

From Field Photosynthesis to Genetic Architecture: Insights from the First Dedicated Photosynthesis Hackathon

Anna Matuszyńska^{1*}, Olakunle Sansa^{2*}, Femi James Adekoya^{3,4}, Olusegun Olaitan Akinyemi⁵, Esther Anokye⁶, Olukunle Babatunde Bashir², Zsa Zsa Friederique Boyny⁷, Mkpuma Kenneth Chukwuka⁸, Elouen Corvest¹, Adenike Oluwaseun Dada^{2,9}, Matteo Dell'Acqua¹⁰, Georgina Lala Ehembamba¹¹, Andris Justin Finkbeiner⁷, Swivia Hamabwe¹², Djidjoho A.T. Hodehou¹³, Olivia Kacheyo¹⁴, Kelvin Kamfwa¹², Kanthu Joseph Mhango³, Wajiha Mu'az Abdullahi^{15,16}, Gideon Munduwe¹², Solomon Ntukidem², Olalekan Kazeem Obisesan¹⁶, Ifeoluwa Simeon Odesina^{17,18}, Nwogu-Uchenna Ogechi^{19,20}, Omotola Dorcas Olaoye¹⁹, Mary Morenike Olayinka^{2,21}, Isaac Osei-Bonsu^{21,22}, Kolawale Olalekan Rilwan², Leandro Stival²³, Zeynu Tehar¹⁰, Regina Mumbua Tende²⁴, Jacky To²³, Uchendu Kelechi Ugochukwu²⁵, Anika Unger⁷, Marvin van Aalst¹, Dominik Vrbic¹⁴, Chao Zhang²³, Tom P.J.M. Theeuwes^{14*}, David M. Kramer^{14*}, Johannes Kromdijk^{26*}

¹ Computational Life Science, RWTH Aachen University, Worringerweg 1, 52074 Aachen, Germany

² International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria

³ Harper Adams University, Newport, United Kingdom

⁴ Integrated Aerial Precision Ltd., Nigeria

⁵ Norwegian University of Life Sciences (NMBU), Ås, Norway

⁶ Cocoa Research Institute of Ghana (CRIG), Tafo, Ghana

⁷ Justus Liebig University Giessen, Giessen, Germany

⁸ Jomo Kenyatta University of Agriculture and Technology (JKUAT), Juja, Kenya

⁹ Institute of Agricultural Research and Training (IART), Ibadan, Nigeria

¹⁰ Scuola Superiore Sant'Anna, Pisa, Italy

¹¹ Institut Sénégalais de Recherches Agricoles (ISRA), Dakar, Senegal

¹² School of Agricultural Sciences, The University of Zambia, Lusaka, Zambia

¹³ Centre d'Étude Régional pour l'Amélioration de l'Adaptation à la Sécheresse (CERAAS), Thiès, Senegal

¹⁴ Jan IngenHousz Institute, Wageningen, Netherlands

¹⁵ Institute for Agricultural Research (IAR), Zaria, Nigeria, Ahmadu Bello University, Zaria, Nigeria

¹⁶ University of Ibadan, Ibadan, Nigeria

¹⁷ Laboratório de Melhoramento Genético Vegetal, Universidade Estadual do Norte Fluminense

Darcy Ribeiro, Campos dos Goytacazes, Rio de Janeiro 28013602, Brazil

¹⁸ Cytogenetics and Plant Breeding Unit, Department of Plant Science and Biotechnology, University of Jos, Jos 930001, Nigeria

¹⁹ University of Ghana / West Africa Centre for Crop Improvement (WACCI), Accra, Ghana

²⁰ Michael Okpara University of Agriculture, Umudike, Nigeria

²¹ Kwame Nkrumah University of Science and Technology (KNUST), Kumasi, Ghana

²² CSIR-Crops Research Institute, Kumasi, Ghana

²³ Wageningen University & Research, Wageningen, Netherlands

²⁴ Kenya Agricultural and Livestock Research Organization (KALRO), Nairobi, Kenya

²⁵ National Root Crops Research Institute (NRCRI), Kuru Outstation, Nigeria

²⁶ Department of Plant Sciences, University of Cambridge, Cambridge, United Kingdom

* Corresponding authors: anna.matuszynska@cpbl.rwth-aachen.de, o.sansa@cgiar.org, tom.theeuwes@jii.org, david.kramer@jii.org and jk417@cam.ac.uk

Abstract

Photosynthesis is among the most consequential yet genetically complex traits in crop plants, and translating its natural variation into actionable genomic targets remains a central challenge for

breeding climate-resilient varieties. To start addressing this, researchers are generating increasingly large, multi-environment field photosynthesis datasets. Yet, these data have been structurally under-analysed since their inception. Here we report the outcomes of the first dedicated hackathon focused on computational mining of such field data held in Accra, Ghana, in March 2026. Bringing together data scientists, plant physiologists, geneticists, and breeders from Europe and Africa, these interdisciplinary teams used photosynthetic data collected with hand-held fluorometers to genome-wide marker data across four crop species: cowpea (*Vigna unguiculata*), barley (*Hordeum vulgare*), common bean (*Phaseolus vulgaris*), and potato (*Solanum tuberosum*).

Despite using different species and methods, independent teams identified the same three key findings. First, mechanism-informed feature engineering and dynamic modelling recover genetic signals that are not detected or discarded in standard analysis pipelines, resulting in traits with improved heritability and meaningful associations with yield. Secondly, machine learning methods proved effective at uncovering genetic associations, with temporally resolved features substantially outperforming single time-point measurements. Third, raw chlorophyll fluorescence and absorbance traces consistently contained more information and predictive power than the extracted parameters currently used. A defining feature of this event was having experimentalists and data scientists working together, enabling AI approaches to be grounded in domain knowledge and biological mechanisms rather than relying on data alone.

Keywords: *photosynthesis, GWAS, genetic architecture, phenomics, cowpea, common bean, barley, potato, hackathon, MultispeQ, Africa, plant breeding, Machine Learning, openJII*

Abbreviations

AUC:	area-under-the-curve
DIRK:	dark-interval relaxation kinetics
ECS:	electrochromic shift
FDR:	false discovery rate
GWAS:	Genome-wide association study
H ² :	broad-sense heritability
IITA:	International Institute of Tropical Agriculture
ILR	isometric log-ratio
LEF:	linear electron flow
MAGIC:	Multiparent advanced generation inter-cross
MDA:	mean decrease accuracy
MDG:	Mean decrease gini
MTA:	Marker-trait association
NPQ:	non-photochemical quenching
NPQ _t :	total NPQ
OOB:	out-of-bag
PAR:	photosynthetically active radiation
qL:	fraction of open PSII centers
RIL:	Recombinant inbred line
SNP:	single nucleotide polymorphism
SPAD:	relative chlorophyll content

Φ_{II} : PSII quantum yield
 Φ_{NO} : non-regulated energy loss quantum yield
 Φ_{NPQ} : NPQ quantum yield

1. Introduction

Photosynthesis is the primary source of organic carbon for crop biomass, and improving its efficiency in field conditions is one of the most pursued ambitions in current plant science. Theoretical analyses place the maximum solar energy conversion efficiency of C3 photosynthesis at 4.6% under natural conditions, while most crops achieve less than half this figure in the field (Monteith, 1977; Zhu et al., 2008). The gap reflects biological inefficiencies in light harvesting, energy dissipation, and carbon fixation, all of which are genetically variable, potentially heritable, and, in principle, reducible (Smith et al., 2023). Engineering demonstrations have confirmed this potential: accelerating non-photochemical quenching (NPQ) relaxation improved tobacco biomass by up to 20% in field conditions (Kromdijk et al., 2016); photorespiratory bypass increased tobacco yield by 40% under controlled conditions and other bypasses are predicted to increase photosynthetic efficiency by 27% (South et al., 2019; Smith et al., 2025); ribulose-1,5-bisphosphate carboxylase/oxygenase modifications improved carbon fixation rates (Salesse-Smith et al., 2018; Yoon et al., 2020); and manipulation of Calvin-Benson-Bassham cycle flux improved growth in multiple species (Ainsworth & Long, 2021).

Despite these advances, progress toward genetic improvement of field photosynthesis has remained limited (Kromdijk & Long, 2016; Long et al., 2025; Ort et al., 2015; Theeuwes et al., 2022). First, phenotyping photosynthesis at scale in the field is technically demanding: gas exchange measurements are slow and expensive, and earlier fluorescence-based methods required laboratory conditions incompatible with large breeding nurseries. This enabling factor has been substantially addressed through the development of affordable handheld devices for measuring photosynthetic traits. Low-cost deployment of sufficient parallel devices of these handheld sensors has made high-throughput assessment under field conditions achievable (provided an equal number of operators can also be arranged). The MultispeQ device (Kuhlgert et al., 2016) delivers leaf-level measurements of, among others, chlorophyll fluorescence parameters (PSII quantum yield (Φ_{II}), NPQ quantum yield (Φ_{NPQ}), non-regulated energy loss quantum yield (Φ_{NO}), fraction of open PSII centers (qL), and linear electron flow (LEF)) (Kramer et al., 2004; Tietz et al., 2017), relative chlorophyll content (SPAD), ATP synthase conductivity, and proton motive force (ECSt, vH+, gH+) (Avenson et al., 2005; Kanazawa & Kramer, 2002), PSI activity (active centers, open/oxidized/over-reduced states), as well as environmental parameters (Photosynthetically active radiation (PAR), temperature, and humidity) in approximately 15 seconds per measurement. Users are also able to develop their own protocols based on specific research interests by incorporating different light intensities or newly defined parameters. The short measuring time enables phenotyping campaigns at the scale required for multi-environment trials and genetic mapping studies. With the MultispeQ more than 6.2 million such measurements had been recorded by June 2026 (PhotosynQ and openJII platforms). Deployment of this infrastructure has enabled genome-wide association studies linking photosynthetic traits to genomic regions in cowpea (*Vigna unguiculata*) (Sansa et al., 2024) and maize (*Zea mays*) (Ali et al., 2025; Li et al., 2025), as well as the use of genomic prediction models of yield in barley (*Hordeum vulgare*) (Gao et al., 2026). This shows that the phenotypic, genomic, and computational resources necessary for this class of analysis are increasingly available.

