Massive outsourcing of energetically costly amino acids at the origin of animals

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

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

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

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

*Corresponding authors. Emails: mirjana.domazet@fer.hr, tdomazet@irb.hr

Animals are generally capable of synthesizing eleven amino acids, while the remaining nine, often referred to as essential, have to be acquired through diet. This has a profound impact on animals by defining their ecological lifestyle and evolutionary trajectory. Recent phylogenomic studies reveal that this phenotype is a result of gene losses that occurred at the root of animal tree. However, it remains unclear which selective forces, if any, were directing this far-reaching metabolic simplification event. Here, we show that the essential amino acids are energetically far more expensive to synthesize than non-essential ones and that animals, by outsourcing their production, significantly removed constraints on their usage. This suggests that the selective pressures on energy management drove this metabolic outsourcing, which in turn allowed animal genes to evolve more freely through protein sequence space.

Introduction

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 are 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 a part of a broader genome simplification episode that preceded metazoan radiation, which included a wide range of metabolic simplifications (*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*). However, there is considerable variability among AAs in the energy required for their production, with the cheapest and most expensive AAs differing by almost an order of magnitude (*9–11*). On the other hand, free energy is a critical and limited resource that sustains biological systems, meaning that its harnessing and efficient usage is under continuous selective pressure (*12*, *13*). For instance, energetically expensive AAs sparingly occur in protein sequences, especially in highly expressed genes (*9*, *14*, *15*). This selectively favored bias leads to the net reduction of energy costs related to AA synthesis in cells (*12*, *13*).

We reasoned that the massive loss of EAAs 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 the energy budget that could subsequently have been reallocated to other functional needs. We further assumed that the outsourcing of expensive EAAs would allow animals to use them more freely in their proteomes, compared to their unicellular choanozoan ancestor, since the energetic constraints on their usage would be reduced. Surprisingly, these possibilities have never been tested, although the estimates of AA biosynthesis energy costs have been available for a long time (*9–11*) and AA biosynthetic capabilities have been studied from various angles (*1–3*, 5, *11*).

Results

Essential amino acids are expensive

To address these questions, we first retrieved AAs biosynthesis costs that are currently available in the literature, which includes estimates for *Escherichia coli*, *Saccharomyces cerevisiae*, and *Homo sapiens* (*9–11*). These energy costs are estimated as a number of high-energy phosphate bonds (~P) required for the synthesis of an AA from its metabolic precursors (*9–11*). We then sorted 20 AAs in increasing order according to the energy cost estimates in yeast (Fig. 1A). We chose yeast because it is phylogenetically the closest organism to metazoans which possesses an intact AA synthesis machinery and for which AA energy synthesis cost estimates are available (*10*). If humans are used as a proxy for the AA essentiality status in animals, AAs cluster 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, 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% most expensive AAs are outsourced (Fig. 1A).

To statistically evaluate observed differences in energy synthesis costs between EAAs and NEAAs in animals we compared these two groups using a non-parametric test (Fig. 1B–E). We made comparisons using either *E. coli* or *S. cerevisiae* AA energy cost estimates. In addition, we considered AA dispensability in animals in two versions; in the first classical one we grouped 9 EAAs, whereas in the second one, we extended this set to 11 members by treating cysteine (C) and tyrosine (Y) as EAAs. Regardless of which test version we performed, the average energy costs of EAAs and NEAAs differ significantly with large effect sizes ($r > 0.5$) (Fig. 1B–E). This suggests that the high energy costs of EAAs were imposing substantial selective pressures, which eventually were resolved by the loss of the underlying biosynthesis pathways at the root of Metazoans.

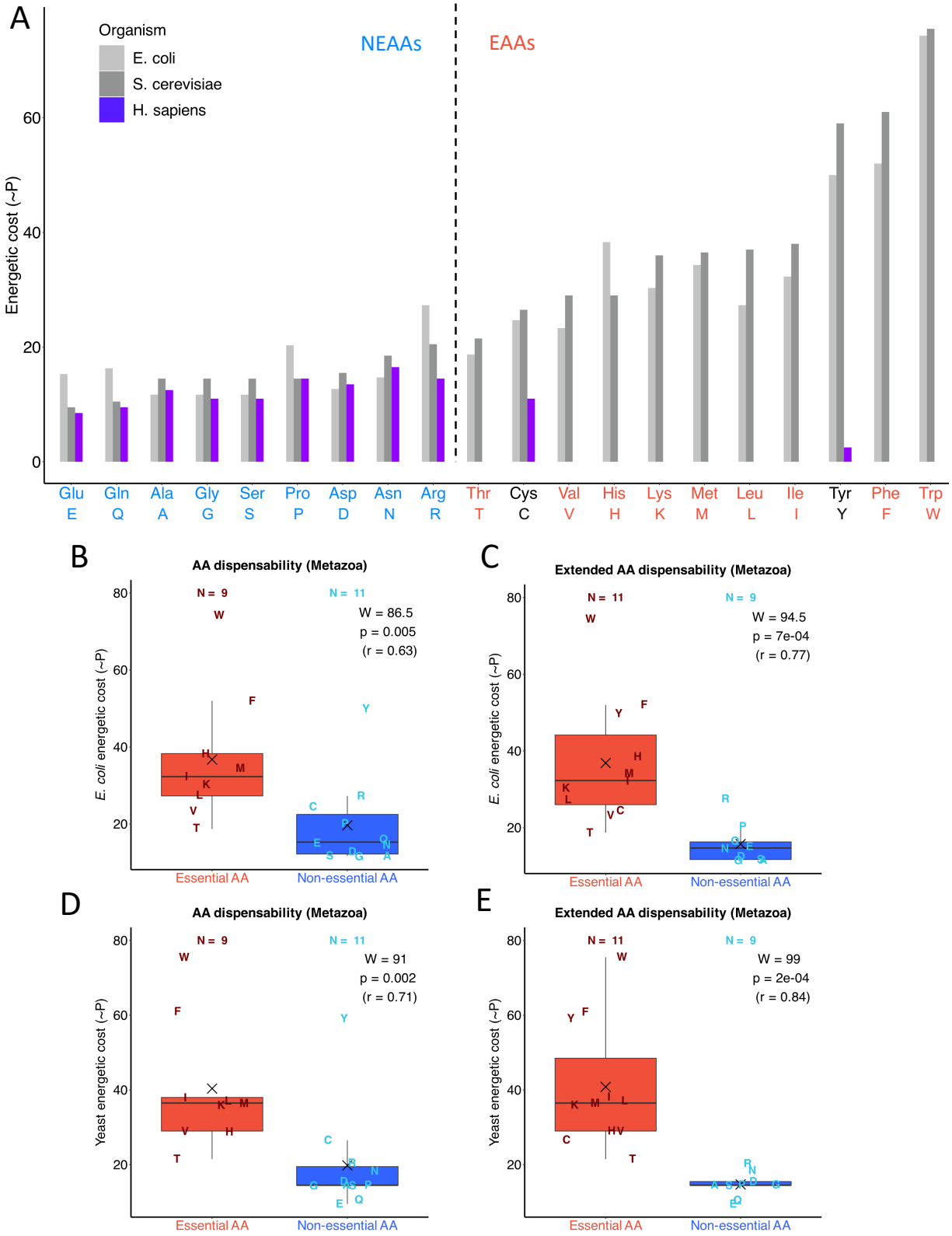


Fig. 1. The comparison of amino acid synthesis costs. **(A)** The energy costs of amino acids

