Received: 27 October 2023 | Revised: 9 February 2024 | Accepted: 14 February 2024

DOI: 10.1002/ppp3.10507

Plants People Planet PPP Open Access

RESEARCH ARTICLE

Small proteins, great promises: Geographic bioprospecting of Bowman–Birk protease inhibitors and domestication side-effects in African cowpea (*Vigna unguiculata* L.)

Davide Panzeri[1] | Elisa Toini[1] | Jacopo Vertemara[1] | Giuseppe Silvestri[1] | Victor Vladut Bunea[1,2] | Giovanni Zecca[1] | Werther Guidi Nissim[1] | Robert Philipp Wagensommer[2] | Giuseppe Zampella[1] | Massimo Labra[1] | Fabrizio Grassi[1]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Milan, Italy

[2]Faculty of Education, Free University of Bozen-Bolzano, Brixen-Bressanone, Italy

Correspondence
Davide Panzeri and Fabrizio Grassi, Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126 Milan, Italy.
Email: davide.panzeri@unimib.it and fabrizio.grassi@unimib.it

Funding information
This work was funded under two projects: 'National Biodiversity Future Center – NBFC' - the National Recovery and Resilience Plan (NRRP), Mission 4 Component 2 Investment 1.4 - Call for tender No. 3138 of 16 December 2021, rectified by Decree n.3175 of 18 December 2021 of Italian Ministry of University and Research funded by the European Union – NextGenerationEU; Project code CN_00000033, Concession Decree No. 1034 of 17 June 2022 adopted by the Italian Ministry of University and Research, CUP H43C22000530001. Spoke 5 and Spoke 6. PRIN (Project of Relevant National Interest) project Plants Bioprospecting Of Serine Proteases Inhibitors To Develop An Innovative Colon Cancer Prevention Strategy, 'PRESERVE' by the 'Ministero dell'Istruzione dell'Università e della Ricerca' (MIUR), PRIN 2020 - CUP H45E21000160001.

Societal Impact Statement

The legume crop cowpea is grown worldwide, but 90% of the world's total share is produced in Africa. It is a promising species due to its resilience properties, balance of macro and micronutrients and presence of health-promoting bioactive compounds. In African countries, cowpea has a crucial role in guaranteeing food security as a subsistence crop for families and commercial income for small farmers. The discovery of compounds with high nutraceutical value and bioactive properties supports socio-economic policies to improve health and nutrition, especially in low- and middle-income countries. In turn, this encourages biodiversity protection and crop enhancement programmes.

Summary

- Bowman–Birk protease inhibitors (BBIs) are a restricted group of small proteins in plants mainly involved in defence mechanisms against pests. BBIs are demonstrated to be active components capable of reducing the viabilities of different cancer cell lines. BBI bioactivity is directly linked to the inhibition capacity, but the variability and the efficiency against the physiological targets of different BBI isoforms remain still unexplored.
- We analysed the natural genetic diversity of two main genes encoding BBI trypsin-trypsin (BBI-TT) and trypsin-chymotrypsin (BBI-TC) in wild and domesticated cowpea samples mainly spread in Sub-Saharan Africa. We analysed DNA sequences and respective amino acidic isoforms/isoproteins to explore signs of natural selection and haplotype relationships. Moreover, we calculated the binding energy between BBIs and their biological targets to identify which are the most efficient inhibitors and their geographical locations.

Davide Panzeri and Elisa Toini contributed equally.

- We found a high level of haplotype diversity for both genes, almost exclusively in wild accessions and detected positive and negative selection signals in the amino acid sequences. Furthermore, in the wild diversity pool, some BBI-TT and BBI-TC mature proteins were potentially better interactors with the physiological targets.
- The long interaction between plant-pathogen has selected new and useful isoforms in wild lineages that have allowed the chances of survival of the species to improve. On the other hand, the domestication process has produced an intense bottleneck leaving only poorly efficient BBI variants. In addition to providing information on the natural diversity and evolution of BBIs, our work discusses the potential applications in agriculture and human health.

KEYWORDS

binding energy calculations, bioactive compounds, biodiversity, Bowman–Birk protease inhibitors, domestication, legumes, selective pressure, *Vigna*

1 | INTRODUCTION

Plants are a fundamental source of natural protease inhibitors which have the function of discouraging the attacks of invading insect herbivores by inactivating digestive proteases (Pandey et al., 2022). Protease inhibitors are generally small proteins containing one or more inhibitory domains and are classified based on the chemical nature of the inhibited groups (Laskowski & Kato, 1980). Among the serine protease inhibitors, Bowman–Birk protease inhibitors (BBIs) are one of the best characterised groups. Molecular studies have suggested that BBIs appeared in the most ancient tracheophytes (James et al., 2017), but were lost by many families during the evolution process. However, in Poaceae and Fabaceae families, BBIs show a highly conserved structure composed of about 70 amino acids, 14 of which are cysteines that create 7 disulphide bonds folding the protein (Mello et al., 2003). In Fabaceae, BBIs show a 'double headed' conformation because of the presence of two independent inhibitory domains (Domoney, 1999; Mello et al., 2003). Affinity for its specific target is determined by a single amino acid (called P_1) located in the interactive domain of the protein that initiates the interaction with the target enzyme (Schecter & Berger, 1967). According to the intrinsic physicochemical characteristics of the P_1 amino acid, BBI can inhibit trypsin (with positively charged amino acids, such as lysine or arginine) or chymotrypsin (with aromatic or apolar amino acids, such as phenylalanine, tyrosine and leucine) (Clemente & Domoney, 2006) widely present in the digestive system of target insects.

Although BBIs are mainly involved in defence mechanisms against pests, they are believed to also be involved in some abiotic stress control mechanisms (Dramé et al., 2013; Shan et al., 2008). These inhibitors have shown to be resistant to high temperatures, typically required in cooking recipes, and, therefore, diets based on these nutrients are recognised to be beneficial in the treatment of many diseases for their anticarcinogenic, antioxidant and anti-inflammatory properties (Gitlin-Domagalska et al., 2020). Chemopreventive activities have been tested showing encouraging results for several legumes, fundamentals for human alimentation, such as soybean, pea and bean (Clemente & del Carmen Arques, 2014; Conti et al., 2021; Sánchez-Chino et al., 2015). Some researchers have proposed the employment of BBIs or enriched extracts as therapeutic agents in various human cancers (Gitlin-Domagalska et al., 2020; Srikanth & Chen, 2016; Tripodi et al., 2020) in particular against colorectal cancer (CRC) models (Clemente et al., 2005, 2010; Lima et al., 2016; Olías et al., 2019; Panzeri et al., 2020). Recently, BBI extracts recovered from *Vigna unguiculata* (L.) Walp. (cowpea) water-boiled seeds were assumed to be the main active component able to reduce the viability of different CRC cell lines reducing the levels of phosphorylated epidermal growth factor receptor (EGFR) (Panzeri et al., 2020). These findings confirm the importance of including cowpea in legume-based diets to prevent CRC occurrence. Considering all these beneficial properties, it is considered important to study the variability of BBI with the aim of identifying more effective forms both for plant pest resistance and human wellbeing.

Cowpea has been considered an orphan crop for several decades, but its cultivation is spreading in many semi-arid countries where other crops are poorly suitable (Boukar et al., 2019; Guzzetti et al., 2019; Panzeri et al., 2022). Some studies that have compared traditional and modern landraces have shown different concentrations of proteins and minerals, proposing that some underutilised lineages could be used to select new varieties (Abadassi, 2015; Boukar et al., 2011; Dakora & Belane, 2019; Horn et al., 2015). Given that cowpea is an increasingly widespread crop with great plasticity and adaptability, wild lineages could become a crucial resource for future breeding and agricultural programmes providing useful genes and traits that could greatly improve current cultivars (Harouna et al., 2018; Padulosi & Ng, 1997). Taxonomic analyses have identified over 10 subspecies with a high morphological diversity suggesting that *V. unguiculata* originated in the South-East regions of Africa where the greatest wild diversity is found (Coulibaly et al., 2002; Pasquet, 1996, 1997, 1999; Pasquet et al., 2021; Pasquet & Padulosi, 2013; Xu et al., 2012). Unfortunately, these studies consider

only morphological or non-coding molecular markers and the natural variability of genes that encode for BBIs remains unknown (Muñoz-Amatriaín et al., 2021; Panzeri et al., 2022).

In our study, we have explored the natural genetic diversity of two main genes encoding BBI in cowpea able to inhibit trypsin and chymotrypsin enzymes (hereafter BBI-TT for trypsin/trypsin gene and BBI-TC for trypsin/chymotrypsin gene). These two genes are the main ones responsible for BBI proteins in cowpea as observed in various studies (Clemente et al., 2010; Hilder et al., 1989; Mehdad et al., 2016; Panzeri, 2023). We hypothesise that wild accessions during the evolution process have kept the majority of diversity, and they are currently harbouring the BBI proteins with high energetic efficiency. To address our hypothesis, we: (i) characterised the haplotypes and respective amino acidic isoforms and assessed the geographical distribution of BBIs variability; (ii) identified the codons targeted by natural selection; (iii) finally, we calculated the energy of interaction between the BBI proteins and their respective biological targets to identify which are the most efficient protease inhibitors and in which accessions they are located.

2 | METHODS

2.1 | Definitions

For greater clarity, we define 'accession' as the unique identifier assigned to collected organisms by the germplasm seed banks in their specific collection, while we define 'individual' as the plant gathered after germination and from which we have extracted the genomic DNA. On average, 1–3 individuals per accession were analysed in this study. Furthermore, we will use the term 'haplotype' to refer to all possible combinations of genetic variants found on the same locus, whereas the term 'isoform' will refer to the corresponding translated amino acid sequences. Eventually, we use the term 'isoprotein' to refer to the different forms of mature protein sequence, starting from the first maturation cleavage (Kumar et al., 2002).

