基因體學於研究生物多樣性的進展

Isheng Jason Tsai 蔡怡陞
生態基因體學實驗室

Biodiversity Research Center, Academia Sinica
中央研究院生物多樣性中心
2023.10.16

It is an exciting time to be in (to do research)!

There’s so much that I want to share with you

What is genomics?

Genomics

"is an interdisciplinary field of biology focusing on the structure, function, evolution, mapping, and editing of genomes." (wiki)

There's also these disciplines …

Bioinformatics

Systems Biology?

Computational Biology

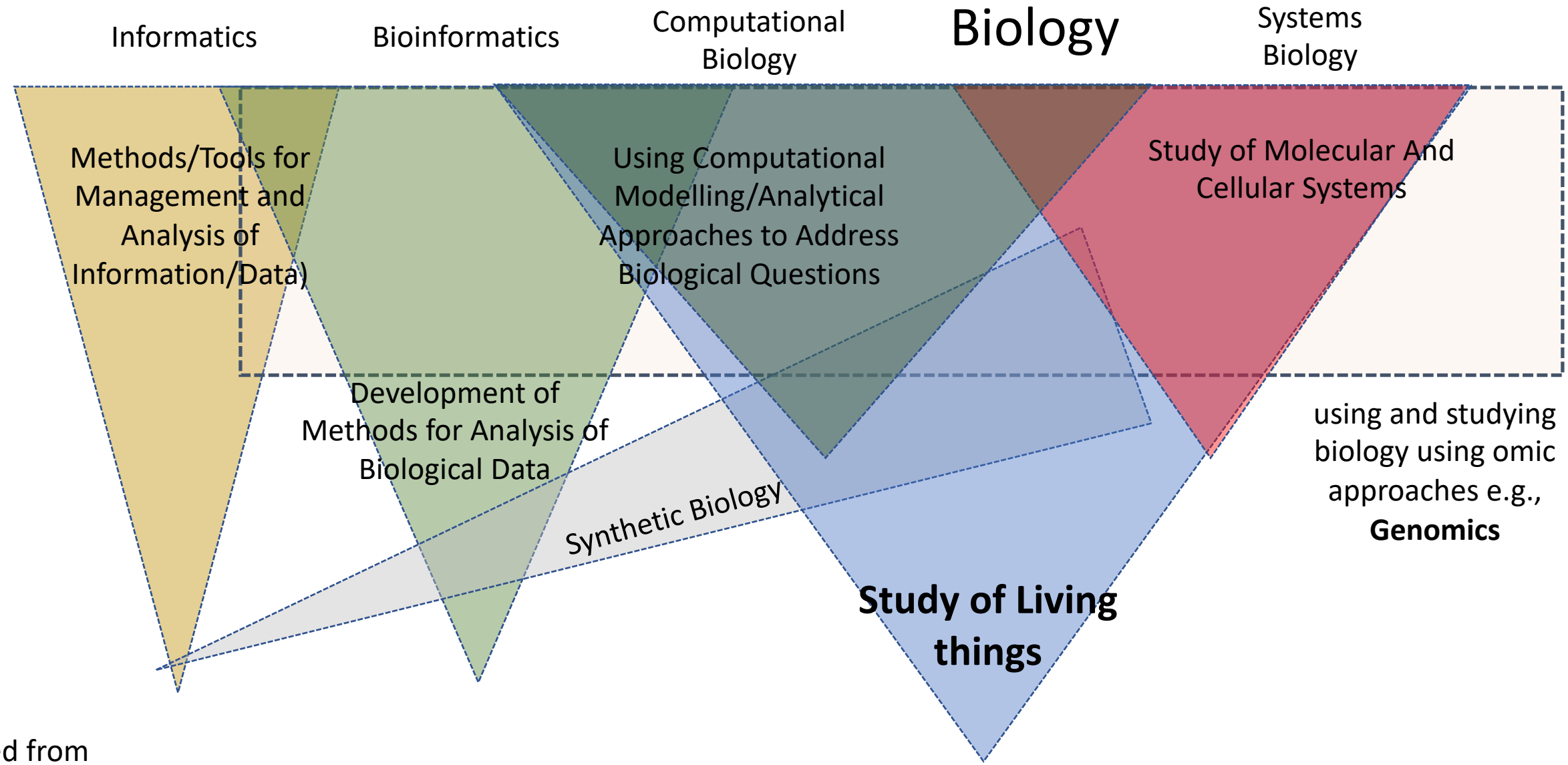
Synthetic Biology

Biology?

Any others? How are they different?

[1] https://irp.nih.gov/catalyst/v19i6/systems-biology-as-defined-by-nih
https://www.genome.gov/genetics-glossary/Bioinformatics