synthesis in *E. coli*, *S. cerevisiae*, and *H. sapiens* are shown as a number of high-energy phosphate bonds (~P). Amino acids are sorted in ascending order according to the energy costs in yeast. (**B–E)** The comparison of energy costs between essential and non-essential amino acids in Metazoa. We took energy cost estimates form *E. coli* (*9*) and yeast (*10*) (Supplementary Table 1). We performed the comparison by considering 9 essential and 11 non-essential amino acids (**B, D**), and by considering 11 essential and 9 non-essential amino acids (**C, E**). In the extended AA dispensability analyses (**C, E**), 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. The differences in energy costs between the two groups were shown by boxplots and the significance of these differences was tested by the Mann-Whitney U test with continuity correction. We depicted the corresponding W-value, p-value, and effect size (r) in each panel. The X symbol represents the mean and individual AAs are shown by one-letter symbols.

Animal proteomes are expensive

However, these metabolic simplifications were only possible under the condition that the last common ancestor of animals (animal LCA) was able to secure in surplus these costly EAAs from the environment. If we assume that the unicellular ancestor of animals was a prototroph for AAs, like the present-day choanoflagellates, then some important shift in the 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 (*16*, *17*) to more efficient multicellular-style suspension/filter feeding, or possibly to some sort of grazing behavior or macrophagous predation (*18*, *19*).

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. To test this idea, we compared the usage of EEAs between Choanoflagellata and Metazoa (Fig. 2, Supplementary Data 1). This comparison showed that animals use on average more EAAs in their proteins than choanoflagellates (Fig. 2A and B). Similarly, we also found that the energy cost of an average AA is higher in metazoan than choanoflagellate proteomes (Fig. 2C and D). These results are in line with our prediction that the reduction of selective constraints on energetically expensive AAs allowed animals to more freely explore protein sequence space (*20–22*).

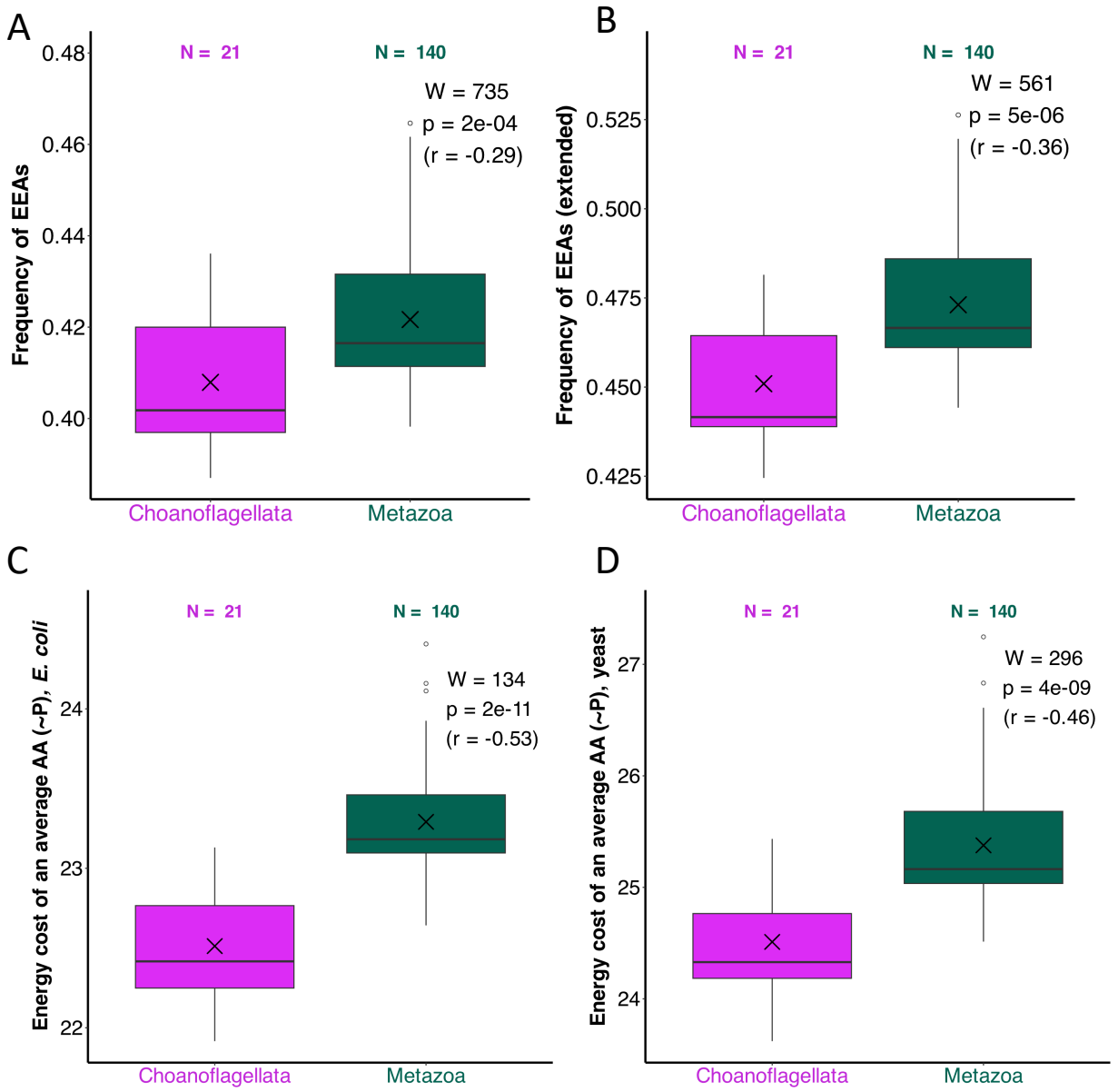


Fig. 2. The comparison of EAAs usage and the energy costs of an average AA in Choanoflagellata and Metazoa. (**A**) The cumulative frequency of nine AAs which are generally considered to be essential in Metazoans (Thr, Val, Leu, Lys, Ile, Met, His, Phe, Trp). (**B**) The cumulative frequency of eleven EAAs set. This extended dataset contains two AAs which could be considered conditionally essential (Cys, Tyr). (**C**–**D**) The energy cost of an average AA computed using *E. coli* (*9*) and yeast (*10*) estimates. This value represents a weighted mean of AAs energy biosynthesis costs where the frequencies of twenty AAs in a proteome act as weights. The differences in energy costs between the two groups were shown by boxplots and the significance of these differences was tested by the Mann-Whitney U test with continuity correction. We

depicted the corresponding W-value, p-value, and effect size (r) in each panel. The X symbol represents the mean. The list of choanoflagellates (21) and metazoans (140) whose proteomes are included in calculations are available in Supplementary Data 1 (*7*)

