nature communications

Article

https://doi.org/10.1038/s41467-026-68724-6

Outsourcing of energetically costly amino acids at the origin of animals

Received: 29 November 2024

Accepted: 9 January 2026

Niko Kasalo[1,2], Mirjana Domazet-Lošo[3] ✉ & Tomislav Domazet-Lošo[1,4] ✉

Animals are generally capable of synthesizing eleven amino acids, while the remaining nine—often referred to as essential—must be acquired through diet. This metabolic trait profoundly shapes animal ecology and evolutionary trajectories, and recent phylogenomic studies indicate that it arose from gene losses at the root of the animal tree. However, it remains unclear which selective forces, if any, directed this far-reaching metabolic simplification. Here, we show that essential amino acids are energetically far more expensive to synthesize than non-essential ones, particularly under high respiratory conditions—a hallmark of the animal lifestyle. Using a probabilistic selection test, we find support for the hypothesis that energy-related selection, counterbalanced by pleiotropic effects, favored the outsourcing of essential amino acid production under well-oxygenated and nutrient-rich conditions. Remarkably, we also find that extant animals use expensive amino acids more frequently than their closest unicellular relatives, suggesting that constraints on their usage were relaxed once their synthesis was externalized, thereby enabling animal genes to evolve more freely across protein sequence space. In this light, we propose that the origin of animals was tightly linked to energy-related adaptations rather than to unpredictable stochastic events, as recently suggested.

Proteins are made of 20 proteinogenic amino acids (AAs). The biosynthesis of these AAs is an indispensable biochemical process on which all life depends. Many organisms, such as plants, fungi, various bacteria, and many unicellular eukaryotes are capable of synthesizing all 20 AAs[1–4]. However, other lineages lost the capability to synthesize some AAs, which means that they have to harvest these essential amino acids (EAAs) from their environment[1,2]. A prominent example is animals (Metazoa), which lost the capability to synthesize nine EAAs[1,3,5]. Recent phylogenomic studies showed that this simplification occurred at the root of the animal tree, after the split from choanoflagellates[6,7]. This metabolic loss is part of a broader simplification event that preceded metazoan radiation and included a wide range of metabolic reductions[6–8].

However, which selective forces, or random processes, were driving this simultaneous loss of EAA biosynthetic pathways in metazoans remains a great mystery[3]. The biosynthesis of AAs is generally energetically costly[9], but there is also considerable variability among them in the energy required for their production, with the cheapest and most expensive AAs differing by almost an order of magnitude[9–12]. On the other hand, free energy is a critical and limited resource that sustains biological systems, meaning that its harnessing and efficient usage are under continuous selective pressure[13,14]. For instance, energetically expensive AAs sparingly occur in protein sequences, especially in highly expressed genes[9,15–17]. It has been shown that this selectively favored bias leads to a net reduction in energy costs related to AA synthesis in cells[13,14].

---

[1]Laboratory of Evolutionary Genetics, Division of Molecular Biology, Ruđer Bošković Institute, Zagreb, Croatia. [2]Department of Biology, Faculty of Science, University of Zagreb, Zagreb, Croatia. [3]Department of Applied Computing, Faculty of Electrical Engineering and Computing, University of Zagreb, Zagreb, Croatia. [4]School of Medicine, Catholic University of Croatia, Zagreb, Croatia. ✉e-mail: mirjana.domazet@fer.hr; tdomazet@irb.hr

Inspired by these findings, we reasoned that the massive loss of EAA biosynthesis capability at the root of the animal tree could also be connected to energy management. A plausible hypothesis is that the biosynthesis of EAAs, as a group, incurs higher energy costs than the biosynthesis of non-essential AAs (NEAAs). By outsourcing the production of energetically costly AAs, an animal ancestor could have freed a part of its energy budget, which could then be reallocated to other functional needs. We further assumed that the outsourcing of expensive EAAs would enable animals to use them more freely in their proteomes—compared to their unicellular holozoan ancestors—as the energetic constraints on their usage would be reduced. Surprisingly, these possibilities have never been tested, even though estimates of AA biosynthesis energy costs have been available for a long time[9–11] and AA biosynthetic capabilities have been studied from various perspectives[1–5,11]. It is also puzzling that no effort has yet been made to conceptually link AA synthesis costs with their essentiality status, nor has any study considered the potential role of selection in shaping AA auxotrophies in animals.

In this work, we identify the potential selective forces and ecophysiological conditions that led to the essential/non-essential AA dichotomy. Although alternative explanations cannot be definitively ruled out, our analyses support the hypothesis that energy management and pleiotropic effects exert antagonistic selective pressures that —together with ecophysiological factors such as respiration mode and AA availability in the environment—played a significant role in driving the outsourcing of EAA synthesis in the metazoan lineage. Moreover, we find that animal proteomes contain a higher proportion of energetically costly AAs than those of their closest unicellular relatives, indicating that EAAs were more freely integrated into proteomes once their synthesis was outsourced. Accordingly, we argue that the emergence of animals was driven by energy-management adaptations rather than stochastic events.

Results

Essential amino acids are expensive

We first retrieved the most recent estimates[12] on the required number of energy-bearing metabolites (ATP, NADH, and NADPH) for AA biosynthesis (Supplementary Data 1). These most recent estimates corrected some inadvertent errors that had propagated in previous studies using these values to calculate AA biosynthesis energy costs (see "Methods"). With these corrected values, we then calculated the direct cost of AA biosynthesis, representing the number of high-energy phosphate bonds (~P) required for synthesizing an AA from its metabolic precursor (Supplementary Data 1 and Supplementary Table 1).

It is important to note that direct costs quantify the energy used for AA production from metabolic precursors without considering how this metabolic mode impacts the overall energy balance of the cell. To address this, we also calculated the opportunity cost (see "Methods"), defined as the total energy that would have been produced if the precursors had been metabolized, plus the energy lost during the synthesis of AAs (Supplementary Data 1 and Supplementary Table 1). This alternative cost measure, which has been extensively used in previous work[9–11,18], quantifies the impact of the choice to synthesize an AA on the overall energy balance of the cell by accounting for the unrealized energy gain from precursors (lost opportunity). In essence, it shows how much energy is given up by an organism when producing AAs[18].

However, actual values of both opportunity costs and direct costs are directly dependent on the respiratory mode of a cell, which can differ within the life cycle of an organism as well as between different species[10,19–21]. Surprisingly, the fundamental importance of respiration mode for AA energy cost estimation is rarely recognized[10,12]. To account for this effect, we calculated here both direct and opportunity costs for three respiratory modes: fermentation, low respiration, and high respiration (see "Methods", Supplementary Data 1 and Supplementary Table 1). This range of respiration modes encompasses all major metabolic states across the tree of life, enabling us to identify the respiratory conditions that most strongly favor the outsourcing of AAs.

For comparative purposes, we first sorted the 20 AAs in increasing order based on the opportunity cost calculated for high respiratory conditions, which correspond to the metabolic lifestyle of Metazoans (Fig. 1a). To illustrate how the dispensability status of AAs in animals is distributed in this representation, we marked the nine EAAs and eleven NEAAs[1,3,5] in different colors (Fig. 1a). Remarkably, the AAs are arranged almost perfectly into two groups: (1) energetically cheaper NEAAs and (2) energetically more expensive EAAs (Fig. 1a).

The only exceptions are cysteine (C) and tyrosine (Y), which have relatively high energy costs and are usually considered NEAAs. However, in animals they are synthesized directly from other EAAs, which are secured from the environment; i.e., cysteine (C) from methionine (M) and tyrosine (Y) from phenylalanine (F)[4]. This suggests that cysteine (C) and tyrosine (Y) should be considered partially essential, as most of their synthesis costs are externally covered. In this context, it is evident that nearly all energy costs related to the top 55% (11 out of 20) most expensive AAs are outsourced (Fig. 1a). The analyses of opportunity costs under low respiration and fermentation conditions revealed very similar patterns, although overall values were lower and the difference between the cheapest and most expensive AAs was reduced (Supplementary Figs. 1 and 2).

While opportunity cost has been the prevailing energy cost measure in previous studies[9–11,18], we also considered direct cost here. In contrast to opportunity costs, which account for the alternative option of respiring AA precursors, direct costs strictly reflect the expenses of AA synthesis. This measure is therefore particularly useful in situations where cells have a surplus of energy, making alternative metabolic options less relevant. A simple visual inspection of the direct cost chart (Fig. 1b) revealed that the direct costs of NEAAs are typically lower than those of EAAs; however, the difference between the two groups, although evident, is not as clear-cut as it is for opportunity costs (Fig. 1b and Supplementary Figs. 1b and 2b). Thus, to assess these differences more quantitatively, we statistically compared average direct costs as well as opportunity costs between NEAAs and EAAs using a non-parametric test (Fig. 2; Supplementary Figs. 3 and 4). We considered AA dispensability in two versions; in the first classical one, we grouped standard 9 EAAs (Fig. 2a, c), whereas in the second one, we extended this set to 11 members by treating cysteine (C) and tyrosine (Y) as EAAs (Fig. 2b, d).

Regardless of which cost measure is considered, under high respiration conditions the average energy costs of EAAs are significantly higher than the average energy costs of NEAAs with mostly large effect sizes ($r > 0.5$) (Fig. 2). Interestingly, although average EAA costs are always higher than average NEAA costs, this difference becomes less apparent and significant under low respiration, and even less so under a fermentation lifestyle (Supplementary Figs. 3 and 4). To better present this trend, we plotted the sum of AA opportunity costs as well as direct costs across three respiratory conditions (Fig. 3). This chart revealed a clear pattern in which the differences in total AA synthesis costs between NEAAs and EAAs are the lowest under fermentation, increase under low respiration, and are the highest under high respiration (Fig. 3). This trend is obvious for both direct and opportunity costs but is much stronger for opportunity costs, showing that more efficient respiration disproportionally increases the energy costs of EAAs in contrast to NEAAs. If we assume that energy consumption related to AA synthesis influences fitness, the observed differences between NEAAs and EAAs should be most strongly exposed to selection under high respiration. Subsequently, this selective pressure could lead to the eventual loss of EAA biosynthesis pathways, as observed at the root of Metazoa.

