



**HAL**  
open science

# DNA metabarcoding illuminates the black box of soil animal biodiversity

Thibaud Decaëns

► **To cite this version:**

Thibaud Decaëns. DNA metabarcoding illuminates the black box of soil animal biodiversity. *Molecular Ecology*, 2021, 30 (1), pp.33-36. <10.1111/mec.15761>. <hal-03648916>

**HAL Id: hal-03648916**

**<https://hal.science/hal-03648916v1>**

Submitted on 22 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

1 **DNA metabarcoding illuminates the black box of soil animal**  
2 **biodiversity**

3 Thibaud Decaëns

4 CEFE, Univ Montpellier, CNRS, EPHE, IRD, Univ Paul Valéry Montpellier 3, Montpellier,  
5 France. Email : [Thibaud.Decaens@cefe.cnrs.fr](mailto:Thibaud.Decaens@cefe.cnrs.fr)

6 **Identifying the determinants of the distribution of soil organism**  
7 **communities at different spatial scales is essential to our understanding of the**  
8 **functioning of terrestrial ecosystems. Indeed, spatial patterns of soil biota, and the**  
9 **factors that determine them, influence spatial patterns of many ecological**  
10 **processes such as the decomposition of organic detritus, the release of nutrients**  
11 **and, ultimately, the spatial structure of plant communities (Ettema & Wardle,**  
12 **2002). However, our ability to describe and understand the spatial distribution of**  
13 **soil communities is severely hampered by taxonomic and methodological**  
14 **limitations, significantly reducing our ability to infer ecological and evolutionary**  
15 **mechanisms from observed patterns. In a From the Cover article in this issue of**  
16 **Molecular Ecology, Arribas, Andujar, Salces-Castellano, Emerson, & Vogler (2020)**  
17 **used a high-throughput-sequencing approach, i.e. haplotype-level-COI**  
18 **metabarcoding, to bypass these roadblocks and analyse the spatial arrangements**  
19 **of soil arthropods communities at different phylogenetic levels, from haplotypes**  
20 **to nested operational units. They found that local assemblage composition was**  
21 **primarily explained by habitat, and that community similarity decreased with**  
22 **distance in a similar way in all habitats and geographical regions, and at all**  
23 **hierarchical levels. This highlights that both habitat filtering and dispersal**  
24 **limitation are major drivers of local-scale assembly process in soil arthropod**

25 **communities. Beyond these significant results, their study also highlights how**  
26 **DNA-based approaches, by producing combined description of genetic and**  
27 **taxonomic diversity, opens up new perspectives to soil spatial ecology and**  
28 **macroecology.**

29 Soil biological communities are among the most diversified on Earth, comprising  
30 as much as a quarter of currently known global biodiversity (Decaëns, 2010). They are  
31 involved in the regulation of a large number of ecological processes and in the provision  
32 of critical ecosystem services for human well-being. Paradoxically, belowground biota  
33 suffers from a conspicuous deficit of research when compared to aboveground  
34 organisms, and soils are still considered to be the third remaining biotic frontier after  
35 oceanic abysses and tropical forest canopies. To date, only a very small fraction of soil  
36 species diversity has been formally documented, and only a few studies have considered  
37 spatial ecology of soil communities. The paucity of our taxonomic and ecological  
38 knowledge for soil biota has historically limited its integration into the development of  
39 theoretical ecology, which has remained below what it should be given the quantitative  
40 and functional importance of these organisms (Thakur et al., 2020).

