

1 DNA barcoding and community assembly - A simple solution to a complex problem

2
3 Susan R. Kennedy^{1,2*}

4 Henrik Krehenwinkel²

5
6 ¹Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate
7 University, Onna, Okinawa, Japan

8 ²Department of Biogeography, Trier University, Trier, Germany

9
10 *susanrkennedy@gmail.com

11
12 **Identifying the current and past processes driving community assembly is critical in**
13 **the effort to understand the Earth's biodiversity and its response to future environmental**
14 **change. But while studies on community assembly often emphasize the role of contemporary**
15 **ecological drivers, it has been particularly challenging to account for the effects of past**
16 **processes in shaping present-day communities. In this issue of *Molecular Ecology*, Hao *et al.***
17 **(2020) provide a holistic analysis of factors driving the assembly of diverse communities of**
18 **Lepidoptera in two mountain ranges in Northeastern China. The authors use an impressively**
19 **large dataset and exceptionally comprehensive analyses to test how processes of range**
20 **expansion and gene flow, speciation and extinction, dispersal limitation, environmental**
21 **filtering and competition have led to present-day diversity patterns. A key novelty of this**
22 **work is the exhaustive use of DNA barcodes, relatively simple yet powerful molecular**
23 **markers, to tackle complex biological questions. The authors elegantly show the utility of**
24 **DNA barcoding data for research beyond simple taxonomic assignment. Their approach is**
25 **remarkable as it manages to integrate population genetics, phylogenetic history, species**
26 **diversity and ecology into a well-rounded picture of community assembly. With this work,**
27 **Hao *et al.* demonstrate the great promise of DNA barcoding for exhaustive community**
28 **analysis of even highly diverse and complex systems, raising the bar for future research.**

29
30 Community assembly is defined as “the study of the processes that shape the identity and
31 abundance of species within ecological communities” (Kraft & Ackerly 2014). This simple
32 statement summarizes a burgeoning, yet complex, research field (Fukami 2015). Biological
33 communities are the result of multiple historical and contemporary processes operating on both
34 the ecological and evolutionary scale, which affect biodiversity at multiple levels, from individuals
35 and populations to species. In order to explain present-day biodiversity patterns, it is important to
36 take all of these factors into account. However, this requires vast amounts of data as well as the
37 ability to reliably infer past events. Consequently, most studies on community assembly are limited
38 in temporal scale and taxonomic breadth.

39
40 In this issue, Hao *et al.* (2020) provide an exemplary solution to this problem. The authors
41 focus on the moths (Order Lepidoptera) inhabiting the Yanshan and Taihang mountains of
42 Northeastern China (Fig. 1), a region characterized by high species richness and abundance.
43 Lepidopterans are taxonomically and ecologically diverse, and as such comprise a large and
44 complex biological community in the study area. Hao *et al.* (2020) use an extensive data set of
45 nearly 8000 specimens representing over 800 species collected from the two mountain ranges to
46 identify which historic and contemporary processes – on both the ecological and evolutionary

47 scales – drive community assembly. The authors use a large DNA barcode dataset to characterize
48 diversity assembly from the population to the community scale. In the current era of high-
49 throughput sequencing technologies, ecological questions are often approached by in-depth studies
50 of large, genome-wide sequence datasets (e.g. Jones *et al.* 2012). It is both remarkable and
51 refreshing to see how much information can be obtained from a simple, single-locus marker like a
52 DNA barcode, as Hao *et al.* (2020) demonstrate by extending the utility of barcode information
53 far beyond its typical applications for species identification and delimitation.

54
55 DNA barcoding, when initially proposed, was praised by ecologists, who saw its huge
56 potential to quickly characterize the species composition of a community (Moritz & Cicero 2004).
57 Yet, DNA barcoding is often not used to its full potential, with many studies merely providing lists
58 of local species assemblages. DNA barcodes are variable enough to distinguish even recently
59 diverged species and usually also contain ample intraspecific variation. They are thus well suited
60 for phylogenetic and even population genetic analyses. Here, Hao *et al.* (2020) use DNA barcoding
61 of their moth communities to build species-level phylogenetic trees, reconstruct haplotype
62 networks, and extract population genetic signatures for a considerable set of taxa. While the short
63 *COI* barcodes used here can resolve recent phylogenetic divergences, they are less suited to support
64 deeper branches. Hao *et al.* (2020) handle this problem by building a coarser-scale phylogeny at
65 the family level based on nuclear reference sequences, and grafting species-level phylogenies onto
66 this backbone. This simple yet elegant approach could easily be implemented in other systems.