AAs dispensability in animals

Nine AAs are canonically considered to be essential in animals, although there are many exceptions to this rule, often related to further loss in biosynthesis capabilities. To get a more inclusive overview of AAs dispensability in animals, we compiled the available information on AAs dispensability across the metazoan tree, alongside several other eukaryotes and *E. coli* as outgroups (Fig. 3, Supplementary Data 2). It is evident that the set of nine EAAs is occasionally expanded by the complete loss of the biosynthesis capability for the two partially essential amino acids (cysteine, C and tyrosine, Y) (Fig. 3). Interestingly, the capability for self-production of arginine (R) — energetically the costliest AA among canonical NEAAs — is also commonly absent in many metazoan groups (Fig. 3). Whether some animal group will retain the potential to synthesize arginine (R) is correlated with the nitrogen excretion mechanism used by that particular clade.

For instance, ureotelic animals (mainly amphibians and mammals), which excrete excess nitrogen as urea, must retain arginine (R) synthesis because this process is a part of the urea cycle (*23*). Conversely, ammoniotelic and uricotelic animals, which excrete excess nitrogen as ammonia and uric acid respectively, do not require arginine for these processes (*23*). This lack of pleiotropic effects related to nitrogen excretion allowed ammoniotelic and uricotelic animals (i.e., most metazoans in Fig. 3) to further optimize their energy management via the loss of arginine biosynthesis capability (*5*, *23*). Another interesting pattern can be seen with glutamine, which is the only AA that is non-essential in all of the examined species (Fig. 3). This could be due to the fact that glutamine is vital for many fundamental metabolic functions, such as nitrogen metabolism, nucleotide synthesis, NEAA synthesis, and the control of epigenetic changes (*24*). Glutamine synthetase is one of the oldest known functioning genes, which was already present in the last universal common ancestor and is thus likely a fundamental element of cellular metabolism (*25*).

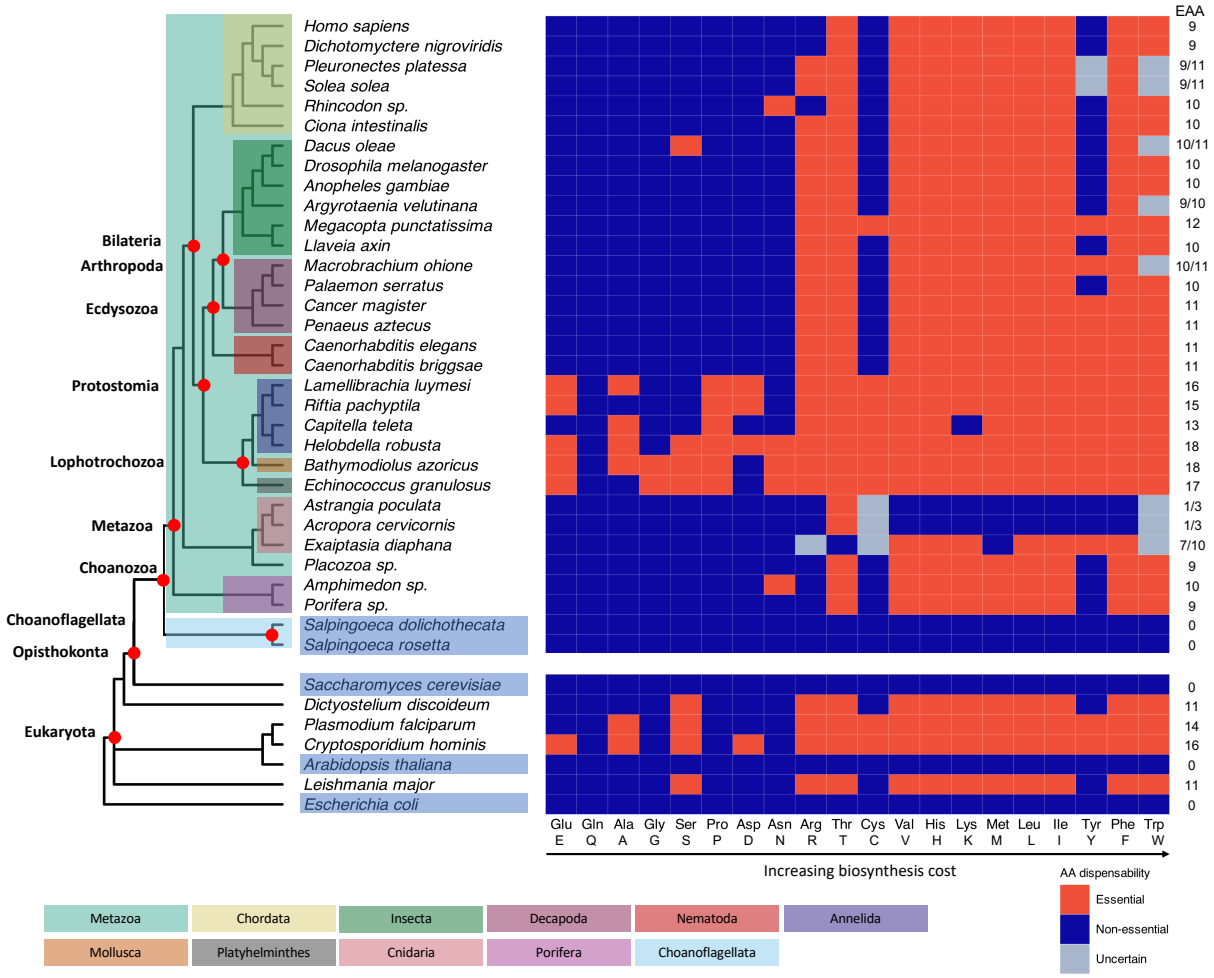


Fig. 3. An overview of AA dispensability in Metazoa. We extracted the dispensability status of amino acids (i.e., essential vs. non-essential) for various organisms that are available in the literature (Supplementary Data 2) and mapped it to the consensus phylogeny. The major animal clades are labeled on the phylogeny and color-coded. The names of taxa with no AA auxotrophies are shaded in blue.