2.2 | Sampling

In order to explore the geographic distribution of gene diversity of BBI-TT and BBI-TC genes in *V. unguiculata* a total of 426 individuals were sampled. In particular, 303 individuals (126 accessions) were obtained from International Institute of Tropical Agriculture (https://my.iita.org/accession2/, IITA, Ibadan, Nigeria), 116 individuals (51 accessions) were obtained from Meise Botanical Garden (http://db.plantentuinmeise.be/RESEARCH/COLLECTIONS/LIVING/PHASEOLUS/index.html, Meise, Belgium) and 7 individuals (5 accessions) were obtained from Centro De Recursos Fitogenéticos (https://www.inia.es/en-en/units/Institutes%20and%20Centres/CRF/Paginas/Home.aspx, CRF, Madrid, Spain). The list of all accessions, including origins, taxonomical classification and geographic coordinates is reported in Table S1. Overall, 399 individuals (168 accessions) came from Sub-Saharan African countries, whereas 27 individuals (14 accessions) came from out of Africa countries.

As a phylogenetic outgroup, 3 accessions belonging to *Vigna mungo* (L.) Hepper, *Vigna hirtella* Ridl. and *Vigna trinervia* (B.Heyne ex Wight & Arn.) Tateishi & Maxted were downloaded from the National Center for Biotechnology Information (NCBI) databank (https://www.ncbi.nlm.nih.gov, Table S2).

2.3 | DNA isolation and BBI genes characterisation

The choice of the candidate genes was carried out carefully based on literature and published data. In particular, the screening of existing protein crystals published in public databases (InterPRO, UniProtKB and PDB) led us to the identification of two BBI mature proteins: trypsin–trypsin (Q4VVG2, https://www.uniprot.org/uniprotkb/Q4VVG2/entry, crystal entry 2g81) and trypsin–chymotrypsin (Q9S9H8, https://www.uniprot.org/uniprotkb/Q9S9H8/entry, crystal entry 2r33). Secondly, the specific sequence of these BBI entries was blasted into the reference *V. unguiculata* genome to obtain the complete gene sequence (Figure S1).

The genomic DNA of 426 samples was extracted from dried leaves with the E.Z.N.A.® Plant DNA kit (Omega Bio-Tek, https://omegabiotek.com/) according to the manufacturer's instructions. BBI-TT and BBI-TC genes were PCR-amplified using Esco Healthcare Swift-MaxPro thermocycler and in 25 µL total volume reactions containing 5–30 ng DNA, GoTaq® G2 Green Master Mix (Promega, https://ita.promega.com/), 5 pmol of each primer and purified water until the final reaction volume. Details on PCR protocols and primers are given in Tables S3 and S4. Products of PCR were visualised by 2% agarose gel electrophoresis stained with EuroSafe Nucleic Acid Stain (EuroClone, https://www.euroclonegroup.it/) and purified prior to DNA sequencing with the QIAquick PCR Purification Kit (Qiagen, https://www.qiagen.com/it). Successively, DNA sequencing was performed by Eurofins Genomics (Vimodrone, https://eurofinsgenomics.eu/), and sequences obtained were deposited in the NCBI GenBank database.

2.4 | Data preparation and gene diversity

Multiple sequence alignment was generated with the online version of MAFFT v. 7 (Katoh et al., 2019; Kuraku et al., 2013) producing two different datasets for BBI-TT and BBI-TC genes. DNA sequences were compared visually and those sequences containing heterozygous sites were probabilistically resolved in the corresponding haplotypes using the PHASE software (number of iterations = 500, thinning interval = 1, burn-in = 100) implemented in DnaSP v.6.12.03 (Rozas et al., 2017). Only individuals with resulting phase probabilities >.95 were considered for successive analysis. Successively, parsimony informative sites (P), nucleotide diversity (π) and haplotype diversity (HD) were calculated for both genes using the programme DnaSP v.6.12.03.

2.5 | AMOVA analyses

Analysis of molecular variance (AMOVA) statistical analyses on the geographic distribution of haplotypes was performed using Arlequin v3.5.2.2 (Excoffier et al., 1992; Excoffier & Lischer, 2010; Weir & Cockerham, 1984). The analysis was conducted by dividing accessions into two groups, West versus South-East Africa, excluding non-African accessions from the grouping and the subsequent analysis (Table S1). The dividing region was set in correspondence with the tropical Congolese rainforest. Analyses considered polymorphic sites including indels and were run with 1000 standard permutations. Moreover, P, π and HD were calculated for geographic groups (West and South-East Africa) and wild and domesticated groups.

2.6 | Phylogenetic and network analysis

We inferred a maximum likelihood (ML) tree using RAxML-HPC BlackBox for both datasets produced with and without outgroups performed under the general time-reversible with gamma distribution (GTRGAMMA) model of substitution (Stamatakis, 2014). All runs were performed on the Cyberinfrastracture for Phylogenetic Research (CIPRES) Science Gateway (Miller et al., 2010). Moreover, a neighbour joining (NJ) tree was produced using the p-distance method and 1000 bootstrap replicates by MEGA11 software (Tamura et al., 2021). Finally, to understand the relationships among the identified haplotypes the Popart 1.7 software (Leigh & Bryant, 2015) was chosen to perform network analyses using the parsimony method (Templeton–Crandall–Sing network (TCS) network). A network is generated by the integrated algorithm by collapsing identical haplotypes and separating them by single base mutations with probability >95% (Clement et al., 2002).

2.7 | Positive and negative selection analyses

In order to detect the sites under selection the haplotypes were prepared using Codon Alignment v2.1.0 (https://www.hiv.lanl.gov) and all stop codons were manually removed. Fast, unconstrained Bayesian approximation (FUBAR) (Murrell et al., 2013), fixed effects likelihood (FEL) (Kosakovsky Pond & Frost, 2005), mixed effects model of evolution (MEME) (Murrell et al., 2012) and single-likelihood ancestral counting (SLAC) (Kosakovsky Pond & Frost, 2005) methods were performed using the DATAMONKEY web server (https://www.datamonkey.org/) (Weaver et al., 2018), while site model (SM) tests were performed using EasyCodeML v1.4 software (Gao et al., 2019). Each dataset was analysed to explore signs of positive and negative selection, and for each software, the analysis parameters were set as described below.

2.8 | Site model test

SM uses a statistical distribution to account for variation in the ω ratio among codons (Nielsen & Yang, 1998). The ratio (ω) of the non-synonymous substitution rate (dN) to the synonymous substitutions rate (dS) was used to determine the selective pressure (Yang & Nielsen, 2002). Codon substitution models were investigated under the preset mode: M7 versus M8 and M8a versus M8. To minimise the possibility of stumbling in local optima, we ran the M8 model with different starting ω values ($\omega = 2$ and 0.5) as suggested by Anisimova et al. (2002). Positive selection was reported in results if the estimated ω parameter was greater than 1 and if the likelihood ratio test (LRT) of the comparison (M7 vs. M8 and M8a vs. M8) showed p-value smaller than .05. In order to identify the codons under positive selection, we performed the Bayes empirical Bayes (BEB) analysis implemented in EasyCodeML (Yang et al., 2005). For each codon, positive selection was reported when BEB analysis showed $p > .9$.

2.9 | FUBAR analysis

FUBAR uses a Bayesian approach to infer dN and dS substitution rates on a per-site basis for a given coding alignment assuming that the selection pressure for each site is constant along the entire phylogeny (Murrell et al., 2013). FUBAR reports evidence for pervasive positive selection using posterior probabilities and the authors suggest that $p > .9$ is strongly suggestive of positive selection (Weaver et al., 2018). FUBAR was run using the following default options: number of Markov chain Monte Carlo (MCMC) chains to run = 5, length of each chain = 2,000,000, burn-in = 1,000,000 and samples drawn from each chain = 100. Moreover, we selected the advanced options: number of grid points = 50; concentration parameter of the Dirichlet prior = 0.5.

2.10 | FEL analysis

FEL uses an ML approach to infer dN and dS substitution rates on a per-site basis for a given coding alignment along the entire phylogeny (Kosakovsky Pond & Frost, 2005). FEL fits an MG94xREV model to each codon site to infer site-specific dN and dS substitution rates. We selected 500 bootstrap resamples. The selection signal of each codon estimated was reported for p-value < .1 (Poon et al., 2009).

2.11 | MEME analysis

MEME employs a mixed-effects ML approach to test the hypothesis that individual sites have been subject to positive selection (Murrell et al., 2012). MEME allows the distribution of ω to vary from site to site and also from branch to branch at a site and it can capture the molecular footprints of both episodic and pervasive positive selection. The significance threshold was set at $p = .1$ (Poon et al., 2009).

2.12 | SLAC analysis

SLAC uses ML to infer the most likely ancestral sequence at each node of the phylogeny (Kosakovsky Pond & Frost, 2005). SLAC then

employs a modified version of the Suzuki–Gojobori counting method (Suzuki & Gojobori, 1999) to directly count the total number of dN and dS changes that have occurred at each site. The significance threshold was set at $p = .1$ (Poon et al., 2009) and is ascertained at each site using an extended binomial distribution even if it may not be accurate for data sets that show high divergence levels.

2.13 | Geographical distribution

The maps of the accession distributions were generated with QGIS Desktop v. 3.28.3-Firenze (https://www.qgis.org/it/site/). Coordinates given by germplasm databanks were used to localise every accession. When coordinates were not available or uncertain, details provided in the accession description were instead used to increase the localisation precision. Furthermore, accessions were coloured based on the corresponding haplotypes, isoforms and isoproteins for each gene. The point size was determined by the number of samples. To avoid multiple overlaps on the same point, the software was instructed to shift the point around the precise position with a distance set at 1.

