Plant Biotechnology Journal

Plant Biotechnology Journal (2024), pp. 1–19 doi: 10.1111/pbi.14340

Site-directed genotype screening for elimination of antinutritional saponins in quinoa seeds identifies TSARL1 as a master controller of saponin biosynthesis selectively in seeds

Mai Duy Luu Trinh[1,*,†], Davide Visintainer[1,†], Jan Günther[1,†], Jeppe Thulin Østerberg[2,†], Rute R. da Fonseca[3], Sara Fondevilla[4], Max William Moog[1], Guangbin Luo[1], Anton F. Nørrevang[1], Christoph Crocoll[1], Philip V. Nielsen[1], Sven-Erik Jacobsen[5], Toni Wendt[2], Søren Bak[1], Rosa Laura López-Marqués[1] and Michael Palmgren[1,*]

[1]*Department of Plant and Environmental Sciences, University of Copenhagen, Frederiksberg, Denmark*
[2]*Traitomic A/S, Copenhagen, Denmark*
[3]*Section for Biodiversity, Globe Institute, University of Copenhagen, København Ø, Denmark*
[4]*Institute for Sustainable Agriculture-CSIC, Córdoba, Spain*
[5]*Quinoa Quality ApS, Regstrup, Denmark*

Received 18 December 2023;
revised 1 March 2024;
accepted 4 March 2024.
**Correspondence* (Tel +45 3533 2362; Fax +45 3533 3365; email mdlt@plen.ku.dk; Tel +45 2398 8444; Fax +45 3533 3365; email palmgren@plen.ku.dk)
[†]These authors made an equal contribution to the work.
Keywords: *Chenopodium quinoa*, genome-wide association study, mutagenesis, quinoa, saponin biosynthesis, site-directed genotype screening.

Summary

Climate change may result in a drier climate and increased salinization, threatening agricultural productivity worldwide. Quinoa (*Chenopodium quinoa*) produces highly nutritious seeds and tolerates abiotic stresses such as drought and high salinity, making it a promising future food source. However, the presence of antinutritional saponins in their seeds is an undesirable trait. We mapped genes controlling seed saponin content to a genomic region that includes *TSARL1*. We isolated desired genetic variation in this gene by producing a large mutant library of a commercial quinoa cultivar and screening the library for specific nucleotide substitutions using droplet digital PCR. We were able to rapidly isolate two independent *tsarl1* mutants, which retained saponins in the leaves and roots for defence, but saponins were undetectable in the seed coat. We further could show that TSARL1 specifically controls seed saponin biosynthesis in the committed step after 2,3-oxidosqualene. Our work provides new important knowledge on the function of TSARL1 and represents a breakthrough for quinoa breeding.

Introduction

Global agriculture faces a major challenge: to attain food security amidst continued human population growth, limited arable land and water and the predicted effects of climate change, such as increased salinization and aridity (FAO, 2021; Mukhopadhyay *et al*., 2021; Negacz *et al*., 2022; Ruiz *et al*., 2014). Major crops, including maize (*Zea mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*) and soybean (*Glycine max*), are not resilient to abiotic stresses such as extreme temperature, frost, drought, soil salinization and flooding, which are increasing in frequency as a result of rapid climate change (Jägermeyr *et al*., 2021). Therefore, breeding novel crops that can tolerate harsh environmental conditions is crucial (López-Marqués *et al*., 2020; Luo *et al*., 2022).

Quinoa (*Chenopodium quinoa* Willd.) is a tetraploid crop indigenous to the Andes mountains, where it has been domesticated since 5800–4400 BC (Pearsall, 2008). Quinoa is adapted to the harsh climatic conditions in this region and tolerates adverse abiotic factors (Hinojosa *et al*., 2018) such as frost (Jacobsen *et al*., 2005, 2007), drought (Imamura *et al*., 2020; Lin and Chao, 2021) and high salinity (Adolf *et al*., 2013; Hariadi *et al*., 2011; Moog *et al*., 2022). In addition, quinoa seed has a remarkably high nutritional value due to its exceptional protein profile (accounting for ~15% of the dry weight and containing a balanced set of essential amino acids), high-quality oil content (6%–9% of the dry weight), containing oleic acid, linolenic acid (belonging to the omega-3 fatty acid family), linoleic acid (belonging to the omega-6 fatty acid family) and high mineral content (e.g. zinc, potassium, copper, magnesium and iron) (Lim, 2013; Suárez-Estrella *et al*., 2018). Therefore, quinoa is an excellent candidate crop for increasing food security and supporting a transition to plant-based diets. However, there is still substantial room for further genetic improvement in quinoa, especially to adapt the crop to different environments. Relevant breeding targets in quinoa are increased seed size (a character appreciated by the consumer), downy mildew resistance, short plant height (to facilitate mechanical harvesting), heat tolerance,

Please cite this article as: Trinh, M D L., Visintainer, D., Günther, J., Østerberg, J.T., Fondevilla, R R da Fonseca, S., Moog, M.W., Luo, G., Nørrevang, A.F., Crocoll, C., Nielsen, P.V., Jacobsen, S.-E., Wendt, T., Bak, S., López-Marqués, R.L. and Palmgren, M. (2024) Site-directed genotype screening for elimination of antinutritional saponins in quinoa seeds identifies TSARL1 as a master controller of saponin biosynthesis selectively in seeds. *Plant Biotechnol. J.*, https://doi.org/10.1111/pbi.14340.

earliness, reduced seed shattering and pre-harvested sprouting and reduced seed saponin content (López-Marqués *et al*., 2020; Zurita-Silva *et al*., 2014).

Although quinoa has been successfully transformed (Komari, 1990; Xiao *et al*., 2022), regenerating the transformed cells and tissues is a bottleneck in applying CRISPR/Cas technology to quinoa. Recently, FIND-IT (fast identification of nucleotide variants by droplet digital PCR) technology was developed as a sensitive non-GM approach for screening variant populations (Knudsen *et al*., 2022). As the quinoa genome is now fully sequenced, assembled and annotated (Jarvis *et al*., 2017; Li and Lightfoot, 2021; Yasui *et al*., 2016; Zou *et al*., 2017), we aimed to use this technique to precisely isolate targeted mutations potentially resulting in desired traits in quinoa.

Saponins account for up to 4% of the dry mass of quinoa seeds, accumulating mainly in the outer layers of the seed (Jarvis *et al*., 2017) and leading to a bitter taste (Suárez-Estrella *et al*., 2018). Furthermore, saponins are antinutrients that interfere with vitamin and mineral absorption, resulting in cytotoxicity and haemolysis when overconsumed (Kuljanabhagavad *et al*., 2008; Kuljanabhagavad and Wink, 2009; Satheesh and Fanta, 2018; Suárez-Estrella *et al*., 2018; Woldemichael and Wink, 2001) and must, therefore, be removed from quinoa products before consumption creating an environmentally problematic waste product. Saponins are glycosylated triterpenoid- or steroid-derived specialized metabolites widely distributed in the plant kingdom (Vincken *et al*., 2007; Wink, 2004). Quinoa mainly produces triterpenoid saponins (saponins), including various derivatives of the four main aglycones: oleanolic acid, hederagenin, phytolaccagenic acid and serjanic acid (Kuljanabhagavad *et al*., 2008; Madl *et al*., 2006; Mastebroek *et al*., 2000; Ruiz *et al*., 2017; Woldemichael and Wink, 2001). At least 40 triterpenoid saponins have been isolated and structurally characterized in the past 30 years (El Hazzam *et al*., 2020). Quinoa ubiquitously accumulates saponins at different concentrations in all plant tissues and organs examined, including roots, stems, leaves, flowers, fruits and seeds (Fiallos-Jurado *et al*., 2016; Kuljanabhagavad *et al*., 2008; Lim *et al*., 2020; Mastebroek *et al*., 2000).

In this study, we carried out a genome-wide association study (GWAS) involving 100 cultivars of quinoa and identified a genomic region associated with seed saponin content including *TSARL1*. Subsequently, we generated a large mutant library from a commercial cultivar of quinoa and investigated whether site-directed genotype screening could identify mutants of potential agronomic interest in a population of this cultivar. By screening the library developed, we identified mutants with specific nucleotide substitutions in *TSARL1*. Strikingly, it was not possible to detect saponins in the seeds of the mutant plants, whereas the saponin content did not differ in other parts of the plant compared with the wild type (WT). This provides a solution for the environmentally problematic disposal of waste resulting from dehulling quinoa seeds (Ruiz *et al*., 2017). Further, we found that TSARL1 controls the expression of the genes encoding β-amyrin synthase (BAS) and β-amyrin 28-oxidase, the activity of which represents the committed steps in the saponin biosynthesis pathway. These findings not only provide insights into the functions of TSARL1 in quinoa saponin biosynthesis, but also provide proof of concept that the mutant library developed is well adapted for accelerating quinoa improvement.

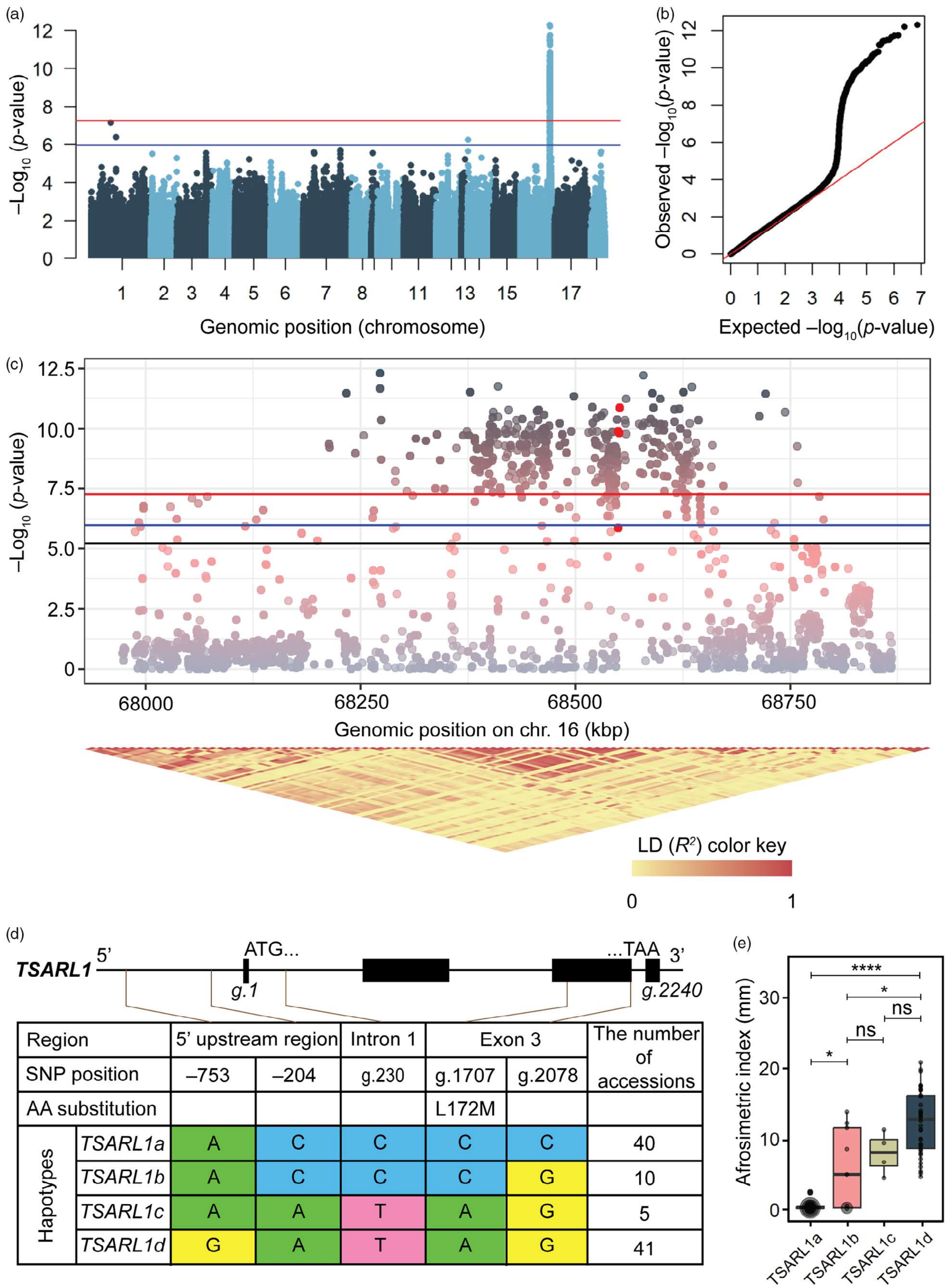
Results

TSARL1 SNPs are associated with saponin content in quinoa

To identify target genes controlling seed saponin content in quinoa, we first performed next-generation genome sequencing of 124 quinoa accessions and cultivars and obtained sequences from an additional 34 accessions in public databases. The sequence reads were mapped to the reference genome (ASM168347v1) (Jarvis *et al*., 2017), and after removing duplicates, 20 million unfiltered single-nucleotide polymorphisms (SNPs) were identified. These SNPs were then filtered for quality, resulting in 3 637 746 high-confidence SNPs.

Saponin content was scored in 88 out of 124 accessions using the afrosimetric method (Figure S1a,b). Linear mixed-model analysis using the likelihood ratio test in GEMMA identified significant SNPs (above Bonferroni threshold ($-\log_{10}$(*P*-value) = 7.27), red line in Figure 1a,c) associated with saponin content localized on chromosome 16, with the most significant SNP observed at position 68 622 294 bp (Figure 1; Figure S1c-e). Within this region, there were 64 genes, 41 of which contained 140 SNPs above the suggestive threshold ($-\log_{10}$(*P*-value) = 5.97) (Table S1). Among these genes, we identified AUR62017204 (*TSARL1*) and AUR62017206 (*TSARL2*), encoding bHLH transcription factors that have previously been reported to be associated with saponin content (Jarvis *et al*., 2017; Maldonado-Taipe *et al*., 2022; Patiranage *et al*., 2022). An uncharacterized gene, AUR62017205, may also contribute to controlling this trait in quinoa as it contained seven missense SNPs associated with the trait (Table S1). Among the identified genes, 23 contained missense mutations and 4 contained splice site mutations predicted to alter the expressed protein sequence (Table S1). Among these genes, *TSARL1*, AUR62017213 (encoding a putative lysine-specific demethylase 5D-like protein), AUR62017198 (encoding a putative kinesin-like protein) and

Figure 1 Genome-wide association study (GWAS) to detect genomic regions related to saponin content in 100 quinoa accessions. (a) A $-\log_{10}$(*P*-value) genome-wide association plot (Manhattan plot) showing genomic regions associated with saponin content in quinoa. Significant threshold (red line) and suggested threshold (blue line) correspond to adjusted Bonferroni ($-\log_{10}$(*P*-value) = 7.27) and suggestive ($-\log_{10}$(*P*-value) = 5.97) thresholds, respectively. The most significant SNP associated with the saponin content trait in quinoa is located at position Chr16:68622294. (b) Quantile–quantile (QQ) plot of the data shown in the Manhattan plot. Most of the observed *P*-values for each SNP correspond to the expected values from a theoretical χ^2-distribution excepting significant *P*-values of SNPs associated with the saponin content trait in quinoa. (c) Manhattan plot of the 900-kbp region including the most significant SNPs on chromosome 16. Red dots mark SNPs within the *TSARL1* coding region. Linkage plot for the region is included below the graph. Red and blue lines represent the adjusted Bonferroni and suggestive significance thresholds, respectively. Black line represents the significance threshold for FDR-adjusted *P*-values ($-\log_{10}$(*P*-value) = 5.22). (d) Haplotypes of *TSARL1* associated with saponin content in quinoa. (e) Differences in saponin content among four haplotypes of *TSARL1*. *$P < 0.05$, ****$P < 0.0001$ (Mann–Whitney non-parametric test).