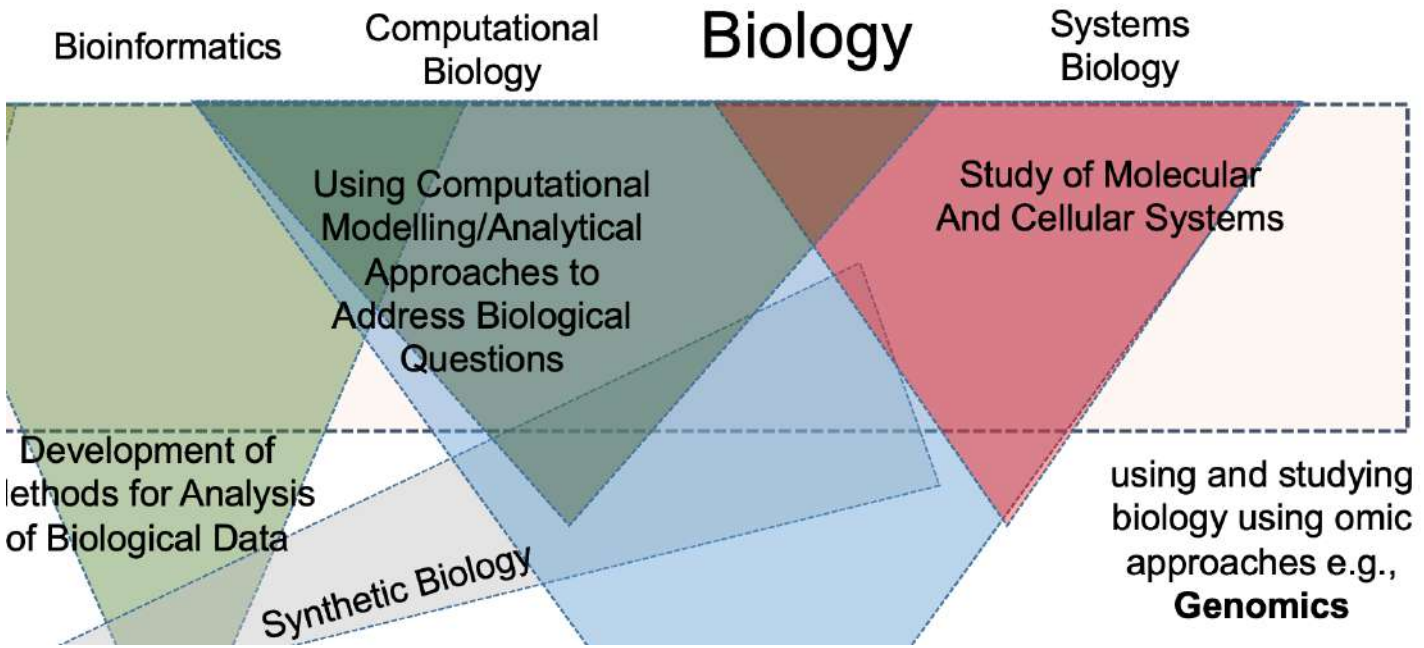
Overlapping Fields



Adapted from
https://www.youtube.com/watch?v=lJzybEXmIj0

Course focus

- Not an algorithm class (although some need mentioned)
- Not a data science class
- About **nine** weeks of omics, **five** weeks of practical class in R, **one** week of discussion one week of exam



Students are expected to learn the **core understanding** behind omic methods and know where to start to integrate these approaches into their own research.

Grade

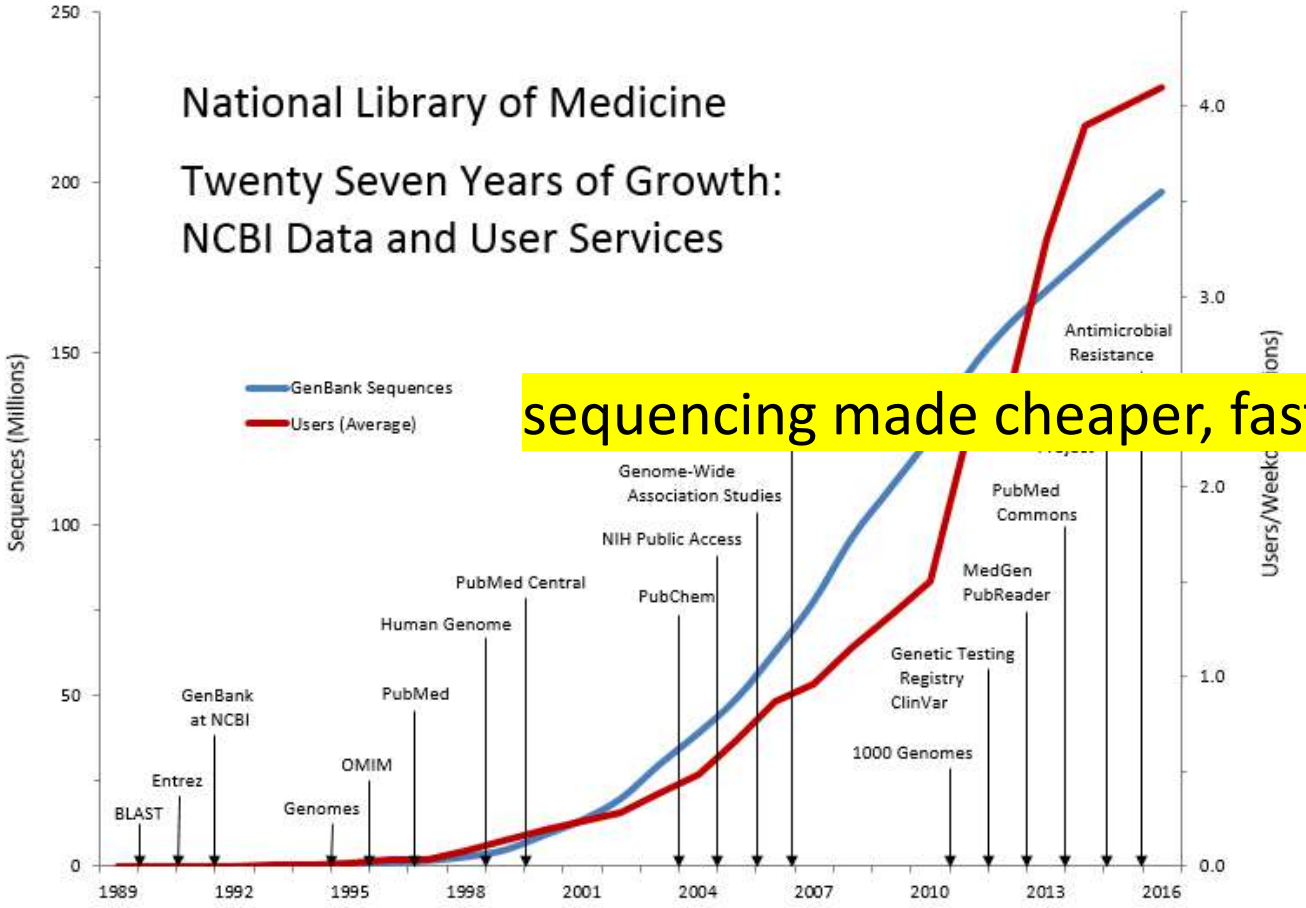
Based on attendance (20%), homework (30%) and final exam (50%)

