The biosynthetic potential of yet uncharted ecosystems in an era of declining biodiversity

Maarten Boneschansker

March 2023

NP's are defined as chemical (small molecule) products derived from nature. Examples include caffeine, nicotine and taxol from plants, estradiol and conotoxins from animals, mycotoxins from fungi, and antibiotics like penicilline and ampicilline or antiparasiticals like avermectin from bacteria. NP's are also sometimes loosely defined as any chemical produced by an organism that is not used in primary metabolism, but is a benefit to overall fitness, the term secondary metabolite is therefore often used interchangeably. There are many organisms that produce NP's spanning all three domains of life, but most are produced by plants and micro-organisms like fungi and bacteria. Animals do produce secondary metabolites; toxins and venoms are common NP's in animals, some being used clinically, and hormones like estradiol and oestrogen are also NP's and are used by humans on a vast scale. Plants however are the undisputed champions of NP diversity with around 145,000 chemicals derived from plants described in literature. Especially plants and fungi tend to form symbiotic networks and it is expected that many interesting molecules can be found at the intersection between these domains. Bacteria and fungi are the main producers of NP's in medicine though. For example penicillin - the first antibiotic - from the *Pennicillium* fungus. And the bacterial genus *Actinomycetes* that has provided the basis for almost 2/3rd of clinical antibiotics in use today [1].

The story of the discovery of penicillin is well-known [2], the accidental fungal infection of a petri dish, less known is that it took another serendipitous discovery, this time on a rotten cantaloupe by an assistant named Mary Hunt to reveal a strain of *Pennicillium* that produced six times as much of the drug. Even less well-known is that the original discoverers and Nobel-prize recipients did not patent penicillin production as they thought it unethical, their American industrial collaborators were not that morally burdened and got incredibly rich.

Serendipitous discovery of NP's by scientists and laymen/women has been one the main sources of novel drugs for most of modern history. Discoveries of antibiotics on rotten fruit or anti-parasiticals from a Japanese professors' local golf-course soil - the late Satoshi Ōmura, an avid golfer, discovered the anti-river blindness compound class of the Avermectins from a soil sample he took while golfing - have saved many millions, but essentially have come down to looking to nature really well.

This essay argues that we still depend on serendipitous NP discovery, that new discoveries can be found in diverse habitats, and that as nature is under pressure and biodiversity is lost at a rapid tempo, we are also quickly losing large amounts of biosynthetic potential forever. This is evermore pressing as new NP's are not only able to accelerate world development, they are also needed to prevent our development backsliding.

Natural Products

Natural products (NP's) have been used for millennia by humans, but only in recent decades has the rate of discovery and use really accelerated. Traditional medicine from all over the globe is often NP based and indeed many traditional applications are now called modern medicine. An example is aspirin from the bark of the willow tree, although now chemically produced, a potent natural painkiller. Another example is Taxol, an anti-cancer agent, from the yew tree. NP's spanning all domains of life have found many diverse applications.

One use of NP's that stands out is their use as antibiotics. Antibiotics have saved countless lives and propelled humanity into an unprecedented disease free era, which has allowed us to build prospering and stable societies, relatively free of pandemics. Bacteria are the second largest killer of humans throughout history - only second to viruses. It was the smallpox and measles viruses that decimated the new world, but during multiple epidemics in (pre)medieval times up to 50% of the European population was lost due to Bubonic plague, a bacterial infection. Also, even though the Spanish Flu of 1918-1920 - caused by the H1N1 influenza virus - caused the most deaths in the last century, easily more than the first world war, it was probably secondary bacterial pneumonia that was the direct cause of death in most cases [4]. In recent times, covid-19 caused millions to perish worldwide, with sepsis - bacterial infection of the blood - being a main contributor to mortality. Unsurprisingly, one of the main treatments of covid-19 is antibiotics [5]. It is thus not only bacterial infection that kills readily directly, bacterial secondary infection often underlies lethality in the case of a viral infection. It could thus be argued that it is bacteria that are actually humanities largest killer.

This makes it noteworthy that today death by bacterial infection is relatively rare in high-income countries and those that do die are often old or weakened by another affection [6]. A great achievement this may be, our weapon against one of our greatest enemies is running out as bacteria have started to develop antimicrobial resistance (AMR) on a large scale, with as much as 4,95 million deaths associated with AMR in 2019 [7]. And this is expected to rise sharply. AMR is estimated to be one of the major threats to modern civilization [8]. It is also disturbing to imagine a new viral pandemic spreading the globe, but this time accompanied by resistant bacteria. It is not known exactly how many deaths from covid-19 can be attributed to AMR, but it is likely, since for example the methicilin resistant *Staphylococcus aureus* (MRSA) bacterium is now common in hospitals, that it is appreciable.

