What are the applications of DSI?

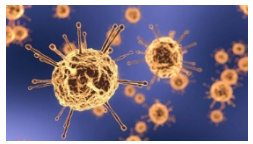
DSI 101 webinar
December 1, 2020





Dr. Amber Hartman Scholz, Deputy to the Director

DSI enables a LOT of public good



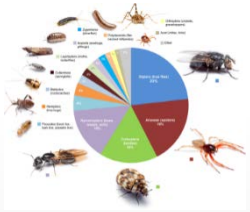
Epidemiology

Wildlife conservation

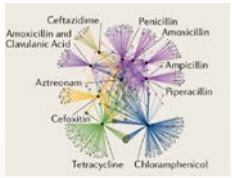
Pollutant effects in fish

Reducing waste in pig feed

PANGAEA: Data Publisher for Earth & Environmental Science



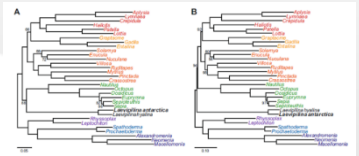
Insect taxonomy & Loss of pollinators



Antimicrobial resistance

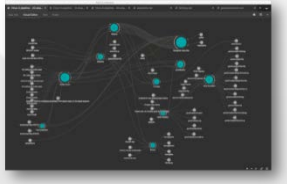
Tropical ecosystem research & climate change adaptation

Poaching & Wildlife trade



Mollusk barcoding

Fungal diversity & nutrient cycling



Viral metagenomics & bioprospecting

Freshwater Biodiversity and Evolution

3 GOALS OF THE CBD → 3 EXAMPLES

FOR EACH EXAMPLE:

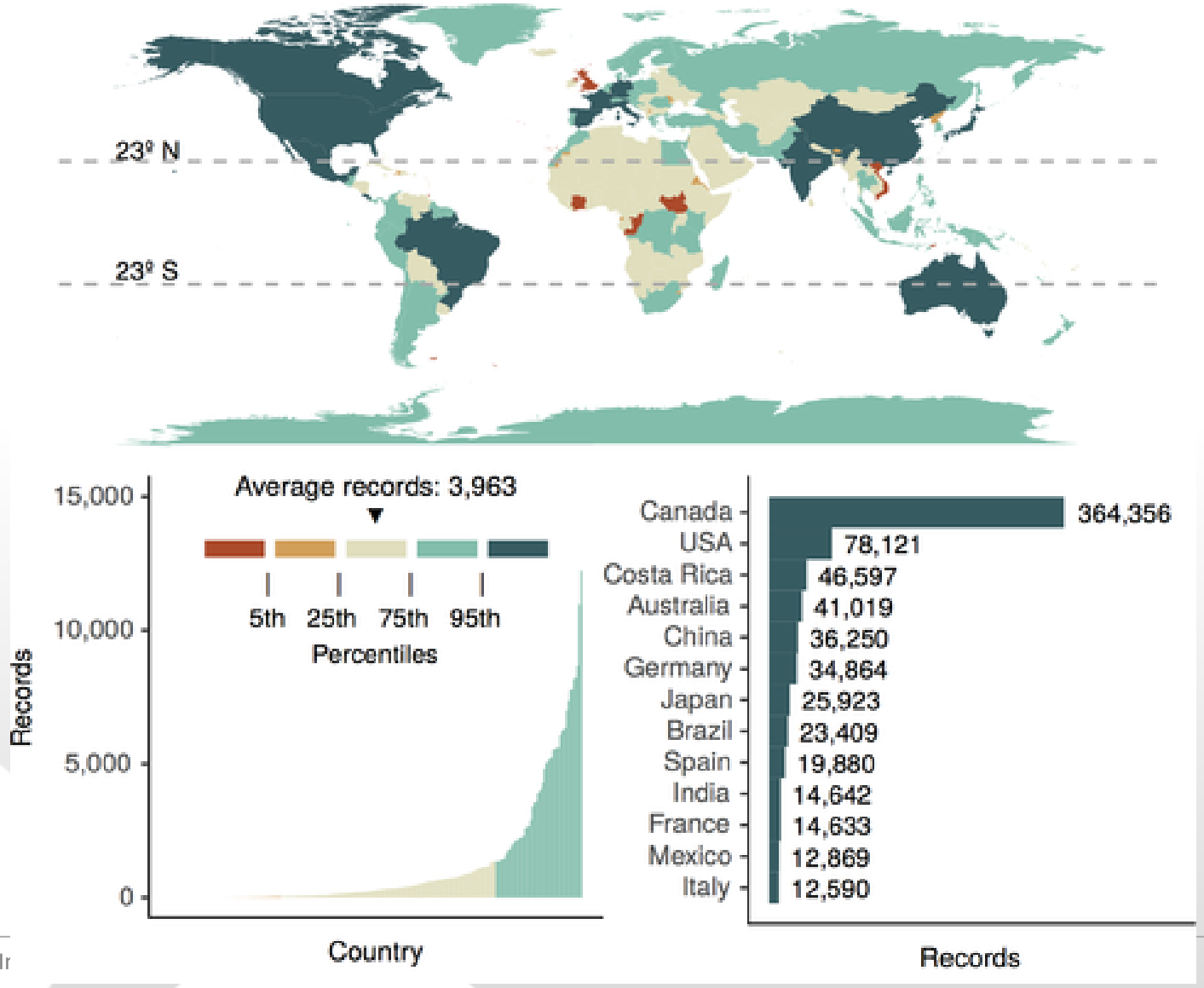
WHAT IS DSI USED FOR?