2.14 | Binding energy calculations

Structures of the 17 BBI mature proteins (7 BBI-TT and 10 BBI-TC, see Section 3) were predicted using AlphaFold v2.0 (Jumper et al., 2021) and then refined through molecular dynamics (MD) simulations with Groningen Machine for Chemical Simulation (GROMACS, Abraham et al., 2015; Páll et al., 2015). Five simulations were carried out for 100 ns each with an integration step of 2 fs in an NPT ensemble at a temperature of 300 K and pressure of 1 atm using Chemistry at Harvard Macromolecular Mechanics (CHARMM36, Huang & MacKerell, 2013) force field. For every mature protein, trajectories were concatenated and processed using the cluster analysis tool of GROMACS to identify the most representative structure for each system. Cluster analysis was carried out with gromos algorithm (0.6 Å cut-off for both BBI-TT and BBI-TC) (Daura et al., 1999). The structure representing the centroid of the most populated cluster was used as the starting point for the binding energy calculations. For the computational analysis, two crystals were used as reference: the BBI from *Vigna radiata* (L.) R. Wilczek, in complex with porcine trypsin (PDB id: 3MYW) and BBI from *Glycine max* (L.) Merr. in complex with bovine alpha-chymotrypsin (PDB id: 5J4S). These crystals were used to build a total of 14 BBI-TT trypsin/trypsin complexes (7 for each inhibitory domain) and 20 BBI-TC trypsin/chymotrypsin complexes (10 for each inhibitory domain), 2 for each isoform, in order to obtain a dimeric structure for each of the two protease reactive sites with trypsin/chymotrypsin, according to the standard mechanism for serine proteinase inhibition. Each complex was obtained by superimposing and replacing the original crystallographic BBI structure with the BBIs described in this work. Also, these complexes were refined using the MD protocol previously described (i.e. gromos cutoff: 0.5 Å for both BBI-TT-trypsin/trypsin complex and BBI-TC-trypsin/chymotrypsin complex). Trypsin and chymotrypsin structures were isolated from the aforementioned X-ray structure and optimised using the same MD protocol illustrated above for the BBI isoforms (gromos cutoff: 0.4 Å for porcine trypsin and 0.35 Å for bovine chymotrypsin). The binding energy is defined as the energy difference between the bound conformation of BBI to an enzyme (trypsin or chymotrypsin, $E_{enz\text{-}BBI}$) and the unbound BBI and enzyme in water ($E_{BBI} + E_{enz}$) according to this formula (Kuhn & Kollman, 2000):

$$\Delta E = E_{enz-BBI} - (E_{BBI} + E_{enz}).$$

A single-point energy calculation was performed on each structure using MacroModel from Schrödinger Suite (Release 2021-1, 2021) in order to obtain the potential energy of the system.

3 | RESULTS

3.1 | Molecular characterisation, gene diversity and evolution

Two multiple sequence alignments of 324 bp for BBI-TT and 345 bp for BBI-TC were produced and after the analyses with the Phase algorithm, a total of 24 haplotypes for BBI-TT and 29 haplotypes for BBI-TC were identified. These haplotypes resulted in 13 isoforms and 7 isoproteins for BBI-TT and 20 isoforms and 10 isoproteins for BBI-TC. Sequences corresponding to haplotypes, isoforms and isoproteins are summarised in Tables S5–S10. The list of individual samples and relative haplotypes, isoforms and isoproteins is summarised in Table S11. The distribution of haplotypes, isoforms and isoproteins for each gene in wild and cultivated subspecies is summarised in Table 1. Notably, only three isoforms and two isoproteins and only two isoforms and one isoprotein were found in domesticated subspecies for the BBI-TT and BBI-TC genes, respectively. Polymorphic sites for haplotypes, isoforms and isoproteins are listed in Tables S12–S17.

The diversity indices showed higher variability for BBI-TC ($P = 32$, $\pi = 0.00672 \pm 1 \times 10^{-5}$ and $HD = 0.685 \pm 0.016$) than for BBI-TT ($P = 18$, $\pi = 0.00403 \pm 1 \times 10^{-5}$ and $HD = 0.677 \pm 0.016$).

TABLE 1 The table summarises the distribution of haplotypes, isoforms and isoproteins for the BBI-TT and BBI-TC genes in wild and domesticated cowpea. The number of variants for each category is indicated in brackets.

Gene	Group	Haplotype	Isoforms	Isoproteins
BBI-TT	Wild	H1–H14, H16–H24 (23)	Iso1–Iso13 (13)	P1–P7 (7)
	Domesticated	H1, H2, H15 (3)	Iso1, Iso2, Iso3 (3)	P1, P2 (2)
BBI-TC	Wild	H1–H29 (29)	Iso1–Iso20 (20)	P1–P10 (10)
	Domesticated	H1, H7 (2)	Iso1, Iso6 (2)	P1 (1)

These indices were calculated to compare wild and domesticated accessions remarking the greatest variability within wild ones (BBI-TT: $P = 18$, $\pi = 0.00420 \pm 1 \times 10^{-5}$ and $HD = 0.7350 \pm 0.012$, BBI-TC: $P = 32$, $\pi = 0.00876 \pm 1 \times 10^{-5}$ and $HD = 0.799 \pm 0.0002$) than into domesticated ones (BBI-TT: $P = 4$, $\pi = 0.00259 \pm 0.00039$ and $HD = 0.245 \pm 0.035$, BBI-TC: $P = 1$, $\pi = 0.00056 \pm 0.00009$ and $HD = 0.2 \pm 0.032$) Furthermore, they were also performed for geographic groups and biological origin, showing higher value of variability for South Eastern Africa (Tables S18–S20). For both genes, the AMOVA analysis on wild accessions showed a 2.2% (BBI-TT) and 9% (BBI-TC) genetic differentiation among geographic groups, within which most of the observed variance was partitioned. The same analysis was performed on domesticated accessions of the two geographic groups showing 16% and 18.5% variation among populations, respectively (Tables S21 and S22).

Statistical parsimony networks that explain the phylogenetic relationships between haplotypes for each gene are shown in Figures 1a and 2a. Different colours used in the networks correspond to 13 and 20 isoforms obtained for BBI-TT and BBI-TC, respectively. The Sub-Saharan distribution of the different isoforms is shown in Figures 1b and 2b. The same networks painted for biological origin (wild vs. domesticated) and for different isoforms and isoproteins are available in the Supporting Information (Figures S2 and S3). Geographical distributions of accessions and their relative haplotypes, isoforms and isoproteins are reported in the Supporting Information (Figures S4–S10).

3.2 | Selection pressure on genes

Because SM test and BEB analyses implemented in EasyCodeML requires a phylogram as the input file, alternative phylogenetic trees were produced for the different analyses (see Section 2). Trees generated from RAxML and NJ analyses are listed in Newick format in Table S23. We found a considerable number of codons under negative selection in both genes (up to 9 sites for BBI-TT and up to 8 for BBI-TC). Moreover, signatures of positive selection were found in 2 residues for BBI-TT and in 3 residues for BBI-TC. All codons targeted by natural selection are summarised in Figure 3, which highlights their location on the amino acid sequences and the different analysis methods used. The generated trees are shown in Newick format in Table S19 and the complete list of all outputs obtained from SM, FEL, FUBAR, MEME and SLAC analyses is reported in the Supporting Information (Tables S25–S30).

3.3 | Binding energy calculations

Computational approaches were performed to assess and predict the inhibition capability of the proteins found by the genetic diversity exploration. In particular, among each calculation, some mature proteins resulted in needing less energy to interact with the physiological targets. Figure 4 shows the calculated relative energies for each mature protein compared with the most common BBI (P1) used as a reference. We highlighted that P2 and P7 for BBI-TT and P3 and P6 for BBI-TC were the mature proteins examined in this study that needed less binding energy (raw calculations are provided in Tables S31–S34). Taking the sequence of P1 as a reference, the mutations of P2 and P7 (BBI-TT) are two amino acid mutations near the first trypsin-interacting loop (isoleucine to methionine at Residue 50 and threonine to serine at Residue 52, respectively). The mutations of P3 and P6 (BBI-TC) are, respectively, a glycine to alanine (pos. 92) found near the chymotrypsin-interacting loop, while the P6 has a mutation in the N-terminus tail (glutamate to valine in pos. 40).

4 | DISCUSSION

A growing number of crystal structures of plant protease inhibitors are becoming available and an increasing number of pharmaceutical companies are showing interest in their application in human clinical trials (Srikanth & Chen, 2016). Although protease inhibitors can be produced by synthetic processes, they can be more easily ingested through a diet based on traditional cereals or legumes, reducing their costs and avoiding negative impacts on the human organism. Therefore, in this study, we have explored the natural diversity developed during the evolution of the BBI-TT and BBI-TC genes in cowpea, evidencing which accessions conserved the mature BBIs with the highest potential for agriculture and human health.

