Crossing the Rubicon: the BMBF WiLDSI project

“Science-based solutions for Digital Sequence Information (DSI) in preparation for COP 15”

„Wissenschaftsbasierte Lösungsansätze für Digitale Sequenzinformation (DSI)“

Dr. Amber Hartman Scholz, Deputy to the Director

1969 2019

Allianz d
ganisatio

Stellungn

forscher Leopoldina
schaften

auschdienst

09. Febr

Allianz s
fährdet

Stellung
kussion,
Ressour
vention

VBIO ~ GS Berlin ~ Luisenstraße 58/59 ~ 10117 Berlin ~ Germany

Cristiana Paşca Palmer
Executive Secretary, UN Assistant Secretary-General

Secretariat of the Convention on Biological Diversity
413, Saint Jacques Street, suite 800
Montreal QC H2Y 1N9
Canada

Dr. Kerstin Elbing
Geschäftsstelle Berlin
Luisenstraße 58/59
10117 Berlin

Telefon: 030-27891916
e-Mail: elbing@vbio.de

May 31st, 2019

Joint submission on DSI (Ref. SCBD/NPU/DC/VN/KG/RKi/87804)

Die Alliar
chenden
Biodiversi
quenzinfo
Nagoya-P
unterliege

Dear Executive Secretary Cristiana Paşca Palmer,

with its Notification SCBD/NPU/DC/VN/KG/RKi/87804 the CBD secretariat has asked for the submission of views and information concerning Digital Sequence Information on Genetic Resources.

Please find attached the joint submission of Consortium of German Natural History Collections, DNFS (Deutsche Naturwissenschaftliche Forschungssammlungen), German Life Sciences Association (Verband Biowissenschaften, Biologie und Biomedizin in Deutschland, VBIO e. V.)

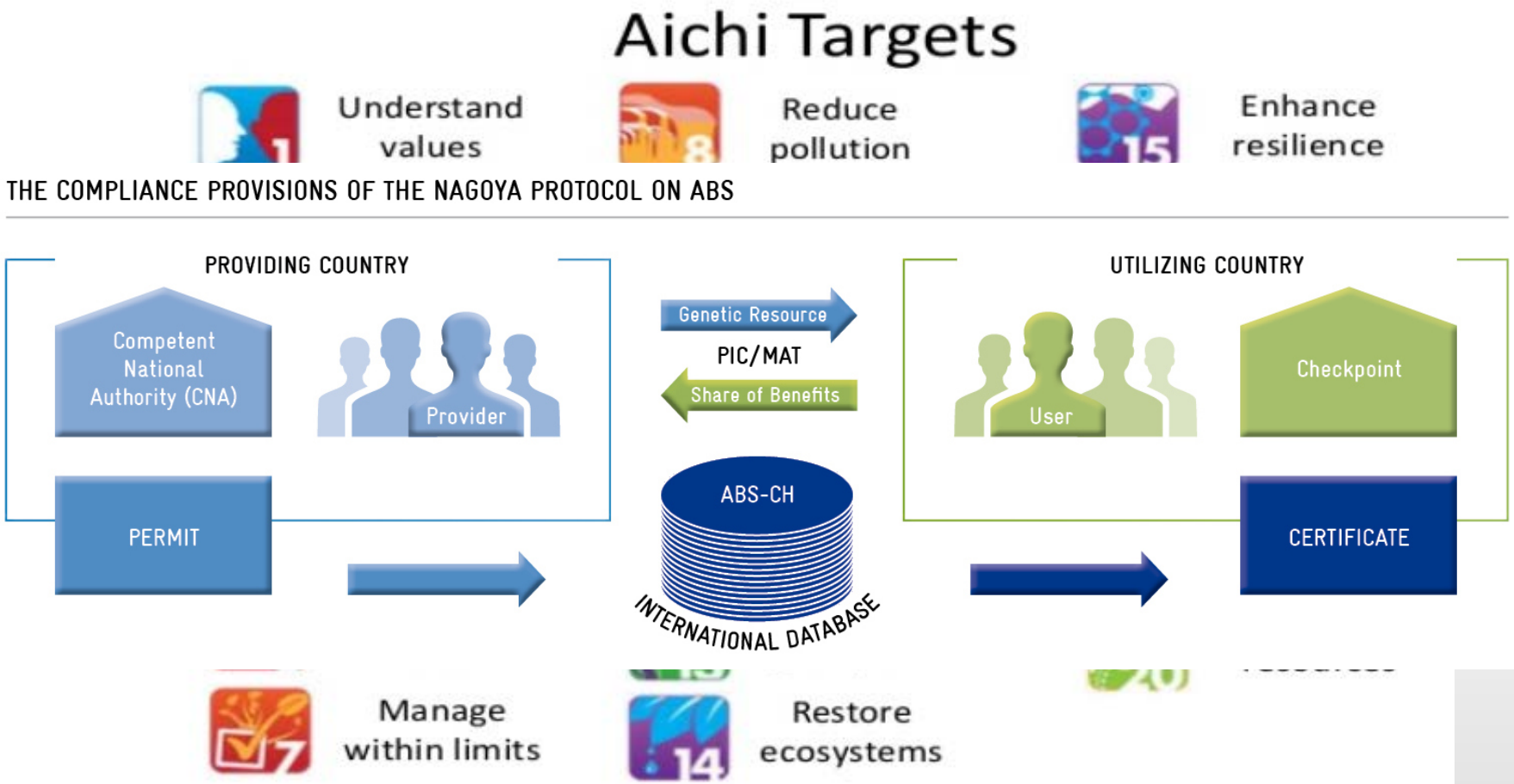
Crossing the Rubicon??

Source: https://www.onthisday.com/photos/julis-caesar-crosses-the-rubicon

So why engage now in the DSI process?

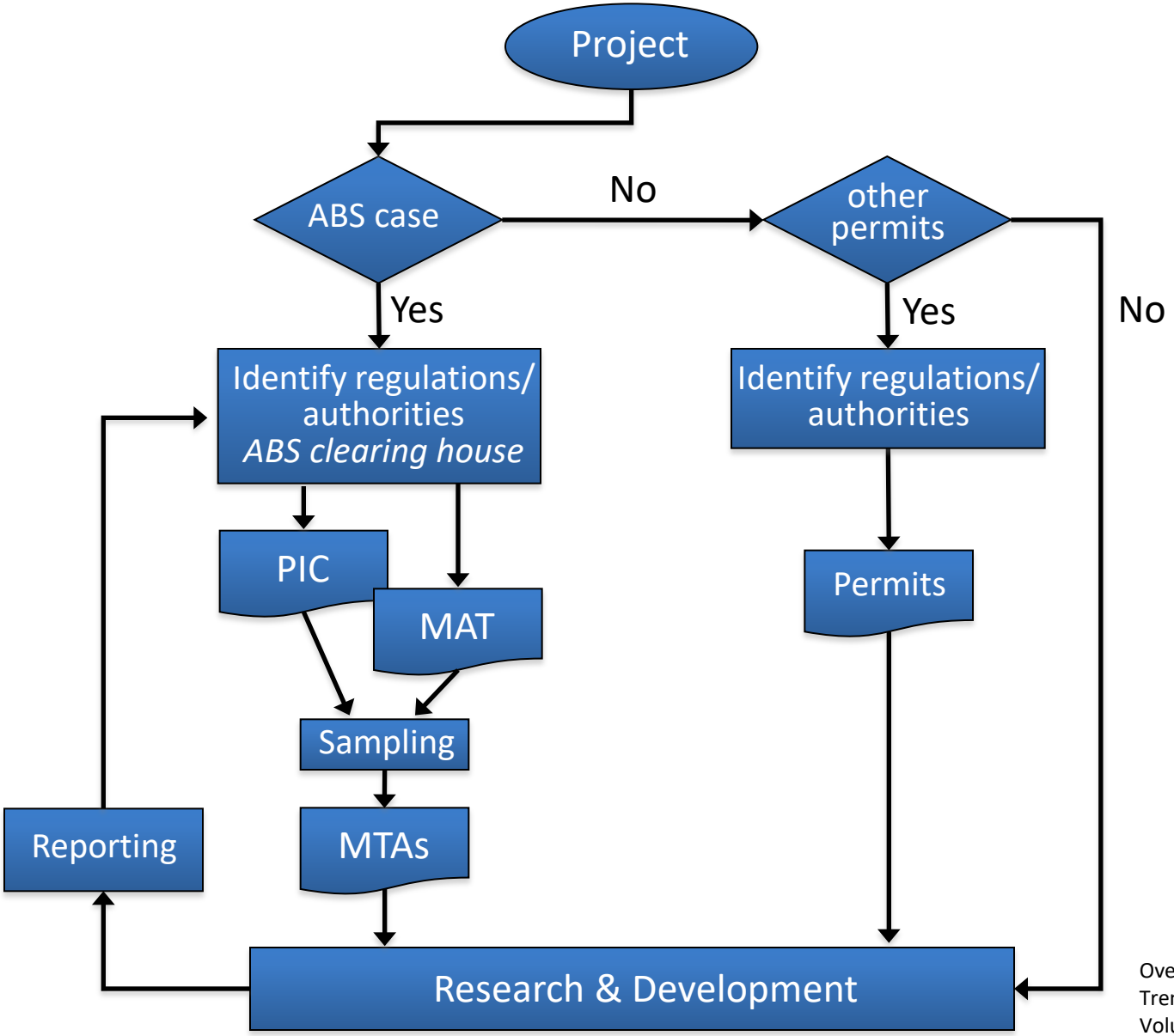
1. Lessons learned from Nagoya
2. „Biopiracy“ accusations vs. reality
3. Science & Biodiversity depend on sequence data!
4. Put out the facts
5. Prevent bad ideas

