PNAS RESEARCH ARTICLE | PLANT BIOLOGY OPEN ACCESS

Species-wide inventory of *Arabidopsis thaliana* organellar variation reveals ample phenotypic variation for photosynthetic performance

Tom P. J. M. Theeuwen[a,1,2], Raúl Y. Wijfjes[b,3], Delfi Dorussen[a,4], Aaron W. Lawson[a,5], Jorrit Lind[a], Kaining Jin[a,6], Janhenk Boekeloo[a], Dillian Tijink[a], David Hall[c], Corrie Hanhart[a], Frank F. M. Becker[a], Fred A. van Eeuwijk[d], David M. Kramer[c,1], Erik Wijnker[a,7], Jeremy Harbinson[e], Maarten Koornneef[a,2], and Mark G. M. Aarts[a]

Affiliations are included on p. 11.

Contributed by Maarten Koornneef; received July 12, 2024; accepted October 28, 2024; reviewed by Wolfgang Busch and Johannes Kromdijk

Efforts to improve photosynthetic performance are increasingly employing natural genetic variation. However, genetic variation in the organellar genomes (plasmotypes) is often disregarded due to the difficulty of studying the plasmotypes and the lack of evidence that this is a worthwhile investment. Here, we systematically phenotyped plasmotype diversity using *Arabidopsis thaliana* as a model species. A reanalysis of whole-genome resequencing data of 1,541 representative accessions shows that the genetic diversity among the mitochondrial genomes is eight times lower than among the chloroplast genomes. Plasmotype diversity of the accessions divides the species into two major phylogenetic clusters, within which highly divergent subclusters are distinguished. We combined plasmotypes from 60 *A. thaliana* accessions with the nuclear genomes (nucleotypes) of four *A. thaliana* accessions to create a panel of 232 cytonuclear genotypes (cybrids). The cybrid plants were grown in a range of different light and temperature conditions and phenotyped using high-throughput phenotyping platforms. Analysis of the phenotypes showed that several plasmotypes alone or in interaction with the nucleotypes have significant effects on photosynthesis and that the effects are highly dependent on the environment. Moreover, we introduce Plasmotype Association Studies (PAS) as a method to reveal plasmotypic effects. Within *A. thaliana,* several organellar variants can influence photosynthetic phenotypes, which emphasizes the valuable role this variation has on improving photosynthetic performance. The increasing feasibility of producing cybrids in various species calls for further research into how these phenotypes may support breeding goals in crop species.

organellar variation | photosynthesis | cybrids | high-throughput phenotyping

Significance

Photosynthesis is one of the few crop traits that has been largely unaddressed by plant breeding which can contribute to increasing crop yield potential. Exploiting genetic variation within organellar genomes presents a promising, yet untapped resource to improve photosynthesis. However, the extent of organellar variation and its impact on photosynthesis within a species remain largely unknown. Using *Arabidopsis thaliana* as a model species, we revealed highly divergent clusters of organellar variation. We constructed 232 combinations of species-representative organellar and nuclear genomes, referred to as cybrids. High-throughput phenotyping of these cybrids revealed that organellar variants can substantially impact photosynthesis in some environments. These findings indicate that organellar genome variation may be a valuable resource for improving photosynthesis in crops.

Improving photosynthetic performance has become an important goal for increasing crop yield to meet global food demands in the coming decades (1–4). Genetic variation is often overlooked in studies on improving photosynthetic performance (5, 6) due to the limited variation in genes encoding the core photosynthetic machinery (3). Recently, evidence is accumulating that genetic variation can contribute to photosynthetic differences within species (7–18). Even though this research area is rapidly gaining more attention, the contribution of the organellar genomes to photosynthetic variation remains unclear. In plants, this organellar genetic variation is divided between two organelles that each have their own genome, the chloroplast and the mitochondrion. Derived from ancient endosymbiotic ancestors, the genomes of the extant organelles are drastically reshaped by evolution. Thousands of genes were lost or moved to the nuclear genome, while only some genes have been retained in the organellar genomes (19). The chloroplast and mitochondrial genomes of *Arabidopsis thaliana* only harbor 129 and 57 genes, respectively (20, 21). The retention of genes in the organellar genomes may be explained by the need for cross talk between the organellar and nuclear genomes; suppressed mutation rate in organellar genomes; difficulty in importing proteins over the organellar membranes; or the requirement of redox-regulated gene expression (22). The retained genes primarily encode components of the photosynthetic and respiration machinery and essential components needed for organellar gene expression.

Genetic variation in the organellar genomes may be limited as a result of purifying selection and neutrality of variants (23). This would explain the strong conservation of organellar genes across plant species and the observations that mutations in organellar genes often lead to severe, debilitating phenotypes (24, 25). The reduced number of nonsynonymous mutations in comparison to synonymous mutations is further evidence

[1]Present address: Jan IngenHousz Institute, Wageningen 6708 PB, The Netherlands.

[2]To whom correspondence may be addressed. Email: tom.theeuwen@wur.nl or maarten.koornneef@wur.nl.

[3]Present address: Institute of Genetics, Faculty of Biology, Ludwig-Maximilians-Universität Munich, Planegg-Martinsried 82152, Germany.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2414024121/-/DCSupplemental.

Published November 27, 2024.

that purifying selection is prominent in organellar genomes (26, 27). The mutation rate of a plant mitochondrial genome is generally regarded to be approximately sixfold lower than that of a nuclear genome, while the chloroplast genome mutation rate is about twofold lower than that of the nuclear genome (28). Over time though, mutations are expected to occur, causing organellar genetic variation that can be selected for in natural environments. An example of organellar adaptation is the mutation in *PsbA*, which encodes the D1 protein of photosystem II (PSII), that confers herbicide tolerance to *A. thaliana* (29). This mutation has been selected for on the British Railway network through application of herbicides (30). Another example is the *RbcL* gene, encoding the Rubisco large subunit, where several amino acid residues are under positive selection in land plants (31). Organellar variation in *Arabidopsis lyrata* is also found to increase fitness in specific environments (32). Systematic analyses of organellar variations within a species are rare, but studies using *A. thaliana* suggest that cytoplasmic variation is widespread (33–36). This shows that organellar variation can contribute to plant adaptation, even though it is slower and more limited than adaptation based on nuclear variation.

To assess the role of organellar genetic variation in photosynthetic performance, a systematic analysis of organellar genotype and phenotype variation at the species level is needed. However, the cytoplasm is uniparentally inherited, which complicates efforts to separate the nuclear and cytoplasmic origins of phenotypic variation. Even though reciprocal hybrids and segregating populations can give a good indication of cytoplasmic effects, maternal effects and genomic imprinting are known to play a significant role and could be interpreted as false-positive cytoplasmic effects (37–39). To exclude maternal effects and genomic imprinting, recurrent backcrossing can be used. This approach allows the production of novel combinations (cybrids) of nuclear genomes (nucleotypes) and organellar genomes (plasmotypes). Analysis of the resulting cybrids has indeed shown that phenotypic variation can be associated with organellar genomes in different plant species (40–42). Backcrossing approaches are particularly useful in plant breeding programs, but they are lengthy and residual nuclear introgressions that can influence phenotypic variation may be remaining. Flood et al. (43) showed that in *A. thaliana*, a maternal haploid induction system can be used as an attractive alternative method to create cybrids without a lengthy backcrossing procedure (43). Such maternal haploid induction system is based on a *GFP-tailswap* mutant, where the centromere-specific histone 3 (*CenH3*) gene is replaced by a *GFP*-tagged *CenH3* gene. A *GFP-tailswap* mutant can be used to replace the complete nucleotype within one generation (43, 44). This method can efficiently be used to systematically separate phenotypic variation caused by nucleotype variation from that caused by plasmotype variation.

Flood et al. (43) showed that plasmotypic variation can result in significant differences in photosynthetic performance (43). These so far unknown phenotypic differences were found to be caused either by variation in the plasmotype alone or as a result of an interaction between plasmotype and nucleotype variation. The phenotypic differences were most often observed in dynamic light conditions. This showed that photosynthesis is affected by plasmotype variation, but the seven accessions of the cybrid panel represented a limited fraction of the plasmotype diversity to be found in *A. thaliana*. As *A. thaliana* is native to most of the Eurasian landmass and high-altitude regions of Africa, accessions originating from vastly different environmental conditions are available (35, 45–49). The availability of accessions that potentially represent the global species-wide genetic diversity makes *A. thaliana* an excellent model system to explore the extent to which plasmotypic variation contributes to overall phenotypic variation and more specifically photosynthetic variation.

Here, we reanalyze the whole-genome sequencing data of 1,541 publicly available *A. thaliana* accessions to obtain high-quality organellar genetic variation. Using this dataset, we reveal how much genetic variation is present within the species and how it is distributed over the organelles. From this dataset, a selection of 60 genetically diverse, species-representative plasmotypes were combined with four distinct and diverse nucleotypes to construct a panel of 232 cybrids. To characterize and understand the contribution of plasmotype variation to photosynthetic variation, we phenotyped this cybrid panel in detail using three different experimental setups in which they were exposed to a range of light and temperature conditions.

Results

Plasmotypic Variation Analysis. To quantify the plasmotype diversity among *A. thaliana* accessions, we reanalyzed available sequencing data of 1,541 accessions collected in Europe, Asia, and Africa. Variant calling for the organellar genomes of a subset was performed previously, but quality control filtering steps were applied as tailored to nuclear genome analyses. Such filtering can result in low-quality variant data, since organellar genomes have different intrinsic properties, for example, higher copy number and heterozygosity. To compare variants among accessions without introducing biases from separate variant calling projects, we performed variant calling on all accessions simultaneously, with quality filters separately tailored to the chloroplast and mitochondria. The updated version of the mitochondrial reference genome was used for variant calling (21).

We first filtered out poor-quality raw sequencing data and accessions with a high percentage of heterozygous calls in the chloroplast genome, which are indicative of sample contaminants. This analysis resulted in variant calling data for the organellar genomes of 1,531 *A. thaliana* accessions. The dataset comprises 5,015 variants in the chloroplast genomes and 1,430 variants in the mitochondrial genomes, compared to the Col-0 reference genome. Resultingly, 3.2% of the nucleotides in the chloroplast genome and 0.4% of the nucleotides in the mitochondrial genome showed genetic variation within at least one accession. Therefore, the genetic diversity within the chloroplast genomes is eightfold higher compared to the mitochondrial genomes. Specifically, in chloroplast genes, the ratio between predicted nonsynonymous mutations versus synonymous mutations is 0.9, while for mitochondrial genic regions, this is 1.8. Thus, we conclude that the genetic diversity in the chloroplasts is higher than that in the mitochondria, but that the chloroplast genes are more conserved than the mitochondrial genes.