To further evaluate the hypothesis that selective forces related to energy management contributed to the outsourcing of EAAs under

https://doi.org/10.1038/s41467-026-68724-6

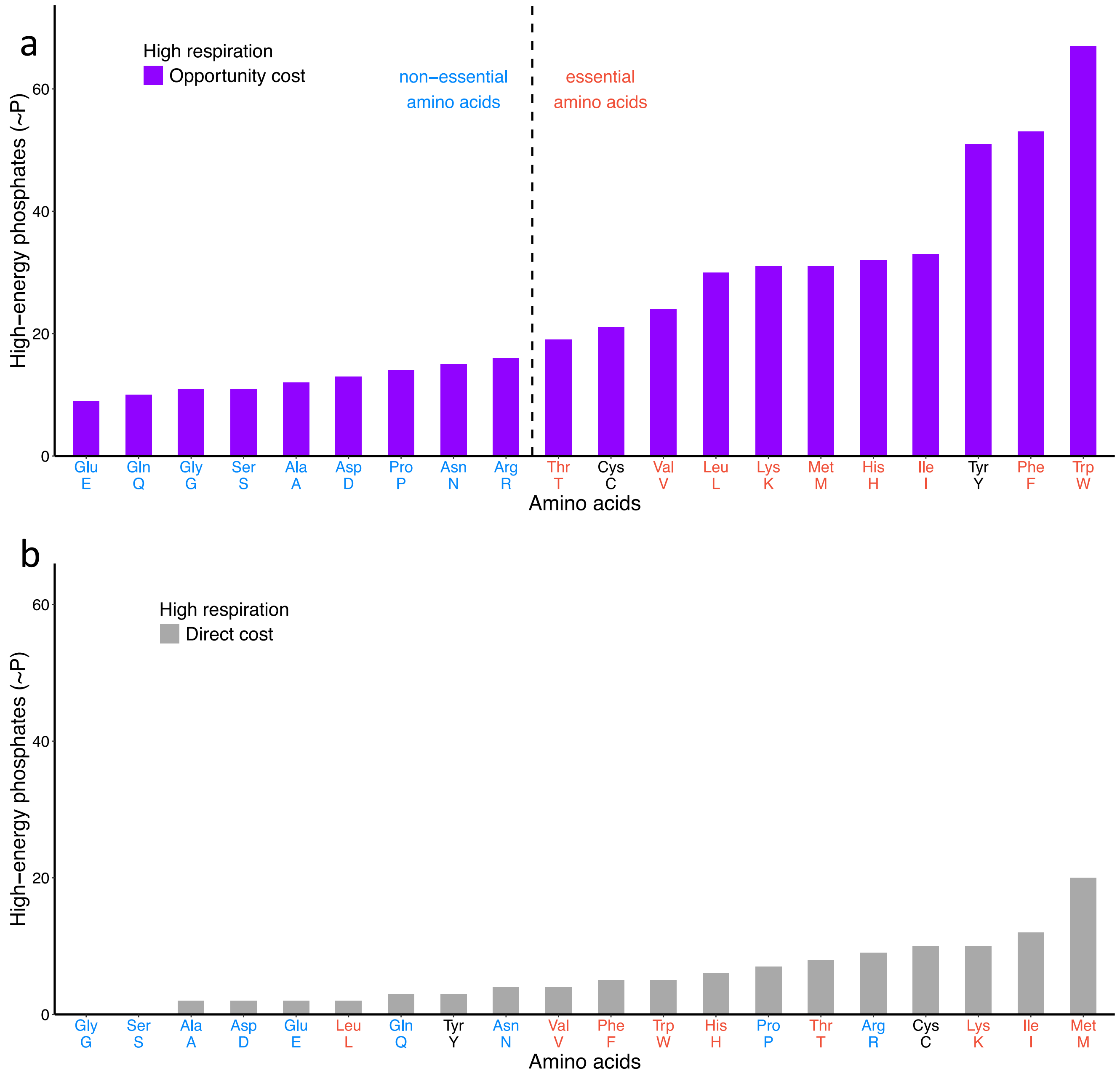


Fig. 1 | Comparison of amino acid synthesis costs under high respiration. **a** Opportunity cost (OC) and **b** direct cost (DC) are shown as the number of high-energy phosphate bonds (-P) required for synthesis. Amino acids are sorted in ascending order according to their biosynthesis costs. Essential amino acids (EAAs) in animals are shown in red; non-essential amino acids (NEAAs) in blue. Source data are provided as a Source data file.

high-respiration conditions, we developed the combinatorial phenotype selection (CPS) test, which probabilistically assesses whether a quantitative factor (e.g., biosynthetic cost) exerts directional selection on a combinatorial phenotypic trait (e.g., the auxotrophic/prototrophic status of amino acids), using neutral evolution as the null hypothesis (see “Methods”). We first generated all possible permutations of AA dispensability status under the assumption that 9 (or 11) out of 20 AAs are essential and then calculated the average opportunity (direct) cost of EAAs for every permutation. Based on the obtained distribution of average opportunity (direct) costs of EAAs, we calculated the matching empirical probability mass function (PMF) which was then used to retrieve the probability that the real-life set of EAAs in animals appeared randomly (Fig. 4). A low *p*-value would indicate that selection related to the AA synthesis costs shaped the observed set of EAAs in animals more strongly than random processes.

All performed CPS tests showed that the mean cost of the observed EAAs in animals falls at the right tail of the empirical distribution (Fig. 4 and Supplementary Data 2). In fact, in some analyses, the observed EAA combination in animals has the highest average opportunity cost in comparison to 167,960 possible permutations (Fig. 4b). This suggests that the observed EAA combination is highly unlikely to have arisen by chance alone and that our results are consistent with the hypothesis that AA production costs shaped the evolution of the EAA set in animals. Although animal lifestyle is universally linked to high respiration, it is interesting to test if selection could shape the EAA set observed in animals under low respiration and fermentation conditions. Similar to high respiration data, the CPS test of low respiration and fermentation data returned quite low *p*-values, which suggests that even in these metabolic situations, it is unlikely that random processes were a major force in shaping the EAA set.

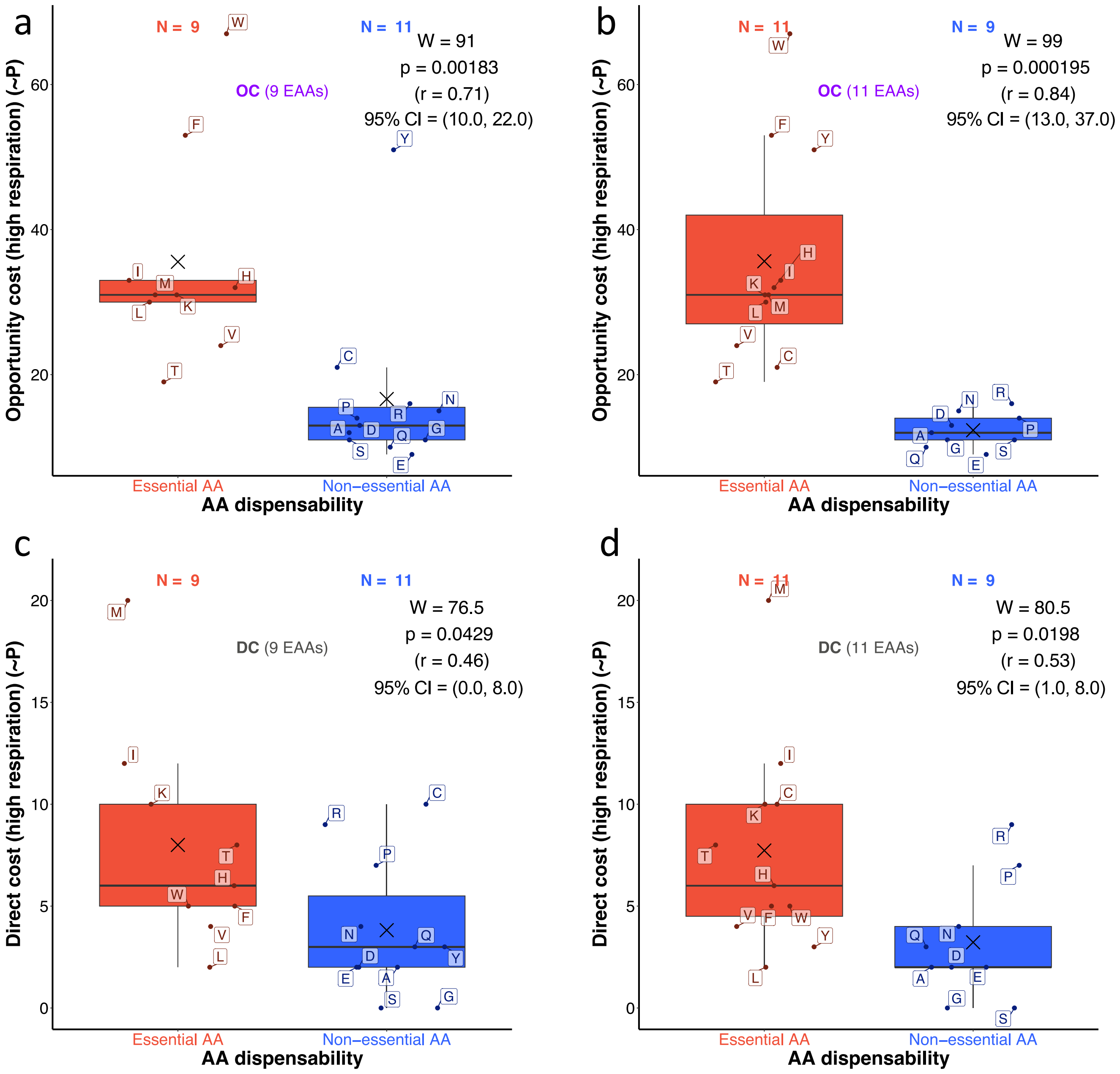


Fig. 2 | Comparison of energy costs under high respiration between essential and non-essential AAs in animals. The opportunity cost (**a**, **b**) represents the energy expended on AA synthesis combined with the energy that could be produced from precursor molecules, while the direct cost (**c**, **d**) represents the energy expended on AA synthesis only (see Supplementary Table 1). We performed the comparison by considering (**a**) opportunity costs and (**c**) direct costs of 9 essential and 11 non-essential amino acids, and by considering (**b**) opportunity costs and (**d**) direct costs of 11 essential and 9 non-essential amino acids. In the extended AA dispensability analyses (**b**, **d**), we considered cysteine (C) and tyrosine (Y) as conditionally essential amino acids, given that their synthesis directly depends on the essential amino acids methionine (M) and phenylalanine (F), respectively. The differences in energy costs under high respiration between the two groups were shown using boxplots and the significance of these differences was tested by the two-sided Mann-Whitney U test with continuity correction. The corresponding W-value, *p*-value, effect size (*r*), and the 95% non-parametric confidence intervals are displayed. X symbols indicate means, bars represent medians, lower and upper hinges correspond to the first and third quartiles, and whiskers extend from the hinges to the largest or smallest values within 1.5 times the interquartile range. Individual AAs are represented by one-letter symbols. Source data are provided as a Source data file.

However, these *p*-values are less significant compared to those obtained for cost measures under high respiration (Supplementary Figs. 5 and 6), which highlights that the selective loss of EAAs is most probable under high-respiration conditions.