(a)
−Log10 (p-value)
Genomic position (chromosome)
(b)
Observed −log10(p-value)
Expected −log10(p-value)
(c)
−Log10 (p-value)
Genomic position on chr. 16 (kbp)
LD (R2) color key
0
1
(d)
TSARL1
5'
ATG...
g.1
...TAA
3'
g.2240
Region | 5' upstream region | Intron 1 | Exon 3 | The number of accessions
SNP position | −753 | −204 | g.230 | g.1707 | g.2078
AA substitution | L172M
Hapotypes
TSARL1a | A | C | C | C | C | 40
TSARL1b | A | C | C | C | G | 10
TSARL1c | A | A | T | A | G | 5
TSARL1d | G | A | T | A | G | 41
(e)
Afrosimetric index (mm)
****
*
ns
ns
*
TSARL1a
TSARL1b
TSARL1c
TSARL1d

AUR62017191 (encoding a putative signal peptide endopeptidase-like protein) contained both missense and splice site mutations (Table S1).

In *TSARL2*, a homoeolog of *TSARL1* (Figure S2 and S3) (Jarvis *et al*., 2017; Suzuki *et al*., 2021), only one missense SNP was identified, which was predicted to result in a conservative substitution of Leu292 to Val (Figure S3, red arrow, red square) within the C-terminal (ACT-like) domain. This substitution is not specific to either bitter or sweet quinoa accessions as it was found in six sweet (CHEN-125, −161, −183, −211, −262 and −345) and five bitter (CHEN-115, −148, −218, −360 and −593) accessions. Other identified SNPs in *TSARL2* were in noncoding regions (introns, upstream and downstream regions), and splice site mutation events were not detected (Table S1).

Within the *TSARL1* gene region, we identified five significant SNPs that segregated in our population (Figure 1d; Figure S4a), of which two, C1707A and G2078C, localized to exon 3. C1707A is a missense SNP, which is predicted to result in a conservative substitution of Leu172 to Met (Figure S3, yellow arrow, yellow square) within the bHLH domain. The SNP G2078C was reported to cause alternative splicing, which results in truncated *TSARL1* transcripts and a premature STOP codon in the coding sequence (Jarvis *et al*., 2017). We also identified four major haplotypes of *TSARL1* (*TSARL1a*, *b*, *c* and *d*; Figure 1d), which were further confirmed by Sanger sequencing of *TSARL1* in six sweet and five bitter quinoa accessions (Figure S4b). Quinoa accessions that inherited *TSARL1a* produced sweet seeds, and those that inherited *TSARL1d* produced bitter seeds with a high saponin content (Figure 1e).

Using site-directed genotype screening to isolate specific mutations in quinoa

To generate a SNP library, we used seeds from the bitter commercial quinoa cultivar Titicaca, which has been bred to withstand North European conditions (Jacobsen, 2017; Jacobsen *et al*., 1996; Taaime *et al*., 2023). In addition to being robust and high yielding (De Bock *et al*., 2021; Thiam *et al*., 2021), a major reason for this cultivar's success is that it is insensitive to photoperiod (Patiranage *et al*., 2021). In our greenhouse experiments, we found that quinoa cv. Titicaca plants exhibited an earlier flowering time and seed maturation than low-saponin (sweet) accessions and cultivars such as CHEN-125, CHEN-161, CHEN-183, CHEN-211, CHEN-262, CHEN-345, Vikinga, Atlas and Riobamba when grown under long-day conditions (Figure S5). Therefore, seeds from quinoa cv. Titicaca were considered an ideal primary material (M_0) to generate our SNP library.

Ethyl methanesulfonate (EMS) caused random point mutations in the quinoa genome with mutation frequencies ranging from 0.5 to 16.6 mutations per Mb when concentrations increased from 0.1 to 0.5% (v/v), respectively (Figure S6a,b). To avoid a high level of background mutations, 0.1 and 0.2% (v/v) EMS treatments were chosen for library generation (Figure S6c).

Mutated seeds (M_1) were planted and grown to maturity under field conditions in Denmark during the 2019 and 2020 growth seasons. M_2 seeds from approximately 100 000 plants were harvested and pooled (100 plants per pool; Figure S7a). The procedure to extract mutants and validate identified mutants is detailed in Materials and Methods (see Figure S7b for a schematic explanation). In initial screens to determine whether site-directed genotype screening can be employed using this library, we identified 10 loss-of-function mutants for 10 genes of interest (Figure S6d; Table S2). Thus, the library had broad coverage and was useful for identifying specific genetic variation.

Isolation of *tsarl1* mutants

Employing the established EMS mutant library, we used site-directed genotype screening (Knudsen *et al*., 2022) to identify two heterozygous *tsarl1* mutants (M_2): one mimicking the previously described *tsarl1-1* allele and the other in which the fourth codon had been converted to a STOP codon (Figure 2a; Figure S8b). The mutants were self-pollinated, and homozygous mutants were identified by sequencing *TSARL1* genomic regions isolated from the M_3 progeny (Figure S8a). The two new homozygous *tsarl1* mutant alleles contained a G2079A substitution (*tsar1-3*) and a G12A substitution (*tsarl1-4*) (Figure 2a). The *tsarl1-4* substitution replaced a UGG codon encoding tryptophan (W4) with a premature STOP codon (TGA), causing a predicted complete loss of function of TSARL1 (Figure 2a). The G2079A substitution in *tsarl1-3* resulted in a truncated *TSARL1* transcript (Figure 2b,c), presumably as a result of alternative splicing. The cryptic splice site (G1875|G1876) was identified by sequencing the *TSARL1* transcript from *tsarl1-3* and was similar to that identified in *tsarl1-1* (Figure 2b,c; Figure S8b,c). The protein encoded by *tsarl1-1* and *tsarl1-3* is predicted to be C-terminally truncated (Figure 2c; Figures S9 and S10).

TSARL1 loss of function abolishes saponin accumulation in seeds

An afrosimetric assay (Koziol, 1991) was performed to estimate the saponin content in seeds harvested from WT, *tsarl1-3* and *tsarl-4* plants. Upon shaking seeds with water, seeds of the *tsarl1-3* and *tsarl-4* mutants did not produce a stable foam layer above the water, as was obtained with WT seeds (Figure S11). As the mutant seeds were sweet according to definition (Koziol, 1991), the *tsarl1-3* and *tsarl1-4* mutants were renamed *sweet seed1* (*sws1*) and *sws2*, respectively.

We extracted total metabolites from plant material and the extracts were analysed via a Dionex UltiMate 3000 Quaternary Rapid Separation UHPLC$^+$ coupled to a Compact micrOTOF-Q mass spectrometer equipped with an electrospray ion source operated in negative ion mode (LC-qToF-MS/MS), and principal component analysis (PCA) was performed using samples of quinoa seeds, leaves and roots (Figure 3a). Seed samples formed two separate groups, indicating significant differences in metabolite composition. The total abundance of 38 previously identified saponins in seeds was found to have decreased by more than 10-fold in the mutant lines compared to the WT (Figure 3b). Base peak chromatograms indicated the reduction of specific saponins in *sws1* and *sws2* compared to those of WT seeds (Figure 3c). Among 39 detected seed saponins (numbered according to their retention time; Table S3), five major saponins (i.e. CQ-15, CQ-17, CQ-22, CQ-30 and CQ-X) accumulated to greater amounts in the WT extracts (Figure 3c,d; Table S3). CQ-15, CQ-17, CQ-22 and CQ-30 have previously been characterized (Table S3) (Jarvis *et al*., 2017; Tabatabaei *et al*., 2022), but the exact structure of CQ-X remains elusive. In *sws1* and *sws2* seeds, the molar concentration of CQ-15 saponin was reduced to 15% of WT levels, CQ-17 was reduced to trace amounts (~4% of WT), and CQ-22 (~0.2% of WT), CQ-30 (~0.02% of WT) and CQ-X (~0.001% of WT) saponins were even further reduced (Figure 3d).

PCA of total metabolite features and base peak chromatograms indicated no visible difference in the relative density of metabolites in leaves (Figure 3a; Figure S12a) and roots

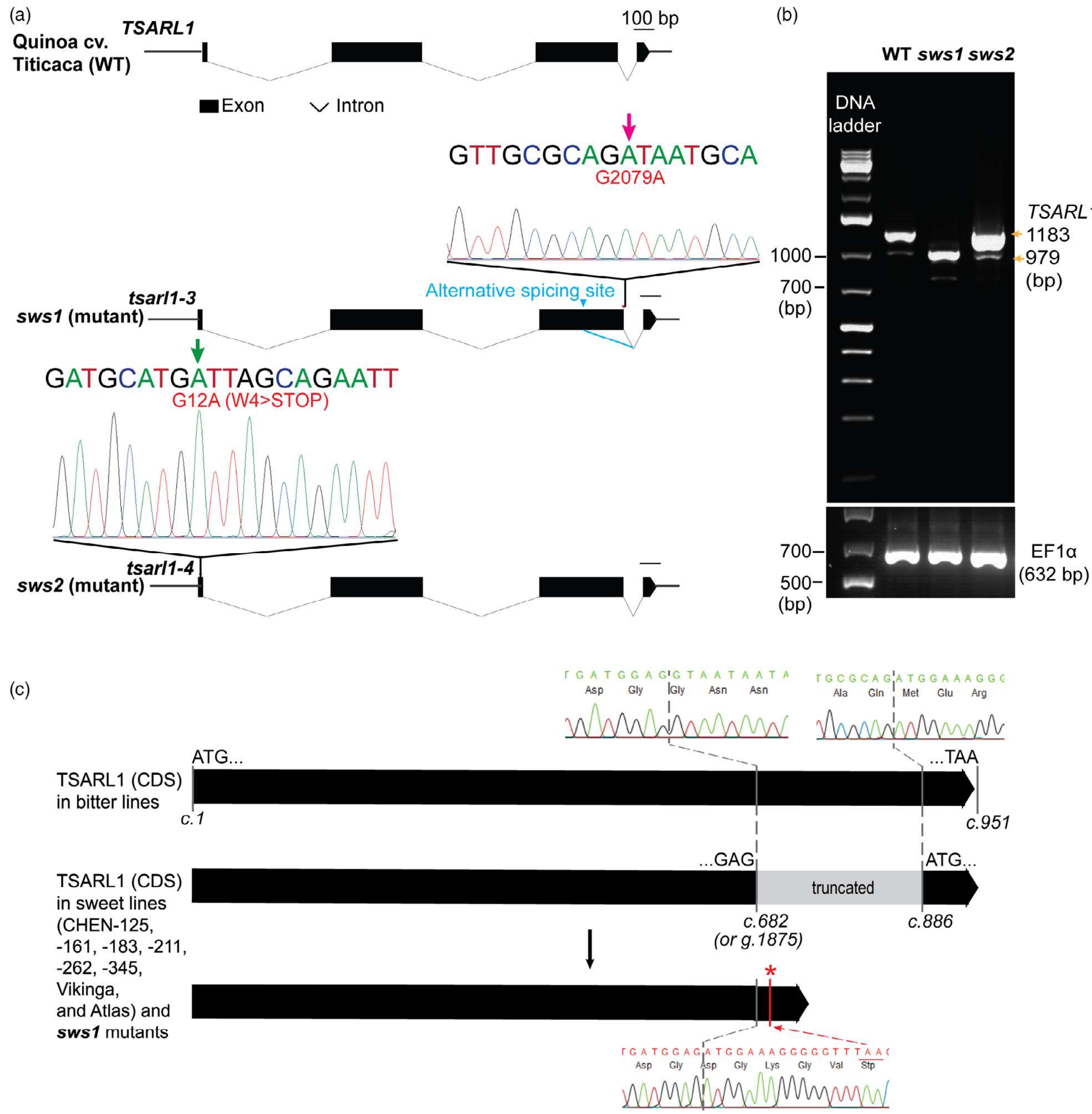
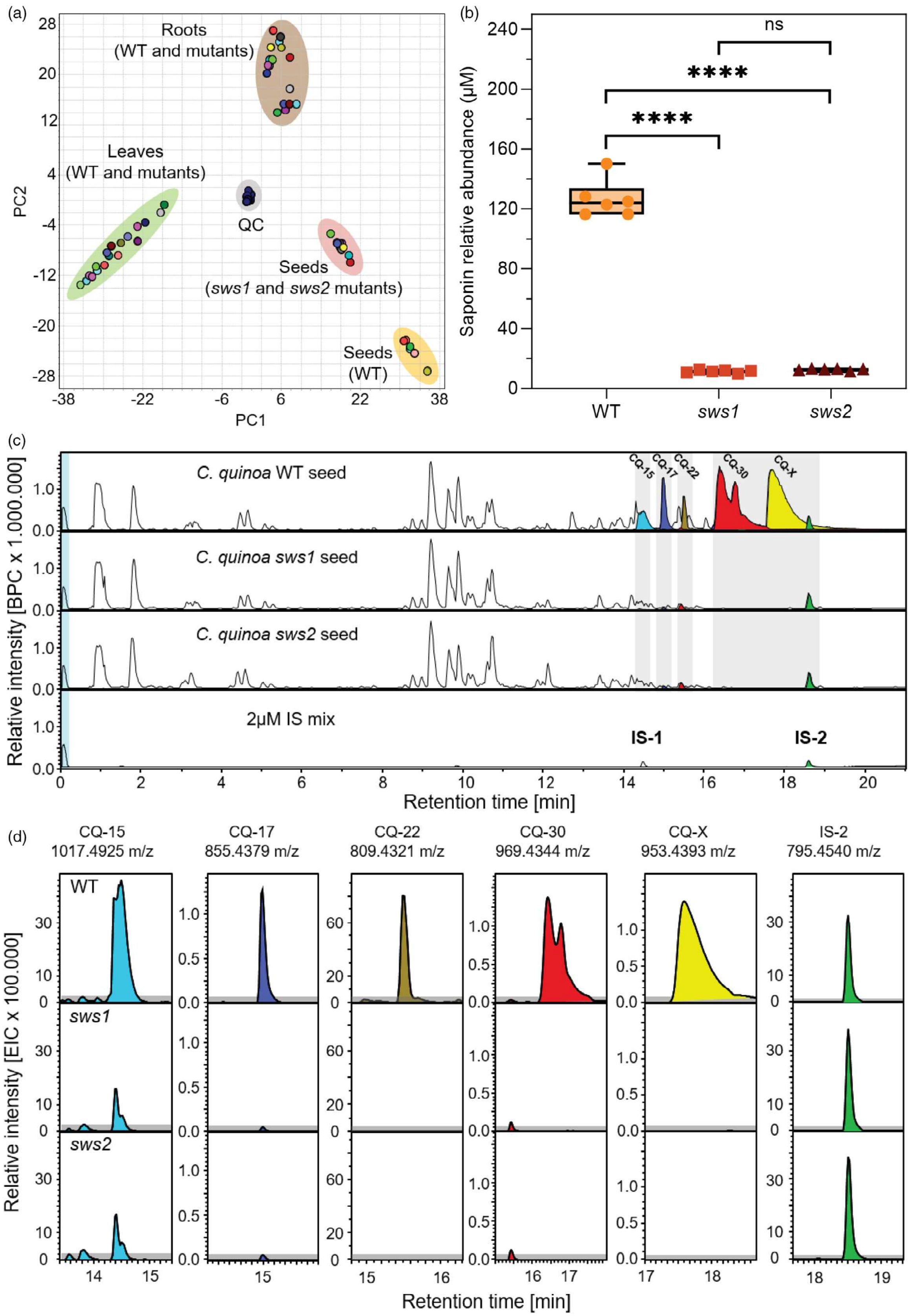