Calculating the economic impact of the Human Genome Project

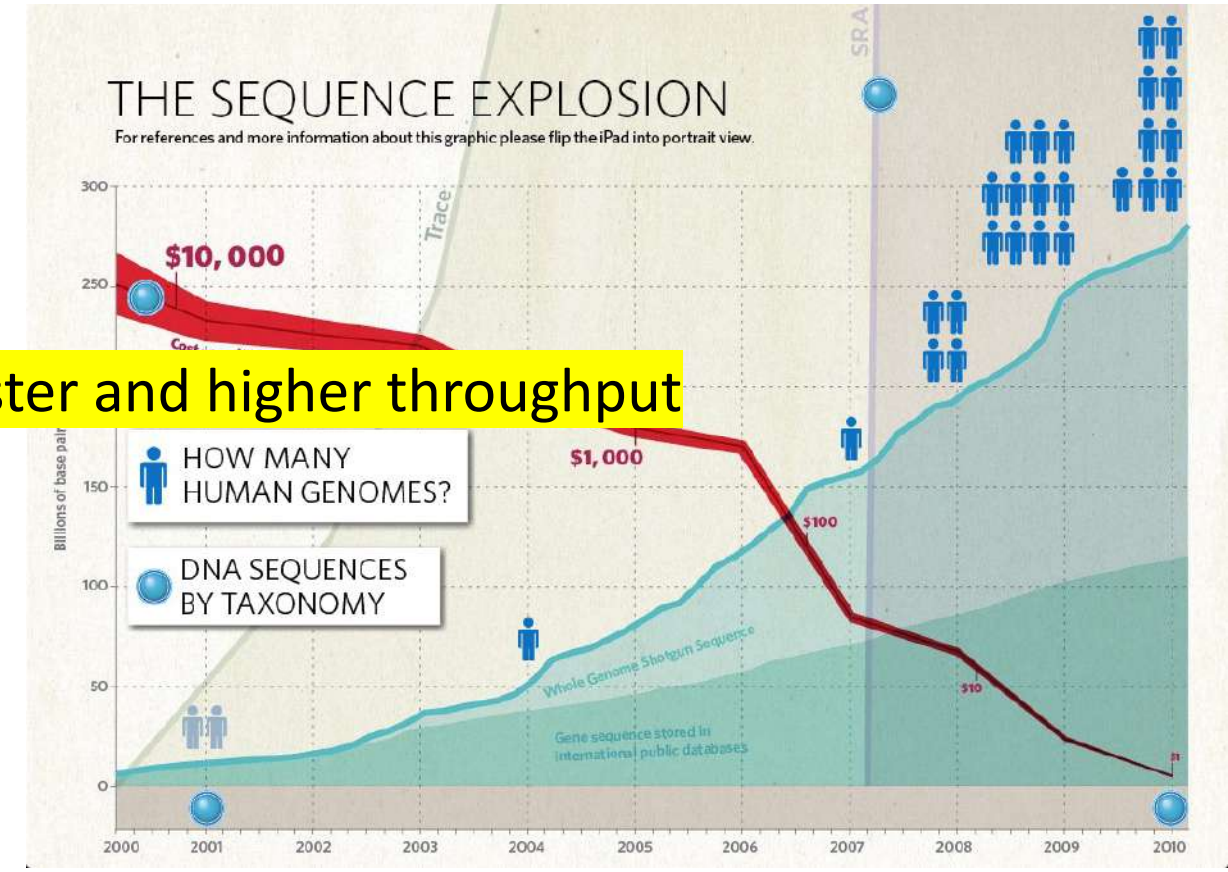
Public funding of scientific R&D has a significant positive impact on the wider economy, but quantifying the exact impact of research can be difficult to assess. A new report by research firm Battelle Technology Partnership Practice estimates that **between 1988 and 2010, federal investment in genomic research generated an economic impact of $796 billion**, which is impressive considering that Human Genome Project (HGP) spending **between 1990-2003 amounted to $3.8 billion**. This figure equates to a return on investment (ROI) of 141:1 (that is, every $1 invested by the U.S. government generated $141 in economic activity). The report was commissioned by Life Technologies Foundation.

https://www.genome.gov/27544383/calculating-the-economic-impact-of-the-human-genome-project/

2000-2010s – Second generation sequencing and associated challenges



sequencing made cheaper, faster and higher throughput



https://www.nlm.nih.gov/about/https://www.nlm.nih.gov/about/2018CJ.html

http://www.nature.com/news/2010/100331/full/464670a.html

A brief history of bioinformatics
Jeff Gauthier, Antony T Vincent, Steve J Charette, Nicolas Derome
Briefings in Bioinformatics (2018) https://doi.org/10.1093/bib/bby063

Oxford Nanopore

Key	SmidgION	Flongle	MinION	GridION	PromethION
System Price	TBC	Included in $5K Starter Pack	Included in $1K Starter Pack	Included in $50K Starter Pack	Included in $135K Starter Pack
Number of channels	200 channels	128 channels	512 channels	5 x 512 = 2,560*	48 x 3,000* = 144,000
Per flow cell Current Data – Max Data	TBC	1 - 3.3 Gb	17 - 40 Gb	17 - 40 Gb	125 - 311 Gb
Per Device Current Data – Max Data				85 - 200 Gb	3/6 - 20 Tb
Price per Gb Current Data – Max Data	**TBC**	**$90 - $30**	**$30 - $12.5**	**$17.5 - $7.5**	**$5 - $2**