Selective trade-off between AA energetics and pleiotropy

If one assumes that AA energy management is an important selective force driving the loss of EAA production, the question arises why only half of the most energetically expensive AAs were outsourced in animals. Theoretically, if AA energetics exert strong selective pressure, all AAs could potentially be outsourced. A likely explanation is that AA pleiotropy counteracts AA energetics, leading to a set of EAAs shaped by a selective trade-off between these two factors. For example, the biosynthesis of arginine is part of the urea cycle, which is essential for the excretion of excess nitrogen in ureotelic animals (primarily amphibians and mammals)[22]. Despite its relatively high energetic cost,

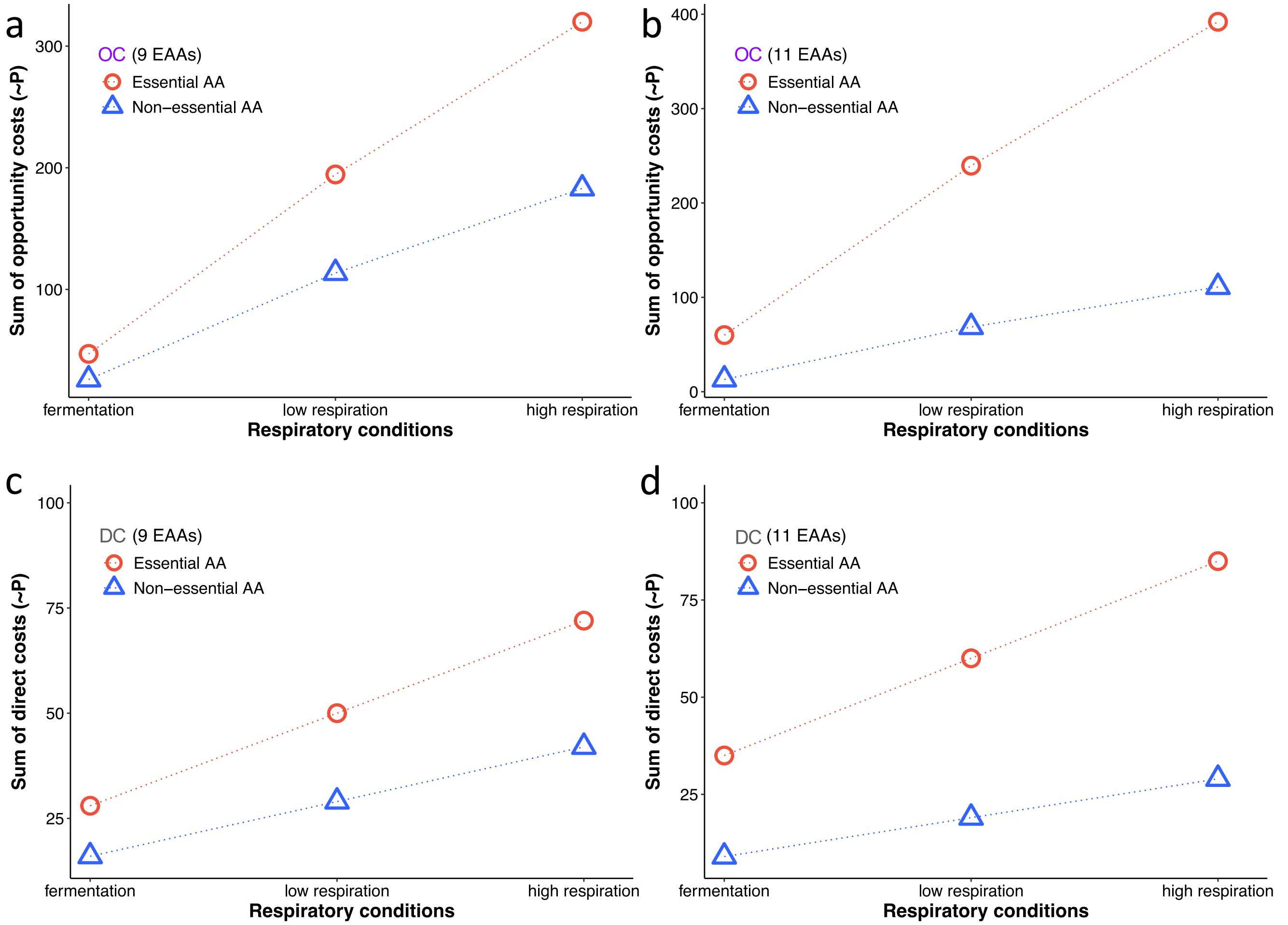


Fig. 3 | Dependence of amino acid biosynthesis costs on respiratory conditions. Summed opportunity costs (OC) (**a**, **b**) and direct costs (DC) (**c**, **d**) across all 20 AAs were calculated under fermentative, low respiratory, and high respiratory conditions (see Supplementary Table 1). Trends for EAAs (red line, circles) and NEAAs (blue line, triangles) are shown. The significance of differences between NEAAs and EAAs within a particular respiration mode are shown in Fig. 1, Supplementary Figs. 1 and 2. We performed the calculations by considering 9 EAAs (**a**, **c**), and 11 EAAs (**b**, **d**). In the extended AA dispensability analyses (**b**, **d**), we considered cysteine (C) and tyrosine (Y) as conditionally essential amino acids given that their synthesis depends on the essential amino acids methionine (M) and phenylalanine (F), respectively. Source data are provided as a Source data file.

an animal lacking an alternative means of nitrogen excretion cannot afford to lose the ability to synthesize arginine.

To test the role of pleiotropy, we applied the same battery of tests used in the analyses of AA energy costs. To visually inspect the data, we first mapped the number of KEGG pathways and biochemical reactions on individual AAs distributed according to their increasing opportunity costs (Fig. 5a). It can be seen that most of the cheapest AAs, and especially glutamate, take part in a large number of reactions and pathways. However, two conditionally essential AAs (cysteine and tyrosine) and half of the essential AAs also display relatively high involvement in various cellular processes (Fig. 5a).

The comparison using non-parametric tests showed that EAAs have on average lower pleiotropy than NEAAs, however with borderline significance (Supplementary Fig. 7). To test if these differences could be governed by selection as opposed to random processes, we again applied CPS test where the tested factor, instead of energy costs, was pleiotropy measured by the number of KEGG pathways and biochemical reactions in which AAs participate. This test showed that the observed EAA pleiotropies fall within the left tails of the empirical PMF, with *p*-values fluctuating around the 0.05 threshold (Fig. 5b–e). These results indicate that pleiotropy may have imposed selective pressure to maintain the biosynthetic capabilities of certain amino acids. In turn, this implies that pleiotropy counteracts energy-related selection to some extent, although it appears to be the weaker force in this interaction.

AA dispensability in animals

Nine AAs are canonically considered essential in animals, although there may be exceptions to this rule, often related to further losses in biosynthetic capabilities[5,23,24]. However, the information on the AA dispensability status in animals and their allies is rather cursory and scattered over methodologically disparate studies[5,6,23–28]. To get a coherent overview of AAs dispensability in animals and their closest relatives (non-animal holozoans), we explored the completeness score (CS) of AA synthesis pathways on a large holozoan tree comprising 167 species using reciprocal best hit as well as clustering approaches at different significance cutoffs (see Methods). We also note that the completeness score of AA synthesis pathways is only a proxy of AA dispensability because it is possible that some taxa evolved or co-opted alternative pathways. Nevertheless, regardless of the approach and parameters used, we obtained highly consistent results (Fig. 6, Supplementary Figs. 8–13).

Most notably, we recovered an almost universal trend of reduction of pathways involved in the biosynthesis of the 9 canonically essential AAs in animals. This trend is not apparent in the non-animal holozoans where pathway completeness does not follow a clear

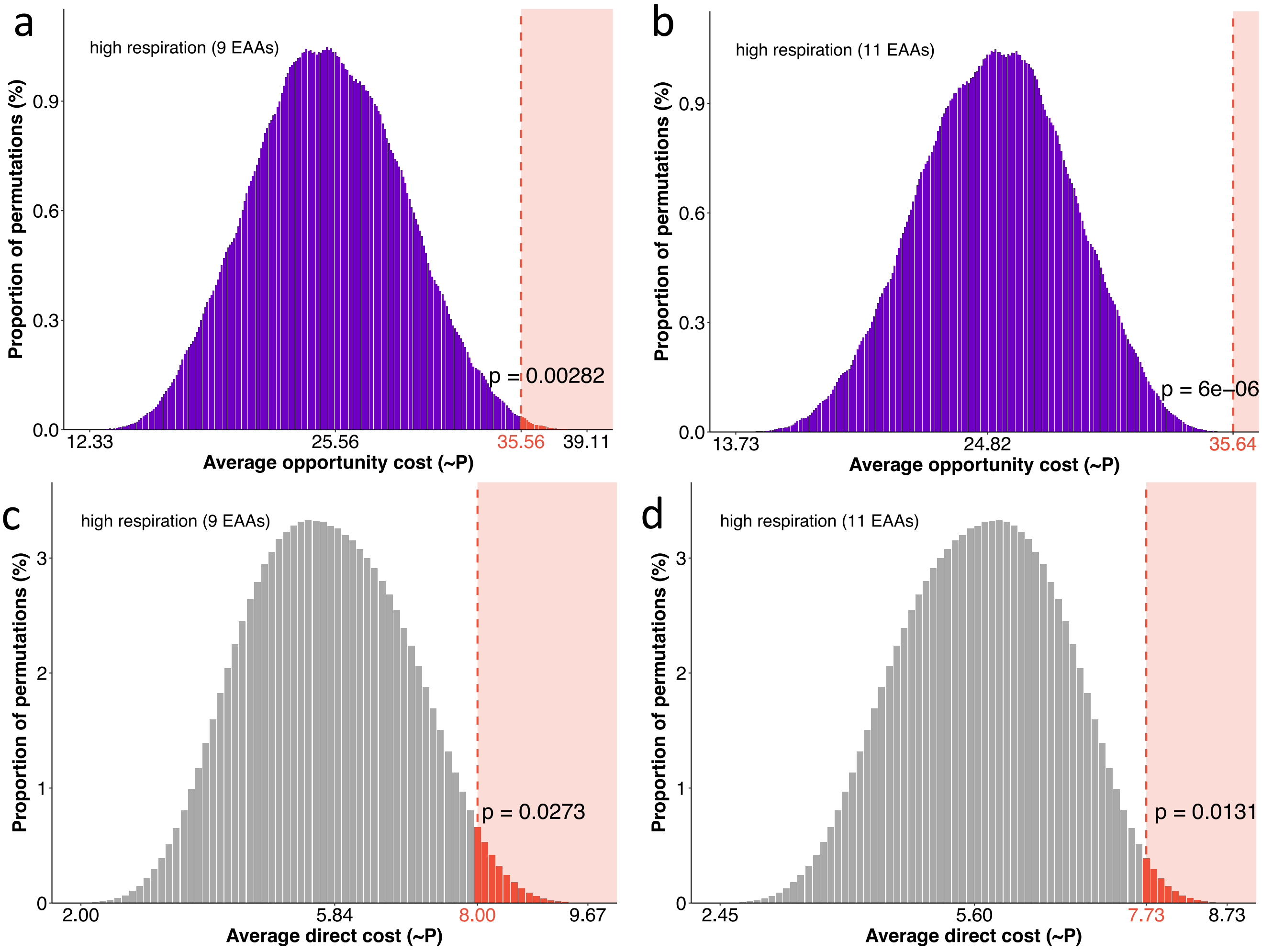


Fig. 4 | Probability that energy-related selective pressure under high respiration shaped the EAA set in animals. We used the combinatorial phenotype selection (CPS) test protocol, which first assigned essentiality status to all possible combinations of 9 or 11 AAs (a total of 167,960 permutations). For each AA group in each permutation, average opportunity (**a**, **b**) and direct costs (**c**, **d**) under high respiration were calculated, and the proportions of these averages are shown in histograms. The obtained distribution represents the empirical probability mass function (PMF). The value in red denotes the average value of the EAA sets actually observed in animals. The *p*-value was calculated in a one-sided fashion by summing the proportions of average values equal to, or more extreme than, the observed average. Low *p*-values support the hypothesis that selection pressures related to AA energy management drove the loss of essential amino acid biosynthetic capabilities in animals (see CPS test section in “Methods”). Source data are provided as a Source data file.

pattern (Fig. 6, Supplementary Figs 8–13). Since the choanoflagellate group Acanthoecida has in our analysis high completeness of biochemical pathways, and the same is true for Teretosporea and *Capsaspora*, we conclude that the ability to synthesize all 20 AAs is likely a plesiomorphic feature of holozoans, which agrees with previous work[6]. To statistically evaluate whether non-animal holozoans are significantly more prototrophic than animals, we applied a non-parametric test to compare the average AA auxotrophy index (AI_{avg}) calculated for each species[29]. Briefly, the AI_{avg} quantifies the overall extent of incompleteness in AA biosynthesis pathways in a given species, with values ranging from 0 (fully prototrophic) to 1 (fully auxotrophic) (see “Methods”). Indeed, regardless of the method used to detect the presence of AA biosynthesis enzymes, animals consistently appear significantly more auxotrophic than other holozoans (Fig. 6, Supplementary Figs. 8–13).

