Review

Bioinformatics for Marine Products: An Overview of Resources, Bottlenecks, and Perspectives

Luca Ambrosino [1], **Michael Tangherlini** [1], **Chiara Colantuono** [1], **Alfonso Esposito** [2], **Mara Sangiovanni** [1], **Marco Miralto** [1], **Clementina Sansone** [3] and **Maria Luisa Chiusano** [1,4,*]

[1] Department of Research Infrastructures for Marine Biological Resources, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy; luca.ambrosino@szn.it (L.A.); michael.tangherlini@szn.it (M.T.); chiara.colantuono@szn.it (C.C.); mara.sangiovanni@szn.it (M.S.); marco.miralto@szn.it (M.M.)

[2] Department of Cellular, Computational and Integrative Biology - CIBIO, University of Trento, 38123 Povo, Trento, Italy; alfonso.esposito@unitn.it

[3] Department of Marine Biotechnology, Stazione Zoologica Anton Dohrn, Villa Comunale, 80121 Naples, Italy; clementina.sansone@szn.it

[4] Department of Agriculture, University of Naples Federico II, Portici, 80055 Naples, Italy

* Correspondence: chiusano@unina.it; Tel.: +39-(0)812539492

check for updates

Abstract: The sea represents a major source of biodiversity. It exhibits many different ecosystems in a huge variety of environmental conditions where marine organisms have evolved with extensive diversification of structures and functions, making the marine environment a treasure trove of molecules with potential for biotechnological applications and innovation in many different areas. Rapid progress of the omics sciences has revealed novel opportunities to advance the knowledge of biological systems, paving the way for an unprecedented revolution in the field and expanding marine research from model organisms to an increasing number of marine species. Multi-level approaches based on molecular investigations at genomic, metagenomic, transcriptomic, metatranscriptomic, proteomic, and metabolomic levels are essential to discover marine resources and further explore key molecular processes involved in their production and action. As a consequence, omics approaches, accompanied by the associated bioinformatic resources and computational tools for molecular analyses and modeling, are boosting the rapid advancement of biotechnologies. In this review, we provide an overview of the most relevant bioinformatic resources and major approaches, highlighting perspectives and bottlenecks for an appropriate exploitation of these opportunities for biotechnology applications from marine resources.

Keywords: bioinformatics; omics; marine resources; biotechnological applications; marine observatories

1. Introduction

The origin of life has been traced from the sea about 1.5 billion years before the evolution of mankind. Since then, marine organisms have diversified in structure and functions, making the marine environment the largest and most variable ecosystem on Earth, comprising more than 70% of the planet surface and adapting to a wide range of conditions, from the extreme cold of polar seas to the extreme high temperatures and pressures of deep-sea hydrothermal vents [1].

The first living organisms appeared in the sea more than 3.5 billion years ago [2,3] and the evolutionary processes have molded marine organisms, which range from viruses to eukaryotes, to survive extreme temperatures, variable salinity and pressure, and attacks by other species, including prokaryotic and viral invaders [4–13].

The adaptation to a variety of conditions featured by extremely different marine environments determines an enormous amount of genetic and functional diversity [14], offering a precious source

of biological materials and molecules which are contributing to innovation in many fields [1], including medicine and pharmacology [15], nutrition [16–18], agriculture [18–21], biofuels [22–24], cosmetics [25–27], innovations for sustainability (e.g., bioremediation [28] and bioplastics [29]), and other industrial sectors. As other examples, the marine microbiota appears to be a promising and endless source for new drug development [30–32], with new chemotherapeutants, novel antibiotics and health products to prevent and combat diseases [15], cancer [17,33,34], and drug-resistant pathogens [35], which are becoming a significant threat to public health. In health sciences, many marine natural products were revealed to be toxins or bioactive compounds, and were deeply studied to understand their action [15,17,33,36] and possible applications. In food sciences and agriculture, the marine environment has always been a gold mine [16–21], even when exploited as by-products or waste materials [19,20]. Marine products such as the algae-derived polysaccharides (e.g., agar), which have been used in food processing and preservation since the first half of the last century [21,37], are now widely used in nutrition but also for the delivery of bioactive compounds and nutraceuticals [38], or even for innovative opportunities (e.g., to produce degradable bioplastics [29]) for sustainable products.

Nevertheless, the marine habitat is still poorly explored. It is estimated that, despite 250 years of taxonomic classification and over 1.2 million species already catalogued in reference databases such as the World Register of Marine Species [39–42], 91% of species in the ocean still await description [43].

One of the reasons for the expanding interest in tools and approaches for observing and exploring the marine environment is to identify novel molecular entities as sources for new compounds for innovation in health, nutrition, agriculture, care, goods, and energetics.

Today, about 7000 molecules extracted from the sea are already used or are being validated for several purposes, ranging from medicine to industrial applications. The number of compounds isolated from marine species increases annually by almost 400 to 500 newly discovered products, and many more are still to be discovered [44]. Such compounds can be natively produced as secondary metabolites and become part of the organisms or secreted in the extracellular milieu [45]. Bioactive compounds can either be polypeptides or small molecules (lipopolysaccharides, polyphenols, alkaloids, etc.), but also nonribosomal peptides (e.g., vancomycin or daptomycin, actinomycin D, and cyclosporine) [46], polyketides [47], and nucleic acids [48–50]. However, the number of approved and marketed marine natural products is still very limited (11 approved drugs, five of which have anticancer activity, and more than 20 other natural products in clinical phase, as of 2018 [51]).

The trend to discovery novel products is focused mainly on the study of target species that are useful for the isolation of new active compounds, following well-established step-by-step approaches. Companies and enterprises are, therefore, strongly investing in all the methodologies that show potential to effectively shorten the pipelines for drug identification and development [52,53]. Research has focused on opportunities to generate more innovation in a short time and, as a consequence, is resulting in a remarkable interest in bioinformatics [54]. Indeed, bioinformatics offers methodologies for efficiently extracting value-added information from omics experimental data and for modeling, and these tools are useful not only to accelerate the identification of biologically active candidates, but also to investigate their action and effects on other living systems, such as on specific species or particular ecosystems (for example, in pharmacogenomics [55], microbial ecology, and agriculture [56,57]).

Since the capacity for producing bioactive compounds is encoded in the genome of a species, the identification of novel compounds and the molecular mechanisms for their synthesis often starts from genome or transcriptome sequencing, which takes advantage of the flourish of advanced methodologies which have been revolutionized by the advent of the next-generation sequencing (NGS) technologies, in the framework of the "isolate and then test" instead of "test and then isolate" approach [58]. Marine biotechnology, and biology in general, have largely profited in recent years from the advent of cost effective NGS, therefore, expanding the sequencing projects, which has resulted in major advances in the field [59], as well as other omics approaches (e.g., proteomics and metabolomics) [60,61], all supporting the understanding of structure and functionality of molecules. After the identification of novel compounds and possibly the elucidation of the metabolic pathways

leading to these products, the organisms can be isolated and further investigated, or only the genes encoding the involved compounds can be isolated and then expressed in heterologous hosts, for controlled production. Alternatively, heterologous expression of biosynthetic genes or gene clusters (identified, for example, through metagenomics libraries) also produces compounds derived from yet-to-be isolated microorganisms [62] with cost effective techniques. On the one hand, isolated species, or the recombinant species, can be grown under controlled conditions (for example, in bioreactors) in order to obtain large amounts of the target compounds without harvesting the original wild population (which in some cases could lead to ecosystems unbalance) or using synthetic production that is often more expensive, thus improving sustainability of the productivity chain [25,63–65]. On the other hand, only a fraction of the marine diversity can be aptly cultured in a laboratory, and therefore this approach must be complemented by alternative techniques in order to explore a larger portion of this diversity. Indeed, in particular, molecular techniques and advanced sequencing approaches have been used for this purpose to capture the genetic and genomic diversity of the unculturable fraction of marine biological diversity (especially regarding prokaryotes).

The first massive exploration of marine diversity using a molecular approach resulted from the Craig Venter's Global Ocean Sampling expedition. This scientific endeavor consisted of a worldwide voyage, inspired by Darwin's voyage on the "Beagle", undertaken to sample marine organisms and assess their diversity through DNA sequencing [66,67]. By analyzing mainly nucleic acid sequenced data and building in silico models, in fact, novel species were identified and the biosynthetic production of fluxes of compounds were inferred, understanding the definition of pathways of interest, and, eventually, redesigned [68]. For example, by means of advanced bioinformatic pipelines it is possible to identify thousands of possible biosynthetic gene clusters along DNA sequences, which can be explored and investigated by computational analyses before experimental characterization [69].

The ability to perform genome and gene data mining as an essential complementary approach to traditional experimental methods has significantly sped up the process of natural compounds discovery [70]. A closer look at the scientific production related to natural compounds obtained from marine species is found in the MarinLit database (Table 1).

In addition, recent advances have strongly improved the biotechnological tools to enhance and manipulate the production of natural molecules. As an example, advanced experimental techniques such as those related to genome editing have been recently introduced and are being widely exploited to directly modify the genome sequence in regions of interest [71]. As a main novel example, approaches that take advantage of the CRISPR/Cas9 machinery [72] are being adopted to generate mutant genomes that target specific genes. Additionally, for these approaches, bioinformatics provides tools able to predict CRISPR/Cas9 targets, even in novel or partial genome sequences [73–75].

In this review, we focus on the main resources of bioinformatics and methodologies, discussing their role in supporting and accelerating the discovery of new marine-derived products, describing major applications, and highlighting opportunities, bottlenecks, challenges, and perspectives in the field.

Table 1. General or marine-specific reference resources/repositories per section, listed in alphabetical order.

Name	Section	Website
Scientific literature		
MarinLit	Marine natural products literature	http://pubs.rsc.org/marinlit/
Genomics and Transcriptomics		
AmiGO	GO functional annotation repository and analyses services	http://amigo.geneontology.org/amigo
Aniseed	Genome browser and multi-omics repository for Ascidiacea	https://www.aniseed.cnrs.fr/aniseed/
ArrayExpress	Next-generation-sequencing (NGS) data repository	https://www.ebi.ac.uk/arrayexpress/
BLAST	Local alignment versus sequence database service	https://blast.ncbi.nlm.nih.gov/Blast.cgi
CCTop	CRISPR/Cas9 target prediction tool	https://crispr.cos.uni-heidelberg.de/
CHOPCHOP	CRISPR/Cas9 and TALEN target Prediction Tool	http://chopchop.cbu.uib.no/
dbEST	Expressed sequence tag (EST) sequence repository	https://www.ncbi.nlm.nih.gov/nucleotide/
DDBJ	General multi-omics repository and analyses services	https://www.ddbj.nig.ac.jp/index-e.html
DRA	General NGS data repository	https://www.ddbj.nig.ac.jp/dra/index-e.html
Echinobase	Genome browser and multi-omics repository for Echinoderms	http://www.echinobase.org/Echinobase/
Ensembl	General multi-omics repository and analyses services	https://www.ensembl.org/
Gene Ontology	GO functional annotation repository and analyses services	http://geneontology.org/
IMG/ER	Prokaryotic sequence and function repository	https://img.jgi.doe.gov/cgi-bin/mer/main.cgi
JGI	Multi-omics repository and analyses services	https://jgi.doe.gov/
KEGG Genome	Genome sequence repository	https://www.genome.jp/kegg/genome.html
LAST	Long sequence alignment service	http://last.cbrc.jp/
Mauve	Genome alignment via homolog blocks detection	http://darlinglab.org/mauve/
MicroPan	Bacterial pangenome analysis library for R environment	https://cran.r-project.org/web/packages/micropan/index.html
NCBI	General multi-omics repository and analyses services	https://www.ncbi.nlm.nih.gov/
OIST MGU	Genome browser and analyses services for 19 marine species	https://marinegenomics.oist.jp/
PanFP	Bacterial pangenome-based functional profiles	https://github.com/srjun/PanFP
PGAWeb	Bacterial pangenome analyses service	http://pgaweb.vlcc.cn
ProGenomes	Prokaryotic sequence and functional repository	http://progenomes.embl.de/
SRA	General NGS data repository	https://www.ncbi.nlm.nih.gov/sra
Metagenomics and metatranscriptomics		
dbCAN	Automated carbohydrate-active enzyme annotation	http://bcb.unl.edu/dbCAN2/
EBI Metagenomics	Microbiome sequence repository and analyses services	https://www.ebi.ac.uk/metagenomics/
Geotraces	Marine key trace elements and isotopes data repository	http://www.geotraces.org/
GLOSSary	Marine microbial sequence repository and analyses services	https://bioinfo.szn.it/glossary/
KEGG MGENES	Annotated environmental gene catalog and analyses service	https://www.genome.jp/mgenes
Marine Metagenomics Portal	Marine microbiome repository and analyses services	https://mmp.sfb.uit.no/
MG-RAST	Phylogenetic and functional analysis for metagenomics	https://www.mg-rast.org/
Ocean Gene Atlas	Analytical service for marine planktonic organisms	http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/

Table 1. *Cont.*

Name	Section	Website
Tara Oceans Database	Expedition specific raw reads sequence repository	https://www.ebi.ac.uk/services/tara-oceans-data
Proteomics and structural biology		
AMBER	Molecular dynamics simulation program	http://ambermd.org/
AutoDock	Molecular docking program	http://autodock.scripps.edu/
AutoDock Vina	Multithreading program for molecular docking	http://vina.scripps.edu
CHARMM	Molecular dynamics simulations program	https://www.charmm.org/charmm/
Desmond	Molecular dynamics simulations server	https://www.schrodinger.com/desmond
DOCK	Molecular docking server	http://dock.compbio.ucsf.edu/
FlexX	Molecular docking server	https://www.biosolveit.de/FlexX/
Glide	Molecular docking server	https://www.schrodinger.com/glide
GOLD	Molecular docking program	https://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/
GROMACS	Molecular dynamics simulations program	http://www.gromacs.org
HHpred	Homology modelling server	https://toolkit.tuebingen.mpg.de/#/tools/hhpred
I-TASSER	Ab-initio structure prediction server	https://zhanglab.ccmb.med.umich.edu/I-TASSER/
ICM	Molecular docking program	http://www.molsoft.com/docking.html
InterPro	Protein function repository and analytical services	https://www.ebi.ac.uk/interpro/
LeDock	Molecular docking program	http://www.lephar.com/download.htm
Modeller	Homology modelling program	https://salilab.org/modeller/
MOE-Dock	Molecular docking server	https://www.chemcomp.com/index.htm
NAMD	Molecular dynamics simulations program	http://www.ks.uiuc.edu/Research/namd/
OpenMM	Molecular dynamics simulations program	http://openmm.org/
PDB	Protein structure repository	https://www.rcsb.org/
PFAM	Protein family repository	https://pfam.xfam.org/
Phyre2	Threading and ab-initio structure prediction server	http://www.sbg.bio.ic.ac.uk/~{}phyre2/html/page.cgi?id=index
RaptorX	Homology modelling and threading structure prediction server	http://raptorx.uchicago.edu
rDock	Molecular docking program	http://rdock.sourceforge.net/
Robetta	Homology modelling and ab-initio structure prediction server	http://www.robetta.org/
Surflex	Molecular docking program	http://www.jainlab.org/downloads.html
Swiss-model	Homology modelling server	https://swissmodel.expasy.org
SwissDock	Molecular docking server	http://www.swissdock.ch
UniProt	Protein sequence and function repository	https://www.uniprot.org/
Metabolomics		
Anti-smash	Annotation and analysis of secondary metabolite biosynthesis	https://antismash.secondarymetabolites.org/#!/start
ChemSpider	Compound repository	http://www.chemspider.com/
GNPS	Tandem mass (MS/MS) spectrometry data repository	https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash.jsp
KEGG	Metabolism data repository and analyses services	https://www.genome.jp/kegg/
MEROPS	Compound repository and analyses services	https://www.ebi.ac.uk/merops/
MetaCyc	Metabolism data repository and analyses services	https://metacyc.org/
NaPDoS	Compound repository and analyses services	https://www.biokepler.org/use_cases/napdos
Reactome	Metabolism data repository and analyses services	https://reactome.org/
The Super Natural II database	Compound repository	http://bioinf-applied.charite.de/supernatural_new/index.php

2. Bioinformatics Applications and Resources in Marine Omics

A number of different approaches can be used to explore novel and useful compounds from marine resources, including metabolites, enzymes, or other molecules, and to investigate the molecular mechanisms involved in their production and functional properties. These methodologies range from whole-genome sequencing (a stand-alone research line which also provides a reference baseline for further omics approaches such as transcriptomics and proteomics used to investigate the functional activity of species or tissues) to metabolomics, used to understand the phenotypical effects of genome expression. Their meta-omics counterparts (e.g., metagenomics and metatranscriptomics) are able to tackle similar issues at the community level.