It might then be an unpleasant surprise to learn that pharmaceutical com-

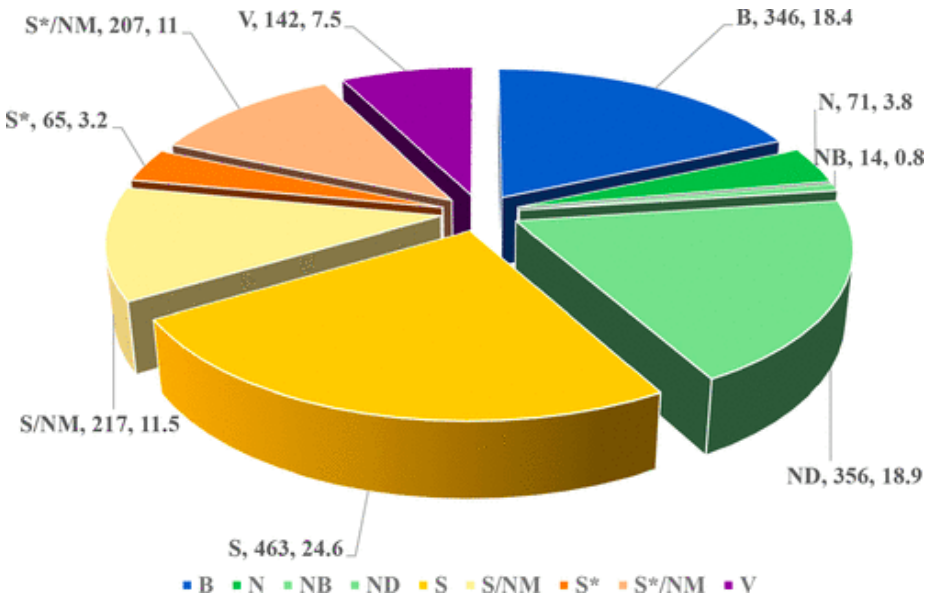

Figure 1: All newly approved drugs from and their respective sources spanning 1981 to and including 2019. “B”: Biological; usually a large (¿50 residues) peptide or protein either isolated from an organism/cell line or produced by biotechnological means in a surrogate host. “N”: NP “NB”: NP “Botanical” (in general these have been recently approved) “ND”: Derived from a NP and is usually a semisynthetic modification. “S”: Totally synthetic drug, often found by random screening/modification of an existing agent. “S*”: Made by total synthesis, but the pharmacophore is/was from a NP “V”: Vaccine. “NM”: NP mimic. Note that the only small molecule drug category that is not a vaccine that is not (derived from) a NP (S) is a minority with only quarter of new compounds. Image taken from Newman et al. [3].

panies are currently not researching new antibiotics on a mass scale. Most of those that are being researched are close to current ones and novel ones do not target critical AMR species [9]. We are in fact set for disaster, as the incoming wave of AMR is not matched by antibiotic research investment. Reasons for this are legion but one important one is the relatively low discovery rate of new compounds [10].

A new genomic golden age

From the 1940's up until the 1970's it had been relatively easy to pluck the low hanging NP fruit from culture broths, leading to 1000's of newly described compounds. Discovery consisted of culturing samples from the environment, often soil, and screening for bacterial growth inhibition or, in the case of cancer-screenings, cell death. This approach was wildly successful, so successful that at some point this 'golden-age' of NP discovery came to an end when most accessible compounds had been discovered. NP discovery was then sidelined during the advent of combinatorial chemistry in the 1990's. Combinatorial chemistry describes a process where a lead compound is varied at multiple sites, for example 3 variable sites with 10 variations leads to a 1000 total compounds screened. This however was a dead end, with no viable drugs reported at all. As it turns out, bio-active compounds are not randomly spread through chemical space, rather they are clustered in islands of bioactivity, which are hard to find. In practice combinatorial chemistry only scratched the surface of possible compounds, the majority of which showed no bio-activity at all. It is thus the millions of years of natural evolution that provide a crucial guide to these islands of bioactivity that we should harness by turning towards NP's again. Indeed pharmaceutical companies and academics alike have seen renewed interest in NP research [11, 12, 3].

Even though the golden age method of broth fermentation had exhausted its initial potential, genome sequencing has revealed that there is still plenty of so called biosynthetic potential in NP discovery [1]. Since the early 2000's we have entered the genomic age and studies into the genome of even well-studied NP producers came to the conclusion that the reservoir of yet undiscovered NP's is huge.

Genome Mining

The widespread availability in combination with improvements in bioinformatic software and hardware led to the creation of an entirely new field in NP discovery: genome mining, where an organism is not screened by phenotype, but by genotype. This approach has strong advantages over traditional approaches, the first of which is to overcome the limited expression of secondary metabolites.

In fact most producers of NP's only do so under very specific circumstances, which is why they never showed up in the culture broths of the 1970's [13, 14]. Genome mining circumvents this because even if an organism does not produce a compound, genes encoding it are still present and can be discovered. Another

a

	Genera in NCBI	Example of well-studied genus	Specialized metabolites isolated from this genus	Extrapolated number of specialized metabolites
land plants	15,573	*Brassica*	349	~5,400,000
algae	2,206	*Laurencia*	902	~2,000,000
fungi	716	*Aspergillus*	2,034	~1,500,000
bacteria	3,980	*Pseudomonas*	318	~1,300,000
sponges	499	*Dysidea*	515	~250,000
cnidarians	1,152	*Sinularia*	807	~900,000
arthropods	41,922	*Drosophila*	104	~4,400,000
other invertebrates	9,706	*Caenorhabditis*	52	~500,000
vertebrates	9,838	*Dendrobates*	142	~1,400,000
			Total:	~18,000,000

b

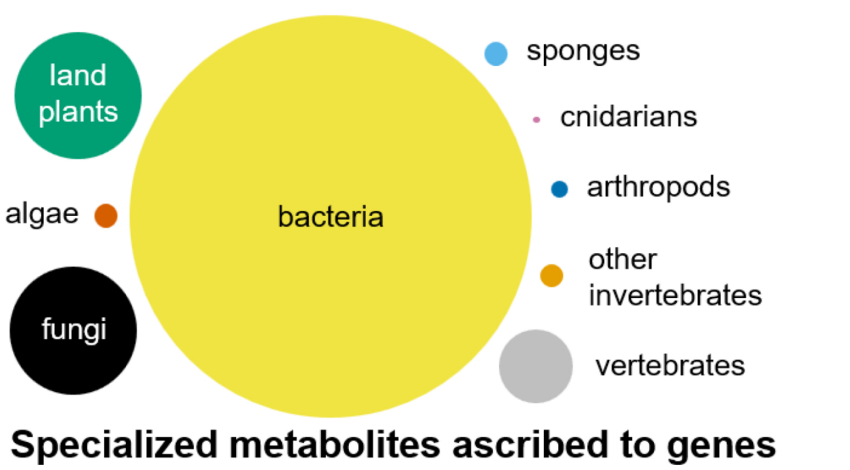

Figure 2: a) An extrapolation of specialized metabolites out there based on known genera and NP's. Note that the number of genera is expected to rise sharply, especially for microbes, so 18,000,000 might even be an underestimation. b) Most specialized metabolites are ascribed to bacterial origin, this expected to change as plant and especially fungal bioinformatic screening tools come of age. Image taken from Medema et al. [1].

problem with culturing is that bacteria that feel at home in culture broths will dominate the environment, whereas in the original sampling environment the composition of species might have been completely different. This can be overcome by sensitive sequencing techniques that catch even the lowest amount of DNA.

Thirdly, the genome mining approach inherently connects any discovered molecules to their biosynthetic genes, allowing for heterologous expression and bulk production. Especially in microorganisms, genes encoding secondary metabolites are clustered together on the genome and follow specific patterns of expression. This makes it relatively easy to detect, isolate, and express such a cluster of genes and isolate the resulting product. This process, called dereplication, is a significant hurdle in regular NP discovery.

