Reading the Book of Life – Omics as a Universal Tool Across Disciplines

Jan David Brüwer and Hagen Buck-Wiese

Abstract

In the last centuries, new high-throughput technologies, including sequencing and mass-spectrometry, have emerged and are constantly refurbished in order to decipher the molecular code of life. In this review, we summarize the physiological background from genes via transcriptome to proteins and metabolites and discuss the variety of dimensions in which a biological entity may be studied. Herein, we emphasize regulatory processes which underlie the plasticity of molecular profiles on different ome layers. We discuss the four major fields of omic research, namely genomics, transcriptomics, proteomics, and metabolomics, by providing specific examples and case studies for (i) the assessment of functionality on molecular, organism, and community level; (ii) the possibility to use omic research for categorization and systematic efforts; and (iii) the evaluation of responses to environmental cues with a special focus on anthropogenic influences. Thereby, we exemplify the knowledge gains attributable to the integration of information from different omes and the enhanced precision in predicting the phenotype. Lastly, we highlight the advantages of combining multiple omics layers in assessing the complexity of natural systems as meta-communities and -organisms.

J. D. Brüwer (✉)
Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

Max Planck Institute for Marine Microbiology, Bremen, Germany
e-mail: bruewer_j@gmx.de

H. Buck-Wiese (✉)
Faculty of Biology and Chemistry, University of Bremen, Bremen, Germany

Max Planck Institute for Marine Microbiology, Bremen, Germany
e-mail: h.buckwiese@googlemail.com

Introduction and Historical Background

The discovery of nucleic acids in 1896 by Friedrich Miescher and the suggestion of deoxyribonucleic acid (DNA) as the genetic material by Avery, MacLeod, and McCarty in 1943 revolutionized the life sciences (Avery et al. 1943; Dahm 2005). Genomics, from the suggested word "genome" for haploid chromosome sets by Hans Winkler (Noguera-Solano et al. 2013), arose with the aim to decipher the molecular language. It took another 10 years before Franklin, Wilkins, Watson, and Crick unraveled the double-helical structure of DNA in 1953 (Dahm 2005). The conversion from nucleotide sequence into amino acid was first recognized, when Heinrich Matthaei and Marshall Nirenberg discovered that the RNA sequence of three Uracil bases codes for the amino acid phenylalanine with their so-called Poly-U experiment (Nirenberg 2004; Dahm 2005). Five years later, in 1966, the translation of all base combinations into the 20 protein-forming amino acids had been resolved (Nirenberg 2004). For nucleotide sequence analysis, Frederick Sanger and colleagues developed the first widely applied method, the Sanger sequencing, in 1977 and, thus, established the foundation for modern genomic and transcriptomic research (Box 1) (Sanger et al. 1977). In more recent years, high-throughput molecular technologies, e.g., next-generation sequencing (NGS) (Box 1) and mass spectrometry (Box 2) have developed, enabling genome-scale deciphering of the molecular signatures, which encode life on earth.

These technologies provide the opportunity for a wide range of studies which can be divided into four major fields according to the targeted molecules: genomics, transcriptomics, proteomics, and metabolomics. In definition, genomics describes the analysis of any genetic material (DNA) isolated from an organism or the environment. It includes, for example, whole genome sequencing and detection methods such as environmental DNA (eDNA). Transcriptomics is the study of any form of RNA, including messenger RNA (mRNA), transfer RNA (tRNA), ribosomal RNA (rRNA),

Box 1: Nucleic Acid Sequence Analysis

Background

The nucleic acids contain information in the shape of a code constituting of two purines, Adenine A and Guanine G, and two pyrimidine bases, Cytosine C and either Thymine T in DNA or Uracil U in RNA. Selective pairing of A with T and G with C gives rise to the stable double strand structure of DNA and confers a mechanism to pass on the information in the coded sequence via polymerization, i.e. DNA replication and RNA transcription (in this case, substituting U for T) (Alberts et al. 2008; Klug et al. 2012). The widely applied DNA/RNA sequencing methods to read the nucleotide code are based on this selective binding.

The first sequencing developed by Sanger in the 1970s required four separate polymerizations, each with a fraction of dideoxynucleotides (ddNTPs) which would terminate the elongation – hence the name 'chain-termination method' (Lu et al. 2016). Parallel size separation (using gel-electrophoresis) of the synthesized strands, each with a specific dd-nucleotide at the end, and subsequent radioactive detection allowed to infer the order of the different bases in the template's sequence. Modern techniques for Sanger sequencing are based on fluorescently labeled ddNTPs, emitting differentiable signals, which can be detected by a laser and evaluated electronically (Schuster 2008). Recently developed second-generation sequencing (such as Illumina) use dNTPs which emit a base-specific fluorescent signal when the phosphordiester bond is formed and the DNA elongated. Different to the traditional Sanger sequencing, the process does not require termination and every elongation process yields a signal per nucleic acid. The advantages of these sequencing methods lie in high-throughput through the simultaneous sequencing of multiple DNA/RNA fragments (e.g., from environmental samples) from a variety of organisms with usually reliable high-quality results (Schuster 2008). The drawbacks belay in comparatively short sequence strands (about 100–300 bp), demanding assemblies to solve the 'puzzle' of different short fragments. However, third-generation sequencing (such as offered by PacBio with the SMRT cell) make use of double-stranded DNA with two hairpin structures at the end, the so-called SMRTbell. This way, fragments of several thousand base pairs may be sequenced, which may subsequently be complemented by shorter fragments to maintain the quality standard via high coverage (Rhoads and Au 2015).

The emerging fourth generation sequencing technique, the nanopore sequencing (such as the MiniION by Oxford Nanopore Techniques), does not require previous amplification but aims at directly sequencing single molecules and promises to sequence tens of kilobases (kb). A membrane is equipped with nanopores that is selectively permeable for DNA and RNA. An electric force is driving the electrophoresis of the negatively charged fragments towards the anode and, thus, into the membrane. A motor protein is ratcheting the fragment through the membrane. This causes different perturbations of the membrane current depending on the nucleotide, which may be computationally translated into base sequences (Cherf et al. 2012; Feng et al. 2015). Different from previous sequencing methods, the fourth generation nanopore sequencing may even be used to analyze proteins, polymers, and other single-strand macromolecules (Feng et al. 2015).

