# Contents


Acknowledgments xiii

**1 Introduction to environmental DNA (eDNA) 1**

1.1 Definitions 1
1.2 A brief history of eDNA analysis 2
1.3 Constraints when working with eDNA 3
1.4 Workflow in eDNA studies and main methods used 4
1.5 Environmental DNA as a monitoring tool 5

**2 DNA metabarcode choice and design 7**

2.1 Which DNA metabarcode? 7
2.2 Properties of the ideal DNA metabarcode 8
2.3 *In silico* primer design and testing 9
2.3.1 Prerequisites 10
2.3.2 Reference sequences: description, filtering, and formatting for `ecoPrimers` 10
2.3.3 *In silico* primer design with `ecoPrimers` 11
2.3.3.1 The `ecoPrimers` output 11
2.3.4 *In silico* primer testing with `ecoPCR` 11
2.3.4.1 The `ecoPCR` output 14
2.3.4.2 Filtering of the `ecoPCR` output 16
2.3.4.3 Evaluation of primer conservation 16
2.3.4.4 Taxonomic resolution and *Bs* index 17
2.4 Examples of primer pairs available for DNA metabarcoding 19

**3 Reference databases 21**

3.1 Extracting reference databases from EMBL/GenBank/DDBJ 21
3.1.1 Downloading a local copy of EMBL 21
3.1.2 Identifying sequences corresponding to the relevant metabarcode 23
3.2 Marker-specific reference databases 23
3.2.1 Nuclear rRNA gene reference databases 23
3.2.2 Eukaryote-specific databases 24
3.3 Building a local reference database 25
3.3.1 PCR-based local reference database 26
3.3.2 Shotgun-based local reference database 27
3.4 Current challenges and future directions 27

**4 Sampling** 28

4.1 The cycle of eDNA in the environment 28
4.1.1 State and origin 28
4.1.2 Fate 29
4.1.3 Transport 29
4.2 Sampling design 30
4.2.1 Focusing on the appropriate DNA population 31
4.2.2 Defining the sampling strategy 32
4.3 Sample preservation 33

**5 DNA extraction** 35

5.1 From soil samples 35
5.2 From sediment 39
5.3 From litter 39
5.4 From fecal samples 39
5.5 From water samples 40

**6 DNA amplification and multiplexing** 41

6.1 Principle of the PCR 41
6.2 Which polymerase to choose? 43
6.3 The standard PCR reaction 44
6.4 The importance of including appropriate controls 45
6.4.1 Extraction negative controls 45
6.4.2 PCR negative controls 45
6.4.3 PCR positive controls 46
6.4.4 Tagging system controls 46
6.4.5 Internal controls 46
6.5 PCR optimization 46
6.6 How to limit the risk of contamination? 48
6.7 Blocking oligonucleotides for reducing the amplification of undesirable sequences 50
6.8 How many PCR replicates? 51
6.9 Multiplexing several metabarcodes within the same PCR 52
6.10 Multiplexing many samples on the same sequencing lane 52
6.10.1 Overview of the problem 52
6.10.2 Strategy 1: single-step PCR with Illumina adapters 54
6.10.3 Strategy 2: two-step PCR with Illumina adapters 55
6.10.4 Strategy 3: single-step PCR with tagged primers 55

**7 DNA sequencing** 58

7.1 Overview of the first, second, and third generations of sequencing technologies 58
7.2 The Illumina technology 59
7.2.1 Library preparation 59
7.2.2 Flow cell, bridge PCR, and clusters 60

7.2.3 Sequencing by synthesis 62
7.2.4 Quality scores of the sequence reads 63

**8 DNA metabarcoding data analysis 65**

8.1 Basic sequence handling and curation 65
8.1.1 Sequencing quality 65
8.1.1.1 The pros and cons of read quality-based filtering 65
8.1.1.2 Quality trimming software 68
8.1.2 Paired-end read pairing 68
8.1.3 Sequence demultiplexing 69
8.1.4 Sequence dereplication 69
8.1.5 Rough sequence curation 70
8.2 Sequence classification 70
8.2.1 Taxonomic classification 71
8.2.2 Unsupervised classification 73
8.2.3 Chimera identification 75
8.3 Taking advantages of experimental controls 76
8.3.1 Filtering out potential contaminants 76
8.3.2 Removing dysfunctional PCRs 78
8.4 General considerations on ecological analyses 80
8.4.1 Sampling effort and representativeness 81
8.4.1.1 Evaluating representativeness of the sequencing per PCR 81
8.4.1.2 Evaluating representativeness at the sampling unit or site level 81
8.4.2 Handling samples with varying sequencing depth 83
8.4.3 Going further and adapting the ecological models to metabarcoding 84

**9 Single-species detection 85**

9.1 Principle of the quantitative PCR (qPCR) 85
9.1.1 Recording amplicon accumulation in real time via fluorescence measurement 85
9.1.2 The typical amplification curve 86
9.1.3 Quantification of target sequences with the Ct method 86
9.2 Design and testing of qPCR barcodes targeting a single species 87
9.2.1 The problem of specificity 87
9.2.2 qPCR primers and probe 88
9.2.3 Candidate qPCR barcodes 88
9.3 Additional experimental considerations 88
9.3.1 General issues associated with sampling, extraction, and PCR amplification 88
9.3.2 The particular concerns of contamination and inhibition 88