The modern field of genome mining is growing at a rapid pace and works together narrowly with phenotype based discovery efforts. The field is rich with tools and databases and in more recent years, efforts have been made to bridge the gap between specialized bioinformaticians and lab-researchers. One tool of great importance is antiSMASH [15], a software package made to run from an easy-to-use web interface, but with the latest in genome mining tools behind it. Thus, with the public availability of high end bioinformatic tools, any wet lab biologist can scan her/his genome for interesting secondary metabolites with ease.

The other way around, if a compounds is detected in the genome, there are plenty of resources around to try and express it in either it's natural host or heterologous. Databases like actinobase provide detailed instructions on how to best handle Streptomyces bacteria for example, some of the most avid producers of NP's. A golden standard in NP research is now the genome mining of sequences, possibly rare, and then validating the most promising compounds in the lab. Some nice examples of this are given in references [16] and [17].

Whereas most of the exploratory power and novelty once came from culturing, the wet lab, this is now more the domain of the dry lab biologists. Most genome mining tools do indeed prioritize novelty in a great manner, sometimes even at the cost of accuracy. Only when the genome mining approach provides a novel and likely interesting NP candidate does the wet lab come in nowadays, whereas culturing used to be the first step, not the last. This reverse in order of things, first sequence and analyze and then culture instead of the other way around, is only expected to grow as the influx of genomic data and improvement in bioinformatics will continue to provide orders of magnitude more leads than can be practically validated. In fact, two main bottlenecks in the current sequence to product pipeline are actually prioritization efforts within genome mining and low throughput at the lab end.

It is expected that with increased feedback and lab-validation of compounds the bioinformatic tools can be sharpened, especially those based on machine learning as currently learning data is relatively rare. Golden standard data on which machine learning algorithms can be trained are especially rare, but increasing steadily. An example is the leading database for NP's called MIBiG, which contains minimal information on gene clusters, including products, genes,

modifications and raw sequences. The recent 3,0 version has been a community effort which not only added new entries, but also updated existing ones to a higher quality. However, even as MIBiG is important and a community standard, its current 2,502 entries are dwarfed by the estimated millions of suspected gene clusters. Every round of updates of MIBiG and similar databases will accelerate discovery though and hopes are that within the decade MIBiG and others will contain orders of magnitude more data.

Other improvements on the practical side include universal expression tools to circumvent the problem of non-expressed gene clusters. Although still relatively new, bioinformatic approaches using machine learning and large scale data analyses have led to new methods of promoter designs which could solve the cryptic gene expression problem [18]. Better designed promoters are more precise and allow researchers to accurately and strongly express gene clusters in their respective native organisms.

Another important approach to expression of gene clusters is heterologous expression in other life forms than the native gene cluster host. Heterologous expression is a keystone in biological research and bio-industry and especially crucial if a gene cluster for a lead compound is to be expressed at large scales. The relatively recent and much discussed Crispr-Cas9 gene editing is transforming heterologous expression in that it is now cheaper and easier to edit genes than ever and this is only expected to increase. Moreover, new applications of the Cas9 system are piling up in a rapid pace and the impact on biology is expected to grow into other fields such as expression and regulation quickly [19, 20]. Even whole synthetic life forms have been created, a major milestone, which might one day prove to be very useful in creating organisms specifically targeted at expressing biosynthetic gene clusters [21].

Increases in tool accuracy and validation methods will provide useful feedback on itself and to new tools, but one of the main challenges in NP discovery is still how to generate truly novel results. To go back to a golden age of discovery requires true novelty, compounds nothing alike has ever been seen before are the holy grail and this job, once up to the culture biologists, is now up to bioinformaticians. As things currently stand machine learning tools are gaining popularity in the field quickly for their ability to detect subtle patterns in data. Indeed some approaches have revealed novelty inaccessible to classical genome mining tools, albeit often at the cost of decreased accuracy [17]. However, machine learning has its fundamental drawbacks too in that algorithms will always have to be trained on previously known data, which makes these approaches weaker at detecting something that is not remotely like anything known before. It is therefore also crucial that keep trying to construct theoretical frameworks on how and why secondary metabolites work. A recent example of this is from Wisecaver et al., where the definition of a gene cluster is expanded to transcriptomic space, which leads to the discovery of novel compound classes [22].

Further down the line it is expected that peptide prediction tools will increase in accuracy so much that functional prediction based on structure becomes possible. Tools like AlphaFold can help, but maybe also language models trained on biological sequence as the recent chat-GPT has shown to be possible. For now

however, AlphaFold is not useful for peptides and secondary metabolites and so far functional prediction of compounds is possible, but still quite inaccurate [16]. Thus, as a torrent of genomic data and NP predictions is expected, lab validation will remain the golden standard for years to come. This is especially true for de novo discoveries, which are of course the most interesting.

Undescribed species dominate

The huge potential of genome mining relies not only on the fact that known microbial species can produce much more NP's than previously thought, the vast majority of microorganisms have never been cultured and characterized thoroughly, and thus no metabolic profile or genome sequence is available [23]. It is estimated that from a 100 bacterial cells seen under the microscope in any sample, only one can be cultured [24]. Recently, new techniques like (meta)genomic long-read sequencing [25, 26], single cell sequencing, and more sophisticated culturing methods have allowed for the elucidation of these rarer microorganisms and their potential NP's. A recent report estimates that about 25% of metagenomic data belongs to undescribed phyla. Moreover this *microbial dark matter* is often unrelated to known species. As similar species often share similar NP's this novelty in phylogeny promises novelty in NP's too [27].

Genomic techniques have also enabled scientists to reconstruct the tree of life, with a stark increase in microbial diversity as compared to Eukaryotic macroscopic life [28]. Bacteria dominate the Earth in terms of diversity, then Archaea, third come the Eukaryotes. That is not to say that there are not plenty of Eukaryotes yet undiscovered. It has been estimated that the current number of generally accepted species of fungus of 120,000 is a gross underestimation and that the actual number of fungal species lies somewhere between 2,2 and 3,8 million species [29]. It is thought that there are roughly 1,5 million animal species, about 80% of which are Arthropods. It is further estimated that there are roughly 8,7 million Eukaryotes on the planet, although this number is probably an underestimation as the increased number of estimated fungal species mentioned above was not incorporated in this paper by Mora et al. [30]. Such an estimate could not be made for prokaryotes because as the authors note there seems to be no sign of decrease in discovery rates from which an asymptotic limit could be estimated. Indeed, another study estimated the number of microbial species of all three domains of life at 1 trillion species. [31] Interestingly, while only a minority of animals, fungi, and especially prokaryotes are known, plants are relatively well described according to Mora et al. Even so, as plants harbour a chemical diversity many times over that of other Eukaryotes per species, a smaller number of undiscovered species still holds great biosynthetic potential.