1. Learning from history: flashback to 2010 Kuhhandel / Horse trade



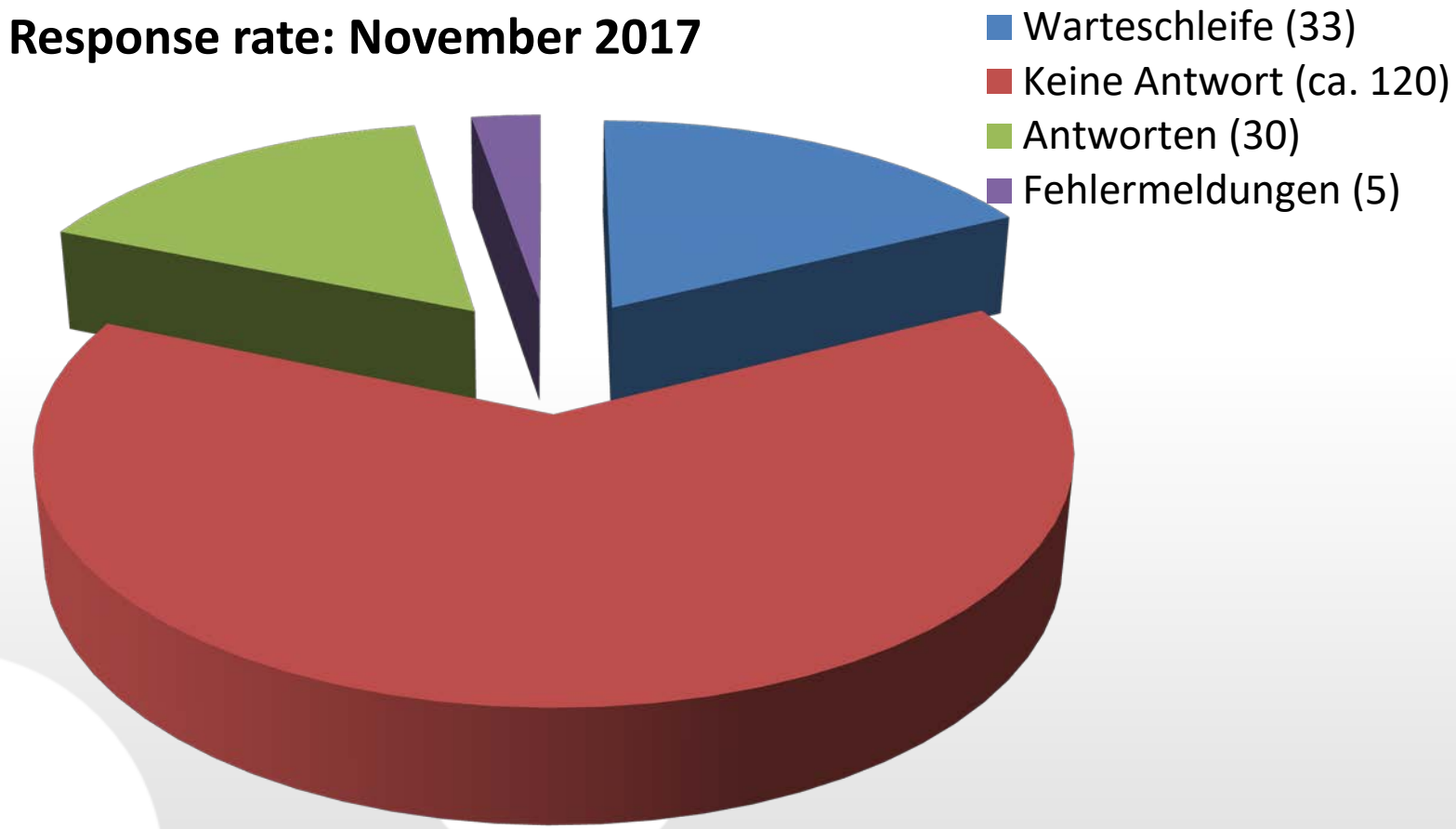
http://www.abs-initiative.info/fileadmin/media/Knowledge_Center/Multimedia/Graphics/ABS_Grafik_The_compliance_provisions.jpg

