

Depicting the genetic basis of grass pea seeds primary metabolism through a genome-wide association study

Ana Margarida Sampaio¹, Letice Gonçalves¹, Ana Margarida Rodrigues¹, Carla António¹, Maria Carlota Vaz Patto¹

¹Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal.

DOI: 10.5281/zenodo.15262705

High-quality food is nowadays one of the consumers top priorities. With plant-based eating habits being exceedingly popular, grain legumes are becoming more demanded namely due to their high protein content. However, not all grain legumes have been characterized to the same extent and limited data on seed quality is available on some promising species hampering their use and improvement. *Lathyrus sativus* (grass pea) is one such understudied example. Grass pea is an important source of protein and calories in drier areas, with a growing interest over three continents due to its ability to respond to climate challenges.

To reduce the gap of knowledge on the nutritional quality of grass pea seeds, the primary metabolism of a worldwide collection of about 150 grass pea genotypes was analyzed, allowing also to study their genetic basis and environmental influence. Mature seeds from these genotypes were harvested during three growing seasons (2015-16, 2016-17 and 2017-18), from a repeated field trial, in a Portuguese traditional grass pea growing region (Alvaiázere) and milled into flour. Using GC-TOF-MS (gas chromatography time-of-flight mass spectrometry metabolomics platform) primary metabolites present in the flour were relatively quantified. Forty-seven primary metabolites have been identified, including 23 amino acids, 11 sugars and derivatives, 7 organic acids, and 2 polyamines, among others. From those, 43 revealed significant differences between genotypes, with only two of them (glutamine and sucrose) revealing a genotype $\times$ environment ($G \times E$) interaction variance component superior to the genotypic one.

To unravel the genetic basis of these metabolites content, best linear unbiased estimators (BLUEs) were combined with previously generated 5,651 single nucleotide polymorphisms (SNP) and a genome-wide association study (GWAS) was performed using linear mixed models accounting for the genotypes' genetic relatedness.

Seventy-four genomic regions were identified associated with the 43 different nutritional quality-related metabolite variation, with six regions associated with multiple traits. Potential candidate genes will be explored, allowing to better understand the molecular mechanisms regulating the grass pea seeds primary metabolism. The generated knowledge on metabolite content and associated genomic regions will support future precision breeding efforts to answer consumers' nutritional quality concerns and increase the use of grass pea in plant-based diets.