4.1 | Distribution of variability and natural selection in BBI genes

Our results show that during evolution several mutations have been accumulated on both BBI genes examined. As evidenced by different tests (Figure 3), signatures of negative selection are spread along the amino acid sequences indicating that several residues are probably fundamental in maintaining the structure of proteins. On the other hand, codons targeted by pervasive positive selection may be promoting the emergence of different and more efficient mature isoproteins, suggesting that evolutionary forces could be still in progress. Although different stressors such as past climate changes, pests or agriculture can have influenced the evolution of BBI genes, the interaction with insects should be the main candidate (Bussotti et al., 2014; Pandey et al., 2022; Panzeri et al., 2022). Different genes that encode for BBIs are frequently found in plants (Panzeri et al., 2022; Srikanth & Chen, 2016) and the strategy of increasing the number of variants is generally used in plants to reinforce the defences efficiency, minimising the hydrolysis risk by the enzymes of insects (Gitlin-Domagalska et al., 2020). Our results, showing a remarkable number of different isoforms within the wild subspecies (13 isoforms in BBI-TT and 20 isoforms in BBI-TC), would therefore be consistent with the 'arms race' paradigm. In this perspective, new isoforms produced in the populations could be targeted by natural selection during the evolution, improving the chances of survival of the species. Although this

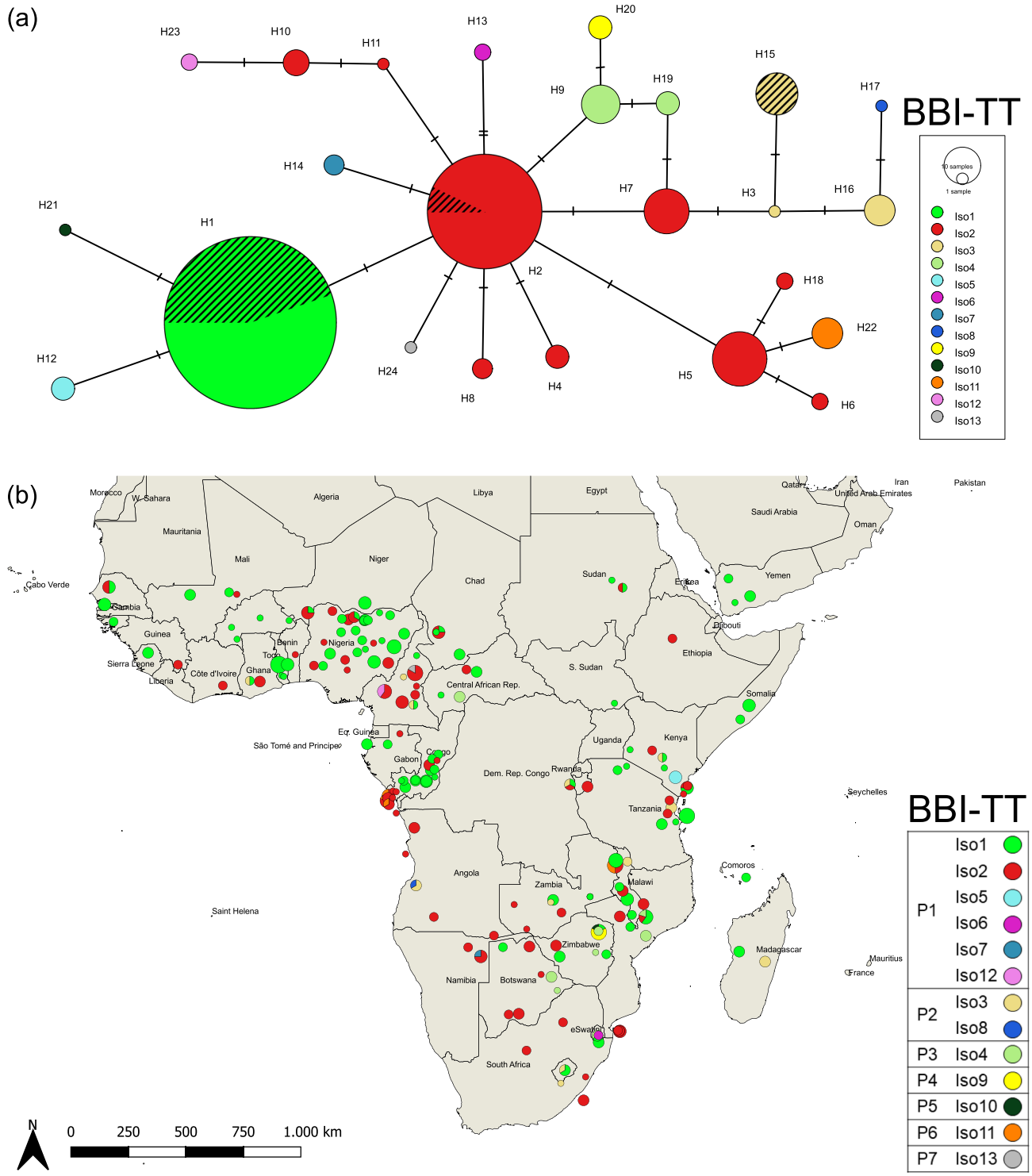

FIGURE 1 (a) Network and geographical distribution in Sub-Saharan Africa of haplotypes and corresponding isoforms of the gene encoding the trypsin–trypsin Bowman–Birk protease inhibitor (BBI-TT) in *Vigna unguiculata* accessions. Haplotype networks were generated using the PopArt v.1.7 programme from BBI-TT gene. Colours indicate the corresponding isoforms (Iso1–Iso13) and portions with lined texture highlight samples from domesticated accessions. Black hatch marks along the branches indicate the numbers of mutations. (b) Map showing the distribution of BBI-TT gene isoforms (grouped by corresponding mature proteins) in Sub-Saharan Africa.

explanation seems plausible, we cannot completely exclude that other causes contributed to shaping the current pattern of observed isoforms. On the other hand, regarding the cultivars included in this work, our findings do not support the idea that the domestication process had a positive role in the selection and conservation of isoforms (Larson et al., 2014; Panzeri et al., 2022). Among all landraces analysed

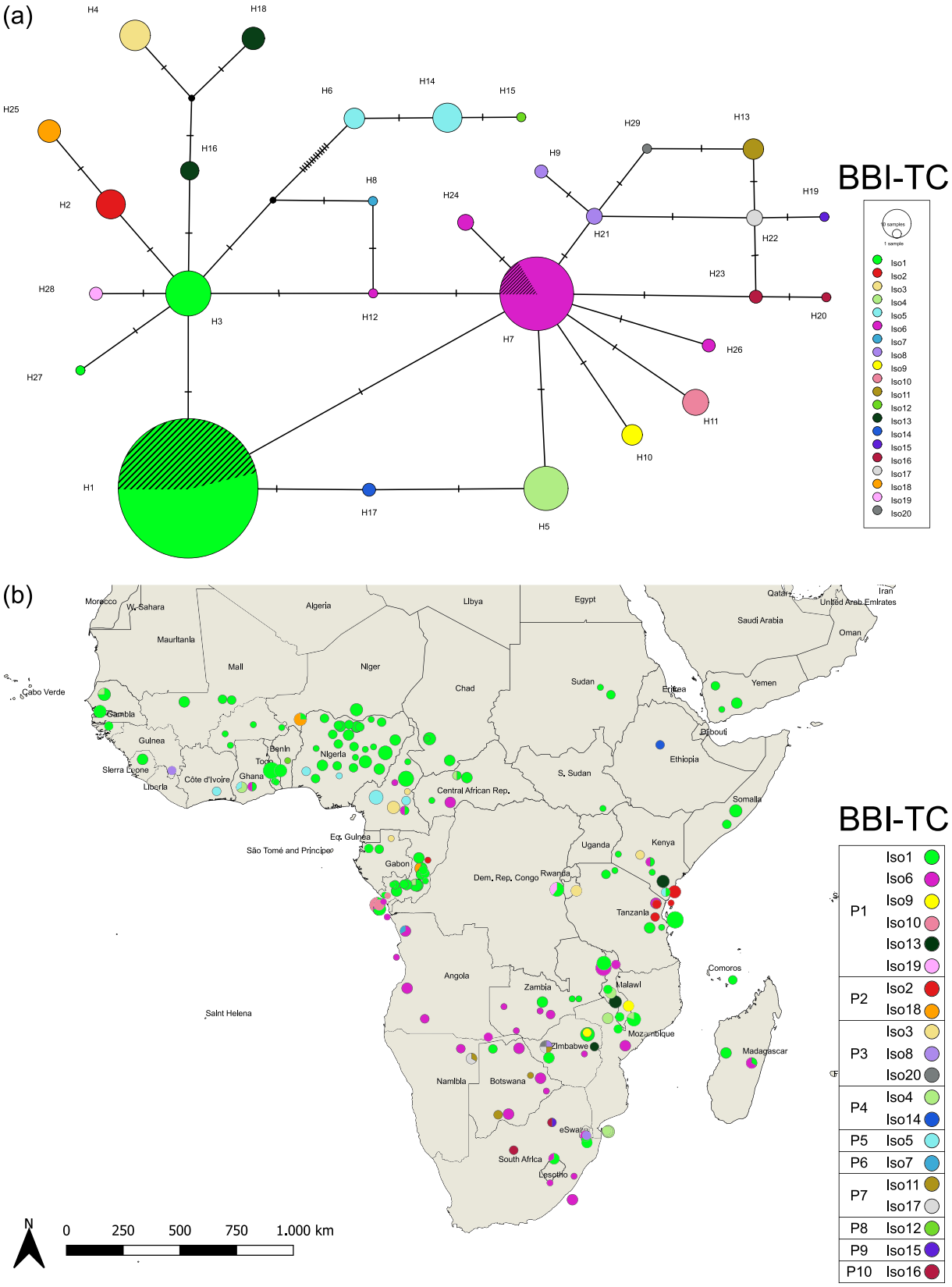

FIGURE 2 (a) Network and geographical distribution in Sub-Saharan Africa of haplotypes and corresponding isoforms of the gene encoding the trypsin–chymotrypsin Bowman–Birk protease inhibitor (BBI-TC) in *Vigna unguiculata* accessions. Haplotype networks were generated using the PopArt v.1.7 programme from BBI-TC gene. Colours indicate the corresponding isoforms (Iso1–Iso20) and portions with lined texture highlight samples from domesticated accessions. Black hatch marks along the branches indicate the numbers of mutations. (b) Map showing the distribution of BBI-TC gene isoforms (grouped by corresponding mature proteins) in Sub-Saharan Africa.

