*Arabidopsis*) capacity. What sequences underlie these differences in response? Is *Arabidopsis* PsbS a relatively “slow” pH sensor but efficient quencher? What role may protein-protein and protein-pigment interactions play (e.g. in efficiency of interaction, sites of quenching) in attenuating this response? Identifying substitutions that underlie MqE-dominant versus MqTot-dominant PsbS may provide the opportunity to physiologically uncouple the two phenotypes.

A surprising but reasonable outcome came from the relatively low quenching capacity of monocot PsbS, despite robust NPQ *in vivo*. Lou et al., report on complementation of

Arabidopsis npq4 by the monocot bamboo PsbS (*Phyllostachys edulis*)¹⁷⁸, however, complemented lines only ever reached 50% of WT NPQ capacity. This is surprising when placed in the context of published *Arabidopsis npq4* complementation data¹¹⁹, which saw NPQ capacities that far exceed WT NPQ capacity in hemizygous, single-copy functional complemented lines in which PsbS expression is driven by the same 35S promoter (**Figure 4-7B**). Has PsbS diverged in its function or interaction partners between monocots and dicots? Quantitative complementation in a monocot background, such as in protoplasts of rice *psbs* knockout mutants generated in Chapter 3, may substantiate this hypothesis.

However, the throughput achieved here does come with caveats. For example, the use of a cleaved fluorescence reporter fails to capture changes in PsbS protein stability which may bias interpretations. And while instability is still an interesting phenotype in this dissection of PsbS structure-function relationships, knowing when structural instability is present or how stability may be biased by the specificity of species-specific chloroplastic housekeeping proteases will be essential in understanding NPQ phenotypes. Stability was a challenge for Beraldo et. al. who report that 17 of their 18 native base-edited substitutions destabilized PsbS in *Physcomitrium* entirely¹⁷⁹. However, one significant advantage of our study is in the use of extant species that, by its nature, avoids areas of strong deleterious selection within PsbS¹⁸⁰ like the transmembrane domains edited in their study. As we see in our MSA (**Figure 4-16**), we identify and interrogate little functional diversity at those transmembrane sites. Regardless, a rigorous assessment of PsbS accumulation across constructs may further substantiate our dataset.

Our study has provided a high-throughput, quantitatively rigorous way to rapidly test LHCSR and PsbS structure-function-expression relationships. The broad swath of phylogenetic diversity covered here is just the starting point—future experiments interrogating co-evolution of PsbS with other putative interactors, and site-directed mutagenesis studies to fine-tune PsbS-dependent NPQ based on the outcomes of this dataset will be exceptionally worthwhile. Following more robust statistical interrogation of the evolutionary dynamics behind our tested and total sequences, this work has the potential to significantly boost our understanding of drivers that modulate PsbS-dependent NPQ. And with that basic biological knowledge in hand, this research may also guide more efficient VPZ approaches specific to target species and desired kinetics of quenching.

Chapter 5 : Future Outlooks

I started this dissertation discussing the limitations of genetic studies—with loss-of-function NPQ phenotypes plagued by epistasis, functional complexity, and dynamic plasticity. But even now, decades away from the mutagenized *Arabidopsis* seeds that sparked our understanding of NPQ *in vivo*, we can still turn to genetics to ask new and exciting questions. We just have to be creative enough in asking them, and desperate enough to get them to work. It is my hope that the mutants, gene-editing approaches, and structure-function relationships that I have begun to describe here will advance our basic and functional understanding of PsbS-dependent NPQ.