But it‘s not quite so simple in practice

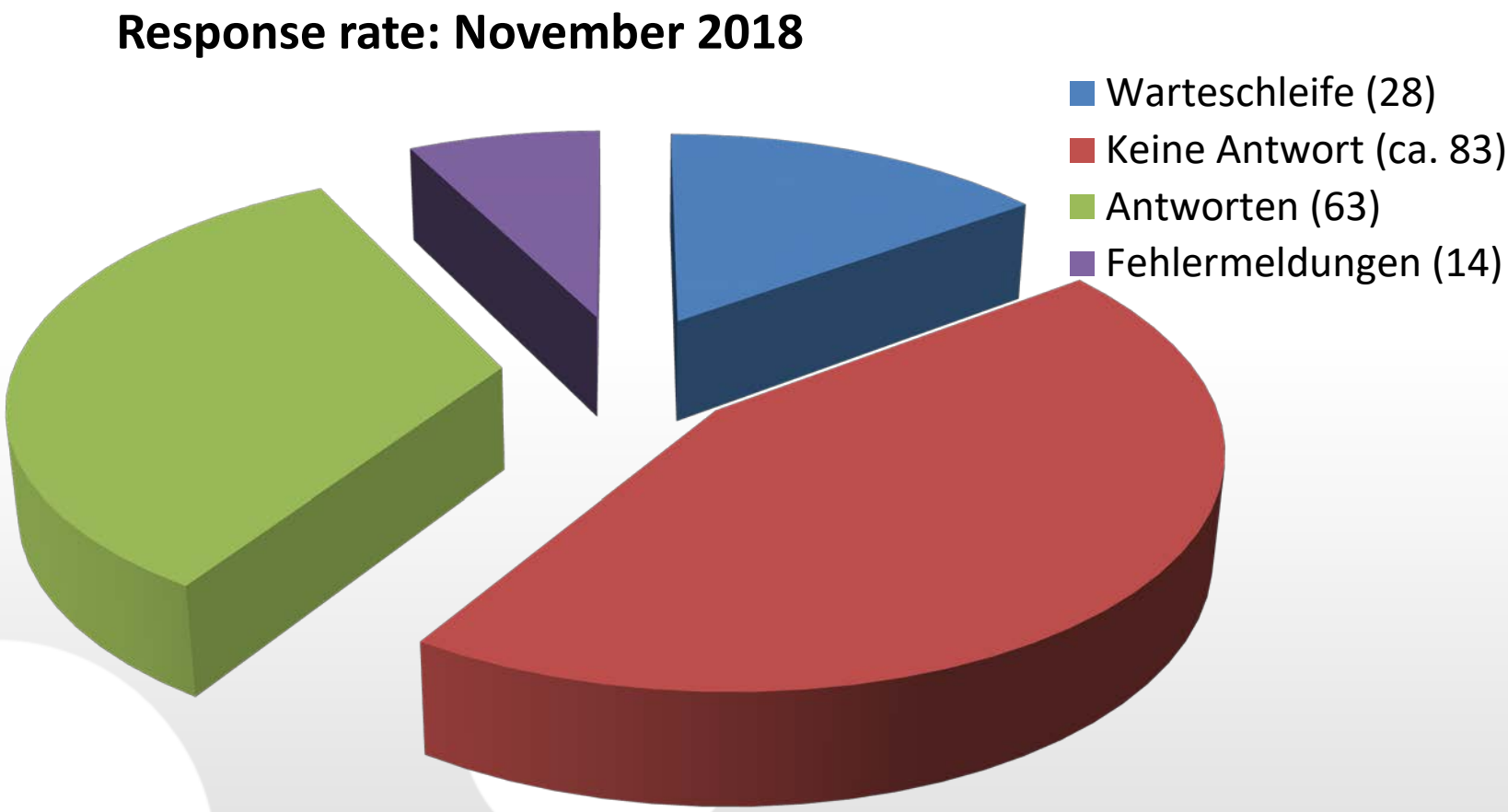


Overmann and Scholz.
Trends in Microbiology.
Volume 25, ISSUE 2, 2017

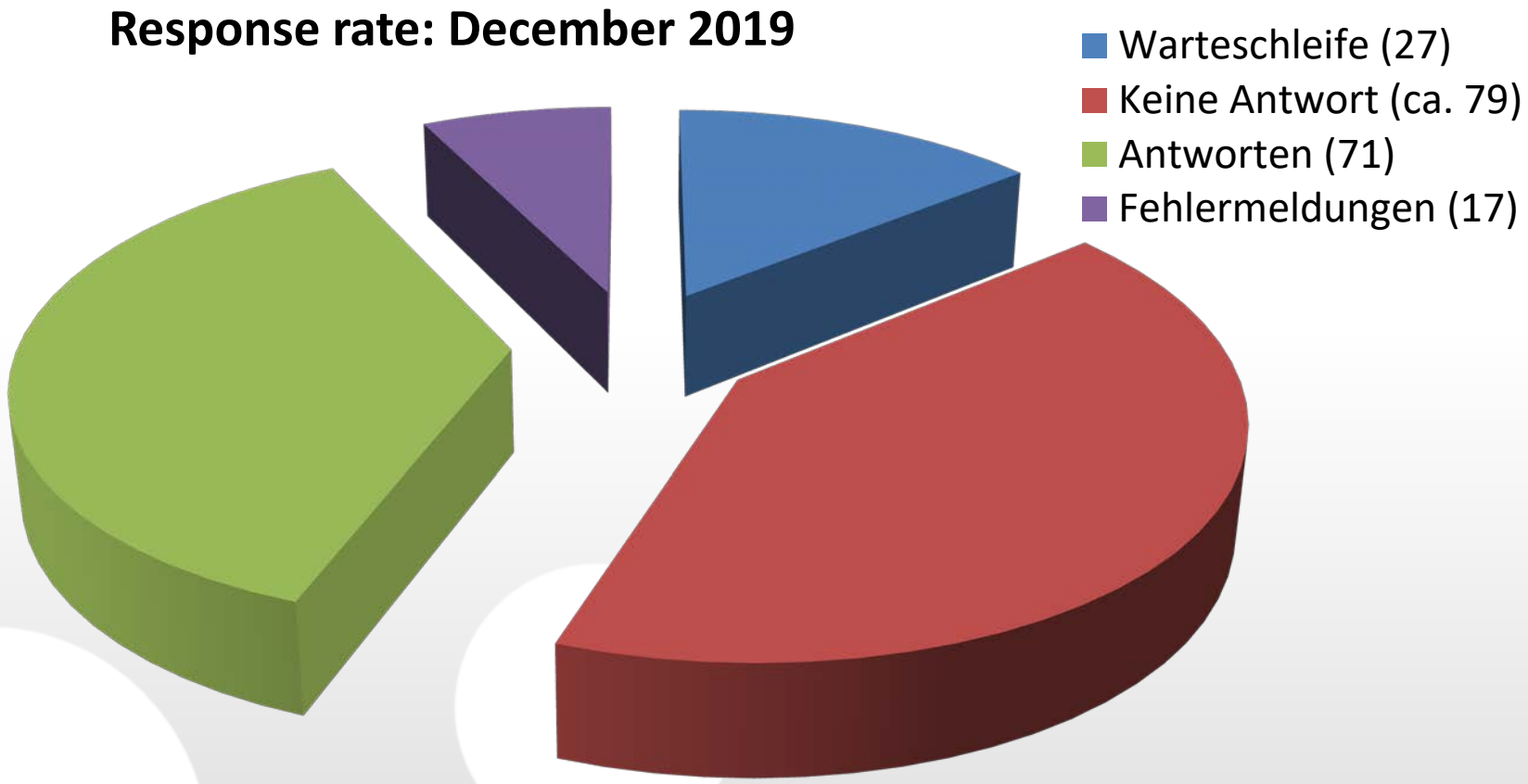
It is challenging to get information on CBD/NP compliance



It is challenging to get information on CBD/NP compliance



It is challenging to get information on CBD/NP compliance

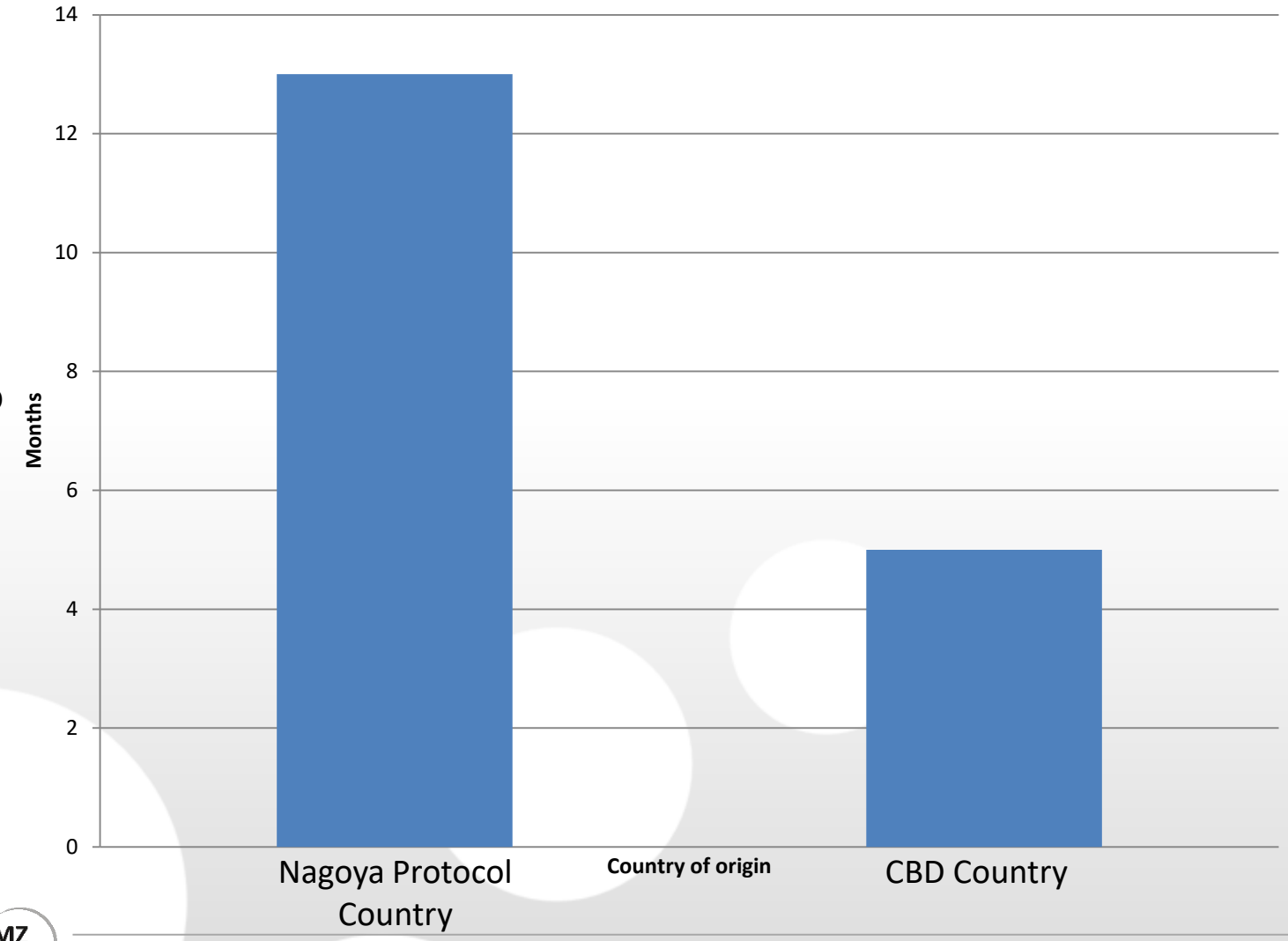


Survey results: the NP compliance problem

- Short survey of 20 institutes involved in biodiversity research
- 40 projects
- 23 countries
- Average:
  - 3 year project length
  - 65,000 Eur

The Nagoya Protocol and CBD cause delays in biodiversity research

Time to permit



The NP system is clunky at best and leads to jurisdiction shopping, but was it worth it for biodiversity conservation?