Strategies

To target a particular portion of the queried nucleotide sequence, e.g., targeting the 16S rRNA/rDNA of microorganisms for phylogenetic assessment, specific primer sequences can be used.

A variety of techniques grouped under the description of restriction site-associated DNA sequencing (RADseq) is currently in scope for assessing genotypic differences of a range of organisms, including those with largely unknown genomes. These techniques are based on digestion of isolated DNA with one or few restriction enzymes and subsequent sequencing of resulting fragments. As most restriction sites prevail among specimen and closely related species, predominantly similar sets of loci are sequenced, at which different alleles can be identified (Andrews et al. 2016).

In case of whole genome sequencing using NGS, short fragments of DNA of few hundred base pair length are inserted into vectors, called library. To aid in later assembly, libraries with shotgun mate pair fragments of specified greater lengths complement the short vector sequences, which consist of a high fragment coverage. After standard quality controls of the reads (including adapter and primer removal), the assembly of the genome from the multitude of small sequences relies on overlapping regions and mate pairs (e.g., Baumgarten et al. 2015).

Prior to RNA sequencing, the RNA-template has to be transcribed into a cDNA, using a reverse transcriptase. A quantitative interpretation of transcriptome and

Box. 1: (continued)

metagenome data analyses has to be treated with caution due to exponential amplification steps. However, normalization steps to account for differential amplification within samples, as well as differential sequencing depth across samples, may be used to gain better estimates of quantities as well as maintaining data comparable. This may be achieved by calculation of "Fragments Per Kilobase of exon model per million Mapped reads" (FPKM). Further biostatistic normalization to eliminate sequencing biases, e.g., using nCounter (Geiss et al. 2008), may be helpful in evaluation of the data (Liu et al. 2016).

lution achieved by retention. The experimental approach requires a comparison of the metabolic profile yielded by the mass spectrometer either to a standard or between two or more samples. A bioinformatic overlay of the produced profiles provides information on significant differences in abundance and thereby delineates molecules of interest. Their mass-per-charge ratios now serve to find reference molecules in databases. However, due to the novelty of metabolome-wide studies, there is a considerable number of molecules, which remains to be identified and entered into the repositories. If there is a mass-to-charge hit and standards are available for the molecules of interest, the identity can be confirmed via retention times and MS/MS profiles (Patti et al. 2012).

Box 2: Mass-Per-Charge of Peptides and Metabolites

Protein and targeted metabolite analyses, including antibody, ionization, and spectroscopy approaches, date back more than a century. Technical advances in the field of mass spectrometry (MS) are, however, revolutionizing the possibilities in these fields, now supporting proteome-wide peptide sequence identification and untargeted metabolome characterization and comparison.

Protein studies have traditionally been relying on the usage of antibodies on a small scale but as a precisely localizing method. Nevertheless, limited availability of antibodies for different protein structures, comparatively low throughput, high costs for antibody production, and low quantitative comparability due to lacking standards have hampered proteome-scale assessments. Deep high-throughput MS has emerged as an opportunity to read-out relative and absolute concentrations of proteins genome-wide. Label-free quantification via tandem mass spectrometry (MS/MS) allows the recognition of individual peptide spectra. These are compared to entries in databases, optimally containing all peptide sequences expected to be present, but few irrelevant ones. Current developmental and research efforts, though, target the *de novo* determination only from the peptide's spectrum (Liu et al. 2016; Ruggles et al. 2017).

Current-standard for untargeted metabolome analysis is a liquid chromatography coupled with mass spectrometry. Since theoretically every type of small molecule possesses a unique retention time and a unique mass-per-charge ratio, this procedure separates and characterizes each metabolite. Adjustments in liquid phases regarding hydrophilic and hydrophobic components and their directions can improve the reso-

and micro RNA (miRNA). The study of the protein content of an organism and its respective functions is comprised by proteomics. Metabolomics deals with any small molecules that are produced or ingested by an organism (Handelsman 2004; Patti et al. 2012; Pascault et al. 2015; Beale et al. 2016; Liu et al. 2016).

In this review, we will delineate the physiological background of omics research and will exemplify the wide spectrum of applicability under aspects of functionality, systematics, and response to environmental cues. Finally, we aim to highlight the significance of multi-omics for an in-depth understanding of complex systems.

Physiological Background

The genome depicts the inherited foundation within a cell and is – apart from epigenetic changes – consistent in almost every healthy somatic cell of a multicellular organism. It encodes for the high variety of proteins, as well as non-protein coding sequences, such as **r**ibosomal RNA (rRNA), **t**ransfer RNA (tRNA), and **mi**cro RNA (miRNA) (Alberts et al. 2008).

Gene expression begins with the transcription of a DNA sequence into a pre-mRNA. The newly synthesized nucleotide sequence constitutes a reverse complement of the coding strand with ribose phosphates instead of deoxyribose phosphates forming the backbone, and Uracil pairing with Adenine instead of Thymine (Alberts et al. 2008).

Promoter sequences upstream of open reading frames, the DNA region to be transcribed, contribute significantly to expression by recruiting the RNA polymerase. However, expression profiles remain a complex puzzle due to influences of *cis*- and *trans*-regulatory motifs and binding of transcription factors. Further, epigenetic modifications as

cytosine methylation, histone acetylation, and changes in chromatin structure may lead to a subsequently altered transcriptome (Alberts et al. 2008).

Due to the translation of mRNA into amino acids via the triplet code, proteins are in a qualitative sense direct product of genes with mRNA transcripts as intermediates. This allows functional predictions of genes via comparison of sequence similarities to annotated genes in a highly curated database, such as NCBI RefSeq (O'Leary et al. 2016).