Chapter 6 : References

1. Niyogi, K. K. PHOTOPROTECTION REVISITED: Genetic and Molecular Approaches. *Annu. Rev. Plant. Physiol. Plant. Mol. Biol.* **50**, 333–359 (1999).
2. Durrant, J. R., Giorgi, L. B., Barber, J., Klug, D. R. & Porter, G. Characterisation of triplet states in isolated Photosystem II reaction centres: Oxygen quenching as a mechanism for photodamage. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1017**, 167–175 (1990).
3. Krieger-Liszkay, A., Fufezan, C. & Trebst, A. Singlet oxygen production in photosystem II and related protection mechanism. *Photosynth Res* **98**, 551–564 (2008).
4. Liu, Z. *et al.* Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* **428**, 287–292 (2004).
5. Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. Architecture of the Photosynthetic Oxygen-Evolving Center. *Science* **303**, 1831–1838 (2004).
6. Telfer, A., Bishop, S. M., Phillips, D. & Barber, J. Isolated photosynthetic reaction center of photosystem II as a sensitizer for the formation of singlet oxygen. Detection and quantum yield determination using a chemical trapping technique. *Journal of Biological Chemistry* **269**, 13244–13253 (1994).
7. Krieger-Liszkay, A. Singlet oxygen production in photosynthesis. *Journal of Experimental Botany* **56**, 337–346 (2005).
8. Demmig-Adams, B., Winter, K., Krüger, A. & Czygan, F.-C. Light Response of CO₂ Assimilation, Dissipation of Excess Excitation Energy, and Zeaxanthin Content of Sun and Shade Leaves 1. *Plant Physiology* **90**, 881–886 (1989).
9. Valladares, F. & Pearcy, R. W. The functional ecology of shoot architecture in sun and shade plants of *Heteromeles arbutifolia* M. Roem., a Californian chaparral shrub. *Oecologia* **114**, 1–10 (1998).
10. Mooney, H. A., Ehleringer, J. & Björkman, O. The energy balance of leaves of the evergreen desert shrub *Atriplex hymenelytra*. *Oecologia* **29**, 301–310 (1977).
11. DeLucia, E. H., Nelson, K., Vogelmann, T. C. & Smith, W. K. Contribution of intercellular reflectance to photosynthesis in shade leaves. *Plant, Cell & Environment* **19**, 159–170 (1996).
12. Park, Y. I., Chow, W. S. & Anderson, J. M. Chloroplast Movement in the Shade Plant *Tradescantia albiflora* Helps Protect Photosystem II against Light Stress. *Plant Physiology* **111**, 867–875 (1996).
13. Havaux, M. & Tardy, F. Loss of chlorophyll with limited reduction of photosynthesis as an adaptive response of Syrian barley landraces to high-light and heat stress. *Functional Plant Biol.* **26**, 569–578 (1999).
14. Melis, A. Solar energy conversion efficiencies in photosynthesis: Minimizing the chlorophyll antennae to maximize efficiency. *Plant Science* **177**, 272–280 (2009).
15. Bellafiore, S., Barneche, F., Peltier, G. & Rochaix, J.-D. State transitions and light adaptation require chloroplast thylakoid protein kinase STN7. *Nature* **433**, 892–895 (2005).

16. Boardman, N. t. Comparative photosynthesis of sun and shade plants. *Annual review of plant physiology* **28**, 355–377 (1977).
17. Slattery, R. A., Walker, B. J., Weber, A. P. M. & Ort, D. R. The Impacts of Fluctuating Light on Crop Performance. *Plant Physiology* **176**, 990–1003 (2018).
18. Müller, P., Li, X. P. & Niyogi, K. K. Non-photochemical quenching. A response to excess light energy. *Plant physiology* **125**, 1558–1566 (2001).
19. Li, X.-P. *et al.* A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* **403**, 391–395 (2000).
20. Nilkens, M. *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 - Bioenergetics* **1797**, 466–475 (2010).
21. Yamamoto, H. Y., Nakayama, T. O. M. & Chichester, C. O. Studies on the light and dark interconversions of leaf xanthophylls. *Archives of Biochemistry and Biophysics* **97**, 168–173 (1962).
22. Malnoë, A. *et al.* The Plastid Lipocalin LCNP Is Required for Sustained Photoprotective Energy Dissipation in Arabidopsis. *The Plant Cell* **30**, 196–208 (2018).
23. Long, S. P., Humphries, S. & Falkowski, P. G. Photoinhibition of photosynthesis in nature. *Annual review of plant biology* **45**, 633–662 (1994).
24. Kanazawa, A. & Kramer, D. M. In vivo modulation of nonphotochemical exciton quenching (NPQ) by regulation of the chloroplast ATP synthase. *Proceedings of the National Academy of Sciences* **99**, 12789–12794 (2002).
25. Takizawa, K., Kanazawa, A. & Kramer, D. M. Depletion of stromal Pi induces high 'energy-dependent' antenna exciton quenching (qE) by decreasing proton conductivity at CFO-CF1 ATP synthase. *Plant, Cell & Environment* **31**, 235–243 (2008).
26. Li, X.-P., Phippard, A., Pasari, J. & Niyogi, K. K. Structure-function analysis of photosystem II subunit S (PsbS) in vivo. *Functional Plant Biol.* **29**, 1131 (2002).
27. Li, X.-P. *et al.* Regulation of Photosynthetic Light Harvesting Involves Intrathylakoid Lumen pH Sensing by the PsbS Protein *. *Journal of Biological Chemistry* **279**, 22866–22874 (2004).
28. Krishnan-Schmieden, M., Konold, P. E., Kennis, J. T. M. & Pandit, A. The molecular pH-response mechanism of the plant light-stress sensor PsbS. *Nat Commun* **12**, 2291 (2021).
29. Pfundel, E. E. & Dilley, R. A. The pH Dependence of Violaxanthin Deepoxidation in Isolated Pea Chloroplasts. *Plant Physiology* **101**, 65–71 (1993).
30. Arnoux, P., Morosinotto, T., Saga, G., Bassi, R. & Pignol, D. A Structural Basis for the pH-Dependent Xanthophyll Cycle in Arabidopsis thaliana. *The Plant Cell* **21**, 2036–2044 (2009).
31. Müller-Moulé, P., Conklin, P. L. & Niyogi, K. K. Ascorbate Deficiency Can Limit Violaxanthin De-Epoxidase Activity in Vivo. *Plant Physiology* **128**, 970–977 (2002).
32. Bethmann, S., Melzer, M., Schwarz, N. & Jahns, P. The zeaxanthin epoxidase is degraded along with the D1 protein during photoinhibition of photosystem II. *Plant Direct* **3**, e00185 (2019).