Oxford Nanopore – how it works

Introduction to nanopore *
https://vimeo.com/297106166

Voltrax
https://vimeo.com/297106291

Sequencing for farmers
https://vimeo.com/294216876

Rainforest
https://www.youtube.com/watch?v=6RRSxWtJPUw

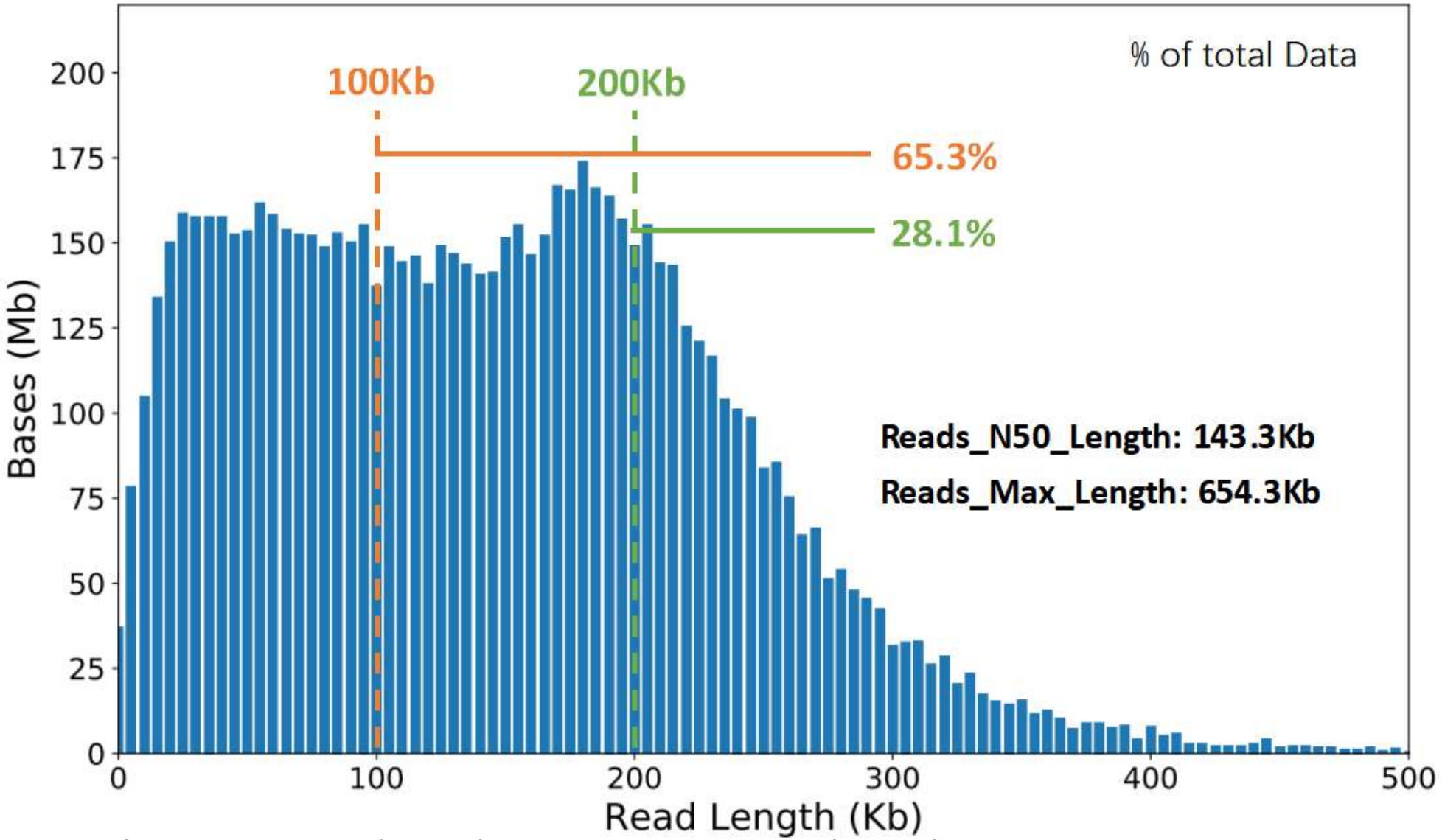
@ Oceans
https://vimeo.com/294744892

From Extreme to everyday
https://www.youtube.com/watch?v=tQ_oo7_36r8

Reference
https://nanoporetech.com/how-it-works

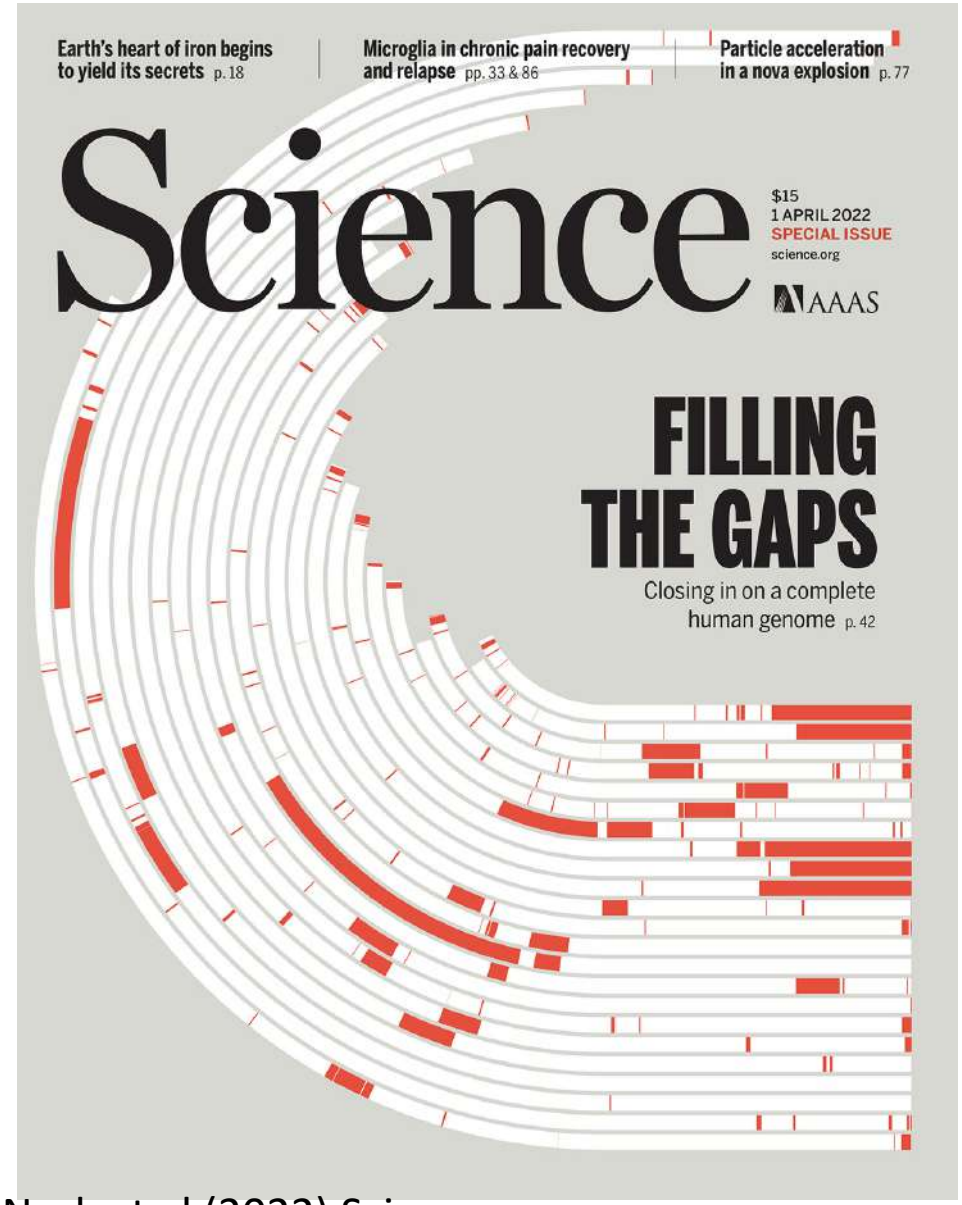
Nanopore Sequencing of Ebola Viruses Under Outbreak Conditions
https://www.youtube.com/watch?v=SYBzPEoENWI ; https://www.nature.com/articles/nature16996

