Genetic Basis of Metabolic Trait Variation in the Giant Duckweed *Spirodela polyrhiza* (Lemnaceae)

Dissertation

Zur Erlangung des Grades

Doktor der Naturwissenschaften

Am Fachbereich Biologie

Der Johannes-Gutenberg-Universität Mainz

Martin Höfer

geb. am 10.03.1995 in Saalfeld (Germany)

Mainz, November 2024

Dekan: Prof. Dr. Eberhard Thines

1. Berichterstatter: [redacted]

2. Berichterstatter: [redacted]

Tag der mündlichen Prüfung: 27.02.2025

Summary

Pleiotropy describes a phenomenon in which two or more traits are controlled by a single genetic factor. Since several traits can be negatively associated with each other, resulting in trade-offs, studying pleiotropic gene functions is of great importance for understanding the evolution of plant metabolic traits and biotechnological optimization of plants. Using the giant duckweed *Spirodela polyrhiza* as a model organism for investigating pleiotropic gene functions in plants, this thesis aims to identify gene candidates controlling herbicide resistance, metabolite contents and sexual reproduction in *S. polyrhiza*. For this we quantified the intra-specific variation in free metabolite contents, growth and diquat resistance in *S. polyrhiza*. Candidate genes associated with these traits were identified through genome-wide association studies (GWAS). The first article focuses on quantifying the phenotypic variation in diquat resistance, revealing 8.5-fold differences between the most resistant and most susceptible genotypes. Genotypes showed variation in superoxide dismutase activity, ratios of oxidized to reduced antioxidant contents and uptake kinetics of diquat. On a genetic level, genes involved in biotic stress responses (*SpLOX2.1* and *SpGBPL2*), apoptosis control (*SpDLH*) and starch metabolism (*SpSBE3*) were associated with diquat resistance. The second article focuses on identifying candidate genes controlling variation in free metabolite contents and growth in *S. polyrhiza*. Overall, secondary metabolites showed stronger intra-specific variation than primary metabolites and growth parameters. Often primary metabolites showed strong correlations with plant biomass. A GWAS conducted on free metabolite levels identified candidate genes controlling photosynthesis (*SpLHCB5*), protein-degradation (*SpUBP7*), cell-cycle (*SpPUB4*) and secondary metabolite synthesis (*SpMYBC1*) to be associated with biomass-correlated traits. The third study investigates the migration history and genetic diversity of *S. polyrhiza*. Analysis of population structure revealed four genetic populations for this duckweed species, which were largely defined by their geographic locations as American, SE-Asian, Indian and European populations. A genome-wide scan identified 69 genes, whose homologs were mainly linked to sexual reproduction processes to be under selection. Together, this thesis studies gene candidates controlling herbicide resistance, free metabolite contents and sexual reproduction in *S. polyrhiza*, which can be used for biotechnological optimization of this plant. The results further suggest that pleiotropic genes involved in stress responses confer non-targeted site herbicide resistance.

Zusammenfassung

Pleiotropie beschreibt die genetische Kontrolle einer oder mehrerer Eigenschaften durch einen einzigen genetischen Faktor. Hierdurch können negative Assoziation von Eigenschaften, sogenannte Trade-offs entstehen. Daher ist die Erforschung von pleiotropen Effekten von zentraler Bedeutung für evolutionäre Studien zum Thema pflanzlicher Metabolismus und für die biotechnologische Optimierung von Pflanzen. Die vorliegende Arbeit befasst sich mit der Beschreibung pleiotroper Gene im Modellorganismus *Spirodela polyrhiza*, wobei Genfunktionen im Zusammenhang mit Herbizidresistenz, sexueller Reproduktion und Metabolitkonzentrationen identifiziert werden sollen. Hierfür wurden intraspezifische Variationen von Metabolitkonzentration, Wachstumrate und Diquatresistenz in Genotypen von *S. polyrhiza* bestimmt. Mit diesen Eigenschaften assoziierte Gene wurden mittels Genomweiter Assoziationsstudien (GWAS) identifiziert. Der erste Artikel dokumentiert Variationen von Diquatresistenz in *S. polyrhiza*, wobei der resistenteste Genotyp eine 8.5-fach höhere Resistenz gegenüber des niedrigsten Resistenzniveaus aufwies. Weiterhin unterschieden sich die Genotypen hinsichtlich antioxidativer Eigenschaften und ihrer Diquataufnahmekapazität. Auf genetischer Ebene konnten Gene im Zusammenhang mit biotischen Stressantworten (*SpLOX2.1* and *SpGBPL2*), Apoptose (*SpDLH*) und Stärkemetabolismus (*SpSBE3*) mit Diquatresistenz assoziiert werden. Der zweite Artikel befasst sich mit der Identifizierung von Genen, welche Wachstum und Metabolitkonzentrationen in *S. polyrhiza* kontrollieren. Insgesamt zeigten Sekundärmetabolite eine geringere intraspezifische Varianz als Primärmetabolite und Wachstumsparameter. Primärmetabolite waren oft stark mit pflanzlicher Biomasse korreliert. Mittels GWAS wurden Gene, welche Prozesse wie Photosynthese (*SpLHCB5*), Proteinabbau (*SpUBP7*), Zell-Zyklus (*SpPUB4*) oder Sekundärmetabolitsynthese (*SpMYBC1*) regulieren, mit Biomasse-korrelierten Eigenschaften assoziiert. Die dritte Studie ergründet die historische Migration und genetische Vielfalt von *S. polyrhiza*. Mittels demographischer Analyse wurden vier genetische Populationen, welche nach ihrem Verbreitungsgebiet als Amerika, SO-Asien, Indien und Europa – Populationen benannt wurden, identifiziert. Ein genomweiter Scan ergab, dass 69 Gene, welche überwiegend mit sexueller Reproduktion assoziiert waren, unter Selektion stehen. Zusammenfassend identifizierte diese Arbeit zahlreiche Gene mit Bezug zu Herbizidresistenz, Metabolitkonzentrationen und sexueller Reproduktionen, welche zur biotechnologischen Optimierung von *S. polyrhiza* genutzt werden können. Aufbauend darauf scheinen Gene, welche an Stressantworten beteiligt sind, non-targeted site Resistenz gegen Diquat zu vermitteln.

List of Publications

Thesis publications

Article I:

Höfer, M., Schäfer, M., Wang, Y., Wink, S. & Xu, S. (2024). Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in *Spirodela polyrhiza*. Plants, 13(6), 845, https://doi.org/10.3390/plants13060845

Article II:

Höfer, M., Schäfer, M., Wink, S., Wang, Y. & Xu, S. (2024). Genome-Wide Association Study of Metabolic Traits in the Duckweed *Spirodela polyrhiza*. bioRxiv:2024.2007.2026.605148. https://doi.org/10.1101/2024.07.26.605148

Article III:

Wang, Y., Duchen, P., Chávez, A., Sree, K.S., Appenroth, K.J., Zhao, H., Höfer, M., Huber, M. & Xu, S. (2024). Population genomics and epigenomics of *Spirodela polyrhiza* provide insights into the evolution of facultative asexuality. Communications Biology, 7(1), 581, https://doi.org/10.1038/s42003-024-06266-7

Additional publications

Huber, M., Gablenz, S. & Höfer, M. (2021). Transgenerational non-genetic inheritance has fitness costs and benefits under recurring stress in the clonal duckweed *Spirodela polyrhiza*. Proceedings of the Royal Society B, 288(1955), 20211269, https://doi.org/10.1098/rspb.2021.1269

List of Figures

List of Tables

List of Abbreviation

ABC – ATP-binding cassette

AGL6 – AGAMOUS - LIKE 6

AGL62 – AGAMOUS - LIKE 62

AGO1 – ARGONAUT 1

AOS – ALLENE OXIDE SYNTHASE

AP3 – APETALA 3

ATP – Adenosine-trisphosphate

BB – BIG BROTHER

BLINK – Bayesian-information and linkage-disequilibrium iteratively nested keyway

cDNA – Complementary desoxyribonucleic acid

CYP450 – CYTOCHROME P450

DNA – Desoxyribonucleic acid

DTH8 – DAYS TO HEADING 8

EDDHA – Ethylenediamine-*N*,*N*'-bis(2-hydroxyphenylacetic acid)

EDS1 – ENHANCED DISEASE SUSCEPTIBILITY 1

FarmCPU – Fixed and random model circulating probability unification

FLK – FLOWERING LOCUS K

Fm – Maximum fluorescence

Fm` – Maximum fluorescence (light adapted state)

Fv – Variable fluorescence

Fv` – Variable fluorescence (light adapted state)

GLM – Generalized linear model

GSH – Glutathione

GWAS – Genome-wide association study

HPLC-MS – High performance liquid chromatography – mass spectrometry

IPA1 – IDEAL-PLANT ARCHITECTURE 1

K_a – Substitutions on non-synonymous sites

K_s – Substitutions on synonymous sites

LC-MS – Liquid chromatography mass spectrometry

MADS – MCM1 AGAMOUS DEFICIENS SRF

MCM1 – MINI CHROMOSOME MAINTENANCE 1

MIKC^{C} – MADS - intervening, keratin-like and C-like domain

miR – microRNA

MLM – Mixed linear model

MLMM – Multi locus mixed model

NLR – NUCLEOTIDE-BINDING LEUCINE-RICH REPEAT

NPR1 – NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1

NTSR – Non-targeted-site resistance

PAL – PHENYLALANIENE AMMONIA LYASE

PCA – Principal component analysis

pH – Pondus hydrogenii

PSI – Photosystem I

qPCR – Quantitative polymerase chain reaction

QTN – Quantitative trait nucleotide

RGR – Relative growth rate

RNA – Ribonucleic acid

ROS – Reactive oxygen species

SA – Salicylic acid

SE – South-East

SNP – Single nucleotide polymorphism

SO – Süd-Ost

SOD – SUPEROXIDE DISMUTASE

SOC1 – SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1

SpDLH – DIENELACTONE HYDROLASE

SpETHE1 – MITOCHONDRIAL SULFUR DIOXYGENASE 1

SpGBPL2 – GUANYLATE-BINDING PROTEIN 2

SpHPL – HYDROPEROXIDE LYASE

SpLHCB5 – LIGHT-HARVESTING-CHLOROPHYLL-A/B-BINDING PROTEIN 5

SpLOX2.1 – LIPOXYGENASE2.1

SpMYBC1 – MYELOBLASTOSIS C 1

SpPUB4 – U-BOX CONTAINING PROTEIN 4

SpSBE3 – STARCH-BRANCHING-ENZYME 3

SpSCAMP3 – SECRETORY CARRIER-ASSOCIATED MEMBRANE PROTEIN 3

SpUBP7 – UBIQUITIN-CARBOXYTERMINAL HYDROLASE 7

SpYPEL – YIPPEE-LIKE PROTEIN

SRF – SERUM RESPONSE FACTOR

TOC1 – TIME OF CAP EXPRESSION 1

TSR – Target-site resistance

UV – Ultraviolet

π – Pairwise nucleotide diversity

2,4-D – 2,4-Dichlorophenoxyacetic acid

Table of Contents

1. Introduction

Studying the genetic background of life history traits is challenging since the pathways controlling morphogenesis and reproduction in organisms are often very complex. Also, many pathways display strong interaction patters, meaning that manipulation of one component leads to physiological changes in many other organismal traits (Mauro and Ghalambor 2020). Hormonal pathways are famous examples of the strong interconnectivity of developmental traits. For instance, injection of the alpha-melanocyte-stimulating hormone in the blue morphs of the cichlid species *Astatotilapia burtoni* simultaneously increases pigmentation and aggressiveness, showing that behavior and habitus can be jointly regulated through changes of a single component (Dijkstra et al. 2017). On a genetic level, the joint connection of two or more traits is called pleiotropy (Mauro and Ghalambor 2020). Pleiotropy mostly occurs through one gene controlling several traits. A famous example of pleiotropy is *DAYS TO HEADING* on chromosome *8* (*DTH8*) in *Arabidopsis thaliana*, which decreases flowering rate and increases plant height through regulation of photoperiod sensitivity (Wang et al. 2019b). Alternatively, pleiotropy can also occur through strong linkage between genes. As an example of this, the comb color of a chicken was strongly correlated with egg production - likely through a strong linkage between two genes controlling for ovulation and pigmentation independently (Dong et al. 2019).

Pleiotropy can be of great biotechnological interest, since negative correlations between traits caused by antagonistic pleiotropy, are a major reason why certain traits cannot be optimized simultaneously (Mauro and Ghalambor 2020). Such negative correlations between several traits are also referred to as trade-offs. On the phenotype level, trade-offs determine resource allocation between reproduction and survival (Stearns 1989), or between growth and specialized metabolites (Huot et al. 2014).

A detailed understanding of the genetic basis underlying trade-offs is therefore crucial for improving nutritional value and crop yield in plants. Due to their adaption to various environments plants show overall great amounts of metabolic variation (Kartseva et al. 2024; Schulz et al. 2015; Chen et al. 2016). Since studying trade-offs mainly relies on correlations of phenotypic parameters (Stearns 1989), high amounts of phenotypic variation are favorable for elucidating their genetic background. Because plants play an important role in global food production, research on plant metabolic variation has been shown to be a valuable resource for understanding pleiotropic gene functions and trade-offs (Cu et al. 2021; Angelovici et al. 2013; Alvarez et al. 2010; Živanović et al. 2020; Fang and Luo 2019).

However, in many cases, such metabolic variations are continuous traits and are therefore reflected by complex genetic architectures (Cu et al. 2021; Zhang et al. 2019). Consequently, the exact contribution of certain gene candidates to a trait is often difficult to determine in many cases, compromising marker-based breeding strategies and the genetic engineering of crops (Fang and Luo 2019). Since environmental factors influence the expression of trade-offs (Stearns 1989), another important question is how different environments have specifically selected for pleiotropic gene functions during evolution.

Based on these challenges this thesis will focus on the elucidation of the genetic basis of metabolic pleiotropy using the duckweed *Spirodela polyrhiza* (Lemnaceae) as model organism. In the first section of the introduction (1.1), I will focus which on how phenotypic traits are affected by variations in plant metabolism. Next, it will be shown which evolutionary principles shaped plant metabolism (1.2). The subsequent section (1.3) will then focus on genome-wide association studies (GWAS) as a tool to determine genetic markers associated with quantitative traits. In the last section (1.4), *S. polyrhiza* will be introduced as a model organism for studying pleiotropic gene function in plants.

1.1 Mechanisms of Metabolic Adaptions in Plants

As sessile organisms, plants must deal with a broad array of environmental changes. As a result of this, plant metabolic traits strongly vary in adaptation to environmental gradients. For instance, parameters like leaf thickness or contents of chlorophyll and carotin increase with elevation, whereas photosynthesis rate, stomatal density and leaf size were shown to decrease (Halbritter et al. 2018). In the case of the duckweed *Lemna minor*, the contents of phenolic compounds and C:N ratios increased with latitude and mean average temperature (Chou et al. 2019). Other metabolic changes are lifestyle dependent. For instance free floating plants were shown to have higher nitrogen and free amino acid contents than submerged plants (Chou et al. 2019). As a consequence of alterations of amino acid composition through age caused by increased rates of proteolysis, many metabolic traits change during plant development (Tilsner et al. 2005). The adaptation to this vast array of stress factors is facilitated through a broad multifunctionality of plant metabolites. Therefore, the first section will focus on how the functional diversity of free amino acids and derived compounds facilitates plant adaptation to different stress factors (1.1.1). The next section will then focus on genetic factors controlling metabolic pleiotropy using herbicide resistance as an example (1.1.2).

1.1.1 Adaptive Functions of Free Metabolites to Natural Stresses

On the molecular level many adaption processes to environmental conditions are facilitated through changes in free metabolite contents (Hui et al. 2023). The adaptive function of free metabolite levels is reflected by their high genotype-specific differences (Chen et al. 2014; Chen et al. 2016; Fang and Luo 2019). As a consequence of their high importance in stress responses, variations of specialized metabolites were shown to be higher than those for primary metabolites (Chen et al. 2016). In plants, amino acids and derived compounds constitute a major class of free metabolites controlling metabolic functions (Schulz et al. 2015; Schäfer et al. 2016; Hildebrandt et al. 2015; Yang et al. 2020). Since 62% of total ammonium and 32% of total fixed carbon are used for amino acid biosynthesis (Smith et al. 1961), amino acids are major sinks of carbon and nitrogen in plants, explaining their central role in growth dynamics. As such free amino acids fulfill a broad range of systemic functions such as pH regulation, growth regulation, energy supply and defense against biotic and abiotic stressors (Hildebrandt et al. 2015).

Due to their central function in energy metabolism free amino acids are often released during plant developmental processes such as leaf development and seed germination (Tilsner et al. 2005). Since the contents of free amino acids can reach high levels, they substantially contribute to the regulation of turgor within the plant cell (Fukutoku and Yamada 1981; Živanović et al. 2020). This function is especially relevant in conditions when plants suffer strong osmotic stress. During drought, plants were shown to increase intra-cellular contents of L-proline to compensate for water deficit (Živanović et al. 2020). The involvement of amino acids in herbivore defense is mainly explained by their function as precursors for specialized metabolites. Most importantly, L-phenylalanine serves as a precursor for phenolic acids, flavonoids, and stress hormones like salicylic acid (SA).

Among these derived metabolites, flavonoids were shown to have high functional diversity. Through their hydroxy groups they function as chelators of toxic heavy metal ions (Brown et al. 1998) and antioxidants (Muhlemann et al. 2018; Kumar and Pandey 2013). Through their conjugated π-electron system flavonoids are further capable of absorbing UV-light, protecting the plant from photodamage (Tattini et al. 2004). Additionally, certain flavonoids like anthocyanins show a cold-induced biosynthesis and were shown to promote freezing tolerance in plants (Schulz et al. 2015). By controlling ROS-levels, flavonoids interfere with various developmental processes such as pollen tube growth (Muhlemann et al. 2018; Ringli et al. 2008).

1.1.2 Genetic Mechanisms of Non-Targeted-Site Resistance

As a consequence of the functional diversity of free metabolites, several metabolic traits can be influenced by the simultaneous manipulation of a single gene. Among plants, resistance to herbicides constitutes a famous example of this metabolic pleiotropy. Mutations in resistance-conferring genes can be caused by amplification of target gene copies in the genome (Gaines et al. 2010), promotor rearrangements that increase gene expression (Mernke et al. 2011), epigenetic mechanisms (Pan et al. 2019), non-synonymous point mutations or deletions within coding sequences of genes (Yu et al. 2015; Perotti et al. 2019; Maeda et al. 2019). Depending on the complexity of the underlying mechanism, resistance cases can be grouped into two different categories:

1. Target site resistance describes mutations within the molecular target of the herbicide. Per definition, these mutations can either lead to an overexpression of the molecular target (Yanniccari et al. 2017) or to a structural change in the protein encoded by the target gene through non-synonymous mutation in the coding sequence (Yu et al. 2015). Since target site resistance only influences the function of a single gene, it is defined as a monogenic trait, conferring a discrete phenotype (Jugulam and Shyam 2019).
2. The counterpart of target-site resistance, non-targeted site resistance (NTSR), describes systemic changes in plant metabolism leading to increased herbicide resistance (Jugulam and Shyam 2019). The non-target-site mechanism can therefore be understood as a metabolic compensation of herbicide-mediated stress through mechanisms like reduced uptake or translocation of herbicide (Dinelli et al. 2006), herbicide efflux (Pan et al. 2021; Prosecka et al. 2009), sequestration (Visnovitz et al. 2008; Ge et al. 2010; Jóri et al. 2007), chemical inactivation of herbicide (Pan et al. 2019; Chun et al. 2002), quenching of reactive oxygen species produced by herbicide (Shaaltiel and Gressel 1986; Lewinsohn and Gressel 1984; Jansen et al. 1989) or changes in signaling/apoptosis (Moretti et al. 2018). Since it requires complex interplay of various pathways that act synergistically together to compensate herbicide-mediated stress, NTSR is often polygenic (Jugulam and Shyam 2019; Keith et al. 2017).

Looking at the gene functions underlying NTSR, we see that a high number of NTSR-associated genes are pleiotropic. A prominent example of this is the CYP450 (cytochrome P450) gene family, which is represented by around 16000 genes within the plant kingdom (Nelson 2018). CYP450 genes encode monooxygenases, catalyzing oxidation reactions and are key players in

the metabolism of defense metabolites like terpenes (Geisler et al. 2013), phytohormones like jasmonic acid (Koo et al. 2011) and flavonoids (Ahmad et al. 2019; Steele et al. 1999). As a consequence of their central function in specialized metabolism many CYP450 enzymes are stress-induced (Steele et al. 1999; Pandian et al. 2020). Pleiotropic effects conferred by CYP450 genes are mainly related to the lack of strong substrate specificity for some of these enzymes. Besides their natural substrates, CYP71A10 metabolizes phenylurea herbicides (Siminszky et al. 2003), whereas CYP71A11 catalyzes the breakdown of chlorotoluron herbicides (Gion et al. 2014). Similar to CYP450 genes, ABC (ATP-binding cassette) transporters, many of which confer resistance to herbicides, show great functional diversity. Through catalyzing transport of alkaloids (Shitan et al. 2003) or the excretion of heavy metals and secondary metabolites (Kim et al. 2007; Khare et al. 2017), ABC transporters show strong functional involvement in drought stress (Kim et al. 2010), heavy metal stress (Kim et al. 2007) or pathogen defense (Khare et al. 2017). Due to their transport function, ABC transporters can allocate resources to herbicide-damaged tissues or catalyze herbicide efflux of herbicides like glyphosate (Tani et al. 2015; Tani et al. 2016; Pan et al. 2021). Since ABC transporters and CYP450 enzymes are crucial for secondary metabolism, resistance mechanisms associated with these genes are likely explained by a structural similarity of their natural target molecules with the herbicide (Chun et al. 2002; Pan et al. 2019; Gion et al. 2014).

Besides chemical similarities between herbicides and secondary metabolites, mechanistic similarities between herbicide-generated stress and environmental stress can explain the pleiotropic effects of NTSR conferring genes. A common stress mechanism of drought, salt and heavy metal exposure is the induction of reactive oxygen species (ROS) in plants (Xiong et al. 2002; Schützendübel and Polle 2002). Since many herbicides act via the production of ROS, plant cell death caused by herbicide exposure is likely occurring via ROS-induced apoptosis similar to many other environmental stress factors (Chen and Dickman 2004; Foyer and Noctor 2005). Genotype-specific differences in apoptosis signaling or quenching of ROS were therefore central for explaining differences in herbicide resistance levels in plants (Van Horn et al. 2018; Melicher et al. 2022; Shaaltiel and Gressel 1986).

Since herbicide resistance is mechanistically associated with stress responses in many cases, high fitness costs can be linked with resistance, because resistance-conferring mutations often compromise the catalytic efficiency of the corresponding gene product (Yu et al. 2015). However, for two reasons, many NTSR mechanisms were not associated with resistance costs (Comont et al. 2019). First, in NTSR compensatory mutations, which recover fitness costs and

resistance-conferring mutations are thought to evolve simultaneously (Comont et al. 2019). Second, in contrast to TSR mechanisms, which are mostly constitutive traits (Yu et al. 2015; Yanniccari et al. 2017), some NTSR mechanisms are thought to be inducible traits (Keith et al. 2017; Lewinsohn and Gressel 1984; Moretti et al. 2018), which is a more resource-efficient way to cope with metabolic costs as we will see in section 1.2.

1.2 Evolutionary Origin of Plant metabolic Diversity

Depending on the strength and duration of environmental stimuli plants may express defense constitutively or make it inducible. Inducible defenses provide more general but unspecific protection against stressors compared to constitutive defenses (Murren et al. 2015).

Phenotypic costs for the expression of traits were shown to be strongly environmentally dependent (Murren et al. 2015), whereas the evolution of plastic responses such as inducible defenses requires heterogenous environmental setups. In agreement with this concept, exposure to complex stresses, like simultaneous applications of several herbicides to plants were shown to favor generalist NTSR evolution, whereas applications of a single herbicide favored the evolution of TSR mechanisms (Comont et al. 2020). Since NTSR to herbicides is generally defined by a fast or constitutive upregulation of the plants general, unspecific stress response it has likely evolved from defense strategies against complex stress factors (Keith et al. 2017; Dyer 2018).

Because, as a consequence of their immobility, plants have to cope with fast environmental fluctuations, making traits plastic might be seen as a common strategy to save costs in the absence of natural stressors (Campbell and Kessler 2013). Consequently, a remarkably large portion of plant life-history traits are very plastic compared to other organisms (Sun et al. 2023; Campbell and Kessler 2013). Approximately 54% of variations in plant size are thought to be explained through phenotypic plasticity (Chevin and Lande 2011).

Since fitness costs can be associated with pleiotropic gene functions, the expression of many defense-related genes is inducible or tissue-specific (Upadhyay et al. 2020; Wada et al. 2014; Wang et al. 2020). This results in characteristic temporal and spatial expression patterns of defense traits (Brütting et al. 2017). The following section (1.2.1) deals with strategies of plants to maintain spatial resource allocation patterns of metabolites. Next, I will introduce how environmental factors influence the evolution of inducible plant-life history traits, using the mating system as an example (1.2.2).

1.2.1 Resource Allocation Theories in Plant Defense

On a systemic level, plants keep the balance between growth and defense through different production rates of defensive metabolites and their allocation to tissues as part of their stress response, known as growth-defense trade-off (1.2.1.1). Consequently, resistance to stress factors was often associated with reduced allocation of resources for reproduction (Comont et al. 2019). Also, plants save resistance-associated costs by efficient allocation of defense resources to metabolically more valuable tissues, which is known as optimal defense theory.

1.2.1.1 Growth-Defense Trade-offs

Fitness costs originating from stress responses can occur through different ways. Since plant growth can be interpreted as the conversion of primary metabolites into cells (Hilty et al. 2021; Züst and Agrawal 2017), any diversion of primary metabolites for secondary metabolism would compromise plant growth. According to the resource-drain hypothesis, carbon for growth and primary metabolite biosynthesis becomes less available, when plants produce specialized metabolites (Foyer et al. 2007). Besides this, the biosynthesis of specialized metabolites often requires many elements like sulfur and nitrogen, which are limiting factors for plant growth (Foyer et al. 2007). In agreement with this concept increased production of anthocyanins was associated with reduced growth in *S. polyrhiza* genotypes (Oláh et al. 2009). Also, many specialized metabolites are auto-toxic to the plant and require energy-intense sequestration (Halkier and Gershenzon 2006) or detoxification mechanisms (Dick et al. 2012; Li et al. 2021) to prevent cell damage.

1.2.1.2 Optimal Defense Theory

Since stressors like herbivores often attack only certain parts of the plant specifically, defense reactions must be coordinated in a tissue or organ-specific manner. The central idea of this so-called optimal defense theory is that defense metabolites are transported to tissues with high metabolic importance for the plant, like flowering tissue (Brütting et al. 2017). The resulting tissue-specific differences in specialized metabolite contents ensure increased protection against stress factors. Since glucosinolates are produced in older leaves and are transported to younger leaves, herbivores were showing stronger avoidance of younger leaves than for older leaves (Hunziker et al. 2021). On a regulatory level the allocation of defense resources to tissues with increased metabolic importance is mediated through cytokinins (Brütting et al. 2017). These differences in specialized metabolite contents between tissues were caused by different

genetic control mechanisms, resulting in tissue-specific expression patterns of many genes (Chen et al. 2016).

1.2.2 Environment-Driven Evolution of Mating System

Besides plastic responses, trade-offs and resource allocation patterns are also of central importance for plant morphogenesis and evolution. An example of this is the evolution of plant mating systems. Since the formation of reproductive organs and seed production are costly, an increased resource allocation to defense is thought to compromise the development of the sexual reproductive system in a similar way like growth (Ramos and Schiestl 2019, 2020). Genetically reproductive traits and defense traits can be associated via pleiotropy, since mutations changing reproduction might also change resource allocation between growth and defense (Johnson et al. 2015). This pleiotropy can be conferred by one gene functioning in several biosynthetic pathways or transport processes of metabolites controlling flower development and defense simultaneously. For instance, through their function as floral color pigments, anthocyanins promote the attraction of pollinators and herbivores to plants (Johnson et al. 2015) and simultaneously mediate plant resistance against cold stress (Schulz et al. 2015). Increased allocation of anthocyanins to flower tissues might, therefore, compromise plant resistance to environmental stress factors.

In the case of *Brassica rapa* increased resource allocation to defense as a consequence of herbivory led to the evolution of fewer flowers that were less attractive to pollinators (Ramos and Schiestl 2019). As a consequence of the resulting pollen limitation, plants evolved higher rates of selfing for reproductive assurance through reduced herkogamy (Ramos and Schiestl 2019). Since selfing leads to inbreeding depression through decreasing gene flow and effective population size (Eckert et al. 2010; Kuester et al. 2017), herbivory can trigger coevolution of mating system and herbivore resistance at the costs of reduced plant fitness through inbreeding. Reduction of gene flow and inbreeding can also promote the spread of herbicide resistance conferring mutations in plant populations. For *Ipomoea purpurea* self-pollination rates through reduced herkogamy were higher in genotypes with high herbicide-resistant levels compared to the ones with lower resistance (Kuester et al. 2017). Through the reproductive isolation of resistant genotypes, alleles driving selfing associate with loci enhancing herbicide resistance, which in turn drives faster elimination of susceptible alleles since allelic segregation in selfing populations increases homozygosity levels (Kuester et al. 2017). On an environmental scale, selfing was therefore more frequent in fragmented or disturbed environments where outcrossing

is limited by the availability of pollen (Eckert et al. 2010). Conversely, increased outcrossing occurred preferentially in environments with relaxed growth-defense trade-offs, allowing more resource allocation to pollinator attraction (Eckert et al. 2010).

Besides this, many plant species can also reproduce clonally to avoid costs associated with sexual reproduction (Wang et al. 2024). The avoidance of such costs is especially helpful during the colonization of new environments since it allows faster reproduction, but might decrease fertility due to the accumulation of deleterious mutations as long-term consequences (Barrett 2015). Clonal reproduction is often connected to aquatic environments where genotypes show long-distance dispersion, because the production of vegetative shoots minimizes the risk of death and leads to an increased division of labor (Barrett 2015). Also, carbon resources are more limited in aquatic environments disfavoring the production of costly reproductive organs (Santamaría 2002).

1.3 Genome-Wide Association Studies for Studying Metabolism

As shown in the previous section plant metabolism is characterized by strong trait correlation patterns. Since strong trait correlation easily leads to false positive signals, genetic association studies aiming to elucidate the genetic basis of plant metabolism often require large sample sizes to have sufficient power to detect true causative loci (Angelovici et al. 2017; Chen et al. 2014; Chen et al. 2016). Due to this high degree of complexity and since many inducible metabolic traits are heritable (Campbell and Kessler 2013), GWAS can be used as a tool to study the genetic background of plant metabolism. A GWAS conducted on shade avoidance in *A. thaliana* lead to the identification of markers associated with the phytochrome C gene and phytohormone signaling and biosynthesis (Filiault and Maloof 2012), providing valuable insights into the genetic basis of a plastic trait.

The concept of genetic association studies first emerged around 2000 and has received lots of methodological improvements since then. The first association studies were conducted with small sets of genetic markers on regions of interest identified in other studies (Tibbs Cortes et al. 2021; Thornsberry et al. 2001). After the availability of SNP datasets representative for entire genomes, created in the human genome project, GWAS was finally possible for *Homo sapiens*. Since association tests in large sets of genetic markers from individuals with different degrees of relatedness would require methods to control population stratification, mathematical methods for estimating population structure were already developed before the first genome-wide SNP datasets were available for *Homo sapiens* (Tibbs Cortes et al. 2021). The aim of

population structure estimates is to correct for loci correlated by relatedness of individuals to prevent an inflation of *P*-values caused by relatedness (Pritchard et al. 2000a). As an example, the STRUCTURE software works through the definition of subpopulations in the dataset based on null markers (Pritchard et al. 2000a).

The generalized linear model (GLM) was then among the first approaches which allowed correcting for population structure by adding subpopulations defined by STRUCTURE as cofactors into the model (Pritchard et al. 2000b). After the identification of population structure, tests for significant association were conducted within each subpopulations structure (Pritchard et al. 2000b) (Figure 1). After publication of the draft human genome the first GWAS was successfully conducted on the myocardial infarct risk of *Homo sapiens* (Ozaki et al. 2002). However, the first GWAS conducted in plants on the traits flowering time and pathogen resistance in *A. thaliana* in 2005 identified many false positives signals (Aranzana et al. 2005). Since STRUCTURE population analysis assumes random mating and strong population admixture with individuals jointly belonging to two or more subpopulations (Pritchard et al. 2000a), it is not suitable for GWAS in selfing species, which have a long history of isolation by distance and inbreeding, explaining the high amounts of false positives (Aranzana et al. 2005).

The Mixed Linear Model (MLM), developed in 2006, splits individuals relatedness into two components: kinship, which estimates relatedness between individuals as identity by descent and population structure, therefore assuming no homogenous subpopulations (Yu et al. 2006). Due to this improvement, MLM performed better in GWAS conducted for *A. thaliana* than GLM (Yu et al. 2006).

A further improvement of computational power on GWAS was achieved through the development of multi-locus models, which allowed the identification of small effect loci (Tibbs Cortes et al. 2021). Running through several iterations, the most significant SNP is added as covariate in each iteration (Tibbs Cortes et al. 2021; Segura et al. 2012). The model is then iterated till covariance explained by the remaining markers is 0 or till a user-defined threshold (Tibbs Cortes et al. 2021; Segura et al. 2012). The first multi-locus model MLMM (multi-locus mixed model), developed in 2012, performed stepwise MLM, incorporating significant markers simultaneously as covariates, but still being computationally expensive (Segura et al. 2012).

The problem of lacking computational efficiency was solved by the development of the Fixed and random model Circulating Probability Unification (FarmCPU) (Liu et al. 2016). Being also a multi-locus model, FarmCPU divided the genome into bins of equal sizes and selected the

marker with the lowest *P*-value from each bin as quantitative trait nucleotide (QTN). The best combination of bin and QTNs is then determined via a random effect model (Liu et al. 2016).

The major problem with FarmCPU is that it assumes significant markers to be homogenously distributed across the genome and its random model, which is used for kinship calculation is computationally very expensive (Liu et al. 2016; Huang et al. 2019). Like FarmCPU, the Bayesian-information and linkage-disequilibrium iteratively nested keyway method (BLINK) also uses maximum likelihood for calculating iteration thresholds (Huang et al. 2019). However, by replacing the bin method with Bayesian criteria on linkage disequilibrium information, BLINK eliminates the computationally expensive random effect model, allowing significant markers to be non-homogenously distributed across the genome (Huang et al. 2019) (Figure 1).

Since GWAS usually tests for associations among huge amounts of markers, significance thresholds have to be adjusted to counter the multiple testing problem. To correct for this, previous GWAS used Bonferroni corrections or minor allele frequencies to correct significance thresholds according to the number of markers studied (Tibbs Cortes et al. 2021). More recently, permutation-based thresholds have become increasingly popular in GWAS (Che et al. 2014; John et al. 2022). Permutation-based approaches calculate significance thresholds through multiple randomizations of phenotypic values. The significance score for permutation-based GWAS is then calculated as the fraction of all permutations, which gives more significant results than the original dataset (Backes et al. 2014). Permutation, therefore, allows a non-parametric estimate of significance thresholds in situations when the sample size is low, or the distribution of phenotypic values is unclear for another reason (Che et al. 2014).

Together, the recent methodological improvements in increasing power and computational efficiency render GWAS as an exceptionally suitable method for studying the genetic basis of metabolic traits in plants.

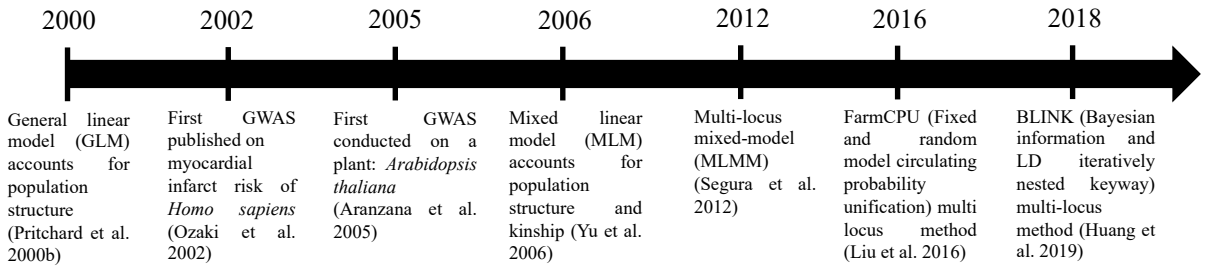


Figure 1: Timeline depicting selected key events in the development of genome-wide association studies (GWAS).

1.4 Metabolism of *Spirodela polyrhiza*

Studying the genetic mechanism of metabolic traits in plants using genetic association studies requires the selection of a suitable model organism. Due to their strongly reduced body plans but high metabolic and genetic diversity reflected by their worldwide distribution (Acosta et al. 2021; Tippery and Les 2020), duckweeds are good candidates for studying the genetics underlying plant metabolism. In this chapter, I aim to introduce the physiological features of the duckweed *Spirodela polyrhiza*, qualifying it as a model for studying metabolic pleiotropy in plants.

Duckweeds are small, free-floating, aquatic macrophytes inhabiting freshwater systems. Phylogenetically duckweeds are monocots and were previously classified as a subfamily of Araceae (Lemnoidea) (Nauheimer et al. 2012). However, the branching from the remaining Araceae happened 103.6 million years ago (Nauheimer et al. 2012) and duckweeds form a physiologically and morphologically well-defined class within monocots (Sree et al. 2016). Because of that, they were grouped as an own family named Lemnaceae, which contains 36 species and five genera: *Spirodela*, *Landoltia*, *Lemna*, *Wolffiella* and *Wolffia*, more recently (Sree et al. 2016; Bog et al. 2020). The evolutionary trajectory of Lemnaceae led to a simplification of their body plan, with the two most ancestral genera *Spirodela* and *Landoltia* (Tippery et al. 2015) showing the greatest morphological differentiation.

Since *S. polyrhiza* has the smallest genome and the most complex body plan among duckweeds (Wang et al. 2014; Xu et al. 2019), it became an interesting model system for studying the genetic basis of duckweed metabolism (Wang et al. 2014). In the case of *S. polyrhiza*, its body consists of a roundish leaf/stem-like structure, which is referred to as a frond and several adventitious roots (Acosta et al. 2021; Landolt 1988). The high amounts of aerenchyma in frond tissue create enough buoyancy to let them float on the water surface (Ziegler et al. 2023). Each frond has two meristematic pockets from which new daughter fronds emerge via asexual budding. Clusters of several fronds are called colonies and are connected via stipes (Landolt 1988). Further, *S. polyrhiza* shows a very reduced shoot structure at the base of its stipe (Acosta et al. 2021; Landolt 1988). Like other duckweeds, *S. polyrhiza* is capable of flowering but almost entirely reproduces asexually, remaining in the juvenile stage for its entire life, which is why duckweeds were seen as neotenous plants (Wang et al. 2014; Hoang et al. 2019). *S. polyrhiza* is world-widely distributed and present in all climate zones except of deserts (Ziegler

et al. 2023). The adaptation to these different environmental setups required large metabolic adaptions for this species.

Caused by differences in climatic factors, aquatic environments strongly differ in nutrient levels (McDowell et al. 2020), requiring high metabolic flexibility of *S. polyrhiza*. As an adaptation to different nutrient levels in water *S. polyrhiza* can switch between heterotrophy, mixotrophy and autotrophy (Sun et al. 2023). For instance, the presence of an organic carbon source favors the mixotrophic growth of duckweed (Firmin et al. 2022), which is characterized by reduced photosynthesis rate, protein content and higher contents of rubisco compared to phototropic growth (Sun et al. 2023). Under phototropic conditions, *S. polyrhiza* shows elevated contents of secondary metabolites such as flavonoids (Sun et al. 2023). Together, these findings reveal high levels of adaption to environmental conditions involving large metabolic reprogramming. In the following sections, we discuss how metabolic and genetic features shape growth (1.4.1), development (1.4.2) and stress response (1.4.3) in *S. polyrhiza*, before finally looking into the genetic diversity underlying these traits (1.4.4).

1.4.1 Growth and Reproduction

In its asexual reproduction mode *S. polyrhiza* reproduces through alternating detachment of daughter fronds from its vegetative pouches. The second daughter frond detaches from the other pocket as the first daughter with some delay (Landolt 1988) (Figure 2). Since asexual reproduction does not require costly investments in reproductive organs and seed development, reproduction can go at much shorter time scales, as shown in section 1.2.2, making *S. polyrhiza* one of the world's fastest-growing angiosperms with a biomass duplication time of less than two days (Ziegler et al. 2015). For overwintering, *S. polyrhiza* produces turions, which sink to the ground and germinate in spring. The rate of sexual reproduction is likely less than 1% for this species (Ho et al. 2019). However, this value might also reflect genetic variation due to past sexual reproduction events (Ho et al. 2019).

The high amount of clonal reproduction is thought to have been beneficial during recent migration events allowing fast colonization of new aquatic habitats (Wang et al. 2024; Xu et al. 2019). Since nutrient levels, light and temperature are relatively constant over a day within aquatic environments, loss of genes associated with environmental synchronization of growth might have favored the evolution of an asexual reproduction style, as discussed in section 1.2.2 (Michael et al. 2021). Supporting this argument *S. polyrhiza* has 1/3 less genes involved in light signaling and circadian clock regulation compared to *Oryza sativa* (Michael et al. 2021).

These asexual growth rates further show large intraspecific variation in duckweed (Sree et al. 2015; Ziegler et al. 2015). Physiologically, these growth rates were positively correlated with increased metabolic activities such as the photosynthesis rate (Sree et al. 2015). Furthermore, intraspecific differences in growth rates of *S. polyrhiza* were shown to be caused by differences in resource allocation (Sun et al. 2023). In this regard, lower starch content was connected to increased growth rates (Sun et al. 2023). Since plants with larger frond sizes were shown to have lower frond duplication rates than plants with smaller frond sizes, a second trade-off exists between the duplication of fronds and the increase of frond area (Strzałek and Kufel 2021). Because frond enlargement is connected to shady environments, growth investment to larger frond size might represent a shade adaption strategy (Strzałek and Kufel 2021).

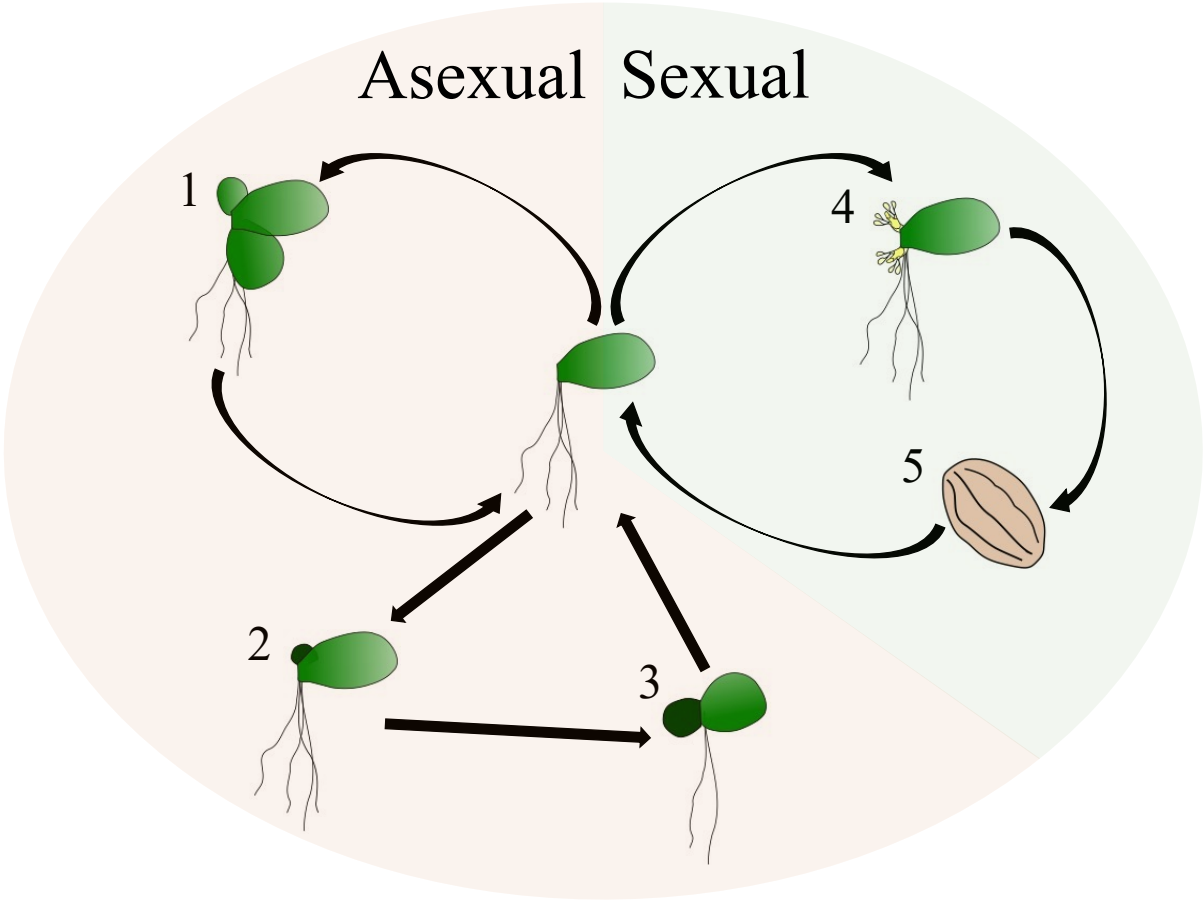


Figure 2: Sexual and asexual reproductive modes of *Spirodela polyrhiza*. Verified connections between life stages are shown with black arrows: The by far largest amount of reproduction occurs via asexual budding of daughter fronds from the mother frond (1). For overwintering, *S. polyrhiza* produces dark green turions (2), which sink to the ground and germinate in spring, giving rise to new daughter fronds (3). Under stressful environmental conditions, flowers can be formed from the meristematic pockets (4). Seed formation (5) and

germination processes were only described as extremely rare natural observations (Saeger 1929; Landolt 1986).

1.4.2 Regulation of Organ Development and Morphogenesis

The simplification of the body plan and decrease in developmental complexity is reflected by reduced contents of protein-coding genes in the *S. polyrhiza* genome. With a total of number of 19623 genes, *S. polyrhiza* has 28% and 50% fewer protein-coding genes compared to *Arabidopsis thaliana* and *O. sativa*, respectively (Wang et al. 2014). Among these gene losses were mainly genes involved in lignin biosynthesis and cell wall organization, which have critical functions for the formation of complex tissues in higher plants (Wang et al. 2014), explaining the overall morphological simplification of *S. polyrhiza*.

Genetically the suppression of flowering might be explained through an increased copy number of genes and microRNAs repressing the transition from juvenile to adult stage increase (Wang et al. 2014). In addition to that, duckweeds seem to be characterized by an exceptionally high loss of genes involved in flower formation. The genome of *S. polyrhiza* lost five out of 16 clades of MIKC^{C}-group MADS-box genes (Gramzow and Theißen 2020), making it the genome with the highest simultaneous loss of MADS-box genes among 26 tested plant species (Gramzow and Theißen 2015).

Consequently, the floral organs of *S. polyrhiza* are of very rudimentary shape (Figure 3) and are only rarely observed in nature. In contrast to most plants, where flowering rates are mostly linked to seasonal changes in day length and irradiance rather than stress, flowering in duckweeds can be induced by environmental stress such as lack of nutrients (Shimakawa et al. 2012; Takeno 2016). Possibly this decoupling of flowering from daylength is explained by the loss of many genes associated with light signaling (Michael et al. 2021), as mentioned in the previous section (1.4.1).

Indeed, *S. polyrhiza* flowers could only be induced through the treatment of high-density cultures with SA (Fourounjian et al. 2021; Khurana and Maheshwari 1980). Since SA functions as a stress hormone in plants (Huot et al. 2014), flowering might be naturally induced under stressful conditions to promote species survival under conditions unlikely for the individual to survive (Shimakawa et al. 2012; Wada et al. 2014). On a molecular level this happens through a synchronization of SA biosynthesis with that of flavonoids (Shimakawa et al. 2012), which constitute a major class of defense metabolites in *S. polyrhiza* (Böttner et al. 2021; Oláh et al. 2009). Through catalyzing the first step in both SA and flavonoid biosynthesis out of L-

phenylalanine the phenylalanine ammonia lyase (PAL) plays a central role in this synchronization process (Gordon and Koukkari 1978). PAL is stress-inducible, showing increased expression under poor nutritional status (Shimakawa et al. 2012) and low temperature (Wada et al. 2014), explaining the joint induction of flowering and flavonoid biosynthesis by stress in duckweed under these conditions (Takeno 2016).

However, older studies described seed production in *S. polyrhiza* after pollination as an extremely rare biological event (Landolt 1986; Saeger 1929). A more recent attempt to experimentally produce seed in *S. polyrhiza* has not been successful (Fourounjian et al. 2021). Eventually, these conflicting results indicate, that some genotypes of *S. polyrhiza* have lost the ability to complete their sexual reproduction cycle entirely. Low rates of seed production might also refer to self-sterility in *S. polyrhiza*, as shown for *L. minor* (Landolt 1986). A possible explanation for the rarity of seed development is that many genes whose orthologs were involved in seed formation were upregulated during the transition from the vegetative frond stage to the resting turion stage (Pasaribu et al. 2023). This suggests a re-functionalization of genes controlling seed development for turion production in *S. polyrhiza* (Pasaribu et al. 2023).

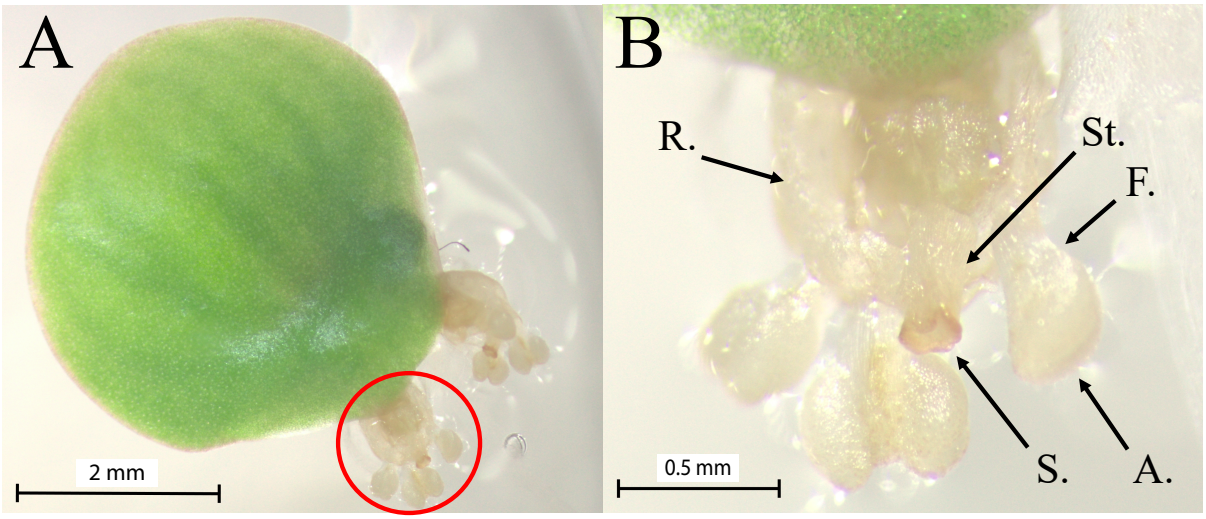


Figure 3: Flower morphology in *Spirodela polyrhiza*. A: Two flowers emerge from the meristematic pockets of a mother frond after six weeks induction with 1 µM Salicylic acid and 25 µM EDDHA in Hoagland medium. The red circle highlights a single flower, illustrated in greater detail (B) B: *S. polyrhiza* flowers consist of four stamina, which are further divided into anthers (A.) and filaments (F.) and one pistil with pigmented stigma (S.) and short style (St.). The floral organs are attached to a receptacle (R.). Pictures: ® Martin Höfer

1.4.3 Stress-Response to Environmental Factors

Gene loss and re-functionalization events also seem to be characteristic of stress responses in this duckweed species. As a consequence of its aquatic lifestyle, *S. polyrhiza* is confronted with large amounts of bacterial and viral pathogens, which is reflected by an increase of gene families involved in antimicrobial defense and immune response on the genome level (Wang et al. 2014). Intriguingly, Lemnaceae were shown to lack the EDS1 (ENHANCED DISEASE SUSCEPTIBILITY 1) pathway, which is a key pathway for microbial defense in higher plants (Baggs et al. 2022). Instead defense reactions towards pathogen infection were mainly characterized through the upregulation of CYP450 genes, which are involved in secondary metabolite biosynthesis and genes encoding antimicrobial peptides (Baggs et al. 2022). Together, this suggests that specialized metabolites play a central role in duckweed defense towards pathogens.

This hypothesis is further supported by the antimicrobial activities of duckweed exudates, which are rich in flavonoids (Ziegler et al. 2023). Also, the overall high concentrations of flavonoids in duckweed extracts might contribute to preventing systemic infections of duckweed tissue (Dan et al. 2022). Flavonoids were further shown to play a central role in stress responses towards environmental factors (Zhao et al. 2015). Different flavonoid classes seem to have stress-specific functions in duckweed. Whereas anthocyanins were shown to protect *S. polyrhiza* specifically against UV light and Cr (VI) mediated stress (Oláh et al. 2009; Böttner et al. 2021), luteolins protect the plant from copper-induced stress (Böttner et al. 2021). At an organismic level, high amounts of reactive oxygen species produced by these heavy metals are compensated through a temporal increase of antioxidant levels such as glutathione (GSH), allowing *S. polyrhiza* to accumulate high amounts of copper and cadmium (Zhao et al. 2017).

