frontiers | Frontiers in Horticulture

TYPE Technology and Code
PUBLISHED 11 December 2025
DOI 10.3389/fhort.2025.1686134

OPEN ACCESS

EDITED BY
Jose G. Vallarino,
University of Malaga, Spain

REVIEWED BY
Aamir W. Khan,
The University of Missouri, United States
Wang Xiukang,
Yan'an University, China

*CORRESPONDENCE
Christina Stonoha-Arther
christina.arther@usda.gov

†These authors share first authorship

RECEIVED 14 August 2025
REVISED 19 November 2025
ACCEPTED 25 November 2025
PUBLISHED 11 December 2025

CITATION
Stonoha-Arther C and Panke-Buisse K (2025) resido: an R package for exploring amino acid residue composition of peptide and protein sequences.
Front. Hortic. 4:1686134.
doi: 10.3389/fhort.2025.1686134

resido: an R package for exploring amino acid residue composition of peptide and protein sequences

Christina Stonoha-Arther*† and Kevin Panke-Buisse†

U.S. Dairy Forage Research Center, USDA-ARS, Madison, WI, United States

Amino acid composition may be used in synthetic biology for fine-tuning amino acid content of plants to improve human or animal diets. One barrier to plant-based diet adoption is insufficiency of key amino acids such as methionine and lysine. Legumes, for example, are a good source of dietary protein, but lack sufficient amounts of these essential amino acids. Despite the potential utility of exploring amino acid content in proteomes or subsets of sequences, purpose-built bioinformatic tools are largely lacking. Here, we present an R package, resido, that facilitates the characterization and discovery of proteins based on amino acid content. We used resido to investigate the sulfur-containing amino acids of white lupin (*Lupinus albus* L.). White lupin has the potential to be a valuable component of plant-based diets, although there are some drawbacks, one being the lack of sulfur-containing amino acids in the seeds. Using resido, we identified several protein sequences that had a high percentage of sulfur amino acids; one collagen-like protein was 25% methionine. Overall, resido is a straightforward bioinformatic tool that leverages the ubiquity of R to analyze amino acid content of user-supplied peptide sequences and may be useful to efforts modulating overall amino acid content of crops for refining nutrition in animal or human diets.

KEYWORDS

bioinformatics and computational biology, legumes (Fabaceae), amino acids (AA), pulses, R package, lupin (*Lupinus albus* L.)

1 Introduction

Amino acid content in human and animal diets is important for overall nutrition, especially when those diets consist of mostly plants. Humans, for example, cannot synthesize nine essential amino acids and therefore must obtain these from dietary sources. Providing all nine essential amino acids in the amounts required for human nutrition from single-plant sources is currently impractical. Plant-based diets must be managed in order to ensure that all of these essential amino acids are available for protein synthesis.

Legumes are a staple source of protein in both human and animal diets, but lack certain amino acids such as methionine and lysine, which must be supplemented or obtained from other sources. Much research has been done in this area regarding dairy cow diets. Several studies and meta-analyses have shown that supplemental methionine and lysine in dairy cow diets increases milk protein (Vyas and Erdman, 2009; Patton, 2010; Zanton et al., 2014). Milk yield has also been shown to increase with added methionine, but is somewhat dependent on the source (Patton, 2010; Zanton et al., 2014). Therefore, some work has been done, using various techniques, to modify the amino acid content of legumes to improve their nutritional profile (Avraham et al., 2005; Guo et al., 2020; Rushovich and Weil, 2021).