Figure 2 Identification of the two independent *tsarl1* mutants (*sws1* and *sws2*). (a) Genomic structure of *TSARL1* and its alleles. The *sws1* and *sws2* mutants contain homozygous *tsar1-3* and *tsarl1-4* alleles, respectively. *tsarl1-3* contains a G2079A substitution (magenta arrow), resulting in a cryptic splice site (G1875|G1876, blue arrowhead) and a truncated *TSARL1* transcript. *tsarl1-4* contains a G12A substitution (green arrow), resulting in a premature stop codon in the *TSARL1* transcript (W4➔STOP). (b) DNA electrophoresis of RT-PCR products. RT-PCR was performed to specifically amplify *TSARL1* transcripts from quinoa flowers harvested from cv. Titicaca wild-type (WT), *sws1* and *sws2* plants. RT-PCR products amplified from transcripts of the *elongation factor 1α* (*EF-1α*) gene were used as a positive control. The Thermo Scientific GeneRuler 1 kb Plus DNA ladder was used for sizing double-stranded DNA. (c) Identification of the cryptic splice site shown in (a) using Sanger sequencing of *TSARL1* coding sequences (CDSs) amplified from cDNA of bitter and sweet quinoa lines. The red asterisk indicates the position of a premature STOP codon in the *TSARL1* coding sequence after alternative spicing. The typical trace peaks were extracted from full sequencing data of TSARL1 coding sequences from tested sweet accessions/cultivars and *sws1* mutants.

(Figure 3a; Figure S12b) between the WT and the two mutants. Thus, TSARL1 appears to specifically influence saponin biosynthesis in seeds but not in other tissues of quinoa.

TSARL1 controls saponin biosynthesis in the seed coat

To determine how TSARL1 is involved in saponin biosynthesis in quinoa, we first performed RNA-seq analysis with total RNA isolated from developing fruits (consisting of developing seeds and fruit carpels) of WT, *sws1* and *sws2* plants. More than 2600 and 1600 differentially expressed genes were identified in the groups *sws1* versus WT and *sws2* versus WT, respectively (Figure S13a). Among the 692 shared genes (Figure S13b), we identified 22 genes encoding metabolic enzymes in the saponin biosynthesis pathway that were consistently down-regulated in



(a)
Roots
(WT and mutants)
Leaves
(WT and mutants)
QC
Seeds
(sws1 and sws2 mutants)
Seeds
(WT)
PC2
PC1
(b)
ns
****
****
Saponin relative abundance (μM)
WT
sws1
sws2
(c)
C. quinoa WT seed
C. quinoa sws1 seed
C. quinoa sws2 seed
2μM IS mix
CQ-15
CQ-17
CQ-22
CQ-30
CQ-X
IS-1
IS-2
Relative intensity [BPC x 1.000.000]
Retention time [min]
(d)
CQ-15
1017.4925 m/z
CQ-17
855.4379 m/z
CQ-22
809.4321 m/z
CQ-30
969.4344 m/z
CQ-X
953.4393 m/z
IS-2
795.4540 m/z
WT
sws1
sws2
Relative intensity [EIC x 100.000]
Retention time [min]

Figure 3 Metabolite profiling of wild-type (WT), *sws1* and *sws2* seeds, leaves and roots. (a) Principal component analysis of total metabolite features of leaf (green shade), root (brown shade), seed (orange and pink shade), harvested from WT and *sws* mutants, and quality control (grey shade) samples from 5 to 25 min retention time are shown. Groups were separated based on their tissue of origin. Seed samples from two separate groups highlighting different metabolite accumulation in WT seeds in comparison to *sws1* and *sws2* seeds. (b) Total relative saponin abundance in seeds from WT, *sws1* and *sws2*. Identified saponins were quantified, and total molar concentrations are shown to be dramatically decreased for *sws1* (green) and *sws2* (red) in comparison to WT (blue) seeds. Six biological replicates were analysed for all seed tissues (n = 6). ****P < 0.0001 (one-way ANOVA followed by Tukey's test with a 95% CI). ns, statistically not significant. (c) Base peak reverse-phase LC-qToF-MS/MS chromatograms of extracts from WT, *sws1* and *sws2* seeds. Chromatograms were normalized to the signal intensity of sodium formate calibrant (blue shading). The extraction buffer with internal standard mix of the internal standards (IS-1 and IS-2) was used as a relative retention time reference for bidesmosidic saponins (IS-1; 2 μM hederacoside C) and monodesmosidic saponins (IS-2; 2 μM α-hederin). Five identified saponins CQ-15 (turquoise), CQ-17 (blue), CQ-22 (khaki), CQ-30 (red) and CQ-X (yellow), and the internal standard (IS-2, green) are shown in colour. (d) Relative abundance of saponins in WT (top), *sws1* (middle) and *sws2* (bottom) seeds. Extracted representative chromatograms for each of the highly comparable biological replicate sample groups are shown (n = 6). Grey shading represents the noise level relative to the maximum peak height of 5%.

both *sws1* and *sws2* mutants, except for a gene encoding a squalene-synthase-like protein (LOC110718883), which was up-regulated (Figure 4a; Table S4). In both *sws1* and *sws2* mutants, two homologous genes encoding β-amyrin synthase (BAS; LOC110687932 and LOC110726782) were strongly down-regulated (Figure 4a; Table S4), consistent with data for *tsarl1-1* reported by Jarvis *et al*. (2017). LOC110726782/*CqbAS1* was previously described to contribute to saponin biosynthesis (Fiallos-Jurado *et al*., 2016). The genes encoding CYP716A78 and CYP716A79 (LOC110726788 and LOC110687936, respectively) were also strongly down-regulated in both the *sws1* and *sws2* mutants. These two enzymes have previously been characterized as β-amyrin 28-oxidases that, in quinoa, synthesize oleanolic acid from β-amyrin (Fiallos-Jurado *et al*., 2016), a committed step in the saponin biosynthesis pathway.

To confirm the transcriptomic data at the protein level, total proteins were extracted and purified from developing fruits harvested from *sws* mutants and WT plants and subjected to proteomic analysis. Approximately 11 000 proteins were identified in all protein samples (Figure S14a). PCA of protein abundance (Figure S14b) indicated that the *sws* mutants grouped close together and distinct from the WT. Among the down-regulated proteins, 30 were shared between the two comparison groups (*sws1* versus WT and *sws2* versus WT; Figure 4b,c; Table S5; Figure S14c), 22 of which likely contribute to saponin biosynthesis (Table 1). These 22 proteins included 10 enzymes that function in the triterpenoid biosynthetic pathway, two putative cytochrome P450 monooxygenases, eight glycosyltransferases and two ABC-type xenobiotic transporters (Table 1), and 20 of them were also identified in our transcriptomic data (Figure 4a; Table S4). Among these enzymes, BAS (CqbAS1; encoded by LOC110726782) and β-amyrin 28-oxidase (CYP716A78; encoded by LOC110726788) exhibited the strongest decrease in protein levels in *sws* mutants ($\log_2$ fold change from −6.14 to −7.08; Table 1; Figure 4c; Table S5), suggesting that they may play significant roles in controlling saponin biosynthesis in quinoa seeds.

TSARL1 is expressed in reproductive tissues only and localizes to the nucleus

We conducted reverse transcription-PCR (RT-PCR) to specifically amplify *TSARL1* transcripts extracted from roots, stems, panicle leaves, mature leaves, fruits, flowers, fruit coats and seeds of quinoa cv. Titicaca. *TSARL1* transcripts could only be amplified from fruits, flowers, fruit coats and seeds (Figure S15a). A recombinant green fluorescent protein (GFP) fused with TSARL1 transiently expressed in *Nicotiana benthamiana* leaves localized only to the nucleus of mesophyll cells, while free GFP, the negative control, also localized to the cytosol (Figure S15b).

Mutation of *TSARL1* does not influence normal plant growth and reproduction

We compared the growth and yield of the mutant plants with those of the WT under both greenhouse and field conditions and did not observe significant differences in shoot biomass (fresh and dry weight, Figure 5a,b), shoot height (both inbred M_4 generation mutant lines and backcrossed BCF_2 generation) (Figure 5c,d; Figures S16a–c), flower morphology (Figure 5e) or fruit maturation (Figure 5f) between the *sws* mutants and WT plants. Both *sws* mutant lines and WT plants completed their life cycles normally and produced seeds of similar size (Figure 5h; Figure S16d) and weight (Figure 5i). Seed yield per plant was scored after backcrossing *sws* mutants to quinoa cv. Titicaca (WT). In the BCF_2 generation, seed production between individual plants varied considerably, presumably due to inbreeding depression; however, there was no significant difference in seed production between groups of tested genotypes (*tsarl1-3* or *tsarl1-4* homozygotes, *tsarl1-3* or *tsarl1-4* heterozygotes, and *TSARL1* homozygotes; Figure 5j). Furthermore, when plants were treated with the generalist insect herbivore species *Spodoptera exigua*, the development of caterpillars was comparable when larvae were allowed to feed on foliage of the WT and *sws* mutants (Figure 5k).

To test the performance of *tsarl1* mutants in field conditions, seeds harvested from BCF_2 plants were sown in the 2023 growth season in Zealand, Denmark. The backcrossed line BCF_3 *tsarl1-3/tsarl1-3* produced comparable seed yields to those produced by the control line BCF_3 *TSARL1/TSARL1* (1) in both sector and bulk harvested experiments (Figure 5l; Figure S16e). The BCF_3 *tsarl1-4/tsarl1-4* line produced slightly less seeds than the control line BCF_3 *TSARL1/TSARL1* (2) in the sector experiments (Figure 5l); however, the overall seed yield of the BCF_3 *tsarl1-4/tsarl1-4* line showed an ~3-fold reduction in comparison to the control line in the bulk harvested experiment (Figure S16e). Overall, the consistency between the preliminary field trial and greenhouse experiments with the *tsarl1-3* mutant suggested that the sweet mutant line may perform in the field as well as its parent, the bitter quinoa cv. Titicaca. The negative performance observed in the BCF_3 *tsarl1-4/tsarl1-4* line may be due to unknown environmental factors and further field trial experiments need to be conducted with this line.