Lacking the possibility of sexual reproduction almost entirely, duckweeds seem to possess a trans-generational stress memory via methylation signals. For instance, CHG methylation levels showed temperature effects with high transgenerational stability (Van Antro et al. 2023). This suggests that duckweeds can inherit their responses to environmental stress through an epigenetic mechanism, allowing fast adaptions to new conditions for the following generations.

1.4.4 Genetic Diversity

As a consequence of this asexual reproduction-mode, favoring non-genetic inheritance, *S. polyrhiza* has an exceptionally low mutation rate (Xu et al. 2019). Its low genetic diversity is also reflected by low methylation levels (Wang et al. 2024) and pairwise nucleotide diversity

($\pi = 0.00093$) (Xu et al. 2019). These features are explained by a high portion of clonality even for genotypes separated through large geographic distances. Even though the mutation rate is quite low, *S. polyrhiza* has a large effective population with a size of 9.8×10^5, possibly due to its high census population size and worldwide distribution (Xu et al. 2019).

A genetic analysis of the *S. polyrhiza* diversity panel revealed the existence of 159 clonal families (Wang et al. 2024). These genotypes could be grouped into four genetic populations named after their preferred geographic distribution: Europe, SE-Asia, India and America (Xu et al. 2019). A reason for this high clonality might be that the *S. polyrhiza* genome seems to have experienced only weak purifying selection, as indicated by the high K_a/K_s ratio of 0.495 (Ho et al. 2019).

In addition to that *S. polyrhiza* has one of the smallest genomes among duckweeds with only 158 Mbp containing a high density of protein-coding genes (Wang et al. 2014). Since large sample sizes are required to detect common variants with small effects (Duncan et al. 2019), the combination of low mutation rate, large effective population size and compact genome make *S. polyrhiza* an ideal candidate for genetic association studies to explore the genetic foundation explaining heritable metabolic adaptions.

2. Aims of this Thesis

Due to its small body size and high reproduction rate, *S. polyrhiza* is highly suitable for large-scale phenotypic screening procedures to identify phenotypic variation. From a genetic perspective the low genetic diversity and low mutation rate of *S. polyrhiza* (Xu et al. 2019) reduce the number of false positives in GWAS.

Also, due to their high susceptibility to bipyridinium herbicides (Tyler et al. 2006; Cedergreen et al. 2007), duckweeds were used as bioindicator species for herbicide pollution in freshwater systems (Funderburk and Lawrence 1963). For duckweeds, resistance to bipyridinium herbicides was described previously (Tyler et al. 2006), suggesting the existence of great variation in resistance levels. Mechanistically, bipyridinium herbicides like diquat act through diversion of electrons from photosystem I (PSI), creating reactive oxygen species through the Mehler-like reaction (Farrington et al. 1973; Hawkes 2014). Bipyridinium herbicides have several binding places on PSI, whereas the Fe-S cluster at center B within PSI are the main site of electron acceptation (Fujii et al. 1990).

Since the binding place of bipyridinium herbicide is unspecific, the evolution of target-site resistance mechanisms through structural modifications of PSI is unlikely. As consequence of

this, resistance to bipyridinium herbicides is mainly conferred via non-targeted-site mechanisms (Visnovitz et al. 2008; Shaaltiel and Gressel 1986; Yu et al. 2010), which is why the genetic architecture underlying these resistance traits is likely very complex. Due to the above-mentioned advantages, we aim to use *S. polyrhiza* as a model system to explore NTSR mechanisms to the bipyridinium herbicide diquat.

Besides its high susceptibility to herbicides, *S. polyrhiza* has considerable nutritional value, reflected by its high contents in flavonoids, protein and starch (Smith et al. 2024). Because of that, *S. polyrhiza* is used for bioethanol production as well as human and animal feedstock (Acosta et al. 2021). Further biotechnological optimization of *S. polyrhiza* requires identification of quantitative trait nucleotides (QTNs) to improve free metabolite contents.

Related to that, this thesis has three central goals:

1. We first aim to elucidate the phenotypic variation in resistance levels to the bipyridinium herbicide diquat and to identify candidate genes associated with this NTSR trait using GWAS.
2. Next, we aim to identify genetic associations explaining variation in free metabolite contents and growth using GWAS.
3. Using population genetics methods, we plan to identify candidate genes under species and population-wide selection, determining metabolic and life-history features in *S. polyrhiza*. Among the candidates identified in the previous GWAS to gain insights into the evolution of metabolic traits in *S. polyrhiza*.

Since the identification of gene candidates controlling metabolism using GWAS involves high-throughput phenotyping, several pre-tests will be conducted to determine the herbicide screening concentration and the endpoint of the toxicity assays. The screening concentration for diquat will be determined using a two-step approach. First, we will narrow down the range of potential concentrations in a dose-response experiment with two genotypes. Furthermore, duration and inoculum size will be adapted with this first test. Next, we estimate the broad-sense heritability of diquat resistance through repeated testing of 20 different genotypes at two different diquat concentrations in two batches. For the herbicide screening, we will select the concentration causing the highest heritable variation within this pre-liminary set of 20 genotypes.

For the screening, genotypes were screened under diquat and control conditions in triplicate (Figure 4). Fitness under diquat was expressed relative to the untreated control. Due to the high number of genotypes tested, the assay must be split up into five batches. To correct for batch

effects, the resistance of each genotype was standardized to that of genotype SP028, which was analyzed throughout all batches as an internal standard.

The sampled plant material will be freeze-dried for biomass measurement and metabolite analysis. The diquat treated samples were used for measurement of the diquat tissue concentration. Control samples were used for measurement of free metabolite contents, quantifying the concentrations of 42 free metabolites, including eleven phytohormones, eleven specialized metabolites and 20 amino acids and derived compounds, using HPLC-MS. Broad sense heritability of growth and free metabolite concentrations were calculated from repeatedly analyzed genotypes. To further characterize the physiological basis of the phenotypic variation of resistance, we determined the eight most resistant and eight most susceptible genotypes from the herbicide screening procedure in a dose-response experiment measuring fitness under diquat exposure. The freeze-dried plant material from these tests will be used for the quantification of diquat uptake kinetics and glutathione and ascorbate contents to evaluate the physiological effects of diquat resistance in genotypes. Since differences in growth speed between genotypes might have affected the diquat resistance in our fitness test, we measured chlorophyll fluorescence as Fv/Fm and Fv`/Fm` in a time course experiment under high diquat concentrations to confirm results from screening experiments. We will finally conduct GWAS on growth, free metabolite contents and diquat resistance to discover the genetic basis of metabolic traits using the vcf2gwas platform (Vogt et al. 2021). To gain insights into evolution metabolic traits in *S. polyrhiza*, we will identify candidate genes under species-wide or population-wide selection using results from our population genetics study on this plant. The expression of the most important candidate genes identified using GWAS or selection analysis will be checked using qPCR.

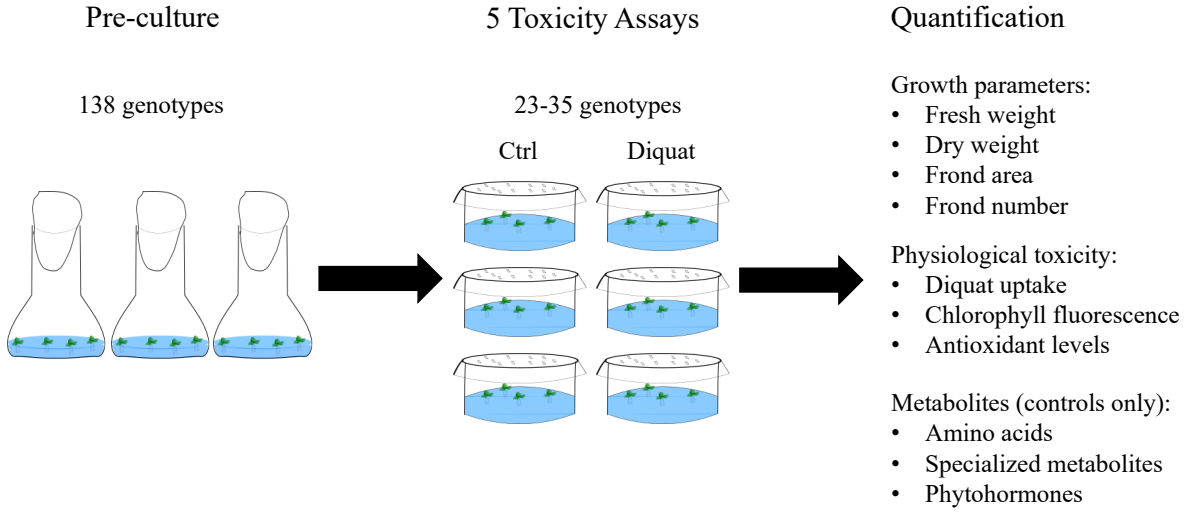


Figure 4: Schematic workflow of phenotypic characterization of 138 *Spirodela polyrhiza* genotypes. Genotypes were propagated in Erlenmeyer flasks for 14 days, prior to the toxicity assays. For measuring diquat resistance each genotype was cultivated in triplicates under control conditions and at 10 nM diquat for 7 days. For each genotype, diquat toxicity was determined as growth inhibition measured through frond area, frond number, fresh weight and dry weight relative to an untreated control. In separate experiments genotype-specific differences in herbicide uptake, chlorophyll fluorescence kinetics and antioxidant levels were determined under diquat exposure. For each genotype, constitutive levels of free amino acids, specialized metabolites and phytohormones were measured using LC-MS.

3. Articles

3.1 Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in *Spirodela polyrhiza*

3.1.1 Summary

In this project, we focused on exploring the physiological and genetic basis of diquat resistance in the giant duckweed *Spirodela polyrhiza*. To examine intraspecific variation in resistance levels, we screened 138 genotypes of *S. polyrhiza* for their fitness under diquat treatment. The screening revealed differences in resistance levels by factor 8.5 between the most resistant genotype SP050 and the most susceptible genotype SP011. To check the involvement of transport-associated processes in diquat resistance, we measured the diquat tissue concentration for each genotype at the end of the screening procedure using LC-MS. Diquat tissue concentration was positively correlated with susceptibility to diquat. Analyzing genotype-specific uptake kinetics, we found that SP050 showed reduced herbicide uptake after 24h of diquat exposure compared to SP011. Genotype-specific antioxidant protection towards diquat-induced oxidative stress was measured through examination of superoxide dismutase (SOD) activity and measurements of glutathione and ascorbate contents in SP050 and SP011. SP050 showed increased antioxidant protection in terms of elevated SOD activity and increased contents of reduced glutathione and ascorbate relative to their oxidized counterparts compared to SP011. Conducting a genome-wide association study (GWAS) on resistance levels of the screened genotypes using SNPs and structure variations as input, we identified three genetic markers located within a *MITOCHONDRIAL SULFUR DIOXYGENASE* (*SpETHE1*), the *LIPOXYGENASE2.1* (*SpLOX2.1*) and *GUANYLATE-BINDING PROTEIN* (*SpGBPL2*) to be associated with resistance. Studying the expression of *SpETHE1* using qPCR we did not find any difference for this gene between SP050 and SP011, making a direct involvement of *SpETHE1* in diquat resistance unlikely. Studying genes in close distance to *SpETHE1*, we found *DIENELACTONE HYDROLASE* (*SpDLH*) and the *STARCH-BRANCHING-ENZYME 3* (*SpSBE3*) showing reduced expression in SP050 compared to SP011. Taken together, this study suggests that uptake and antioxidant processes jointly contributed to diquat resistance in *S. polyrhiza*. Whereas antioxidant mechanisms seemingly provide short-term protection against diquat-induced oxidative stress before transport-associated processes enable a long-term fitness benefit under diquat exposure. Further, the study associated genes involved in stress-response pathways (*SpETHE1*, *SpDLH* and *SpGBPL2*) and starch metabolism (*SpSBE3*) with diquat resistance.

3.1.2 Zusammenfassung

Der Fokus dieses Projektes lag in der Charakterisierung von physiologischen und genetischen Mechanismen, die Resistenz gegen das Herbizid Diquat in *Spirodela polyrhiza* vermitteln.

Zur Bestimmung der intraspezifischen Varianz an Resistenzen wurden 138 Genotypen von *S. polyrhiza* hinsichtlich der Inhibition von Fitnessparametern charakterisiert. Der resistenteste Genotyp SP050 zeigte eine um den Faktor 8.5 höhere Resistenz als der am wenigsten resistente Genotyp SP011. Um eine Beteiligung von Transport-assoziierten Mechanismen an der ermittelten Resistenz zu testen, wurde die Gewebekonzentration von Diquat für sämtlich Genotypen nach Ende der Diquat Behandlung mittels LC-MS bestimmt. Die Diquatkonzentration des Pflanzengewebes zeigte eine positive Korrelation mit der Inhibition durch das Herbizid. Eine Analyse der Genotyp-spezifischen Aufnahmekinetik des Herbizids ergab, dass SP050 eine niedrigere Akkumulation nach 24-stündiger Exposition zeigte als SP011. Weiterhin wurden Genotyp-spezifische antioxidative Mechanismen durch Messungen der Aktivität der Superoxiddismutase (SOD) und der Konzentrationen von L-Ascorbinsäure und Glutathion untersucht. Verglichen mit SP011 zeigte SP050 erhöhte SOD Aktivität sowie höhere Werte an reduzierten Glutathion und L-Ascorbinsäure relativ zum Gesamtantioxidantienpool nach Exposition durch Diquat. Eine Genomweite Assoziationsstudie (GWAS) für SNPs und strukturelle Variationen, durchgeführt an Diquatresistenzleveln von 138 Genotypen, assoziierte drei Marker innerhalb von Gensequenzen der *MITOCHONDRIAL SULFUR DIOXYGENASE 1* (*SpETHE1*), der *LIPOXYGENASE2.1* (*SpLOX2.1*) und des *GUANYLATE-BINDING PROTEIN 2* (*SpGBPL2*) mit Diquat Resistenz. Eine Quantifikation der Genexpression von *SpETHE1* mittels qPCR zeigte keinerlei Unterschiede zwischen SP050 und SP011, wodurch eine direkte Beeinflussung der Diquat Resistenz durch dieses Gen als unwahrscheinlich erscheint. Eine Messung der Expression von Genen in unmittelbarer Nähe zu *SpETHE1* zeigte eine verminderte Expression der *DIENELACTONE HYDROLASE* (*SpDLH*) und des *STARCH-BRANCHING-ENZYME 3* (*SpSBE3*) in SP050 verglichen mit SP011. Zusammenfassend wurden unterschiedliche Akkumulationskinetiken des Herbizids sowie antioxidative Stressantworten als ursächlich für die observierte Variation von Diquat Resistenzen ermittelt. Hierbei scheinen antioxidative Mechanismen einen kurzfristigen Schutz gegen Diquat-induzierten oxidative Stress zu vermitteln, bevor Transport-assoziierte Mechanismen ein langfristiges Überleben unter Diquat Exposition vermitteln. Weiterhin assoziierte die Studie Gene mit Bezug zu Stressantworten (*SpETHE1*, *SpDLH* and *SPGBPL2*) und Stärkeabbau (*SpSBE3*) mit Diquat Resistenz.

3.1.3 Statement of Contribution

In this project, Martin Höfer planned all experiments together with his supervisor Shuqing Xu. Together with Samuel Wink and Yangzi Wang, Martin Höfer conducted all toxicity studies. The analysis of SOD activity was done by Martin Höfer. Together with Martin Schäfer, Martin Höfer extracted and quantified antioxidants and diquat in plant tissue. Furthermore, Martin Höfer conducted all analyses, prepared all figures, and wrote the first draft of the manuscript. All co-authors provided suggestions for improving the content of the manuscript.

Supervision confirmation

Prof. Shuqing Xu

iomE
Institut für Organismische und Molekulare Evolutionsbiologie
JG|U Mainz

plants

MDPI

Article

Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in *Spirodela polyrhiza*

Martin Höfer [1], Martin Schäfer [1], Yangzi Wang [1], Samuel Wink [2] and Shuqing Xu [1,*]

[1] Institute for Organismic and Molecular Evolution (iomE), Johannes Gutenberg University, 55128 Mainz, Germany; mschaefer@uni-mainz.de (M.S.)
[2] Institute for Evolution and Biodiversity, University of Münster, 48149 Münster, Germany
* Correspondence: shuqing.xu@uni-mainz.de

Abstract: Understanding non-target-site resistance (NTSR) to herbicides represents a pressing challenge as NTSR is widespread in many weeds. Using giant duckweed (*Spirodela polyrhiza*) as a model, we systematically investigated genetic and molecular mechanisms of diquat resistance, which can only be achieved via NTSR. Quantifying the diquat resistance of 138 genotypes, we revealed an 8.5-fold difference in resistance levels between the most resistant and most susceptible genotypes. Further experiments suggested that diquat uptake and antioxidant-related processes jointly contributed to diquat resistance in *S. polyrhiza*. Using a genome-wide association approach, we identified several candidate genes, including a homolog of dienelactone hydrolase, that are associated with diquat resistance in *S. polyrhiza*. Together, these results provide new insights into the mechanisms and evolution of NTSR in plants.

Keywords: non-targeted-site herbicide resistance; diquat; *Spirodela polyrhiza*; duckweed; GWAS; dose–response measurements

Citation: Höfer, M.; Schäfer, M.; Wang, Y.; Wink, S.; Xu, S. Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in *Spirodela polyrhiza*. *Plants* **2024**, *13*, 845. https://doi.org/10.3390/plants13060845

Academic Editor: Dayong Zhang

Received: 15 January 2024
Revised: 27 February 2024
Accepted: 12 March 2024
Published: 14 March 2024

1. Introduction

More than 500 unique cases of herbicide resistance have been discovered worldwide [1]. Many of these herbicide resistance cases are caused by genetic changes other than the target site and are therefore referred as non-targeted-site resistance (NTSR) cases [2]. Since many NTSR cases constitute complex and polygenic traits, our current understanding of the evolution and mechanism of NTSR remains limited.

Due to their lack of target site resistance mechanisms, bipyridinium herbicides, such as diquat and paraquat, constitute an ideal system for deciphering the genetic basis of non-target-site resistance mechanisms. These herbicides act by diverting electrons from the photosystem I (PSI), thereby competing with ferredoxin for its binding place on PSI [3,4]. Upon electron acceptation from PSI, they transfer one electron to molecular oxygen, generating superoxide anions. The reactive oxygen species (ROS) produced by this Mehler-like reaction lead to the destruction of thylakoid membranes and a decrease in the photosynthesis rate of the plant, which ultimately causes cell death by the induction of apoptosis [5]. With the widespread usage of bipyridinium herbicides for weed and algae control in agriculture and aquaculture [6–9], numerous resistance cases have been reported since the 1980s [10,11]. Most of these resistance cases exhibited a cross-resistance between paraquat and diquat [6,10,12–14]. From a mechanistic point of view, the vast majority of these cases were due to the vacuolar sequestration/chloroplast exclusion of the herbicide [13,15–19], an increased activity of the plant's antioxidant radical scavenging system [20,21], or an intracellular-uptake-associated mechanism [6,22].

Most of the known herbicide resistance mechanisms are also shown to be involved in stress responses to environmental factors such as salinity [23,24], oxidative stress [24], heavy metal exposure [25], and cold stress [24]. Most prominently, ABC transporters, a class

of proteins involved in various stress responses, such as those to drought [26], exposure to heavy metals [27], and pathogens [28], are involved in paraquat resistance [25,29]. These findings suggest that NTSR might have a pleiotropic origin from stress response or secondary metabolite pathways, thus explaining the short evolutionary time scale between the commercialization of herbicides and the appearance of resistant weed species. Yet, the underlying genetic principles remain poorly understood for many cases of bipyridinium herbicide resistance.

Here, we used giant duckweed *Spirodela polyrhiza* (L.) Schleid. (Arales: Lemnaceae) as a model system to study the mechanisms of diquat resistance. Because of their high susceptibility towards bipyridinium herbicides [30], their small body, and fast clonal reproduction [31], duckweeds are often used for toxicological studies on herbicides. Among duckweeds, *S. polyrhiza* has the smallest genome size and low genetic variation [32,33]. The available large genetic diversity panel and re-sequencing data on *S. polyrhiza* will also enable the identification of genetic mechanisms using a genome-wide association (GWA) approach. In this study, we focused on two main questions: (1) To what extent do *S. polyrhiza* genotypes vary in terms of their diquat resistance levels? (2) What are the underlying molecular and genetic mechanisms of diquat resistance?

2. Results

2.1. Intraspecific Variations in Diquat Resistance in S. polyrhiza

As the intra-specific variations in resistance are dependent on the applied diquat concentration, we used a stepwise approach to select the diquat concentration for screening. First, we used two randomly selected genotypes to perform a dose–response curve in *S. polyrhiza*. The results suggested that the EC10 and EC90 values for inhibiting FW (fresh weight) biomass in *S. polyrhiza* are ~1.8 nM and ~11.1 nM, respectively. Measuring the inhibition of the relative growth rate of the frond number, we obtained EC10 and EC90 values of ~2.2 nM and ~53.0 nM, respectively (Figure A1, Table A1). We then screened diquat resistance among 19 genotypes using 5 and 10 nM of diquat. We found the broad-sense heritability of diquat resistance to be higher at 10 nM (H2 = 0.98) than at 5 nM (H2 = 0.90). Furthermore, several genotypes showed hormetic effects at 5 nM of diquat concentration, whereas at 10 nM, we found a clear inhibition ranging from 22% to 84% across all genotypes (Figure A2). Therefore, we chose 10 nM as the concentration for screening diquat resistance among 138 genotypes.

Among the screened genotypes, diquat resistance varied significantly, either estimated based on the RGR (relative growth rate) of the frond number, the RGR of the frond area, or the decrease in biomass (FW) (Figure A3). Of all the quantified parameters, the RGR of the frond number was the most reproducible for quantifying diquat resistance (Figure A4). We did not detect significant differences in diquat resistance between the four genetic populations of *S. polyrhiza* (Tukeys HSD) identified using genome-wide markers (Figure A5). To further validate the variation in diquat resistance in our genotype accession, we performed dose-response experiments and estimated the EC50 values of the eight most resistant and eight most susceptible genotypes identified in the screening experiment (Table A2). The estimated EC50 value of the most resistant genotype (SP050, registered four-digit code: 0090) is 8.5-fold higher than that of the most susceptible genotype (SP011, registered four-digit code: 9242) (Figure 1 and Table A2). The differences in diquat resistance between SP050 and SP011 are also reflected by the decrease in two chlorophyll fluorescence parameters: Fv/Fm and Fv'/Fm', in response to diquat. During the first 8 h of diquat treatment, SP050 has a much slower decay of Fv/Fm and Fv'/Fm' than SP011 (Figure 1C,D). Together, these results suggest that diquat resistance greatly varied among different genotypes of *S. polyrhiza*.

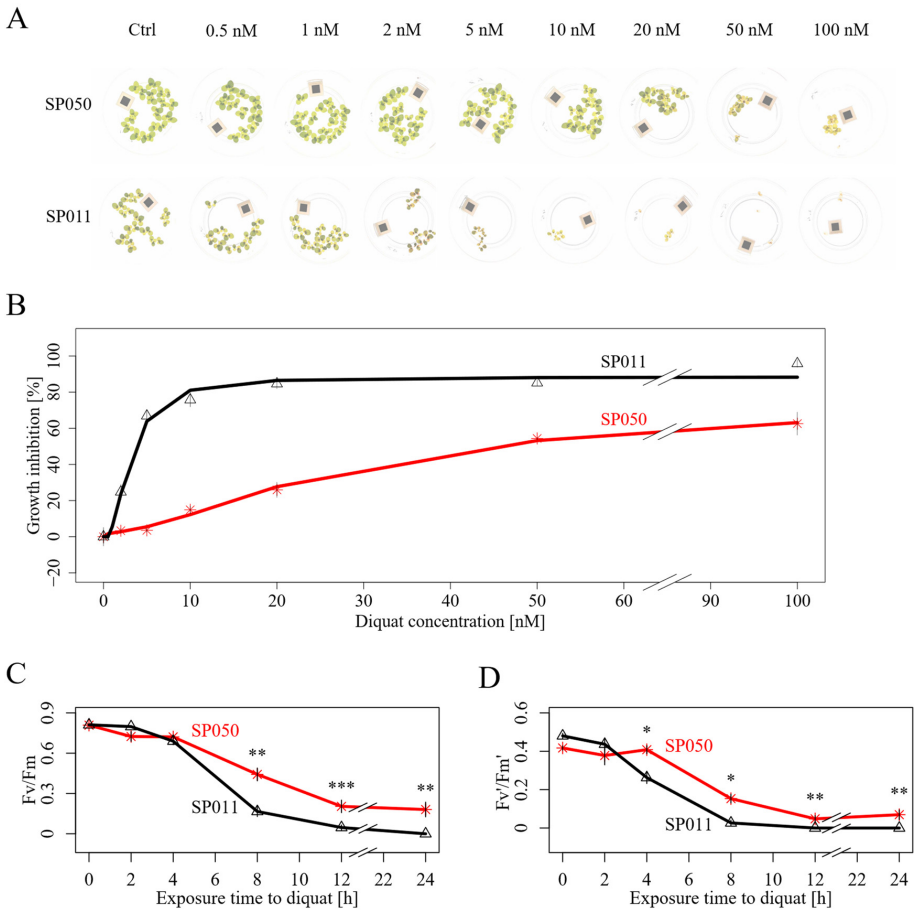


Figure 1. Comparison of growth kinetics of the most resistant genotype SP050 and the most susceptible genotype SP011 under diquat treatment (**A**) *S. polyrhiza* genotypes SP011 and SP050 grown under various diquat concentrations for 8 days. (**B**) Dose-response curves depicting the mean IE values with their corresponding standard errors calculated from three biological replicates for various diquat concentrations. We used the UCRS4c hormetic model from the drc R-package to estimate the EC50 values. The predicted functions used to derive this value were drawn as colored lines. (**C**,**D**) The development of the chlorophyll fluorescence parameters Fv/Fm (**C**) and Fv'/Fm' (**D**) are shown over time. The data points constitute mean values from three biological replicates; error bars indicate ± standard error, and different significance levels were marked with asterisks: *—0.05 ≥ *p*-adjust > 0.01, **—0.01 ≥ *p*-adjust > 0.001, ***—*p*-adjust < 0.001 (linear mixed effect model, Tukeys HSD).

2.2. Changes in Diquat Uptake and Antioxidant Capacity Contribute to Diquat Resistance

We compared the diquat uptake and antioxidant capacities among different genotypes to understand the molecular mechanisms of variations in diquat resistance. We found that diquat concentration in whole plants significantly correlates with inhibition effects, suggesting that reduced diquat uptake capacity increases diquat resistance (Figure 2A). However, the diquat concentration in frond tissue might be lower in highly resistant genotypes due to their larger biomass accumulation compared to susceptible genotypes after the 7 days of bioassay. To address this issue, we further compared the diquat uptake kinetics between SP050 and SP011 within 48 h, during which diquat did not affect plant biomass considerably. The results showed that SP011 accumulated significantly more diquat than SP050 after 24 h (Figure 2B), indicating their differences in diquat uptake capacity. A comparison of the diquat concentration between root and frond tissue suggested that roots have an approximately three-fold higher diquat concentration than fronds (Figure 2C).

However, differences in diquat uptake between SP011 and SP050 were likely not due to their differences in root/frond diquat concentration ratio, which are similar between the two genotypes (Figure 2C). A measurement of the medium concentration of diquat did not reveal large fluctuations in cultures of SP050 and SP011 within a period of 7 days (Table A3). Due to the high stability of diquat in plant tissue and medium samples, the measured uptake kinetics are unlikely to be confounded by degradation processes.

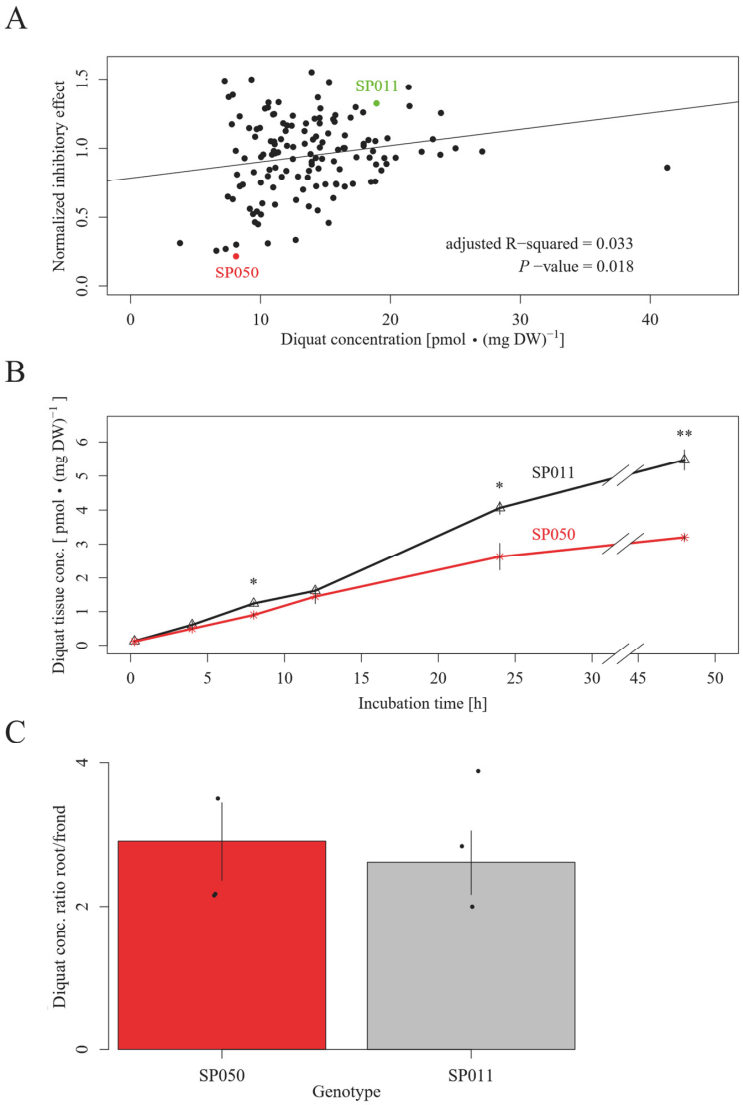


Figure 2. Uptake and translocation mechanisms of diquat. (**A**) The diquat resistance calculated based on the inhibition of the RGR of the frond number is significantly correlated (F-test) with the diquat tissue concentration of 137 genotypes. (**B**) After 24 and 48 h of diquat exposure, SP011 accumulated significantly more diquat than SP050. *—$0.05 \geq p$-adjust > 0.01, **—$0.01 \geq p$-adjust > 0.001 (two-sided Student's *t*-test). (**C**) Both genotypes showed an approximately three-fold higher concentration of diquat in root tissue compared to frond tissue. No significant difference between the two genotypes was found (two-sided Student's *t*-test). The diquat tissue concentration was measured after 24 h of diquat exposure.

We then measured the antioxidant capacity by quantifying the glutathione and ascorbate content levels among eight *S. polyrhiza* genotypes. Glutathione and ascorbate play a

central function in protecting the plant from oxidative stress via the Asada–Halliwell–Foyer cycle (Figure 3A), which controls the redox hub by detoxifying superoxide anions generated by diquat. The quenching process of ROS occurs through the synergistic action of several enzymes, including the superoxide dismutase (SOD) via the conversion of the reductive metabolites AsA and GSH to their oxidized counterparts (Figure 3A), dehydroascorbate (DHA) and oxidized glutathione (GSSG), respectively. Therefore, we used the activity of SOD together with the ratios of AsA/DHA and GSH/GSSG as proxies to estimate antioxidant stress responses to diquat [23,24,34,35]. Diquat exposure decreased the GSH/GSSG and AsA/DHA levels in SP011 but not in SP050 (Figure 3B,C). For SP011, both ratios reached a minimum after 4 h of diquat treatment, followed by a recovery to their initial value after 24 h. The decline in the ratios was mainly reflected by a decrease in GSH and AsA concentrations in plant tissue (Figure A6). Additionally, we found a significant correlation between diquat tissue concentration and GSH/GSSG levels (Figure A7A) but not AsA/DHA levels (Figure A7B). However, the GSH/GSSG and AsA/DHA values were not correlated with resistance levels (Figure A7C,D). The enhanced maintenance of antioxidant ratios in SP050 is also reflected by the genotype's elevated SOD activity under diquat treatment over SP011 (Figure 3D). Together, these data suggest that SOD activity under diquat treatment is correlated with resistance levels to diquat (Figure A8A). However, no such association was observed for the constitutive SOD activity (Figure A8B).

2.3. Genetic Basis of Diquat Resistance in S. polyrhiza

To identify the genetic basis of diquat resistance, we applied a GWA approach using both the single-nucleotide polymorphisms (SNPs) and structure variations (SVs) that were identified in our previous study [37]. Among all 42,462 SNPs, only one synonymous SNP (C to T) located within the coding sequence of the guanylate-binding-protein-like 2 *SpGBPL2* (SpGA2022_050909) showed a significant association with diquat resistance quantified based on FW only (Figures A9 and A10, Table A4). Among the 842 SVs, we found that a 94 bp deletion located in the seventh intron of the mitochondrial sulfur dioxygenase *SpETHE1* (SpGA2022_053619) is associated with diquat resistance quantified based on all growth parameters (Figures 4A–C and A11, Table A5). Genotypes that are homozygous at the deletion locus were significantly more resistant to diquat than heterozygous genotypes or genotypes lacking the deletion. Also, heterozygous genotypes were on average more resistant than genotypes lacking the deletion (Figure 4D). In addition, we found a 184 bp deletion within the fourth exon and the fourth intron of the lipoxygenase 2.1 *SpLOX2.1* (SpGA2022_008887) (Figures 4B, A11 and A12, Table A5) and a 56-bp deletion in an intergenic region (Figures 4C and A11, Table A5) that is ~60 Kb far from the next open reading frame to be associated with diquat resistance. However, the 184 bp and 56 bp deletions are specific to only one of the growth parameters that were used for quantifying resistance effects.

To further validate the association between the 94 bp deletion in the *SpETHE1* intron and diquat resistance, we amplified the deletion region using PCR in the genomes SP050 and SP011. These two genotypes showed different levels of diquat resistance. The results confirmed that SP050 is homozygous for the identified deletion and that SP011 is lacking the deletion completely (Figure A13A). We then aimed to examine whether the deletion affected the expression or splicing of *SpETHE1* using qPCR. To this end, we conducted a time-course experiment and analyzed the expression and splicing of SpETHE1 in the roots and fronds, respectively. We could not find any differences in the splicing pattern between SP050 and SP011, neither for the control nor for the diquat-treated samples (Figure A13B). We found that *SpETHE1* expression was elevated in the root tissue relative to the frond tissue, approximately by a factor of three in both genotypes (Figure A14B). However, under control conditions and within 4 h of diquat treatment, the expression levels were similar between SP011 and SP050 (Figures 5 and A14). After 72 h of diquat treatment, in the fronds, *SpETHE1* expression was 21% lower in SP050 compared to SP011, and it showed similar expression in the roots (Figures 5A and A14A). The differential expression is likely due to the differences in diquat uptake between the two genotypes.

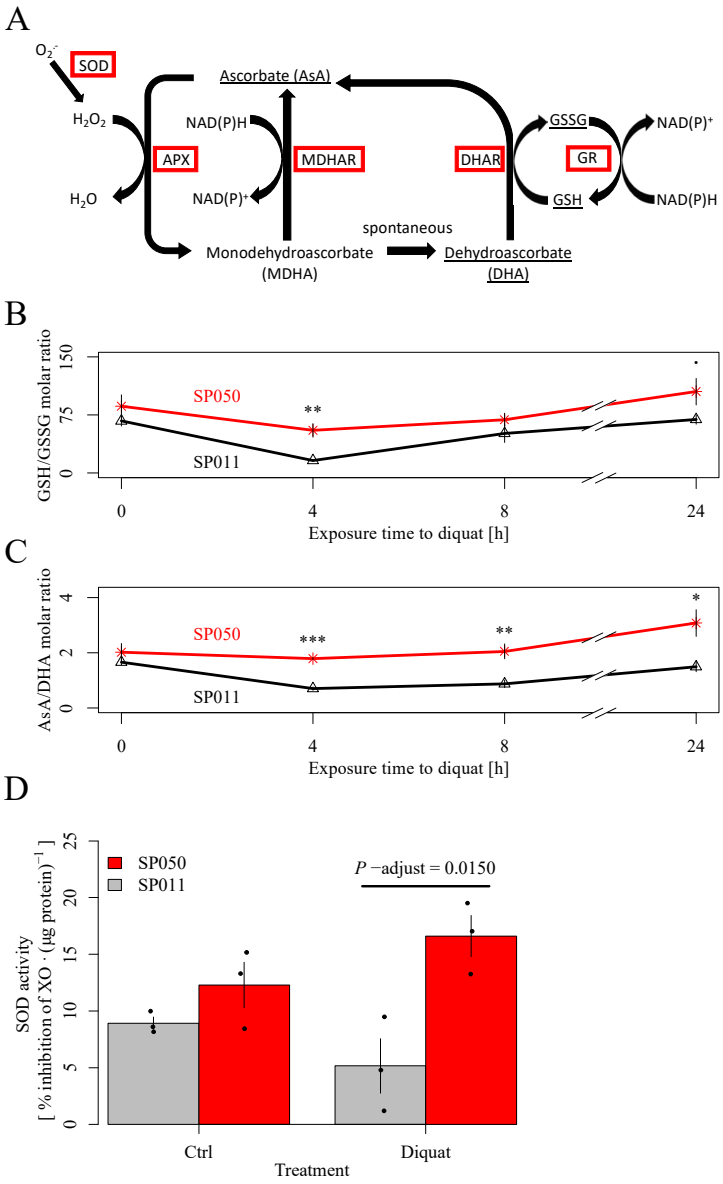


Figure 3. (**A**) Foyer–Asada–Halliwell cycle according to Noctor and Foyer, 1998 [36]. Enzymes are highlighted by red margins. Metabolites analyzed in this study are underlined. SOD converts superoxide anions to H_2O_2. Through oxidation of AsA, H_2O_2 is degraded to H_2O by APX. AsA is regenerated by oxidation of GSH to GSSG. The latter is reconverted to GSH by GR, consuming NADPH. SOD—superoxide dismutase, APX—ascorbate peroxidase, MDHAR—monodehydroascorbate reductase, DHAR—dehydroascorbate reductase, GR—glutathione reductase. The time course of mean GSH/GSSG (**B**) and AsA/DHA (**C**) ratios of four to five replicates with their respective standard errors are shown for SP050 and SP011. SP050 exhibited a significantly increased GSH/GSSG ratio over SP011 after 4 h of diquat treatment and an increased AsA/DHA ratio after 4, 8, and 24 h of diquat treatment (unpaired two-sided Student's *t*-test). (**D**) Mean SOD activity after 1-day incubation at 10 nM of diquat or control conditions for three biological replicates ± standard error. Genotype SP050 exhibited a significantly higher SOD activity than SP011 after 24 h of diquat exposure. *—0.05 ≥ *p*-adjust > 0.01, **—0.01 ≥ *p*-adjust > 0.001, ***—*p*-adjust < 0.001 (two-sided Student's *t*-test).

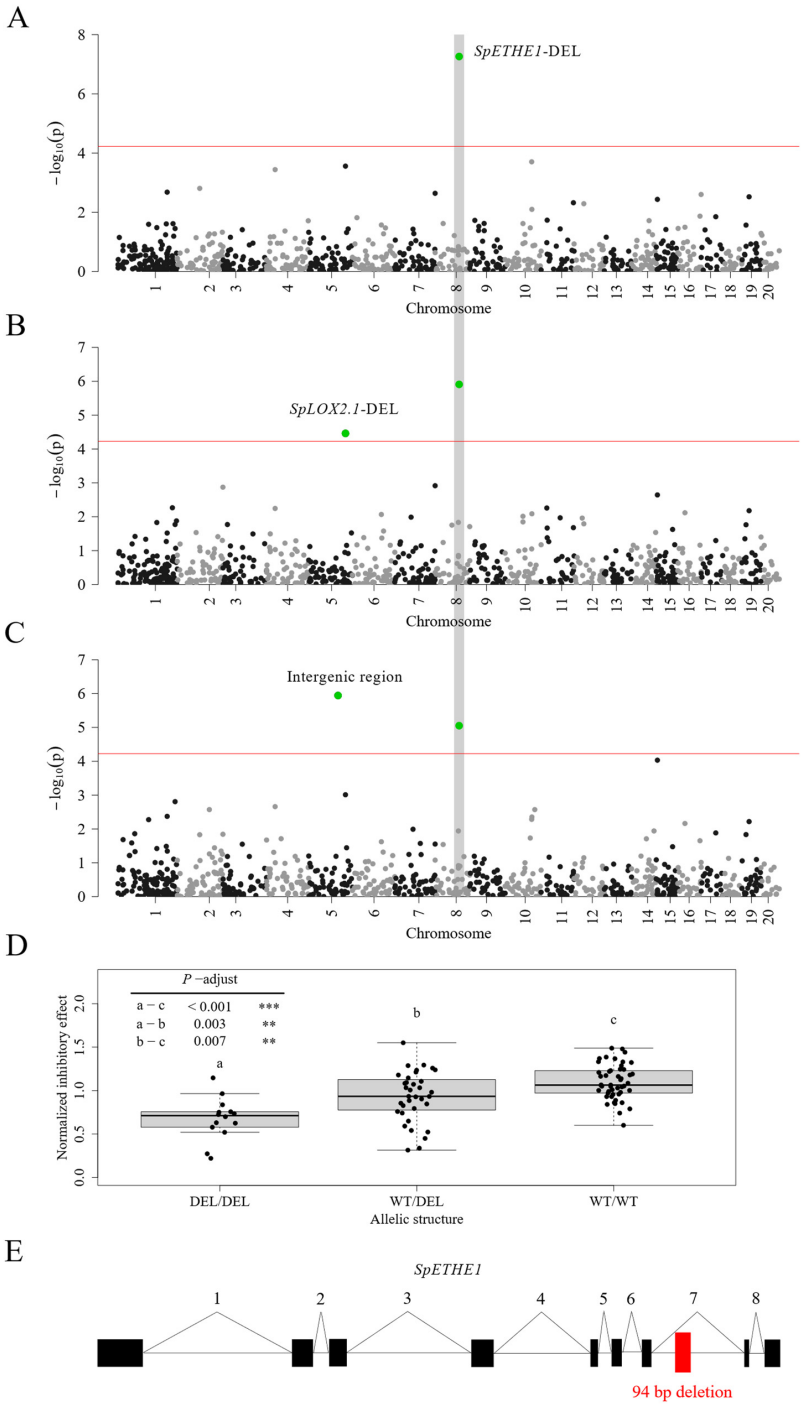


Figure 4. (**A**–**C**) Manhattan plots from a GWAS conducted on structure variations, highlighting the diquat resistance-associated 94 bp deletion in *SpETHE1*. All significant markers (Wald test) were highlighted as green dots. The Bonferroni corrected significance threshold (red line) at $p < 0.05$ is 5.94×10^{-5}. The GWAS was conducted using normalized inhibition effects calculated based on frond number (**A**), frond area (**B**), and fresh weight (**C**). (**D**) Box plot of mean values of the norm. IE RGR of

the frond number of genotypes carrying the respective alleles. On average, genotypes homozygous for the deletion were more resistant to diquat than heterozygous genotypes or genotypes without the deletion. **—0.01 ≥ *p*-adjust > 0.001, ***—*p*-adjust < 0.001 (Tukeys HSD). Heterozygous genotypes exhibited a greater mean resistance than genotypes without deletion. (**E**) Gene structure of *SpETHE1* with the 94 bp deletion in the seventh intron highlighted as a red box.

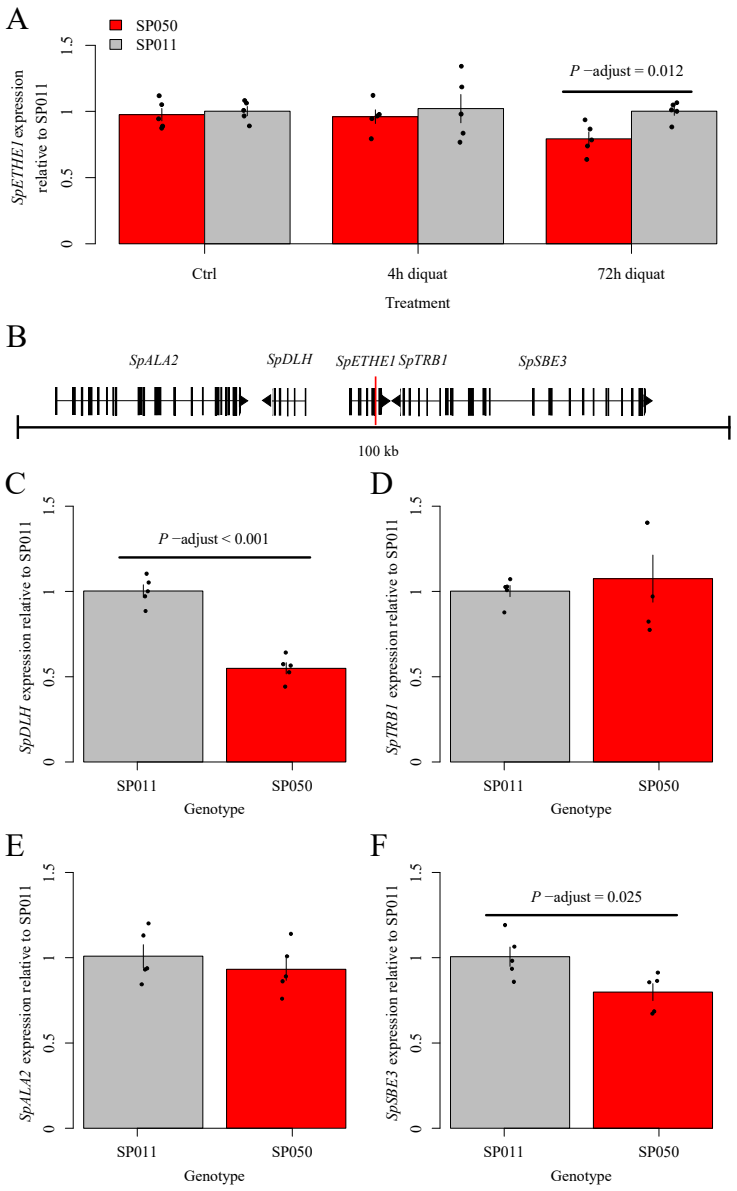


Figure 5. Relative transcript abundance of candidate genes in the frond tissue of *S. polyrhiza*. Mean values of five biological replicates are shown with their respective standard errors. (**A**) *SpETHE1* expression in SP050 and SP011 relative to its mean expression in SP011 without diquat treatment (Ctrl) or treated with diquat for 4 h and 72 h. After 72 h, *SpETHE1* was less expressed in SP050 than in SP011 (two-sided Student's *t*-test). (**B**) The location of *SpDLH*, *SpTRB*, *SpALA*, *SpSBE3* and the deletion in *SpETHE1*. The deletion is depicted as a red line. Transcript abundance of *SpDLH* (**C**), *SpTRB1* (**D**), *SpALA2* (**E**), and *SpSBE3* (**F**) in SP050 relative to their mean expression in SP011. The transcript abundances of *SpDLH* and *SpSBE3* are lower in SP050 than in SP011 (two-sided Student's *t*-test).

Because the genome of *S. polyrhiza* is compact and cis-regulatory elements can be located further away from the gene, we examined the expression of four genes located near *SpETHE1*: *SpDLH* (SpGA2022_012161), *SpALA2* (SpGA2022_053617), *SpSBE3* (SpGA2022_053621), and *SpTRB1* (SpGA2022_012163) (Figure 5B–F and Figure A14). Among them, the expression of *SpDLH*, the start codon of which is located 9.6 Kb downstream of the deletion (Figure 5B), is ~67% lower in the root tissue of SP050 compared to SP011 (Figure A14C). In the frond tissue, the expression was ~45% lower in SP050 compared to SP011 (Figure 5C). *SpDLH* is a homolog of dienelactone hydrolase, which is involved in the degradation of xenobiotics such as chlorocatechol [38] and might be linked to apoptotic processes as well [39]. In addition, the expression of *SpSBE3*, the start codon of which is located 7.7 Kb from the deletion, is 21% lower in the frond tissue of SP050 than in SP011 (Figure 5C). The expression of the other two genes was not different between SP050 and SP011.

3. Discussion

Here, we investigated the diquat resistance in *S. polyrhiza* using the worldwide diversity panel and identified gene candidates associated with the resistance. We found that the levels of resistance varied among genotypes by a factor of 8.5, likely due to the combination of changes in diquat uptake kinetics and antioxidant capacity. We used a GWAS approach to show that several candidate genes involved in metabolic processes or stress responses to biotic or abiotic factors are likely associated with diquat resistance.

Our results did not reveal any gene candidates with a transporter or carrier function that might have caused the observed differences in diquat uptake. Yet, since bipyridinium herbicides were also shown to bind to plant cell walls, restricted herbicide movement into the apoplast might not be exclusively explained through the plasma membrane transporter but also by the genotype's cell wall composition as shown in paraquat-resistant biotypes of *Hordeum glaucum* (L.) Steud. (Poales: Poaceae) and in *Rehmannia glutinosa* (Gaertn.) Steud. (Lamiales: Orobanchaceae) [13,40]. Since we did not check for uptake differences on apoplast and protoplast levels separately, we cannot rule out the involvement of such extracellular movement barriers in diquat uptake. On a systemic level, increased retention of diquat in root tissue seems unlikely to explain differences in resistance levels since the root/frond ratio of the diquat concentration was similar across resistance levels. Alternatively, the observed difference in the diquat concentrations can be due to different capacities in metabolizing the herbicide.

From a physiological point of view, the increased diquat resistance of SP050 was characterized by an increased SOD activity and stable antioxidant ratios under diquat exposure compared to susceptible genotypes. Such genotypic-specific changes in AsA/DHA ratios were previously related to differences in SOD activity [35]. Since the GSH/GSSG ratio and diquat tissue concentration were significantly correlated after 24h of diquat exposure, some of the observed differences in antioxidant ratios might be influenced by diquat uptake kinetics as well. This suggests that diquat resistance in *S. polyrhiza* is likely facilitated by a synergistic involvement of uptake-related and antioxidant processes. Related to this, it was proposed that antioxidant processes only provide short-term protection against herbicide-mediated oxidative stress before more potent mechanisms such as sequestration or reduced uptake of the herbicide secure the long-term survival of resistant genotypes [41], which might explain why diquat-caused changes in the antioxidant ratios were only transient across different resistance levels.

From a genetic perspective, a 94-bp intronic deletion in *SpETHE1* was the only marker associated with several resistance parameters. Since we did not find significant differences in gene expression or splicing patterns associated with the deletion, the direct functional involvement of *SpETHE1* in diquat resistance seems unlikely. Yet, the differential expression of *SpDLH* and *SpSBE3*, two genes downstream of *SpETHE1*, as a potential consequence of the deletion, might suggest the involvement of long-distance gene regulation mechanisms. Such regulatory mechanisms seem to be more common in plant species with large or middle-sized genomes, such as *Zea mays* L. (Poales: Poaceae) [42,43] or *Oryza sativa* L.

(Poales: Poaceae) [44] than in species with small genomes like *Arabidopsis thaliana* (L.) Heynh. (Brassicales: Brassicaceae) [45,46]. Yet, enhancement of transcription through chromatin looping was also shown to be established by introns [47], which might be a process more common in species with compact genomes that lack long intergenic sequences like *S. polyrhiza*.

Many homologs of our candidate genes were involved in various metabolic or stress response pathways. Whereas GBPLs and lipoxygenase like *SpLOX2.1* are involved in response mechanisms to pathogen infection [48–50], *SBE3* is involved in starch metabolism [51]. *DLH* functions in xenobiotic degradation [38,52] and apoptosis [39]. Their association with diquat tolerance suggests a pleiotropic function in herbicide resistance for these genes.

The conceptual connection of pleiotropic gene function to bipyridinium herbicide resistance was first established in *Conyza canadensis* (L.) Cronquist. (Asterales: Asteraceae) [53], where resistance to paraquat was inherited by a single gene that controls the regulation of multiple antioxidant enzymes [53]. In *Avena fatua* L. (Poales: Poaceae), multiple genes with pleiotropic functions that were involved in xenobiotic catabolism, redox maintenance, secondary metabolite pathways, and stress response pathways were differentially expressed at constitutive levels between resistant and susceptible biotypes [54]. These findings suggest that resistant genotypes benefit from a faster upregulation of their stress response when they experience herbicide exposure [54,55]. Since herbicide treatments trigger a systemic stress response, it has been argued that sub-lethal doses of herbicides will select genotypes with a constitutive upregulation of stress response-associated genes [55]. Yet, since many of our genotypes have likely never experienced direct selection through sub-lethal doses of diquat, this concept is only partially suitable for explaining the observed variation in diquat resistance. Alternatively, non-targeted-site resistance to herbicides can also be induced by heterogenous environments [56], suggesting that the observed resistance to diquat in *S. polyrhiza* also might have evolved from a simultaneous adaption to multiple environmental stress factors [57].