It is also worth noting that most animals and plants have their own unique microbiome/rhizome so any increase in animal species might very well be accompanied by yet another increase in microbial diversity. Animals tend to live in symbiosis with bacteria for food digestion, while plants often partner up with fungi and bacteria in their root systems. Especially with fungi this symbiosis can go so far that the fungal mycelium is actually partially located within(!)

the plant tissue.

A recent overview of NP discovery and genome mining specifically put the estimated potential compounds at 18,000,000 [1]. However, this was based on genera currently represented in genome databases. As the number of yet undescribed species is estimated to be several factors higher than currently described species - with the exception of plants, the amount of potential compounds could be 18,000,000 many times over. It is not only the untapped biosynthetic potential of known species that contributes to this number, it is also yet undiscovered species, which can mainly be found in four theoretical biomes.

The rare biosphere

Microbial dark matter is often found in the *rare biosphere* [32], where uncommon organisms reside, essentially in plain sight but hard to detect because of their low frequency. Some of these organisms might only be encountered once in a 100,000 cells. Interestingly it is exactly these organisms that are of special interest to NP discovery as these organisms are more likely to be involved in specialty roles in the ecosystem, requiring specialty secondary metabolites. Rare organisms do not always play a vital role in an ecosystem, they tend to be metabolically inactive, very small, and sparsely populated, which makes them resistant to cellular breakdown, grazing, and viral infection respectively. In other words, even organisms that do not play an active role can still linger as their population decline is inversely related to the part they play in an ecosystem. The result is that most microbial ecosystems are dominated by a few dozen species involved in carbon and nitrogen cycles, with up to two orders of magnitudes of species more in the rare biosphere within a single sample. Plenty of novel microbes have even been found in the most well-studied microbial ecosystem known: the human gut [33]. They have also been described in soil [34], another well-studied ecosystem. But also in relatively underexplored biomes like the (deep) ocean [16, 35], deep soil (deep biosphere) [36, 37, 38], and extreme environments [39].

The deep biosphere

Animals such as nematodes have been shown to thrive deep underground, but they are a minority in terms of biomass. Plants contribute a significant part of the biomass underground, but only down to the upper layer of the deep biosphere. It is actually bacteria and fungi that dominate this biome, all the way down to 5km. Interestingly also in extreme environments without oxygen, under high pressure, and up to 100 degrees Celsius. The total amount of biomass of microorganisms in the deep biosphere is actually thought to be much larger than regular terrestrial and marine biospheres [40]. The microbiome of the earth is also very diverse and samples taken from different sites often have a unique fingerprint [41, 42, **?**, **?**]. This means that there is probably huge biosynthetic potential to be found underground and makes a case to start genomic sampling from drill wells and mines. The separation between underground and above ground life is not only spatial, but also temporal. Underground life moves at

a different pace and cells can be 1000's to a million years old, as energy is so sparse that metabolic rates are only a fraction of those above ground [43]. The energy sources also differ, the main energy source comes from rocks directly by reduction of hydrogen rich compounds like methane and sulfates, in essence, these bacteria eat rocks [44, 45]. Perhaps surprisingly, most microbes in these environments are not specialists to underground life, only a minority are primary producers, while the majority live off those primary producers and have 'normal' metabolism based on oxidation. It is thought that many organisms at this depth were once above ground species and that they seeped down with ground water into aquifers. It is possible that these organisms harbour antibiotic genes that were common in a different era, antibiotics to which bacteria are not resistant today. Underground life moves at a different tempo and is relatively unaffected from what happens above, this can not be said from most other biomes on Earth. Life under gletsjers and antarctic life also moves very slowly, but is these biomes are under great threat from climate change. It might even be argued that these ice-bound life forms should be sampled as soon as possible as ice mass on earth is being lost at a rapid pace.

The ocean biosphere

The ocean covers about 2/3rd of the surface of the Earth and is volume-wise only second to the deep Earth biome in size. It is also however relatively empty and aside from some hotspots has the bioproductivity of a terrestrial desert, as a popular pop song once put it 'the ocean is a desert with its life underground'. Indeed terrestrial life that dominates Earth by biomass, mostly in the form of plants [40]. Most marine animals in the non-coastal zones are more defined by lattitude than longitude, this is because the big oceans can be considered one big biome. Notable plants are seaweeds and kelp, but most of the photosynthesis is performed by marine algae and cyanobacteria, both known to be NP producers. Together these two groups of organisms are called phytoplankton and they are the basis of the food chain in many of the worlds oceans. A collection of animals called zooplankton feed on phytoplankton, which is the basis for most larger marine life. Microbial sampling of the ocean's open water columns has recently revealed plenty of biosynthetic potential, as well as a great diversity and novelty in ocean microbe ecosystems [16, **?**]. Most of the biosynthetic potential from the ocean comes from free-floating organisms like dinoflagelates, cyanobacteria, and algae, but one group of animals stands out: sponges. One of the oldest types of animals and relatively simple in structure they are the source of 30% of ocean NP's. Whether they are the sources directly, are in symbiosis with bacteria, or that they simply filter feed NP producing organisms and are thus DNA-hotspots [46], is under discussion though. Nevertheless, sponges are a rich source of NP's [47].

The terrestrial biosphere

Rich as the ocean, rare, and deep biospheres may be, they are all dwarfed by the mass and diversity of terrestrial life [48]. From the earlier mentioned 8,7 millions estimated species a minority of 2,1 million are marine [49]. Interestingly, species diversity also tends to increase from the poles to the equator. An effect that is attributed to multiple factors, but temperature being a dominant one [50, 51]. This is a relatively recent phenomenon, which indicates that it is not anchored to the planets orbit or other permanent factors, but rather environmental, which means this could change quickly due to, amongst others, climate change [52, 53]. Today the tropics are the most diverse habitats on Earth for plant and animal life [54]. Microbial life follows a similar pattern in latitudinal species diversity distribution as other organisms, albeit a bit weaker, thus the tropics are also hotspots in microbial life. This is not only due to increased soil- and air-borne microbe diversity, but also because of host-associated diversity as there are more species of plants and animals in the tropics, with unique accompanying rhizomes and microbiomes [55]. Species diversity on a macroscopic scale is thus a useful indicator of diversity on microscopic scale and the biodiversity in the tropics could be yet another hotspot of biosynthetic potential.