HOW WAS IT USED?

CBD GOAL 1. DSI FOR BIODIVERSITY CONSERVATION

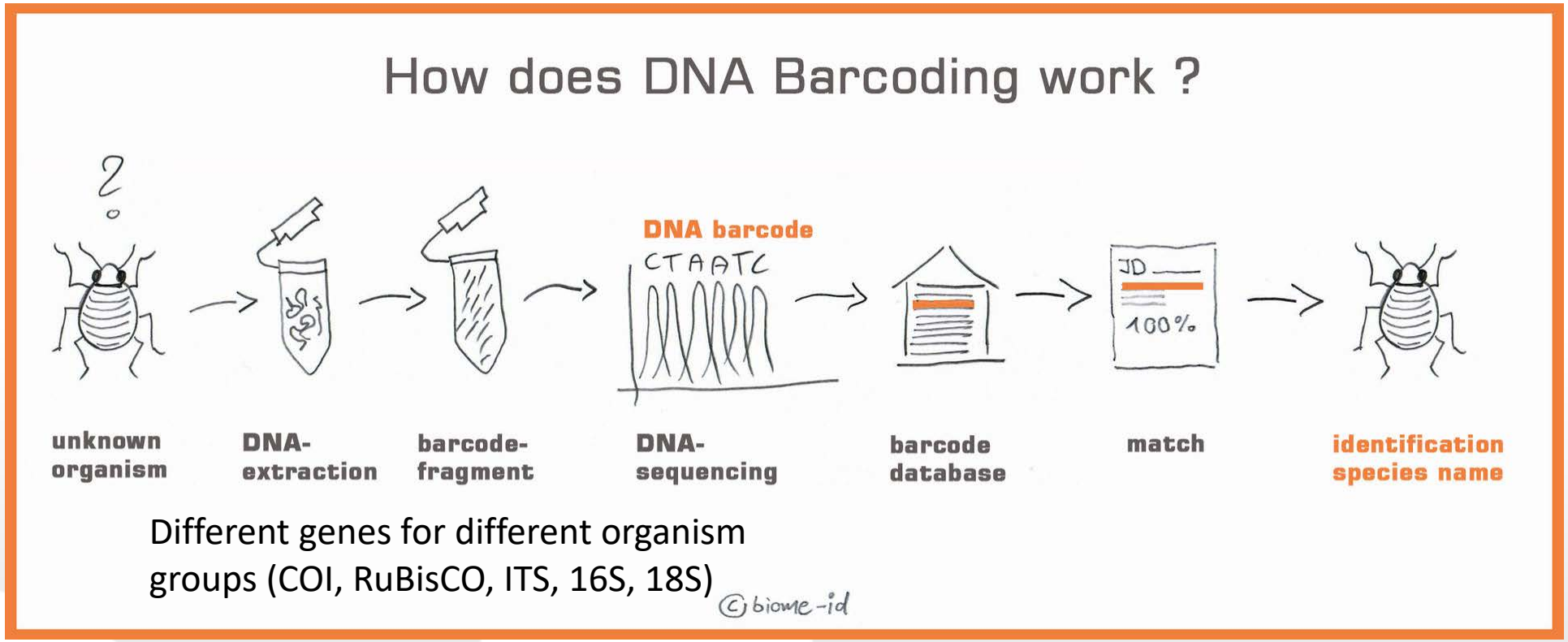
IDENTIFY SPECIES → BARCODING
UNKOWN SEQUENCES → BLAST

You want to conserve biodiversity? You can only protect what you know you have. How? Barcoding.