41 DNA-based approaches, such as DNA barcoding and metabarcoding, now offer the  
42 opportunity to remove these impediments and improve our capacity to describe soil  
43 biota patterns (Rougerie et al., 2009). In soil invertebrates, DNA barcodes are used in  
44 support of morphological diagnoses to increase our ability to detect and describe new  
45 species. They allow automated taxonomic assignments, including for cryptic species or  
46 for individuals not presenting the necessary diagnostic characters for morphological  
47 identifications (e.g. immature individuals). They can also be used to delineate  
48 operational taxonomic units (OTUs), which represent species surrogates to describe

49 diversity patterns for taxonomically unresolved groups, or in poorly studied regions.  
50 Finally, high-throughput sequencing approaches such as DNA and eDNA metabarcoding  
51 (*sensu* Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012) are now revolutionizing soil  
52 fauna community ecology. eDNA metabarcoding has been applied to soil environmental  
53 samples to describe community patterns for single taxonomic groups (e.g. earthworms  
54 in Pansu et al., 2015) or for whole communities (e.g. Zinger et al., 2019). However, these  
55 studies generally use specific markers, rather than universally recognized barcodes,  
56 thus limiting the completion of taxonomic assignments and the possibilities of  
57 comparisons between studies. Conversely, DNA metabarcoding approaches analyse  
58 invertebrate bulk samples (Yu et al., 2012), and typically use these standard barcodes,  
59 allowing them to rely on the growing taxonomic information contained in global  
60 databases (Andujar, Arribas, Yu, Vogler, & Emerson, 2018; Porter & Hajibabaei, 2018).

61         The study by Arribas et al. (2020) is a perfect example of the added value that DNA  
62 metabarcoding can represent for soil invertebrate community ecology. The use of whole  
63 community metabarcoding (WCM) first allowed the authors to bypass the difficulties of  
64 taxonomic assignments in three highly diverse groups of soil arthropods (i.e. springtails,  
65 mites and beetles). In their study, they were able to consider tens of thousands of  
66 specimens belonging to >1000 species, thus covering a broader taxonomic range than  
67 what would have been possible through a classical approach. A key point of their study  
68 is that it exemplifies the importance of using COI in metabarcoding approaches, as this  
69 locus provides intraspecific (genetic) variation that can be used for subsequent analyses  
70 (Andujar et al., 2018). This enabled the authors to analyse the variation of beta-diversity  
71 with distance among communities at different nested phylogenetic levels, from unique  
72 haplotypes to lineages delimited with decreasing identity thresholds (here from 99% to  
73 92 %). Their results nicely demonstrate how niche and neutral processes combine to

74 explain the assemblage of soil arthropod communities. While community composition  
75 was primarily explained by habitat (i.e. pasture versus forest), they also found a  
76 distance-decay in community similarity, which was parallel among hierarchical levels  
77 and consistent among habitats and regions. This last pattern clearly underpins limited  
78 dispersal as an important driver of local community assembly. Finally, they described a  
79 high rate of microendemism, reinforcing the hypothesis of short dispersal distances  
80 driving community composition, and suggesting that species richness of soil arthropods  
81 at broader scales could be higher than expected.

82       Beyond these significant results, Arribas et al. (2020) lay the methodological  
83 foundations of a general analytical workflow to streamline soil invertebrate community  
84 studies (Fig. 1). As exemplified in their study, WCM using COI gene first generate sets of  
85 sequences that can be used as such or after aggregating them into OTUs. To go further,  
86 taxonomic assignments of these OTUs, when made possible by the choice of standard  
87 barcodes and the availability of existing libraries, can enable to link species names to  
88 ecological information, e.g. through queries in functional trait databases. Therefore, the  
89 proposed workflow has the potential to produce dataset triplets combining  
90 phylogenetic, taxonomic and functional descriptors of communities. Hereafter, I propose  
91 some examples of how such datasets could be analysed, in complementary way to the  
92 study of Arribas et al. (2020), to provide insights into our understanding of the ecology  
93 and evolutionary biology of soil invertebrates.