Among major animal clades, the largest reduction in the AA biosynthesis capability is apparent in Lophotrochozoa (Fig. 3, Supplementary Fig. 1 and 2). For instance, the sediment-dwelling annelid *Capitella teleta* is auxotrophic for 14 AAs. Even more extensive reductions are evident in the annelids that inhabit deep-sea vents, *Riftia pachyptila* (15 EAAs) and *Lamellibranchia luymesi* (16 EAAs), while the bivalve *Bathymodiolus azoricus*, which is also associated with deep-sea vents, is auxotrophic for 18 AAs (*26*). Deep-sea vents are environments in which primary production is carried out by chemoautotrophs. Most of the fauna present at deep-sea vents harbor

169 chemoautotroph bacterial endosymbionts, which are entangled with their hosts in an array of
170 interdependent metabolic interactions (*27*, *28*). These tight biological interactions obviously
171 allowed deep-sea animals to almost entirely outsource AA production to their endosymbionts (*7*),
172 which consequently lowers their energy requirements in these unique ecosystems.

173 Extreme metabolic reductions can also be observed in lophotrochozoans which are obligate
174 parasites (Fig 3). In our analysis these are the leech *Helobdella robusta* (18 EAAs) and the
175 tapeworm *Echinococcus granulosus* (17 EAAs) (Fig. 3). The life cycles of these species are
176 intimately linked with their hosts and their biosynthetic capabilities are likely reduced due to the
177 direct availability of nutrients (*5*). Like the symbiotic interactions at the deep-sea vents, parasitism
178 also creates biological interactions that favor severe functional outsourcing of energetically costly
179 metabolic pathways (*7*). It should be noted that currently there is no data on whether these losses
180 in the AA biosynthetic capabilities of lophotrochozoas are solely ecologically dependent, or if this
181 group has some metabolic predisposition that favors this evolutionary trajectory.

182 Among the outgroups (Fig. 3, Supplementary Fig. 1 and 2), there are some interesting cases of
183 convergence: the soil-dwelling amoeba *Dictyostelium discoideum* (11 EAAs) and the euglenozoan
184 parasite *Leishmania major* (11 EAAs) are nearly identical to metazoans in their AA dispensability,
185 with only serine being additionally essential. This parallelism implies that a similar selective force
186 (i.e., energy management) drives the loss of the canonical EAAs. More extreme reductions are
187 apparent in the parasitic alveolates *Plasmodium falciparum* (14 EAAs) and *Cryptosporidum*
188 *hominis* (16 EAAs). A likely explanation for this pattern is that these two parasites are exposed in
189 their life cycles to very homogenous environments that allow them to further reduce their
190 metabolic independence (*29*).

191

192 **Discussion**

193 Several studies examined the genes involved in AAs biosynthesis pathways and found that many
194 of them are missing in animals (e.g., *5*, *1, 28*). It was suggested that these reductions are caused
195 by animal heterotrophy that relaxed selective pressures on maintaining AA biosynthetic
196 capabilities, implying that the loss of individual AAs should be a random process (*5*, *1*). However,
197 this interpretation does not consider the energy required for AA production and does not provide
198 an explanation for many independent reductions in different clades resulting in approximately the

199 same sets of EAAs. It should be also stressed that all animals still utilize all AAs, but that the
200 energetically expensive ones (EAAs) are externally sourced, and thus the energy required for their
201 production is saved. We previously introduced the term “functional outsourcing” for this
202 phenomenon since the simplified organisms still depend on the lost biosynthetic pathways which
203 are merely outsourced to the environment (*7*).

204 Following this argumentation line, we think that the terms "essential" and "non-essential" AAs,
205 which were coined within the context of animal dietary science, are quite misleading. All twenty
206 proteinogenic amino acids are indispensable for every cellular organism on earth, and therefore
207 essential. In fact, most of the “non-essential” AAs have to be supplemented by feeding (*30*).
208 Therefore, the only relevant question is whether they are internally or externally synthesized, or in
209 other words, built in-house or outsourced (*7*). We thus advocate that the terms auxotrophic amino
210 acids (AAAs) and prototrophic amino acids (PAAs) should be adopted when the dispensability of
211 AAs is discussed, regardless which focal organism on the tree of life is in focus.

212 In summary, we have shown that the biosynthesis of nutritionally essential AAs is far more
213 expensive than the non-essential ones. This suggests that natural selection, driven by energy-
214 saving requirements, shaped this important synapomorphy of animals. This is further supported by
215 similar, but convergent, outsourcing of EAAs in other eukaryotic lineages. Actually, it was
216 experimentally shown that the bacterial strains which are auxotrophic for different amino acids
217 gain fitness advantage over their prototrophic counterparts (*2*). Around 575 million years ago,
218 Earth’s ecosystems were profoundly impacted by the new ecological interactions that appeared
219 with animals and their heterotrophic way of life (*12*). We propose that selection driven by energy
220 economy initially favored the loss of expensive AA biosynthesis until an equilibrium between
221 functional outsourcing and functional autonomy was reached, resulting in the so-called essential
222 and non-essential AA sets.

223 This important selection-driven metabolic outsourcing allowed animals to more freely use
224 energetically expensive AAs in already existing and newly arisen genes (*7, 20–22*). Whether this
225 opportunity also allowed animals to evolve functionally more optimized proteins has yet to be
226 explored. We also suspect that the first animals, by outsourcing the production of expensive amino
227 acids, opened a window of opportunity which allowed them to relocate a substantial part of their
228 energy budget from costly metabolic synthesis to more animal-specific and energy-demanding
229 functions, such as muscle movements and electric signaling. In this context, we think that the

230 recent suggestion that the origin of animals was a stochasticity-guided black swan event (*17*) is
231 probably premature.

232

233 **Methods**

234 We gathered the AAs biosynthesis costs, estimated as a number of high-energy phosphate bonds
235 (~P) required for the synthesis of an AA from its metabolic precursors, for *Escherichia coli*,
236 *Saccharomyces cerevisiae*, and *Homo sapiens* from the literature *(9–11*); Supplementary Table 1.
237 The data on AA dispensability in different taxa was also collected from the literature (*1, 3, 5, 6, 9*,
238 *31–44*) and mapped on the consensus phylogeny assembled following the relevant phylogenetic
239 papers (*45–49*). The proteomes of choanoflagellates (21 species) and metazoans (140 species)
240 were taken from our previous study (Supplementary Data 1) (*7*). Using this dataset, we calculated
241 for each considered proteome the frequency of 20 AAs. By taking these frequencies, we calculated
242 the cumulative frequency of nine AAs which are generally considered to be essential in Metazoans
243 (Thr, Val, Leu, Lys, Ile, Met, His, Phe, Trp). We repeated this analysis by adding to this set
244 additional two AAs which could be considered conditionally essential (Cys, Tyr). This set of 11
245 essential amino acids also represents the top 11 AAs with the highest biosynthesis costs. Using
246 AAs frequencies, we also calculated the cost of an average AA in each proteome. This value
247 represents a weighted mean of AAs energy biosynthesis costs where the frequencies of twenty
248 AAs in a proteome are acting as weights (Supplementary Data 1). The significance of differences
249 in the energy costs between EAAs and NEAAs, differences in the frequency of EAAs usage, and
250 differences in the energy cost of an average AA was tested by the two-tailed Mann-Whitney U test
251 with continuity correction. For each plot, we report the Mann-Whitney statistic (W-value), the p-
252 value, and the effect size r. All analyses were conducted in the R environment using packages
253 rcompanion (Mann Whitney U test implementation) and ggtree (phylogeny mapping).