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

Step 1. Identifying biodiversity through barcoding

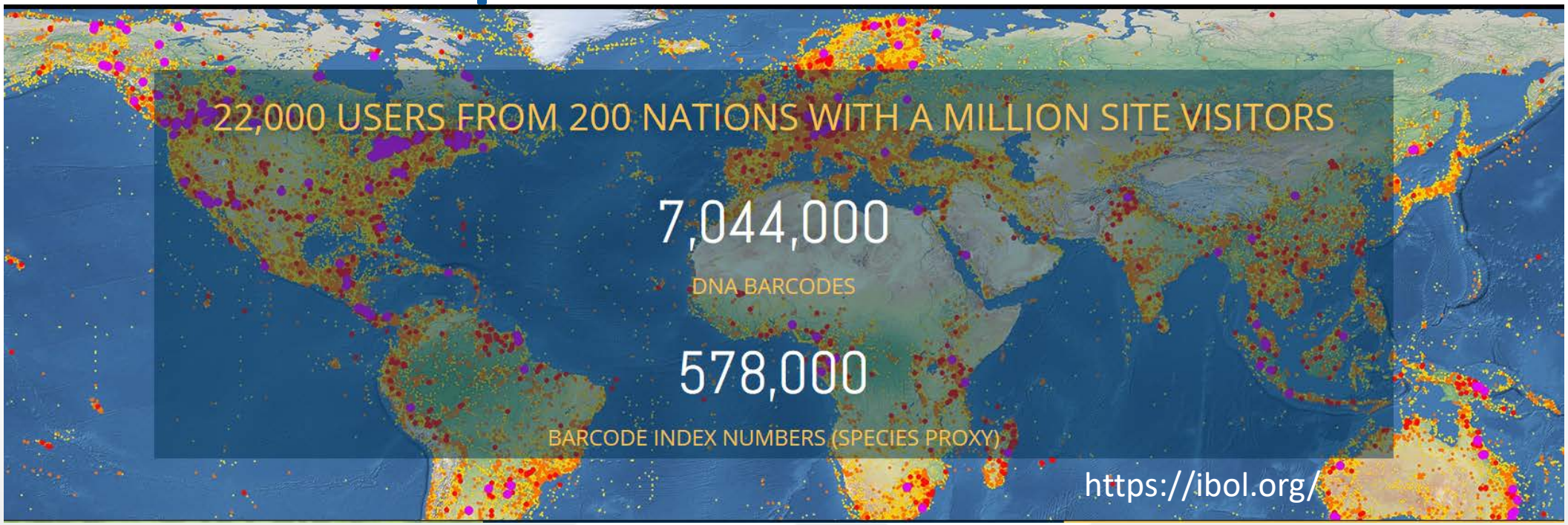


Barcoding is not looking at the *function* of the genes but rather variations in the letters to give a "*name tag*"

https://www.biome-id.com/english-1/molecular-services-1/dna-barcoding/

Large barcoding databases

iBOL Example



Other than identifying new species, barcoding can also be used to identify:

