## scientific data



### **DATA DESCRIPTOR**

# **OPEN** The diversity of quinoa morphological traits and seed metabolic composition

Iman Tabatabaei<sup>1,2</sup>, Saleh Alseekh 1,3, Mohammad Shahid<sup>4</sup>, Ewa Leniak<sup>1</sup>, Mateusz Wagner 5, Henda Mahmoudi<sup>4</sup>, Sumitha Thushar<sup>4</sup>, Alisdair R. Fernie<sup>1,3</sup>, Kevin M. Murphy<sup>6</sup>, Sandra M. Schmöckel<sup>7</sup>, Mark Tester (10<sup>8</sup>, Bernd Mueller-Roeber (10<sup>1,2,3</sup>, Aleksandra Skirycz (10<sup>1,5 ⋈</sup> & Salma Balazadeh<sup>1,9</sup>

Quinoa (Chenopodium quinoa Willd.) is an herbaceous annual crop of the amaranth family (Amaranthaceae). It is increasingly cultivated for its nutritious grains, which are rich in protein and essential amino acids, lipids, and minerals. Quinoa exhibits a high tolerance towards various abiotic stresses including drought and salinity, which supports its agricultural cultivation under climate change conditions. The use of quinoa grains is compromised by anti-nutritional saponins, a terpenoid class of secondary metabolites deposited in the seed coat; their removal before consumption requires extensive washing, an economically and environmentally unfavorable process; or their accumulation can be reduced through breeding. In this study, we analyzed the seed metabolomes, including amino acids, fatty acids, and saponins, from 471 quinoa cultivars, including two related species, by liquid chromatography - mass spectrometry. Additionally, we determined a large number of agronomic traits including biomass, flowering time, and seed yield. The results revealed considerable diversity between genotypes and provide a knowledge base for future breeding or genome editing of quinoa.

#### **Background & Summary**

Quinoa (Chenopodium quinoa Willd.) is increasingly attracting global attention because of its unusually high grain nutritional value including high protein content, the composition and quantity of lipids, a good balance of essential amino acids, as well as isoflavones and interesting antioxidant functional properties<sup>1-3</sup>. Quinoa was first domesticated in the Lake Titicaca basin about 7,000 years ago, from where it spread to other regions in South America and the world4.

An agriculturally important asset of quinoa is its remarkable ability to adapt to diverse agroecological zones, which allows growth in hot dry deserts and in tropical areas with up to 88% relative humidity, from -8 °C to  $40\,^{\circ}\text{C}^{5}$ , and from sea level to  $4,000\,\text{m}$  high mountainous regions. Its adaptability to sodic and alkaline soils is also remarkable allowing cultivation from pH 4.5 to 9.06. Quinoa is a highly drought tolerant crop that fares well in regions of below 200 mm yearly rainfall (7 and references therein). It is tolerant against high salinity and considered a facultative halophyte<sup>8-10</sup>. In 2013, the Food and Agricultural Organization (FAO) declared the International Year of Quinoa' in recognition of the capacity of the crop to help mitigate hunger and malnutrition in food-insecure countries, and in recognition of the ancestral efforts of the Andean people to preserve quinoa as a crop (http://www.fao.org/quinoa-2013/en/).

<sup>1</sup>Max Planck Institute of Molecular Plant Physiology (MPI-MP), Am Muehlenberg 1, 14476, Potsdam, Germany. <sup>2</sup>University of Potsdam, Department Molecular Biology, Karl-Liebknecht-Strasse 24-25, Haus 20, 14476, Potsdam, Germany. 3Center of Plant Systems Biology and Biotechnology (CPSBB), 139 Ruski Blvd., 4000, Plovdiv, Bulgaria. International Center for Biosaline Agriculture (ICBA), Academic City, Near Zayed University, Dubai, United Arab Emirates. <sup>5</sup>Boyce Thompson Institute, 533 Tower Rd., Ithaca, NY, 14853, USA. <sup>6</sup>Washington State University, Sustainable Seed Systems Lab, 273 Johnson Hall, PO Box 646420, Pullman, WA, 99164-6420, USA. <sup>7</sup>Department Physiology of Yield Stability, Institute of Crop Science, University of Hohenheim, Fruwirthstr. 21, 70599, Stuttgart, Germany. 8King Abdullah University of Science and Technology (KAUST), Division of Biological and Environmental Sciences and Engineering, Thuwal, Saudi Arabia. 9Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE, Leiden, The Netherlands. <sup>™</sup>e-mail: as4258@cornell.edu; s.balazadeh@biology.leidenuniv.nl