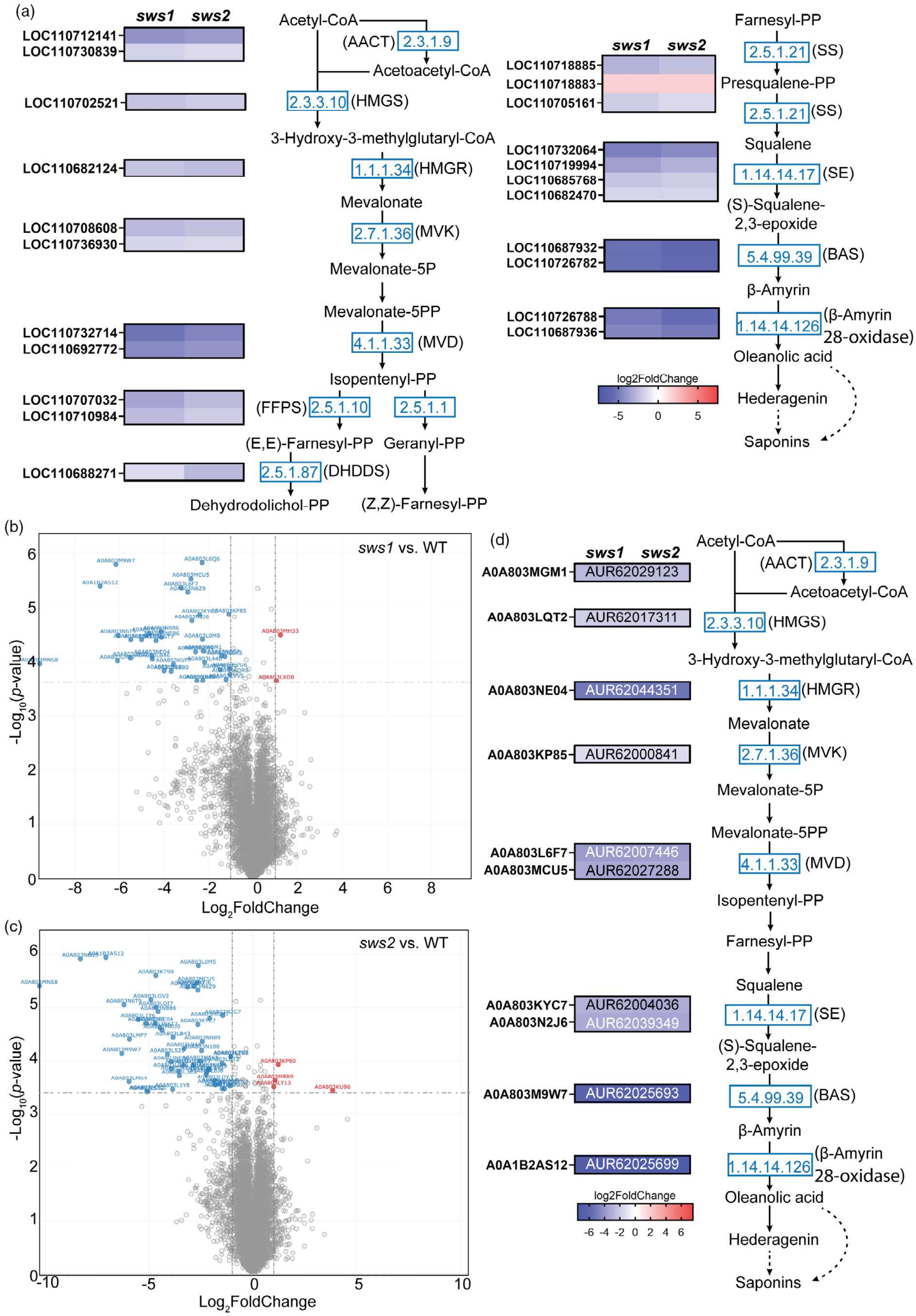

(a)
sws1 sws2
LOC110712141
LOC110730839
LOC110702521
LOC110682124
LOC110708608
LOC110736930
LOC110732714
LOC110692772
LOC110707032
LOC110710984
LOC110688271
Acetyl-CoA
(AACT) 2.3.1.9
Acetoacetyl-CoA
2.3.3.10 (HMGS)
3-Hydroxy-3-methylglutaryl-CoA
1.1.1.34 (HMGR)
Mevalonate
2.7.1.36 (MVK)
Mevalonate-5P
Mevalonate-5PP
4.1.1.33 (MVD)
Isopentenyl-PP
(FFPS) 2.5.1.10
2.5.1.1
(E,E)-Farnesyl-PP
Geranyl-PP
2.5.1.87 (DHDDS)
Dehydrodolichol-PP
(Z,Z)-Farnesyl-PP
LOC110718885
LOC110718883
LOC110705161
LOC110732064
LOC110719994
LOC110685768
LOC110682470
LOC110687932
LOC110726782
LOC110726788
LOC110687936
log2FoldChange
-5 0 5
Farnesyl-PP
2.5.1.21 (SS)
Presqualene-PP
2.5.1.21 (SS)
Squalene
1.14.14.17 (SE)
(S)-Squalene-2,3-epoxide
5.4.99.39 (BAS)
β-Amyrin
1.14.14.126 (β-Amyrin 28-oxidase)
Oleanolic acid
Hederagenin
Saponins
(b)
sws1 vs. WT
-Log10(p-value)
Log2FoldChange
(c)
sws2 vs. WT
-Log10(p-value)
Log2FoldChange
(d)
sws1 sws2
A0A803MGM1 AUR62029123
A0A803LQT2 AUR62017311
A0A803NE04 AUR62044351
A0A803KP85 AUR62000841
A0A803L6F7 AUR62007446
A0A803MCU5 AUR62027288
A0A803KYC7 AUR62004036
A0A803N2J6 AUR62039349
A0A803M9W7 AUR62025693
A0A1B2AS12 AUR62025699
log2FoldChange
-6 -4 -2 0 2 4 6
Acetyl-CoA
(AACT) 2.3.1.9
Acetoacetyl-CoA
2.3.3.10 (HMGS)
3-Hydroxy-3-methylglutaryl-CoA
1.1.1.34 (HMGR)
Mevalonate
2.7.1.36 (MVK)
Mevalonate-5P
Mevalonate-5PP
4.1.1.33 (MVD)
Isopentenyl-PP
Farnesyl-PP
Squalene
1.14.14.17 (SE)
(S)-Squalene-2,3-epoxide
5.4.99.39 (BAS)
β-Amyrin
1.14.14.126 (β-Amyrin 28-oxidase)
Oleanolic acid
Hederagenin
Saponins

Figure 4 Differences in gene expression at transcript and protein levels in *sws* mutants compared to the wild-type (WT). (a) Differences in expression of genes encoding enzymes in mevalonate pathway and saponin biosynthesis pathway. Gene names (with the prefix LOC) are indicated at the left side of each panel. Rectangles next to the gene names are coloured according to their fold change in expression ($\log_2$) in *sws1* and *sws2* mutants compared to those in WT plants. The experiment was performed with four biological replicates, and the fold changes in expression are all statistically significant with post hoc adjusted *P*-value (P_{adj}) < 0.05. Colour gradient bar showing fold change in expression ($\log_2$) is indicated in the figure. (b) and (c) Volcano plots comparing protein abundance in developing fruits harvested from *sws* mutants and WT plants. Proteins that display a significantly reduced or increased expression in *sws* mutants relative to the WT are highlighted in blue and red, respectively. Dashed lines indicate thresholds ($-1 > \log_2$FC (*sws* mutant/WT) > 1; $-\log_{10}$(*P*-value) for the protein with the highest significant post hoc adjusted *P*-value (P_{adj}) below 0.05). (d) Differences in protein/enzyme abundance in mevalonate pathway and saponin biosynthesis pathway. Protein names are indicated to the left of each rectangle, and the corresponding gene names (with the prefix AUR) are indicated inside each rectangle. Rectangles next to the protein names are coloured according to their fold change in expression ($\log_2$) in *sws1* and *sws2* mutants compared to those in WT plants. The experiment was performed with three biological replicates, and the fold changes in expression are all statistically significant with post hoc adjusted *P*-value (P_{adj}) < 0.05. Colour gradient bar showing fold change in expression ($\log_2$) is indicated in the figure. The metabolic pathways were adapted from the terpenoid backbone biosynthesis pathway (cqi00900) and the sesquiterpenoid and triterpenoid biosynthesis pathway (cqi00909) of the KEGG database (https://www.genome.jp/kegg/).

Discussion

In this work, we developed a very large quinoa mutant library that allowed for the identification of specific nucleotide substitutions in the quinoa genome. We used this method to identify mutants in several genes of interest, among them two loss-of-function mutants of *TSARL1*, which have an improved agronomical trait, namely, the absence of antinutritional saponins in the seed coat.

Loss of function of TSARL1 caused a dramatic decrease in the expression of genes encoding *Cq*bAS1 and β-amyrin 28-oxidases (CYP716A78 and CYP716A79), which catalyse the committed step of the saponin biosynthesis pathway post 2,3-oxidosqualene ((*S*)-squalene-2,3-epoxide; Figure 4). The downregulation of genes in the mevalonate (MVA) pathway in *sws* mutants may also be due to the disruption of TSARL1, as most of the identified down-regulated genes are predicted to contain *cis*-elements recognized by TSARL1 (Table S5).

Disruption of the MVA pathway in vegetative tissues severely inhibits plant growth and reproduction, as observed in Arabidopsis *hmg1* mutants (Suzuki *et al*., 2004). The observed downregulation of the MVA pathway in *sws* mutants may thus be specific for seeds, as we did not detect phenotypic changes in the vegetative tissues of these plants.

Saponins have been reported to play a role in the defence mechanisms of various plant species (Augustin *et al*., 2011, 2012; Osbourn, 1996). Specifically, the monodesmosidic saponin, 3-O-β-D-glucopyramosyl hederagenin, found in quinoa, exhibits molluscicidal and fungicidal properties in the roots of *Dolichos kilmandscharicus* (Kuljanabhagavad and Wink, 2009). Triterpenoid saponins from quinoa have been shown to be highly toxic to cold-blooded animals (Joshi *et al*., 2008; San Martín *et al*., 2008). While previous research has suggested a positive correlation between seed saponin content and fitness parameters such as viability and germination rate in other quinoa genotypes (Granado-Rodríguez *et al*., 2021), our studies indicate that the *TSARL1* gene primarily regulates saponin content in seeds. Thus, the loss of function of this gene should not significantly impact resistance to pathogens and pests that primarily feed on plant parts other than seeds. Indeed, our findings demonstrate that downregulation of *TSARL1*-controlled saponin accumulation in quinoa seeds does not affect resistance to the generalist insect herbivore *S. exigua* when feeding on quinoa leaves (Figure 5k). Furthermore, quinoa fitness remains unaffected in *sws* mutants compared to the wild type under both greenhouse and field conditions (Figure 5). Additionally, the successful cultivation of sweet quinoa varieties worldwide suggests that the loss of function of the TSARL1 gene may be advantageous for obtaining saponin-free seeds without compromising yield significantly.

In summary, we removed saponin, an important antinutrient, from quinoa seeds of a commercial variety of quinoa. Plant growth and reproduction appeared to be unaffected in *sws1* and *sws2* mutants. Thus, using site-directed genotype screening, we made a major step forward in improving quinoa seed quality without compromising yield. The library, therefore, appears to be a promising tool to accelerate quinoa breeding.

Methods

Plant materials

The *Chenopodium quinoa* Willd. (quinoa) cultivar Titicaca (QuinoaQuality, Denmark) was used to generate a SNP library. The commercial quinoa cultivars Vikinga (QuinoaQuality, Denmark), Atlas and Riobamba (Radicle, the Netherlands) were used for phenotypic analysis. For the phenotypic and genome-wide association study (GWAS), an additional 98 quinoa accessions (Table S6) available from GenBank Information System of the IPK Gatersleben (GBIS/I, https://gbis.ipk-gatersleben.de/gbis2i/faces/index.jsf) were used.

Generation of a SNP library in quinoa

A SNP library in quinoa was generated as described previously (Knudsen *et al*., 2022) with some modifications. Quinoa M_0 seeds were soaked in either 0.1 or 0.2% (w/v) EMS solution for 16 h, then air-dried in a fume hood for 24 h. The mutagenized M_1 seeds were sown at a density of 6 kg ha^{-1} and grown to maturity under field conditions in Denmark during the 2019 and 2020 field seasons. When all quinoa plants were mature, brown and dry, M_2 seeds were harvested from pools of approximately 100 plants (Figure S7a). The M_2 seeds were then screened to identify targeted specific nucleotide substitutions (Figure S7b).

Determination of mutation load using (n)GBS/ddRAD analysis

Normalized genotyping by sequencing/double digest restriction-site associated DNA [(n)GBS/ddRAD] and subsequent analysis on mutagenized quinoa plants was performed on samples of more than 19 M_2 progeny plants of individual selfed M_1 parent plants along with more than five M_2 plants from different individual

Table 1 Proteins that show altered abundance in *sws* mutants

Protein ID (UniProt)	Gene ID/Name	Protein description	Log$_2$ difference		*P*-value	
			sws1 versus WT	*sws2* versus WT	*sws1* versus WT	*sws2* versus WT
Triterpenoid biosynthetic pathway						
A0A1B2AS12	AUR62025699 (LOC110726788)	β-Amyrin 28-oxidase (CYP716A78)	−6.85	−7.08	4.04E-06	1.15E-06
A0A803M9W7	AUR62025693 (LOC110726782)	β-Amyrin synthase (CqbAS1)	−6.14	−6.31	1.61E-06	7.23E-05
A0A803NE04	AUR62044351 (LOC110718708)	HMG-CoA reductase (HMGR)	−4.53	−4.71	7.82E-05	1.98E-05
A0A803L6F7	AUR62007446 (LOC110732714)	Diphosphomevalonate decarboxylase	−3.23	−3.14	4.32E-06	4.10E-06
A0A803MCU5	AUR62027288 (LOC110692772)	Diphosphomevalonate decarboxylase	−2.78	−2.72	2.98E-06	3.41E-06
A0A803N2J6	AUR62039349 (LOC110732064)	Squalene epoxidase	−2.74	−2.85	1.76E-05	3.94E-06
A0A803KYC7	AUR62004036 (LOC110682470)	Squalene epoxidase	−2.39	−2.65	1.36E-05	2.09E-05
A0A803MGM1	AUR62029123 (LOC110730839)	Acetyl-CoA acetyltransferase	−2.23	−2.09	6.35E-05	1.60E-05
A0A803LQT2	AUR62017311 (LOC110702521)	HMG-CoA synthase	−1.47	−1.49	1.41E-04	1.11E-04
A0A803KP85	AUR62000841 (LOC110736930)	Mevalonate kinase	−1.08	−1.07	1.33E-05	8.43E-05
Other cytochrome P450 monooxygenases						
A0A803NE86	AUR62044433 (LOC110719310 or LOC110721253)	CYP72A219-like	−4.10	−3.94	3.59E-05	1.04E-04
A0A803LB44	AUR62009083 (LOC110705288)	CYP72A219-like	−3.98	−3.32	1.46E-04	5.88E-05
UDP-glycosyltransferases						
A0A803N6T9	AUR62041447 (LOC110711362)	Anthocyanin 3′-O-beta-glucosyltransferase-like (UGT73B-like)	−6.04	−6.20	3.26E-05	8.88E-06
A0A803LQI7	AUR62017216 (LOC110702441)	Anthocyanidin 3-O-glucoside 2″-O-glucosyltransferase-like (UGT79B-like)	−4.65	−4.66	3.03E-05	9.91E-06
A0A803N6T7	AUR62041445 (LOC110711361)	Anthocyanin 3′-O-beta-glucosyltransferase-like (UGT73B-like)	−4.34	−4.47	4.12E-05	2.40E-05
A0A803N886	AUR62042030 or/and AUR62041446 (LOC110711361 or/and LOC110738265)	Anthocyanin 3′-O-beta-glucosyltransferase-like (UGT73-like)	−4.12	−4.57	2.83E-05	1.21E-05
A0A803KVF7	AUR62003015 (LOC110720877)	Scopoletin glucosyltransferase-like (UGT73B-like)	−3.57	−2.88	1.12E-04	1.20E-04
A0A803NB89	AUR62043317 (LOC110719072)	UDP-glycosyltransferase 74E1-like (UGT74E-like)	−2.57	−2.44	6.59E-05	4.34E-05
A0A803LMF7	AUR62015468 (LOC110689329)	Cellulose-synthase superfamily-derived glycosyltransferase	−6.07	−5.94	9.60E-05	3.94E-05
A0A803KT98	AUR62002256 (LOC110717430)	Cellulose-synthase superfamily-derived glycosyltransferase	−4.99	−4.67	3.95E-05	2.50E-05
Transporters						
A0A803N6Z9	AUR62041507 (LOC110697519)	ABC-type xenobiotic transporter	−2.92	−2.65	5.20E-06	4.71E-06
A0A803NA49	AUR62042848 (LOC110694020)	ABC-type xenobiotic transporter	−2.51	−2.10	2.23E-04	1.47E-04

parent M_1 plants for the same mutagenesis treatment. The (n) GBS/ddRAD analysis, including DNA extraction, library construction, sequencing and initial bioinformatic analysis, was performed at LGC Genomics GmbH (Germany) as described by Knudsen *et al*. (2022).

After SNP identification, the following steps were used to identify SNPs induced by the mutagenesis treatment: (1) removal of loci having any neighbouring SNP loci within 50 bp, (2) disregarding loci whose alternate fraction was lower than 35% in a single sample, (3) removal of SNP loci occurring in more than one group originating from different M_1 plants, (4) removal of SNP loci occurring only once in M_2 samples from the same M_1 parent and (5) removal of outlier samples from the analysis based on interquartile range. The estimated mutation density number was based on the genomic coverage in individual (n)GBS/ddRAD analyses.