In summary, NP's have been and are important sources of useful compounds for humans, especially and most pressingly in antibiotics. The *golden age* of discovery is over, but a new one awaits thanks to the advent of genome mining. Two factors contribute to the potential success of genome mining: the first is deeper and higher quality (meta) genome sequencing, the second is the sequencing of novel DNA, as genomic novelty promises biosynthetic novelty. This novelty can be found in many places, right under our noses in common genomes and well-studied biomes, elucidated by better sequencing, but also crucially in underexplored habitats like the ocean, the deep Earth, and terrestrial biodiversity hotspots, elucidated by exploration. Finally, the fact that culturing alters species composition and changes metabolic profiles, and the fact that many species remain unculturable or undescribed, makes a strong case for *in situ* (meta)genomic sampling.

The sixth mass extinction event

Natural diversity is under threat, this has been dubbed the 'other natural crisis' next to climate change, though both are very much intertwined. There is discussion about whether the current events of species loss are truly the beginning of an anthropogenic sixth extinction event and if so what to do about it, but it is sure that we are losing diversity at a rapid pace. It is speculated that the relative loss of species is as much of 13% since 1500AD, with a strong acceleration since the industrial revolution. Today, it is thought that the amount of species lost per year exceeds the number of newly reported species, thus we are actually losing known extant species faster than they are added. Even very high estimates of natural species extinction rates - as species naturally go extinct in

abrupt manners too - would only add up to about 4 species lost per year for mammals. It is estimated that we have so far averaged 300-520 species lost per year since 1500AD [56, 57].

Islands have seen the most species loss of all habitats, some islands have only been inhabited for a couple thousands of years, but have already lost significant proportions of their species diversity. The relatively small time scale of extinction events on those islands has allowed for modelling of species extinction and the introduction of a concept called 'extinction debt'. The idea being that before a species finally goes extinct, it lingers with at least a couple times a species maximum age [58]. In practice this means that even though species are still present, if their habitat is destroyed they are doomed for extinction. Worryingly it also exactly the hotspots of diversity in the world that harbour most endemic species and separate ecosystems, which means they can be considered islands and are therefore extra vulnerable. Only 25 of such hotspots exist on the planet and some argue that focus of conservation efforts should be on these areas as we simply cannot save all anymore [59].

All the previous five mass extinction events are clearly visible in the marine fossil record, while this presumed one is more land-based. This is sometimes referred to as evidence against the existence of a mass extinction event, but it is more likely that the one we are living through is unique because it is happening at such unprecedented speeds. It also follows that since we humans are a terrestrial species, the sixth mass extinction caused by us will be a terrestrially based. For example, 1/3 of land today is unwilded and used for agriculture or building, the ocean floor and pelagic zones are largely untouched.

Anthropogenic change; Homo Sapiens' footprint

The cause of this massive Anthropogenic extinction is simple: us. The human population has grown exponentially and is thought to reach 11 billion by 2050. Even though were are not dominant in terms of biomass, from an ecological perspective we are wildly successful, no species has ever contributed so much to global environmental change as we have. In fact, today, it is exactly our changing of the environment that is killing of species as destruction of habitat is the main contributor to extinction (and future extinction). Land-use for agriculture has by far had the greatest impact on species diversity so far, about 32% of land is used for agriculture. What's more is that a global shift in the North towards reforestation and a global shift in the South towards deforestation and agricultural expansion has caused considerate net biodiversity loss as recently regenerated Northern temperate forests are not nearly as rich as yet untouched tropical rainforests. Cynically, this means that reforestation efforts in the West might have the net opposite effect as intended [60, 61].

Some species have been hunted to extinction, but this is a rare minority. Examples include the dodo, Tasmanian tiger, and great auk, often driven to extinction by secondary effects like introduction of rats from ships to which the original inhabitants were defenseless. Pollution and spreading of exotic species have also had serious impacts, but mostly localized. It is also worth noting

that climate change is yet to make a significant impact on species diversity. As the world heats up, oceans will acidify and land mass will dry, undoubtedly leading to further extinction events, but it should be noted that most of current extinction has been achieved by habitat destruction directly. Perhaps current carbon dioxide levels and predicted heating are in fact giant extinction debts and will, even if we stop habitat destruction right now, push us into the sixth extinction event anyway [62].

The largest cause of habitat destruction on land has been agriculture, but even where we do not directly destroy the environment, the ocean is still subject to ecological collapse. Species diversity collapse has not been documented (yet) in the oceans and the amount of fish considered extinct remains limited. That does not mean that the ocean remains unaffected and fish stock collapse has been widely acknowledged. acidification, hypoxic regions and warming of the ocean are all factors that have contributed to a depletion of pelagic fish, but overfishing seems to be the major culprit [63, 64]. Whereas fish are rapidly declining in numbers, plankton life seems to be adapting to new conditions rather than to go extinct. Species composition is expected to change in for example diatoms, that need calcium carbonate for their shells, which will be harder to come by in a more acidic ocean [52, 65, 66]. This is not unexpected as plankton and microbial life has a turnover rate of days, not years, so the current climate and biodiversity crises relatively span a lot more generations for microbial life than for animal life, which offers species more opportunity to adapt.

Microbial biodiversity

It is clear that macrobial biodiversity is in crisis, as much as a potential six mass extinction event, and this is not even due to climate change yet, as these effects will come further down the road, but due to direct human activities such as hunting and polluting, but mostly agriculture. Microbial diversity is orders of magnitude greater than macrobial diversity but it unknown whether this diversity is increasing or decreasing [67]. In part we could not have known as only recent techniques like single-cell and metagenome sampling have even elucidated the existence and frequency of microbial species, in another part we may have conveniently overlooked them. Even though there is no consensus to the direction microbial species diversity, based on current knowledge of the relationship between macrobial and microbial life and we can make an estimation.

First of all there is probably very little effect of Anthropogenic human activity on deep biome microbes. They are shielded from the environment at a physical scale and often not connected to above ground ecosystems anyway. It is exactly for these reasons that astrobiology is interested in these microbes as it is those that could have survived for eons under the surface of certain planets, moons, or even Mars.