254

255 **Data availability**

256 All data are available in the main text or the supplementary materials.

257

258 **Acknowledgments**

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 KK.01.1.1.01.0009 DATACROSS (M.D.-L., T.D.-L.). We used the computational resources of the University Computing Center -SRCE (Isabella) and the Institute Ruđer Bošković.

Author contribution

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., T.D.-L. wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

1. R. Guedes, F. Prosdocimi, G. Fernandes, L. Moura, H. Ribeiro, J. Ortega, Amino acids biosynthesis and nitrogen assimilation pathways: a great genomic deletion during eukaryotes evolution. *BMC Genomics* **12**, S2 (2011) doi: 10.1186/1471-2164-12-S4-S2.

2. G. D'Souza, S. Waschina, S. Pande, K. Bohl, C. Kaleta, C. Kost, Less is more: selective advantages CAN explain the prevalent loss of biosynthetic genes in bacteria. *Evolution* **68**, 2559–2570 (2014) doi: 10.1111/evo.12468.

3. J. Trolle, R. M. McBee, A. Kaufman, S. Pinglay, H. Berger, S. German, L. Liu, M. J. Shen, X. Guo, J. A. Martin, M. E. Pacold, D. R. Jones, J. D. Boeke, H. H. Wang, Resurrecting essential amino acid biosynthesis in mammalian cells. *eLife* **11**, e72847 (2022) doi: 10.7554/eLife.72847.

4. M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Research* **51**, D587–D592 (2023) doi: 10.1093/nar/gkac963.

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

6. D. J. Richter, P. Fozouni, M. B. Eisen, N. King, Gene family innovation, conservation and loss on the animal stem lineage. *eLife* **7**, e34226 (2018) doi: 10.7554/eLife.34226.

7. M. Domazet-Lošo, T. Široki, K. Šimičević, T. Domazet-Lošo, Macroevolutionary dynamics of gene family gain and loss along multicellular eukaryotic lineages. *Nat Commun* **15**, 2663 (2024) doi: 10.1038/s41467-024-47017-w.

8. C. Guijarro-Clarke, P. W. H. Holland, J. Paps, Widespread patterns of gene loss in the evolution of the animal kingdom. *Nat Ecol Evol* **4**, 519–523 (2020) doi: 10.1038/s41559-020-1129-2.

9. H. Akashi, T. Gojobori, Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. *Proceedings of the National Academy of Sciences* **99**, 3695–3700 (2002) doi: 10.1073/pnas.062526999.

10. A. Wagner, Energy Constraints on the Evolution of Gene Expression. *Molecular Biology and Evolution* **22**, 1365–1374 (2005) doi: 10.1093/molbev/msi126.

11. H. Zhang, Y. Wang, J. Li, H. Chen, X. He, H. Zhang, H. Liang, J. Lu, Biosynthetic energy cost for amino acids decreases in cancer evolution. *Nat Commun* **9**, 4124 (2018) doi: 10.1038/s41467-018-06461-1.

12. O. P. Judson, The energy expansions of evolution. *Nat Ecol Evol* **1**, 1–9 (2017) doi: 10.1038/s41559-017-0138.

13. K. P. Kepp, Survival of the cheapest: how proteome cost minimization drives evolution. *Quarterly Reviews of Biophysics* **53**, e7 (2020) doi: 10.1017/S0033583520000037.

310 14. H. Seligmann, Cost-Minimization of Amino Acid Usage. *J Mol Evol* **56**, 151–161 (2003) doi:
311 10.1007/s00239-002-2388-z.

312 15. J. Swire, Selection on Synthesis Cost Affects Interprotein Amino Acid Usage in All Three
313 Domains of Life. *J Mol Evol* **64**, 558–571 (2007) doi: 10.1007/s00239-006-0206-8.

314 16. N. Ros-Rocher, A. Pérez-Posada, M. M. Leger, I. Ruiz-Trillo, The origin of animals: an
315 ancestral reconstruction of the unicellular-to-multicellular transition. *Open Biology* **11**,
316 200359 (2021) doi: 10.1098/rsob.200359.

317 17. I. Ruiz-Trillo, K. Kin, E. Casacuberta, The Origin of Metazoan Multicellularity: A Potential
318 Microbial Black Swan Event. *Annual Review of Microbiology* **77**, 499–516 (2023) doi:
319 10.1146/annurev-micro-032421-120023.

320 18. S. M. Stanley, An Ecological Theory for the Sudden Origin of Multicellular Life in the Late
321 Precambrian. *Proceedings of the National Academy of Sciences* **70**, 1486–1489 (1973) doi:
322 10.1073/pnas.70.5.1486.

323 19. S. Bengtson, Origins and Early Evolution of Predation. *The Paleontological Society Papers* **8**,
324 289–318 (2002) doi: 10.1017/S1089332600001133.

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

327 21. T. Domazet-Lošo, J. Brajković, D. Tautz, A phylostratigraphy approach to uncover the
328 genomic history of major adaptations in metazoan lineages. *Trends in Genetics* **23**, 533–539
329 (2007) doi: 10.1016/j.tig.2007.08.014.

330 22. D. Tautz, T. Domazet-Lošo, The evolutionary origin of orphan genes. *Nat Rev Genet* **12**, 692–
331 702 (2011) doi: 10.1038/nrg3053.

332 23. J. I. M. Fernandes, A. E. Murakami, Arginine metabolism in uricotelic species. *Acta*
333 *Scientiarum. Animal Sciences* **32**, 357–366 (2010) doi: 10.4025/actascianimsci.v32i4.10990.

334 24. H. C. Yoo, Y. C. Yu, Y. Sung, J. M. Han, Glutamine reliance in cell metabolism. *Exp Mol*
335 *Med* **52**, 1496–1516 (2020) doi: 10.1038/s12276-020-00504-8.

25. G. de Carvalho Fernandes, A. C. Turchetto-Zolet, L. M. Pereira Passaglia, Glutamine synthetase evolutionary history revisited: Tracing back beyond the Last Universal Common Ancestor. *Evolution* **76**, 605–622 (2022) doi: 10.1111/evo.14434.