However, even when field phenotypic data exist, their analysis is often constrained by workflows developed for simpler, static quantitative traits and the measurements may contain more informative but unexplored parameters. Such workflows are poorly suited to photosynthetic measurements, which are dynamic, environmentally sensitive, and mechanistically structured, which may, in turn, leave substantial genetic information unused. Addressing this gap requires more than applying additional statistical methods: it requires close interaction between plant physiologists, geneticists, breeders,

and computational scientists to define traits that are both analytically robust and biologically meaningful. Hackathons are increasingly used as a structured format for this type of interdisciplinary problem-solving, bringing together participants from diverse backgrounds for short, intensive, goal-oriented collaborations (Chau & Gerber, 2023). They are typically time-bounded events designed to accelerate the development of ideas, prototypes, and even scientific insights, while simultaneously fostering collaboration and community-building across disciplinary boundaries. In educational and research contexts, hackathons have been shown to promote experiential learning, multidisciplinary teamwork, and problem-solving under realistic constraints, enabling participants to integrate domain knowledge with data-driven and computational approaches (Sotaquirá-Gutiérrez et al., 2025). As highlighted in our previous work (Van Aalst et al., 2025), although most empirical evidence to date stems from broader STEM contexts, the hackathon format is particularly well suited to life sciences, where fragmented expertise and heterogeneous data necessitate tightly integrated, short-cycle collaboration.

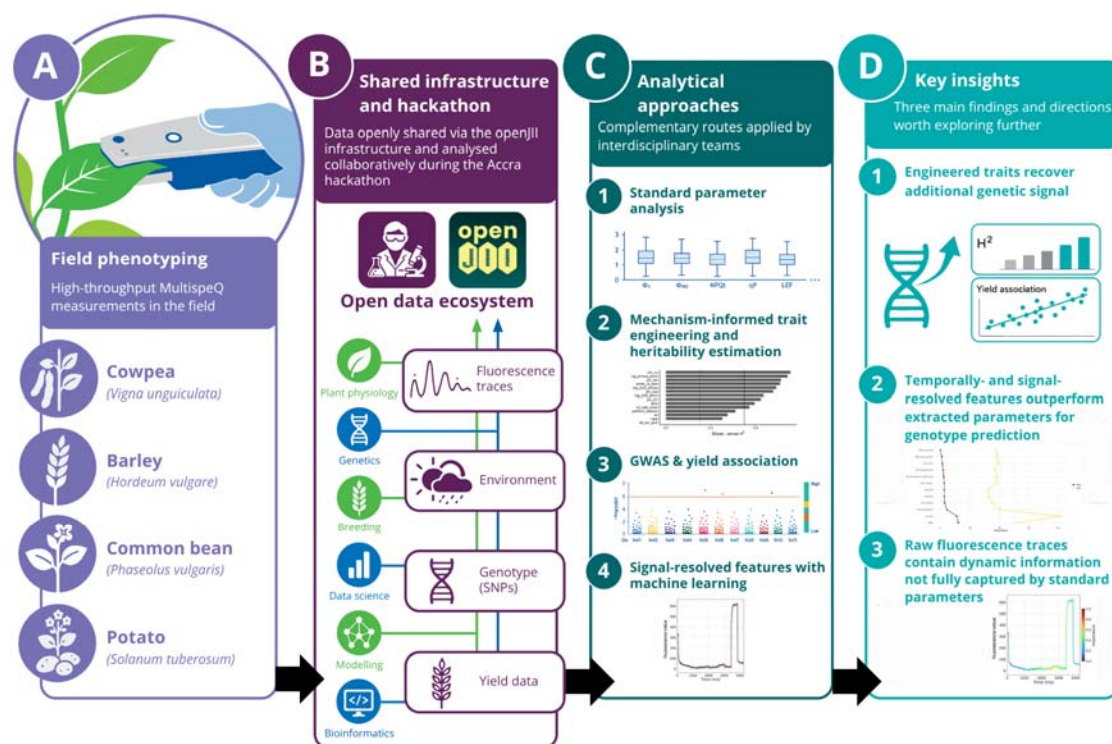


Fig. 1 Workflow of the first dedicated photosynthesis hackathon linking field phenotyping, shared infrastructure, and genetic interpretation. Field MultispeQ measurements collected from cowpea, barley, common bean and potato included data on extracted photosynthetic parameters, raw fluorescence traces, and environmental covariates (A). They were complemented with the yield records and genome-wide marker data. During the Accra hackathon, these data layers were made available through the openJII infrastructure and analysed by interdisciplinary teams combining plant physiology, genetics, breeding, agronomy, data science, modelling and bioinformatics expertise (B). The analytical routes included standard parameter analysis, mechanism-informed trait engineering, stress-response differentials, raw-trace analysis, heritability estimation, genome-wide association studies (GWAS), yield association and machine-learning-based classification (C). The resulting insights highlight three main directions for field photosynthesis genetics: engineered traits can recover additional genetic signals, temporally- and signal-resolved features outperform extracted parameters for genotype predictions and raw fluorescence traces contain dynamic information not fully captured by standard extracted parameters (D).

To address this gap and to benchmark competing analytical approaches on identical datasets, we organised the first hackathon dedicated entirely to the photosynthesis–genomics intersection. Held in Accra, Ghana, in March 2026, the event brought together 35 participants from various fields. Five

independent teams applied distinct computational strategies to identical curated datasets for cowpea, barley, common bean, and potato, generating convergent and complementary insights that would be difficult to obtain within a single research group (summarised in Fig. 1). This paper reports the scientific outputs of that event.

Across independent analytical streams, convergent findings emerged around three themes: the heritability gains achievable through mechanism-informed feature engineering; the underexploited information content of raw fluorescence and absorbance trace data relative to already extracted parameters; and the power of machine learning to recover genetic signal when grounded in biological domain knowledge. These results are reported here as a community resource; detailed analytical treatments will follow in a series of original research articles.

2. The Photosynthesis Hackathon: Design and Scientific Challenge

The hackathon posed a single, well-defined challenge: identify photosynthetic parameters that reliably predict genetic components by using field-measured MultispeQ photosynthetic phenotypes paired with recorded ambient environmental data (e.g., PAR, air temperature, relative humidity) and genome-wide marker data. Teams worked across four crop species chosen to maximise biological and methodological breadth: cowpea (*Vigna unguiculata*), a drought-tolerant legume central to food security in sub-Saharan Africa (Silva et al., 2018); barley (*Hordeum vulgare*), a major temperate cereal which is often regarded as a model for cereals especially its close hexaploid relative wheat (Langridge, 2018); potato (*Solanum tuberosum*), a major staple crop with belowground storage organs (Lutaladio & Castaldi, 2009); and common bean (*Phaseolus vulgaris*) a nutritionally important grain legume and a dietary staple for millions of people (Broughton et al., 2003). All datasets were curated and delivered at the start of the event via openJII (an open-source IoT platform for recording and analysing large photosynthesis datasets, developed by the Jan IngenHousz Institute, www.openjii.org).

The programme ran for five days in Accra, Ghana, hosted by the International Institute of Tropical Agriculture (IITA): Day 1 established a shared biological and statistical foundation across disciplinary backgrounds; Days 2-4 were devoted to intensive team analysis; and Day 5 culminated in group presentations followed by planning for publication, of which this paper is the first output. Participants were 35 researchers at Master's level and above from 14 countries across Europe and Africa, spanning data science, plant physiology, genetics, breeding, bioinformatics, and science policy. Selection prioritised geographic and disciplinary diversity, with particular encouragement for African institutions; with all participants becoming authors of this paper.

3. Datasets and Shared Analytical Infrastructure

All measurements were collected using MultispeQ v2 devices (PhotosynQ Inc., East Lansing, MI, USA), and experiments were designed using the PhotosynQ platform by Kuhlger et al. (2016). Two measurement protocols were used depending on the dataset (see Fig. 2): the *Photosynthesis RIDES 2.0 protocol* (developed by David M. Kramer laboratory at the MSU-DOE Plant Research Laboratory, Michigan State University (MSU)), an often used protocol for MultispeQ v2 instruments running firmware ≥ 2.34 , and the *UNZA_PIRK_DIRK_LightPotential_14 protocol*, based on the light potential protocol developed by Kanazawa et al. (Kanazawa et al., 2021), which is a more elaborate protocol developed by the Jan IngenHousz Institute for usage at the 2025 University of Zambia and Jan IngenHousz Institute field trials.