In eukaryotes, the RNA sequence is, nevertheless, subject to possible modifications, which may impede the recognition of a gene-protein pair. Variable intron removal from maturing mRNAs by splicing may lead to multiple isoforms from a single pre-mRNA (Alberts et al. 2008). Further, RNA editing (see example in section "Response to environmental cues") may introduce sequence alterations as a co- or post-transcriptional modification, not to be confused with de-capping, splicing, and poly(A)-removal (see e.g., Klug et al. 2012; Liew et al. 2017).

Sequence Alterations Influence Protein Functioning

Non-synonymous sequence alterations, i.e. single nucleotide exchanges, deletions, or insertions, may significantly influence or disrupt protein functioning. Firstly, a protein's physiological role is sensitive to secondary and tertiary structure formation and stability (e.g., α-helix and cysteine double bounds, respectively), which may be significantly altered due to aforementioned non-synonymous sequence alterations. Secondly, the phosphorylation of serine, threonine, and tyrosine, as well as acetylation and ubiquitylation of lysine are major post-translational modifications, which are involved in triggering activation and degradation (reviewed in Klug et al. 2012; Ruggles et al. 2017). Thus, sequence alterations which lead to the exchange of one of these four amino acids are likely to affect the protein's performance. Lastly, guiding and localization sequences are essential to position proteins in cellular compartments or membranes. For example, the nuclear membrane of most eukaryotic cells is freely permeable to molecules up to 9 nm. Macromolecules of greater sizes depend on a specific nuclear localization sequence (NLS), which mediates the transport. Alteration of a single amino acid may result in a dysfunction of the NLS and the decreased transport efficiency of the macromolecule into the nucleus (Zanta et al. 1999).

Consequently, complex reactions such as protein-protein interactions, transcription cascades, signaling networks, and metabolic pathways may be altered by single nucleotide exchanges (Kim et al. 2016).

Quantitative Regulation of the Proteome

The physiological roles of RNA reach far beyond the gene to protein transmission, where (pre)-mRNA, rRNA, and tRNA are allocated. For instance, the translation-regulatory roles of miRNAs have been discovered in 1993 (Almeida et al. 2011; see section "Functionality"). In humans, for example, at least 70% of the genome is transcribed into RNA, but only about 2% are effectively translated into protein (Pheasant and Mattick 2007). Consequently, immense proportions of the genome are suggested to encode for quantitative regulation, which can be detected with current omics approaches (Klug et al. 2012). The current state of knowledge considers the abundance of mRNA transcripts to explain up to 84% of the respective protein concentration. This value may vary depending on the respective mRNA, mainly attributable to sequence- or splice isoform-dependent translation rates (Liu et al. 2016). Additionally, induced changes in gene expression, e.g., due to environmental cues, may only be detectable in the proteome after a lag phase (e.g., 6–7 h in mammals; see also section "Response to environmental cues").

The number of copies per gene does not generally define respective transcript nor protein abundances. Genetic diseases or tumors may induce gene copy number alterations (CNAs). In such cases, transcriptome and proteome do mostly not exhibit the same fold changes as could be expected from the CNAs in the genome. Negative feedback loops, called buffering, may occur on the transcriptional and translational level. There are, however, plenty of sequence-specific exceptions to this general pattern, which are, therefore, possibly involved in the symptomatic (Liu et al. 2016 and references therein).

Metabolomics

The entirety of small molecules within an organism, the metabolome, constitutes a biochemical representation. It is substance to continuous turn-over, alteration, and relocation by the physiological machinery of RNAs and, most of all, proteins (e.g., Patti et al. 2012; Beale et al. 2016). While targeted metabolomics assesses only a fraction of particular interest, newly emerged technologies enable untargeted detection and quantification of almost the entire metabolome (Patti et al. 2012).

Untargeted metabolomics combined with genomic and/or transcriptomic data may allow the inference of gene and protein function, as well as metabolic cascades and pathways. It becomes possible to detect physiological attributes such as the use of substrates, secondary metabolite secretion, or possible inter-individual signaling, and connect these to the presence or expression of genes (Freilich et al. 2011;

Llewellyn et al. 2015; Kim et al. 2016). In combination with information on intrinsic or even environmental ontology, it may provide insights into the plastic phenotypic range and might suggest possible adaptation or acclimatization responses (Dick 2017).

Functionality

A genome-wide survey on potential open reading frames and prediction of gene function can help to characterize an organism or study its ecological background. An example from marine plant genetics is the recently published genome of the true seaweed *Zostera marina* (commonly referred to as eelgrass). It contains 20,450 genes, of which a majority (86.6%) were validated using a transcriptomic approach (Olsen et al. 2016). Functional annotation revealed gene losses and gains that could be attributed to the marine habitat. Those included losses of stomatal differentiation, airborne communication, and immune-system-related genes, to name only three examples (Olsen et al. 2016).

Using next-generation sequencing or quantitative PCR (qPCR) approaches, transcript abundances may be assessed (Liu et al. 2016; see also Box 1). As such, this provides a good possibility to estimate biological activity rather than the mere presence and abundance. In microbial ecology, for example, the *nifH* gene is a common biomarker for nitrogen-fixing bacteria, i.e. diazotrophs (Gaby and Buckley 2012). Pogoreutz et al. (2017) queried gene and transcript abundance of *nifH* in order to investigate nitrogen fixation in the coral holobiont (see Box 3 for details on the metaorganism/holobiont concept). They detected autotrophic corals to exhibit a higher *nifH* gene abundance, correlated with increased expression rates. Consequently, the authors suggested that low nitrogen-uptake via heterotrophy may be compensated by the microbial component of the holobiont.