In our analysis, we detected a few additional AA biosynthesis pathways with low completeness in animals, beyond the canonical nine. The most notable example is the arginine pathway, which is partially complete in many non-vertebrate taxa (Fig. 6, Supplementary Figs. 8–13). This result aligns well with previous studies that report arginine auxotrophy in some insects and tunicates[5,30]. This volatility in auxotrophic/prototrophic status within animals is intriguing, given that arginine is the most expensive NEAA; that is, it lies at the boundary between NEAAs and EEAs when ranked by opportunity costs (Fig. 1a). Its dispensability status in animals is most likely dependent on the sensitive interplay between taxon-dependent differences in its environmental availability, the energy burden it poses on the metabolism and its pleiotropic effects[5,30].

The analysis of AA pathway completeness allowed us to estimate AA auxotrophies for each metazoan species individually (Supplementary Data 4). Based on the completeness score estimates (Fig. 6), we assigned auxotrophy status to any AA whose biosynthesis pathway was incomplete. This enabled us to compare the opportunity costs between the EAA and NEAA sets and perform the CPS test in each metazoan species independently (Supplementary Data 2 and 4). We found that, across species, EAAs consistently exhibited higher opportunity costs than NEAAs, supporting the hypothesis that AA outsourcing was driven by energy-related selection (Supplementary Data 2 and 4).

Although the significance of these patterns gradually declines at higher levels of auxotrophy, it remains statistically significant in all species, with only two showing *p*-values slightly above 0.05. However, even in those two cases (*Choloepus hoffmanni* and *Mnemiopsis leidyi*),

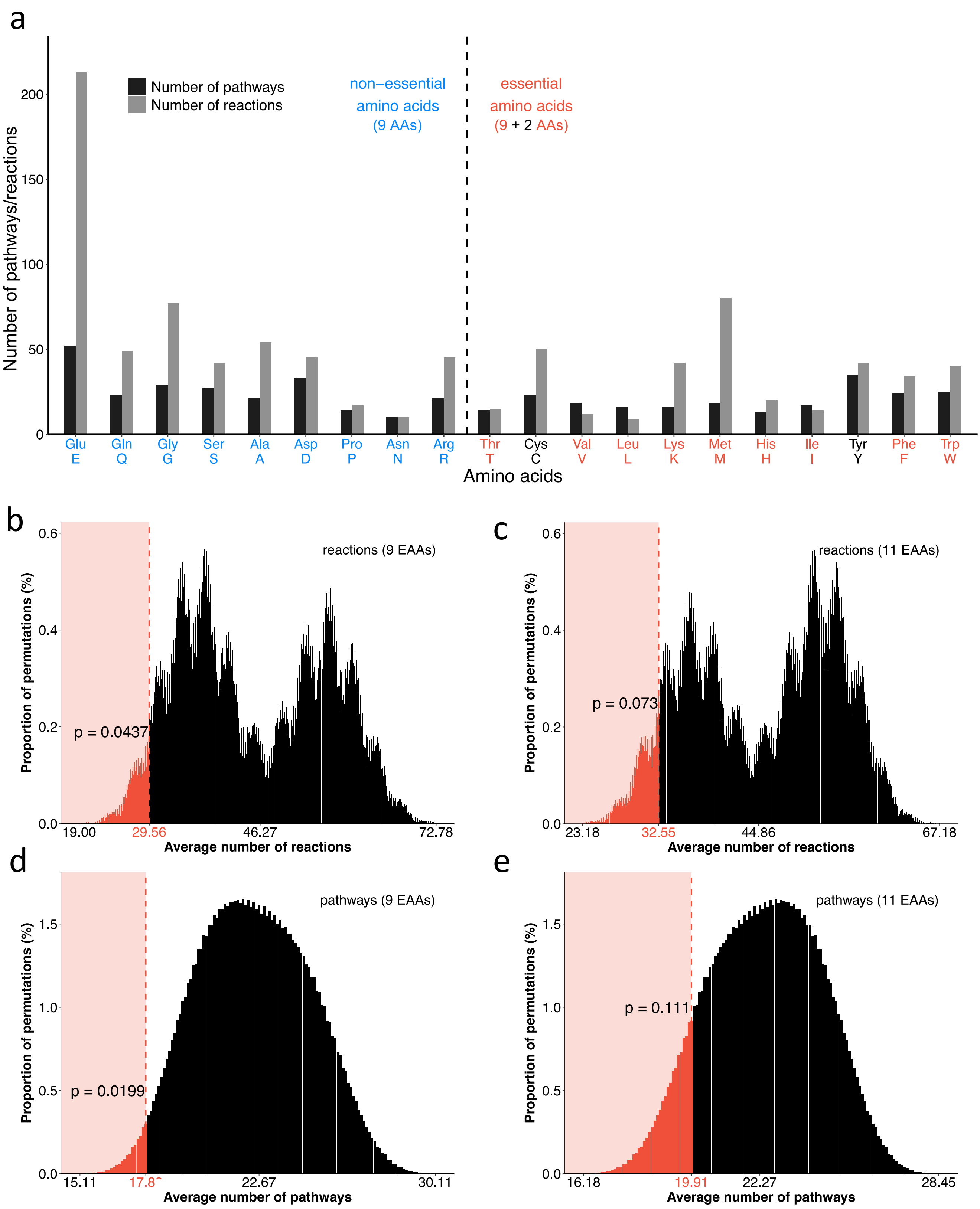
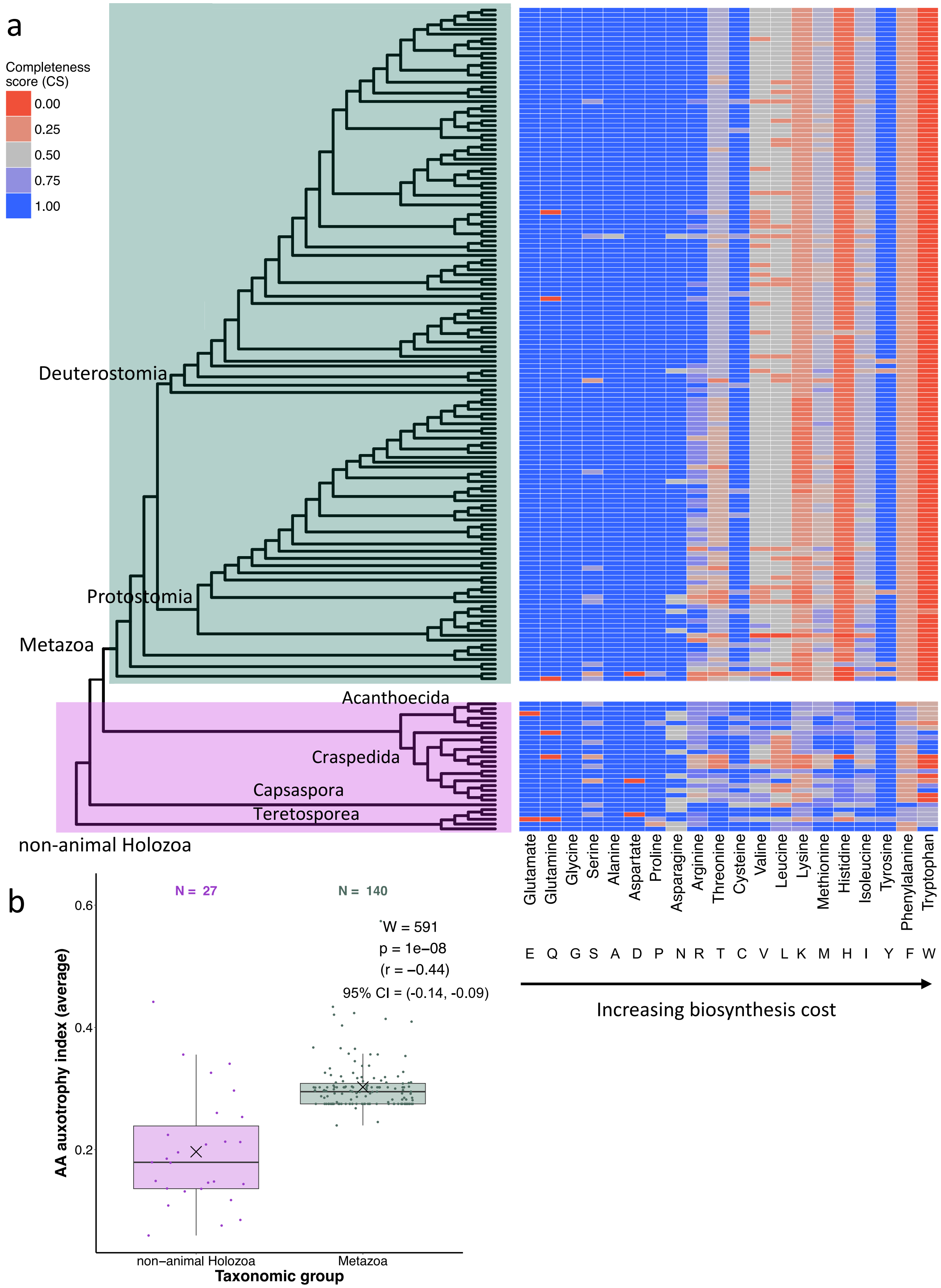


Fig. 5 | Pleiotropy measures and probability that pleiotropy-related selective pressure shaped the EAA set in animals. a For each amino acid, the number of pathways and reactions encoded in the KEGG database are shown. Amino acids are sorted in ascending order according to their opportunity costs. **b–e** We used the combinatorial phenotype selection (CPS) test protocol, which first assigned essentiality status to all possible combinations of 9 or 11 AAs (167,960 permutations). For each AA group in each permutation, the average number of reactions (**b, c**) and pathways (**d, e**) were calculated and the proportions of these averages are shown in histograms (empirical PDF). The value in red denotes the average pleiotropy value of the EAA set observed in animals. The *p*-value was calculated in a one-sided fashion, by summing the proportions of average values equal to or lower than the observed average. Relatively low *p*-values suggest that selection pressures associated with AA pleiotropy may have partially counteracted those driven by energy management (see CPS test section in "Methods"). Source data are provided as a Source data file.



the trend of EAAs being more energetically costly than NEAAs is still evident (Supplementary Data 2 and 4). This is an expected effect, as the canonical (and most prevalent) EAA set (Fig. 1) represents one of the most energetically costly combinations. Adding more auxotrophic AAs to this canonical set naturally lowers the average EAA opportunity cost. However, such extensions of the canonical EAA set are relatively uncommon—only a little over 6% of species in our metazoan dataset exhibit more than 11 auxotrophies (Supplementary Data 2 and 4).

