

Predicting Ecosystem Metaphenome from Community Metagenome: A Grand Challenge for Environmental Biology

Neo Martinez¹

¹Indiana University Bloomington

March 07, 2024

Abstract

Elucidating how an organism's characteristics emerge from its DNA sequence has been one of the great triumphs of biology. This triumph has cumulated in sophisticated computational models that successfully predict how an organism's detailed phenotype emerges from its specific genotype. Inspired by that effort's vision and empowered by its methodologies, this Viewpoint describes a grand challenge to predict the biotic characteristics of an ecosystem, its metaphenome, from nucleic acid sequences of all the species in its community, its metagenome. Meeting this challenge would integrate rapidly advancing abilities of environmental nucleic acids (eDNA and eRNA) to identify organisms, their ecological interactions, and their evolutionary relationships with advances in mechanistic models of complex ecosystems. Addressing the challenge aims to help integrate ecology and evolutionary biology into a more unified and successfully predictive science that can better help describe and manage ecosystems and the services they provide to humanity.

1 **Predicting Ecosystem Metaphenome from Community Metagenome:**

2 **A Grand Challenge for Environmental Biology**

3
4 **Neo D. Martinez***

5 Center for Complex Networks and Systems

6 School of Informatics, Computing, and Engineering

7 919 E. 10th Street, Room 302

8 Bloomington, Indiana 47408

9 Indiana University, Bloomington, IN USA

10 Email: neo@PEaCELab.net Phone: +1 510 295 7624

11
12 Running title: **A Challenge to Predict Metaphenome from Metagenome**

13
14 10 Keywords: **synthesis, prediction, ecology, ecosystem, evolution, theory, computation, data science,**
15 **networks, environmental nucleic acids.**

16
17 Type of article: **Viewpoint**

18
19 Content: **146 words in abstract, 2000 words in main text, 1 figure, 1 table, 44 references**

20
21 NDM designed the study, performed the research, and wrote the manuscript.

22
23 *to whom correspondence should be sent

24 **ABSTRACT: Elucidating how an organism’s characteristics emerge from its DNA sequence has been**
25 **one of the great triumphs of biology. This triumph has cumulated in sophisticated computational**
26 **models that successfully predict how an organism’s detailed phenotype emerges from its specific**
27 **genotype. Inspired by that effort’s vision and empowered by its methodologies, this Viewpoint**
28 **describes a grand challenge to predict the biotic characteristics of an ecosystem, its metaphenome,**
29 **from nucleic acid sequences of all the species in its community, its metagenome. Meeting this**
30 **challenge would integrate rapidly advancing abilities of environmental nucleic acids (eDNA and eRNA)**
31 **to identify organisms, their ecological interactions, and their evolutionary relationships with advances**
32 **in mechanistic models of complex ecosystems. Addressing the challenge aims to help integrate**
33 **ecology and evolutionary biology into a more unified and successfully predictive science that can**
34 **better help describe and manage ecosystems and the services they provide to humanity.**

35 INTRODUCTION

36 “Grand Challenges” have emerged as one of the most compelling tools to motivate, engage, and
37 organize major research programs across the sciences and engineering (Omenn 2006; Kaldewey 2018).
38 This is especially true for biology. Grand challenges in evolution include assembling a great tree of life
39 summarizing the evolution of all life on our planet (Hinchliff *et al.* 2015). Grand challenges in ecology
40 include understanding the relationship between biodiversity and ecosystem function (National Research
41 Council 2001; van der Plas 2019). One of the largest and most successful grand challenges since
42 sequencing the human genome includes molecular and cell biologists’ project to predict individual
43 organisms’ characteristics from their DNA sequence otherwise known as predicting an organism’s
44 phenotype from its genotype (Fig. 1, National Research Council 2010). This challenge was largely met
45 for one species by a whole-cell simulation of a human pathogen (Karr *et al.* 2012) as proposed a decade
46 earlier (Tomita 2001).