3.1 Protocols

The *Photosynthesis RIDES 2.0* protocol delivers a complete measurement in approximately 15 seconds per leaf without requiring pre-acclimation by capturing ambient PAR and replicating it inside the leaf chamber using calibrated internal actinic LEDs, a key design feature that enables rapid sequential sampling across large field populations (Kuhlgert et al., 2016). The measurement sequence uses pulsed orange excitation light (605 nm, 10 μ s pulses at 100 Hz), a multiphase saturating flash via the 650 nm red LED (maximum intensity 8,000 μ mol photons $m^{-2} s^{-1}$), and a far-red illumination step (730 nm), producing a time-resolved fluorescence trace from which Φ_{II} , Φ_{NPQ} , Φ_{NO} , NPQ_i , and LEF are extracted, alongside electrochromic shift (ECS) signals recorded via dark-interval relaxation kinetics (DIRK) that estimate thylakoid proton motive force and ATP synthase conductance (Baker et al., 2007). The SPAD is recorded simultaneously via multi-wavelength absorbance. *Photosynthesis RIDES 2.0* has been validated and used for large-scale field phenotyping across multiple crop species, including common bean (Keller et al., 2024) and cowpea (Sansa et al., 2024). Throughout each measurement, ambient PAR, air temperature, relative humidity, leaf angle, and GPS coordinates are recorded automatically and stored with each observation, providing the environmental covariate dataset used in downstream modelling.

The *UNZA_PIRK_DIRK_LightPotential_14* protocol is an extension of the *Photosynthesis RIDES 2.0* protocol, adding additional light-response and relaxation blocks. Also, the multiphase saturating flash had a maximum of 6,000 μ mol photons $m^{-2} s^{-1}$. The initial actinic light period reproduces the ambient PAR, while the second actinic light period is fixed at 2500 μ mol photons $m^{-2} s^{-1}$. This allows assessment of how quickly chlorophyll fluorescence parameters respond to a sudden change from ambient to 2500 μ mol photons $m^{-2} s^{-1}$. During the far-red illumination, different light intensities are used, while chlorophyll fluorescence (excitation at 580nm, measurement at wavelengths > 730 nm) and absorbance (at about 820nm) traces are recorded. These additions extend the measurement from a rapid snapshot lasting approximately 15 s to a more detailed dynamic protocol that lasts approximately 25 s, which trades throughput for richer information and the short-term response of photosynthetic regulation to changes to high light conditions (see Fig. 2).

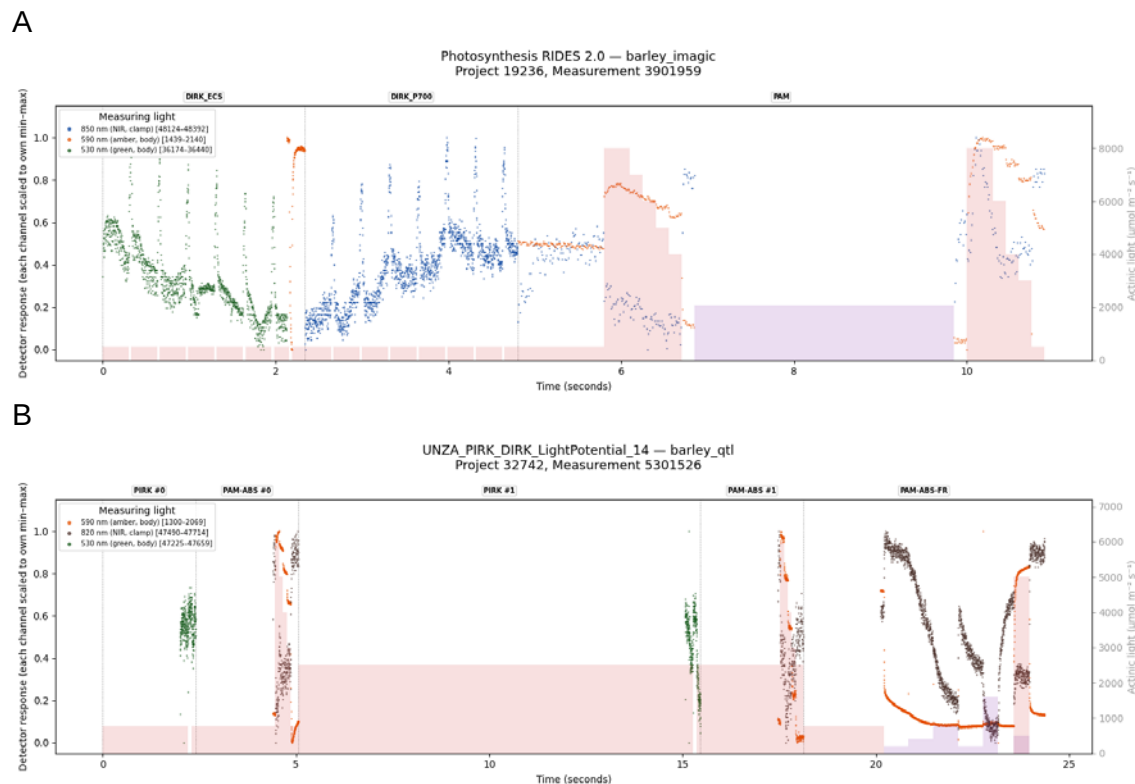


Fig. 2. MultispeQ measurement protocols used for the field photosynthesis phenotyping datasets used during the Hackathon. Schematic representation of the A) *Photosynthesis RIDES 2.0* and B) *UNZA_PIRK_DIRK_LightPotential_14*

protocols used in the analysed datasets. Both protocols record time-resolved chlorophyll fluorescence and absorbance traces together with environmental metadata, but differ in the sequence and duration of light pulses, saturating flashes and relaxation phases. Blocks of red (650nm) and purple (730nm) represent measuring light intensity and the dots represent the traces measured in green (530nm), amber (590nm) and NIR (820/850nm). Plots generated on the openJII platform ([openJII.org](https://openjii.org)).

3.2 Data Sets

Table 1 Summary of five data sets used during the 2026 photosynthesis Hackathon in Accra, Ghana.

	Cowpea (<i>Vigna unguiculata</i>)	Barley (<i>Hordeum vulgare</i>)	Barley (<i>Hordeum vulgare</i>)	Potato (<i>Solanum tuberosum</i>)	Bean (<i>Phaseolus vulgaris</i>)
Location	Ibadan, Ikenne, and Kano, Nigeria	Nergena, Netherlands	Wollo, and Gondar, Ethiopia	Wageningen, Netherlands	Chisamba, Zambia
Protocol	Photosynthesis_RIDES_2.0	UNZA-PIRK-DIRK_LightPotential_14	Photosynthesis_RIDES_2.0	UNZA-PIRK-DIRK_LightPotential_14	UNZA-PIRK-DIRK_LightPotential_14
Measurements	3,360	1,680	13,178	3,681	9,716
Genotypes	112	272	508	48	484
Type	Diversity population	3 RIL populations, subset of MAGIC population	MAGIC population	Diversity population	4 RIL populations
Conditions	3 locations Control & Drought stress	Population trial	2 locations 2 replicates	Population trial	Control & Drought stress
Yield data	Yes	Yes	No	Yes	No
Link to dataset on openJII	Sansa et al., 2024	https://openjii.org/en-US/platform/experiments/00c11895-cb24-4ffb-8497-cd8870cb2db5	Available after publication	https://openjii.org/en-US/platform/experiments/3f274b42-d006-4166-9c15-a501964920cd	Available after publication

Cowpea (Vigna unguiculata), Nigeria

The MultispeQ measurements collected from a cowpea diversity panel were evaluated across multiple environments in Nigeria using the *Photosynthesis RIDES 2.0*, as described by Sansa et al. (2024). This panel provides matched pairs of before- and during-stress measurements of two years and three locations, enabling the derivation of differential traits that capture stress-response dynamics independently of baseline values. Yield data of the 112 accessions were available as a separate file and linked to photosynthetic observations via genotype × environment identifiers. DArT-Seq single nucleotide polymorphism (SNP) markers were quality-filtered as part of the published study, and participants worked only with this filtered SNP dataset for genome-wide association studies (GWAS), consisting of 9210 markers distributed across the eleven chromosome of cowpea. The final dataset for GWAS analysis of photosynthetic traits comprised ninety-eight cowpea accessions evaluated across 10 genotype × environment combinations defined by location, year, and stress conditions.

Barley (Hordeum vulgare), Netherlands and Ethiopia

Two independent barley datasets were analysed. The dataset from the Netherlands was collected on three recombinant inbred line (RIL) populations, which are a subset of a multiparent advanced generation inter-cross (MAGIC) population (Casale et al., 2022). The populations consisted of 272 genotypes (3 parental lines, 267 progenies, and 2 elite genotypes used as repeated checks) and phenotyped using the *UNZA_PIRK_DIRK_LightPotential_14* protocol (see Fig. 2B). The Ethiopian dataset comprised measurements of 508 genotypes from a multiparental population from three environments across two locations (Wollo and Gondar), differing substantially in altitude and ambient

pressure. The data was collected using the *Photosynthesis_RIDES_2.0 protocol*. Both datasets included SNP marker data for association analysis.

Potato (Solanum tuberosum), Netherlands

The potato dataset was collected at eight time points within a single growing season in Wageningen, The Netherlands, providing a time-series structure absent from the other datasets. For this experiment, the *UNZA_PIRK_DIRK_LightPotential_14* protocol was used. A diverse population of forty-eight diploid potato genotypes was measured, enabling within-season trajectory analysis as well as static summaries. Due to single-season and single-location design, this dataset was used primarily for methodological demonstration of time-series approaches rather than definitive genetic mapping.