Animal proteomes are expensive

The metabolic simplifications through the loss of EAA synthesis capabilities were only possible under the condition that the last common

Fig. 6 | Overview of AA dispensability in Metazoa. a We used a reference database of 167 holozoans (Supplementary Data 4)[7] to obtain a comprehensive overview of AA dispensability in this group. All enzymes involved in AA biosynthesis were retrieved from the KEGG and MetaCyc databases. We searched for their homologs within our reference database using the reciprocal best hit method implemented in MMSeqs2 with an e-value cutoff of 10^{-60} (see "Methods"). For each AA, we calculated a completeness score (CS) representing the percentage of enzymes within a pathway that returned significant sequence similarity matches to our collection of AA biosynthesis enzymes. For AAs with multiple alternative pathways, only the most complete pathway is shown. Fully resolved trees and MMSeqs2 analyses using reciprocal best hits and clustering methods at different e-value thresholds are provided in Supplementary Figs. 8–13. **b** Comparison of average AA auxotrophy index (AI_{avg}) between animals and non-animal holozoans. For each species (27 non-animal holozoans, 140 metazoans), the auxotrophy index for each AA was calculated as 1−CS. These values were averaged to obtain AI_{avg}, representing the average proportion of missing AA biosynthesis pathways per species (0 = fully prototrophic; 1 = fully auxotrophic). Differences between groups are shown as boxplots, with statistical significance assessed using the two-sided Mann–Whitney U test with continuity correction. The corresponding W-value, p-value, effect size (r), and the 95% non-parametric confidence intervals are displayed. X symbols indicate means, bars represent medians, lower and upper hinges correspond to the first and third quartiles, and whiskers extend from the hinges to the largest or smallest values within 1.5 times the interquartile range. Source data are provided as a Source data file.

ancestor of animals (animal LCA) was able to secure these costly EAAs in surplus from the environment. If we assume that the unicellular ancestor of animals was a prototroph for AAs[6] (Fig. 6), then an important shift in feeding ecology occurred in the animal LCA that allowed this massive metabolic outsourcing. This might be a switch from a predominantly bacteria-eating lifestyle, as seen in present-day choanoflagellates[31,32], to more efficient multicellular-style suspension/filter feeding, or possibly to some sort of grazing behavior or macrophagous predation[33,34].

Irrespective of which of these feeding adaptations occurred, it should have allowed animals to use EAAs more freely in their proteomes, since the energy for their production was largely outsourced. Previous studies have consistently shown that AA synthesis cost tends to negatively correlate with the AA frequency in the proteome[11,15]. Consistent with these findings, we observed the same trend in our data (Supplementary Data 3). However, no study has compared the AA frequencies between the proteomes of prototrophic and auxotrophic organisms. To test this, we compared the usage of EEAs between non-animal holozoans and Metazoa (Fig. 7, Supplementary Figs. 14 and 15). This comparison showed that animals use on average more EAAs in their proteins than non-animal holozoans (Fig. 7a, b; Supplementary Figs. 14 and 15). Similarly, we found that the energetic cost of an average AA in the proteome ($OC_{proteome}$ and $DC_{proteome}$; see Eqs. 1 and 2) is higher in metazoans than in non-animal holozoans (Fig. 7c, d).

However, when all genes within a genome are considered, these observed trends could depend on the internal evolutionary dynamics of a particular lineage. To account for this, we reduced the dataset to include only the conserved genes shared between animals and non-animal holozoans (see "Methods"). The patterns obtained for these conserved genes are congruent with those obtained using all genes within a genome (Supplementary Fig. 14), suggesting that, in animals, the restriction on the usage of expensive AAs is also relaxed for conserved genes. Together, these patterns indicate that the usage of expensive AAs in animal proteomes generally increases once their production is outsourced. All these results are consistent with our prediction that the reduction of selective constraints on energetically expensive AAs allowed animals to more freely explore protein sequence space[35–37].

Discussion

Several studies examined the genes involved in the AA biosynthesis pathways and found that many of them are missing in animals[1,5–7,24]. It was suggested that these reductions are caused by animal heterotrophy that relaxed selective pressures on maintaining AA biosynthetic capabilities[1,5]. While improved heterotrophic feeding strategies likely played a role in the outsourcing of EAAs in animals, this is just one factor in the presumably more complex scenario that resulted in this metabolic simplification. For instance, it has not been investigated which evolutionary forces primarily drove the loss of AA synthesis capabilities after purifying selection for their maintenance was reduced due to improved dietary access to amino acids. Some studies vaguely suggest that random processes contributed to the loss of dispensable amino acid synthesis pathways[1,5]; however, the possibility of selectively driven EAA loss has not been considered. To address this, we propose an amino acid outsourcing model based on the findings of this study (Fig. 8).

According to our model, the availability of environmental amino acids and an organism's respiration mode are primary ecophysiological determinants that set the stage for the outsourcing of AAs (Fig. 8). A surplus of amino acids in the environment reduces selective pressures on in-house AA production (Fig. 8) by lowering the net energy costs associated with this production. However, the extent of these savings largely depends on the respiration physiology of an organism that resides in an AA-rich environment (Fig. 8). The total costs of AA production substantially increase with the shift from fermentation to high respiration (Fig. 3 and Supplementary Table 1). This increase applies to both direct costs and opportunity costs, with the totals for all 20 AAs rising by 2.6-fold and 6.9-fold, respectively, as the respiration mode shifts from fermentation to high respiration (Fig. 3, Supplementary Table 1). This dramatic increase supports the hypothesis that selective pressures related to AA production costs are considerably greater in organisms—such as animals—with high-respiration lifestyles. Additionally, the metabolic shift from fermentation to high respiration widens the differences between the cheapest and most expensive AAs (Figs. 3 and 8).

Although both average opportunity costs (Fig. 2a, b) and direct costs (Fig. 2c, d) follow the same trend—with EAAs consistently showing significantly higher values than NEAAs—this pattern is more pronounced for opportunity costs (Fig. 3). This difference likely arises because opportunity costs more accurately reflect the biological phenotype under selection, accounting not only for the direct energy required for synthesis but also for the energetic trade-offs involved in diverting metabolic precursors toward AA production. While direct costs may be informative when cells have a continuous surplus of energy—making alternative metabolic options less relevant—this scenario is likely unrealistic in natural ecosystems. Therefore, we suggest that opportunity costs generally provide a better measure for testing ecologically relevant scenarios than direct costs.

It is possible that the environmental abundance of AAs is an additional driver of AA auxotrophy evolution. Testing this, however, would require data on the relative concentrations of individual AAs in the ancient environment and the assumption that these proportions remained stable over evolutionary timescales. To our knowledge, such exact paleoenvironmental data are not available, and we therefore could not directly test this factor as a selective pressure in our analysis. However, if one assumes that AA availability at the origin of metazoans mirrored their present-day genomic frequencies, one might expect cheaper AAs to have been outsourced due to their presumed environmental surplus, and expensive AAs to have been retained due to their environmental scarcity. Instead, we observe the opposite: expensive AAs are outsourced. This suggests that, while environmental

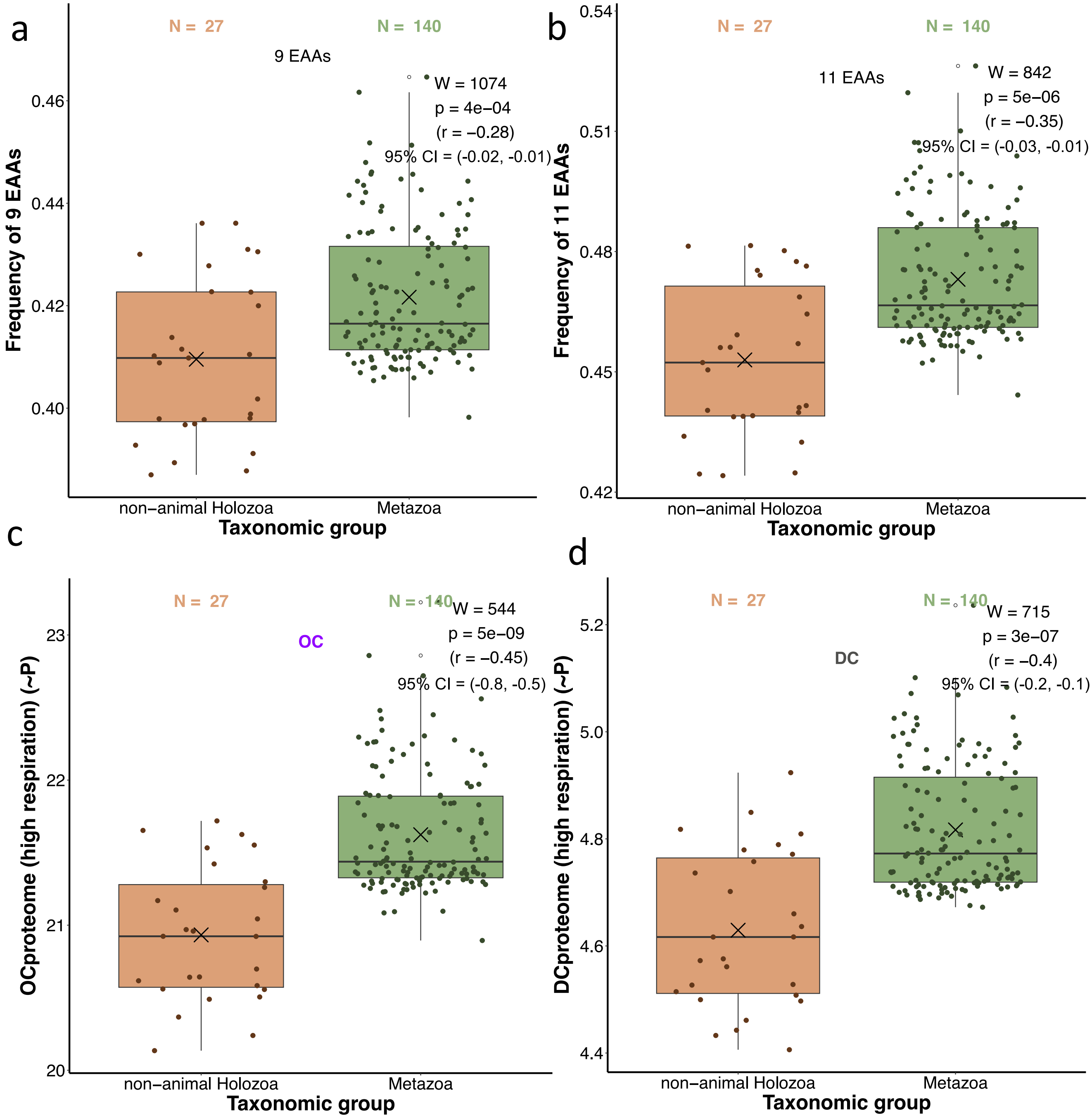


Fig. 7 | Comparison of EAA usage and the energy costs of an average AA between non-animal Holoza and Metazoa. a The cumulative frequency of nine AAs which are generally considered to be essential in Metazoans (Thr, Val, Leu, Lys, Ile, Met, His, Phe, Trp). **b** The cumulative frequency of eleven EAAs. This extended dataset contains two AAs, which could be considered conditionally essential (Cys, Tyr). **c**, **d** The opportunity cost ($OC_{proteome}$) and the direct cost ($DC_{proteome}$) of an average AA in a proteome (see Supplementary Table 1). This value represents a weighted mean of AA biosynthesis energy costs where the frequencies of twenty AAs in a proteome act as weights. The differences in energy costs between the two groups were shown using boxplots and the significance of these differences was tested by the two-sided Mann-Whitney U test with continuity correction. The corresponding W-value, *p*-value, effect size (*r*), and the 95% non-parametric confidence intervals are displayed. X symbols indicate means, bars represent medians, lower and upper hinges correspond to the first and third quartiles, and whiskers extend from the hinges to the largest or smallest values within 1.5 times the interquartile range. The list of non-animal holozoans (27) and metazoans (140) whose proteomes are included in calculations is available in Supplementary Data 4[7]. Source data are provided as a Source data file.

amino acid availability may have been a necessary precondition for outsourcing, energy-related selection was likely an important factor shaping the pattern of amino acid auxotrophies in animals.