White lupin (*Lupinus albus* L.) is a pulse that has been traditionally grown in the Mediterranean region for centuries and is gaining popularity in other parts of the world such as Australia and Canada. It has the potential to be a valuable component of plant-based diets, although the paucity of sulfur (S)-containing amino acids in the seeds is a limitation (Duranti and Cerletti, 1979). Much of the previous research in this area on white lupin and other legumes has relied on transgenic overexpression of a methionine-rich albumin protein from sunflower (Kortt and Caldwell, 1990; Tabe and Droux, 2002; Girija et al., 2020). This protein was discovered to be methionine-rich (16%) incidentally during a biochemical characterization of the proteins within sunflower seeds (Kortt and Caldwell, 1990). Some of the drawbacks of overexpressing this albumin for increasing methionine content is that, at least in legumes, the resulting plant will always be classified as transgenic, and the methionine content of the overexpressed protein is relatively low.

It may be possible to achieve increased methionine levels in legumes through cisgenic overexpression of native methionine-rich genes or motifs. However, none of the existing bioinformatic tools, including ExPASy ProtParam and EMBOSS pepstats, for identifying and categorizing these proteins are available as an R-native workflow (Gasteiger et al., 2005; Cock et al., 2009; Madeira et al., 2024). To leverage the flexibility and versatility of R, we developed resido, an R package that facilitates the characterization, classification, and visualization of peptide sequences based on amino acid content. We used resido to explore the white lupin proteome (v1) for proteins rich in the S-containing amino acids, methionine and cysteine.

2 Methods

2.1 Purpose

The R package, resido, was designed to characterize, sort, and visualize the amino acid content of peptide and protein sequences from user-input fasta files or from a character string of a single sequence (Figure 1).

2.2 Features

resido was created in R statistical software (R Core Team, 2024). There are eight individual resido-specific functions that are included in the package. In addition, there are four functions imported from other R packages, and six other packages which are imported dependencies (Figure 1; Supplementary File 2). The definitions of the amino acids and amino acid groups that are used for various resido functions are listed in Supplementary File 3. AI

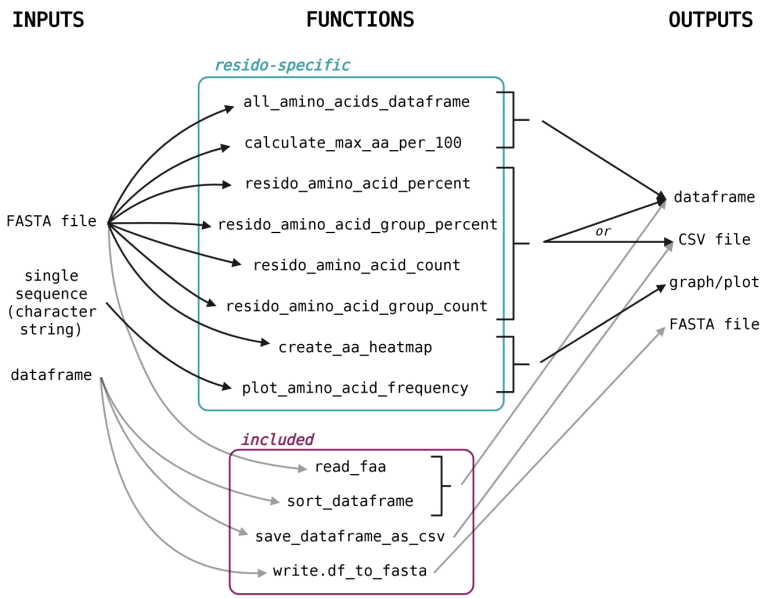

FIGURE 1
Overview of the inputs, functions, and outputs of resido. Figure created in https://BioRender.com.

TABLE 1 Comparison of total percent methionine and maximum percent methionine per 100 amino acids in a selection of methionine-rich proteins from this study and previous studies.

Protein	Percent methionine	Maximum percent methionine per 100 amino acids
Collagen-like protein from lupin (Lalb_Chr03g0040411)	25.20	32.00
Unknown protein from lupin (Lalb_Chr18g0049981)	21.43	22.00
ABC-like transporter from lupin (Lalb_Chr14g0367601)	18.31	18.31
Seed albumin from sunflower (HanXRQChr11g0337861)	12.77	16.00
β-zein from maize (GenBank: AAA33543.1)	11.11	14.00

tools were used for expediting code generation, but all generated code was debugged, added manually, and tested by the authors.