Mixed success (at best)

https://www.cbd.int/images/publications/gbo3/figures/table2.png

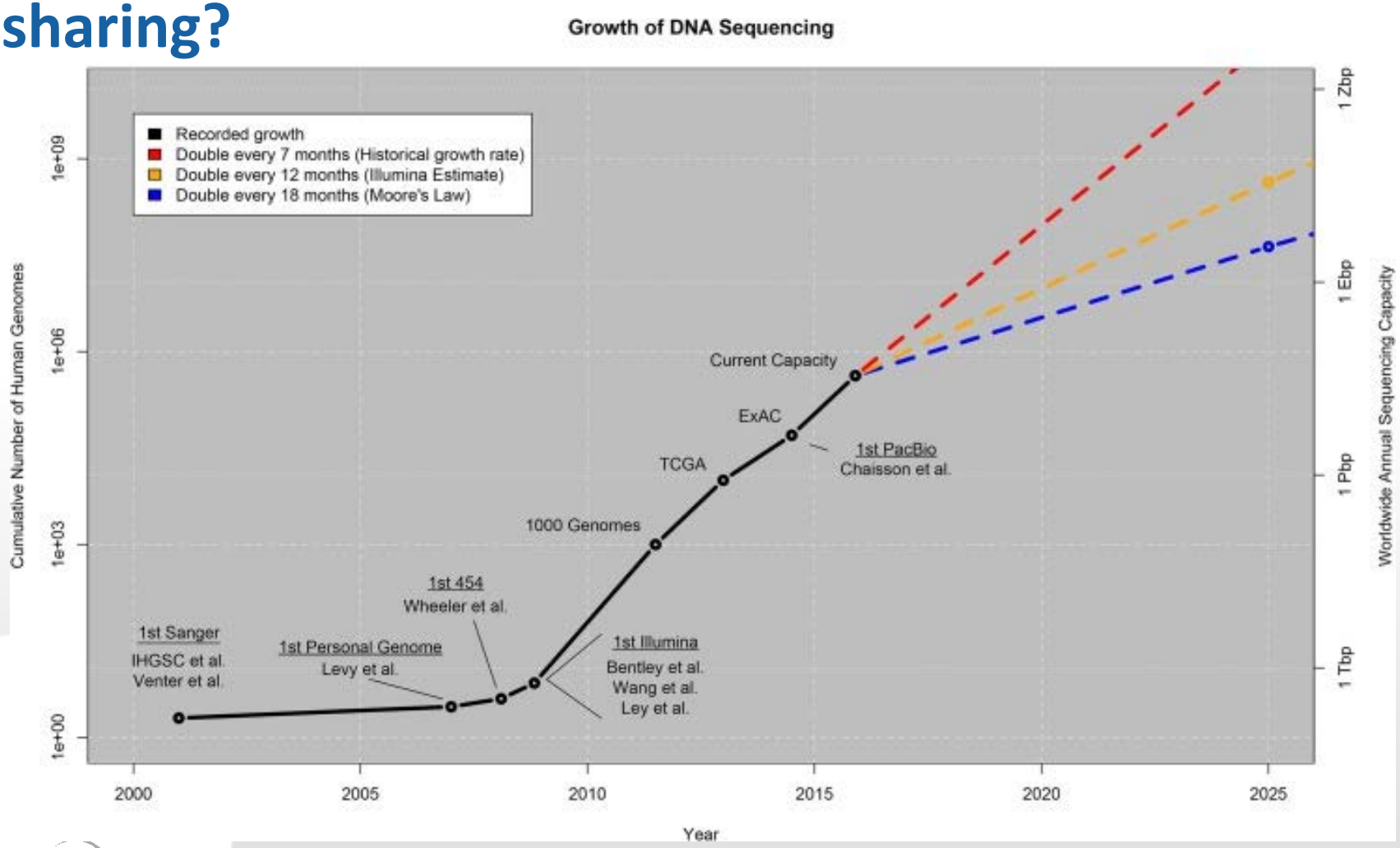
TABLE 2 Trends shown by agreed indicators of progress towards the 2010 biodiversity target

Status and trends of the components of biological diversity	
Trends in extent of selected biomes, ecosystems, and habitats	Most habitats in most parts of the world are declining in extent, although forest area expands in some regions, and the loss of mangroves has slowed significantly, except in Asia.
Trends in abundance and distribution of selected species	Most species with limited population size and distribution are being further reduced, while some common and invasive species become more common. (but limited number of taxa assessed)
Change in status of threatened species	The risk of extinction increases for many threatened species, although some species recovery programmes have been very successful. (for those species assessed)
Trends in genetic diversity of domesticated animals, cultivated plants, and fish species of major socio-economic importance	It is likely that the genetic variety of cultivated species is declining, but the extent of such decline and its overall impacts are not well understood. (although many case studies with a high degree of certainty are available)
Coverage of protected areas	There has been a significant increase in coverage of protected areas, both terrestrial and marine, over the past decade. However, many ecological regions, particularly in marine ecosystems, remain underprotected, and the management effectiveness of protected areas remains variable.
Ecosystem integrity and ecosystem goods and services	
Marine Trophic Index	Despite intense pressure the Marine Trophic Index has shown a modest increase globally since 1970. However there is substantial regional variation with declines being recorded in half of the marine areas with data. Although the global increases may indicate a recovery it is more likely a consequence of fishing fleets expanding their areas of activity, thus encountering fish stocks in which larger predators have not yet been removed in large numbers.
Connectivity – fragmentation of ecosystems	Most terrestrial and aquatic ecosystems are becoming increasingly fragmented, despite an increased recognition of the value of corridors and connections, especially in climate change adaptation.
Water quality of aquatic ecosystems	Most parts of the world are likely to be suffering from declines in water quality, although quality in some areas has improved through control of point-source pollution.
Threats to biodiversity	
Nitrogen deposition	Human activity has doubled the rate of creation of reactive nitrogen on the planet's surface. Pressure on biodiversity from nutrient pollution continues to increase, although some measures to use nutrients more efficiently, to reduce their release into water and the atmosphere, are beginning to show positive effects.
Trends in invasive alien species	The number and rate of spread of alien species is increasing in all continents and all ecosystem types. (although many case studies with a high degree of certainty are available)
Sustainable use	
Area of forest, agricultural and aquaculture ecosystems under sustainable management	There are considerable efforts under way to increase the extent of areas of land under sustainable management. Regional efforts on sustainable forest management are expected to contribute to this. Traditional agricultural practices are being maintained and revitalized as the demand for ethical and healthy products increases. However, these are still relatively small niches and major efforts are required to substantially increase the areas under sustainable management.
Ecological footprint and related concepts	The ecological footprint of humanity is increasing. Efforts at increasing resource efficiency are more than compensated by increased consumption by a growing and more prosperous human population.
Status of traditional knowledge, innovations and practices	
Status and trends of linguistic diversity and numbers of speakers of indigenous languages	A large number of minority languages are believed in danger of disappearing, and linguistic diversity is very likely declining. (although case studies with a high degree of certainty are available)
Status of access and benefit sharing	
? Indicator of access and benefit-sharing to be developed	The need and possible options for additional indicators are being examined by the Ad Hoc Open-ended Working Group on Access and Benefit-sharing.
Status of resources transfers	
Official development assistance (ODA) provided in support of the Convention	The volume of ODA for biodiversity has increased over the past few years.

Negative changes · Positive changes · No clear global trend. Positive and negative changes are occurring depending on the region or biome considered · ? Insufficient information to reach a definitive conclusion.

Degree of certainty: Low · Medium · High

2. Biopiracy accusations
Why is sequencing the next target for benefit sharing?



PLoS Biol. 2015 Jul; 13(7): e1002195. doi: 10.1371/journal.pbio.1002195

2. Bio

News | Published: 20 May 2010

Researchers start up cell with synthetic genome

Alla Katsnelson

Nature (2010) | Cite this article

29 Accesses | 1 Citations | 121 Altmetric | Metrics

vw.downtoearth.org.in/blo
pen-season-for-biopiracy-

Ebol
oblig

A fully synthesized genome transforms one species of bacterium into another.

by Edwa

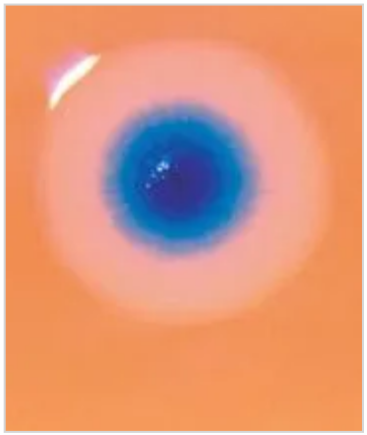
Scientists have built a bacterial genome from scratch and used it to 'reboot' a cell from a different species of bacterium.

A US com
developed
tion (DSI)
African ep
heavy atter
discussions
versity (CB
tion on acc
and other g

The Ebola
European
datory ben
of physical
sequences
strings attached.

DS

Daniel Gibson and his colleagues at the J. Craig Venter Institute in Rockville, Maryland, synthesized the genome of the bacterium *Mycoplasma mycoides*, consisting of about 1.1 million base pairs. Having assembled the genome inside a yeast cell, they transplanted it



A chemical marker (blue) shows the ability of a single bacterium with synthesized

2. Biopiracy accusations

Biological properties and existing technical capacities for synthetic biology of plastids compared to bacteria, yeast and the plant nucleus.