All of this suggests that the strongest energy-related selective pressures will act on organisms thriving in oxygenated environments where AAs are readily available and whose ecophysiology relies on high respiration. Beyond animals, at least two other examples of convergence in eukaryotic lineages further support this model. The soil-dwelling ameba *Dictyostelium discoideum* (11 EAAs) and the euglenozoan *Leishmania major* (also 11 EAAs) demonstrate a striking similarity to metazoans regarding amino acid dispensability, with only serine and arginine being additionally essential[5]. Both organisms thrive

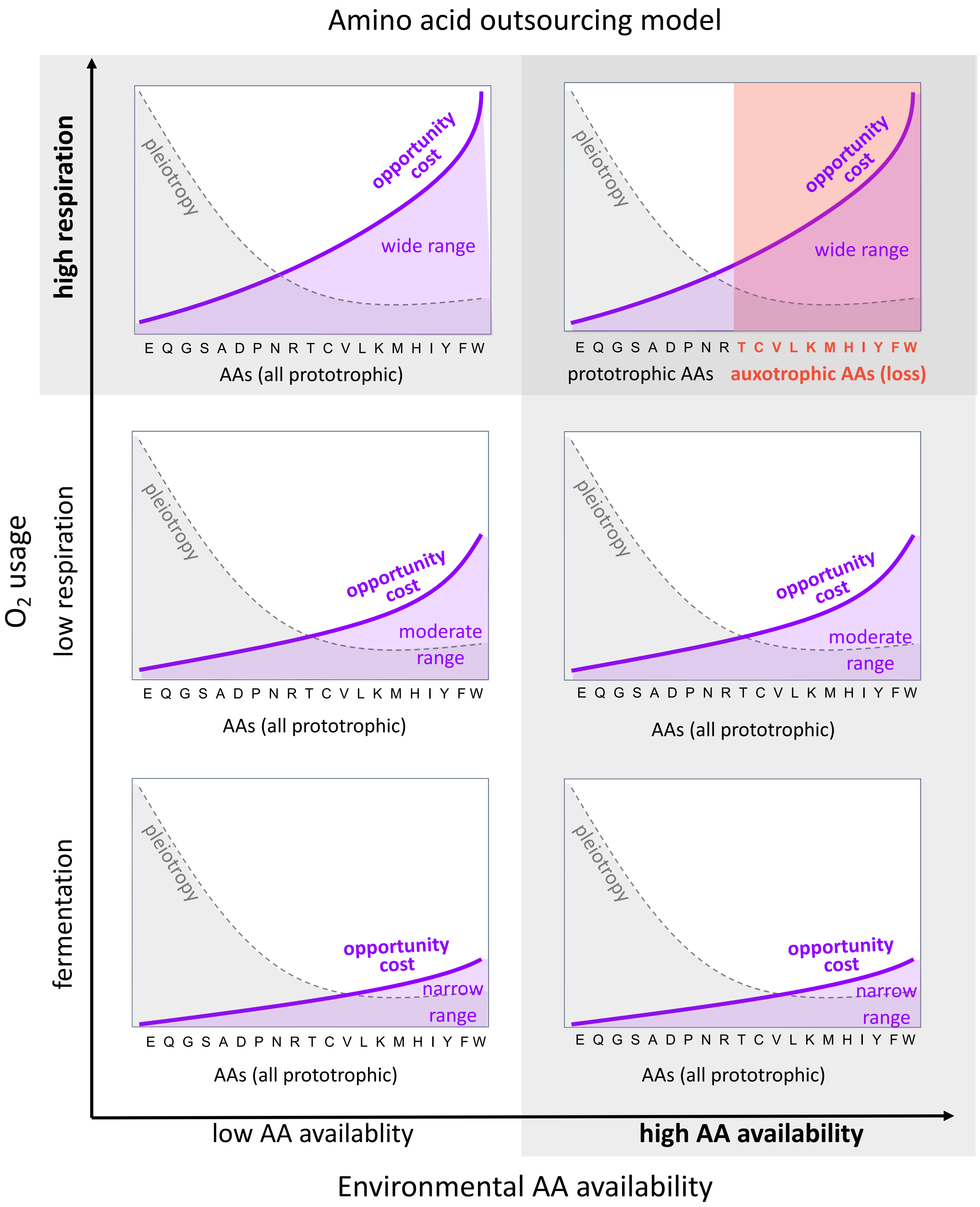


Fig. 8 | Amino acid outsourcing model. Six charts depict the interplay between the four key predictors of AA auxotrophy: AA biosynthesis costs, pleiotropy, the availability of AAs in the environment, and respiration mode. On each chart, AAs are ordered by increasing opportunity costs (OC), which are depicted by the purple line. The area below the purple line depicts the total AA opportunity costs. The degree of change between the lowest and highest OC values is designated as "narrow," "moderate," and "high range". The dashed line depicts the pleiotropic effects of AAs, with higher values indicating more reactions and pathways in which the AA takes part. Shaded backgrounds indicate ecophysiological conditions that favor the loss of AA prototrophy. All graphs are simplified drawings of data presented in this study.

in oxygen-rich environments[38,39], and their feeding ecologies—*L. major* as a parasite and *D. discoideum* as a generalist predator[40]—resemble those of certain animal lineages.

Another example of convergence with animals was recently reported in prokaryotes[29,41]. Bacteria from the phylum Bdellovibrionota resemble animals in their obligate predatory and aerobic lifestyles. These ecophysiological similarities created conditions for the convergent evolution of a similar set of AA auxotrophies and for the increased use of energetically expensive AAs in the proteomes of both Bdellovibrionota[29] and animals. Furthermore, an experimental study has shown that bacterial strains auxotrophic for various AAs gain a fitness advantage over their prototrophic counterparts in AA-rich cultures[2], indicating that selective pressure on energy management may universally underlie AA auxotrophies. On the other hand, fully AA-prototrophic *E. coli* strains utilize an "overflow" mechanism to genetically switch from aerobic to anaerobic metabolism[19]. This approach enables cheaper AA production during periods of high cell proliferation and protein synthesis, when demand is high. This highlights that, while an aerobic lifestyle generally allows for more efficient energy production, it also incurs the cost of more expensive AA synthesis.

Taken together, there are at least three solutions to this persistent trade-off between the overall energy budget and AA production under aerobic conditions. One strategy is to outsource the production of energetically expensive AAs, as seen in all animals, *D. discoideum*, *L. major*, and Bdellovibrionota. Another solution is the "overflow" mechanism based on gene regulation, as observed in *E. coli*. A third approach is a shift to a predominantly fermentative lifestyle with occasional aerobic episodes featuring reduced respiration, as seen in some yeasts[7,42].

In line with this, our CPS tests—which probabilistically assess whether a quantitative factor influences an observed phenotype—support the hypothesis that selective pressures related to energy cost optimization drove the loss of EAA biosynthetic capability (Figs. 4 and 8). However, additional CPS analyses suggest that this energy-driven selection was partially counteracted by pleiotropy—another likely selective force—that helped to maintain the biosynthetic capabilities of NEAAs with important pleiotropic functions (Fig. 5, Fig. 8). For instance, glutamate (E), the most pleiotropic amino acid, serves as a major metabolic hub, playing a critical role in nitrogen assimilation, nucleoside and amino acid biosynthesis[43]. Related glutamine (Q), which is synthesized by glutamine synthetase from glutamate and ammonia, also has diverse metabolic functions, including nitrogen metabolism, nucleotide synthesis, non-EAA synthesis, and the regulation of epigenetic changes[44]. Interestingly, glutamine synthetase is one of the oldest known functional genes, present in the last universal common ancestor, suggesting its fundamental role in cellular metabolism[45]. Comparably to glutamate (E) and glutamine (Q), other NEAAs are also known to perform a variety of important metabolic functions besides protein synthesis[30,46–49].

It must be emphasized that, although an obligate predatory lifestyle, efficient respiration, heightened environmental AA availability, the energetic costs of AAs, and their pleiotropy were crucial in the evolution of AA auxotrophy in animals, it is unlikely that they were the only factors driving this complex evolutionary simplification. Other, yet untested, factors may also have contributed to the outsourcing of EAAs. Most notably, while multiple cases of convergence support our model of selection-driven metabolic reduction, stochastic events may have created the necessary preconditions for energy-related selection to act[32]. For example, although increased AA availability was necessary for metabolic outsourcing to occur, the underlying cause of this increased availability remains unclear. A specific constellation of evolutionary preconditions must evidently be met before energy-related selection on AA auxotrophies becomes operational, as evidenced by the fact that many organisms remain prototrophic. However, even with these conditions met, we cannot exclude the influence of additional selective forces unrelated to AA biosynthetic costs, as well as neutral processes, in shaping animal AA metabolism and proteome composition.

It should also be emphasized that animals still utilize all AAs. However, energetically expensive ones (EAAs) are sourced externally, allowing animals to conserve the energy that would be required for their production. We previously introduced the term "functional outsourcing" to describe this phenomenon, as simplified organisms still rely on the lost biosynthetic pathways, which are now outsourced to the environment[7]. This phenomenon is modeled by the Black Queen Hypothesis, which explains the evolution of ecological dependencies arising from gene loss. According to this hypothesis, some free-living organisms can lose a costly function if it is continuously provided by other organisms in the accessible environment[50,51].

Following this argumentation line, we think that the terms "essential" and "non-essential" AAs, which were coined within the context of animal dietary science, are quite misleading. All twenty proteinogenic amino acids are indispensable for every cellular organism on earth, and therefore essential. In fact, most of the non-essential AAs have to be supplemented to some extent by feeding[52]. In some cases, non-essential AA supplementation is necessary to ensure optimal health and fitness, leading some researchers in nutritional science to question the validity of the terms "essential" and "non-essential" [53,54]. Therefore, the only relevant question is whether they are internally or externally synthesized, or in other words, built in-house or outsourced[7]. We thus advocate that the terms auxotrophic amino acids (AAAs) and prototrophic amino acids (PAAs) should be adopted when the dispensability of AAs is discussed, regardless of which organism on the tree of life is in focus.

In summary, we have shown that the biosynthesis of nutritionally essential AAs is far more expensive than that of non-essential ones, regardless of the cost measures applied. We also found evidence consistent with the idea that the loss of EAA biosynthesis capability is an important synapomorphy of animals, driven by natural selection related to energy-saving requirements. Around 575 million years ago, Earth's ecosystems were fundamentally impacted by the oxygenation of the oceans[55] and by new ecological interactions that emerged alongside the rise of animals and their heterotrophic lifestyle[13]. Although alternative possibilities cannot be definitively excluded, we propose that the selective trade-off between energy economy and pleiotropy favored the loss of costly AA biosynthesis until an equilibrium between functional outsourcing and functional autonomy was reached, resulting in the so-called essential and non-essential AA sets.

We further propose that this important metabolic outsourcing enabled animals to use energetically expensive AAs more freely in both existing and newly arisen genes[7,35–37]. Whether this opportunity also enabled animals to evolve functionally more optimized proteins remains to be explored. We also suspect that the first animals, by outsourcing the production of expensive amino acids, opened a window of opportunity which allowed them to relocate a substantial part of their energy budget from costly metabolic synthesis to more animal-specific and energy-demanding functions, such as muscle movements, electric signaling, and increased protein glycosylation[56]. In this context, we believe that the recent suggestion that the origin of animals was a stochasticity-guided black swan event[32] is likely premature.

Methods

AA cost estimates and pleiotropy

We retrieved data on the number of ATP and NAD(P)H used in the biosynthesis of AAs from Kaleta et al.[12]. Previous studies primarily relied on estimates from Craig and Weber[18], which contain some inaccuracies. For instance, ATP values for asparagine and serine were erroneously transcribed from the original source[57]. Additionally, minor updates were made to the ATP and NAD(P)H values for arginine, cysteine, histidine, methionine, and tryptophan based on the most recent biochemical findings[12]. We also retrieved the number of ATP and NAD(P)H utilized in the metabolism of AA precursors[58]. To approximate ATP generation under different respiratory conditions, we converted the reducing equivalents to ATP as follows: (i) "high respiration," representing fully functional oxidative phosphorylation: 1 NAD(P)H = 2 FADH2 = 2 ATP[12]; (ii) "low respiration," representing oxidative phosphorylation without proton pumping at complex I: 1 NAD(P)H = 2 FADH2 = 1 ATP, which corresponds, for instance, to the metabolism of *S. cerevisiae* and some *E. coli* strains[21,59]; (iii) "fermentation," representing anaerobic conditions without conversion of reducing equivalents to ATP.