Common bean (Phaseolus vulgaris), Zambia

The common bean dataset focussed the photosynthetic behavior during drought stress. Four RIL populations with a total of 484 genotypes were grown in Chisamba, Zambia, during the dry season. The whole trial was irrigated, but at grain filling one half of the trial got progressively less irrigated, until no irrigation at all. For 8 days during this drought stress, MultispeQ measurements were done using the *UNZA_PIRK_DIRK_LightPotential_14* protocol.

4. Analytical Strategies Developed During the Hackathon

Five independent teams applied distinct but complementary analytical strategies to identical curated datasets, enabling cross-validation of findings that no single approach could have produced alone. Heritability and yield-association analyses (Section 5.1) combined linear mixed-effects models — with environment, replicate, and genotype $\times$ environment interaction as random effects, and ambient PAR, air temperature, and relative humidity as fixed-effect covariates — with genome-wide association (GWAS) analysis using Bayesian multi-locus methods that control for population stratification and kinship (Huang et al., 2019). A central methodological contribution was the treatment of PSII energy partitioning components (Φ_{II} , Φ_{NPQ} , and Φ_{NO}) as compositional data: because these three quantum yields sum to a fixed total, standard analysis of their absolute values can obscure genuine differences in energy allocation strategy. Isometric log-ratio (ILR) transformation was therefore applied prior to genetic analysis, alongside derivation of stress-response differential traits from paired before-stress and during-stress observations. Genotype-prediction analyses (Section 5.2) used random forest classifiers (Breiman, 2001) applied to both single time-point and temporally aggregated (AUC-transformed) photosynthetic traits. Raw-trace classification analyses (Section 5.3) used numerous machine learning approaches ranging from gradient-boosted classifiers and multi-layer perceptrons trained on extracted parameter sets to time-series foundation models applied directly to raw fluorescence traces.

5. Preliminary Results and Findings

Three themes structured the analyses and are reported here as the central outputs of this community effort. Detailed analyses, extended methods, and species-specific results will be reported in a series of follow-up papers currently in preparation.

5.1 Mechanism-informed feature engineering recovers heritable genetic signal

In both barley datasets, heritabilities for most isolated traits were near zero before environmental correction, consistent with strong masking by within- and between-day PAR, temperature, and humidity variation. After incorporating these environmental factors as fixed-effect covariates, heritabilities went up, implying a stronger recovered genetic signal. This confirms that environmental

correction is not optional in field fluorescence phenotyping, it is a prerequisite for genetic signal recovery.

Analysis of the cowpea dataset showed that engineered traits designed to represent the biological mechanisms governing fluorescence signals were strongly enriched among the highest-ranked traits for broad-sense heritability (H^2). In the cowpea diversity panel, comprising over 600 genotype $\times$ environment combinations across multiple Nigerian environments, ILR-transformed energy partitioning traits and log-ratio trade-off indices consistently occupied the top positions in the heritability ranking, alongside the extracted component Φ_{NO} which was the single most heritable trait ($H^2 = 0.557$). The extracted component NPQ_t , the parameter most commonly reported in field photosynthesis studies, ranked substantially lower ($H^2 \approx 0.25$ versus $H^2 > 0.50$ for the leading engineered traits) (see Fig. 3A). The most heritable traits reflected the $\Phi_{NPQ}:\Phi_{NO}$ trade-off (the balance between regulated photoprotection and unregulated energy loss) consistent with the interpretation that genotypes differ in energy allocation strategy rather than in the absolute magnitude of any single flux. Among LEF-derived traits, the composite index `lef_per_phi2` was near zero in heritability, suggesting that this particular representation of linear electron flow does not capture heritable genetic variation in this population under these conditions; however, `lef` and `lef_safe_index` (explained in Table S1) showed low-to-moderate heritability and should not be generalised to the same conclusion.

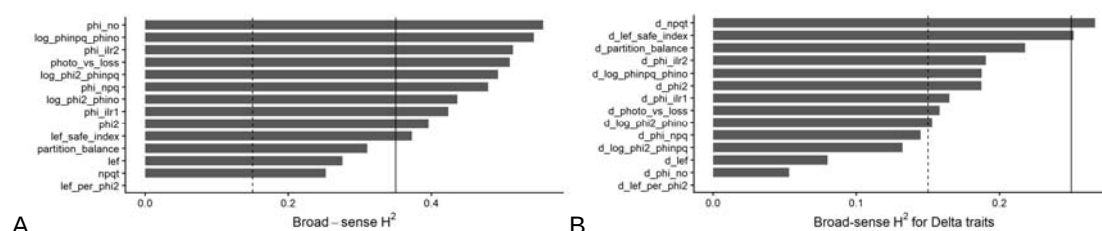


Fig. 3: Heritability of extracted versus engineered traits. (A) Broad-sense heritability for the baseline traits as well as (B) the estimated heritability for stress-response traits. Compositional data (`phi_2`, `phi_no`, `phi_npq`) were transformed into ILR coordinates using an orthonormal basis defined by sequential binary partition. The first balance (`ilr1`) contrasts `phi_2` against the geometric mean of Φ_{NO} and Φ_{NPQ} with a scaling factor of $\sqrt{2/3}$, while the second balance (`ilr2`) contrasts Φ_{NO} against Φ_{NPQ} with a scaling factor of $\sqrt{1/2}$.

The paired before and during stress measurement structure of the cowpea trial enabled the derivation of fourteen stress-response differential traits. These differential traits were heritable in their own right (H^2 reaching 0.27 for ΔNPQ_t), and their genetic architecture was demonstrably distinct from that of traits for either before or during stress separately, referred to as the baseline values. This can be seen in the difference between the heritability ranking across the Δ -trait panel as compared to the baseline ranking, with ΔNPQ_t , the magnitude of enhanced non-photochemical quenching induced under stress, outranking the ILR-based differential traits that dominated the baseline panel (see Fig. 3B). This reversal is physiologically coherent. Under ambient conditions, genotypic variation in the $\Phi_{NPQ}:\Phi_{NO}$ partition indicates differences in how absorbed energy is allocated between regulated photoprotective dissipation and non-regulated energy loss. Under acute stress, the primary variable distinguishing genotypes becomes whether they can activate sufficient total quenching capacity. The two trait classes therefore capture complementary aspects of photosynthetic performance that would be conflated under a single-measurement design. Genotype rankings for stress responsiveness and constitutive efficiency were not strongly correlated, providing a preliminary suggestion that breeding programmes targeting both dimensions would need to apply selection on both trait classes independently.

The engineered trade-off traits that led to the heritability ranking also showed the strongest associations with yield in the cowpea dataset. The $\Phi_{NPQ}:\Phi_{NO}$ log-ratio and its ILR equivalent both correlated negatively with yield ($r \approx -0.33$), indicating that genotypes with a higher proportion of regulated versus unregulated dissipation tend to yield less. Stress-response traits contributed independent yield-relevant signal beyond baseline values, including delta ILR traits and ΔNPQ_t , with fluorescence fixed effects collectively absorbing approximately 8% of genotypic yield variance in mixed-model analyses, a non-trivial proportion for a single phenotyping instrument applied in 15-second measurements per plant.

Genome-wide association analysis in cowpea recovered genomic regions exceeding the false discovery rate (FDR) significance threshold ($-\log_{10}(P) > 3.3$) for the engineered traits d_NPQt, d_LEF Safe Index, and d_Partition Balance. While GWAS revealed significant marker-trait associations (MTAs) for these engineered traits, including some strong MTA signals at a lower significance threshold, a GWAS hit for d_npqt was identified on chromosome 7 (Vu07_4745348), corresponding to the same genomic region previously reported by Sansa *et al.* (2024) for NPQt during stress (see Fig. 4). The candidate gene (*Vigun07g046450*), located within 71 bp of the associated SNP, has functional roles in regulating plant responses to abiotic stress. The independent recovery of this locus using a different trait, here an ILR-transformed stress-response differential rather than an extracted NPQ measurement, provides biological validation that the engineered transformations preserve and in some cases amplify genuine genetic signals. Several additional candidate regions not previously reported were also identified; these likely reflect biological information specific to the relative partitioning between regulated and unregulated dissipation that is inaccessible when analysing raw NPQ values, and are prioritised in a follow-up paper focused on GWAS results.