The functions 'all_amino_acid_dataframe' and 'calculate_max_aa_per_100' take a user-supplied fasta file of amino acid sequences and returns a dataframe as the output. The former function returns the proportion of every canonical amino acid for each sequence in the fasta file. The latter-mentioned function returns the highest proportion of an amino acid of interest within any 100 amino acids per sequence in the fasta file. If a protein sequence is shorter than 100 amino acids, this function will return the percent amino acid for the entire length of the protein (see the ABC-like transporter from lupin (Lalb_Chr14g0367601) in Table 1).

The 'create_aa_heatmap' and 'plot_amino_acid_frequency' functions both have visual outputs that rely on ggplot2 (Supplementary File 2). 'create_aa_heatmap' makes a heatmap of the amino acid frequency of each sequence within a user-input fasta file. 'plot_amino_acid_frequency' takes a single character string as the input and plots the amino acid frequencies as a bar chart.

The following functions take a user-supplied fasta file as the input, along with an amino acid or amino acid group of interest and returns a dataframe (default), or a csv file: 'resido_amino_acid_percent,' 'resido_amino_acid_group_percent,' 'resido_amino_acid_count,' 'resido_amino_acid_group_count.'

2.3 White lupin proteome

The white lupin proteome was downloaded and used for the initial resido input (Hufnagel et al., 2020). The fasta file was used as the input for the function 'resido_amino_acid_group_percent' with the amino acid group of interest designated as 'S' (sulfur-containing amino acid group). The resulting dataframe was sorted in decreasing order from highest percent S-containing amino acids to lowest percent per sequence. The top ten highest S-containing sequences were printed into a new fasta file using the included function 'write.df_to_fasta' (Vijayan and Sreekumar, 2023). This new fasta file was used as the input for 'create_aa_heatmap.' Percent cysteine and percent methionine were calculated in these top-ten S-containing amino acids protein sequences using 'resido_amino_acid_percent.' The number one ranked S-containing amino acids sequence and the number three ranked sequence (Lalb_Chr14g0367601 and Lalb_Chr03g0040411, respectively) were used as the input for 'plot_amino_acid_frequency' function. Lalb_Chr03g0040411 was also used as the input for 'calculate_max_aa_per_100.' All functions listed above ran to completion in less than one second.

2.4 Availability and use

resido is available and maintained on GitHub and Zenodo by the authors (https://github.com/cstonoha/resido; https://doi.org/10.5281/zenodo.15802550). There is a detailed README file available with the package and there is an R Markdown file supplementary to this manuscript (Supplementary File 1). resido is freely available under a Creative Commons license and will be updated by the authors for added functionality as needed.

3 Results/discussion

We used resido to identify and characterize proteins within the proteome of white lupin that had a relatively high percent of S-containing amino acids. A fasta file of all of the sequences that make up the white lupin proteome was used to identify the top ten sequences with the highest percent S-containing amino acids. This list of ten was then written to a fasta file and used as the subsequent input in order to visualize the amino acid densities in a heatmap (Figure 2).

The heatmap indicated that the densities of all the amino acids varied widely between these sequences, even those of cysteine and methionine. One sequence, a collagen-like protein, Lalb_Chr03g0040411, had relatively high methionine, but no cysteine. We then characterized sequences by percent of cysteine and methionine separately using the resido function 'resido_amino_acid_percent' (Table 2). This supported the findings from the heatmap visualization regarding Lalb_Chr03g0040411; it has no cysteine residues but contains over 25% methionine. On the other hand, the top S-containing amino acid sequence, Lalb_Chr14g0367601, had similar methionine and cysteine content (18.31% and 19.70%, respectively).