Read length and capacity go beyond



https://twitter.com/GrandOmics_Intl/status/1188724248480608257/photo/1

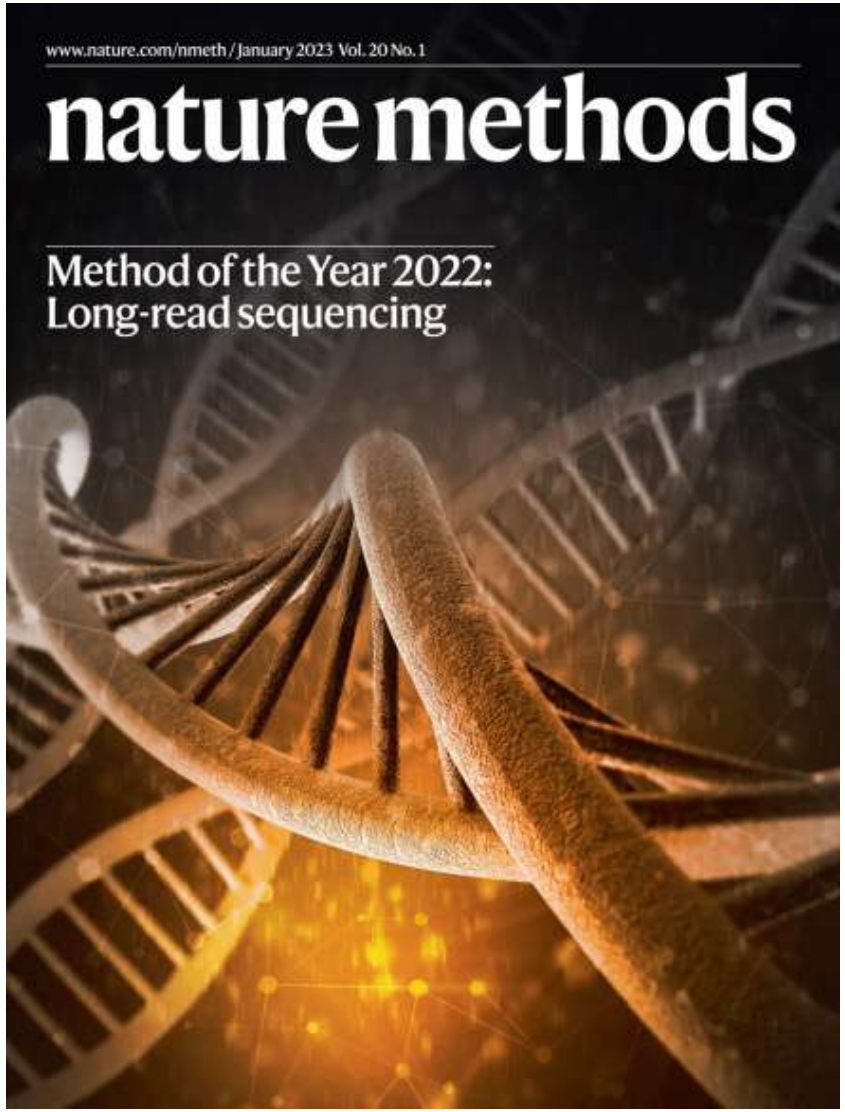
(Real) Completion of human genome



STATISTICS	GRCH38	T2T-CHM13	DIFFERENCE (±%)
Summary			
Assembled bases (Gbp)	2.92	3.05	+4.5
Unplaced bases (Mbp)	11.42	0	−100.0
Gap bases (Mbp)	120.31	0	−100.0
Number of contigs	949	24	−97.5
Contig NG50 (Mbp)	56.41	154.26	+173.5
Number of issues	230	46	−80.0
Issues (Mbp)	230.43	8.18	−96.5
Gene annotation			
Number of genes	60,090	63,494	+5.7
Protein coding	19,890	19,969	+0.4
Number of exclusive genes	263	3,604	
Protein coding	63	140	
Number of transcripts	228,597	233,615	+2.2
Protein coding	84,277	86,245	+2.3
Number of exclusive transcripts	1,708	6,693	
Protein coding	829	2,780	
Segmental duplications			
Percentage of segmental duplications (%)	5.00	6.61	
Segmental duplication bases (Mbp)	151.71	201.93	+33.1
Number of segmental duplications	24097	41528	+72.3
RepeatMasker			
Percentage of repeats (%)	51.89	53.94	
Repeat bases (Mbp)	1,516.37	1,647.81	+8.7
Long interspersed nuclear elements	626.33	631.64	+0.8
Short interspersed nuclear elements	386.48	390.27	+1.0
Long terminal repeats	267.52	269.91	+0.9
Satellite	76.51	150.42	+96.6
DNA	108.53	109.35	+0.8
Simple repeat	36.5	77.69	+112.9
Low complexity	6.16	6.44	+4.6
Retroposon	4.51	4.65	+3.3
rRNA	0.21	1.71	+730.4

Nurk et al (2022) Science

Very repetitive sequences

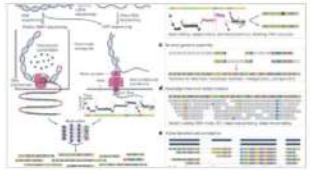


Comment
12 Jan 2023
Nature Methods

Approaching complete genomes, transcriptomes and epi-omes with accurate long-read sequencing

The year 2022 will be remembered as the turning point for accurate long-read sequencing, which now establishes the gold standard for speed and accuracy at competitive costs. We discuss the key bioinformatics techniques needed to power long reads across application areas and close with our vision for long-read sequencing over the coming years.