**10 Environmental DNA for functional diversity 90**

10.1 Functional diversity from DNA metabarcoding 90
10.1.1 Functional inferences 90
10.1.2 Targeting active populations 92

10.2 Metagenomics and metatranscriptomics: sequencing more than a barcode 93
10.2.1 General sampling constraints 94
10.2.1.1 Optimization of the number of samples 94
10.2.1.2 Enrichment in target organisms 94
10.2.1.3 Enrichment in functional information 95
10.2.2 General molecular constraints 96
10.2.3 From sequences to functions 96
10.2.3.1 Assembling (or not) a metagenome 97
10.2.3.2 Sorting contigs or reads in broad categories 97
10.2.3.3 Extracting functional information via taxonomic inferences 98
10.2.3.4 Functional annotation of metagenomes 98

**11 Some early landmark studies 99**

11.1 Emergence of the concept of eDNA and first results on microorganisms 99
11.2 Examining metagenomes to explore the functional information carried by eDNA 100
11.3 Extension to macroorganisms 102

**12 Freshwater ecosystems 104**

12.1 Production, persistence, transport, and detectability of eDNA in freshwater ecosystems 104
12.1.1 Production 104
12.1.2 Persistence 104
12.1.3 Transport/diffusion distance 105
12.1.4 Detectability 106
12.2 Macroinvertebrates 106
12.3 Diatoms and microeukaryotes 106
12.4 Aquatic plants 107
12.5 Fish, amphibians, and other vertebrates 107
12.5.1 Species detection 107
12.5.2 Biomass estimates 108
12.6 Are rivers conveyer belts of biodiversity information? 108

**13 Marine environments 110**

13.1 Environmental DNA cycle and transport in marine ecosystems 110
13.2 Marine microbial diversity 111
13.3 Environmental DNA for marine macroorganisms 112

**14 Terrestrial ecosystems 114**

14.1 Detectability, persistence, and mobility of eDNA in soil 114
14.2 Plant community characterization 116
14.3 Earthworm community characterization 117
14.4 Bacterial community or metagenome characterization 117
14.5 Multitaxa diversity surveys 119

**15 Paleoenvironments** 121

15.1 Lake sediments 121
15.1.1 Pollen, macrofossils, and DNA metabarcoding 121
15.1.2 Plants and mammals from Lake Anterne 121
15.1.3 Viability in the ice-free corridor in North America 122
15.2 Permafrost 125
15.2.1 Overview of the emergence of permafrost as a source of eDNA 125
15.2.2 Large-scale analysis of permafrost samples for reconstructing past plant communities 125
15.3 Archaeological midden material 126
15.3.1 Bulk archaeological fish bones from Madagascar 126
15.3.2 Midden from Greenland to assess past human diet 126

**16 Host-associated microbiota** 127

16.1 DNA dynamics 127
16.2 Early molecular-based works 127
16.3 Post-holobiont works 128

**17 Diet analysis** 131

17.1 Some seminal diet studies 131
17.1.1 Proof of concept—analyzing herbivore diet using next-generation sequencing 131
17.1.2 Assessing the efficiency of conservation actions in Białowieża forest 132
17.1.3 Characterizing carnivore diet, or how to disentangle predator and prey eDNA 133
17.1.4 Analyzing an omnivorous diet, or integrating several diets in a single one 133
17.2 Methodological and experimental specificities of eDNA diet analyses 135
17.2.1 eDNA sources 135
17.2.1.1 Feces 135
17.2.1.2 Gut content 135
17.2.1.3 Whole body 135
17.2.2 Quantitative aspects 136
17.2.2.1 Relationship between the amount of ingested food and DNA quantity in the sample 136
17.2.2.2 Quantifying DNA with PCR and next-generation sequencing 137
17.2.2.3 Empirical correction of abundances 138
17.2.3 Diet as a sample of the existing biodiversity 138
17.2.4 Problematic diets 139

**18 Analysis of bulk samples** 140

18.1 What is a bulk sample? 140
18.2 Case studies 140

18.2.1 Bulk insect samples for biodiversity monitoring 140
18.2.2 Nematode diversity in tropical rainforest 141
18.2.3 Marine metazoan diversity in benthic ecosystems 141
18.3 Metabarcoding markers for bulk samples 141
18.4 Alternative strategies 143

**19 The future of eDNA metabarcoding 144**

19.1 PCR-based approaches 144
19.1.1 Single-marker approach 144
19.1.2 Multiplex approach 144
19.2 Shotgun-based metabarcoding 145
19.2.1 Without enrichment by capture 145
19.2.2 With enrichment by capture 146
19.3 Toward more standardization 146
19.3.1 For sound comparisons across studies 146
19.3.2 For environmental monitoring 147
19.4 Next-generation reference databases 148
19.5 Open questions 148
19.5.1 What will be the impact of new sequencing technologies on eDNA analysis? 148
19.5.2 Will some specific repositories be developed for DNA metabarcoding? 148
19.5.3 Will metabarcoding provide quantitative results? 149
19.5.4 Will metabarcoding be fully integrated into ecological models and theories? 150
19.5.5 How do we train students and managers to effectively integrate this tool into academic and operational ecological research and monitoring? 150

Appendix 1 151
Appendix 2 217
Appendix 3 220
References 223
Index 247