26. Y. Li, M. G. Tassia, D. S. Waits, V. E. Bogantes, K. T. David, K. M. Halanych, Genomic adaptations to chemosymbiosis in the deep-sea seep-dwelling tubeworm Lamellibrachia luymesi. *BMC Biology* **17**, 91 (2019) doi: 10.1186/s12915-019-0713-x.

27. Y. Lan, J. Sun, C. Chen, Y. Sun, Y. Zhou, Y. Yang, W. Zhang, R. Li, K. Zhou, W. C. Wong, Y. H. Kwan, A. Cheng, S. Bougouffa, C. L. Van Dover, J.-W. Qiu, P.-Y. Qian, Hologenome analysis reveals dual symbiosis in the deep-sea hydrothermal vent snail Gigantopelta aegis. *Nat Commun* **12**, 1165 (2021) doi: 10.1038/s41467-021-21450-7.

28. A. L. de Oliveira, J. Mitchell, P. Girguis, M. Bright, Novel Insights on Obligate Symbiont Lifestyle and Adaptation to Chemosynthetic Environment as Revealed by the Giant Tubeworm Genome. *Molecular Biology and Evolution* **39**, msab347 (2022) doi: 10.1093/molbev/msab347.

29. CDC - NCHS - National Center for Health Statistics (2024). https://www.cdc.gov/nchs/index.htm.

30. G. Wu, F. W. Bazer, Z. Dai, D. Li, J. Wang, Z. Wu, Amino Acid Nutrition in Animals: Protein Synthesis and Beyond. *Annual Review of Animal Biosciences* **2**, 387–417 (2014) doi: 10.1146/annurev-animal-022513-114113.

31. G. C. Rock, K. W. King, Amino Acid Synthesis from Glucose-U-14C in *Argyrotaenia velutinana* (Lepidoptera: Tortricidae) Larvae. *The Journal of Nutrition* **95**, 369–373 (1968) doi: 10.1093/jn/95.3.369.

32. L. M. Fitzgerald, A. M. Szmant, Biosynthesis of “essential” amino acids by scleractinian corals. *Biochem J* **322**, 213–221 (1997).

33. J. T. Wang, A. E. Douglas, Essential amino acid synthesis and nitrogen recycling in an alga–invertebrate symbiosis. *Marine Biology* **135**, 219–222 (1999) doi: 10.1007/s002270050619.

362 34. L. S. Miyajima, G. A. Broderick, R. D. Reimer, Identification of the essential amino acids of
363 the freshwater shrimp, Macrobrachium ohione. *Proceedings of the annual meeting - World*
364 *Mariculture Society* **8**, 245–250 (1977) doi: 10.1111/j.1749-7345.1977.tb00122.x.

365 35. C. B. Cowey, J. R. M. Forster, The essential amino-acid requirements of the prawnPalaemon
366 serratus. The growth of prawns on diets containing proteins of different amino-acid
367 compositions. *Mar. Biol.* **10**, 77–81 (1971) doi: 10.1007/BF02026769.

368 36. K. L. Shewbart, W. L. Mies, P. D. Ludwig, Identification and quantitative analysis of the
369 amino acids present in protein of the brown shrimp Penaeus aztecus. *Marine Biology* **16**, 64–
370 67 (1972) doi: 10.1007/BF00347849.

371 37. G. W. Lasser, W. V. Allen, The essential amino acid requirements of the Dungeness crab,
372 Cancer magister. *Aquaculture* **7**, 235 (1976).

373 38. C. B. Cowey, J. Adron, A. Blair, Studies on the Nutrition of Marine Flatfish the Essential
374 Amino Acid Requirements of Plaice and Sole. *Journal of the Marine Biological Association*
375 *of the United Kingdom* **50**, 87–95 (1970) doi: 10.1017/S0025315400000618.

376 39. A. Zečić, I. Dhondt, B. P. Braeckman, The nutritional requirements of Caenorhabditis elegans.
377 *Genes & Nutrition* **14**, 15 (2019) doi: 10.1186/s12263-019-0637-7.

378 40. R. Ponnudurai, M. Kleiner, L. Sayavedra, J. M. Petersen, M. Moche, A. Otto, D. Becher, T.
379 Takeuchi, N. Satoh, N. Dubilier, T. Schweder, S. Markert, Metabolic and physiological
380 interdependencies in the Bathymodiolus azoricus symbiosis. *ISME J* **11**, 463–477 (2017) doi:
381 10.1038/ismej.2016.124.

382 41. H. Zheng, W. Zhang, L. Zhang, Z. Zhang, J. Li, G. Lu, Y. Zhu, Y. Wang, Y. Huang, J. Liu, H.
383 Kang, J. Chen, L. Wang, A. Chen, S. Yu, Z. Gao, L. Jin, W. Gu, Z. Wang, L. Zhao, B. Shi,
384 H. Wen, R. Lin, M. K. Jones, B. Brejova, T. Vinar, G. Zhao, D. P. McManus, Z. Chen, Y.
385 Zhou, S. Wang, The genome of the hydatid tapeworm Echinococcus granulosus. *Nat Genet*
386 **45**, 1168–1175 (2013) doi: 10.1038/ng.2757.

387 42. M. Moriyama, T. Fukatsu, Host's demand for essential amino acids is compensated by an
388 extracellular bacterial symbiont in a hemipteran insect model. *Front. Physiol.* **13** (2022) doi:
389 10.3389/fphys.2022.1028409.

390 43. T. Rosas-Pérez, M. Rosenblueth, R. Rincón-Rosales, J. Mora, E. Martínez-Romero, Genome
391 Sequence of "Candidatus Walczuchella monophlebidarum" the Flavobacterial
392 Endosymbiont of Llaveia axin axin (Hemiptera: Coccoidea: Monophlebidae). *Genome Biol*
393 *Evol* **6**, 714–726 (2014) doi: 10.1093/gbe/evu049.

394 44. G. J. Tsiropoulos, Amino-acid synthesis in adult Dacus oleae (Gmelin) (Diptera Tephritidae)
395 determined with [U-14C] glucose. *Arch Int Physiol Biochim* **92**, 313–316 (1984) doi:
396 10.3109/13813458409071172.

397 45. A. Weigert, C. Bleidorn, Current status of annelid phylogeny. *Org Divers Evol* **16**, 345–362
398 (2016) doi: 10.1007/s13127-016-0265-7.

399 46. C. E. Laumer, R. Fernández, S. Lemer, D. Combosch, K. M. Kocot, A. Riesgo, S. C. S.
400 Andrade, W. Sterrer, M. V. Sørensen, G. Giribet, Revisiting metazoan phylogeny with
401 genomic sampling of all phyla. *Proc Biol Sci* **286**, 20190831 (2019) doi:
402 10.1098/rspb.2019.0831.