Sam Kovaka, Shujun Ou ... Michael C. Schatz

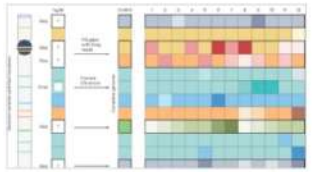


Comment
12 Jan 2023
Nature Methods

Comprehensive variant discovery in the era of complete human reference genomes

Advances in long-read sequencing technologies have broadened our understanding of genetic variation in the human population, uncovered new complex structural variants and offered an opportunity to elucidate new variant associations with disease.

Monika Cechova & Karen H. Miga

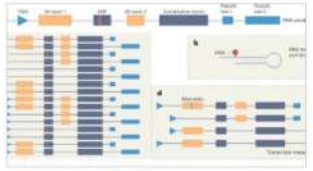


Comment
12 Jan 2023
Nature Methods

The variables on RNA molecules: concert or cacophony? Answers in long-read sequencing

Long-read sequencing has become a widely employed technology that enables a comprehensive view of RNA transcripts. Here, we discuss the importance of long-read sequencing in interpreting the variables along RNA molecules, such as polyadenylation sites, transcription start sites, splice sites and other RNA modifications. In addition, we highlight the history of short-read and long-read technologies and their advantages and disadvantages, as well as future directions in the field.

Careen Foord, Justine Hsu ... Hagen U. Tilgner

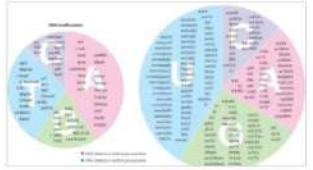


Comment
12 Jan 2023
Nature Methods

Long-read sequencing in the era of epigenomics and epitranscriptomics

As long-read sequencing technologies continue to advance, the possibility of obtaining maps of DNA and RNA modifications at single-molecule resolution has become a reality. Here we highlight the opportunities and challenges posed by the use of long-read sequencing technologies to study epigenetic and epitranscriptomic marks and how this will affect the way in which we approach the study of health and disease states.

Morghan C. Lucas & Eva Maria Novoa



Comment
12 Jan 2023
Nature Methods

Long-read metagenomics paves the way toward a complete microbial tree of life

Long-read sequencing has made closed microbial genomes a routine task, and the dramatic increase in quality and quantity now paves the way to a complete microbial tree of life through genome-centric metagenomics.

Mads Albertsen

https://www.nature.com/collections/eibbdadhga
Volume 20 Issue 1, January 2023 ; nature methods

First take home message: Sequencing are now much cheaper, longer and more accurate

First take home message: Sequencing are now much cheaper, longer and more accurate. We can now use this to study biodiversity!

生物多樣性 Biodiversity (Biological Diversity)

The Biological Diversity Crisis

Despite unprecedented extinction rates, the extent of biological diversity remains unmeasured

Certain measurements are crucial to our ordinary understanding of the universe. What, for example, is the mean diameter of the Earth? It is 12,742 kilometers. How many stars are there in the Milky Way? Approximately 10^{11}. How many genes are there in a small virus particle? There are 10 (in ϕX174 phage). What is the mass of an electron? It is 9.1×10^{-28} grams. And how many species of organisms are there on Earth? We do not know, not even to the nearest order of magnitude.

The Biological Diversity Crisis

https://www.jstor.org/stable/43310356

Edward O. Wilson
(1929-2021)
"Father of biodiversity"

愛德華・奧斯本・威爾遜
E. O.威爾森

Definition of term biodiversity

Over the weekend, two of the country's leading naturalists, E. O. Wilson and Tom Lovejoy, died a day apart. Wilson, who was perhaps best known for his work on ants, was a pioneer in the field of conservation biology; Lovejoy was one of the founders of the field. The two men were friends—part of an informal network that Wilson jokingly referred to as the "rain-forest mafia"—and there was something eerie about their nearly synchronous passing. "I'm trying very hard not to imagine a greater planetary message in the loss of these biodiversity pioneers right now," Joel Clement, a senior fellow at the Harvard Kennedy School's Belfer Center for Science and International Affairs, tweeted on Monday.

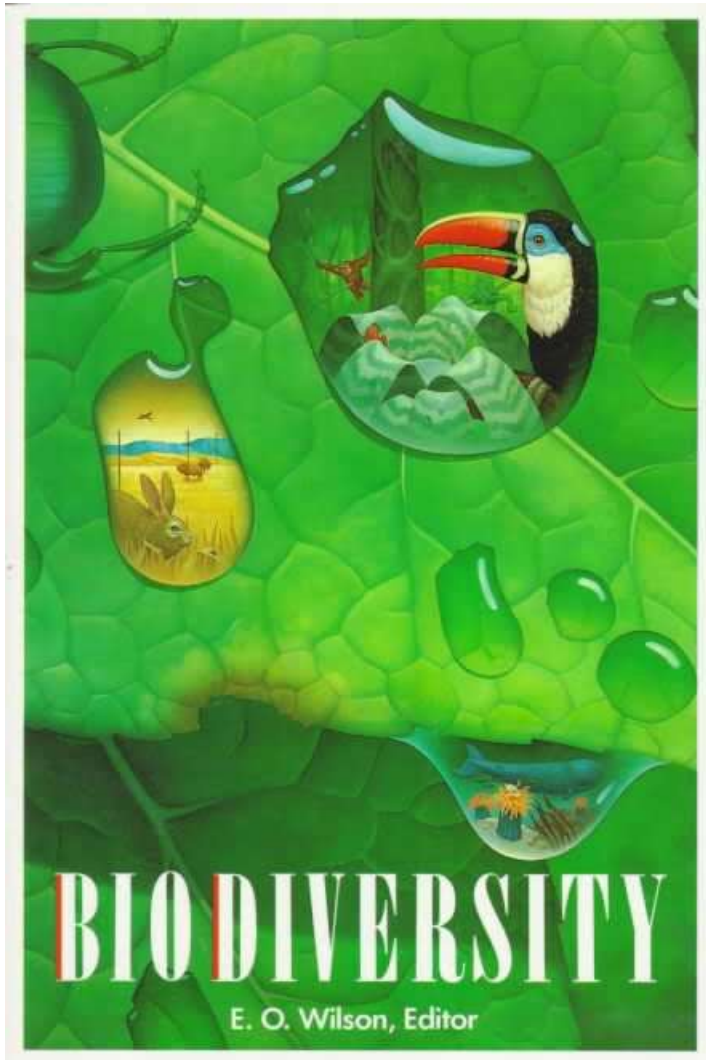
The two scientists first met in the mid-nineteen-seventies. At that point, Wilson was in his mid-forties, and teaching biology at Harvard. Lovejoy, a dozen years younger, was working for the World Wildlife Fund. Over lunch, they got to talking about where the W.W.F. should focus its efforts. They agreed that it should be in the tropics, because the tropics are where most species actually live. There wasn't a good term for what they were trying to preserve, so they tossed one around—"biological diversity"—and put it into circulation. "People just started using it," Lovejoy recalled, in an interview in 2015. (Later, the phrase would be shortened to "biodiversity.")