94       Even if DNA barcodes only capture a small fraction of phylogenetic signal, barcode  
95 trees are relevant to predict metrics of phylogenetic community structure obtained from  
96 multi-gene phylogenies (Boyle & Adamowicz, 2015). When available, backbone  
97 phylogenies can further be used to improve the phylogenetic context of the massive

98 community level datasets generated by WCM. Thus, WCM datasets can be analysed  
99 through combining in a complementary way trait-based and phylogenetic approaches,  
100 which has been emphasised as a powerful strategy to dissect multi-scales biodiversity  
101 patterns (Mouquet et al., 2012). Indeed, non-random patterns in functional and  
102 phylogenetic community structure can be tracked to disentangle the relative importance  
103 of niche and neutral processes, and how they interact with species evolutionary history,  
104 in the processes of community assembly. DNA barcodes obtained from WCM can also be  
105 used in phylogeographic and population genetic analyses, which, when performed  
106 comparatively for large numbers of species across ecological guilds, can reveal general  
107 patterns of differentiation and allow testing hypotheses regarding the origin and  
108 maintenance of community diversity (Craft et al., 2010; Turon, Antich, Palacin, Praebel,  
109 & Wangensteen, 2020).

110         Analysing the turnover component of beta diversity among WCM datasets  
111 obtained from large-scale sampling designs, is expected to reveal macroecological  
112 patterns in the taxonomic, functional and phylogenetic components of soil biodiversity,  
113 and to allow reassessing global species numbers as suggested by Arribas et al. (2020).  
114 Finally, analysing WCM datasets through diversity partitioning approaches, can reveal  
115 the relative contributions of alpha and beta components to the overall diversity (Crist,  
116 Veech, Gering, & Summerville, 2003). The use of Hill's equivalent numbers of species  
117 offers the possibility of jointly analysing the taxonomic, phylogenetic and functional  
118 components of diversity through a nested hierarchy of multiple scales, in order to  
119 identify the spatial scales that most contribute to the diversity of species, genes and  
120 functional traits (Alberdi & Gilbert, 2019; Pavoine, Marcon, & Ricotta, 2016). The scope  
121 of such an approach applied to whole communities of soil invertebrates would be  
122 considerable, as it would result in a comprehensive understanding of the mechanisms

123 supporting soil biological diversity within and between communities, and would help to  
124 reach a better integration of these functionally important organisms in the definition of  
125 conservation strategies.

## 126 **References**

127 Alberdi, A., & Gilbert, M. T. P. (2019). A guide to the application of Hill numbers to DNA-  
128 based diversity analyses. *Molecular Ecology Resources*, 19(4), 804–817. doi:  
129 10.1111/1755-0998.13014

130 Andujar, C., Arribas, P., Yu, D. W., Vogler, A. P., & Emerson, B. C. (2018). Why the COI  
131 barcode should be the community DNA metabarcode for the metazoa. *Molecular Ecology*,  
132 27(20), 3968–3975. doi: 10.1111/mec.14844

133 Arribas, P., Andujar, C., Salces-Castellano, A., Emerson, B. C., & Vogler, A. P. (2020). The  
134 limited spatial scale of dispersal in soil arthropods revealed with whole-community  
135 haplotype-level metabarcoding. *Molecular Ecology*, *in press*. doi: 10.1111/mec.15591

136 Boyle, E. E., & Adamowicz, S. J. (2015). Community Phylogenetics: Assessing Tree  
137 Reconstruction Methods and the Utility of DNA Barcodes. *Plos One*, 10(6), e0126662.  
138 doi: 10.1371/journal.pone.0126662

139 Craft, K. J., Pauls, S. U., Darrow, K., Miller, S. E., Hebert, P. D. N., Helgen, L. E., ... Weiblen, G.  
140 D. (2010). Population genetics of ecological communities with DNA barcodes: An  
141 example from New Guinea Lepidoptera. *Proceedings of the National Academy of Sciences*  
142 *of the United States of America*, 107(11), 5041–5046. doi: 10.1073/pnas.0913084107

143 Crist, T. O., Veech, J. A., Gering, J. C., & Summerville, K. S. (2003). Partitioning species  
144 diversity across landscapes and regions: A hierarchical analysis of alpha, beta, and  
145 gamma diversity. *American Naturalist*, 162(6), 734–743. doi: 10.1086/378901

146 Decaëns, T. (2010). Macroecological patterns in soil communities. *Global Ecology and*

147 *Biogeography*, 19, 287–302.