Screening of quinoa libraries for specific nucleotide substitutions

Genomic DNA (gDNA) was extracted from 25% of the seeds from each seed pool, where each pool represented approximately 100 individual plants. First, all seed pools were screened for possible specific nucleotide substitutions occurring in the M_1 generation. As described previously (Knudsen *et al*., 2022), gDNA extracted from each pool was diluted, mixed with droplet digital PCR (ddPCR) reagents and a master mixture of the TaqMan probe solution, then applied to 96-well PCR plates, where each well represented one screened seed pool. The reaction mixture was prepared on a QX200 AutoDG Droplet Digital PCR system (Bio-Rad, USA). In the system, 20 μL of the reaction mixture was separated and encapsulated into approximately 20 000 droplets; therefore, each droplet encased approximately one genome equivalent of gDNA. The droplets were dispensed into 96-well plates and heat-sealed at 180 °C for 5 s with pierceable foil using a PX1 PCR plate sealer (Bio-Rad, USA). The 96-well plates were then inserted into a thermal cycler (Uno96, VWR, USA) for PCR amplification using pre-set thermal cycles: 95 °C for 10 min, followed by 40 cycles of 94 °C for 30 s and 55 °C for 1 min, ending with 98 °C for 10 min. The 96-well plates were next placed into a QX200 Droplet reader (Bio-Rad) to count positive and negative droplets for each fluorophore, and the raw data obtained were analysed using Bio-Rad QuantaSoft software (Bio-Rad). Second, after detecting seed pools containing the targeted specific nucleotide substitution events, approximately 1000 seeds from the remaining 75% of the seeds were sown in soil. gDNA was extracted from the first leaves of quinoa seedlings and grouped into sub-pools containing gDNA from approximately 100 seedlings. These samples were diluted and subjected to the ddPCR process and analysis as described above. If positive amplification (fluorescence signals) was detected from a seedling sub-pool, each gDNA sample from individual seedlings of that sub-pool was subjected to ddPCR screening and analysis, as described above, to detect seedlings (M_2) having specific nucleotide substitutions in their genomes. These specific nucleotide substitutions were further validated using specific PCR amplification and Sanger sequencing of purified PCR products (Eurofins Genomics, Germany). See Figure S7b for a schematic explanation.

Isolation of *tsarl1* mutants

Site-directed genotype screening was used to identify two quinoa parent plants (M_2): one having a heterozygous G2079A nucleotide substitution and the other one having a heterozygous G12A nucleotide substitution in the *TSARL1* gene. These plants were grown and selfed under greenhouse conditions (Carlsberg Research Laboratory, Denmark), and dried M_3 seeds were collected from fully mature and dried plants. M_3 seeds were sterilized using ethanol (70% and 95%, v/v), placed on agar containing half-strength MS medium and grown under long-day conditions (8-h-dark (16 °C)/16-h-light (23 °C) cycle, 110 μmol photons m^{-2} s^{-1}). Small leaf cuttings from 2-week-old seedlings were used for PCR amplification using a Phire® Plant Direct Kit (FINNZYMES, Finland). CqTSARL1-G2079A-F/CqTSARL1-G2079A-R and CqTSARL1-W4Stop-F/CqTSARL1-W4Stop-R primer pairs (Table S7) were used for PCR amplification and Sanger sequencing (Eurofins Genomics, Germany) to detect G2079A and G12A nucleotide substitutions, respectively, in the *TSARL1* gene in M_3 quinoa plants. Homozygous *tsarl1* mutants, containing G2079A or G12A nucleotide substitutions, were identified and transferred to soil for growth and seed production under greenhouse conditions (long-day conditions (8-h-dark/16-h-light cycle), 200 μmol photons m^{-2} s^{-1}, University of Copenhagen, Denmark). See Figure S8a for a schematic explanation.

Genomic DNA extraction

DNA was extracted using a version of the cetyltrimethylammonium bromide (CTAB) method. Plant tissue (50–100 mg) was snap frozen. Adapters and 24x tissue lyser racks were stored at −80 °C and prepared on dry ice. Frozen tubes were placed in the cold racks, and a 4-mm metal sphere was added to each tube. CTAB extraction buffer [3% w/v CTAB, 28% w/v NaCl, 4% w/v EDTA pH 8, 10% w/v Tris–HCl pH 8, 3% w/v polyvinylpyrrolidone (PVP, 40 kDa), 0.2% w/v β-mercaptoethanol] was prepared by mixing chemicals sequentially under a fume hood and stored in a 60 °C water bath. Tissue disruption was carried out in a Retsch-QIAGEN Tissuelyser II at 25 Hz for 1 min, followed by an additional 2 min on dry ice before another round of shaking. Once processed, samples were kept on dry ice and transported to the laboratory. Pre-warmed CTAB extraction buffer (500 μL) was added to each sample, and the tubes were mixed gently by inverting before being placed into a 60 °C bath for 60 min. The homogenates were then centrifuged for 10 min at 10 000 × ***g*** at room temperature. The supernatant of each sample was transferred into a clean tube and 5 μL of RNase (10 mg/mL in water) was added, which was incubated at room temperature for 15 min. Subsequently, 500 μL of chloroform: isoamyl alcohol (24:1, v/v) was added, and the tubes were gently mixed by inverting before being centrifuged for 10 min at 14 000 × ***g*** at room temperature to separate the phases. The upper phase was transferred to a clean tube (approximately 400–500 μL), and 0.7 volume of isopropanol (for 450 μL, 315 μL isopropanol) was mixed in and incubated at −20 °C for 15 min. The mixture was then centrifuged for 10 min at 10 000 × ***g*** 4 °C. The supernatant was decanted, and the pellet was washed using 500 mL ice-cold 70% (v/v) ethanol (stored in the freezer at −20°C). After centrifuging for 5 min at 10 000 × ***g*** at 4 °C, another ethanol wash was performed. Any remaining ethanol was removed, and if the pellet obstructed this process, it was centrifuged again. The pellet was then dried in a chemical fume hood for approximately 30 min until it became white, making sure not to overdry the DNA. Finally, the DNA was gently resuspended in 50 μL of nuclease-free water and stored at −20 °C.

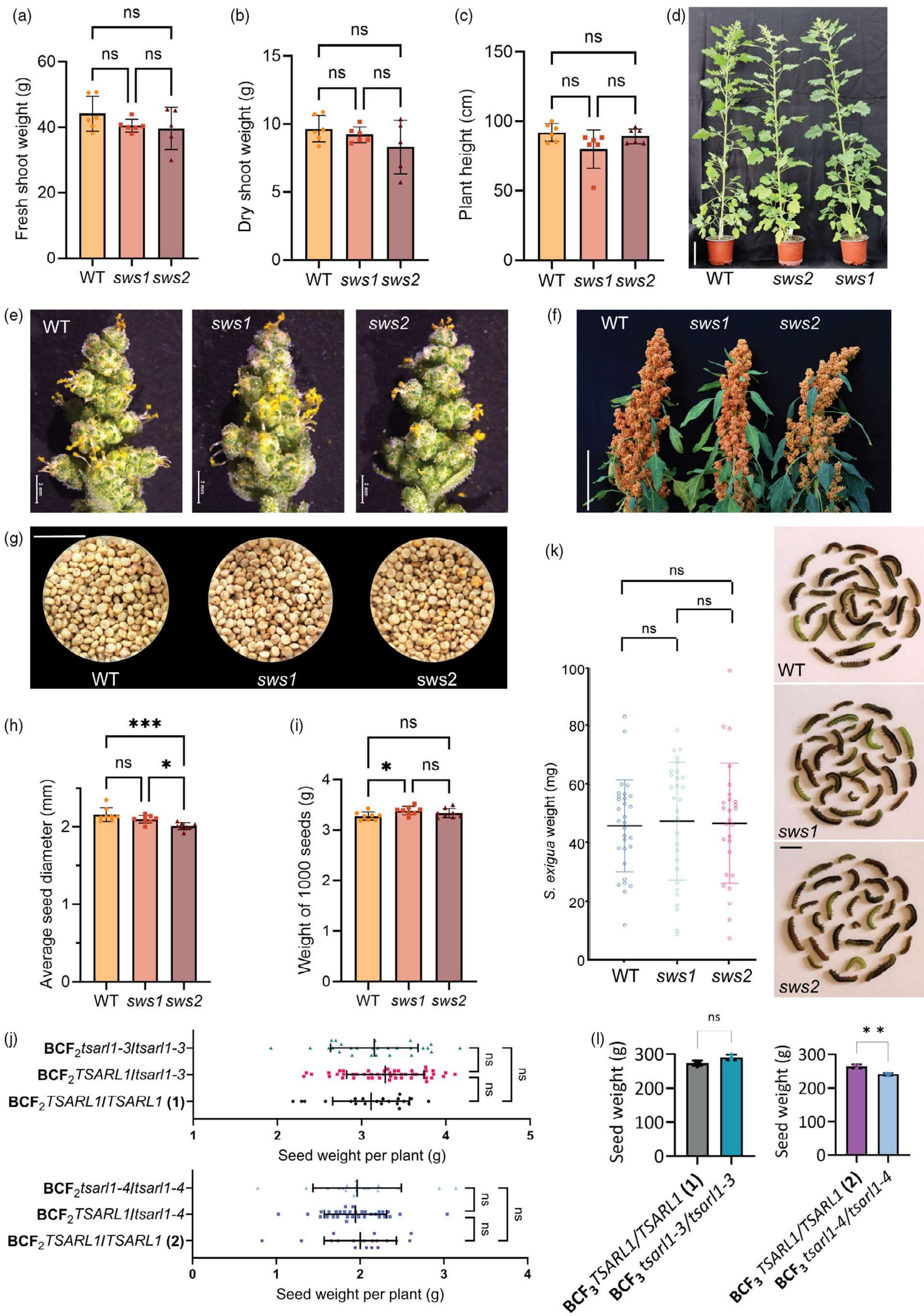

(a)
Fresh shoot weight (g)
WT sws1 sws2
ns
(b)
Dry shoot weight (g)
(c)
Plant height (cm)
(d)
WT sws2 sws1
(e)
WT
sws1
sws2
(f)
WT sws1 sws2
(g)
WT sws1 sws2
(h)
Average seed diameter (mm)
***
*
(i)
Weight of 1000 seeds (g)
(j)
BCF2tsarl1-3/tsarl1-3
BCF2TSARL1/tsarl1-3
BCF2TSARL1/TSARL1 (1)
Seed weight per plant (g)
BCF2tsarl1-4/tsarl1-4
BCF2TSARL1/tsarl1-4
BCF2TSARL1/TSARL1 (2)
Seed weight per plant (g)
(k)
S. exigua weight (mg)
WT sws1 sws2
(l)
Seed weight (g)
BCF3 TSARL1/TSARL1 (1)
BCF3 tsarl1-3/tsarl1-3
BCF3 TSARL1/TSARL1 (2)
BCF3 tsarl1-4/tsarl1-4
**

Figure 5 Phenotypic analysis of the wild-type (WT) and *sws* mutants. (a) Fresh shoot weight per plant. 48 days after sowing (DAS) plants were used. Data are the mean ± SD (n = 6). (b) Dry shoot weight per plant. 48 DAS plants were used. Data are the mean ± SD (n = 6). (c) Height of plants 40 DAS. Data are the mean ± SD (n = 6). (d) 40-DAS WT, *sws1* and *sws2* plants. Scale bar, 10 cm. (e) Flowers of WT, *sws1* and *sws2*. Scale bar, 2 mm. (f) Panicles of WT, *sws1* and *sws2* plants at 91 DAS. Scale bar, 10 cm. (g) Dried seeds from WT, *sws1* and *sws2* plants. Scale bar, 10 mm. (h) Average seed diameter (~64–112 seeds per sample were measured). Data are the mean ± SD (n = 8). (i) Weight of 1000 seeds. Data are the mean ± SD (n = 8). *$P < 0.05$ (one-way ANOVA followed by Tukey's test with a 95% CI). ns, statistically not significant. (j) Total seed weight produced per plant. Seeds were harvested from individual plants with different genotypes in the F_2 generation after backcrossing *sws* mutants to quinoa cv. Titicaca WT (BCF_2 generation). BCF_2 *TSARL1/TSARL1* (1) or (2) plants were the second progeny after backcrossing *sws1* or *sws2* mutants, respectively, to quinoa cv. Titicaca WT. Plants were grown under greenhouse conditions (long-day [8-h-dark/16-h-light cycle] conditions, 200 μmol photons m^{-2} s^{-1}). Plants were irrigated with nutrient water once a week in the first 2 weeks after sowing seeds. Data are the mean ± SD. ns, statistically not significant. (k) Non-choice test for herbivore susceptibility. Larvae of *Spodoptera exigua* were allowed to feed on young plants of the bitter quinoa cv. Titicaca WT and *sws* mutants. After 8 days, the larvae were collected and their weight was scored. (l) Weight (g) of total quinoa seeds per sector harvested in the field. One plant sector included 25 individual plants. Eight grams of BCF_3 seeds were sown in the field (Zealand, Denmark) in April 2023 and harvested in September 2023. BCF_3 *TSARL1/TSARL1* (1) or (2) plants were the third progeny after backcrossing *sws1* or *sws2* mutants, respectively, to quinoa cv. Titicaca WT. Data are the mean ± SD (n = 3). **$P < 0.01$ (Student's t-test). ns, statistically not significant. All plants from (a) to (i) were grown under greenhouse conditions (long-day conditions [8-h-dark/16-h-light cycle], 200 μmol photons m^{-2} s^{-1}). Plants were irrigated with nutrient water once a week until anthesis. *$P < 0.05$, ***$P < 0.001$ (one-way ANOVA followed by Tukey's test with a 95% CI).

Genomic data processing

Genotypes were obtained from 124 individuals. Of these, 34 datasets were generated in a previous study (Patiranage *et al.*, 2022), and the raw data were downloaded from the NCBI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra). Accession numbers and SRA identifiers are given in Table S6. These samples were sequenced using the NovaSeq Illumina platform with 150-bp paired-end reads, with 8× to 10× coverage. The remaining 90 datasets were generated in this study using the same technique to reduce bias. Sequencing was performed at Novogene Europe (Novogene, China). Between 11 and 16 Gb of reads was generated per sample, aiming for a coverage of 6–10× across the quinoa haploid genome (1.4 Gbp). The raw fastq files were filtered and trimmed using fastp (Chen *et al.*, 2018) using the following parameters: base phred quality (−q 20), maximum percentage of unqualified bases (−u 30), read minimum length after trimming (−l 50), fixed trimming from 5′ end (−5 20), sliding window (-W 5), mean quality for sliding window (−M 20) and low-quality trimming for 3′ end enabled (−g). More than 94% of reads per sequenced sample passed the filters, and the total number of reads per sample ranged from 71.5 to 104.2 million.