The following sections provide an overview of the main bioinformatic methodologies, with details on possible applications in marine biotechnology. A discussion of general topics that are not specific to marine biology, such as genomic or transcriptomic assembly and annotation, is beyond the scope of this paper, however interested readers can find pointers to relevant literature in the related sections.

2.1. Genomics and Transcriptomics

The advent of novel technologies, such as the introduction of next generation sequencing (NGS) techniques, favored the shift from the less efficient Sanger methodology to the sequencing of huge numbers of DNA fragments, due to fast and cheaper high-throughput technologies. The BAC-by-BAC (Bacterial Artificial Chromosome) based genome sequencing was almost replaced by the whole genome shotgun (WGS) approach [76]. This transition increased the need for new methods of data processing, mining, and management and further challenged bioinformatic research to provide advanced technologies to support the sequencing efforts [77]. This change resulted in the establishment of several genome-sequencing projects, expanding the activities that were mainly focused on reference model species for marine biology, such as *Ciona robusta* or *Strongylocentrotus purpuratus* [78,79], to other species, with the release of many draft genomes obtained from the sequencing of new species and the resequencing and genotyping of already available genomes [80–84] (Figure 1). Resequencing efforts were, in some cases, necessary to improve the poor quality genomes obtained with older or inadequate technologies, which were useful to detect candidate new compounds but not for comparative genomic analyses.
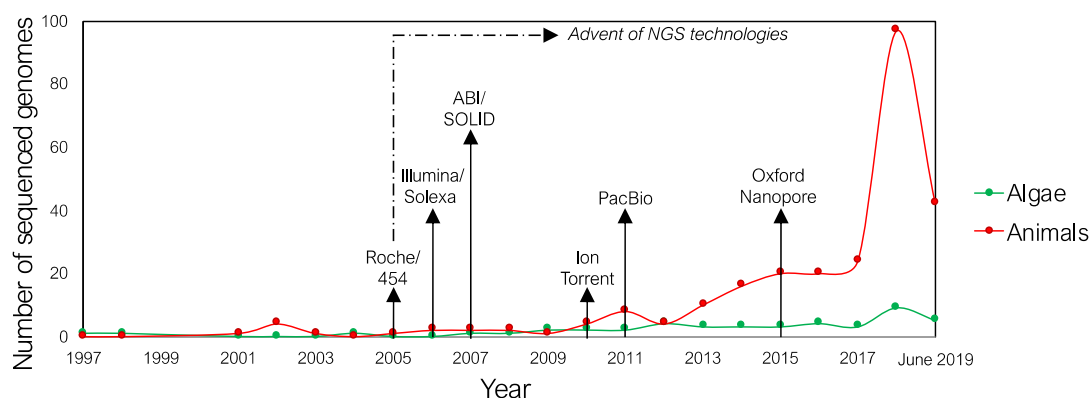

Figure 1. Number of sequenced genomes per year since 1997 until June 2019. Sequenced marine algae (green) and animals (red) genomes are shown. The years of the advent of next-generation sequencing (NGS) technologies as well as of the launches of the principal platforms on the market are indicated.

Under the umbrella of the International Nucleotide Sequence Database Collaboration (INSDC) [85], all the information related to biological sequences, including those from marine resources, is flowing into general databases (Table 1). The Reference sequence database at NCBI [86–88]; the EMBL-EBI sequence collection, including the vertebrates, prokaryotes, protists, fungi, plants, and metazoan partitions [89,90]; and the DNA Data Bank of Japan (DDBJ) [91] are the three reference sites in the

consortium. In addition, due to the generation and release of huge numbers of sequences (raw reads) produced and released by next-generation sequencing efforts, the INSDC system built specialized archives to store data either as raw or processed, such as the Sequence Read Archive (SRA) [92] and Gene Expression Omnibus (GEO) [93] at NCBI, ArrayExpress [94] and European Nucleotide Archive (ENA) [95] at EMBL-EBI, and the DDBJ sequence Read Archive (DRA) [96].

Beyond the INSDC project, the Integrated Microbial Genomes with Microbiome Samples (IMG/M) and Expert Review (IMG/ER) [97] at the Joint Genome Institute (JGI) and proGenomes [98] are parallel efforts organizing reference resources for microbial genomes and microbiomes, providing information on genes, genomes, and functions, and providing tools for comparative analyses. In more detail, the IMG/M and IMG/ER partitions at the JGI provide highly specialized repositories for curated microbial, viral, and fungal genomes with taxonomic affiliation and specific tools for exploring their characteristics (e.g., assembly quality and completion levels, potential markers for auxotrophy, and geographic localization). The proGenomes initiative represents an attempt to provide a highly accurate prokaryotic genome database with curated taxonomic affiliations and functional annotations based on different collections, including CAZymes and dbCAN [99,100], and markers for antibiotic resistances, which represent useful references for the selection of organisms of biotechnological interest.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) genome partition within the KEGG resource [101] also acts as a reference repository for sequence data, which can be queried by users to characterize enzyme pathways and explore potential genes of biotechnological interest in complete reference genomes.

In order to annotate genomes, several resources are available to provide functional information about gene and gene clusters, beyond the gene association to pathways as provided by KEGG. One example is the Gene Ontology (GO) international consortium, which aims to provide reliable gene classification based on their functional descriptions and on the establishment of a reference vocabulary of molecular functions, cellular locations, and biological processes that gene products may be involved in [102]. As an example, one of the most used platforms for searching and browsing the Gene Ontology database is represented by AmiGO [103]. A widespread use of the GO is to perform gene sets enrichment analysis. Given a set of genes, for example those that may be expressed in specific conditions, an enrichment analysis can detect the over- or under-represented GO terms within the selected dataset as compared to a species-specific GO collection for the whole gene complement [102]. However, even using different tools to functionally annotate genomes, many genes still remain undefined. The percentage of anonymous genes can be very different among different species or taxa.

Additionally, the transcriptome sequencing shifted from the low-throughput Sanger-based expressed sequence tags (ESTs) production [104] to NGS approaches. Among these, RNA-seq produces a more detailed and quantitative overview of a transcriptome and the associated level of expression per gene, also providing—thanks to dedicated bioinformatics pipelines [105–107]—deep details on alternative splicing and allele-specific information [108], even in the absence of a reference sequenced genome (de novo transcriptome analyses). Compared to other transcriptome-based approaches such as ESTs and microarray analyses [109], the throughput of the RNA-seq techniques, together with lower experimental costs, allowed the spread of many projects that either accompanied the genome sequencing of many species to define their representative gene expression atlases [110] or independently allowed characterization of the transcriptome complement of a novel species exploiting a de novo assembly approach [111–120].

Nucleic acid sequencing techniques are also used a great deal in marine biotechnology and, in particular, to search for new marine drugs, due to the combination and integration of genomics and transcriptomic approaches that aim to find and quickly annotate genes producing interesting compounds [121,122]. Moreover, genomics and transcriptomics have been proven to be useful for the characterization of marine species that are important in the production of secondary metabolites and enzymes of interest for industrial, pharmaceutical, and green biotechnology applications [123,124]. Some recent examples of such enzymes include the new flavin-dependent halogenase, isolated from

a marine sponge metagenome [125] and several α-amylases isolated from a sea anemone microbial community [126], whereas metabolites range from derivatives of amino acids and nucleosides, macrolides, porphyrins, terpenoids to aliphatic cyclic peroxides, and sterols [127].

Transcriptomics can also help determine whether biosynthetic gene clusters are transcriptionally silent or not [128], by revealing their regulatory machinery, and possibly, the type of post-translational modification that can be amended to the proteins [129]. To this aim, we also mention some of the alternative NGS-based sequencing approaches that are being increasingly used to better support these investigations, such as small RNA, epigenome, or single cell sequencing [130–133]. We do not include specific details for associated repositories, although all the associated public sequencing production is collected in the general NGS-related resources previously mentioned in this review.

The data collections from transcriptomic projects are also all available through reference databases. In particular, ESTs libraries are stored and easily retrievable in the dbEST partition of NCBI [104] (which was included in the nucleotide section since the beginning of 2019), while raw RNA-seq data are included in SRA [92], ArrayExpress [94], and DRA databases [96].

In some cases, it is also possible to exploit both genomic and transcriptomic data through dedicated web pages that can be species-, genera-, or clade-specific. For examples of marine species, remarkable marine-specific multi-omics resources are Aniseed [134], fully dedicated to sea squirts (Ascidiacea), and Echinobase [135], which include genomes and transcriptome data of five different echinoderms, and the genome projects of the OIST Marine Genomics Unit (Table 1), which includes information concerning 19 different marine species. All these resources allow the user to access both annotated genome assemblies, as well as raw reads, and are accompanied by genome sequence browsers [89,134] to visualize structures of genes and transcripts, and, when available, to retrieve information on the encoded proteins [136].

Because sequence-like data are the major reference product in molecular biology, the development of bioinformatic methodologies has focused extensively on the design of techniques to detect sequence similarities. Most computational methods for sequence similarities are based on global or local similarity searches that are based on alignment tools [137,138]. Currently, the methodology used most to detect similarities at the nucleotide or amino acid level is the basic local alignment search tool (BLAST) [139], which compares nucleotide or protein sequences to sequence databases. Since some particular BLAST searches can be very sophisticated and involve intense computations, such as tBLASTx analyses of entire transcriptome collections that consider all six potential ORFs of each sequence, similar efforts spread. Complete genome alignments can be carried out using different approaches, from nucleotide alignments (e.g., through the LAST sequence alignment toolkit, which also can be utilized for alignment of very large mammalian genomes [140]) to block alignments (e.g., through Mauve [141]). The genome alignment tools can highlight conserved regions, rearrangements, and differences between large genome sequences, allowing the researchers to analyze peculiarities at the genomic level (e.g., sequence insertions, duplications, or potential horizontal gene transfer events). This information can be used to search for potential novelty genes or operons, reorganization, and regions encoding for the production of novel molecules of biotechnological interest.

Similarity search methods are also the basis of the approaches that focus on the definition of computationally based homologs, comparing genes or genomes based on orthology inference [142–146], analyzing gene families mainly based on the detection of computationally defined paralogs [143,147,148], and highlighting peculiarities due to the selection of those genes that are species-specific [149].

The aforementioned approaches were also adapted to detect highly related conserved portions of genomes, even in the same species. This is generally common in prokaryotes, in which genome plasticity, mosaicism, and high rates of horizontal gene transfer drove strain differentiation [150], although present also in more complex species such as plants and vertebrate that show variable levels of genome duplications [151,152]. This deluge of genomes belonging to the same taxon led to the development of the concept of pan-genomics [153], which refers to "the entire genomic repertoire of a given phylogenetic clade, encoding for all possible lifestyles carried out by its organisms."

To this extent, several different pipelines that perform these analyses have been developed over time. Some of them are available as tools for local installations (e.g., micropan for R [154] or PanFP [155]), others are instead available as webservices (e.g., PGAweb [156]). Tentative attempts at defining public databases to explore microbial pangenomes have also been devised (e.g., PanGeneHome [157]).

A comparison among several genomes of the same taxon helps researchers determine to which extent the actual genetic diversity has been sampled. For example, species such as *Bacillus anthracis* have a closed pangenome, with 2893 core genes and only 85 accessory genes after just nine individuals sequenced [158]. Conversely, species such as *Pseudomonas aeruginosa* have a relatively small core genome as compared with the large accessory genome (665 genes constituting the 1% of the whole pan-genome), and therefore its pangenome is defined as "open", meaning that its diversity is still not sampled thoroughly [159]. From a biotechnological point of view, pangenomic analyses can highlight whether genes or gene clusters of biotechnological interest can be found in specific strains of well-known organisms, and thus potentially introduced in novel screening approaches or easily transferred to well-characterized organisms for their high-throughput production [160,161]. Indeed, this approach has been applied before to identify, clone, and express candidate antibiotic resistance genes in the *Salinispora* genus by screening more than 80 strains. This approach also correctly identified previously "orphaned" gene clusters, for which function could not be assigned, by inspecting the function of their orthologs and by analyzing their products through heterologous expression in suitable hosts [162].

Due to the diffusion of NGS technologies and the constant decrease in the sequencing costs a plethora of genomics and transcriptomic datasets have been generated. Although associated with the same species, they often exhibit high dissimilarities in terms of data quality, curation, and methodologies employed (e.g., same species with different genome annotation versions). This is due to several factors which include the following: (1) data production is several order of magnitude faster than the release of exhaustively annotated and curated datasets; (2) the opportunity to publicly release even partial and still uncomplete data [163], which can be of interest for the scientific research, but often remain in a preliminary version; (3) the run for releasing dedicated resources for specific targets, which often causes uncoordinated parallel efforts, resulting in similar resources covering partially overlapping information; (4) the lack of rules for the withdrawal of obsolete public collections; and (5) the presence of software errors, such as in automatic annotation pipelines, that might generate errors not easily detected and that are, if not curated, inherited in subsequent versions. Navigating this overwhelming amount of resources can cause confusion for non-expert users and lead to limited scientific applications, and thus determining one of the major bottlenecks in bioinformatics [163]. Examples of heterogeneity in terms of data content are evident even in reference platforms such as NCBI and Ensembl [87,89]. As an example, for the echinoderms reference species, i.e., *Strongylocentrotus purpuratus* (purple sea urchin), the latest genome assembly versions in NCBI and in Ensembl are different (Spur. version 4.2 and 3.1, respectively), misleading the users and affecting the reproducibility and the comparability of the generated results outside of the used platform. Moreover, as in the case of the diatom *Phaeodactylum tricornutum*, although the latest available genome assembly version is the same in both NCBI and Ensembl, the genome annotation versions refer to different analytical pipelines, the NCBI genome annotation pipeline [164], and the Ensembl gene annotation approach [165], respectively. Indeed, annotation pipelines have different sensitivity for determining gene structures and predicting CDSs, giving results that are resource dependent and do not necessarily fully overlap.

The sources of heterogeneity, in terms of genome assemblies and gene annotation versions, and the quality of the annotation, are severe limiting factors for the sharing of comparable results from public web-based services, which affect the reliability of the available information resources and subsequent results such as gene expression, gene family analysis, and comparative genomics. These issues are worsened when considering prokaryotic genome annotation. Thousands of prokaryotic genomes are released yearly and annotated with automatic tools, whose accuracy did not improve

accordingly [166]. Actually, the sources of biases are amplified as an effect of draft annotations and a solution to this problem has not been found yet [166].

All these aspects may also mislead non-expert users, who need education in the field in order to appropriately move through the overwhelming amount of resources. To mitigate this issue, the straightforward direction should be that reference websites (e.g., NBCI and Ensembl) should share, cross-reference, and integrate more information, even coming from smaller consortia efforts, and clearly report updates and errors, if any. A complementary approach could be provided by experts in the field that could produce smaller but effective resources by releasing manually curated information on selected species. As an example, the GENOMA platform [167], an ongoing project, at the moment collects and integrates genomic information about four different marine species, reporting statistics and comparisons among several genomics resources, also through dedicated genome browsers.

2.2. Metagenomics and Metatranscriptomics

To date, it has been established that a very minimal fraction of all the microorganisms inhabiting marine environments can be successfully isolated through traditional culturing methods, and that the vast majority of the potential functional diversity in the ocean is currently not exploited [168,169]. However, the emergence of culture-independent techniques coupled with metagenomic approaches has provided researchers with additional and valuable tools to analyze the functional potential of a community of species [168,169], which is a key approach to detect novel opportunities for biotechnological applications.