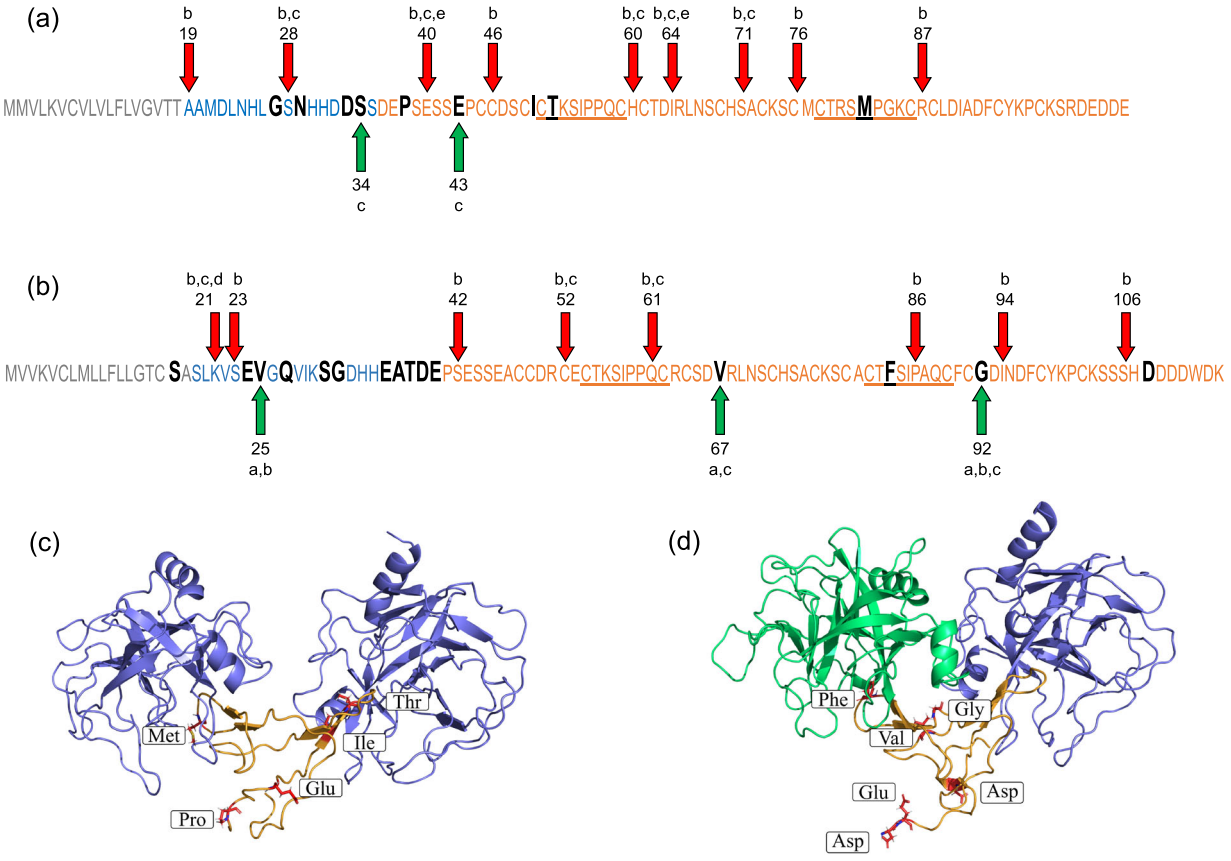

FIGURE 3 Graphic summary of the *in-silico* analyses conducted on the whole amino acid sequence and residue mutations on the mature protein in the interaction with their physiological targets. Selection analyses on the whole translated gene sequences for (a) trypsin–trypsin Bowman–Birk protease inhibitor (BBI-TT) and (b) trypsin–chymotrypsin Bowman–Birk protease inhibitor (BBI-TC). The shown sequences and proteins, used as reference, correspond to P1 of both BBI-TT and BBI-TC. Portions of the amino acid sequence are coloured differently: signal peptide in grey, propeptide in blue and mature protein in orange. Interactive domains are underlined while residues that change among all isoforms found are represented in bold-black font (see Tables S13 and S16). Sites detected by selection analyses are indicated by arrows (red = negative selection and green = positive selection). Letters explain which method detected the corresponding site under pressure (a = site model [SM], b = fixed effects likelihood [FEL], c = fast, unconstrained Bayesian approximation [FUBAR] and d = single-likelihood ancestral counting [SLAC]) and the numbers explain the corresponding residue. Graphical representation of BBI-TT (c) and BBI-TC (d) mature isoproteins in interaction with the physiological targets. Trypsins are coloured in light purple, chymotrypsin in green and BBIs in orange while the mutated amino acids are shown in red and labelled with the three letters code (in alphabetical order: Asp-aspartate, Glu-glutamate, Gly-glycine, Ile-isoleucine, Met-methionine, Phe-phenylalanine, Pro-proline, Thr-threonine, Val-valine).

in this study, only two isoforms for BBI-TC and three isoforms for BBI-TT were identified (see Table 1). This notable loss of variation compared with wild subspecies can be seen as a side effect of the domestication process. Likewise other crops, the accessions extensively cultivated in the world today have originated from a small number of plants. The marked reduction in population size associated with the domestication bottleneck would have caused a reduction of the gene pool variability. In addition, reducing population size would have increased the strength of genetic drift, leading to further loss of variability.

Finally, we highlighted that the geographic distribution of genetic diversity is unequally distributed between West and South/east Sub-Saharan Africa and the higher variability found is located in South-eastern regions. Because it is generally recognised that the centre of origin of a species is characterised by higher variability, our results reinforce the idea that the most ancient wild lineages of *V. unguiculata* are originated in the South-Eastern area (Padulosi & Ng, 1997; Pasquet, 1997, Pasquet & Padulosi, 2013; Pasquet et al., 2021). Therefore, we suggest that extensive sampling, especially in the areas where the wild progenitors originated, could greatly increase the probability of finding new unknown genetic variants.

4.2 | BBI bioactivity and potential applications

We have evaluated the interaction energies of Bowman–Birk protease inhibitors with trypsin and chymotrypsin targets demonstrating that the mature proteins that need less binding energy are spread mostly in wild accessions of African cowpea. P2 and P7 for BBI-TT and P3 and P6 for BBI-TC are the most efficient BBI forms examined

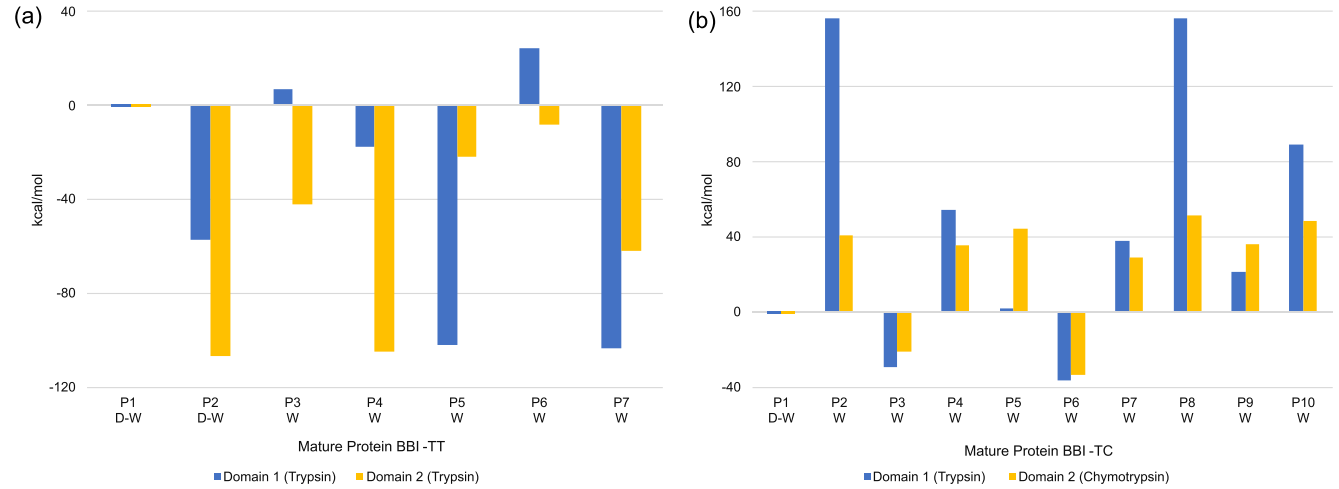

FIGURE 4 Calculation results of the found mature proteins in interaction with physiological targets (trypsin or chymotrypsin) for each interacting domain. Binding energy calculations of the (a) trypsin–trypsin Bowman–Birk protease inhibitor (BBI-TT) and (b) trypsin–chymotrypsin Bowman–Birk protease inhibitor (BBI-TC) mature proteins in comparison to the most common protein (P1 for each graph). Blue column explains the calculated energy expressed in kcal/mol at the first protein interaction domain, while the yellow column at the second interaction domain, enzyme affinity is summarised in brackets. Protein belonging to domesticated and/or wild accessions are specified with D for domesticated and W for wild.