Generation of binary alignment map (BAM) files

After filtering and trimming, reads were mapped to the quinoa reference genome V1 (Genome ID: 33827, v1.4; CoGe database, https://genomevolution.org/coge/). This is a linkage-map-anchored pseudomolecule assembly based on the unscaffolded assembly in Genome id28667 (CoGe database). Minimap2 was used for mapping, which has been shown to perform well with bwa-mem on short (>100 bp), paired, non-spliced reads, with higher speed (Li, 2018, 2021). Reads with mapping quality above 30 were selected using SAMtools v1.8 (Danecek *et al.*, 2021). Processing of the resulting sorted BAM files was performed using GATK4 (version 4.4.0.0) (Depristo *et al.*, 2011; McKenna *et al.*, 2010). Individual reads from each file were then assigned read group labels, based on the library, platform and sample name, using the GATK4 AddOrReplaceReadGroups (Picard) tool. Duplicate reads (PCR artefacts) were flagged within BAM files using the GATK4 MarkDuplicates (Picard) tool. BAM files belonging to the same samples but originating from different lanes were later merged.

Variant calling and filtering

A subset of 100 accessions was used for calling variants by running bcftools (version 1.8) mpileup piped into bcftools call (Lefouili and Nam, 2022; Li and Barrett, 2011). Calling was performed for each chromosome, filtering for read mapping quality (MQ > 30) and base quality (BQ > 20). Each chromosome variant call format (VCF) file was then examined in R (v4.3.0) by extracting the statistics contained within the INFO field. Based on these statistics, bcftools view (bcftools v.1.8) was then used to perform a first round of filtering, using the following parameters: total depth (500 < DP < 950), mapping quality (MQ > 40), Mann–Whitney U test of Mapping Quality Bias (MQB), Base Quality Bias (BQB), Read Position Bias (RPB) and Mapping Quality versus Strand Bias (MQSB) all set to >0.05. This step produced a main SNP dataset for each chromosome. This dataset was filtered further by minimal allele frequency (MAF) and maximum fraction of missing calls for each position (MISS) using vcftools (v0.1.16) (Danecek *et al.*, 2011). A preliminary SNP-based population structure was computed from this filtered set using Tassel PCA (Bradbury *et al.*, 2007). MAF filtering was performed separately on coastal (4 accessions, MAF >0.5) and highland (96 accessions, MAF >0.04) accessions to avoid losing coastal-specific SNPs. Sets from coastal and highland accessions were then merged. To phase genotypes and infer missing calls, the tool Beagle (v5.0) was used with default settings (Browning *et al.*, 2018, 2021). At this point, all chromosome VCF files were concatenated into a genome-level high-confidence SNP set, which contained 3 637 746 SNPs.

Linkage disequilibrium (LD)

The LDkit (Tang *et al.*, 2020) tool was used for calculating LD plots in the selected region.

Population structure

Genotype likelihoods were calculated using ANGSD (v0.940-stable) (Korneliussen *et al.*, 2014), filtering by MAF >0.05, mapping quality (−minMaQ 40) and base quality (−minQ 20). The general syntax used is as follows:

```
angsd -b $bam_list -out $output_prefix -ref $ref_genome -
doMaf 1 -minMaf 0.05 -minQ 20 -minMapQ 40 -uniqueOnly 1
-remove_bads 1 -only_proper_pairs 1 -C 50 -doGlf 2 -
```

doCounts 1 -doDepth 1 -minInd 75 -GL 1 -SNP_pval 1e-6 -doMajorMinor 1 -nThreads $num_threads >& $output_prefix.log.

Next, PCAngsd (v1.2) (Meisner and Albrechtsen, 2018) was used to generate eigenvectors and eigenvalues. Population structure was visualized in R using ggplot2 v4.3.1 (https://CRAN.R-project.org/package=ggplot2).

GWAS

GWAS was conducted by performing linear mixed-model analysis using the likelihood ratio test in GEMMA (v0.98.5) (Zhou, 2017) and in GAPIT (a Genome Association and Predicted Integrated Tool) (v3.4) (Wang and Zhang, 2021). For GAPIT, the analysis was run using GLM ('general' linear model), MLM (Mixed linear model), MLMM (Multi-locus mixed-model), FarmCPU (Fixed and random model Circulating Probability Unification) and BLINK (Bayesian-information and Linkage disequilibrium Iteratively Nested Keyway) methods. For both tools, five principal components from the PCAngsd output were included as covariates and in the internally generated centred kinship matrix to account for the observed population stratification. To perform correction for multiple tests, the actual number of tested variants was estimated using the Genetic Type I error calculator, and the suggestive and Bonferroni significance thresholds were adjusted accordingly (Li *et al*., 2012). The significance threshold based on false discovery rate (FDR)–adjusted *P*-values was also calculated. This provides another conservative estimate that is more flexible than Bonferroni, as it takes into account the distribution of actual *P*-values. Regions around significant SNPs were annotated using a custom database generated using the QQ74 reference genome, using SnpEff 4.3 T in Galaxy(https://usegalaxy.org/) (Cingolani *et al*., 2012; The Galaxy Community, 2022). SNPs were categorized according to their position with respect to gene-coding sequences. The putative functions of genes within the regions were predicted using the online tools DAVID (https://david.ncifcrf.gov/tools.jsp) (Huang *et al*., 2009) and Plaza (https://bioinformatics.psb.ugent.be/plaza/versions/plaza_v4_5_dicots/) (Van Bel *et al*., 2022).

Afrosimetric assay to estimate saponin content in quinoa seeds

Dried quinoa seeds were used for the afrosimetric assay. A standard protocol was described previously (Koziol, 1991), in which 0.5 g seeds was added to a test tube containing 5 mL of distilled water, followed by shaking vigorously for 30 s. After a 30-min rest, the tube was shaken again for 30 min. This shaking-and-resting step was repeated once more before letting the tube rest for 5 min and then observing the height of the stable foam to the nearest 0.1 cm. A positive control tube containing 0.2 g saponins (Sigma, USA) in 5 mL of distilled water was included. This standard protocol was used for post-harvest estimation of bitterness or sweetness of quinoa seeds. According to Koziol (1991), quinoa seeds that cannot produce a stable foam layer in the afrosimetric assay have less than 0.11% (w/w) saponins and can be considered sweet seeds.

Alternatively, a modified afrosimetric assay was used as described previously (Otterbach *et al*., 2021). In this modified afrosimetric assay, five quinoa seeds were added to a microtube containing 0.5 mL of distilled water. This modified assay was used to score saponin content in quinoa seeds from 88 quinoa accessions used for GWAS. Scores were assigned by measuring the height in millimetres of foam formed for each sample after treatment. Measurements were carried out manually for each sample using ImageJ v1.53 (Schneider *et al*., 2012).

Total saponin extraction and metabolite analysis using LC-qToF-MS/MS

Quinoa plants were grown in soil under greenhouse conditions (long-day conditions (8-h-dark/16-h-light cycle), 200 μmol photons m^{-2} s^{-1}; University of Copenhagen, Denmark) until they fully matured and produced seeds; these dried seeds were used for saponin extraction and analysis. Quinoa plants were also grown in a hydroponic system using air-supplied water-based mineral nutrient solutions [containing approximately 0.2 mM KH_2PO_4, 0.2 mM K_2SO_4, 0.3 mM $MgSO_4.7H_2O$, 0.1 mM NaCl, 0.3 mM $Mg(NO_3)_2.6H_2O$, 0.9 mM $Ca(NO_3)_2.4H_2O$, 0.6 mM KNO_3, 50 μM Fe(III)-EDTA-Na, 1 μM $MnCl_2.4H_2O$, 0.7 μM $ZnSO_4$, 0.8 μM $CuSO_4.5H_2O$, 20 μM H_3BO_3 and 0.8 μM $Na_2MoO_4.2H_2O$; pH 5.5–6.5, adjusted using NaOH or HCl solutions] under greenhouse conditions (long-day conditions; 8-h-dark/16-h-light cycle; 200 μmol photons m^{-2} s^{-1}; University of Copenhagen, Denmark) from 07/10/2022 to 29/11/2022; total leaves and roots were harvested, freeze-dried, homogenized and used for saponin extraction and analysis.

Approximately 0.1 g dried quinoa seed, leaf or root sample was placed into a 2-ml microtube containing four 5-mm stainless steel beads (Qiagen, Germany). All samples were homogenized using a tissue lyser (Retsch®, Germany) to achieve fine powder. Total saponins in homogenized tissue samples were extracted with 1 mL of 80% (v/v) methanol solution containing a mix of 2 μM hederacoside C and 2 μM α-hederin (internal standards; Sigma, USA), assisted by ultrasound in an ultrasonic bath sonicator (Branson 5510, Branson Ultrasonics, USA) for 30 min, as previously described (Gómez-Caravaca *et al*., 2011). Tissue particles were pelleted using high-speed centrifugation (18.407 × ***g***, 15 min, 4 °C; Eppendorf centrifuge 5430/5430 R), and the liquid extracts were collected. The extracts were stored at −20 °C overnight for further precipitation of proteins and small particles, followed by high-speed centrifugation (18.407 × ***g***, 15 min, 4 °C). Extracts in supernatant fraction were then collected. Samples were randomly numbered, diluted with milli-Q water and filtered (mobile phase purification) for LC-qToF-MS/MS analysis.

To identify individual saponins, metabolite analysis by LC-qToF-MS/MS was performed essentially as described (Giavalisco *et al*., 2011; Jarvis *et al*., 2017; Tabatabaei *et al*., 2022). LC-qToF-MS/MS was performed on a Dionex UltiMate 3000 Quaternary Rapid Separation UHPLC+ focused system (Thermo Fisher Scientific, Germering, Germany). Separation was achieved on a Kinetex 1.7 μm XB-C18 column (100 × 2.1 mm, 1.7 μm, 100 Å, Phenomenex). For eluting 0.05% (v/v) formic acid in milli-Q grade water and acetonitrile [supplied with 0.05% (v/v) formic acid] were employed as mobile phases A and B, respectively. Gradient conditions were as follows: 0.0–1.0 min 5% B; 1.0–9.0 min 5%–20% B, 9.0–23.0 min 20%–75% B, 23.0–25.0 min 75%–100% B, 25.0–26.5 min 100% B, 26.5–27.0 min 100%–5% B and 27.0–32.0 min 5% B. The flow rate of the mobile phase was 300 μL/min. The column temperature was maintained at 30 °C. The UHPLC was coupled to a Compact micrOTOF-Q mass spectrometer (Bruker, Bremen, Germany) equipped with an electrospray ion source (ESI) operated in negative ion mode. The ion spray voltage was maintained at −3900 V in negative ion mode. Dry temperature was set to 250 °C, and the dry gas flow was set to 8 L/min. Nitrogen was used as the dry gas, nebulizing gas and collision gas. The nebulizing gas was set to 2.5 bar and collision energy to 10 eV. MS spectra were acquired in an *m/z* range from 50 to 1400 and MS/MS spectra in a range from 200 to 1400 *m/z*. Sampling rate was set to 3 Hz. Na-formate clusters were used for mass calibration. All files were calibrated by postprocessing.

Base peak chromatograms were normalized to the dried weight of input materials (shoots, roots and seeds) and are shown using the same scale in the figures (Figure 3c; Figure S12). Detected saponins are listed in Table S3 according to their retention time (12–20 min) and specific m/z values determined from previously published data (Jarvis *et al*., 2017; Tabatabaei *et al*., 2022). Specific peaks of the major seed saponins detected in the base peak chromatograms were coloured, as shown in Figure 3c,d. Concentrations of these major saponins were calculated by relative comparison of their peak areas to those of the internal standards for which the concentration was known. PCA of LC-qToF-MS/MS data of extracts from quinoa seeds, leaves and roots were plotted using MZmine 3 (Schmid *et al*., 2023).

RNA-seq analysis

Quinoa plants were grown on soil under greenhouse conditions (long-day conditions (8-h-dark/16-h-light cycle), 200 μmol photons m^{-2} s^{-1}; University of Copenhagen, Denmark) from 07/07/2022 to 12/09/2022. Young fruits were used to extract total RNA using SV Total RNA Isolation System (Promega, USA). The integrity of total RNA extracts (2.7–15.5 μg) was tested using the Agilent 2100 Bioanalyzer system (Agilent Technologies, USA) and agarose gel electrophoresis. High-qualified total RNA samples (total RNA amount ≥200 ng, RNA integrity number ≥4 (with flat base line), no degradation and no contamination) were sent to Novogene Europe (Novogene, China) for RNA-seq analysis.

Messenger RNA (mRNA) was purified from total RNA using poly-T oligo-attached magnetic beads after fragmentation, and the first-strand cDNA was synthesized using random hexamers, followed by the second-strand cDNA synthesis. The library was checked with Qubit (Invitrogen, USA) and real-time PCR for quantification and bioanalyzer for size distribution detection. Quantified libraries were pooled and sequenced on an Illumina platform (Illumina®, USA). Raw data were stored in FASTQ format files containing sequences of reads and corresponding base quality.

Raw data were processed through fastq software to obtain clean data (clean reads) after removing reads containing adapter, reads containing poly-N and low-quality reads. Phred score (error rate, Q20 and Q30) and GC content in the clean data were calculated. Only clean data with high quality were processed to next steps.

The quinoa reference genome assembly ASM168347v1 (GCF_001683475.1 (58 754 genes, 49 138 protein-coding genes)) was downloaded from GenBank (NCBI, https://www.ncbi.nlm.nih.gov/). An index of the reference genome was built, and paired-end clean reads were aligned to the reference genome using Hisat2 v2.05 (Kim *et al*., 2015, 2019; Mortazavi *et al*., 2008; Pertea *et al*., 2016).

FeatureCounts v1.5.0-p3 (Liao *et al*., 2014) was used to count the reads mapped to each gene. For estimating gene expression levels, fragments per kilobase of transcript sequence per millions base pairs sequenced (FPKM) of each gene was calculated based on the length of the gene and read counts mapped to this gene (Mortazavi *et al*., 2008).

Differential expression analysis of two groups (*sws1* versus WT or *sws2* versus WT) was performed using the DESeq2 R package v1.20.0 (Love *et al*., 2014). DESeq2 provides statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution (Love *et al*., 2014). The bigger the number of genes, the higher the cumulative degree of false positives in the hypothesis test. The resulting *P*-values were adjusted using the Benjamini and Hochberg (BH) approach for controlling the FDR. Genes with an adjusted *P*-value (P_{adj}) ≤ 0.05 found by DESeq were assigned as differently expressed. A standard screening criterium for a biologically meaningful threshold was applied with $|\log_2(\text{FoldChange})| \geq 1$ and $P_{adj} \leq 0.05$ to identify significantly regulated differential genes between two groups.