Taken together, this study paved the foundation for understanding the evolution of diquat resistance in *S. polyrhiza*. Further studies validating the function and mechanisms of the identified candidate genes will provide further insights into the evolution of herbicide resistance in plants and help develop more sustainable weed management strategies.

4. Materials and Methods

4.1. Plant Material and Cultivation Procedure

For pre-cultivation, we propagated plant material of *S. polyrhiza* in Erlenmeyer flasks for 14 days before each assay. For all assays and pre-cultivations we grew our *S. polyrhiza* genotypes in N-medium [58] at 26 °C, with 135 µmol photons·m^{-2}·s^{-1} of light per day, a light/dark rhythm of 16h/8h and 50% humidity (GroBanks, model BB-XXL.3+cLED, CLF Plant Climatics, Wertingen, Germany) if not stated otherwise. Before each assay, we sterilized all of the analyzed genotypes via a combined protocol involving Klorix and Cefotaxime [59]. For all of the conducted assays, we grew the plants in plastic beakers (Verpackungsbecher PP, transparent, round, 250 mL, Plastikbecher.de GmbH, Giengen, Germany) filled with 150 mL N-medium (control conditions). We covered the cultures using perforated lids to allow air exchange. For the herbicide treatments, we supplemented the N-medium with the indicated concentration of diquat (Diquat-dibromide monohydrate, CAS:6385-62-2, Supelco, St. Louis, MO, USA). For all diquat measurements, the herbicide concentration is always expressed as the concentration of diquat-dibromide. Diquat was always applied as an aliquot from aqueous stocks. At the starting point of each experiment, we added fronds as colonies to the medium. If not stated otherwise, we triplicated each condition and genotype in all experiments.

4.2. Quantification of Diquat Toxicity

We expressed the diquat resistance of each genotype based on the changes in the fitness parameter relative growth rate (RGR) of frond number and frond area as well as

the inhibition of fresh weight (FW) relative to an untreated control. To estimate the first two parameters, we counted the frond number and determined the surface area overgrown with fronds for each culture at the beginning and end of each assay. To this end, we took a picture of each culture at the beginning and end of each toxicity assay. To quantify the frond area, we added a reference stone with a black square of an edge size of 1 cm to each culture. At the endpoint of the cultivation, we harvested the plant material, dried it briefly with tissue paper, and stored it in pre-scaled reaction tubes to determine the FW. The procedure mentioned above applies to all toxicity assays if not stated otherwise. Based on these data, we calculated the RGR using a published method [31]:

$RGR = \frac{lnNE - lnN0}{t}$ where NE is the frond number/area at the endpoint of the toxicity test and $N0$ is the frond number/area at the starting point of the toxicity test.

We expressed the inhibition of the herbicide on the growth of the genotypes as an inhibitory effect (IE) defined by the following formula for all fitness parameters mentioned above:

$IE = \frac{M-X}{M} \times 100$ where M is the mean of an untreated control and X is the mean of the treatment.

4.2.1. Determination of Screening Concentration

First, we aimed to identify a suitable diquat concentration that affects the growth in most of the genotypes. To make a rough estimate, we applied a diquat concentration range of 2 nM to 500 nM to two randomly selected genotypes, SP035 (registered four-digit code: 0109) and SP077 (registered four-digit code: 9508). For each replicate, we inoculated ten fronds at the start of cultivation and harvested the plant material after 10 days of cultivation. For each concentration, we quantified the growth inhibition as the IE of the FW and the RGR of the frond number. To estimate a screening concentration, we calculated the EC10 and EC90 values for both parameters and genotypes using the EC function from the "drc" R-package [60,61]. Based on the information obtained from this experiment, we further narrowed down the range of potential screening concentrations by testing the inhibition of 19 genotypes to 5 nM and 10 nM of diquat. Our criterion for choosing a suitable screening concentration was that no hormetic effects were visible for any of the 19 genotypes. To avoid the overgrowth of cultures, we decreased the cultivation time to 7 days and inoculated only six fronds at the starting point of the experiment. We quantified diquat resistance based on the IE of the RGR frond number (IE of RGR frond number), the IE of the RGR frond area (IE of RGR frond area) and the IE of the fresh weight (IE of FW). We further estimated the broad-sense heritability of diquat resistance at 5 and 10 nM according to a published method [62] using the H2cal function of the inti R-package [63].

4.2.2. Diquat Resistance Screening

We assayed 138 genotypes in five batches, each encompassing 23 to 35 genotypes, regarding their growth inhibition at a diquat concentration of 10 nM for 7 days. A list of all screened genotypes is provided in the Supplementary Materials (Table S1). At the beginning of the cultivation, we inoculated six fronds in each culture. To control for batch effects, we cultivated genotype SP028, which continuously served as an internal standard throughout all batches. We then normalized the inhibition of all genotypes to that of SP028 in the respective batch. We used these normalized the values of the IE of RGR the frond number (norm. IE RGR of frond number) and area (norm. IE of RGR frond area) and the IE of the FW (norm. IE of FW) to identify the most resistant and susceptible genotypes.

4.2.3. Dose–Response Measurements

To quantify the variation in diquat resistance levels in our 138-genotype accession, we estimated the half maximal effective concentration (EC50) values for the eight most resistant and the eight most susceptible genotypes identified in our resistance screening. We applied a concentration range from 0.5 nM to 100 nM of diquat to a population of six fronds at the starting point of the assay. The duration of the toxicity test was 8 days.

To correct for the occurrence of hormetic effects, we calculated all dose–response curves using the drc R-package, applying the hormetic Cedergreen–Ritz–Streibig model UCRS.4c, which has previously been used in toxicity tests of herbicide formulations on duckweed species [64]. In contrast to normal logistic models, which are strictly monotonous functions, the UCRS4.c model takes stimulatory effects, that are caused by hormesis into account [64]. Therefore, the UCRS.4c model estimates the potency of the herbicide via a lower bound on the ED50 value [64]. Since the hormetic effects were minimal in all cases, the lower bound on ED50 provides a close approximation of the EC50. Therefore, we refer to the lower bound on ED50 as an estimate of EC50 in this manuscript.

4.3. Chlorophyll Fluorescence Measurement

We evaluated diquat resistance on a physiological level, measuring the decay of the chlorophyll fluorescence parameters Fv/Fm and Fv'/Fm' in response to high levels of diquat. The Fv/Fm ratio indicates the decrease in the quantum yield of linear electron flow through PSII [65] and the Fv'/Fm' ratio is an estimate of the maximum quantum yield of PSII in a light-adapted state [66]. We quantified Fv/Fm and Fv'/Fm' for the previously identified four most resistant (R) and four most susceptible genotypes (S). For this, we acclimated all of the plant cultures at 26 °C at 110 µmol photons·m^{-2}·s^{-1} of light per day and a light/dark rhythm of 16 h/8 h (Percival, model AR-41L3, CLF Plant Climatics) for two days before measurement. We used 12-well plates (Nunc, Roskilde, Denmark) filled with 6 mL of N-medium containing 40 µM of diquat for herbicide exposure of the plant material. Each well was inoculated with a duckweed colony composed of three to four fronds. Each genotype was analyzed with three biological replicates for each time point. For each replicate, three areas of interest were defined. We chose durations of 0 min, 30 min, 1 h, 2 h, 4 h, 8 h, 12 h, and 24 h as endpoints for the incubation. We conducted fluorescence measurements using an I-PAM system (M-series, MAXI version, Heinz Walz GmbH, Effeltrich, Germany), applying all settings published previously for chlorophyll fluorescence measurements in *Lemna minor* L. (Arales: Lemnaceae) [67]. For dark adaption, we incubated our samples in the dark for ten min before the saturation pulse. After recording Fv/Fm, we adapted the plants to actinic light at 84 PAR (photosynthetically active radiation) for 10 min, before measuring the Fv'/Fm' ratio.

4.4. Superoxide Dismutase Activity

We measured SOD activity for the four most susceptible (S) and four most resistant (R) genotypes identified in our dose–response experiment. To determine the superoxide dismutase (SOD) activity from the plant material, we cultivated five colonies of each genotype under 10 nM of diquat or under control conditions for 24 h. Next, we suspended 10 mg (FW) of plant material from each sample in 500 µL of 0.01 M MOPS buffer (Carl Roth, Karlsruhe, Germany) at pH 7.2. We homogenized the suspension by pulsed mixing for 5 s and centrifuged each sample at 14,000× *g* for 1 min before analysis. For the quantification of the SOD activity, we diluted 150 µL of supernatant 1:4 in MOPS buffer pH 7.2. We measured SOD activity following the instructions of the SOD determination kit (Product 19160, SOD Assay Kit, Sigma-Aldrich, St. Louis, MO, USA). To determine the protein concentration, we diluted the supernatant 1:2 in MOPS buffer at pH 7.2 and analyzed the samples according to the manufacturer's instructions using the Rapid Gold BCA Protein Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). We standardized SOD activity to the measured protein content. To check for the involvement of SOD activity in diquat resistance, we further correlated the SOD activity with our estimates of EC50 values.

4.5. Ascorbate and Glutathione Measurements

To measure the impact of diquat exposure on shifts in the concentration ratios of ascorbate (AsA) to dehydroascorbate (DHA) and reduced (GSH) to oxidized glutathione (GSSG), we cultivated 30 fronds from each R and S genotype for 4, 8 and 24 h under 10 nM of diquat. We grew the control samples for 24 h in N-medium without diquat. For each

genotype and measurement point, we made five biological replicates. At the endpoint of the assay, we harvested all of the plants, dried them with tissue paper, transferred them into pre-scaled reaction tubes, and froze them immediately in liquid nitrogen. We stored all of the samples at −80 °C until analysis. We extracted antioxidants from samples of 20 mg FW each with 250 µL of 5% (*w*/*v*) metaphosphoric acid (Carl Roth). The extracts were diluted 1:10 with 5% (*w*/*v*) metaphosphoric acid containing 1.11 µg/mL of isotope-labeled GSH (Glutathione-Glycine-$^{13}C_2$,^{15}N, Sigma-Aldrich) (labGSH). The further steps of the extraction protocol are based on a method published previously with the modifications listed below [68].

We analyzed all of the antioxidants via LC-MS. For separation, we used a Nexera X3 UHPLC system (Shimadzu, Kyoto, Japan) equipped with a Nucleodur Sphinx RP column (250 × 4.6 mm, 5 µm, Macherey-Nagel, Düren, Germany) and a Nucleodur Sphinx RP EC 4/3 guard column (5 µm, Macherey-Nagel) maintained at 25 °C. For the mobile phase, we used an aqueous solution of 0.2% (*v*/*v*) formic acid, 0.05% (*v*/*v*) acetonitrile as buffer A, and acetonitrile as buffer B. For elution, we applied a gradient mode using the following program: 0.0 min/2% B, 3.5 min/2% B, 5.0 min/100% B, 8.0 min/100% B, 8.2 min/2% B, 12.5 min/2% B at a constant rate of 1 mL/min. The injection volume was 1 µL of extract per sample. For quantification we used an LC-MS/MS-8060 (Shimadzu) with an electrospray ionization source with the following parameters: nebulizing gas flow: 3 L/min, heating gas flow: 10 L/min, drying gas flow: 10 L/min, interface temperature: 350 °C, DL temperature: 250 °C, heat block temperature: 400 °C, CID gas flow: 270 kPa, interface voltage: 4000 V. GSH and labGSH were quantified in the positive ionization mode, all other metabolites were quantified in the negative ionization mode. We operated with a mass spectrometer in multiple reaction monitoring (MRM) mode using the following settings: Ch1: 308.09 → 76.1 *m*/*z* (CE −26 V), Ch2: 308.09 → 162.1 *m*/*z* (CE −16 V) for GSH, Ch1: 311.09 → 181.95 *m*/*z* (CE −11 V), Ch2: 311.09 → 165.2 *m*/*z* (CE −14 V) for labGSH, Ch1: 173.01 → 113.20 *m*/*z* (CE 10 V), Ch2: 173.01 → 143.05 *m*/*z* (CE 12 V) for DHA, Ch1: 175.02 → 86.85 *m*/*z* (CE 22 V), Ch2: 175.02 → 115.05 *m*/*z* (CE 13 V) for AsA and Ch1: 611.14 → 305.85 *m*/*z* (CE 24 V), Ch2: 611.14 → 271.90 *m*/*z* (CE 28 V) for GSSG. For all of the metabolites, Ch1 was used as the quantifier and channel Ch2 as the qualifier. We facilitated the quantification of GSH based on the internal standard labGSH. In the case of GSSG, AsA, and DHA, we quantified them using the internal standard and a conversion factor (CF) that was determined based on external standard curves ($n_{Analyte} = n_{labGSH} \times CF$). The empirically determined CFs were 13.1 for AsA, 652.6 for DHA and 2.2 for GSSG.

4.6. Diquat Measurement

To check the possibility of an involvement of transport-related mechanisms in diquat resistance, we quantified the diquat concentration in plant tissue for all R and S genotypes. To relate diquat resistance to tissue concentration, we extracted diquat from freeze-dried samples of 138 genotypes. We pooled all triplicates of the same genotype treated with diquat. Due to a sample being lost during diquat extraction, the total number of samples was reduced to 137. The extraction procedure follows a previously described method with minor modifications [69]. We homogenized the freeze-dried plant material (DW) in a tissue lyser (Qiagen, Venlo, The Netherlands) by adding two to three metal beads and shaking for 1 min at 25 Hz. We aliquoted 10 mg of the ground plant material in 96-well biotubes (catalog number: MTS-11-C, MTS-11-C-R, Axygen, New York, NY, USA) and resuspended each sample with 800 µL of 40% (*v*/*v*) methanol, that was acidified with 0.0375 N HCl and contained 20 ng [$^2H_4$]-diquat dibromide (03627-5 mg, Supelco) as an internal standard. We homogenized the samples in a tissue lyser (Qiagen) at 20 Hz for 10 min, followed by an incubation step at 80 °C for 15 min and a second homogenization step at 20 Hz for 10 min. After two centrifugation steps at 2000× *g* for 20 min at 4 °C, we transferred the supernatant to a PCR plate for analysis. Due to the adhesion of diquat to glass surfaces, we did not use glass tubes throughout the whole extraction process. Before the measurement, we performed another centrifugation step at 2000× *g* for 20 min at 4 °C.

Since tissue dilution effects caused by the higher propagation rates of resistant compared to susceptible genotypes in the diquat treatments might have confounded the correlation between diquat resistance and tissue concentration, we established an uptake kinetic for the R and S genotypes. We cultivated the genotypes in N-medium supplemented with 10 nM of diquat for 15 min, 30 min, 1 h, 2 h, 4 h, 8 h, 12 h, and 24 h. We triplicated each treatment, with each replicate containing 15 fronds growing in colonies. We extracted 2 mg of homogenized plant material (DW) applying the workflow described above but using only 200 µL of extraction buffer containing 1 ng of the [$^2H_4$]-diquat dibromide as the internal standard per sample.

We measured the diquat concentration in the root and frond tissue separately to look for tissue-specific translocation patterns. We cultivated the genotypes at 10 nM of diquat for 24 h. Each replicate consisted of 30 fronds growing as colonies. We dissected samples into the root and frond tissue for separate analysis. The extraction of each 2 mg of tissue per sample was performed as described in the previous paragraph.

To monitor changes in the diquat concentration of the medium, we cultivated R and S genotypes for 7 days at 10 nM of diquat. Each day, a medium sample of 400 µL was taken from each replicate and immediately frozen in liquid nitrogen. After lyophilization, we resuspended the precipitate in 400 µL extraction buffer containing 4 ng of [$^2H_4$]-diquat dibromide. We then homogenized 300 µL of the supernatant in a tissue lyser as mentioned above and continued with the extraction as described for plant tissue samples.

We analyzed the diquat via LC-MS. For separation, we used a Nexera X3 UHPLC system (Shimadzu) coupled to an LC-MS/MS-8060 with an Acclaim Trinity Q1 column (100 × 2.1 mm, 3 µm, Thermo Fisher Scientific) and a Trinity Q1 guard column (10 × 2.1 mm, 5 µm, Thermo Fisher Scientific) maintained at 30 °C. The injection volume of the sample was 0.5 µL for all measurements. For the mobile phase, we used a buffer containing 75% (v/v) acetonitrile and 25 mM ammonium acetate with a pH adjusted to 5.0 with acetic acid in an isocratic mode at a flow rate of 0.4 mL/min. For quantification, we used an LCMS-8060 (Shimadzu) with an electrospray ionization source with the following parameters: nebulizing gas flow: 3 L/min, heating gas flow: 10 L/min, drying gas flow: 10 L/min, interface temperature: 300 °C, DL temperature: 250 °C, heat block temperature: 400 °C, CID gas flow: 270 kPa, interface voltage: 1000 V. The mass spectrometer operated in multiple reaction monitoring (MRM) mode using the following settings: Ch1: 183.09 → 157.00 m/z (CE −20 V), Ch2: 183.09 → 130.10 m/z (CE −31 V) for diquat and Ch1: 186.11 → 158.00 m/z (CE −20 V), Ch2: 186.11 → 131.10 m/z (CE −33 V) for [$^2H_4$]-diquat. For this measurement channel, Ch1 was the quantifier and Ch2 the qualifier. To check for potential associations between the uptake of diquat or the diquat resistance and the antioxidant levels, we correlated the diquat tissue concentration after 24 h and the resistance levels with the GSH/GSSG and AsA/DHA levels.

4.7. GWAS

To reduce the false positive signals due to clonality, we selected all sequenced genotypes from our 138 genotype accession and grouped them into 98 genets according to a recently published method [70]. From each genet the genotype with the highest sequencing coverage was selected as representative genotype. All of the analyzed genotypes are listed in the Table S1. We used the normalized IE of FW, frond number and frond area as input parameters for the GWAS. We conducted all GWAS on the 98 representative genotypes using an univariate linear mixed model [71] implemented in the vcf2gwas package [72]. We used structural variants (SVs) (>50 nucleotides) and SNPs as input for our GWAS analyses. All genotypic data for SNPs and SVs, as well as the annotation, were part of a recent publication [37] and can be found under https://github.com/Xu-lab-Evolution/Great_duckweed_popg (accessed on 9 January 2024).

We formatted the genotype SV dataset as the input for vcf2gwas according to the published protocol [73]. Before analysis, we pruned our SNP and SV datasets using the PLINK package, which is integrated in the vcf2gwas platform, applying a window size

of 100 markers, a step size of 10 markers, and allowing a phased r^2-threshold of 0.33. We removed all SNPs and SVs with a minor allele frequency (MAF) of less than 5%, resulting in 842 SVs and 42,462 SNPs.

To estimate differences in diquat resistance on a population level, we sorted all genotypes into four genetic populations based on previously published information [37].

4.8. RT-PCR Analysis of Splicing Pattern in SpETHE1

To validate the presence of the identified mutations in the most resistant (SP050) and the most susceptible genotype (SP011), we isolated genomic DNA from these *S. polyrhiza* genotypes using the DNeasy Plant mini kit (Qiagen) and genotyped them using primers 1_ETHE1_fwd and 2_ETHE1_rev (Table A6). The presence of the 94 bp deletion would yield a 705 bp product, whereas a wild-type sample would give a 799 bp product. The PCR program included an initial denaturation step at 95 °C for 3 min followed by 31 cycles of 95 °C—15 s, 60 °C—15 s, 72 °C—70 s with a single final elongation step at 72 °C for 5 min. We visualized the products on a 1% agarose gel.

To check whether *SpETHE1* (SpGA2022_053619) shows different splicing patterns between SP050 and SP011, we cultivated 30 fronds of each genotype for 4, 8, and 24 h at a 10 nM diquat concentration. For each genotype, 30 fronds each were cultivated for 24 h in a medium without diquat were used as control. For all RNA sample preparations, we extracted <20 mg of FW aliquots using the InnuPREP RNA mini kit (Analytik Jena, Jena, Germany) and checked RNA integrity on a 1% agarose gel afterwards. We performed all cDNA synthesis reactions following the RevertAid First Strand cDNA synthesis kit (Takara, Shiga, Japan) using random hexameric primers to ensure cDNA synthesis from low-stability RNAs lacking polyA-tails. Next, we amplified fragments from the *SpETHE1* cDNA samples with the primers 3_ETHE1_fwd and 2_ETHE1_rev (Table A6), that bind on the fourth and nineth exon. The primers are flanking the intronic sequence carrying the deletion. Correct splicing is indicated by a fragment size of 330 bp. We carried out an amplification from genomic DNA using the same primer pair as the control, expecting a product of 1848 bp. The time program for amplification from cDNA template included an initial denaturation step at 95 °C for 3 min followed by 30 cycles of 95 °C for 15 s, 60 °C for 15 s, 72 °C for 30 s with a single final elongation step at 72 °C for 5 min. For amplification from genomic DNA, the duration of the extension step at 72 °C was increased to 150 s. We separated the products on a 2% agarose gel. For all amplifications, we used the PrimeSTAR HS DNA polymerase (Takara).

4.9. RT- qPCR for Monitoring Gene Expression in Duckweed Tissue

We measured the expression differences in *SpETHE1* and genes in close distance to *SpETHE1* in SP050 and SP011. To analyze the diquat-induced differences in expression levels of *SpETHE1* in root and frond tissue separately, we cultivated 40 fronds from each genotype for 4 and 72 h at 10 nM of diquat. As a control, we cultivate 40 fronds from each genotype at N-medium without diquat for 24 h. We made five biological replicates for each genotype and treatment. Since *ETHE1* showed a potential circadian and diurnal regulation pattern [74,75], we always harvested the treated samples and controls at the same time of day. Root and frond tissue were collected from each culture separately. The root and frond tissues of the control cultures were also used to determine the constitutive expression levels of the flanking genes *SpDLH* (SpGA2022_012161), *SpTRB1* (SpGA2022_012163), *SpALA2* (SpGA2022_053617), and *SpSBE3* (SpGA2022_053621). We conducted all cDNA syntheses as described in the previous section but with oligo-dT primers instead of random hexameric primers. For all qPCR experiments, we quantified *SpETHE1* according to a previously published method [76] using the glycerin-aldehyde-3-phosphate dehydrogenase (*SpGAPDH*, SpGA2022_054082) and alpha elongation factor one (*SpaEF*, SpGA2022_005771) as reference genes. The primer sequences of the two reference genes were published previously [37].

We calculated the primer efficiencies based on a dilution series of cDNA templates (Table A7) and evaluated their specificity by loading the reaction mixtures on a 2% agarose gel. Before starting the RT-qPCR, we diluted all cDNA samples 1:100 in water. We carried out the RT-qPCR using a RotorGene Q system (Qiagen), using a master mix from the KAPPA SYBR FAST kit (Roche, Basel, Switzerland). For each gene and sample, we made three technical replicates. The time program for qPCR included an initial denaturation at 98 °C for 3 min followed by 40 cycles of 98 °C for 3 s of denaturation and 60 °C for 20 s of annealing/extension.

4.10. Software and Statistics

We performed all statistical analyses using R version 4.2.0. We analyzed all of the pictures used to quantify the frond areas and numbers using ImageJ (version 1.53g) running on the Fiji platform. We accomplished the measurement of chlorophyll fluorescence using ImagingWinGigE version 2.47. For all LCMS measurements, we recorded the signals using LabSolutions version 5.97. We integrated peaks using Lab Solutions Insight version 3.5 SP2. We used the FSA package of R [77] to calculate the standard errors. For the statistical analyses of differences in chlorophyll fluorescence parameters, we applied the lme4 package [78] and the multcomp R packages [79] to implement a linear mixed effect model and a Tukeys post hoc test for multiple comparisons. For all other experiments involving the most resistant and most susceptible genotype, we compared mean values using the Student's *t*-test. We used the *F*-test to examine the significance of our regressions. We used one-way ANOVA followed by a Tukey's HSD post hoc test for comparisons of multiple mean values. If the criterion of homogeneity of variances was violated, we used the Games Howell test to compare multiple mean values. To conduct our GWAS, we used version 0.8.7. of the vcf2gwas platform. To determine significant genetic markers with a Bonferroni-corrected $p < 0.05$, we used a Wald test. We converted genotype files into vcf format, which is the only accepted format in vcf2gwas, using the TASSEL platform version 5.0 [80].

5. Conclusions

The findings of this study provide insights into the genetic and physiological principles of non-targeted-site resistance to diquat in *S. polyrhiza*. Since the analyzed genotypes showed gradual variation in their diquat resistance that could be associated with herbicide uptake kinetics and antioxidant responses, we suggest a polygenic mechanism behind the observed phenotypic variation in resistance levels. The association of gene candidates involved in apoptosis, pathogen response, starch metabolism and xenobiotic degradation points to a pleiotropic origin of diquat resistance from stress response pathways. Since most of the analyzed genotypes have likely never experienced selection through diquat, we suggest that the detected intra-specific variation in diquat resistance might originate from an adaption to environmental stress factors such as drought, UV stress, and herbivory.

Taken together our results contribute to a better understanding of the rapid evolution of novel resistance cases around the world. Association studies on the genetic basis of herbicide resistance are required to identify novel genetic signatures which can be used for the marker-associated breeding of herbicide resistance crop species or for the development of molecular screening methods for resistant weed species.

Supplementary Materials: The following supporting information can be downloaded at: https://doi.org/10.5061/dryad.2fqz612ww. Table S1: List of genotypes screened for their diquat resistance.

Author Contributions: Conceptualization, M.H. and S.X.; methodology, M.H., S.W., and M.S.; software, Y.W. and M.H.; validation, M.H.; formal analysis, M.H.; investigation, M.H. and S.W. and Y.W.; resources, S.X.; data curation, S.X.; writing—original draft preparation, M.H.; writing—review and editing, S.X.; visualization, M.H.; supervision, S.X.; project administration, S.X.; funding acquisition, S.X. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the German Research Foundation (DFG), grant number 427577435 to S.X. The LC-MS instrument was funded by the German Research Foundation (DFG), grant number 435681637 to S.X.

Data Availability Statement: Publicly available datasets including Supplementary Table S1 were analyzed in this study. These data can be found here: https://doi.org/10.5061/dryad.2fqz612ww (accessed on 3 February 2024).

Acknowledgments: We thank Marie Sarazova for providing technical assistance and Alexander Kröger for his support in the toxicity assays.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Appendix A

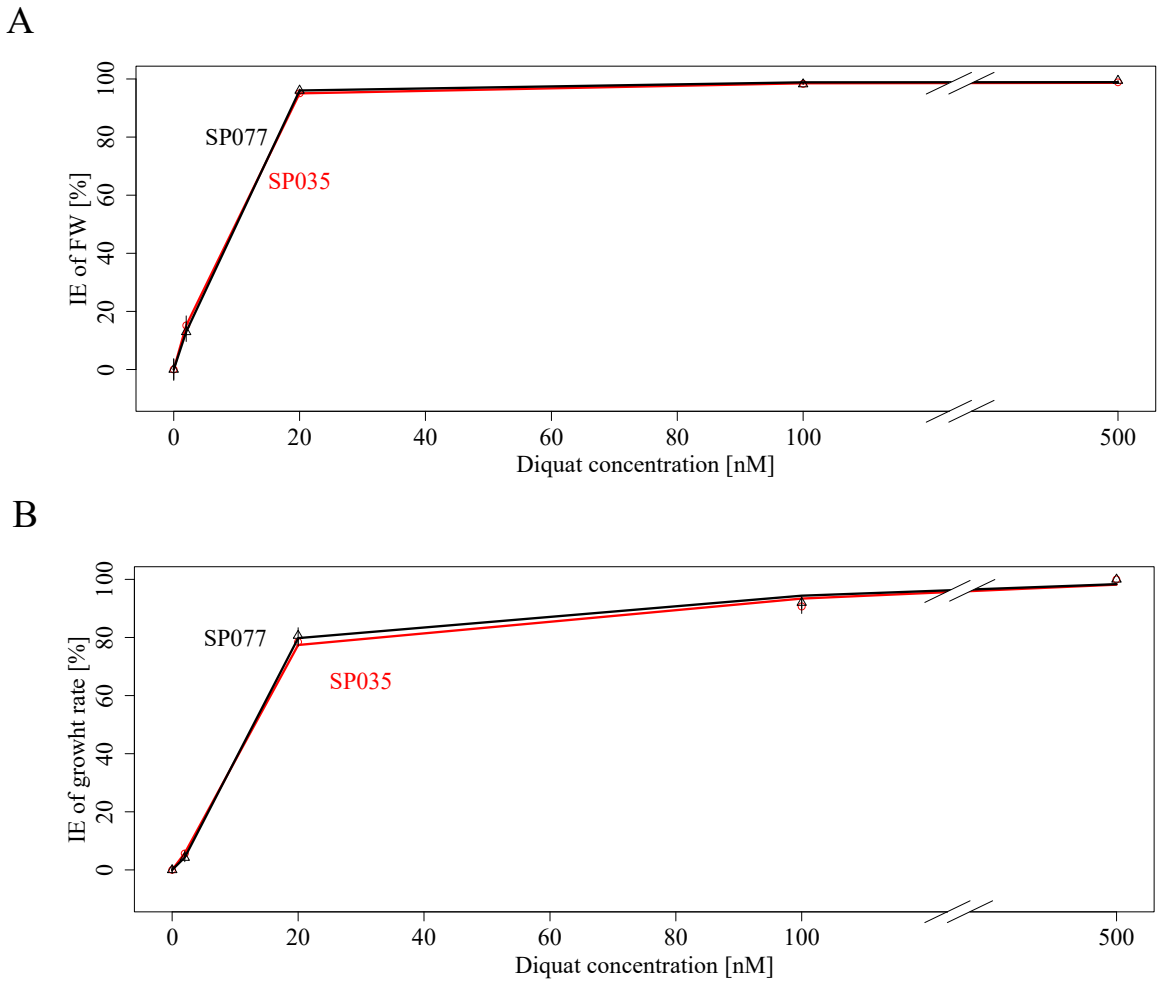


Figure A1. Dose–response curve of two *S. polyrhiza* genotypes (SP077 and SP035) for various diquat concentrations. Mean IE values for FW (**A**) and RGR of frond number (**B**) are shown with their respective standard errors. The curves were fitted using the hormetic UCRS4c.

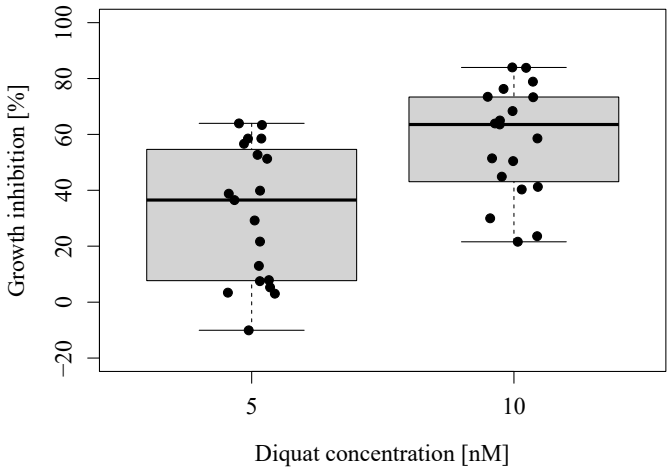


Figure A2. Distribution of IE of RGR of frond number values, here shown as growth inhibition in % from 19 genotypes cultivated at diquat concentrations of 5 nM and 10 nM. We observed hormetic effects for one genotype in the 5 nM treatment.

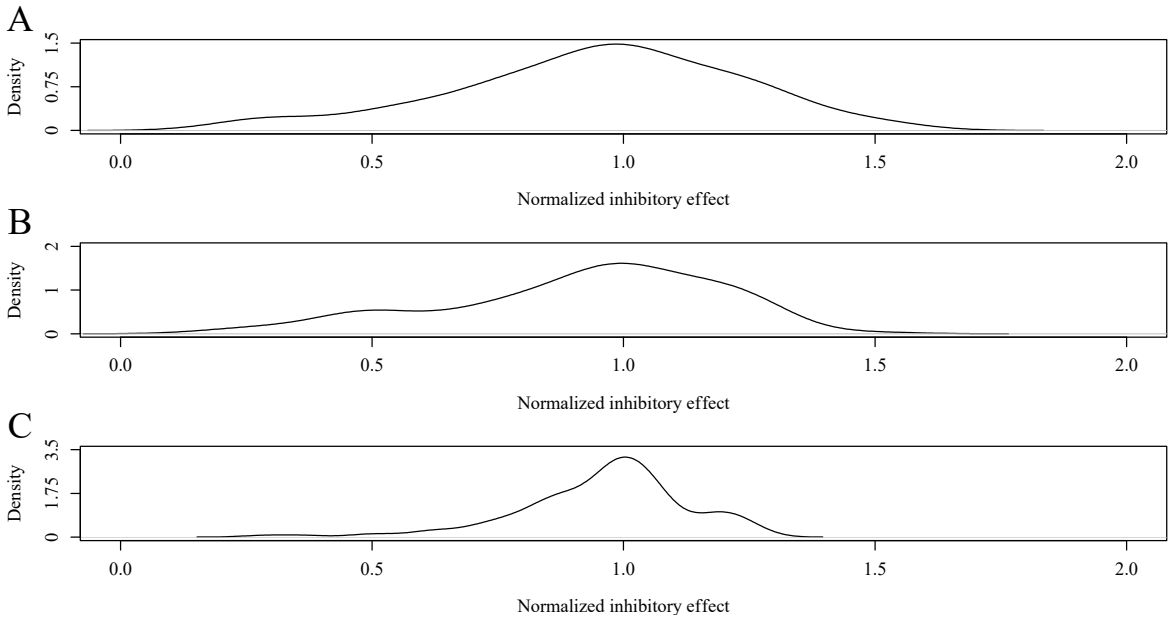


Figure A3. Density function of all diquat resistance cases shown by their normalized growth inhibition measured for the parameter's norm. IE of RGR frond number (**A**), norm. IE of RGR frond area (**B**), and the norm. IE of FW (**C**). The growth inhibition was normalized to the internal standard genotype SP028, whose diquat resistance was set at 1.0.

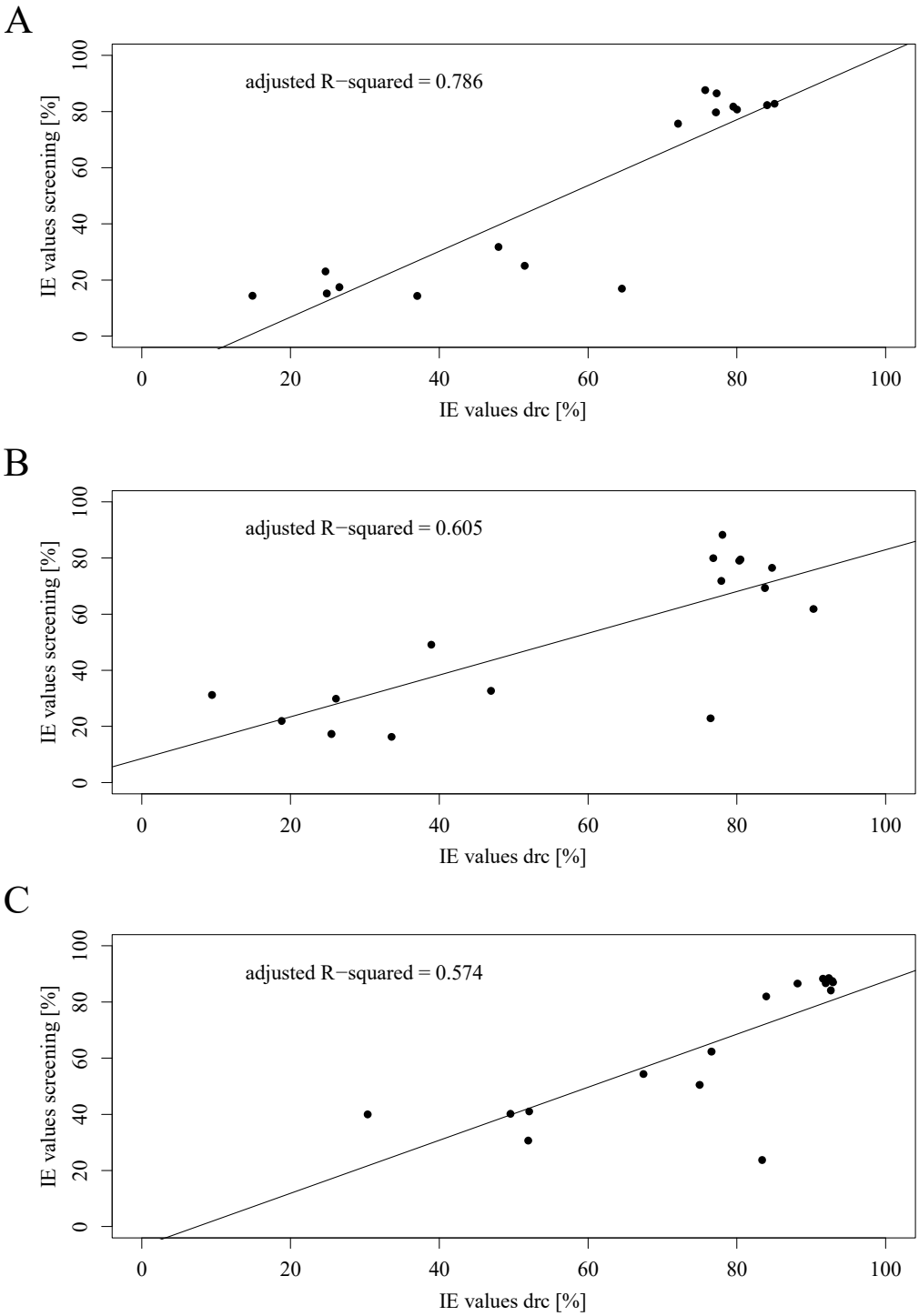


Figure A4. Correlation of inhibitory effect of diquat on several fitness parameters from the screening experiment with that of the dose–response curve experiment (drc) at a concentration of 10 nM: (**A**) RGR frond number, (**B**) RGR frond area, (**C**) FW.

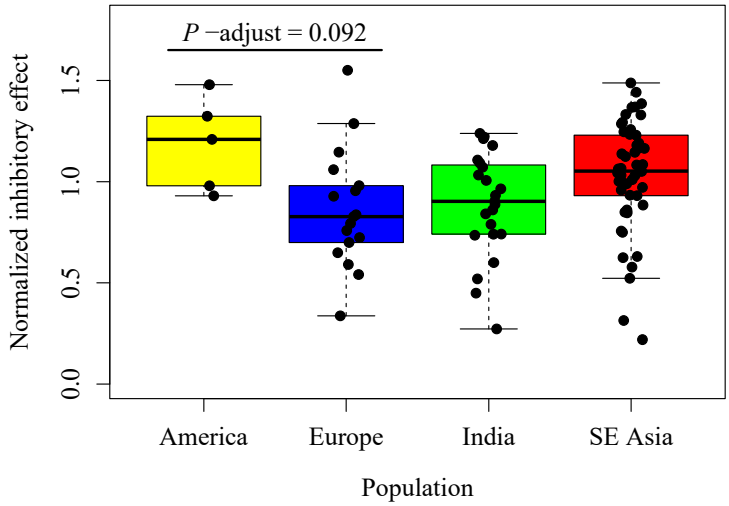


Figure A5. Population-wise distribution of resistance levels shown as normalized IE of RGR frond number for the 98 representative genotypes used in the GWAS.

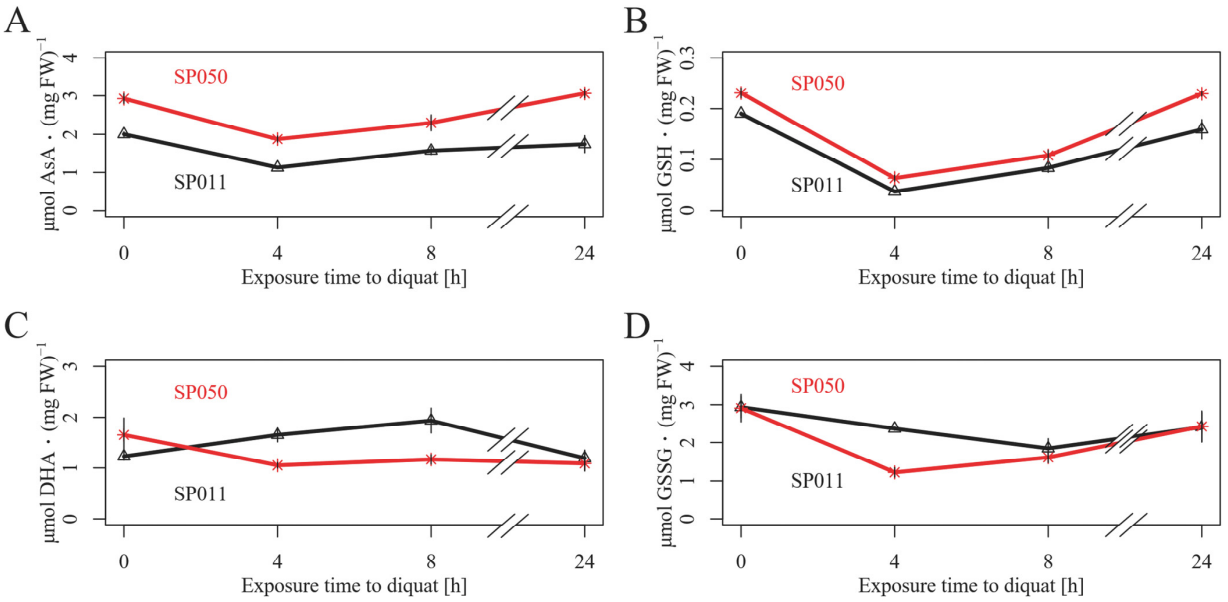


Figure A6. Ascorbic acid (AsA), dehydroascorbic acid (DHA), reduced (GSH), and oxidized glutathione (GSSG) plant tissue concentrations in response to diquat are shown with their respective standard errors for SP050 and SP011. Whereas the reduction of AsA concentration was compensated by an increase in DHA, both GSH and GSSG decreased after 4 h of diquat exposure. (**A**) AsA concentration, (**B**) GSH concentration, (**C**) DHA concentration, (**D**) GSSG concentration. Mean values and standard errors were derived from four or five biological replicates.

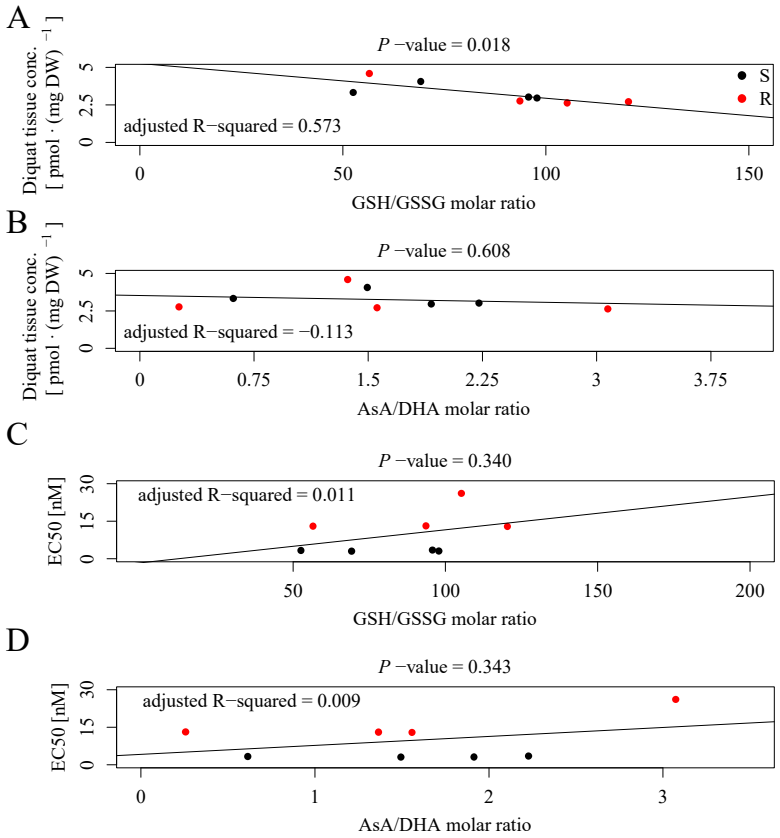


Figure A7. The diquat plant tissue concentration after 24 h of diquat treatment was correlated to the corresponding GSH/GSSG (**A**) and AsA/DHA (**B**) molar ratios for the four R and four S genotypes (Table A2). Only the GSH/GSSG molar ratio was significantly associated with the diquat tissue concentration (F-test). In (**C**,**D**), the estimates of EC50 were correlated to the GSH/GSSG (**C**) and ASA/DHA ratios (**D**).

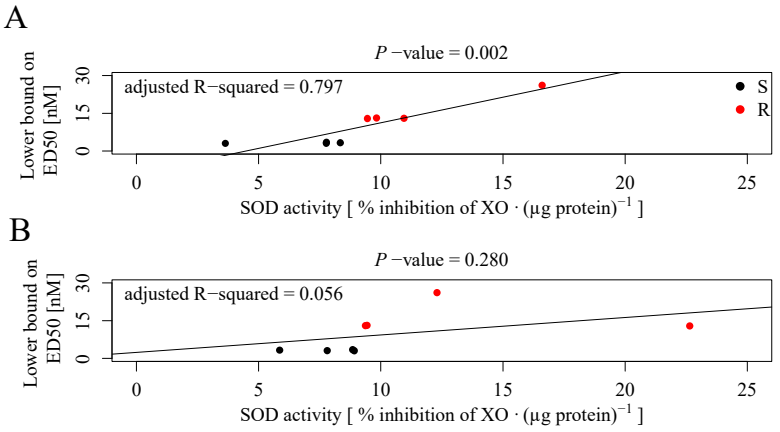


Figure A8. The resistance levels expressed as estimates of EC50 were correlated with the SOD activity measured under 24 h diquat treatment (**A**) or constitutive conditions (**B**). We used the four S and four R genotypes shown in Table A2 for this analysis. Diquat resistance correlated significantly with the SOD activity under diquat treatment (F-test). No such effect was found for the constitutive SOD activity.

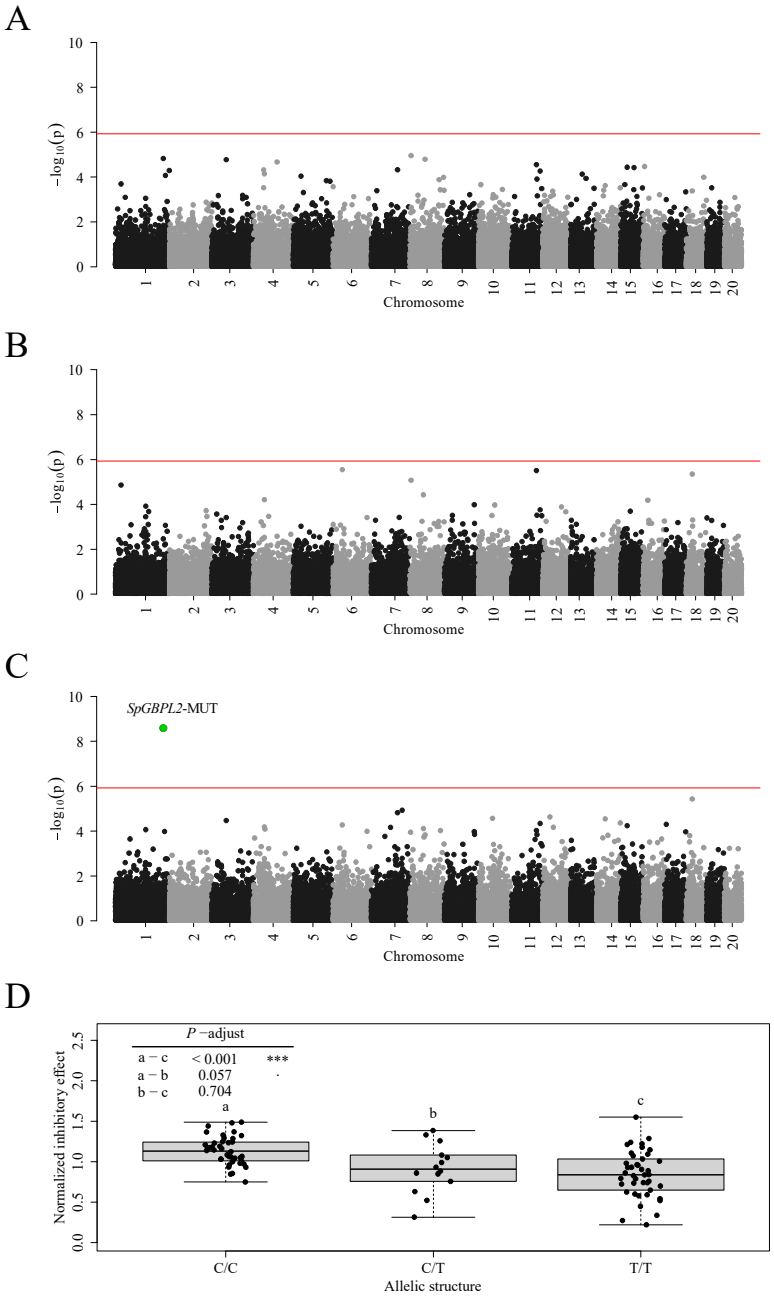


Figure A9. Manhattan plots for a GWAS of diquat resistance calculated based on frond number (**A**) frond area (**B**) and freshweight (**C**). A synonymous SNP, causing a C-T transition in the coding sequence of *SpGBPL2*, is associated with an increased resistance measured by norm. IE of FW (Wald test). The Bonferroni corrected significance threshold (red line) was at $p < 0.05$ is 1.18×10^{-6}. (**D**) Genotypes that are homozygous for the allele containing thymine instead of cytosine were significantly more resistant to diquat. ***—*p*-adjust < 0.001 (Games Howell test).

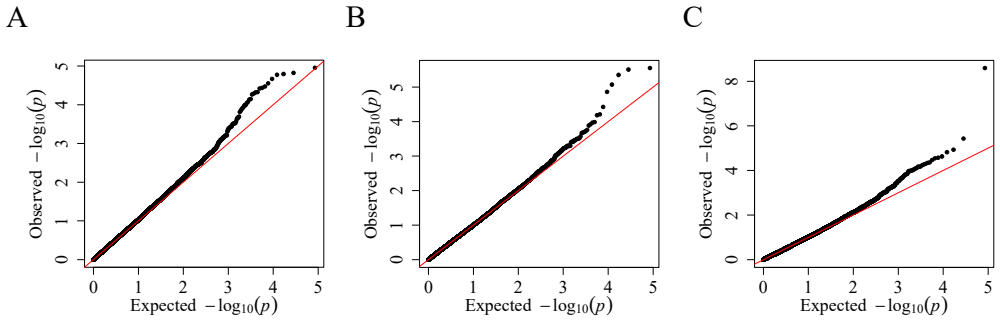


Figure A10. QQ plots for GWAS conducted on SNPs associated with diquat resistance traits. (**A**) norm. IE of RGR frond number (**B**) norm. IE of RGR frond area (**C**) norm. IE of FW.

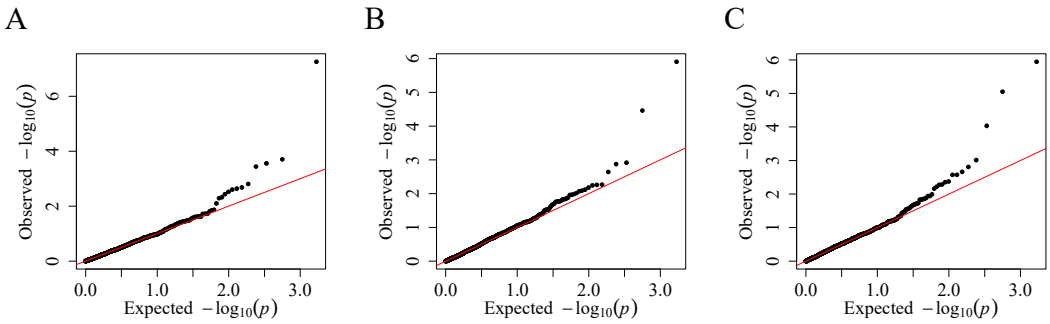


Figure A11. QQ plots from GWAS conducted for structure variations on traits expressing diquat resistance. (**A**) norm. IE of RGR frond number (**B**) norm. IE of RGR frond area (**C**) norm. IE of FW.

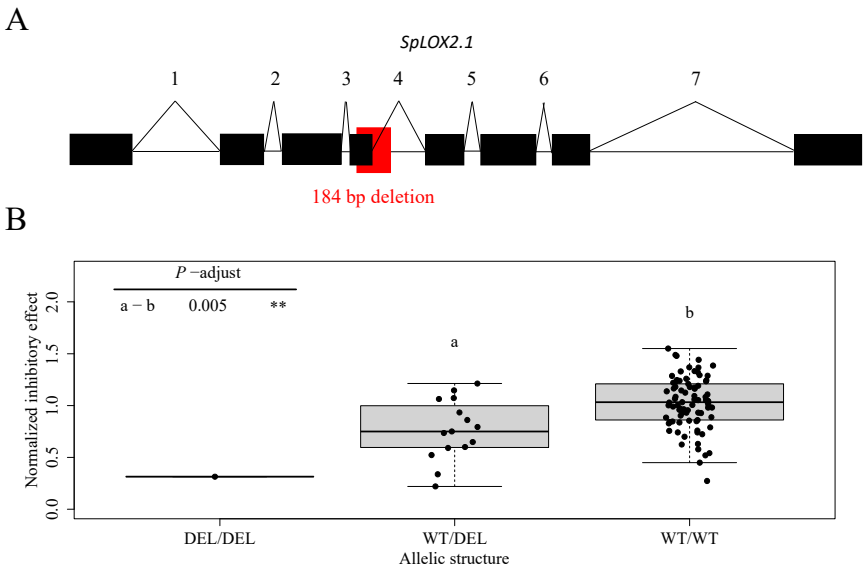


Figure A12. (**A**) Exon—intron structure of *SpLOX2.1*. The 184 bp deletion is at the border between the fourth exon and fourth intron. (**B**) Genotypes that are heterozygous for the deletion were significantly more resistant than genotypes lacking the deletion. **—0.01 ≥ *p*-adjust > 0.001 (unpaired two-sided Student's *t*-test). Since only one genotype was homozygous for the deletion, we did not include this sample in our statistical analysis.