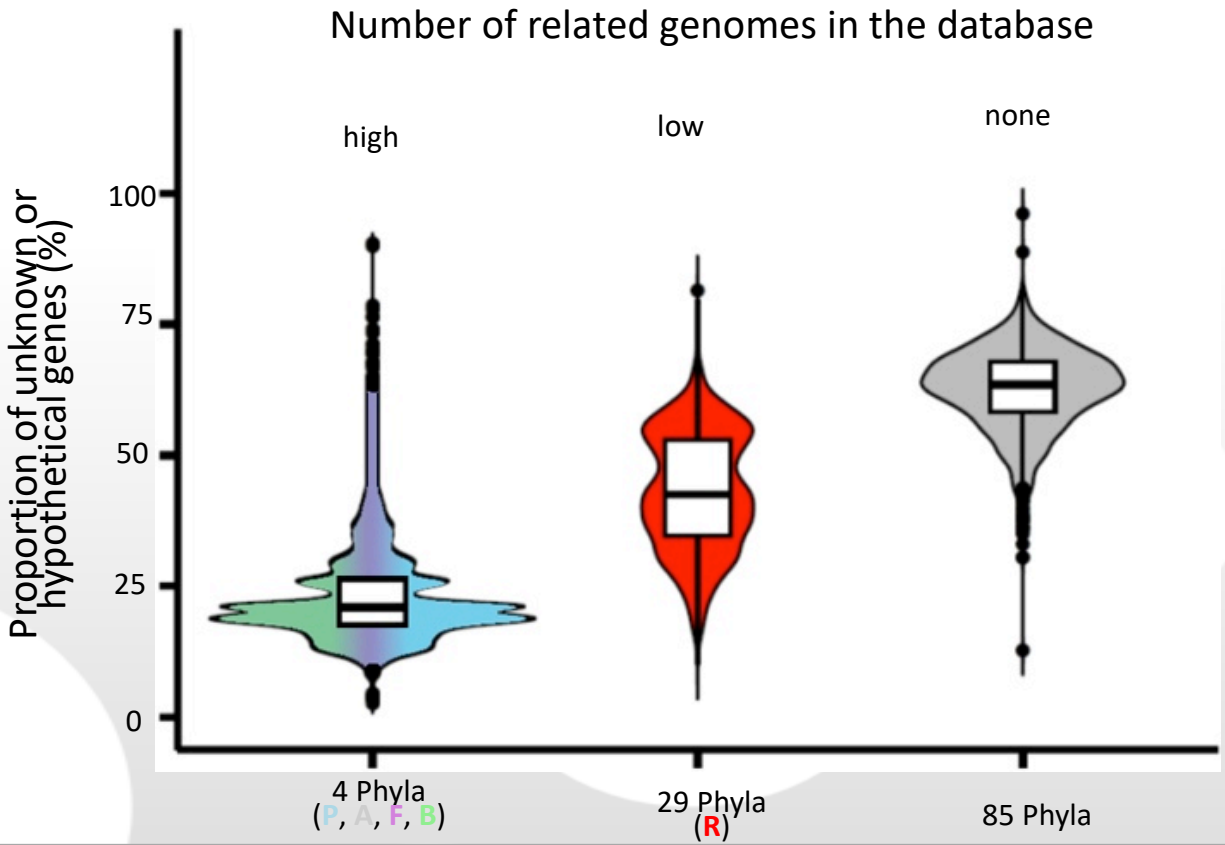
- **Invasive species**
- **Pests & pathogens**
- **Contaminants**

Step 2. Understanding new, unknown biodiversity by BLASTing it against all DSI in INSDC

Pop quiz:

What does this mean?

AACCGTCTACGGCCCGATCTCCCTGCACCCTGGCCCCACCCTTCCAAATGCGGCATATCCAGT



Unknown genes = unusable DNA baby babble.

The bigger and more complete the DSI dataset is, the better the chance we will understand novel biodiversity.

This means science needs one big, central database!

Sikorski, Overmann. BioSpektrum. July 2017.

CBD GOAL 2. SUSTAINABLE USE OF BIOLOGICAL DIVERSITY

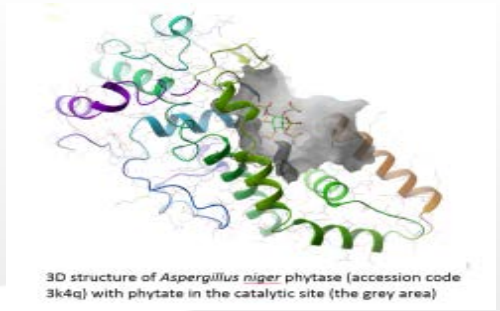
COMPARING DSI TO REDUCE POLLUTION

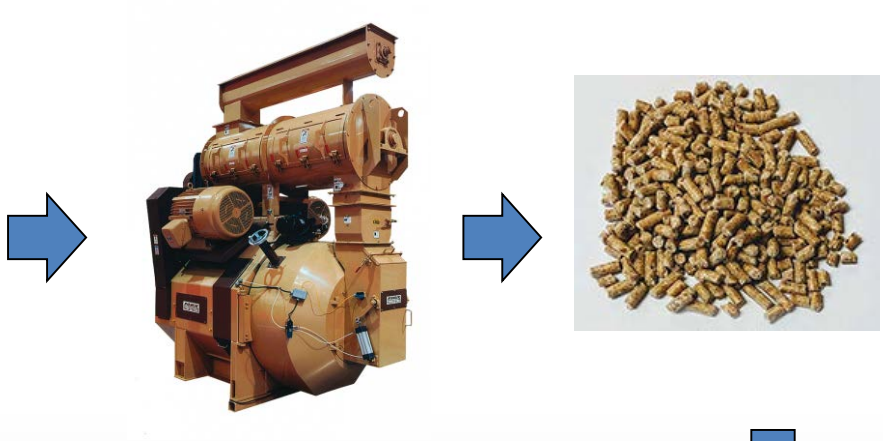
THANKS TO MARKUS WYSS, DSM

Goal: improve water quality, decrease phosphorous pollution

https://www.usgs.gov/special-topic/water-science-school/science/phosphorus-and-water?qt-science_center_objects=0#qt-science_center_objects