Next, we studied the plasmotype diversity between accessions to reveal phylogenetic relationships. The mitochondrial genome is known to be more dynamic than the chloroplast genome (50, 51). This results in extensive interspecies modifications in the mitochondrial genome (52), and therefore may not represent the true phylogenetic origin (33, 53, 54). Therefore, we focused on the chloroplast genomes to reconstruct the phylogeny of *A. thaliana*. We found that among the 1,531 accessions, 1,495 unique chloroplast genotypes could be identified, meaning that most accessions have unique chloroplast genomes. Principal component (PC) analysis of the chloroplast variation revealed three distinct main clusters diverging from one central point, but the fraction of variation explained by PC1 (2.1%) and PC2 (1.5%) is low (Fig. 1*A*). The low fraction explained by the main PCs indicates there is substantial genetic variation between the chloroplast genomes. Rooting the neighbor-joining tree based on the chloroplast genome with *A. lyrata*, *Arabidopsis halleri*, *Arabidopsis carpatica,* and *Capsella rubella* shows that *A. thaliana*

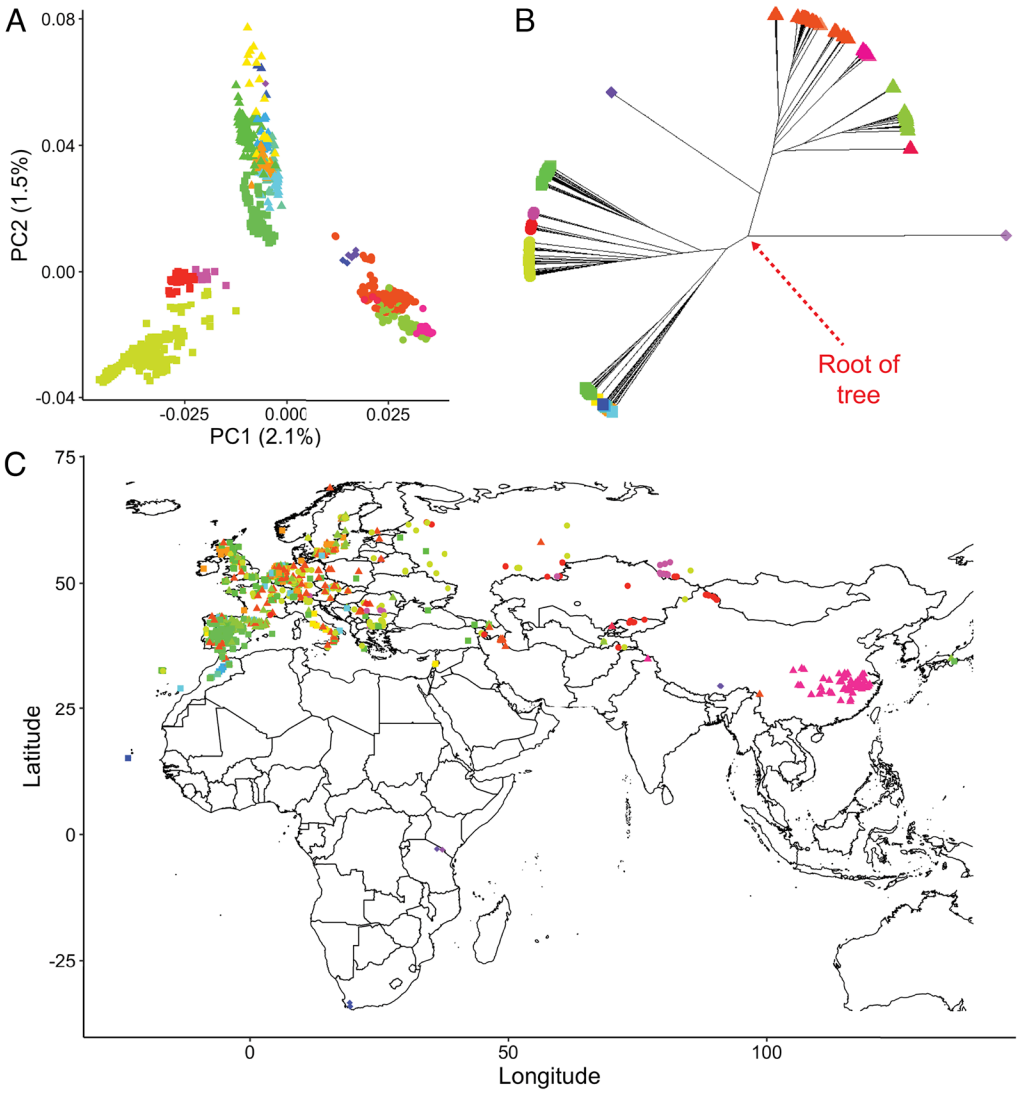

Fig. 1. Phylogenetic analyses of organellar diversity in *A. thaliana* based on chloroplast genome variation. (*A*) The first two principal components (PCs) showing the major components that separate the organelle diversity into three main branches. Colors are based on hierarchical clustering on the basis of a neighbor-joining tree with k = 20. Accessions within each of the three branches are depicted with a triangle, circle, or square, with the exception of accessions close to the root of the neighbor-joining tree. The same color codes and symbols are used in panels *B* and *C*. For additional PCs, *SI Appendix*, Fig S5. (*B*) Neighbor-joining tree displaying the genetic relationship between accessions, based on chloroplast genetic diversity. The red arrow indicates the root of the tree, where the four related species are connected. For clarity, these are left out. (*C*) Accessions with colors and symbols as defined in panels *A* and *B* mapped to the geographic positions where they have been collected. Accessions sampled in North America are not shown as these are found to have originated on the European continent (45).

accessions have diverged into two main clusters (Fig. 1*B*). The deep branch lengths in the neighbor-joining tree are due to substantial genetic variation between different clusters. The neighbor-joining trees of the chloroplast and mitochondrial genomes have substantial overlap (with a cophenetic correlation coefficient of 0.65, *SI Appendix*, Figs. S1 and S2). The accessions that are clustered differently between the chloroplast and mitochondria could be indicative of paternal transmission of organellar genomes (*SI Appendix*, Fig. S2). However, most rearrangements between clusters are due to 1 or 2 variants except for 29 Chinese accessions that are an outgroup only in the mitochondrial neighbor-joining tree. These rearrangements are likely caused by mutation accumulation, rather than by paternal transmission. Therefore, we conclude there is no evidence for paternal transmission.

Much of the observed organellar variation could be attributed to population structure and historical species expansion. To define different subclusters, we split the plasmotypes into 20 different clusters (k = 20) (*SI Appendix*, Fig. S4). Organellar subclusters could arise in genetic isolation without geographic isolation due to the absence of genetic recombination of plasmotypes between accessions due to the largely maternal inheritance. Plotting the accessions onto their geographic locations shows how accessions from West Asia and Europe are largely a mixture of the most common organellar clusters observed within *A. thaliana* (Fig. 1*C*). In other geographic locations, only specific organellar subclusters occur: the Cape Verde islands and Madeira each represent their own subcluster, and also the Yangtze River basin, the Northern and Southern Altai mountains, Uzbekistan, the Iberian Peninsula, Tanzania, South Africa, and several regions within Morocco are home to unique accessions from one subcluster. This shows that in some parts of the native range, organellar subclusters are specific to one geographic location and absent in other regions. The geographic isolation of clusters may be due to neutral mutation accumulation, but it can also result in genetic adaptation.

Cybrid Panel Construction. We constructed a cybrid panel to generate a representative overview of the impact of organellar variation on plant phenotypes beyond the previous cybrid panel we examined (43). We selected 53 accessions from 18 of the 20 chloroplastic subclusters we defined. The two subclusters that are not represented contain accessions from South Africa and Tanzania and are only known from herbarium material (46). Among the 53 accessions is the Staro-2 accession, which represents an outgroup based on the mitochondrial neighbor-joining tree (*SI Appendix*, Figs. S1 and S2). We added seven additional accessions for a total of 60, including three accessions from Africa (i.e., ET2, Tanz-2, and Toufl) which had not been sequenced previously. We combined the plasmotypes of all 60 accessions with four diverged nuclear accessions (i.e., Bur-0, Col-0, Tanz-1, and Cvi-0; *SI Appendix*, Fig. S3). We expected that these will result in many

novel nucleotype–plasmotype interactions. In the process of cybrid production, we observed substantial variation in the percentage of haploids recovered from the total number of germinating seeds with fractions ranging between 20.1% haploids for the Bur-0 nucleotype to 51.8% haploids for the Col-0 nucleotype. This means there is genetic variation for the haploidization due to the loss of maternal chromosomes during postzygotic mitotic division, based on the *GFP-tailswap* mutant system (55). The resulting 240 cybrids were whole-genome-sequenced to verify their genotypes, and 232 cybrids were found to have the expected genotype (*SI Appendix*, Fig. S6). These 232 cybrids were subsequently used to assess the plasmotypic contribution to photosynthetic variation.

Cybrids in Dynamic Environments. Previous work showed that plasmotypes, and the interaction between plasmotypes and nucleotypes, primarily contributed to the overall phenotypic variation when grown under high and dynamic light conditions (43). Therefore, we exposed the cybrids to a range of different conditions including steady-state light intensities, dynamic light conditions, low temperatures, and combinations of these. During these treatments, we phenotyped the photosynthetic response using two different high-throughput chlorophyll fluorescence phenotyping systems (Fig. 2 and *SI Appendix*, Figs. S7 and S8). We phenotyped for F_v/F_m, NPQ, $NPQ_{(t)}$, Φ_{PSII}, Φ_{NPQ}, Φ_{NO}, q_E, $q_{E(t)}$, q_I, $q_{I(t)}$, q_L, and projected leaf area with the Dynamic Environment Phenotyping Instrument (DEPI) and Φ_{PSII} and projected leaf area in the Phenovator system (56, 57). Broad sense heritability (H^2) was used to estimate how much of the phenotypic variation was explained by the total genetic component. The contribution of the nucleotype, plasmotype, and nucleotype–plasmotype interaction to phenotypic variation was calculated as a fraction of the total H^2. The Ely plasmotype was excluded from H^2 analysis, because its large-effect *psbA* mutation may influence the H^2 more than any other plasmotype or nucleotype variant. This would mask observations on the contributions by the separate genomes for the rest of the cybrid panel. To determine whether these phenotyping systems allowed us to assess genotype–environment interactions, we first analyzed overall H^2. NPQ was found to have an average H^2 of 9.2% under steady-state low light, while under fluctuating light this increased to 15% (Fig. 2*B* and *SI Appendix*, Table S1). The H^2 increased further to 34.6% when combining steady-state low light with a low temperature (12 °C). Fluctuating light in combination with a low temperature resulted in H^2 of 27.5%, with outliers during the high light fluctuations of 61.1%. This shows that our phenotyping setup acts as a reliable platform for observing genotype–environment interactions.

The average H^2 was 31.6% across all environmental conditions, for a total of 1,986 phenotypes in the DEPI experiment. The plasmotype and nucleotype–plasmotype interaction components explained 1.3% and 1.9%, respectively (*SI Appendix*, Fig. S7). Regarding NPQ, the fraction of H^2 for the additive plasmotype accounts for on average 1.1%, and the fraction H^2 for the nucleotype–plasmotype interaction accounts for 2.5% (Fig. 2*C*). For some phenotypes, under certain conditions, these components account for substantially more, even though the average H^2 is low. This was most pronounced for the rapid-relaxing component of NPQ, q_E, for which the additive plasmotype effect explains on average 4.7%, and the nucleotype–plasmotype interaction explains 2.5% (Fig. 2*D*). However, up to 36.6% of the total H^2 for q_E is explained by the nucleotype–plasmotype interaction under fluctuating light