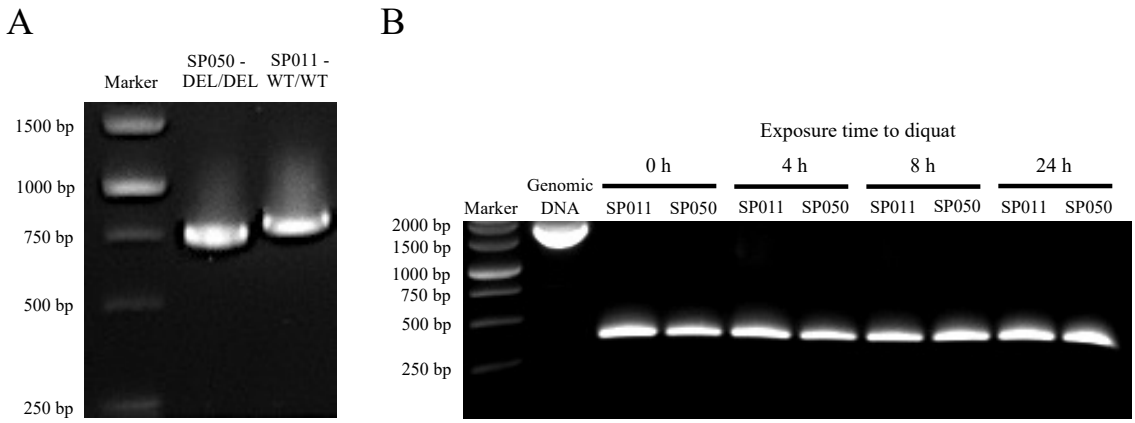


Figure A13. Genotyping of the deletion in *SpETHE1* and analysis of the gene splicing pattern in the *S. polyrhiza* genotypes SP050 and SP011. (**A**) Genomic DNA extracted from SP050 and SP011 was genotyped for the presence of the deletion in *SpETHE1*. The absence of the 94-bp deletion (WT/WT) results in a product of 799 bp length. As indicated by the shift of the amplicon band obtained for SP050, this genotype was homozygous for the deletion in *SpETHE1*. (**B**) A 330-bp fragment was amplified from cDNA pools with primers binding on exons flanking the intron with the 94 bp deletion for SP050 and SP011. The cDNA was synthesized with random hexameric primers to account for unstable RNA species. No products that deviated from the predicted 330-bp fragment were amplified from both genotypes. This indicates no differences in the splicing pattern of *SpETHE1* between the two genotypes.

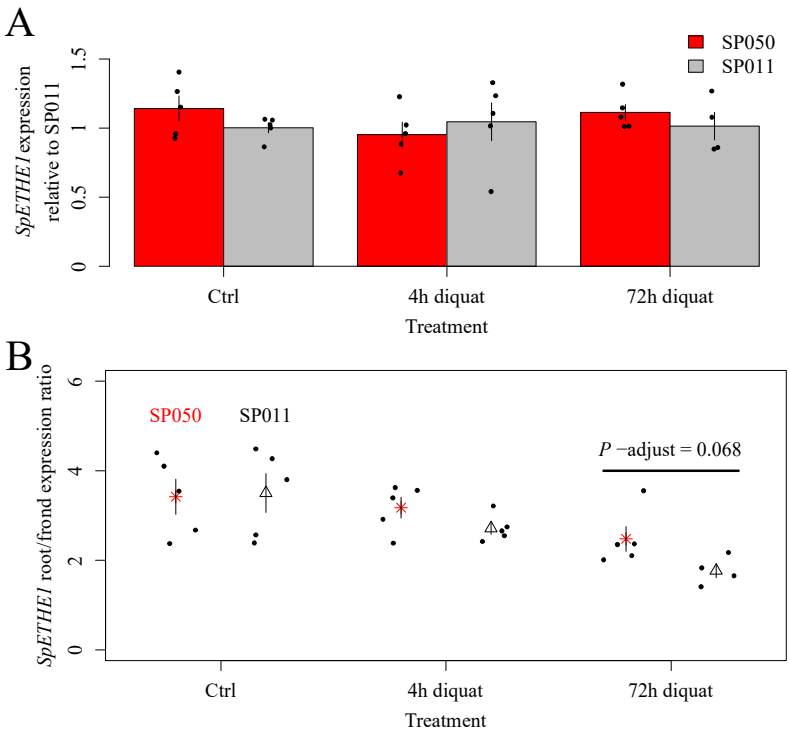


Figure A14. *Cont*.

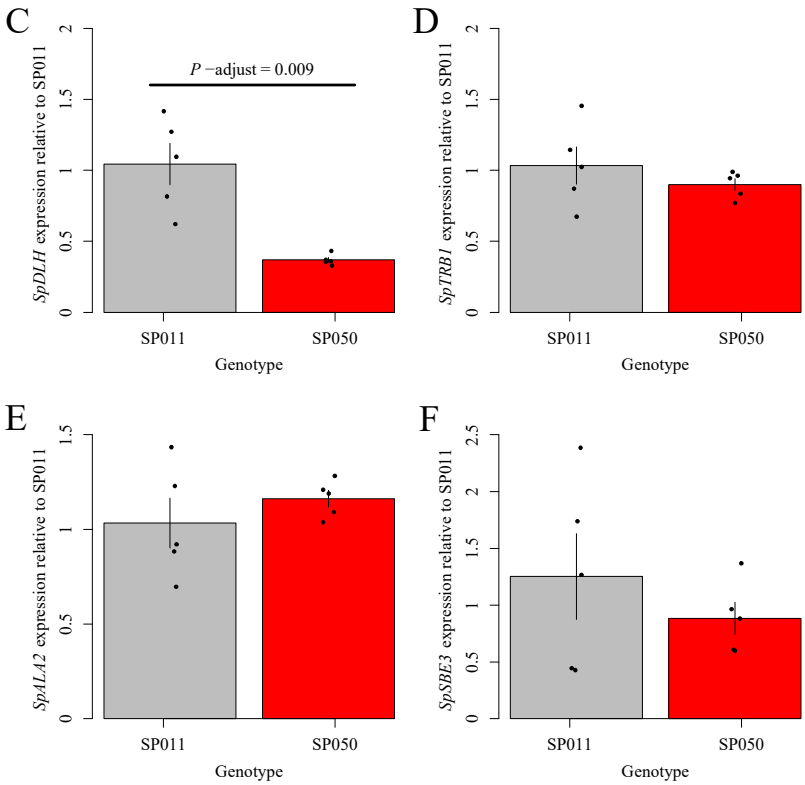


Figure A14. Relative expression of candidate genes in the roots. (**A**) *SpETHE1* time course expression under control conditions or after 4 and 72 h of diquat exposure. No significant differences in *SpETHE1* expression were found between SP011 and SP050. (**B**) *SpETHE1* root–frond expression ratio in SP050 and SP011 under control conditions and after diquat treatment for 4 and 72 h. *SpETHE1* expression was stronger in the root than in frond tissue by factors two to four in both genotypes. Expression differences in *SpDLH* (**C**), *SpTRB1* (**D**), *SpALA2* (**E**), and *SpSBE3* (**F**) between SP050 and SP011 are shown under control conditions. We detected a significantly reduced expression of *SpDLH* in SP050 compared to SP011.

Appendix B

Table A1. Estimates of EC10 and EC90 values for the two *S. polyrhiza* genotypes SP035 and SP077. We derived the EC values for the inhibition of the fitness parameters biomass (FW) and growth rate (RGR of frond number) from a hormetic model (USCR4c).

	Biomass		Growth Rate	
Genotype	**EC90 [nM] ± Standard Error [nM]**	**EC10 [nM] ± Standard Error [nM]**	**EC90 [nM] ± Standard Error [nM]**	**EC10 [nM] ± Standard Error [nM]**
SP035	11.45 ± 1.65	1.76 ± 0.27	59.22 ± 31.7	2.20 ± 0.17
SP077	10.71 ± 8.10	1.85 ± 0.82	46.80 ± 20.6	2.25 ± 0.15

Table A2. The estimated EC50 values with their respective standard error are shown for 16 genotypes. According to this, we identified the four most resistant (R) and four most susceptible genotypes (S) for future experiments. The most resistant and most susceptible genotypes were SP050 and SP011, respectively.

Genotype	Four-Digit Code	Origin of Genotype	Estimated EC50 RGR Frond Number [nM] ± Standard Error [nM]	Fold Change to the Most Susceptible Genotype	Classification
SP007	7674	Nepal, Kathmandu	3.31 ± 0.47	1.08	S
SP011	9242	Ecuador, Guayas Yaguachi Nuevo	3.06 ± 0.52	1.00	S
SP031	9514	Austria, Vienna	4.47 ± 0.53	1.46	-
SP050	0090	China, Sichuan	26.13 ± 2.61	8.54	R
SP078	9607	Switzerland, Zurich	12.93 ± 0.56	4.23	R
SP109	-	Switzerland	10.24 ± 1.08	3.35	-
SP165	-	China, Dali, Yunan	13.16 ± 0.64	4.30	R
SP171	-	China, Hunan, Hengyang	3.11 ± 0.49	1.02	S
SP185	-	China, Jiangsu, Hongze	4.16 ± 0.41	1.36	-
SP196	-	China, Jiangxi	10.69 ± 0.17	3.49	-
SP199	-	China, Hainan	4.61 ± 0.57	1.51	-
SP202	-	China, Shandong, Qingdao	4.46 ± 0.25	1.46	-
SP211	-	China, Shandong, Jinan	3.49 ± 0.25	1.14	S
SP221	5588	India	13.03 ± 0.97	4.26	R
SP236	5610	India	7.38 ± 1.13	2.41	-
SP239	5613	Israel	7.87 ± 0.78	2.57	-

Table A3. Diquat medium concentration measured daily within one week (d0–d7) in cultures of SP011 and SP050. The starting point of the cultivation is termed as d0.

	Diquat Medium Concentration [pmol/mL] ± Standard Error [pmol/mL]		
Timepoint	**Ctrl**	**SP011**	**SP050**
d0	8.11 ± 0.18	8.23 ± 0.13	8.58 ± 0.02
d1	8.47 ± 0.32	8.62 ± 0.18	8.55 ± 0.08
d2	7.86 ± 0.14	8.65 ± 0.18	8.60 ± 0.10
d3	8.04 ± 0.10	8.50 ± 0.06	9.00 ± 0.13
d4	8.67 ± 0.36	9.13 ± 0.03	8.73 ± 0.34
d5	8.73 ± 0.18	9.02 ± 0.17	9.38 ± 0.12
d6	9.00 ± 0.22	8.81 ± 0.46	9.12 ± 0.40
d7	8.84 ± 0.08	8.48 ± 0.54	9.60 ± 0.41

Table A4. Significant SNPs associated with diquat resistance in *S. polyrhiza*.

Marker	Parameter(s)	Chromosome	Location	*p*-Wald	Comment
SpGBPL2-MUT	norm. IE of FW	1	10117635	2.55×10^{-9}	Location in the coding sequence of Guanylate binding protein-like 2 (*SpGBPL2*)

Table A5. Significant SVs associated with diquat resistance in *S. polyrhiza*.

Marker	Parameter(s)	Size	Chromosome	Location	*p*-Wald	Comment
SpETHE1-DEL	norm. IE of RGR frond number, norm. IE of RGR area, norm. IE of FW	94 bp	8	5199142	5.49×10^{-8} (norm. IE of RGR frond number), 1.23×10^{-6} (norm. IE of RGR frond area), 8.89×10^{-6} (norm. IE of FW)	Deletion located in the intronic region of *SpETHE1*: mitochondrial ETHE1 persulfide dioxygenase
Intergenic region	norm. IE of FW	56 bp	5	5617685	1.14×10^{-6}	Insertion intergenic region
SpLOX2.1-DEL	norm. IE of RGR area	184 bp	5	7048033	3.45×10^{-5}	Deletion located in *SpLOX2.1*: chloroplastic lipoxygenase 2.1

Table A6. Primers used for genotyping *S. polyrhiza* genotypes to check the presence or absence of the intronic deletion in *SpETHE1*.

Name	Sequence Primer 5′-3′
1_ETHE1_fwd	GGTGGAGGAGGAGAGCAAGTACAAC
2_ETHE1_rev	GCACGGCCACGTCGATCATCTTC
3_ETHE1_fwd	GCTGCGTAACGTATGTCACAGGCG

Table A7. Primers used for RT-qPCR studies. The sequences of the primer pairs for *SpaEF* and *SpGAPDH* were published previously [37].

Fwd Primer 5′-3′	Rev Primer 5′-3′	Gene	Gene ID	Product Size	Efficiency
TCGAAGCCGGCATTTCC AAGGACG	TCGCCTTCGAGTACTTGGG	*SpaEF*	SpGA2022_005771	129 bp	99.8%
AGCATCCAAGAAGGTG AAGATCGGC	TTGTAGTCGGTCG TGATGAAGGGG	*SpGAPDH*	SpGA2022_054082	132 bp	104.4%
ACCTACACCTA CTTGCTTGCCGACG	GCCTGTTCCAGTGACATGG	*SpETHE1*	SpGA2022_053619	171 bp	99.0%
GCTTCAGCCGA GTGGCTCAAGTC	AGCAGCATCCA CTTCAGGGACCAG	*SpDLH*	SpGA2022_012161	117 bp	106.8%
CGAAGCAGAAG TGGCGACCTGAAG	CCCGAGATCCCCAACCGTT	*SpTRB1*	SpGA2022_012163	189 bp	102.1%
GTCAGTGGTTGTCA GAGACTGTCAG	TCACGCCATGCCA AACATAGCGTAC	*SpALA2*	SpGA2022_053617	171 bp	83.8%
ACCGCCACAGGCAACACA	CATCTGCTTCTTGTCCCTCT	*SpSBE3*	SpGA2022_053621	192 bp	97.1%

References

1. Heap, I. The International Survey of Herbicide Resistant Weeds. Available online: www.weedscience.org (accessed on 25 February 2024).
2. Jugulam, M.; Shyam, C. Non-Target-Site Resistance to Herbicides: Recent Developments. *Plants* **2019**, *8*, 417. [CrossRef] [PubMed]
3. Hawkes, T.R. Mechanisms of resistance to paraquat in plants. *Pest Manag. Sci.* **2014**, *70*, 1316–1323. [CrossRef] [PubMed]
4. Fujii, T.; Yokoyama, E.-I.; Inoue, K.; Sakurai, H. The sites of electron donation of Photosystem I to methyl viologen. *Biochim. Biophys. Acta Bioenerg.* **1990**, *1015*, 41–48. [CrossRef]
5. Chen, S.; Dickman, M.B. Bcl-2 family members localize to tobacco chloroplasts and inhibit programmed cell death induced by chloroplast-targeted herbicides. *J. Exp. Bot.* **2004**, *55*, 2617–2623. [CrossRef] [PubMed]
6. Tyler, J.K.; Haller, W.T.; Les, G. Documentation of Landoltia (*Landoltia punctata*) Resistance to Diquat. *Weed Sci.* **2006**, *54*, 615–619.
7. Chiconela, T.; Haller, W. Herbicide combinations for the enhancement of diquat phytotoxicity for hydrilla control. *ARPN J. Agric. Biol. Sci.* **2013**, *8*, 555–562.
8. Blackburn, R.D. Evaluating Herbicides Against Aquatic Weeds. *Weeds* **1963**, *11*, 21–24. [CrossRef]
9. Ganascini, D.; Mendes, I.; Caon, I.; Cattani, C.; Mercante, E.; Coelho, S.; Viana, O.; Prudente, V. Evaluation of bean desiccation plants with diquat and glufosinate-ammonium using terrestrial hyperspectral sensor. *Aust. J. Crop. Sci.* **2022**, *16*, 216. [CrossRef]
10. Itoh, K. Paraquat Resistance in *Erigeron philadelphicus* L. *Weed Res.* **1984**, *29*, 301–307. [CrossRef]
11. Fuerst, E.; Nakatani, H.; Dodge, A.; Penner, D.; Arntzen, C. Paraquat Resistance in Conyza. *Plant Physiol.* **1985**, *77*, 984–989. [CrossRef]

12. Powles, S.B.; Edwin, S.T.; Tom, R.M. A Capeweed (*Arctotheca calendula*) Biotype in Australia Resistant to Bipyridyl Herbicides. *Weed Sci.* **1989**, *37*, 60–62. [CrossRef]
13. Preston, C.; Holtum, J.A.; Powles, S.B. On the Mechanism of Resistance to Paraquat in *Hordeum glaucum* and *H. leporinum*: Delayed Inhibition of Photosynthetic O_2 Evolution after Paraquat Application. *Plant Physiol.* **1992**, *100*, 630–636. [CrossRef] [PubMed]
14. Purba, E.; Preston, C.; Powles, S. Paraquat resistance in a biotype of *Vulpia bromoides* (L.) S. F. Gray. *Weed Res.* **2006**, *33*, 409–413. [CrossRef]
15. Yu, Q.; Huang, S.; Powles, S. Direct measurement of paraquat in leaf protoplasts indicates vacuolar paraquat sequestration as a resistance mechanism in *Lolium rigidum*. *Pestic. Biochem. Physiol.* **2010**, *98*, 104–109. [CrossRef]
16. Brunharo, C.; Hanson, B.D. Vacuolar Sequestration of Paraquat Is Involved in the Resistance Mechanism in *Lolium perenne* L. spp. multiflorum. *Front. Plant Sci.* **2017**, *8*, 1485. [CrossRef] [PubMed]
17. Halász, K.; Soós, V.; Jóri, B.; Racz, I.; Lasztity, D.; Szigeti, Z. Effect of transporter inhibitors on paraquat resistance of horseweed (*Conyza canadensis*/L./Cronq.). *Acta Biol. Szeged.* **2002**, *46*, 23–24.
18. Jóri, B.; Soós, V.; Szegő, D.; Páldi, E.; Szigeti, Z.; Rácz, I.; Lásztity, D. Role of transporters in paraquat resistance of horseweed *Conyza canadensis* (L.) Cronq. *Pestic. Biochem. Physiol.* **2007**, *88*, 57–65. [CrossRef]
19. Visnovitz, T.; Soós, V.; Jóri, B.; Racz, I.; Szigeti, Z. Staying alive: Insight into the resistance mechanism of *Conyza canadensis* to xenobiotic paraquat. *Acta Herbol.* **2008**, *17*, 173–178.
20. Shaaltiel, Y.; Gressel, J. Multienzyme oxygen radical detoxifying system correlated with paraquat resistance in *Conyza bonariensis*. *Pestic. Biochem. Physiol.* **1986**, *26*, 22–28. [CrossRef]
21. Jansen, M.A.; Shaaltiel, Y.; Kazzes, D.; Canaani, O.; Malkin, S.; Gressel, J. Increased Tolerance to Photoinhibitory Light in Paraquat-Resistant *Conyza bonariensis* Measured by Photoacoustic Spectroscopy and CO_2-Fixation. *Plant Physiol.* **1989**, *91*, 1174–1178. [CrossRef]
22. Dong, S.; Hu, H.; Wang, Y.; Xu, Z.; Zha, Y.; Cai, X.; Peng, L.; Feng, S. A pqr2 mutant encodes a defective polyamine transporter and is negatively affected by ABA for paraquat resistance in *Arabidopsis thaliana*. *J. Plant Res.* **2016**, *129*, 899–907. [CrossRef] [PubMed]
23. Lee, Y.P.; Kim, S.H.; Bang, J.W.; Lee, H.S.; Kwak, S.S.; Kwon, S.Y. Enhanced tolerance to oxidative stress in transgenic tobacco plants expressing three antioxidant enzymes in chloroplasts. *Plant Cell Rep.* **2007**, *26*, 591–598. [CrossRef] [PubMed]
24. Kwon, S.-Y.; Choi, S.-M.; Ahn, Y.-O.; Lee, H.-S.; Lee, H.-B.; Park, Y.-M.; Kwak, S.-S. Enhanced stress-tolerance of transgenic tobacco plants expressing a human dehydroascorbate reductase gene. *J. Plant Physiol.* **2003**, *160*, 347–353. [CrossRef] [PubMed]
25. Xi, J.; Xu, P.; Xiang, C.B. Loss of AtPDR11, a plasma membrane-localized ABC transporter, confers paraquat tolerance in *Arabidopsis thaliana*. *Plant J.* **2012**, *69*, 782–791. [CrossRef] [PubMed]
26. Kim, D.Y.; Jin, J.Y.; Alejandro, S.; Martinoia, E.; Lee, Y. Overexpression of AtABCG36 improves drought and salt stress resistance in Arabidopsis. *Physiol. Plant* **2010**, *139*, 170–180. [CrossRef] [PubMed]
27. Kim, D.Y.; Bovet, L.; Maeshima, M.; Martinoia, E.; Lee, Y. The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. *Plant J.* **2007**, *50*, 207–218. [CrossRef] [PubMed]
28. Khare, D.; Choi, H.; Huh, S.U.; Bassin, B.; Kim, J.; Martinoia, E.; Sohn, K.H.; Paek, K.H.; Lee, Y. Arabidopsis ABCG34 contributes to defense against necrotrophic pathogens by mediating the secretion of camalexin. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5712–E5720. [CrossRef]
29. Prosecka, J.; Orlov, A.V.; Fantin, Y.S.; Zinchenko, V.V.; Babykin, M.M.; Tichy, M. A novel ATP-binding cassette transporter is responsible for resistance to viologen herbicides in the cyanobacterium *Synechocystis* sp. PCC 6803. *FEBS J.* **2009**, *276*, 4001–4011. [CrossRef]
30. Cedergreen, N.; Abbaspoor, M.; Sørensen, H.; Streibig, J.C. Is mixture toxicity measured on a biomarker indicative of what happens on a population level? A study with *Lemna minor*. *Ecotoxicol. Environ. Saf.* **2007**, *67*, 323–332. [CrossRef]
31. Ziegler, P.; Adelmann, K.; Zimmer, S.; Schmidt, C.; Appenroth, K.J. Relative in vitro growth rates of duckweeds (Lemnaceae)—The most rapidly growing higher plants. *Plant Biol.* **2015**, *17*, 33–41. [CrossRef]
32. Wang, W.; Haberer, G.; Gundlach, H.; Gläßer, C.; Nussbaumer, T.; Luo, M.C.; Lomsadze, A.; Borodovsky, M.; Kerstetter, R.A.; Shanklin, J.; et al. The Spirodela polyrhiza genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat. Commun.* **2014**, *5*, 3311. [CrossRef] [PubMed]
33. Xu, S.; Stapley, J.; Gablenz, S.; Boyer, J.; Appenroth, K.J.; Sree, K.S.; Gershenzon, J.; Widmer, A.; Huber, M. Low genetic variation is associated with low mutation rate in the giant duckweed. *Nat. Commun.* **2019**, *10*, 1243. [CrossRef] [PubMed]
34. Katerova, Z.I.; Miteva, L.P.-E. Glutathione and Herbicide Resistance in Plants. In *Ascorbate-Glutathione Pathway and Stress Tolerance in Plants*; Anjum, N.A., Chan, M.-T., Umar, S., Eds.; Springer: Dordrecht, The Netherlands, 2010; pp. 191–207.
35. Melicher, P.; Dvořák, P.; Krasylenko, Y.; Shapiguzov, A.; Kangasjärvi, J.; Šamaj, J.; Takáč, T. Arabidopsis Iron Superoxide Dismutase FSD1 Protects Against Methyl Viologen-Induced Oxidative Stress in a Copper-Dependent Manner. *Front. Plant Sci.* **2022**, *13*, 823561. [CrossRef] [PubMed]
36. Noctor, G.; Foyer, C.H. Ascorbate and Glutathione: Keeping Active Oxygen Under Control. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1998**, *49*, 249–279. [CrossRef] [PubMed]
37. Wang, Y.; Duchen, P.; Chávez, A.; Sree, S.K.; Appenroth, K.J.; Zhao, H.; Widmer, A.; Huber, M.; Xu, S. Population genomics and epigenomics provide insights into the evolution of facultative asexuality in plants. *bioRxiv* **2023**. [CrossRef]

38. Liao, Y.; Zhou, X.; Yu, J.; Cao, Y.; Li, X.; Kuai, B. The Key Role of Chlorocatechol 1,2-Dioxygenase in Phytoremoval and Degradation of Catechol by Transgenic Arabidopsis. *Plant Physiol.* **2006**, *142*, 620–628. [CrossRef] [PubMed]
39. Basmaciyan, L.; Jacquet, P.; Azas, N.; Casanova, M. A novel hydrolase with a pro-death activity from the protozoan parasite *Leishmania major*. *Cell Death Discov.* **2019**, *5*, 99. [CrossRef]
40. Chun, J.C.; Ma, S.Y.; Kim, S.E.; Lee, H.J. Physiological Responses of *Rehmannia glutinosa* to Paraquat and Its Tolerance Mechanisms. *Pestic. Biochem. Physiol.* **1997**, *59*, 51–63. [CrossRef]
41. Szigeti, Z. Mechanism of paraquat resistance—From the antioxidant enzymes to the transporters. *Acta Biol. Szeged.* **2005**, *49*, 177–179.
42. Liu, L.; Du, Y.; Shen, X.; Li, M.; Sun, W.; Huang, J.; Liu, Z.; Tao, Y.; Zheng, Y.; Yan, J.; et al. KRN4 Controls Quantitative Variation in Maize Kernel Row Number. *PLoS Genet.* **2015**, *11*, e1005670. [CrossRef]
43. Ricci, W.A.; Lu, Z.; Ji, L.; Marand, A.P.; Ethridge, C.L.; Murphy, N.G.; Noshay, J.M.; Galli, M.; Mejía-Guerra, M.K.; Colomé-Tatché, M.; et al. Widespread long-range cis-regulatory elements in the maize genome. *Nat. Plants* **2019**, *5*, 1237–1249. [CrossRef]
44. Ishii, T.; Numaguchi, K.; Miura, K.; Yoshida, K.; Thanh, P.T.; Htun, T.M.; Yamasaki, M.; Komeda, N.; Matsumoto, T.; Terauchi, R.; et al. OsLG1 regulates a closed panicle trait in domesticated rice. *Nat. Genet.* **2013**, *45*, 462–465. [CrossRef] [PubMed]
45. Liu, C.; Wang, C.; Wang, G.; Becker, C.; Zaidem, M.; Weigel, D. Genome-wide analysis of chromatin packing in Arabidopsis thaliana at single-gene resolution. *Genome Res.* **2016**, *26*, 1057–1068. [CrossRef] [PubMed]
46. Marand, A.P.; Eveland, A.L.; Kaufmann, K.; Springer, N.M. cis-Regulatory Elements in Plant Development, Adaptation, and Evolution. *Annu. Rev. Plant Biol.* **2023**, *74*, 111–137. [CrossRef] [PubMed]
47. Moabbi, A.M.; Agarwal, N.; El Kaderi, B.; Ansari, A. Role for gene looping in intron-mediated enhancement of transcription. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 8505–8510. [CrossRef] [PubMed]
48. Huang, S.; Zhu, S.; Kumar, P.; MacMicking, J.D. A phase-separated nuclear GBPL circuit controls immunity in plants. *Nature* **2021**, *594*, 424–429. [CrossRef] [PubMed]
49. Zabbai, F.; Jarosch, B.; Schaffrath, U. Over-expression of chloroplastic lipoxygenase RCI1 causes PR1 transcript accumulation in transiently transformed rice. *Physiol. Mol. Plant Pathol.* **2004**, *64*, 37–43. [CrossRef]
50. Mochizuki, S.; Sugimoto, K.; Koeduka, T.; Matsui, K. Arabidopsis lipoxygenase 2 is essential for formation of green leaf volatiles and five-carbon volatiles. *FEBS Lett.* **2016**, *590*, 1017–1027. [CrossRef]
51. Shim, K.-C.; Adeva, C.; Kang, J.-W.; Luong, N.H.; Lee, H.-S.; Cho, J.-H.; Kim, H.; Tai, T.H.; Ahn, S.-N. Interaction of starch branching enzyme 3 and granule-bound starch synthase 1 alleles increases amylose content and alters physico-chemical properties in japonica rice (*Oryza sativa* L.). *Front. Plant Sci.* **2022**, *13*, 968795. [CrossRef]
52. Fukao, Y.; Ferjani, A.; Fujiwara, M.; Nishimori, Y.; Ohtsu, I. Identification of Zinc-Responsive Proteins in the Roots of Arabidopsis thaliana Using a Highly Improved Method of Two-Dimensional Electrophoresis. *Plant Cell Physiol.* **2009**, *50*, 2234–2239. [CrossRef]
53. Shaaltiel, Y.; Chua, N.H.; Gepstein, S.; Gressel, J. Dominant pleiotropy controls enzymes co-segregating with paraquat resistance in *Conyza bonariensis*. *Theor. Appl. Genet.* **1988**, *75*, 850–856. [CrossRef]
54. Keith, B.K.; Burns, E.E.; Bothner, B.; Carey, C.C.; Mazurie, A.J.; Hilmer, J.K.; Biyiklioglu, S.; Budak, H.; Dyer, W.E. Intensive herbicide use has selected for constitutively elevated levels of stress-responsive mRNAs and proteins in multiple herbicide-resistant *Avena fatua* L. *Pest Manag. Sci.* **2017**, *73*, 2267–2281. [CrossRef] [PubMed]
55. Dyer, W.E. Stress-induced evolution of herbicide resistance and related pleiotropic effects. *Pest Manag. Sci.* **2018**, *74*, 1759–1768. [CrossRef] [PubMed]
56. Comont, D.; Lowe, C.; Hull, R.; Crook, L.; Hicks, H.L.; Onkokesung, N.; Beffa, R.; Childs, D.Z.; Edwards, R.; Freckleton, R.P.; et al. Evolution of generalist resistance to herbicide mixtures reveals a trade-off in resistance management. *Nat. Commun.* **2020**, *11*, 3086. [CrossRef] [PubMed]
57. Mittler, R.; Zandalinas, S.I.; Fichman, Y.; Van Breusegem, F. Reactive oxygen species signalling in plant stress responses. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 663–679. [CrossRef] [PubMed]
58. Appenroth, K.J. Media for in vitro-cultivation of duckweed. *Duckweed Forum* **2015**, *3*, 180–186.
59. Lam, E.; Acosta, K. Useful methods: Cefotaxime: A useful antibiotic for duckweed culture management. *Duckweed Forum* **2019**, *7*, 52–54.
60. Streibig, J.; Ritz, C. Bioassay Analysis Using R. *J. Stat. Softw.* **2010**, *12*, 1–22. [CrossRef]
61. Ritz, C.; Baty, F.; Streibig, J.C.; Gerhard, D. Dose-Response Analysis Using R. *PLoS ONE* **2016**, *10*, e0146021. [CrossRef]
62. Cullis, B.R.; Smith, A.B.; Coombes, N.E. On the Design of Early Generation Variety Trials with Correlated Data. *J. Agric. Biol. Environ. Stat.* **2006**, *11*, 381–393. [CrossRef]
63. Lozano-Isla, F. *Inti: Tools and Statistical Procedures in Plant Science, R package version 0.6.3*; R Package Vignette: Madison, WI, USA, 2023.
64. Cedergreen, N.; Ritz, C.; Streibig, J.C. Improved empirical models describing hormesis. *Environ. Toxicol. Chem.* **2005**, *24*, 3166–3172. [CrossRef] [PubMed]
65. Genty, B.; Briantais, J.-M.; Baker, N.R. The relationship between the quantum yield of photosynthetic electron transport and quenching of chlorophyll fluorescence. *Biochim. Biophys. Acta Gen. Subj.* **1989**, *990*, 87–92. [CrossRef]
66. Baker, N.R. Chlorophyll Fluorescence: A Probe of Photosynthesis In Vivo. *Annu. Rev. Plant Biol.* **2008**, *59*, 89–113. [CrossRef]
67. Küster, A.; Pohl, K.; Altenburger, R. A fluorescence-based bioassay for aquatic macrophytes and its suitability for effect analysis of non-photosystem II inhibitors. *Environ. Sci. Pollut. Res. Int.* **2007**, *14*, 377–383. [CrossRef]

68. Asudi, G.O.; Omenge, K.M.; Paulmann, M.K.; Reichelt, M.; Grabe, V.; Mithöfer, A.; Oelmüller, R.; Furch, A.C.U. The Physiological and Biochemical Effects on Napier Grass Plants Following Napier Grass Stunt Phytoplasma Infection. *Phytopathology* **2021**, *111*, 703–712. [CrossRef] [PubMed]
69. Chamkasem, N.; Harmon, T. Determination of Paraquat and Diquat in Potato by Liquid Chromatography/Tandem Mass Spectrometer. *J. Regul. Sci.* **2017**, *5*, 1–8. [CrossRef]
70. Ho, E.K.H.; Bartkowska, M.; Wright, S.I.; Agrawal, A.F. Population genomics of the facultatively asexual duckweed *Spirodela polyrhiza*. *New Phytol.* **2019**, *224*, 1361–1371. [CrossRef]
71. Zhou, X.; Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **2012**, *44*, 821–824. [CrossRef]
72. Vogt, F.; Shirsekar, G.; Weigel, D. Vcf2gwas-python API for comprehensive GWAS analysis using GEMMA. *Bioinformatics* **2021**, *38*, 839–840. [CrossRef]
73. Lemay, M.-A.; Malle, S. A Practical Guide to Using Structural Variants for Genome-Wide Association Studies. In *Genome-Wide Association Studies*; Torkamaneh, D., Belzile, F., Eds.; Springer: New York, NY, USA, 2022; pp. 161–172.
74. Kaur, C.; Singla-Pareek, S.L.; Sopory, S.K. Stress response of OsETHE1 is altered in response to light and dark conditions. *Plant Signal. Behav.* **2014**, *9*, e973820. [CrossRef]
75. Krüßel, L.; Junemann, J.; Wirtz, M.; Birke, H.; Thornton, J.D.; Browning, L.W.; Poschet, G.; Hell, R.; Balk, J.; Braun, H.P.; et al. The mitochondrial sulfur dioxygenase ETHYLMALONIC ENCEPHALOPATHY PROTEIN1 is required for amino acid catabolism during carbohydrate starvation and embryo development in Arabidopsis. *Plant Physiol.* **2014**, *165*, 92–104. [CrossRef]
76. Hellemans, J.; Mortier, G.; De Paepe, A.; Speleman, F.; Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol.* **2007**, *8*, R19. [CrossRef]
77. Ogle, D.H.; Doll, J.C.; Wheeler, P.; Dinno, A. *FSA: Fisheries Stock Analysis, R package version 0.9.5*; R Package Vignette: Madison, WI, USA, 2023.
78. Bates, D.; Mächler, M.; Bolker, B.; Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **2015**, *67*, 1–48. [CrossRef]
79. Hothorn, T.; Bretz, F.; Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **2008**, *50*, 346–363. [CrossRef]
80. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef]

3.2 Genome-Wide Association Study of Metabolic Traits in the Duckweed *Spirodela polyrhiza*

3.2.1 Summary

In this study we aimed to elucidate the genetic basis of free metabolite levels and growth in *Spirodela polyrhiza*.

For this, we characterized the profiles of 42 metabolites and growth in 137 *S. polyrhiza* genotypes and identified SNPs and structure variations associated with these traits using a GWA approach. Growth was quantified as fresh weight, dry weight, relative growth rate (RGR) of frond area and RGR of frond number. Metabolites, including 20 amino acids and related compounds, 11 specialized metabolites and 11 phytohormones were quantified by LC-MS.

Among all quantified metabolites L-Asparagine and L-Glutamine were found to be most abundant in *S. polyrhiza*. On average, specialized metabolites showed higher amounts of intraspecific variation than amino acids and phytohormones. Whereas correlations of most amino acids with biomass were positive, specialized metabolites were mostly negatively associated with biomass. Among all metabolites, Cyanidine-3-O-glycoside showed the strongest negative correlation patterns with growth, either measured by increase of frond number or area.

We applied a GWA approach to identify the genetic background controlling growth and metabolite levels. Although no SNPs were associated with growth traits, we identified many genetic markers jointly associated with the contents of several metabolites. Among them, we identified markers linked to the *LIGHT-HARVESTING-CHLOROPHYLL-A/B-BINDING PROTEIN 5* (*SpLHCB5*), the *UBIQUITIN-CARBOXYTERMINAL HYDROLASE 7* (*SpUBP7*), the *U-BOX CONTAINING PROTEIN 4* (*SpPUB4*) and the transcription factor *SpMYBC1* to be associated with metabolites, positively correlated to plant biomass. Marker presence was associated with an increased expression level of *SpUBP7* in root tissue, studied by qPCR.

Together, this study revealed strong positive correlation patterns between contents of various amino acids and biomass. Further, multiple genes controlling photosynthesis (*SpLHCB5*), protein-degradation (*SpUBP7*), cell-cycle (*SpPUB4*) and secondary metabolite synthesis (*SpMYBC1*) were associated with biomass-associated traits. The identified candidate genes are therefore of potential use for the simultaneous optimization of biomass and metabolite contents in duckweed.

3.2.2 Zusammenfassung

Das Ziel dieser Studie lag in der Ergründung der genetischen Basis metabolischer Eigenschaften und des Wachstums von *Spirodela polyrhiza*.

Hierfür wurde die Wachstumsgeschwindigkeit und die Konzentrationen von 42 Metaboliten in 137 Genotypen von *S. polyrhiza* bestimmt. Anschließend wurden mit diesen Eigenschaften assoziierte SNPs und strukturelle Variationen ermittelt. Wachstum wurde in Form der relativen Wachstumsrate gemessen an Frondanzahl und -fläche, sowie anhand der Biomasse als Frisch- oder Trockengewicht bestimmt. Die Metaboliten, bestehend aus 20 Aminosäuren und strukturell eng verwandten Verbindungen, 11 Sekundärmetaboliten und 11 Phytohormonen wurden mittels LC-MS quantifiziert.

Innerhalb aller quantifizierten Metaboliten zeigten L-Asparagin und L-Glutamin die höchsten Konzentrationen im Pflanzengewebe. Im Durchschnitt zeigten Konzentrationen von Sekundärmetaboliten ein höheres Maß an intraspezifischer Variabilität verglichen mit Aminosäuren und Phytohormonen. Während Aminosäurelevel vorwiegend positiv mit der pflanzlichen Biomasse korrelierten, waren Sekundärmetabolitlevel vorwiegend negativ mit der Biomasse korreliert. Innerhalb aller quantifizierter Metabolite zeigte Cyanidin-3-O-glykosid die stärksten negativen Korrelationen zu den gemessenen Wachstumsraten. Zur Identifikation der genetischen Basis des pflanzlichen Wachstums und der Metabolitkonzentrationen wurde eine Genomweite Assoziationsstudie durchgeführt. Insgesamt konnten keine SNPs mit Wachstumsparametern assoziierte werden. Allerdings sind zahlreiche Marker, die mit mehreren Metabolitkonzentrationen simultan assoziiert sind, gefunden worden. Beispielsweise sind Marker innerhalb oder in starker Nähe zu dem *LIGHT-HARVESTING-CHLOROPHYLL-A/B-BINDING PROTEIN 5* (*SpLHCB5*), dem *UBIQUITIN-CARBOXYTERMINAL HYDROLASE 7* (*SpUBP7*), dem *U-BOX CONTAINING PROTEIN 4* (*SpPUB4*) oder dem Transkriptionsfaktor *SpMYBC1* mit gleichzeitig mit mehreren Metaboliten assoziiert, die positiv mit pflanzlicher Biomasse korrelieren. Die innerhalb von *SpUBP7* gefundene Strukturvariation war mit einer erhöhten Expression des Gens assoziiert.

Insgesamt identifizierte die Studie starke positive Korrelationen zwischen verschiedensten Aminosäuren und pflanzlicher Biomasse. Weiterhin wurden Gene, welche funktional in Prozessen wie Photosynthese (*SpLHCB5*), Proteinabbau (*SpUBP7*), Zell-Zyklus Kontrolle (*SpPUB7*) und Sekundärmetabolitbiosynthese (*SpMYBC1*) involviert sind mit Metaboliten assoziiert, die positiv mit Biomasse korrelieren. Aufgrund dessen, können die identifizierten Kandidatengene zur simultanen Optimierung von Biomasse und Metabolitleveln dienen.

3.2.3 Statement of Contribution

This project was planned and coordinated by Shuqing Xu and Martin Höfer. Growth rates of the 137 genotypes were recorded by Yangzi Wang, Samuel Wink and Martin Höfer. Martin Schäfer and Samuel Wink conducted all metabolite extractions and LC-MS based quantifications. Martin Höfer analyzed all data, created all figures and wrote the first draft of this manuscript. All co-authors provided feedback on the manuscript.

Supervision confirmation

Prof. Shuqing Xu

iomE
Institut für Organismische und Molekulare Evolutionsbiologie
JG|U Mainz

Genome-Wide Association Study of Metabolic Traits in the Duckweed *Spirodela polyrhiza*

Höfer, Martin[1]; Schäfer, Martin[1]; Wang, Yangzi[1,2]; Wink, Samuel[2]; Xu, Shuqing[1]

[1] Institute for Organismic and Molecular Evolution (iomE), Johannes Gutenberg University, 55128 Mainz, Germany

[2] Institute for Evolution and Biodiversity, University of Münster, 48148 Münster, Germany

Abstract

The exceptionally high growth rate and high contents of flavonoids make the giant duckweed *Spirodela polyrhiza* (L.) Schleid. (Arales: Lemnaceae) an ideal organism for food production and metabolic engineering. To facilitate this, identification of the genetic basis underlying growth and metabolic traits is essential. Here, we characterized the genetics underpinning the growth and contents of 42 metabolites in *S. polyrhiza* using a genome-wide association (GWA) approach. We found that biomass positively correlates with contents of many free amino acids, including L-Glutamine, L-Tryptophane and L-Serine, but negatively correlates with many specialized metabolites, such as flavonoids. GWA analysis showed that several candidate genes were simultaneously associated with several metabolic traits, qualifying them as targets for growth manipulation and metabolic engineering. Together, this study provides insights into the metabolic diversity of *S. polyrhiza* and its underlying genetic architecture, paving the way for industrial applications of this plant via targeted breeding or genetic engineering.

Keywords: mGWAS, trait variation, *Spirodela polyrhiza*, targeted metabolomics, duckweed metabolism

Introduction

With a biomass duplication rate of less than two days, *S. polyrhiza* is one of the world's fastest-growing angiosperms (Ziegler et al. 2015). Due to its high levels of flavonoids and amino acids, *S. polyrhiza* is one of the best organisms for producing food resources and pharmaceuticals (Acosta et al. 2021; Baek et al. 2021; Zhao et al. 2015; Smith et al. 2024). To fully realise its industrial potential, specific improvements on its growth and metabolite production are needed, which requires a detailed understanding of the genetic mechanisms controlling these traits. Although previous forward genetic studies on metabolic traits in species like *Zea mays* L. (Poales: Poaceae) (Chen et al. 2016), *Oryza sativa* L. (Poales: Poaceae) (Chen et al. 2014; Zhang et al. 2019; Cu et al. 2021) and *Arabidopsis thaliana* (L.) Heynh. (Brassicales: Brassicaceae) (Angelovici et al. 2013; Angelovici et al. 2017) identified many promising candidate genes, the genetic principles controlling growth and plant metabolisms in *S. polyrhiza* largely remain unknown, hampering its further biotechnological optimization.

For most plants, to survive from natural stress factors, such as herbivory, they must carefully allocate their resources to growth or defense pathways. On a molecular level, free amino acids play a pivotal role in balancing resource distribution between biomass production and the synthesis of defense metabolites. Biosynthesis of amino acids can account for 50% of all carbon compound syntheses and consume 32% of total fixed carbon dioxide (Smith et al. 1961), making them a major carbon sink for biomass production in plants (Noctor and Foyer 1998). The central metabolic function of free amino acids is highlighted by their association with key metabolic pathways such as glycolysis, tri-carboxylic-acid (TCA) cycle, pentose-phosphate-pathway, urea-

cycle and photorespiration (Figure 1) (Noctor and Foyer 1998). Besides their function in biomass gain, amino acids are precursors of many specialized metabolites like flavonoids and phenolic acids (Figure 1), which are involved in plant defense. Yet, the biosynthesis of specialized metabolites often negatively impacts plant biomass, primarily through the consumption of large nutrient resources (Züst et al. 2011) and autotoxicity (Dick et al. 2012; Li et al. 2021). Consequently, optimization for high crop yields often leads to high susceptibilities to pests due to a lack of defense metabolites, a phenomenon known as growth-defense trade-off (Züst and Agrawal 2017; Huot et al. 2014). On a regulatory level resource allocation from growth to defense pathways is controlled by phytohormones such as Jasmonic acid (JA), Abscisic acid (ABA), Salicylic acid (SA) and Auxin (IAA) (Huot et al. 2014; Hui et al. 2023; Aftab et al. 2010; Živanović et al. 2020a). Therefore, understanding the genetic mechanism controlling levels of phytohormones, free amino acids and related metabolites represents the key for optimizing the biomass yield and specialized metabolite contents in plants.

Here, we focus on understanding the genetic control regulating levels of free amino acids and their related specialized metabolites in the giant duckweed. We quantified the contents of 42 metabolites (Figure 1) in 137 genotypes of *S. polyrhiza* and performed a genome-wide association study on the contents of these metabolites. We aim to address the following questions: 1) To what extent do growth and metabolism vary in *S. polyrhiza*? 2) which metabolites are associated with biomass and growth? and 3) which genes control these metabolic traits?

Materials and Methods

Data collection and sample preparation

In a previous study we estimated the growth of 138 *S. polyrhiza* genotypes under herbicide treatment and under control conditions (Höfer et al. 2024b). We quantified

79 growth as relative growth rates (RGR) of frond area and frond number and as fresh
80 weight and dry weight at the end of a seven-day growth period. For each genotype
81 growth was estimated from three technical replicates. The quantified growth
82 parameters are publicly available (Höfer et al. 2024a). A table documenting the
83 geographic origins of all genotypes analyzed in this study can be found at:
84 https://datadryad.org/stash/downloads/file_stream/2894056.
85 All analyses included in this study were based on plant material grown under control
86 conditions (Höfer et al. 2024b). After harvesting, we freeze-dried plant material and
87 stored it at room temperature (RT) till metabolite extraction. To have sufficient plant
88 material for metabolite analysis, triplicates were pooled for each genotype. The RGRs
89 were estimated (Höfer et al. 2024b) based on a published method (Ziegler et al. 2015),
90 applying the formula: $RGR = \frac{\ln(Nt) - \ln(N0)}{t}$, where N_t is the frond number / area at the
91 endpoint of the assay, N_0 is the frond number at the start of the cultivation and t as the
92 duration of the bioassay.
93
94 **Metabolite extraction**
95 To investigate genotypic variation in metabolite contents via LC-MS, we extracted 42
96 metabolites from pooled triplicates of 137 genotypes (Table S1, Figure 1). The number
97 of analyzed genotypes was reduced from 138 to 137 due to mislabelling of one
98 genotype, which was excluded from all analyses. The protocols used for LC-MS
99 analysis of metabolite levels from different fractions are listed as Methods S1-5 in the
100 supplemental part. The entire extraction procedure and quantification procedure is
101 based on a method published previously (Schäfer et al. 2016).
102 The samples were ground with two steel balls at 30 Hz for 30 s in a TissueLyser
103 (Qiagen, Venlo, Netherlands). Next, we aliquoted 10 mg of the homogenized samples
104 in 96-well BioTubes (Axygen, New York, NY, USA). We then applied 800 µL of

105 extraction buffer (methanol:water:formic acid (15:4:1) (v/v)) mixture containing [$^{13}C_6$]-
106 IAA (1 ng), [$^{2}H_5$]-JA (10 ng), [$^{2}H_6$]-ABA (10 ng), [$^{2}H_4$]-SA (10 ng), [$^{2}H_5$]-tZ (0.5 ng), [$^{2}H_6$]-
107 iP (0.2 ng) and [$^{2}H_6$]-iPR (0.2 ng) to each tube. We sealed the plate with a pre-cooled
108 mat and homogenized the samples in a TissueLyser at 25 Hz for 60 s. Samples were
109 incubated at -20 °C overnight. After that, we repeated the homogenization step at 25
110 Hz for 60 s. We then centrifuged the samples at 2250 g for 20 min at 4 °C and
111 transferred 600 µL of the supernatant in new BioTubes, that were pre-cooled to -20 °C
112 previously. The pellet was kept for re-extraction. We diluted an aliquot of 1 µL of the
113 supernatant in 99 µL of algae amino acid standard mix -^{13}C, ^{15}N (0.5 ng/µL) (Sigma-
114 Aldrich, St. Louis, MO, USA) and transferred the dilutions in 96-well PCR plates
115 (Kisker, Steinfurt, Germany). The PCR plates were then stored at -20 °C till analysis
116 of all amino acids and several secondary metabolites (Methods S1 and S2, Table S1).
117 The pellet was reextracted with 600 µL of extraction buffer without standards. We
118 homogenized the reextracted samples at 25 Hz for 60 s before incubation at -20 °C for
119 30 min. After another centrifugation step at 2250 g for 20 min at 4 °C, we mixed 600
120 µL of the supernatant from the reextracted samples with the remaining supernatant
121 from the first part of the extraction. We centrifuged the combined supernatants at 2250
122 g for 20 min at 4 °C. Before loading our samples, we conditioned the HR-X columns
123 (Macherey-Nagel, Düren, Germany) with 600 µL of MeOH, followed by 600 µL of
124 extraction buffer and discarded the flow through. We collected the flow-through of our
125 samples in Nunc 96-well Deep Well Plates (Thermo Fisher Scientific, Waltham,
126 Massachusetts, USA). Next, we applied another 200 µL of extraction buffer and
127 collected the flow through in the same well plate. We then evaporated the MeOH at 45
128 °C. Next, we added 850 µL of 1 N formic acid to each sample, mixed the samples at
129 20 Hz for 30 s using a TissueLyser and centrifuged them at 2250 g for 20 min at 4 °C.
130 Before starting the next purification step, we conditioned HR-XC columns with 600 µL

MeOH and afterwards with 600 µL 1 N formic acid. The flow-through was discarded. We loaded the samples on the conditioned HR-XC columns and subsequently washed the columns with 1.2 mL of 1 N formic acid. We discarded the flow through and eluted the sample with 1 mL of 0.2 N formic acid in 80% (v/v) MeOH. We collected the eluates in BioTubes. After sample homogenization by inversion, we transferred 50 µL of the supernatant into a 96-well PCR plate for analysis of phytohormones (Method S3, Table S1). The sealed plates were stored at -20 °C till analysis.

For analysis of Indole-3-acetic acid (IAA) and phenolic acids (Method S4), the remaining eluate was evaporated under a constant nitrogen stream at 45 °C and samples were reconstituted to 50 µL with 0.2 N HCOOH in 80% (v/v) MeOH. We covered the plates before mixing at 20 Hz for 30 s. After centrifugation, samples were transferred to a 96-well PCR plate and were stored till analysis at -20 °C. For extraction of cytokinins (Method S5), we washed the HR-X column twice with 1 mL of 0.35 N NH_4OH. Cytokinins were finally eluted with 1 mL of 0.35 N NH_4OH in 60% (v/v) methanol. The samples were then dried under constant nitrogen flow. Next, we reconstituted sample volume to 50 µL with 0.1% (v/v) acetic acid. The samples were then homogenized at 20 Hz for 30 s, before incubation in the ultrasonic bath for 5 min. The samples were centrifuged at 2250 g for 20 min at 4 °C, and the supernatant was transferred to a 96-well PCR plate and stored at -20 °C until analysis (Method S5).

LC-MS analysis of metabolite contents

We quantified all metabolites via LC-MS. For separation and quantification, we used a Nexera X3 UHPLC system (Shimadzu, Kyoto, Japan) coupled to an LCMS-8060 system (Shimadzu). For separation of all metabolites, we used a Zorbax RRHD Eclipse XDB-C_{18} column (50 x 3 mm, 1.8 µm) (Agilent, Santa Clara, CA, USA) with a 1290 Infinity II inline filter (0.3 µm) (Agilent). For all measurements, the autosampler was

pre-cooled to 5 °C during the entire measurement procedure. As a mobile phase, we used a mixture of 0.05% (v/v) of formic acid and 0.1% (v/v) acetonitrile as solvent A and 100% methanol as solvent B, that were applied in gradient mode at a constant flow rate of 0.5 mL/min in all methods. The column oven temperature was always 42 °C. For analyses, we used an electrospray ionization source with the following parameters: nebulizing gas flow: 3 L/min, heating gas flow: 10 L/min, drying gas flow: 10 L/min, interface temperature: 300 °C, DL temperature: 250 °C, heat block temperature: 400 °C and CID gas flow: 270 kPa. For metabolites measured in the positive and negative ionization modes, we applied interface voltages of 4000 V and -3000 V, respectively (Table S2-6). A detailed description of individual metabolite analysis can be found in the supplemental (Method S1-5).