Metagenomics is a widely explored approach, promoting major shifts in understanding marine ecology. Specifically, metagenomics refers to the genetic and genomic analysis of microorganisms recovered from mixed communities from a specific environment [170] and can be utilized for the taxonomic and functional characterization of that environment. One example was the identification of proteorhodopsin, which led to the discovery of new trophic strategies in the ocean surfaces [171,172].

The advent of massive DNA and RNA sequencing technologies has enabled the development of large-scale research endeavors. After Craig Venter's seminal Global Ocean Sampling expedition [66,67], many others have been carried out, such as the Tara Oceans and Malaspina expeditions [173], which are leading marine researchers that explore novel tools to appropriately study this huge diversity. Although promising, these approaches are inherently complex. Depth of sequencing, library construction technique, and sequencing technology, for instance, can either enrich for specific fractions of the marine community or generate biases [174–176]. In addition, the bioinformatic approaches used in these studies are still evolving, constantly being improved to adapt to the challenges imposed by the complexity of this research and to the updating of reference information which is accumulating quickly. To support the validation of the best practices for achieving reproducible and reliable results, a comprehensive evaluation of such methods is being carried out in the framework of the initiative for the Critical Assessment of Metagenome Interpretation (CAMI) [177].

Recent investigations of the metabolic potential of genomes reconstructed from metagenomes defined thousands of genomes or their fragments from several marine environments [178–180]. Indeed, genome mining can be helpful for the identification of new compounds [181], whereas comparative genomics may lead to the inference of ecological or evolutionary patterns [182]. As an example, such approaches revealed an interesting functional partitioning between surface and deep-ocean populations of the clade SAR11 (which is one of the most abundant components of bacterioplankton) [183], which had important consequence on the global nitrogen balance [184].

As both an alternative and a complement to metagenomics, metatranscriptomics has recently been expanded and better explored to characterize complex natural communities from a functional point of view [185]. This approach, which has been used to characterized different ecosystems, including marine deep-sea sediments [186], provides advantages as compared to the DNA-based sequencing, which include minor susceptibility to amplification biases and the possibility to only capture the living fraction of the organisms inhabiting the community. Nevertheless, it is also

characterized by important hindrances and potential biases, including but not limited to the lack of reference genomes (which are needed for the evaluation of potential taxonomical and genomic novelties) and standardized laboratory procedures and bioinformatics pipelines [187]. Therefore, there is a compelling need for the development of standardized reference collections and protocols for metatranscriptomic annotations and analyses, which are still quite pioneering but might yield important results for bioprospecting [185].

The massive flow of meta-omics sequence data highlights the need for comprehensive databases to collect the accumulating information and, additionally, appropriate curation and tools to exploit this information for taxonomical assignments and functional analyses. Until now, meta-omics-specific reference databases did not exist. However, efforts have been implemented for the creation of sequence databases which could act as both repository and data analysis reference sites, most providing either access to specific sequence databases or to more generalized repositories. One example is represented by the MGnify tool within the EBI metagenomics portal [188]. This tool makes it possible for users to characterize raw sequence data and assembled contigs using either a taxonomical (through analyses of sequences related to small and large ribosomal subunits) or a functional-based approach (through gene finding and analysis of potential protein coding nucleotide sequences as compared with data from the InterPro database and GO [189,190], see also the proteomics section for more details), thus utilizing publicly-available reference databases. Another reference example to the scope is provided by the MG-RAST server [191], which is a storage and analysis dedicated web server that allows the processing and storage of raw sequence metagenome data (Table 1). Established in 2008, MG-RAST currently stores up to 203.43 Tbp of 385,064 metagenomes [192]. It allows users to search for taxonomic and functional annotations of the submitted sequence samples using a combination of sequence databases. RefSeq for taxonomic annotation of shotgun reads and contigs, the SEED Subsystem architecture, KO, NOG, and COG databases for functional annotation and the RDP, GreenGenes, and SILVA databases for ribosomal subunit similarity.

More multipurpose databases and repositories are represented by the KEGG MGENES partition [101], which is an attempt to store and organize a collection of genes reconstructed from metagenomes, allowing the search for specific genes, the browsing of gene annotations, the comparison among samples, and the BLAST-based comparisons against the database, and by the MMP (marine metagenomics portal), a specialized repository for (meta)genomic data for marine microbial organisms [193]. Currently, this system provides the MAR databases (contextual and sequence databases of complete and draft marine prokaryotic genomes, as well as genes and proteins from metagenomic samples, which can be downloaded to be deployed locally for other purposes), the META-PIPE pipeline [194] (a workflow for the analysis of metagenomics data, not yet available to the general public) and MAR BLAST (a basic BLAST search tool against the MAR databases). This repository can provide useful information concerning taxonomic and functional data of marine prokaryotes, which can be further investigated and tailored to gain insights on species and genes of biotechnological interest.

More recently, the two expeditions Tara Oceans [195] and bioGEOTRACES [196–199] started collecting marine water samples with data generated from both projects, which were organized in different, independent databases. In particular, the Tara Ocean expedition gave rise to more general bioinformatic resources, as well as to specialized sequence repositories. As an example, the Ocean Gene Atlas [200], a web server organizing a collection of 40 million prokaryotic genes and greater than 110 million eukaryotic transcripts, which have been produced in the framework of the Tara Ocean investigations, allows for the query and comparison of nucleotide and amino acid sequences against the built-in databases. This server has been used, for instance, to investigate a new class of potentially widely distributed subclass of carbonic anhydrase which might play important roles in the global carbon cycle [201]. The GLOSSary (GLobal Ocean 16S subunit web accessible resource) [202] represents an effort to appropriately organize reprocessed taxonomic data from prokaryotes extracted from published Tara Oceans sequence datasets. The platform allows users to explore and query

the underlying dataset to obtain indications on the distribution of prokaryotic organisms across the major oceanic basins. Although this specific platform currently exclusively relies on ribosomal data to investigate the taxonomic information within the Tara Ocean sequence data, it allows researchers to analyze the geospatial distribution of species of interest and also to gain insights into potential relationships, beyond providing data to complement experimental efforts and link genomic resources (including genes and gene clusters of biotechnological interest) to specific environments.

As addressed by these examples, the reported attempts to organize (meta)genomic data are disparate and heterogeneous, and none of them are specifically focused on bioprospection and biotechnological developments, even though separate tools are being made available to this aim, such as, for example, the dbCAN meta-server [203], which was designed for the investigation of carbohydrate-active enzymes. Possible approaches for bioprospection and identification of novel useful compounds in metagenomics include the identification of genes or gene clusters for the discovery of secondary metabolites and catalysts for their synthesis [204] or novel enzymes at the whole-community level, or in silico isolation and characterization of genomes or associated portions ("genome-resolved metagenomics"). After screening through bioinformatic pipelines, mainly based on similarity searches, potential genes and clusters can be identified, cloned, and expressed in heterologous hosts [62,205], as previously introduced.

2.3. Proteomics and Structural Biology

Proteins are the main actors in functional processes carried out by a biological system. They act in response to the development of internal or external stimuli, and to environmental changes [206,207]. Proteomics aims to identify and quantify proteins, a systems-based perspective of how organisms mount their molecular responses.

The two-dimensional gel electrophoresis (2-DE) technique, which separates mixtures of proteins based on their properties, enables the dissemination of different proteomic approaches [208]. Bottom-up proteomics procedures include the proteolysis of protein mixtures, and the analysis of the generated fragments by liquid chromatography–mass spectrometry (LC-MS) [209–211]. In top-down procedures, proteins are directly subjected to gas-phase fragmentation, followed by MS analysis [212,213]. Middle-down proteomic approaches [214], instead, generate longer peptide fragments as compared to bottom-up strategies that use protocols involving single-residue specific proteases such as Lys-C [215–217], Glu-C [218,219], Asp-N [220], and Lys-N [221,222].

Proteomic studies led to the discovery of peptides and toxins useful for biomedical research from sea anemone [223], sea sponge [224], cone snails [225], and cyanobacteria [30,226]. In a focused special issue of 2015, entitled "Proteomics in marine organisms" [227], 20 contributions about different species ranging from Bacteria and mammals, to microalgae and flowering plants, provided a representative compendium on marine proteomics.

All proteomics studies, as all omics approaches, rely on bioinformatic resources that enable the analysis of the raw data, as well as the exploitation of the produced outcomes. The reference resource of protein sequences and their annotation is UniProt, The Universal Protein Resource [228], collecting more than 16,000 reference proteomes (updated to July 2019). This general resource offers a BLAST server to sequence similarities detection by scanning the entire UniProt database, a multiple alignment tool based on the Clustal Omega program [229], and a text search by keywords.

One of the most comprehensive resources in terms of information related to protein sequences is InterPro [190], a database containing different kinds of classifications of protein-related features, including, as an example, protein family information from PFAM, the protein families database [230], accompanied by further detailed descriptions such as protein domains or sequence conserved signatures. Users can perform a similarity-based functional annotation, and also list all the proteins across all the species in the InterPro database having the same functional annotations. InterPro developers, moreover, freely distribute an associated software to enable the users to retrieve information about thousands of protein sequences in one analysis [231].

An important branch of proteomics for biotechnological applications is the so-called structural biology [232]. Structural studies on protein data are important to understand which and how amino acid sequences contribute to a specific protein folding, revealing structure–function relationships, a fundamental step for the elucidation of cellular processes. Protein structure information, as examples, is essential to address challenges in enzyme discovery and to identify ligand receptor properties, favoring protein design. The prediction of three-dimensional (3D) structures, the investigation of structural peculiarities, the simulation of functional and structural behavior of biomolecules, as well as their interactions provide valuable predictive tools as an alternative to expensive screening experiments, which are crucial to the search for lead compounds in biotechnological applications, drug discovery, and design. The main bioinformatic applications downstream of structural proteomics techniques, indeed, are in the fields of (1) prediction of protein structures, (2) molecular dynamics simulation, and (3) molecular docking.

Prediction of protein structures is a fundamental approach to highlight conformational aspects of molecules of biotechnological interest, for example, elucidating structural features related to environmental adaptation, such as warm or cold-adapted mechanisms that confer thermostability in extremophilic enzymes [233,234], or specifying enzymatic action useful for biotechnological applications [235]. The prediction of protein structures follows two main strategies: (1) comparative approaches, based on homology modeling [236] or protein threading techniques [237,238], which predict new structures by modeling sequences from unknown structures using solved structures from homolog sequences as templates, or by recognizing common protein folds in protein sequences that lack homolog sequences; and (2) ab initio approaches [239], based on intrinsic chemical and physical characteristics of amino acid sequences rather than previously solved structures. Major protein prediction programs and web resources are summarized in the related section in Table 1.

Molecular dynamics simulations enable evaluation of the biotechnological potential of molecules of interest by providing an in silico estimation of the stability of enzymes and protein complexes even before performing in vitro studies [240]. Widely used software packages for performing molecular dynamics simulations are GROMACS [241] and NAMD [242], which simulate the Newtonian equations of motion for biological systems with hundreds to millions of particles. Other widely used molecular dynamics simulation programs [243] are listed in the related section in Table 1.

Molecular docking is an in silico drug design approach to leverage 3D structures for ligand discovery, fitting one or more compounds into binding sites [244,245], predicting the bound conformations and the binding affinity. AutoDock [246] is one of the most-used programs for molecular docking and virtual screening, particularly after the speed up deriving from the implementation of multithreading in the AutoDock Vina [247] update. SwissDock [248] is a public webserver based on EADock DSS software [249] and on S3DB—a database of manually curated target and ligand structures [250] that is able to predict complexes between proteins and small ligands. These bioinformatics approaches, together with cutting-edge technologies able to highlight physical interaction with the target protein (e.g., the cellular thermal shift assay (CETSA) [251]), are essential in the design and development of new drugs. Other molecular docking programs [252] are listed in the related section in Table 1.

A fundamental reference resource for structural bioinformatics applications is the Protein Data Bank (PDB), a database of tridimensional structure data. It stores structures from X-ray crystallography, nuclear magnetic resonance (NMR), cryo-electron microscopy, and theoretical modeling [253]. The expansion of these useful collections thanks to novel technologies in high-throughput structure determination is also going to provide a consistent boost to the current information [254].

The major bottleneck of proteomics studies derives from their nature, that is, from the complexity of biological structures and of the physiological processes in which they are involved [255,256]. In particular, in structural biology applications, this complexity has an impact on the resolution of the protein structure data generated by crystallographic or NMR experiments [257], affecting all the downstream bioinformatics procedures described above, such as homology modelling,

molecular dynamics and drug design techniques. Structures with resolution of 3 Å or higher show only the basic conformation of the protein chain, lacking any information about their atomic structure [253]. Moreover, the need to isolate and study molecules through structure outside from their natural functional context, inhibiting their typical changes, can affect the right conformational assignments and mislead associated investigations on their behavior in biological environments.

2.4. Metabolomics

Metabolomics, together with other omics sciences, provides large and complex datasets, fundamental to understanding a wide variety of cellular processes. From this perspective, the extreme variability of chemical and physical conditions in the marine environment have made metabolomics a key field for the study of marine diversity [127]. The metabolome of an organism, in fact, directly correlates with gene expression and the associated protein production, affecting downstream functional pathways and representing the phenotypical responses of the organism to a vast range of physiological and environmental stimuli [258]. Often, organism reactions to a changing condition include the remodeling of their metabolism and regulating the levels of specific metabolites, which can potentially represent markers of a particular response (e.g., biotic or abiotic stresses). In this context, metabolomics helps the evaluation of the impact of climate changes on marine organisms, unraveling contributions that marine systems could play in mitigating the effects of global warming [259,260].

While MS-based proteomic approaches still require the separation of protein mixtures and the analysis of fragmented peptides, metabolomic approaches are based on the direct profiling of nonfragmented molecules via MS techniques [261,262].

Although bottlenecks make these technologies spread and improve their throughput, studies from metabolomics has helped, for example, to identify compounds with inhibitory effects against common human pathogens from the sponge bacterium *Rhodococcus* sp. UA13 [263], and from a panel of marine myxobacteria [264]. Other studies have suggested the use of marine-adapted fungi as biocontrol agents in agriculture [265].

Bioinformatics remarkable resources exploited in such efforts are the global natural products social molecular networking (GNPS) [266], which represents an open-access knowledge basis for organization and sharing of raw, processed, or identified tandem mass (MS/MS) spectrometry data, and the antiSMASH server [70], which allows genome-wide identification, annotation, and analysis of gene clusters related to secondary metabolite biosynthesis.

KEGG [101], Reactome [267], and MetaCyc [268] are the reference databases for enzymes, reactions, and metabolic and regulatory pathways, respectively. These resources include tools to highlight and interact with specific sub-paths or enzymes within the maps of the metabolic pathways of the selected species, enabling the download of maps in different file formats. An innovative software implemented in MetaCyc is the Pathway Tools software [269], which permits the computational prediction of the metabolic networks of any organism that has a sequenced and annotated genome [270].

ChemSpider [271] and The Super Natural II database [272] are two public resources providing access to the structure information of a huge diversity of compounds, including element composition, molecular weight, monoisotopic mass, and pharmacological activity.

NaPDoS [273] and MEROPS [274] are specific resources exclusively dedicated to secondary metabolite genes associated with polyketide synthase and non-ribosomal peptide synthesis pathways, and to peptidases, their substrates, and inhibitors.

All the presented resources rely on the correct definition of the biological function of the collected bioactive compounds and metabolites. An accurate annotation is necessary for data interpretation; however, metabolite identification is still a major bottleneck in untargeted metabolomics [275]. Computational workflows for metabolomic interpretation, including high-throughput metabolite profiling and annotation, are highly challenging tasks, with fast evolving metabolomics datasets specifically generated by dedicated service centers [274,275]. The main delicate issues are due to the variability of resolution and the difficulty to establish generalized standards from different specialized

laboratories and technologies. Although community guidelines for the detection of metabolites were established years ago, the adaptation to recommended standards is still far from being achieved. The complexity of metabolomic data from different combinations of various chromatographic and mass spectrometric acquisition methods has resulted in the establishment of diverse pipelines and workflows, which often involve nonstandardized manual curation. Furthermore, bioinformatics tools in the field still need to better address the problem of enrichment analyses, accurately linking metabolomic data to the most reliable similar compounds and building exhaustive pathway diagrams. These approaches need to be integrated into customizable workflows, such as the ones based on R or Python programming languages for the design of reusable and shared software [276].