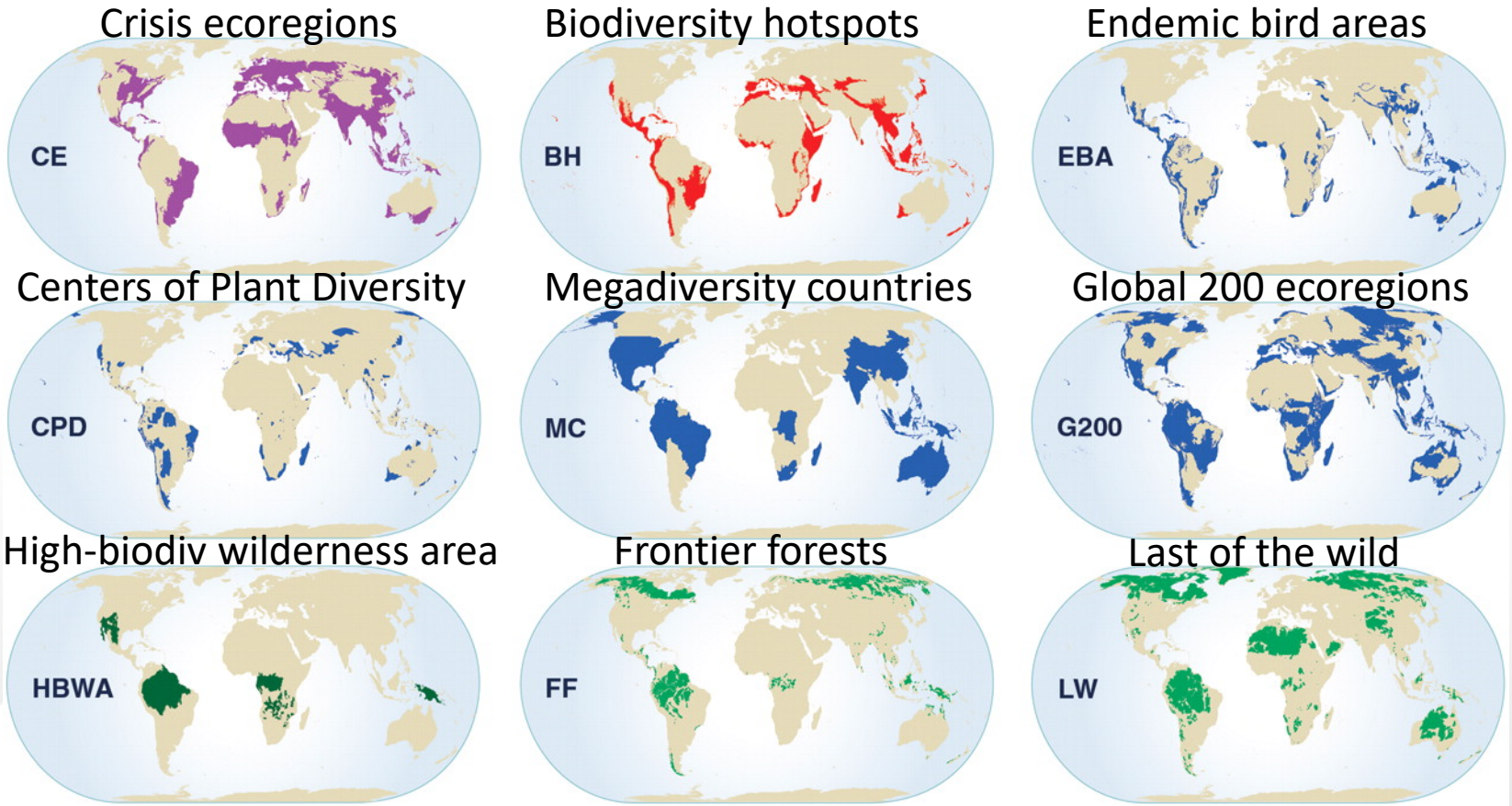
	Bacteria	Yeast	Plant (nucleus)	Plant (plastid)
Biological features				
Genomic simplicity	**	*	*	***
Physiological simplicity	***	**	*	**
Homologous recombination	**	***	*	***
Polycistronic gene expression	***	*	*	***
Gene expression strength	**	**	*	***
Absence of transgene silencing	***	**	*	***
Biological transgene containment	*	*	*	**
Photoautotrophic growth	**	*	***	***
Available tools and techniques				
Ease of genetic manipulation	***	***	**	*
Speed of genetic manipulation	***	***	**	*
Genetic parts	***	**	*	*
Genetic part characterization	***	**	*	*
Genetic part assembly	***	***	**	**
Genetic part insulation	***	**	**	*
Gene expression control	***	**	**	*
Genetic & metabolic network models	***	**	*	*
Demonstrated applications				
Synthetic metabolic pathways	***	***	**	***
Synthetic genetic circuits	***	**	*	*
Synthetic subcompartments	**	**	*	*
Synthetic genomes	**	*	*	**

DSMZ Lei

©2019 by American Society of Plant Biologists

Christian R. Boehm, and Ralph Bock Plant Physiol. 2018;179:794-802

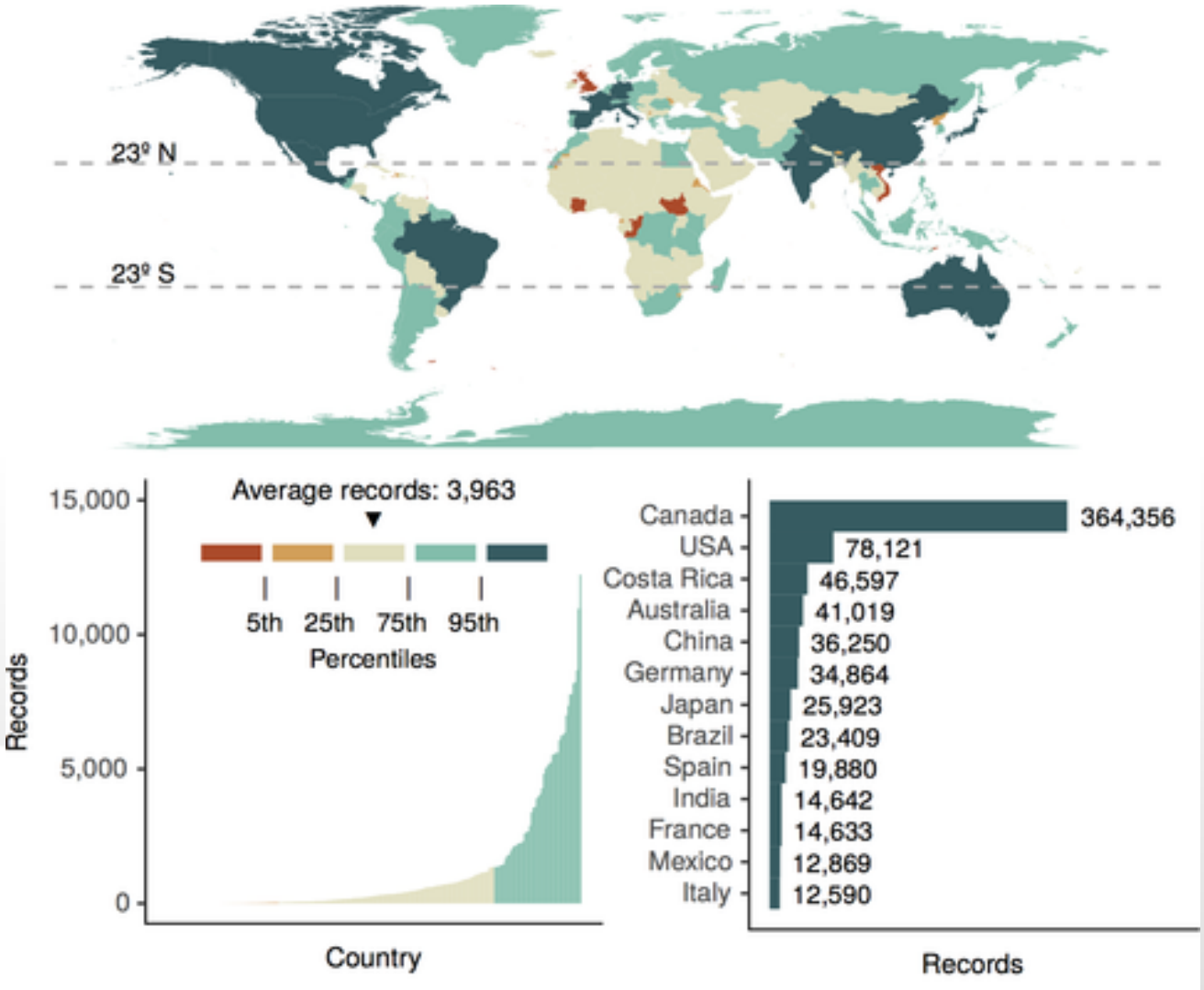
3. Biodiversity is global



T. M. Brooks et al. Science 2006;313:58-61

Published by AAAS

3. Biodiversity research needs NSD databases



Porter TM, Hajibabaei M (2018) Over 2.5 million COI sequences in GenBank and growing. PLOS ONE 13(9): e0200177. https://doi.org/10.1371/journal.pone.0200177