47 Mechanistic models of phenotypes emerging from their genotypes continue to be developed
48 for different organisms using a variety of approaches. Several employ empirically parameterized models
49 of complex networks comprised of modules using mathematical and other algorithms to represent DNA
50 transcription, translation of RNA into proteins, and metabolic processes involving those proteins (Fig. 1,
51 Karr *et al.* 2012; Fang *et al.* 2020). Others model individual molecules and the cell's physical structure
52 created by those molecules to generate much more highly resolved representations of cellular structure
53 such as membranes, organelles, and proteins as well as their interactions (Feig & Sugita 2019). Both
54 approaches include many types of interactions involving thousands of biochemical species, interactions,
55 and parameters operating on multiple scales from molecular dynamics to cell division. The
56 sophistication and coordinated research and engineering applied to modeling whole cells vastly exceeds
57 that dedicated to similarly detailed modelling of whole ecosystems. Both inspired and informed by
58 whole-cell modeling as well as a long history of more modest behavioral, population, community and
59 ecosystem theory and modeling, the grand challenge to predict ecosystem metaphenome from
60 community metagenome seeks to motivate a multiscale mechanistic understanding of how the detailed
61 structure and function of ecosystems emerge from the interactions among organisms coexisting within
62 an environment (Jansson & Hofmockel 2018).

63 Addressing this challenge helps integrate behavioral, population, community, and ecosystem
64 ecology, several of the largest subdisciplines of ecology, with phylogenetics and population genetics,
65 two of the largest subdisciplines of evolutionary biology. Together, ecology and evolution comprise the
66 lion's share of environmental biology which has yet to achieve the synthetic and predictive successes
67 enjoyed by physics, chemistry, and molecular and cell biology. Such limitations may be surmounted by
68 predicting ecosystem metaphenomes from the metagenomes within the ecosystem's communities
69 starting first with relatively simple experimental ecosystems in the lab and eventually extending to
70 ecosystems more generally. Ecological communities are described by the diversity and identity of

71 species within a habitat and are identifiable by their genetic “barcodes.” The community metagenome
72 consists of the genomes of all species within the ecosystem’s communities. A key metaphenome is the
73 distribution of organisms among all trophic levels within a habitat and the dynamics of their populations
74 and biomass over time. Cell and molecular biologists’ success at meeting their genotype-to-phenotype
75 challenge points towards the tractability of a similar challenge at the ecosystem level and contributes
76 powerful social (e.g., structured collaborations), scientific (e.g., networks of networks) and technical
77 (e.g., sequencers and software) methodologies for addressing the challenge. What the grand challenge
78 is, how it may be met, and why it is worth pursuing are described further below.

79 THE CHALLENGE

80 To scale up our understanding of organismal behavior and interactions to the structure and
81 function of ecosystems, it is eminently clear that knowing which organisms and interactions occur in an
82 ecosystem is a useful place to start. Documenting these ecosystem traits has been a priority since well
83 before Darwin’s voyages. More recently, rapidly emerging technologies based on environmental nucleic
84 acids (eNA) including eDNA and eRNA have greatly increased our ability to describe these traits at a
85 vastly higher degree of resolution with extraordinarily lower effort and cost (Beng & Corlett 2020;
86 Deiner *et al.* 2021). These technologies illuminate the community metagenome constituted by the
87 genomes within ecosystems which provides a uniquely powerful description of the current state and
88 dynamic potential of the ecosystem (Jansson & Hofmockel 2018; Deiner *et al.* 2021). Sequences of
89 eDNA within metagenomes may identify organisms at virtually any level of taxonomic resolution from
90 the species level and above down through the population level (Luck *et al.* 2003) to the individual
91 organism itself. Such sequences also provide phylogenetic information about organisms’ evolutionary
92 history. Thus, metagenomes elucidate who organisms are, where they come from, their evolutionary
93 potential, and help leverage existing data on well-known organisms (e.g., metabolic capabilities) that
94 often accurately describe less familiar close relatives (Davies 2021). The location of metagenomes such