Phytases break down when heated





Many sequences contributed a little bit to a more heat stable phytase

Protein Engineering vol.13 no.1 pp.49–57, 2000

From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase

Martin Lehmann[1], Dirk Kostrewa, Markus Wyss, Roland Brugger, Allan D'Arcy, Luis Pasamontes and Adolphus P.G.M. van Loon

F.Hoffmann-La Roche Ltd, Grenzacherstrasse 124, CH-4070 Basel, Switzerland

[1]To whom correspondence should be addressed; email: Martin.Lehmann@Roche.com

Naturally-occurring phytases having the required level of thermostability for application in animal feeding have not been found in nature thus far. We decided to *de novo* construct consensus phytases using primary protein sequence comparisons. A consensus enzyme based on 13 fungal phytase sequences had normal catalytic properties, but showed an unexpected 15–22°C increase in unfolding temperature compared with each of its parents. As a first step towards understanding the molecular basis of

and 63°C (see below), we were interested in developing a rapid procedure to increase the unfolding temperature and thus the thermostability of phytases.

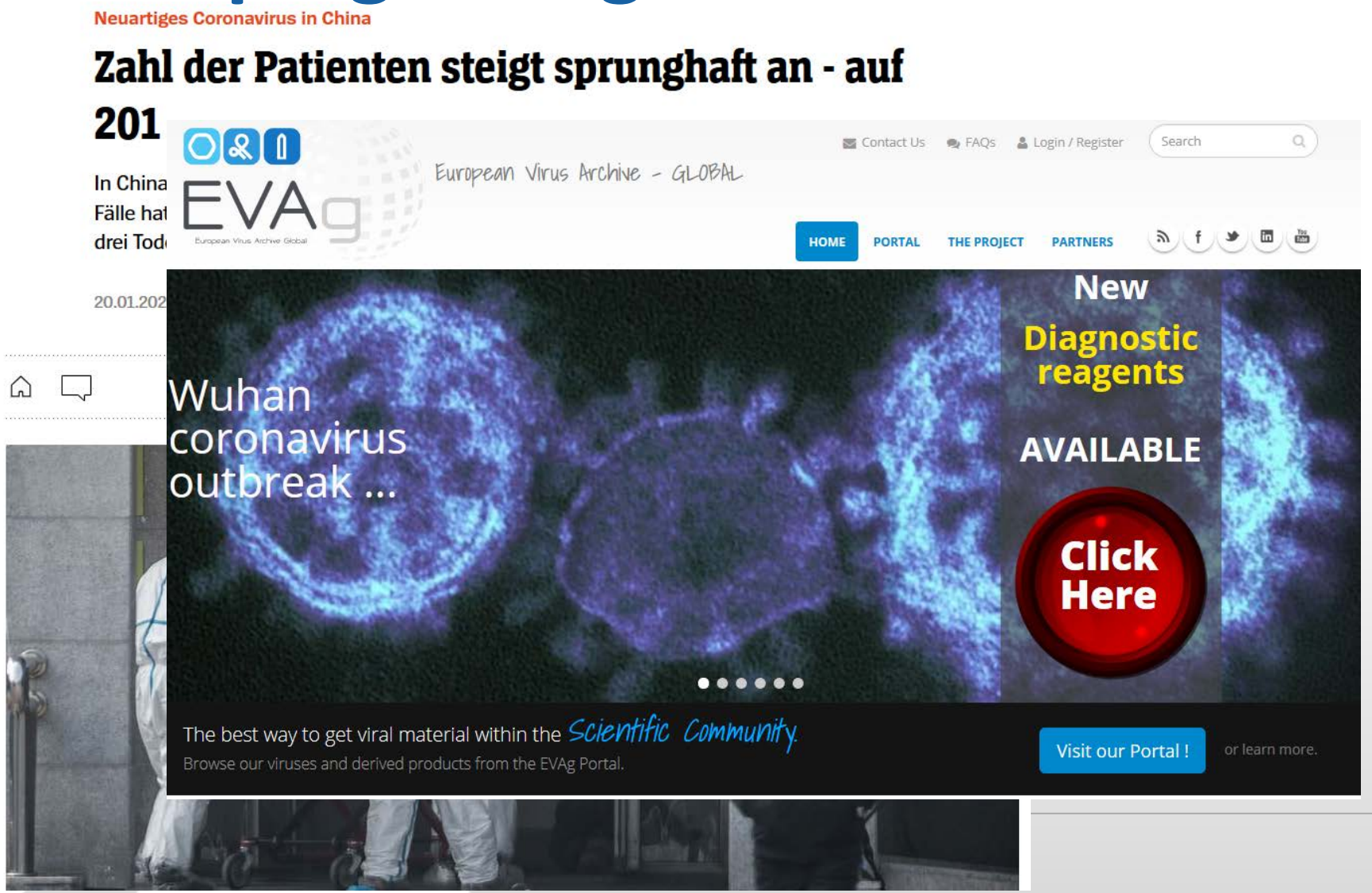
Increasing the thermostability of an enzyme usually requires combining multiple amino acid exchanges, each of which slightly increases the unfolding temperature of the protein. The main problem, however, is the identification of the relevant amino acid residues. In general terms mutations that increase thermostability may, for example, result in formation of hydrogen bonds, salt or disulphide bridges, increase the hydrophobic packing or the α-helix or β-sheet stability or stabilize β-turns or flexible termini or loops (for review see Jaenicke *et al.*, 1996). Despite extensive knowledge of the general mechanisms governing protein stability (Dill *et al.*, 1989; Dill, 1990; Fersht and Serrano, 1993; Matthews, 1993; Cordes *et al.*, 1996) no rapid and reliable procedures are available for increasing the thermostability of a given protein. For successful thermostabil-