We calculated the costs of AA production in two ways: direct cost and opportunity cost. The direct cost was calculated by summing ATP and ATP equivalents of energy-bearing metabolites (NADH and NADPH) used in the production of an AA, starting from the precursor molecules (Supplementary Table 1 and Supplementary Data 1). The opportunity cost was calculated by summing 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 (Supplementary Table 1 and Supplementary Data 1). The energy gained from the catabolism of precursors was calculated by summing produced ATP and ATP equivalents of energy-bearing metabolites (NADH, NADPH, and FADH2) (Supplementary Table 1 and Supplementary Data 1). The opportunity cost measure, which has been extensively used in previous work[9–11,18], quantifies how a decision to synthesize an AA impacts the overall energetic balance of a cell by taking into account an unrealized energy gain from precursors (lost opportunity). In essence, opportunity cost shows how much energy is given up by an organism when producing AAs[18], and is analogous to the "opportunity cost" concept in economics[60].

Data on pleiotropy were collected from the KEGG database[4] and are represented by two metrics: (i) the number of KEGG reactions, and (ii) the number of KEGG pathways in which each amino acid (AA) is involved.

Combinatorial phenotype selection (CPS) test

We define the CPS test as a statistical method used to evaluate whether a quantitative factor (e.g., synthesis cost or pleiotropy) exerts directional selection on a combinatorial phenotypic trait (e.g., the auxotrophic/prototrophic status of amino acids), using neutral evolution as the null hypothesis. The CPS test estimates the probability (*p*-value) that an observed combinatorial phenotypic state arose by chance with respect to the tested factor, based on an empirically derived PMF obtained through permutations. In other words, the permutations represent all possible alternative phenotypes that could have arisen during the evolution of the trait, allowing us to compare these potential outcomes with the phenotype observed in nature. In this way, we explored the entire combinatorial space of possible phenotypic states.

A similar statistical approach was previously used to assess the positional evolutionary bias of enzymes on the endoplasmic reticulum membrane[56]. The ability to synthesize AAs is a unique combinatorial phenotype because 20 biogenic AAs make a relatively large set of biochemically similar molecules that all forms of life use as building blocks for proteins. A binary state of prototrophy/auxotrophy can be easily assigned to the AA set, allowing for the straightforward exploration of all possible combinations through permutations. Although the CPS test could theoretically be applied to other similar data, it was specifically designed for this particular phenotype.

The CPS tests (Figs. 4 and 5; Supplementary Figs. 5 and 6) were performed by calculating the average value of each factor (direct cost, opportunity cost, number of pathways, number of reactions) across all possible permutations for a given number of EAAs. For instance, in the case of nine or eleven EAAs, there are 167,960 possible permutations. Because the number of distinct average values is finite, we grouped them into bins containing identical values. This produced a distribution of averages for each factor, conditional on the EAA group size. We then calculated the proportion of permutations falling into each bin by dividing the number of elements in the bin by the total number of permutations. The resulting distribution forms an empirical PMF for a given factor and EAA group size, which was used to estimate the probability that the observed EAA set could have arisen randomly. These probabilities (*p*-values) were calculated by summing the proportions of permutations from the observed real-world value to the most extreme value at the nearest tail of the distribution.

The central idea of the CPS test is to estimate the likelihood that a combinatorial phenotypic state (e.g., the EAA set) could have arisen by chance with respect to a given quantitative factor. If the factor had no influence on the composition of the observed group, the group's average value would be expected to fall near the center of the distribution—yielding a high *p*-value. Conversely, if the tested factor contributed to shaping the group's composition, the average value would likely fall near one of the distribution's tails, resulting in a low *p*-value. Importantly, a low *p*-value for one factor does not exclude the possibility that other factors, acting in concert, also exerted selective pressure on the group's composition. The code used to perform the CPS test is available in a GitHub repository (see "Code availability").

Analyses of AA frequencies in proteomes

The proteomes of non-animal holozoans (27 species) and metazoans (140 species) were taken from our previous study (Supplementary Data 4)[7]. Using this dataset, we calculated the frequency of each of the 20 AAs in each proteome. By summing frequencies, we calculated the cumulative frequency of nine AAs, which are generally considered to be essential in Metazoa (Thr, Val, Leu, Lys, Ile, Met, His, Phe, Trp). We repeated this analysis by adding two additional AAs that could be considered conditionally essential (Cys, Tyr) (Fig. 7a, b; Supplementary Fig. 15).

Using AA frequencies, we also calculated the opportunity cost and the direct cost of an average AA in a proteome ($OC_{proteome}$ and $DC_{proteome}$) (Fig. 7c, d). This value represents a weighted mean of AA biosynthesis costs where the frequencies of twenty AAs in a proteome are acting as weights (Supplementary Data 4), using the following equations[29,41]:

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

$$DC_{proteome} = \frac{\sum_{i=1}^{n=20} DC_i \times N_i}{\sum_{i=1}^{n=20} N_i} = \sum_{i=1}^{n=20} DC_i \times F_i \tag{2}$$

In these equations, OC_i represents the opportunity cost of amino acid i, DC_i represents the direct cost of amino acid i, 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.

To control for the presence of lineage-specific genes, we repeated this procedure using only genes conserved between animals and non-animal holozoans (Supplementary Figs. 14 and 15). We first clustered all 167 proteomes using the following MMSeqs2 (version 14-7e284) parameters: -e 0.001 -c 0.8 --max-seqs 400 --cluster-mode 1 --cov-mode 0. By using a cutoff of $c = 0.8$, we obtained clusters of proteins with highly similar architectures[7]. We then extracted only clusters that contained both metazoan and non-metazoan holozoan proteins and repeated the frequency and energy cost analyses. The initial dataset contained 3,514,971 proteins, while the dataset with only conserved genes contained 1,460,551 proteins (42%).

To examine correlations between AA costs and AA frequencies (Supplementary Data 3), we calculated average AA frequencies and compared them with AA opportunity costs. We calculated correlations between AA costs and AA frequencies for each species individually (Supplementary Data 3) using the Pearson correlation coefficient.

AA biosynthesis pathway completeness

To construct the heatmaps depicting the completeness of AA biosynthetic pathways (Fig. 6, Supplementary Figs. 8–13), we first assembled a list of all enzymes involved in these pathways from the MetaCyc[61] and KEGG[4] 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 protein sequences from the KEGG database for 1200 available eukaryotic organisms resulting in a total of 645,363 protein sequences (Supplementary Data 4).

We used this collection of AA biosynthesis enzyme sequences as a query to perform a reciprocal best hit search, as implemented in MMSeqs2[62], against our database of 167 holozoan proteomes[7]. The search was conducted using the "mmseqs easy-rbh" option with default parameters. To test the robustness of our results, we considered the reciprocal best hit results at three e-value cutoffs: 10^{-10}, 10^{-30}, and 10^{-60} (Supplementary Figs. 8–10). In line with earlier study, which has shown that functional orthologs always have *e*-values lower than 10^{-60}[5], we presented in the main text the results at this *e*-value cutoff (Fig. 6).

To further verify results obtained by reciprocal best hit search, we performed MMSeqs2 clustering for each species separately. For every species, we created an MMSeqs2 database consisting of a proteome of that single species and our collection of AA biosynthesis enzyme sequences. In the next step we preformed MMSeqs2 clustering using the following parameters: cluster -c 0.8, --cov-mode 0, --cluster-reassign (Supplementary Figs. 11–13). By using the -c 0.8 parameter, we generated clusters of proteins with highly similar architectures with at least 80% shared alignment length[7,29,41]. Similar to reciprocal best hit search, to test the robustness of our findings, we repeated clustering at three e-value cutoffs: 10^{-10}, 10^{-30}, and 10^{-60} (Supplementary Figs. 11–13).

Finally, for each AA and species, we divided the number of detected enzymes in an AA synthesis pathway *i* (DE_i) by the expected total number of enzymes present in that pathway (E_i), resulting in the completeness score of the AA synthesis pathway *i* (CS_i) ranging from 0 to 1. If alternative AA biosynthetic pathways were present in a species, we selected the most complete one. The CS_i was calculated using the following equation:

$$CS_i = \frac{DE_i}{E_i} \tag{3}$$

For all considered holozoan species (27 non-animal holozoans, 140 metazoans), we computed the average auxotrophy index (AI_{avg}) as described in Kasalo et al.[29,41]. For each species, we first calculated the AA auxotrophy index of each amino acid *i* (AI_i), where *i* denotes one of 20 AAs, by subtracting the completeness score (CS_i) from 1. We then averaged AI_i values across all amino acids in a given species to obtain AI_{avg}, a measure of the species' overall auxotrophy for AA biosynthesis pathways. In theory, AI_{avg} values range from 0 (fully prototrophic) to 1 (fully auxotrophic) and are calculated using the following equation:

$$AI_{avg} = \frac{\sum_{i=1}^{i=20} AI_i}{20} = \frac{\sum_{i=1}^{i=20}(1 - CS_i)}{20} \tag{4}$$

We calculated AI_{avg} values for all six versions of our AA biosynthesis completeness search and used the Mann-Whitney U test to determine if there is a statistically significant difference in AI_{avg} values between metazoans and non-animal holozoans (Fig. 6, Supplementary Figs. 8–13). The code used to detect AA biosynthesis pathway completeness is available in a GitHub repository (see "Code availability').

AA auxotrophy in individual metazoan species

Building on the pathway completeness analysis, we next examined AA auxotrophy in each metazoan species individually. For every species, we classified an AA as auxotrophic (i.e., part of the EAA set) if its completeness score was below 1, as determined by the reciprocal best hit protocol with an *e*-value cutoff of 10^{-60} (Fig. 6 and Supplementary Fig. 8). Using these species-specific EAA sets, we compared the opportunity costs of EAAs and NEAAs with the Mann-Whitney U test. We also estimated the probability that energy-related selection was implicated in amino acid outsourcing using the CPS test (Supplementary Data 2 and 4).

Statistical analyses and tools

The significance of differences in the energy costs between EAAs and NEAAs, differences in the frequency of EAA usage, differences in the energy cost of an average AA, and differences in the number of reactions and pathways was tested by the two-tailed Mann-Whitney U with continuity correction. For each plot, we report the Mann-Whitney statistic (W), the *p*-value, the effect size (*r*), and the 95% non-parametric confidence intervals. The Mann-Whitney U tests were conducted in the R environment using the package rcompanion (version 2.4.36).

To calculate Pearson correlations, we used the cor.test() function in the R stats package (version 3.6.2). Permutations were calculated using the RcppAlgos package (version 2.9.3). Phylogeny mapping was performed using the ggtree package (version 3.10.1)[63].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data generated in this study are available in the supplementary materials and Figshare[64] at: 10.6084/m9.figshare.30861554. Genome accession codes are available in the Supplementary Data 4 file. Source data are provided as a Source Data file. Source data are provided with this paper.

Code availability

The code used to perform the combinatorial phenotype selection test and auxotrophy detection is available in a GitHub repository[65]: https://github.com/nkasalo/MassiveOutsourcing.