in this study, probably obtained through a long interaction between plants and different pathogens spread in Africa. The past lack of human-driven selective pressure for some isoforms over others, combined with the domestication bottleneck and the random effect of genetic drift can be argued as the main causes of the small number of isoproteins observed within the domesticated group (Panella & Gepts, 1992; Xiong et al., 2016). Unfortunately, these isoproteins largely correspond to the less energy-efficient BBI variants. This event is clearly observable in P1 of BBI-TT, which, despite being the most common protein in domesticated accessions, shows a higher binding energy when interacting with physiological targets where the calculated differences with P2 are −57.2 kcal/mol and −106.8 kcal/mol for the two trypsin domains, respectively. Single amino acid substitutions can have huge impacts on the protein stability or target affinity, due to their physicochemical properties. At the protein level, this change can involve a total target shift, such as the domain affinity in the BBI for trypsin or chymotrypsin, or a change in overall net charge or interaction energy. MD coupled with binding energy calculations have allowed us to consider the entire 3D protein model resulting in a clearer and wider perspective. We report that improving mutated amino acids had a direct effect on the energy needed for target interaction, probably due to the proximity to the interaction loops and the physicochemical properties of the single amino acids. Comparing the best BBIs (Figure 4, Tables S14 and S17) with the most common isoprotein (P1) diffused in almost the totality of domesticated accessions, we present a panel of mutations that have impacted the protein functionality. For example, the change of a methionine into an isoleucine in Position 50 (P2 to P1, BBI-TT) involves the loss of a sulphur atom and prevents a further salt bridge, affecting the protein–protein interaction. The change from a serine to a threonine in Position 52 (P7 to P1, BBI-TT) involves an addition of a methyl group, moderately decreasing the polarity of the loop. For the BBI-TC gene, the mutation in Position 92 (P3 to P1) implies a loss of a methyl group (a glycine instead of an alanine), disfavouring the chymotrypsin hydrophobic interaction. In addition, the codon in Position 92 turned out to be targeted by positive selection indicating that pervasive forces of selection are still influencing the evolution of protein in this site. Although the most efficient BBI forms examined in this study are located especially in wild lineages, a restricted group of domesticated accessions show a BBI form characterised by low binding energy (P2 for BBI-TT). These landraces are mostly cultivated in South-eastern regions of Africa (Kenya, TVU11422; Tanzania, TVU13237; Lesotho, TVU15404; Madagascar, TVU6901) and in India (NI778), suggesting that at least a part of the Asian cowpea could have arrived directly across the Indian ocean or across the Sabaean Lane in modern Yemen from South-Eastern Africa to India (Blench, 2003; Herniter et al., 2020). This result does not contradict the hypothesis of diffusion from Africa to Asia proposed by different authors (Herniter et al., 2020; Padulosi & Ng, 1997; Panzeri et al., 2022), and, in addition, it highlights that some domestic lineages are also a valuable pool of genetic resources (Ramanatha Rao & Hodgkin, 2002) that can be exploited to integrate or ameliorate features of other existing cultivars. The physiological role of BBIs is crucial, especially for the first developmental stages of the plant (Clemente & Domoney, 2006), contributing to the rise of the innate defensive plant capability, germination success and harvest outcome. In this perspective, de novo domestication approaches could be applied to develop new cultivars with elite traits that improve both the nutritional quality and the resistance to biotic/abiotic stresses (Fernie & Yan, 2019; Renzi et al., 2022; Smỳkal et al., 2018). Regarding human health potential applications, BBIs inhibition is not limited to trypsin and chymotrypsin, but can also interact with other molecular targets (Rawlings

et al., 2004). In particular, BBIs can successfully and selectively interact with protein complexes involved in cancer growth and development, such as proteasome 20S and metalloproteinases 2 and 9 (Fereidunian et al., 2014; Mehdad et al., 2016). Although the mechanism of interaction is still unclear, we hypothesise that some BBI variants identified in this study could exert stronger inhibition on such targets, resulting in improved anticancer activity. However, in vitro and in vivo experiments are consequently needed to appraise the real capacity of the best variants. Although this step is time costly (Cui et al., 2020), it is fundamental to fully characterise the mechanism of action and the target interaction in the models. In the long term these results, if properly confirmed, could open a lot of applications, starting from the use as supplements coming directly from everyday diet but also the adoption of these molecules as drug synergic adjuvants in combination with election drugs (such as cetuximab), reducing or overcoming the drug-resistance phenomena in many tumours.

AUTHOR CONTRIBUTIONS

Conceptualization: Davide Panzeri, Massimo Labra and Fabrizio Grassi. *Methodology*: Davide Panzeri, Elisa Toini, Jacopo Vertemara and Giuseppe Silvestri. *Software*: Davide Panzeri, Jacopo Vertemara, Giuseppe Silvestri and Giuseppe Zampella. *Formal analysis*: Davide Panzeri, Elisa Toini, Jacopo Vertemara, Giuseppe Silvestri, Giovanni Zecca and Fabrizio Grassi. *Investigation*: Davide Panzeri, Elisa Toini, Victor Vladut Bunea, Jacopo Vertemara and Giuseppe Silvestri. *Resources*: Robert Philipp Wagensommer, Giuseppe Zampella, Massimo Labra and Fabrizio Grassi. *Data curation*: Davide Panzeri, Elisa Toini, Jacopo Vertemara and Giuseppe Silvestri. *Writing—original draft*: Davide Panzeri and Fabrizio Grassi. *Writing—review and editing*: Elisa Toini, Victor Vladut Bunea, Jacopo Vertemara, Giovanni Zecca, Werther Guidi Nissim, Robert Philipp Wagensommer and Massimo Labra. *Visualisation*: Davide Panzeri. *Supervision*: Massimo Labra and Fabrizio Grassi. *Project administration*: Davide Panzeri, Massimo Labra and Fabrizio Grassi. *Funding acquisition*: Giuseppe Zampella, Massimo Labra and Fabrizio Grassi.

ACKNOWLEDGEMENTS

We wish to thank Andrea Copetta, Mogahid Sabeel, Rosanna Giacovelli and Caio Chiodi for technical assistance during the experimental phase. We thank Emily Iwu (IITA), Filip Vandelook (Meise Plantentuin) and Alfonso Clemente (CSIC-EEZ) for their support during the gathering of the seed bank accessions. We also wish to thank Penelope Barrington and Dr. Emily Palm for revising the paper.

CONFLICT OF INTEREST STATEMENT

No conflict of interest is declared.

DATA AVAILABILITY STATEMENT

Haplotype sequences and relative translation of BBI-TT and BBI-TC are published on NCBI databank (https://www.ncbi.nlm.nih.gov/). The provided codes are: from OR398407 to OR398430 for BBI-TT and from OR398431 to OR398459 for BBI-TC. All websites in the main text are up to date (21-02-2024).

ORCID

Davide Panzeri https://orcid.org/0000-0003-2672-5846
Elisa Toini https://orcid.org/0000-0001-5668-9976
Jacopo Vertemara https://orcid.org/0000-0002-3352-1000
Giuseppe Silvestri https://orcid.org/0000-0002-7180-9734
Victor Vladut Bunea https://orcid.org/0009-0007-9893-1602
Giovanni Zecca https://orcid.org/0000-0003-2334-8897
Werther Guidi Nissim https://orcid.org/0000-0001-6738-0122
Robert Philipp Wagensommer https://orcid.org/0000-0002-1614-4821
Giuseppe Zampella https://orcid.org/0000-0003-0517-6016
Massimo Labra https://orcid.org/0000-0003-1065-5804
Fabrizio Grassi https://orcid.org/0000-0003-3606-6469

REFERENCES

Abadassi, J. (2015). Cowpea (*Vigna unguiculata* (L.) Walp.) agronomic traits needed in tropical zone. *International Journal of Pure & Applied Bioscience*, *3*(4), 158–165. https://worldveg.tind.io/record/56137/ (access date 21-02-2024)

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001 (access date 21-02-2024)

Anisimova, M., Bielawski, J. P., & Yang, Z. (2002). Accuracy and power of Bayes prediction of amino acid sites under positive selection. *Molecular Biology and Evolution*, *19*, 950–958. https://doi.org/10.1093/oxfordjournals.molbev.a004152 (access date 21-02-2024)

Blench, R. (2003). The movement of cultivated plants between Africa and India in prehistory. *Food, Fuel and Fields: Progress in African Archaeobotany, in Africa Prehistorica 15 (Koln 2003)*, 273–292.

Boukar, O., Belko, N., Chamarthi, S., Togola, A., Batieno, J., Owusu, E., Haruna, M., Diallo, S., Umar, M. L., Olufajo, O., & Fatokun, C. (2019). Cowpea (*Vigna unguiculata*): Genetics, genomics and breeding. *Plant breeding* (Vol. 138) (pp. 415–424). https://doi.org/10.1111/pbr.12589 (access date 21-02-2024)

Boukar, O., Massawe, F., Muranaka, S., Franco, J., Maziya-Dixon, B., Singh, B., & Fatokun, C. (2011). Evaluation of cowpea germplasm lines for protein and mineral concentrations in grains. *Plant Genetic Resources*, *9*, 515–522. https://doi.org/10.1017/s1479262111000815 (access date 21-02-2024)

Bussotti, F., Ferrini, F., Pollastrini, M., & Fini, A. (2014). The challenge of Mediterranean sclerophyllous vegetation under climate change: From acclimation to adaptation. *Environmental and Experimental Botany*, *103*, 80–98. https://doi.org/10.1016/j.envexpbot.2013.09.013 (access date 21-02-2024)

Clemente, A., & del Carmen Arques, M. (2014). Bowman-Birk inhibitors from legumes as colorectal chemopreventive agents. *World Journal of Gastroenterology: WJG*, *20*, 10305–10315. https://doi.org/10.3748/wjg.v20.i30.10305 (access date 21-02-2024)

Clemente, A., & Domoney, C. (2006). Biological significance of polymorphism in legume protease inhibitors from the Bowman-Birk family. *Current Protein and Peptide Science*, *7*, 201–216. https://doi.org/10.2174/138920306777452349 (access date 21-02-2024)

Clemente, A., Gee, J. M., Johnson, I. T., MacKenzie, D. A., & Domoney, C. (2005). Pea (*Pisum sativum* L.) protease inhibitors from the Bowman–Birk class influence the growth of human colorectal adenocarcinoma HT29 cells *in vitro*. *Journal of Agricultural and Food Chemistry*, *53*, 8979–8986. https://doi.org/10.1021/jf051528w (access date 21-02-2024)

Clemente, A., Moreno, F. J., Marín-Manzano, M. D. C., Jiménez, E., & Domoney, C. (2010). The cytotoxic effect of Bowman-Birk isoinhibitors, IBB1 and IBBD2, from soybean (*Glycine max*) on HT29

human colorectal cancer cells is related to their intrinsic ability to inhibit serine proteases. *Molecular Nutrition & Food Research*, *54*, 396–405. https://doi.org/10.1002/mnfr.200900122 (access date 21-02-2024)

Conti, M. V., Guzzetti, L., Panzeri, D., De Giuseppe, R., Coccetti, P., Labra, M., & Cena, H. (2021). Bioactive compounds in legumes: Implications for sustainable nutrition and health in the elderly population. *Trends in Food Science & Technology*, *117*, 139–147. https://doi.org/10.1016/j.tifs.2021.02.072 (access date 21-02-2024)

Coulibaly, S., Pasquet, R. S., Papa, R., & Gepts, P. (2002). AFLP analysis of the phenetic organization and genetic diversity of *Vigna unguiculata* L. Walp. reveals extensive gene flow between wild and domesticated types. *Theoretical and Applied Genetics*, *104*, 358–366. https://doi.org/10.1007/s001220100740 (access date 21-02-2024)