Correlation and Principal Component Analysis

Next, we used regression and principal component analysis (PCA) to identify correlation patterns between metabolic parameters and to identify the influences of population structure on metabolite levels and growth. We established multiple correlations for all metabolite and growth data using the ggcorrplot R-package (version 0.1.4.1) (Kassambara 2023). We used the Pearson correlation coefficient to measure the strength of our linear correlations. All PCAs were conducted using the R package pcaMethods (version 1.88.0) (Stacklies et al. 2007). We calculated 95% confidence intervals based on standard deviations using the vegan package (version 2.6-2) (Oksanen et al. 2022) for all PCAs. All three-dimensional scatter plots were created using the R package scatterplot3d (version 0.3.42) (Ligges et al. 2003). For identifying metabolite categories, which strongly cluster together, we sorted all 42 metabolites into nine groups based on their chemical characteristics and biochemical precursors: Flavonoids, Phenolic acids, Aromatic amino acids, Glutamate family, Jasmonates,

183 Serine family, Pyruvate family, Aspartate family and Cytokinins (Figure 1). The PCA
184 on metabolite concentrations was subsequently done for all 137 genotypes.

185 Since *S. polyrhiza* has four genetic populations, as previously described by our
186 lab (Wang et al. 2024; Xu et al. 2019), we next investigated to which extent differences
187 in metabolite concentrations and growth are explained through population level.
188 Assignment of the genotypes to different genetic populations was done in one of our
189 previous studies (Wang et al. 2024). Due to the high proportion of clonal phenotypes
190 in our accession, 137 genotypes were grouped into 97 clonal families for PCA on
191 population structure of growth parameters and metabolite levels, respectively. For
192 analysis, each clonal family was represented by the genotype with the highest
193 sequencing coverage, as described previously (Höfer et al. 2024b).

194

195 **GWAS**

196 To explore the genetic background underlying metabolic traits, we conducted GWAS
197 on growth data and metabolite contents of all representative genotypes. We used
198 single nucleotide polymorphisms (SNPs) and structural variations (SVs) (> 50 bp) as
199 genetic markers for all GWAS analyses. To allow their usage for GWAS, SVs were
200 recoded according to a previously published method (Lemay and Malle 2022). To
201 correct for missing data, we performed an imputation of both SNP and SV data sets
202 using beagle 5.4 version 22Jul22.46e (Browning et al. 2018).

203 We conducted GWAS on SNP and SV data using the vcf2gwas platform
204 (version 0.8.7) (Vogt et al. 2021). Markers were pruned using the Plink software
205 integrated into vcf2gwas, with phased r^2 thresholds of 0.33 and 0.15 for SVs and SNPs,
206 respectively. For filtering low abundant alleles, we applied a minor allele frequency
207 (MAF) threshold of 5%, giving 10,057 SNPs and 1182 SVs for analysis. We corrected
208 for population structure through principal component analysis (PCA) from the input

209 genotype files. Here, we estimated population structure based on four PCs accounting
210 for the four genetic populations of *S. polyrhiza*. All GWAS were conducted using a
211 univariate linear mixed model (Zhou and Stephens 2012). The genotypic data,
212 including the annotation, were published previously (Wang et al. 2024) and can be
213 found under: https://github.com/Xu-lab-Evolution/Great_duckweed_popg (accessed
214 on 22 March 2024).
215
216 **RT-qPCR quantification of candidate genes**
217 Our GWAS detected a deletion in *SpUBP7* that was associated with increased
218 contents of L-Glutamine and L-Serine. To validate the effect of the deletion on gene
219 function, we studied the expression of *SpUBP7* (*SpGA2022_056000*) in two different
220 allelic backgrounds.
221 For this we cultivated genotypes SP012 (low contents of L-Serine and L-
222 Glutamine, no deletion in *SpUBP7*) and SP187 (high contents of L-Serine and L-
223 Glutamine, homozygous for deletion in *SpUBP7*) in N-medium (Appenroth 2015) at 26
224 °C, 135 µmol photons · m^{-2} · s^{-1} 16 h/8 h light/dark rhythm. For each genotype, five
225 replicates were made, each consisting of ten fronds growing as colonies. The fronds
226 were cultivated in plastic beakers (Verpackungsbecher PP, transparent, round, 250
227 mL, Plastikbecher.de GmbH, Giengen, Germany), which were covered with perforated
228 lids to allow gas exchange. Each beaker was filled with 150 mL N-medium.
229 After five days of cultivation, we harvested the root and frond tissue of each
230 replicate separately for RNA extraction. We extracted RNA from < 20 mg fresh weight
231 using the InnuPREP RNA mini kit (Analytik Jena, Jena, Germany). For each sample,
232 RNA concentration was measured using a Nanodrop and integrity was checked via
233 gel-electrophoresis on a 1% agarose gel. We used 600 ng of RNA for each cDNA
234 synthesis following the instructions of the RevertAid First Strand cDNA synthesis kit

(Takara, Shiga, Japan). For all cDNA syntheses, we used Oligo-dT primers. qPCRs were carried out using a RotorGene Q system (Qiagen, Venlo, The Netherlands) applying a time program with an initial denaturation of 98 °C – 3 min, followed by 40 cycles with 98 °C for 3 s and 60 °C for 20 s. Before qPCR, we estimated primer efficiency for *SpUBP7* by using serial dilutions of pooled cDNA templates from root and frond tissue samples. Primer specificity was checked by evaluating qPCR products on a 2% agarose gel and by melting curve analysis. We conducted all qPCR reactions according to the instructions of the KAPPA SYBR FAST kit (Roche, Basel, Switzerland). All cDNA samples were diluted in a 1:100 ratio in nuclease-free water before analysis. We used the alpha-elongation factor one gene *SpaEF* (*SpGA2022_005771*) and the glycerinaldehyd-3-phosphate dehydrogenase gene *SpGAPDH* (*SpGA2022_054082*) as references for visualization of candidate gene expression according to the delta-delta Ct method with multiple reference genes as published previously (Hellemans et al. 2007). The primers used for this qPCR study can be found in Table S7. The primers for the reference genes *SpaEF* and *SpGAPDH* were used in previous publications (Höfer et al. 2024b; Wang et al. 2024).

Software and Statistics

We conducted all statistical analyses with R version 4.2.0. Standard errors of means were calculated using the plotrix R package (Lemon 2006). For all LC-MS measurements, we recorded and quantified analyte and standard peaks with LabSolutions software version 5.97. We estimated the broad sense heritability (H2.c) from our comprehensive data set, which includes repeatedly analyzed genotypes with the lme4 R package (Bates et al. 2015) following a method previously published by Cullis et al. (2006).

The significance of correlations was evaluated using the F-test. We used Bonferroni corrected *P*-value < 0.05 to determine significant genetic markers associated with our traits. The *P*-values for GWAS were calculated using the Wald test. In case of *SpUBP7*, we compared the SV effect of heterozygous and WT samples using a two-sided Student's *t*-test. We used Students *t*-test for comparison of gene expression between two genotypes. We conducted the Levene test to check for homogeneity of variances using the car R package (Fox and Weisberg 2019).

Results

Intraspecific variation of growth rate and metabolic traits

We quantified 42 metabolites, including 20 free amino acids and their derivatives, 11 secondary metabolites and 11 phytohormones (Figure 1) among 137 *S. polyrhiza* genotypes. We found L-Asparagine and L-Glutamine were the most abundant free amino acid in *S. polyrhiza*, having average genotype tissue concentrations of 114.4 µmol/mg DW and 57.0 µmol/mg DW, respectively (Table S8). Abscisic acid (ABA) was the least abundant metabolite, showing an average concentration of 37 pmol/mg DW (Table S9).

Among all quantified metabolic features contents of flavonoids and phenolic acids showed the highest amounts of intraspecific variation (Table S10, Figure S1-4). Chlorogenic acid and Luteolin-7-O-glycoside showed 214.9 and 183.8-fold differences, respectively, among genotypes (Table S10, Figure S3). Growth rate, either quantified as RGRs of frond number or frond area, had less than 2.4-fold differences (Figure S1).

On average, growth parameters also appeared to have a higher broad-sense heritability than metabolites (Table 1, Table S8-10). All growth parameters showed overall high broad-sense heritability values between 0.6 and 0.9, with fresh weight being the most heritable (Table 1). We found 25 metabolites reached high heritability

values (H2.c > 0.3). The remaining 16 metabolites showed low heritability (H2.c < 0.3) (Table S8-10). Together, these findings suggest that growth rate and metabolic traits are variable within *S. polyrhiza*, and the majority of these traits are heritable.

Metabolic traits correlate with biomass and growth

Plant metabolic traits are often associated with growth and biomass (Angelovici et al. 2013; Chen et al. 2016; Živanović et al. 2020b; Cu et al. 2021). In *S. polyrhiza*, we found plant metabolites showed higher correlations with biomass than with growth rate. In total, we found 19% and 29% of quantified metabolites showed significant correlation with growth rate (RGR of frond number and frond area) and biomass (fresh weight and dry weight) (Figure 2A, Figure S5 and S6), respectively. Among 20 quantified free amino acids, 11 showed a significant positive correlation with biomass (dry weight), including L-Glutamine, L-Tyrosine, L-Serine and L-Threonine (Figure 2A, Figure S6), whereas five were positively correlated with growth rate (RGR of frond number). Among phytohormones, only IAA and JA-Ile showed a positive correlation with biomass. Interestingly, most of the quantified specialized metabolites, while positively correlated with cytokinins, were negatively correlated with biomass (Figure 2A). Among all metabolites, Cyanidine-3-C-glycoside showed the strongest negative correlation with both growth rate and biomass (fresh weight) (Figure 2A, Figure S5 and S7).

Among the quantified metabolites, we found strong correlations within each metabolite group (Figure S8). Most of the free amino acids are positively correlated with each other and with flavonoids. Phenolic acids often were positively correlated with flavonoids (Figure 2A), reflecting their biosynthetic relationships (Figure 1).

Because *S. polyrhiza* has a strong population structure (Wang et al. 2024; Xu et al. 2019), which can confound the phenotypic variations and reduce the power of

312 GWAS. We further investigated whether metabolisms differed among populations,
313 using a principal component analysis (PCA). The results showed that there are no
314 population-specific differences at the metabolic level (Figure 2B), indicating GWAS can
315 be used for most of the metabolic traits in *S. polyrhiza*.
316
317 **GWAS of metabolite traits**
318 We used a GWA approach to identify the genetic basis underlying different metabolic
319 traits. Here, we focused on 10057 unlinked SNPs and 1182 SVs. A total of 75 SNPs
320 and 15 SVs were significantly associated with the 42 metabolites (Table S11-13).
321 Among them, 13 SNPs and three SVs were found to be associated with several
322 metabolites, indicating their pleiotropic role in metabolisms. Surprisingly, we did not
323 find any SNPs and only one SV, located in an intergenic region (> 30 Kb from next
324 open reading frame), to be associated with growth rate and biomass (Table S12 and
325 S13).
326 Among metabolic traits, we identified several loci that are associated with the
327 contents of L-Glutamine, L-Tyrosine, L-Tryptophane, L-Serine, Chlorogenic acid and
328 IAA. Most notably, several SNPs located within the *light harvesting protein 5*
329 (*SpLHCB5*), which functions as a structural component of photosystem II (de Bianchi
330 et al. 2008) were associated with contents of L-Glutamine, L-Valine, L-Tryptophan and
331 L-Tyrosine (Figure 3 A-C, Table S13). Further, a SNP located upstream of a *U-box*
332 *containing protein 4* (*SpPUB4*), which functions as a regulator of cell division
333 processes in meristematic regions (Kinoshita et al. 2015), was associated with L-
334 Tyrosine content. Three SNPs associated with the contents of L-Tyrosine, Chlorogenic
335 acid and IAA were linked to *SpUGT89B1*, *SpCYP71AU50* and *SpMYBC1*, three genes
336 functioning in secondary metabolite biosynthesis (Yamaguchi et al. 2014; Caputi et al.
337 2012; Ke et al. 2021) (Figure 3 B, D, E). A SNP within *SpDPE2*, which is functionally

involved in starch metabolism is associated with IAA content (Li et al. 2022) (Figure 3E). For SVs, contents of L-Glutamine and L-Serine were associated with a 94 bp intronic deletion in the *Ubiquitin-carboxyterminal hydrolase 7 SpUBP7* (Figure 4A-C). Homologs of *SpUBP7* were shown to stabilize ubiquitin upon proteasome binding and regulate proteasomal activity (Leggett et al. 2002; Wu et al. 2019).The presence of the intronic deletion was associated with increased contents of L-Glutamine and L-Serine (Figure 4D and E) and an increased expression of *SpUBP7* in roots (Figure 4F and G). These findings suggest that biomass and metabolite contents are coordinated jointly through photosynthesis, starch metabolism, cell division, secondary metabolite biosynthesis and protein degradation.

Discussion

Using targeted metabolomics on 137 *S. polyrhiza* genotypes, we identified strong positive correlations between chemically related metabolites. Additionally, candidate genes involved in photosynthesis (*SpLHCB5*), protein degradation (*SpUBP7*) and organ development (*SpLOB*, *SpAGL62*) were repeatedly associated with different metabolite levels, suggesting strong genetic co-regulation. Since levels of free amino acids and secondary metabolites correlated with plant biomass in many cases, some of these candidate genes are of potential use for joint optimization of plant growth and nutrient level.

Free amino acids have large pool sizes and turnover rates in duckweed and were shown to serve as major carbon and nitrogen sources for starch and storage protein production (Evans et al. 2018). Previous work showed that supplementation of growth medium with certain amino acids like L-Glutamine was associated with increased protein content and biomass in duckweed (Shi et al. 2023). In agreement with these observations, the associations between free amino acid levels and biomass

364 are mostly positive. Compared to free amino acids, phenylpropanoids and
365 phytohormones are characterized by a higher functional diversity. In *S. polyrhiza*,
366 anthocyanins were shown to alleviate Cr(VI) stress (Oláh et al. 2009), whereas
367 luteolins and apigenins are involved in stress responses to copper and UV-light,
368 respectively (Böttner et al. 2021). Previous studies found the contents of these three
369 flavonoid classes to be inversely correlated with biomass in *S. polyrhiza* (Böttner et al.
370 2021; Oláh et al. 2009) and explained this by increased allocation of carbon resources
371 to flavonoid biosynthesis (Oláh et al. 2009). Compared to this, our study revealed a
372 more divers association pattern of flavonoid levels with biomass. While the contents of
373 Sinapic acid, luteolins and Cyanidin-3O-glycoside were negatively associated with
374 biomass, contents of Apigenin-8-C-glycoside and Chlorogenic acid showed positive
375 correlations with biomass. In other plant species, changes in profiles of glycosylated
376 flavonoids were shown to regulate plant growth by influencing auxin transport (Ringli
377 et al. 2008), possibly explaining different association patterns of glycosylated luteolins
378 and apigenins to plant biomass. Contents of chlorogenic acid were associated with
379 increased leaf growth and increased contents of antioxidants and proteins, suggesting
380 a regulatory function in plant growth (Zhang et al. 2024). The overall positive
381 associations of cytokinin levels with most phenylpropanoids, were previously explained
382 by their function in coordinating resource allocation to defense metabolism in plant
383 development (Brütting et al. 2017). For phytohormones, levels of JA-Ile were positively
384 correlated with biomass. In duckweeds, jasmonates show bimodal dose-response
385 patterns, with small concentrations stimulating growth and turion germination and high
386 concentrations delaying growth (Appenroth et al. 1991; Piotrowska et al. 2010). Since
387 metabolites were measured in unstressed plant material, JA contents likely never
388 exceeded constitutive levels, which would be required for significant growth inhibition.

389 Together these correlation patterns suggest that *S. polyrhiza* can be simultaneously
390 optimized for higher biomass and increased contents of certain phenylpropanoids.
391 On a genetic level, our study identified genetic loci that are associated with most
392 secondary metabolite and phytohormone contents, whereas for many free amino
393 acids, no significant associations were found. For growth rate and biomass, we did not
394 find significant associations for any markers located within the genic region, likely due
395 to their highly polygenic feature. Previous GWAS conducted on complex traits
396 highlighted, that the success in identifying genetic associations greatly depends on
397 sample size and phenotypic variation (Schizophrenia Working Group of the Psychatric
398 Genomics Consortium 2014; Duncan et al. 2019; Chen et al. 2014). In line with
399 previous reports (Chen et al. 2016; Pham et al. 2023), we found that secondary
400 metabolite levels showed the highest average phenotypic variation across all
401 metabolic traits, whereas variation for growth parameters was lowest. For secondary
402 metabolites, high variation in tissue concentration was explained by their low trait
403 complexity (Pham et al. 2023; Haghi et al. 2022; Chen et al. 2014). Whereas primary
404 metabolites and growth seem to be regulated by many small effect loci, secondary
405 metabolites are often controlled by just a few major-effect loci (Chen et al. 2014; Chen
406 et al. 2016). Since small effect loci are often diffusely distributed across the genome,
407 GWAS on primary metabolite levels and growth often require large sample sizes to
408 detect them (Duncan et al. 2019). Compared to other association studies on plant
409 metabolic traits (Chen et al. 2014; Chen et al. 2016; Cu et al. 2021; Angelovici et al.
410 2013), our GWAS suffered from a low sample size, likely explaining the lack of genetic
411 associations for growth traits and several metabolite contents.
412 On a physiological level, differences in growth were correlated with the
413 photosynthesis rate for duckweed (Sree et al. 2015). Connected to this, the metabolism
414 of storage molecules plays a major role in plant growth. Interestingly, many biomass-

415 correlated metabolites were associated with genes influencing photosynthetic
416 efficiencies (*SpLHCB6*), protein turn over (*SpUBP7*), starch metabolism (*SpDPE*) or
417 cell-cycle control (*SpPUB4*). These genes provide suitable candidates for future
418 reverse genetic studies, that will be required for further elucidation of their exact
419 function in duckweed metabolism. Taken together, this study provides insights into the
420 molecular basis of growth regulation and metabolite homeostasis in *S. polyrhiza*.
421 Through the identification of gene candidates associated with metabolic traits, this
422 study contributes to laying the foundations for further optimization strategies in *S.*
423 *polyrhiza*.
424

425 **Acknowledgement**

426 We thank Marie Serwaty-Sárazová for maintaining the giant duckweed collection.
427 This project was funded by the German Research Foundation (project numbers
428 427577435 to SX). The LC-MS/MS instrument was funded by the German Research
429 Foundation (project number 435681637 to SX).
430

431 **Main Tables**

432 Table 1: Phenotypic variation and heritability of fitness parameters in *S. polyrhiza*

Fitness trait	Mean	SE	Range		Fold change	H2.c
			Minimum	Maximum		
Fresh weight	227.7 mg	7.3 mg	71.6 mg	507.7 mg	10.4	0.865
Dry weight	16.67 mg	0.5 mg	3.4 mg	35.2 mg	10.4	0.602
RGR of frond area	0.323 d^{-1}	0.004 d^{-1}	0.165 d^{-1}	0.401 d^{-1}	2.4	0.841
RGR of frond number	0.259 d^{-1}	0.003 d^{-1}	0.152 d^{-1}	0.328 d^{-1}	2.2	0.802

433 Abbreviations: SE - standard error, H2.c – broad sense heritability according to Cullis
434 et al. (2006)

Main Figures

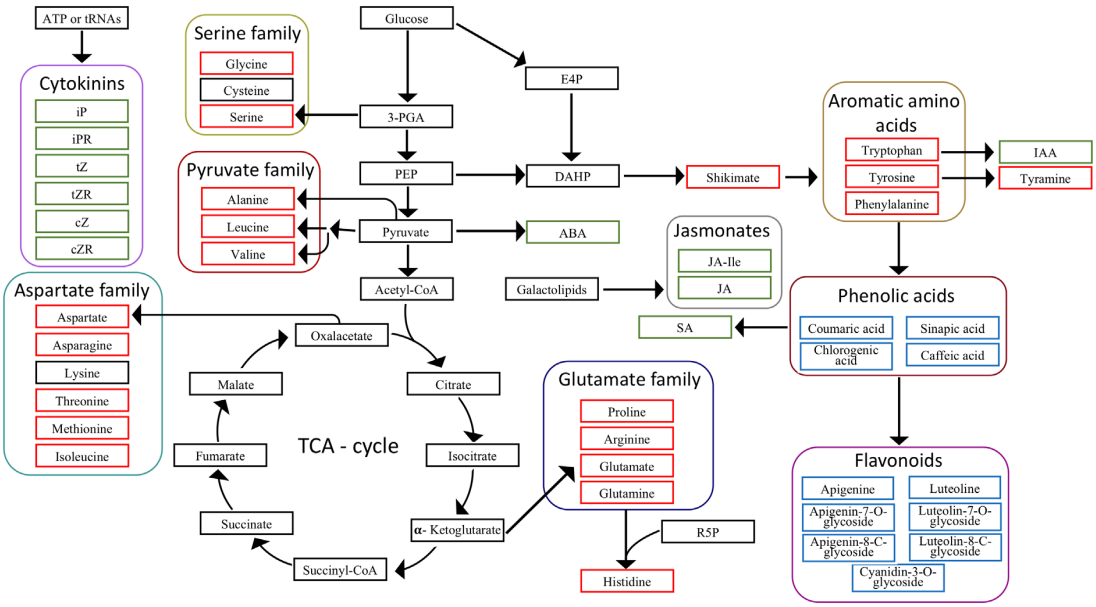


Figure 1: Overview of the biosynthesis of metabolites analyzed in this study. Metabolites were connected to their biochemical precursors with arrows. All quantified compounds were sorted into the three categories amino acid metabolism, specialized metabolism and phytohormones and shown in boxes colored red, blue, and green, respectively. Direct degradation products and biosynthetic precursors of amino acids such as tyramine and shikimic acid were grouped together with all quantified amino acids in the group amino acid metabolism. Amino acids were subcategorized into five families based on their last common precursor (serine, pyruvate, aspartate and glutamate) or on common chemical properties (aromatic amino acids). All specialized metabolites quantified in this study are phenylpropanoids derived from phenylalanine. Phenolic acids are also precursors of flavonoids. In contrast to amino acids and secondary metabolites, compounds with a major function in signalling were characterized as phytohormones. Chemically derived metabolites with the same

452 precursors were further grouped in biosynthetic families, illustrated by frames of
453 different colors.
454 Abbreviations: TCA – tricarboxylic acid, iP – Isopentenyladenine, iPR –
455 Isopentenyladenine riboside, tZ – trans-Zeatin, cZ – cis-Zeatin, tZR – trans-Zeatin
456 riboside, cZR – cis-Zeatin riboside, 3-PGA – 3-Phosphoglycerate, PEP –
457 Phosphoenolpyruvate, CoA - Coenzyme A, E4P – Erythrose-4-phosphate, R5P –
458 Ribose-5-phosphate, DAHP – 3-Deoxyarabinoheptulonate-7-phosphate, JA –
459 Jasmonic acid, JA-Ile – Jasmonic acid-Isoleucine conjugate, IAA – Indole-3-acetic
460 acid, SA – Salicylic acid, ABA – Abscisic acid
461
462

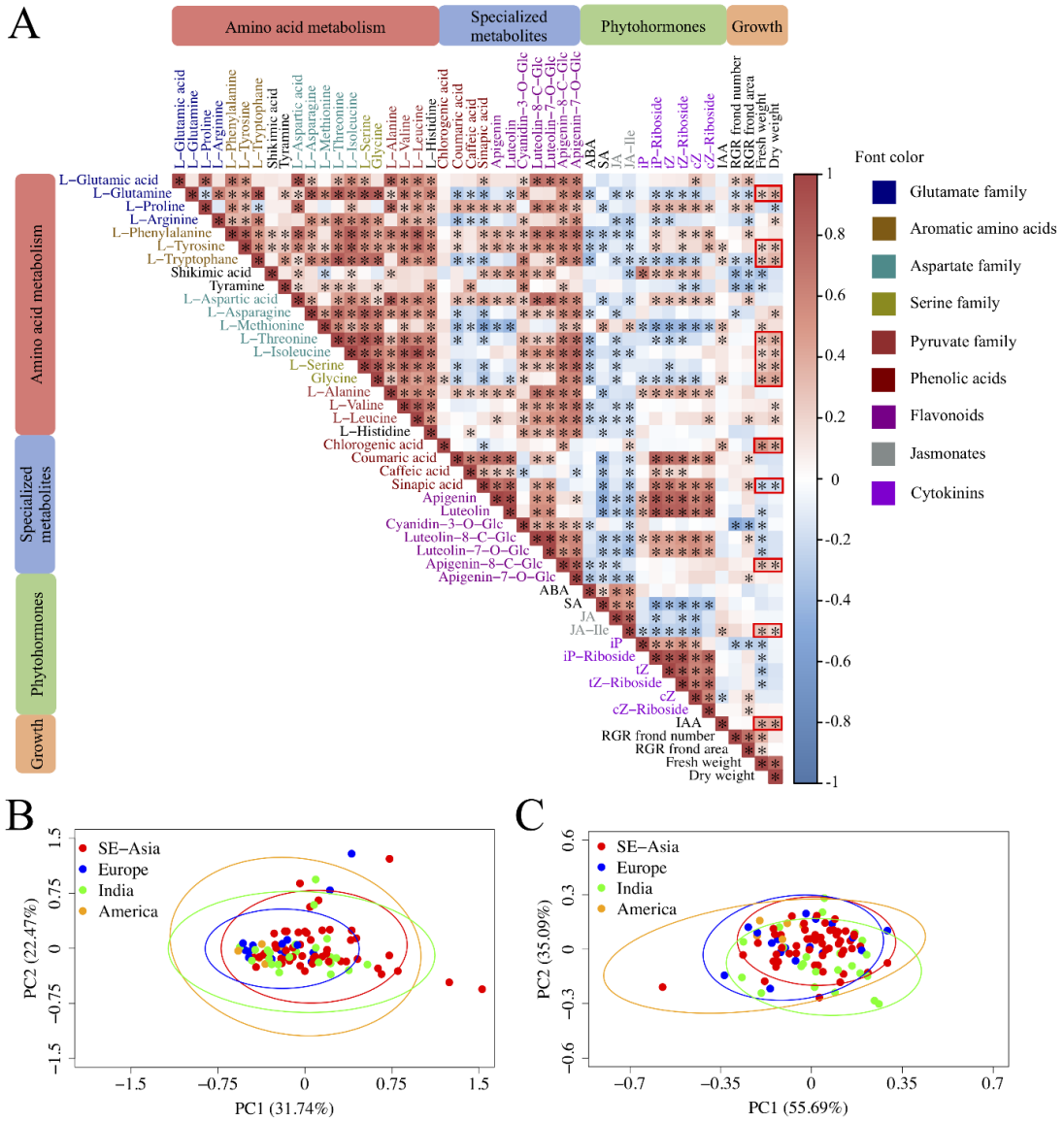
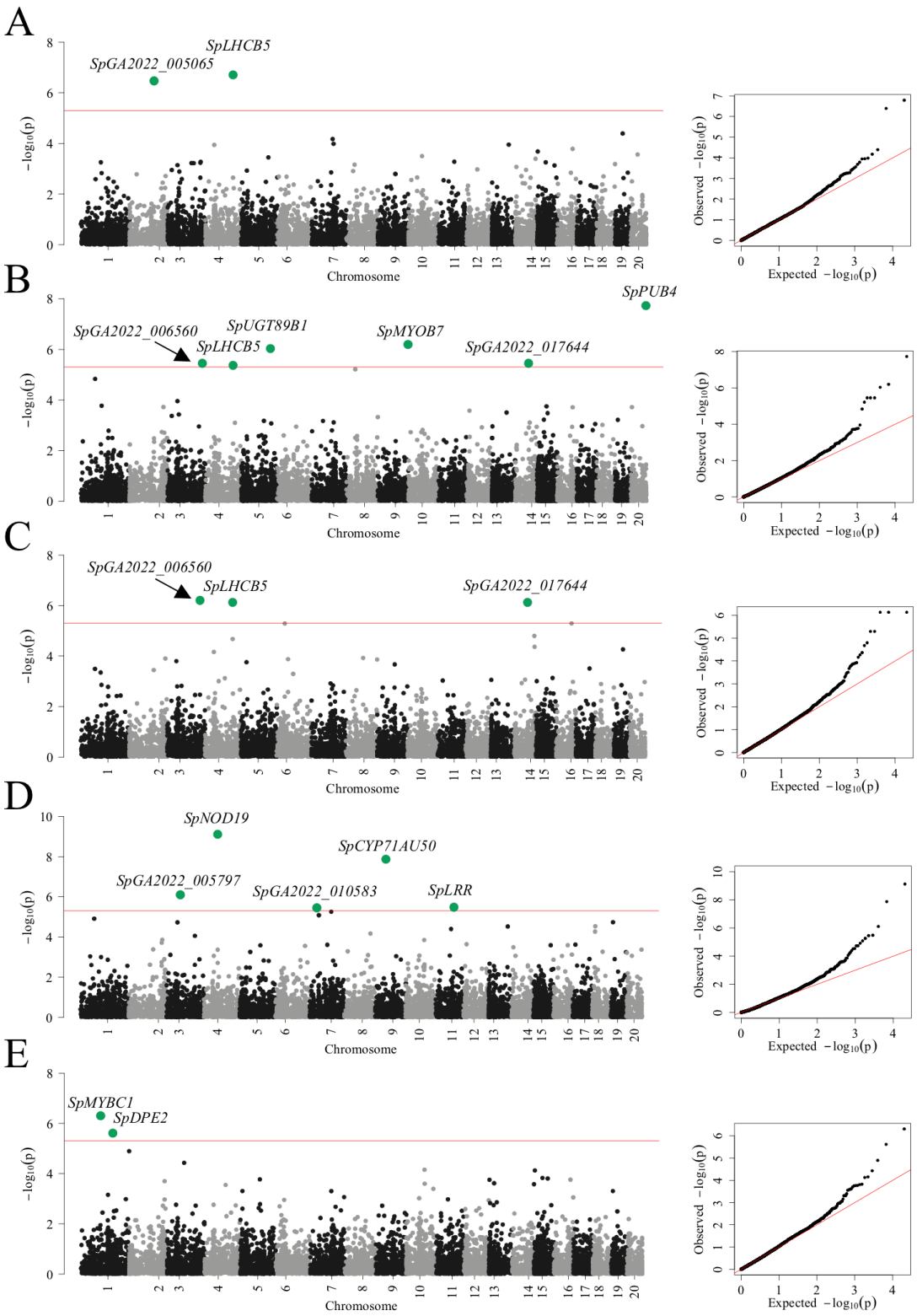


Figure 2: Correlation matrix (A) and principal coordinate analyses (PCA) on growth parameters (B) and metabolite content (C). A: Tissue concentrations of amino acids showed strong positive correlation patterns. In most cases, contents of amino acids were positively correlated with flavonoid content and biomass. Metabolite concentrations and growth rates were measured for 137 genotypes. All metabolites were categorized accordheing to Figure 1. The strength of the correlation was estimated using tPearson correlation coefficient and is reflected by the intensities of

472 red color for positive and blue color for negative correlations. Asterisks highlight
473 significant correlations (*F*-test, $P < 0.05$). Simultaneous correlations of dry and fresh
474 weight were highlighted with red margins. B and C: population differences on growth
475 (B) and metabolisms (C). All PCAs were conducted on 97 representative genotypes
476 assigned to four previously defined genetic populations (Xu et al. 2019). The ellipses
477 show 95 % confidence intervals calculated based on standard deviation. The color of
478 the ellipses corresponds to the genetic populations. Abbreviations: iP –
479 Isopentenyladenine, iPR – Isopentenyladenine riboside, tZ – trans-Zeatin, cZ – cis-
480 Zeatin, tZR – trans-Zeatin riboside, cZR – cis-Zeatin riboside, JA – Jasmonic acid, JA-
481 Ile – Jasmonic acid-Isoleucine conjugate, IAA – Indole-3-acetic acid, SA – Salicylic
482 acid, ABA – Abscisic acid
483
484

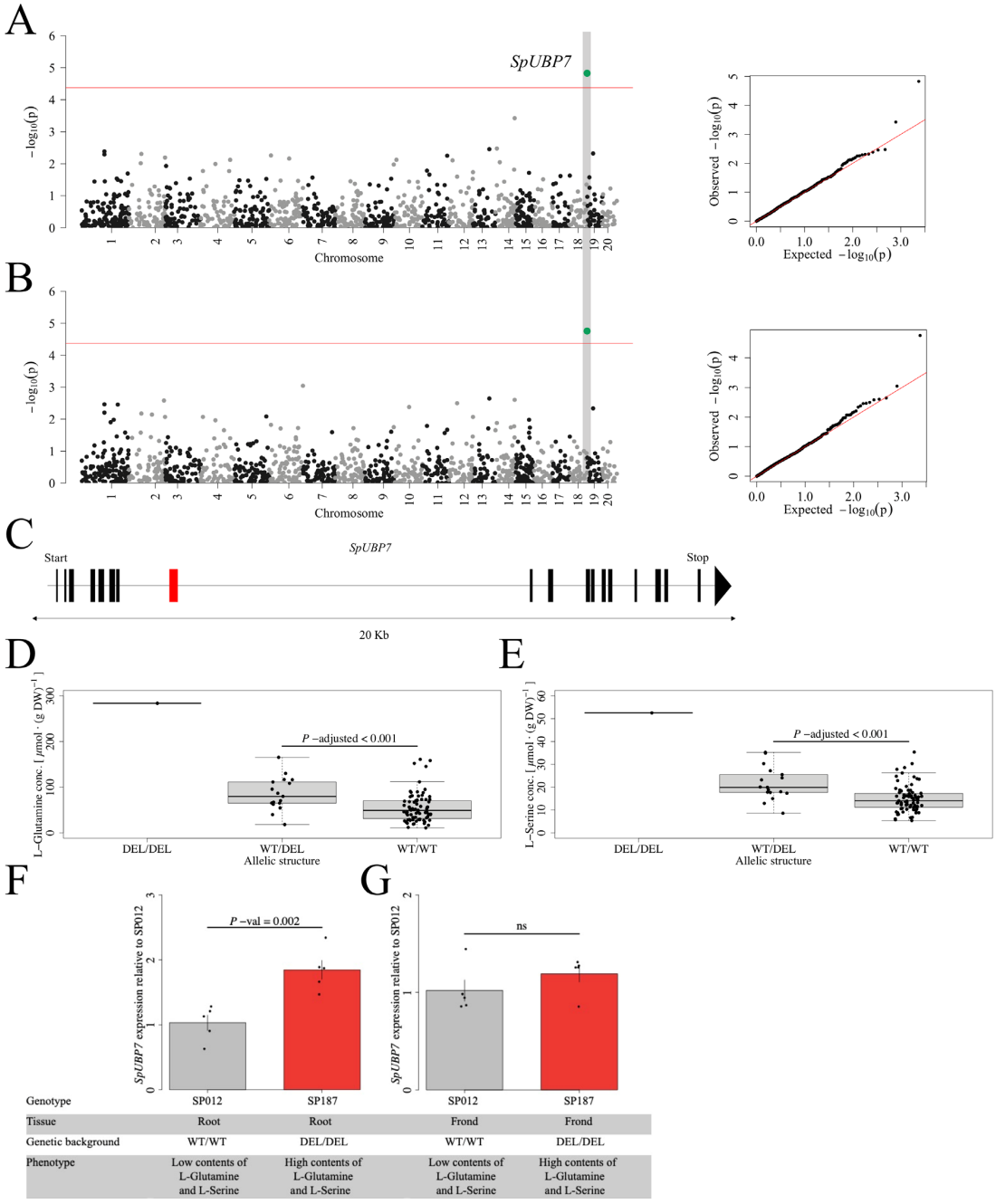


485

486 **Figure 3: Manhattan plots of SNP-based GWAS for free amino acids.** Each plot

487 shows the GWAS result on the concentrations of L-Glutamine (A), L-Tyrosine (B), L-

488 Tryptophan (C) and Chlorogenic acid (D) and Indole-3-acetic acid (IAA) (E) with their
489 corresponding QQ-plots on the right side. The Bonferroni corrected *P*-value = 0.05
490 (Wald-test) at $4.97 \cdot 10^{-6}$ was used as significance threshold and is shown as red line.
491 All significant markers are labelled with the names of gene candidates.
492



493

Figure 4: The presence of an intronic deletion in *SpUBP7* is associated with increased contents of L-Glutamine and L-Serine. A and B: The SV-based GWAS on L-Glutamine (A) and L-Serine (B). The red line represents the Bonferroni corrected *P*-value = 0.05 (Wald-test) at $4.23 \cdot 10^{-5}$. C: The associated 94 bp deletion (red bar) is located within the 7th intron of *SpUBP7*. D and E: The presence of the deletion was associated with increased contents in L-Glutamine (D) and L-Serine (E). F: *SpUBP7* shows genotype-dependent expression differences on root level. SP187, a genotype that is homozygous for the deletion, shows enhanced *SpUBP7* expression compared to genotype SP012 that lacks the deletion. G: No significant differential expression was observed between SP187 and SP012 in the fronds.

References

Acosta K, Appenroth KJ, Borisjuk L, Edelman M, Heinig U, Jansen MAK, Oyama T, Pasaribu B, Schubert I, Sorrels S, Sree KS, Xu S, Michael TP, Lam E (2021) Return of the Lemnaceae: duckweed as a model plant system in the genomics and postgenomics era. *The Plant Cell*, **33**, 3207-3234. https://doi.org/10.1093/plcell/koab189

Aftab T, Masroor M, Khan A, Idrees M, Naeem M, Moinuddin (2010) Salicylic acid acts as potent enhancer of growth, photosynthesis and artemisinin production in Artemisia annua L. *Journal of Crop Science and Biotechnology*, **13**, 183-188. https://doi.org/10.1007/s12892-010-0040-3

Angelovici R, Batushansky A, Deason N, Gonzalez-Jorge S, Gore MA, Fait A, DellaPenna D (2017) Network-Guided GWAS Improves Identification of Genes Affecting Free Amino Acids. *Plant Physiology*, **173**, 872-886. https://doi.org/10.1104/pp.16.01287

Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin H, Cepela J, Buell R, Gore MA, Dellapenna D (2013) Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. *The Plant Cell*, **25**, 4827-4843. https://doi.org/10.1105/tpc.113.119370

Appenroth KJ (2015) Media for in vitro-cultivation of duckweed. *Duckweed Forum* **3**, 180-186.

Appenroth KJ, Dathe W, Hertel W, Augsten H (1991) Photophysiology of Turion Germination in Spirodela polyrhiza (L.) SCHLEIDEN. VII. Action of Jasmonic acid. *Journal of Plant Physiology*, **138**, 345-349. https://doi.org/10.1016/S0176-1617(11)80298-1

Baek G, Saeed M, Choi H-K (2021) Duckweeds: their utilization, metabolites and cultivation. *Applied Biological Chemistry*, **64**, 73. https://doi.org/10.1186/s13765-021-00644-z

Bates D, Mächler M, Bolker B, Walker S (2015) Fitting Linear Mixed-Effect Models Using lme4. *Journal of Statistical Software*, **67**, 1-48, https://doi.org/10.18637/jss.v067.i01

Böttner L, Grabe V, Gablenz S, Böhme N, Appenroth KJ, Gershenzon J, Huber M (2021) Differential localization of flavonoid glucosides in an aquatic plant implicates different functions under abiotic stress. *Plant, Cell & Environment*, **44**, 900-914. https://doi.org/10.1111/pce.13974

Browning BL, Zhou Y, Browning SR (2018) A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, **103**, 338-348. https://doi.org/10.1016/j.ajhg.2018.07.015

Brütting C, Schäfer M, Vanková R, Gase K, Baldwin IT, Meldau S (2017) Changes in cytokinins are sufficient to alter developmental patterns of defense metabolites in Nicotiana attenuata. *The Plant Journal*, **89**, 15-30. https://doi.org/10.1111/tpj.13316

Caputi L, Malnoy M, Goremykin V, Nikiforova S, Martens S (2012) A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. *The Plant Journal*, **69**, 1030-1042. https://doi.org/10.1111/j.1365-313X.2011.04853.x

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, Zhang W, Zhang L, Yu S, Wang G, Lian X, Luo J (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nature Genetics*, **46**, 714-721. https://doi.org/10.1038/ng.3007

Chen W, Wang W, Peng M, Gong L, Gao Y, Wan J, Wang S, Shi L, Zhou B, Li Z, Peng X, Yang C, Qu L, Liu X, Luo J (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. *Nature Communications*, **7**, 12767. https://doi.org/10.1038/ncomms12767

Cu ST, Warnock NI, Pasuquin J, Dingkuhn M, Stangoulis J (2021) A high-resolution genome-wide association study of the grain ionome and agronomic traits in rice Oryza sativa subsp. indica. *Scientific Reports*, **11**, 19230. https://doi.org/10.1038/s41598-021-98573-w

Cullis BR, Smith AB, Coombes NE (2006) On the Design of Early Generation Variety Trials with Correlated Data. *Journal of Agricultural, Biological and Environmental Statistics*, **11**, 381-393. https://doi.org/10.1198/108571106X154443

de Bianchi S, Dall'Osto L, Tognon G, Morosinotto T, Bassi R (2008) Minor antenna proteins CP24 and CP26 affect the interactions between photosystem II subunits and the electron transport rate in grana membranes of Arabidopsis. *The Plant Cell*, **20**, 1012-1028. https://doi.org/10.1105/tpc.107.055749

Dick R, Rattei T, Haslbeck M, Schwab W, Gierl A, Frey M (2012) Comparative Analysis of Benzoxazinoid Biosynthesis in Monocots and Dicots: Independent Recruitment of Stabilization and Activation Functions. *The Plant Cell*, **24**, 915-928. https://doi.org/10.1105/tpc.112.096461

Duncan LE, Ostacher M, Ballon J (2019) How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. *Neuropsychopharmacology*, **44**, 1518-1523. https://doi.org/10.1038/s41386-019-0389-5

Evans EM, Freund DM, Sondervan VM, Cohen JD, Hegeman AD (2018) Metabolic Patterns in Spirodela polyrhiza Revealed by (15)N Stable Isotope Labeling of Amino Acids in Photoautotrophic, Heterotrophic, and Mixotrophic Growth Conditions. *Frontiers in Chemistry*, **6**, 191. https://doi.org/10.3389/fchem.2018.00191

Fox J, Weisberg S (2019) An R Companion to Applied Regression. Third edition, SAGE Publications, Thousand Oaks, CA, USA

Haghi R, Ahmadikhah A, Fazeli A, Shariati V (2022) Candidate genes for anthocyanin pigmentation in rice stem revealed by GWAS and whole-genome resequencing. *The Plant Genome*, **15**, e20224. https://doi.org/10.1002/tpg2.20224

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology*, **8**, R19. https://doi.org/10.1186/gb-2007-8-2-r19

Höfer M, Schäfer M, Wang Y, Wink S, Xu S (2024a) Data from: Genetic mechanism of non-targeted-site resistance to diquat in Spirodela polyrhiza. https://doi.org/10.5061/dryad.2fqz612ww

Höfer M, Schäfer M, Wang Y, Wink S, Xu S (2024b) Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in Spirodela polyrhiza. *Plants*, **13**, 845. https://doi.org/10.3390/plants13060845

Hui T, Zhang Y, Jia R, Hu Y, Wang W, Wang Y, Wang Y, Zhu Y, Yang L, Xiang B (2023) Metabolomic analysis reveals responses of Spirodela polyrhiza L. to salt stress. *Journal of Plant Interactions*, **18**. https://doi.org/10.1080/17429145.2023.2210163

Huot B, Yao J, Montgomery BL, He SY (2014) Growth–Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness. *Molecular Plant*, **7**, 1267-1287. https://doi.org/10.1093/mp/ssu049

Kassambara A (2023) ggcorrplot: Visualization of a Correlation Matrix using 'ggplot2'.

Ke Y, Abbas F, Zhou Y, Yu R, Fan Y (2021) Auxin-Responsive R2R3-MYB Transcription Factors HcMYB1 and HcMYB2 Activate Volatile Biosynthesis in Hedychium coronarium Flowers. *Frontiers in Plant Science*, **12**, 710826. https://doi.org/10.3389/fpls.2021.710826

Kinoshita A, ten Hove CA, Tabata R, Yamada M, Shimizu N, Ishida T, Yamaguchi K, Shigenobu S, Takebayashi Y, Iuchi S, Kobayashi M, Kurata T, Wada T, Seo M, Hasebe M, Blilou I, Fukuda H, Scheres B, Heidstra R, Kamiya Y, Sawa S (2015) A plant U-box protein, PUB4, regulates asymmetric cell division and cell proliferation in the root meristem. *Development*, **142**, 444-453. https://doi.org/:10.1242/dev.113167

Leggett DS, Hanna J, Borodovsky A, Crosas B, Schmidt M, Baker RT, Walz T, Ploegh H, Finley D (2002) Multiple associated proteins regulate proteasome structure and function. *Molecular Cell*, **10**, 495-507. https://doi.org/10.1016/s1097-2765(02)00638-x

Lemay M-A, Malle S (2022) A Practical Guide to Using Structural Variants for Genome-Wide Association Studies. In: Torkamaneh D, Belzile F (eds) Genome-Wide Association Studies. Springer US, New York, NY, USA, pp 161-172. https://doi.org/10.1007/978-1-0716-2237-7_10

Lemon J (2006) Plotrix: a package in the red light distric of R. *R-News*, **6**, 8-12.

Li J, Halitschke R, Li D, Paetz C, Su H, Heiling S, Xu S, Baldwin IT (2021) Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. *Science*, **371**, 255-260. https://doi.org/10.1126/science.abe4713

Li X, Apriyanto A, Castellanos JF, Compart J, Muntaha SN, Fettke J (2022) Dpe2/phs1 revealed unique starch metabolism with three distinct phases characterized by different starch granule numbers per chloroplast, allowing insights into the control mechanism of granule number regulation by gene co-regulation and metabolic profiling. *Frontiers in Plant Science*, **13**. https://doi.org/10.3389/fpls.2022.1039534

Ligges U, Mächler M (2003) Scatterplot3d – an R Package for Visualizing Multivariate Data. *Journal of Statistical Software*, **8**, 1-20. https://doi.org/10.18637/jss.v008.i11

Noctor G, Foyer CH (1998) A re-evaluation of the ATP :NADPH budget during C3 photosynthesis: a contribution from nitrate assimilation and its associated respiratory activity? *Journal of Experimental Botany*, **49**, 1895-1908. https://doi.org/10.1093/jxb/49.329.1895

Oksanen J, Simpson GL, Blanchet GF, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, Wagner H, Barbour M, Bedward M, Bolker B, Borcard D, Carvalho G, Chirico M, De Caceres M, Durand S, Evangelista HBAF, Rich, Friendly M, Furneaux B, Hannigan G, Hill MO, Lahti L, McGlinn D, Ouellette M-H, Ribeiro Cunha E, Smith T, Stier A, Ter Braak CJF, Weedon J (2022) vegan: Community Ecology Package.

Oláh V, Combi Z, Szöllösi E, Kanalas P, Mészáros I (2009) Anthocyanins: Possible

antioxidants against Cr (VI) induced oxidative stress in Spirodela polyrrhiza. *Cereal Research Communications*, **37**, 533-536. https://doi.org/10.1556/CRC.37.2009.Suppl.4

Pham CH, Do TD, Nguyen HTL, Hoang NT, Tran TD, Vu MTT, Doi HH, Bui T-GT, Henry RJ (2023) Genome-wide association mapping of genes for anthocyanin and flavonoid contents in Vietnamese landraces of black rice. *Euphytica*, **220**, 11. https://doi.org/10.1007/s10681-023-03268-0

Piotrowska A, Bajguz A, Czerpak R, Kot K (2010) Changes in the Growth, Chemical Composition, and Antioxidant Activity in the Aquatic Plant Wolffia arrhiza (L.) Wimm. (Lemnaceae) Exposed to Jasmonic Acid. *Journal of Plant Growth Regulation*, **29**, 53-62. https://doi.org/10.1007/s00344-009-9113-8

Ringli C, Bigler L, Kuhn BM, Leiber R-M, Diet A, Santelia D, Frey B, Pollmann S, Klein M (2008) The Modified Flavonol Glycosylation Profile in the Arabidopsis rol1 Mutants Results in Alterations in Plant Growth and Cell Shape Formation. *The Plant Cell*, **20**, 1470-1481. https://doi.org/10.1105/tpc.107.053249

Schäfer M, Brütting C, Baldwin IT, Kallenbach M (2016) High-throughput quantification of more than 100 primary- and secondary-metabolites, and phytohormones by a single solid-phase extraction based sample preparation with analysis by UHPLC–HESI–MS/MS. *Plant Methods*, **12**, 30. https://doi.org/10.1186/s13007-016-0130-x

Schizophrenia Working Group of the Psychatric Genomics Consortium (2014) Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, **511**, 421-427. https://doi.org/10.1038/nature13595

Shi H, Ernst E, Heinzel N, McCorkle S, Rolletschek H, Borisjuk L, Ortleb S, Martienssen R, Shanklin J, Schwender J (2023) Mechanisms of metabolic adaptation in the duckweed Lemna gibba: an integrated metabolic, transcriptomic and flux analysis. *BMC Plant Biology*, **23**, 458. https://doi.org/10.1186/s12870-023-04480-9

Smith DC, Bassham JA, Kirk M (1961) Dynamics of the photosynthesis of carbon compounds II. Amino acid synthesis. *Biochimica et Biophysica Acta*, **48**, 299-313. https://doi.org/10.1016/0006-3002(61)90478-4

Smith KE, Schäfer M, Lim M, Robles-Zazueta CA, Cowan L, Fisk ID, Xu S, Murchie EH (2024) Aroma and metabolite profiling in duckweeds: Exploring species and ecotypic variation to enable wider adoption as a food crop. *Journal of Agriculture and Food Research*, **18**, 101263. https://doi.org/10.1016/j.jafr.2024.101263

Sree KS, Sudakaran S, Appenroth K-J (2015) How fast can angiosperms grow? Species and clonal diversity of growth rates in the genus Wolffia (Lemnaceae). *Acta Physiologiae Plantarum*, **37**, 204. https://doi.org/10.1007/s11738-015-1951-3

Stacklies W, Redestig H, Scholz M, Walther D, Selbig J (2007) pcaMethods—a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, **23**, 1164-1167. https://doi.org/10.1093/bioinformatics/btm069

Vogt F, Shirsekar G, Weigel D (2021) vcf2gwas-python API for comprehensive GWAS analysis using GEMMA. *Bioinformatics*, **38**, 839-840. https://doi.org/10.1093/bioinformatics/btab710

Wang Y, Duchen P, Chávez A, Sree KS, Appenroth KJ, Zhao H, Höfer M, Huber M, Xu S (2024) Population genomics and epigenomics of Spirodela polyrhiza provide insights into the evolution of facultative asexuality. *Communications Biology*, **7**, 581. https://doi.org/10.1038/s42003-024-06266-7

Wu R, Zheng W, Tan J, Sammer R, Du L, Lu C (2019) Protein partners of plant ubiquitin-specific proteases (UBPs). *Plant Physiology and Biochemistry*, **145**, 227-236. https://doi.org/10.1016/j.plaphy.2019.08.032

Xu S, Stapley J, Gablenz S, Boyer J, Appenroth KJ, Sree KS, Gershenzon J, Widmer A, Huber M (2019) Low genetic variation is associated with low mutation rate in the giant duckweed. *Nature Communications*, **10**, 1243. https://doi.org/10.1038/s41467-019-09235-5

Yamaguchi T, Yamamoto K, Asano Y (2014) Identification and characterization of CYP79D16 and CYP71AN24 catalyzing the first and second steps in l-phenylalanine-derived cyanogenic glycoside biosynthesis in the Japanese apricot, Prunus mume Sieb. et Zucc. *Plant Molecular Biology*, **86**, 215-223. https://doi.org/10.1007/s11103-014-0225-6

Zhang M, Xiao Q, Li Y, Tian Y, Zheng J, Zhang J (2024) Exploration of exogenous chlorogenic acid as a potential plant stimulant: enhancing physiochemical properties in Lonicera japonica. *Physiology and Molecular Biology of Plants*, **30**, 453-466. https://doi.org/10.1007/s12298-024-01435-8

Zhang P, Zhong K, Zhong Z, Tong H (2019) Genome-wide association study of important agronomic traits within a core collection of rice (Oryza sativa L.). *BMC Plant Biology*, **19**, 259. https://doi.org/10.1186/s12870-019-1842-7

Zhao Y, Fang Y, Jin Y, Huang J, Bao S, Fu T, He Z, Wang F, Wang M, Zhao H (2015) Pilot-scale comparison of four duckweed strains from different genera for potential application in nutrient recovery from wastewater and valuable biomass production. *Plant Biology*, **17**, 82-90. https://doi.org/10.1111/plb.12204

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821-824. https://doi.org/10.1038/ng.2310

Ziegler P, Adelmann K, Zimmer S, Schmidt C, Appenroth KJ (2015) Relative in vitro growth rates of duckweeds (Lemnaceae) – the most rapidly growing higher plants. *Plant Biology*, **17**, 33-41. https://doi.org/10.1111/plb.12184

719 Živanović B, Milić Komić S, Tosti T, Vidović M, Prokić L, Veljović Jovanović S (2020a)
720 Leaf Soluble Sugars and Free Amino Acids as Important Components of Abscisic Acid-
721 Mediated Drought Response in Tomato. *Plants*, **9**, 9.
722 https://doi.org/10.3390/plants9091147
723 Živanović B, Milić Komić S, Tosti T, Vidović M, Prokić L, Veljović Jovanović S
724 (2020b) Leaf Soluble Sugars and Free Amino Acids as Important Components of
725 Abscisic Acid—Mediated Drought Response in Tomato. *Plants*, **9**, 9.
726 https://doi.org/10.3390/plants9091147
727 Züst T, Agrawal AA (2017) Trade-Offs Between Plant Growth and Defense Against
728 Insect Herbivory: An Emerging Mechanistic Synthesis. *Annual Review of Plant Biology*,
729 **68**, 513-534. https://doi.org/10.1146/annurev-arplant-042916-040856
730 Züst T, Joseph B, Shimizu KK, Kliebenstein DJ, Turnbull LA (2011) Using knockout
731 mutants to reveal the growth costs of defensive traits. *Proceedings of the Royal Society
732 B: Biological Sciences*, **278**, 2598-2603. https://doi.org/10.1098/rspb.2010.2475
733
734

735 **Data availability statement**

736 All raw data and scripts used for analysis can be found under:
737 https://doi.org/10.17632/xwsxpfcysd.1 or
738 https://data.mendeley.com/preview/xwsxpfcysd?a=2d1b7639-12ef-4665-bc72-
739 5096a149004e

740

741

742 **Funding**

743 This research was funded by the "Deutsche Forschungsgemeinschaft", grant
744 number 427577435 to SX.

745

746 **Author contributions**

747 MH analyzed the data. MH, MS, SW and YW performed the experiments. YW and
748 SX provided resources and technical infrastructure. SX and MH conceived and
749 supervised the project. MH wrote the manuscript. All authors contributed to the final
750 version of the manuscript.

