Bacterial amino acid auxotrophies enable energetically costlier proteomes

Niko Kasalo[1], Tomislav Domazet-Lošo[1,2,*] and Mirjana Domazet-Lošo[3,*]

[1]Laboratory of Evolutionary Genetics, Division of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, HR-10000 Zagreb, Croatia

[2]School of Medicine, Catholic University of Croatia, Ilica 242, HR-10000 Zagreb, Croatia

[3]Department of Applied Computing, Faculty of Electrical Engineering and Computing, University of Zagreb, Unska 3, HR-10000 Zagreb, Croatia

*Correspondence: T.D.-L. tdomazet@irb.hr; M. D.-L. mirjana.domazet@fer.hr

Abstract

The outsourcing of amino acid (AA) production to the environment is relatively common across the tree of life. We recently showed that the massive loss of AA synthesis capabilities in animals is governed by selective pressure linked to the energetic costs of AA production. Paradoxically, these AA auxotrophies facilitated the evolution of costlier proteomes in animals by enabling the increased use of energetically expensive AAs. Experiments in bacteria have shown that AA auxotrophies can provide a fitness advantage in competition with prototrophic strains. However, it remains unclear whether energy-related selection also drives the evolution of bacterial AA auxotrophies and whether this affects the usage of expensive AAs in bacterial proteomes. To investigate these questions, we computationally determined AA auxotrophy odds across 980 bacterial genomes representing diverse taxa and calculated the energy costs of all their proteins. Here, we show that auxotrophic AAs are generally more expensive to synthesize than prototrophic AAs in bacteria. Moreover, we found that the cost of auxotrophic AAs significantly correlates with the cost of their respective proteomes. Interestingly, out of all considered taxa, *Mollicutes* and *Borreliaceae*—chronic pathogens highly successful in

immune evasion—have the most AA auxotrophies and code for the most expensive proteomes. These findings indicate that AA auxotrophies in bacteria, similar to those in animals, are shaped by selective pressures related to energy management. Our study highlights bacterial AA auxotrophies as costly outsourced functions that allowed bacteria to more freely explore protein sequence space. It remains to be investigated whether this relaxed use of expensive AAs also enabled auxotrophic bacteria to evolve proteins with improved or novel functionality.

Keywords: bacteria; amino acids; auxotrophy; energetics; synthesis; selection; evolution; functional outsourcing

Introduction

The concept of functional outsourcing posits that organisms streamline their genomes by losing costly functions that can be substituted through external biological interactions [1]. This principle clearly applies to the outsourcing of amino acid (AA) production, a phenomenon observed widely across the tree of life [1–5]. A particularly striking example is animals, which almost universally lack the ability to synthesize about half of the proteinogenic AAs [2,5]. Recent research demonstrates that this phenomenon is driven by selective pressure related to the high energy cost of AA production, enabling animals to evolve costlier proteomes that incorporate expensive AAs more frequently [5]. These findings raise the possibility that similar correlations between proteome energetics and AA biosynthetic capabilities exist in other major clades on the tree of life, including bacteria.

Compared to animals, the understanding of bacterial AA biosynthesis remains relatively incomplete. Earlier studies indicated that most bacteria possess at least a few auxotrophies [6,7]. However, recent research has revealed that the prevalence of AA auxotrophies may have been overestimated due to gaps in knowledge regarding bacterial biosynthetic pathways, though many bacteria still lack the ability to synthesize the full set of AAs [4,8]. Experimental evidence shows that AA auxotrophies can confer a fitness advantage in competition with

prototrophic strains [7,9], and they may serve as an evolutionary strategy to reduce biosynthetic burdens via cooperative interactions within microbial communities [10].

The adaptive benefits of AA auxotrophies in bacteria remain under debate. D'Souza & Kost [9] speculated that outsourcing AA biosynthesis might reduce cellular metabolic costs. While it is well-established that AAs differ in their biosynthetic costs [10–12] and that costlier AAs foster stronger microbial cross-feeding interactions [10], a recent study found no significant correlation between AA biosynthesis costs and the prevalence of AA auxotrophies among bacteria [4]. However, alternative approaches could be employed to rigorously test whether energy-related selection drives the evolution of AA auxotrophies. For instance, if such selection influences the loss of AA biosynthetic capabilities, it should also manifest in the integration of more expensive AAs into bacterial proteomes [5]. To our knowledge, no study has yet examined bacterial proteome energetics in this context.

Detecting auxotrophies is itself an active area of research, employing various methodologies. Many rely on in silico approaches, including genome-scale metabolic modeling [13,14] or homology-guided annotation of enzymes [4,7,15]. These computational approaches are typically validated against a limited number of datasets obtained by experiments [4,14]. However, no simple and standardized protocol currently exists for comparing AA auxotrophy estimates across studies.

To address these gaps, we assembled a taxonomically diverse dataset of bacterial proteomes and analyzed whether AA biosynthesis costs can explain trends in AA auxotrophy composition. Using a simple methodology for auxotrophy detection based on MMseqs2 clustering, we achieved results comparable in quality to those of previous approaches. Our findings reveal that more expensive AAs are more frequently lost and that bacteria with greater numbers of expensive AA auxotrophies encode costlier proteomes, paralleling patterns observed in animals. These results suggest that energy-driven selection plays a key role in shaping auxotrophic phenotypes in bacteria, allowing them to explore protein sequence space more freely during evolution.