Thomas Lovejoy
(1931-2021)

Edward O. Wilson
(1929-2021)
"Father of biodiversity"

https://www.newyorker.com/news/daily-comment/honoring-the-legacy-of-e-o-wilson-and-tom-lovejoy

Definition of term biodiversity



1. CHALLENGES TO THE PRESERVATION OF BIODIVERSITY
2. HUMAN DEPENDENCE ON BIOLOGICAL DIVERSITY
3. DIVERSITY AT RISK: TROPICAL FORESTS
4. DIVERSITY AT RISK: THE GLOBAL PERSPECTIVE
5. THE VALUE OF BIODIVERSITY
6. HOW IS BIODIVERSITY MONITORED AND PROTECTED?
7. SCIENCE AND TECHNOLOGY: HOW CAN THEY HELP?
8. RESTORATION ECOLOGY: CAN WE RECOVER LOST GROUND?
9. ALTERNATIVES TO DESTRUCTION
10. POLICIES TO PROTECT DIVERSITY
11. PRESENT PROBLEMS AND FUTURE PROSPECTS
12. WAYS OF SEEING THE BIOSPHERE

https://nap.nationalacademies.org/read/989/chapter/1

Biodiversity (= biological diversity) definition

Biodiversity studies comprise the systematic examination of the full array of different kinds of organisms together with the technology by which the diversity can be maintained and used for the benefit of humanity. Current basic research at the species level focuses on the process of species formation, the standing levels of species numbers in various higher taxonomic categories, and the phenomena of hyperdiversity and extinction proneness. The major practical concern is the massive extinction rate now caused by human activity, which threatens losses in the esthetic quality of the world, in economic opportunity, and in vital ecosystem services.

目前，生物多樣性研究是系統地探索各種生物，並研究如何保持這種多樣性，讓它為人類帶來好處。研究集中在物種是如何形成的、在不同的高級分類中有多少物種，以及某些物種為何特別多或容易滅絕。最讓人擔憂的是，由於人類的活動，許多物種正面臨滅絕的威脅，這不僅影響我們的環境美觀，還可能削弱經濟發展和生態系統所提供的重要功能。

PR Ehrlich and EO Wilson (1991) Science

https://www.science.org/doi/10.1126/science.253.5021.758

Biodiversity (= biological diversity) definition

Biodiversity studies comprise the systematic examination of the full array of different kinds of organisms together with the technology by which the diversity can be maintained and used for the benefit of humanity. Current basic research at the species level focuses on the process of species formation, the standing levels of species numbers in various higher taxonomic categories, and the phenomena of hyperdiversity and extinction proneness. The major practical concern is the massive extinction rate now caused by human activity, which threatens losses in the esthetic quality of the world, in economic opportunity, and in vital ecosystem services.