751

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

3.3 Population Genomics and Epigenomics of *Spirodela polyrhiza* Provide Insights into the Evolution of Facultative Asexuality

3.3.1 Summary

The focus of this study lies in the characterization of the genomic/epigenomic diversity, population structure and migration history of *Spirodela polyrhiza*.

The genomes of 228 sequenced genotypes of *S. polyrhiza* were used to identify SNPs, insertions/deletions (< 50bp) and structure variations (> 50bp). CpG, CHH and CHG - methylation levels were determined through bisulfite sequencing of five genotypes from each genetic population of *S. polyrhiza*. SNPs were used to infer of population structure via PCA and phylogeny, applying a maximum-likelihood approach. The demographic history of the genetic populations of *S. polyrhiza* was predicted using an Approximate Bayesian Computation (ABC) approach. A genome-wide selection scan was used to discover genomic regions under positive selection.

This study identified a total of 1241981 SNPs, 166075 insertions/deletions and 3205 structure variations. The identified genome-wide nucleotide diversity of *S. polyrhiza* was 0.0016. Further, *S. polyrhiza* was shown to have the lowest average species-wide methylation rate among all studied angiosperms. Analysis of population structure identified four genetic populations, which were largely defined by their geographic locations as American, SE-Asian, Indian and European populations. Approximate bayesian computation modelling predicts that American and SE-Asian populations are derived from a common ancestral population. Indian and European populations later diverged from the SE-Asian population. Populations differed regarding their linkage-disequilibrium decay and heterozygosity rate, suggesting differences in rates of sexual reproduction. Different levels of heterozygosity were associated with two MADS-box genes: *AGL62* and *SOC1*, whose orthologs control the development of floral organs. A genome-wide scan identified 69 genes whose homologs function predominantly in processes such as gametogenesis, embryogenesis, flowering time and organ development to be under selection. Many genes involved in reproduction and organ development, like *BB*, *FLK*, *AGL6* and *AP3* were under selection in the European or Indian population.

Together, this study suggests that natural selection shapes the plants reproduction mode and organ development, explaining strong population-specific variations in genetic and epigenetic markers.

3.3.2 Zusammenfassung

Diese Studie charakterisiert die genetische/epigenetische Variation, sowie die Demographie und Migrationsgeschichte von *Spirodela polyrhiza*. Hierfür wurden die Genome von insgesamt 228 *S. polyrhiza* Genotypen zur Identifikation von SNPs, Insertionen/Deletionen (<50 bp) und strukturelle Variationen (>50 bp) gefiltert. Cytosinmethylierungen (CpG, CHH und CHG) wurden mittels Bisulfidsequenzierung von jeweils fünf Genotypen aus jeder genetischen Population von *S. polyrhiza* identifiziert. Demographische oder phylogenetische Analysen wurden mittels PCA oder Maximum-likelihood basierenden Methoden auf Basis von SNP-Daten durchgeführt. Migrationsprozesse der verschiedenen Populationen wurde mittels Approximate bayesian computation vorhergesagt. Genomische Regionen, welche Ziel von Selektionsereignissen waren, wurden in einem Genom-weiten Scan auf Basis von SNP-Daten identifiziert. Insgesamt wurden in dieser Studie 1241981 SNPs, 166075 Insertionen/Deletionen und 3205 strukturelle Variationen identifiziert. Die Genom-weite Nukleotid-Diversität betrug 0.0016. Weiterhin konnte für *S. polyrhiza* die niedrigste durchschnittliche Spezies-weite Methylierungsrate innerhalb aller analysierter Angiospermen ermittelt werden. Durch Analyse der Populationsstruktur wurden insgesamt vier genetische Populationen, welche nach ihren überwiegenden Verbreitungsgebieten als Amerika, SO-Asien, Indien und Europa Populationen bezeichnet worden sind, identifiziert. Gemäß der Vorhersage des Approximate bayesian computation -Modells, ist ein Szenario, bei dem die Amerika und SO-Asien Populationen einer gemeinsamen Urpopulation entspringen und die Indien und Europa Populationen sich später aus der SO-Asien Population entwickelt haben am wahrscheinlichsten. Genetisch unterschieden sich die Populationen hinsichtlich der Größe ihrer Kopplungsblöcke und ihres Heterozygotiegrades, was auf Unterschiede in der sexuellen Reproduktionsrate zurückzuführen sein könnte. Unterschiede im Heterozygotieindex wurden mit zwei MADS-box Genen: *AGL62* und *SOC1*, deren Orthologe die Entwicklung von Blütenorganen steuern, assoziiert. Ein Genom-weiter Scan identifizierte 69 Gene, welche Ziel von Selektionsereignissen waren. Homologe dieser Gene sind funktional vorwiegend an Prozessen wie Gametogenese, Embryogenese, Blühinduktion und Organentwicklung beteiligt. Viele Gene, die an Reproduktions- oder Entwicklungsprozessen beteiligt sind wie *BB*, *FLK*, *AGL6* und *AP3* standen in der Europa- oder Indien-Population unter Selektion. Zusammenfassend liefert dieses Studie Erkenntnisse darüber, wie natürliche Selektion reproduktive Eigenschaften und Entwicklungsprozesse formen kann. Aus derartigen Selektionsprozesse resultieren starke Populations-spezifische Unterschiede in der Verteilung genetischer und epigenetischer Marker.

3.3.3 Statement of Contribution

The project was supervised by Meret Huber and Shuqing Xu. All data analyses were conducted by Yangzi Wang, Pablo Duchen and Shuqing Xu. Klaus-Jürgen Appenroth, Sowjanja Sree, Hai Zhao and Shuqing Xu provided duckweed genotypes and further resources to the project. Alexandra Chàvez conducted all DNA extractions and DNA sample preparations. Martin Höfer conducted RNA-extractions and generated cDNA libraries used for confirming the expression of the identified candidate genes via qPCR in two genotypes. Shuqing Xu, Yangzi Wang, Pablo Duchen, Alexandra Chàvez and Meret Huber wrote the first draft of the manuscript and prepared all figures. All authors provided feedback on the final version of the manuscript or contributed to the supplemental part of the manuscript.

Supervision confirmation

Prof. Shuqing Xu

iomE
Institut für Organismische und Molekulare Evolutionsbiologie
JG|U Mainz

communications biology

Article

https://doi.org/10.1038/s42003-024-06266-7

Population genomics and epigenomics of *Spirodela polyrhiza* provide insights into the evolution of facultative asexuality

Yangzi Wang[1,2,8], Pablo Duchen[1,2,8], Alexandra Chávez[1,2,3], K. Sowjanya Sree[4], Klaus J. Appenroth[5], Hai Zhao[6], Martin Höfer[1,2], Meret Huber[1,3] & Shuqing Xu[1,2,7]

Many plants are facultatively asexual, balancing short-term benefits with long-term costs of asexuality. During range expansion, natural selection likely influences the genetic controls of asexuality in these organisms. However, evidence of natural selection driving asexuality is limited, and the evolutionary consequences of asexuality on the genomic and epigenomic diversity remain controversial. We analyzed population genomes and epigenomes of *Spirodela polyrhiza*, (L.) Schleid., a facultatively asexual plant that flowers rarely, revealing remarkably low genomic diversity and DNA methylation levels. Within species, demographic history and the frequency of asexual reproduction jointly determined intra-specific variations of genomic diversity and DNA methylation levels. Genome-wide scans revealed that genes associated with stress adaptations, flowering and embryogenesis were under positive selection. These data are consistent with the hypothesize that natural selection can shape the evolution of asexuality during habitat expansions, which alters genomic and epigenomic diversity levels.

Understanding the evolution of sexual reproduction has long been at the center of evolutionary biology. Theories suggest that asexual reproduction is beneficial for the short term but costly for the long term, mainly due to accumulations of deleterious mutations and low effective population size[1–5]. Facultative asexuality, where organisms can reproduce both sexually and asexually depending on environmental conditions, should be optimal for one individual's lifespan[6,7]. While rather few animals such as aphids (Aphidoidea)[8], water fleas (Cladocerans)[9], and rotifers[10] reproduce facultatively asexually, up to ~80% of the flowering plants, including important crops and keystone species, can reproduce both sexually and asexually[11]. Asexual reproduction in plants involves different types of vegetative reproduction (e.g. runners, tubers, bulbs, corms, suckers, plantlets), as well as apomixis, the formation of seeds without fertilization[12]. Because changes between sexual and asexual reproduction affect the ability to persist in the short and long term, natural selection might act on the genetic controls of sexual and asexual reproduction in facultative asexual organisms, which in turn can alter the levels of genomic diversity, heterozygosity and effectiveness of selection in the population[2,5,13–15]. However, direct evidence supporting this prediction remains scarce, mainly due to the lack of a suitable facultative asexually reproducing system in which the signature of selection can be detected at genomic levels.

Evolutionary changes in sexual and asexual reproduction might also affect the maintenance and dynamics of chromatin marks, e.g., epigenetic markers such as DNA methylations. In plants, cytosine methylation can occur in three sequence contexts: CpG, CHG, and CHH (H = A, T, or C), which are controlled by different mechanisms and have different dynamics during reproduction[16]. Typically, CpG and CHG methylation are maintained by methyltransferases1 (*MET1*) and CHROMOMETHYLASE3 (*CMT3*), respectively, whereas CHH methylation is mostly maintained by *CMT2*[17]. During sexual reproduction, DNA methylations are highly dynamic[18]. In both male and female gametogenesis, the megaspore mother cell and microspore mother cell experience dramatic chromatin changes during cell specification, such as heterochromatin decondensation and an enlarged nuclear volume[19,20]. During male gametogenesis, sperm DNA is highly methylated in the CpG and CHG context but has low CHH methylation in retrotransposons[18,21,22]. During female gametogenesis, CpG

[1]Institute of Organismic and Molecular Evolution, University of Mainz, 55128 Mainz, Germany. [2]Institute for Evolution and Biodiversity, University of Münster, 48161 Münster, Germany. [3]Institute of Plant Biology and Biotechnology, University of Münster, 48161 Münster, Germany. [4]Department of Environmental Science, Central University of Kerala, Periya 671320, India. [5]Matthias Schleiden Institute — Plant Physiology, Friedrich Schiller University of Jena, 07743 Jena, Germany. [6]Chengdu Institute of Biology, Chinese Academy of Sciences, 6100641 Chengdu, China. [7]Institute for Quantitative and Computational Biosciences, University of Mainz, 55218 Mainz, Germany. [8]These authors contributed equally: Yangzi Wang, Pablo Duchen. e-mail: shuqing.xu@uni-mainz.de

and CHH methylation remains largely steady[23]. After fertilization, CHH methylation increases during embryogenesis and can approach 100% at individual cytosines, which then decreases likely through a passive mechanism after germination[24–26]. In contrast, during vegetative reproduction, DNA methylation is likely steady since meiosis and embryogenesis are lacking[27–29]. Although Niederhuth, C. E. et al.[30]. comparing DNA methylations among 34 angiosperm species suggested that clonally propagated species often have low CHH methylation, the extent to which asexual reproduction affects genome-wide methylation levels remains unclear.

Here, we investigated the population genome and epigenome of a facultatively asexual plant, *Spirodela polyrhiza* (the giant duckweed; Lemnaceae), using samples from a global collection. This species, like other duckweeds from the genera *Spirodela*, *Landoltia* and *Lemna*, is characterized by leaf-like fronds derived from fused stems[31] and, with multiple roots on each frond[32] and with a highly reduced vascular system[33]. *Spirodela polyrhiza* reproduces vegetatively via budding under normal conditions but very rarely switches to sexual reproduction under unfavorable conditions[34,35]. Recent studies showed that despite its global distribution in diverse habitats, the genomic diversity, spontaneous mutation rates and DNA methylation levels in *S. polyrhiza* are very low[36–41], which might be associated with its overall low frequency of sexual reproduction. DNA methylation profiling of two genotypes suggests that DNA methylation in *S. polyrhiza*, which is substantially lower than in other plants, varied between genotypes[41]. Further insights into the evolutionary origin and consequences of asexuality on genomic and epigenomic variation in *S. polyrhiza* are required to understand the demographic history and to identify the footprint of selection on the genome.

Results

Extremely low genomic variations in *S. polyrhiza*

We sequenced the genomes of 131 globally distributed *S. polyrhiza* genotypes with an average of ~25 X coverage. Together with previously published samples[36,37], we analyzed the genomic diversity of 228 *S. polyrhiza* individuals across five continents (Supplementary Data 1). We identified 1,241,981 high-quality biallelic single-nucleotide polymorphisms (SNPs) and 166,075 short insertions and deletions (INDELs, less than 50 bp of length). Based on an updated genome annotation of *S. polyrhiza* (see Supplementary Results Methods 1.1 and Supplementary Results Section 2.1), we found that most of the SNPs (70.3%) are in the intergenic regions (Supplementary Fig. 1). Of all the SNPs located in the protein-coding regions, 61,039 were identified as nonsynonymous and 44,287 as synonymous (Supplementary Data 2). Consistent with our previous study[36], the genome-wide nucleotide diversity is 0.0016 (Supplementary Table 1), which falls within the lower range of genome-wide nucleotide diversity of other tested multicellular eukaryotes (Supplementary Table 2 and Supplementary Fig. 2). The species-wide efficacy of selection (π_N/π_S ratio) is 0.37, the highest among studied organisms[42], indicating a relatively relaxed purifying selection in *S. polyrhiza*, despite its large effective population size[36,37].

In addition to SNPs and small INDELs, we also characterize the genome-wide structural variations (SVs, ≥50 bp in length) in *S. polyrhiza* (see Supplementary Methods Section 1.2 and Supplementary Results Section 2.2). We identified 3,205 high-quality SVs, including 2,089 deletions, 291 duplications, and 825 insertions. Among all identified SVs, 155 duplications and 169 deletions affected protein-coding sequences (Supplementary Table 3 and Supplementary Data 3). Using a permutation approach at a genome-wide level, we identified gene families that are significantly enriched with SVs and small INDELs (see Supplementary Methods Section 1.3, 1.4, and 1.5, Supplementary Results Section 2.3, and Supplementary Data 4 and 5), respectively. We found several gene families related to defences, such as *RPP8*[43] and the glycoside hydrolase[44], are enriched with both SVs and small INDELs. This is consistent with findings from *Arabidopsis* and other plant species, which show that SVs are enriched in stress and pathogen resistance[45,46] (Supplementary Data 5). Interestingly, we also found SVs and small INDELs are also enriched in gene families that are involved in organ development and reproduction, such as the receptor-like protein kinases gene family[47] and MADS-box gene family that has been shown to have substantial gene losses and copy number variations in duckweeds[48–50].

Population structure and demographic history of *S. polyrhiza*

Because *S. polyrhiza* is facultatively asexual, genotypes collected from the geographic proximity can be derived from the same clonal family. Using a previously established grouping threshold that was developed in *S. polyrhiza*[2], we identified 159 likely clonal families in the sampled population (Supplementary Data 6).

Population structure and principal component analyses revealed four populations in the sampled *S. polyrhiza* (Fig. 1a and b). Consistent with our previous study, the four populations are largely concordant with their geographic origins, namely America, Southeast Asia (SE-Asia), Europe, and India (Supplementary Fig. 3), with a few exceptions that can be due to recent migration events or artifacts during long-term duckweed maintenance.

We inferred the population history with a Maximum Likelihood (ML) phylogeny and Approximate Bayesian Computation (ABC). For ML, we used *Colocasia esculenta* (from the Araceae family) as an outgroup. The maximum likelihood phylogeny of all 228 genotypes indicated an early split of the American population from the other populations and subsequent splits of the Indian and European populations from SE-Asia (Fig. 1c). The European population constitutes the most recent split (Fig. 1c and d). Here, genotypes collected from the transcontinental region (e.g. Russia) showed intermediate features of SE-Asian and European populations, suggesting this as a likely migration route. Furthermore, the Indian population possibly originated via Thailand and Vietnam, as genotypes from these countries show intermediate features between Indian and SE-Asian populations.

We modeled the demographic history using an ABC modeling approach to further validate the evolutionary history of the four populations in *S. polyrhiza* (see Supplementary Methods Section 1.6 and Supplementary Table 4). Based on the phylogenetic analysis, we simulated three plausible demographic scenarios, allowing for either the SE-Asian, American or an additional putative population to function as the ancestral population (Supplementary Fig. 4). We found that the scenario, in which the American population and Asian population were derived from an additional putative ancestral population, constituted the most supported model (Fig. 1d). While the American population was separated from other populations around one million generations ago, the European population was derived from the SE-Asian population only 12,000 generations ago (see Supplementary Results Section 2.4 and Supplementary Table 5).

Determinants of genomic diversity among populations

Among the four populations, nucleotide diversity (π) and the efficacy of selection (π_N/π_S ratio) varied among populations (Fig. 2b). While the SE-Asian population has the highest π and lowest π_N/π_S ratio, the American population has the lowest π and highest π_N/π_S ratio. Interestingly, while the European population has a much smaller π compared to the SE-Asian population, the π_N/π_S ratio of the European population remains similar to the latter, likely due to its recent split from the SE-Asian population.

Using genome-wide SNPs, we found that linkage disequilibrium (LD) is comparable to *Arabidopsis thaliana*[51], suggesting considerable historical sexual reproduction in *S. polyrhiza*. However, the extent of LD decay varied substantially among populations (Fig. 2b and Supplementary Fig. 5). While the Asian population showed the most rapid LD decay (about 12 kb at $r^2 = 0.2$), the European population had very long LD blocks (>100 kb). The Indian and American populations had intermediate LD decay. Consistently, the Asian population had the highest recombination rate compared to the other three (Fig. 2b). Different LDs and recombination rates found among populations indicate that the frequencies of sexual reproduction varied among populations. In addition, we found that the variations of heterozygosity in *S. polyrhiza* showed a similar pattern with the genomic diversity and recombination rate among four populations (Fig. 2b and Supplementary Fig. 6).

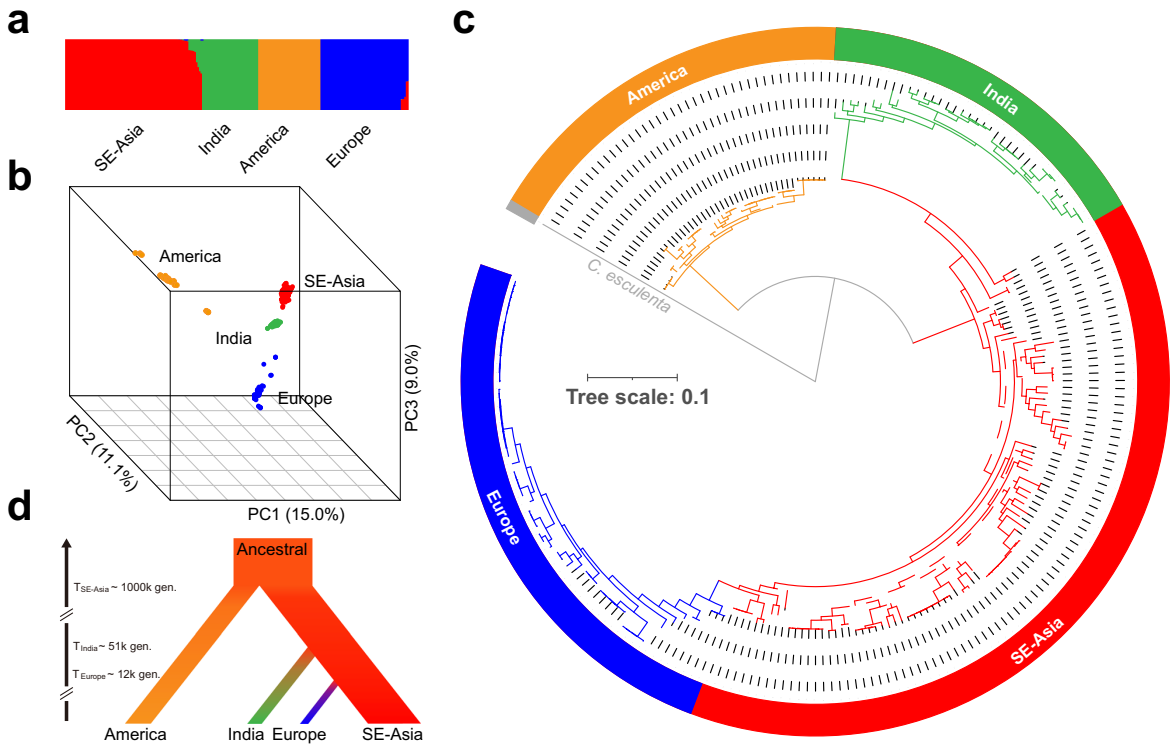


Fig. 1 | Phylogeny, population structure and demographic model of 228 *S. polyrhiza*. **a** The population structure of 228 *S. polyrhiza* genotypes. **b** The principal component (PC) analysis of the SNPs from 228 *S. polyrhiza* genotypes. The three coordinates indicate the first three PCs. **c** The Maximum Likelihood phylogenetic tree of 228 *S. polyrhiza* genotypes. The gray branch represents the outgroup - *C. esculenta*. Dashed branches represent internal nodes with supporting values lower than 0.75 (the max is 1). **d** The demographic model of *S. polyrhiza* populations. "$T_{population}$" indicates the estimated divergence time in generations.

Interestingly, the changes in genomic diversity and levels of heterozygosity are associated with two SVs involving MADS-box genes that are involved in sexual reproduction. One SV is an 84 bp insertion at the last coding sequence (CDS) of gene SpGA2022_005278, a homolog of *AGL62* from the Mα subclade of MADS-box genes (Supplementary Fig. 7). In *A. thaliana*, *AGL62* is a transcription factor that suppresses endosperm cellularization by activating the expression of a putative invertase inhibitor, *InvINH1*, in the micropylar region of the endosperm[52,53] (Supplementary Fig. 8). The insertion may potentially disrupt the function of the *AGL62*-like gene, suggesting a possible reduction in the suppression of endosperm development, which might be required for sexual reproduction (Fig. 2a). Consistently, we found the insertion was at a higher abundance in the SE-Asian population (87.5%) than in other populations (Fig. 2c, d). In addition, the insertion positively correlates with heterozygosity within the European population (Supplementary Table 6 and Supplementary Fig. 9).

Another SV is a 69 bp deletion at 1.8 kb upstream of SpGA2022_007306, (Supplementary Fig. 7), a gene that show homology to *SOC1* (but shorter than *SOC1*, Supplementary Data 7), which is a positive regulator of the flowering process in *A. thaliana*[54] (Fig. 2a). Conserved protein domain analyses suggested that SpGA2022_007306 has *SRF*-like MADS domain but lacks the K-box region (Supplementary Fig. 10), which is similar to *Os03g03100* (*OsMADS50*), a *SOC1* homology that are involved in regulating flowering time in rice[55–58]. The deletion was exclusively found in the Indian population with the alternate allele frequency of 73% (Fig. 2c). It is plausible that the deletion, due to its disruption potential at the *cis*-regulatory region, reduces the ability of this *SOC1*-like gene to respond to the upstream floral activators (e.g. *CO*) in *S. polyrhiza*, thus reducing the frequency of sexual reproduction in the Indian population (Fig. 2d). Consistently, this deletion negatively correlates with heterozygosity in the Indian population (Supplementary Table 6, Supplementary Fig. 11). However, future functional validations on SV of the two MADS-box genes are needed to provide further mechanistic insights into the observed patterns.

Population epigenomic diversity in *S. polyrhiza*

As changes in sexual reproduction can also alter epigenomic dynamics, we further investigated the patterns of population epigenomic diversity in *S. polyrhiza*. We selected five individuals from each population and quantified their shoot DNA methylation levels at single-base resolution using whole genome bisulfite sequencing (Supplementary Table 7). Similar to a recent study[39], we found that only 1.6% of cytosines are methylated in *S. polyrhiza* (7.6% of CpG, 2.3% of CHG, and 0.1% of CHH; Supplementary Table 8), and the average species-wide methylation level is the lowest among all studied angiosperms (Supplementary Fig. 12)[30,59]. The hierarchical clustering of 20 methylomes in CHG and CHH contexts in gene bodies show overall consistency with their genetic similarity (Supplementary Fig. 13 and 14) with few discrepancies were mostly found within the same population or between the recently diverged SE-Asian and European populations. While in the CpG context, we did not observe clear correlations between genetic and methylation distances (Supplementary Fig. 15).

We then compared the genome-wide weighted methylation level (wML) among populations. For CpG methylation, no differences were found among four populations at genome-wide, gene body, or TE levels (Fig. 3a, d and g). For CHG, the Indian population had the lowest genome-wide methylation level among all four populations (Fig. 3b, e, and h). Interestingly, for CHH, the SE-Asia and Europe populations had the higher genome-wide methylation levels compared to American and India populations ($P < 0.05$, pairwise Wilcoxon test; Fig. 3c), while the European and Indian populations showed a gradual reduction of methylation in comparison to the SE-Asian population. The pattern was the same for both gene

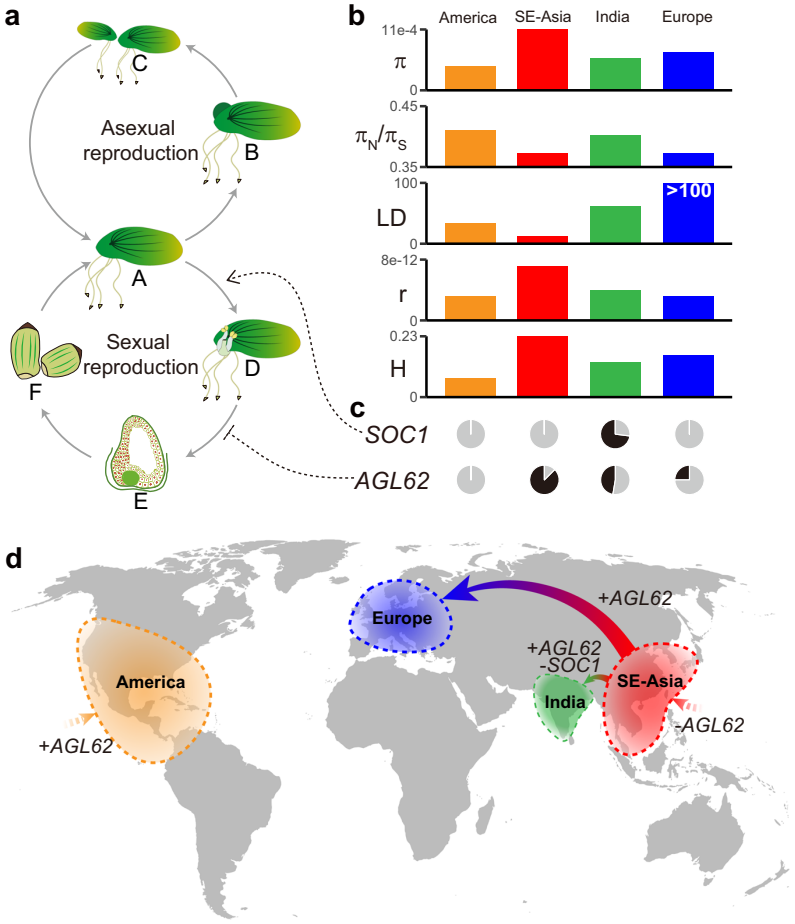


Fig. 2 | Genomic diversity variation among four populations might result from the switching between sexual and asexual propagation in *S. polyrhiza*. **a** Scheme of asexual and sexual propagation cycles in *S. polyrhiza*. (**A**) Vegetative stage of *S. polyrhiza*; (**B**) Budding; (**C**) Offspring from clonal propagation; (**D**) *S. polyrhiza* flowering; (**E**) Putative schematic of ovule at endosperm cellularization stage; (**F**) Putative schematic of seeds. **b** Bar plots show the differences among the four populations in terms of "π": genome-wide nucleotide diversity; "LD": the physical extent (in kb) of pairwise SNPs at r^2 of 0.2 (Europe does not yet reach $r^2 = 0.2$ at 100 kb, Supplementary Fig. 5); "π_N/π_S": the efficacy of linked selection; "r": genome-wide recombination rate; "H": the median of per population genome-wide heterozygosity rate. **c** Two panels of pie charts indicate the allele frequencies of the functional allele of *SOC1*-like and *AGL62*-like genes among populations. Gray: functional allele frequency; Black: SVs allele frequency. **d** The distribution and migration world map of the four *S. polyrhiza* populations. "+" suggests the increased functional allele frequency, while "-" suggests the decreased functional allele frequency.

bodies and TEs ($P < 0.05$, pairwise Wilcoxon test; Fig. 3f, i). The genome-wide reduction of CHH methylation is consistent with the hypothesis that clonal reproduction reduces CHH methylation, and the effects gradually accumulate over clonal generations[60].

The footprint of selection on the genome

To identify the genomic signature of selection at the species level, we performed genome-wide scans. To reduce false positives, we used the μ-statistics from RAiSD[61], the composite likelihood ratio CLR statistic from SweeD[62], and the T statistic from LASSI[63]. We found 69 genes showed strong signatures of selection using all three methods (Supplementary Fig. 16 and Supplementary Data 8). Manual inspection indicated that several orthologs of these genes are related to gametogenesis (e.g., *NOTCHLESS*) and embryogenesis (e.g., *NUP214*, *CPSF*, *CDK*, *AGP*, and *ACR4*) in *Arabidopsis thaliana*[64–68]. Further enrichment analysis indeed showed that embryo lethal genes were enriched in these 69 genes ($P = 0.016$, χ^2 test). In addition, the *A. thaliana* orthologs of several genes under selection are also associated with controlling sexual reproduction, including floral development (*DRMY1* and *ACR4*)[64,69], flowering time (*NF-Y AT2G27470*, *NF-YAT1G72830*, and *CPSF*), pollen development (*EFOP3*, *ELMOD*, and *CLC*)[66,70–74], seed development (*NUP214*, *NF-Y AT2G27470* and *NF-YAT1G72830*, and *Transducin/WD40*)[65,70,75]. Furthermore, among these 69 genes, we also found several genes involved in leaf development and vascularity (*SECA2*, *RbgA*, *PHABULOSA/PHAVOLUTA*)[76–78], light signaling (*NF-Y*, *CCR4-NOT*, and *PPP*)[70,79,80], root development (*GEND1*, *WAVY*, and *ACR4*)[64,81,82], DNA damage repair (*ATM* and *Xrcc3*)[83,84], and stress tolerance (*phospholipase D*, *histone superfamily protein*, *RabGAP*, *FC1*, *NUDX2*) (Supplementary Data 8).

To further understand the selection that drove the evolution within individual populations, we identified the signature of positive selection in a three-population tree using patterns of linked allele frequency differentiation and calculating the corresponding composite-likelihood ratio (CLR, see Methods). In total, we found 1,883 genes on the SE-Asian branch, 593 genes on the Indian branch and 401 genes on the European branch (Fig. 4a; see Supplementary Results Section 2.6, and Supplementary Data 9) which showed strong signatures of selection (top 1% of CLR values). We did not find evidence supporting the hypothesis that differentially methylated genes were under positive selection (see Supplementary Methods Section 1.7, Supplementary Results Section 2.5, and Supplementary Data 10).

We found that genes under positive selection in the European branch are enriched with reproduction and development-related GO terms (Supplementary Fig. 17). Among these, SpGA2022_013448, in chromosome 9, is an ortholog of FLOWERING LOCUS KH DOMAIN (*FLK*) that delays flowering by up-regulating *FLC* family members in *A. thaliana*[85]. This gene showed a strong signature of selection in the European branch but not in other branches (Fig. 4c, d). Similarly, SpGA2022_006111, on chromosome 3, is an ortholog of the *A. thaliana* BIG BROTHER (*BB*) that negatively regulates floral organ size and is also under selection in Europe[86] (Fig. 4c).

In the SE-Asian population, we found that gene SpGA2022_051517, a CHROMOMETHYLASE3 (*CMT3*) ortholog in *A. thaliana* that is likely associated with maintaining CHG methylation[17], was under positive selection. This is consistent with the higher CHG methylation levels observed in the SE-Asian population when compared to the European and Indian populations (Figs. 3a, b). Within the Indian population, we found that five MADS-box genes have been under selection exclusively along this branch. Given that there are 43 MADS-box genes in the genome, the fact that five of

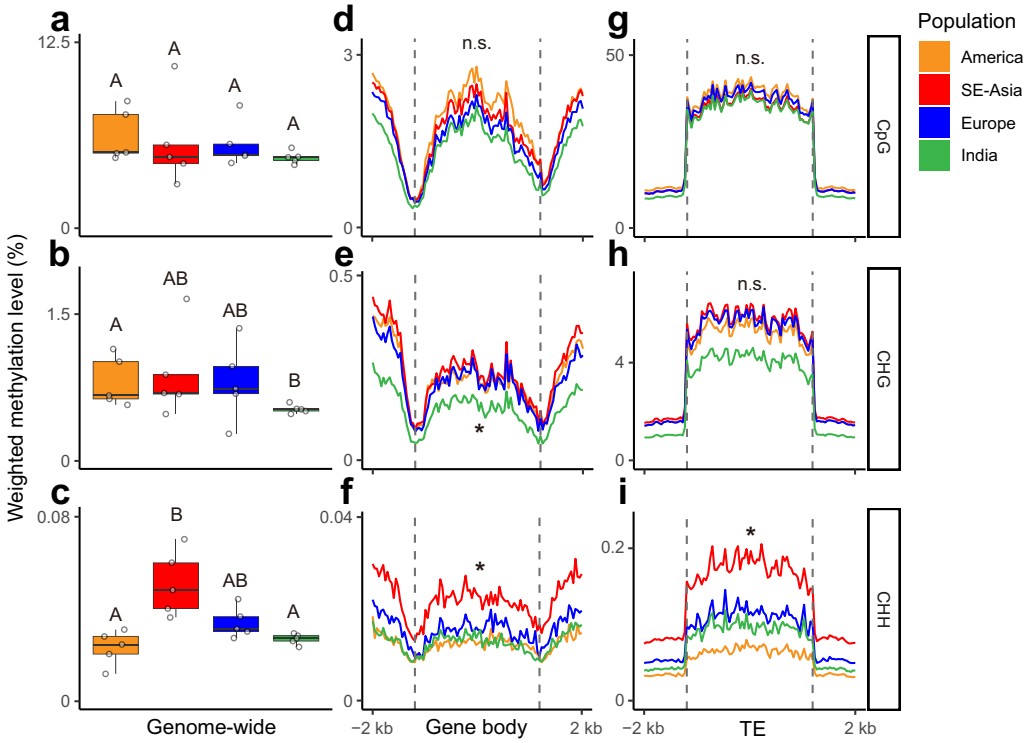


Fig. 3 | Weighted methylation level (wML) among four populations. a–c Box plots of genome-wide weighted methylation level (wML) in (**a**) CpG, (**b**) CHG, and (**c**) CHH context ($N = 5$). Uppercase letters indicate statistical differences among populations using Wilcoxon test (with bonferroni method for multiple tests corrections). **d–f** wML in the gene body and its flanking 2 kb regions in (**d**) CpG, (**e**) CHG, and (**f**) CHH context. **g–i** wML in TE and its flanking 2 kb regions in (**g**) CpG, (**h**) CHG, and (**i**) CHH context. The asterisks indicate significant differences between populations ($P < 0.05$; Wilcoxon test), while "n.s." indicates no significant difference.

them have been targeted by selection, constitutes a significant enrichment of such genes under selection ($P = 0.0075$, Fisher's Exact test). For example, SpGA2022_013078 is an homolog of AGAMOUS-LIKE6 (*AGL6*), which is involved in flower and meristem identity specification in rice[87]; SpGA2022_052274, a homolog of *APETALA3* (*AP3*), is involved in the petal and stamen specification in *A. thaliana*[88]; and SpGA2022_006905 belongs to the SHORT VEGETATIVE PHASE (*SVP*-group) which controls the time of flowering and meristem identity[89].

We found 77 genes under positive selection (top 1% CLR values) in both the European and Indian populations (Fig. 4b), significantly more genes than expected by chance ($P < 2.2e\text{-}16$, Fisher's Exact Test). Among these, gene SpGA2022_055195, an ortholog to *CYP78A9* of cytochrome P450 monooxygenases in *A. thaliana*, belongs to a highly conserved gene family CYP78A. Previous studies in *A. thaliana* and other species found that *CYP78A9* plays a critical role in promoting cell proliferation during flower development and further impacts seed size[90–92]. In addition, the RNA-seq data indicates that *CYP78A9* is differentially expressed between India and Europe populations (see Supplementary Methods Section 1.8, Supplementary Results Section 2.7, and Supplementary Data 11). Overall, these data consistently suggest that genes involved in reproduction and development were under selection in Indian and European populations, which might have led to reduced sexual reproduction in these two populations.

Discussion

Here, we characterized the genomic and epigenomic diversity, as well as the demographic history of a facultative asexual flowering plant, *S. polyrhiza*. We found that among populations of *S. polyrhiza*, demographic history and reproductive system jointly determine the population's genomic and epigenomic diversity. Analyses on the footprint of selection suggest that natural selection drove the reduced vascular system and increased asexuality in *S. polyrhiza*.

Theory predicts that asexual reproduction reduces genomic diversity and the efficiency of purifying selection[93]. Consistent with this prediction, at the species level, we found that *S. polyrhiza* has very low genomic diversity and reduced purifying selection (seen as an increased π_N/π_S ratio), when compared to a wide range of spermatophyte plants[42]. Within species, the SE-Asian population, which has the highest frequency of sexual reproduction based on the estimated recombination rate, has the highest genomic diversity, the lowest π_N/π_S ratio and the highest heterozygosity (Fig. 2b), supporting the theoretical prediction[2,5,13–15]. The low π_N/π_S ratio found in the European population, which has the lowest sexual reproduction and genomic diversity, is most likely due to its migration history. The demographic model suggested that the European population derived from the SE-Asian population very recently (Fig. 1d). It is likely that the π_N/π_S ratio in the European population remained the same as its ancestral population and has not reached an equilibrium level yet.

While there are fewer genome-wide SVs in *S. polyrhiza* compared to other species[94,95], we found these variants and small INDELs are in tendency enriched in stress responses and reproduction, such as MADS-box genes. This indicates that the loss-of-function of genes involved in flower development and sexual reproduction, is under natural selection. The results are consistent with the observation that the number of functional MADS-box genes was dramatically reduced in *S. polyrhiza*[49].

Single-base resolution methylomes of 20 individuals showed that the overall CpG, CHG and CHH methylation levels in *S. polyrhiza* shoots are very low, consistent with previous studies[39,41]. The low levels of DNA methylation might be associated with reduced sexual reproduction: while CpG and CHG methylations in plants are important for controlling cross-overs during

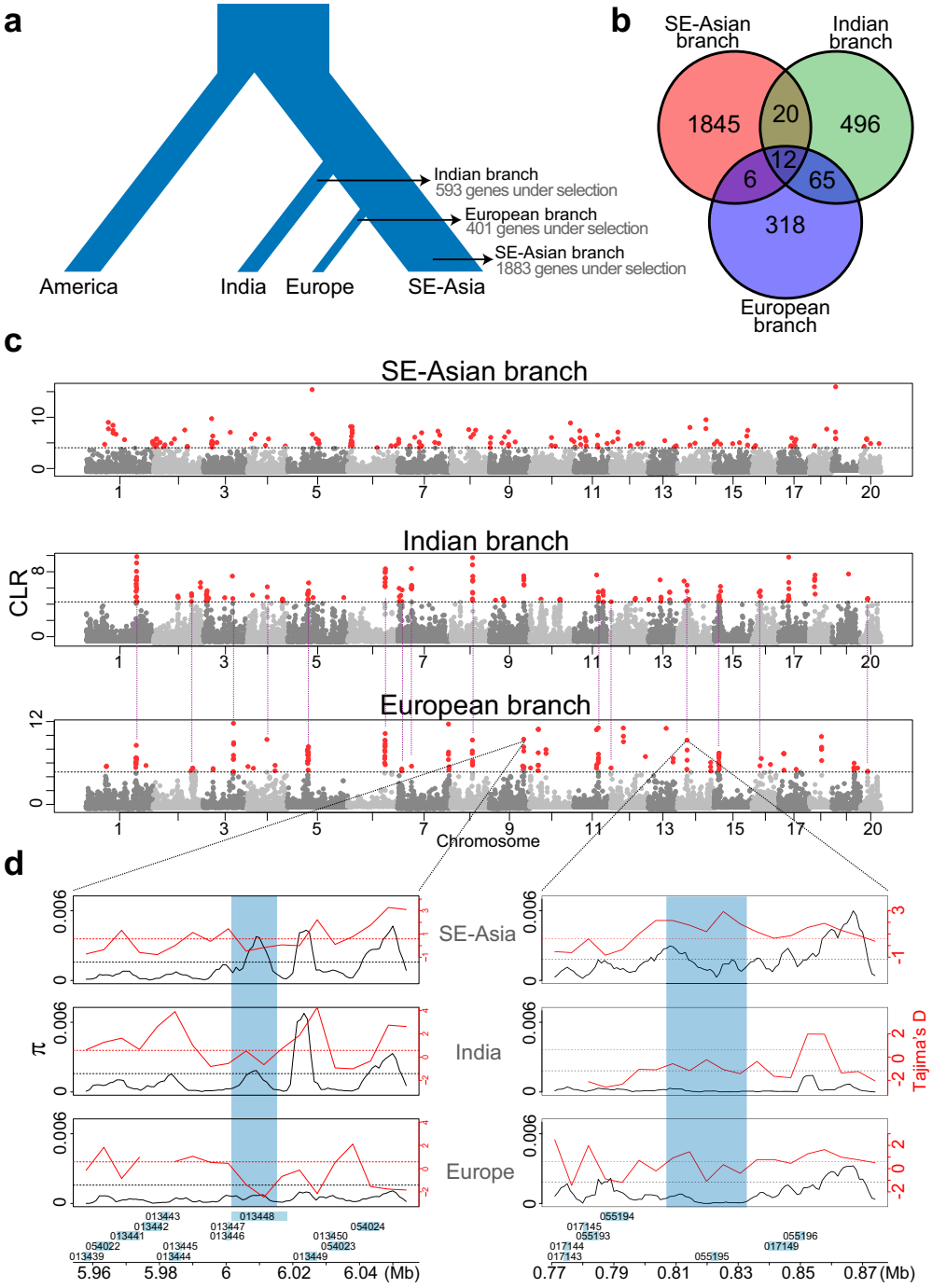


Fig. 4 | Branch-specific selection signature scans. a Population tree is used for the population-specific selection analyses. Input consists of an "outgroup" population (America) and two "ingroup" populations (India and SE-Asia, or Europe and SE-Asia). We present the number of genes in the ingroup populations belonging to the top 1% CLR scores reported by 3P-CLR. **b** Venn diagram showing the genes under selection that are common between the populations of SE-Asia, India, and Europe. **c** Branch-specific genome-wide scan of selection: each dot represents a chromosomal locus picked by 3P-CLR for which a CLR score is calculated; loci on the top 1% CLR are shown in red. The dashed purple lines between European and Indian panels indicate common peaks of selection detected at the same genomic regions in both populations. **d** Genomic diversity as depicted by π (solid black lines) and Tajima's D (solid red lines) for two selected loci: the *FLK* (left) and the *CYP78A9* (right). *FLK* shows a signature of selection (low π and negative Tajima's D) in the European population, while *CYP78A9* shows a signature of selection in both Europe and India. Dashed lines indicate genome-wide averages for π (black) and Tajima's D (red).

meiosis[96] and are increased during male gametogenesis, CHH methylation is highly accumulated during embryogenesis[18,24–26]. In facultative asexual plants, due to reduced sexual reproduction and meiosis, the selection of genetic mechanisms maintaining or increasing the CpG, CHG and CHH methylation is reduced or absent, which might have led to the reduced CpG, CHG and CHH methylation levels. Consistently, a recent study suggests that *S. polyrhiza* has lost several genes in the RdDM pathway[41]. Interestingly, within species, the CHG and CHH methylation profile of the 20 individuals largely correlates with their genetic distance (Supplementary Fig. 13 and 14), indicating a gradual neutral evolution of DNA methylomes in *S. polyrhiza*. For example, the Indian and European populations, which diverged from SE-Asian populations around 51,000 and 12,000 generations ago, gradually decreased their CHH methylations (Fig. 3a–c).

At the species level, using a genome-wide scan approach, we found a strong signature of natural selection on genes involved in flower and seed development, indicating that the evolution of reproduction, likely, an increased clonal propagation in *S. polyrhiza*, was driven by natural selection. This is consistent with the pattern that many aquatic organisms reproduce clonally[97]. In addition, several genes related to vascularity, root development and DNA damage repair were also under strong selection, suggesting the reduced root and vascular development and low mutation rate in *S. polyrhiza* were likely also driven by natural selection.

Among populations, we found strong positive selection on genes involved in sexual reproduction and development in India and Europe populations, two recently evolved populations that showed reduced genomic recombination. These results are consistent with the hypothesis that natural selection favors clonal reproduction in *S. polyrhiza* during the recent colonization process, a pattern that was frequently found in many invasive species[98,99]. However, despite strong selection favoring clonal reproduction, substantial recombination in the *S. polyrhiza* genome, mostly in the SE-Asian population, remained, reflecting that sexual reproduction is essential to overcome the costs involved in clonal reproduction in the long term.

Taken together, the structure of population genomes and epigenomes of *S. polyrhiza* suggest that demography and natural selection acting on the reproduction system and organ development can shape genome-wide genomic and epigenomic variations.

Materials and Methods

DNA sample preparation and sequencing

We sequenced 131 genotypes that were primarily collected from Asia and Europe (Supplementary Data 1). These samples were cultivated in N-medium[100] until DNA isolation using a CTAB method. Library preparations were carried out following the protocol described in Xu et al.[36]. All libraries were sequenced either on Illumina HiSeq X Ten or Illumina Hiseq 4000 platforms for paired-end sequencing with a read size of 150 bp. Low-quality reads and adapter sequences were trimmed with AdapterRemoval (v2.033)[101]. On average, 33.8 million reads per genotype were obtained. The clean reads were aligned to the *S. polyrhiza* reference genome[48,102] using BWA-MEM (https://github.com/lh3/bwa) with default parameters. Reads without alignment hits or with multiple alignment positions were removed. SAMtools "rmdup" function was used to remove PCR duplicates[103].

Genetic variant identification and gene family annotation

After filtered out low-quality SNPs using GATK[104] (v4.1.4.1, Java 11) with options: "QD < 2.0 | QUAL < 30.0 | SOR > 3.0 | FS > 60.0 | MQ < 40.0 | MQRankSum < -12.5 | ReadPosRankSum < −8.0", we identified 8,363,387 SNPs. Then, VCFtools (v0.1.13)[105] and GATK were used to remove SNPs that have the following features: (1) SNPs from organelle genomes (9,278 SNPs); (2) missing genotypes >20% (85,645 SNPs); (3) mean sequencing depth <8 or >41 (179,920 SNPs); (4) non-biallelic (448,404 SNPs); (5) minor allele frequency (MAF) <1% (6,102,027 SNPs); and finally, (6) located in small SNP clusters (≥ 3 SNPs in a ten base-pair window, accounted for 296,132 SNPs). We updated the protein-coding gene annotation of *S. polyrhiza* based on recently published transcriptomes and Iso-seq data (see Supplementary Methods Section 1.1, Supplementary Results Section 2.1, Supplementary Table 9, and Supplementary Figs. 18–20). We used SnpEff (version 5.0c)[106] to annotate SNPs and INDELs. To exam whether SNP cluster filtering criterion affects the estimation of genomic diversity and selection, we performed additional analyses based on a more relaxed filtering parameter (≥200 SNPs in 1 Kb region). Although the second SNP cluster filtering criterion resulted in 18.7% more SNPs, which are mostly (>88%) located in TE regions or nearby the SV or INDELs, the patterns of genomic diversity and selection did not change. In addition to SNPs and INDELs, We identified SVs using a joint genotyping pipeline and stringent quality filtration processes (see Supplementary Methods Section 1.2, Supplementary Results Section 2.2, and Supplementary Fig. 21-24).

We estimated genome-wide nucleotide diversity (π) and genome-wide π_N/π_S ratios using SNPgenie (v2019.10.31)[107]. The SNPs overlapping with the structure variations were excluded from the calculation to minimize potential interference caused by misalignments, ensuring a more accurate and reliable analysis.

The genome-wide heterozygosity for each individual was calculated using VCFtools (v0.1.13)[105]. We estimated the genetic associations between heterozygosity and the SVs of *AGL62* and *SOC1* using RVTESTS[108] with the single variant Wald test.

To study the potential genetic factors related to the variation of sexual reproduction frequency in *S. polyrhiza*, we annotated the MADS-box gene family (see Supplementary Methods Section 1.3 and Supplementary Results Section 2.3, Supplementary Fig. 25-27, Supplementary Table 10, and Supplementary Data 12). Other gene families that were annotated in Arabidopsis were also identified in *S. polyrhiza* using an orthology-based method (see Supplementary Methods Section 1.4 and Supplementary Data 4).

Population structure and linkage disequilibrium (LD)

We grouped genetically similar genotypes by defining clonal genotype pairs that have no more than 0.01% different homozygous sites and no more than 2% different heterozygous sites. These thresholds were previously adopted by Ho et al.[37].

Prior to the population structure analysis, we removed SNPs that (1) deviated from Hardy-Weinberg Equilibrium (Fisher exact test, $P < 0.01$) or (2) linked loci (each pair of SNP have correlation coefficient $r^2 > 0.33$ in a sliding window with a size of 50 SNPs and step of 5 SNPs), using VCFtools (v0.1.13)[105] and Plink (v1.9)[109].

Principal component analysis (PCA) and population structure analysis were carried out using Plink (v1.9)[109] and fastStructure (v1.0)[110], respectively. The simple mode (as default) from fastStructure was used for the population structure analysis. The K value was estimated using a heuristic function in fastStructure.

For each of the 159 clonal families, we selected the least missingness genotype (i.e. the genotype with the highest sequencing coverage of that clonal family) as the representative genotype. SNP information from all 159 representative genotypes was used to estimate the linkage disequilibrium decay for each of the four populations. PopLDdecay (v3.41)[111] was used to measure LD decay. For each population, we used the following filters: SNP of missing allele > 20% and MAF < 0.05. The allele frequency correlation (denoted as r^2) of pairwise SNPs within 100 kb physical distance was calculated.

Phylogenetic tree reconstruction

We used BLAST+ version 2.9.0[112] to identify orthologous fragments between the genomes of *S. polyrhiza* and *Colocasia esculenta* (Araceae). For each SNP from the core set, the reference allele and its flanking 300 bp (upstream 150 bp and downstream 150 bp, respectively) sequences were extracted from the *S. polyrhiza* genome and then aligned to the *C. esculenta* reference genome[113]. The hit thresholds were set as (1) alignment identity >70%; (2) e-value > 1e − 6; (3) minimum aligned sequence length ≥50 (the aligned sequence must cover SNP position); (4) keep the best hit; and (5) ignore short deletions from *C. esculenta*. The orthologous alleles from *C. esculenta* were used as the outgroup genotype. We identified only 13,120 SNPs that have orthologous fragments in the *C. esculenta* genome. Those data were further used to infer the maximum-likelihood

(ML) phylogenetic tree using RAxML-ng (v1.0.1)[114]. The best hit model was estimated to be 'TVM + G4' using Modeltest-ng (0.1.6)[115,116]. The bootstrapping converged after 700 iterations of the ML tree search. ITOL v5[117] and the Python package ETE2[118] were used for tree visualization.

Selection analysis

Genome-wide scans of selection were performed on all 20 chromosomes of all sampled populations. Selective sweeps were inferred by three programs: RAiSD[61], SweeD[62] and LASSI[63]. RAiSD uses the μ statistic, which provides information on the SFS, LD, and genomic diversity to evaluate the presence of positive selection[61]. SweeD calculates the traditional composite likelihood ratio (CLR) to infer loci under selection[62]. LASSI employs the *T* statistic, which uses a likelihood model based on the haplotype frequency spectrum to detect hard and soft sweeps[63]. As recommended by the authors of LASSI, we selected the top 5% *T* scores as candidates for selection. For RAiSD and SweeD we selected the top 1% scores. After finding the common genes under selection according to all three programs, we reported the genes that have orthologs in *A. thaliana*. The embryo lethal genes from *A. thaliana*[119] were used for the enrichment analysis.