Host-borne microbes both marine and terrestrial are at serious risk of decrease in biodiversity, as this is directly coupled to host diversity. With every mammal, bird, and plant species lost, we lose a unique corresponding microbiome or rhizome respectively.

Free-living microbes are probably hugely affected by human activity. As land has changed greatly, so have probably the accompanying microbes in the soil, air, and water. The warming and acidification and also species loss in the ocean will also probably have a large effect on microbial species diversity. What this effect will be is harder to say though, microbes have turnover rates orders of magnitude smaller than macrobial life forms and can thus adapt very quickly (with the exception of deep Earth and those living in icy worlds). Also, because microbes can exist in a fastened spore-like state for long times - the rare biosphere - the extinction of entire species is not as likely as with macrobial life. Rather it is more likely that as conditions change, the composition of free-living microbial ecosystems will change with it, rather than diminish. This can already be seen in plankton.

Thanks to recent techniques like next generation sequencing and genome-mining, NP's are re-emerging as a rich source of chemical potential. Using nature as an inspiration is a tremendously exciting idea, something known since the 70's, but recent decades have revealed that this excitement - with a 1000 compounds discovered that changed the world - has even been a major under-estimation of Nature' s potential as a guide. However, in a cruel twist of our own devise, it is exactly Nature, our Earth biome that is rapidly lost due to our activities, potentially taking all the biosynthetic potential we had recently rediscovered with it. Yet, not all is lost, due to the way microbial species can keep themselves in a fastened rare state for many years there could be rare instances of almost extinct rare microbial life everywhere. This makes a strong case for ultra-deep sequencing efforts and deep analysis of samples to be able to catch the rare biosphere. At the same time there are some resilient environments like deep ocean and deep Earth biomes who will probably linger even long after humans have gone extinct. Thus, areas we have already destroyed should be sampled with a focus on quality because remnants of long-gone diversity landscapes might still be present. At the same time microbial life coupled to macrobial life probably faces a similar fate: rapid extinction of many. Therefore it is also crucial that we direct efforts towards sequencing as many unique microbiomes as possible before host-species are lost forever. This applies to bacteria in the guts of animals, but also very much to fungi in the root systems of plants. Unfortunately it is all but inevitable that we will lose a tremendous amount of biodiversity, as has already happened, and the effects of climate change are yet to come. The future looks very bleak from a macrobial diversity perspective. In order to maximize microbial diversity retention and thus their biosynthetic potential, we should make sequencing standard-practice wherever conservationists are active. Especially biological hotpots need to be sequenced as soon as possible, otherwise we run the risk of losing many microbes we do not even know yet, and will never know. Fortunately this is doable. New genomic sequencing techniques are increasing in ease-of-use and decreasing in cost at astonishing rates and the first long-read pocket devices are already on the market. Many conservationists and other trying to chart biodiversity networks are in place, they should be expanded. Furthermore, these networks should be equipped with sequencing techniques so they can extend their activities into the

microbial world and save its biosynthetic potential.

References

[1] Marnix H. Medema, Tristan de Rond, and Bradley S. Moore. Mining genomes to illuminate the specialized chemistry of life. *Nature Reviews Genetics*, 22(9):553–571, Sep 2021.

[2] Unknown author(s). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of b. influenzæ. *British journal of experimental pathology*, 10(3):226—236, June 1929.

[3] David J. Newman and Gordon M. Cragg. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Products*, 83(3):770–803, 2020. PMID: 32162523.

[4] David M Morens, Jeffery K Taubenberger, and Anthony S Fauci. Predominant role of bacterial pneumonia as a cause of death in pandemic influenza: implications for pandemic influenza preparedness. *J Infect Dis*, 198(7):962–970, October 2008.

[5] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, Lulu Guan, Yuan Wei, Hui Li, Xudong Wu, Jiuyang Xu, Shengjin Tu, Yi Zhang, Hua Chen, and Bin Cao. Clinical course and risk factors for mortality of adult inpatients with covid-19 in wuhan, china: a retrospective cohort study. *The Lancet*, 395(10229):1054–1062, Mar 2020.

[6] WHO. The top 10 causes of death. 2019.

[7] Murray et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet*, 399(10325):629–655, Feb 2022.

[8] World Health Organization. Global antimicrobial resistance and use surveillance system (glass) report 2022. 2022.

[9] Antibacterial agents in clinical and preclinical development: an overview and analysis. geneva: World health organization; 2021. licence: Cc by-nc-sa 3.0 igo.

[10] Roberto Kolter and Gilles P. van Wezel. Goodbye to brute force in antibiotic discovery? *Nature Microbiology*, 1(2):15020, Jan 2016.

[11] Philip Hunter. Harnessing nature's wisdom. turning to nature for inspiration and avoiding her follies. *EMBO reports*, 9(9):838–840, Sep 2008. 18762775[pmid].

[12] Leonard Katz and Richard H Baltz. Natural product discovery: past, present, and future. *Journal of Industrial Microbiology and Biotechnology*, 43(2-3):155–176, 03 2016.

[13] Essaid Ait Barka, Parul Vatsa, Lisa Sanchez, Nathalie Gaveau-Vaillant, Cedric Jacquard, Hans-Peter Klenk, Christophe Clément, Yder Ouhdouch, and Gilles P. van Wezel. "taxonomy, physiology, and natural products of actinobacteria". *Microbiology and Molecular Biology Reviews*, 80(1):1–43, 2016.

[14] Christian S. Riesenfeld, Robert M. Goodman, and Jo Handelsman. Uncultured soil bacteria are a reservoir of new antibiotic resistance genes. *Environmental Microbiology*, 6(9):981–989, 2004.

[15] Kai Blin, Simon Shaw, Alexander M Kloosterman, Zach Charlop-Powers, Gilles P van Wezel, Marnix H Medema, and Tilmann Weber. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research*, 49(W1):W29–W35, 05 2021.

[16] Lucas Paoli, Hans-Joachim Ruscheweyh, Clarissa C. Forneris, Florian Hubrich, Satria Kautsar, Agneya Bhushan, Alessandro Lotti, Quentin Clayssen, Guillem Salazar, Alessio Milanese, Charlotte I. Carlström, Chrysa Papadopoulou, Daniel Gehrig, Mikhail Karasikov, Harun Mustafa, Martin Larralde, Laura M. Carroll, Pablo Sánchez, Ahmed A. Zayed, Dylan R. Cronin, Silvia G. Acinas, Peer Bork, Chris Bowler, Tom O. Delmont, Josep M. Gasol, Alvar D. Gossert, André Kahles, Matthew B. Sullivan, Patrick Wincker, Georg Zeller, Serina L. Robinson, Jörn Piel, and Shinichi Sunagawa. Biosynthetic potential of the global ocean microbiome. *Nature*, 607(7917):111–118, Jul 2022.