33. Tikkanen, M., Mekala, N. R. & Aro, E. M. Photosystem II photoinhibition-repair cycle protects Photosystem I from irreversible damage. *Biochimica et Biophysica Acta - Bioenergetics* **1837**, 210–215 (2014).
34. Zhu, X., Ort, D. R., Whitmarsh, J. & Long, S. P. The slow reversibility of photosystem II thermal energy dissipation on transfer from high to low light may cause large losses in carbon gain by crop canopies: a theoretical analysis. *Journal of Experimental Botany* **55**, 1167–1175 (2004).
35. Long, S. P. *et al.* Into the Shadows and Back into Sunlight: Photosynthesis in Fluctuating Light. *Annual Review of Plant Biology* **73**, 617–648 (2022).
36. Acevedo-Siaca, L. G. *et al.* Variation in photosynthetic induction between rice accessions and its potential for improving productivity. *New Phytologist* **227**, 1097–1108 (2020).
37. Wang, Y., Burgess, S. J., de Becker, E. M. & Long, S. P. Photosynthesis in the fleeting shadows: an overlooked opportunity for increasing crop productivity? *The Plant Journal* **101**, 874–884 (2020).
38. Kromdijk, J. *et al.* Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science* **354**, 857–862 (2016).
39. De Souza, A. P. *et al.* Soybean photosynthesis and crop yield are improved by accelerating recovery from photoprotection. *Science* **377**, 851–854 (2022).
40. Garcia-Molina, A. & Leister, D. Accelerated relaxation of photoprotection impairs biomass accumulation in Arabidopsis. *Nat. Plants* **6**, 9–12 (2020).
41. Lehretz, G. G., Schneider, A., Leister, D. & Sonnewald, U. High non-photochemical quenching of VPZ transgenic potato plants limits CO₂ assimilation under high light conditions and reduces tuber yield under fluctuating light. *Journal of Integrative Plant Biology* **64**, 1821–1832 (2022).
42. Zhu, Z. Overexpression of Non-Photochemical Quenching (NPQ) Related Genes Alters Plant Development and Enhances Plant Osmotic Stress Resistance in Creeping Bentgrass. (Clemson University, 2021).
43. Maxwell, K. & Johnson, G. N. Chlorophyll fluorescence—a practical guide. *Journal of Experimental Botany* **51**, 659–668 (2000).
44. Murchie, E. H. & Lawson, T. Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. *Journal of Experimental Botany* **64**, 3983–3998 (2013).
45. Schreiber, U. Pulse-Amplitude-Modulation (PAM) Fluorometry and Saturation Pulse Method: An Overview. in *Chlorophyll a Fluorescence: A Signature of Photosynthesis* (eds. Papageorgiou, G. C. & Govindjee) 279–319 (Springer Netherlands, 2004). doi:10.1007/978-1-4020-3218-9_11.
46. Sylak-Glassman, E. J., Zaks, J., Amarnath, K., Leuenberger, M. & Fleming, G. R. Characterizing non-photochemical quenching in leaves through fluorescence lifetime snapshots. *Photosynth Res* **127**, 69–76 (2016).
47. Kalaji, H. M. *et al.* Frequently asked questions about in vivo chlorophyll fluorescence: practical issues. *Photosynth Res* **122**, 121–158 (2014).

48. Niyogi, K. K., Grossman, A. R. & Björkman, O. Arabidopsis Mutants Define a Central Role for the Xanthophyll Cycle in the Regulation of Photosynthetic Energy Conversion. *The Plant Cell* **10**, 1121–1134 (1998).
49. Li, X.-P., Müller-Moulé, P., Gilmore, A. M. & Niyogi, K. K. PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. *Proceedings of the National Academy of Sciences* **99**, 15222–15227 (2002).
50. Zia, A., Johnson, M. P. & Ruban, A. V. Acclimation- and mutation-induced enhancement of PsbS levels affects the kinetics of non-photochemical quenching in Arabidopsis thaliana. *Planta* **233**, 1253–1264 (2011).
51. Bergantino, E. *et al.* Light- and pH-dependent structural changes in the PsbS subunit of photosystem II. *Proceedings of the National Academy of Sciences* **100**, 15265–15270 (2003).
52. Niyogi, K. K. *et al.* Photoprotection in a zeaxanthin- and lutein-deficient double mutant of Arabidopsis. *Photosynthesis Research* **67**, 139–145 (2001).
53. Ghanotakis, D. F. *et al.* Comparative structural and catalytic properties of oxygen-evolving photosystem II preparations. *Photosynth Res* **14**, 191–199 (1987).
54. Teardo, E. *et al.* Evidences for interaction of PsbS with photosynthetic complexes in maize thylakoids. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1767**, 703–711 (2007).
55. Correa-Galvis, V., Poschmann, G., Melzer, M., Stühler, K. & Jahns, P. PsbS interactions involved in the activation of energy dissipation in Arabidopsis. *Nature Plants* **2**, 1–8 (2016).
56. Sacharz, J., Giovagnetti, V., Ungerer, P., Mastroianni, G. & Ruban, A. V. The xanthophyll cycle affects reversible interactions between PsbS and light-harvesting complex II to control non-photochemical quenching. *Nature Plants* **3**, 1–9 (2017).
57. Caffarri, S., Kouřil, R., Kereiche, S., Boekema, E. J. & Croce, R. Functional architecture of higher plant photosystem II supercomplexes. *The EMBO Journal* **28**, 3052–3063 (2009).
58. Funk, C. *et al.* The PSII-S Protein of Higher Plants: A New Type of Pigment-Binding Protein. *Biochemistry* **34**, 11133–11141 (1995).
59. Funk, C., Adamska, I., Green, B. R., Andersson, B. & Renger, G. The Nuclear-encoded Chlorophyll-binding Photosystem II-S Protein Is Stable in the Absence of Pigments (*). *Journal of Biological Chemistry* **270**, 30141–30147 (1995).
60. Dominici, P. *et al.* Biochemical Properties of the PsbS Subunit of Photosystem II Either Purified from Chloroplast or Recombinant *. *Journal of Biological Chemistry* **277**, 22750–22758 (2002).
61. Fan, M. *et al.* Crystal structures of the PsbS protein essential for photoprotection in plants. *Nat Struct Mol Biol* **22**, 729–735 (2015).
62. Daskalakis, V. Protein–protein interactions within photosystem II under photoprotection: the synergy between CP29 minor antenna, subunit S (PsbS) and zeaxanthin at all-atom resolution. *Phys. Chem. Chem. Phys.* **20**, 11843–11855 (2018).