Results

To uncover global trends in the evolution of bacterial auxotrophies, we compiled a database of 980 high-quality proteomes representing the diversity of bacteria (Table S1). We assessed the completeness of amino acid (AA) biosynthesis pathways in these species using a MMseq2 clustering approach [16]. This method clusters, in a single step, a representative sample of bacterial enzymes known to catalyze reactions in AA biosynthesis pathways together with all proteomes in our database (see Materials and Methods). Based on the composition of the recovered clusters, functional information on the AA anabolism is then transferred between cluster members allowing us to determine the completeness of 20 AA biosynthesis pathways for each species. Unlike previous studies that rely on strictly defined cutoff values to designate auxotrophies—thereby losing part of the available information—we analyzed the pathway completeness values directly. This approach provides a more accurate representation of the odds that a particular pathway is present.

For each of the twenty amino acids, completeness scores range from 0 to 1, where 0 indicates that all enzymes in the pathway are absent, and 1 indicates that all enzymes are present in a given species (Fig. 1, File S1). We found that the total pathway completeness for the 20 amino acids has an average value of 17.33, with a median of 19.5, for the whole dataset. Consistent with these values, nearly 66% of the species in our dataset exhibit total pathway completeness above 19. These high completeness values suggest that many bacteria are prototrophic for most amino acids, aligning with findings from recent studies [4,8]. However, certain taxonomic groups—including Lactobacillales, *Mollicutes*, and Borreliaceae—show a significant reduction in multiple amino acid biosynthetic pathways (Fig. 1, File S1), indicating that auxotrophies are common in these groups.

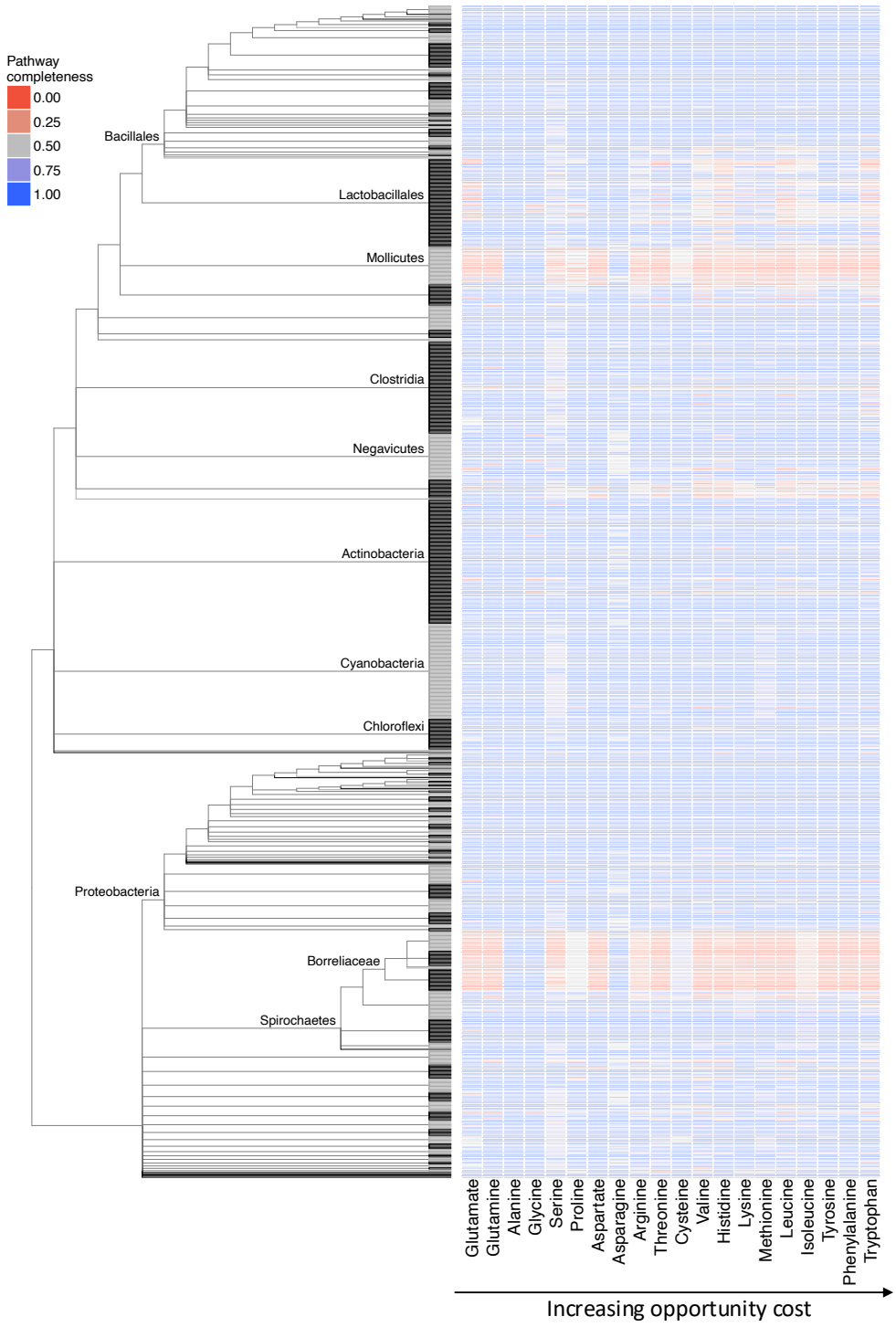

Figure 1. Completeness of AA biosynthesis pathways in bacteria. We created a database of 980 bacterial species to get a comprehensive overview of AA dispensability in this group. Fully resolved tree is shown File S1. We retrieved data on enzymes involved in AA biosynthesis pathways from the KEGG and MetaCyc databases. We searched for their homologs within our