The R&D process is non-linear. Lots of actors contribute a little bit. Determining what is decisive for the value-generation is difficult!

CBD GOAL 3. FAIR & EQUITABLE BENEFIT SHARING

USING SYNTHETIC BIOLOGY TO FIGHT A PANDEMIC FAST

Goal: respond to a pandemic by developing a diagnostic kit.



1. Chinese researchers upload SARS-CoV-2 DSI to INSDC on January 10, 2020

GenBank ▾ Send to: ▾

Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome

NCBI Reference Sequence: NC_045512.2

FASTA Graphics

Go to: ⊙

```
LOCUS       NC_045512              29903 bp ss-RNA     linear   VRL 18-JUL-2020
DEFINITION  Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1,
            complete genome.
ACCESSION   NC_045512
VERSION     NC_045512.2
DBLINK      BioProject: PRJNA485481
KEYWORDS    RefSeq.
SOURCE      Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)
  ORGANISM  Severe acute respiratory syndrome coronavirus 2
            Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes;
            Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae;
            Betacoronavirus; Sarbecovirus.
REFERENCE   1  (bases 1 to 29903)
  AUTHORS   Wu,F., Zhao,S., Yu,B., Chen,Y.M., Wang,W., Song,Z.G., Hu,Y.,
            Tao,Z.W., Tian,J.H., Pei,Y.Y., Yuan,M.L., Zhang,Y.L., Dai,F.H.,
            Liu,Y., Wang,Q.M., Zheng,J.J., Xu,L., Holmes,E.C. and Zhang,Y.Z.
  TITLE     A new coronavirus associated with human respiratory disease in
            China
  JOURNAL   Nature 579 (7798), 265-269 (2020)
   PUBMED   32015508
  REMARK    Erratum:[Nature. 2020 Apr;580(7803):E7. PMID: 32296181]
REFERENCE   2  (bases 13476 to 13503)
  AUTHORS   Baranov,P.V., Henderson,C.M., Anderson,C.B., Gesteland,R.F.,
            Atkins,J.F. and Howard,M.T.
  TITLE     Programmed ribosomal frameshifting in decoding the SARS-CoV genome
  JOURNAL   Virology 332 (2), 498-510 (2005)
   PUBMED   15680415
REFERENCE   3  (bases 29728 to 29768)
  AUTHORS   Robertson,M.P., Igel,H., Baertsch,R., Haussler,D., Ares,M. Jr. and
            Scott,W.G.
  TITLE     The structure of a rigorously conserved RNA element within the SARS
            virus genome
  JOURNAL   PLoS Biol 3 (1), e5 (2005)
   PUBMED   15630477
REFERENCE   4  (bases 29609 to 29657)
  AUTHORS   Williams,G.D., Chang,R.Y. and Brian,D.A.
  TITLE     A phylogenetically conserved hairpin-type 3' untranslated region
            pseudoknot functions in coronavirus RNA replication
  JOURNAL   J Virol 73 (10), 8349-8355 (1999)
   PUBMED   10482585
REFERENCE   5  (bases 1 to 29903)
  CONSRTM   NCBI Genome Project
  TITLE     Direct Submission
```