63. Guardini, Z., Bressan, M., Caferri, R., Bassi, R. & Dall'Osto, L. Identification of a pigment cluster catalysing fast photoprotective quenching response in CP29. *Nat. Plants* **6**, 303–313 (2020).
64. Johnson, M. P. & Ruban, A. V. Restoration of Rapidly Reversible Photoprotective Energy Dissipation in the Absence of PsbS Protein by Enhanced ΔpH *. *Journal of Biological Chemistry* **286**, 19973–19981 (2011).
65. Johnson, M. P., Zia, A. & Ruban, A. V. Elevated ΔpH restores rapidly reversible photoprotective energy dissipation in Arabidopsis chloroplasts deficient in lutein and xanthophyll cycle activity. *Planta* **235**, 193–204 (2012).
66. Petrou, K., Belgio, E. & Ruban, A. V. pH sensitivity of chlorophyll fluorescence quenching is determined by the detergent/protein ratio and the state of LHCII aggregation. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1837**, 1533–1539 (2014).
67. Ruban, A. V. & Wilson, S. The Mechanism of Non-Photochemical Quenching in Plants: Localization and Driving Forces. *Plant and Cell Physiology* **62**, 1063–1072 (2021).
68. Ruban, A. V. Light harvesting control in plants. *FEBS Letters* **592**, 3030–3039 (2018).
69. Dall'Osto, L. *et al.* Two mechanisms for dissipation of excess light in monomeric and trimeric light-harvesting complexes. *Nature Plants* **3**, 1–9 (2017).
70. Ma, Y.-Z., Holt, N. E., Li, X.-P., Niyogi, K. K. & Fleming, G. R. Evidence for direct carotenoid involvement in the regulation of photosynthetic light harvesting. *Proceedings of the National Academy of Sciences* **100**, 4377–4382 (2003).
71. Son, M., Moya, R., Pinnola, A., Bassi, R. & Schlau-Cohen, G. S. Protein–Protein Interactions Induce pH-Dependent and Zeaxanthin-Independent Photoprotection in the Plant Light-Harvesting Complex, LHCII. *J. Am. Chem. Soc.* **143**, 17577–17586 (2021).
72. Nicol, L. & Croce, R. The PsbS protein and low pH are necessary and sufficient to induce quenching in the light-harvesting complex of plants LHCII. *Sci Rep* **11**, 7415 (2021).
73. Son, M., Pinnola, A. & Schlau-Cohen, G. S. Zeaxanthin independence of photophysics in light-harvesting complex II in a membrane environment. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1861**, 148115 (2020).
74. Barros, T., Royant, A., Standfuss, J., Dreuw, A. & Kühlbrandt, W. Crystal structure of plant light-harvesting complex shows the active, energy-transmitting state. *The EMBO Journal* **28**, 298–306 (2009).
75. Park, S., Steen, C. J., Fischer, A. L. & Fleming, G. R. Snapshot transient absorption spectroscopy: toward in vivo investigations of nonphotochemical quenching mechanisms. *Photosynth Res* **141**, 367–376 (2019).
76. Park, S. *et al.* Chlorophyll-Carotenoid Excitation Energy Transfer in High-Light-Exposed Thylakoid Membranes Investigated by Snapshot Transient Absorption Spectroscopy. *J. Am. Chem. Soc.* **140**, 11965–11973 (2018).
77. Park, S. *et al.* Snapshot Transient Absorption Spectroscopy of Carotenoid Radical Cations in High-Light-Acclimating Thylakoid Membranes. *J. Phys. Chem. Lett.* **8**, 5548–5554 (2017).