95 as within the tissue, gut, or pollen sacks of an organism, combined with the organism's known biology
96 such whether it's an animal, plant, parasite, pollinator, etc., can illuminate interactions such as
97 predation, parasitism, herbivory, symbiosis, and mutualism (Kennedy *et al.* 2020). Innovative methods,
98 such as separating differently sized DNA molecules, can distinguish endosymbionts and parasites from
99 prey (Krehenwinkel *et al.* 2017). The number of eNA copies can even elucidate species' biomass,
100 population size, life-stage structure, eco-evolutionary dynamics (Hao *et al.* 2020), organismal
101 interactions (e.g., feeding, pollination, and growth, Deiner *et al.* 2021; Yates *et al.* 2021) and rates of
102 ecosystem processes (Kennedy *et al.* 2020).

103 Combined with the vast literature on the physiology, natural history, and ecology of identified
104 organisms, the metagenome may unlock most of the empirical information essential to predicting
105 ecosystem behavior including virtually all co-existing organisms' identities and their most important and
106 cryptic interactions. Correlations with species composition and habitat may allow inferences such as the
107 size, location, and hydrologic regime of a lake or the soil type and fertility of a grassland. As such, the
108 metagenome may elucidate the biotic and abiotic structure of the ecosystem needed to be modeled and
109 ecosystem dynamics to be simulated. These dynamics form much of the ecosystem's metaphenome to
110 be predicted from, and test, models that formalize theories of how ecosystem metaphenomes emerge
111 from community metagenomes (Table 1).

113 HOW THE CHALLENGE MAY BE ADDRESSED

114 After deriving the observed and inferred biotic potential and abiotic context of an ecosystem
115 from its metagenome, the next step of the challenge is to describe the mechanisms by which its
116 characteristics persist and change through time. This has been achieved more generically by leveraging
117 the large degree of conformity in organismal behavior and their interactions (Evans *et al.* 2013; Martinez
118 2020). For example, organisms' metabolic, growth, and maximum consumption rates typically increase

119 as a three quarters power law of their body size (Brown *et al.* 2004). Additionally, feeding interactions
120 from herbivory through carnivory and parasitism to decomposition that determine carbon and
121 population dynamics are highly conserved both within evolutionary lineages (Edger *et al.* 2015; Davies
122 2021) and among ecosystems (Williams & Martinez 2008). Mechanisms responsible for this consistency
123 include the trophic hierarchy whereby organic energy necessarily flows from autotrophs through
124 heterotrophs at higher trophic levels and ultimately to decomposers. That constraint plus a species-
125 level mechanism that constrains generalists to consume organisms adjacent in this hierarchy enables the
126 trophic “niche model” to successfully predict food-web characteristics (e.g., means and variances of
127 species’ trophic levels, generality, vulnerability, connectedness, etc.) within a wide range of terrestrial
128 and aquatic ecosystems (Williams & Martinez 2000, 2008; Dunne *et al.* 2004) including Cambrian
129 ecosystems over a half billion years old (Dunne *et al.* 2008). Yet more constraints emerge from the
130 relatively consistent body size ratios between consumers and resources such as predators and their prey
131 (Brose *et al.* 2019) and parasites and their hosts (Hechinger 2013). Allometric trophic network (ATN)
132 models integrate these constraints with the identities and abundance of species as inputs into relatively
133 comprehensive models of the structure and dynamics of complex ecosystems (Table 1). ATN model’s
134 output richly describes ecosystems’ metaphenomes including how the abundance of multiple species
135 change over time when subjected to environmental change, biodiversity loss, and extraction of
136 ecosystem services (Table 1).

