Chapter 1

Introduction

FRANS J. DE BRUIJN

In this first volume of the Handbook, metagenomics is introduced, together with computer-assisted analysis, information on consortia and databases, and as a number of complementary methods, such as microarrays, metatranscriptomics, metaproteomics, metabolomics, phenomics (the “omics”), and single-cell analysis.

Part 1, “Background Chapters,” contains a number of chapters on nonmetagenomic methods, such as different genomic fingerprinting techniques and their analysis and level of resolution, as well as the first approach to metagenomics (Chapter 2). All these methods are still used today.

In Part 2, “The Species Concept,” several experts examine the parameters to call something a new species and provide suggestions to authors when it is proper to call a novel isolate (operating taxonomic unit (OTU)) a new species. The recommendations of two expert meetings on the topic are summarized in another chapter in this part describing the 70% DNA–DNA hybridization level as essential in the species concept. This discussion is very relevant to all phylogenetic studies in both volumes of the Handbook.

In Part 3, metagenomics is introduced and a number of practical parameters of this technique are outlined. An introduction to metagenomics and the other “omics” is presented in Chapter 14. Three subsequent chapters deal with the 16S rRNA gene as phylogenetic marker and also examine the pitfalls of its use. Three chapters describe the impact of next-generation sequencing on metagenomics, examine its accuracy and quality of reads, and review the potential and challenges of environmental shotgun sequences for studying the hidden world of microbes. Metagenomics can involve (a) the generation and analysis of clone libraries which can be screened for particular properties and (b) random sequencing of metagenomic DNA. The former is discussed in an article on vector tools and functional screening of metagenomic libraries (see also Parts 6 and 7, Vol. II). The latter is used in many other articles in the Handbook. The remaining articles in this section introduce various technical aspects of metagenomics, as well as novel approaches such as gene-targeted metagenomics, using homing endonuclease restriction and marker insertion for phylogenetic studies, finding integrons, arrayOme- and rRNA-facilitated mobilome discovery, and improved serial analysis of V1 ribosomal sequence tags (SARST-V1) to study bacterial diversity. A plethora of other studies in various habitats are presented in Volume II of this Handbook.

In Part 4, some consortia and databases are discussed, including the Metacontrol consortium focusing on the metagenomics of suppressive soils, the Terragenome consortium to provide a metagenomic shotgun and fosmid sequencing analysis of a “reference” soil, and the Argentinian BIOSPAS consortium aimed at bringing together a group of scientists employing metagenomic and associated approaches. This is followed by a description of the Human Gut Microbiome Initiative (HGMI) and the related Human Microbiome Project (HMP). Chapter 36 in this part describes the Ribosomal Database Project, an irreplaceable source for phylogenetic studies, using the rRNA genes as target (see Chapter 15, Vol. I). The final chapter in this part describes the Metagenomics RAST server as a public resource for automated phylogenetic and functional analysis of Metagenomes.

In Part 5, a smorgasbord of computer programs is presented essential for the analysis of (meta)genomic data. Clearly, computer-assisted analysis is a crucial component of every metagenomic project, and progress in the field is dependent on creating programs and databases for ever-growing datasets and can be the limiting factor for large metagenomic, transcriptomic,
proteomic, and metabolomic projects. It equals in importance to the development of higher throughput novel sequencing methods (see Chapter 18, Vol. I). The authors in Part 5, as well as all other authors, have been asked to highlight the programs and web sites used in their chapters; therefore in addition to the limited programs highlighted in Part 5, a wealth of further information and other programs can be found in the chapters in Volumes I and II.

In Part 6 a number of complementary approaches to metagenomics are presented, including metagenomics approaches in systems biology, the use of stable isotope probing, and subtractive hybridization.

In Part 6A the use of microarrays, including phylochips and geochips and metagenomic arrays, is discussed and examples in different habitats, such as NASA rocket cleanrooms, are given. This part also contains a chapter on phenotypic arrays or “phenomics,” another “omic” technique, which can reveal the metabolic capacity of microbes in microplates.

In Part 6B, some examples of metatranscriptomic analysis are presented, which permit a glimpse into the metagene expression profile in various environments, such as the symbiotic protist community in Reticulitermes and comparative day and night metatranscriptomics of microbial communities in the North Pacific. In addition a “double RNA” approach is presented to simultaneously assess the structure and function of microbial communities, and one chapter on the metatranscriptomics of eukaryotes is included.

In Part 6C, metaproteomics approaches are highlighted, and examples are presented on the proteomics of microbial stress responses, the metaproteomic analysis of Chesapeake Bay microbial communities, high-throughput proteomics in cyanobacteria, and global proteomic analysis of the chromate response in Arthrobacter.

In Part 6D, metabolomics is highlighted, which requires more sophisticated tools such as mass spectrometry. Examples include (a) two chapters that review the small molecule dimension and high-resolution tools to monitor bacterial growth on a molecular level, (b) one chapter on metabolomics in plants, where the metabolomics techniques are well established, and (c) a chapter on metabolite identification, pathways and “omic” integration using databases and other tools.

In Part 6E, a highly specialized complementary approach is described, namely the isolation and use of single cells for metagenomic and other analysis. None of the parts described above are comprehensive. They mainly give a short insight about what one can do in addition to metagenomics to extract more functional data from the system under study to answer the following questions: “Who is there?” and “What are they doing?” An attempt was made to select studies in very different habitats, and a variety of approaches are highlighted. This is continued and expanded upon in Volume II.