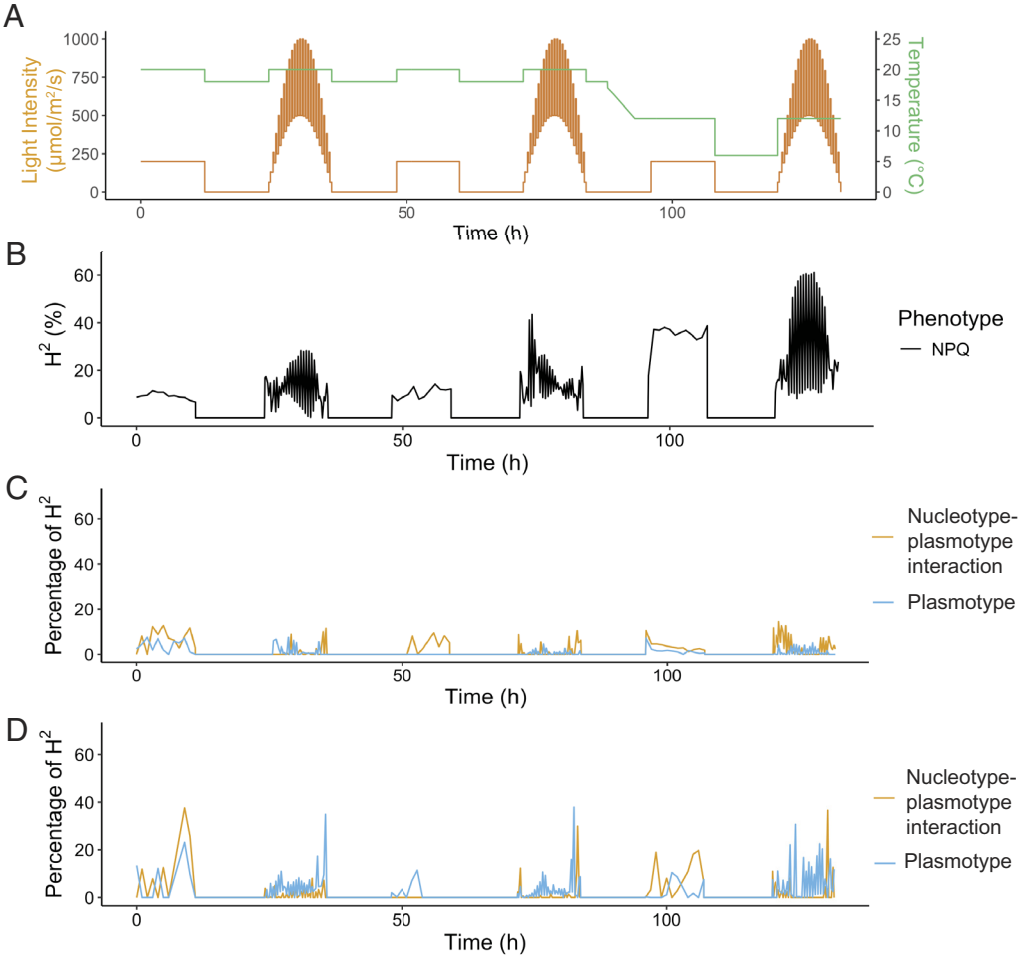

Fig. 2. Overview of overall broad sense heritability and the fractions explained by the plasmotype and nucleotype–plasmotype interaction components of nonphotochemical quenching (NPQ) in response to different environmental conditions in the DEPI system. (*A*) Environmental conditions to which the plants were exposed after being grown under steady-state light conditions of 200 µmol m^{-2} s^{-1}, with t = 0 being the start of the photoperiod 21 d after sowing. (*B*) Broad sense heritability (H^2; shown as percentage) of NPQ in response to the different environmental conditions, as shown in panel *A*. (*C*) Percentage H^2 for NPQ in response to the different environmental conditions for the additive plasmotype effect (blue) and nucleotype–plasmotype interaction effect (yellow). (*D*) The same as panel C, but for the percentage H^2 for q_E in response to the different environmental conditions.

conditions in combination with a low temperature (Fig. 2*D*). For the additive plasmotype effect, this was up to 37.9% of the total H^2 for q_E at the end of the fluctuating light days (Fig. 2*D*). This shows that plasmotype variation, either via additive effects or via an interaction with the nucleotype, on average accounts for relatively little of the H^2 in comparison to the nucleotype. However, under specific conditions, the plasmotype plays a substantial role and therefore can influence overall photosynthetic performance.

The DEPI experiment allowed us to stress the cybrids with high and fluctuating light conditions in combination with cold temperatures at the end of the growing period. However, the impact of dynamic environmental conditions over the whole growing period remained unknown. We grew the entire cybrid panel outside, in semiprotected conditions during the spring of 2020 and 2021 in Wageningen (The Netherlands, 51°59'20.0"N 5°39'43.2"E, Fig. 3*A*). In such conditions, the cybrids were constantly exposed to a dynamic environment. Analysis of the data showed that the resulting 57 photosynthetic and morphological phenotypes had an average H^2 of 16.6% (*SI Appendix*, Fig. S9), of which the plasmotype explained 2.5% and the nucleotype–plasmotype interaction explained 5.6% (Fig. 3*C*). The plasmotype and nucleotype–plasmotype interaction explained 37.7% of the H^2 for NPQ (measured as $NPQ_{(t)}$) after a high- to low-light transition. This is substantially more than observed for NPQ in the DEPI experiment (Fig. 2*C*), which means that for the same phenotype, the contribution of the plasmotype depends heavily on the environment in which plants are grown.

Plasmotypes Causing Phenotypic Effects. Next, we focused on revealing the plasmotypes that caused phenotypic differences. Using PC analyses, we revealed plasmotypes with deviating effect sizes from the 1,986 phenotypes from the DEPI system including 376 phenotypes from the Phenovator system and 57 phenotypes from the semiprotected experiments. The PCs are calculated for the overall plasmotype effects (i.e., regardless of the nucleotype), and the nucleotype–plasmotype interactions (Fig. 4). Based on photosynthetic parameters, the PC show that the relationship between plasmotypes and nucleotypes changes completely between the three different experiments (Fig. 4 *B*, *D*, and *F*). For example, the Bur-0 plasmotype is separated from the others in the plasmotype PC analysis of the semiprotected experiment (Fig. 4*E* and *SI Appendix*, Fig. S10), but is grouping centrally in the PCs of the other two experiments (Fig. 4 *A* and *C*). The Bur-0 plasmotype was previously found to cause significant phenotypic effects (43), and subsequent in-depth analysis further confirmed a persistent photosynthetic effect under fluctuating light conditions (58). In the fluctuating light conditions of the DEPI experiment, the Bur-0 plasmotype effect can indeed be distinguished (*SI Appendix*, Fig. S11), although the PC analysis did not identify it as one of the outliers (Fig. 4*A*). This means that environmental conditions strongly influence the differences between PC analyses, emphasizing the strong plasmotype–nucleotype–environment interaction.

The Bur-0 plasmotype effect showed to be environmentally dependent; however, the PC analyses revealed specific plasmotypes to have stronger overall phenotypic effects. The Staro-2 plasmotype effect stands out most persistently in all three experiments (Fig. 4 *A*, *C*, and *D*). The Staro-2 plasmotype is consistently showing increased Φ_{PSII} in primarily stable low-light conditions (200 µmol m^{-2} s^{-1}), as can be seen in the Phenovator experiment (*SI Appendix*, Fig. S12). Other plasmotypes convey a less consistent phenotypic effect and are limited to specific environmental conditions. The plasmotypes of IP-Sne and IP-Boa are responsible to the high fraction of H^2 in q_E in fluctuating light combined with low temperatures (Fig. 2*D*) when the PC analysis focuses on just these conditions (*SI Appendix*, Fig. S13). The differences in q_E are primarily caused by changes in the basal dissipation parameter Φ_{NO} (59; *SI Appendix*, Fig. S13). IP-Sne reduces Φ_{NO} by 10.5% in comparison to the average and IP-Boa increases Φ_{NO} by 11.3% in comparison to the average (*SI Appendix*, Fig. S14). Both accessions originate from the Iberian Peninsula and they are found in

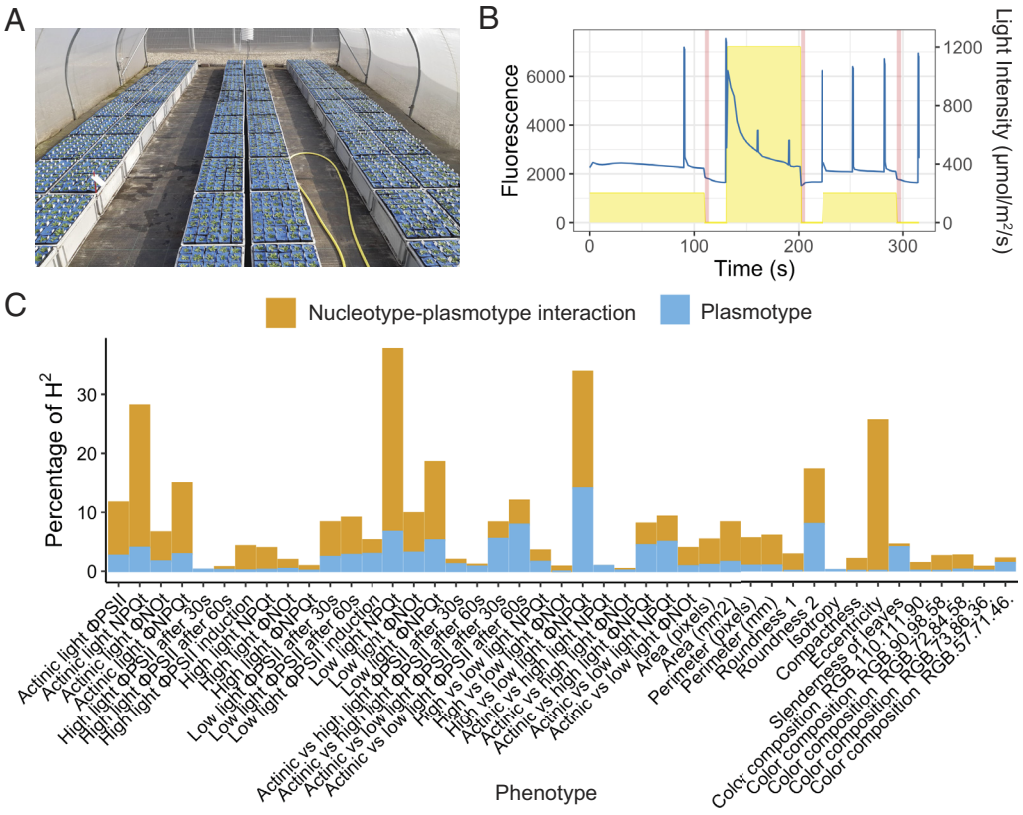

Fig. 3. Overview of the experimental design of the cybrids grown in a gauze tunnel, the phenotyping system, and H^2 of the measured phenotypes. (*A*) Image of the gauze tunnel experiment in spring 2022. (*B*) The 6-min light regime and chlorophyll fluorescence measuring regime used to phenotype the cybrid panel. The blue line shows the fluorescence signal of one plant, in response to the light conditions (in yellow, with the y-axis on the right). The saturating flashes during the light are used to measure F'_m at each timepoint, with an F' measurement just before that. In the dark periods, after the far-red application, F'_0 is measured. The application of far-red light is indicated by the red bars. (*C*) Percentage of H^2 explained by the plasmotype (blue) and nucleotype–plasmotype interaction (orange) for phenotypes measured over two consecutive years. Ten phenotypes with H^2 values lower than 5% are not displayed, as fractions of low H^2 can be inflated.

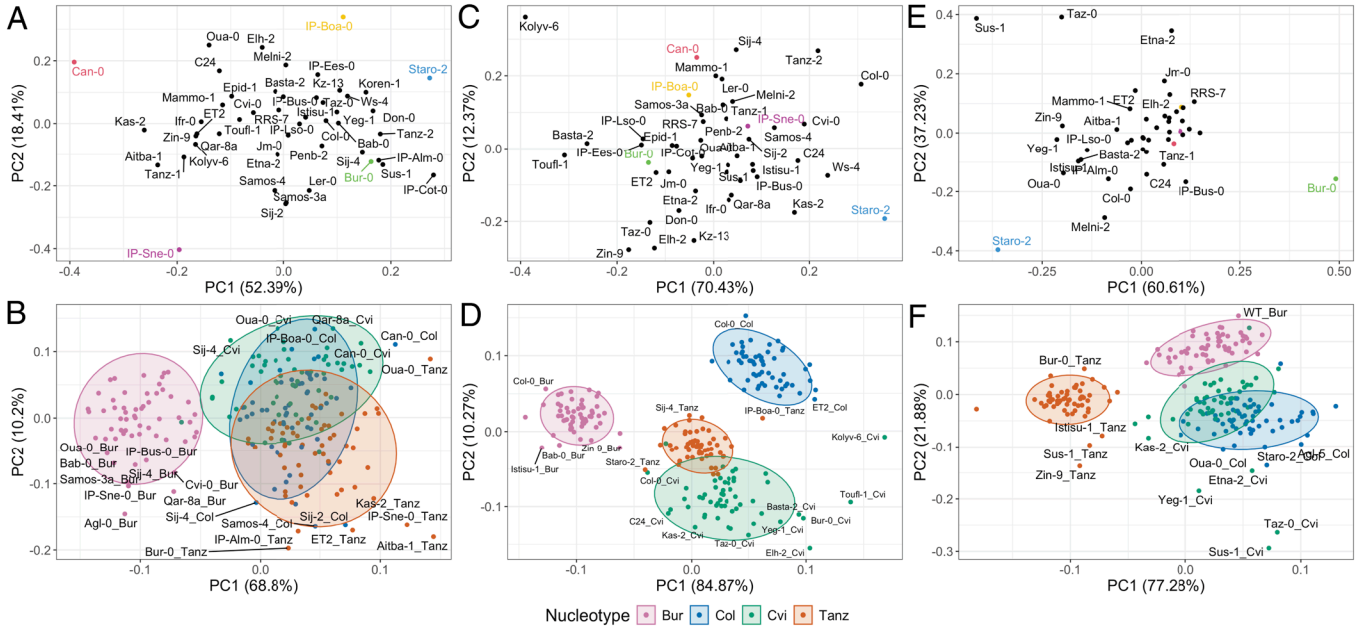

Fig. 4. Principal component analysis of the photosynthetic phenotypes in three different experiments. (*A* and *B*) PC analysis of the photosynthetic phenotypes in the DEPI experiments. (*C* and *D*) PC analysis of the photosynthetic phenotypes Phenovator experiments. (*E* and *F*) PC analysis of the semiprotected experiments. Panels *A*, *C*, and *E* show the additive analysis, and the outlying plasmotypes are colored. Panels *B*, *D*, and *F* show the analysis of the cybrids separately and are color coded per nucleotype. The ovals represent a multivariate t-distribution. Cybrids in panels *B*, *D*, and *F* are annotated as "plasmotype_nucleotype." For all PCs, only the data for photosynthetic parameters are used. In panels *A*, *C*, and *E*, the outlying plasmotypes are colored; these plasmotypes are discussed in the text.

the same subcluster of the neighbor-joining tree, but they respond in opposite manners in terms of Φ_{NO}. This implies that even closely related plasmotypes can cause opposing phenotypes.