reference database using MMseqs2 clustering (see Methods). For each AA, we showed a completeness score, which represents the percentage of enzymes within a pathway that returned significant sequence similarity matches to our reference collection of AA biosynthesis enzymes. In the case of AAs with multiple alternative pathways, we showed the results only for the most complete one.

If energy-driven selection generally influences the evolution of bacterial amino acid (AA) auxotrophies, one would expect that the ability to synthesize energetically expensive AAs is more frequently lost. To test this hypothesis, we devised an AA auxotrophy index (AI), defined as 1 minus the completeness score, and compared it against the opportunity cost [5], which estimates the energy expenditure associated with AA synthesis (see Materials and Methods). We observed a significant moderate correlation between opportunity cost and auxotrophy index when opportunity cost values for respiratory metabolism were applied (Fig. 2A, B). However, under fermentative conditions, this correlation was much weaker and not statistically significant (Fig. 2C). Together, these findings suggest that the evolution of AA auxotrophies in bacteria, similar to animals, is generally driven by selection favoring energetic savings during AA synthesis, a pattern most pronounced under respiratory conditions.

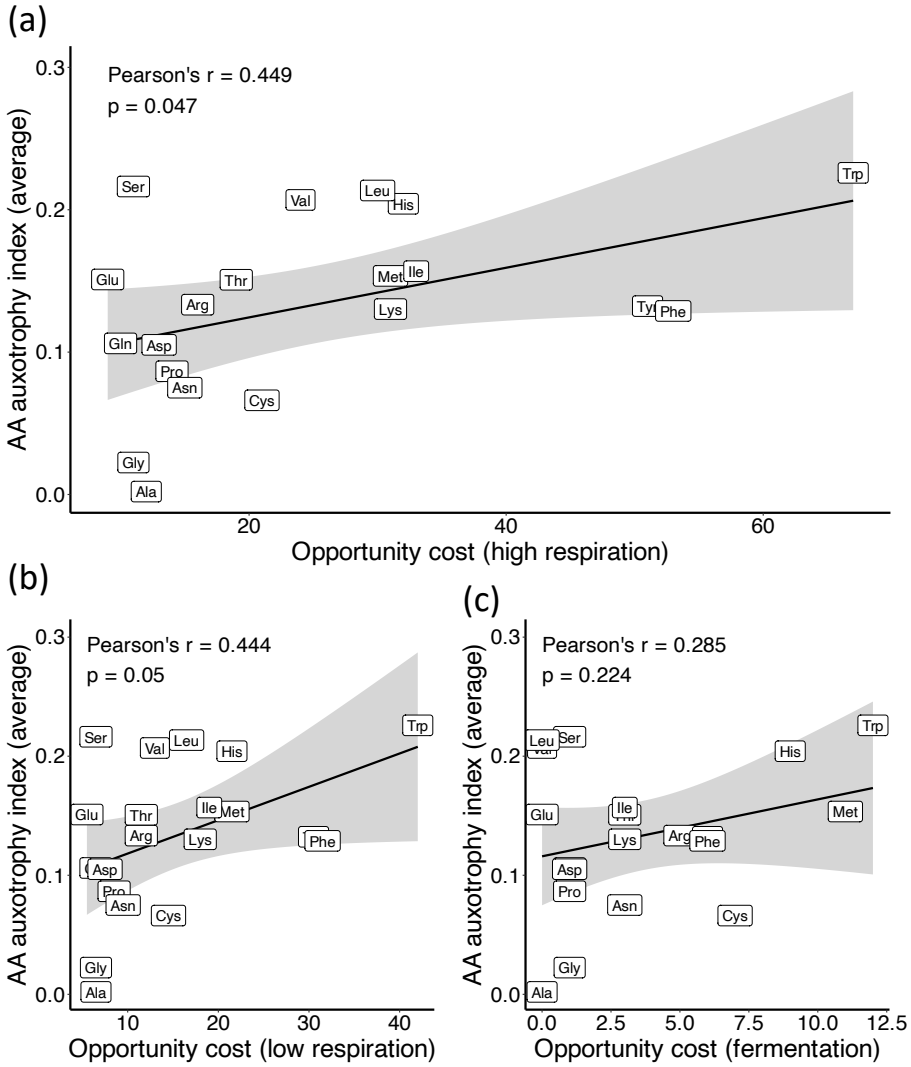

Figure 2. Correlation between AA biosynthesis cost and the average AA auxotrophy index. We estimated the AA auxotrophy index (AI), a measure which equals one minus completeness score, for 980 bacterial species and calculated the average value for each AA. We correlated this value with the opportunity cost of each AA, calculated for three different respiratory modes (see Materials and Methods). Pearson correlation coefficient and p-value are shown on the graph.