References

1. Guedes, R. et al. Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. *BMC Genom.* **12**, S2 (2011).
2. D'Souza, G. et al. Less is more: selective advantages can explain the prevalent loss of biosynthetic genes in bacteria. *Evolution* **68**, 2559–2570 (2014).
3. Trolle, J. et al. Resurrecting essential amino acid biosynthesis in mammalian cells. *eLife* **11**, e72847 (2022).
4. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
5. Payne, S. H. & Loomis, W. F. Retention and loss of amino acid biosynthetic pathways based on analysis of whole-genome sequences. *Eukaryot. Cell* **5**, 272–276 (2006).
6. Richter, D. J., Fozouni, P., Eisen, M. B. & King, N. Gene family innovation, conservation and loss on the animal stem lineage. *eLife* **7**, e34226 (2018).

7. Domazet-Lošo, M., Široki, T., Šimičević, K. & Domazet-Lošo, T. Macroevolutionary dynamics of gene family gain and loss along multicellular eukaryotic lineages. *Nat. Commun.* **15**, 2663 (2024).
8. Guijarro-Clarke, C., Holland, P. W. H. & Paps, J. Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat. Ecol. Evol.* **4**, 519–523 (2020).
9. Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proc. Natl. Acad. Sci. USA* **99**, 3695–3700 (2002).
10. Wagner, A. Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* **22**, 1365–1374 (2005).
11. Zhang, H. et al. Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat. Commun.* **9**, 4124 (2018).
12. Kaleta, C., Schäuble, S., Rinas, U. & Schuster, S. Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnol. J.* **8**, 1105–1114 (2013).
13. Judson, O. P. The energy expansions of evolution. *Nat. Ecol. Evol.* **1**, 1–9 (2017).
14. Kepp, K. P. Survival of the cheapest: how proteome cost minimization drives evolution. *Q. Rev. Biophys.* **53**, e7 (2020).
15. Seligmann, H. Cost-minimization of amino acid usage. *J. Mol. Evol.* **56**, 151–161 (2003).
16. Swire, J. Selection on synthesis cost affects interprotein amino acid usage in all three domains of life. *J. Mol. Evol.* **64**, 558–571 (2007).
17. Heizer, E. M., Raymer, M. L. & Krane, D. E. Amino acid biosynthetic cost and protein conservation. *J. Mol. Evol.* **72**, 466–473 (2011).
18. Craig, C. L. & Weber, R. S. Selection costs of amino acid substitutions in ColE1 and ColIa gene clusters harbored by Escherichia coli. *Mol. Biol. Evol.* **15**, 774–776 (1998).
19. Basan, M. et al. Overflow metabolism in Escherichia coli results from efficient proteome allocation. *Nature* **528**, 99–104 (2015).
20. Ferguson, S. J. ATP synthase: from sequence to ring size to the P/O ratio. *Proc. Natl. Acad. Sci. USA* **107**, 16755–16756 (2010).
21. Kok, S., Kozak, B. U., Pronk, J. T. & Maris, A. J. A. Energy coupling in Saccharomyces cerevisiae: selected opportunities for metabolic engineering. *FEMS Yeast Res.* **12**, 387–397 (2012).
22. Fernandes, J. I. M. & Murakami, A. E. Arginine metabolism in uricotelic species. *Acta Sci. Anim. Sci.* **32**, 357–366 (2010).
23. Li, Y. et al. Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm Lamellibrachia luymesi. *BMC Biol.* **17**, 91 (2019).
24. de Oliveira, A. L., Mitchell, J., Girguis, P. & Bright, M. Novel insights on obligate symbiont lifestyle and adaptation to chemosynthetic environment as revealed by the giant tubeworm genome. *Mol. Biol. Evol.* **39**, msab347 (2022).
25. Lasser, G. W. & Allen, W. V. The essential amino acid requirements of the Dungeness crab. *Cancer Magister Aquac.* **7**, 235 (1976).
26. Cowey, C. B., Adron, J. & Blair, A. Studies on the nutrition of marine flatfish the essential amino acid requirements of plaice and sole. *J. Mar. Biol. Assoc. U. K.* **50**, 87–95 (1970).
27. Zečić, A., Dhondt, I. & Braeckman, B. P. The nutritional requirements of Caenorhabditis elegans. *Genes Nutr.* **14**, 15 (2019).
28. Lan, Y. et al. Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail Gigantopelta aegis. *Nat. Commun.* **12**, 1165 (2021).
29. Kasalo, N., Domazet-Lošo, T. & Domazet-Lošo, M. Bacterial amino acid auxotrophies enable energetically costlier proteomes. *IJMS* **26**, 2285 (2025).
30. Scaraffia, P. Y. et al. Discovery of an alternate metabolic pathway for urea synthesis in adult *Aedes aegypti* mosquitoes. *Proc. Natl. Acad. Sci. USA* **105**, 518–523 (2008).
31. Ros-Rocher, N., Pérez-Posada, A., Leger, M. M. & Ruiz-Trillo, I. The origin of animals: an ancestral reconstruction of the unicellular-to-multicellular transition. *Open Biol.* **11**, 200359 (2021).
32. Ruiz-Trillo, I., Kin, K. & Casacuberta, E. The origin of metazoan multicellularity: a potential microbial black swan event. *Annu. Rev. Microbiol.* **77**, 499–516 (2023).
33. Stanley, S. M. An ecological theory for the sudden origin of multicellular life in the late precambrian. *Proc. Natl. Acad. Sci. USA* **70**, 1486–1489 (1973).
34. Bengtson, S. Origins and early evolution of predation. *Paleontol. Soc. Pap.* **8**, 289–318 (2002).
35. Domazet-Loso, T. & Tautz, D. An evolutionary analysis of orphan genes in drosophila. *Genome Res.* **13**, 2213–2219 (2003).
36. Domazet-Lošo, T., Brajković, J. & Tautz, D. A phylostratigraphy approach to uncover the genomic history of major adaptations in metazoan lineages. *Trends Genet.* **23**, 533–539 (2007).
37. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**, 692–702 (2011).
38. Cochet-Escartin, O. et al. Hypoxia triggers collective aerotactic migration in Dictyostelium discoideum. *eLife* **10**, e64731 (2021).
39. Duarte, M. et al. *Leishmania* type II dehydrogenase is essential for parasite viability irrespective of the presence of an active complex I. *Proc. Natl. Acad. Sci. USA* **118**, e2103803118 (2021).
40. Shreenidhi, P. M., Brock, D. A., McCabe, R. I., Strassmann, J. E. & Queller, D. C. Costs of being a diet generalist for the protist predator *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA* **121**, e2313203121 (2024).
41. Kasalo, N., Domazet-Lošo, M. & Domazet-Lošo, T. Convergence in amino acid outsourcing between animals and predatory bacteria. *IJMS* **26**, 3024 (2025).
42. Marcet-Houben, M., Marceddu, G. & Gabaldón, T. Phylogenomics of the oxidative phosphorylation in fungi reveals extensive gene duplication followed by functional divergence. *BMC Evol. Biol.* **9**, 295 (2009).
43. Walker, M. C. & van der Donk, W. A. The many roles of glutamate in metabolism. *J. Ind. Microbiol. Biotechnol.* **43**, 419–430 (2016).
44. Yoo, H. C., Yu, Y. C., Sung, Y. & Han, J. M. Glutamine reliance in cell metabolism. *Exp. Mol. Med.* **52**, 1496–1516 (2020).
45. de Carvalho Fernandes, G., Turchetto-Zolet, A. C. & Pereira Passaglia, L. M. Glutamine synthetase evolutionary history revisited: tracing back beyond the Last Universal Common Ancestor. *Evolution* **76**, 605–622 (2022).
46. Krall, A. S., Xu, S., Graeber, T. G., Braas, D. & Christofk, H. R. Asparagine promotes cancer cell proliferation through use as an amino acid exchange factor. *Nat. Commun.* **7**, 11457 (2016).
47. Zhu, Y. et al. A critical role of glutamine and asparagine γ-nitrogen in nucleotide biosynthesis in cancer cells hijacked by an oncogenic virus. *mBio* **8**, e01179-17 (2017).
48. Holeček, M. Serine metabolism in health and disease and as a conditionally essential amino acid. *Nutrients* **14**, 1987 (2022).
49. Christgen, S. L. & Becker, D. F. Role of proline in pathogen and host interactions. *Antioxid. Redox Signal.* **30**, 683–709 (2019).
50. Morris, J. L. et al. The timescale of early land plant evolution. *Proc. Natl. Acad. Sci. USA* **115**, E2274–E2283 (2018).
51. Mas, A., Jamshidi, S., Lagadeuc, Y., Eveillard, D. & Vandenkoornhuyse, P. Beyond the black queen hypothesis. *ISME J.* **10**, 2085–2091 (2016).
52. Wu, G. et al. Amino acid nutrition in animals: protein synthesis and beyond. *Annu. Rev. Anim. Biosci.* **2**, 387–417 (2014).
53. Wu, G. et al. Dietary requirements of "nutritionally non-essential amino acids" by animals and humans. *Amino Acids* **44**, 1107–1113 (2013).
54. Hou, Y., Yin, Y. & Wu, G. Dietary essentiality of 'nutritionally non-essential amino acids' for animals and humans. *Exp. Biol. Med.* **240**, 997–1007 (2015).

55. Brocks, J. J. et al. The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578–581 (2017).
56. Kifer, D. et al. Contrasting Macroevolutionary Patterns in the Human N-Glycosylation Pathway. *Engineering* S2095809925003819 https://doi.org/10.1016/j.eng.2025.06.039 (2025).
57. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the Bacterial Cell: An Molecular Approch* (Sinauer, Sunderland (Mass), 1990).
58. Voet, D., Voet, J. G. & Pratt, C. W. *Fundamentals of Biochemistry: Life at the Molecular Level* (John Wiley & Sons, 2016).
59. Schuetz, R., Kuepfer, L. & Sauer, U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Mol. Syst. Biol.* **3**, 119 (2007).
60. Mankiw, G. *Principles of Economics* (Cengage Learning, 2023).
61. Caspi, R. et al. The MetaCyc database of metabolic pathways and enzymes - a 2019 update. *Nucleic Acids Res.* **48**, D445–D453 (2020).
62. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
63. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. ggtree: an package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
64. Kasalo, N., Domazet-Lošo, M. & Domazet-Lošo, T. Outsourcing of energetically costly amino acids at the origin of animals supplementary data. https://doi.org/10.6084/m9.figshare.30861554 (2025).
65. Kasalo, N., Domazet-Lošo, M. & Domazet-Lošo, T. Outsourcing of energetically costly amino acids at the origin of animals. *MassiveOutsourcing* https://doi.org/10.5281/zenodo.17900057 (2025).

Acknowledgements

We thank M. Futo, A. Tušar, S. Koska, D. Franjević, and G. Klobučar for discussions. This work was supported by the Croatian Science Foundation under the project IP-2016-06- 5924 (T.D.-L.), the City of Zagreb (T.D.-L.), the Adris Foundation (T.D.-L.), the European Regional Development Fund, PK.1.1.10.0007 DATACROSS (M.D.-L. and T.D.-L.). We used the computational resources of the University Computing Center—SRCE (Padobran) and the Institute Ruđer Bošković. N.K. is enrolled in the doctoral program in Biology at the Department of Biology, Faculty of Science, University of Zagreb.

Author contributions

T.D.-L. and M.D.-L. initiated the study. N.K., M.D.-L., T.D.-L. conceptualized and performed the analyses. N.K. prepared the figures and tables for publication. N.K., M.D.-L., and T.D.-L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-68724-6.

Correspondence and requests for materials should be addressed to Mirjana Domazet-Lošo or Tomislav Domazet-Lošo.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

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