https://www.ncbi.nlm.nih.gov/nuccore/NC045512

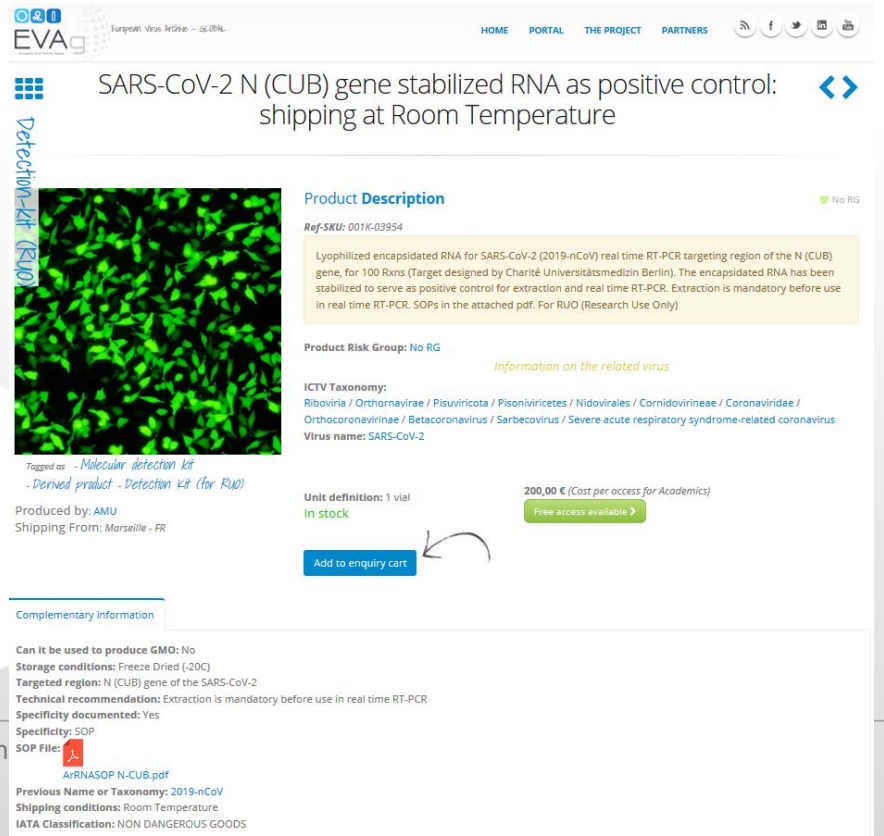
2. German researchers synthesize SARS-CoV-2

Driouich et al. 2018

“A virus is a piece of bad news wrapped in protein“

Dorothy H. Crawford

Direct synthesis is only possible for viruses!



3. Identify a stable but unique gene for diagnosis

- Unique: only possible by comparing to other viruses.
- Stable: only possible by comparing to multiple patient samples of same virus. (Large dataset or prior knowledge needed)

4. Diagnostic kits shipped Feb.9
Commercial test kits available 1-2 months later.

Bundle

Primers, Probe and encapsidated RNA pos. control - SARS-CoV-2 N (CUB) gene, small packaging

Tagged as - Derived product - Bundle - Molecular detection kit - Detection Kit (for RUO)

Produced by: AMU
Shipping From: Marseille - FR

Product **Description**

No RG

Ref-SKU: 001B-03965

Lyophilized encapsidated RNA for SARS-CoV-2 real time RT-PCR targeting region of the N(CUB) gene (for 100 Rxns) AND Lyophilized ready-to-use primers and probe (Lyoph-P&P) for real time RT-PCR (96 rxns in the form of 4x24 rnxs). Target, primers and probe designed by Charité (CUB) Berlin. SOPs in the attached pdf.