The nucleotype–plasmotype interaction explains a higher fraction of H^2 compared to the additive plasmotype effect. The PC analyses show different plasmotypes as deviant when comparing the four nucleotypes (Fig. 4 *B*, *D*, and *F*). For example, the Oua-0 plasmotype is only deviating when combined with the Tanz-1 nucleotype, whereas the Agl-0 plasmotype is only deviating when combined with the Bur-0 nucleotype (Fig. 4*B*). Furthermore, while the PC analysis of additive plasmotype effects suggests that Can-0 has large overall phenotypic effects, its position within each nucleotype distribution varies (Fig. 4 *A* and *B*). Can-0 is obviously deviating when combined with the Col-0 nucleotype, but less so when combined with of the Bur-0 and Cvi-0 nucleotypes. For the Tanz-1 nucleotype, on the other hand, Can-0 is found to group centrally. This emphasizes the role of nucleotype–plasmotype interactions in determining photosynthetic phenotypes.

Plasmotype Association Studies. Overall, there is considerable plasmotypic variation conveying photosynthetic differences, but it is generally associated with small effect sizes. This makes the detection of true-positive plasmotype effects complicated. Moreover, methods to narrow down on the causal genetic variations in the plasmotypes are largely lacking. For both cases, the concept underlying genome-wide association studies (GWAS) can be useful. However, GWAS is developed for mapping in the nucleotype and depends on recombination, which is absent in the plasmotypes. Therefore, we explored an alternative method, which we termed Plasmotype Association Studies (PAS), to avoid confusion with GWAS. In PAS, a variant that is unique to a plasmotype or group of plasmotypes, with a significant phenotypic effect, will be identified. The photosynthetic phenotypes analyzed are quantitative traits, thus the variation for it may be expected to be normally distributed. However, because plasmotypes causing phenotypic variation may be caused by only one variant, it does not necessarily reflect a normal distribution (*SI Appendix*, Fig. S15). To efficiently correct for population structure, a univariate linear mixed model was used in GEMMA (60). The significance threshold is determined based on the total number of unique combinations of variants, which is 313 for the chloroplast genomes of the 60 plasmotypes. With $\alpha = 0.05$, this results in a Bonferroni-corrected significance threshold of $-\log_{10}(p) = 3.8$. To test this approach, we performed PAS for traits at timepoints where the Bur-0 and Can-0 plasmotypes were found to differ phenotypically. This revealed unique genetic variants for the Can-0 and Bur-0 plasmotypes (Fig. 5 *A* and *B* and *SI Appendix*, Fig. S15). Theeuwen et al. (58) reported a single nucleotide polymorphism (SNP) within the *NdhG* gene of the Bur-0 plasmotype to cause the photosynthetic effect. This causal SNP is among the variants with significant associations with the phenotype, thus showing the potential usefulness of PAS.

The phenotypic changes due to the Bur-0 and Can-0 plasmotypes are unique, and therefore a minor allele frequency threshold would remove these effects from the analyses. However, for phenotypic effects caused by a variant that is shared between several plasmotypes, PAS has increased statistical powers as compared to pairwise comparisons between plasmotypes. The minor allele frequency threshold was set at 4% to find associations caused by variants in at least two plasmotypes. Using PAS we found two groups of plasmotypes, each with one shared chloroplast SNP, associated with phenotypic differences. The plasmotypes of IP-Sne and Sij-2 revealed a SNP at base pair position 27,256, a variant that is associated with differences in q_E, NPQ, q_L, $q_{E(t)}$, and $NPQ_{(t)}$ in fluctuating light conditions during low temperatures (Fig. 5*C*). The other group, with the plasmotypes of Kas-1, Kas-2, and Etna-2, share a SNP at position 66,114, associated with differences in NPQ in high light conditions (Fig. 5*D*). Variants at positions 27,256 and 66,114 are upstream mutations in *RpoC1* and *PsbJ*, respectively. Determining whether these are indeed causal for the phenotypic differences observed requires further experimentation, but it demonstrates the potential power of PAS, as these groups of plasmotypes did not show up when plasmotype effects were examined one by one.

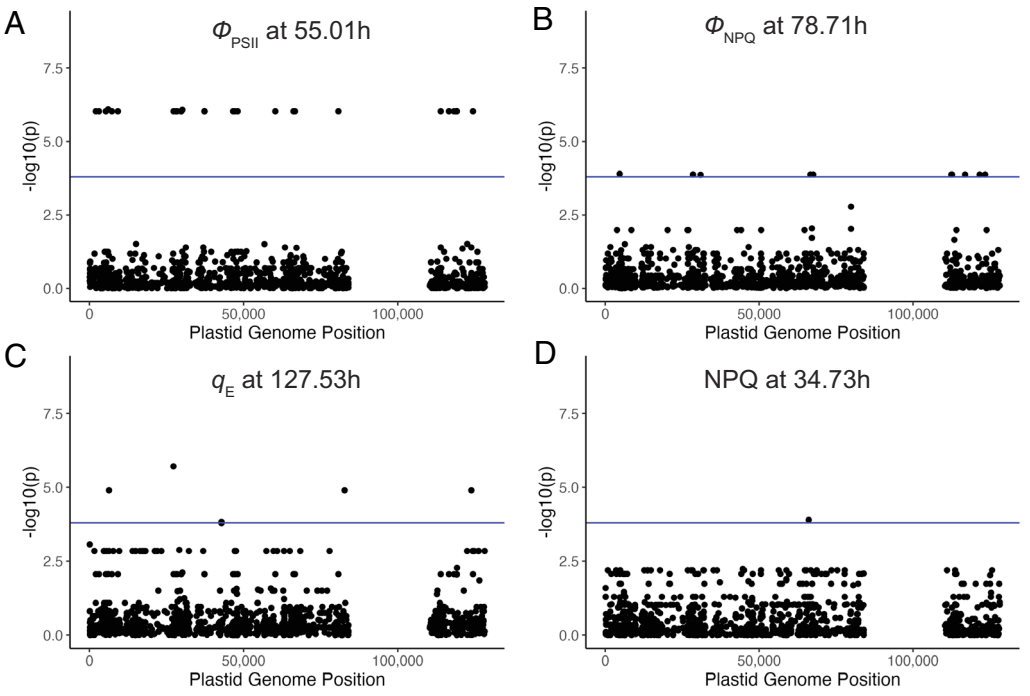

Fig. 5. Plasmotype Association Studies (PAS) performed for specific phenotypes acquired in the DEPI experiment. (*A*) LOD-score plot for Φ_{PSII} at 55.01 h, revealing variants with LOD-scores exceeding the Bonferroni-corrected threshold (blue line), unique to the Bur-0 plasmotype. (*B*) LOD-score plot for Φ_{NPQ} at 78.71 h, revealing a series of significant variants unique to the Can-0 plasmotype. (*C*) LOD-score plot for q_E at 127.53 h, revealing a significant variant at position 27256 of the chloroplast genome that is shared between IP-Sne and Sij-2. (*D*) LOD-score plot for NPQ at 34.73 h, revealing a significant variant at position 66114 of the chloroplast genome that is shared by Kas-1, Kas-2, and Etna-2. The timepoints mentioned here match the light conditions in Fig. 2*A*. The Bonferroni-corrected threshold ($\alpha = 0.05$) is set at $-\log_{10}(p) = 3.8$.

Discussion

Previous studies using *A. thaliana* have shown that there is genetic variation among organellar genomes. However, these studies were based on a limited number of accessions and genetic loci (33, 35, 61, 62). A recent systematic phylogenetic analysis using whole-genome sequencing data confirmed widespread organellar variation (36). We extended this analysis to include more accessions and used the updated version of the mitochondrial reference genome (21). The updated mitochondrial reference genome is based on Col-0 instead of C24 in the earlier reference genome versions. This ensures that chloroplast, mitochondrial, and nuclear diversity can all be compared among accessions. This dataset allowed us to study the characteristics of organellar evolution and diversity at a species level.

We show that the neighbor-joining trees of the mitochondria and chloroplast genomes are highly correlated, and there is no evidence for any large-scale paternal transmission of organellar genomes in *A. thaliana*, even though paternal transmission of chloroplasts is shown to happen in experimental studies (63). Furthermore, we found that the genetic diversity in the chloroplast genomes was eight times that in the mitochondrial genome. This is likely caused by a difference in mutation rate, meaning that the mutation rate difference between the chloroplast and mitochondria is more than twice as high as previously reported (64, 65). The difference between the mutation rates is hypothesized to result from a more efficient repair mechanism in the mitochondria (51). In contrast to the mutation rate being higher in the chloroplast, chloroplast genes appear to be more conserved than the mitochondrial genes. However, this is based on relatively few mutations in the mitochondria due to the low mutation rate. We found 1,495 unique plasmotypes among the 1,531 accessions, and deep branching between subclusters in the neighbor-joining tree. Together, the mutation accumulation and lack of recombination resulting from the absence of paternal transmission have resulted in diverging plasmotypic subclusters. Therefore, we conclude that there is a high degree of genetic variation between *A. thaliana* plasmotypes and that most of the genetic diversity is present among the chloroplast genomes.

Organellar genomes are thought to be under purifying selection. While this does not exclude positive selection, organellar variants have long been considered neutral (23). Organellar variation that is adaptive would result in specific plasmotypes confined to specific environmental conditions. An example of this is the spread of the *PsbA* mutation conferring resistance to the herbicide triazine along British railway lines (30). The main plasmotype clusters of the 1,531 *A. thaliana* accessions are spread through most of West Asia and Europe, suggesting that these plasmotypes were not specifically adapted to their environment. This spread is in line with the postglacial recolonization from Southern European refugia, especially from the Balkans (66). Some smaller subclusters are more geographically bound to specific regions, such as a few subclusters from sub-Saharan Africa. In the Southern Altai mountains in Central Asia, a different subcluster is observed than in the Northern Altai mountains. Similarly, several regions within the Atlas Mountains in North Africa are native to only one subcluster. These observations could indicate there is selection for local adaptation to specific environments, though this can also result from genetic drift.

To investigate the impact that organellar genetic variation can have on phenotypes at a species level, we made cybrids with 60 diverse plasmotypes, which is a considerable expansion of the previous cybrid panel we made that consisted of the reciprocal cybrids of seven accessions (43). That study and another one found that nucleotype–plasmotype interactions determined more variation than plasmotypes alone (43, 67). We therefore now combined the 60 plasmotypes with four diverged nucleotypes, to capture as much nucleotype–plasmotype diversity. Flood et al. (43) found that the plasmotype accounted for 2.9% of the H^2 and the nucleotype–plasmotype interaction accounted for 5.2% (43). Here, we find on average 1.3% of the variation to be explained by the plasmotype and 2.0% by the nucleotype–plasmotype interaction. Since all seven plasmotypes used in the previous cybrid panel (43) are also included here, the lower percentage of explained H^2 is either due to the larger genetic variation in the four nucleotypes or due to the wider range of environmental conditions to which the cybrid panel was exposed. Overall, we find that the plasmotype and nucleotype–plasmotype interaction

accounted on average for 3.3% of the total H^2 of the phenotypes investigated here, while the organellar genomes together make up only 0.4% of the *A. thaliana* genome. Under specific environments, this can go up to 37.9% of the H^2 to be explained by the plasmotype and 36.6% by nucleotype–plasmotype interaction. Also, the PC analyses showed that several plasmotypes and nucleotype–plasmotype interactions contribute to phenotypic differences in different environmental conditions. This suggests that organellar genetic variation could play a role in adaptation to new environments.