Trait	Variation/range	
Panicle color (at physiological maturity)	White, Purple, Red, Pink, Yellow, Orange, Brown, Grey, Black,	
	Red and White, Red and Pink, Red and Yellow, Green, Red and	
	Green	
Panicle shape	<sup>1</sup> Glomerulate, <sup>2</sup> Intermediate, <sup>3</sup> Amarantiform	
Panicle density	Lax, intermediate, compact	
Lodging	Horizontal, semi-erect, compact	
Days to flowering*	Min. 50 days, max. 70 days	
Plant height (cm)**	Min. 17.2 cm - max. 145.25 cm	
Number of primary branches/plant***	Min. 1.5 - max. 22.5	
Number of panicles/plant (at physiological	Min. 1.5 - max. 22.5	
maturity)		
Panicle length (cm)	Min. 7.05 cm - max. 71.75 cm	
Plant dry weight (g)	Min. 6.6 g - max. 365.5 g	
Seed weight (g)/plant	Min. 1.15 g - max. 77.9 g	





Fig. 1 Morphological characteristics, and phenological and reproductive traits in quinoa genotypes. (A) Four qualitative and seven quantitative traits were measured and analyzed. ¹Glomerulate: glomerules are inserted in the primary axis showing a globose shape; ²Intermediate: showing both shapes; ³Amarantiform: glomerules are inserted directly in the secondary axis and have an elongated shape. Asterisks indicate: \* number of days from sowing until 50% of plants have started flowering; \*\* recorded at physiological maturity, from root collar to panicle apex; \*\*\* number of branches counted from the base to the second third of the stem from plants at physiological maturity. (B) Representative genotypes with diversity in seeds. 1. CHEN 274, 2. CHEN 210, 3. D-12038, 4. CHEN 109, 5. D-9998, 6. AMES 13219, 7. D-12020, 8. AMES 19046, 9. PI-433378, 10. D-12377, 11. Co-Ka-1821, 12. D-11999, 13. CHEN 70, 14. D-12140, 15. CHEN 243, 16. D-11980, 17. D-12275, 18. PI-433231. (C) Representative genotypes with diversity in panicle shape/color; 1. Ames 13727, 2. Ames 13749, 3. PI 634924, 4. Ames 13742, 5. Ames 13761, 6. Ames 22157, 7. PI 665276, 8. Ames 13731.

Although quinoa grains have an exceptional nutritional value, the seed coat typically contains bitter-tasting and potentially anti-nutritional saponins<sup>11</sup>. Therefore, quinoa seeds require substantial processing (water-extensive washing) to remove saponins before consumption. Reduction of saponins has been a breeding target and in the future may also be achieved with biotechnological methods, such as genome editing. The quinoa saponins occur predominantly in the form of triterpenoid glycosides<sup>12–14</sup>. Their large structural diversity renders analyses non-trivial<sup>15</sup>.

The biological functions of saponins in quinoa remain to be investigated. Saponins may play a role in seed germination, and in deterring birds or fungal infections (reviewed in <sup>16</sup>). Evidence indicates that not only the total amount of saponins is regulated (e.g., by the bHLH transcription factor CqTSARL1)<sup>17</sup>, but also the saponin profile <sup>17</sup>. However, to date, seed saponin profiles of only few quinoa genotypes have been determined <sup>17</sup>. As some saponins may even be beneficial to human health <sup>18</sup>, the diversity in saponin composition poses a great resource for breeding new and more healthy quinoa cultivars.