Disclaimer:

LIST OF

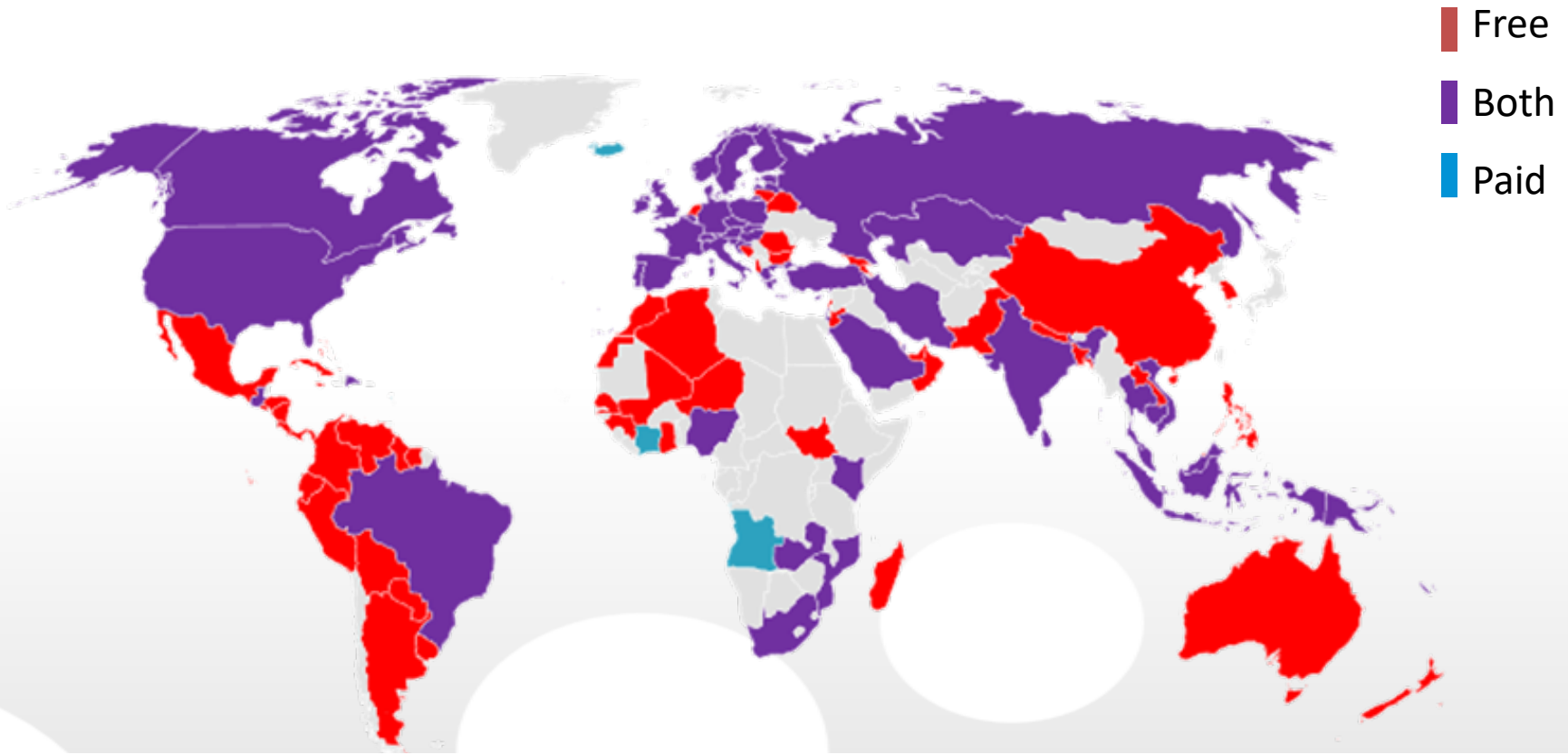
Ref-SKU
001K-03964
001K-03954

Product Risk

ICTV Taxonon
Riboviria / Ort
Orthocoronav
Virus name: S

https://www.youtube.com/watch?v=tgyzdgf66eM

5. Send it around the world for FREE. (Non-) monetary benefit-sharing?



- All customers had the free option available but some declined.
- African and Oceania not as severely affected by covid19 pandemic
- Funded by the EU Commission = €### million.
- 114 countries have received diagnostic kits to-date.

DSMZ

Implications for DSI policymakers

1. DSI is essential for different types of public good and all 3 CBD goals.
2. Modern-day biodiversity research depends on DSI
   - Species identification (barcoding)
   - Learning about unknown biology
3. DSI can be commercially utilized but it is a complex, iterative process based on huge datasets.
4. Determining benefit-sharing “credit“ is very hard. (Lots of things contribute. Some more, some less.)

Thank you!

amber.h.scholz@dsmz.de

WiLDSI

Bundesministerium für Bildung und Forschung

EVAg

Horizon 2020 European Union Funding for Research & Innovation

https://www.dsmz.de/collection/nagoya-protocol/digital-sequence-information