137 Augmenting this ATN approach (Martinez 2020) with the greatly increased resolution and
138 precision of network interactions derived from metagenomes may predict metaphenomes as modular,
139 differential-equation-based, whole-cell modeling efforts have predicted phenotypes (Karr *et al.* 2012).
140 One ATN module may determine network structure from empirical observations assisted by the
141 probabilistic niche model to suggest mistaken interactions (Williams *et al.* 2010). Another module
142 parameterizes the equations with metabolic and consumption rates again either specifically from

143 empirical observations or from those allometrically derived from the metabolic theory of ecology. A
144 third module runs the simulations with a 1-day time step in contrast to the 1-second time step of whole-
145 cell models. Other approaches may mimic physical models of whole cells that simulate the spatial
146 structure and dynamics of each molecule within a cell by doing the same for each organism within an
147 ecosystem (Katz *et al.* 2011). More plausible may be hybrid approaches where small organisms
148 including microbes are simulated using differential equations (Weitz *et al.* 2015; Jansson & Hofmøckel
149 2018) while large organisms are simulated with individual-based models (DeAngelis 2018). Rapid
150 advances in automated observation (Dell *et al.* 2014), large scale computing, and eNA-based analyses
151 will make current challenges much less prohibitive in the near future when modeling whole
152 environmental systems may eventually focus on socio-ecological sustainability (Martinez *et al.* 2012;
153 Davies *et al.* 2016).

154 WHY ADDRESS THE CHALLENGE? POSSIBILITY AND UTILITY

155 Systems biologists have already shown how efforts connecting genotype to phenotype can far
156 surpass those that ecologists employ in terms of empirical and mechanistic richness, computational
157 sophistication, and predictive success. Much of conceptual and technical heavy lifting has already been
158 achieved or at least initiated in the modeling of both whole cells and whole ecosystems. The challenge
159 to predict ecosystem metaphenome from community metagenome aims to focus environmental
160 biologists as the genotype-to-phenotype challenge has focused molecular and cell biologists.

161 At least since Darwin famously concluded his Origin of Species describing a “tangled bank” of
162 species “so different from each other, and dependent on each other in so complex a manner, [that] have
163 all been produced by laws acting around us,” scientists have sought to discover how the “integrity of
164 eco-systems” (Sir David Attenborough’s interpretation of “tangle bank,” Thorpe 2000) emerges from the
165 complex interactions among diverse organisms in nature. While many biologists focused on how

166 molecules such as DNA determine the structure and function of organisms, environmental biologists
167 focus more on how organisms determine the structure (e.g., diversity) and function (e.g., stability) of
168 ecosystems. Both sets of scientists focus on highly complex interdependent biological systems with
169 many different types of interactions operating on vastly different scales. However, cell and molecular
170 biologists have more ambitiously and successfully achieved whole system understanding than have
171 environmental biologists. The grand challenge to predict ecosystem metaphenomes from community
172 metagenomes aims to collaboratively leverage the talents and achievements of the former group for the
173 benefit of the latter. While retaining a focus on nucleic acids, the shift in scale from cell to ecosystem
174 shifts genome analyses from identifying genes, their interactions and their products, and their products'
175 interactions to a more straightforward identification of organisms and their interactions (Fig. 1).
176 Instead of interactions among genes and biochemical species (Fang *et al.* 2020), environmental
177 biologists focus on interactions within and among taxonomic species. Despite this shift, much of the
178 conceptual, mathematical, and computational methods are remarkably similar (Fig. 1). Both endeavors
179 also benefit from simultaneously achieving other key challenges while pursuing a grander one. They
180 include "integration of heterogenous databases, identification of the limits of our knowledge, predicting
181 complex, multi-network phenotypes, and suggesting future experiments that may lead to new
182 knowledge" (Carrera & Covert 2015). Perhaps most importantly, both pursue more comprehensive and
183 synthetically predictive understanding of the biological systems they study (Fig. 1). Interdisciplinary
184 collaborations that leverage these similarities may advance environment biology to where it more
185 effectively helps solve some of the most difficult environmental problems on the planet.