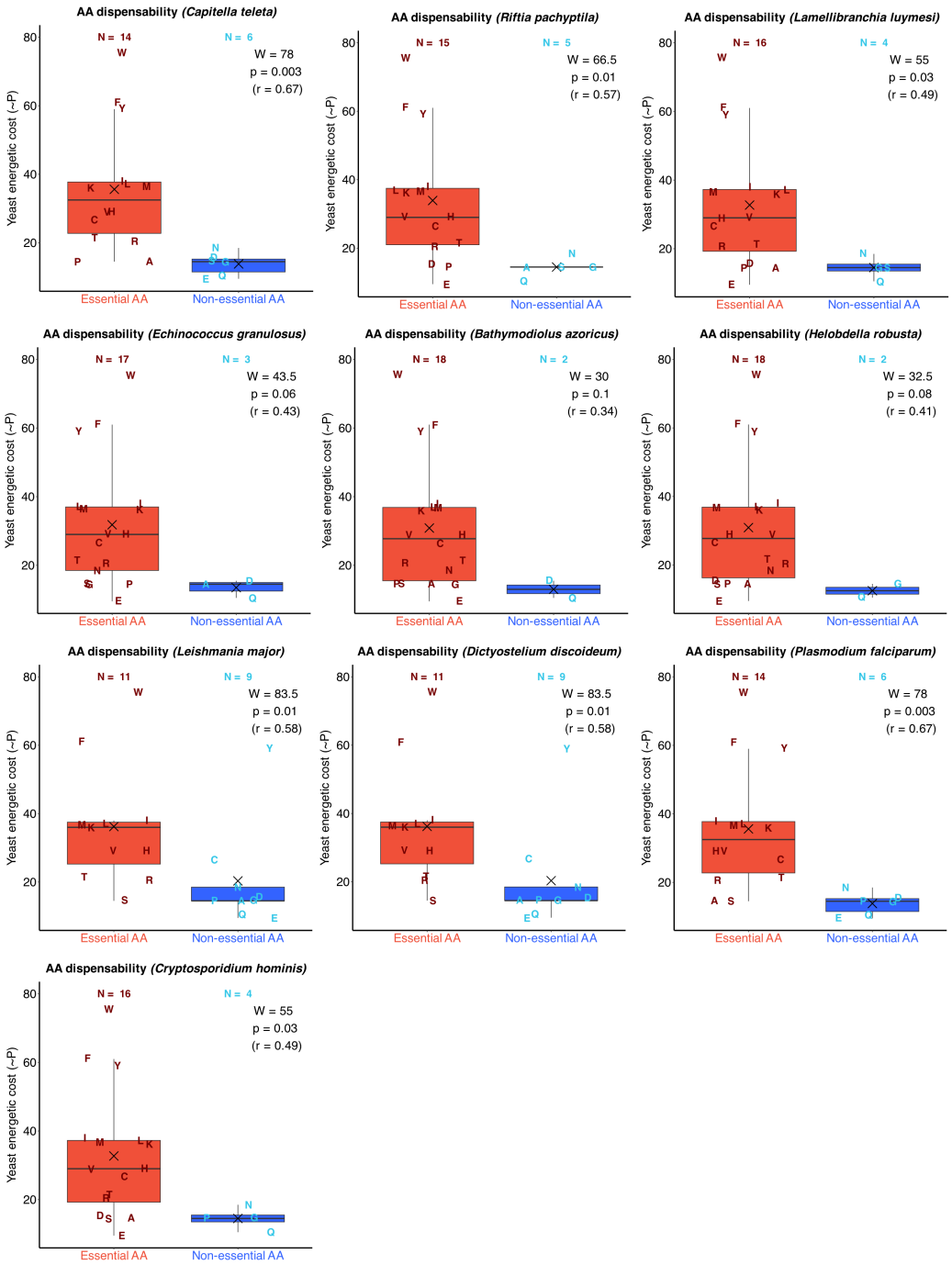
403 47. J. M. Wolfe, J. W. Breinholt, K. A. Crandall, A. R. Lemmon, E. M. Lemmon, L. E. Timm, M.
404 E. Siddall, H. D. Bracken-Grissom, A phylogenomic framework, evolutionary timeline and
405 genomic resources for comparative studies of decapod crustaceans. *Proc Biol Sci* **286**,
406 20190079 (2019) doi: 10.1098/rspb.2019.0079.

407 48. B. Misof, S. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware,
408 T. Flouri, R. G. Beutel, O. Niehuis, M. Petersen, F. Izquierdo-Carrasco, T. Wappler, J. Rust,
409 A. J. Aberer, U. Aspöck, H. Aspöck, D. Bartel, A. Blanke, S. Berger, A. Böhm, T. R.
410 Buckley, B. Calcott, J. Chen, F. Friedrich, M. Fukui, M. Fujita, C. Greve, P. Grobe, S. Gu,
411 Y. Huang, L. S. Jermiin, A. Y. Kawahara, L. Krogmann, M. Kubiak, R. Lanfear, H. Letsch,
412 Y. Li, Z. Li, J. Li, H. Lu, R. Machida, Y. Mashimo, P. Kapli, D. D. McKenna, G. Meng, Y.
413 Nakagaki, J. L. Navarrete-Heredia, M. Ott, Y. Ou, G. Pass, L. Podsiadlowski, H. Pohl, B. M.
414 von Reumont, K. Schütte, K. Sekiya, S. Shimizu, A. Slipinski, A. Stamatakis, W. Song, X.

415 Su, N. U. Szucsich, M. Tan, X. Tan, M. Tang, J. Tang, G. Timelthaler, S. Tomizuka, M.
416 Trautwein, X. Tong, T. Uchifune, M. G. Walzl, B. M. Wiegmann, J. Wilbrandt, B. Wipfler,
417 T. K. F. Wong, Q. Wu, G. Wu, Y. Xie, S. Yang, Q. Yang, D. K. Yeates, K. Yoshizawa, Q.
418 Zhang, R. Zhang, W. Zhang, Y. Zhang, J. Zhao, C. Zhou, L. Zhou, T. Ziesmann, S. Zou, Y.
419 Li, X. Xu, Y. Zhang, H. Yang, J. Wang, J. Wang, K. M. Kjer, X. Zhou, Phylogenomics
420 resolves the timing and pattern of insect evolution. *Science* **346**, 763–767 (2014) doi:
421 10.1126/science.1257570.