The future of metabolite identification depends on the use of metabolome data repositories and associated data analysis tools, enabling data sharing and downstream analyses in an automated fashion, overcoming the lack of standardized methods or procedures [277–279].

3. Bottlenecks and Perspectives

3.1. Bottlenecks

It is evident from the many examples given above that marine organisms represent an ever-increasing subject of scientific investigation. Of course, marine organisms are also of great commercial interest for big (chemical, pharmacological, biotech) companies, especially those species belonging to the so-called marine areas beyond national jurisdiction (ABNJ), which cover 64 percent of the surface of the oceans and nearly 95 percent of its volume. Usually rich entities, such as big companies or first-world universities, which have invested money in the collection and analysis of marine organisms, secure their findings by national or international patents [280], thus denying the scientific community free access to their results. Although the Nagoya protocol (2010) has somehow addressed the need for regulation in the field, the question of who is the owner of the ocean's biodiversity data still remains open, and what kind of legal and political action are needed to prevent an unfair appropriation of marine data is still a subject of debate [280].

The establishment of general centralized data repositories through reference sites and exploitable through methodologies publicly accessible due to extended bioinformatic efforts in the different fields of the omics technologies, was an incredible achievement in biosciences, favoring the sharing and the spreading of information fundamental for the fast advancement of the scientific research. These efforts reduced the scientific costs, due to the redistributions of methods and results and paved the way for further investigations offering general benefits for all scientific communities. On the one hand, flourishing community-specific collections and the accessibility to these approaches, even for non-expert users, need a conscious set up also of shared rules and appropriate education to avoid spreading limited quality and poorly reusable datasets. On the other hand, prior to any bioinformatic approach, correct taxonomic information on the specimens used is paramount for the production of high-quality research: sadly, the number of expert taxonomists in many fields of biologic research is dwindling, in some cases due to the difficulties taxonomists face in the identification of specimens. In addition, specialized backup facilities are also required to maintain voucher collections for future inquiries.

Although, on one hand, the fast evolving performances of sequencing technologies are the core of the incredible acceleration of molecular data production at affordable costs, on the other hand the fast production and release of novel sequence data, such as those from genome or transcriptome assemblies, faces the bottleneck of slower bioinformatic research to establish curated collections and updated information, even in reference platforms. The constant fast release of novel sequenced genomes or different assemblies from the same genome, as an example, reduces the efficiency of data curation, resource updating, and, as a consequence, may affect the quality of the subsequent analyses, such as gene family or gene expression assessments, as well as comparative genomics [144]. This holds

for all the sections described in this review. Instead, specific bottlenecks that are exclusively related to a particular area of omics sciences, are discussed within each section of this review.

Bioinformatic tools need to follow the fast development of novel technologies and to adapt to the larger data size, but there is also a need for expert curators and of coordinated community feedback to validate the spreading results. Indeed, limits in data annotation and its updating and curation represent the main challenges in this field of research in biology.

3.2. Perspectives

Because of the precious amount of information that the sea can offer, the establishment of integrated marine data collections from multiple observations is emerging as a compelling need in the scientific community, primarily with the aim of assessing and monitoring the health status of marine ecosystems, but also as a multifaceted approach to unravel the complexity of marine species of biological and biotechnological interest [281,282]. These marine observatories are expected to support the collection, in both time and space, of several different types of information, ranging from classical biogeochemical and oceanographic measurements, to satellite images and multi-omics molecular data.

The enormous flow of data generated by the heterogeneous technologies which are being employed could initially appear as overwhelming when focusing on comparisons between different datasets. Big data should always suggest focusing on data acquisition, collection, and organization rather than data quality assessment as a first step. Data should always be considered to be an important factor of richness and one of the major opportunities for advancement, as long as efforts are made in the direction of their collection, integration, standardization, interrogation, and interpretation. In the past, the budget requirements for storing data have been quite disadvantageous; however, during the last decade, storage costs have become cheaper, due to hardware technologies and cloud commercial offerings, making a catch-all approach possible or even desirable at the moment. To this aim, great attention is being focused on current state-of-the-art of infrastructures to collect, organize, and share the data. Computing system architectures are undergoing rapid growth due to the establishment of cloud, virtualization, and orchestration technologies, which allows extremely complex, resilient, redundant, distributed, and almost infinitely scalable setups. At the same time, modern DBMS (Database management System) allow for heterogeneous big data, better metadata and, above all, extremely simple scalability.

The explosion of the so-called Internet of Things, allows for extended nomadic sensor networks and pervasive computing, providing fine grained, cost-effective, real-time data acquisition that represents a real boost in metadata enrichment and augmentation. All these factors, together with the growing trends in artificial intelligence (machine learning and deep learning, mostly) to build knowledge from the data, represent an intriguing challenging scenario to be exploited to disentangle the intrinsic complexity of big biological collections.

Apart from collecting, storing, and quality checking all the achievable useful data resources, the high heterogeneity of the data pushes the need for a systematic design of statistical, mathematical, computational, and bioinformatic tools aimed at analyzing and integrating them while exploiting multidisciplinary competences. Moreover, to let these data be comparable across time and space, shared protocols, reliable references, computing pipelines, and standard metadata must be established and agreed by the involved scientific communities [163], giving place to reliable long standing coordinated efforts.

Exploiting this spatial-temporal information to build new in silico models and predictors is of great importance to widen the knowledge on marine organisms and on their biotechnological relevance. A necessary step towards this direction is also ensuring the availability and accessibility of the information generated by the scientific community, by adopting the vision of data fairness [283]. The challenge of omics data integration is pivotal to understanding biological systems, i.e., transcriptomic and proteomic data can help improve the resolution of genome annotation [284], while coupling meta-metabolomics to metagenomics will link bioactive compounds and their

producers [285]. Another interesting perspective is the meta-analytical approach [286]. Plenty of data have been produced from marine resources, especially genomic and metagenomic datasets. Such data are usually analyzed to profile taxonomy and provide an overview on the main detectable functionalities. However, no platforms exist to provide standardized data processing for marine biology, such as the curatedMetagenomicData for the human metagenome [287].

A critical comparison and appropriate integration of results across different efforts will be an added value to identify real trends and sources of biases in the evolving area of marine biotechnology research, further leading scientific discovery forward.

References

1. Danovaro, R.; Corinaldesi, C.; Dell'Anno, A.; Fuhrman, J.A.; Middelburg, J.J.; Noble, R.T.; Suttle, C.A. Marine viruses and global climate change. *FEMS Microbiol. Rev.* **2011**, *35*, 993–1034. [CrossRef] [PubMed]

2. Argulis, L.; Schwartz, K.V. *Five Kingdoms: An Illustrated Guide to the Phyla of Life on Earth*; Freeman WH and Company: New York, NY, USA, 1982.

3. Macdougall, J.D. *A Short History of Planet Earth*; John Wiley (Ed.): New York, NY, USA, 1996; p. 274.

4. Bernhard, J.M.; Kormas, K.; Pachiadaki, M.G.; Rocke, E.; Beaudoin, D.J.; Morrison, C.; Visscher, P.T.; Cobban, A.; Starczak, V.R.; Edgcomb, V.P. Benthic protists and fungi of Mediterranean deep hypsersaline anoxic basin redoxcline sediments. *Front. Microbiol.* **2014**, *5*, 605. [CrossRef] [PubMed]

5. CAREX. Roadmap for Research on Life in Extreme Environment. Available online: http://commercialspace.pbworks.com/f/2011.01+CAREX_Roadmap_Final.pdf (accessed on 20 August 2019).

6. Chen, L.; DeVries, A.L.; Cheng, C.-H.C. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3817. [CrossRef] [PubMed]

7. Corinaldesi, C.; Tangherlini, M.; Luna, G.M.; Dell'Anno, A. Extracellular DNA can preserve the genetic signatures of present and past viral infection events in deep hypersaline anoxic basins. *Proc. R. Soc. B Biol. Sci.* **2014**, *281*, 20133299. [CrossRef] [PubMed]

8. Danovaro, R.; Gambi, C.; Dell'Anno, A.; Corinaldesi, C.; Pusceddu, A.; Neves, R.C.; Kristensen, R.M. The challenge of proving the existence of metazoan life in permanently anoxic deep-sea sediments. *BMC Biol.* **2016**, *14*, 43. [CrossRef]

9. Danovaro, R.; Molari, M.; Corinaldesi, C.; Dell'Anno, A. Macroecological drivers of archaea and bacteria in benthic deep-sea ecosystems. *Sci. Adv.* **2016**, *2*, e1500961. [CrossRef] [PubMed]

10. Gagnière, N.; Jollivet, D.; Boutet, I.; Brélivet, Y.; Busso, D.; Da Silva, C.; Gaill, F.; Higuet, D.; Hourdez, S.; Knoops, B.; et al. Insights into metazoan evolution from Alvinella pompejana cDNAs. *BMC Genom.* **2010**, *11*, 634. [CrossRef]

11. Hu, Y.; Ghigliotti, L.; Vacchi, M.; Pisano, E.; Detrich, H.W.; Albertson, R.C. Evolution in an extreme environment: Developmental biases and phenotypic integration in the adaptive radiation of antarctic notothenioids. *BMC Evol. Biol.* **2016**, *16*, 142. [CrossRef]

12. Jimeno, J.; Faircloth, G.; Sousa-Faro, J.M.F.; Scheuer, P.; Rinehart, K. New Marine Derived Anticancer Therapeutics—A Journey from the Sea to Clinical Trials. *Mar. Drugs* **2004**, *2*, 14–29. [CrossRef]

13. Zeppilli, D.; Leduc, D. Biodiversity and ecology of meiofauna in extreme and changing environments. *Mar. Biodivers.* **2018**, *48*, 1–4. [CrossRef]

14. Barone, G.; Rastelli, E.; Corinaldesi, C.; Tangherlini, M.; Danovaro, R.; Dell'Anno, A. Benthic deep-sea fungi in submarine canyons of the Mediterranean Sea. *Prog. Oceanogr.* **2018**, *168*, 57–64. [CrossRef]

15. Lauritano, C.; Andersen, J.H.; Hansen, E.; Albrigtsen, M.; Escalera, L.; Esposito, F.; Helland, K.; Hanssen, K.Ø.; Romano, G.; Ianora, A. Bioactivity Screening of Microalgae for Antioxidant, Anti-Inflammatory, Anticancer, Anti-Diabetes, and Antibacterial Activities. *Front. Mar. Sci.* **2016**, *3*, 68. [CrossRef]

16. Cherry, P.; Yadav, S.; Strain, C.R.; Allsopp, P.J.; McSorley, E.M.; Ross, R.P.; Stanton, C. Prebiotics from Seaweeds: An Ocean of Opportunity? *Mar. Drugs* **2019**, *17*, 6. [CrossRef] [PubMed]

17. Galasso, C.; Gentile, A.; Orefice, I.; Ianora, A.; Bruno, A.; Noonan, D.M.; Sansone, C.; Albini, A.; Brunet, C. Microalgal Derivatives as Potential Nutraceutical and Food Supplements for Human Health: A Focus on Cancer Prevention and Interception. *Nutrients* **2019**, *11*, 6. [CrossRef] [PubMed]

18. Overland, M.; Mydland, L.T.; Skrede, A. Marine macroalgae as sources of protein and bioactive compounds in feed for monogastric animals. *J. Sci. Food Agric.* **2019**, *99*, 13–24. [CrossRef] [PubMed]

19. Chatterjee, A.; Singh, S.; Agrawal, C.; Yadav, S.; Rai, R.; Rai, L.C. Chapter 10—Role of Algae as a Biofertilizer. In *Algal Green Chemistry*; Rastogi, R.P., Madamwar, D., Pandey, A., Eds.; Elsevier: Amsterdam, The Netherlands, 2017; pp. 189–200.

20. López-Mosquera, M.E.; Fernández-Lema, E.; Villares, R.; Corral, R.; Alonso, B.; Blanco, C. Composting fish waste and seaweed to produce a fertilizer for use in organic agriculture. *Procedia Environ. Sci.* **2011**, *9*, 113–117. [CrossRef]

21. Mohan, K.; Ravichandran, S.; Muralisankar, T.; Uthayakumar, V.; Chandirasekar, R.; Seedevi, P.; Abirami, R.G.; Rajan, D.K. Application of marine-derived polysaccharides as immunostimulants in aquaculture: A review of current knowledge and further perspectives. *Fish Shellfish Immunol.* **2019**, *86*, 1177–1193. [CrossRef] [PubMed]

22. Chi, Z.; Liu, G.-L.; Lu, Y.; Jiang, H.; Chi, Z.-M. Bio-products produced by marine yeasts and their potential applications. *Bioresour. Technol.* **2016**, *202*, 244–252. [CrossRef]

23. Maeda, Y.; Yoshino, T.; Matsunaga, T.; Matsumoto, M.; Tanaka, T. Marine microalgae for production of biofuels and chemicals. *Curr. Opin. Biotechnol.* **2018**, *50*, 111–120. [CrossRef]

24. Swain, M.R.; Natarajan, V.; Krishnan, C. Chapter Nine—Marine Enzymes and Microorganisms for Bioethanol Production. In *Advances in Food and Nutrition Research*; Kim, S.-K., Toldrá, F., Eds.; Academic Press: Cambridge, MA, USA, 2017; Volume 80, pp. 181–197.

25. Corinaldesi, C.; Barone, G.; Marcellini, F.; Dell'Anno, A.; Danovaro, R. Marine Microbial-Derived Molecules and Their Potential Use in Cosmeceutical and Cosmetic Products. *Mar. Drugs* **2017**, *15*, 118. [CrossRef]

26. Kim, J.H.; Lee, J.E.; Kim, K.H.; Kang, N.J. Beneficial Effects of Marine Algae-Derived Carbohydrates for Skin Health. *Mar. Drugs* **2018**, *16*, 459. [CrossRef] [PubMed]

27. Venkatesan, J.; Anil, S.; Kim, S.K.; Shim, M.S. Marine Fish Proteins and Peptides for Cosmeceuticals: A Review. *Mar. Drugs* **2017**, *15*, 143. [CrossRef] [PubMed]

28. Paniagua-Michel, J.; Rosales, A. Marine bioremediation-A sustainable biotechnology of petroleum hydrocarbons biodegradation in coastal and marine environments. *J. Bioremediation Biodegradation* **2015**, *6*, 1.

29. Costa, S.S.; Miranda, A.L.; de Morais, M.G.; Costa, J.A.V.; Druzian, J.I. Microalgae as source of polyhydroxyalkanoates (PHAs)—A review. *Int. J. Biol. Macromol.* **2019**, *131*, 536–547. [CrossRef] [PubMed]

30. Engene, N.; Rottacker, E.C.; Kastovsky, J.; Byrum, T.; Choi, H.; Ellisman, M.H.; Komarek, J.; Gerwick, W.H. Moorea producens gen. nov. sp. nov. and Moorea bouillonii comb. nov. tropical marine cyanobacteria rich in bioactive secondary metabolites. *Int. J. Syst. Evol. Microbiol.* **2012**, *62 Pt 5*, 1171–1178. [CrossRef]

31. Imhoff, J.F.; Labes, A.; Wiese, J. Bio-mining the microbial treasures of the ocean: New natural products. *Biotechnol. Adv.* **2011**, *29*, 468–482. [CrossRef] [PubMed]

32. Tan, L.T. Bioactive natural products from marine cyanobacteria for drug discovery. *Phytochemistry* **2007**, *68*, 954–979. [CrossRef]

33. Long, S.; Sousa, E.; Kijjoa, A.; Pinto, M.M. Marine Natural Products as Models to Circumvent Multidrug Resistance. *Molecules* **2016**, *21*, 892. [CrossRef]

34. Xiong, Z.Q.; Wang, J.F.; Hao, Y.Y.; Wang, Y. Recent advances in the discovery and development of marine microbial natural products. *Mar. Drugs* **2013**, *11*, 700–717. [CrossRef]

35. Xiong, Z.-Q.; Zhang, Z.-P.; Li, J.-H.; Wei, S.-J.; Tu, G.-Q. Characterization of Streptomyces padanus JAU4234, a producer of actinomycin X_2, fungichromin, and a new polyene macrolide antibiotic. *Appl. Environ. Microbiol.* **2012**, *78*, 589–592. [CrossRef]

36. Li, J.W.; Vederas, J.C. Drug discovery and natural products: End of an era or an endless frontier? *Science* **2009**, *325*, 161–165. [CrossRef] [PubMed]

37. Badawy, M.E.I.; Rabea, E.I. Current Applications in Food Preservation Based on Marine Biopolymers. In *Polymers for Food Applications*; Gutiérrez, T.J., Ed.; Springer International Publishing: Cham, Switzerland, 2018; pp. 609–650.