Furthermore, using the resido function, 'calculate_max_aa_per_100,' we found that the Lalb_Chr03g0040411 sequence had a 100-amino acid stretch that contained 32% methionine (Table 1). We then plotted the frequency of each amino acid within these two sequences for comparison (Figure 3), confirming observations in Figure 2 and Table 2: Lalb_Chr14g0367601 is rich in methionine and cysteine and Lalb_Chr03g0040411 has high glycine, methionine, and proline, but no cysteine.

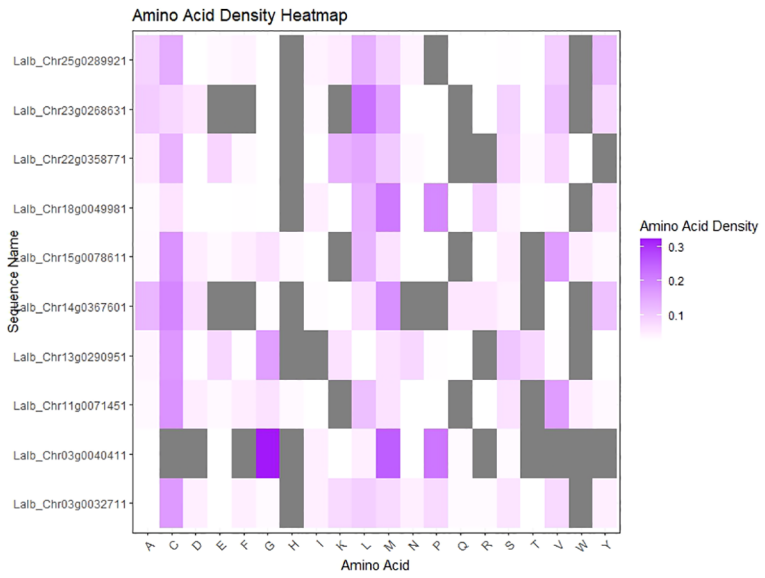

FIGURE 2
Amino acid density heatmap, generated with resido, of the top-ten highest percent of sulfur-containing amino acids protein sequences from white lupin. Shades of purple indicate amino acid frequency/density; white indicates a frequency of 0.0 and dark purple indicates a frequency of 0.3, which is the highest density in this subset of sequences. Gray boxes indicate the amino acid is not present in the sequence.

Using resido, we identified three lupin proteins that have higher methionine content than the often-used sunflower albumin protein (Tables 1, 2). These proteins, or specific motifs within these proteins, may be interesting candidates for overexpression with the goal of cisgenically increasing methionine content in the seeds.

Comparing the methionine content between three of the proteins found in the white lupin proteome and two proteins that have been used for increasing methionine content in plants (seed albumin from sunflower and β-zein from maize) showed that the three proteins identified with resido have higher overall methionine content and higher percentage of methionine within 100 amino acids (Table 1) (Duranti and Cerletti, 1979; Kortt and Caldwell, 1990; Guo et al., 2020). Furthermore, this provides examples for testing the accuracy of resido: the β-zein protein is the same one that was used to increase methionine content in soybean and the methionine content that the authors calculated was the same as what we report here (Guo et al., 2020).

TABLE 2 Percent cysteine and methionine of the top-ten S-containing amino acids protein sequences in white lupin.

Rank	Sequence name	Percent methionine	Percent cysteine
1	Lalb_Chr14g0367601	18.31	19.72
2	Lalb_Chr18g0049981	21.43	6.35
3	Lalb_Chr03g0040411	25.20	0.00
4	Lalb_Chr03g0032711	7.69	16.92
5	Lalb_Chr11g0071451	6.56	18.03
6	Lalb_Chr15g0078611	6.56	18.03
7	Lalb_Chr13g0290951	6.67	17.33
8	Lalb_Chr23g0268631	15.32	8.06
9	Lalb_Chr22g0358771	10.00	13.33
10	Lalb_Chr25g0289921	8.62	14.66