Proteomic analysis

Quinoa plants were grown on soil under greenhouse conditions (long-day conditions (8-h-dark/16-h-light cycle), 200 μmol photons m^{-2} s^{-1}; University of Copenhagen, Denmark) from 18/01/2023 to 04/04/2023. Developing fruits (approximately 100–300 mg) were used to extract total protein according to Wang et al. (Wang *et al*., 2006) Total protein pellets were sent to the Proteomics Research Infrastructure (PRI, University of Copenhagen, Denmark) for mass spectrometry (MS) analysis. Protein samples were processed as previously described (Wiśniewski *et al*., 2009). Protein pellets were dissolved, diluted in a digestion buffer containing 1 μg Trypsin/Lys-C Mix (Promega, USA) and 0.01% ProtemaseMax™ surfactant (Trypsin enhancer, Promega, USA), then incubated overnight at 37 °C. Tryptic peptide mixtures were desalted using stage-tips containing styrene divinylbenzene reverse-phase sulfonate material (SDB-RPS, 3 M, USA). Desalted peptide mixtures were separated on 15-cm columns packed with best-in-class ReproSil C18 1.9-μm beads (PepSep Columns, Bruker Daltonik GmbH, Germany) on a HPLC device (Evosep One HPLC, Ecosep Biosystems, Denmark), then injected into a timsTOF Pro mass spectrometer (Bruker Daltonik GmbH, Germany) that was operated in Parallel Accumulation Serial Fragmentation (PASEF) mode as previously described (Meier *et al*., 2018). Raw MS data, obtained by data-independent acquisition (DIA) and data-dependent acquisition (DDA) methods, were analysed with MaxQuant v1.6.15.0 (Max Planck Institute of Biochemistry, Germany) and Spectronaut® (BIOGNOSYS, Switzerland). Peak lists were searched against the UniProt FASTA database (UP000596660_63459.fasta (33 949 entries), May 2023). FDR was set to 5% for both peptides (minimum length of seven amino acids) and proteins. Label-free quantification (LFQ) was performed using Spectronaut® (BIOGNOSYS, Switzerland) using default manufacturer settings.

Statistical and bioinformatic analyses of MaxQuant v1.6.15.0 (Max Planck Institute of Biochemistry, Germany) and Spectronaut® (BIOGNOSYS, Switzerland) outputs were performed in Perseus v1.6.14.0 (Max Planck Institute of Biochemistry, Germany) (Tyanova *et al*., 2016) and R v4.0.3. LFQ intensity values were normalized by $\log_2$ transformation, and proteins with less than two valid values (two unique peptides) in at least one group were filtered out. The missing values were imputed with the MinProb method (random draws from a Gaussian distribution; width = 0.3 and downshift = 1.8) (Lazar *et al*., 2016). Differentially regulated proteins in each group comparison were identified by one-way ANOVA test with Benjamini-Hochberg (BH) correction from multiple hypotheses, followed by post hoc pairwise comparison t-tests using the same parameters and BH correction. In the volcano plots, proteins that displayed a significantly reduced or increased expression are highlighted in blue and red, respectively; dashed lines indicate thresholds ($-1 > \log_2$FoldChange (*sws* mutant/WT) > 1; $-\log_{10}$(*P*-value) for the protein with the highest significant post hoc adjusted *P*-value (P_{adj}) below 0.05).

Subcellular localization analysis using heterologous transient expression of *TSARL1* in *Nicotiana benthamiana*

TSARL1 coding sequences (CDSs) were specifically amplified from a cDNA library using the attB1-CqTSARL1-F/attB2-TSARL1-NoEnd_R primer pair (Table S7) and PrimeSTAR® GXL DNA Polymerase (Takara, Japan). The obtained *TSARL1* CDS fragments were cloned into pGWB505 destination empty vector (Addgene, USA) (Nakagawa *et al.*, 2009) via the Gateway cloning system using pDONR/Zeo (Invitrogen, USA) as donor vectors to generate an in-frame *TSARL1-eGFP* fusion construct. This fusion construct was transfected into *Agrobacterium* C58C1 strain (Gold Biotechnology, USA) via an electroporation method. Also, the pDGB3α2_35S:P19:Tnos vector (Addgene, USA) (Sarrion-Perdigones *et al.*, 2013) was transfected into *Agrobacterium* C58C1 strain (Gold Biotechnology, USA) via an electroporation method. This transfected *Agrobacterium* strain, containing recombinant pGWB505 (having *TSARL1-eGFP* fusion construct) or pDGB3α2_35S:P19:Tnos vectors (Addgene, USA), was co-infiltrated to *N. benthamiana* to induce transient expression of *TSARL1-eGFP* in leaf mesophyll and epidermal cells. As a control, a transfected *Agrobacterium* GV3101 strain, containing pGFPGUSPlus vector (Addgene, USA) (Vickers *et al.*, 2007) expressing eGFP and GUS, and a transfected *Agrobacterium* C58C1 strain, containing pDGB3α2_35S:P19:Tnos vector expressing P19, were also co-infiltrated into *N. benthamiana* leaves. eGFP fluorescence signals from mesophyll and epidermal cells of infiltrated *N. benthamiana* plants were observed using a confocal laser scanning microscope (Leica SP5-X system, Leica, Germany) with a water 20x objective (NA 0.7) at an excitation wavelength of 488 nm. Emission signals were recorded at 500–525 nm (GFP) and 611–680 nm (autofluorescence).

Tissue-specific expression analysis using RT-PCR

Total RNA from quinoa roots, stems, leaves, flowers, fruits, fruit coats (perianths) and young seeds, consisting of seed coat (pericarp), curved embryo and perisperm, was extracted using an RNeasy Kit (Qiagen, Germany). Fruit coats and young seeds were manually dissected from young fruits using forceps and tweezers under the microscope (Leica EZ4, Leica, Germany). cDNA libraries were synthesized using an iScript cDNA Synthesis Kit (Bio-Rad, USA). RT-PCR reactions to amplify full lengths (1183 bp) or fragments (573 bp) of *TSARL1* CDS were performed using Pl_TSARL1_F1/ Pl_TSARL1_R1 or CqTSARL1_CDS_F/CqTSARL1_CDS_R primer pairs (Table S7), respectively, and PrimeSTAR® GXL DNA Polymerase (Takara, Japan). As a positive control, the CDS of the AUR62027932 gene, a housekeeping gene encoding a putative *elongation factor 1α* in quinoa, was amplified using Fp1_Cqef1a/Rp1_Cqef1a primer pairs and PrimeSTAR® GXL DNA Polymerase (Takara, Japan).

Phenotypic analysis of *tsarl1* mutants

Quinoa plants were grown on soil under greenhouse conditions (short-day (16-h-dark/8-h-light cycle) or long-day (8-h-dark/16-h-light cycle) conditions, 200 μmol photons m^{-2} s^{-1}). Plants were irrigated with nutrient water once every week in the first 2 weeks after sowing seeds or until anthesis. Plant development was observed at 5, 12, 19, 25 and 40 days after sowing (DAS; seedling). Shoot heights were measured at 25 and 40 DAS. Shoot weights (fresh and dried weight) were measured at 48 DAS. Flower morphology was observed using a Leica M205FA stereomicroscope (Leica, Germany). Seeds were purified and weighed from individual plants, and seed sizes were measured using ImageJ (https://imagej.net/).

Non-choice test for herbivore susceptibility

Larvae of *Spodoptera exigua* were allowed to feed on young plants of the bitter quinoa cv. Titicaca WT and *sws* mutants. After 8 days, the larvae were collected and their weight was scored.

Field trial experiments

Testing BCF_2 quinoa lines (8 g seed per line) were planted in separate rows in the field at research station Flakkebjerg, University of Aarhus, Zealand, Denmark, during the 2023 field season. Seeds were sown in 1- to 2-cm depths in a prepared seed bed just after frost in Denmark (the first month of spring, April) with a soil temperature above 0 °C. The space between two seed rows was kept around 25–50 cm so that hoeing could be used for weed control. Manure was used to provide nitrogen (N) to quinoa plants with a relative amount corresponding to 80–120 kg N ha^{-1}. Disease and pest controls were applied, if necessary, especially when the humidity was high (~85% RH) with temperatures of 15–20 °C. Early harvest of quinoa occurred yearly in September and was performed with a combine harvester. In each testing quinoa line, we randomly selected three sectors when 25 plants were sampled per sector. Total seed weight per sector was measured, and the average seed weight produced per sector was calculated.

Acknowledgements

We thank the Proteomics Research Infrastructure (PRI) at University of Copenhagen for technical assistance in proteomic analysis. This work was supported by the Carlsberg Foundation (RaisingQuinoa CF18-1113 M.P.) and the Novo Nordisk Foundation (NovoCrops NNF19OC0056580 to M.P. and R.L.L.-M. and EcoSap NNF20OC0060298 to S.B. and J.G.). This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 801199 to DV.

Author contributions

M.P. conceived the study. M.D.L.T., D.V., J.G. and J.T.Ø. designed the experiments. M.D.L.T., D.V., J.G., J.T.Ø., M.W.M., G.L., A.F.N., C.C., P.V.N., S.E.-J. performed the experiments. M.D.L.T., R.R.F., S.F., T.W., S.B., R.L.L.-M. and M.P. supervised the project. M.D.L.T. and M.P. wrote the first draft of the manuscript.

Competing interests

T.W. and J.T.Ø are inventors on a patent application related to this work filed by Carlsberg A/S no. WO2021/069614, filed on 8 October 2020, published on 15 April 2021; T.W. is inventor on patent applications related to this work filed by Carlsberg A/S (no. WO2018/001884, filed on 23 June 2017, published on 4 January 2018; no. WO2019/129736, filed on 21 December 2018, published on 4 July 2019; no. WO2019/129739, filed on 21 December 2018, published on 4 July 2019). The authors declare no other competing interests. J.T.Ø. and T.W. are Traitomic A/S employees. The methodology presented in the manuscript is applied commercially by Traitomic A/S (www.traitomic.com).

Data availability statement

The data that support the findings of this study are openly available on request.

References

Adolf, V.I., Jacobsen, S.E. and Shabala, S. (2013) Salt tolerance mechanisms in quinoa (*Chenopodium quinoa* Willd.). *Environ. Exp. Bot.* **92**, 43–54.

Augustin, J.M., Kuzina, V., Andersen, S.B. and Bak, S. (2011) Molecular activities, biosynthesis and evolution of triterpenoid saponins. *Phytochemistry* **72**, 435–457.

Augustin, J.M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J.K., Khakimov, B., Olsen, C.E. *et al*. (2012) UDP-glycosyltransferases from the UGT73C subfamily in Barbarea vulgaris catalyze sapogenin 3-O-glucosylation in saponin-mediated insect resistance. *Plant Physiol.* **160**, 1881–1895.

Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y. and Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Browning, B.L., Zhou, Y. and Browning, S.R. (2018) A one-penny imputed genome from next-generation reference panels. *Am. J. Hum. Genet.* **103**, 338–348.

Browning, B.L., Tian, X., Zhou, Y. and Browning, S.R. (2021) Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.* **108**, 1880–1890.

Chen, S., Zhou, Y., Chen, Y. and Gu, J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890.

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J. *et al*. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E. *et al*. (2011) The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A. *et al*. (2021) Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008.

De Bock, P., Van Bockstaele, F., Muylle, H., Quataert, P., Vermeir, P., Eeckhout, M. and Cnops, G. (2021) Yield and nutritional characterization of thirteen quinoa (*Chenopodium quinoa* willd.) varieties grown in north-west europe—part I. *Plan. Theory* **10**, 2689.

Depristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C. *et al*. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.

El Hazzam, K., Hafsa, J., Sobeh, M., Mhada, M., Taourirte, M., Kacimi, K.E.L. and Yasri, A. (2020) An insight into saponins from Quinoa (*Chenopodium quinoa* Willd): A review. *Molecules* **25**, 1–22.

FAO (2021) *Global map of salt-affected soils: GSASmap V1.0*. Rome: FAO. 20.

Fiallos-Jurado, J., Pollier, J., Moses, T., Arendt, P., Barriga-Medina, N., Morillo, E., Arahana, V. *et al*. (2016) Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in *Chenopodium quinoa* leaves. *Plant Sci.* **250**, 188–197.

Giavalisco, P., Li, Y., Matthes, A., Eckhardt, A., Hubberten, H.M., Hesse, H., Segu, S. *et al*. (2011) Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* **68**, 364–376.

Gómez-Caravaca, A.M., Segura-Carretero, A., Fernández-Gutiérrez, A. and Caboni, M.F. (2011) Simultaneous determination of phenolic compounds and saponins in quinoa (*Chenopodium quinoa* Willd) by a liquid chromatography-diode array detection-electrospray ionization-time-of-flight mass spectrometry methodology. *J. Agric. Food Chem.* **59**, 10815–10825.

Granado-Rodríguez, S., Aparicio, N., Matías, J., Pérez-Romero, L.F., Maestro, I., Gracés, I., Pedroche, J.J. *et al*. (2021) Studying the impact of different field environmental conditions on seed quality of Quinoa: The case of three different years changing seed nutritional traits in Southern Europe. *Front. Plant Sci.* **12**, 1–21.

Hariadi, Y., Marandon, K., Tian, Y., Jacobsen, S.E. and Shabala, S. (2011) Ionic and osmotic relations in quinoa (*Chenopodium quinoa* Willd.) plants grown at various salinity levels. *J. Exp. Bot.* **62**, 185–193.

Hinojosa, L., González, J.A., Barrios-Masias, F.H., Fuentes, F. and Murphy, K.M. (2018) Quinoa abiotic stress responses: A review. *Plan. Theory* **7**, 106.

Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.

Imamura, T., Yasui, Y., Koga, H., Takagi, H., Abe, A., Nishizawa, K. *et al*. (2020) A novel WD40-repeat protein involved in formation of epidermal bladder cells in the halophyte quinoa. *Commun. Biol.* **3**, 1–14.

Jacobsen, S.E. (2017) The scope for adaptation of quinoa in Northern Latitudes of Europe. *J. Agron. Crop. Sci.* **203**, 603–613.

Jacobsen, S.E., Hill, J. and Stølen, O. (1996) Stability of quantitative traits in quinoa (*Chenopodium quinoa*). *Theor. Appl. Genet.* **93**, 110–116.

Jacobsen, S.E., Monteros, C., Christiansen, J.L., Bravo, L.A., Corcuera, L.J. and Mujica, A. (2005) Plant responses of quinoa (*Chenopodium quinoa* Willd.) to frost at various phenological stages. *Eur. J. Agron.* **22**, 131–139.

Jacobsen, S.E., Monteros, C., Corcuera, L.J., Bravo, L.A., Christiansen, J.L. and Mujica, A. (2007) Frost resistance mechanisms in quinoa (*Chenopodium quinoa* Willd.). *Eur. J. Agron.* **26**, 471–475.

Jägermeyr, J., Müller, C., Ruane, A.C., Elliott, J., Balkovic, J., Castillo, O., Faye, B. *et al*. (2021) Climate impacts on global agriculture emerge earlier in new generation of climate and crop models. *Nat. Food* **2**, 873–885.

Jarvis, D.E., Ho, Y.S., Lightfoot, D.J., Schmöckel, S.M., Li, B., Borm, T.J.A., Ohyanagi, H. *et al*. (2017) The genome of *Chenopodium quinoa*. *Nature* **542**, 307–312.

Joshi, R.C., San Martín, R., Saez-Navarrete, C., Alarcon, J., Sainz, J., Antolin, M.M., Martin, A.R. *et al*. (2008) Efficacy of quinoa (*Chenopodium quinoa*) saponins against golden apple snail (Pomacea canaliculata) in The Philippines under laboratory conditions. *Crop Prot.* **27**, 553–557.