186 ACKNOWLEDGEMENTS

187 Anna Kuparinen and Hongseok Ko provided useful comments. Funding was provided by the USA
188 Department of Energy grant DE-SC0016247 and National Science Foundation grants 1241253, 1313830,
189 1642894, 1754207, and 1934817 to NDM.

190 REFERENCES

- 191 Allhoff, K.T., Weiel, E.M., Rogge, T. & Drossel, B. (2015). On the interplay of speciation and dispersal: An
192 evolutionary food web model in space. *J. Theor. Biol.*, 366, 46–56.
- 193 Bathe, O. & Farshidfar, F. (2014). From Genotype to Functional Phenotype: Unraveling the Metabolomic
194 Features of Colorectal Cancer. *Genes (Basel)*, 5, 536–560.
- 195 Beng, K.C. & Corlett, R.T. (2020). *Applications of environmental DNA (eDNA) in ecology and conservation:
196 opportunities, challenges and prospects*. Biodivers. Conserv. Springer Netherlands.
- 197 Bland, S., Valdovinos, F.S., Hutchings, J.A. & Kuparinen, A. (2019). The role of fish life histories in
198 allometrically scaled food-web dynamics. *Ecol. Evol.*, 1–10.
- 199 Boit, A., Martinez, N.D., Williams, R.J. & Gaedke, U. (2012). Mechanistic theory and modelling of
200 complex food-web dynamics in Lake Constance. *Ecol. Lett.*, 15, 594–602.
- 201 Brose, U., Archambault, P., Barnes, A.D., Bersier, L.F., Boy, T., Canning-Clode, J., *et al.* (2019). Predator
202 traits determine food-web architecture across ecosystems. *Nat. Ecol. Evol.*, 3, 919–927.
- 203 Brown, H., Gillooly, J.F., Allen, A.P., Savage, V.M. & West, G.B. (2004). Toward a metabolic theory of
204 ecology. *Ecology*, 85, 1771–1789.
- 205 Carrera, J. & Covert, M.W. (2015). Why Build Whole-Cell Models? *Trends Cell Biol.*, 25, 719–722.
- 206 Davies, N., Field, D., Gavaghan, D., Holbrook, S.J., Planes, S., Troyer, M., *et al.* (2016). Simulating social-
207 ecological systems: The Island Digital Ecosystem Avatars (IDEA) consortium. *Gigascience*, 5, 1–4.
- 208 Davies, T.J. (2021). Ecophylogenetics redux. *Ecol. Lett.*, 24, 1073–1088.
- 209 DeAngelis, D.L. (2018). *Individual-Based Models and Approaches in Ecology*. Chapman and Hall/CRC.
- 210 Deiner, K., Yamanaka, H. & Bernatchez, L. (2021). The future of biodiversity monitoring and conservation
211 utilizing environmental DNA. *Environ. DNA*, 3, 3–7.
- 212 Dell, A.I., Bender, J.A., Branson, K., Couzin, I.D., de Polavieja, G.G., Noldus, L.P.J.J., *et al.* (2014).
213 Automated image-based tracking and its application in ecology. *Trends Ecol. Evol.*, 29, 417–428.
- 214 Dunne, J.A., Williams, R.J. & Martinez, N.D. (2004). Network structure and robustness of marine food
215 webs. *Mar. Ecol. Prog. Ser.*, 273.
- 216 Dunne, J.A., Williams, R.J., Martinez, N.D., Wood, R.A. & Erwin, D.H. (2008). Compilation and network
217 analyses of Cambrian food webs. *PLoS Biol.*, 6.
- 218 Edger, P.P., Heidel-Fischer, H.M., Bekaert, M., Rota, J., Glöckner, G., Platts, A.E., *et al.* (2015). The
219 butterfly plant arms-race escalated by gene and genome duplications. *Proc. Natl. Acad. Sci.*, 112,
220 8362–8366.
- 221 Evans, M.R., Bithell, M., Cornell, S.J., Dall, S.R.X., Díaz, S., Emmott, S., *et al.* (2013). Predictive systems
222 ecology. *Proc. R. Soc. B Biol. Sci.*, 280, 20131452.
- 223 Fang, X., Lloyd, C.J. & Palsson, B.O. (2020). Reconstructing organisms in silico: genome-scale models and
224 their emerging applications. *Nat. Rev. Microbiol.*, 18, 23–26.
- 225 Feig, M. & Sugita, Y. (2019). Whole-cell models and simulations in molecular detail. *Annu. Rev. Cell Dev.*