Our study reports the variability of the metabolome of mature quinoa seeds of a large number of genotypes (471 in total; Supplementary Dataset 1, available at figshare<sup>19</sup>). Additionally, we determined agronomic traits such as plant height, total biomass, panicle density, days to flowering, and seed weight (Fig. 1 and Supplementary Table 1<sup>19</sup>). The experimental pipeline employed for liquid chromatography – mass spectrometry (LC-MS)-based metabolome analysis of seeds is represented in Fig. 2. Metabolites were annotated using a library of authentic reference compounds, and in-source fragmentation patterns, and the data are reported in compliance with established standards<sup>20</sup> (Supplementary Table 2<sup>19</sup> and MetaboLights database, MTBLS2382). We detected and quantified 400 seed metabolites representing diverse chemical classes: 37 triterpenoid saponins, 14 flavonoids,

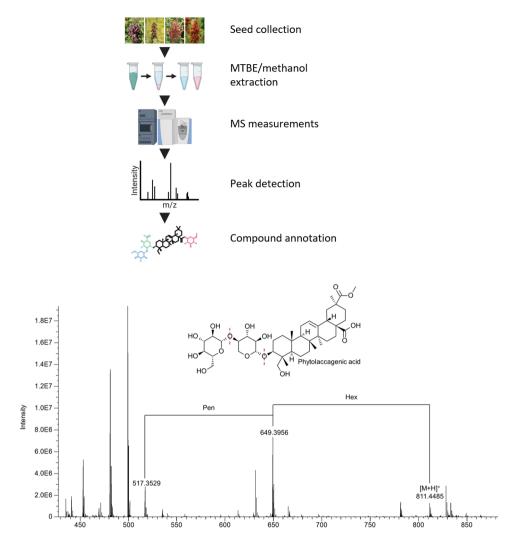
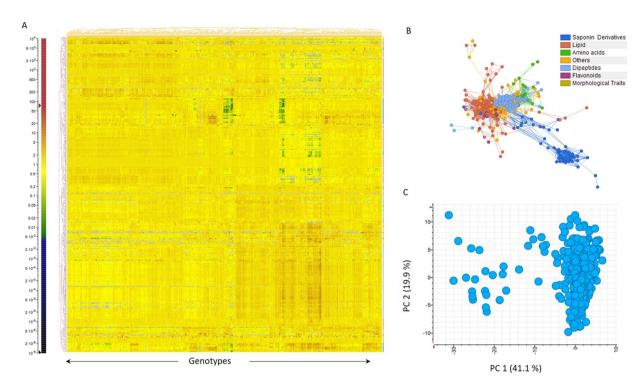


Fig. 2 Schematic representation of the experimental pipeline. Seed samples were extracted using the MTBE/methanol method, followed by LC-MS analysis of the lipid and polar fractions, peak detection using GeneData software, and compound annotations using in-house reference libraries, and in-source fragmentation. The bottom panel displays an example of saponin annotation using in-source fragmentation in the positive mode. Given is a chromatogram and putative structure. The figure was prepared using BioRender (www.biorender.com).

15 amino acids, 117 dipeptides, 126 lipids, and 91 other metabolites. To explore the variation between genotypes, principal component analysis (PCA) and a hierarchical cluster (HCA) heatmap were established on metabolic and phenotypic traits (Fig. 3A,B). The heatmap revealed considerable differences in metabolite abundances across genotypes (Fig. 3A), which was confirmed by PCA (Fig. 3B), in which the first and the second components explained 41.1% and 19.9% respectively, of saponin variance. PCA analysis identified 21 genotypes, mostly originating from Peru/Latin America, whose position in the PCA plot largely correlates with their geographical origin (Supplementary Table 3). Finally, we investigated correlations between and within different metabolite classes and phenotypic traits (Fig. 3C). This showed that many metabolites are highly associated within the network. However, no significant strong correlations were found between saponin content and morphological traits, indicating that genotypes with low saponin content can be selected in the future by breeding or genome editing without an impact on yield-related traits.