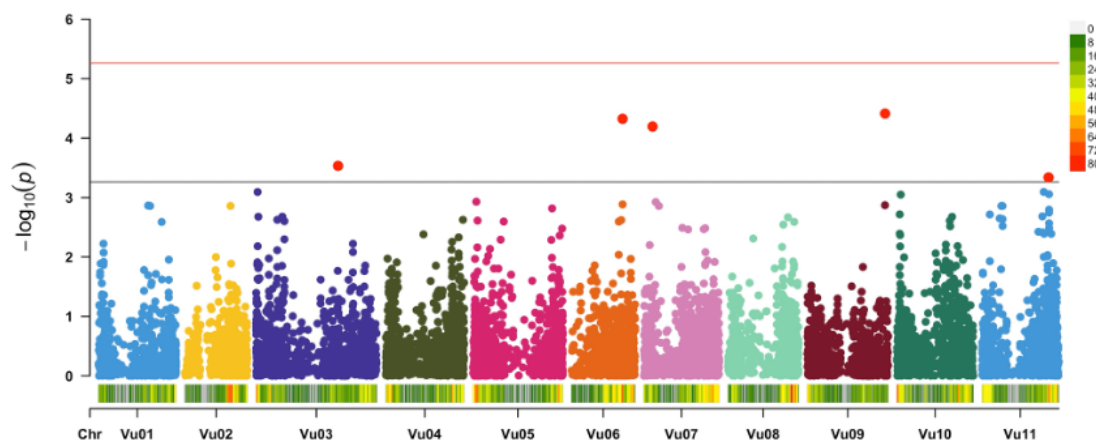


Fig. 4 Manhattan plot showing significant MTAs for d_npqt in the cowpea dataset. 9,210 DArT-Seq SNP markers distributed across the eleven cowpea chromosomes. The red line denotes Bonferroni threshold while black line denotes the false discovery rate (FDR) threshold for declaring significant MTAs. Vu: *Vigna Unguiculata*, chr: Chromosome

5.2 Temporally- and signal-resolved features outperform extracted parameters for genotype prediction

Another way to assess the genetic signal is to determine whether photosynthesis-related traits contain sufficient information for genotype prediction. To this end, a random forest classifier was applied (Breiman, 2001). The random forest classifier was configured with 500 trees and validated using 100 iterations of random subsampling, with 80% of the genotypes selected for training in each iteration.

The out-of-bag (OOB) accuracy was calculated as error, where the OOB error is the proportion of incorrectly classified OOB samples.

For each tree, a bootstrap sample of the same size as the original dataset was generated by sampling with replacement. Consequently, approximately one-third of the samples were excluded from the bootstrap sample and served as OOB observations for that tree. Each sample could therefore be evaluated using only those trees for which it was not included in the corresponding bootstrap sample. The OOB error was estimated from these classifications, ensuring that each evaluated sample was classified exclusively by trees that had not been trained on it. As a result, the OOB accuracy provides an unbiased estimate of the classifier's generalization performance.

For this approach, the potato dataset was used, which contains timeseries data on 48 genotypes (Table 1). Assuming random assignment, the probability of correctly classifying a sample by chance is $\frac{1}{48}$. Consequently, the expected number of correct classifications under random guessing is $3681 \times \frac{1}{48} \approx 77$ samples, that is corresponding to an expected accuracy of approximately 2.1%.

When using scaled features, the mean OOB accuracy across 100 random sub-samplings reached approximately 6.7%, that is already better than random guessing.

Additionally, feature importance was assessed using mean decrease in accuracy (MDA) and mean decrease in Gini impurity (MDG) (Nembrini et al., 2018). For MDA, each feature was randomly permuted in the OOB samples for each tree while all other features remained unchanged. The resulting decrease in OOB accuracy, relative to the unpermuted predictions, was calculated and averaged across all trees. Features causing larger decreases in OOB accuracy were considered more important for accurate classification.

MDG was derived from the decrease in Gini impurity during tree construction. The Gini impurity was defined as $1 - \sum_{k=1}^K p_k^2$, where p_k is the proportion of observations in class k . For each split, the reduction in Gini impurity attributable to the selected feature was calculated and accumulated across all trees. MDG therefore provided an alternative measure of feature importance, with larger values indicating greater contributions to class discrimination. Across all runs, SPAD consistently emerged as the most important feature, ranking first under both MDA and MDG (see Fig. 5).

Compared to single time-point measurements of photosynthetic traits, temporally resolved dynamic features collected over the course of a measurement day appear to contain more genomic information (Kaiser et al., 2015). Ambient light, in particular, can vary substantially within minutes, leading to measurements that increasingly reflect environmental fluctuations rather than purely genotypic effects (Gao et al., 2024). Measuring all genotypes within a narrow time window captures their response to the prevailing environmental conditions at that moment, embedding the genotypic effects within that environmental context. Repeating this across multiple time windows throughout the day then allows the genotype-specific responses to continuously changing conditions to be resolved. As measuring all genotypes at precisely the same time point is not achievable in practice with handheld devices such as the MultispeQ v2, residual noise from within-window environmental fluctuations cannot be fully avoided. However, under the assumption of sufficient environmental stability within each time window, this noise may remain acceptable. Therefore, an area-under-the-curve (AUC) transformation of traits within this interval may better capture genotype-dependent photosynthetic responses. This approach was demonstrated using Φ_{II} , which was measured repeatedly within a single day.

After smoothing the temporal trajectories and computing daily AUC values for Φ_{II} , inclusion of these aggregated measures in the random forest model increased the OOB accuracy to approximately 28%. The AUC feature ranked highest under the MDG criterion and second highest under MDA. This indicates that time-dependent variation captures more relevant genotypic information than individual

time-point measurements. However, the overall improvement in predictive performance remains limited, at least in part due to constraints in the dataset's temporal resolution.

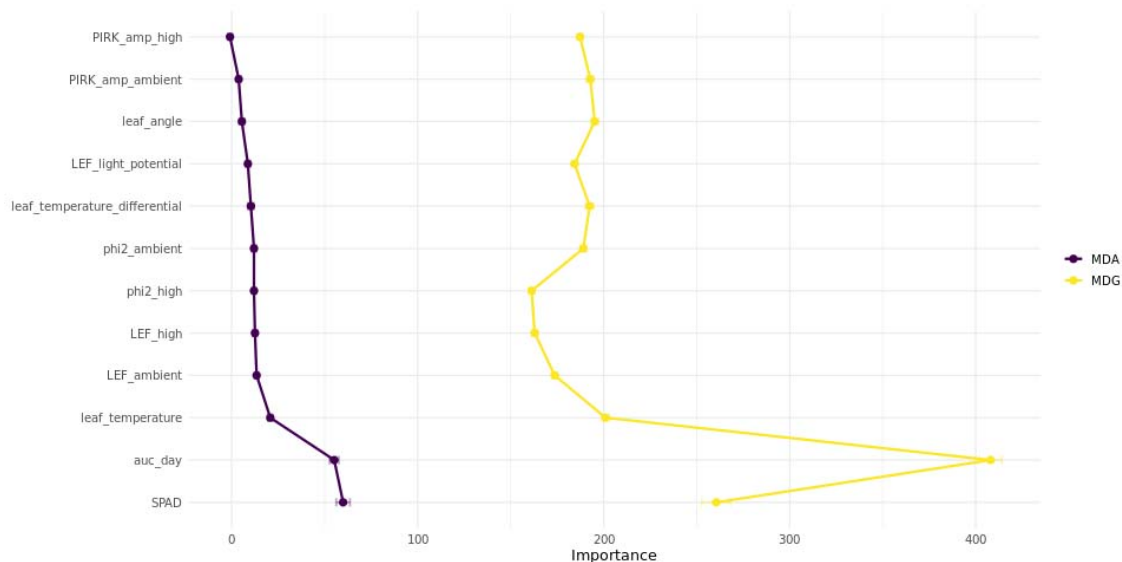


Fig. 5: Mean Decrease Accuracy (MDA) and Mean Decrease Gini (MDG) for all features included in the random forest classifier. For Φ_{II} (phi2_high), the area-under-the-curve (AUC) was calculated separately for each day. Data from one day (2025-08-14) were excluded due to insufficient observations.

In addition, the substantial covariance among fluorescence-derived traits, low heritabilities, and potentially complex interactions among physiological and environmental variables suggest that relevant genotype $\times$ environment interactions may not be fully captured by individual engineered features. Future analyses may therefore benefit from integrating harmonised multi-environment datasets through shared computational infrastructures such as OpenJII. Such datasets could facilitate factor-analytic approaches that incorporate photosynthetic traits as covariates to improve the characterisation of latent structures underlying genotype $\times$ environment interactions. These frameworks represent the covariance structure of genotype $\times$ environment effects through a reduced set of latent factors, enabling efficient dimensionality reduction while capturing major interaction patterns (Kelly et al., 2007). Initial exploratory analyses from the hackathon using the Netherlands barley dataset indicated that a factor-analytical model incorporating photosynthetic covariates, including LEF light potential, PAR, SPAD, and leaf temperature differential, explained a greater proportion of variance than a baseline model. PAR emerged as the dominant source of explained variance. However, only a single latent factor could be robustly estimated due to the dataset's environmental structure. Further evaluation across harmonised multi-environment datasets will be required to assess the broader applicability of this approach.

5.3 Raw fluorescence traces as a source of underused physiological information

The most striking finding across independent teams was that raw fluorescence traces retained substantially more predictive information than the extracted parameters extracted from them. In the common bean dataset, binary classifiers trained to distinguish stress-treated from control plants achieved 86% balanced accuracy using the 8 signal-separated raw traces from the *UNZA_PIRK_DIRK_LightPotential_14* protocol. For this prediction we captured performance variation across different raw trace combinations evaluated in the grid search, where *raw_pam_ambient* and *raw_pam_fr* contained most predictive power (see Fig. 6A). To see how these traces compare to the extracted parameter Φ_{II} , we compared the predictive power of Φ_{II} during ambient and high light with their respective raw traces. With Φ_{II} during ambient light we reach 60% balanced accuracy, while the

raw_pam_ambient trace reached 80% (see Fig. 6B). Combining both Φ_{II} during ambient and high light reaches 69%, which is increased to 83% when both respective traces are used. Although this result still requires validation across independent datasets, genotypes, and environments, it suggests that extracted parameters may discard dynamic information relevant to stress classification.