Legumes in general have relatively low methionine levels compared to their total protein. This may be because legume nodules require sulfur in order to function and fix nitrogen; leaving less sulfur for S-containing amino acids in the aboveground portion of the plant (Becana et al., 2018). Lupin has especially low methionine levels possibly because it doesn't form arbuscular mycorrhizal (AM) associations or at least has lost the AM-specific genes required to maintain these associations (Delaux et al., 2014). Perhaps this lack of AM fungi affects the amount of sulfur that is taken up by the roots, making sulfur less available for the plant (Giovannetti et al., 2014; Sieh et al., 2013). This is important because overexpression of a methionine-rich protein would most likely also require sulfur supplementation in order to produce increased methionine levels in the seeds or leaves.

resido has many applications for plant synthetic biology and addressing fundamental plant biology questions. For example, it may be helpful for discovering and identifying classes of peptides or proteins that are partly defined based on their amino acid content. For example, nodule-specific cysteine-rich peptides are small, defensin-like peptides that typically have four to six cysteine residues. These peptides are used by some legumes to coax intracellular rhizobia within the nodules to terminally differentiate and ultimately to fix nitrogen (Stonoha-Arther and Wang, 2018). Identifying these peptides is broadly useful in various studies, from molecular biology and genetics to ecology and evolutionary biology.