3. Biodiversity research needs NSD databases

PNAS

GenBank is a reliable resource for 21st century biodiversity research

Matthieu Leray[a], Nancy Knowlton[b,1], Shian-Lei Ho[c], Bryan N. Nguyen[b,d,e], and Ryuji J. Machida[c,1]

[a]Smithsonian Tropical Research Institute, Smithsonian Institution, Panama City, 0843-03092, Republic of Panama; [b]National Museum of Natural History, Smithsonian Institution, Washington, DC 20560; [c]Biodiversity Research Centre, Academia Sinica, 115-29 Taipei, Taiwan; [d]Department of Biological Sciences, The George Washington University, Washington, DC 20052; and [e]Computational Biology Institute, Milken Institute School of Public Health, The George Washington University, Washington, DC 20052

Contributed by Nancy Knowlton, September 15, 2019 (sent for review July 10, 2019; reviewed by Ann Bucklin and Simon Creer)

Traditional methods of characterizing biodiversity are increasingly being supplemented and replaced by approaches based on DNA sequencing alone. These approaches commonly involve extraction and high-throughput sequencing of bulk samples from biologically complex communities or samples of environmental DNA (eDNA). In such cases, vouchers for individual organisms are rarely obtained, often unidentifiable, or unavailable. Thus, identifying these sequences typically relies on comparisons with sequences from genetic databases, particularly GenBank. While concerns have been raised about biases and inaccuracies in laboratory and analytical methods, comparatively little attention has been paid to the taxonomic reliability of GenBank itself. Here we analyze the metazoan mitochondrial sequences of GenBank using a combination of distance-based clustering and phylogenetic analysis. Because of their comparatively rapid evolutionary rates and consequent high taxonomic resolution, mitochondrial sequences represent an invaluable resource for the detection of the many small and often undescribed organisms that represent the bulk of animal diversity. We show that metazoan identifications in GenBank are surprisingly accurate, even at low taxonomic levels (likely <1% error rate at the genus level). This stands in contrast to previously

(13), the largest repository of genetic data for biodiversity (14, 15). In many cases, no vouchers are available to independently confirm identification, because the organisms are tiny, very difficult or impossible to identify, or lacking entirely (in the case of eDNA). While concerns have been raised about biases and inaccuracies in laboratory and analytical methods used in metabarcoding (16–18), comparatively little attention has been paid to the taxonomic reliability of GenBank itself, whose >4.7 million mitochondrial gene sequences representing >170,000 metazoan species have never been comprehensively assessed for taxonomic accuracy.

Mitochondrial sequences represent an invaluable resource because of their comparatively rapid evolutionary rates and consequent high taxonomic resolution (19). The majority of metazoan mitochondrial genomes contain 37 genes, including 2 coding for ribosomal RNA, 13 coding for proteins, and 22 tRNAs (20). The tRNAs generally are very short (approximately 60 to 70 bp) and thus not useful as taxonomic markers. However, the other 15 genes are generally much longer than 100 bp, making them potentially useful markers. For the many small and often undescribed organisms

ENVIRONMENTAL SCIENCES

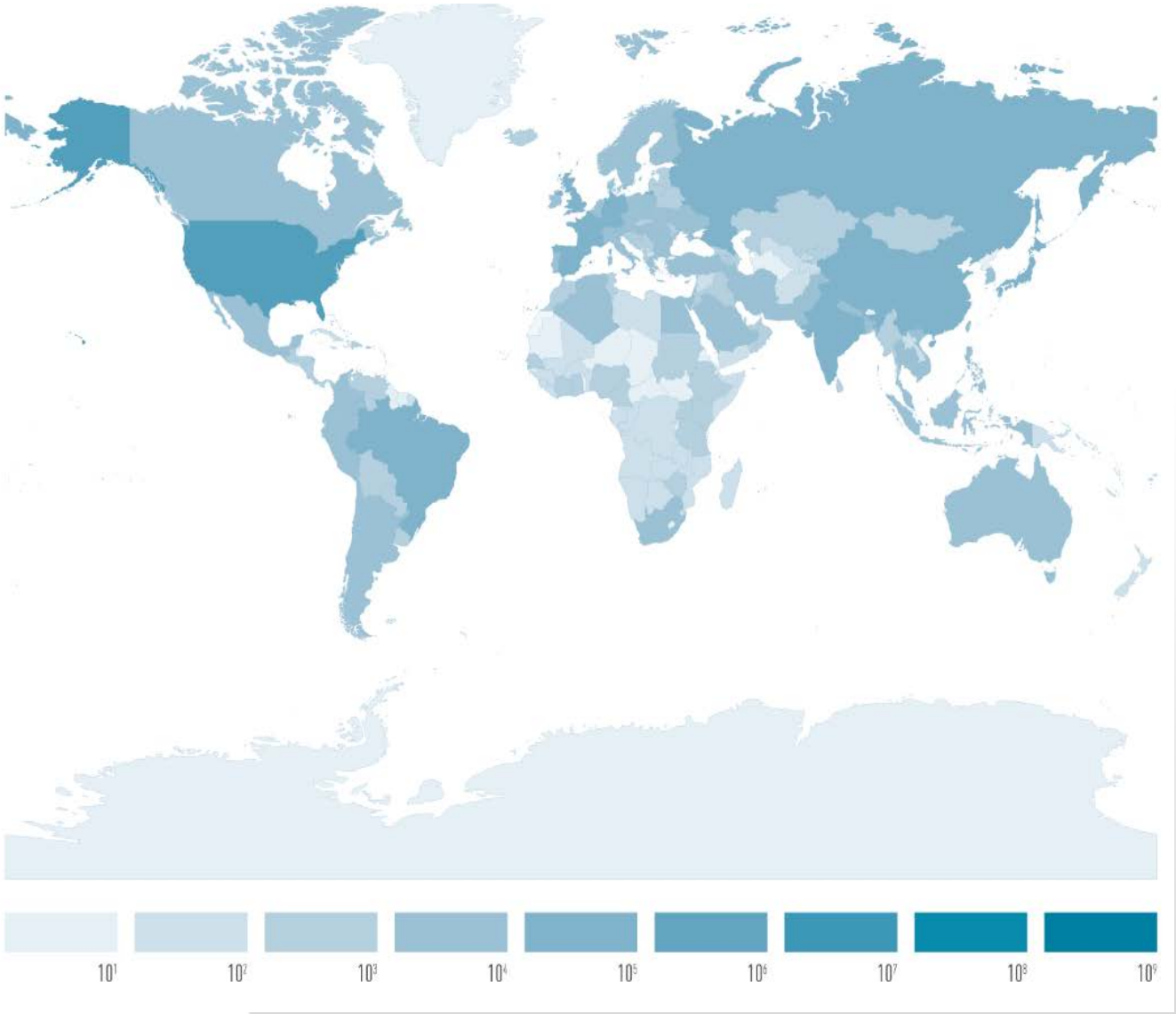
4. Is the Global North taking advantage of the Global South by putting all this DSI online? Who is actually benefitting?

There are 10-15 million total users of INSDC. They live in every country in the world.