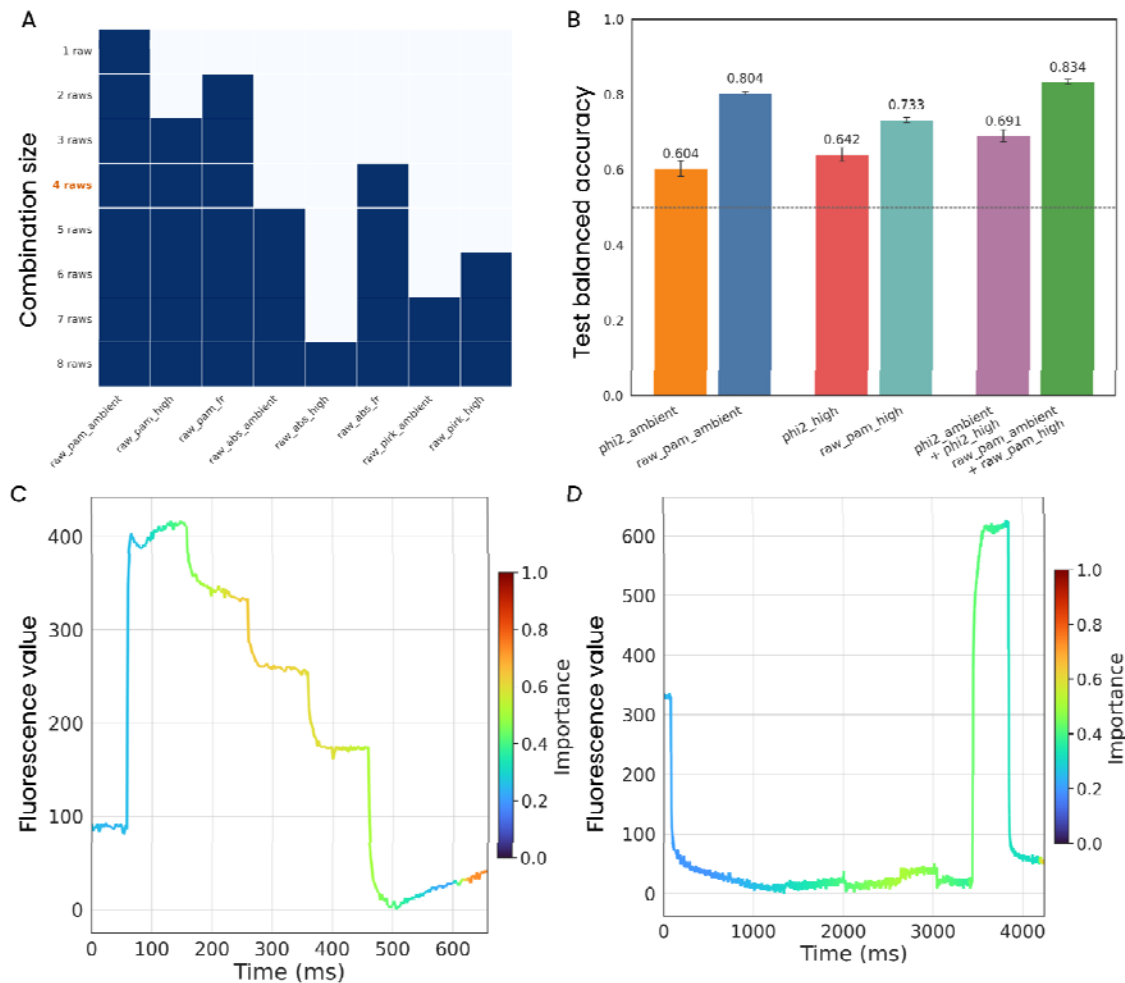


Fig. 6: The importance of different parts of the raw fluorescence traces for binary classifiers distinguishing between stress-treated and control plants in the bean dataset. Binary classifiers were trained with all possible combinations of different sections from the traces. (A) For each total number of sections included for training (combination size), the presence of a section in the best model of that combination size is denoted by the blue rectangles. The overall best model with a classification accuracy of 86% is highlighted in orange at four sections. The importance is indicated by the frequency of a section being present in the best models, with *raw_pam_ambient* being the most frequently present. (B) The predictive power of the extracted Φ_{II} during ambient and high light as compared to their respective raw traces. (C) The feature importance of the *raw_pam_ambient* trace, which is used to calculate the extracted Φ_{II} during ambient light. The feature importance is calculated as an average over all samples, when just this trace is used for the prediction. (D) The feature importance of the *raw_pam_fr_data* trace during the far-red associated section of the *UNZA_PIRK_DIRK_LightPotential_14* protocol. The feature importance is averaged over all samples, like in panel C.

By mapping the relative importance per trace, by averaging the importance over all samples, the middle section of the multi-phase flash pulse in the *raw_pam_ambient* trace and the rise kinetics of the saturating pulse and the response from low to high far-red light in the *raw_pam_fr_data* trace contain the most information (see Fig. 2, 6C, and 6D). Its temporal structure reflects defined phases of the measurement protocol, including baseline fluorescence, light-induced changes in PSII operating efficiency, NPQ induction or relaxation, and far-red-associated dynamics. These phases are linked to processes described by dynamic models of light harvesting, photoprotection, and electron transport (Stirbet et al., 2020, 2024). The fact that machine-learning models assigned high predictive

value to regions outside the standard extracted parameters, therefore, points to a biologically meaningful source of information, not merely to improved statistical classification. A promising next step is therefore to fit fluorescence traces using model-constrained learning approaches, such as physics-informed neural networks (Cuomo et al., 2022) or universal differential equations (Kidger, 2022). These methods could combine the structure of existing ordinary differential equation models with data-driven components for processes that are incompletely specified or difficult to parameterise (see Fig. 7). In practice, this would allow raw traces to be converted into model-derived traits, such as NPQ induction or relaxation rates, apparent electron transport regulation, or far-red relaxation kinetics. Such traits would be more suitable than black-box classifier outputs for heritability analysis, GWAS, and breeding applications.

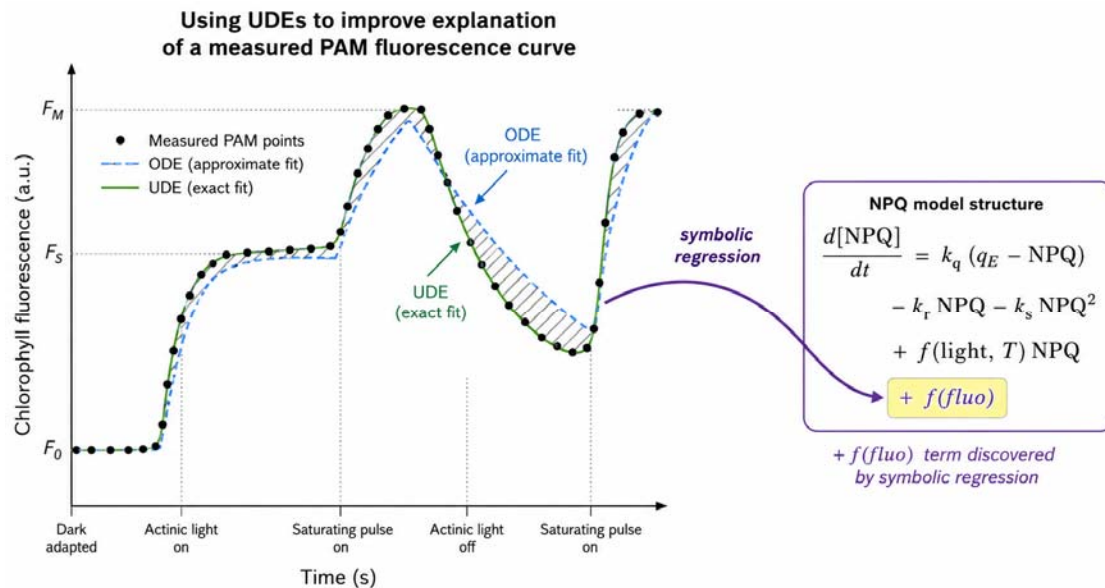


Fig. 7: Schematic overview of using universal differential equations (UDEs) to improve the interpretation of measured PAM fluorescence traces. Measured PAM fluorescence points are first fitted with a conventional ordinary differential equation model, which captures the overall trajectory but may miss parts of the dynamic response. A UDE-based model extends the mechanistic description by learning unresolved components of the fluorescence dynamics, resulting in a closer fit to the measured trace. Symbolic regression can then be applied to the learned UDE term to recover an interpretable functional relationship, as illustrated here as an additional light-dependent term in an NPQ model structure.

6. Lessons for Future Photosynthesis Data Challenges

Beyond the specific biological findings, the Accra hackathon highlighted several practical requirements for productive community analysis of field photosynthesis data. First, substantial pre-hackathon data curation was essential. Phenotypic, environmental, yield, and marker datasets had to be harmonised before participants could meaningfully compare analytical strategies. In a short collaborative event, unresolved differences in file formats, genotype identifiers, environmental metadata, and trait definitions can otherwise consume the time needed for analysis.

Second, the dedicated first day was critical for establishing a shared working language. Participants differed in their familiarity with fluorescence nomenclature, MultispeQ protocols, genetic analysis, and computational workflows. It is worth noting that many participants arrived with limited prior exposure to photosynthesis as a discipline. For several, the hackathon represented their first structured engagement with fluorescence-based measurements and the underlying biophysics. Yet this was paired with a strong motivation to incorporate photosynthetic traits into their own breeding and data

science work. For many, the event served as an eye-opening experience in both directions: revealing the genuine complexity of interpreting dynamic fluorescence signals, while simultaneously demonstrating that meaningful analysis is achievable within days when the right tools, data, and expertise are brought together. Short introductory sessions aligned the group on the biological meaning of the measured traits, the structure of the datasets, and the available analytical tools, allowing the following days to focus on analysis rather than disciplinary translation.