#### Methods

**Quinoa germplasm.** Four-hundred and sixty-eight quinoa genotypes, plus one accession from djulis (*Chenopodium formosanum* Koidz.) and one from goosefoot (*Chenopodium album* L.) were selected for the field experiment (Supplementary Table 1<sup>19</sup>). Seeds from quinoa accession QQ74, for which a reference genome sequence is available<sup>17</sup>, were included as well. The source of the seeds is given in Supplementary Table 1<sup>19</sup>. Seeds of the different genotypes were propagated at the International Center for Biosaline Agriculture (ICBA) fields in years 2016 and 2017, and stored in a cold chamber at 2 °C and a relative humidity of 30%. The seeds were sown by hand by dibbling 2-3 seeds for each hole/location into the ground, to a depth of 1-2 cm near the dripper. Plants were thinned after about two weeks by removing unusually weak or strong individuals to leave one plant per location.



**Fig. 3** Phenotypic and metabolic variation across quinoa accessions. **(A)** Heatmap showing the relative levels of metabolites and phenotypic data measured on quinoa genotypes (full data set provided in Supplementary Dataset 1 and Supplementary Table 1<sup>19</sup>). **(B)** Correlation network between phenotypic traits and metabolic traits, each node represents a metabolite or a plant phenotypic trait, edges connecting two nodes show an association between two traits. **(C)** Principal component analysis (PCA) on saponin data across all genotypes.

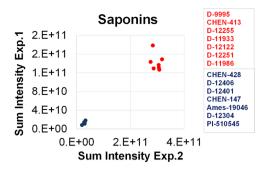
**Experimental site and design.** Experiments were carried out at the field research facilities of the International Center for Biosaline Agriculture, ICBA (N 25° 05.847; E 055° 23.464), Dubai, the United Arab Emirates, from November 2016 to April 2017. The soils at ICBA experimental fields are sandy in texture, that is, fine sand (sand 98%, silt 1%, and clay 1%), calcareous (50–60% CaCO<sub>3</sub> equivalents), porous (45% porosity), and moderately alkaline (pH 8.2). The saturation percentage of the soil is 26 with a very high drainage capacity, while electrical conductivity of its saturation extract (ECe) is 1.2 dS m<sup>-1</sup>. According to the American Soil Taxonomy<sup>21</sup>, the soil is classified as typic Torripsamments, carbonatic and hyperthermic<sup>22</sup>. Prior to sowing, poultry manure (Al Yahar Organic Fertilizers, UAE) was added at 40 tons per hectare (t ha<sup>-1</sup>) in the field chosen for the experiments. After four weeks of sowing, urea (nitrogen-phosphorus-potassium (NPK) content: 46-0-0) was applied at 40 kg ha<sup>-1</sup>, while NPK (20-20-20) was applied at 30 kg ha<sup>-1</sup> after eight weeks of planting. Fertigation technique was used for the application of chemical fertilizers. The experimental plots were randomized following an augmented design<sup>23</sup>, with each accession harboring a plot size of 1 m x 1 m. The distance between both, rows and plants was 25 cm.

**Irrigation system.** A drip irrigation system was used for the experiment, with drippers at 25 cm distance, which was part of SCADA (Supervisory Control and Data Acquisition) system. Irrigation was provided twice a day for 10 min each time. For irrigation, about 13.3 L of water was used daily per plot. Data on relative humidity, temperature, and rainfall at the experimental site were recorded by the meteorological station at ICBA (Supplementary Table 4).

**Data collection.** Eleven different morphological traits were recorded to assess the variation among the quinoa genotypes (including two species related to quinoa). For days to flowering, the data were recorded when about 50% of plants were flowering (Supplementary Table 1<sup>19</sup>). Data on plant height, number of primary branches, number of panicles, main panicle length, plant dry weight, and seed weight were collected after plant maturity. For dry weight measurements, plants were kept in a drier electric oven (Model-PF 30, Carbolite, United Kingdom) at 40°C for 48 hours.

**Extraction of lipids and polar metabolites.** The extraction protocol was adapted and modified from Giavalisco *et al.*<sup>24</sup>. Metabolites were extracted from the quinoa seeds using a methyl-tert-butyl ether (MTBE)/ methanol/water solvent system. Equal volumes of the lipid and polar fractions were dried in a centrifugal evaporator and stored at -20 °C until processed further.