However, the central question remains how these reductions in AA biosynthetic pathways influence overall proteome energetics. If energy-driven selection underpins the evolution of

AA auxotrophies in bacteria, one would expect auxotrophic species to maintain more expensive proteomes than their prototrophic counterparts. This is because auxotrophs consume considerably less energy on AA biosynthesis and acquire the missing AAs at a relatively low cost from the environment. In essence, this pattern would suggest that the energy saved on the synthesis of costly AAs offsets part of the proteome's energy expenditure, enabling auxotrophs to incorporate more expensive proteins, potentially contributing to novel functions [5].

To test this idea, we calculated the opportunity cost of an average amino acid (AA) for each bacterial proteome ($OC_{proteome}$) as well as the overall biosynthetic cost savings achieved through AA auxotrophy ($OC_{savings}$, see Materials and Methods). These calculations were performed across three respiratory modes: fermentation, low respiration, and high respiration [5]. We then assessed the correlation between $OC_{proteome}$ and OC_{saving} (Fig. 3). Regardless of the respiratory mode used to estimate energy expenditure, our results showed that species achieving the greatest energy savings in AA production also maintain the most expensive proteomes. Similar to animals, bacterial auxotrophies appear to influence not only the immediate energy budget related to AA production but also reduce selective pressure against the use of energetically costly AAs in proteomes.

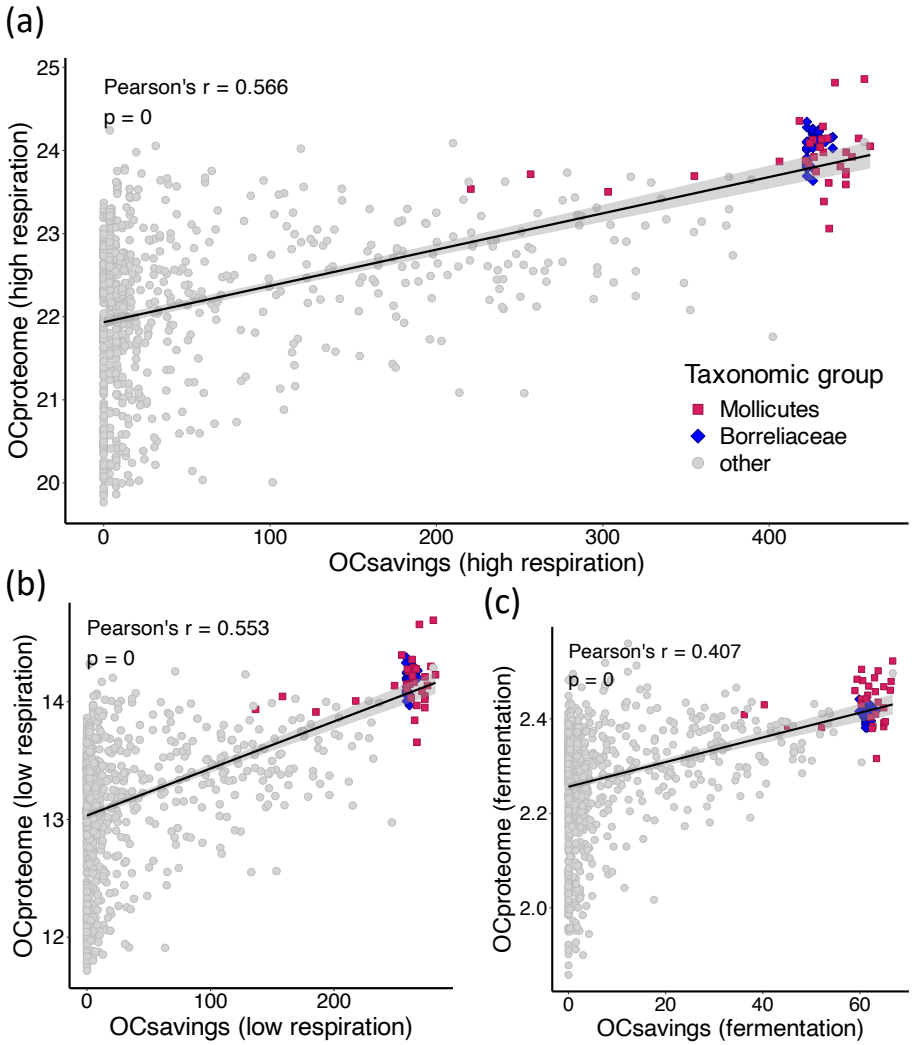

Figure 3. Correlation between the cost of an average AA per proteome and the amount of outsourced energy for AA biosynthesis. We estimated the AA biosynthesis pathway completeness for 980 bacterial species. For each proteome, we calculated its opportunity cost ($OC_{proteome}$) by calculating for each AA the product of its opportunity cost and its frequency in the proteome and then by taking the sum of the obtained values (see Materials and Methods). For each species, we also estimated the energy that was saved by outsourcing AA biosynthesis ($OC_{savings}$). To obtain this value we first multiplied auxotrophy index with opportunity cost for every AA and then we summed the obtained values for all 20 AAs (see Materials and Methods). We repeated all calculations for three different respiratory modes. Pearson correlation coefficient (r) and p-value are shown on the graph.