Third, shared computational infrastructure was useful, but not sufficient on its own. Providing curated datasets and analysis tools through openJII reduced installation barriers and gave participants a common starting point. However, the experience also showed that overly prescriptive environments can limit participation by requiring researchers to abandon their preferred analytical tools. Future events should therefore prioritise harmonised, analysis-ready datasets that can be easily downloaded and used in Python, R, or other environments while preserving a common data standard.

Fourth, mixed disciplinary teams shaped the analysis itself, not only the interpretation. Physiological and genetic expertise helped define biologically meaningful feature transformations, identify relevant environmental covariates, and reject statistically attractive but physiologically implausible outputs. Computational expertise enabled participants to test approaches not yet standard in field photosynthesis genetics, including compositional transformations, time-series analysis, and machine-learning methods applied directly to raw fluorescence traces.

The event also exposed limitations. Hackathons are well-suited to generating analytical prototypes, comparing alternative approaches, and identifying cross-cutting signals, but they are not a substitute for full validation. Several findings reported here therefore require extension in companion studies with formalised pipelines, independent validation sets, and fully documented model specifications. The main value of the format was to reveal, across independent teams and datasets, where current analytical practice is likely losing information. These lessons provide a practical template for future community efforts aimed at converting high-throughput photosynthesis phenotyping into genetically and physiologically interpretable traits.

7. Outlook and Conclusions

The Accra hackathon showed that field photosynthesis data, already collected at scale across research projects worldwide, contain substantially more analytical potential than is captured by standard parameter-based workflows. The event brought together complementary approaches to the same problem: *how to convert dynamic, environmentally sensitive fluorescence measurements into traits that are useful for genetic analysis and breeding*. Together, our results suggest that field photosynthesis genetics should move beyond fixed parameter extraction toward trait definitions that retain dynamic, environmental, and physiological structure. Yet, these findings should be interpreted as a foundation for further work rather than as the final analysis of each dataset. The hackathon format was valuable because it allowed independent teams to compare analytical strategies quickly and generate testable hypotheses, but did not provide enough time for formal validation or provision of fully documented pipelines. Clearly, independent test sets and species-specific follow-up studies are needed. The most immediate opportunities are to standardise analysis-ready photosynthesis phenotyping datasets and to develop interpretable features from raw fluorescence traces, and make them openly available via platforms such as OpenJII.

Finally, the long-term impact of this work will depend not only on improved analytical methods, but also on expanding the community able to use them. While high-throughput field photosynthesis measurements are becoming increasingly accessible, extracting biologically meaningful information from these datasets still requires specialised expertise in plant physiology, quantitative analysis, and computational methods. Realising the full potential of these datasets will therefore require both shared

expertise and effective mechanisms for transferring knowledge across disciplines and institutions. Broadening training in quantitative photosynthesis, computational analysis, and data interpretation will enable a much wider community of plant scientists and breeders to exploit these datasets, increasing the return on existing phenotyping efforts. Community teaching resources that integrate education with interactive analysis and mechanistic modelling such as the one proposed by Philipps *et al.* (Philipps *et al.*, 2025) or Ross *et al.* (Ross *et al.*, 2006), provide one route towards this goal, lowering the barrier to entry while promoting reproducible and physiologically informed analyses.

More broadly, this work argues for a shift in field photosynthesis genetics from extracting a fixed set of standard parameters toward designing traits that preserve physiological dynamics and environmental context. High-throughput fluorescence phenotyping can already generate the necessary data. The next advance will depend on whether these data can be converted into traits that are statistically robust, mechanistically interpretable, and useful for selection.

Acknowledgements

The authors gratefully acknowledge funding from the Jan IngenHousz Institute, the Mastercard Foundation, University of Cambridge Climate Resilience and Sustainability Research Fund. We thank the International Institute of Tropical Agriculture (IITA) for hosting the event in Accra, Ghana, and the entire IITA local team for their generous support throughout. We acknowledge Yanrong Gao, Po-Ya Wu, Shizue Matsubara and Benjamin Stich for providing datasets that did not get used during the Hackathon. We also thank Lutz Kupferschläger (RWTH Aachen University) for the design of the overview illustration (Figure 1) for this manuscript.

References

- Ainsworth, E. A., & Long, S. P. (2021). 30 years of free-air carbon dioxide enrichment (FACE): What have we learned about future crop productivity and its potential for adaptation? *Global Change Biology*, 27(1), 27–49. <https://doi.org/10.1111/gcb.15375>
- Ali, W., Grzybowski, M., Torres-Rodríguez, J. V., Li, F., Shrestha, N., Mathivanan, R. K., De Bernardeaux, G., Hoang, K., Mural, R. V., Roston, R. L., Schnable, J. C., & Sahay, S. (2025). Quantitative genetics of photosynthetic trait variation in maize. *Journal of Experimental Botany*, 76(14), 4141–4153. <https://doi.org/10.1093/jxb/eraf198>
- Avenson, T. J., Cruz, J. A., Kanazawa, A., & Kramer, D. M. (2005). Regulating the proton budget of higher plant photosynthesis. *Proceedings of the National Academy of Sciences*, 102(27), 9709–9713. <https://doi.org/10.1073/pnas.0503952102>
- Baker, N. R., Harbinson, J., & Kramer, D. M. (2007). Determining the limitations and regulation of photosynthetic energy transduction in leaves. *Plant, Cell & Environment*, 30(9), 1107–1125. <https://doi.org/10.1111/j.1365-3040.2007.01680.x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>

538 Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., & Vanderleyden, J. (2003). Beans
539 (*Phaseolus* spp.) – model food legumes. *Plant and Soil*, 252(1), 55–128.
540 <https://doi.org/10.1023/A:1024146710611>

541 Casale, F., Van Inghelandt, D., Weisweiler, M., Li, J., & Stich, B. (2022). Genomic prediction of the
542 recombination rate variation in barley – A route to highly recombinogenic genotypes. *Plant*
543 *Biotechnology Journal*, 20(4), 676–690. <https://doi.org/10.1111/pbi.13746>

544 Chau, C. W., & Gerber, E. M. (2023). On Hackathons: A Multidisciplinary Literature Review.
545 *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21.
546 <https://doi.org/10.1145/3544548.3581234>

547 Cuomo, S., Di Cola, V. S., Giampaolo, F., Rozza, G., Raissi, M., & Piccialli, F. (2022). Scientific Machine
548 Learning Through Physics-Informed Neural Networks: Where we are and What's Next. *Journal of*
549 *Scientific Computing*, 92(3), 88. <https://doi.org/10.1007/s10915-022-01939-z>

550 Edwards, G. E., & Baker, N. R. (1993). Can CO₂ assimilation in maize leaves be predicted accurately
551 from chlorophyll fluorescence analysis? *Photosynthesis Research*, 37(2), 89–102.
552 <https://doi.org/10.1007/BF02187468>

553 Gao, Y., Stein, M., Oshana, L., Zhao, W., Matsubara, S., & Stich, B. (2024). Exploring natural genetic
554 variation in photosynthesis-related traits of barley in the field. *Journal of Experimental Botany*, 75(16),
555 4904–4925. <https://doi.org/10.1093/jxb/erae198>

556 Genty, B., Briantais, J.-M., & Baker, N. R. (1989). The relationship between the quantum yield of
557 photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochimica et Biophysica*
558 *Acta (BBA) - General Subjects*, 990(1), 87–92. [https://doi.org/10.1016/S0304-4165\(89\)80016-9](https://doi.org/10.1016/S0304-4165(89)80016-9)

559 Genty, B., Goulas, Y., Dimon, B., & Peltier, G. (1992, January 1). *Modulation of efficiency of primary*
560 *conversion in leaves, mechanisms involved at PS2*. TIB-Portal. 9th International congress on
561 photosynthesis, Research in photosynthesis. <https://www.tib.eu/de/suchen/id/BLCP%3ACN000864547>

562 Huang, M., Liu, X., Zhou, Y., Summers, R. M., & Zhang, Z. (2019). BLINK: A package for the next level
563 of genome-wide association studies with both individuals and markers in the millions. *GigaScience*, 8(2),
564 giy154. <https://doi.org/10.1093/gigascience/giy154>

565 Kaiser, E., Morales, A., Harbinson, J., Kromdijk, J., Heuvelink, E., & Marcelis, L. F. M. (2015). Dynamic
566 photosynthesis in different environmental conditions. *Journal of Experimental Botany*, 66(9), 2415–2426.
567 <https://doi.org/10.1093/jxb/eru406>

568 Kanazawa, A., Chattopadhyay, A., Kuhlger, S., Tuitupou, H., Maiti, T., & Kramer, D. M. (2021). Light
569 potentials of photosynthetic energy storage in the field: What limits the ability to use or dissipate rapidly
570 increased light energy? *Royal Society Open Science*, 8(12), 211102.
571 <https://doi.org/10.1098/rsos.211102>

Kanazawa, A., & Kramer, D. M. (2002). *In vivo* modulation of nonphotochemical exciton quenching (NPQ) by regulation of the chloroplast ATP synthase. *Proceedings of the National Academy of Sciences*, 99(20), 12789–12794. <https://doi.org/10.1073/pnas.182427499>