1.	United States	22.69 %
2.	China	15.42 %
3.	India	6.16 %
4.	Japan	3.97 %
5.	Germany	3.67 %
6.	United Kingdom	3.45 %
7.	France	2.84 %
8.	Brazil	2.83 %
9.	Spain	2.31 %
10.	Russian Federation	2.25 %

Costs: $3-5 per user

50% of users live in countries that do not contribute to NSD infrastructure costs

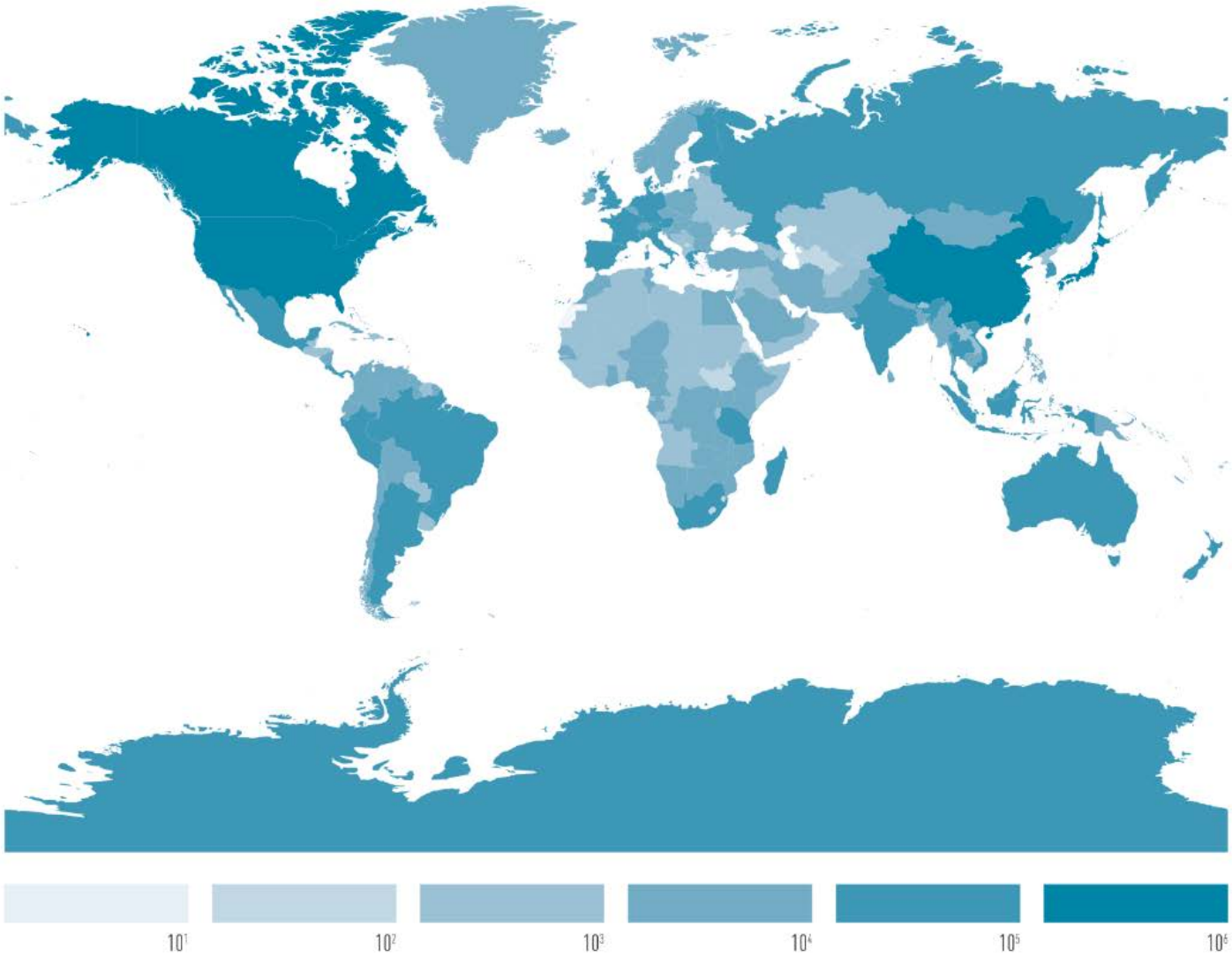


Most DSI does NOT come from the Global South!

What is the country of origin for non-human NSD?

1.	China	18.23 %
2.	United States	17.39 %
3.	Canada	9.10 %
4.	Japan	7.24 %
5.	India	3.46 %
6.	Australia	2.66 %
7.	Mexico	2.54 %
8.	Brazil	2.30 %
9.	Germany	1.83 %
10.	Spain	1.58 %