38. Venugopal, V. *Marine Polysaccharides: Food Applications*; CRC Press: Boca Raton, FL, USA, 2016.

39. Appeltans, W.; Decock, W.; Vanhoorne, B.; Hernandez, F.; Bouchet, P.; Boxshall, G.; Fauchald, K.; Gordon, D.; Poore, G.; Van Soest, R. The World Register of Marine Species: An Authoritative, Open-Access Web-Resource for All Marine Species. Available online: http://marinespecies.org/ (accessed on 20 August 2019).

40. Bisby, F.; Roskov, Y.; Culham, A.; Orrell, T.; Nicolson, D.; Paglinawan, L.; Bailly, N.; Appeltans, W.; Kirk, P.; Bourgoin, T.; et al. Species 2000 & ITIS Catalogue of Life, 2012 Annual Checklist. Available online: www.catalogueoflife.org/col/ (accessed on 20 August 2019).

41. Patterson, D.; Mozzherin, D.; Shorthouse, D.P.; Thessen, A. Challenges with using names to link digital biodiversity information. *Biodivers. Data J.* **2016**, *4*, e8080. [CrossRef] [PubMed]

42. Horton, T.; Kroh, A.; Ahyong, S.; Bailly, N.; Boyko, C.B.; Brandão, S.N.; Gofas, S.; Hooper, J.N.A.; Hernandez, F.; Holovachov, O.; et al. World Register of Marine Species. Available online: http://www.marinespecies.org (accessed on 20 August 2019).

43. Mora, C.; Tittensor, D.P.; Adl, S.; Simpson, A.G.; Worm, B. How many species are there on Earth and in the ocean? *PLoS Biol.* **2011**, *9*, e1001127. [CrossRef] [PubMed]

44. Greco, G.R.; Cinquegrani, M. Firms Plunge into the Sea. Marine Biotechnology Industry, a First Investigation. *Front. Mar. Sci.* **2016**, *2*, 124. [CrossRef]

45. Katz, L.; Baltz, R.H. Natural product discovery: Past, present, and future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155–176. [CrossRef] [PubMed]

46. Leclère, V.; Weber, T.; Jacques, P.; Pupin, M. Bioinformatics Tools for the Discovery of New Nonribosomal Peptides. *Methods Mol. Biol.* **2016**, *1401*, 209–232. [PubMed]

47. Lorente, A.; Makowski, K.; Albericio, F.; Álvarez, M. Bioactive marine polyketides as potential and promising drugs. *Ann. Mar. Biol. Res.* **2014**, *1*, 1–10.

48. Carteni, F.; Bonanomi, G.; Giannino, F.; Incerti, G.; Vincenot, C.E.; Chiusano, M.L.; Mazzoleni, S. Self-DNA inhibitory effects: Underlying mechanisms and ecological implications. *Plant Signal. Behav.* **2016**, *11*, e1158381. [CrossRef]

49. Mazzoleni, S.; Bonanomi, G.; Incerti, G.; Chiusano, M.L.; Termolino, P.; Mingo, A.; Senatore, M.; Giannino, F.; Carteni, F.; Rietkerk, M.; et al. Inhibitory and toxic effects of extracellular self-DNA in litter: A mechanism for negative plant-soil feedbacks? *New Phytol.* **2015**, *205*, 1195–1210. [CrossRef]

50. Mazzoleni, S.; Carteni, F.; Bonanomi, G.; Senatore, M.; Termolino, P.; Giannino, F.; Incerti, G.; Rietkerk, M.; Lanzotti, V.; Chiusano, M.L. Inhibitory effects of extracellular self-DNA: A general biological process? *New Phytol.* **2015**, *206*, 127–132. [CrossRef]

51. Liang, X.; Luo, D.; Luesch, H. Advances in exploring the therapeutic potential of marine natural products. *Pharmacol. Res.* **2019**, *147*, 104373. [CrossRef] [PubMed]

52. Iskar, M.; Zeller, G.; Zhao, X.M.; van Noort, V.; Bork, P. Drug discovery in the age of systems biology: The rise of computational approaches for data integration. *Curr. Opin. Biotechnol.* **2012**, *23*, 609–616. [CrossRef] [PubMed]

53. Whittaker, P.A. What is the relevance of bioinformatics to pharmacology? *Trends Pharmacol. Sci.* **2003**, *24*, 434–439. [CrossRef]

54. Ortega, S.S.; Cara, L.C.; Salvador, M.K. In silico pharmacology for a multidisciplinary drug discovery process. *Drug Metab. Drug Interact.* **2012**, *27*, 199–207. [CrossRef] [PubMed]

55. Katara, P. Role of bioinformatics and pharmacogenomics in drug discovery and development process. *Netw. Modeling Anal. Health Inform. Bioinform.* **2013**, *2*, 225–230. [CrossRef]

56. Ambrosino, L.; Colantuono, C.; Monticolo, F.; Chiusano, M.L. Bioinformatics Resources for Plant Genomics: Opportunities and Bottlenecks in The -omics Era. *Curr. Issues Mol. Biol.* **2017**, 71–88. [CrossRef]

57. Esposito, A.; Colantuono, C.; Ruggieri, V.; Chiusano, M.L. Bioinformatics for agriculture in the Next-Generation sequencing era. *Chem. Biol. Technol. Agric.* **2016**, *3*, 1–12. [CrossRef]

58. Trindade, M.; van Zyl, L.J.; Navarro-Fernández, J.; Abd Elrazak, A. Targeted metagenomics as a tool to tap into marine natural product diversity for the discovery and production of drug candidates. *Front. Microbiol.* **2015**, *6*, 890. [CrossRef] [PubMed]

59. Lauritano, C.; Ianora, A. Grand Challenges in Marine Biotechnology: Overview of Recent EU-Funded Projects. In *Grand Challenges in Marine Biotechnology*; Springer: New York, NY, USA, 2018; pp. 425–449.

60. Hartmann, E.M.; Durighello, E.; Pible, O.; Nogales, B.; Beltrametti, F.; Bosch, R.; Christie-Oleza, J.A.; Armengaud, J. Proteomics meets blue biotechnology: A wealth of novelties and opportunities. *Mar. Genom.* **2014**, *17*, 35–42. [CrossRef]

61. Lacerda, C.M.; Reardon, K.F. Environmental proteomics: Applications of proteome profiling in environmental microbiology and biotechnology. *Brief. Funct. Genom. Proteom.* **2009**, *8*, 75–87. [CrossRef]

62. Huo, L.; Hug, J.J.; Fu, C.; Bian, X.; Zhang, Y.; Müller, R. Heterologous expression of bacterial natural product biosynthetic pathways. *Nat. Prod. Rep.* **2019**. [CrossRef] [PubMed]

63. de Pascale, D.; De Santi, C.; Fu, J.; Landfald, B. The microbial diversity of Polar environments is a fertile ground for bioprospecting. *Mar. Genom.* **2012**, *8*, 15–22. [CrossRef] [PubMed]

64. Kennedy, J.; O'Leary, N.D.; Kiran, G.S.; Morrissey, J.P.; O'Gara, F.; Selvin, J.; Dobson, A.D. Functional metagenomic strategies for the discovery of novel enzymes and biosurfactants with biotechnological applications from marine ecosystems. *J. Appl. Microbiol.* **2011**, *111*, 787–799. [CrossRef] [PubMed]

65. Parte, S.; Sirisha, V.L.; D'Souza, J.S. Chapter Four—Biotechnological Applications of Marine Enzymes from Algae, Bacteria, Fungi, and Sponges. In *Advances in Food and Nutrition Research*; Kim, S.-K., Toldrá, F., Eds.; Academic Press: Cambridge, MA, USA, 2017; Volume 80, pp. 75–106.

66. Gross, L. Untapped Bounty: Sampling the Seas to Survey Microbial Biodiversity. *PLoS Biol.* **2007**, *5*, e85. [CrossRef] [PubMed]

67. Venter, J.C.; Remington, K.; Heidelberg, J.F.; Halpern, A.L.; Rusch, D.; Eisen, J.A.; Wu, D.; Paulsen, I.; Nelson, K.E.; Nelson, W.; et al. Environmental Genome Shotgun Sequencing of the Sargasso Sea. *Science* **2004**, *304*, 66. [CrossRef] [PubMed]

68. Kim, S.-K.; Venkatesan, J. Introduction to Marine Biotechnology. In *Springer Handbook of Marine Biotechnology*; Kim, S.-K., Ed.; Springer Berlin Heidelberg: Berlin/Heidelberg, Germany, 2015; pp. 1–10.

69. Medema, M.H.; Fischbach, M.A. Computational approaches to natural product discovery. *Nat. Chem. Biol.* **2015**, *11*, 639–648. [CrossRef]

70. Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S.Y.; Medema, M.H.; Weber, T. antiSMASH 5.0: Updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res.* **2019**. [CrossRef]

71. Lin, W.R.; Tan, S.I.; Hsiang, C.C.; Sung, P.K.; Ng, I.S. Challenges and opportunity of recent genome editing and multi-omics in cyanobacteria and microalgae for biorefinery. *Bioresour. Technol.* **2019**, *291*, 121932. [CrossRef]

72. Doudna, J.A.; Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **2014**, *346*, 1258096. [CrossRef]

73. Heigwer, F.; Kerr, G.; Boutros, M. E-CRISP: Fast CRISPR target site identification. *Nat. Methods* **2014**, *11*, 122. [CrossRef]

74. Montague, T.G.; Cruz, J.M.; Gagnon, J.A.; Church, G.M.; Valen, E. CHOPCHOP: A CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res.* **2014**, *42*, W401–W407. [CrossRef] [PubMed]

75. Stemmer, M.; Thumberger, T.; del Sol Keyer, M.; Wittbrodt, J.; Mateo, J.L. CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS ONE* **2015**, *10*, e0124633. [CrossRef] [PubMed]

76. Saraswathy, N.; Ramalingam, P. 7—Genome sequencing methods. In *Concepts and Techniques in Genomics and Proteomics*; Woodhead Publishing: Cambridge, UK, 2011; pp. 95–107.

77. Magi, A.; Benelli, M.; Gozzini, A.; Girolami, F.; Torricelli, F.; Brandi, M.L. Bioinformatics for next generation sequencing data. *Genes* **2010**, *1*, 294–307. [CrossRef] [PubMed]

78. Dehal, P.; Satou, Y.; Campbell, R.K.; Chapman, J.; Degnan, B.; De Tomaso, A.; Davidson, B.; Di Gregorio, A.; Gelpke, M.; Goodstein, D.M.; et al. The Draft Genome of *Ciona intestinalis*: Insights into Chordate and Vertebrate Origins. *Science* **2002**, *298*, 2157. [CrossRef]

79. Sea Urchin Genome Sequencing, C.; Sodergren, E.; Weinstock, G.M.; Davidson, E.H.; Cameron, R.A.; Gibbs, R.A.; Angerer, R.C.; Angerer, L.M.; Arnone, M.I.; Burgess, D.R.; et al. The genome of the sea urchin Strongylocentrotus purpuratus. *Science* **2006**, *314*, 941–952. [CrossRef]

80. Carreras, C.; Ordóñez, V.; Zane, L.; Kruschel, C.; Nasto, I.; Macpherson, E.; Pascual, M. Population genomics of an endemic Mediterranean fish: Differentiation by fine scale dispersal and adaptation. *Sci. Rep.* **2017**, *7*, 43417. [CrossRef]

81. Igarashi, Y.; Zhang, H.; Tan, E.; Sekino, M.; Yoshitake, K.; Kinoshita, S.; Mitsuyama, S.; Yoshinaga, T.; Chow, S.; Kurogi, H.; et al. Whole-Genome Sequencing of 84 Japanese Eels Reveals Evidence against Panmixia and Support for Sympatric Speciation. *Genes* **2018**, *9*, 10. [CrossRef]

82. Malde, K.; Seliussen, B.B.; Quintela, M.; Dahle, G.; Besnier, F.; Skaug, H.J.; Øien, N.; Solvang, H.K.; Haug, T.; Skern-Mauritzen, R.; et al. Whole genome resequencing reveals diagnostic markers for investigating global migration and hybridization between minke whale species. *BMC Genom.* **2017**, *18*, 76. [CrossRef]

83. Xu, S.; Song, N.; Zhao, L.; Cai, S.; Han, Z.; Gao, T. Genomic evidence for local adaptation in the ovoviviparous marine fish Sebastiscus marmoratus with a background of population homogeneity. *Sci. Rep.* **2017**, *7*, 1562. [CrossRef]

84. Xu, S.; Zhao, L.; Xiao, S.; Gao, T. Whole genome resequencing data for three rockfish species of Sebastes. *Sci. Data* **2019**, *6*, 97. [CrossRef]

85. Cochrane, G.; Karsch-Mizrachi, I.; Takagi, T. The International Nucleotide Sequence Database Collaboration. *Nucleic Acids Res.* **2016**, *44*, D48–D50. [CrossRef] [PubMed]

86. CRIBI Database. Available online: http://genomes.cribi.unipd.it (accessed on 25 January 2018).

87. NCBI_Resource_Coordinators, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13. [CrossRef] [PubMed]

88. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef] [PubMed]

89. Cunningham, F.; Achuthan, P.; Akanni, W.; Allen, J.; Amode, M.R.; Armean, I.M.; Bennett, R.; Bhai, J.; Billis, K.; Boddu, S.; et al. Ensembl 2019. *Nucleic Acids Res.* **2019**, *47*, D745–D751. [CrossRef] [PubMed]

90. Li, W.; Cowley, A.; Uludag, M.; Gur, T.; McWilliam, H.; Squizzato, S.; Park, Y.M.; Buso, N.; Lopez, R. The EMBL-EBI bioinformatics web and programmatic tools framework. *Nucleic Acids Res.* **2015**, *43*, W580–W584. [CrossRef] [PubMed]

91. Mashima, J.; Kodama, Y.; Kosuge, T.; Fujisawa, T.; Katayama, T.; Nagasaki, H.; Okuda, Y.; Kaminuma, E.; Ogasawara, O.; Okubo, K.; et al. DNA data bank of Japan (DDBJ) progress report. *Nucleic Acids Res.* **2016**, *44*, D51–D57. [CrossRef]

92. Kodama, Y.; Shumway, M.; Leinonen, R. The Sequence Read Archive: Explosive growth of sequencing data. *Nucleic Acids Res.* **2012**, *40*, D54–D56. [CrossRef]

93. Barrett, T.; Clark, K.; Gevorgyan, R.; Gorelenkov, V.; Gribov, E.; Karsch-Mizrachi, I.; Kimelman, M.; Pruitt, K.D.; Resenchuk, S.; Tatusova, T.; et al. BioProject and BioSample databases at NCBI: Facilitating capture and organization of metadata. *Nucleic Acids Res.* **2012**, *40*, D57–D63. [CrossRef]

94. Kolesnikov, N.; Hastings, E.; Keays, M.; Melnichuk, O.; Tang, Y.A.; Williams, E.; Dylag, M.; Kurbatova, N.; Brandizi, M.; Burdett, T.; et al. ArrayExpress update–simplifying data submissions. *Nucleic Acids Res.* **2015**, *43*, D1113–D1116. [CrossRef]

95. Leinonen, R.; Akhtar, R.; Birney, E.; Bower, L.; Cerdeno-Tárraga, A.; Cheng, Y.; Cleland, I.; Faruque, N.; Goodgame, N.; Gibson, R.; et al. The European Nucleotide Archive. *Nucleic Acids Res.* **2011**, *39*, D28–D31. [CrossRef]