Kim, D., Langmead, B. and Salzberg, S.L. (2015) HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12**, 357–360.

Kim, D., Paggi, J.M., Park, C., Bennett, C. and Salzberg, S.L. (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915.

Knudsen, S., Wendt, T., Dockter, C., Thomsen, H.C., Rasmussen, M., Egevang Jørgensen, M. *et al*. (2022) FIND-IT: Accelerated trait development for a green evolution. *Sci. Adv.* **8**, eabq2266.

Komari, T. (1990) Transformation of cultured cells of *Chenopodium quinoa* by binary vectors that carry a fragment of DNA from the virulence region of pTiBo542. *Plant Cell Rep.* **9**, 303–306.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R. (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356.

Koziol, M.J. (1991) Afrosimetric estimation of threshold saponin concentration for bitterness in quinoa (*Chenopodium quinoa* Willd). *J. Sci. Food Agric.* **54**, 211–219.

Kuljanabhagavad, T. and Wink, M. (2009) Biological activities and chemistry of saponins from *Chenopodium quinoa* Willd. *Phytochem Rev.* **8**, 473–490.

Kuljanabhagavad, T., Thongphasuk, P., Chamulitrat, W. and Wink, M. (2008) Triterpene saponins from *Chenopodium quinoa* Willd. *Phytochemistry* **69**, 1919–1926.

Lazar, C., Gatto, L., Ferro, M., Bruley, C. and Burger, T. (2016) Accounting for the multiple natures of missing values in label-free quantitative proteomics data sets to compare imputation strategies. *J. Proteome Res.* **15**, 1116–1125.

Lefouili, M. and Nam, K. (2022) The evaluation of Bcftools mpileup and GATK HaplotypeCaller for variant calling in non-human species. *Sci. Rep.* **12**, 11331–11338.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100.

Li, H. (2021) New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574.

Li, H. and Barrett, J. (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993.

Li, B. and Lightfoot, D.J. (2021) A chromosome-scale quinoa reference genome assembly. In *The Quinoa Genome. Compendium of Plant Genomes* (Schmöckel, S.M., ed), pp. 65–80. Cham: Springer.

Li, M.X., Yeung, J.M.Y., Cherny, S.S. and Sham, P.C. (2012) Evaluating the effective numbers of independent tests and significant p-value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum. Genet.* **131**, 747–756.

Liao, Y., Smyth, G.K. and Shi, W. (2014) FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930.

Lim, T.K. (2013) *Edible Medicinal and Non-Medicinal Plants*. Dordrecht: Springer Netherlands.

Lim, J.G., Park, H.M. and Yoon, K.S. (2020) Analysis of saponin composition and comparison of the antioxidant activity of various parts of the quinoa plant (*Chenopodium quinoa* Willd). *Food Sci. Nutr.* **8**, 694–702.

Lin, P.H. and Chao, Y.Y. (2021) Different drought-tolerant mechanisms in quinoa (*Chenopodium quinoa* willd) and djulis (*Chenopodium formosanum* koidz.) based on physiological analysis. *Plan. Theory* **10**, 2279.

López-Marqués, R.L., Nørrevang, A.F., Ache, P., Moog, M., Visintainer, D., Wendt, T., Østerberg, J.T. *et al*. (2020) Prospects for the accelerated improvement of the resilient crop quinoa. *J. Exp. Bot.* **71**, 5333–5347.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.

Luo, G., Najafi, J., Correia, P.M.P., Trinh, M.D.L., Chapman, E.A., Østerberg, J.T., Thomsen, H.C. *et al*. (2022) Accelerated domestication of new crops: Yield is key. *Plant Cell Physiol.* **63**, 1624–1640.

Madl, T., Sterk, H., Mittelbach, M. and Rechberger, G.N. (2006) Tandem mass spectrometric analysis of a complex triterpene saponin mixture of *Chenopodium quinoa*. *J. Am. Soc. Mass Spectrom.* **17**, 795–806.

Maldonado-Taipe, N., Barbier, F., Schmid, K., Jung, C. and Emrani, N. (2022) High-density mapping of quantitative trait loci controlling agronomically important traits in quinoa (*Chenopodium quinoa* Willd). *Front. Plant Sci.* **13**, 1–18.

Mastebroek, H.D., Limburg, H., Gilles, T. and Marvin, H.J.P. (2000) Occurrence of sapogenins in leaves and seeds of quinoa (*Chenopodium quinoa* Willd). *J. Sci. Food Agric.* **80**, 152–156.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K. *et al*. (2010) The Genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303.

Meier, F., Brunner, A.D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N. *et al*. (2018) Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545.

Meisner, J. and Albrechtsen, A. (2018) Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* **210**, 719–731.

Moog, M.W., Trinh, M.D.L., Nørrevang, A.F., Bendtsen, A.K., Wang, C., Østerberg, J.T., Shabala, S. *et al*. (2022) The epidermal bladder cell-free mutant of the salt-tolerant quinoa challenges our understanding of halophyte crop salinity tolerance. *New Phytol.* **236**, 1409–1421.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L. and Wold, B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* **5**, 621–628.

Mukhopadhyay, R., Sarkar, B., Jat, H.S., Sharma, P.C. and Bolan, N.S. (2021) Soil salinity under climate change: Challenges for sustainable agriculture and food security. *J. Environ. Manage.* **280**, 111736.

Nakagawa, T., Ishiguro, S. and Kimura, T. (2009) Gateway vectors for plant transformation. *Plant Biotechnol.* **26**, 275–284.

Negacz, K., Vellinga, P., Barrett-Lennard, E.G., Choukr-Allah, R. and Elzenga, T. (2022) *Future of sustainable agriculture in saline environments*. Boca Raton, FL: CRC Press.

Osbourn, A. (1996) Saponins and plant defence — a soap story. *Trends Plant Sci.* **1**, 4–9.

Otterbach, S., Wellman, G. and Schmöckel, S.M. (2021) *Saponins of Quinoa: Structure, Function and Opportunities*, pp. 119–138. Berlin: Springer International Publishing.

Patiranage, D.S.R., Asare, E., Maldonado-Taipe, N., Rey, E., Emrani, N., Tester, M. and Jung, C. (2021) Haplotype variations of major flowering time genes in quinoa unveil their role in the adaptation to different environmental conditions. *Plant Cell Environ.* **44**, 2565–2579.

Patiranage, D.S., Rey, E., Emrani, N., Wellman, G., Schmid, K., Schmöckel, S.M., Tester, M. *et al*. (2022) Genome-wide association study in quinoa reveals selection pattern typical for crops with a short breeding history. *Elife* **11**, e66873.

Pearsall, D.M. (2008) Plant domestication and the shift to agriculture in the Andes. In *Handbook of South American Archeology*(Silverman, H. and Isbell, W.H., eds). New York, NY: Springer New York.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T. and Salzberg, S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667.

Ruiz, K.B., Biondi, S., Oses, R., Acuña-Rodríguez, I.S., Antognoni, F., Martinez-Mosqueira, E.A., Coulibaly, A. *et al*. (2014) Quinoa biodiversity and sustainability for food security under climate change. A review. *Agron. Sustain Dev.* **34**, 349–359.

Ruiz, K.B., Khakimov, B., Engelsen, S.B., Bak, S., Biondi, S. and Jacobsen, S.E. (2017) Quinoa seed coats as an expanding and sustainable source of bioactive compounds: An investigation of genotypic diversity in saponin profiles. *Ind. Crop Prod.* **104**, 156–163.

San Martín, R., Ndjoko, K. and Hostettmann, K. (2008) Novel molluscicide against Pomacea canaliculata based on quinoa (*Chenopodium quinoa*) saponins. *Crop Prot.* **27**, 310–319.

Sarrion-Perdigones, A., Vazquez-Vilar, M., Palací, J., Castelijns, B., Forment, J., Ziarsolo, P. *et al*. (2013) Goldenbraid 2.0: A comprehensive DNA assembly framework for plant synthetic biology. *Plant Physiol.* **162**, 1618–1631.

Satheesh, N. and Fanta, S.W. (2018) Review on structural, nutritional and anti-nutritional composition of Teff (*Eragrostis tef*) in comparison with Quinoa (*Chenopodium quinoa* Willd). *Cogent. Food Agric.* **4**, 1546942.

Schmid, R., Heuckeroth, S., Korf, A., Smirnov, A., Myers, O., Dyrlund, T.S., Bushuiev, R. *et al*. (2023) Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **41**, 447–449.

Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **97**, 671–675.

Suárez-Estrella, D., Torri, L., Pagani, M.A. and Marti, A. (2018) Quinoa bitterness: causes and solutions for improving product acceptability. *J. Sci. Food Agric.* **98**, 4033–4041.

Suzuki, M., Kamide, Y., Nagata, N., Seki, H., Ohyama, K., Kato, H., Masuda, K. *et al*. (2004) Loss of function of 3-hydroxy-3-methylglutaryl coenzyme A reductase 1 (HMG1) in Arabidopsis leads to dwarfing, early senescence and male sterility, and reduced sterol levels. *Plant J.* **37**, 750–761.

Suzuki, H., Seki, H. and Muranaka, T. (2021) Insights into the diversification of subclade IVa bHLH transcription factors in Fabaceae. *BMC Plant Biol.* **21**, 1–10.

Taaime, N., Rafik, S., El Mejahed, K., Oukarroum, A., Choukr-Allah, R., Bouabid, R. and El Gharous, M. (2023) Worldwide development of agronomic management practices for quinoa cultivation: a systematic review. *Front Agron.* **5**, 1215441.

Tabatabaei, I., Alseekh, S., Shahid, M., Leniak, E., Wagner, M., Mahmoudi, H., Thushar, S. *et al*. (2022) The diversity of quinoa morphological traits and seed metabolic composition. *Sci. Data* **9**, 323.

Tang, Y., Li, Z., Wang, C., Liu, Y., Yu, H., Wang, A. and Zhou, Y. (2020) LDkit: a parallel computing toolkit for linkage disequilibrium analysis. *BMC Bioinformatics* **21**, 461.

The Galaxy Community (2022) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. *Nucleic Acids Res.* **50**, W345–W351.

Thiam, E., Allaoui, A. and Benlhabib, O. (2021) Quinoa productivity and stability evaluation through varietal and environmental interaction. *Plan. Theory* **10**, 714.

Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M.Y., Geiger, T., Mann, M. *et al*. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740.

Van Bel, M., Silvestri, F., Weitz, E.M., Kreft, L., Botzki, A., Coppens, F. and Vandepoele, K. (2022) PLAZA 5.0: extending the scope and power of comparative and functional genomics in plants. *Nucleic Acids Res.* **50**, D1468–D1474.

Vickers, C.E., Schenk, P.M., Li, D., Mullineaux, P.M. and Gresshoff, P.M. (2007) pGFPGUSPlus, a new binary vector for gene expression studies and optimising transformation systems in plants. *Biotechnol. Lett.* **29**, 1793–1796.

Vincken, J.P., Heng, L., de Groot, A. and Gruppen, H. (2007) Saponins, classification and occurrence in the plant kingdom. *Phytochemistry* **68**, 275–297.

Wang, J. and Zhang, Z. (2021) GAPIT version 3: Boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics* **19**, 629–640.

Wang, W., Vignani, R., Scali, M. and Cresti, M. (2006) A universal and rapid protocol for protein extraction from recalcitrant plant tissues for proteomic analysis. *Electrophoresis* **27**, 2782–2786.

Wink, M. (2004) Phytochemical diversity of secondary metabolites. In *Encyclopedia of Plant & Crop Science*, pp. 915–919. New York: Routledge.

Wiśniewski, J.R., Zougman, A., Nagaraj, N. and Mann, M. (2009) Universal sample preparation method for proteome analysis. *Nat. Methods* **6**, 359–362.

Woldemichael, G.M. and Wink, M. (2001) Identification and biological activities of triterpenoid saponins from *Chenopodium quinoa*. *J. Agric. Food Chem.* **49**, 2327–2332.

Xiao, X., Meng, F., Satheesh, V., Xi, Y. and Lei, M. (2022) An Agrobacterium-mediated transient expression method contributes to functional analysis of a transcription factor and potential application of gene editing in *Chenopodium quinoa*. *Plant Cell Rep.* **41**, 1975–1985.

Yasui, Y., Hirakawa, H., Oikawa, T., Toyoshima, M., Matsuzaki, C., Ueno, M., Mizuno, N. *et al*. (2016) Draft genome sequence of an inbred line of *Chenopodium quinoa*, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. *DNA Res.* **23**, 535–546.

Zhou, X. (2017) A unified framework for variance component estimation with summary statistics in genome-wide association studies. *Ann. Appl. Stat.* **11**, 2027–2051.

Zou, C., Chen, A., Xiao, L., Muller, H.M., Ache, P., Haberer, G., Zhang, M. *et al*. (2017) A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. *Cell Res.* **27**, 1327–1340.

Zurita-Silva, A., Fuentes, F., Zamora, P., Jacobsen, S.E. and Schwember, A.R. (2014) Breeding quinoa (*Chenopodium quinoa* Willd.): Potential and perspectives. *Mol. Breed* **34**, 13–30.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Mapping of SNPs associated with seed saponin content.
Figure S2 TSARL1 homologs in quinoa.
Figure S3 Amino acid alignment of TSARL1 homologs.
Figure S4 Characterization of SNPs associated with seed saponin content.
Figure S5 Comparison between quinoa cv. Titicaca and other sweet cultivars during later stages of plant development.
Figure S6 Generation of a quinoa SNP library using EMS mutagenesis.
Figure S7 Site-directed genotype screening procedure in quinoa.
Figure S8 Isolation of the two independent tsarl1 mutants (sws1 and sws2) and genomic structure of TSARL1 and its alleles.
Figure S9 Amino acid sequence alignment of TSARL1, TSARL1-1 and TSARL1-3 encoded by TSARL1 and its alleles (tsarl1-1 and tsarl1-3), respectively.
Figure S10 Predicted tertiary structures of TSARL1, TSARL1-1 and TSARL1-3.
Figure S11 Standard afrosimetric assay for quality of saponin content in WT, sws1 and sws2 seeds.
Figure S12 Metabolite profiling of methanol extracts from quinoa leaves and roots.
Figure S13 Differential gene expression in fruit tissues between sws1 or sws2 mutants and WT plants.
Figure S14 Differential protein abundance in fruit tissues between sws1 or sws2 mutants and WT plants.
Figure S15 Tissue-specific expression and subcellular localization of TSARL1 and its protein.
Figure S16 Phenotypic analysis of the WT and sws mutants.
Table S1 List and properties of mutants used in this study.
Table S2 Summary of the effect of the identified SNPs within the identified genomic region of chromosome 16 on saponin content.
Table S3 List of detected saponins via reverse-phase LC-qToF-MS/MS.
Table S4 Differential expression of genes encoding metabolic enzymes involved in saponin biosynthesis.
Table S5 Regulated proteins in sws mutants.
Table S6 List of quinoa accessions used in this study.
Table S7 List of primers used in this study.