Cui, W., Aouidate, A., Wang, S., Yu, Q., Li, Y., & Yuan, S. (2020). Discovering anti-cancer drugs via computational methods. *Frontiers in Pharmacology*, *11*, 733. https://doi.org/10.3389/fphar.2020.00733 (access date 21-02-2024)

Dakora, F. D., & Belane, A. K. (2019). Evaluation of protein and micronutrient levels in edible cowpea (*Vigna unguiculata* L. Walp.) leaves and seeds. *Frontiers in Sustainable Food Systems*, *3*, 70. https://doi.org/10.3389/fsufs.2019.00070 (access date 21-02-2024)

Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F., & Mark, A. E. (1999). Peptide folding: When simulation meets experiment. *Angewandte Chemie International Edition*, *38*, 236–240. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M (access date 21-02-2024)

Domoney, C. (1999). Inhibitors of legume seeds. In P. R. Shewry & R. Casey (Eds.), *Seed proteins* (pp. 635–655). Springer Netherlands. https://doi.org/10.1007/978-94-011-4431-5_27 (access date 21-02-2024)

Dramé, K. N., Passaquet, C., Repellin, A., & Zuily-Fodil, Y. (2013). Cloning, characterization and differential expression of a Bowman–Birk inhibitor during progressive water deficit and subsequent recovery in peanut (*Arachis hypogaea*) leaves. *Journal of Plant Physiology*, *170*, 225–229. https://doi.org/10.1016/j.jplph.2012.09.005 (access date 21-02-2024)

Excoffier, L., & Lischer, H. E. L. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x (access date 21-02-2024)

Excoffier, L., Smouse, P. E., & Quattro, J. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*, 479–491. https://doi.org/10.1093/genetics/131.2.479 (access date 21-02-2024)

Fereidunian, A., Sadeghalvad, M., Oscoie, M. O., & Mostafaie, A. (2014). Soybean Bowman-Birk protease inhibitor (BBI): Identification of the mechanisms of BBI suppressive effect on growth of two adenocarcinoma cell lines: AGS and HT29. *Archives of Medical Research*, *45*, 455–461. https://doi.org/10.1016/j.arcmed.2014.07.001 (access date 21-02-2024)

Fernie, A. R., & Yan, J. (2019). *De novo* domestication: An alternative route toward new crops for the future. *Molecular Plant*, *12*, 615–631. https://doi.org/10.1016/j.molp.2019.03.016 (access date 21-02-2024)

Gao, F., Chen, C., Arab, D. A., Du, Z., He, Y., & Ho, S. Y. W. (2019). EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecology and Evolution*, *9*, 3891–3898. https://doi.org/10.1002/ece3.5015 (access date 21-02-2024)

Gitlin-Domagalska, A., Maciejewska, A., & Dębowski, D. (2020). Bowman-Birk inhibitors: Insights into family of multifunctional proteins and peptides with potential therapeutical applications. *Pharmaceuticals*, *13*, 421. https://doi.org/10.3390/ph13120421 (access date 21-02-2024)

Guzzetti, L., Fiorini, A., Panzeri, D., Tommasi, N., Grassi, F., Taskin, E., Misci, C., Puglisi, E., Tabaglio, V., Galimberti, A., & Labra, M. (2019). Sustainability perspectives of *Vigna unguiculata* L. Walp. cultivation under no tillage and water stress conditions. *Plants*, *9*, 48. https://doi.org/10.3390/plants9010048 (access date 21-02-2024)

Harouna, D. V., Venkataramana, P. B., Ndakidemi, P. A., & Matemu, A. O. (2018). Under-exploited wild *Vigna* species potentials in human and animal nutrition: A review. *Global Food Security*, *18*, 1–11. https://doi.org/10.1016/j.gfs.2018.06.002 (access date 21-02-2024)

Herniter, I. A., Muñoz-Amatriaín, M., & Close, T. J. (2020). Genetic, textual, and archeological evidence of the historical global spread of cowpea (*Vigna unguiculata* [L.] Walp.). *Legume Science*, *2*, e57. https://doi.org/10.1002/leg3.57 (access date 21-02-2024)

Hilder, V. A., Barker, R. F., Samour, R. A., Gatehouse, A. M. R., Gatehouse, J. A., & Boulter, D. (1989). Protein and cDNA sequences of Bowman-Birk protease inhibitors from the cowpea (*Vigna unguiculata* Walp.). *Plant Molecular Biology*, *13*, 701–710. https://doi.org/10.1007/bf00016025 (access date 21-02-2024)

Horn, L., Shimelis, H., & Laing, M. (2015). Participatory appraisal of production constraints, preferred traits and farming system of cowpea in the northern Namibia: Implications for breeding. *Legume Research: An International Journal*, *38*(5), 691–700.

Huang, J., & MacKerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry*, *34*, 2135–2145. https://doi.org/10.1002/jcc.23354 (access date 21-02-2024)

James, A. M., Jayasena, A. S., Zhang, J., Berkowitz, O., Secco, D., Knott, G. J., Whelan, J., Bond, C. S., & Mylne, J. S. (2017). Evidence for ancient origins of Bowman–Birk inhibitors from *Selaginella moellendorffii*. *The Plant Cell*, *29*, 461–473. https://doi.org/10.1105/tpc.16.00831 (access date 21-02-2024)

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., & Potapenko, A. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2 (access date 21-02-2024)

Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*, 1160–1166. https://doi.org/10.1093/bib/bbx108 (access date 21-02-2024)

Kosakovsky Pond, S. L., & Frost, S. D. (2005). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*, 1208–1222. https://doi.org/10.1093/molbev/msi105 (access date 21-02-2024)

Kuhn, B., & Kollman, P. A. (2000). Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *Journal of Medicinal Chemistry*, *43*, 3786–3791. https://doi.org/10.1021/jm000241h (access date 21-02-2024)

Kumar, P., Sreerama, Y. N., & Gowda, L. R. (2002). Formation of Bowman–Birk inhibitors during the germination of horsegram (*Dolichos biflorus*). *Phytochemistry*, *60*, 581–588. https://doi.org/10.1016/s0031-9422(02)00178-4 (access date 21-02-2024)

Kuraku, S., Zmasek, C. M., Nishimura, O., & Katoh, K. (2013). aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. *Nucleic Acids Research*, *41*, W22–W28. https://doi.org/10.1093/nar/gkt389 (access date 21-02-2024)

Larson, G., Piperno, D. R., Allaby, R. G., Purugganan, M. D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., Dobney, K., Doust, A. N., Gepts, P., Gilbert, M. T. P., Gremillion, K. J., Lucas, L., Lukens, L., Marshall, F. B., Olsen, K. M., Pires, J. C., … Fuller, D. Q. (2014). Current perspectives and the future of domestication studies. *Proceedings of the National Academy of Sciences*, *111*,

6139–6146. https://doi.org/10.1073/pnas.1323964111 (access date 21-02-2024)

Laskowski, M., & Kato, I. (1980). Protein inhibitors of proteinases. *Annual Review of Biochemistry*, *49*, 593–626. https://doi.org/10.1146/annurev.bi.49.070180.003113 (access date 21-02-2024)

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*, 1110–1116. https://doi.org/10.1111/2041-210x.12410 (access date 21-02-2024)

Lima, A. I. G., Mota, J., Monteiro, S., & Ferreira, R. (2016). Legume seeds and colorectal cancer revisited: Protease inhibitors reduce MMP-9 activity and colon cancer cell migration. *Food Chemistry*, *197*, 30–38. https://doi.org/10.1016/j.foodchem.2015.10.063 (access date 21-02-2024)

Mehdad, A., Brumana, G., Souza, A. A., Barbosa, J., Ventura, M. M., & De Freitas, S. M. (2016). A Bowman–Birk inhibitor induces apoptosis in human breast adenocarcinoma through mitochondrial impairment and oxidative damage following proteasome 20S inhibition. *Cell Death Discovery*, *2*, 15067. https://doi.org/10.1038/cddiscovery.2015.67 (access date 21-02-2024)

Mello, M. O., Tanaka, A. S., & Silva-Filho, M. C. (2003). Molecular evolution of Bowman–Birk type proteinase inhibitors in flowering plants. *Molecular Phylogenetics and Evolution*, *27*, 103–112. https://doi.org/10.1016/s1055-7903(02)00373-1 (access date 21-02-2024)

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 gateway computing environments workshop (GCE)* (pp. 1–8). IEEE. https://doi.org/10.1109/gce.2010.5676129 (access date 21-02-2024)

Muñoz-Amatriaín, M., Lo, S., Herniter, I. A., Boukar, O., Fatokun, C., Carvalho, M., Castro, I., Guo, Y., Huynh, B., Roberts, P. A., et al. (2021). The UCR Minicore: A resource for cowpea research and breeding. *Legume Science*, *3*, e95. https://doi.org/10.1002/leg3.95 (access date 21-02-2024)

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*, 1196–1205. https://doi.org/10.1093/molbev/mst030 (access date 21-02-2024)

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*, e1002764. https://doi.org/10.1371/journal.pgen.1002764 (access date 21-02-2024)

Nielsen, R., & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics*, *148*, 929–936. https://doi.org/10.1093/genetics/148.3.929 (access date 21-02-2024)

Olías, R., Becerra-Rodríguez, C., Soliz-Rueda, J. R., Moreno, F. J., Delgado-Andrade, C., & Clemente, A. (2019). Glycation affects differently the main soybean Bowman–Birk isoinhibitors, IBB1 and IBBD2, altering their antiproliferative properties against HT29 colon cancer cells. *Food & Function*, *10*, 6193–6202. https://doi.org/10.1039/c9fo01421g (access date 21-02-2024)

Padulosi, S., & Ng, N. Q. (1997). Origin, taxonomy, and morphology of *Vigna unguiculata* (L.) Walp. *Advances in Cowpea Research*, 1, 1–12. https://hdl.handle.net/10568/95988 (access date 21-02-2024)