Discussion

In our previous work, we applied the concept of functional outsourcing [1] to AA biosynthesis in animals and developed a model describing the conditions necessary for the evolution of AA auxotrophies [5]. This model predicts that the loss of AA biosynthetic pathways is selectively favored in respiring organisms with abundant environmental availability of AAs. Crucially, the loss of AA production capabilities is not random—energy-optimizing selection favors the loss of pathways for energetically costly AAs. This, in turn, enables auxotrophs to more freely explore protein sequence space by reducing selective constraints on the use of expensive AAs in their proteomes [5]. To test the broader validity of this model across major clades, we investigated the patterns of AA auxotrophies in bacteria.

While animals exhibit nearly identical sets of auxotrophies among each other, bacterial metabolisms are far more diverse, making bacterial auxotrophies more challenging to detect and interpret [4,7,8]. Although it is known that AA auxotrophies can confer a fitness advantage in competition with prototrophic strains [7,9], it was unclear whether this advantage arises solely from the immediate increase in available energy or if, in the long run, it also influences proteome composition [5].

Our results reveal a significant positive correlation between the AA auxotrophy index averaged across tested bacteria and the AA opportunity cost, as predicted by the AA outsourcing model [5]. This finding aligns with earlier observations that costlier AAs promote stronger microbial cross-feeding interactions [10] and that AA auxotrophies confer a fitness advantage [7,9]. Apart from our study, only one prior investigation has explicitly examined bacterial AA auxotrophies from an energetics perspective, reporting no significant correlation between the frequency of an AA being auxotrophic and its biosynthetic cost [4].

The reasons for this discrepancy are unclear and may stem from differences in bacterial genome datasets, auxotrophy detection pipelines, or biosynthetic cost estimates. Thus, to further test the robustness of our findings, we analyzed data from another comparable study that calculates

AA auxotrophies across the bacterial tree of life using metabolic modeling but lacks energetic calculations [14]. Similar to our study, this dataset also exhibited a significant correlation between the frequency of an AA being auxotrophic and its biosynthetic cost (Figure S1), with an even higher correlation coefficient than observed in our results. Collectively, these findings suggest that, in at least some bacterial groups, AA auxotrophies are influenced by energy savings at the level of AA biosynthesis.

The second prediction of our model is that bacterial species with more auxotrophies should have more expensive proteomes. Our results confirm this prediction, showing that auxotrophic species indeed maintain more expensive proteomes. This finding indicates that AA auxotrophy fundamentally influences proteome composition. The energy savings achieved through the outsourcing of AA biosynthesis relax the constraints on the incorporation of costly AAs into the proteome, thereby enabling auxotrophic organisms to explore protein sequence space more freely [1,5,17–19].

Based on these findings, we hypothesize that the increase in the frequencies of costly AAs in auxotrophic species' proteomes could result in the evolution of proteins with novel functions [1,5,17–19]. Interestingly, the most auxotrophic groups with the most expensive proteomes in our analysis are *Mollicutes* and *Borreliaceae*, the members of which are notorious for causing severe infections that are difficult to manage [20–22]. It is also known that *Borreliella* (syn. *Borrelia*) *burgdorferi* (Spirochaetales) harbors many Borreliaceae-specific genes with unknown functions, which may be implicated in the development of Lyme disease [23]. The evolution and function of these novel genes may be explainable in terms of auxotrophy-related energetic shifts, which should be explored in future studies.

In conclusion, our results suggest a global macroevolutionary trend where AA biosynthesis capability is shaped by energy-saving selection, ultimately leading to the evolution of more expensive proteomes. However, bacteria exhibit remarkable ecological and metabolic diversity, thriving in an array of physically and chemically unique habitats [8,24,25]. This diversity raises the possibility that factors beyond energetic costs influence reductions in AA biosynthetic

pathways. Future studies focusing on specific bacterial lineages and ecologies will be crucial to uncover the role of such factors in shaping the evolution of AA biosynthetic capabilities.

Materials and Methods

To create the database of bacterial proteomes (Table S1), we combined datasets previously assembled for the phylostratigraphic analyses of *Bacillus subtilis* [26] and *Borrelia burgdorferi* [23], along with a resolved phylogeny of *Escherichia coli*. This resulted in a database of 980 species representing most major bacterial lineages. The proteomes, primarily retrieved from the NCBI database and supplemented by the Ensembl database, were evaluated for contamination using BUSCO [27], and all were confirmed to be free of contamination [1].

In our previous study [5], we retrieved pathways and enzyme codes involved in amino acid (AA) biosynthesis from the KEGG and MetaCyc databases. For Aas that can be synthesized via multiple alternative pathways, we treated each pathway separately, even when they shared some enzymes. Using this collection of enzyme codes associated with AA biosynthesis, we retrieved bacterial protein sequences from the KEGG database [28]. For each genus with representatives annotated in KEGG, we selected the species with the largest number of enzymes catalyzing reactions in AA biosynthesis pathways. This process resulted in a protein sequence reference database comprising 387,892 enzymes across 2,095 species (Table S2).