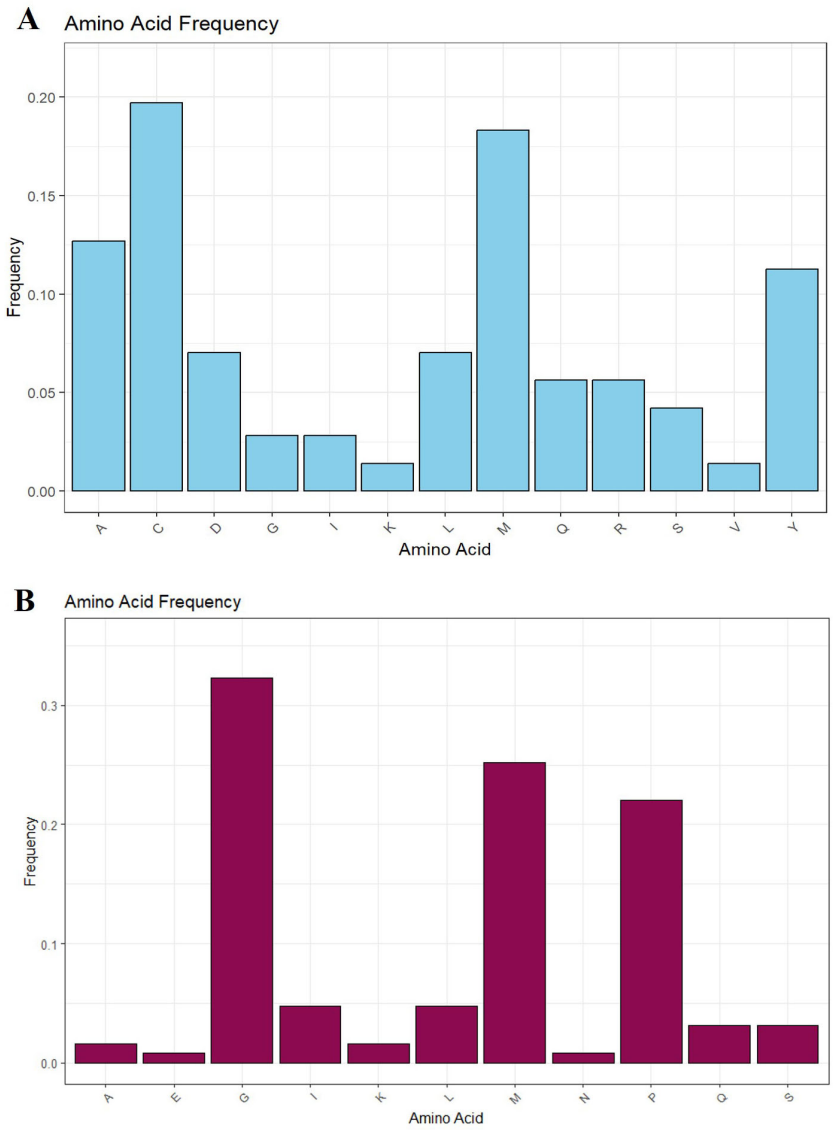

FIGURE 3
Amino acid frequency plots, generated with resido, of two white lupin protein sequences. **(A)** The white lupin protein sequence with the highest percent S-containing amino acids, Lalb_Chr03g0040411. **(B)** The white lupin protein sequence with the highest percent methionine within the highest percent S-containing amino acids sequences, Lalb_Chr14g0367601 (color modified from resido output).

While our interests and foci are predominantly aimed at plant proteins, resido as an application is useful for all domains of life. Amino acid composition of peptides, proteins, proteomes, or entire clades may reveal important underlying themes that affect protein evolution and amino acid usage (Hormoz, 2013; Morimoto and Pietras, 2024).

4 Conclusion

The R package presented here, resido, is a straightforward and broadly useful bioinformatic tool for protein characterization, identification, and categorization. One application for this tool is identifying proteins for possible overexpression in crop plants to modify amino acid content, but resido can be applied to any peptide or protein sequence across all domains of life.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author/s. The R package is available here: https://github.com/cstonoha/resido, https://doi.org/10.5281/zenodo.15802550.

Author contributions

CS-A: Methodology, Supervision, Investigation, Conceptualization, Software, Writing – review & editing, Project administration, Writing – original draft. KP-B: Validation, Conceptualization, Methodology,

Project administration, Supervision, Investigation, Writing – original draft, Software, Writing – review & editing.

Funding

The author(s) declared financial support was received for the research and/or publication of this article. This work was funded through congressional allocation to USDA ARS project 5090-21500-001-000D.

Conflict of interest

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

Generative AI statement

The author(s) declare that Generative AI was used in the creation of this manuscript. Google Gemini version 1.5 was used for some coding and editing of the R scripts within the resido package.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fhort.2025.1686134/full#supplementary-material

References

Avraham, T., Badani, H., Galili, S., and Amir, R. (2005). Enhanced levels of methionine and cysteine in transgenic alfalfa (Medicago sativa L.) plants over-expressing the Arabidopsis cystathionine γ-synthase gene. *Plant Biotechnol. J.* 3, 71–79. doi: 10.1111/j.1467-7652.2004.00102.x

Becana, M., Wienkoop, S., and Matamoros, M. A. (2018). Sulfur transport and metabolism in legume root nodules. *Front. Plant Sci.* 9. doi: 10.3389/fpls.2018.01434

Cock, P. J. A., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25, 1422–1423. doi: 10.1093/bioinformatics/btp163

Delaux, P.-M., Varala, K., Edger, P. P., Coruzzi, G. M., Pires, J. C., and Ané, J.-M. (2014). Comparative phylogenomics uncovers the impact of symbiotic associations on host genome evolution. *PloS Genet.* 10, e1004487. doi: 10.1371/journal.pgen.1004487

Duranti, M., and Cerletti, P. (1979). Amino acid composition of seed proteins of Lupinus albus. *J. Agric. Food Chem.* 27, 977–978. doi: 10.1021/jf60225a038

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein Identification and Analysis Tools on the ExPASy Server," in *The Proteomics Protocols Handbook*. Ed. J. M. Walker (Humana Press, Totowa, NJ), 571–607. doi: 10.1385/1-59259-890-0:571

Giovannetti, M., Tolosano, M., Volpe, V., Kopriva, S., and Bonfante, P. (2014). Identification and functional characterization of a sulfate transporter induced by both sulfur starvation and mycorrhiza formation in Lotus japonicus. *New Phytol.* 204, 609–619. doi: 10.1111/nph.12949