Transcriptomes are interesting in another regard, as some RNA species have regulatory functions, e.g. miR-NAs which are short (about 22 base pairs) single-stranded RNA molecules. They have the potential to align with mRNA via sequence identity and thereby either inhibit the translation or induce degradation (Gottlieb 2017). A single miRNA may bind to several different mRNAs and *vice versa* (Selbach et al. 2008). In humans, the Chromosome 19 miRNA cluster (C19mc) is almost exclusively expressed in the extra-embryonic tissue of the placenta (Luo et al. 2009) and seems to be an important component of the immune system during viral infections (Delorme-Axford et al. 2013). C19mc has been suggested to be a key component of embryonic-maternal communication, as well as essential to suppress a maternal immune response (Gottlieb 2017).

Box 3: Metaorganisms

Evidence supports the notion that all multicellular organisms live in synergistic interdependence with a variety of microorganisms, including bacteria, archaea, and viruses. Together, they form a complex entity, termed holobiont or metaorganism (McFall-Ngai et al. 2013; Bosch and Miller 2016). The microbial community of a metaorganism constantly influences the performance of a metaorganism. The human gastrointestinal tract (GI) microbiota, for example, is of great importance for the digestion and ingestion of nutrients and metabolites. In addition, the human microbiota have been suggested to have great impacts on the behavior and even neurological functions (Turnbaugh et al. 2007; Biagi et al. 2012; Sampson and Mazmanian 2015).

Various studies can show that a metaorganism's microbiome changes and that at least part of the metaorganism can compensate for environmental stressors (Bosch and Miller 2016; Buck-Wiese et al. 2016; Hernandez-Agreda et al. 2016; Ziegler et al. 2017). In order to study the biology of a multicellular organism, the whole metaorganism should therefore be respected.

In a metagenomics and -proteomics study, Leary et al. (2014) assessed the microbial community of biofilms on two different navy ships. The metagenomics data revealed prokaryotic signature to be most abundant on both ships. However, the meta-proteome on the first ship hull was dominated by eukaryotic cytoskeleton proteins, while diatom carbon fixation and photosystem related proteins were most abundant on the second hull. The authors argue that observed differences between metagenomics and -proteomic results may be due to retention of prokaryotic DNA in the biofilm, especially of inactivated or dead bacteria. Further, the eukaryotic proteome is usually larger in size and may exhibit a higher dynamic range of gene expression. In this case, a single omics approach may have provided misleading results (Leary et al. 2014; Beale et al. 2016).

The relatively novel field of untargeted metabolomics can provide sufficiently broad information to infer previously unknown functions. For example, stony corals are in constant association with a variety of microorganisms (Rohwer et al. 2002). About a decade ago, a study by Ritchie (2006) could show that bacteria isolated from the coral mucus microbiome inhibit the growth of several gram-positive and -negative bacteria, and suggested an antimicrobial activity. In addition, Shnit-Orland et al. (2012) could show that *Pseudoalteromonas* spp., a frequent coral mucus symbiont, secrets antimicrobial agents against gram-positive strains (see also Brüwer et al. within the abstracts related to the

chapter "Tropical Aquatic Ecosystems across Time, Space, and Disciplines"). Untargeted metabolomics could be applied to investigate these inferred substances as putative components for medicinal use.

In fact, omic and multi-omic studies on functionality elucidate the significance of the molecular code on the phenotype level. They thereby contribute to the body of knowledge by which we can extrapolate information from molecular reads. Ultimately, they enable us to "write" in the book of life.

Systematics

Scientists often aim to explain complex natural phenomena with comprehensive models, which are constantly adapted and expanded. The species model, for example, is – if not updated – at least constantly discussed in the scientific literature, especially regarding prokaryotes (e.g., Stackebrandt et al. 2002; Wilmes et al. 2009; Amann and Rosselló-Móra 2016).

Mutations are an essential source of genetic variability, which are estimated to occur at (region-specific) constant rates per replication for closely related species (Gillooly et al. 2005) and are used to resolve phylogenetic relationships. Polymorphisms (mostly single nucleotide polymorphism, SNP) in the genome create different alleles that may be targeted by specific restriction enzymes. Some techniques and methods, such as restriction site-associated DNA sequencing (RADseq), enable the detection of many SNPs on a population level in order to study the genetic background of populations and migratory dynamics (Andrews et al. 2016; see also Box 1).

As some parts of the genome are more prone to mutations which could lead to lethal dysfunctions of the encoded molecule, they depict highly preserved regions with significantly lower mutation rates compared to other parts of the genome. Non-lethal mutations that do happen within these regions usually remain in the genome and may be queried by amplification and/or sequencing for phylogenetic assignment.

The DNA regions coding for the small subunit of ribosomal RNA in prokaryotes, the 16S rRNA gene, and the mitochondrially encoded cytochrome c oxidase I in eukaryotes, the COI gene, constitute such highly conserved areas of genetic information, which are commonly used for taxonomic characterization by barcoding approaches (Pimm et al. 2014; see also McCarthy et al. within the abstracts related to this chapter for an ancient DNA example).

Traditional microbial characterization approaches require a cultivation prior to phenotypic classification. However, most marine microbes are very challenging to or not at all cultivable (Pedrós-Alió 2012; Epstein 2013; Amann and Rosselló-Móra 2016). The advances of next-generation sequencing methods provide the possibility to detect and

phylogenetically classify a vast amount of microbes simultaneously, including those that are not cultivable (see also Weinheimer et al. within the abstracts related to this chapter). In a recently published microbial ecology study, Röthig et al. (2016) aimed to characterize the microbiome of the model metaorganism *Aiptasia*. Besides a metagenomics approach by DNA isolation of *Aiptasia* tissue and subsequent high throughput 16S rRNA gene sequencing, the authors applied a culture dependent approach. They detected 295 different taxa in the metagenomic data, while they were only able to culture 14 of those (with a 100% match in the 16S rRNA) (Röthig et al. 2016; see also Slaby et al. within the abstracts related to this chapter for a marine sponge microbiome).