[17] Alexander M. Kloosterman, Peter Cimermancic, Somayah S. Elsayed, Chao Du, Michalis Hadjithomas, Mohamed S. Donia, Michael A. Fischbach, Gilles P. van Wezel, and Marnix H. Medema. Expansion of ripp biosynthetic space through integration of pan-genomics and machine learning uncovers a novel class of lanthipeptides. *PLoS biology*, 18(12):e3001026–e3001026, Dec 2020. 33351797[pmid].

[18] William C. W. Chen, Leonid Gaidukov, Yong Lai, Ming-Ru Wu, Jicong Cao, Michael J. Gutbrod, Gigi C. G. Choi, Rachel P. Utomo, Ying-Chou Chen, Liliana Wroblewska, Manolis Kellis, Lin Zhang, Ron Weiss, and Timothy K. Lu. A synthetic transcription platform for programmable gene expression in mammalian cells. *Nature Communications*, 13(1):6167, Oct 2022.

[19] Mazhar Adli. The crispr tool kit for genome editing and beyond. *Nature Communications*, 9(1):1911, May 2018.

[20] F. Ann Ran, Patrick D. Hsu, Jason Wright, Vineeta Agarwala, David A. Scott, and Feng Zhang. Genome engineering using the crispr-cas9 system. *Nature Protocols*, 8(11):2281–2308, Nov 2013.

[21] Daniel G. Gibson, John I. Glass, Carole Lartigue, Vladimir N. Noskov, Ray-Yuan Chuang, Mikkel A. Algire, Gwynedd A. Benders, Michael G.

Montague, Li Ma, Monzia M. Moodie, Chuck Merryman, Sanjay Vashee, Radha Krishnakumar, Nacyra Assad-Garcia, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Lei Young, Zhi-Qing Qi, Thomas H. Segall-Shapiro, Christopher H. Calvey, Prashanth P. Parmar, Clyde A. Hutchison, Hamilton O. Smith, and J. Craig Venter. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329(5987):52–56, 2010.

[22] Jennifer H. Wisecaver, Alexander T. Borowsky, Vered Tzin, Georg Jander, Daniel J. Kliebenstein, and Antonis Rokas. A global coexpression network approach for connecting genes to specialized metabolic pathways in plants. *The Plant cell*, 29(5):944–959, May 2017. 28408660[pmid].

[23] Karen G. Lloyd, Andrew D. Steen, Joshua Ladau, Junqi Yin, and Lonnie Crosby. Phylogenetically novel uncultured microbial cells dominate earth microbiomes. *mSystems*, 3(5):e00055–18, Sep 2018. 30273414[pmid].

[24] Staley, j. t., and konopka, a. (1985). measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. annual review of microbiology, 39, 321–346. https://doi.org/10.1146/annurev.mi.39.100185.001541.

[25] Mads Albertsen. Long-read metagenomics paves the way toward a complete microbial tree of life. *Nature Methods*, 20(1):30–31, Jan 2023.

[26] Micheal C. Wilson and Jörn Piel. Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chemistry & Biology*, 20(5):636–647, 2013.

[27]

[28] Laura A. Hug, Brett J. Baker, Karthik Anantharaman, Christopher T. Brown, Alexander J. Probst, Cindy J. Castelle, Cristina N. Butterfield, Alex W. Hernsdorf, Yuki Amano, Kotaro Ise, Yohey Suzuki, Natasha Dudek, David A. Relman, Kari M. Finstad, Ronald Amundson, Brian C. Thomas, and Jillian F. Banfield. A new view of the tree of life. *Nature Microbiology*, 1(5):16048, Apr 2016.

[29] David L. Hawksworth and Robert Lücking. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiology Spectrum*, 5(4):5.4.10, 2017.

[30] Camilo Mora, Derek P. Tittensor, Sina Adl, Alastair G. B. Simpson, and Boris Worm. How many species are there on earth and in the ocean? *PLOS Biology*, 9(8):1–8, 08 2011.

[31] Kenneth J. Locey and Jay T. Lennon. Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences*, 113(21):5970–5975, 2016.

[32] C. Pedrós-Alió. Rare biosphere. In *Reference Module in Life Sciences*. Elsevier, 2017.

[33] Edoardo Pasolli, Francesco Asnicar, Serena Manara, Moreno Zolfo, Nicolai Karcher, Federica Armanini, Francesco Beghini, Paolo Manghi, Adrian Tett, Paolo Ghensi, Maria Carmen Collado, Benjamin L. Rice, Casey DuLong, Xochitl C. Morgan, Christopher D. Golden, Christopher Quince, Curtis Huttenhower, and Nicola Segata. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell*, 176(3):649–662.e20, Jan 2019.

[34] Alexander Crits-Christoph, Spencer Diamond, Cristina N. Butterfield, Brian C. Thomas, and Jillian F. Banfield. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature*, 558(7710):440–444, Jun 2018.

[35] J. Craig Venter, Karin Remington, John F. Heidelberg, Aaron L. Halpern, Doug Rusch, Jonathan A. Eisen, Dongying Wu, Ian Paulsen, Karen E. Nelson, William Nelson, Derrick E. Fouts, Samuel Levy, Anthony H. Knap, Michael W. Lomas, Ken Nealson, Owen White, Jeremy Peterson, Jeff Hoffman, Rachel Parsons, Holly Baden-Tillson, Cynthia Pfannkoch, Yu-Hui Rogers, and Hamilton O. Smith. Environmental genome shotgun sequencing of the sargasso sea. *Science*, 304(5667):66–74, 2004.

[36] Daniel R Colman, Saroj Poudel, Blake W Stamps, Eric S Boyd, and John R Spear. The deep, hot biosphere: Twenty-five years of retrospection. *Proc Natl Acad Sci U S A*, 114(27):6895–6903, July 2017.

[37] Frederick S. Colwell and Steven D'Hondt. Nature and Extent of the Deep Biosphere. *Reviews in Mineralogy and Geochemistry*, 75(1):547–574, 01 2013.

[38] T Gold. The deep, hot biosphere. *Proceedings of the National Academy of Sciences*, 89(13):6045–6049, 1992.