Girija, A., Shotan, D., Hacham, Y., and Amir, R. (2020). The level of methionine residues in storage proteins is the main limiting factor of protein-bound-methionine accumulation in arabidopsis seeds. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.01136

Guo, C., Liu, X., Chen, L., Cai, Y., Yao, W., Yuan, S., et al. (2020). Elevated methionine content in soybean seed by overexpressing maize β-zein protein. *Oil Crop Sci.* 5, 11–16. doi: 10.1016/j.ocsci.2020.03.004

Hormoz, S. (2013). Amino acid composition of proteins reduces deleterious impact of mutations. *Sci. Rep.* 3, 2919. doi: 10.1038/srep02919

Hufnagel, B., Marques, A., Soriano, A., Marquès, L., Divol, F., Doumas, P., et al. (2020). High-quality genome sequence of white lupin provides insight into soil exploration and seed quality. *Nat. Commun.* 11, 492. doi: 10.1038/s41467-019-14197-9

Kortt, A. A., and Caldwell, J. B. (1990). Low molecular weight albumins from sunflower seed: identification of a methionine-rich albumin. *Phytochemistry* 29, 2805–2810. doi: 10.1016/0031-9422(90)87080-E

Madeira, F., Madhusoodanan, N., Lee, J., Eusebi, A., Niewielska, A., Tivey, A. R. N., et al. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. *Nucleic Acids Res.* 52, W521–W525. doi: 10.1093/nar/gkae241

Morimoto, J., and Pietras, Z. (2024). Differential amino acid usage leads to ubiquitous edge effect in proteomes across domains of life that can be explained by amino acid secondary structure propensities. *Sci. Rep.* 14, 25544. doi: 10.1038/s41598-024-77319-4

Patton, R. A. (2010). Effect of rumen-protected methionine on feed intake, milk production, true milk protein concentration, and true milk protein yield, and the factors that influence these effects: A meta-analysis. *J. Dairy Sci.* 93, 2105–2118. doi: 10.3168/jds.2009-2693

R Core Team (2024). *R: The R Project for Statistical Computing*. Available online at: https://www.r-project.org/ (Accessed December 19, 2024).

Rushovich, D., and Weil, R. (2021). Sulfur fertility management to enhance methionine and cysteine in soybeans. *J. Sci. Food Agric.* 101, 6595–6601. doi: 10.1002/jsfa.11307

Sieh, D., Watanabe, M., Devers, E. A., Brueckner, F., Hoefgen, R., and Krajinski, F. (2013). The arbuscular mycorrhizal symbiosis influences sulfur starvation responses of Medicago truncatula. *New Phytologist* 197, 606–616. doi: 10.1111/nph.12034

Stonoha-Arther, C., and Wang, D. (2018). Tough love: accommodating intracellular bacteria through directed secretion of antimicrobial peptides during the nitrogen-fixing symbiosis. *Curr. Opin. Plant Biol.* 44, 155–163. doi: 10.1016/j.pbi.2018.04.017

Tabe, L. M., and Droux, M. (2002). Limits to sulfur accumulation in transgenic lupin seeds expressing a foreign sulfur-rich protein. *Plant Physiol.* 128, 1137–1148. doi: 10.1104/pp.010935

Vijayan, A., and Sreekumar, J. (2023). *baseq: Basic Sequence Processing Tool for Biological Data*. Available online at: https://github.com/ambuvjyn/baseq (Accessed August 14, 2025).

Vyas, D., and Erdman, R. A. (2009). Meta-analysis of milk protein yield responses to lysine and methionine supplementation. *J. Dairy Sci.* 92, 5011–5018. doi: 10.3168/jds.2008-1769

Zanton, G. I., Bowman, G. R., Vázquez-Añón, M., and Rode, L. M. (2014). Meta-analysis of lactation performance in dairy cows receiving supplemental dietary methionine sources or postruminal infusion of methionine. *J. Dairy Sci.* 97, 7085–7101. doi: 10.3168/jds.2014-8220