In the next step, we combined the downloaded enzyme sequences known to be involved in amino acid (AA) biosynthesis pathways with all sequences from our bacterial proteomes (4,230,625 sequences across 980 species) into a single database. We then clustered this combined database using MMseqs2 [16] with the following parameters: -cluster-mode 0, -cov-mode 0, -c 0.8, and -e 0.001. Clustering with these parameters generated clusters whose

members exhibit highly similar architectures, as the alignment between query and target sequences covered at least 80% of their length [1].

Based on the presence of enzymes involved in AA synthesis, we functionally annotated the remaining members of the respective clusters. For each AA biosynthesis pathway and species in the database, we calculated a pathway completeness score (i.e., prototrophy index) by dividing the number of detected enzymes by the total number of enzymes in that pathway, resulting in values ranging from 0 to 1. If a species contained alternative biosynthetic pathways for an AA, the pathway with the highest completeness score was selected. We then calculated the AA auxotrophy index (AI_i) using the formula:

$$AI_i = 1 - completness\ score$$

In this formula i denotes one of 20 AAs (i = 1, ..., *20*).

To evaluate the performance of our method, we utilized a previously established testing set of experimentally identified prototrophies and auxotrophies [4,14] to estimate pathway completeness scores. The first tested dataset comprised 160 fully prototrophic species [4]. Our approach exhibited an error rate of 0.012, indicating that approximately 1.2% of amino acids were incorrectly identified as auxotrophic (Data S1). The second dataset included 15 species with at least one known auxotrophy [14]. In this case, we detected an error rate of 0.188 for false prototrophs, meaning that around 19% of amino acids were incorrectly classified as prototrophic (Data S2). Taken together, these error rates suggest that our approach to auxotrophy detection is conservative and aligns with error rates reported for similar methods [4,14].

In our earlier study, we calculated the opportunity cost of biosynthesis for each AA depending on the three respiration modes: high respiration, low respiration and fermentation [5]. The opportunity cost is calculated as the sum of the energy lost in the synthesis of AAs (direct cost) and the energy that would have been produced if a cell catabolized precursors instead of making AAs [5]. This measure reflects the overall impact of AA synthesis on the cell's energetic budget.

Using the AA opportunity cost, we also calculated the opportunity cost of an average AA in each proteome ($OC_{proteome}$) using the following formula:

$$OC_{proteome} = \frac{\sum_{i=1}^{n=20} OC_i \times N_i}{\sum_{i=1}^{n=20} N_i} = \sum_{i=1}^{n=20} OC_i \times f_i$$

In this formula, OC_i represents the opportunity cost of a given AA, N_i denotes the total number of occurrences of this AA in the entire proteome, and f_i represents the frequency of the AA in the proteome (calculated as the number of occurrences of the AA divided by the total number of AAs in the proteome).

We also introduced a new measure to quantify the energetic savings of a species by linking the AA pathway completeness to the opportunity cost. This measure estimates the energy saved by outsourcing AA production to the environment, relative to the energy required to synthesize the full set of 20 AAs. It is calculated as follows:

$$OC_{savings} = \sum_{i=1}^{n=20} OC_i \times AI_i$$

In this formula, OC_i denotes the opportunity cost of a given AA, while AI_i denotes the AA auxotrophy index.

To calculate correlations, we used the cor.test() function in the R stats (v. 3.6.2) package. The heatmap was visualized using the ggtree R package [29].

Author Contributions: Conceptualization, N.K., T.D.-L. and M.D.-L.; methodology, N.K., T.D.-L. and M.D.-L.; software, N.K. and M.D.-L.; validation, N.K., T.D.-L. and M.D.-L.; formal analysis, N.K., T.D.-L. and M.D.-L.; writing—original draft preparation, N.K., T.D.-L. and M.D.-L.; writing—review and editing, N.K., T.D.-L. and M.D.-L.; visualization, N.K.; supervision, T.D.-L. and M.D.-L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the European Regional Development Fund KK.01.1.1.01.0009 DATACROSS (M.D.-L., T.D.-L.).

Data Availability Statement: All data are available in the supplementary materials.

Acknowledgments: We thank M. Futo, A. Tušar, S. Koska, D. Franjević and G. Klobučar for discussions. We used the computational resources of the University Computing Center - SRCE (Padobran) and the Institute Ruđer Bošković.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Domazet-Lošo, M.; Široki, T.; Šimičević, K.; Domazet-Lošo, T. Macroevolutionary Dynamics of Gene Family Gain and Loss along Multicellular Eukaryotic Lineages. *Nat Commun* **2024**, *15*, 2663, doi:10.1038/s41467-024-47017-w.

2. Payne, S.H.; Loomis, W.F. Retention and Loss of Amino Acid Biosynthetic Pathways Based on Analysis of Whole-Genome Sequences. *Eukaryotic Cell* **2006**, *5*, 272–276, doi:10.1128/ec.5.2.272-276.2006.