78. Ruban, A. V., Horton, P. & Robert, B. Resonance Raman Spectroscopy of the Photosystem II Light-Harvesting Complex of Green Plants: A Comparison of Trimeric and Aggregated States. *Biochemistry* **34**, 2333–2337 (1995).
79. Townsend, A. J. *et al.* The causes of altered chlorophyll fluorescence quenching induction in the Arabidopsis mutant lacking all minor antenna complexes. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1859**, 666–675 (2018).
80. Külheim, C., Ågren, J. & Jansson, S. Rapid Regulation of Light Harvesting and Plant Fitness in the Field. *Science* **297**, 91–93 (2002).
81. Pogson, B., McDonald, K. A., Truong, M., Britton, G. & DellaPenna, D. Arabidopsis carotenoid mutants demonstrate that lutein is not essential for photosynthesis in higher plants. *Plant Cell* **8**, 1627–1639 (1996).
82. Pogson, B. J., Niyogi, K. K., Björkman, O. & DellaPenna, D. Altered xanthophyll compositions adversely affect chlorophyll accumulation and nonphotochemical quenching in Arabidopsis mutants. *Proceedings of the National Academy of Sciences* **95**, 13324–13329 (1998).
83. Jahns, P. & Holzwarth, A. R. The role of the xanthophyll cycle and of lutein in photoprotection of photosystem II. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1817**, 182–193 (2012).
84. Dall'Osto, L. *et al.* Lutein is needed for efficient chlorophyll triplet quenching in the major LHCII antenna complex of higher plants and effective photoprotection in vivo under strong light. *BMC Plant Biology* **6**, 32 (2006).
85. Leuenberger, M. *et al.* Dissecting and modeling zeaxanthin- and lutein-dependent nonphotochemical quenching in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences* **114**, E7009–E7017 (2017).
86. Mozzo, M., Dall'Osto, L., Hienerwadel, R., Bassi, R. & Croce, R. Photoprotection in the Antenna Complexes of Photosystem II: ROLE OF INDIVIDUAL XANTHOPHYLLS IN CHLOROPHYLL TRIPLET QUENCHING *. *Journal of Biological Chemistry* **283**, 6184–6192 (2008).
87. Niyogi, K. K., Björkman, O. & Grossman, A. R. The roles of specific xanthophylls in photoprotection. *Proceedings of the National Academy of Sciences* **94**, 14162–14167 (1997).
88. Liu, H. *et al.* CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Molecular Plant* **10**, 530–532 (2017).
89. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences* **112**, 3570–3575 (2015).
90. Hsiao, T. *et al.* Inference of CRISPR Edits from Sanger Trace Data. <http://biorxiv.org/lookup/doi/10.1101/251082> (2018) doi:10.1101/251082.
91. Bombarely, A. *et al.* A Draft Genome Sequence of Nicotiana benthamiana to Enhance Molecular Plant-Microbe Biology Research. *MPMI* **25**, 1523–1530 (2012).

92. Duckham, S. C., Linforth, R. S. T. & Taylor, I. B. Absciscic-acid-deficient mutants at the aba gene locus of *Arabidopsis thaliana* are impaired in the epoxidation of zeaxanthin. *Plant Cell Environ* **14**, 601–606 (1991).
93. Long, S. P., Marshall-Colon, A. & Zhu, X.-G. Meeting the Global Food Demand of the Future by Engineering Crop Photosynthesis and Yield Potential. *Cell* **161**, 56–66 (2015).
94. Ort, D. R. *et al.* Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences* **112**, 8529–8536 (2015).
95. South, P. F., Cavanagh, A. P., Liu, H. W. & Ort, D. R. Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science* **363**, eaat9077 (2019).
96. Cavanagh, A. P., South, P. F., Bernacchi, C. J. & Ort, D. R. Alternative pathway to photorespiration protects growth and productivity at elevated temperatures in a model crop. *Plant Biotechnology Journal* **20**, 711–721 (2022).
97. Wei, S. *et al.* A transcriptional regulator that boosts grain yields and shortens the growth duration of rice. *Science* **377**, eabi8455 (2022).
98. Wu, J. *et al.* Overexpression of zmm28 increases maize grain yield in the field. *Proceedings of the National Academy of Sciences* **116**, 23850–23858 (2019).
99. James, V. A., Avart, C., Worland, B., Snape, J. W. & Vain, P. The relationship between homozygous and hemizygous transgene expression levels over generations in populations of transgenic rice plants. *Theoretical and Applied Genetics* **104**, 553–561 (2002).
100. Kasajima, I. *et al.* Molecular distinction in genetic regulation of nonphotochemical quenching in rice. *Proceedings of the National Academy of Sciences* **108**, 13835–13840 (2011).
101. Hubbart, S. *et al.* Enhanced thylakoid photoprotection can increase yield and canopy radiation use efficiency in rice. *Commun Biol* **1**, 22 (2018).
102. Głowacka, K. *et al.* Photosystem II Subunit S overexpression increases the efficiency of water use in a field-grown crop. *Nat Commun* **9**, 868 (2018).
103. Wang, Q. *et al.* Genetic Architecture of Natural Variation in Rice Nonphotochemical Quenching Capacity Revealed by Genome-Wide Association Study. *Frontiers in Plant Science* **8**, (2017).
104. Rodríguez-Leal, D., Lemmon, Z. H., Man, J., Bartlett, M. E. & Lippman, Z. B. Engineering Quantitative Trait Variation for Crop Improvement by Genome Editing. *Cell* **171**, 470–480.e8 (2017).
105. Liu, L. *et al.* Enhancing grain-yield-related traits by CRISPR–Cas9 promoter editing of maize CLE genes. *Nature Plants* **7**, 287–294 (2021).
106. Song, X. *et al.* Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nat Biotechnol* **40**, 1403–1411 (2022).
107. Wada, N., Ueta, R., Osakabe, Y. & Osakabe, K. Precision genome editing in plants : state- of-the-art in CRISPR / Cas9-based genome engineering. 1–12 (2020).
108. Schatz, M. C. *et al.* Whole genome de novo assemblies of three divergent strains of rice, *Oryza sativa*, document novel gene space of aus and indica. 16 (2014).