Similarly, the gastrointestinal (GI) microbial community is an important component of a vertebrate metaorganism (see Box 3). Dewar et al. (2013) assessed the residual GI microbiota in feces of the king, gentoo, macaroni, and little penguins by 16S rRNA gene sequencing. The authors detected a diverse microbial community (>5 k operational taxonomic units (OTUs) identified) with significant differences in relative abundances of microbial taxa across penguin species. They further identified known human pathogens (including *Helicobacter, Veillonella, Mycoplasma*, etc.), although their virulence in penguins or other sea-birds remains questionable (Dewar et al. 2013, 2014).

Samples of environmental DNA (eDNA) may be subject to similar analysis of highly conserved genome regions. Such studies would aim to detect DNA traces in the environment to extrapolate on the presence and potentially even abundance of the corresponding organisms (Taberlet et al. 2012; Valentini et al. 2016; see also Mauvisseau et al. within the abstracts related to this chapter). A recent comparison of the efficiency of traditional surveys and eDNA approaches aimed to detect amphibians and fishes in natural aquatic environments by designing group (i.e. amphibian and fish) specific primers of mitochondrial DNA (mtDNA) (Valentini et al. 2016). Two amphibian species (*Triturus marmoratus* and *Pelophylax* sp.) were observed with conventional methods but not detected via barcoding, whilst a total of 64 species could exclusively be recorded in the eDNA samples. The fish survey prompted a similar result. Thus, ecological surveys using eDNA appear to be more thorough and accurate. In addition, they are less destructive, are more (cost) efficient, and do not fully rely on the taxonomic knowledge and species identification of experts, compared to commonly applied surveys. Furthermore, it will be less challenging to establish standardized protocols, thus, making survey studies more comparable, especially if they are conducted across various scientific laboratories (Valentini et al. 2016).

In virology, characterization efforts demand the verification of newly identified viruses by visual evidence (e.g.,

electron microscopy), as well as molecular methods (e.g., sequencing). However, a panel of leading virologists recently proposed that viruses only known from metagenomic samples need to be incorporated into the official classification scheme by the International Committee on Taxonomy of Viruses (ICTV) (Fauquet and Mayo 2001; Simmonds et al. 2017). Paez-Espino et al. (2016), for example, analyzed around 5 terabyte metagenomics data of diverse samples, which led to the discovery of 125,000 new predicted viral genomes as well as a massive increase in the number of putatively identified viral genes (Paez-Espino et al. 2016; Simmonds et al. 2017).

To conclude, metagenomic and metatranscriptomic data may be used to detect, characterize, and taxonomically rank all 'lifeforms', including previously unknown and uncultured organisms and viruses.

Metabolomic analyses, for example of lipids, are commonly performed in marine plankton research to identify the dietary preferences of plankton species. These analyses are based on the fact that some fatty acids (biomarkers) are characteristic of specific groups of phyto- or zooplankton (e.g., 16:1(n-7) for diatoms, 18:4(n-3) for dinoflagellates and 18:1(n-9) for metazoans) and are incorporated into the consumers body tissue largely unmodified, thus retaining a dietary signature (Graeve et al. 1994; Dalsgaard et al. 2003). For example, the fatty acid pattern of the Arctic copepod *Calanus finmarchicus* typically reflects a dinoflagellate nutrition. However, in *in vitro* feeding experiments with the diatom *Thalassiosira antarctica* over several weeks, the fatty acid composition depicted a change towards a diatom-like signature/profile (Graeve et al. 1994) showing the unchanged incorporation of dietary fatty acids. Thus, the fatty acid biomarker concept might allow differentiating the source of phytoplankton (diatom vs. dinoflagellate) and if a species mostly feeds on phytoplankton (herbivory) or other zooplankton (carnivory) giving hints on its trophic position and role in the food web. In this regard, studying the proteome or metabolome can provide clues on the functional role of an organism and might even provide a glimpse on the phylogeny (Jones et al. 2014; Llewellyn et al. 2015).

Functional groups of proteins are not necessarily linked to phylogeny. The Gene Ontology (GO) provides a hierarchical structure based on functional grouping of the gene products. Thus, it helps to identify the physiological role based on molecular function, cellular component, and biological process (The Gene Ontology Consortium 2004, 2017). The Kyoto Encyclopedia of Genes and Genomes (KEGG) has been established to interrelate genes based on their high-level function. Identifying a queried gene in the KEGG PATHWAY database integrates it in a corresponding pathway and may show its connectivity to other genes, thus allows to extrapolate on its physiological impact (Kanehisa et al. 2016, 2017). In summary, GO-terms and the KEGG database may help to identify involved physiological processes, as well as potentially easing the comparison amongst different organisms or species.

Response to Environmental Cues

The recent geological era is called the Anthropocene, because human activities have severely impacted geology and ecosystems, including the alteration of carbon fluxes. Long-term temperature rise and increased short-term fluctuations (IPCC 2007, 2014), as well as pollution accidents, such as oil spills (e.g., McNutt et al. 2012), challenge the adaptive capacities of species. However, technological advances allow us to utilize those capacities to increase the yield of biological products (e.g., Park et al. 2015). To subsequently suggest strategies for conservation and bioengineering, the different omics fields can be a useful tool to identify genetic, transcriptomic, proteomic, or metabolic responses to environmental cues.

A consortium of marine geneticists proposed to investigate the adaptability and resilience to environmental stress in a three-step approach (Voolstra et al. 2015). Firstly, species along a natural gradient should be queried for genetic, epigenetic, or transcriptional differences, to secondly experimentally test the resilience of specimen from each extremum to the corresponding opposite poles. Thirdly, possible (epi-) genetic or transcriptional differences should be investigated and their impacts regarding the environmental parameter evaluated (Voolstra et al. 2015). Following this approach, Ziegler et al. (2017) assessed the plasticity of the microbial community within a metaorganism in response to temperature regimes (Box 3). They were able to show that the microbiome significantly contributed to thermal-stress resilience (Ziegler et al. 2017). This underlines the role of a metaorganism's microbial community when reacting to and coping with environmental change (Bosch and Miller 2016; Buck-Wiese et al. 2016).