In the current study, we primarily show the impact of organellar genetic variation on photosynthesis parameters, and we assume that any differences in photosynthetic phenotypes are most probably associated with chloroplast genes. Nevertheless, also mitochondrial genes are known to play a role in photosynthesis (68). In the cybrid panel, we found that the Staro-2 plasmotype induced cytoplasmic male sterility (CMS) due to a partially duplicated mitochondrial ORF (69). CMS is often associated with reduced mitochondrial respiration (70, 71). A previous study showed that genotypes displaying CMS also suffer from reduced photosynthetic performance (72). Remarkably, the Staro-2 plasmotype showed the most consistent and highest positive effect on Φ_{PSII} of all plasmotypes, suggesting there is a correlation between mitochondrial respiration and photosynthesis. Further analysis of the Staro-2 plasmotype could provide unique insights into how reduced mitochondrial respiration can result in increased photosynthetic performance.

Upon closer examination of specific plasmotypes that caused photosynthetic differences, several other insights into the phenotypic impact of plasmotype variation were revealed. The parameter that showed the largest fraction of H^2 explained by the plasmotype is q_E. As q_E is a photoprotection mechanism induced under fluctuating light conditions (73), plasmotype variation for q_E may play a substantial role in light adaptation. One of the plasmotypes influencing the H^2 for q_E is the Bur-0 plasmotype. It is found to have faster recovery of Φ_{PSII} than the Col-0 plasmotype after a high-light to low-light transition (58). Even though this plasmotype results in a substantial impact on overall photosynthetic performance and biomass accumulation, it is not one of the consistently deviating plasmotypes in the current study. Can-0 is the plasmotype with the biggest impact on q_E. Different from the Bur-0 plasmotype, the Can-0 plasmotype increases q_E and decreases Φ_{PSII}. The Bur-0 effect is caused by allelic variation of the chloroplast *NdhG* gene (58), encoding for a subunit of the NDH-like complex. Can-0 has mutations in *NdhF* and *NdhD* (in addition to a variant in *Rps15*), hinting at the importance of the NDH complex in q_E and Φ_{PSII}. The NDH complex is involved in one of the pathways for cyclic electron transport, an essential mechanism in balancing ATP and NADPH production (74). With 12 of the 85 protein-coding genes in the chloroplast encoding components of the NDH complex, variation in these genes may explain a substantial part of the observed photosynthetic differences.

Revealing the genetic variation underlying phenotyping differences remains difficult due to the small effect sizes and multiple testing corrections needed. Statistical power can be increased by using variants that are shared by all plasmotypes showing a phenotypic effect. For this, we propose to use the PAS approach on the plasmotypes. PAS is similar to GWAS, but as organellar genomes do not recombine like the nuclear genomes, the resulting associations are not subject to linkage disequilibrium decay (75). Consequently, upon PAS it is not possible to identify which of the variants with the same LOD score could be the cause of any observed phenotypic effect. This approach resembles how GWAS has been used to associate genetic variants with phenotypic differences in bacterial genomes (76–78). We performed PAS only on the chloroplast genome due to the dynamic properties of the mitochondrial genome and apparent heterozygosity of some variants, making correct associations difficult (53, 54). Nevertheless, significant associations found in the chloroplast may be caused by genetic variation in the mitochondrial genome, due to both organellar genomes of a plasmotype inheriting together. Minor allele frequency thresholds are used in GWAS to avoid the chance of identifying false-positive associations between alleles and phenotypic variation. Due to the small population size of our cybrid panel and the high degree of genetic diversity, a minor allele frequency threshold of 5% excludes most variants. Nevertheless, several significant associations are identified. These associations will need to be experimentally confirmed to identify the causal variant. For organellar genomes, transformation approaches would be the most promising (79–83). Although, at the moment, most of the methods that allow changing any genetic variant or generating knockouts in all copies of the organellar genomes are still under development. An alternative would be to use a genetic exclusion approach, where plasmotypes with shared variants are phenotyped for the same trait. In the absence of a common phenotypic difference among these variants compared to the rest, that specific variant can be excluded (58). Overall, we show that PAS has the potential to identify candidate alleles associated with phenotypic effects.

Previous research already showed that different phenotypes can arise due to plasmotypic variation (23, 43, 67, 84), but a systematic analysis on how plasmotypic variation contributes to photosynthetic performance was lacking. Here, we used *A. thaliana* as a model species to conclude that plasmotype diversity is widespread and that several plasmotypes harbor variants that can significantly impact photosynthetic performance. The contribution of the plasmotype variation as part of total phenotypic variation is relatively small. However, it can be substantial especially in dynamic environmental conditions. As *A. thaliana* is an inbreeding species, it would be valuable to see whether the plasmotype contribution to phenotypes may differ more for outbreeding species. Currently, the range of crops in which cybrids can be made via haploid induction is quickly expanding (85–88). Given the widespread plasmotypic effects on photosynthetic performance and the advancing capabilities to produce cybrids, the time has now come to tap into organellar diversity for crop improvement.

Materials and Methods

Analysis of *A. thaliana* Organellar Diversity. The recalling of the variant was done using publicly available sequencing data. The data were obtained from ENA for the following datasets; 1,135 worldwide accessions (PRJNA273563) (45), 117 Chinese accessions (PRJNA293798) (48), 75 African accessions (PRJEB19780) (46), 14 Madeiran accessions (PRJEB23751) (49), 1 Tibetan accession (47), 7 global accession (PRJEB29654) (43), 192 Dutch accessions (89) and 4 related species (*A. lyrata*, *A. halleri*, *A. carpatica,* and *Capsella rubella*). The variant calling was done with the same pipeline as described in (43). One accession was removed due to a missing .fastq file. All accessions were mapped to the *A. thaliana* Col-0 reference genome (TAIR10.1, GCA_000001735.2) (21). To remove probable false-positive calls, the organelle genome variants were filtered based on their quality-by-depth score (QD). The chloroplast call set was filtered with a QD of minimally 25, leaving 4,095 variants and for the mitochondrial call set was filtered with a QD of minimally 20, leaving 1,152 variants. Ten accessions with more than 50% of the chloroplast genome variants called heterozygous were removed.

Subsequent analyses were done in R (version 4.0). The four related species were left out of the analysis, unless stated explicitly. Ape (version 5.4-1) (90) and vcfR (version 1.12.0) (91) were used to perform hierarchical clustering, based on complete-linkage clustering. The dendogram was cut using a k = 20 as an arbitrary cutoff, but based on the elbow method. Principal component analyses

where done with SeqArray (version 1.28.1) (92) and SNPRelate (version 1.22.0), as these packages allowed to include multiallelic variants (93). Geographic locations were taken from the respective papers and plotted on the map using rworldmap (version 1.3-6) and rworldxtra (version 1.01). Figures where made using ggplot2 (version 3.3.2) and GGally (version 2.0.0).

Plant Material and Creation of the Cybrid Panel. The wild-type plants were either present in the laboratory, obtained via the Nottingham Arabidopsis Stock Centre, or sent to us via colleagues (*SI Appendix*, Table S3). We used the *GFP-tailswap* line as described in Flood et al. (43), as haploid inducer line. The cybrid panel was made as described in Flood et al. (43). The *GFP-tailswap* line was crossed as paternal line to all 60 accessions, and the F_2 genotypes homozygous for *GFP-tailswap* were selected. This resulted in 60 *GFP-tailswap* genotypes, all having a different plasmotype. Subsequently, these 60 *GFP-tailswap* genotypes were crossed to the four nucleotype donors, with the *GFP-tailswap* as the maternal parent. 14 cybrids in the cybrid panel were overlapping with a previous cybrid panel and were obtained from Flood et al. (43). All cybrids were propagated in the same controlled greenhouse, to exclude possible batch effects.

Genotyping of the Cybrid Panel. DNA extraction was done in 96 deep well plates. Single rosette leaves were harvested from individual plants and placed in the deep well plates, snap-frozen in liquid nitrogen, and ground with a Retsch MM300 TissueLyser. 100 mL extraction buffer (200 mM Tris-HCl, 25 mM EDTA, and 1% SDS) was prepared by adding 40 µL of 20 mg/mL RNase A. 500 µL of extraction buffer including RNase A was added to each well, and incubated at 37 ˚C for 1 h, and inverted every 15 min. To pellet the debris, the plates were centrifuged for 5 min at 3,000×*g*. In a new deep well plate, 130 µL KAc buffer (98.14 g potassium acetate and 3.5 mL Tween were added to 160 mL H_2O, and H_2O was added to reach 200 mL) and 400 µL lysate were added. The plates were sealed and inverted for 2 min and incubated on ice for 10 min. To pellet the debris, the plates were centrifuged for 5 min at 3,000×*g*, and 400 µL of supernatant was transferred to a new plate containing 440 µL Sera-Mag Speedbeads (Cytiva Europe) diluted in PEG buffer. Plates were sealed and placed on a shaking table for 30 min. Next, the plates were placed on a magnet for 5 min and the supernatant was removed. The pellet was washed with 500 µL 80% EtOH three times. The plates were left to evaporate in the fume hood for 1 h and the DNA was resuspended in 50 µL milliQ.

Consecutively, the Hackflex protocol was used to do the library preparation (94). Samples were pooled, and the fraction showing strands of roughly 300 to 500 bp in size were selected and sequenced for on average 8X whole-genome coverage sequencing by Novogene (United Kingdom) Ltd. The reads were trimmed using Cutadapt (95), removing the adaptor sequences and bases from the 5′ and 3′ ends with a Phred quality score below 20. Reads shorter than 75 bp were also removed. The reads were then mapped to the *A. thaliana* Col-0 reference genome, TAIR10.1, using speedseq and the Burrow-Wheeler aligner, BWA-MEM (96, 97). The alignments were sorted and indexed using Samtools (98). Duplicate reads were marked using GATK (99). Variant calling was performed using freebayes, and the resulting variant call format (VCF) file was separated into three VCF files–one for each of the nuclear, mitochondrial, and plastid genomes (100). Distance matrices were calculated for each of the three genomes using PLINK (101), and the ape package in R was used to produce neighbor-joining trees (102). The cybrids with incorrect genotypes were removed from all statistical analyses.

Phenotyping and Data Analysis.

Semiprotected condition experiments. Plant growth took place in an outdoor, gauze-covered tunnel at Unifarm, Wageningen University and Research, the Netherlands (51.9882583, 5.66119897). The footprint of the tunnel measured 8 × 5 m. The tunnel was enclosed by synthetic gauze material that was largely penetrable by rainwater, sunlight and wind so to provide conditions similar to those encountered in the field. Rain gauges placed inside and outside the tunnel confirmed that all rainwater was able to penetrate the gauze. Light irradiance was measured both inside the tunnel and at a metrological station within 500 m from the tunnel (*SI Appendix*, Figs. S16 and S17). A comparison of readings from both locations, during the growing period of spring 2020, indicated that the gauze decreased the light irradiance that penetrated through to the growing area on average by 94.3 µmol m^{-2} s^{-1}. The floor of the tunnel was covered in black landscape material and the tunnel contained a zipper door to prevent the entry of any undesired interferents.