109. Chu, C. C. *et al.* Establishment of an efficient medium for anther culture of rice, through comparative experiments on the nitrogen sources. *Sci Sin* **18**, (1975).
110. Murashige, T. & Skoog, F. A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiologia Plantarum* **15**, (1962).
111. Vandesompele, J. *et al.* Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3**, (2002).
112. Hellemans, J., Mortier, G., De Paepe, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19 (2007).
113. Fu, X. *et al.* The coordination of OsbZIP72 and OsMYBS2 with reverse roles regulates the transcription of *OsPsbS1* in rice. *New Phytol* **229**, 370–387 (2021).
114. Jain, M., Nijhawan, A., Tyagi, A. K. & Khurana, J. P. Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochemical and Biophysical Research Communications* **345**, 646–651 (2006).
115. Karlsson, J. O., Ostwald, K., Kabjorn, C. & Andersson, M. A Method for Protein Assay in Laemmli Buffer. *Analytical Biochemistry* **219**, 144–146 (1994).
116. Ouyang, S. *et al.* The TIGR Rice Genome Annotation Resource: improvements and new features. *Nucleic Acids Research* **35**, D883–D887 (2007).
117. D-GENIES: dot plot large genomes in an interactive, efficient and simple way [PeerJ]. <https://peerj.com/articles/4958/>.
118. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24–26 (2011).
119. Niyogi, K. K. Is PsbS the site of non-photochemical quenching in photosynthesis? *Journal of Experimental Botany* **56**, 375–382 (2004).
120. Hubbart, S., Ajigboye, O. O., Horton, P. & Murchie, E. H. The photoprotective protein PsbS exerts control over CO₂ assimilation rate in fluctuating light in rice: Photoprotection and CO₂ assimilation in rice. *The Plant Journal* no-no (2012) doi:10.1111/j.1365-313X.2012.04995.x.
121. Soyk, S. *et al.* Bypassing Negative Epistasis on Yield in Tomato Imposed by a Domestication Gene. *Cell* **169**, 1142–1155.e12 (2017).
122. Zhang, H. *et al.* Genome editing of upstream open reading frames enables translational control in plants. *Nat Biotechnol* **36**, 894–898 (2018).
123. Alonge, M. *et al.* Major Impacts of Widespread Structural Variation on Gene Expression and Crop Improvement in Tomato. *Cell* **182**, 145–161.e23 (2020).
124. Song, J.-M. *et al.* Eight high-quality genomes reveal pan-genome architecture and ecotype differentiation of *Brassica napus*. *Nat. Plants* **6**, 34–45 (2020).
125. Sun, S. *et al.* Extensive intraspecific gene order and gene structural variations between Mo17 and other maize genomes. *Nat Genet* **50**, 1289–1295 (2018).