To test for population/branch-specific signals of selection, we ran a composite likelihood ratio (CLR) approach as implemented in 3P-CLR[120]. Briefly, this method uses three-population trees coupled with genomic data as input, from which patterns of linked allele frequency differentiation are calculated. By doing this, this algorithm can tell apart signals of selection that happened in either branch of the tree or in the ancestral lineage, as well as outputting the loci with the highest CLR[120]. In our case, we used either a North America-Asia-Europe, or a North America-Asia-India population tree as input, and 3P-CLR output the CLR across windows along each chromosome in the *S. polyrhiza* genome. We then selected the top 1% windows for each branch of the input tree and reported the genes that are present in each window. To further validate the evidence of positive selection on the regions with the highest CLR, we ran scans of Tajima's D and genomic diversity along the same windows and contrasted them with the same signal along the other population branches. We expect negative Tajima's D and low genomic diversity values along the populations with high CLR values. For authenticity validation of genes under selection, we used RT-qPCR to check the expression of eight genes (see Supplementary Methods 1.9, Supplementary Results 2.8, Supplementary Fig. 28, and Supplementary Table 11 and 12). Another expanded list that includes 37 candidate genes was also created, and these genes' expression (RNA-seq) and orthology alignments against their Arabidopsis counterparts were examined (Supplementary Data 7).

DNA methylation in *S. polyrhiza*

We selected five genotypes from each of the four populations (America, India, SE-Asia, Europe) for single-base whole-genome bisulfite sequencing (WGBS). The genotypes originated from distinct clonal families, except for two European genotypes that came from the same clonal family (Supplementary Table 7).

FastQC (v0.11.5, https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to summarize statistics of the sequencing reads. Trimmomatic (v 0.36)[121] was used to filter out low-quality reads with the parameters "SLIDINGWINDOW: 4:15, LEADING:3, TRAILING:3, ILLUMINACLIP: adapter.fa: 2: 30: 10, MINLEN:36". To account for the genetic variations among genotypes, we generated pseudo-reference genome for each genotype by substituting SNP from the *S. polyrhiza* reference genome using GATK, using a similar strategy to previous studies[122,123]. Bismark (v 0.16.3)[124] was used to align bisulfite-treated reads to pseudo-reference genomes. Identical reads aligned to the same genomic regions were deemed as duplicated reads and thus were removed. Cytosines covered by less than five sequencing reads were excluded from the study. Only after applying these filters the sequencing depth and coverage were then summarized. The sodium bisulfite non-conversion rate was calculated as the percentage of non-converted cytosines to all cytosines in the reads that mapped to the chloroplast genome[125] (GenBank: JN160603.2). For each cytosine site, a binomial test was performed to determine if the cytosine was methylated. If the methylation frequency at the site was lower than the background, which was estimated as the non-conversion rate, then the site was considered unmethylated, and the reads supporting methylation at this site were excluded[126].

We calculated two different methylation parameters: the proportion of methylated cytosines (mC methylation) and weighted methylation level (wML)[126]. For both parameters, only cytosines covered by more than four sequencing reads were involved in the calculation. Those cytosines with low reads supporting methylation but not passing the binomial test were considered as un-methylated cytosines. The mC proportion was calculated by dividing the number of methylated cytosines by the total number of cytosines. Genomic regional wML was calculated using the methylKit (v1.17.5)[127] and the regioneR (v1.28.0)[128], with input based on the cytosine report generated with the Bismark pipeline. Line plots that show the wML patterns across the gene body and transposable elements, as well as their 2 kb flanking regions, were generated using ViewBS (v0.1.11)[129]. The hierarchical clustering, based on the methylation profiles' similarity, was done using methylKit. The comparison between the genetic phylogenetic tree and hierarchical clustering based on the methylome was made using the R packages ggtree (v3.4.4)[130], treeio (v1.20.2)[131], ape (v5.6.2)[132], and phytools (v1.2.0)[133].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw genomic and bisulfite sequencing reads involved in this study can be retrieved from NCBI under accession numbers Bioproject PRJNA701543 and Bioproject PRJNA934173. The scripts for the data analyses are deposited in https://github.com/Xu-lab-Evolution/Great_duckweed_popg. The authors declare that the data and corresponding computational codes supporting the conclusions of this study are available within the article and its supplementary information file.

Received: 1 August 2023; Accepted: 30 April 2024;
Published online: 16 May 2024

References

1. Kondrashov, A. S. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**, 435–440 (1988).
2. Muller, H. J. The relation of recombination to mutational advance. *Mutat. Res.* **1**, 2–9 (1964).
3. Case, T. J. & Taper, M. L. On the coexistence and coevolution of asexual and sexual competitors. *Evolution* **40**, 366–387 (1986).
4. Doncaster, C. P., Pound, G. E. & Cox, S. J. The ecological cost of sex. *Nature* **404**, 281–285 (2000).
5. Hartfield, M. Evolutionary genetic consequences of facultative sex and outcrossing. *J. Evol. Biol.* **29**, 5–22 (2016).
6. Green, R. F. & Noakes, D. L. G. Is a little bit of sex as good as a lot. *J. Theor. Biol.* **174**, 87–96 (1995).
7. Lynch, M. & Gabriel, W. Phenotypic evolution and parthenogenesis. *Am. Nat.* **122**, 745–764 (1983).
8. Simon, J. C., Rispe, C. & Sunnucks, P. Ecology and evolution of sex in aphids. *Trends Ecol. Evol.* **17**, 34–39 (2002).
9. Hebert, P. D. N. Population biology of *Daphnia* (Crustacea, Daphnidae). *Biol. Rev.* **53**, 387–426 (1978).
10. Wallace, R. L. Rotifers: Exquisite metazoans. *Integr. Comp. Biol.* **42**, 660–667 (2002).
11. Klimeš, L., Klimešová, J., Hendriks, R. & van Groenendael, J. in *The Ecology and Evolution of Clonal Plants* (eds H. de Kroon & J. van Groenendael) 1–29 (Backhuys Publishers, 1997).
12. de Meeus, T., Prugnolle, F. & Agnew, P. Asexual reproduction: genetics and evolutionary aspects. *Cell Mol. Life Sci.* **64**, 1355–1372 (2007).
13. Keightley, P. D. & Otto, S. P. Interference among deleterious mutations favours sex and recombination in finite populations. *Nature* **443**, 89–92 (2006).

14. Jaron, K. S. et al. Convergent consequences of parthenogenesis on stick insect genomes. *Sci. Adv.* **8**, eabg3842 (2022).
15. Tucker, A. E., Ackerman, M. S., Eads, B. D., Xu, S. & Lynch, M. Population-genomic insights into the evolutionary origin and fate of obligately asexual *Daphnia pulex*. *Proc. Natl Acad. Sci. USA* **110**, 15740–15745 (2013).
16. Niederhuth, C. E. & Schmitz, R. J. Covering your bases: inheritance of DNA methylation in plant genomes. *Mol. Plant* **7**, 472–480 (2014).
17. Matzke, M. A. & Mosher, R. A. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat. Rev. Genet* **15**, 394–408 (2014).
18. Gehring, M. Epigenetic dynamics during flowering plant reproduction: evidence for reprogramming? *N. Phytol.* **224**, 91–96 (2019).
19. She, W. et al. Chromatin reprogramming during the somatic-to-reproductive cell fate transition in plants. *Development* **140**, 4008–4019 (2013).
20. She, W. J. & Baroux, C. Chromatin dynamics in pollen mother cells underpin a common scenario at the somatic-to-reproductive fate transition of both the male and female lineages in *Arabidopsis*. *Front. Plant Sci*. **6**, 294 (2015).
21. Slotkin, R. K. et al. Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461–472 (2009).
22. Calarco, J. P. et al. Reprogramming of DNA methylation in pollen guides epigenetic inheritance via small RNA. *Cell* **151**, 194–205 (2012).
23. Ingouff, M. et al. Live-cell analysis of DNA methylation during sexual reproduction in *Arabidopsis* reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev.* **31**, 72–83 (2017).
24. Bouyer, D. et al. DNA methylation dynamics during early plant life. *Genome Biol.* **18**, 179 (2017).
25. Kawakatsu, T., Nery, J. R., Castanon, R. & Ecker, J. R. Dynamic DNA methylation reconfiguration during seed development and germination. *Genome. Biol.* **18**, 171 (2017).
26. Narsai, R. et al. Extensive transcriptomic and epigenomic remodelling occurs during *Arabidopsis thaliana* germination. *Genome. Biol.* **18**, 172 (2017).
27. Verhoeven, K. J. F., Jansen, J. J., van Dijk, P. J. & Biere, A. Stress-induced DNA methylation changes and their heritability in asexual dandelions. *N. Phytol.* **185**, 1108–1118 (2010).
28. Verhoeven, K. J. & Preite, V. Epigenetic variation in asexually reproducing organisms. *Evolution* **68**, 644–655 (2014).
29. Van Antro, M. et al. DNA methylation in clonal duckweed (*Lemna minor* L.) lineages reflects current and historical environmental exposures. *Mol. Ecol.* **32**, 428–443 (2023).
30. Niederhuth, C. E. et al. Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* **17**, 194 (2016).
31. Landolt, E., Jäger-Zürn, I. & Schnell, R. *Extreme Adaptations in Angiospermous Hydrophytes*, 290 (Gebrüder Borntraeger, 1998).
32. Bog, M., Appenroth, K. J. & Sree, K. S. Key to the determination of taxa of lemnaceae: an update. *Nordic. J. Botany* **38**, e02658 (2020).
33. Kim, I. Structural differentiation of the connective stalk in *Spirodela polyrhiza* (L.) schleiden. *Appl. Microsc.* **46**, 83–88 (2016).
34. Hicks, L. E. Flower production in the lemnaceae. *Ohio J. Sci.* **32**, 115–132 (1932).
35. Fourounjian, P., Slovin, J. & Messing, J. Flowering and seed production across the lemnaceae. *Int J. Mol. Sci.* **22**, 2733 (2021).
36. Xu, S. et al. Low genetic variation is associated with low mutation rate in the giant duckweed. *Nat. Commun.* **10**, 1243 (2019).
37. Ho, E. K. H., Bartkowska, M., Wright, S. I. & Agrawal, A. F. Population genomics of the facultatively asexual duckweed *Spirodela polyrhiza*. *N. Phytol.* **224**, 1361–1371 (2019).
38. Sandler, G., Bartkowska, M., Agrawal, A. F. & Wright, S. I. Estimation of the SNP mutation rate in two vegetatively propagating species of duckweed. *G3-Genes Genom. Genet.* **10**, 4191–4200 (2020).
39. Michael, T. P. et al. Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies. *Plant J.* **89**, 617–635 (2017).
40. Bog, M. et al. Strategies for intraspecific genotyping of duckweed: comparison of five orthogonal methods applied to the giant duckweed *Spirodela polyrhiza*. *Plants (Basel)* **11**, 3033 (2022).
41. Harkess, A. et al. The unusual predominance of maintenance DNA methylation in spirodela polyrhiza. *G3 Genes Genomes Genet.* **14**, jkae004 (2024).
42. Chen, J., Glemin, S. & Lascoux, M. Genetic diversity and the efficacy of purifying selection across plant and animal species. *Mol. Biol. Evol.* **34**, 1417–1428 (2017).
43. McDowell, J. M. et al. Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the *RPP8* locus of *Arabidopsis*. *Plant Cell* **10**, 1861–1874 (1998).
44. Xu, Z. W. et al. Functional genomic analysis of glycoside hydrolase family 1. *Plant Mol. Biol.* **55**, 343–367 (2004).
45. Pinosio, S. et al. Characterization of the poplar pan-genome by genome-wide identification of structural variation. *Mol. Biol. Evol.* **33**, 2706–2719 (2016).
46. Zmienko, A. et al. Athcnv: A map of DNA copy number variations in the Arabidopsis genome. *Plant Cell* **32**, 1797–1819 (2020).
47. Cui, Y., Lu, X. & Gou, X. Receptor-like protein kinases in plant reproduction: current understanding and future perspectives. *Plant Commun.* **3**, 100273 (2022).
48. Wang, W. et al. The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. *Nat. Commun.* **5**, 3311 (2014).
49. Gramzow L., Theissen G. Stranger than fiction: Loss of MADS-box genes during evolutionary miniaturization of the duckweed body plan. Loss of MADS-box genes in duckweeds. In: *The Duckweed Genomes, Compendium of Plant Genomes*. (eds. Cao X.H., Fourounjian, P. & Wang, W.) (Springer Nature; Cham, Switzerland, 2020).
50. Yoshida, A. et al. Characterization of frond and flower development and identification of ft and fd genes from duckweed *Lemna aequinoctialis* Nd. *Front. Plant Sci.* **12**, 697206 (2021).
51. Cao, J. et al. Whole-genome sequencing of multiple *Arabidopsis thaliana* populations. *Nat. Genet.* **43**, 956–963 (2011).
52. Kang, I. H., Steffen, J. G., Portereiko, M. F., Lloyd, A. & Drews, G. N. The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* **20**, 635–647 (2008).
53. Hoffmann, T. et al. The identification of type I MADS box genes as the upstream activators of an endosperm-specific invertase inhibitor in *Arabidopsis*. *BMC Plant Biol.* **22**, 18 (2022).
54. Lee, J. & Lee, I. Regulation and function of SOC1, a flowering pathway integrator. *J. Exp. Bot.* **61**, 2247–2254 (2010).
55. Norton, G. J. et al. Genome wide association mapping of grain and straw biomass traits in the rice Bengal and Assam Aus panel (baap) grown under alternate wetting and drying and permanently flooded irrigation. *Front. Plant Sci.* **9**, 1223 (2018).
56. Ryu, C. H. et al. OsMADS50 and OsMADS56 function antagonistically in regulating long day (LD)-dependent flowering in rice. *Plant Cell Environ.* **32**, 1412–1427 (2009).
57. Lee, S., Kim, J., Han, J. J., Han, M. J. & An, G. Functional analyses of the flowering time gene OsMADS50, the putative suppressor of overexpression of CO 1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in rice. *Plant J.* **38**, 754–764 (2004).
58. Lee, S. & An, G. Diversified mechanisms for regulating flowering time in a short-day plant rice. *J. Plant Biol.* **50**, 241–248 (2007).
59. Cokus, S. J. et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).
60. Ibanez, V. N. & Quadrana, L. Shaping inheritance: how distinct reproductive strategies influence DNA methylation memory in plants. *Curr. Opin. Genet Dev.* **78**, 102018 (2023).

61. Alachiotis, N. & Pavlidis, P. RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Commun. Biol.* **1**, 79 (2018).
62. Pavlidis, P., Zivkovic, D., Stamatakis, A. & Alachiotis, N. SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.* **30**, 2224–2234 (2013).
63. Harris, A. M. & DeGiorgio, M. A likelihood approach for uncovering selective sweep signatures from haplotype data. *Mol. Biol. Evol.* **37**, 3023–3046 (2020).
64. Demko, V., Ako, E., Perroud, P. F., Quatrano, R. & Olsen, O. A. The phenotype of the CRINKLY4 deletion mutant of *Physcomitrella patens* suggests a broad role in developmental regulation in early land plants. *Planta* **244**, 275–284 (2016).
65. Braud, C., Zheng, W. & Xiao, W. Identification and analysis of LNO1-like and AtGLE1-like nucleoporins in plants. *Plant Signal Behav.* **8**, e27376 (2013).
66. Zhao, H., Xing, D. & Li, Q. Q. Unique features of plant cleavage and polyadenylation specificity factor revealed by proteomic studies. *Plant Physiol.* **151**, 1546–1556 (2009).
67. Takatsuka, H., Umeda-Hara, C. & Umeda, M. Cyclin-dependent kinase-activating kinases CDKD;1 and CDKD;3 are essential for preserving mitotic activity in *Arabidopsis thaliana*. *Plant J.* **82**, 1004–1017 (2015).
68. Johnson, K. L., Kibble, N. A., Bacic, A. & Schultz, C. J. A fasciclin-like arabinogalactan-protein (FLA) mutant of *Arabidopsis thaliana*, fla1, shows defects in shoot regeneration. *PLoS One* **6**, e25154 (2011).
69. Zhu, M. et al. Robust organ size requires robust timing of initiation orchestrated by focused auxin and cytokinin signalling. *Nat. Plants* **6**, 686–698 (2020).
70. Zhao, H. et al. The *Arabidopsis thaliana* nuclear factor Y transcription factors. *Front. Plant Sci.* **7**, 2045 (2016).
71. Chantha, S. C., Gray-Mitsumune, M., Houde, J. & Matton, D. P. The MIDASIN and NOTCHLESS genes are essential for female gametophyte development in *Arabidopsis thaliana*. *Physiol. Mol. Biol. Plants* **16**, 3–18 (2010).
72. Chen, X. et al. Full-length EFOP3 and EFOP4 proteins are essential for pollen intine development in *Arabidopsis thaliana*. *Plant J.* **115**, 37–51 (2023).
73. Zhou, Y. et al. Members of the ELMOD protein family specify formation of distinct aperture domains on the *Arabidopsis* pollen surface. *eLife* **10**, e71061 (2021).
74. Jossier, M. et al. The *Arabidopsis* vacuolar anion transporter, AtCLCc, is involved in the regulation of stomatal movements and contributes to salt tolerance. *Plant J.* **64**, 563–576 (2010).
75. Gachomo, E. W., Jimenez-Lopez, J. C., Baptiste, L. J. & Kotchoni, S. O. GIGANTUS1 (GTS1), a member of Transducin/WD40 protein superfamily, controls seed germination, growth and biomass accumulation through ribosome-biogenesis protein interactions in *Arabidopsis thaliana*. *BMC Plant Biol.* **14**, 37 (2014).
76. Skalitzky, C. A. et al. Plastids contain a second sec translocase system with essential functions. *Plant Physiol.* **155**, 354–369 (2011).
77. Jeon, Y., Ahn, H. K., Kang, Y. W. & Pai, H. S. Functional characterization of chloroplast-targeted RbgA GTPase in higher plants. *Plant Mol. Biol.* **95**, 463–479 (2017).
78. McConnell, J. R. et al. Role of PHABULOSA and PHAVOLUTA in determining radial patterning in shoots. *Nature* **411**, 709–713 (2001).
79. Schwenk, P. et al. Uncovering a novel function of the CCR4-NOT complex in phytochrome A-mediated light signalling in plants. *eLife* **10**, e63697 (2021).
80. Farkas, I., Dombradi, V., Miskei, M., Szabados, L. & Koncz, C. *Arabidopsis* PPP family of serine/threonine phosphatases. *Trends Plant Sci.* **12**, 169–176 (2007).
81. Guo, Z. F., Wang, X. Y., Hu, Z. B., Wu, C. Y. & Shen, Z. G. The pentatricopeptide repeat protein GEND1 is required for root development and high temperature tolerance in *Arabidopsis thaliana*. *Biochem. Biophys. Res. Commun.* **578**, 63–69 (2021).
82. Mochizuki, S. et al. The *Arabidopsis* WAVY GROWTH 2 protein modulates root bending in response to environmental stimuli. *Plant Cell* **17**, 537–547 (2005).
83. Liu, C. H. et al. Repair of dna damage induced by the cytidine analog zebularine requires atr and atm in *Arabidopsis*. *Plant Cell* **27**, 1788–1800 (2015).
84. Bleuyard, J. Y. & White, C. I. The *Arabidopsis* homologue of Xrcc3 plays an essential role in meiosis. *EMBO J.* **23**, 439–449 (2004).
85. Lim, M. H. et al. A new *Arabidopsis* gene, FLK, encodes an RNA binding protein with K homology motifs and regulates flowering time via FLOWERING LOCUS C. *Plant Cell* **16**, 731–740 (2004).
86. Disch, S. et al. The E3 ubiquitin ligase BIG BROTHER controls *Arabidopsis* organ size in a dosage-dependent manner. *Curr. Biol.* **16**, 272–279 (2006).
87. Li, H. F. et al. The AGL6-like gene OsMADS6 regulates floral organ and meristem identities in rice. *Cell Res* **20**, 299–313 (2010).
88. Krizek, B. A. & Meyerowitz, E. M. The *Arabidopsis* homeotic genes APETALA3 and PISTILLATA are sufficient to provide the B class organ identity function. *Development* **122**, 11–22 (1996).
89. Lee, S., Choi, S. C. & An, G. Rice SVP-group MADS-box proteins, OsMADS22 and OsMADS55, are negative regulators of brassinosteroid responses. *Plant J.* **54**, 93–105 (2008).
90. Fang, W. J., Wang, Z. B., Cui, R. F., Li, J. & Li, Y. H. Maternal control of seed size by EOD3/CYP78A6 in *Arabidopsis thaliana*. *Plant J.* **70**, 929–939 (2012).
91. Sotelo-Silveira, M. et al. Cytochrome P450 CYP78A9 is involved in *Arabidopsis* reproductive development. *Plant Physiol.* **162**, 779–799 (2013).
92. Qi, X. L., Liu, C. L., Song, L. L., Li, Y. H. & Li, M. *Pacyp78a9*, a cytochrome P450, regulates fruit size in sweet cherry (*Prunus avium* L.). *Front Plant Sci.* **8**, 2076 (2017).
93. Ellegren, H. & Galtier, N. Determinants of genetic diversity. *Nat. Rev. Genet* **17**, 422–433 (2016).
94. Zhou, Y. F. et al. The population genetics of structural variants in grapevine domestication. *Nat. Plants* **5**, 965–979 (2019).
95. Guan, J. et al. Genome structure variation analyses of peach reveal population dynamics and a 1.67 Mb causal inversion for fruit shape. *Genome Biol.* **22**, 13 (2021).
96. Underwood, C. J. et al. Epigenetic activation of meiotic recombination near *Arabidopsis thaliana* centromeres via loss of H3K9me2 and non-CG DNA methylation. *Genome Res.* **28**, 519–531 (2018).
97. Santamaria, L. Why are most aquatic plants widely distributed? dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oecol.* **23**, 137–154 (2002).
98. Wang, Y. J. et al. Invasive alien plants benefit more from clonal integration in heterogeneous environments than natives. *N. Phytol.* **216**, 1072–1078 (2017).
99. Gutekunst, J. et al. Clonal genome evolution and rapid invasive spread of the marbled crayfish. *Nat. Ecol. Evol.* **2**, 567–573 (2018).
100. Appenroth, K.J.; et al. Photophysiology of turion formation and germination in Spirodela polyrhiza. *Biol. Plant.* **38**, 95–106 (1996)
101. Schubert, M., Lindgreen, S. & Orlando, L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res. Notes* **9**, 88 (2016).
102. Cao, H. X. et al. The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. *N. Phytol.* **209**, 354–363 (2016).
103. Li, H. et al. The sequence alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
104. McKenna, A. et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

105. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
106. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
107. Nelson, C. W., Moncla, L. H. & Hughes, A. L. SNPGenie: estimating evolutionary parameters to detect natural selection using pooled next-generation sequencing data. *Bioinformatics* **31**, 3709–3711 (2015).
108. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
109. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
110. Raj, A., Stephens, M. & Pritchard, J. K. fastStructure: variational inference of population structure in large SNP data sets. *Genetics* **197**, 573–U207 (2014).
111. Zhang, C., Dong, S. S., Xu, J. Y., He, W. M. & Yang, T. L. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**, 1786–1788 (2019).
112. Camacho, C. et al. BLAST +: architecture and applications. *BMC Bioinform.* **10**, 421 (2009).
113. Yin, J. M. et al. A high-quality genome of taro (*Colocasia esculenta*(L.) Schott), one of the world's oldest crops. *Mol. Ecol. Resour.* **21**, 68–77 (2021).
114. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**, 4453–4455 (2019).
115. Flouri, T. et al. The phylogenetic likelihood library. *Syst. Biol.* **64**, 356–362 (2015).
116. Darriba, D. et al. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
117. Letunic, I. & Bork, P. Interactive tree Of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
118. Huerta-Cepas, J., Dopazo, J. & Gabaldon, T. ETE: a python environment for tree exploration. *BMC Bioinforma.* **11**, 24 (2010).
119. Meinke, D. W. Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in *Arabidopsis*. *N. Phytol.* **226**, 306–325 (2020).
120. Racimo, F. Testing for ancient selection using cross-population allele frequency differentiation. *Genetics* **202**, 733–750 (2016).
121. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
122. Schmitz, R. J. et al. Patterns of population epigenomic diversity. *Nature* **495**, 193–198 (2013).
123. Kawakatsu, T. et al. Epigenomic diversity in a global collection of arabidopsis thaliana accessions. *Cell* **166**, 492–505 (2016).
124. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
125. Wang, W. Q. & Messing, J. High-throughput sequencing of three lemnoideae (duckweeds) chloroplast genomes from total DNA. *PLoS One* **6**, e24670 (2011).
126. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* **28**, 583–585 (2012).
127. Akalin, A. et al. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. *Genome. Biol.* **13**, R87 (2012).
128. Gel, B. et al. regioneR: an R/Bioconductor package for the association analysis of genomic regions based on permutation tests. *Bioinformatics* **32**, 289–291 (2016).
129. Huang, X. S., Zhang, S. L., Li, K. Q., Thimmapuram, J. & Xie, S. J. ViewBS: a powerful toolkit for visualization of high-throughput bisulfite sequencing data. *Bioinformatics* **34**, 708–709 (2018).
130. Yu, G. C., Lam, T. T. Y., Zhu, H. C. & Guan, Y. Two methods for mapping and visualizing associated data on phylogeny using ggtree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).
131. Wang, L. G. et al. Treeio: An R package for phylogenetic tree input and output with richly annotated and associated data. *Mol. Biol. Evol.* **37**, 599–603 (2020).
132. Paradis, E., Claude, J. & Strimmer, K. APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).
133. Revell, L. J. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).

Acknowledgements

We thank Martin Schäfer, Marie Sárazová, and Laura Böttner for supporting plant sample maintenance and DNA isolations. We thank Arturo Mari-Ordonez and Pavlos Pavlidis for valuable comments and Alex Widmer for contributing resources at the early stage of this project. This project is supported by the German Research Foundation (427577435 and 438887884 to S. X, and 422213951 to M. Hu), the Center for Adaptation to a Changing Environment (ACE) at ETH Zurich (to S. X.), the Swiss National Science Foundation (P400PB_186770 to M. Hu.), the Volkswagen Foundation (97236 to M. Hu.) and through career development measures of the University of Münster (to M. Hu.) The project was inspired by discussions with the members of the CRC TRR 212 (NC3) – Project number 316099922, and Research Training Group 2526 (GenEvo) – Project number 407023052. Parts of this research were conducted using the supercomputer Mogon and/or advisory services offered by the Johannes Gutenberg University Mainz (hpc.uni-mainz.de), which is a member of the AHRP (Alliance for High-Performance Computing in Rhineland Palatinate, www.ahrp.info) and the Gauss Alliance e.V. The authors gratefully acknowledge the computing time granted on the supercomputer Mogon at the Johannes Gutenberg University Mainz (hpc.uni-mainz.de) and PALMA-II at the University of Münster.

Author contributions

Y. W., P. D. and S. X. performed data analysis. A. C. and M. Ho. performed the experiments. K. J. A., H. Z., K. S. S., and S.X. contributed to the giant duckweed collections and resources. S. X. and M. Hu. conceived and supervised the project. S. X., Y. W., P. D., A. C. and M. Hu. wrote the manuscript. All authors contributed to the final version of the manuscript.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s42003-024-06266-7.

Correspondence and requests for materials should be addressed to Shuqing Xu.

Peer review information *Communications Biology* thanks Yang Jae Kang, Kent Holsinger and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Primary Handling Editor: George Inglis. A peer review file is available.

4 Conclusion and Outlook

4.1 Concept of Non-Targeted-Site Resistance Evolution in Plants

In line with earlier reports (Goggin et al. 2018; Szigeti 2005), we discovered a large amount of phenotypic variation in herbicide resistance in *S. polyrhiza*. The first report on natural variation in herbicide resistance dates back to the year 1957, documenting a case of resistance to 2,4-D in *Daucus carota* (Switzer 1957; Whitehead and Switzer 1963). Since 2,4-D was commercially introduced in modern agriculture around ten years before (Peterson et al. 2015), the emergence of this first resistance case indicates that herbicide resistance evolution occurs in short evolutionary time scales. In the case of *Raphanus raphanistrum* variation in resistance mechanisms was also discovered in several ecotypes coexisting in one location (Goggin et al. 2018). Since susceptible populations also varied in terms of herbicide degradation speed (Goggin et al. 2018), these findings suggest the possibility of natural variation in herbicide resistance prior to herbicide introduction in modern agriculture.

A possible explanation for this is that adaption to environmental stress factors was also shown to influence the efficiency of herbicide-based weed control (Radchenko et al. 2021). In this regard, several environmental adaptions were shown to influence herbicide resistance. For instance, increased leaf thickness, which is an adaption to elevation (see section 1.1) (Halbritter et al. 2018) and stomatal closure were shown to reduce the uptake of glyphosate (Ziska and Teasdale 2000; Kumar et al. 2023). Since temperature strongly influences degradation kinetics and phloem transport of herbicides (Kumar et al. 2023; Godar et al. 2015), temperature-driven adaption of these processes could have a potentially strong impact on herbicide efficiency. Also, plants were shown to have decreased requirements of aromatic amino acids for protein synthesis under high CO_2 levels (Taub et al. 2008; Kumar et al. 2023). Since glyphosate acts through inhibition of aromatic acid biosynthesis, decreased protein synthesis in adaption to high CO_2 levels might promote the evolution of glyphosate resistance (Kumar et al. 2023).

On a molecular level, plant stress responses to factors like drought, water stress, UV-radiation and herbicides follow the same scheme, which usually involves four steps of stress adaption: alarm, acclimation, resistance, exhaustion/recovery (Keith et al. 2017). Consequently, plants with upregulated stress responses also showed increased resistance to herbicides (Keith et al. 2017; Dyer 2018; Lee et al. 2007; Kwon et al. 2003). This mechanistic similarity suggests an evolution of herbicide resistance through pleiotropy from genes coordinating stress responses (Figure 5). As a consequence of these molecular changes in stress-responses, several cross-

resistances between herbicides and non-anthropogenic stress factors were reported. For instance, paraquat resistant plants showed elevated resistance to heat stress (Tang et al. 2006) and H_2O_2 (Kwon et al. 2003). Together, these strong mechanistic similarities suggest an evolution of herbicide resistance mechanisms from stress-response pathways (Figure 5).

Since changes in stress-response are polygenic and, therefore, evolve gradually (Keith et al. 2017), the identification of multiple intermediate resistance levels in our study (Höfer et al. 2024a) highlights that future studies on the evolution of herbicide resistance cases should focus on a large-scale characterization of genotypes with intermediate resistance levels. As resistance traits can be continuous (Goggin et al. 2018), just studying extreme phenotypes, as done in many previous studies (Dinelli et al. 2006; Ge et al. 2010; Jóri et al. 2007), will therefore give an incomplete picture of the trajectory of resistance evolution. Also, because resistance modes can vary on species level (Szigeti 2005), characterization of intermediate resistant genotypes might reveal which environmental settings favor the evolution of certain resistance modes.

Since herbicide resistance cases are causing huge economical costs (Oerke 2006), several strategies can be applied to avoid resistance evolution. One strategy could be to not or only minimally rely on herbicide-based control and instead focus on traditional techniques like the removal of weeds (Vats 2015) or microbial biological control strategies (Heap and Duke 2018) as alternatives. Since resistant weed populations only cause damage when spreading significantly in natural populations, monitoring of resistant weeds in populations via molecular genetic methods might become crucial in modern agriculture. Once a critical distribution of resistant weeds is reached, stubble burning can be applied to slow down the spread (Zia-Ul-Haq et al. 2019). Frequent crop rotations, which disrupt the life cycles of resistant weeds through tilling of land, can help to prevent the spread of resistant weeds (Vats 2015). In a similar way, the rotation of herbicides with different sites and modes of action can prevent the evolution of target-site resistance cases (Vats 2015; Heap and Duke 2018).

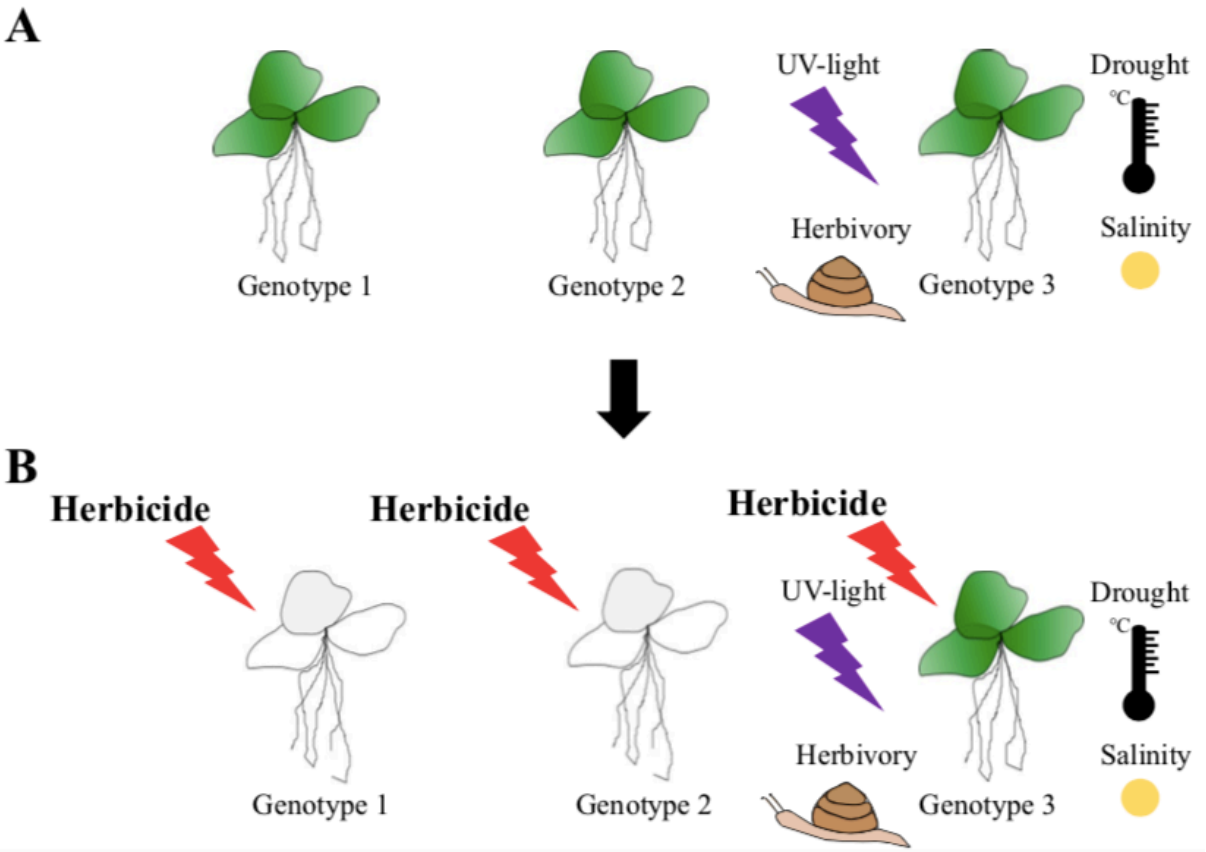


Figure 5: Theory for explaining the origin of natural variation in herbicide tolerance for *Spirodela polyrhiza*. A: Among different genotypes of *S. polyrhiza*, genotype 3 shows the best adaption to environmental stress factors such as UV-light, herbivory, drought, and salinity, due to faster upregulation or higher constitutive expression of intrinsic stress response pathways. B: Through effects like the production of reactive oxygen species, leaf desiccation and chlorosis herbicides cause similar stress symptoms as non-anthropogenic stress factors. Therefore, plants showing increased tolerance to environmental stress factors like genotype 3 are more likely to survive doses of herbicides, lethal to less well-adapted plants like genotypes 1 and 2.

4.2 Optimization Strategies for Growth and Defense

Besides controlling for herbicide-resistant weeds, optimization of balance between growth and specialized metabolite production is of central importance for maximizing crop yield (Huot et al. 2014). According to the resource-drain hypothesis, one would expect strong negative associations between all specialized metabolites on the one site with growth and primary metabolites on the other site. However, many previous studies, including our own study of

metabolite levels in *S. polyrhiza* (Höfer et al. 2024b) found multiple positive associations of specialized metabolite contents with plant growth and primary metabolism (Aftab et al. 2010; Zhang et al. 2024; Ringli et al. 2008), suggesting that the concept of growth-defense tradeoffs does not apply strictly for plants.

An explanation for this is, that costs for the synthesis of secondary metabolites are very unequal among specialized metabolite classes (Neilson et al. 2013; Foyer et al. 2007), which is why certain defensive metabolites like cyanogenic glycosides also play an important role as sinks for carbohydrates and nitrogen (Neilson et al. 2013). Apart from this, growth reduction during exposure to stresses was hypothesized to be caused by a genetic mechanism to protect the replication machinery of DNA from damage during this vulnerable stage (Foyer et al. 2007). Here, often genes regulating plant immunity and pathogen resistance mediate growth-defense tradeoffs (Gao et al. 2024). An example for this are *NLR*s (*NUCLEOTIDE-BINDING LEUCINE-RICH REPEAT*), which encode for receptors recognizing pathogen effectors (Wang et al. 2019a). Overexpression of *NLR*s in susceptible cultivars confers resistance to blast disease in rice but causes dwarf phenotypes through growth delay (Wang et al. 2019a; Gao et al. 2024). The pleiotropic character of some immune genes offers the possibility to improve growth and defense at the same time. The first pleiotropic gene discovered to improve both growth and defense simultaneously was *IDEAL-PLANT ARCHITECTURE 1* (*IPA1*), which encodes a regulator of plant immunity (Wang et al. 2018). During pathogen infection *IPA1* is activated through phosphorylation and initiates the induction of NLR-mediated immunity (Wang et al. 2018). Whereas in the absence of pathogens, IPA1 is dephosphorylated and promotes plant fertility through increasing grain yield (Wang et al. 2018). Overexpression of *IPA1* leads to an increase in the magnitude of this response, increasing both yield and plant resistance to fungal infection (Wang et al. 2018).

Another possibility for overcoming growth-defense tradeoffs is to express pathogen resistance-conferring genes under the control of disease-inducible promotors. Since permanent expression of pathogen resistance genes decreases plant fitness, growth-defense tradeoffs can be overcome through transcriptional or translation suppression of defense-associated genes under constitutive conditions. As an example for this, the facultative translational suppression during pathogen absence of the immune gene *NONEXPRESSOR OF PATHOGENESIS-RELATED GENES 1* (*NPR1*) leads to increased resistance without additional fitness costs (Xu et al. 2017). Growth-defense tradeoffs can also be influenced through the suppression of certain regulatory micro-RNAs targeting RNA-induced silencing complex proteins. Here, suppression of miR168

which targets the argonaut protein *AGO1* was shown to improve yield, shorten flowering time and increase immunity simultaneously (Wang et al. 2021).

Together, these findings suggest that growth-defense trade-offs can be optimized in plants through reverse engineering of pleiotropic gene candidates controlling plant immune responses.

4.3 Genetic Control of Immunity and Growth in *S. polyrhiza*

From all GWAS candidates associated with metabolic traits, we identified a total of five candidate genes to be under population-wide selection within the SE-Asia or India duckweed populations (Table 1). For two of these candidates (*SpGA2022_010583*, *SpGA2022_007797*) no functional homologs were identified (Table 1). Homologs of the remaining three candidate genes showed potential involvement in pathogen resistance (*SpHPL*, *SpSCAMP3*) (Wang et al. 2020; Lu et al. 2021) and cell division (*SpYPEL*) (Hosono et al. 2010). All three candidates were associated with cytokinin levels. Since cytokinins play a crucial role in the allocation of defense resources (Brütting et al. 2017), these candidate genes might explain natural variation in duckweed pathogen resistance. This would make them potential candidates for optimization of growth-defense tradeoffs in *S. polyrhiza*.

Knock out of the *HYDROXY-PEROXIDE LYASE 3* (*HPL3*), an homolog of *SpHPL*, leads to the induction of defense gene expression and increased resistance to herbivorous insects at the expense of decreased seed size and seed weight (Wang et al. 2020). Functionally, HPL3 plays a central role in the oxylipin pathway, producing green-leave volatiles (Farmer and Goossens 2019). Thereby, HPL competes with the allene oxide synthase (AOS) for their common substrate 13-hydroperoxy octadecatrienoic (Farmer and Goossens 2019). Since AOS is involved in the biosynthesis of jasmonic acid, the loss of function of HPL3 is thought to trigger the observed allocation of resources from growth to defense through an increase in jasmonate levels (Wang et al. 2020).

In contrast to this *SCAMP3* (*SECRETORY CARRIER MEMBRANE PROTEIN 3*) was shown to function as a host factor regulating virus replication (Lu et al. 2021). Located within Golgi membranes, SCAMP3 participates in intracellular protein transport through vesicles (Lu et al. 2021). Such vesicular transport processes are relevant for defense functions in two ways. First, vesicles can act through the secretion of immune molecules in extracellular space for pathogen defense (Yun et al. 2023). Second, vesicles can function as replication organelles for viruses (Lu et al. 2021). Since suppression of *SCAMP3* expression decreases virus replication, it is

thought to play a central role in the interaction between viral surface protein and the vesicles functioning as replication organelles (Lu et al. 2021).

With *SpYPEL* (*YIPPEE-LIKE PROTEIN*), we identified a gene whose homologs were functionally involved in cell cycle progression and growth (Hosono et al. 2010; Han et al. 2018). *YPEL*s are thought to have important functions in mitosis, since they were shown to move from the nucleus to the spindle midzone during initiation of cell division (Hosono et al. 2010). Because of that, knockouts of *YPEL*s were shown to lead to growth defects through reduction of cell division rate on a systemic level (Hosono et al. 2010; Han et al. 2018).

Since *SpHPL*, *SpSCAMP3* and *SpYPEL* were identified to be under population-wide selection and markers in close distance to these genes were associated with cytokinin levels, which control resource allocation to either growth or defense, our results indicate that functional variation in these genes might be relevant for genotype-specific differences in growth and defense for *S. polyrhiza*. However, further reverse genetic studies are required to elucidate the exact potential of these genes for the optimization of growth-defense tradeoffs in *S. polyrhiza*.

Table 1: GWAS candidates under population-wide selection. All candidate genes were identified in GWAS conducted on free metabolite contents of *S. polyrhiza* (Höfer et al. 2024b).

Gene	Annotation	Marker-type	Marker-location	*P*-Wald	Associated metabolite(s)	Population-wide selection
SpGA2022_009939	*SpYPEL* (*YIPPEE-LIKE PROTEIN*)	54 bp deletion	10473 bp upstream of *SpGA2022_009939*	$2.16 \cdot 10^{-6}$	N^6-Isopentenyl adenin	SE-Asia
SpGA2022_054953	*SpSCAMP3* (*SECRETORY CARRIER-ASSOCIATED MEMBRANE PROTEIN* 3)	76 bp duplication	3'-UTR of *SpGA2022_054953*	$6.73 \cdot 10^{-7}$	cis-Zeatin	India
SpGA2022_010583	-	G/C - transversion	Intronic region of *SpGA2022_010583*	$3.51 \cdot 10^{-6}$	Chlorogenic acid	SE-Asia
SpGA2022_007797	-	T/C - transition	1567 bp upstream of *SpGA2022_007797*	$1.45 \cdot 10^{-6}$	N^6-Isopentenyl adenin	India
SpGA2022_008667	*SpHPL* (*HYDROPEROXIDE LYASE*)	T/C - transition	13696 bp upstream of *SpGA2022_008667*	$2.29 \cdot 10^{-9}$, $2.65 \cdot 10^{-9}$, $7.34 \cdot 10^{-11}$, $1.14 \cdot 10^{-7}$	N^6-Isopentenyl adenin, N^6-Isopentenyl adenin – riboside, trans-Zeatin, trans-Zeatin – riboside	SE-Asia

4.4 Studying the Evolution of Plant Mating Systems

Overall, our work identified many genes controlling mating system to be under population- or species-wide selection, suggesting population-specific differences in flowering rates in S. *polyrhiza* (Wang et al. 2024). Strong population-based variation in sexual reproduction rates was already described previously (Whitehead et al. 2018), indicating that most plants don't follow a single mating strategy, exclusively.

Such intraspecific variations in plant mating systems provide the basis for future selection pressures can act on. For example, human land-use often affects habitat fragmentation in natural ecosystems, impacting the evolution of plant mating systems through changes in plant-pollinator interactions (Eckert et al. 2010). Increased grazing during agricultural land-use was shown to decrease outcrossing (Stout 2014). Inbreeding caused by reduced outcrossing rates might then promote herbicide resistance by applying stronger purifying selection on susceptible alleles (Kuester et al. 2017)

Also, global warming-driven migration events of warm-adapted species in temperate climates cause an overall biotic homogenization (Montràs-Janer et al. 2024). During rapid species

expansion, clonally reproducing species might be favored (Wang et al. 2024), as discussed in section 1.2.2. A consequence of clonal reproduction is a low mutation rate and reduced diversity (Baggs et al. 2022). The accumulation of deleterious mutations due to clonal growth causes gene losses, leading to trait reduction in plants (Baggs et al. 2022; Johnson et al. 2015).

Besides these above-mentioned predictions on future evolutionary trajectories, many aspects of mating system evolution are still unclear, hampering our understanding on evolution of new plant phenotypes.

Many plant reproductive traits are plastic and are strongly influenced by environmental features (Stearns 1989). However, the identification of natural flowering conditions in *S. polyrhiza* might provide valuable insights into the evolution of mating systems in Lemnaceae. Since *S. polyrhiza* is world-wide distributed (Tippery and Les 2020), flowering traits might be induced by various environmental setups depending on the specific genotype, making it difficult to explore environmental conditions that promote flowering in nature on the species level. One way to experimentally study the impact of environmental conditions on flowering in *S. polyrhiza* would be through a translocation experiment (Whitehead et al. 2018) to expose various genotypes from different genetic populations to a broad array of natural environments and monitor flowering rates in nature.

Another important aspect determining the success of sexual reproduction in plants is the seed germination rate, which was shown to vary greatly among duckweed species (Landolt 1986). Since germinating seeds differ from non-germinating seeds regarding their microbial compositions (Acuña et al. 2023), the presence of microbiota benefitting seed germination might be a key factor controlling mating system evolution in plants. In *D. carota*, bacteria associated with pathogen defense and nitrate reduction were associated with seed germination (Acuña et al. 2023), indicating that microbial communities can shape the duration and efficiency of plant sexual reproduction through seed germination. Future work should, therefore, also focus on how bacterial communities shape seed germination in duckweed.

Another important question is how stochasticity shapes the evolution of mating systems. For instance, bottle neck effects were shown to increase stochasticity in mating system evolution, promoting inbreeding effects (Karron et al. 2011). This suggests that reproductive systems can also evolve randomly in nature depending on the population size.

Finally, deeper understatement in the genetic basis of the duckweed mating system will require access to reverse genetic methods, allowing knock-out and overexpression of the identified gene candidates to verify their exact functions in duckweed sexual reproduction.