Black plastic pots measuring 7 × 7 × 18 cm (Bestebreurtje B.V., Huissen, the Netherlands) were used for individual plants as they allowed sufficient depth for unlimited root development. Gray plastic trays measuring 40 × 60 × 20 cm were used to hold 40 of the aforementioned pots. The trays were organized in five rows of 11 trays and one row of 13 trays. A small seedling tray was placed in the bottom of each large gray tray to raise the pots above the edge of the gray tray. The plant pots were filled with a mixture of 40 % sand and 60 % peat provided by Lensli Substrates (Katwijk, the Netherlands). The substrate includes YARA PG MIX™ which contains 15 − 10 − 20 + 3 of N, P_2O_5, K_2O, and MgO. The added fertilizer is in powder form and results in complete substrate values of 1.0 and 5.7 for electrical conductivity and pH, respectively. No additional nutrition was applied during the experiment.

In spring of 2020, plastic wrapping was placed over the trays during the night for the first 14 d of growth to protect from cold temperatures. The same was done in spring 2021 for the first five days of growth. In spring 2020, gray rubber covers were placed on each pot after seedlings had established, leaving 0.5 cm space for water to reach the soil. In spring 2021, the soil was covered with blue Friedola Mega Stop mats as described in (103). The pots were evenly watered as needed according to weather conditions and rainfall. Anti-slug/snail pellets were placed in small piles on the ground around the perimeter of the tunnel. Remote sensors from 30 MHz (Amsterdam, the Netherlands) monitored and recorded light irradiance and temperature every minute at the plant level.

Phenotyping was primarily done using the high-throughput phenotyping system PlantScreen™ SC System provided by Photon Systems Instruments spol. s r.o (Drásov, Czech Republic). Using the chlorophyll fluorescence camera and an RGB camera, we obtained a range of photosynthetic and morphological parameters. A custom-made 6-min fluctuating light protocol (Fig. 3*B*) allowed us to calculate Φ_{PSII}, $\Phi_{NO(t)}$, $\Phi_{NPQ(t)}$, and $NPQ_{(t)}$ at different moments during the 6-min protocol. During the light, every 30 s, F' and F'_m were measured, and at the end of each light period, the lights were turned off and after a 6 s far-red application F'_0 is measured. These chlorophyll fluorescence parameters are used to calculate the following phenotypes Eqs. **1**–**5**.

$$\Phi_{PSII} = \frac{F'_m - F'}{F'_m}, \quad [1]$$

$$NPQ_{(t)} = \frac{4.88}{\left(\frac{F'_m}{F'_0}\right) - 1} - 1, \quad [2]$$

$$q_L = \frac{F'_m - F'}{F'_m - F'_0} * \frac{F'_0}{F'}, \quad [3]$$

$$\Phi_{NO(t)} = \frac{1}{NPQ_{(t)} + 1 + q_L * 4.88}, \quad [4]$$

$$\Phi_{NPQ(t)} = 1 - \left(\Phi_{NO(t)} + \Phi_{PSII}\right). \quad [5]$$

A custom R script converted these measurements into 37 parameters capturing the dynamic response to the fluctuating light protocol. A custom mask using the Schedular software was generated to phenotype 20 plants, with a 7 × 7 cm grid size. In spring and autumn 2020, the gray rubber plates were complemented with additional rubber strips, to mask soil-grown algae being registered. In spring 2021, the blue mats ensured the algal growth was not recorded. Automatic masking by the Data Analyser software ensured background noise was removed, and the plant mask was generated. The RGB camera generated an additional 20 parameters quantifying the morphological characteristics and color properties. In spring 2020, the shoots were also harvested to measure the shoot dry weight when the first flower of a plant opened.

During spring 2020, 221 cybrids were ready and could be sown. The cybrids were sown in an unbalanced, incomplete block design to randomize the cybrids among the pots meaning the nucleotypes being randomized among the trays and the plasmotypes being randomized within the trays. The number of replicates ranged between 10 to 12 for the cybrid genotypes and 60 to 80 for the four wild types. Due to the number of pots, sowing was spread over three days, with cybrids with the Bur-0 nucleotype being sown on March 18th 2020, cybrids with Col-0 and Cvi nucleotypes on March 19th, and cybrids with Tanz nucleotypes on March 20th. Approximately 4 pregerminated

seedlings were sown per pot, and all but the healthiest seedling was removed 20 d after sowing. The trays containing nucleotypes Cvi-0, Col-0, Bur-0, and Tanz-1 were phenotyped in the PlantScreen™ system at 35, 38, 38, and 38 d of growth, respectively.

During spring 2021, all 240 cybrids were ready and could be sown. A complete randomized block design was used, in a way that all 240 cybrids were sown once every six trays. Such six trays were arranged in a 2 × 3 orientation, and together formed one big block, with the total experiment having n = 12. As phenotyping in the PlantScreen™ system allowed only 20 plants, we ensured that all cybrids within a block were measured back-to-back to correct over the day and time as best as possible. Again, the sowing was split over three days, cybrids with the Bur-0 nucleotype were sown on March 17th 2021, cybrids with the Col-0 nucleotype on March 18th, and cybrids with the Cvi and Tanz nucleotype on March 19th. After 24 d of growth, all but the healthiest seedling were removed. Phenotyping in the PlantScreen™ system happened on May 3rd and 4th, meaning the plants ranged between 45 and 48 d old.

Growth measurements, red-green-blue (RGB) measurements, and chlorophyll fluorescence parameters from the two seasons were compiled. Plants that did not establish well were removed via outlier detection on the basis of leaf area. This was done by calculating the mean leaf area per genotype and individuals with a leaf area less than the mean–1.5 × SD were removed. Plants with the Ely plasmotype were removed from all subsequent analyses as this plasmotype confers a particularly large reduction in Φ_{PSII}, thus obscuring smaller contributions of the plasmotype to the heritability of photosynthetic traits (43). Linear mixed models were constructed for each response variable using the lme4 package in R (version 4.1.0) (104). The contribution of the nucleotype, plasmotype, and the nucleotype–plasmotype interaction to each phenotypic variance was quantified by the variance components for these model terms in the random model Eq. **6**, using a restricted maximum likelihood approach.

$$\underline{Y} = \underline{Nucleotype} + \underline{Plasmotype} + \left(\underline{Nucleotype * Plasmotype}\right) + \left(\underline{Block * Experiment}\right) + \underline{error}. \quad [6]$$

The broad sense heritability (H^2) was calculated as the sum of the genetic variance components (Nucleotype, Plasmotype, and Nucleotype × Plasmotype) relative to the total phenotypic variance (105). The contribution of each genetic component to the H^2 was summarized as a percentage of the total H^2. Phenotypes with a H^2 lower than 5% were removed from all subsequent analyses, as variation in these phenotypes cannot be accurately predicted based on genotype (43).

To investigate the main effect of plasmotype and its interaction with nucleotype, a mixed model was fitted with fixed main effects for nucleotype and plasmotype and their fixed interaction and random terms for blocks and error. This resulted in the model given in Eq. **7**.

$$\underline{Y} = Nucleotype + Plasmotype + \left(Nucleotype * Plasmotype\right) + \left(\underline{Block * Experiment}\right) + \underline{\varepsilon}. \quad [7]$$

DEPI experiment. Phenotyping screens were carried out using the Dynamic Environmental Photosynthetic Imaging (DEPI) system (56), with modifications as described in (106). The DEPI system allows all of the plants to be measured simultaneously, allowing repeated measurements under fluctuating light conditions. The plants were grown in climate-controlled conditions for 18 d and moved into the DEPI system to acclimate for three days. The DEPI experiment started (t = 0 h) at midnight on day 21. Two different light regimes were used, one day the light intensity was stable at 200 μmol m^{-2} s^{-1} and the other day it was fluctuating light, with sinusoidal increases and decreases over the day. These fluctuations were characterized as 18 min low light, 2 min darkness, 8 min high light, and 2 min darkness (56).

Growth measurements and images of chlorophyll fluorescence parameters (F_v/F_m, NPQ, $NPQ_{(t)}$, Φ_{PSII}, Φ_{NPQ}, Φ_{NO}, q_E, $q_{E(t)}$, q_I, $q_{I(t)}$, q_L) from five experiments were calculated using Eqs. **8**–**15**. Images of relative fluorescence yield values were recorded as in Cruz et al. (56). The minimal fluorescence was measured with weak light pulses in extensively dark-adapted leaves (F_0), leaves exposed to continuous illumination (F'), following a light–dark transition and 6 s of far-red illumination to oxidize Q_A (F'_0) and at the end of the 2 min darkness (F''_0). The maximal fluorescence was measured during brief (approximately 1 s) saturation pulses to fully reduce Q_A in the fully dark-adapted plants (F_m), light-acclimated leaves (F'_m), and 2 min following transition from light to dark in light-acclimated leaves (F''_m).

$$\frac{F_v}{F_m} = \frac{F_m - F_0}{F_m}, \quad [8]$$

$$NPQ = \frac{F_m}{F'_m} - 1, \quad [9]$$

$$\Phi_{NO} = \frac{F'}{F_m}, \quad [10]$$

$$\Phi_{NPQ} = 1 - \Phi_{PSII} - \Phi_{NO}, \quad [11]$$

$$q_E = \frac{F_m}{F'_m} - \frac{F_m}{F''_m}, \quad [12]$$

$$q_I = \frac{F_m - F''_m}{F'_m}, \quad [13]$$

$$q_{I(t)} = \frac{4.88}{\left(\frac{F''_m}{F''_0}\right) - 1} - 1, \quad [14]$$

$$q_{E(t)} = NPQ - q_{I(t)}. \quad [15]$$

In total, we ran five DEPI experiments, each with 224 plants. Using an incomplete block design, the cybrids were distributed so that they were sown in four out of five experiments (n = 4). Every experiment had eight replicates of the four wild-type parents. To remove plants that performed poorly in comparison to the other individuals of the same genotype, outlier detection was done on the dark-adapted yield of PSII photochemistry (F_v/F_m). The mean F_v/F_m for all plants was calculated and plants with an F_v/F_m less than the mean–1.5 × SD were removed. As in the semicontrolled experiments, plants with the Ely plasmotype were removed from all subsequent analyses.

Statistical modeling was carried out as with the data from the semicontrolled experiments, using Eq. **16** for the estimation of variance components and the calculation of H^2 and Eq. **17** for the calculation of the main effects. Phenotypes with a H^2 lower than 5% were removed for all subsequent analyses.

$$\underline{Y} = \underline{Nucleotype} + \underline{Plasmotype} + \left(\underline{Nucleotype * Plasmotype}\right) + \underline{Block} + \underline{Experiment} + \underline{\varepsilon}, \quad [16]$$

$$\underline{Y} = Nucleotype + Plasmotype + \left(Nucleotype * Plasmotype\right) + \underline{Block} + \underline{Experiment} + \underline{\varepsilon}. \quad [17]$$

Phenovator experiment. The cybrid panel was phenotyped using the Phenovator, a high-throughput phenotyping platform in which 1440 *A. thaliana* plants can be grown (57). Plants have been measured seven times a day for photosystem II efficiency and growth, except during the three days with fluctuating light. On fluctuating-light days, the plants have only been measured once on the middle of the day. The plants were grown with a 12-hour photoperiod (*SI Appendix*, Fig. S8). For the first 17 d, light intensity was set at 200 μmol m^{-2} s^{-1}. On day 18 fluctuating light conditions were started for three days. During these fluctuating light regimes, every twenty minutes the light switched between 500 μmol m^{-2} s^{-1} and 100 μmol m^{-2} s^{-1}, except during measurements. Only one measurement was performed during these days, during one hour. At the end of day 21, the temperature was decreased to from 20 °C to 12 °C during the day, and from 18 °C to 6 °C in the night. Throughout the experiment, the relative humidity was kept at 70%. Plants were grown on a 4 × 4 × 4 cm rockwool substrate provided by Grodan B.V. (Roermond, the Netherlands), and irrigated weekly with a nutrient solution (*SI Appendix*, Table S3). The cybrid panel was screened using this system four times. One run included only the cybrids with the Col-0 nucleotype, and a randomized complete block design was used (n = 24). The other three runs included the full cybrid panel, which was sown in a randomized complete block design (n = 6 in each run).