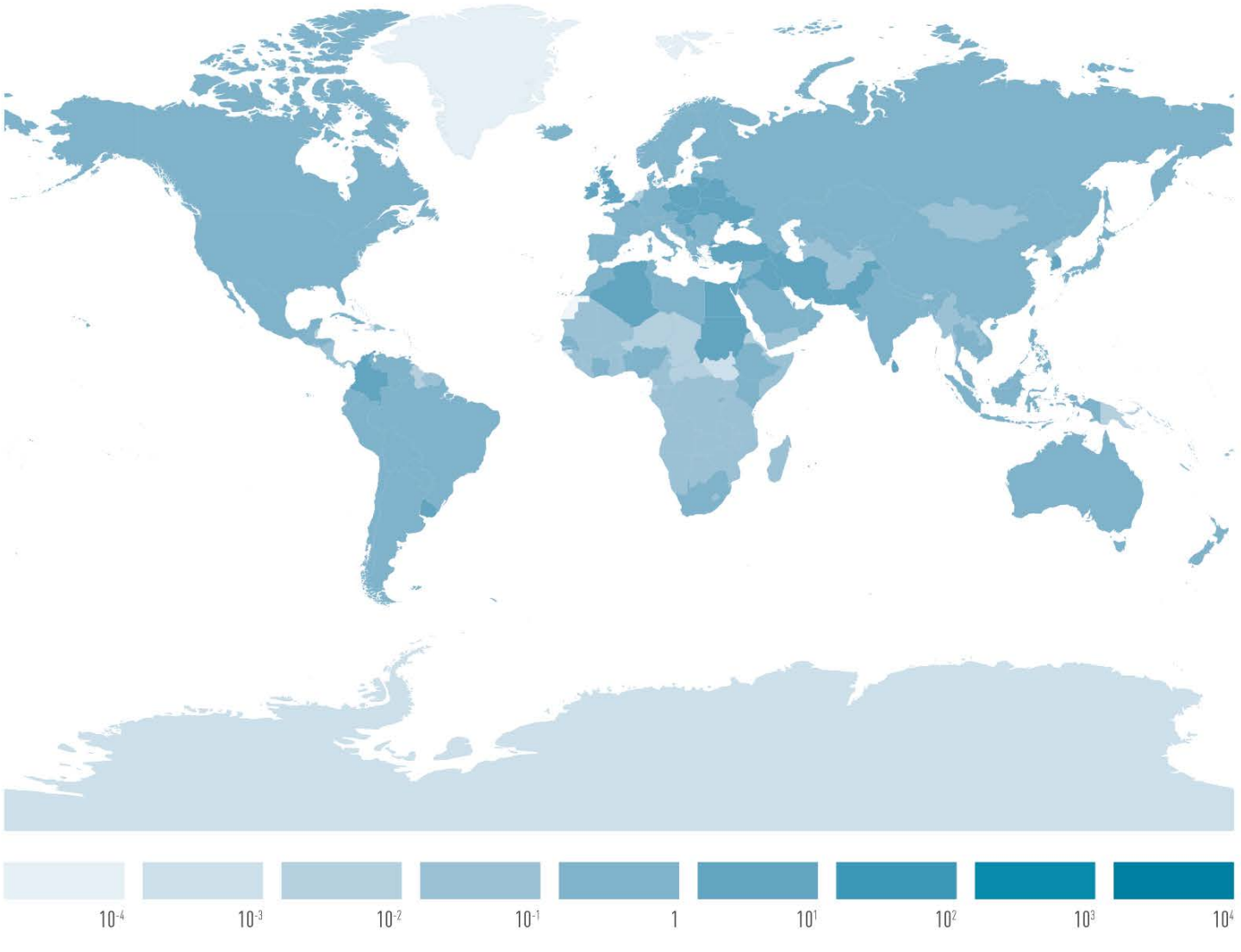
52% of NSD comes from 4 countries



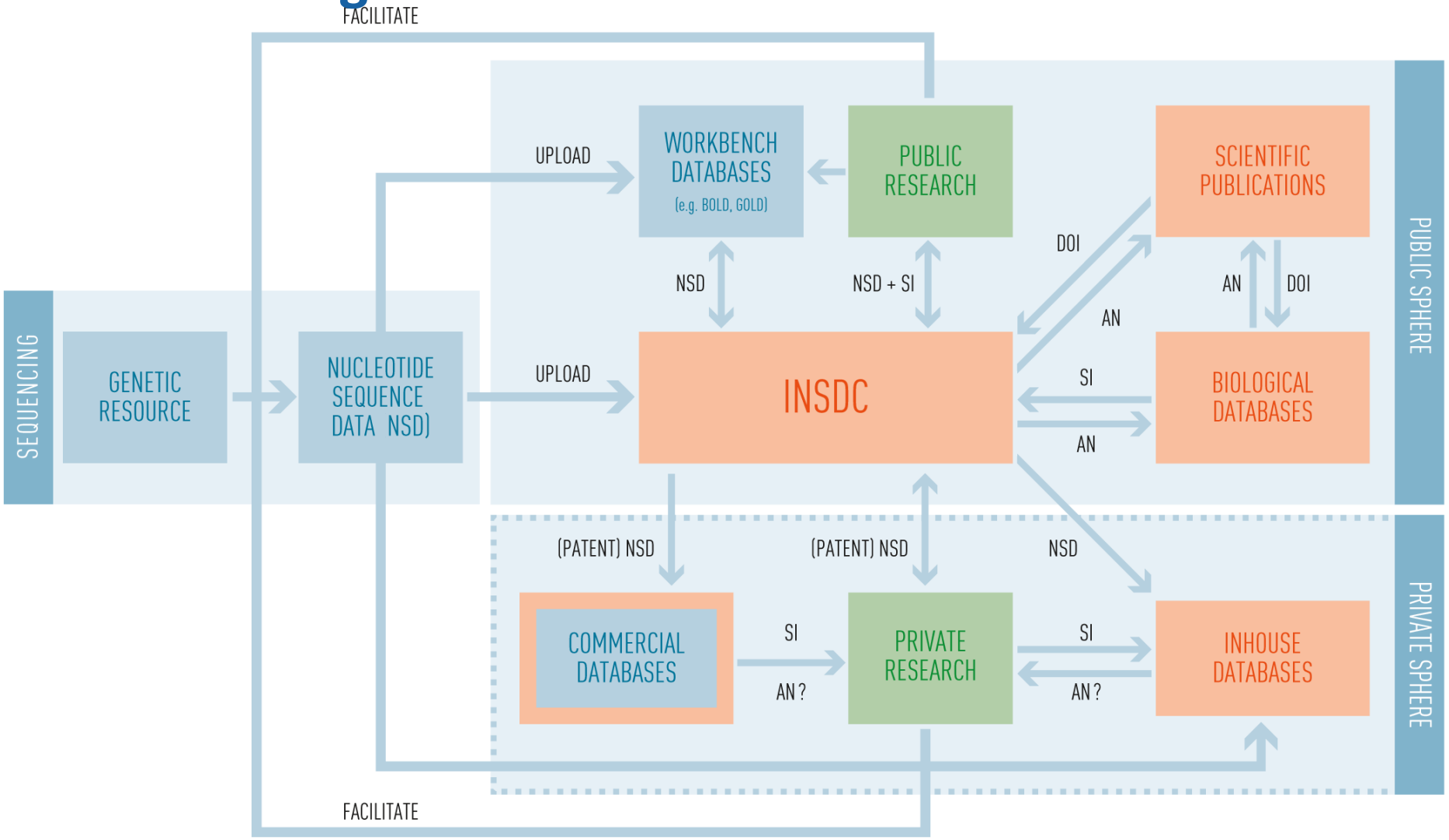
How many users are there per sequence in the INSDC? Benefit sharing is already happening!

How does database usage compare to provided sequences?

1.	Lebanon	27.68
2.	Ukraine	23.84
3.	Belarus	22.88
4.	Iraq	21.41
5.	Colombia	19.79
6.	Algeria	19.54
7.	Pakistan	19.37
8.	Republic of Korea	18.81
9.	Belgium	17.83
10.	Poland	17.47



5. Why would blockchain be so bad?
The existing data infrastucture could not be used

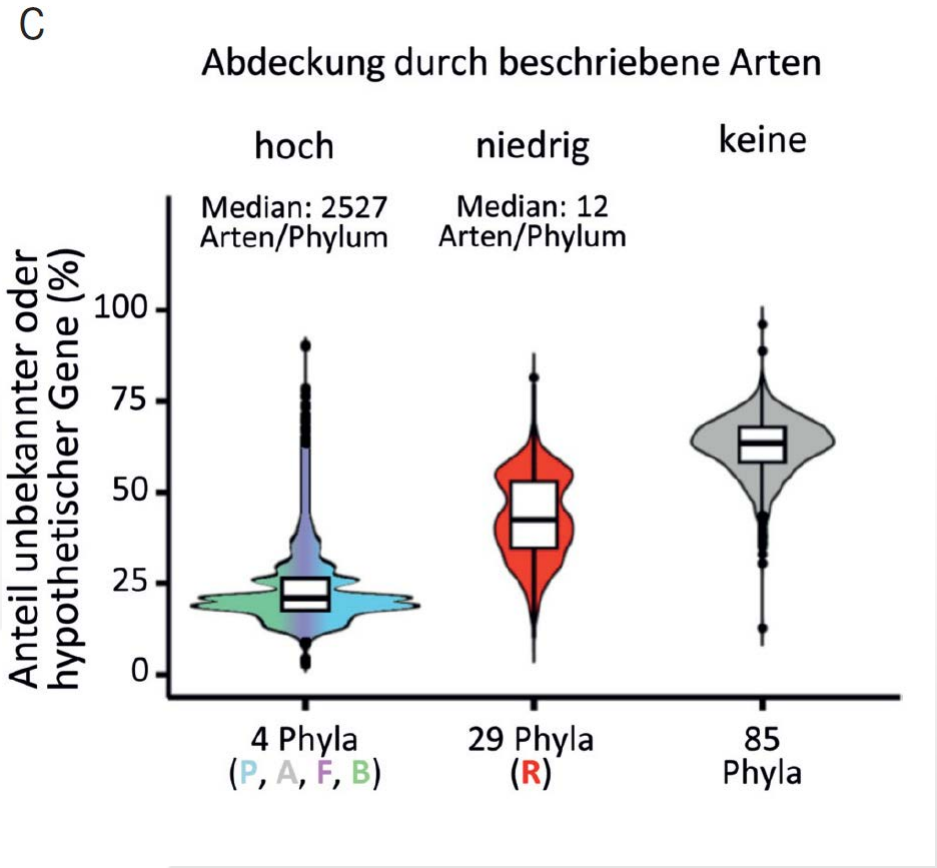


5. Why would a separate database be so bad? We need the FULL sequence dataset to understand new biodiversity!

Pop quiz:

What does this mean?

AACCGTCTACGGCCCGATCTC
CCTGCACCCTGGCCCCACCCT
TCCAAATGCGGCATATCCAGT



Sikorski, Overmann. BioSpektrum. July 2017.

Where did the WiLDSI project start (Aug. 2019)?

1. A framework decision on DSI will happen in China in 2020
2. DSI benefit sharing will need to be non-monetary (easy) & monetary (hard)
3. The current DSI (NSD) system is open, free, and central to biological research & should stay that way
4. We need a better system/deal (for everyone) than the NP

A wish list: a new DSI ABS system should be....

1. Compatible with existing free, open access INSDC
2. Administratively nimble or invisible for scientists
3. Integrated and NOT a stand-alone system (e.g., blockchain)
4. Income-generating in a painless way without explicit public sector funds
5. Clear in legal scope and compatible with CBD/NP and other legal mechanisms
6. Satisfy the demands of the developing world
7. Stop being „DSI“ and become something else

Timeline

- Research (6 months)
  - Sept: Discuss ideas, assign sub-contracts
  - Nov: Integrate results
  - Feb: finalize scenarios
- Scientific community input (4 months)
  - Jan. 21-22: German scientific stakeholder workshop
  - March 10-11: EU scientific stakeholder workshop
  - April (TBD): scientific industry stakeholder workshop
- Socialize ideas (6 months)
  - March: release first project summary report
  - July: *Science* Policy Forum article planned
  - OEWG3 & COP15 Side events

Research areas & Day 2

1. How can you make money from an open system?
2. What could the open system (INSDC) do to adapt without compromising open access? (use policies, country designation, legal permits)
3. What are the non-monetary benefits that arise from the open system?
4. What do scientists in developing countries think about the open system?

→ Scenario development

CBD DSI Study 2/3 co-authors

Fabian Rohden
Germany

Sixing Huang
Germany / China

Gabriele Dröge
Germany

Contributing authors (alphabetical):

- Katharine Barker, USA
- Walter G. Berendsohn, Germany
- Jonathan A. Coddington, USA
- Manuela da Silva, Brazil
- Jörg Overmann, Germany
- Ole Seberg, Denmark
- Michelle van der Bank, South Africa
- Xun Xu, China

Convention on Biological Diversity

WiLDSI Steering Committee

Dr. Guy Cochrane
EMBL-EBI

Database expertise

Dominic Muyldermans
ABSint

Prof. Dr. Claudia Seitz
Uni Bonn

Legal expertise

Dr. Amber H. Scholz
DSMZ

Dr. Jens Freitag
IPK Gaterseleben

Fabian Rohden
DSMZ (former)

Dr. Upneet Hillebrand
DSMZ

Scientific expertise

Dr. Carmen Richerzhagen
DLR/DIE

Dr. Jean Carlos Rodriguez
DIE

Development expertise

Prof. Dr. Esther van Zimmeren
Uni Antwerp

Torsten Thiele
IASS

Financial expertise

Thomas Greiber
BfN

Marliese von den Driesch
BLE

Regulatory expertise (external advisors)

Thank you!

Bundesministerium
für Bildung
und Forschung

Amber.h.scholz@dsmz.de

PS – Dinner is hosted!
Get to work on the questions…. :-)

Dr. Hilke Marie Püschner

Prof. Dr. Jörg Overmann