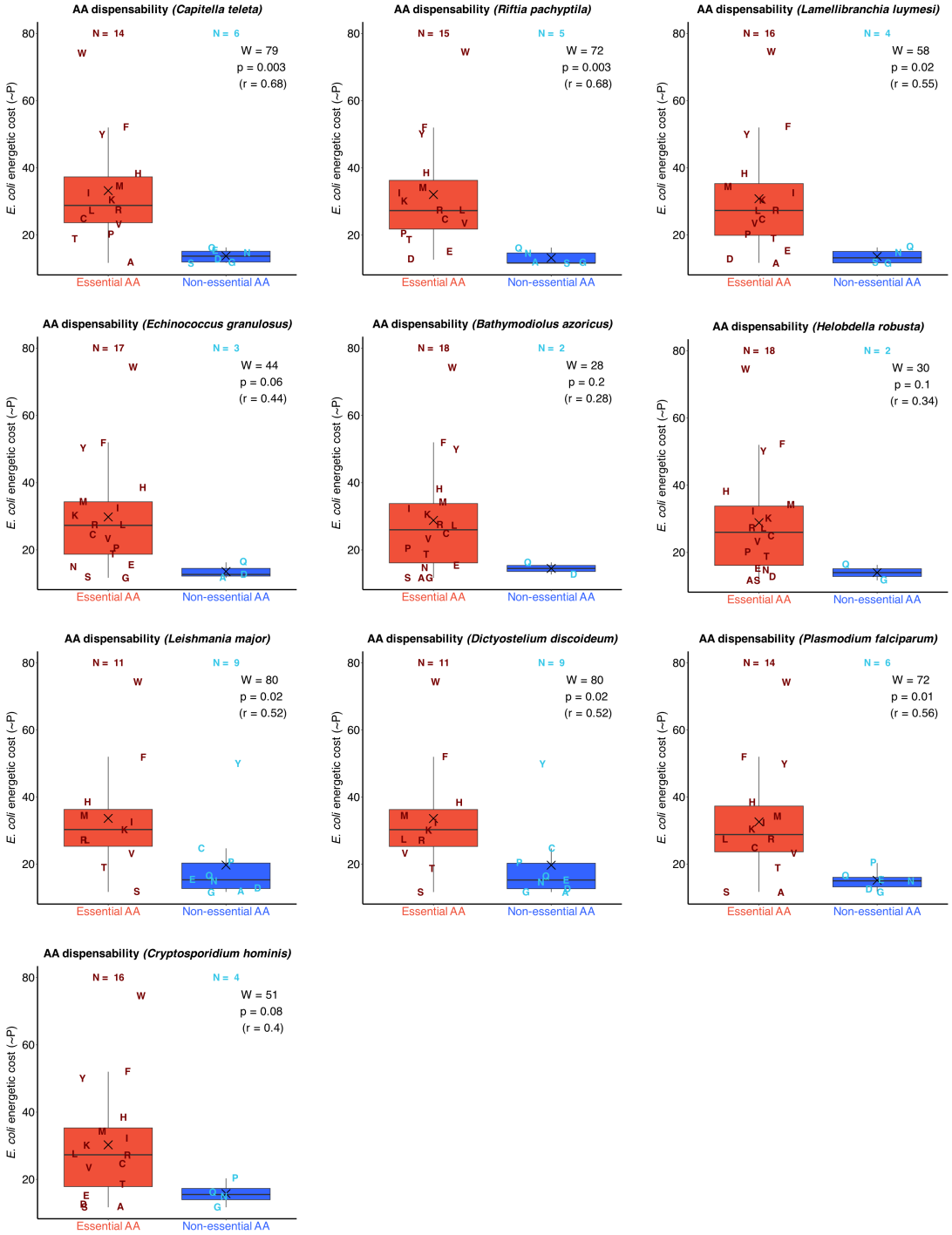
422 49. D. K. Yeates, B. M. Wiegmann, G. W. Courtney, R. Meier, C. Lambkin, T. Pape, Phylogeny
423 and systematics of Diptera: Two decades of progress and prospects*. *Zootaxa* **1668** (2007)
424 doi: 10.11646/zootaxa.1668.1.27.

425
426



Supplementary Fig. 1. The comparison of energy costs between essential and non-essential amino acids for selected organisms depicted in Fig. 3 (yeast energy estimates). We used AA energy costs from yeast (*10*) estimated as a number of high-energy phosphate bonds (~P) required for the synthesis of an AA from its metabolic precursors. The differences in energy costs between the two groups were shown by boxplots and the significance of these differences was tested by the

Mann-Whitney U test with continuity correction. We depicted the corresponding W-value, p-value, and effect size (r) in each panel. The X symbol represents the mean and individual AAs are shown by one-letter symbols.



Supplementary Fig. 2 The comparison of energy costs between essential and non-essential amino acids for selected organisms depicted in Fig. 3 (*E. coli* energy estimates). We used AA energy costs from *E. coli* (*9*) estimated as a number of high-energy phosphate bonds (~P) required

443 for the synthesis of an AA from its metabolic precursors. The differences in energy costs between
444 the two groups were shown by boxplots and the significance of these differences was tested by the
445 Mann-Whitney U test with continuity correction. We depicted the corresponding W-value, p-
446 value, and effect size (r) in each panel. The X symbol represents the mean and individual AAs are
447 shown by one-letter symbols.
448

449 **Supplementary Table 1.** Amino acid biosynthesis costs for *Escherichia coli*, *Saccharomyces*
450 *cerevisiae*, and *Homo sapiens* gathered from the literature (*9–11*). AA energy costs are estimated
451 as a number of high-energy phosphate bonds (~P) required for the synthesis of an AA from its
452 metabolic precursors. Generally accepted dietary essentiality of amino acids is denoted by
453 abbreviations NE (non-essential) and E (essential).
454

Amino acid name	**Three-letter symbol**	**One-letter symbol**	**Animals**	**Animals (extended)**	***E. coli* AA costs, ~P (*9*)**	**Yeast AA costs, ~P (*10*)**	**Human AA costs, ~P (*11*)**
Alanine	Ala	A	NE	NE	11.7	14.5	12.5
Arginine	Arg	R	NE	NE	27.3	20.5	14.5
Asparagine	Asn	N	NE	NE	14.7	18.5	16.5
Aspartate	Asp	D	NE	NE	12.7	15.5	13.5
Cysteine	Cys	C	NE	E	24.7	26.5	11
Glutamate	Glu	E	NE	NE	15.3	9.5	8.5
Glutamine	Gln	Q	NE	NE	16.3	10.5	9.5
Glycine	Gly	G	NE	NE	11.7	14.5	11
Histidine	His	H	E	E	38.3	29	0
Isoleucine	Ile	I	E	E	32.3	38	0
Leucine	Leu	L	E	E	27.3	37	0
Lysine	Lys	K	E	E	30.3	36	0
Methionine	Met	M	E	E	34.3	36.5	0
Phenylalanine	Phe	F	E	E	52	61	0
Proline	Pro	P	NE	NE	20.3	14.5	14.5
Serine	Ser	S	NE	NE	11.7	14.5	11
Threonine	Thr	T	E	E	18.7	21.5	0
Tryptophan	Trp	W	E	E	74.3	75.5	0
Tyrosine	Tyr	Y	NE	E	50	59	2.5
Valine	Val	V	E	E	23.3	29	0

455