**LC-MS metabolomics.** Polar and semipolar metabolites: After extraction, the dried aqueous phase was measured using ultra-performance liquid chromatography coupled to a Q-Exactive mass spectrometer (Thermo



**Fig. 4** Seeds of 14 quinoa genotypes characterized by the lowest and highest saponin content. Metabolites were extracted and measured twice to assess metabolic profiling's reproducibility. Data express a sum of all metabolic features detected in the positive mode in the retention time window between 8.14 and 14.00 min, which corresponds to the saponin elution and is used here as a proxy for total saponin content.

Fisher Scientific) in positive and negative ionization modes, as described<sup>24</sup>. Samples were run in ten consecutive sets of 50 samples and one set of 10 samples. Lipids: After extraction, the dried organic phase was measured using ultra-performance liquid chromatography coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific) in positive mode, as described<sup>24</sup>. Samples were run in ten consecutive sets of 50 samples and one set of 10 samples.

Data pre-processing: LC-MS metabolite data. Expressionist Refiner MS 12.0 (Genedata, Basel, Switzerland) was used for processing the LC-MS data (https://www.genedata.com/products/expressionist). Repetition was used to reduce the volume of data and to speed up processing. All types of data except Primary MS Centroid Data were removed using Data Sweep. Chemical Noise Subtraction activity was used to remove artefacts caused by chemical contamination. Snapshots of chromatograms were saved for further processing. Further processing of chromatogram snapshots was performed as follows: chromatogram alignment (Retention time (RT) search interval 0.5 min), peak detection (minimum peak size 0.03 min, gap/peak ratio 50%, smoothing window 5 points, centre computation by intensity-weighted method with intensity threshold at 70%, boundary determination using inflection points), isotope clustering (RT tolerance at 0.02 min, m/z tolerance 5 ppm, allowed charges 1–4), filtering for a single peak not assigned to an isotope cluster, charge and adduct grouping (RT tolerance 0.02 min, m/z tolerance 5 ppm). A detailed description of the software usage and possible settings was published before<sup>25</sup>.

An MPI-MP in-house reference library was used to identify molecular features allowing 0.005 Da mass deviation and dynamic retention time deviation (maximum 0.2 min). Processing of fractionated samples resulted in annotation of 400 compounds (Supplementary Table 2)<sup>19</sup>.

Saponin and ecdysteroid annotation was based on the fragmentation behavior of the parent ion characteristic for the positive mode, and the mass of the main adduct measured in the negative mode.

**Data processing.** Data represent normalised intensities of the main adduct measured in either the positive or negative mode. Normalisation was done to the median of a given metabolite calculated across a set.

#### **Data Records**

In this study, we generated for the first time a large repertoire of the seed metabolome for 471 quinoa genotypes. Morphological data of plants are presented in Supplementary Table 1. Using high-resolution mass spectrometry, we annotated and provided normalized metabolite data of 400 compounds across the genotypes. Data are presented in EXCEL files (Supplementary Dataset 1). For each compound, we present m/z, retention time, ion detection mode, and annotation confidence (Supplementary Table 2)<sup>24</sup>. The data are hosted and available at figshare<sup>19</sup>). The primary access site for raw metabolic data of the 471 samples is MetaboLights<sup>26</sup>.

#### **Technical Validation**

To validate data reducibility, we chose 14 genotypes representing low and high saponin contents, and again analyzed the saponin content of their seeds (Fig. 4; Supplementary Table 5). The data showed high correlation (Pearson correlation coefficient, 0.98) between the sum of the saponin peaks identified in the two experiments validating metabolomics analysis.