Keller, B., Soto, J., Steier, A., Portilla-Benavides, A. E., Raatz, B., Studer, B., Walter, A., Muller, O., & Urban, M. O. (2024). Linking photosynthesis and yield reveals a strategy to improve light use efficiency in a climbing bean breeding population. *Journal of Experimental Botany*, 75(3), 901–916. <https://doi.org/10.1093/jxb/erad416>

Kelly, A. M., Smith, A. B., Eccleston, J. A., & Cullis, B. R. (2007). The Accuracy of Varietal Selection Using Factor Analytic Models for Multi-Environment Plant Breeding Trials. *Crop Science*, 47(3), 1063–1070. <https://doi.org/10.2135/cropsci2006.08.0540>

Kidger, P. (2022). *On Neural Differential Equations* (Version 1). arXiv. <https://doi.org/10.48550/ARXIV.2202.02435>

Kramer, D. M., Johnson, G., Kiirats, O., & Edwards, G. E. (2004). New Fluorescence Parameters for the Determination of QA Redox State and Excitation Energy Fluxes. *Photosynthesis Research*, 79(2), 209–218. <https://doi.org/10.1023/B:PRES.0000015391.99477.0d>

Kromdijk, J., Glowacka, K., Leonelli, L., Gabilly, S. T., Iwai, M., Niyogi, K. K., & Long, S. P. (2016). Improving photosynthesis and crop productivity by accelerating recovery from photoprotection. *Science*, 354(6314), 857–861. <https://doi.org/10.1126/science.aai8878>

Kromdijk, J., & Long, S. P. (2016). One crop breeding cycle from starvation? How engineering crop photosynthesis for rising CO₂ and temperature could be one important route to alleviation. *Proceedings of the Royal Society B: Biological Sciences*, 283(1826), 20152578. <https://doi.org/10.1098/rspb.2015.2578>

Kuhlgert, S., Austic, G., Zegarac, R., Osei-Bonsu, I., Hoh, D., Chilvers, M. I., Roth, M. G., Bi, K., TerAvest, D., Weebadde, P., & Kramer, D. M. (2016). MultispeQ Beta: A tool for large-scale plant phenotyping connected to the open PhotosynQ network. *Royal Society Open Science*, 3(10), 160592. <https://doi.org/10.1098/rsos.160592>

Langridge, P. (2018). Economic and Academic Importance of Barley. In N. Stein & G. J. Muehlbauer (Eds.), *The Barley Genome* (pp. 1–10). Springer International Publishing. https://doi.org/10.1007/978-3-319-92528-8_1

Li, F., Grzybowski, M., Roston, R. L., & Schnable, J. C. (2025). Nighttime fluorescence phenotyping reduces environmental variability for photosynthetic traits and enables the identification of candidate loci in maize. *Frontiers in Plant Science*, 16, 1595339. <https://doi.org/10.3389/fpls.2025.1595339>

Long, S. P., Wang, Y., Carmo-Silva, E., Cavanagh, A. P., Jonikas, M. C., Kromdijk, J., Long, B. M., Marshall-Colón, A., Shukla, D., Wilson, R. H., Zhu, X.-G., & Ainsworth, E. A. (2025). Feeding from the

sun—Successes and prospects in bioengineering photosynthesis for food security. *Cell*, 188(24), 6700–6719. <https://doi.org/10.1016/j.cell.2025.10.033>

Lutaladio, N., & Castaldi, L. (2009). Potato: The hidden treasure. *Journal of Food Composition and Analysis*, 22(6), 491–493. <https://doi.org/10.1016/j.jfca.2009.05.002>

Monteith, J. L. (1977). Climate and the efficiency of crop production in Britain. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 281(980), 277–294. <https://doi.org/10.1098/rstb.1977.0140>

Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics*, 34(21), 3711–3718. <https://doi.org/10.1093/bioinformatics/bty373>

Ort, D. R., Merchant, S. S., Alric, J., Barkan, A., Blankenship, R. E., Bock, R., Croce, R., Hanson, M. R., Hibberd, J. M., Long, S. P., Moore, T. A., Moroney, J., Niyogi, K. K., Parry, M. A. J., Peralta-Yahya, P. P., Prince, R. C., Redding, K. E., Spalding, M. H., Van Wijk, K. J., ... Zhu, X. G. (2015). Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proceedings of the National Academy of Sciences*, 112(28), 8529–8536. <https://doi.org/10.1073/pnas.1424031112>

Philipps, S., Pfennig, T., Corvest, E., Van Aalst, M., Fürtauer, L., & Matuszyńska, A. (2025). Computational Photosynthesis (ComPhot): Simulation-Based Learning Platform to Study Photosynthesis. *The Plant Cell*, 37(6), koae101. <https://doi.org/10.1093/plcell/koae101>

Ross, P. auline, Tronson, D., & Ritchie, R. aymond J. (2006). Modelling Photosynthesis to Increase Conceptual Understanding. *Journal of Biological Education*, 40(2), 84–88. <https://doi.org/10.1080/00219266.2006.9656019>

Salesse-Smith, C. E., Sharwood, R. E., Busch, F. A., Kromdijk, J., Bardal, V., & Stern, D. B. (2018). Overexpression of Rubisco subunits with RAF1 increases Rubisco content in maize. *Nature Plants*, 4(10), 802–810. <https://doi.org/10.1038/s41477-018-0252-4>

Sansa, O., Abberton, M. T., Ariyo, J., Paliwal, R., Ige, A., Dieng, I., Ayo-Vaughan, M., Olowe, V. I., & Oyatomi, O. (2024). Genome-wide association studies of photosynthetic and agronomic traits in cowpea collection. *G3: Genes, Genomes, Genetics*, jkae233. <https://doi.org/10.1093/g3journal/jkae233>

Silva, A. C. da, Santos, D. da C., Junior, D. L. T., Silva, P. B. da, & Siviero, R. C. dos S. and A. (2018). Cowpea: A Strategic Legume Species for Food Security and Health. In *Legume Seed Nutraceutical Research*. IntechOpen. <https://doi.org/10.5772/intechopen.79006>

Smith, E. N., van Aalst, M., Tosens, T., Niinemets, Ü., Stich, B., Morosinotto, T., Alboresi, A., Erb, T. J., Gómez-Coronado, P. A., & Tolleter, D. (2023). Improving photosynthetic efficiency toward food security: Strategies, advances, and perspectives. *Molecular Plant*, 16(10), 1547–1563.

Smith, E. N., Van Aalst, M., Weber, A. P. M., Ebenhöf, O., & Heinemann, M. (2025). Alternatives to photorespiration: A system-level analysis reveals mechanisms of enhanced plant productivity. *Science Advances*, 11(13), eadt9287. <https://doi.org/10.1126/sciadv.adt9287>

Sotaquirá-Gutiérrez, R., Beltran, L. M., & Garzon Ruiz, J. P. (2025). Hackathons as experiential learning platforms for engineering design skills. *Cogent Education*, 12(1), 2442187. <https://doi.org/10.1080/2331186X.2024.2442187>

South, P. F., Cavanagh, A. P., Liu, H. W., & Ort, D. R. (2019). Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field. *Science*, 363(6422), eaat9077. <https://doi.org/10.1126/science.aat9077>

Stirbet, A., Guo, Y., Lazár, D., & Govindjee, G. (2024). From leaf to multiscale models of photosynthesis: Applications and challenges for crop improvement. *Photosynthesis Research*, 161(1–2), 21–49. <https://doi.org/10.1007/s11120-024-01083-9>

Stirbet, A., Lazár, D., Guo, Y., & Govindjee, G. (2020). Photosynthesis: Basics, history and modelling. *Annals of Botany*, 126(4), 511–537. <https://doi.org/10.1093/aob/mcz171>

Theeuwes, T. P. J. M., Logie, L. L., Harbinson, J., & Aarts, M. G. M. (2022). Genetics as a key to improving crop photosynthesis. *Journal of Experimental Botany*, 73(10), 3122–3137. <https://doi.org/10.1093/jxb/erac076>

Tietz, S., Hall, C. C., Cruz, J. A., & Kramer, D. M. (2017). NPQ_(T): A chlorophyll fluorescence parameter for rapid estimation and imaging of non-photochemical quenching of excitons in photosystem II-associated antenna complexes. *Plant, Cell & Environment*, 40(8), 1243–1255. <https://doi.org/10.1111/pce.12924>

Van Aalst, M., Pfennig, T., Chirove, F., Ronoh, M., & Matuszyńska, A. (2025). Hackathons as essential tools in an interdisciplinary biological training—Report from trainings for Sub-saharan students. <https://doi.org/10.1101/2025.03.27.645823>

Yoon, D.-K., Ishiyama, K., Suganami, M., Tazoe, Y., Watanabe, M., Imaruoka, S., Ogura, M., Ishida, H., Suzuki, Y., Obara, M., Mae, T., & Makino, A. (2020). Transgenic rice overproducing Rubisco exhibits increased yields with improved nitrogen-use efficiency in an experimental paddy field. *Nature Food*, 1(2), 134–139. <https://doi.org/10.1038/s43016-020-0033-x>

Zhu, X.-G., Long, S. P., & Ort, D. R. (2008). What is the maximum efficiency with which photosynthesis can convert solar energy into biomass? *Current Opinion in Biotechnology*, 19(2), 153–159. <https://doi.org/10.1016/j.copbio.2008.02.004>

671