[39] Wen-Sheng Shu and Li-Nan Huang. Microbial diversity in extreme environments. *Nature Reviews Microbiology*, 20(4):219–235, Apr 2022.

[40] Yinon M. Bar-On, Rob Phillips, and Ron Milo. The biomass distribution on earth. *Proceedings of the National Academy of Sciences*, 115(25):6506–6511, 2018.

[41] C. Magnabosco, L.-H. Lin, H. Dong, M. Bomberg, W. Ghiorse, H. Stan-Lotter, K. Pedersen, T. L. Kieft, E. van Heerden, and T. C. Onstott. The biomass and biodiversity of the continental subsurface. *Nature Geoscience*, 11(10):707–717, Oct 2018.

[42]

[43] James A. Bradley, Jan P. Amend, and Douglas E. LaRowe. Survival of the fewest: Microbial dormancy and maintenance in marine sediments through deep time. *Geobiology*, 17(1):43–59, 2019.

[44] Thomas L. Kieft, Clifford C. Walters, Meytal B. Higgins, Anthony S. Mennito, Catherine F.M. Clewett, Verena Heuer, Michael J. Pullin, Sarah Hendrickson, Esta van Heerden, Barbara Sherwood Lollar, Maggie C.Y. Lau, and T.C. Onstott. Dissolved organic matter compositions in 0.6–3.4km deep fracture waters, kaapvaal craton, south africa. *Organic Geochemistry*, 118:116–131, 2018.

[45] Elizabeth Trembath-Reichert, Yuki Morono, Akira Ijiri, Tatsuhiko Hoshino, Katherine S. Dawson, Fumio Inagaki, and Victoria J. Orphan. Methyl-compound use and slow growth characterize microbial life in 2-km-deep subseafloor coal and shale beds. *Proceedings of the National Academy of Sciences*, 114(44):E9206–E9215, 2017.

[46] Stefano Mariani, Charles Baillie, Giuliano Colosimo, and Ana Riesgo. Sponges as natural environmental dna samplers. *Current Biology*, 29(11):R401–R402, 2019.

[47] Bing-Nan Han, Li-Li Hong, Bin-Bin Gu, Yang-Ting Sun, Jie Wang, Jin-Tang Liu, and Hou-Wen Lin. *Natural Products from Sponges*, pages 329–463. Springer Netherlands, Dordrecht, 2019.

[48] Michael J. Benton. Biodiversity on land and in the sea. *Geological Journal*, 36(3-4):211–230, 2001.

[49] Zhi-Qiang Zhang. Animal biodiversity: An outline of higher-level classification and survey of taxonomic richness (addenda 2013). *Zootaxa*, 3703:1–82, 2013.

[50] Helmut Hillebrand. On the generality of the latitudinal diversity gradient. *The American Naturalist*, 163(2):192–211, 2004. PMID: 14970922.

[51] Tim M. Blackburn and Kevin J. Gaston. A sideways look at patterns in species richness, or why there are so few species outside the tropics. *Biodiversity Letters*, 3(2):44–53, 1996.

[52] Adam Woodhouse, Anshuman Swain, William F. Fagan, Andrew J. Fraass, and Christopher M. Lowery. Late cenozoic cooling restructured global marine plankton communities. *Nature*, 614(7949):713–718, Feb 2023.

[53] Isabel S. Fenton, Tracy Aze, Alexander Farnsworth, Paul Valdes, and Erin E. Saupe. Origination of the modern-style diversity gradient 15 million years ago. *Nature*, 614(7949):708–712, Feb 2023.

[54] Philip D. Mannion, Paul Upchurch, Roger B.J. Benson, and Anjali Goswami. The latitudinal biodiversity gradient through deep time. *Trends in Ecology Evolution*, 29(1):42–50, 2014.

[55] et al. Thompson, Luke R.

[56] Robert H. Cowie, Philippe Bouchet, and Benoît Fontaine. The sixth mass extinction: fact, fiction or speculation? *Biological Reviews*, 97(2):640–663, 2022.

[57] Céline Bellard, Cleo Bertelsmeier, Paul Leadley, Wilfried Thuiller, and Franck Courchamp. Impacts of climate change on the future of biodiversity. *Ecology Letters*, 15(4):365–377, 2012.

[58] Kostas A. Triantis, Paulo A. V. Borges, Richard J. Ladle, Joaquín Hortal, Pedro Cardoso, Clara Gaspar, Francisco Dinis, Enésima Mendonça, Lúcia M. A. Silveira, Rosalina Gabriel, Catarina Melo, Ana M. C. Santos, Isabel R. Amorim, Sérvio P. Ribeiro, Artur R. M. Serrano, José A. Quartau, and Robert J. Whittaker. Extinction debt on oceanic islands. *Ecography*, 33(2):285–294, 2010.

[59] Norman Myers, Russell A. Mittermeier, Cristina G. Mittermeier, Gustavo A. B. da Fonseca, and Jennifer Kent. Biodiversity hotspots for conservation priorities. *Nature*, 403(6772):853–858, Feb 2000.

[60] Virginia H. Dale. The relationship between land-use change and climate change. *Ecological Applications*, 7(3):753–769, 1997.

[61] Karina Winkler, Richard Fuchs, Mark Rounsevell, and Martin Herold. Global land use changes are four times greater than previously estimated. *Nature Communications*, 12(1):2501, May 2021.

[62]

[63] Daniel Pauly, Reg Watson, and Jackie Alder. Global trends in world fisheries: impacts on marine ecosystems and food security. *Philos Trans R Soc Lond B Biol Sci*, 360(1453):5–12, January 2005.

[64] William W. L. Cheung, Reg Watson, and Daniel Pauly. Signature of ocean warming in global fisheries catch. *Nature*, 497(7449):365–368, May 2013.

[65] Andrew D. Barton, Andrew J. Irwin, Zoe V. Finkel, and Charles A. Stock. Anthropogenic climate change drives shift and shuffle in north atlantic phytoplankton communities. *Proceedings of the National Academy of Sciences*, 113(11):2964–2969, 2016.

[66] Lukas Jonkers, Helmut Hillebrand, and Michal Kucera. Global change drives modern plankton communities away from the pre-industrial state. *Nature*, 570(7761):372–375, Jun 2019.

[67] David S. Thaler. Is global microbial biodiversity increasing, decreasing, or staying the same? *Frontiers in Ecology and Evolution*, 9, 2021.