Leaf area and Φ_{PSII} measurements were extracted from image files produced by the Phenovator using custom software (57). During each of the seven chlorophyll fluorescence measurements F' and F'_m were measured to calculate Φ_{PSII} using Eq. **1**. Measurements from the four separate experiments were subsequently compiled. Outlier detection was used to remove plants that did not establish well, by removing individuals with a leaf area (15 d after sowing) that was 1.5 × SD lower than the mean per genotype. Furthermore, complete genotypes that showed poor germination, resulting in stunted growth, were removed. This is because these cybrids were not always removed by the outlier detection as the whole genotype was stunted. This included all cybrids with the Cvi nucleotype in one experiment and all of the Cvi^{Sha} cybrids. Plants with the Ely plasmotype were removed from all subsequent analyses, as in the semi-controlled and DEPI experiments. Statistical modeling was carried out as with the data from the semicontrolled and DEPI experiments, using Eq. **18** for the estimation of variance components and the calculation of H^2 and Eq. **19** for the calculation of the main effects. Phenotypes with a H^2 lower than 5% were removed from all subsequent analyses. The model includes image position, to correct for intensity of saturating light as received by each of the twelve positions. The experiment represents the four individual runs.

$$\underline{Y} = \underline{Nucleotype} + \underline{Plasmotype} + \underline{(Nucleotype * Plasmotype)} + \underline{Block} + \underline{Imageposition} + \underline{Experiment} + \underline{\varepsilon}, \quad [18]$$

$$\underline{Y} = Nucleotype + Plasmotype + \underline{(Nucleotype * Plasmotype)} + \underline{Block} + \underline{Imageposition} + \underline{Experiment} + \underline{\varepsilon}. \quad [19]$$

For all the three experiments described above, a separate principal component (PC) analysis was carried out with R, using the main effects for each plasmotype. Further PCs were performed for each of the four nucleotypes using the main effects for the nucleotype–plasmotype interactions, which are equal to the Best Linear Unbiased Estimate. These PC analyses were done for the photosynthetic parameters only.

Plasmotype Association Studies. A genome-wide association study (GWAS) method was used to test the association between genetic variants in the plastid genome and the phenotypes measured in the DEPI experiments. Due to the lack of recombination within the plastid genome, the interpretation of the GWAS results deviates from a standard GWAS performed on the nuclear genome. Unless genetic variants are shared between different plasmotypes, all of the SNPs within a plasmotype will be equally associated with a phenotype. Thus, to avoid confusion with a nuclear GWAS, this test has been named a Plasmotype Association Study (PAS). To carry out the PAS, the main effect for each plasmotype from the DEPI phenotyping experiments was used alongside the VCF file for the plastid genome on the basis of 40X sequencing coverage. Binary PED files were produced from the plastid VCF file using PLINK (101). The main effect for each of the 1986 phenotypes was added manually to the .fam file. GEMMA was used to produce a centered relationship matrix for the plastid genomes of the 60 progenitor lines and to run a univariate association study (60).

Data, Materials, and Software Availability. Supplementary data have been deposited in Zenodo (107). R scripts and Linux commands for data analyses are available via GitLab (108). Seeds of the cybrids are available upon request.

ACKNOWLEDGMENTS. We thank Andreas Hemp for providing seeds of accession ET2, John Doonan for providing seeds of accession Penbont-1, and Carlos Alonso-Blanco for seeds of Agl-5. We also thank Ben Auxier, Roel van Bezouw, René Boesten, Ramon Botet, Pádraic Flood, Willem Kruijer, and Antoine Languilaume for insightful discussions. We acknowledge José van de Belt and Jordy Litjens for their help in the construction of the cybrids and Corrie Hanhart, Gema Flores Andaluz, René Boesten, Sanne Put, Louise Logie, and Federico Fornaguera for their help in sowing experiments. Ben Auxier is acknowledged for feedback on the manuscript. This work was, in part, supported by the Netherlands Organization for Scientific Research (NWO) through ALWGS.2016.012 (T.P.J.M.T.). Bejo Zaden BV and Rijk Zwaan BV are gratefully acknowledged for their support to this work. Portions of this work appeared as part of the PhD dissertation of T.P.J.M.T.

Author affiliations: [a]Laboratory of Genetics, Wageningen University & Research, Wageningen 6708 PB, The Netherlands; [b]Bioinformatics Group, Wageningen University & Research, Wageningen 6708 PB, The Netherlands; [c]Michigan State University Department of Energy Plant Research Lab, Michigan State University, East Lansing, MI 48824; [d]Biometris, Wageningen University & Research, Wageningen 6708 PB, The Netherlands; and [e]Laboratory of Biophysics, Wageningen University & Research, Wageningen 6708 WE, The Netherlands

Author contributions: T.P.J.M.T., F.A.v.E., D.M.K., E.W., J.H., M.K., and M.G.M.A. designed research; T.P.J.M.T., R.Y.W., D.D., A.W.L., J.L., K.J., J.B., D.T., D.H., C.H., F.F.M.B., and M.K. performed research; T.P.J.M.T., R.Y.W., D.D., A.W.L., J.L., K.J., J.B., D.T., and D.H. analyzed data; and T.P.J.M.T., M.K., and M.G.M.A. wrote the paper.

Reviewers: W.B., Salk Institute for Biological Studies; and J.K., University of Cambridge.

The authors declare no competing interest.

[4]Present address: Crop Genetics, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, United Kingdom.

[5]Present address: Max Planck Institute for Plant Breeding Research, Cologne 50829, Germany.

[6]Present address: Centre for Crop Systems Analysis, Wageningen University & Research, Wageningen 6700 AK, The Netherlands.

[7]Present address: Centre for Genetic Resources, Wageningen University & Research, Wageningen 6708 PB, The Netherlands.

1. S. P. Long, X. G. Zhu, S. L. Naidu, D. R. Ort, Can improvement in photosynthesis increase crop yields? *Plant Cell Environ.* **29**, 315–330 (2006).
2. X. G. Zhu, S. P. Long, D. R. Ort, Improving photosynthetic efficiency for greater yield. *Ann. Rev. Plant Biol.* **61**, 235–261 (2010).
3. D. R. Ort *et al.*, Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 8529–8536 (2015).
4. A. Hitchcock *et al.*, Redesigning the photosynthetic light reactions to enhance photosynthesis–the PhotoRedesign consortium. *Plant J.* **109**, 23–34 (2022).
5. P. J. Flood, J. Harbinson, M. G. M. Aarts, Natural genetic variation in plant photosynthesis. *Trends in Plant Sci.* **16**, 327–335 (2011).
6. T. P. J. M. Theeuwen, L. L. Logie, J. Harbinson, M. G. M. Aarts, Genetics as a key to improving crop photosynthesis. *J. Exp. Botany* **73**, 3122–3137 (2022).
7. S. M. Driever, T. Lawson, P. J. Andralojc, C. A. Raines, M. A. J. Parry, Natural variation in photosynthetic capacity, growth, and yield in 64 field-grown wheat genotypes. *J. Exp. Botany* **65**, 4959–4973 (2014).
8. Q. Wang *et al.*, Genetic architecture of natural variation in rice nonphotochemical quenching capacity revealed by genome-wide association study. *Front. Plant Sci.* **8**, 1773 (2017).
9. R. Van Rooijen *et al.*, Natural variation of YELLOW SEEDLING1 affects photosynthetic acclimation of Arabidopsis thaliana. *Nat. Commun.* **8**, 1–9 (2017).
10. C. G. Oakley *et al.*, Genetic basis of photosynthetic responses to cold in two locally adapted populations of Arabidopsis thaliana. *J. Exp. Botany* **69**, 699–709 (2018).
11. A. E. Prinzenberg, M. Víquez-Zamora, J. Harbinson, P. Lindhout, S. van Heusden, Chlorophyll fluorescence imaging reveals genetic variation and loci for a photosynthetic trait in diploid potato. *Physiologia Plantarum* **164**, 163–175 (2018).
12. A. E. Prinzenberg, L. Campos-Dominguez, W. Kruijer, J. Harbinson, M. G. M. Aarts, Natural variation of photosynthetic efficiency in Arabidopsis thaliana accessions under low temperature conditions. *Plant Cell and Environ.* **43**, 2000–2013 (2020).
13. T. Rungrat *et al.*, A genome-wide association study of non-photochemical quenching in response to local seasonal climates in Arabidopsis thaliana. *Plant Direct* **3**, e00138 (2019).
14. M. Faralli, J. Matthews, T. Lawson, Exploiting natural variation and genetic manipulation of stomatal conductance for crop improvement. *Curr. Opin. Plant Biol.* **49**, 1–7 (2019).
15. S. Adachi *et al.*, Genetic architecture of leaf photosynthesis in rice revealed by different types of reciprocal mapping populations. *J. Exp. Botany* **70**, 5131–5144 (2019).
16. K. Taniyoshi, Y. Tanaka, T. Shiraiwa, Genetic variation in the photosynthetic induction response in rice (Oryza sativa L.). *Plant Production Sci.* **23**, 513–521 (2020). 10.1080/1343943X.2020.1777878.
17. L. G. Acevedo-Siaca *et al.*, Variation in photosynthetic induction between rice accessions and its potential for improving productivity. *New Phytol.* **227**, 1097–1108 (2020).
18. L. G. Acevedo-Siaca, R. Coe, W. P. Quick, S. P. Long, Variation between rice accessions in photosynthetic induction in flag leaves and underlying mechanisms. *J. Exp. Botany* **72**, 1282–1294 (2021).
19. J. N. Timmis, M. A. Ayliff, C. Y. Huang, W. Martin, Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).
20. P. Lamesch *et al.*, The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **40**, D1202–D1210 (2012).
21. D. B. Sloan, Z. Wu, J. Sharbrough, Correction of persistent errors in arabidopsis reference mitochondrial genomes. *Plant Cell* **30**, 525–527 (2018).
22. D. O. Daley, J. Whelan, Why genes persist in organelle genomes. *Genome Biology* **6**, 110 (2005).
23. D. G. Bock, R. L. Andrew, L. H. Rieseberg, On the adaptive value of cytoplasmic genomes in plants. *Mol. Ecol.* **23**, 4899–4911 (2014).
24. D. C. Wallace, A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: A dawn for evolutionary medicine. *Ann. Rev. Genet.* **39**, 359 (2005).
25. S. Greiner, "Plastome mutants of higher plants" in *Genomics of Chloroplasts and Mitochondria*, (Springer Science & Business Media, 2012), pp. 237–266.