96. Kaminuma, E.; Mashima, J.; Kodama, Y.; Gojobori, T.; Ogasawara, O.; Okubo, K.; Takagi, T.; Nakamura, Y. DDBJ launches a new archive database with analytical tools for next-generation sequence data. *Nucleic Acids Res.* **2010**, *38*, D33–D38. [CrossRef] [PubMed]

97. Chen, I.A.; Chu, K.; Palaniappan, K.; Pillay, M.; Ratner, A.; Huang, J.; Huntemann, M.; Varghese, N.; White, J.R.; Seshadri, R.; et al. IMG/M v.5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **2019**, *47*, D666–D677. [CrossRef] [PubMed]

98. Mende, D.R.; Letunic, I.; Huerta-Cepas, J.; Li, S.S.; Forslund, K.; Sunagawa, S.; Bork, P. proGenomes: A resource for consistent functional and taxonomic annotations of prokaryotic genomes. *Nucleic Acids Res.* **2017**, *45*, D529–D534. [CrossRef] [PubMed]

99. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef] [PubMed]

100. Yin, Y.; Mao, X.; Yang, J.; Chen, X.; Mao, F.; Xu, Y. dbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2012**, *40*, W445–W451. [CrossRef] [PubMed]

101. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef] [PubMed]

102. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]

103. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2009**, *25*, 288–289. [CrossRef]

104. Boguski, M.S.; Lowe, T.M.; Tolstoshev, C.M. dbEST—Database for "expressed sequence tags". *Nat. Genet.* **1993**, *4*, 332–333. [CrossRef]

105. Clarke, K.; Yang, Y.; Marsh, R.; Xie, L.; Zhang, K.K. Comparative analysis of de novo transcriptome assembly. *Sci. China Life Sci.* **2013**, *56*, 156–162. [CrossRef]

106. Costa-Silva, J.; Domingues, D.; Lopes, F.M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **2017**, *12*, e0190152. [CrossRef] [PubMed]

107. Hrdlickova, R.; Toloue, M.; Tian, B. RNA-Seq methods for transcriptome analysis. *Wiley Interdiscip. Rev. Rna* **2017**, *8*, e1364. [CrossRef] [PubMed]

108. Kukurba, K.R.; Montgomery, S.B. RNA Sequencing and Analysis. *Cold Spring Harb Protoc* **2015**, *2015*, 951–969. [CrossRef] [PubMed]

109. Heller, M.J. DNA Microarray Technology: Devices, Systems, and Applications. *Annu. Rev. Biomed. Eng.* **2002**, *4*, 129–153. [CrossRef] [PubMed]

110. Bostan, H.; Chiusano, M.L. NexGenEx-Tom: A gene expression platform to investigate the functionalities of the tomato genome. *BMC Plant Biol.* **2015**, *15*, 48. [CrossRef] [PubMed]

111. Garcia-Jimenez, P.; Llorens, C.; Roig, F.J.; Robaina, R.R. Analysis of the Transcriptome of the Red Seaweed Grateloupia imbricata with Emphasis on Reproductive Potential. *Mar. Drugs* **2018**, *16*, 490. [CrossRef] [PubMed]

112. Huerlimann, R.; Wade, N.M.; Gordon, L.; Montenegro, J.D.; Goodall, J.; McWilliam, S.; Tinning, M.; Siemering, K.; Giardina, E.; Donovan, D.; et al. De novo assembly, characterization, functional annotation and expression patterns of the black tiger shrimp (*Penaeus monodon*) transcriptome. *Sci. Rep.* **2018**, *8*, 13553. [CrossRef] [PubMed]

113. Lan, Y.; Sun, J.; Xu, T.; Chen, C.; Tian, R.; Qiu, J.-W.; Qian, P.-Y. De novo transcriptome assembly and positive selection analysis of an individual deep-sea fish. *BMC Genom.* **2018**, *19*, 394. [CrossRef]

114. Lauritano, C.; De Luca, D.; Ferrarini, A.; Avanzato, C.; Minio, A.; Esposito, F.; Ianora, A. De novo transcriptome of the cosmopolitan dinoflagellate Amphidinium carterae to identify enzymes with biotechnological potential. *Sci. Rep.* **2017**, *7*, 11701. [CrossRef]

115. Onimaru, K.; Tatsumi, K.; Shibagaki, K.; Kuraku, S. A de novo transcriptome assembly of the zebra bullhead shark, Heterodontus zebra. *Sci. Data* **2018**, *5*, 180197. [CrossRef]

116. Roncalli, V.; Cieslak, M.C.; Sommer, S.A.; Hopcroft, R.R.; Lenz, P.H. De novo transcriptome assembly of the calanoid copepod Neocalanus flemingeri: A new resource for emergence from diapause. *Mar. Genom.* **2018**, *37*, 114–119. [CrossRef] [PubMed]

117. Lauritano, C.; De Luca, D.; Amoroso, M.; Benfatto, S.; Maestri, S.; Racioppi, C.; Esposito, F.; Ianora, A. New molecular insights on the response of the green alga Tetraselmis suecica to nitrogen starvation. *Sci. Rep.* **2019**, *9*, 3336. [CrossRef] [PubMed]

118. Gao, B.; Peng, C.; Zhu, Y.; Sun, Y.; Zhao, T.; Huang, Y.; Shi, Q. High Throughput Identification of Novel Conotoxins from the Vermivorous Oak Cone Snail (*Conus quercinus*) by Transcriptome Sequencing. *Int. J. Mol. Sci.* **2018**, *19*, 3901. [CrossRef] [PubMed]

119. Hu, H.; Bandyopadhyay, P.K.; Olivera, B.M.; Yandell, M. Elucidation of the molecular envenomation strategy of the cone snail Conus geographus through transcriptome sequencing of its venom duct. *Bmc Genom.* **2012**, *13*, 284. [CrossRef] [PubMed]

120. Yao, G.; Peng, C.; Zhu, Y.; Fan, C.; Jiang, H.; Chen, J.; Cao, Y.; Shi, Q. High-Throughput Identification and Analysis of Novel Conotoxins from Three Vermivorous Cone Snails by Transcriptome Sequencing. *Mar. Drugs* **2019**, *17*, 193. [CrossRef] [PubMed]

121. Rivera-de-Torre, E.; Palacios-Ortega, J.; Gavilanes, J.G.; Martínez-del-Pozo, Á.; García-Linares, S. Pore-Forming Proteins from Cnidarians and Arachnids as Potential Biotechnological Tools. *Toxins* **2019**, *11*, 370. [CrossRef] [PubMed]

122. Xie, B.; Huang, Y.; Baumann, K.; Fry, B.G.; Shi, Q. From Marine Venoms to Drugs: Efficiently Supported by a Combination of Transcriptomics and Proteomics. *Mar. Drugs* **2017**, *15*, 103. [CrossRef]

123. Kumar, A.; Sørensen, J.L.; Hansen, F.T.; Arvas, M.; Syed, M.F.; Hassan, L.; Benz, J.P.; Record, E.; Henrissat, B.; Pöggeler, S.; et al. Genome Sequencing and analyses of Two Marine Fungi from the North Sea Unraveled a Plethora of Novel Biosynthetic Gene Clusters. *Sci. Rep.* **2018**, *8*, 10187. [CrossRef]

124. Morlighem, J.-É.R.L.; Huang, C.; Liao, Q.; Braga Gomes, P.; Daniel Pérez, C.; De Brandão Prieto-da-Silva, Á.R.; Ming-Yuen Lee, S.; Rádis-Baptista, G. The Holo-Transcriptome of the Zoantharian Protopalythoa variabilis (Cnidaria: Anthozoa): A Plentiful Source of Enzymes for Potential Application in Green Chemistry, Industrial and Pharmaceutical Biotechnology. *Mar. Drugs* **2018**, *16*, 207. [CrossRef]

125. Smith, D.R.M.; Uria, A.R.; Helfrich, E.J.N.; Milbredt, D.; van Pee, K.H.; Piel, J.; Goss, R.J.M. An Unusual Flavin-Dependent Halogenase from the Metagenome of the Marine Sponge Theonella swinhoei WA. *ACS Chem. Biol.* **2017**, *12*, 1281–1287. [CrossRef]

126. Sarian, F.D.; Janecek, S.; Pijning, T.; Ihsanawati; Nurachman, Z.; Radjasa, O.K.; Dijkhuizen, L.; Natalia, D.; van der Maarel, M.J. A new group of glycoside hydrolase family 13 alpha-amylases with an aberrant catalytic triad. *Sci. Rep.* **2017**, *7*, 44230. [CrossRef] [PubMed]

127. Romano, G.; Costantini, M.; Sansone, C.; Lauritano, C.; Ruocco, N.; Ianora, A. Marine microorganisms as a promising and sustainable source of bioactive molecules. *Marine Environ. Res.* **2017**, *128*, 58–69. [CrossRef] [PubMed]

128. Amos, G.C.A.; Awakawa, T.; Tuttle, R.N.; Letzel, A.-C.; Kim, M.C.; Kudo, Y.; Fenical, W.; Moore, B.S.; Jensen, P.R. Comparative transcriptomics as a guide to natural product discovery and biosynthetic gene cluster functionality. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E11121. [CrossRef] [PubMed]

129. Gorson, J.; Ramrattan, G.; Verdes, A.; Wright, E.M.; Kantor, Y.; Rajaram Srinivasan, R.; Musunuri, R.; Packer, D.; Albano, G.; Qiu, W.G.; et al. Molecular Diversity and Gene Evolution of the Venom Arsenal of Terebridae Predatory Marine Snails. *Genome Biol. Evol.* **2015**, *7*, 1761–1778. [CrossRef] [PubMed]

130. Buenrostro, J.D.; Wu, B.; Chang, H.Y.; Greenleaf, W.J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **2015**, *109*. [CrossRef]

131. Park, P.J. ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **2009**, *10*, 669–680. [CrossRef]

132. Walker, D.L.; Bhagwate, A.V.; Baheti, S.; Smalley, R.L.; Hilker, C.A.; Sun, Z.; Cunningham, J.M. DNA methylation profiling: Comparison of genome-wide sequencing methods and the Infinium Human Methylation 450 Bead Chip. *Epigenomics* **2015**, *7*, 1287–1302. [CrossRef]

133. Ramsköld, D.; Luo, S.; Wang, Y.-C.; Li, R.; Deng, Q.; Faridani, O.R.; Daniels, G.A.; Khrebtukova, I.; Loring, J.F.; Laurent, L.C.; et al. Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **2012**, *30*, 777–782. [CrossRef]

134. Brozovic, M.; Dantec, C.; Dardaillon, J.; Dauga, D.; Faure, E.; Gineste, M.; Louis, A.; Naville, M.; Nitta, K.R.; Piette, J.; et al. ANISEED 2017: Extending the integrated ascidian database to the exploration and evolutionary comparison of genome-scale datasets. *Nucleic Acids Res.* **2018**, *46*, D718–D725. [CrossRef]

135. Kudtarkar, P.; Cameron, R.A. Echinobase: An expanding resource for echinoderm genomic information. *Database* **2017**, *2017*, bax074. [CrossRef]

136. Wang, J.; Kong, L.; Gao, G.; Luo, J. A brief introduction to web-based genome browsers. *Brief. Bioinform.* **2012**, *14*, 131–143. [CrossRef] [PubMed]

137. Needleman, S.B.; Wunsch, C.D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453. [CrossRef]

138. Smith, T.F.; Waterman, M.S. Identification of common molecular subsequences. *J. Mol. Biol.* **1981**, *147*, 195–197. [CrossRef]

139. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and applications. *BMC Bioinform.* **2009**, *10*, 421. [CrossRef] [PubMed]

140. Kielbasa, S.M.; Wan, R.; Sato, K.; Horton, P.; Frith, M.C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **2011**, *21*, 487–493. [CrossRef] [PubMed]

141. Darling, A.C.; Mau, B.; Blattner, F.R.; Perna, N.T. Mauve: Multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **2004**, *14*, 1394–1403. [CrossRef]

142. Altenhoff, A.M.; Dessimoz, C. Phylogenetic and functional assessment of orthologs inference projects and methods. *Plos Comput. Biol.* **2009**, *5*, e1000262. [CrossRef]

143. Altenhoff, A.M.; Dessimoz, C. Inferring orthology and paralogy. *Methods Mol. Biol.* **2012**, *855*, 259–279.

144. Ambrosino, L.; Chiusano, M.L. Transcriptologs: A Transcriptome-Based Approach to Predict Orthology Relationships. *Bioinform. Biol. Insights* **2017**, *11*, 1–8. [CrossRef]

145. Dolinski, K.; Botstein, D. Orthology and functional conservation in eukaryotes. *Annu. Rev. Genet.* **2007**, *41*, 465–507. [CrossRef]

146. Gabaldon, T.; Koonin, E.V. Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* **2013**, *14*, 360–366. [CrossRef] [PubMed]

147. Koonin, E.V. Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* **2005**, *39*, 309–338. [CrossRef] [PubMed]

148. Sonnhammer, E.L.; Koonin, E.V. Orthology, paralogy and proposed classification for paralog subtypes. *Trends Genet.* **2002**, *18*, 619–620. [CrossRef]

149. Ambrosino, L.; Ruggieri, V.; Bostan, H.; Miralto, M.; Vitulo, N.; Zouine, M.; Barone, A.; Bouzayen, M.; Frusciante, L.; Pezzotti, M.; et al. Multilevel comparative bioinformatics to investigate evolutionary relationships and specificities in gene annotations: An example for tomato and grapevine. *BMC Bioinform.* **2018**, *19*, 435. [CrossRef] [PubMed]

150. Tettelin, H.; Riley, D.; Cattuto, C.; Medini, D. Comparative genomics: The bacterial pan-genome. *Curr. Opin. Microbiol.* **2008**, *11*, 472–477. [CrossRef] [PubMed]

151. Adams, K.L.; Wendel, J.F. Polyploidy and genome evolution in plants. *Curr. Opin. Plant Biol.* **2005**, *8*, 135–141. [CrossRef] [PubMed]

152. Donoghue, P.C.J.; Purnell, M.A. Genome duplication, extinction and vertebrate evolution. *Trends Ecol. Evol.* **2005**, *20*, 312–319. [CrossRef] [PubMed]

153. Vernikos, G.; Medini, D.; Riley, D.R.; Tettelin, H. Ten years of pan-genome analyses. *Curr. Opin. Microbiol.* **2015**, *23*, 148–154. [CrossRef]

154. Snipen, L.; Liland, K.H. micropan: An R-package for microbial pan-genomics. *Bmc Bioinform.* **2015**, *16*, 79. [CrossRef]

155. Jun, S.R.; Robeson, M.S.; Hauser, L.J.; Schadt, C.W.; Gorin, A.A. PanFP: Pangenome-based functional profiles for microbial communities. *BMC Res. Notes* **2015**, *8*, 479. [CrossRef]

156. Chen, X.; Zhang, Y.; Zhang, Z.; Zhao, Y.; Sun, C.; Yang, M.; Wang, J.; Liu, Q.; Zhang, B.; Chen, M.; et al. PGAweb: A Web Server for Bacterial Pan-Genome Analysis. *Front. Microbiol.* **2018**, *9*, 1910. [CrossRef] [PubMed]

157. Loiseau, C.; Hatte, V.; Andrieu, C.; Barlet, L.; Cologne, A.; Oliveira, R.D.; Ferrato-Berberian, L.; Gardon, H.e.; Lauber, D.; Molinier, M.e.; et al. PanGeneHome: A Web Interface to Analyze Microbial Pangenomes. *J. Bioinf. Com. Sys. Bio.* **2017**, *1*, 108.