- 226 *Biol.*, 35, 191–211.
- 227 Hao, M., Jin, Q., Meng, G., Yang, C., Yang, S., Shi, Z., *et al.* (2020). Regional assemblages shaped by
228 historical and contemporary factors : Evidence from a species-rich insect group, 2492–2510.
- 229 Hechinger, R.F. (2013). A Metabolic and Body-Size Scaling Framework for Parasite Within-Host
230 Abundance, Biomass, and Energy Flux. *Am. Nat.*, 182, 234–248.
- 231 Hinchliff, C.E., Smith, S.A., Allman, J.F., Burleigh, J.G., Chaudhary, R., Coghill, L.M., *et al.* (2015). Synthesis
232 of phylogeny and taxonomy into a comprehensive tree of life. *Proc. Natl. Acad. Sci. U. S. A.*, 112,
233 12764–12769.
- 234 Jansson, J.K. & Hofmockel, K.S. (2018). The soil microbiome — from metagenomics to metaphenomics.
235 *Curr. Opin. Microbiol.*, 43, 162–168.
- 236 Kaldewey, D. (2018). The Grand Challenges Discourse: Transforming Identity Work in Science and
237 Science Policy. *Minerva*, 56, 161–182.
- 238 Karr, J.R., Sanghvi, J.C., MacKlin, D.N., Gutschow, M. V., Jacobs, J.M., Bolival, B., *et al.* (2012). A whole-
239 cell computational model predicts phenotype from genotype. *Cell*, 150, 389–401.
- 240 Katz, Y., Tunstrøm, K., Ioannou, C.C., Huepe, C. & Couzin, I.D. (2011). Inferring the structure and
241 dynamics of interactions in schooling fish. *Proc. Natl. Acad. Sci.*, 108, 18720–18725.
- 242 Kennedy, S.R., Prost, S. & Krehenwinkel, H. (2020). High-throughput sequencing for community analysis :
243 the promise of DNA barcoding to uncover diversity , relatedness , abundances and interactions in
244 spider communities, 185–201.
- 245 Krehenwinkel, H., Kennedy, S., Pekár, S. & Gillespie, R.G. (2017). A cost-efficient and simple protocol to
246 enrich prey DNA from extractions of predatory arthropods for large-scale gut content analysis by
247 Illumina sequencing. *Methods Ecol. Evol.*, 8, 126–134.
- 248 Kuparinen, A., Boit, A., Valdovinos, F.S., Lassaux, H. & Martinez, N.D. (2016). Fishing-induced life-history
249 changes degrade and destabilize harvested ecosystems. *Sci. Rep.*, 6, 22245.
- 250 Luck, G.W., Daily, G.C. & Ehrlich, P.R. (2003). Population diversity and ecosystem services. *Trends Ecol.*
251 *Evol.*, 18, 331–336.
- 252 Martinez, N.D. (2020). Allometric trophic networks from individuals to socio-ecosystems: Consumer-
253 resource theory and the ecological elephant in the room. *Front. Ecol. Evol.*, 8, 92.
- 254 Martinez, N.D., Tonnin, P., Bauer, B., Rael, R.C., Singh, R., Yoon, S., *et al.* (2012). Sustaining Economic
255 Exploitation of Complex Ecosystems in Computational Models of Coupled Human-Natural
256 Networks. In: *Proceedings of the National Conference on Artificial Intelligence*. pp. 326–334.
- 257 National Research Council. (2001). *Grand Challenges in Environmental Sciences. Gd. Challenges Environ.*
258 *Sci.* National Academies Press, Washington, D.C.
- 259 National Research Council. (2010). *Research at the Intersection of the Physical and Life Sciences. Res.*
260 *Intersect. Phys. Life Sci.* National Academies Press, Washington, D.C.
- 261 Omenn, G.S. (2006). Grand challenges and great opportunities in science, technology, and public policy.
262 *Science (80-)*, 314, 1696–1704.
- 263 van der Plas, F. (2019). Biodiversity and ecosystem functioning in naturally assembled communities. *Biol.*