148 Ettema, C. H., & Wardle, D. A. (2002). Spatial soil ecology. *Trends in Ecology & Evolution*,  
149 17(4), 177–183. doi: 10.1016/S0169-5347(02)02496-5

150 Mouquet, N., Devictor, V., Meynard, C. N., Munoz, F., Bersier, L.-F., Chave, J., ... Thuiller, W.  
151 (2012). Ecophylogenetics: Advances and perspectives. *Biological Reviews*, 87(4), 769–  
152 785. doi: 10.1111/j.1469-185X.2012.00224.x

153 Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J.-M., Gielly, L., Cordonnier, T., ... Cecillon, L.  
154 (2015). Landscape-scale distribution patterns of earthworms inferred from soil DNA.  
155 *Soil Biology & Biochemistry*, 83, 100–105. doi: 10.1016/j.soilbio.2015.01.004

156 Pavoine, S., Marcon, E., & Ricotta, C. (2016). “Equivalent numbers” for species,  
157 phylogenetic or functional diversity in a nested hierarchy of multiple scales. *Methods in*  
158 *Ecology and Evolution*, 7(10), 1152–1163. doi: 10.1111/2041-210X.12591

159 Porter, T. M., & Hajibabaei, M. (2018). Over 2.5 million COI sequences in GenBank and  
160 growing. *Plos One*, 13(9), e0200177. doi: 10.1371/journal.pone.0200177

161 Rougerie, R., Decaens, T., Deharveng, L., Porco, D., James, S. W., Chang, C.-H., ... Hebert, P.  
162 D. N. (2009). DNA barcodes for soil animal taxonomy. *Pesquisa Agropecuaria Brasileira*,  
163 44(8), 789–802. doi: 10.1590/S0100-204X2009000800002

164 Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA.  
165 *Molecular Ecology*, 21(8), 1789–1793. doi: 10.1111/j.1365-294X.2012.05542.x

166 Thakur, M. P., Phillips, H. R. P., Brose, U., De Vries, F. T., Lavelle, P., Loreau, M., ...  
167 Cameron, E. K. (2020). Towards an integrative understanding of soil biodiversity.  
168 *Biological Reviews*, 95(2), 350–364. doi: 10.1111/brv.12567

169 Turon, X., Antich, A., Palacin, C., Praebel, K., & Wangenstein, O. S. (2020). From  
170 metabarcoding to metaphylogeography: Separating the wheat from the chaff. *Ecological*  
171 *Applications*, 30(2), e02036. doi: 10.1002/eap.2036

172 Yu, D. W., Ji, Y., Emerson, B. C., Wang, X., Ye, C., Yang, C., & Ding, Z. (2012). Biodiversity  
173 soup: Metabarcoding of arthropods for rapid biodiversity assessment and  
174 biomonitoring. *Methods in Ecology and Evolution*, 3(4), 613–623. doi: 10.1111/j.2041-  
175 210X.2012.00198.x