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In S. Markidis & E. Laure (Eds.), *Lecture notes in computer science. Solving software challenges for exascale* (pp. 3–27). Springer International Publishing. https://doi.org/10.1007/978-3-319-15976-8_1 (access date 21-02-2024)

Pandey, A., Yadav, R., & Sanyal, I. (2022). Evaluating the pesticidal impact of plant protease inhibitors: Lethal weaponry in the co-evolutionary battle. *Pest Management Science*, *78*, 855–868. https://doi.org/10.1002/ps.6659 (access date 21-02-2024)

Panella, L., & Gepts, P. (1992). Genetic relationships within *Vigna unguiculata* (L.) Walp. based on isozyme analyses. *Genetic Resources and Crop Evolution*, *39*, 71–88. https://doi.org/10.1007/bf00051226 (access date 21-02-2024)

Panzeri, D. (2023). *A bioprospecting multidisciplinary approach to valorise biodiversity: The case of Bowman-Birk protease inhibitors in Vigna unguiculata (L.) Walp*. PhD Thesis. University of Milan-Bicocca. https://hdl.handle.net/10281/404605 (access date 21-02-2024)

Panzeri, D., Guidi Nissim, W., Labra, M., & Grassi, F. (2022). Revisiting the domestication process of African *Vigna* species (Fabaceae): Background, perspectives and challenges. *Plants*, *11*, 532. https://doi.org/10.3390/plants11040532 (access date 21-02-2024)

Panzeri, D., Guzzetti, L., Sacco, G., Tedeschi, G., Nonnis, S., Airoldi, C., Labra, M., Fusi, P., Forcella, M., & Regonesi, M. E. (2020). Effectiveness of *Vigna unguiculata* seed extracts in preventing colorectal cancer. *Food & Function*, *11*, 5853–5865. https://doi.org/10.1039/d0fo00913j (access date 21-02-2024)

Pasquet, R. and Padulosi, S. (2013). Genus *Vigna* and cowpea (*V. unguiculata* [L.] Walp.) taxonomy: Current status and prospects. In *Innovative research along the cowpea value chain (pp. 66–87). Proceedings of the fifth world cowpea conference on improving livelihoods in the cowpea value chain through advancement in science*. Saly (Senegal), 27 Sep-01 Oct 2010, IITA. https://hdl.handle.net/20.500.11766/5163 (access date 21-02-2024)

Pasquet, R. S. (1996). Wild cowpea (*Vigna unguiculata*) evolution. *Advances in Legume Systematics*, *8*, 95–100.

Pasquet, R. S. (1997). A new subspecies of *Vigna unguiculata* (Leguminosae: Papilionoideae). *Kew Bulletin*, *52*, 840–840. https://doi.org/10.2307/4117815

Pasquet, R. S. (1999). Genetic relationships among subspecies of *Vigna unguiculata* (L.) Walp. based on allozyme variation. *Theoretical and Applied Genetics*, *98*, 1104–1119. https://doi.org/10.1007/s001220051174 (access date 21-02-2024)

Pasquet, R. S., Feleke, Y., & Gepts, P. (2021). Cowpea [*Vigna unguiculata* (L.) Walp.] maternal lineages, chloroplast captures, and wild cowpea evolution. *Genetic Resources and Crop Evolution*, *68*, 2799–2812. https://doi.org/10.1007/s10722-021-01155-y (access date 21-02-2024)

Poon, A. F. Y., Frost, S. D. W., & Pond, S. L. K. (2009). Detecting signatures of selection from DNA sequences using Datamonkey. In D. Posada (Ed.), *Methods in molecular biology. Bioinformatics for DNA sequence analysis* (pp. 163–183). Humana Press. https://doi.org/10.1007/978-1-59745-251-9_8 (access date 21-02-2024)

Ramanatha Rao, V., & Hodgkin, T. (2002). Genetic diversity and conservation and utilization of plant genetic resources. *Plant Cell, Tissue and Organ Culture*, *68*, 1–19. https://doi.org/10.1023/A:1013359015812 (access date 21-02-2024)

Rawlings, N. D., Tolle, D. P., & Barrett, A. J. (2004). Evolutionary families of peptidase inhibitors. *Biochemical Journal*, *378*, 705–716. https://doi.org/10.1042/bj20031825 (access date 21-02-2024)

Renzi, J. P., Coyne, C. J., Berger, J., von Wettberg, E., Nelson, M., Ureta, S., Hernández, F., Smỳkal, P., & Brus, J. (2022). How could the use of crop wild relatives in breeding increase the adaptation of crops to marginal environments? *Frontiers in Plant Science*, *13*, 886162. https://doi.org/10.3389/fpls.2022.886162 (access date 21-02-2024)

Rozas, J., Ferrer-Mata, A., Sánchez-Del Barrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*, 3299–3302. https://doi.org/10.1093/molbev/msx248 (access date 21-02-2024)

Sánchez-Chino, X., Jiménez-Martínez, C., Dávila-Ortiz, G., Álvarez-González, I., & Madrigal-Bujaidar, E. (2015). Nutrient and nonnutrient components of legumes, and its chemopreventive activity: A review. *Nutrition and Cancer*, *67*, 401–410. https://doi.org/10.1080/01635581.2015.1004729 (access date 21-02-2024)

Schecter, I., & Berger, A. (1967). On the size of the active site in proteases. I. Papain. *Biochemical and Biophysical Research Communications*, *27*, 157–162. https://doi.org/10.1016/s0006-291x(67)80055-x (access date 21-02-2024)

Shan, L., Li, C., Chen, F., Zhao, S., & Xia, G. (2008). A Bowman-Birk type protease inhibitor is involved in the tolerance to salt stress in wheat. *Plant, Cell & Environment*, *31*, 1128–1137. https://doi.org/10.1111/j.1365-3040.2008.01825.x (access date 21-02-2024)

Smýkal, P., Nelson, M. N., Berger, J. D., & Von Wettberg, E. J. (2018). The impact of genetic changes during crop domestication. *Agronomy*, *8*, 119. https://doi.org/10.3390/agronomy8070119 (access date 21-02-2024)

Srikanth, S., & Chen, Z. (2016). Plant protease inhibitors in therapeutics-focus on cancer therapy. *Frontiers in Pharmacology*, *7*, 470. https://doi.org/10.3389/fphar.2016.00470 (access date 21-02-2024)

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033 (access date 21-02-2024)

Suzuki, Y., & Gojobori, T. (1999). A method for detecting positive selection at single amino acid sites. *Molecular Biology and Evolution*, *16*, 1315–1328. https://doi.org/10.1093/oxfordjournals.molbev.a026042 (access date 21-02-2024)

Tamura, K., Stecher, G., & Kumar, S. (2021). MEGA11: Molecular evolutionary genetics analysis version 11. *Molecular Biology and Evolution*, *38*, 3022–3027. https://doi.org/10.1093/molbev/msab120 (access date 21-02-2024)

Tripodi, F., Lombardi, L., Guzzetti, L., Panzeri, D., Milanesi, R., Leri, M., Bucciantini, M., Angeloni, C., Beghelli, D., Hrelia, S., Onorato, G., Di Schiavi, E., Falletta, E., Nonnis, S., Tedeschi, G., Labra, M., & Coccetti, P. (2020). Protective effect of *Vigna unguiculata* extract against aging and neurodegeneration. *Aging (Albany NY)*, *12*, 19785. https://doi.org/10.18632/aging.104069 (access date 21-02-2024)

Weaver, S., Shank, S. D., Spielman, S. J., Li, M., Muse, S. V., & Kosakovsky Pond, S. L. (2018). Datamonkey 2.0: A modern web application for characterizing selective and other evolutionary processes. *Molecular Biology and Evolution*, *35*, 773–777. https://doi.org/10.1093/molbev/msx335 (access date 21-02-2024)

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641 (access date 21-02-2024)

Xiong, H., Shi, A., Mou, B., Qin, J., Motes, D., Lu, W., Ma, J., Weng, Y., Yang, W., & Wu, D. (2016). Genetic diversity and population structure of cowpea (*Vigna unguiculata* L Walp). *PLoS ONE*, *11*, e0160941. https://doi.org/10.1371/journal.pone.0160941 (access date 21-02-2024)

Xu, P., Wu, X., Wang, B., Luo, J., Liu, Y., Ehlers, J. D., Close, T. J., Roberts, P. A., Lu, Z., & Wang, S. (2012). Genome wide linkage disequilibrium in Chinese asparagus bean (*Vigna unguiculata* ssp. *sesquipedialis*) germplasm: Implications for domestication history and genome wide association studies. *Heredity*, *109*, 34–40. https://doi.org/10.1038/hdy.2012.8 (access date 21-02-2024)

Yang, Z., & Nielsen, R. (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Molecular Biology and Evolution*, *19*, 908–917. https://doi.org/10.1093/oxfordjournals.molbev.a004148 (access date 21-02-2024)

Yang, Z., Wong, W. S., & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Molecular Biology and Evolution*, *22*, 1107–1118. https://doi.org/10.1093/molbev/msi097 (access date 21-02-2024)

Clement, M., Snell, Q., Walke, P., Posada, D., & Crandall, K. (2002). TCS: estimating gene genealogies. *Proceedings 16th International Parallel and Distributed Processing Symposium*. https://doi.org/10.1109/ipdps.2002.1016585

SUPPORTING INFORMATION

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

How to cite this article: Panzeri, D., Toini, E., Vertemara, J., Silvestri, G., Bunea, V. V., Zecca, G., Nissim, W. G., Wagensommer, R. P., Zampella, G., Labra, M., & Grassi, F. (2024). Small proteins, great promises: Geographic bioprospecting of Bowman–Birk protease inhibitors and domestication side-effects in African cowpea (*Vigna unguiculata* L.). *Plants, People, Planet*, 1–14. https://doi.org/10.1002/ppp3.10507