#### **Usage Notes**

As mentioned above, the profiling data revealed considerable differences in metabolite abundances across genotypes. The data may be used to calculate the fold-change of certain metabolites between selected genotypes. In some cases, the composition of metabolites may influence product quality, as e.g. known for the Maillard reaction in bread making<sup>27</sup>. Hence, this dataset may be used in breeding programs when selecting specific genotypes with desirable metabolite profiles that may benefit product quality. Furthermore, in combination with the availability of genome sequences, the data can be used for functional genomics- and metabolite-based genome-wide association studies (mGWAS) to dissect the genetic basis of quinoa seed metabolism. The information on metabolite presence and quantity may also be used as a basis to design molecular markers to characterize responses to abiotic stresses. The data set is useful in genetic and correlation studies to investigate the relationship between metabolic diversity, geographical distribution, and integration with physiological and phenotypic diversity. The

mass spectrometry raw data are available in MetaboLights, which allows download and re-processing with several commonly available tools such as xcm, GNPs and OpenMS. Furthermore, this enables the community to collect metabolite data of 471 different genotypes to generate a standard metabolome of quinoa.

#### Code availability

The commercially available GeneData software (https://www.genedata.com/) was used for LC-MS data analysis.

Received: 27 July 2021; Accepted: 19 May 2022;

Published online: 20 June 2022

#### References

- 1. Pereira, E. et al. Chenopodium quinoa Willd. (quinoa) grains: A good source of phenolic compounds. Food Res. Int. 137, 109574
- 2. Burrieza, H. P., Rizzo, A. J., Moura Vale, E., Silveira, V. & Maldonado, S. Shotgun proteomic analysis of quinoa seeds reveals novel lysine-rich seed storage globulins. Food Chem. 293, 299-306 (2019).
- 3. Pereira, E. et al. Chemical and nutritional characterization of Chenopodium quinoa Willd (quinoa) grains: A good alternative to nutritious food. Food Chem. 280, 110-114 (2019).
- 4. Pearsall, D.M.: Plant domestication and the shift to agriculture in the Andes, in The Handbook of South American Archaeology, H. Silverman and W.H. Isbell, Editors, Springer New York: New York, NY., 105-120 (2008).
- 5. Tapia, M.: The long journey of quinoa: who wrote its history? In FAO & CIRAD. State of the Art Report on Quinoa in the World in 2013, p. 17-13, Rome (2015).
- 6. Jacobsen, S.-E. The worldwide potential for quinoa (Chenopodium quinoa Willd.). Food Rev. Int. 19, 167–177 (2003).
- 7. Hinojosa, L., González, J. A., Barrios-Masias, F. H., Fuentes, F. & Murphy, K. M. Quinoa abiotic stress responses: A Review. Plants (Basel) 7, 106 (2018).
- 8. Bosque Sanchez, H., Lemeur, R. & van Damme, P. Ecophysiological analysis of drought and salinity stress of quinoa (Chenopodium quinoa Willd.). Food Rev. Int. 19, 111-119 (2003).
- 9. Adolf, V. I., Jacobsen, S.-E. & Shabala, S. Salt tolerance mechanisms in quinoa (Chenopodium quinoa Willd.) Environ. Exp. Bot. 92, 43-54 (2013).
- 10. López-Marqués, R. L. et al. Prospects for the accelerated improvement of the resilient crop quinoa. J. Exp. Bot. 71, 5333-5347 (2020).
- 11. Satheesh, N. & Fanta, S. W. Review on structural, nutritional and anti-nutritional composition of teff (Eragrostis tef) in comparison with quinoa (Chenopodium quinoa Willd.). Cogent Food Agric. 4, 1546942 (2018).
- 12. Kuljanabhagavad, T., Thongphasuk, P., Chamulitrat, W. & Wink, M. Triterpene saponins from Chenopodium quinoa Willd. Phytochemistry 69, 1919-1926 (2008).
- 13. Madl, T., Sterk, H., Mittelbach, M. & Rechberger, G. N. Tandem mass spectrometric analysis of a complex triterpene saponin mixture of Chenopodium quinoa. J. Am. Soc. Mass. Spectrom. 17, 795-806 (2006).
- 14. Woldemichael, G. M. & Wink, M. Identification and biological activities of triterpenoid saponins from Chenopodium quinoa. J. Agric. Food Chem. 49, 2327-2332 (2001).
- 15. El Hazzam, K., et al. An insight into saponins from quinoa (Chenopodium quinoa Willd): A review. Molecules 25 (2020).
- 16. Otterbach, S., Wellmann, G. & Schmöckel, S.M. Saponins of quinoa: structure, function and opportunities, in The Quinoa Genome, S. Schmöckel, Editor. 2021, Springer.
- 17. Jarvis, D. E. et al. The genome of Chenopodium quinoa. Nature 542, 307-312 (2017).
- 18. Man, S., Gao, W., Zhang, Y., Huang, L. & Liu, C. Chemical study and medical application of saponins as anti-cancer agents. Fitoterapia 81, 703-714 (2010).
- 19. Tabatabaei, I. et al. The diversity of quinoa morphological traits and seed metabolic composition. figshare https://doi.org/10.6084/ m9.figshare.19780462.v1 (2022).
- 20. Fernie, A. R. et al. Recommendations for reporting metabolite data. Plant Cell 23, 2477–2482 (2011).
- 21. Soil Survey Staff: Keys to Soil Taxonomy. 11th edition ed. 2010, United States Department of Agriculture.
- 22. Shahid, S. A., Dakheel, A. H., Mufti, K. A. & Shabbir, G. Automated in-situ soil salinity logging in irrigated agriculture. Eur. J. Sci. Res. 26, 288-297 (2009).
- 23. Federer, W. Augmented designs. Hawaiian Planters' Record 55, 191-208 (1956).
- 24. Giavalisco, P. et al. Elemental formula annotation of polar and lipophilic metabolites using (13) C, (15) N and (34) S isotope labelling, in combination with high-resolution mass spectrometry. Plant J. 68, 364-376 (2011).
- 25. Sokolowska, E. M., Schlossarek, D., Luzarowski, M. & Skirycz, A. PROMIS: Global Analysis of PROtein-Metabolite Interactions. Curr. Protoc. Plant Biol. 4, e20101 (2019).
- Tabatabaei, I. et al. The diversity of quinoa morphological traits and seed metabolic composition. MetaboLights https://www.ebi. ac.uk/metabolights/MTBLS2382 (2022).
- 27. Scalone, G. L., Cucu, T., De Kimpe, N. & De Meulenaer, B. Influence of free amino acids, oligopeptides, and polypeptides on the formation of pyrazines in Maillard model systems. J. Agric. Food Chem. 63, 5364-5372 (2015).