126. Zhou, Y. *et al.* The population genetics of structural variants in grapevine domestication. *Nat. Plants* **5**, 965–979 (2019).
127. Qin, P. *et al.* Pan-genome analysis of 33 genetically diverse rice accessions reveals hidden genomic variations. *Cell* **184**, 3542–3558.e16 (2021).
128. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. *Nat Rev Genet* **14**, 125–138 (2013).
129. Beyter, D. *et al.* Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nat Genet* **53**, 779–786 (2021).
130. Hehir-Kwa, J. Y. *et al.* A high-quality human reference panel reveals the complexity and distribution of genomic structural variants. *Nat Commun* **7**, 12989 (2016).
131. Domínguez, M. *et al.* The impact of transposable elements on tomato diversity. *Nat Commun* **11**, 4058 (2020).
132. Gaut, B. S., Seymour, D. K., Liu, Q. & Zhou, Y. Demography and its effects on genomic variation in crop domestication. *Nature Plants* **4**, 512–520 (2018).
133. Zhou, H. *et al.* A 1.7-Mb chromosomal inversion downstream of a PpOFP1 gene is responsible for flat fruit shape in peach. *Plant Biotechnology Journal* **19**, 192–205 (2021).
134. Affer, M. *et al.* Promiscuous MYC locus rearrangements hijack enhancers but mostly super-enhancers to dysregulate MYC expression in multiple myeloma. *Leukemia* **28**, 1725–1735 (2014).
135. Lu, Y. *et al.* A donor-DNA-free CRISPR/Cas-based approach to gene knock-up in rice. *Nat. Plants* **7**, 1445–1452 (2021).
136. Kweon, J. *et al.* Targeted genomic translocations and inversions generated using a paired prime editing strategy. *Molecular Therapy* (2022) doi:10.1016/j.ymthe.2022.09.008.
137. Zhao, L. *et al.* Integrative analysis of reference epigenomes in 20 rice varieties. *Nat Commun* **11**, 2658 (2020).
138. Keeling, P. J. The Number, Speed, and Impact of Plastid Endosymbioses in Eukaryotic Evolution. *Annu. Rev. Plant Biol.* **64**, 583–607 (2013).
139. Corlett, R. T. Plant diversity in a changing world: Status, trends, and conservation needs. *Plant Diversity* **38**, 10–16 (2016).
140. Niyogi, K. K. & Truong, T. B. Evolution of flexible non-photochemical quenching mechanisms that regulate light harvesting in oxygenic photosynthesis. *Current Opinion in Plant Biology* **16**, 307–314 (2013).
141. Peers, G. *et al.* An ancient light-harvesting protein is critical for the regulation of algal photosynthesis. *Nature* **462**, 518–521 (2009).
142. Ballottari, M. *et al.* Identification of pH-sensing Sites in the Light Harvesting Complex Stress-related 3 Protein Essential for Triggering Non-photochemical Quenching in *Chlamydomonas reinhardtii**. *Journal of Biological Chemistry* **291**, 7334–7346 (2016).

143. Camargo, F. V. A. *et al.* The Role of Acidic Residues in the C Terminal Tail of the LHCSR3 Protein of *Chlamydomonas reinhardtii* in Non-Photochemical Quenching. *J. Phys. Chem. Lett.* **12**, 6895–6900 (2021).
144. Bonente, G. *et al.* Analysis of LhcSR3, a Protein Essential for Feedback De-Excitation in the Green Alga *Chlamydomonas reinhardtii*. *PLOS Biology* **9**, e1000577 (2011).
145. Kondo, T. *et al.* Single-molecule spectroscopy of LHCSR1 protein dynamics identifies two distinct states responsible for multi-timescale photosynthetic photoprotection. *Nature Chem* **9**, 772–778 (2017).
146. Alboresi, A., Gerotto, C., Giacometti, G. M., Bassi, R. & Morosinotto, T. Physcomitrella patens mutants affected on heat dissipation clarify the evolution of photoprotection mechanisms upon land colonization. *Proceedings of the National Academy of Sciences* **107**, 11128–11133 (2010).
147. Gerotto, C., Alboresi, A., Giacometti, G. M., Bassi, R. & Morosinotto, T. Coexistence of plant and algal energy dissipation mechanisms in the moss *Physcomitrella patens*. *New Phytologist* **196**, 763–773 (2012).
148. Pinnola, A. *et al.* Zeaxanthin Binds to Light-Harvesting Complex Stress-Related Protein to Enhance Nonphotochemical Quenching in *Physcomitrella patens*. *The Plant Cell* **25**, 3519–3534 (2013).
149. Pinnola, A. *et al.* Light-Harvesting Complex Stress-Related Proteins Catalyze Excess Energy Dissipation in Both Photosystems of *Physcomitrella patens*. *The Plant Cell* **27**, 3213–3227 (2015).
150. Girolomoni, L. *et al.* LHCSR3 is a nonphotochemical quencher of both photosystems in *Chlamydomonas reinhardtii*. *Proceedings of the National Academy of Sciences* **116**, 4212–4217 (2019).
151. Correa-Galvis, V. *et al.* Photosystem II Subunit PsbS Is Involved in the Induction of LHCSR Protein-dependent Energy Dissipation in *Chlamydomonas reinhardtii**. *Journal of Biological Chemistry* **291**, 17478–17487 (2016).
152. Tibiletti, T., Auroy, P., Peltier, G. & Caffarri, S. *Chlamydomonas reinhardtii* PsbS Protein Is Functional and Accumulates Rapidly and Transiently under High Light. *Plant Physiology* **171**, 2717–2730 (2016).
153. Gerotto, C. & Morosinotto, T. Evolution of photoprotection mechanisms upon land colonization: evidence of PSBS-dependent NPQ in late Streptophyte algae. *Physiologia Plantarum* **149**, 583–598 (2013).
154. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**, D1178–D1186 (2012).
155. Grigoriev, I. V. *et al.* PhycoCosm, a comparative algal genomics resource. *Nucleic Acids Research* **49**, D1004–D1011 (2021).
156. Johnson, M. *et al.* NCBI BLAST: a better web interface. *Nucleic Acids Research* **36**, W5–W9 (2008).
157. Tardif, M. *et al.* PredAlgo: A New Subcellular Localization Prediction Tool Dedicated to Green Algae. *Molecular Biology and Evolution* **29**, 3625–3639 (2012).