3. Richter, D.J.; Fozouni, P.; Eisen, M.B.; King, N. Gene Family Innovation, Conservation and Loss on the Animal Stem Lineage. *eLife* **2018**, *7*, e34226, doi:10.7554/eLife.34226.

4. Ramoneda, J.; Jensen, T.B.N.; Price, M.N.; Casamayor, E.O.; Fierer, N. Taxonomic and Environmental Distribution of Bacterial Amino Acid Auxotrophies. *Nat Commun* **2023**, *14*, 7608, doi:10.1038/s41467-023-43435-4.

5. Kasalo, N.; Domazet-Lošo, M.; Domazet-Lošo, T. Massive Outsourcing of Energetically Costly Amino Acids at the Origin of Animals 2024.

6. Mee, M.T.; Wang, H.H. Engineering Ecosystems and Synthetic Ecologies. *Mol. BioSyst.* **2012**, *8*, 2470, doi:10.1039/c2mb25133g.

7. D'Souza, G.; Waschina, S.; Pande, S.; Bohl, K.; Kaleta, C.; Kost, C. Less Is More: Selective Advantages CAN Explain the Prevalent Loss of Biosynthetic Genes in Bacteria. *Evolution* **2014**, *68*, 2559–2570, doi:10.1111/evo.12468.

8. Price, M.N.; Zane, G.M.; Kuehl, J.V.; Melnyk, R.A.; Wall, J.D.; Deutschbauer, A.M.; Arkin, A.P. Filling Gaps in Bacterial Amino Acid Biosynthesis Pathways with High-Throughput Genetics. *PLoS Genet* **2018**, *14*, e1007147, doi:10.1371/journal.pgen.1007147.

9. D'Souza, G.; Kost, C. Experimental Evolution of Metabolic Dependency in Bacteria. *PLoS Genet* **2016**, *12*, e1006364, doi:10.1371/journal.pgen.1006364.

10. Mee, M.T.; Collins, J.J.; Church, G.M.; Wang, H.H. Syntrophic Exchange in Synthetic Microbial Communities. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, doi:10.1073/pnas.1405641111.

11. Kaleta, C.; Schäuble, S.; Rinas, U.; Schuster, S. Metabolic Costs of Amino Acid and Protein Production in *Escherichia Coli*. *Biotechnology Journal* **2013**, *8*, 1105–1114, doi:10.1002/biot.201200267.

12. Zengler, K.; Zaramela, L.S. The Social Network of Microorganisms — How Auxotrophies Shape Complex Communities. *Nat Rev Microbiol* **2018**, *16*, 383–390, doi:10.1038/s41579-018-0004-5.

13. Zimmermann, J.; Kaleta, C.; Waschina, S. Gapseq: Informed Prediction of Bacterial Metabolic Pathways and Reconstruction of Accurate Metabolic Models. *Genome Biol* **2021**, *22*, 81, doi:10.1186/s13059-021-02295-1.

14. Starke, S.; Harris, D.M.M.; Zimmermann, J.; Schuchardt, S.; Oumari, M.; Frank, D.; Bang, C.; Rosenstiel, P.; Schreiber, S.; Frey, N.; et al. Amino Acid Auxotrophies in Human Gut

Bacteria Are Linked to Higher Microbiome Diversity and Long-Term Stability. *The ISME Journal* **2023**, *17*, 2370–2380, doi:10.1038/s41396-023-01537-3.

15. Price, M.N.; Deutschbauer, A.M.; Arkin, A.P. GapMind: Automated Annotation of Amino Acid Biosynthesis. *mSystems* **2020**, *5*, 10.1128/msystems.00291-20, doi:10.1128/msystems.00291-20.

16. Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat Biotechnol* **2017**, *35*, 1026–1028, doi:10.1038/nbt.3988.

17. Domazet-Loso, T.; Tautz, D. An Evolutionary Analysis of Orphan Genes in Drosophila. *Genome Res.* **2003**, *13*, 2213–2219, doi:10.1101/gr.1311003.

18. Domazet-Lošo, T.; Brajković, J.; Tautz, D. A Phylostratigraphy Approach to Uncover the Genomic History of Major Adaptations in Metazoan Lineages. *Trends in Genetics* **2007**, *23*, 533–539, doi:10.1016/j.tig.2007.08.014.

19. Tautz, D.; Domazet-Lošo, T. The Evolutionary Origin of Orphan Genes. *Nat Rev Genet* **2011**, *12*, 692–702, doi:10.1038/nrg3053.

20. Strnad, M.; Rudenko, N.; Rego, R.O.M. Pathogenicity and Virulence of *Borrelia Burgdorferi*. *Virulence* **2023**, *14*, 2265015, doi:10.1080/21505594.2023.2265015.

21. Pilo, P.; Frey, J.; Vilei, E.M. Molecular Mechanisms of Pathogenicity of Mycoplasma Mycoides Subsp. Mycoides SC. *The Veterinary Journal* **2007**, *174*, 513–521, doi:10.1016/j.tvjl.2006.10.016.

22. Waites, K.B.; Talkington, D.F. *Mycoplasma Pneumoniae* and Its Role as a Human Pathogen. *Clin Microbiol Rev* **2004**, *17*, 697–728, doi:10.1128/CMR.17.4.697-728.2004.