264 *Rev.*, 94, brv.12499.

265 Thorpe, V. (2000). Darwin's Theory Was Inspired by a Hillock. *Guard*.

266 Tomita, M. (2001). Whole-cell simulation: A grand challenge of the 21st century. *Trends Biotechnol.*, 19,
267 205–210.

268 Weitz, J.S., Stock, C.A., Wilhelm, S.W., Bourouiba, L., Coleman, M.L., Buchan, A., *et al.* (2015). A
269 multitrophic model to quantify the effects of marine viruses on microbial food webs and
270 ecosystem processes. *ISME J.*, 9, 1352–1364.

271 Williams, R.J., Anandanadesan, A. & Purves, D. (2010). The probabilistic niche model reveals the niche
272 structure and role of body size in a complex food web. *PLoS One*, 5.

273 Williams, R.J. & Martinez, N.D. (2000). Simple rules yield complex food webs. *Nature*, 404, 180–183.

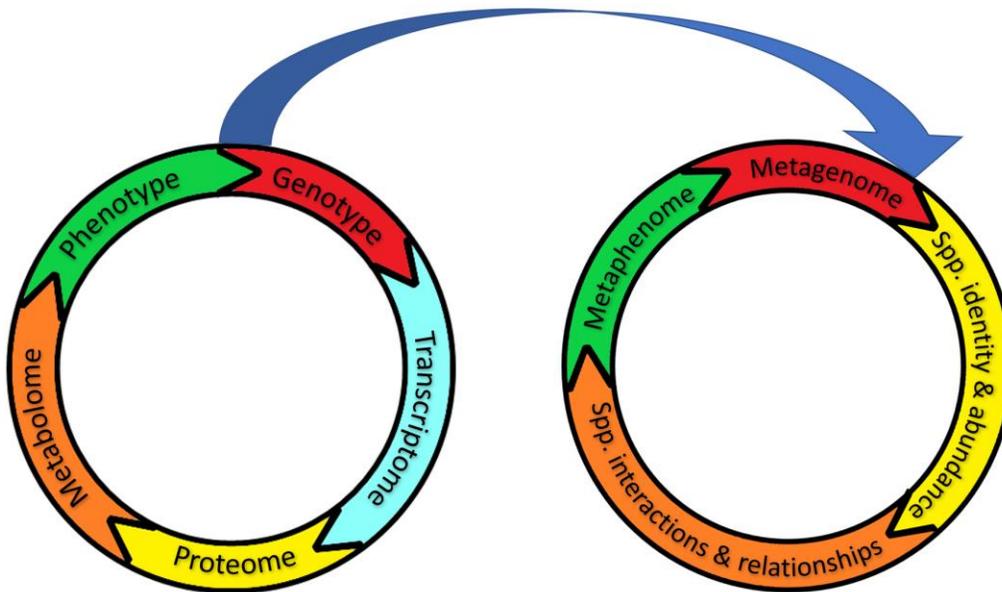
274 Williams, R.J. & Martinez, N.D. (2008). Success and its limits among structural models of complex food
275 webs. *J. Anim. Ecol.*, 77, 512–519.