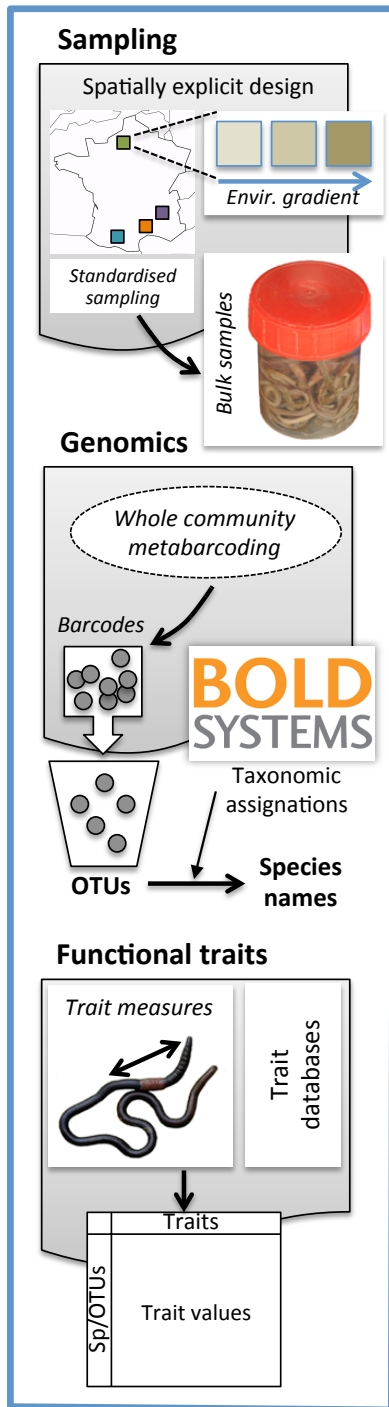
176 Zinger, L., Taberlet, P., Schimann, H., Bonin, A., Boyer, F., De Barba, M., ... Chave, J. (2019).  
177 Body size determines soil community assembly in a tropical forest. *Molecular Ecology*,  
178 28(3), 528–543. doi: 10.1111/mec.14919

179

180 **FIGURE 1** General workflow for a comprehensive analysis of soil faunal communities  
181 through the use of whole community DNA metabarcoding (WCM). Soil invertebrates are  
182 sampled through spatially explicit design and standardised sampling protocols, before  
183 being processed through WCM as described in Arribas et al. (2020). Sequences are  
184 aggregated into operational taxonomic units (OTUs), which are further identified to  
185 species by comparison with available DNA barcode libraries (e.g. Barcode of Life Data  
186 systems – BOLD). Species names are further used to query functional trait databases, or  
187 alternatively traits are described from voucher samples. The complete workflow results  
188 in the production of dataset triplets combining taxonomic, phylogenetic and functional  
189 descriptors of soil invertebrate communities, which can be used to further explore  
190 fundamental aspects of community ecology, evolutionary biology and conservation of  
191 soil invertebrates (see text for details). Envir = environment; Geogr = geographic; TPF  
192 = taxonomic / phylogenetic / functional; PD = phylogenetic dispersion; FD = functional  
193 dispersion; SES = standardised effect size.

194

## Data production

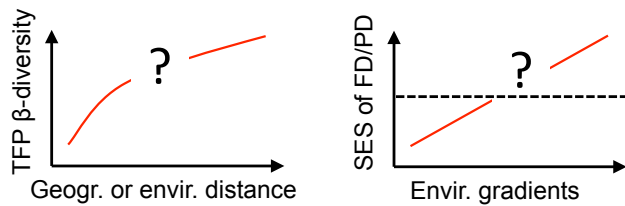


## Dataset triplets

- (T) Taxonomic dataset (OTUs/species)
- (P) Phylogenetic / genetic dataset (barcodes)
- (F) Functional dataset (functional traits)

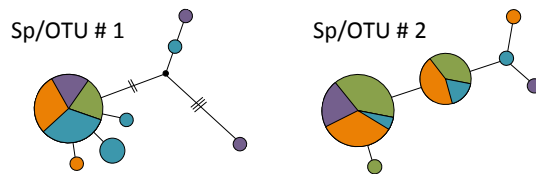
## Community ecology

- Niche-based versus neutral processes



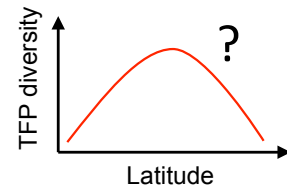
## Comparative phylogeography

- Evolutionary history of species pools
- Trait-related colonisation processes



## Macroecology

- How many species?
- Large scale gradients in TFP diversities



## Conservation

- Critical scales to TFP diversities

$$\gamma = \alpha_1 + \sum_{i=1}^m \beta_i$$