158. Rouli, L.; Mbengue, M.; Robert, C.; Ndiaye, M.; La Scola, B.; Raoult, D. Genomic analysis of three African strains of Bacillus anthracis demonstrates that they are part of the clonal expansion of an exclusively pathogenic bacterium. *New Microbes New Infect.* **2014**, *2*, 161–169. [CrossRef] [PubMed]

159. Freschi, L.; Vincent, A.T.; Jeukens, J.; Emond-Rheault, J.-G.; Kukavica-Ibrulj, I.; Dupont, M.-J.; Charette, S.J.; Boyle, B.; Levesque, R.C. The Pseudomonas aeruginosa Pan-Genome Provides New Insights on Its Population Structure, Horizontal Gene Transfer, and Pathogenicity. *Genome Biol. Evol.* **2018**, *11*, 109–120. [CrossRef] [PubMed]

160. Bosi, E.; Fondi, M.; Orlandini, V.; Perrin, E.; Maida, I.; de Pascale, D.; Tutino, M.L.; Parrilli, E.; Lo Giudice, A.; Filloux, A.; et al. The pangenome of (Antarctic) Pseudoalteromonas bacteria: Evolutionary and functional insights. *BMC Genom.* **2017**, *18*, 93. [CrossRef] [PubMed]

161. Park, C.J.; Andam, C.P. Within-Species Genomic Variation and Variable Patterns of Recombination in the Tetracycline Producer Streptomyces rimosus. *Front. Microbiol.* **2019**, *10*, 552. [CrossRef] [PubMed]

162. Tang, X.; Li, J.; Millan-Aguinaga, N.; Zhang, J.J.; O'Neill, E.C.; Ugalde, J.A.; Jensen, P.R.; Mantovani, S.M.; Moore, B.S. Identification of Thiotetronic Acid Antibiotic Biosynthetic Pathways by Target-directed Genome Mining. *ACS Chem. Biol.* **2015**, *10*, 2841–2849. [CrossRef]

163. Chiusano, M.L. On the Multifaceted Aspects of Bioinformatics in the Next Generation Era: The Run that must keep the Quality. *Next Gener. Seq. Applic* **2015**, *2*, e106. [CrossRef]

164. NCBI. The NCBI Eukaryotic Genome Annotation Pipeline. Available online: https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/ (accessed on 20 August 2019).

165. ENSEMBL. Gene Annotation in Ensembl. Available online: https://www.ensembl.org/info/genome/genebuild/genome_annotation.html (accessed on 20 August 2019).

166. Salzberg, S.L. Next-generation genome annotation: We still struggle to get it right. *Genome Biol.* **2019**, *20*, 92. [CrossRef]

167. Colantuono, C.; Miralto, M.; Sangiovanni, M.; Ambrosino, L.; Chiusano, M.L. *GENOMA: A Multilevel Platform for Marine Biology*; PeerJ Preprints: Madera, CA, USA, 2018.

168. Barone, R.; De Santi, C.; Palma Esposito, F.; Tedesco, P.; Galati, F.; Visone, M.; Di Scala, A.; De Pascale, D. Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. *Front. Mar. Sci.* **2014**, *1*, 38. [CrossRef]

169. Madhavan, A.; Sindhu, R.; Parameswaran, B.; Sukumaran, R.K.; Pandey, A. Metagenome Analysis: A Powerful Tool for Enzyme Bioprospecting. *Appl. Biochem. Biotechnol.* **2017**, *183*, 636–651. [CrossRef] [PubMed]

170. Handelsman, J. Metagenomics: Application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev. Mmbr* **2004**, *68*, 669–685. [CrossRef] [PubMed]

171. Béjà, O.; Aravind, L.; Koonin, E.V.; Suzuki, M.T.; Hadd, A.; Nguyen, L.P.; Jovanovich, S.B.; Gates, C.M.; Feldman, R.A.; Spudich, J.L.; et al. Bacterial Rhodopsin: Evidence for a New Type of Phototrophy in the Sea. *Science* **2000**, *289*. [CrossRef]

172. Béjà, O.; Spudich, E.N.; Spudich, J.L.; Leclerc, M.; DeLong, E.F. Proteorhodopsin phototrophy in the ocean. *Nature* **2001**, *411*, 786–789. [CrossRef] [PubMed]

173. Gregory, A.C.; Zayed, A.A.; Conceicao-Neto, N.; Temperton, B.; Bolduc, B.; Alberti, A.; Ardyna, M.; Arkhipova, K.; Carmichael, M.; Cruaud, C.; et al. Marine DNA Viral Macro- and Microdiversity from Pole to Pole. *Cell* **2019**. [CrossRef] [PubMed]

174. Chistoserdova, L. Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol. Lett.* **2010**, *32*, 1351–1359. [CrossRef]

175. Roumpeka, D.D.; Wallace, R.J.; Escalettes, F.; Fotheringham, I.; Watson, M. A Review of Bioinformatics Tools for Bio-Prospecting from Metagenomic Sequence Data. *Front Genet* **2017**, *8*, 23. [CrossRef]

176. Teeling, H.; Glockner, F.O. Current opportunities and challenges in microbial metagenome analysis—A bioinformatic perspective. *Brief. Bioinform.* **2012**, *13*, 728–742. [CrossRef]

177. Bremges, A.; McHardy, A.C. Critical Assessment of Metagenome Interpretation Enters the Second Round. *mSystems* **2018**, *3*, 4. [CrossRef]

178. Dombrowski, N.; Teske, A.P.; Baker, B.J. Expansive microbial metabolic versatility and biodiversity in dynamic Guaymas Basin hydrothermal sediments. *Nat. Commun.* **2018**, *9*, 4999. [CrossRef]

179. Seitz, K.W.; Dombrowski, N.; Eme, L.; Spang, A.; Lombard, J.; Sieber, J.R.; Teske, A.P.; Ettema, T.J.G.; Baker, B.J. Asgard archaea capable of anaerobic hydrocarbon cycling. *Nat. Commun.* **2019**, *10*, 1822. [CrossRef] [PubMed]

180. Tully, B.J.; Graham, E.D.; Heidelberg, J.F. The reconstruction of 2631 draft metagenome-assembled genomes from the global oceans. *Sci. Data* **2018**, *5*, 170203. [CrossRef] [PubMed]

181. Machado, H.; Sonnenschein, E.C.; Melchiorsen, J.; Gram, L. Genome mining reveals unlocked bioactive potential of marine Gram-negative bacteria. *BMC Genom.* **2015**, *16*, 158. [CrossRef] [PubMed]

182. Coutinho, F.H.; Gregoracci, G.B.; Walter, J.M.; Thompson, C.C.; Thompson, F.L. Metagenomics Sheds Light on the Ecology of Marine Microbes and Their Viruses. *Trends Microbiol.* **2018**, *26*, 955–965. [CrossRef] [PubMed]

183. Cameron Thrash, J.; Temperton, B.; Swan, B.K.; Landry, Z.C.; Woyke, T.; DeLong, E.F.; Stepanauskas, R.; Giovannoni, S.J. Single-cell enabled comparative genomics of a deep ocean SAR11 bathytype. *ISME J.* **2014**, *8*, 1440. [CrossRef] [PubMed]

184. Tsementzi, D.; Wu, J.; Deutsch, S.; Nath, S.; Rodriguez, R.L.; Burns, A.S.; Ranjan, P.; Sarode, N.; Malmstrom, R.R.; Padilla, C.C.; et al. SAR11 bacteria linked to ocean anoxia and nitrogen loss. *Nature* **2016**, *536*, 179–183. [CrossRef] [PubMed]

185. Barone, G.; Varrella, S.; Tangherlini, M.; Rastelli, E.; Dell'Anno, A.; Danovaro, R.; Corinaldesi, C. Marine Fungi: Biotechnological Perspectives from Deep-Hypersaline Anoxic Basins. *Diversity* **2019**, *11*, 113. [CrossRef]

186. Orsi, W.D.; Barker Jorgensen, B.; Biddle, J.F. Transcriptional analysis of sulfate reducing and chemolithoautotrophic sulfur oxidizing bacteria in the deep subseafloor. *Environ. Microbiol. Rep.* **2016**, *8*, 452–460. [CrossRef]

187. Lau, M.C.Y.; Harris, R.L.; Oh, Y.; Yi, M.J.; Behmard, A.; Onstott, T.C. Taxonomic and Functional Compositions Impacted by the Quality of Metatranscriptomic Assemblies. *Front. Microbiol.* **2018**, *9*, 1235. [CrossRef]

188. Mitchell, A.; Bucchini, F.; Cochrane, G.; Denise, H.; ten Hoopen, P.; Fraser, M.; Pesseat, S.; Potter, S.; Scheremetjew, M.; Sterk, P.; et al. EBI metagenomics in 2016–An expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res.* **2016**, *44*, D595–D603. [CrossRef]

189. Consortium, G.O. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* **2015**, *43*, D1049–D1056. [CrossRef] [PubMed]

190. Finn, R.D.; Attwood, T.K.; Babbitt, P.C.; Bateman, A.; Bork, P.; Bridge, A.J.; Chang, H.-Y.; Dosztányi, Z.; El-Gebali, S.; Fraser, M.; et al. InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* **2017**, *45*, D190–D199. [CrossRef] [PubMed]

191. Wilke, A.; Bischof, J.; Gerlach, W.; Glass, E.; Harrison, T.; Keegan, K.P.; Paczian, T.; Trimble, W.L.; Bagchi, S.; Grama, A.; et al. The MG-RAST metagenomics database and portal in 2015. *Nucleic Acids Res.* **2016**, *44*, D590–D594. [CrossRef] [PubMed]

192. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; et al. The metagenomics RAST server—A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **2008**, *9*, 386. [CrossRef] [PubMed]

193. Klemetsen, T.; Raknes, I.A.; Fu, J.; Agafonov, A.; Balasundaram, S.V.; Tartari, G.; Robertsen, E.; Willassen, N.P. The MAR databases: development and implementation of databases specific for marine metagenomics. *Nucleic Acids Res.* **2017**, *46*, D692–D699. [CrossRef]

194. Robertsen, E.; Denise, H.; Mitchell, A.; Finn, R.; Bongo, L.; Willassen, N. ELIXIR pilot action: Marine metagenomics? towards a domain specific set of sustainable services [version 1; peer review: 1 approved, 2 approved with reservations]. *F1000Research* **2017**, *6*. [CrossRef] [PubMed]

195. Bork, P.; Bowler, C.; de Vargas, C.; Gorsky, G.; Karsenti, E.; Wincker, P. Tara Oceans. Tara Oceans studies plankton at planetary scale. Introduction. *Science* **2015**, *348*, 873. [CrossRef]

196. Anderson, R.F.; Mawji, E.; Cutter, G.A.; MEASURES, C.I.; Jeandel, C. GEOTRACES: changing the way we explore ocean chemistry. *Oceanography* **2014**, *27*, 50–61.

197. Biller, S.J.; Berube, P.M.; Dooley, K.; Williams, M.; Satinsky, B.M.; Hackl, T.; Hogle, S.L.; Coe, A.; Bergauer, K.; Bouman, H.A.; et al. Marine microbial metagenomes sampled across space and time. *Scientific Data* **2018**, *5*, 180176. [CrossRef]

198. Karl, D.M.; Church, M.J. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat. Rev. Microbiol.* **2014**, *12*, 699–713. [CrossRef]

199. Steinberg, D.K.; Carlson, C.A.; Bates, N.R.; Johnson, R.J.; Michaels, A.F.; Knap, A.H. Overview of the US JGOFS Bermuda Atlantic Time-series Study (BATS): A decade-scale look at ocean biology and biogeochemistry. *Deep Sea Res. Part II Top. Studies Oceanogr.* **2001**, *48*, 1405–1447. [CrossRef]

200. Villar, E.; Vannier, T.; Vernette, C.; Lescot, M.; Cuenca, M.; Alexandre, A.; Bachelerie, P.; Rosnet, T.; Pelletier, E.; Sunagawa, S.; et al. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res.* **2018**, *46*, W289–W295. [CrossRef] [PubMed]

201. Jensen, E.L.; Clement, R.; Kosta, A.; Maberly, S.C.; Gontero, B. A new widespread subclass of carbonic anhydrase in marine phytoplankton. *The ISME Journal* **2019**. [CrossRef] [PubMed]

202. Tangherlini, M.; Miralto, M.; Colantuono, C.; Sangiovanni, M.; Dell' Anno, A.; Corinaldesi, C.; Danovaro, R.; Chiusano, M.L. GLOSSary: The GLobal Ocean 16S subunit web accessible resource. *BMC Bioinform.* **2018**, *19*, 443. [CrossRef] [PubMed]

203. Zhang, H.; Yohe, T.; Huang, L.; Entwistle, S.; Wu, P.; Yang, Z.; Busk, P.K.; Xu, Y.; Yin, Y. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [CrossRef] [PubMed]

204. Tchigvintsev, A.; Tran, H.; Popovic, A.; Kovacic, F.; Brown, G.; Flick, R.; Hajighasemi, M.; Egorova, O.; Somody, J.C.; Tchigvintsev, D.; et al. The environment shapes microbial enzymes: five cold-active and salt-resistant carboxylesterases from marine metagenomes. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 2165–2178. [CrossRef]

205. Han, X.; Hou, L.; Hou, J.; Zhang, Y.; Li, H.; Li, W. Heterologous Expression of a VioA Variant Activates Cryptic Compounds in a Marine-Derived Brevibacterium Strain. *Marine Drugs* **2018**, *16*, 191. [CrossRef]

206. Feder, M.E.; Walser, J.C. The biological limitations of transcriptomics in elucidating stress and stress responses. *J. Evol. Biol.* **2005**, *18*, 901–910. [CrossRef]

207. Tomanek, L. Proteomics to study adaptations in marine organisms to environmental stress. *J. Proteom.* **2014**, *105*, 92–106. [CrossRef]

208. Slattery, M.; Ankisetty, S.; Corrales, J.; Marsh-Hunkin, K.E.; Gochfeld, D.J.; Willett, K.L.; Rimoldi, J.M. Proteomics: A Critical Assessment of an Emerging Technology. *J. Nat. Prod.* **2012**, *75*, 1833–1877. [CrossRef]

209. Domon, B.; Aebersold, R. Options and considerations when selecting a quantitative proteomics strategy. *Nat. Biotechnol.* **2010**, *28*, 710–721. [CrossRef] [PubMed]

210. Han, X.; Aslanian, A.; Yates, J.R., 3rd. Mass spectrometry for proteomics. *Curr. Opin. Chem. Biol.* **2008**, *12*, 483–490. [CrossRef] [PubMed]

211. Yates, J.R.; Ruse, C.I.; Nakorchevsky, A. Proteomics by mass spectrometry: Approaches, advances, and applications. *Annu. Rev. Biomed. Eng.* **2009**, *11*, 49–79. [CrossRef] [PubMed]

212. Calligaris, D.; Villard, C.; Lafitte, D. Advances in top-down proteomics for disease biomarker discovery. *J. Proteom.* **2011**, *74*, 920–934. [CrossRef] [PubMed]

213. Reid, G.E.; McLuckey, S.A. 'Top down' protein characterization via tandem mass spectrometry. *J. Mass Spectrom.* **2002**, *37*, 663–675. [CrossRef] [PubMed]

214. Cristobal, A.; Marino, F.; Post, H.; van den Toorn, H.W.P.; Mohammed, S.; Heck, A.J.R. Toward an Optimized Workflow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89*, 3318–3325. [CrossRef] [PubMed]

215. Chisolm, D.J.; Klima, J.; Gardner, W.; Kelleher, K.J. Adolescent behavioral risk screening and use of health services. *Adm. Policy Ment. Health Ment. Health Serv. Res.* **2009**, *36*, 374. [CrossRef]

216. Forbes, A.J.; Mazur, M.T.; Patel, H.M.; Walsh, C.T.; Kelleher, N.L. Toward efficient analysis of >70 kDa proteins with 100% sequence coverage. *Proteomics* **2001**, *1*, 927–933. [CrossRef]

217. Wu, S.L.; Kim, J.; Hancock, W.S.; Karger, B. Extended Range Proteomic Analysis (ERPA): A new and sensitive LC-MS platform for high sequence coverage of complex proteins with extensive post-translational modifications-comprehensive analysis of beta-casein and epidermal growth factor receptor (EGFR). *J. Proteome Res.* **2005**, *4*, 1155–1170. [CrossRef]

218. Sidoli, S.; Lin, S.; Karch, K.R.; Garcia, B.A. Bottom-up and middle-down proteomics have comparable accuracies in defining histone post-translational modification relative abundance and stoichiometry. *Anal. Chem.* **2015**, *87*, 3129–3133. [CrossRef]