276 Yates, M.C., Derry, A.M. & Cristescu, M.E. (2021). Environmental RNA: A Revolution in Ecological
277 Resolution? *Trends Ecol. Evol.*

278

279

280



281

282 Figure 1. Simplified depiction (after Bathe & Farshidfar 2014) of how an organism's phenotype emerges
283 from its genotype (left) and its role in an analogous depiction (right) of how an ecosystem's biotic
284 characteristics, its metaphenome, emerge from the genomes of its constituent species, its metagenome.
285 The ability to identify species and quantify their abundance stems from (blue arrow) the ability of
286 genotypes to identify the phenotype (e.g., body type and size, taxonomy, etc.) of organisms. Similar
287 colors between left and right depictions indicate analogous steps of emergence. While proteomes are
288 biologically derived from genotypes via the transcriptome, the identity and abundance of species may
289 be bioinformatically derived from an ecosystem's metagenome. This enables the identification of
290 species' ecological interactions that function similarly to the metabolome in generating essential
291 characteristics of organisms and ecosystems. For example, a key similarity is that both phenotypes and
292 metaphenomes (green) emerge from networks of interactions (orange) among functionally similar
293 molecules (left) and organisms (right) grouped into functionally distinct species (yellow) that interact at
294 concentration- (left) and density- (right) dependent rates modeled by Michaelis-Menten functions (left)
295 also called "functional responses" by ecologists (right, Fang *et al.* 2020; Martinez 2020).

296

297

Step	Challenge	Focus	Allometric trophic network theory
1	Metagenotyping	Habitats, organisms, guts, tissues	30 yr observing L. Constance biota
2	Identify and quantify	Taxa, diversity, phylogenies, populations, growth rates, interactions	Field and lab observations & experiments
3	Assemble and parameterize consumer-resource networks	Feeding, infection, pollination, seed dispersal, nutrient transport, habitat provision, decomposition, preferences, adaptive responses (fear, cooperation, interference, defense, etc.)	25 Nodes: detritus, decomposers, bacteria, phytoplankton, zooplankton, fishes. Empirically observed & allometrically estimated parameters.
4	Code and conduct simulations	Uncertainty, variability, stability, species abundance & distribution. Effects of species loss & invasion, warming & eutrophication, extraction	Seasonal dynamics of a complex food web with 1-day time steps initialized with spring conditions (Boit <i>et al.</i> 2012)
5	Test against observations	Self-consistency, controlled & uncontrolled experiments	Explains 83% & 88% seasonal biomass & production variability of 20 auto- & heterotrophic spp.+ detritus
6	Repeat 1-5 as necessary	Test novel predictions, explore new hypotheses	Effects of age-structure, fishing, environmental noise & mutualism

298

299 Table 1. Grand challenge steps for generally predicting ecosystem metaphenomes from community
300 metagenomes compared to specifically predicting the metaphenome of Lake Constance (Boit *et al.*
301 2012) with Allometric Trophic Network (ATN) theory (Martinez 2020). ATN predictions of ecosystem
302 metaphenomes involve simulations of empirical and niche-model networks of consumer-resource
303 interactions that have been broadly corroborated empirically (Martinez 2020): Vucic-Pestic *et al.* (2011)
304 predicted the recently observed decreases in trophic transfer efficiency (Barneche *et al.* 2021) caused by
305 global warming. Others (Jonsson *et al.* 2018; Curtsdotter *et al.* 2019) predicted observed effects of
306 species loss and invasion (Romanuk *et al.* 2009, 2017; Smith-Ramesh *et al.* 2017). ATN theory (Martinez
307 2020) has been extended to include detritus (Boit *et al.* 2012), evolutionary and spatial ecology (Allhoff
308 *et al.* 2015), ontogenetic niche shifts (Kuparinen *et al.* 2016; Bland *et al.* 2019), environmental variability
309 (Kuparinen *et al.* 2018), and mutualism within pollination networks (Hale *et al.* 2020).