67
68 Using their sequence data alongside locality and morphological measurements of the
69 specimens, the authors perform a comprehensive suite of analyses to illuminate the biogeographic
70 and demographic history of these communities and to assess the relative importance of different
71 ecological factors in determining current biodiversity patterns. They reconstruct ancestral ranges
72 and infer past processes of speciation, extinction and dispersal between sites. Subsequently, they
73 test the relative importance of dispersal limitation, environmental filtering, competition, and
74 combinations thereof. Finally, they measure morphological traits related to dispersal ability, e.g.
75 wing aspect ratio, and test for relationships between these traits and the moths' distribution
76 patterns.

77
78 Based on all of these analyses, Hao *et al.* (2020) uncover an interesting story about the
79 historical processes underlying the assembly of these two moth communities, suggesting the
80 Yanshan mountains as a likely source of diversification for many groups, with prevalent recent
81 expansions into the Taihang mountains. At the same time, they find evidence for strong gene flow
82 between the two sites and touch on possible effects of landscape topography and climate to explain
83 diversity patterns.

84
85 Hao *et al.* (2020) present an exceptionally thorough analysis of a large and comprehensive
86 dataset, bridging the gap between past and current processes on the scales of populations, species
87 and whole communities. To accomplish this for such a diverse group as Lepidoptera (Fig. 2), which
88 contains over 150,000 described species, is impressive. It is challenging enough to characterize
89 present-day assembly patterns of highly diverse communities, such as arthropods, due to
90 limitations in available taxonomic expertise and the substantial time investment required to
91 identify all specimens. This study thus represents a significant step forward in approaches to
92 characterizing community assembly. Moreover, this work demonstrates the considerable potential

93 of simple single-locus sequence data to reveal important insights into the processes underlying
94 community assembly, and sets a high standard for future studies. This latter point is particularly
95 relevant considering the ease with which large community-wide DNA barcode datasets can now
96 be generated using high throughput sequencing-based DNA metabarcoding (Ji *et al.* 2013). This
97 opens up whole new perspectives for the analysis of arthropod communities.

98

99 **References**

- 100 Fukami T (2015) Historical contingency in community assembly: integrating niches, species
101 pools, and priority effects. *Annual Review of Ecology, Evolution, and Systematics*, **46**, 1-
102 23.
- 103 Hao MD, Jin Q, Meng G, *et al.* (2020) Regional assemblages shaped by historical and
104 contemporary factors: evidence from a species-rich insect group. *Molecular Ecology*, **29**,
105 XXX-XXX.
- 106 Ji Y, Ashton L, Pedley SM, *et al.* (2013) Reliable, verifiable and efficient monitoring of
107 biodiversity via metabarcoding. *Ecology Letters*, **16**, 1245-1257.
- 108 Jones FC, Grabherr MG, Chan YF, *et al.* (2012) The genomic basis of adaptive evolution in
109 threespine sticklebacks. *Nature*, **484**, 55-61.
- 110 Kraft NJ, Ackerly DD (2014) Assembly of plant communities. *Ecology and the Environment*, **8**,
111 67-88.
- 112 Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**, 1529-1531.

113

114 **Figure captions**

115

116 Figure 1: Sampling sites in Taihang and Yanshan mountains. WLS - Wulingshan, LBGGM -
117 Labagoumen, SS - Songshan, JF - Jiufeng, JG - Jiangou, LJZ - Longjiazhuang, DLS -
118 Donglingshan, BHS - Baihuashan, LF - Langfang, BD - Baoding. MFS includes JF, JG and LJZ
119 sites because of their close distance.

120

121 Figure 2: Examples of moth species from different families. A. Sphingidae; B. Saturniidae; C.
122 Saturniidae; D. Sphingidae. Fig. 1: Field sites sampled by Hao *et al.* (2020) in the a) Yanshan and
123 b) Taihang mountain ranges of Northeastern China

124

125 Fig. 2: ~~Xxxx xxx~~, one of the lepidopteran species included in the study of Hao *et al.* (2020)