Liew et al. (2017) exposed a facultative endosymbiotic dinoflagellate to acute light and temperature stress and sequenced the transcriptome, to investigate stress-induced RNA editing (see section "Physiological background"). They observed base exchanges from RNA editing most prominently responsive to a heat stress treatment (Liew et al. 2017; see also Olschowsky et al. within the abstracts related to this chapter). RNA editing may induce non-synonymous substitutions which could alter protein functioning as well as stability and thereby contribute to acclimatization (see section "Physiological background").

The exact same dataset was used by Brüwer et al. (2017), who *in silico* detected a diverse viral community associated with the dinoflagellate and observed differential viral gene expression upon heat treatment. As in this example, next-generation sequencing often contains "by-catch" of host-associated microorganisms and viruses, which may be

assessed separately (see also Levin et al. 2017; Brüwer and Voolstra 2018; Brüwer and Voolstra within the abstracts related to this chapter).

Human-induced catastrophes, such as oil spills, result in immediate long-lasting changes of an ecosystem. In April 2010, the off-shore drilling platform Deepwater Horizon sank in the Gulf of Mexico, which resulted in 650 million liters of oil and gas being released into the deep sea (McNutt et al. 2012). All of the emissions combined could cover the Vatican state with an about 147 m thick layer. Kimes et al. (2013) assessed the oil spill affected deep sea-sediments, as commonly observed aerobic oil degradation may be hindered in anaerobic environments. Using a metabolomics approach, they detected an increase of benzyl succinates, a typical product of anaerobic oil biodegradation. An additional metagenomics analysis revealed an aggregation of anaerobic bacteria, in particular *Deltaproteobacteria*, in the respective sediments. This points towards an anoxic catabolism of hydrocarbons, thus suggesting a breakdown of oil (Kimes et al. 2013).

Metabolomics are frequently applied for process optimization and yield increase in bioengineering. In the field of algae-based biofuel production, nitrogen starvation of the algae *Chlamydomonas reinhardtii* has been shown to increase carbon assimilation, nitrogen metabolism, and triacylglycerol production (Park et al. 2015), triacylglycerol being the targeted metabolite.

In conclusion, the omics-toolbox provides the possibility to evaluate adaptation and acclimatization on a metaorganism scale. Due to its broad approach and the high throughput methods, rather unexpected pathways may be revealed.

Complex Systems and Multi-meta-omics

The individual omic techniques can only display a fraction of the biological complexity, as outlined above, since the measurable appearance of life (i.e., an ome) varies greatly depending on the layer accessed (e.g., genome, proteome, etc.). This constitutes, for example, in the turn-over and dynamic changes of the many intracellular components in response to the environment. Single-cell studies aim to tackle this complexity within individual cells by integrating multi-omic data (Bock et al. 2016; see Kalita et al. within the abstracts related to this chapter). On a larger scale, meta-data from many co-existing organisms are valuable to study the composition and interactions of communities.

Metagenomic and metatranscriptomic data may delineate community compositions. This information combined with a metabolic profile can reveal the ecological interactions *exempli gratia* in a microbial consortium (Freilich et al. 2011), or

depict the metabolites in an environment as a functional trait of the respective community (Llewellyn et al. 2015). From such insights, the contributions of species to an ecosystem's functioning and productivity can be deduced. Teeling et al. (2012) collected samples of the North Sea twice a week over the course of a year for multi-omics analysis. The succession of different algal substrate degrading bacterial communities responded to occurring phytoplankton blooms. They concluded that high bacterioplankton diversity in a relatively homogeneous habitat as 'the ocean' may result from temporally distributed niches (Teeling et al. 2012).

Comparisons of species in a community, which are very different for example in size, trophic level, or lifestyle, often lack precision. Although two taxa may share an ecological feature, their relative abundances may be very different in the size of their impact on the environment (e.g., sea-weed grazing of gastropoda and dugongs). And even though two taxa may perform a comparable task, their ecological function can be entirely different (e.g., free-living and endosymbiotic dinoflagellates). Multi-omics can overcome these discrepancies, as it can provide simultaneous information on multiple layers (see Leary et al. 2014; Moran 2015; Thiele et al. 2017 for marine microbiome).

Conclusion

The rapid development of new techniques, software, and decreasing costs for high throughput methods allow unprecedented experimental designs. The rising field of omic studies provides a toolbox for a better understanding of the complexity of life on earth. It is now possible to characterize organisms on a genome-wide scale. Omics have revealed a diversity of previously undetected species and simplify quantitative and functional analyses. However, the transition between different ome-layers is highly variable and requires the integration of multiple omic approaches. This can improve our understanding of the link between the genotype and phenotype. With multi-omics, we could find out how complex biological networks function.

Appendix

This article is related to the YOUMARES 8 conference session no. 2: "Reading the Book of Life – Omics as a Universal Tool Across Disciplines". The original Call for Abstracts and the abstracts of the presentations within this session can be found in the appendix "Conference Sessions and Abstracts", chapter "5 Reading the Book of Life – Omics as a Universal Tool Across Disciplines", of this book.

References

Alberts B, Johnson A, Lewis J et al (2008) Molecular biology of the cell. Garland Science, New York

Almeida MI, Reis RM, Calin GA (2011) MicroRNA history: discovery, recent applications, and next frontiers. Mutat Res – Fundam Mol Mech Mutagen 717:1–8

Amann R, Rosselló-Móra R (2016) After all, only millions? MBio 7(4):e00999–16

Andrews KR, Good JM, Miller MR et al (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. Nat Rev Genet 17:81–92

Avery OT, Macleod CM, McCarty M (1943) Studies on the chemical nature of the substance inducing transformation of pneumococcal types: induction of transformation by a Desoxyribonucleic acid fraction isolated from Pneumococcus type III. J Exp Med 79:137–158

Baumgarten S, Simakov O, Esherick LY et al (2015) The genome of *Aiptasia*, a sea anemone model for coral symbiosis. Proc Natl Acad Sci USA 112:11893–11898