5. References

Acosta K, Appenroth KJ, Borisjuk L, Edelman M, Heinig U, Jansen MAK, Oyama T, Pasaribu B, Schubert I, Sorrels S, Sree KS, Xu S, Michael TP, Lam E (2021) Return of the Lemnaceae: duckweed as a model plant system in the genomics and postgenomics era. The Plant Cell 33 (10):3207-3234. doi:10.1093/plcell/koab189

Acuña JJ, Hu J, Inostroza NG, Valenzuela T, Perez P, Epstein S, Sessitsch A, Zhang Q, Jorquera MA (2023) Endophytic bacterial communities in ungerminated and germinated seeds of commercial vegetables. Scientific Reports 13 (1):19829. doi:10.1038/s41598-023-47099-4

Aftab T, Masroor M, Khan A, Idrees M, Naeem M, Moinuddin (2010) Salicylic acid acts as potent enhancer of growth, photosynthesis and artemisinin production in *Artemisia annua* L. Journal of Crop Science and Biotechnology 13 (3):183-188. doi:10.1007/s12892-010-0040-3

Ahmad N, Jianyu L, Xu T, Noman M, Jameel A, Na Y, Yuanyuan D, Nan W, Xiaowei L, Fawei W, Xiuming L, Haiyan L (2019) Overexpression of a Novel Cytochrome P450 Promotes Flavonoid Biosynthesis and Osmotic Stress Tolerance in Transgenic Arabidopsis. Genes 10 (10). doi:10.3390/genes10100756

Alvarez C, Calo L, Romero LC, García I, Gotor C (2010) An O-acetylserine(thiol)lyase homolog with L-cysteine desulfhydrase activity regulates cysteine homeostasis in Arabidopsis. Plant Physiology 152 (2):656-669. doi:10.1104/pp.109.147975

Angelovici R, Batushansky A, Deason N, Gonzalez-Jorge S, Gore MA, Fait A, DellaPenna D (2017) Network-Guided GWAS Improves Identification of Genes Affecting Free Amino Acids. Plant Physiology 173 (1):872-886. doi:10.1104/pp.16.01287

Angelovici R, Lipka AE, Deason N, Gonzalez-Jorge S, Lin H, Cepela J, Buell R, Gore MA, Dellapenna D (2013) Genome-wide analysis of branched-chain amino acid levels in Arabidopsis seeds. The Plant Cell 25 (12):4827-4843. doi:10.1105/tpc.113.119370

Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005) Genome-wide association mapping in Arabidopsis identifies previously known flowering time and pathogen resistance genes. PLOS Genetics 1 (5):e60. doi:10.1371/journal.pgen.0010060

Backes C, Rühle F, Stoll M, Haas J, Frese K, Franke A, Lieb W, Wichmann HE, Weis T, Kloos W, Lenhof H-P, Meese E, Katus H, Meder B, Keller A (2014) Systematic permutation testing in GWAS pathway analyses: identification of genetic networks in dilated cardiomyopathy and ulcerative colitis. BMC Genomics 15 (1):622. doi:10.1186/1471-2164-15-622

Baggs EL, Tiersma MB, Abramson BW, Michael TP, Krasileva KV (2022) Characterization of defense responses against bacterial pathogens in duckweeds lacking EDS1. New Phytologist 236 (5):1838-1855. doi:10.1111/nph.18453

Barrett SC (2015) Influences of clonality on plant sexual reproduction. Proceedings of the National Academy of Sciences USA 112 (29):8859-8866. doi:10.1073/pnas.1501712112

Bog M, Sree KS, Fuchs J, Hoang PTN, Schubert I, Kuever J, Rabenstein A, Paolacci S, Jansen MAK, Appenroth K-J (2020) A taxonomic revision of Lemna sect. Uninerves (Lemnaceae). TAXON 69 (1):56-66. doi:10.1002/tax.12188

Böttner L, Grabe V, Gablenz S, Böhme N, Appenroth KJ, Gershenzon J, Huber M (2021) Differential localization of flavonoid glucosides in an aquatic plant implicates different functions under abiotic stress. Plant, Cell & Environment 44 (3):900-914. doi:10.1111/pce.13974

Brown JE, Khodr H, Hider RC, Rice-Evans CA (1998) Structural dependence of flavonoid interactions with Cu2+ ions: implications for their antioxidant properties. Biochemical Journal 330 (3):1173-1178. doi:10.1042/bj3301173

Brütting C, Schäfer M, Vanková R, Gase K, Baldwin IT, Meldau S (2017) Changes in cytokinins are sufficient to alter developmental patterns of defense metabolites in *Nicotiana attenuata*. The Plant Journal 89 (1):15-30. doi:10.1111/tpj.13316

Campbell SA, Kessler A (2013) Plant mating system transitions drive the macroevolution of defense strategies. Proceedings of the National Academy of Sciences USA 110 (10):3973-3978. doi:10.1073/pnas.1213867110

Cedergreen N, Abbaspoor M, Sørensen H, Streibig JC (2007) Is mixture toxicity measured on a biomarker indicative of what happens on a population level? A study with *Lemna minor*. Ecotoxicology and Environmental Safety 67 (3):323-332. doi:10.1016/j.ecoenv.2006.12.006

Che R, Jack JR, Motsinger-Reif AA, Brown CC (2014) An adaptive permutation approach for genome-wide association study: evaluation and recommendations for use. BioData Mining 7 (1):9. doi:10.1186/1756-0381-7-9

Chen S, Dickman MB (2004) Bcl-2 family members localize to tobacco chloroplasts and inhibit programmed cell death induced by chloroplast-targeted herbicides. Journal of Experimental Botany 55 (408):2617-2623. doi:10.1093/jxb/erh275

Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, Li Y, Liu X, Zhang H, Dong H, Zhang W, Zhang L, Yu S, Wang G, Lian X, Luo J (2014) Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nature Genetics 46 (7):714-721. doi:10.1038/ng.3007

Chen W, Wang W, Peng M, Gong L, Gao Y, Wan J, Wang S, Shi L, Zhou B, Li Z, Peng X, Yang C, Qu L, Liu X, Luo J (2016) Comparative and parallel genome-wide association studies for metabolic and agronomic traits in cereals. Nature Communications 7:12767. doi:10.1038/ncomms12767

Chevin L-M, Lande R (2011) Adaptation to marginal habitats by evolution of increased phenotypic plasticity. Journal of Evolutionary Biology 24:1462-1476. doi:10.1111/j.1420-9101.2011.02279.x

Chou Q, Cao T, Ni L, Xie P, Jeppesen E (2019) Leaf Soluble Carbohydrates, Free Amino Acids, Starch, Total Phenolics, Carbon and Nitrogen Stoichiometry of 24 Aquatic Macrophyte Species Along Climate Gradients in China. Frontiers in Plant Science 10. doi:10.3389/fpls.2019.00442

Chun JC, Cheol Kim J, Taek Hwang I, Eun Kim S (2002) Acteoside from *Rehmannia glutinosa* nullifies paraquat activity in *Cucumis sativus*. Pesticide Biochemistry and Physiology 72 (3):153-159. doi:10.1016/S0048-3575(02)00008-1

Comont D, Knight C, Crook L, Hull R, Beffa R, Neve P (2019) Alterations in Life-History Associated With Non-target-site Herbicide Resistance in *Alopecurus myosuroides*. Frontiers in Plant Science 10. doi:10.3389/fpls.2019.00837

Comont D, Lowe C, Hull R, Crook L, Hicks HL, Onkokesung N, Beffa R, Childs DZ, Edwards R, Freckleton RP, Neve P (2020) Evolution of generalist resistance to herbicide mixtures reveals a trade-off in resistance management. Nature Communications 11 (1):3086. doi:10.1038/s41467-020-16896-0

Cu ST, Warnock NI, Pasuquin J, Dingkuhn M, Stangoulis J (2021) A high-resolution genome-wide association study of the grain ionome and agronomic traits in rice *Oryza sativa* subsp. *indica*. Scientific Reports 11 (1):19230. doi:10.1038/s41598-021-98573-w

Dan L, Peng L, Zhiqiang Y, Na L, Lunguang Y, Lingling C (2022) Allelopathic inhibition of the extracts of *Landoltia punctata* on *Microcystis aeruginosa*. Plant Signaling & Behaviour 17 (1):2058256. doi:10.1080/15592324.2022.2058256

Dick R, Rattei T, Haslbeck M, Schwab W, Gierl A, Frey M (2012) Comparative Analysis of Benzoxazinoid Biosynthesis in Monocots and Dicots: Independent Recruitment of Stabilization and Activation Functions. The Plant Cell 24 (3):915-928. doi:10.1105/tpc.112.096461

Dijkstra PD, Maguire SM, Harris RM, Rodriguez AA, DeAngelis RS, Flores SA, Hofmann HA (2017) The melanocortin system regulates body pigmentation and social behaviour in a colour polymorphic cichlid fish. Proceedings of the Royal Society B: Biological Sciences 284 (1851):20162838. doi:10.1098/rspb.2016.2838

Dinelli G, Marotti I, Bonetti A, Minelli M, Catizone P, Barnes J (2006) Physiological and molecular insight on the mechanisms of resistance to glyphosate in *Conyza canadensis* (L.) Cronq. biotypes. Pesticide Biochemistry and Physiology 86 (1):30-41. doi:10.1016/j.pestbp.2006.01.004

Dong X, Li J, Zhang Y, Han D, Hua G, Wang J, Deng X, Wu C (2019) Genomic Analysis Reveals Pleiotropic Alleles at EDN3 and BMP7 Involved in Chicken Comb Color and Egg Production. Frontiers in Genetics 10:612. doi:10.3389/fgene.2019.00612

Duncan LE, Ostacher M, Ballon J (2019) How genome-wide association studies (GWAS) made traditional candidate gene studies obsolete. Neuropsychopharmacology 44 (9):1518-1523. doi:10.1038/s41386-019-0389-5

Dyer WE (2018) Stress-induced evolution of herbicide resistance and related pleiotropic effects. Pest Management Science 74 (8):1759-1768. doi:10.1002/ps.5043

Eckert CG, Kalisz S, Geber MA, Sargent R, Elle E, Cheptou P-O, Goodwillie C, Johnston MO, Kelly JK, Moeller DA, Porcher E, Ree RH, Vallejo-Marín M, Winn AA (2010) Plant mating systems in a changing world. Trends in Ecology & Evolution 25 (1):35-43. doi:10.1016/j.tree.2009.06.013

Fang C, Luo J (2019) Metabolic GWAS-based dissection of genetic bases underlying the diversity of plant metabolism. The Plant Journal 97 (1):91-100. doi:10.1111/tpj.14097

Farmer EE, Goossens A (2019) Jasmonates: what ALLENE OXIDE SYNTHASE does for plants. Journal of Experimental Botany 70 (13):3373-3378. doi:10.1093/jxb/erz254

Farrington JA, Ebert M, Land EJ, Fletcher K (1973) Bipyridylium quaternary salts and related compounds. V. Pulse radiolysis studies of the reaction of paraquat radical with oxygen. Implications for the mode of action of bipyridyl herbicides. Biochimica et Biophysica Acta 314 (3):372-381. doi:10.1016/0005-2728(73)90121-7

Filiault DL, Maloof JN (2012) A genome-wide association study identifies variants underlying the *Arabidopsis thaliana* shade avoidance response. PLOS Genetics 8 (3):e1002589. doi:10.1371/journal.pgen.1002589

Firmin A, Selosse M-A, Dunand C, Elger A (2022) Mixotrophy in aquatic plants, an overlooked ability. Trends in Plant Science 27 (2):147-157. doi: 10.1016/j.tplants.2021.08.011

Fourounjian P, Slovin J, Messing J (2021) Flowering and Seed Production across the Lemnaceae. International Journal of Molecular Sciences 22 (5):2733. doi: 10.3390/ijms22052733

Foyer CH, Noctor G (2005) Redox homeostasis and antioxidant signaling: a metabolic interface between stress perception and physiological responses. Plant Cell 17 (7):1866-1875. doi:10.1105/tpc.105.033589

Foyer CH, Noctor G, van Emden HF (2007) An evaluation of the costs of making specific secondary metabolites: Does the yield penalty incurred by host plant resistance to insects result from competition for resources? International Journal of Pest Management 53 (3):175-182. doi:10.1080/09670870701469146

Fujii T, Yokoyama E-i, Inoue K, Sakurai H (1990) The sites of electron donation of Photosystem I to methyl viologen. Biochimica et Biophysica Acta (BBA) - Bioenergetics 1015 (1):41-48. doi:10.1016/0005-2728(90)90213-N

Fukutoku Y, Yamada Y (1981) Diurnal changes in water potential and free amino acid contents of water-stressed and non-stressed soybean plants. Soil Science and Plant Nutrition 27:195-204. doi:10.1080/00380768.1981.10431271

Funderburk HH, Lawrence JM (1963) A Sensitive Method for Determination of Low Concentrations of Diquat and Paraquat. Nature 199 (4897):1011-1012. doi:10.1038/1991011a0

Gaines TA, Zhang W, Wang D, Bukun B, Chisholm ST, Shaner DL, Nissen SJ, Patzoldt WL, Tranel PJ, Culpepper AS, Grey TL, Webster TM, Vencill WK, Sammons RD, Jiang J, Preston C, Leach JE, Westra P (2010) Gene amplification confers glyphosate resistance in *Amaranthus palmeri*. Proceedings of the National Academy of Sciences USA 107 (3):1029-1034. doi:10.1073/pnas.0906649107

Gao M, Hao Z, Ning Y, He Z (2024) Revisiting growth–defence trade-offs and breeding strategies in crops. Plant Biotechnology Journal 22 (5):1198-1205. doi:10.1111/pbi.14258

Ge X, d'Avignon DA, Ackerman JJH, Sammons RD (2010) Rapid vacuolar sequestration: the horseweed glyphosate resistance mechanism. Pest Management Science 66 (4):345-348. doi:10.1002/ps.1911

Geisler K, Hughes RK, Sainsbury F, Lomonossoff GP, Rejzek M, Fairhurst S, Olsen CE, Motawia MS, Melton RE, Hemmings AM, Bak S, Osbourn A (2013) Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proceedings of the National Academy of Sciences USA 110 (35):E3360-3367. doi:10.1073/pnas.1309157110

Gion K, Inui H, Takakuma K, Yamada T, Kambara Y, Nakai S, Fujiwara H, Miyamura T, Imaishi H, Ohkawa H (2014) Molecular mechanisms of herbicide-inducible gene expression of tobacco CYP71AH11 metabolizing the herbicide chlorotoluron. Pesticide Biochemistry and Physiology 108:49-57. doi:10.1016/j.pestbp.2013.12.003

Godar AS, Varanasi VK, Nakka S, Prasad PV, Thompson CR, Mithila J (2015) Physiological and Molecular Mechanisms of Differential Sensitivity of Palmer Amaranth (*Amaranthus palmeri*) to Mesotrione at Varying Growth Temperatures. PLOS One 10 (5):e0126731. doi:10.1371/journal.pone.0126731

Goggin D, Kaur P, Owen M, Powles S (2018) 2,4-D and dicamba resistance mechanisms in wild radish: subtle, complex and population specific? Annals of Botany 122. doi:10.1093/aob/mcy097

Gordon WR, Koukkari WL (1978) Circadian Rhythmicity in the Activities of Phenylalanine Ammonia-Lyase from *Lemna perpusilla* and *Spirodela polyrhiza* 1. Plant Physiology 62 (4):612-615. doi:10.1104/pp.62.4.612

Gramzow L, Theißen G (2015) Phylogenomics reveals surprising sets of essential and dispensable clades of MIKC(c)-group MADS-box genes in flowering plants. Journal of Expermiental Zoology Part B, Molecular and Developmental Evolution 324 (4):353-362. doi:10.1002/jez.b.22598

Gramzow L, Theißen G (2020) Stranger than Fiction: Loss of MADS-Box Genes During Evolutionary Miniaturization of the Duckweed Body Plan. In: Cao XH, Fourounjian P, Wang W (eds) The Duckweed Genomes. Springer International Publishing, Cham, pp 91-101. doi:10.1007/978-3-030-11045-1_9

Halbritter AH, Fior S, Keller I, Billeter R, Edwards PJ, Holderegger R, Karrenberg S, Pluess AR, Widmer A, Alexander JM (2018) Trait differentiation and adaptation of plants along elevation gradients. Journal of Evolutionary Biology 31 (6):784-800. doi:10.1111/jeb.13262

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annual Review of Plant Biology 57:303-333. doi:10.1146/annurev.arplant.57.032905.105228

Han J-H, Shin J-H, Lee Y-H, Kim KS (2018) Distinct roles of the YPEL gene family in development and pathogenicity in the ascomycete fungus *Magnaporthe oryzae*. Scientific Reports 8 (1):14461. doi:10.1038/s41598-018-32633-6

Hawkes TR (2014) Mechanisms of resistance to paraquat in plants. Pest Management Science 70 (9):1316-1323. doi:10.1002/ps.3699

Heap I, Duke SO (2018) Overview of glyphosate-resistant weeds worldwide. Pest Management Science 74 (5):1040-1049. doi:10.1002/ps.4760

Hildebrandt Tatjana M, Nunes Nesi A, Araújo Wagner L, Braun H-P (2015) Amino Acid Catabolism in Plants. Molecular Plant 8 (11):1563-1579. doi:10.1016/j.molp.2015.09.005

Hilty J, Muller B, Pantin F, Leuzinger S (2021) Plant growth: the What, the How, and the Why. New Phytologist 232 (1):25-41. doi:10.1111/nph.17610

Ho EKH, Bartkowska M, Wright SI, Agrawal AF (2019) Population genomics of the facultatively asexual duckweed *Spirodela polyrhiza*. New Phytologist 224 (3):1361-1371. doi:10.1111/nph.16056

Hoang PTN, Schubert V, Meister A, Fuchs J, Schubert I (2019) Variation in genome size, cell and nucleus volume, chromosome number and rDNA loci among duckweeds. Scientific Reports 9 (1):3234. doi:10.1038/s41598-019-39332-w

Höfer M, Schäfer M, Wang Y, Wink S, Xu S (2024a) Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in *Spirodela polyrhiza*. Plants 13 (6):845. doi:10.3390/plants13060845

Höfer M, Schäfer M, Wang Y, Wink S, Xu S (2024b) Genome-Wide Association Study of Metabolic Traits in the Duckweed *Spirodela polyrhiza.* bioRxiv:2024.2007.2026.605148. doi:10.1101/2024.07.26.605148

Hosono K, Noda S, Shimizu A, Nakanishi N, Ohtsubo M, Shimizu N, Minoshima S (2010) YPEL5 protein of the YPEL gene family is involved in the cell cycle progression by interacting with two distinct proteins RanBPM and RanBP10. Genomics 96:102-111. doi:10.1016/j.ygeno.2010.05.003

Huang M, Liu X, Zhou Y, Summers RM, Zhang Z (2019) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8 (2):giy154. doi:10.1093/gigascience/giy154

Hui T, Zhang Y, Jia R, Hu Y, Wang W, Wang Y, Wang Y, Zhu Y, Yang L, Xiang B (2023) Metabolomic analysis reveals responses of *Spirodela polyrhiza* L. to salt stress. Journal of Plant Interactions 18. doi:10.1080/17429145.2023.2210163

Hunziker P, Lambertz SK, Weber K, Crocoll C, Halkier BA, Schulz A (2021) Herbivore feeding preference corroborates optimal defense theory for specialized metabolites within plants. Proceedings of the National Academy of Sciences USA 118 (47):e2111977118. doi:doi:10.1073/pnas.2111977118

Huot B, Yao J, Montgomery BL, He SY (2014) Growth–Defense Tradeoffs in Plants: A Balancing Act to Optimize Fitness. Molecular Plant 7 (8):1267-1287. doi:10.1093/mp/ssu049

Jansen MA, Shaaltiel Y, Kazzes D, Canaani O, Malkin S, Gressel J (1989) Increased Tolerance to Photoinhibitory Light in Paraquat-Resistant *Conyza bonariensis* Measured by Photoacoustic Spectroscopy and CO(2)-Fixation. Plant Physiology 91 (3):1174-1178. doi:10.1104/pp.91.3.1174

John M, Ankenbrand MJ, Artmann C, Freudenthal JA, Korte A, Grimm DG (2022) Efficient permutation-based genome-wide association studies for normal and skewed phenotypic distributions. Bioinformatics 38 (Suppl_2):ii5-ii12. doi:10.1093/bioinformatics/btac455

Johnson MTJ, Campbell SA, Barrett SCH (2015) Evolutionary Interactions Between Plant Reproduction and Defense Against Herbivores. Annual Review of Ecology, Evolution, and Systematics 46 (Volume 46, 2015):191-213. doi:10.1146/annurev-ecolsys-112414-054215

Jóri B, Soós V, Szegő D, Páldi E, Szigeti Z, Rácz I, Lásztity D (2007) Role of transporters in paraquat resistance of horseweed *Conyza canadensis* (L.) Cronq. Pesticide Biochemistry and Physiology 88 (1):57-65. doi:10.1016/j.pestbp.2006.08.013

Jugulam M, Shyam C (2019) Non-Target-Site Resistance to Herbicides: Recent Developments. Plants 8 (10):417. doi:10.3390/plants8100417

Karron JD, Ivey CT, Mitchell RJ, Whitehead MR, Peakall R, Case AL (2011) New perspectives on the evolution of plant mating systems. Annals of Botany 109 (3):493-503. doi:10.1093/aob/mcr319

Kartseva T, Aleksandrov V, Alqudah AM, Arif MAR, Kocheva K, Doneva D, Prokopova K, Börner A, Misheva S (2024) GWAS in a Collection of Bulgarian Old and Modern Bread Wheat Accessions Uncovers Novel Genomic Loci for Grain Protein Content and Thousand Kernel Weight. Plants 13 (8). doi:10.3390/plants13081084

Keith BK, Burns EE, Bothner B, Carey CC, Mazurie AJ, Hilmer JK, Biyiklioglu S, Budak H, Dyer WE (2017) Intensive herbicide use has selected for constitutively elevated levels of stress-responsive mRNAs and proteins in multiple herbicide-resistant *Avena fatua* L. Pest Management Science 73 (11):2267-2281. doi:10.1002/ps.4605

Khare D, Choi H, Huh SU, Bassin B, Kim J, Martinoia E, Sohn KH, Paek KH, Lee Y (2017) Arabidopsis ABCG34 contributes to defense against necrotrophic pathogens by mediating the secretion of camalexin. Proceedings of the National Academy of Sciences USA 114 (28):E5712-e5720. doi:10.1073/pnas.1702259114

Khurana JP, Maheshwari SC (1980) Some effects of salicylic acid on growth and flowering in *Spirodela polyrrhiza* SP20. Plant and Cell Physiology 21 (5):923-927. doi:10.1093/oxfordjournals.pcp.a076066

Kim DY, Bovet L, Maeshima M, Martinoia E, Lee Y (2007) The ABC transporter AtPDR8 is a cadmium extrusion pump conferring heavy metal resistance. The Plant Journal 50 (2):207-218. doi:10.1111/j.1365-313X.2007.03044.x

Kim DY, Jin JY, Alejandro S, Martinoia E, Lee Y (2010) Overexpression of AtABCG36 improves drought and salt stress resistance in Arabidopsis. Physiologia Plantarum 139 (2):170-180. doi:10.1111/j.1399-3054.2010.01353.x

Koo AJK, Cooke TF, Howe GA (2011) Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. Proceedings of the National Academy of Sciences USA 108 (22):9298-9303. doi:10.1073/pnas.1103542108

Kuester A, Fall E, Chang SM, Baucom RS (2017) Shifts in outcrossing rates and changes to floral traits are associated with the evolution of herbicide resistance in the common morning glory. Ecology Letters 20 (1):41-49. doi:10.1111/ele.12703

Kumar S, Pandey AK (2013) Chemistry and biological activities of flavonoids: an overview. The Scientific World Journal 2013:162750. doi:10.1155/2013/162750

Kumar V, Kumari A, Price AJ, Bana RS, Singh V, Bamboriya SD (2023) Impact of Futuristic Climate Variables on Weed Biology and Herbicidal Efficacy: A Review. Agronomy 13 (2):559. doi:10.3390/agronomy13020559

Kwon S-Y, Choi S-M, Ahn Y-O, Lee H-S, Lee H-B, Park Y-M, Kwak S-S (2003) Enhanced stress-tolerance of transgenic tobacco plants expressing a human dehydroascorbate reductase gene. Journal of Plant Physiology 160 (4):347-353. doi:10.1078/0176-1617-00926

Landolt E (1986) Biosystematic Investigations in the Family of Duckweeds, Lemnaceae: The family of Lemnaceae, a monographic study. Vol. 1: Morphology, Karyology, Ecology, Geographic Distribution, Systematic Position, Nomenclature, Descriptions. vol Bd. 2. Eidgenössische Technische Hochschule, Stiftung Rübel, Geobotanisches Institut: Zürich, Switzerland

Landolt E (1988) The family of Lemnaceae - Monographic study, Vols. 1 and 2 - (Vols. 2 and 4 of Biosystematic investigations in the family of duckweeds (Lemnaceae)). Plant Growth Regulation, vol 7. doi:10.1007/BF00037640

Lee YP, Kim SH, Bang JW, Lee HS, Kwak SS, Kwon SY (2007) Enhanced tolerance to oxidative stress in transgenic tobacco plants expressing three antioxidant enzymes in chloroplasts. Plant Cell Reports 26 (5):591-598. doi:10.1007/s00299-006-0253-z

Lewinsohn E, Gressel J (1984) Benzyl viologen-mediated counteraction of diquat and paraquat phytotoxicities. Plant Physiology 76 (1):125-130. doi:10.1104/pp.76.1.125

Li J, Halitschke R, Li D, Paetz C, Su H, Heiling S, Xu S, Baldwin IT (2021) Controlled hydroxylations of diterpenoids allow for plant chemical defense without autotoxicity. Science 371 (6526):255-260. doi:10.1126/science.abe4713

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. PLOS Genetics 12 (2):e1005767. doi:10.1371/journal.pgen.1005767

Lu J-Y, Brewer G, Li M-L, Lin K-Z, Huang C-C, Yen L-C, Lin J-Y (2021) Secretory Carrier Membrane Protein 3 Interacts with 3A Viral Protein of Enterovirus and Participates in Viral Replication. Microbiology Spectrum 9 (1):10.1128/spectrum.00475-00421. doi:10.1128/spectrum.00475-21

Maeda H, Murata K, Sakuma N, Takei S, Yamazaki A, Karim MR, Kawata M, Hirose S, Kawagishi-Kobayashi M, Taniguchi Y, Suzuki S, Sekino K, Ohshima M, Kato H, Yoshida H, Tozawa Y (2019) A rice gene that confers broad-spectrum

resistance to β-triketone herbicides. Science 365 (6451):393-396. doi:10.1126/science.aax0379

Mauro AA, Ghalambor CK (2020) Trade-offs, Pleiotropy, and Shared Molecular Pathways: A Unified View of Constraints on Adaptation. Integrative and Comparative Biology 60 (2):332-347. doi:10.1093/icb/icaa056

McDowell RW, Noble A, Pletnyakov P, Haggard BE, Mosley LM (2020) Global mapping of freshwater nutrient enrichment and periphyton growth potential. Scientific Reports 10 (1):3568. doi:10.1038/s41598-020-60279-w

Melicher P, Dvořák P, Krasylenko Y, Shapiguzov A, Kangasjärvi J, Šamaj J, Takáč T (2022) Arabidopsis Iron Superoxide Dismutase FSD1 Protects Against Methyl Viologen-Induced Oxidative Stress in a Copper-Dependent Manner. Frontiers in Plant Science 13:823561. doi:10.3389/fpls.2022.823561

Mernke D, Dahm S, Walker AS, Lalève A, Fillinger S, Leroch M, Hahn M (2011) Two promoter rearrangements in a drug efflux transporter gene are responsible for the appearance and spread of multidrug resistance phenotype MDR2 in *Botrytis cinerea* isolates in French and German vineyards. Phytopathology 101 (10):1176-1183. doi:10.1094/phyto-02-11-0046

Michael TP, Ernst E, Hartwick N, Chu P, Bryant D, Gilbert S, Ortleb S, Baggs EL, Sree KS, Appenroth KJ, Fuchs J, Jupe F, Sandoval JP, Krasileva KV, Borisjuk L, Mockler TC, Ecker JR, Martienssen RA, Lam E (2021) Genome and time-of-day transcriptome of *Wolffia australiana* link morphological minimization with gene loss and less growth control. Genome Research 31 (2):225-238. doi:10.1101/gr.266429.120

Montràs-Janer T, Suggitt AJ, Fox R, Jönsson M, Martay B, Roy DB, Walker KJ, Auffret AG (2024) Anthropogenic climate and land-use change drive short- and long-term biodiversity shifts across taxa. Nature Ecology & Evolution 8 (4):739-751. doi:10.1038/s41559-024-02326-7

Moretti ML, Van Horn CR, Robertson R, Segobye K, Weller SC, Young BG, Johnson WG, Douglas Sammons R, Wang D, Ge X, d' Avignon A, Gaines TA, Westra P, Green AC, Jeffery T, Lespérance MA, Tardif FJ, Sikkema PH, Christopher Hall J, McLean MD, Lawton MB, Schulz B (2018) Glyphosate resistance in *Ambrosia trifida*: Part 2. Rapid response physiology and non-target-site resistance. Pest Management Science 74 (5):1079-1088. doi:10.1002/ps.4569

Muhlemann JK, Younts TLB, Muday GK (2018) Flavonols control pollen tube growth and integrity by regulating ROS homeostasis during high-temperature stress. Proceedings of the National Academy of Sciences USA 115 (47):E11188-e11197. doi:10.1073/pnas.1811492115

Murren CJ, Auld JR, Callahan H, Ghalambor CK, Handelsman CA, Heskel MA, Kingsolver JG, Maclean HJ, Masel J, Maughan H, Pfennig DW, Relyea RA, Seiter S, Snell-Rood E, Steiner UK, Schlichting CD (2015) Constraints on the evolution of phenotypic plasticity: limits and costs of phenotype and plasticity. Heredity 115 (4):293-301. doi:10.1038/hdy.2015.8

Nauheimer L, Metzler D, Renner SS (2012) Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. New Phytologist 195 (4):938-950. doi:10.1111/j.1469-8137.2012.04220.x

Neilson EH, Goodger JQD, Woodrow IE, Møller BL (2013) Plant chemical defense: at what cost? Trends in Plant Science 18 (5):250-258. doi:10.1016/j.tplants.2013.01.001

Nelson DR (2018) Cytochrome P450 diversity in the tree of life. Biochimica et Biophysica Acta. Proteins and Proteomics 1866 (1):141-154. doi:10.1016/j.bbapap.2017.05.003

Oerke EC (2006) Crop losses to pests. The Journal of Agricultural Science 144 (1):31-43. doi:10.1017/S0021859605005708

Oláh V, Combi Z, Szöllösi E, Kanalas P, Mészáros I (2009) Anthocyanins: Possible antioxidants against Cr (VI) induced oxidative stress in *Spirodela polyrrhiza*. Cereal Research Communications 37:533-536. doi:10.1556/CRC.37.2009.Suppl.4

Ozaki K, Ohnishi Y, Iida A, Sekine A, Yamada R, Tsunoda T, Sato H, Sato H, Hori M, Nakamura Y, Tanaka T (2002) Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction. Nature Genetics 32 (4):650-654. doi:10.1038/ng1047

Pan L, Yu Q, Han H, Mao L, Nyporko A, Fan L, Bai L, Powles S (2019) Aldo-keto Reductase Metabolizes Glyphosate and Confers Glyphosate Resistance in *Echinochloa colona*. Plant Physiology 181 (4):1519-1534. doi:10.1104/pp.19.00979

Pan L, Yu Q, Wang J, Han H, Mao L, Nyporko A, Maguza A, Fan L, Bai L, Powles S (2021) An ABCC-type transporter endowing glyphosate resistance in plants. Proceedings of the National Academy of Sciences USA 118 (16). doi:10.1073/pnas.2100136118

Pandian BA, Sathishraj R, Djanaguiraman M, Prasad PVV, Jugulam M (2020) Role of Cytochrome P450 Enzymes in Plant Stress Response. Antioxidants 9 (5):454. doi: 10.3390/antiox9050454

Pasaribu B, Acosta K, Aylward A, Liang Y, Abramson BW, Colt K, Hartwick NT, Shanklin J, Michael TP, Lam E (2023) Genomics of turions from the Greater Duckweed reveal its pathways for dormancy and re-emergence strategy. New Phytologist 239 (1):116-131. doi:10.1111/nph.18941

Perotti VE, Larran AS, Palmieri VE, Martinatto AK, Alvarez CE, Tuesca D, Permingeat HR (2019) A novel triple amino acid substitution in the EPSPS found in a high-level glyphosate-resistant *Amaranthus hybridus* population from Argentina. Pest Management Science 75 (5):1242-1251. doi:10.1002/ps.5303

Peterson M, McMaster S, Riechers D, Skelton J, Stahlmann P (2015) 2,4-D Past, Present, and Future: A Review. Weed Technology 30(2):303-345. doi:10.1614/WT-D-15-00131.1

Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. Genetics 155 (2):945-959. doi:10.1093/genetics/155.2.945

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000b) Association mapping in structured populations. American Journal of Human Genetics 67 (1):170-181. doi:10.1086/302959

Prosecka J, Orlov AV, Fantin YS, Zinchenko VV, Babykin MM, Tichy M (2009) A novel ATP-binding cassette transporter is responsible for resistance to viologen herbicides in the cyanobacterium *Synechocystis sp.* PCC 6803. The FEBS Journal 276 (15):4001-4011. doi:10.1111/j.1742-4658.2009.07109.x

Radchenko M, Ponomareva I, Pozynych I, Morderer Y (2021) Stress and use of herbicides in field crops. Agricultural Science and Practice 8:50-70. doi:10.15407/agrisp8.03.050

Ramos SE, Schiestl FP (2019) Rapid plant evolution driven by the interaction of pollination and herbivory. Science 364 (6436):193-196. doi:10.1126/science.aav6962

Ramos SE, Schiestl FP (2020) Evolution of Floral Fragrance Is Compromised by Herbivory. Frontiers in Ecology and Evolution 8. doi:10.3389/fevo.2020.00030

Ringli C, Bigler L, Kuhn BM, Leiber R-M, Diet A, Santelia D, Frey B, Pollmann S, Klein M (2008) The Modified Flavonol Glycosylation Profile in the Arabidopsis rol1 Mutants Results in Alterations in Plant Growth and Cell Shape Formation. The Plant Cell 20 (6):1470-1481. doi:10.1105/tpc.107.053249

Saeger A (1929) The Flowering of Lemnaceae. Bulletin of the Torrey Botanical Club 56 (7):351-358. doi:10.2307/2480751

Santamaría L (2002) Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. Acta Oecologica 23 (3):137-154. doi:10.1016/S1146-609X(02)01146-3

Schäfer M, Brütting C, Baldwin IT, Kallenbach M (2016) High-throughput quantification of more than 100 primary- and secondary-metabolites, and phytohormones by a single solid-phase extraction based sample preparation with analysis by UHPLC–HESI–MS/MS. Plant Methods 12 (1):30. doi:10.1186/s13007-016-0130-x

Schulz E, Tohge T, Zuther E, Fernie AR, Hincha DK (2015) Natural variation in flavonol and anthocyanin metabolism during cold acclimation in *Arabidopsis thaliana* accessions. Plant, Cell & Environment 38 (8):1658-1672. doi:10.1111/pce.12518

Schützendübel A, Polle A (2002) Plant responses to abiotic stresses: heavy metal-induced oxidative stress and protection by mycorrhization. Journal Experimental Botany 53 (372):1351-1365

Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nature Genetics 44 (7):825-830. doi:10.1038/ng.2314

Shaaltiel Y, Gressel J (1986) Multienzyme oxygen radical detoxifying system correlated with paraquat resistance in *Conyza bonariensis*. Pesticide Biochemistry and Physiology 26 (1):22-28. doi:10.1016/0048-3575(86)90058-1

Shimakawa A, Shiraya T, Ishizuka Y, Wada KC, Mitsui T, Takeno K (2012) Salicylic acid is involved in the regulation of starvation stress-induced flowering in *Lemna paucicostata*. Journal of Plant Physiology 169 (10):987-991. doi:10.1016/j.jplph.2012.02.009

Shitan N, Bazin I, Dan K, Obata K, Kigawa K, Ueda K, Sato F, Forestier C, Yazaki K (2003) Involvement of CjMDR1, a plant multidrug-resistance-type ATP-binding cassette protein, in alkaloid transport in *Coptis japonica*. Proceedings of the National Academy of Sciences USA 100 (2):751-756. doi:10.1073/pnas.0134257100

Siminszky B, Freytag A, Sheldon B, Dewey R (2003) Co-expression of a NADPH:P450 reductase enhances CYP71A10-dependent phenylurea metabolism in tobacco. Pesticide Biochemistry and Physiology 77:35-43. doi:10.1016/j.pestbp.2003.08.001

Smith DC, Bassham JA, Kirk M (1961) Dynamics of the photosynthesis of carbon compounds II. Amino acid synthesis. Biochimica et Biophysica Acta 48 (2):299-313. doi:10.1016/0006-3002(61)90478-4

Smith KE, Schäfer M, Lim M, Robles-Zazueta CA, Cowan L, Fisk ID, Xu S, Murchie EH (2024) Aroma and metabolite profiling in duckweeds: Exploring species and ecotypic variation to enable wider adoption as a food crop. Journal of Agriculture and Food Research 18:101263. doi:10.1016/j.jafr.2024.101263

Sree KS, Bog M, Appenroth K-J (2016) Taxonomy of duckweeds (Lemnaceae), potential new crop plants. Emirates Journal of Food and Agriculture 28:1. doi:10.9755/ejfa.2016-01-038

Sree KS, Sudakaran S, Appenroth K-J (2015) How fast can angiosperms grow? Species and clonal diversity of growth rates in the genus Wolffia (Lemnaceae). Acta Physiologiae Plantarum 37 (10):204. doi:10.1007/s11738-015-1951-3

Stearns SC (1989) Trade-Offs in Life-History Evolution. Functional Ecology 3 (3):259-268. doi:10.2307/2389364

Steele CL, Gijzen M, Qutob D, Dixon RA (1999) Molecular characterization of the enzyme catalyzing the aryl migration reaction of isoflavonoid biosynthesis in soybean. Archives of Biochemistry and Biophysics 367 (1):146-150. doi:10.1006/abbi.1999.1238

Stout JC (2014) Anthropogenic impacts on pollination networks and plant mating systems. Functional Ecology 28 (1):1-2. doi:10.1111/1365-2435.12220

Strzałek M, Kufel L (2021) Light intensity drives different growth strategies in two duckweed species: *Lemna minor* L. and *Spirodela polyrhiza* (L.) Schleiden. PeerJ 9:e12698. doi:10.7717/peerj.12698

Sun Z, Zhao X, Li G, Yang J, Chen Y, Xia M, Hwang I, Hou H (2023) Metabolic flexibility during a trophic transition reveals the phenotypic plasticity of greater duckweed (*Spirodela polyrhiza* 7498). New Phytologist 238 (4):1386-1402. doi:10.1111/nph.18844

Switzer C (1957) The existence of 2, 4-D-resistant strains of wild carrot. Proceedings of the Northeast Weed Control Conference 11:315-318

Szigeti Z (2005) Mechanism of paraquat resistance – from the antioxidant enzymes to the transporters. Acta Biologica Szegediensis 49 (1-2):177-179

Takeno K (2016) Stress-induced flowering: the third category of flowering response. Journal of Experimental Botany 67 (17):4925-4934. doi:10.1093/jxb/erw272

Tang L, Kwon SY, Kim SH, Kim JS, Choi JS, Cho KY, Sung CK, Kwak SS, Lee HS (2006) Enhanced tolerance of transgenic potato plants expressing both superoxide dismutase and ascorbate peroxidase in chloroplasts against oxidative stress and high temperature. Plant Cell Reports 25 (12):1380-1386. doi:10.1007/s00299-006-0199-1

Tani E, Chachalis D, Travlos IS (2015) A Glyphosate Resistance Mechanism in Conyza canadensis Involves Synchronization of EPSPS and ABC-transporter Genes. Plant Molecular Biology Reporter 33 (6):1721-1730. doi:10.1007/s11105-015-0868-8

Tani E, Chachalis D, Travlos IS, Bilalis D (2016) Environmental Conditions Influence Induction of Key ABC-Transporter Genes Affecting Glyphosate Resistance Mechanism in *Conyza canadensis*. International Journal of Molecular Sciences 17 (4):342. doi:10.3390/ijms17040342

Tattini M, Galardi C, Pinelli P, Massai R, Remorini D, Agati G (2004) Differential accumulation of flavonoids and hydroxycinnamates in leaves of Ligustrum vulgare under excess light and drought stress. New Phytologist 163 (3):547-561. doi:10.1111/j.1469-8137.2004.01126.x

Taub DR, Miller B, Allen H (2008) Effects of elevated CO2 on the protein concentration of food crops: a meta-analysis. Global Change Biology 14 (3):565-575. doi:10.1111/j.1365-2486.2007.01511.x

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) Dwarf8 polymorphisms associate with variation in flowering time. Nature Genetics 28 (3):286-289. doi:10.1038/90135

Tibbs Cortes L, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. Plant Genome 14 (1):e20077. doi:10.1002/tpg2.20077

Tilsner J, Kassner N, Struck C, Lohaus G (2005) Amino acid contents and transport in oilseed rape (*Brassica napus* L.) under different nitrogen conditions. Planta 221 (3):328-338. doi:10.1007/s00425-004-1446-8

Tippery NP, Les DH (2020) Tiny Plants with Enormous Potential: Phylogeny and Evolution of Duckweeds. In: Cao XH, Fourounjian P, Wang W (eds) The Duckweed Genomes. Springer International Publishing, Cham, pp 19-38. doi:10.1007/978-3-030-11045-1_2

Tippery NP, Les DH, Crawford DJ (2015) Evaluation of phylogenetic relationships in Lemnaceae using nuclear ribosomal data. Plant Biology 17 (s1):50-58. doi:10.1111/plb.12203

Tyler JK, Haller WT, Les G (2006) Documentation of Landoltia (*Landoltia punctata*) Resistance to Diquat. Weed Science 54 (4):615-619. doi:10.1614/WS-06-002R.1

Upadhyay RK, Edelman M, Mattoo AK (2020) Identification, Phylogeny, and Comparative Expression of the Lipoxygenase Gene Family of the Aquatic Duckweed, *Spirodela polyrhiza*, during Growth and in Response to Methyl Jasmonate and Salt. International Journal of Molecular Sciences 21 (24):9527. doi:10.3390/ijms21249527

Van Antro M, Prelovsek S, Ivanovic S, Gawehns F, Wagemaker N, Mysara M, Horemans N, Vergeer P, Verhoeven KJF (2023) DNA methylation in clonal duckweed (*Lemna minor* L.) lineages reflects current and historical environmental exposures. Molecular Ecology 32 (2):428-443. doi:10.1111/mec.16757

Van Horn CR, Moretti ML, Robertson RR, Segobye K, Weller SC, Young BG, Johnson WG, Schulz B, Green AC, Jeffery T, Lespérance MA, Tardif FJ, Sikkema PH, Hall JC, McLean MD, Lawton MB, Sammons RD, Wang D, Westra P, Gaines TA (2018) Glyphosate resistance in Ambrosia trifida: Part 1. Novel rapid cell death response to glyphosate. Pest Management Science 74 (5):1071-1078. doi:10.1002/ps.4567

Vats S (2015) Herbicides: History, Classification and Genetic Manipulation of Plants for Herbicide Resistance. In: Lichtfouse E (eds) Sustainable Agriculture Reviews: Volume 15. Springer International Publishing, Cham, pp 153-192. doi:10.1007/978-3-319-09132-7_3

Visnovitz T, Soós V, Jóri B, Racz I, Szigeti Z (2008) Staying alive: Insight into the resistance mechanism of Conyza canadensis to xenobiotic paraquat. Acta Herbologica 17(2):173-178

Vogt F, Shirsekar G, Weigel D (2021) vcf2gwas-python API for comprehensive GWAS analysis using GEMMA. Bioinformatics 38 (3):839-840. doi:10.1093/bioinformatics/btab710

Wada KC, Mizuuchi K, Koshio A, Kaneko K, Mitsui T, Takeno K (2014) Stress enhances the gene expression and enzyme activity of phenylalanine ammonia-lyase and the endogenous content of salicylic acid to induce flowering in pharbitis. Journal of Plant Physiology 171 (11):895-902. doi:10.1016/j.jplph.2014.03.008

Wang H, Li Y, Chern M, Zhu Y, Zhang L-L, Lu J-H, Li X-P, Dang W-Q, Ma X-C, Yang Z-R, Yao S-Z, Zhao Z-X, Fan J, Huang Y-Y, Zhang J-W, Pu M, Wang J, He M, Li W-T, Chen X-W, Wu X-J, Li S-G, Li P, Li Y, Ronald PC, Wang W-M (2021) Suppression of rice miR168 improves yield, flowering time and immunity. Nature Plants 7 (2):129-136. doi:10.1038/s41477-021-00852-x

Wang J, Zhou L, Shi H, Chern M, Yu H, Yi H, He M, Yin J, Zhu X, Li Y, Li W, Liu J, Wang J, Chen X, Qing H, Wang Y, Liu G, Wang W, Li P, Wu X, Zhu L, Zhou JM, Ronald PC, Li S, Li J, Chen X (2018) A single transcription factor promotes both yield and immunity in rice. Science 361 (6406):1026-1028. doi:10.1126/science.aat7675

Wang L, Zhao L, Zhang X, Zhang Q, Jia Y, Wang G, Li S, Tian D, Li WH, Yang S (2019a) Large-scale identification and functional analysis of NLR genes in blast

resistance in the Tetep rice genome sequence. Proceeding of the National Academy of Sciences USA 116 (37):18479-18487. doi:10.1073/pnas.1910229116

Wang P, Gong R, Yang Y, Yu S (2019b) Ghd8 controls rice photoperiod sensitivity by forming a complex that interacts with Ghd7. BMC Plant Biology 19 (1):462. doi:10.1186/s12870-019-2053-y

Wang W, Haberer G, Gundlach H, Gläßer C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, Byrant DW, Mockler TC, Appenroth KJ, Grimwood J, Jenkins J, Chow J, Choi C, Adam C, Cao XH, Fuchs J, Schubert I, Rokhsar D, Schmutz J, Michael TP, Mayer KF, Messing J (2014) The *Spirodela polyrhiza* genome reveals insights into its neotenous reduction fast growth and aquatic lifestyle. Nature Communications 5:3311. doi:10.1038/ncomms4311

Wang Y, Duchen P, Chávez A, Sree KS, Appenroth KJ, Zhao H, Höfer M, Huber M, Xu S (2024) Population genomics and epigenomics of Spirodela polyrhiza provide insights into the evolution of facultative asexuality. Communications Biology 7 (1):581. doi:10.1038/s42003-024-06266-7

Wang Y, Liu M, Ge D, Akhter Bhat J, Li Y, Kong J, Liu K, Zhao T (2020) Hydroperoxide lyase modulates defense response and confers lesion-mimic leaf phenotype in soybean (Glycine max (L.) Merr.). The Plant Journal 104 (5):1315-1333. doi:10.1111/tpj.15002

Whitehead CW, Switzer CM (1963) THE DIFFERENTIAL RESPONSE OF STRAINS OF WILD CARROT TO 2,4-D AND RELATED HERBICIDES. Canadian Journal of Plant Science 43 (3):255-262. doi:10.4141/cjps63-052

Whitehead M, Lanfear R, Mitchell R, Karron J (2018) Plant Mating Systems Often Vary Widely Among Populations. Frontiers in Ecology and Evolution 6:38. doi:10.3389/fevo.2018.00038

Xiong L, Schumaker KS, Zhu JK (2002) Cell signaling during cold, drought, and salt stress. Plant Cell 14 (Suppl_1):S165-183. doi:10.1105/tpc.000596

Xu G, Yuan M, Ai C, Liu L, Zhuang E, Karapetyan S, Wang S, Dong X (2017) uORF-mediated translation allows engineered plant disease resistance without fitness costs. Nature 545 (7655):491-494. doi:10.1038/nature22372

Xu S, Stapley J, Gablenz S, Boyer J, Appenroth KJ, Sree KS, Gershenzon J, Widmer A, Huber M (2019) Low genetic variation is associated with low mutation rate in the giant duckweed. Nature Communications 10 (1):1243. doi:10.1038/s41467-019-09235-5

Yang T, Li H, Tai Y, Dong C, Cheng X, Xia E, Chen Z, Li F, Wan X, Zhang Z (2020) Transcriptional regulation of amino acid metabolism in response to nitrogen deficiency and nitrogen forms in tea plant root (*Camellia sinensis* L.). Scientific Reports 10 (1):6868. doi:10.1038/s41598-020-63835-6

Yanniccari M, Gómez-Lobato ME, Istilart C, Natalucci C, Giménez DO, Castro AM (2017) Mechanism of Resistance to Glyphosate in *Lolium perenne* from Argentina. Frontiers in Ecology and Evolution 5 (123). doi:10.3389/fevo.2017.00123

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics 38 (2):203-208. doi:10.1038/ng1702

Yu Q, Huang S, Powles S (2010) Direct measurement of paraquat in leaf protoplasts indicates vacuolar paraquat sequestration as a resistance mechanism in Lolium rigidum. Pesticide Biochemistry and Physiology 98 (1):104-109. doi:10.1016/j.pestbp.2010.05.007

Yu Q, Jalaludin A, Han H, Chen M, Sammons RD, Powles SB (2015) Evolution of a double amino acid substitution in the 5-enolpyruvylshikimate-3-phosphate synthase in Eleusine indica conferring high-level glyphosate resistance. Plant Physiology 167 (4):1440-1447. doi:10.1104/pp.15.00146

Yun HS, Sul WJ, Chung HS, Lee J-H, Kwon C (2023) Secretory membrane traffic in plant–microbe interactions. New Phytologist 237 (1):53-59. doi:10.1111/nph.18470

Zhang M, Xiao Q, Li Y, Tian Y, Zheng J, Zhang J (2024) Exploration of exogenous chlorogenic acid as a potential plant stimulant: enhancing physiochemical properties in Lonicera japonica. Physiology and Molecular Biology of Plants 30 (3):453-466. doi:10.1007/s12298-024-01435-8

Zhang P, Zhong K, Zhong Z, Tong H (2019) Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.). BMC Plant Biology 19 (1):259. doi:10.1186/s12870-019-1842-7

Zhao Y, Fang Y, Jin Y, Huang J, Bao S, Fu T, He Z, Wang F, Wang M, Zhao H (2015) Pilot-scale comparison of four duckweed strains from different genera for potential application in nutrient recovery from wastewater and valuable biomass production. Plant Biology 17 (Suppl 1):82-90. doi:10.1111/plb.12204

Zhao Z, Shi H, Kang X, Liu C, Chen L, Liang X, Jin L (2017) Inter- and intra-specific competition of duckweed under multiple heavy metal contaminated water. Aquatic Toxicology 192:216-223. doi:10.1016/j.aquatox.2017.09.023

Zia-Ul-Haq M, Khaliq A, Qiang S, Matloob A, Hussain S, Fatima S, Aslam Z (2019) Weed growth, herbicide efficacy, and rice productivity in dry seeded paddy field under different wheat stubble management methods. Journal of Integrative Agriculture 18 (4):907-926. doi:10.1016/S2095-3119(18)62004-0

Ziegler P, Adelmann K, Zimmer S, Schmidt C, Appenroth KJ (2015) Relative in vitro growth rates of duckweeds (Lemnaceae) – the most rapidly growing higher plants. Plant Biology 17 (s1):33-41. doi:10.1111/plb.12184

Ziegler P, Appenroth KJ, Sree KS (2023) Survival Strategies of Duckweeds, the World's Smallest Angiosperms. Plants 12 (11):2215. doi:10.3390/plants12112215

Ziska L, Teasdale J (2000) Sustained growth and increased tolerance to glyphosate observed in a C3 perennial weed, quackgrass (Elytrigia repens), grown at elevated carbon dioxide. Functional Plant Biology 27:159-166. doi:10.1071/PP99099

Živanović B, Milić Komić S, Tosti T, Vidović M, Prokić L, Veljović Jovanović S (2020) Leaf Soluble Sugars and Free Amino Acids as Important Components of Abscisic Acid-Mediated Drought Response in Tomato. Plants 9 (9):1147. doi:10.3390/plants9091147

Züst T, Agrawal AA (2017) Trade-Offs Between Plant Growth and Defense Against Insect Herbivory: An Emerging Mechanistic Synthesis. Annual Review of Plant Biology 68:513-534. doi:10.1146/annurev-arplant-042916-040856

6. Acknowledgements

This thesis occupied a major portion of the past five years of my live. In this process, I greatly enjoyed the highly supportive environment provided by my colleagues and friends. Therefore, I would like to take this opportunity to thank the AG Xu for providing a supportive environment, that enabled me to conduct this research. In this regard, I am especially thankful to [redacted], [redacted] and [redacted]. As supervisor of my PhD project [redacted] provided a very inspiring research environment and a great scientific advice, through which I could improve my communication and writing skills. Furthermore, he provided the basic research concept and organized the funding's for this project. Many thanks also to [redacted], from whom I gained lots of interesting input about GWAS and population genetics through extensive scientific discussions. Through his great reliability and openness [redacted], greatly contributed to most of the research associated with this thesis. He has been of great help during the screening of herbicide resistant genotypes and provided support in the identification of GWAS candidates. Also, outside of the office, he is a great friend to me, joining most of my board game evenings. A large portion of this thesis is associated with chemical metabolite analysis conducted via LC-MS. These metabolite studies would not have been possible without the help of [redacted]. Through establishing extraction and analysis protocols for antioxidants, free metabolite and diquat measurements, he enabled the analysis of metabolite contents in duckweed tissue. I am also thankful to the technicians [redacted], [redacted] and [redacted], who helped maintaining the duckweed collection. Special thanks also go to my student helpers, [redacted] and [redacted], who provided great help during the cultivation and screening of duckweed genotypes. [redacted] also extracted free metabolites from duckweed genotypes together with [redacted]. I am thankful to [redacted] for introducing me into the field of plant genetic manipulation and callus-cultivation. Finally, I would like to thank [redacted] and [redacted] for co-supervising this thesis and providing helpful comments. In the AG Finkemeier, I thank [redacted] for helping we with the chlorophyll measurements in duckweed.

7. Versicherung

Anlage zum Antrag auf Zulassung zur Promotionsprüfung gem. § 12 der Promotionsordnung des Fachbereichs Biologie der Johannes Gutenberg-Universität Mainz vom 01.04.2018

VERSICHERUNG

Name: Martin Höfer

Hiermit versichere ich gemäß § 12, (2) der Promotionsordnung vom 01.04.2018:
(zutreffendes ist angekreuzt.)

- ☒ Ich habe die heute als Dissertation vorgelegte Arbeit selbst angefertigt und ausschließlich die angegebenen Quellen und Hilfsmitte verwendet.
- ☒ Ich habe oder hatte die jetzt als Dissertation vorgelegte Arbeit noch an keiner anderen deutschen oder ausländischen Hochschule oder vergleichbaren Einrichtung zur Erlangung eines akademischen Grades eingereicht.
- ☒ Ich habe noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolglos beendet.
- ☒ Ich habe noch kein Promotions- PhD,- oder ein vergleichbares Graduierungsverfahren im Promotionsfach erfolgreich beendet.
- ☒ Für die Anfertigung der vorgelegten Arbeit wurde keine entgeltliche Hilfe Dritter, insbesondere eine Promotionsberatung oder –vermittlung in Anspruch genommen.

Mainz, den 29.10.2024 M. Höfer

(Unterschrift)

8. Curriculum vitae

Martin Höfer

personal data

address:	Rüsselsheimer Allee 86, 55130 Mainz
telephone:	0174 2511827
place and date of birth:	Saalfeld, 10.03.1995
E-Mail:	mahoefer@uni-mainz.de

education

university

Feb. 2020 – Jan. 2025 (expected)	**PhD studies**, Westfälische Wilhelms-University, Münster and Johannes-Gutenberg University, Mainz, expected finish in Jan. 2025
Oct. 2019 – Jan. 2020	**Vocational orientation** aiming to find a PhD position
ct. 2017 – Oct. 2019	**Master studies in chemical biology**, Friedrich Schiller University, Jena (average mark: 1.5)
Oct. 2014 – Sept. 2017	**Bachelor studies in biochemistry/molecular biology**, Friedrich Schiller University, Jena (average mark: 1.7)

school

2005 – 2014	**graduation: Abitur (A - level)** comprehensive school: IGS Grete Unrein, Jena

professional experiences

Feb. 2020 – Jan. 2025 (expected)	**PhD thesis**, group: Plant Adaption-in-action at the Institute of Evolution and Biodiversity, Münster (till March 2022) and Institute of Organismic and Molecular Evolution, Mainz (since April 2022) Title: Genetic Basis of Metabolic Trait Variation in the Giant Duckweed *Spirodela polyrhiza* (Lemnaceae) Supervision: [redacted], [redacted]
Jan. 2019 – Oct. 2019	**Master thesis**, group: Chemical ecology of cross kingdom interactions at the institute of inorganic and analytical chemistry, Jena

	title: Development of molecular tools to study the genetic basis of the morphogenesis of the marine macroalgae *Ulva mutabilis* supervision: [redacted], [redacted]
Oct. 2019 – Jan. 2020	**Vocational orientation** aiming to find a PhD position
Oct. 2018 – Dec. 2018	**Internship on NGS techniques** in microbiome research in the Department of Host Septomics at the ZIK Septomics, Jena
Sept. 2017 – Feb. 2018	**Research assistant** in the Department of Biochemistry at the Max-Planck Institute of chemical ecology, Jena
Mar. 2017 – Aug. 2017	**Bachelor thesis**, Department of Biochemistry at the Max-Planck Institute for chemical ecology, Jena title: Characteristic of the effects of Copper excess on metabolic and oxidative parameters in *Spirodela polyrhiza* supervision: [redacted], [redacted]
Oct. 2016 – Feb. 2017	**Tutor** for an undergraduate student biochemistry lecture

other experiences

computer skills in Word (very good), Latex (good), Excel (good) and SPSS (basic)

computational skills in R (very good) and Python (basic)

practical experiences in: molecular cloning, cultivation of macroalgae, microbial ecology, LC – MS, quantitative genetics (GWAS), plant tissue cultivation, plant transformation and qPCR

languages: German (native), English (full proficiency), Russian (basic) and Arabic (basic)

conferences

6th International Conference on Duckweed Research and Applications 2022 (poster presentation)

39th Annual Meeting of the International Society of Chemical Ecology 2024 ([redacted] talk)

publications

Höfer, M., Schäfer, M., Wang, Y., Wink, S. & Xu, S. Genetic Mechanism of Non-Targeted-Site Resistance to Diquat in *Spirodela polyrhiza*. *Plants*. **2024**, 13, 845, https://doi.org/10.3390/plants13060845

Höfer, M., Schäfer, M., Wink, S., Wang, Y. & Xu, S. Genome-Wide Association Study of Metabolic Traits in the Duckweed *Spirodela polyrhiza*. *bioRxiv*. **2024**, https://doi.org/10.1101/2024.07.26.605148

Wang, Y., Duchen, P., Chávez, A., Sree, K.S., Appenroth, K.J., Zhao, H., Höfer, M., Huber, M. & Xu, S. Population genomics and epigenomics of *Spirodela polyrhiza* provide insights into the evolution of facultative asexuality. *Commun. Biol.* **2024**, 7, 581, https://doi.org/10.1038/s42003-024-06266-7

Huber, M., Gablenz, S. & Höfer, M. (2021). Transgenerational non-genetic inheritance has fitness costs and benefits under recurring stress in the clonal duckweed *Spirodela polyrhiza*. *Proc. Royal Soc. B*. **2021**, 288, 20211269, https://doi.org/10.1098/rspb.2021.1269

Mainz, 27.10.2024