219. Sidoli, S.; Schwammle, V.; Ruminowicz, C.; Hansen, T.A.; Wu, X.; Helin, K.; Jensen, O.N. Middle-down hybrid chromatography/tandem mass spectrometry workflow for characterization of combinatorial post-translational modifications in histones. *Proteomics* **2014**, *14*, 2200–2211. [CrossRef] [PubMed]

220. Swaney, D.L.; Wenger, C.D.; Coon, J.J. Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J. Proteome Res.* **2010**, *9*, 1323–1329. [CrossRef] [PubMed]

221. Gonzalez-Hidalgo, J.C.; Lopez-Bustins, J.A.; Štepánek, P.; Martin-Vide, J.; de Luis, M. Monthly precipitation trends on the Mediterranean fringe of the Iberian Peninsula during the second-half of the twentieth century (1951–2000). *Int. J. Climatol.* **2009**, *29*, 1415–1429. [CrossRef]

222. Taouatas, N.; Drugan, M.M.; Heck, A.J.; Mohammed, S. Straightforward ladder sequencing of peptides using a Lys-N metalloendopeptidase. *Nat. Methods* **2008**, *5*, 405–407. [CrossRef] [PubMed]

223. Domínguez-Pérez, D.; Campos, A.; Alexei Rodríguez, A.; Turkina, M.V.; Ribeiro, T.; Osorio, H.; Vasconcelos, V.; Antunes, A. Proteomic Analyses of the Unexplored Sea Anemone Bunodactis verrucosa. *Mar. Drugs* **2018**, *16*, 42. [CrossRef]

224. Cassiano, C.; Esposito, R.; Tosco, A.; Zampella, A.; D'Auria, M.V.; Riccio, R.; Casapullo, A.; Monti, M.C. Heteronemin, a marine sponge terpenoid, targets TDP-43, a key factor in several neurodegenerative disorders. *Chem. Commun.* **2014**, *50*, 406–408. [CrossRef]

225. Biass, D.; Dutertre, S.; Gerbault, A.; Menou, J.L.; Offord, R.; Favreau, P.; Stocklin, R. Comparative proteomic study of the venom of the piscivorous cone snail Conus consors. *J. Proteom.* **2009**, *72*, 210–218. [CrossRef]

226. Wase, N.V.; Wright, P.C. Systems biology of cyanobacterial secondary metabolite production and its role in drug discovery. *Expert Opin. Drug Discov.* **2008**, *3*, 903–929. [CrossRef]

227. Knigge, T. Proteomics in Marine Organisms. *Proteomics* **2015**, *15*, 3921–3924. [CrossRef]

228. Uniprot_consortium, UniProt: A hub for protein information. *Nucleic Acids Res.* **2015**, *43*, D204–D212. [CrossRef]

229. Sievers, F.; Higgins, D.G. Clustal omega. *Curr. Protoc. Bioinformatics* **2014**, *48*, 3.13.1–3.13.16. [CrossRef]

230. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]

231. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [CrossRef]

232. Congreve, M.; Murray, C.W.; Blundell, T.L. Structural biology and drug discovery. *Drug Discov. Today* **2005**, *10*, 895–907. [CrossRef]

233. De Santi, C.; Ambrosino, L.; Tedesco, P.; Zhai, L.; Zhou, C.; Xue, Y.; Ma, Y.; de Pascale, D. Identification and characterization of a novel salt-tolerant esterase from a Tibetan glacier metagenomic library. *Biotechnol. Prog.* **2015**, *31*, 890–899. [CrossRef] [PubMed]

234. Russell, N.J. Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extrem. Life Under Extrem. Cond.* **2000**, *4*, 83–90. [CrossRef]

235. De Santi, C.; Tedesco, P.; Ambrosino, L.; Altermark, B.; Willassen, N.P.; de Pascale, D. A New Alkaliphilic Cold-Active Esterase from the Psychrophilic Marine Bacterium Rhodococcus sp.: Functional and Structural Studies and Biotechnological Potential. *Appl. Biochem. Biotechnol.* **2014**. [CrossRef]

236. Muhammed, M.T.; Aki-Yalcin, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem. Biol. Drug Des.* **2019**, *93*, 12–20. [CrossRef]

237. Bhattacharya, S.; Bhattacharya, D. Does inclusion of residue-residue contact information boost protein threading? *Proteins* **2019**, *87*, 596–606. [CrossRef]

238. Bowie, J.U.; Luthy, R.; Eisenberg, D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **1991**, *253*, 164–170. [CrossRef]

239. Delarue, M.; Koehl, P. Combined approaches from physics, statistics, and computer science for ab initio protein structure prediction: Ex unitate vires (unity is strength)? *F1000Res* **2018**, *7*, 1125. [CrossRef] [PubMed]

240. Hata, H.; Nishiyama, M.; Kitao, A. Molecular dynamics simulation of proteins under high pressure: Structure, function and thermodynamics. *Biochim. Et Biophys. Acta. Gen. Subj.* **2019**. [CrossRef] [PubMed]

241. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447. [CrossRef] [PubMed]

242. Phillips, J.C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R.D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802. [CrossRef] [PubMed]

243. Hollingsworth, S.A.; Dror, R.O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99*, 1129–1143. [CrossRef]

244. Lohning, A.E.; Levonis, S.M.; Williams-Noonan, B.; Schweiker, S.S. A Practical Guide to Molecular Docking and Homology Modelling for Medicinal Chemists. *Curr. Top. Med. Chem.* **2017**, *17*, 2023–2040. [CrossRef] [PubMed]

245. Rehman, S.F.; Wasim, A.A.; Iqbal, S.; Khan, M.A.; Lateef, M.; Iqbal, L. Synthesis, lipoxygenase inhibition activity and molecular docking of oxamide derivative. *Pak. J. Pharm. Sci.* **2019**, *32* (Suppl. 3), 1253–1259.

246. Goodsell, D.S.; Morris, G.M.; Olson, A.J. Automated docking of flexible ligands: Applications of AutoDock. *J. Mol. Recognit.* **1996**, *9*, 1–5. [CrossRef]

247. Trott, O.; Olson, A.J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461. [CrossRef]

248. Grosdidier, A.; Zoete, V.; Michielin, O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* **2011**, *39*, W270–W277. [CrossRef]

249. Grosdidier, A.; Zoete, V.; Michielin, O. Fast docking using the CHARMM force field with EADock DSS. *J. Comput. Chem.* **2011**, *32*, 2149–2159. [CrossRef] [PubMed]

250. Grosdidier, A.; Zoete, V.; Michielin, O. Blind docking of 260 protein-ligand complexes with EADock 2.0. *J. Comput. Chem.* **2009**, *30*, 2021–2030. [CrossRef] [PubMed]

251. Seashore-Ludlow, B.; Axelsson, H.; Almqvist, H.; Dahlgren, B.; Jonsson, M.; Lundback, T. Quantitative Interpretation of Intracellular Drug Binding and Kinetics Using the Cellular Thermal Shift Assay. *Biochemistry* **2018**, *57*, 6715–6725. [CrossRef] [PubMed]

252. Pagadala, N.S.; Syed, K.; Tuszynski, J. Software for molecular docking: a review. *Biophys. Rev.* **2017**, *9*, 91–102. [CrossRef] [PubMed]

253. Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242. [CrossRef] [PubMed]

254. Manjasetty, B.A.; Büssow, K.; Panjikar, S.; Turnbull, A.P. Current methods in structural proteomics and its applications in biological sciences. *3 Biotech* **2012**, *2*, 89–113. [CrossRef]

255. Chandramouli, K.; Qian, P.-Y. Proteomics: Challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genom. Proteom.* **2009**, *2009*, 239204. [CrossRef]

256. Yalamanchili, H.K.; Wan, Y.W.; Liu, Z. Data Analysis Pipeline for RNA-seq Experiments: From Differential Expression to Cryptic Splicing. *Curr. Protoc. Bioinformatics* **2017**, *59*, 11.15.1–11.15.21. [CrossRef]

257. Cruickshank, D.W. Remarks about protein structure precision. *Acta Crystallogr. Sect. DBiol. Crystallogr.* **1999**, *55 Pt 3*, 583–601. [CrossRef]

258. Bundy, J.G.; Willey, T.L.; Castell, R.S.; Ellar, D.J.; Brindle, K.M. Discrimination of pathogenic clinical isolates and laboratory strains of Bacillus cereus by NMR-based metabolomic profiling. *FEMS Microbiol. Lett.* **2005**, *242*, 127–136. [CrossRef]

259. Baltar, F.; Bayer, B.; Bednarsek, N.; Deppeler, S.; Escribano, R.; Gonzalez, C.E.; Hansman, R.L.; Mishra, R.K.; Moran, M.A.; Repeta, D.J.; et al. Towards Integrating Evolution, Metabolism, and Climate Change Studies of Marine Ecosystems. *Trends Ecol. Evol.* **2019**. [CrossRef] [PubMed]

260. Brierley, A.S.; Kingsford, M.J. Impacts of Climate Change on Marine Organisms and Ecosystems. *Curr. Biol.* **2009**, *19*, R602–R614. [CrossRef] [PubMed]

261. Fuhrer, T.; Heer, D.; Begemann, B.; Zamboni, N. High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2011**, *83*, 7074–7080. [CrossRef] [PubMed]

262. Zampieri, M.; Sekar, K.; Zamboni, N.; Sauer, U. Frontiers of high-throughput metabolomics. *Curr. Opin. Chem. Biol.* **2017**, *36*, 15–23. [CrossRef] [PubMed]

263. Elsayed, Y.; Refaat, J.; Abdelmohsen, U.R.; Othman, E.M.; Stopper, H.; Fouad, M.A. Metabolomic profiling and biological investigation of the marine sponge-derived bacterium Rhodococcus sp. UA13. *Phytochem. Anal.* **2018**, *29*, 543–548. [CrossRef] [PubMed]

264. Amiri Moghaddam, J.; Crüsemann, M.; Alanjary, M.; Harms, H.; Dávila-Céspedes, A.; Blom, J.; Poehlein, A.; Ziemert, N.; König, G.M.; Schäberle, T.F. Analysis of the Genome and Metabolome of Marine Myxobacteria Reveals High Potential for Biosynthesis of Novel Specialized Metabolites. *Sci. Rep.* **2018**, *8*, 16600. [CrossRef]

265. Oppong-Danquah, E.; Parrot, D.; Blümel, M.; Labes, A.; Tasdemir, D. Molecular Networking-Based Metabolome and Bioactivity Analyses of Marine-Adapted Fungi Co-cultivated With Phytopathogens. *Front. Microbiol.* **2018**, *9*, 2072. [CrossRef] [PubMed]

266. Wang, M.; Carver, J.J.; Phelan, V.V.; Sanchez, L.M.; Garg, N.; Peng, Y.; Nguyen, D.D.; Watrous, J.; Kapono, C.A.; Luzzatto-Knaan, T.; et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.* **2016**, *34*, 828–837. [CrossRef]

267. Fabregat, A.; Jupe, S.; Matthews, L.; Sidiropoulos, K.; Gillespie, M.; Garapati, P.; Haw, R.; Jassal, B.; Korninger, F.; May, B.; et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res.* **2018**, *46*, D649–D655. [CrossRef]

268. Caspi, R.; Billington, R.; Fulcher, C.A.; Keseler, I.M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Midford, P.E.; Ong, Q.; Ong, W.K.; et al. The MetaCyc database of metabolic pathways and enzymes. *Nucleic Acids Res.* **2018**, *46*, D633–D639. [CrossRef]

269. Karp, P.D.; Latendresse, M.; Paley, S.M.; Krummenacker, M.; Ong, Q.D.; Billington, R.; Kothari, A.; Weaver, D.; Lee, T.; Subhraveti, P.; et al. Pathway Tools version 19.0 update: Software for pathway/genome informatics and systems biology. *Brief. Bioinform.* **2016**, *17*, 877–890. [CrossRef] [PubMed]

270. Karp, P.D.; Latendresse, M.; Caspi, R. The pathway tools pathway prediction algorithm. *Stand. Genom. Sci.* **2011**, *5*, 424–429. [CrossRef] [PubMed]

271. The_Royal_Society_of_Chemistry, Editorial: ChemSpider–a tool for Natural Products research. *Nat. Prod. Rep.* **2015**, *32*, 1163–1164. [CrossRef] [PubMed]

272. Banerjee, P.; Erehman, J.; Gohlke, B.-O.; Wilhelm, T.; Preissner, R.; Dunkel, M. Super Natural II—A database of natural products. *Nucleic Acids Res.* **2014**, *43*, D935–D939. [CrossRef] [PubMed]

273. Ziemert, N.; Podell, S.; Penn, K.; Badger, J.H.; Allen, E.; Jensen, P.R. The natural product domain seeker NaPDoS: A phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PLoS ONE* **2012**, *7*, e34064. [CrossRef] [PubMed]

274. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.; Bateman, A.; Finn, R.D. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef]

275. Chaleckis, R.; Meister, I.; Zhang, P.; Wheelock, C.E. Challenges, progress and promises of metabolite annotation for LC–MS-based metabolomics. *Curr. Opin. Biotechnol.* **2019**, *55*, 44–50. [CrossRef]

276. Barupal, D.K.; Fan, S.; Fiehn, O. Integrating bioinformatics approaches for a comprehensive interpretation of metabolomics datasets. *Curr. Opin. Biotechnol.* **2018**, *54*, 1–9. [CrossRef]

277. Meier, R.; Ruttkies, C.; Treutler, H.; Neumann, S. Bioinformatics can boost metabolomics research. *J. Biotechnol.* **2017**, *261*, 137–141. [CrossRef]

278. Riekeberg, E.; Powers, R. New frontiers in metabolomics: From measurement to insight. *F1000Research* **2017**, *6*, 1148. [CrossRef]

279. Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B.S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S. Mass-spectrometry-based metabolomics: Limitations and recommendations for future progress with particular focus on nutrition research. *Metab. Off. J. Metab. Soc.* **2009**, *5*, 435–458. [CrossRef] [PubMed]

280. Blasiak, R.; Jouffray, J.-B.; Wabnitz, C.C.C.; Sundström, E.; Österblom, H. Corporate control and global governance of marine genetic resources. *Sci. Adv.* **2018**, *4*, eaar5237. [CrossRef] [PubMed]

281. Bourlat, S.J.; Borja, A.; Gilbert, J.; Taylor, M.I.; Davies, N.; Weisberg, S.B.; Griffith, J.F.; Lettieri, T.; Field, D.; Benzie, J.; et al. Genomics in marine monitoring: New opportunities for assessing marine health status. *Mar. Pollut. Bull.* **2013**, *74*, 19–31. [CrossRef] [PubMed]

282. Duffy, J.E.; Amaral-Zettler, L.A.; Fautin, D.G.; Paulay, G.; Rynearson, T.A.; Sosik, H.M.; Stachowicz, J.J. Envisioning a Marine Biodiversity Observation Network. *BioScience* **2013**, *63*, 350–361. [CrossRef]

283. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef] [PubMed]

284. Prasad, T.S.K.; Mohanty, A.K.; Kumar, M.; Sreenivasamurthy, S.K.; Dey, G.; Nirujogi, R.S.; Pinto, S.M.; Madugundu, A.K.; Patil, A.H.; Advani, J.; et al. Integrating transcriptomic and proteomic data for accurate assembly and annotation of genomes. *Genome Res.* **2017**, *27*, 133–144. [CrossRef] [PubMed]

285. Tuttle, R.N.; Demko, A.M.; Patin, N.V.; Kapono, C.A.; Donia, M.S.; Dorrestein, P.; Jensen, P.R. Detection of Natural Products and Their Producers in Ocean Sediments. *Appl. Environ. Microbiol.* **2019**, *85*, e02830-e18. [CrossRef]

286. Gurevitch, J.; Koricheva, J.; Nakagawa, S.; Stewart, G. Meta-analysis and the science of research synthesis. *Nature* **2018**, *555*, 175. [CrossRef]

287. Pasolli, E.; Schiffer, L.; Manghi, P.; Renson, A.; Obenchain, V.; Truong, D.T.; Beghini, F.; Malik, F.; Ramos, M.; Dowd, J.B.; et al. Accessible, curated metagenomic data through ExperimentHub. *Nat. Methods* **2017**, *14*, 1023. [CrossRef]