Beale DJ, Karpe AV, Ahmed W (2016) Beyond metabolomics: a review of multi-omics-based approaches. In: Beale DJ, Kouremenos KA, Palombo A (eds) Microbial metabolomics. Springer, Cham, pp 289–312

Biagi E, Candela M, Fairweather-Tait S et al (2012) Ageing of the human metaorganism: the microbial counterpart. Age 34:247–267

Bock C, Farlik M, Sheffield NC (2016) Multi-omics of single cells: strategies and applications. Trends Biotechnol 34:605–608

Bosch TCG, Miller DJ (2016) The Holobiont imperative. Springer, Vienna

Brüwer JD, Voolstra CR (2018) First insight into the viral community of the cnidarian model metaorganism Aiptasia using RNA-Seq data. PeerJ 6:e4449

Brüwer JD, Agrawal S, Liew YJ et al (2017) Association of coral algal symbionts with a diverse viral community responsive to heat shock. BMC Microbiol 17:174

Buck-Wiese H, Voolstra CR, Brüwer JD (2016) The metaorganism frontier – incorporating microbes into the organism's response to environmental change. In: Bode M, Jessen C, Golz V (eds) YOUMARES 7 conference proceedings, pp 94–102

Cherf GM, Lieberman KR, Rashid H et al (2012) Automated forward and reverse ratcheting of DNA in a nanopore at 5-A precision. Nat Biotechnol 30:344–348

Dahm R (2005) Friedrich Miescher and the discovery of DNA. Dev Biol 278:274–288

Dalsgaard J, John MS, Kattner G et al (2003) Fatty acid trophic markers in the pelagic marine environment. Adv Mar Biol 46:225–340

Delorme-Axford E, Donker RB, Mouillet J-F et al (2013) Human placental trophoblasts confer viral resistance to recipient cells. Proc Natl Acad Sci U S A 110:12048–12053

Dewar ML, Arnould JPY, Dann P et al (2013) Interspecific variations in the gastrointestinal microbiota in penguins. MicrobiologyOpen 2:195–204

Dewar ML, Arnould JPY, Krause L et al (2014) Influence of fasting during moult on the faecal microbiota of penguins. PLoS One 9:e99996

Dick GJ (2017) Embracing the mantra of modellers and synthesizing omics, experiments and models. Environ Microbiol Rep 9:18–20

Epstein SS (2013) The phenomenon of microbial uncultivability. Curr Opin Microbiol 16:636–642

Fauquet CM, Mayo MA (2001) The 7th ICTV report. Arch Virol 146:189–194

Feng Y, Zhang Y, Ying C et al (2015) Nanopore-based fourth-generation DNA sequencing technology. Genomics Proteomics Bioinf 13:4–16

Freilich S, Zarecki R, Eilam O et al (2011) Competitive and cooperative metabolic interactions in bacterial communities. Nat Commun 2:589

Gaby JC, Buckley DH (2012) A comprehensive evaluation of PCR primers to amplify the *nifH* gene of nitrogenase. PLoS One 7:e42149

Geiss GK, Bumgarner RE, Birditt B et al (2008) Direct multiplexed measurement of gene expression with color-coded probe pairs. Nat Biotechnol 26:317–325

Gillooly JF, Allen AP, West GB et al (2005) The rate of DNA evolution: effects of body size and temperature on the molecular clock. Proc Natl Acad Sci U S A 102:140–145

Gottlieb A (2017) Untersuchungen zur Expression ausgewählter Stammzellgene an embryonalen Geweben und Tumoren. Dissertation, Universität Bremen

Graeve M, Kattner G, Hagen W (1994) Diet-induced changes in the fatty acid composition of Arctic herbivorous copepods: experimental evidence of trophic markers. J Exp Mar Biol Ecol 182:97–110

Handelsman J (2004) Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev 68:669–685

Hernandez-Agreda A, Leggat W, Bongaerts P, Ainsworth D (2016) The microbial signature provides insight into the mechanistic basis of coral success across reef habitats. MBio 7:e00560–e00516

IPCC (2007) Climate change 2007: The physical science basis. Summary for policymakers. Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge

IPCC (2014) Climate change 2014: Synthesis report. Contribution of working groups I, II and III to the fifth assessment report of the intergovernmental panel on climate change. IPCC, Geneva

Jones OAH, Sdepanian S, Lofts S et al (2014) Metabolomic analysis of soil communities can be used for pollution assessment. Environ Toxicol Chem 33:61–64

Kanehisa M, Sato Y, Kawashima M et al (2016) KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res 44:D457–D462

Kanehisa M, Furumichi M, Tanabe M et al (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45:D353–D361

Kim BM, Kim J, Choi IY et al (2016) Omics of the marine medaka (*Oryzias melastigma*) and its relevance to marine environmental research. Mar Environ Res 113:141–152

Kimes NE, Callaghan AV, Aktas DF et al (2013) Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. Front Microbiol 4:50

Klug WS, Spencer CA, Palladino MA (2012) Concepts of genetics, 10th edn. Pearson Education, San Francisco

Leary DH, Li RW, Hamdan LJ et al (2014) Integrated metagenomic and metaproteomic analyses of marine biofilm communities. Biofouling 30:1211–1223

Levin RA, Voolstra CR, Weynberg KD et al (2017) Evidence for a role of viruses in the thermal sensitivity of coral photosymbionts. ISME J 11:808–812

Liew YJ, Li Y, Baumgarten S et al (2017) Condition-specific RNA editing in the coral symbiont *Symbiodinium microadriaticum*. PLoS Genet 13:1–22

Liu Y, Beyer A, Aebersold R (2016) On the dependency of cellular protein levels on mRNA abundance. Cell 165:535–550

Llewellyn CA, Sommer U, Dupont CL et al (2015) Using community metabolomics as a new approach to discriminate marine microbial particulate organic matter in the western English Channel. Prog Oceanogr 137:421–433