### Acknowledgements

S.B. thanks the Federal Ministry of Education and Research of Germany (BMBF) and the Arab-German Young Academy of Sciences and Humanities (AGYA) for funding of two Research Mobility Program grants to Dubai, United Arab Emirates, which allowed establishing the research reported here. B.M.-R. thanks the University of Potsdam, and S.B. thanks the Max Planck Institute of Molecular Plant Physiology for financial support. A.R.F. and B.M.-R. thank the European Union's Horizon 2020 Research and Innovation Programme, project PlantaSYST (SGA-CSA No. 739582 under FPA No. 664620) for funding. All authors are very grateful to Rostyslav Braginets and Dirk Walther from the Max Planck Institute of Molecular Plant Physiology, Potsdam, for their great help in uploading the metabolomics data to the MetaboLights database.

#### **Author contributions**

S.B. and B.M.R. designed and coordinated the project. S.B. obtained funding. I.T. and E.L. extracted metabolites. A.S. and S.A. analysed metabolite data. M.W. assisted with saponin annotation under the supervision of A.S. S.A. performed PCA and clustering analysis under the supervision of A.R.F. M.S., S.T., and H.M. performed field experiments and collected the morphotrait data. K.M.M., S.M.S., and M.T. contributed to establish the seed collection at ICBA.

#### **Funding**

Open Access funding enabled and organized by Projekt DEAL.

#### **Competing interests**

The authors declare no competing interests.

#### **Additional information**

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01399-y.

**Correspondence** and requests for materials should be addressed to A.S. or S.B.

Reprints and permissions information is available at www.nature.com/reprints.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>.

© The Author(s) 2022