158. Emanuelsson, O., Brunak, S., von Heijne, G. & Nielsen, H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* **2**, 953–971 (2007).
159. One Thousand Plant Transcriptomes Initiative. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**, 679–685 (2019).
160. Zallot, R., Oberg, N. & Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **58**, 4169–4182 (2019).
161. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).
162. Su, G., Kuchinsky, A., Morris, J. H., States, D. J. & Meng, F. GLay: community structure analysis of biological networks. *Bioinformatics* **26**, 3135–3137 (2010).
163. Zhang, B. *et al.* Coordinated protein co-expression in plants by harnessing the synergy between an intein and a viral 2A peptide. *Plant Biotechnology Journal* **15**, 718–728 (2017).
164. Shibuta, M. K. *et al.* A live imaging system to analyze spatiotemporal dynamics of RNA polymerase II modification in *Arabidopsis thaliana*. *Commun Biol* **4**, 1–10 (2021).
165. Sainsbury, F., Thuenemann, E. C. & Lomonossoff, G. P. pEAQ: versatile expression vectors for easy and quick transient expression of heterologous proteins in plants. *Plant Biotechnology Journal* **7**, 682–693 (2009).
166. Sternke, M., Tripp, K. W. & Barrick, D. Consensus sequence design as a general strategy to create hyperstable, biologically active proteins. *Proceedings of the National Academy of Sciences* **116**, 11275–11284 (2019).
167. Xia, L. *et al.* Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. *Journal of Genetics and Genomics* **44**, 235–241 (2017).
168. Pierangelini, M., Ryšánek, D., Lang, I., Adlassnig, W. & Holzinger, A. Terrestrial adaptation of green algae *Klebsormidium* and *Zygnema* (Charophyta) involves diversity in photosynthetic traits but not in CO₂ acquisition. *Planta* **246**, 971–986 (2017).
169. Pierangelini, M., Glaser, K., Mikhailyuk, T., Karsten, U. & Holzinger, A. Light and Dehydration but Not Temperature Drive Photosynthetic Adaptations of Basal Streptophytes (*Hormidiella*, *Streptosarcina* and *Streptofilum*) Living in Terrestrial Habitats. *Microb Ecol* **77**, 380–393 (2019).
170. Permann, C., Pierangelini, M., Remias, D., Lewis, L. A. & Holzinger, A. Photophysiological investigations of the temperature stress responses of *Zygnema* spp (Zygnematophyceae) from subpolar and polar habitats (Iceland, Svalbard). *Phycologia* **61**, 299–311 (2022).
171. Pinnola, A. The rise and fall of Light-Harvesting Complex Stress-Related proteins as photoprotection agents during evolution. *Journal of Experimental Botany* **70**, 5527–5535 (2019).
172. Iwai, M. *et al.* Light-harvesting complex Lhcb9 confers a green alga-type photosystem I supercomplex to the moss *Physcomitrella patens*. *Nature Plants* **1**, 1–7 (2015).

173. Gorski, C. *et al.* The structure of the *Physcomitrium patens* photosystem I reveals a unique Lhca2 paralogue replacing Lhca4. *Nat. Plants* **8**, 307–316 (2022).
174. Gibasiewicz, K. *et al.* Characterization of Low-Energy Chlorophylls in the PSI-LHCI Supercomplex from *Chlamydomonas reinhardtii*. A Site-Selective Fluorescence Study. *J. Phys. Chem. B* **109**, 21180–21186 (2005).
175. Croce, R., Morosinotto, T., Castelletti, S., Breton, J. & Bassi, R. The Lhca antenna complexes of higher plants photosystem I. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1556**, 29–40 (2002).
176. Kunugi, M. *et al.* Evolution of Green Plants Accompanied Changes in Light-Harvesting Systems. *Plant and Cell Physiology* **57**, 1231–1243 (2016).
177. Dikaïos, I. *et al.* Functional analysis of LHCSR1, a protein catalyzing NPQ in mosses, by heterologous expression in *Arabidopsis thaliana*. *Photosynth Res* **142**, 249–264 (2019).
178. Lou, Y. *et al.* Expression and functional analysis of two PsbS genes in bamboo (*Phyllostachys edulis*). *Physiologia Plantarum* **163**, 459–471 (2018).
179. Beraldo, C., Guyon-Debast, A., Alboresi, A., Nogu  , F. & Morosinotto, T. Functional analysis of PsbS transmembrane domains through base editing in *Physcomitrium patens*. 2022.05.24.491476 Preprint at <https://doi.org/10.1101/2022.05.24.491476> (2022).
180. Schultes, N. P. & Peterson, R. B. Phylogeny-directed structural analysis of the *Arabidopsis* PsbS protein. *Biochemical and Biophysical Research Communications* **355**, 464–470 (2007).