Lu H, Giordano F, Ning Z (2016) Oxford Nanopore MinION sequencing and genome assembly. Genomics Proteomics Bioinf 14:265–279

Luo S-S, Ishibashi O, Ishikawa G et al (2009) Human villous trophoblasts express and secrete placenta-specific microRNAs into maternal circulation via exosomes. Biol Reprod 81:717–729

McFall-Ngai M, Hadfield MG, Bosch TCG et al (2013) Animals in a bacterial world, a new imperative for the life sciences. Proc Natl Acad Sci U S A 110:3229–3236

McNutt MK, Camilli R, Crone TJ et al (2012) Review of flow rate estimates of the Deepwater Horizon oil spill. Proc Natl Acad Sci U S A 109:20260–20267

Moran MA (2015) The global ocean microbiome. Science 347:aac8455

Nirenberg M (2004) Historical review: deciphering the genetic code – a personal account. Trends Biochem Sci 29:46–54

Noguera-Solano R, Ruiz-Gutierrez R, Rodriguez-Caso JM (2013) Genome: twisting stories with DNA. Endeavour 37:213–219

O'Leary NA, Wright MW, Brister JR et al (2016) Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. Nucleic Acids Res 44:D733–D745

Olsen JL, Rouzé P, Verhelst B et al (2016) The genome of the seagrass *Zostera marina* reveals angiosperm adaptation to the sea. Nature 530:331–335

Paez-Espino D, Eloe-Fadrosh EA, Pavlopoulos GA et al (2016) Uncovering Earth's virome. Nature 536:425–430

Park JJ, Wang H, Gargouri M et al (2015) The response of *Chlamydomonas reinhardtii* to nitrogen deprivation: a systems biology analysis. Plant J 81:611–624

Pascault N, Loux V, Derozier S et al (2015) Technical challenges in metatranscriptomic studies applied to the bacterial communities of freshwater ecosystems. Genetica 143:157–167

Patti GJ, Yanes O, Siuzdak G (2012) Metabolomics: the apogee of the omics trilogy. Int J Pharm Pharm Sci 13:263–269

Pedrós-Alió C (2012) The rare bacterial biosphere. Annu Rev Mar Sci 4:449–466

Pheasant M, Mattick JS (2007) Raising the estimate of functional human sequences. Genome Res 17:1245–1253

Pimm SL, Jenkins CN, Abell R et al (2014) The biodiversity of species and their rates of extinction, distribution, and protection. Science 344:1246752

Pogoreutz C, Rädecker N, Cárdenas A et al (2017) Nitrogen fixation aligns with *nifH* abundance and expression in two coral trophic functional groups. Front Microbiol 8:1187

Rhoads A, Au KF (2015) PacBio sequencing and its applications. Genomics Proteomics Bioinformatics 13:278–289

Ritchie KB (2006) Regulation of microbial populations by coral surface mucus and mucus associated bacteria. Mar Ecol Prog Ser 322:1–14

Rohwer F, Seguritan V, Azam F et al (2002) Diversity and distribution of coral-associated bacteria. Mar Ecol Prog Ser 243:1–10

Röthig T, Costa RM, Simona F et al (2016) Distinct bacterial communities associated with the coral model *Aiptasia* in Aposymbiotic and symbiotic states with *Symbiodinium*. Front Mar Sci 3:234

Ruggles KV, Krug K, Wang X et al (2017) Methods, tools and current perspectives in proteogenomics. Mol Cell Proteomics 16:959–981

Sampson TR, Mazmanian SK (2015) Control of brain development, function, and behavior by the microbiome. Cell Host Microbe 17:565–576

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A 74:5463–5467

Schuster SC (2008) Next-generation sequencing transforms today's biology. Nat Methods 5:16–18

Selbach M, Schwanhäusser B, Thierfelder N et al (2008) Widespread changes in protein synthesis induced by microRNAs. Nature 455:58–63

Shnit-Orland M, Sivan A, Kushmaro A (2012) Antibacterial activity of *Pseudoalteromonas* in the coral Holobiont. Microb Ecol 64:851–859

Simmonds P, Adams MJ, Benkö M et al (2017) Consensus statement: virus taxonomy in the age of metagenomics. Nat Rev Microbiol 15:161–168

Stackebrandt E, Frederiksen W, Garrity GM et al (2002) Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. Int J Syst Evol Microbiol 52:1043–1047

Taberlet P, Coissac E, Hajibabaei M et al (2012) Environmental DNA. Mol Ecol 21:1789–1793

Teeling H, Fuchs BM, Becher D et al (2012) Substrate-controlled succesion of marine Bacterioplankton populations induced by a phytoplankton bloom. Science 336:608–611

The Gene Ontology Consortium (2004) The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 32:D258–D261

The Gene Ontology Consortium (2017) Expansion of the gene ontology knowledgebase and resources: The gene ontology consortium. Nucleic Acids Res 45:D331–D338

Thiele S, Richter M, Balestra C et al (2017) Taxonomic and functional diversity of a coastal planktonic bacterial community in a river-influenced marine area. Mar Genomics 32:61–69

Turnbaugh PJ, Ley RE, Hamady M et al (2007) The human microbiome project. Nature 449:804–810

Valentini A, Taberlet P, Miaud C et al (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Mol Ecol 25:929–942

Voolstra CR, Miller DJ, Ragan MA et al (2015) The ReFuGe 2020 consortium—using "omics" approaches to explore the adaptability and resilience of coral holobionts to environmental change. Front Mar Sci 2:68

Wilmes P, Simmons SL, Denef VJ et al (2009) The dynamic genetic repertoire of microbial communities. FEMS Microbiol Rev 33:109–132

Zanta MA, Belguise-Valladier P, Behr JP (1999) Gene delivery: a single nuclear localization signal peptide is sufficient to carry DNA to the cell nucleus. Proc Natl Acad Sci U S A 96:91–96

Ziegler M, Seneca FO, Yum LK et al (2017) Bacterial community dynamics are linked to patterns of coral heat tolerance. Nat Commun 8:14213