23. Čorak, N.; Anniko, S.; Daschkin-Steinborn, C.; Krey, V.; Koska, S.; Futo, M.; Široki, T.; Woichansky, I.; Opašić, L.; Kifer, D.; et al. Pleomorphic Variants of Borreliella (Syn. Borrelia)

Burgdorferi Express Evolutionary Distinct Transcriptomes. *IJMS* **2023**, *24*, 5594, doi:10.3390/ijms24065594.

24. Lozupone, C.A.; Knight, R. Global Patterns in Bacterial Diversity. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 11436–11440, doi:10.1073/pnas.0611525104.

25. Flemming, H.-C.; Wuertz, S. Bacteria and Archaea on Earth and Their Abundance in Biofilms. *Nat Rev Microbiol* **2019**, *17*, 247–260, doi:10.1038/s41579-019-0158-9.

26. Futo, M.; Opašić, L.; Koska, S.; Čorak, N.; Široki, T.; Ravikumar, V.; Thorsell, A.; Lenuzzi, M.; Kifer, D.; Domazet-Lošo, M.; et al. Embryo-Like Features in Developing *Bacillus Subtilis* Biofilms. *Molecular Biology and Evolution* **2021**, *38*, 31–47, doi:10.1093/molbev/msaa217.

27. Simão, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing Genome Assembly and Annotation Completeness with Single-Copy Orthologs. *Bioinformatics* **2015**, *31*, 3210–3212, doi:10.1093/bioinformatics/btv351.

28. Kanehisa, M.; Furumichi, M.; Sato, Y.; Kawashima, M.; Ishiguro-Watanabe, M. KEGG for Taxonomy-Based Analysis of Pathways and Genomes. *Nucleic Acids Research* **2023**, *51*, D587–D592, doi:10.1093/nar/gkac963.

29. Yu, G.; Smith, D.K.; Zhu, H.; Guan, Y.; Lam, T.T. ggtree : An r Package for Visualization and Annotation of Phylogenetic Trees with Their Covariates and Other Associated Data. *Methods Ecol Evol* **2017**, *8*, 28–36, doi:10.1111/2041-210X.12628.

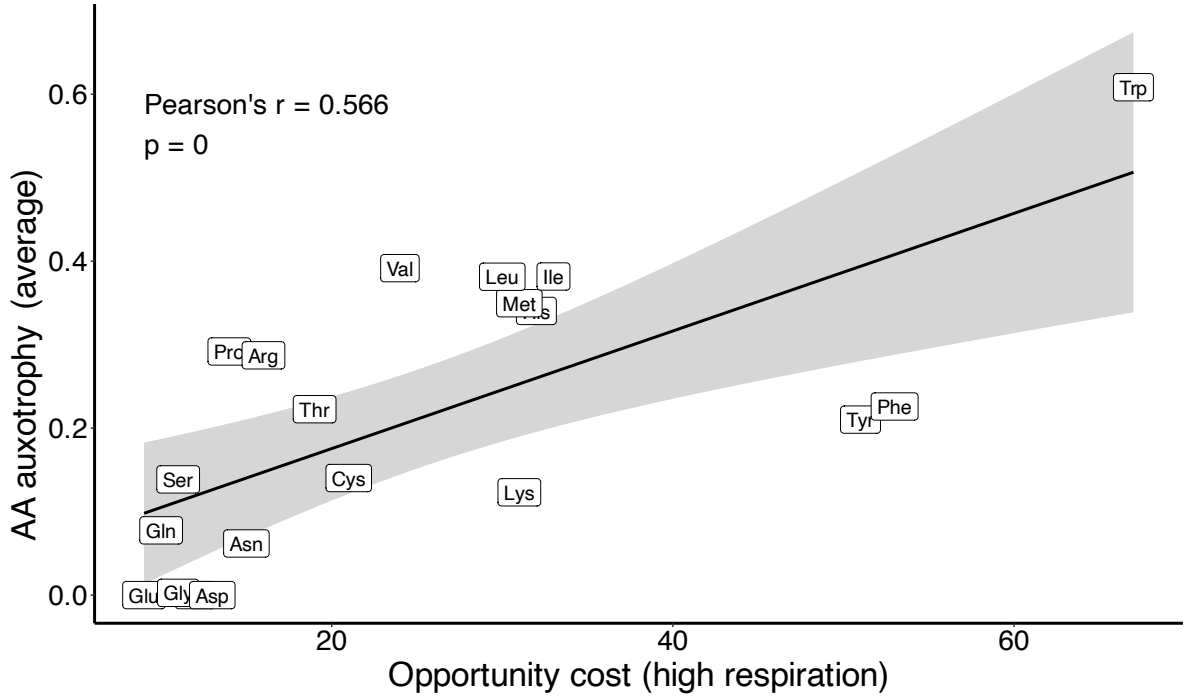

Figure S1. Correlation between AA biosynthesis cost and the average AA auxotrophy based on data from Starke et al. (2023). We calculated the average of AA auxotrophy measures provided in the Starke et al. (2023) study for 3687 bacterial species. We correlated this value with the opportunity cost of each AA, calculated for high respiration mode (see Materials and Methods). Pearson correlation coefficient and p-value are shown on the graph.