26. M. W. Nachman, "Deleterious mutations in animal mitochondrial DNA" in *Mutation and Evolution, Contemporary Issues in Genetics and Evolution*, R. C. Woodruff, J. N. Thompson, Eds. (Springer, Netherlands, 1998), pp. 61–69.
27. D. M. Rand, L. M. Kann, "Mutation and selection at silent and replacement sites in the evolution of animal mitochondrial DNA" in *Mutation and Evolution, Contemporary Issues in Genetics and Evolution*, R. C. Woodruff, J. N. Thompson, Eds. (Springer, Netherlands, 1998), pp. 393–407.
28. K. H. Wolfe, W. H. Li, P. M. Sharp, Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9054–9058 (1987).
29. M. E. El-Lithy *et al.*, Altered photosynthetic performance of a natural Arabidopsis accession is associated with atrazine resistance. *J. Exp. Botany* **56**, 1625–1634 (2005).
30. P. J. Flood *et al.*, Whole-genome hitchhiking on an organelle mutation. *Curr. Biol.* **26**, 1306–1311 (2016).
31. M. V. Kapralov, D. A. Filatov, Widespread positive selection in the photosynthetic Rubisco enzyme. *BMC Evol. Biol.* **17**, 1–10 (2007).
32. P. H. Leinonen, D. L. Remington, O. Savolainen, Local adaptation, phenotypic differentiation, and hybrid fitness in diverged natural populations of arabidopsis lyrata. *Evolution* **65**, 90–107 (2011).
33. M. Moison *et al.*, Cytoplasmic phylogeny and evidence of cyto-nuclear co-adaptation in Arabidopsis thaliana. *Plant J.* **63**, 728–738 (2010).
34. Z. Yin *et al.*, Expression quantitative trait loci analysis of two genes encoding Rubisco activase in soybean. *Plant Physiol.* **152**, 1625–1637 (2010).
35. A. Tyagi *et al.*, Genetic diversity and population structure of Arabidopsis thaliana along an altitudinal gradient. *AoB Plants* **10**, plv145 (2015).
36. C.-W. Hsu, C.-Y. Lo, C.-R. Lee, On the postglacial spread of human commensal Arabidopsis thaliana: Journey to the East. *New Phytol.* **222**, 1447–1457 (2019).
37. L. A. Corey, D. F. Matzinger, C. C. Cockerham, Maternal and reciprocal effects on seedling characters in Arabidopsis thaliana (L.) Heynh. *Genetics* **82**, 677–683 (1976).
38. R. C. Meyer, O. Torjek, M. Becher, T. Altmann, Heterosis of biomass production in Arabidopsis. Establishment during early development. *Plant Physiol.* **134**, 1813–1823 (2004).
39. R. Fujimoto, J. M. Taylor, S. Shirasawa, W. J. Peacock, E. S. Dennis, Heterosis of Arabidopsis hybrids between C24 and Col is associated with increased photosynthesis capacity. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7109–7114 (2012).
40. S. Atwell *et al.*, Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **465**, 627–31 (2010).
41. V. S. Gordon, J. E. Staub, Comparative analysis of chilling response in cucumber through plastidic and nuclear genetic effects component analysis. *J. Am. Soc. Horticultural Sci.* **136**, 256–264 (2011).
42. M. Miclaus *et al.*, Maize cytolines unmask key nuclear genes that are under the control of retrograde signaling pathways in plants. *Genome Biol. Evol.* **8**, 3256–3270 (2016).
43. P. J. Flood *et al.*, Reciprocal cybrids reveal how organellar genomes affect plant phenotypes. *Nature Plants* **6**, 13–21 (2020).
44. M. Ravi, S. W. L. Chan, Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**, 615–618 (2010).
45. C. Alonso-Blanco *et al.*, 1,135 Genomes Reveal the Global Pattern of Polymorphism in Arabidopsis thaliana. *Cell* **166**, 481–491 (2016).
46. A. Durvasula *et al.*, African genomes illuminate the early history and transition to selfing in Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5213–5218 (2017). 10.1073/PNAS.1616736114.
47. L. Zeng *et al.*, Discovery of a high-altitude ecotype and ancient lineage of Arabidopsis thaliana from Tibet. *Sci. Bull.* **62**, 1628–1630 (2017).
48. Y. P. Zou *et al.*, Adaptation of Arabidopsis thaliana to the Yangtze River basin. *Genome Biol.* **18**, 1–11 (2017).
49. A. Fulgione, M. Koornneef, F. Roux, J. Hermisson, A. M. Hancock, Madeiran Arabidopsis thaliana reveals ancient long-range colonization and clarifies demography in eurasia. *Mol. Biol. Evol.* **35**, 564–574 (2018).
50. J.-H. Xu *et al.*, Dynamics of chloroplast genomes in green plants. *Genomics* **106**, 221–231 (2015).
51. J. M. Gualberto, K. J. Newton, Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Ann. Rev. Plant Biol.* **68**, 225–252 (2017).
52. Y. Zou, W. Zhu, D. B. Sloan, Z. Wu, Long-read sequencing characterizes mitochondrial and plastid genome variants in Arabidopsis msh1 mutants. *Plant J.* **112**, 738–755 (2022).
53. V. Shedge, M. Arrieta-Montiel, A. C. Christensen, S. A. Mackenzie, Plant mitochondrial recombination surveillance requires unusual RecA and MutS homologs. *Plant Cell* **19**, 1251–1264 (2007).
54. M. P. Arrieta-Montiel, V. Shedge, J. Davila, A. C. Christensen, S. A. Mackenzie, Diversity of the Arabidopsis mitochondrial genome occurs via nuclear-controlled recombination activity. *Genetics* **183**, 1261–1268 (2009).
55. M. Ravi, S. W. L. Chan, Haploid plants produced by centromere-mediated genome elimination. *Nature* **464**, 615–618 (2010).
56. J. A. Cruz *et al.*, Dynamic environmental photosynthetic imaging reveals emergent phenotypes. *Cell Systems* **2**, 365–377 (2016).
57. P. J. Flood *et al.*, Phenomics for photosynthesis, growth and reflectance in Arabidopsis thaliana reveals circadian and long-term fluctuations in heritability. *Plant Methods* **12**, 14 (2016).
58. T. P. J. M. Theeuwen *et al.*, The NDH complex reveals a trade-off preventing maximizing photosynthesis in Arabidopsis thaliana. bioXriv [Preprint] (2022). https://doi.org/10.1101/2022.11.13.516254 (Accessed 13 November 2022).
59. D. M. Kramer, G. Johnson, O. Kiirats, G. E. Edwards, New fluorescence parameters for the determination of Q_A redox state and excitation energy fluxes. *Photosynthesis Res.* **79**, 209–218 (2004).
60. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).
61. J. Provan, J. J. Campanella, Patterns of cytoplasmic variation in Arabidopsis thaliana (Brassicaceae) revealed by polymorphic chloroplast microsatellites. *Syst. Bot.* **28**, 578–583 (2003).
62. F. X. Picó, B. Méndez-Vigo, J. M. Martínez-Zapater, C. Alonso-Blanco, Natural genetic variation of arabidopsis thaliana is geographically structured in the iberian peninsula. *Genetics* **180**, 1009–1021 (2008).
63. A. K. Azhagiri, P. Maliga, Exceptional paternal inheritance of plastids in Arabidopsis suggests that low-frequency leakage of plastids via pollen may be universal in plants. *Plant J.* **52**, 817–823 (2007).
64. K. H. Wolfe, W. H. Li, P. M. Sharp, Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9054–9058 (1987).
65. G. Drouin, H. Daoud, J. Xia, Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Mol. Phylogenet. Evol.* **49**, 827–831 (2008).
66. C.-R. Lee *et al.*, On the post-glacial spread of human commensal Arabidopsis thaliana. *Nat. Commun.* **8**, 14458 (2017).
67. F. Roux *et al.*, Cytonuclear interactions affect adaptive traits of the annual plant Arabidopsis thaliana in the field. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3687–3692 (2016).
68. Y. Fan *et al.*, The crucial roles of mitochondria in supporting C4 photosynthesis. *New Phytol.* **233**, 1083–1096 (2021).
69. T. P. J. M. Theeuwen, "Exploring the unexplored: Unravelling the cyto-nuclear interactions in Arabidopsis thaliana to improve photosynthesis," PhD thesis, Wageningen University & Research, WU (2023).
70. L. Chen, Y.-G. Liu, Male sterility and fertility restoration in crops. *Annu. Rev. Plant Biol.* **65**, 579–606 (2014).
71. R. Horn, K. J. Gupta, N. Colombo, Mitochondrion role in molecular basis of cytoplasmic male sterility. *Mitochondrion* **19**, 198–205 (2014).
72. M. Sabar, R. De Paepe, Y. de Kouchkovsky, Complex I impairment, respiratory compensations, and photosynthetic decrease in nuclear and mitochondrial male sterile mutants of *Nicotiana sylvestris*. *Plant Physiol.* **124**, 1239–1250 (2000).
73. E. H. Murchie, A. V. Ruban, Dynamic non-photochemical quenching in plants: From molecular mechanism to productivity. *Plant J.* **101**, 885–896 (2020).
74. D. D. Strand, L. D'Andrea, R. Bock, The plastid NAD(P)H dehydrogenase-like complex: Structure, function and evolutionary dynamics. *Biochem. J.* **476**, 2743–2756 (2019).
75. A. Korte, A. Farlow, The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **19**, 1–9 (2013).
76. P. E. Chen, B. J. Shapiro, The advent of genome-wide association studies for bacteria. *Curr. Opin. Microbiol.* **25**, 17–24 (2015).
77. M. M. Maury *et al.*, Uncovering Listeria monocytogenes hypervirulence by harnessing its biodiversity. *Nat. Genet.* **48**, 308–313 (2016).
78. F. Coll *et al.*, PowerBacGWAS: A computational pipeline to perform power calculations for bacterial genome-wide association studies. *Commun. Biol.* **5**, 1–12 (2022).
79. S. Ruf *et al.*, High-efficiency generation of fertile transplastomic Arabidopsis plants. *Nat. Plants* **5**, 282–289 (2019).
80. S. Y. Kwak *et al.*, Chloroplast-selective gene delivery and expression in planta using chitosan-complexed single-walled carbon nanotube carriers. *Nat. Nanotechnol.* **14**, 447–455 (2019).
81. A. Jakubiec *et al.*, Replicating minichromosomes as a new tool for plastid genome engineering. *Nat. Plants* **7**, 932–941 (2021).
82. B. C. Kang *et al.*, Chloroplast and mitochondrial DNA editing in plants. *Nat. Plants* **7**, 899–905 (2021).
83. I. Nakazato *et al.*, Targeted base editing in the plastid genome of Arabidopsis thaliana. *Nat. Plants* **7**, 906–913 (2021).
84. F. Budar, F. Roux, The role of organelle genomes in plant adaptation. *Plant Signaling Behav.* **6**, 635–639 (2011).
85. J. Lv *et al.*, Generation of paternal haploids in wheat by genome editing of the centromeric histone CENH3. *Nat. Biotechnol.* **38**, 1397–1401 (2020).
86. N. Wang, J. I. Gent, R. K. Dawe, Haploid induction by a maize cenh3 null mutant. *Sci. Adv.* **7**, eabe2299 (2021).
87. E. Bortiri *et al.*, Cyto-swapping in maize by haploid induction with a cenh3 mutant. *Nat. Plants* **10**, 567–571 (2024).
88. F. Han *et al.*, One-step creation of CMS lines using a BoCENH3-based haploid induction system in Brassica crop. *Nat. Plants* **10**, 581–586 (2024).
89. R. Y. Wijfjes, "Computational analysis of copy number variation in plant genomes," PhD thesis, Wageningen University & Research, WU (2021).
90. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
91. B. J. Knaus, N. J. Grünwald, vcfr: A package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).
92. X. Zheng *et al.*, SeqArray–a storage-efficient high-performance data format for WGS variant calls. *Bioinformatics* **33**, 2251–2257 (2017).
93. X. Zheng *et al.*, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).
94. D. Gaio *et al.*, Hackflex: Low-cost, high-throughput, Illumina Nextera Flex library construction. *Microbial. Genomics* **8**, 000744 (2022).
95. M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10–12 (2011).
96. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [Preprint] (2013). https://doi.org/10.48550/arxiv.1303.3997 (Accessed 7 November 2022).
97. C. Chiang *et al.*, SpeedSeq: Ultra-fast personal genome analysis and interpretation. *Nat. Methods* **12**, 966–968 (2015).
98. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
99. G. A. Van der Auwera *et al.*, From FastQ data to high confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **11**, 11.10.1–11.10.33 (2013).
100. E. Garrison, G. Marth, Haplotype-based variant detection from short-read sequencing. arXiv [Preprint] (2012). http://arxiv.org/abs/1207.3907 (Accessed 7 November 2022).
101. C. C. Chang *et al.*, Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7 (2015).
102. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
103. A. Junker *et al.*, Optimizing experimental procedures for quantitative evaluation of crop plant performance in high throughput phenotyping systems. *Front. Plant Sci.* **5**, 770 (2015).
104. D. Bates, M. Mächler, B. Bolker, S. Walker, Fitting linear mixed-effects models using lme4. arXiv [Preprint] (2014). http://arxiv.org/abs/1406.5823 (Accessed 7 November 2022).
105. D. S. Falconer, T. F. C. Mackay, *Introduction to Quantitative Genetics* (Longmans Green, 1996).
106. S. Tietz, C. C. Hall, J. A. Cruz, D. M. Kramer, NPQt: A chlorophyll fluorescence parameter for rapid estimation and imaging of non-photochemical quenching of excitons in photosystem-II-associated antenna complexes. *Plant Cell Environ.* **40**, 1243–1255 (2017).
107. T. P. J. M. Theeuwen, Data from "Species wide inventory of Arabidopsis thaliana organellar variation reveals ample phenotypic variation for photosynthetic performance". Zenodo. https://doi.org/10.5281/zenodo.11259864. Deposited 23 May 2024.
108. T. P. J. M. Theeuwen, Novel Cybrids. GitLab. https://git.wur.nl/tom.theeuwen/novel-cybrids. Deposited 12 October 2022.