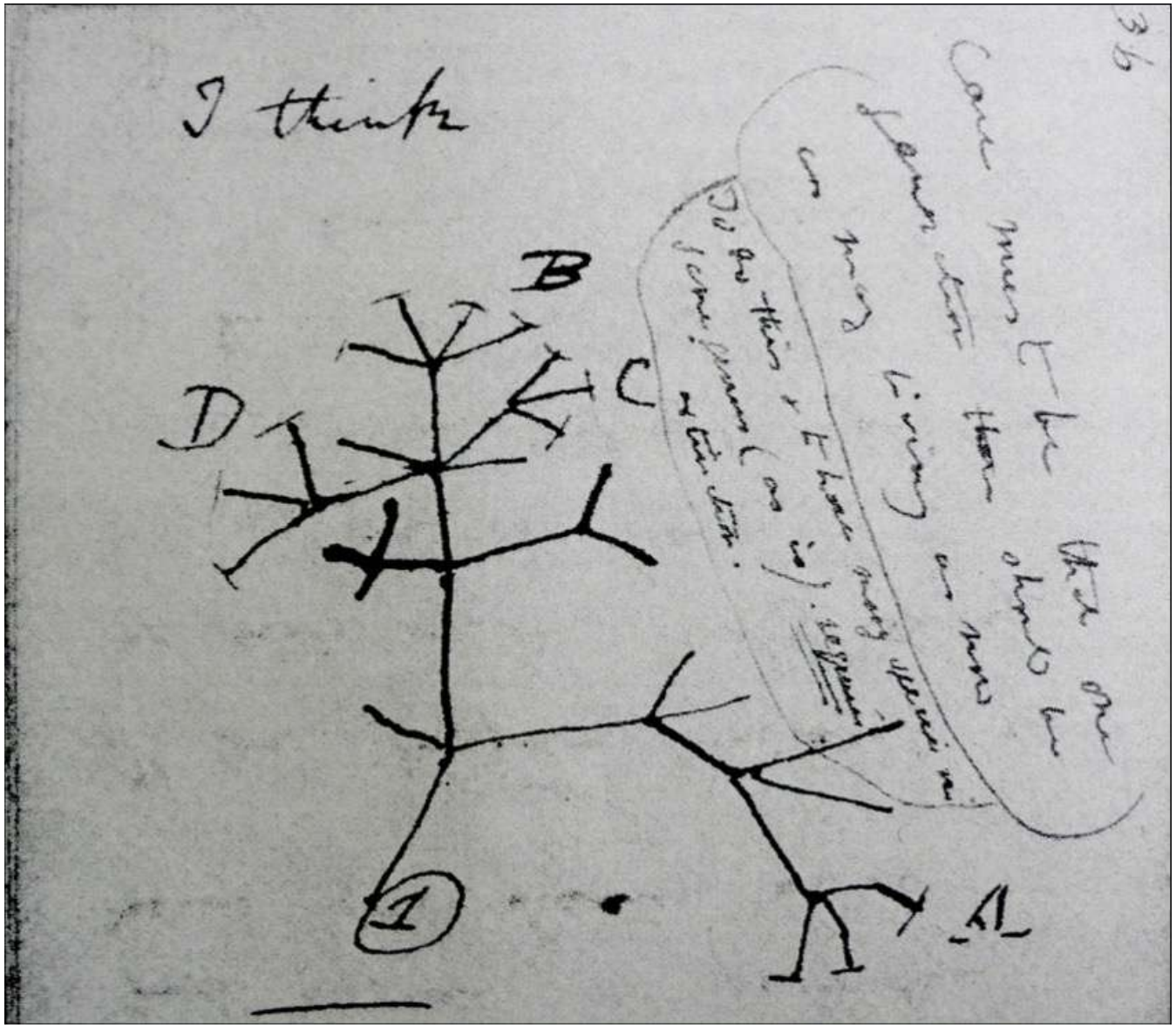
PR Ehrlich and EO Wilson (1991) Science

https://www.science.org/doi/10.1126/science.253.5021.758

目前，生物多樣性研究是系統地探索各種生物，並研究如何保持這種多樣性，讓它為人類帶來好處。**研究集中在物種是如何形成的 (how?)**、在不同的高級分類中有多少物種，以及某些物種為何特別多或容易滅絕。最讓人擔憂的是，由於人類的活動，許多物種正面臨滅絕的威脅，這不僅影響我們的環境美觀，還可能削弱經濟發展和生態系統所提供的重要功能。

Nothing in biology makes sense except in the light of evolution.
- Theodosius Dobzhansky

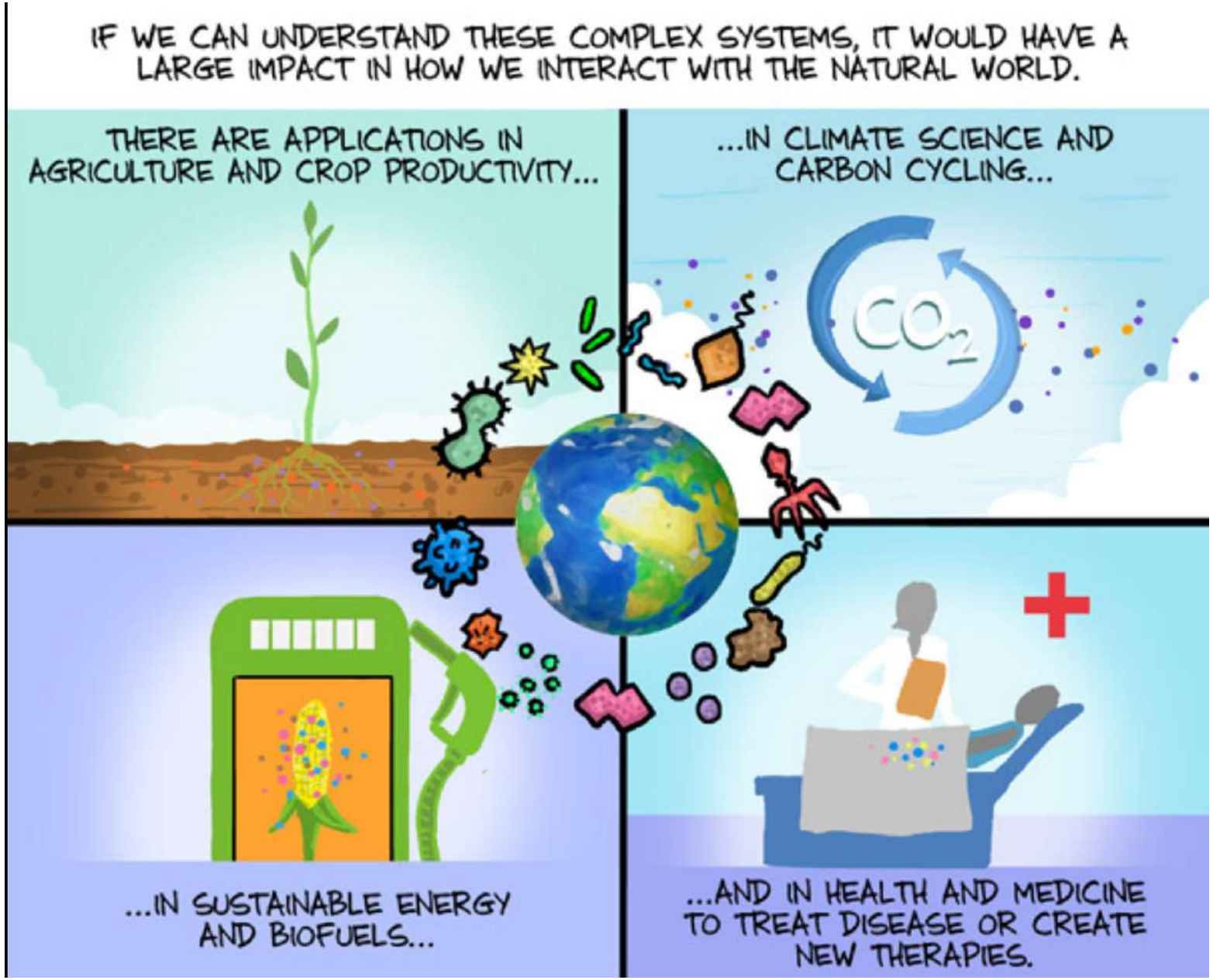
若不透過進化論，生物學的一切都說不通。



Charles Darwin’s notebook

Why should we care (loss of) biodiversity?

- Ethics 道德
- Esthetic 美學
- **Harnessing 利用***
- **Ecosystem functioning 生態系統功能***



Phd comics

PR Ehrlich and EO Wilson (1991) Science

Biodiversity definition

Current Biology
Magazine

Correspondence

How much biodiversity is concealed in the word ‘biodiversity’?

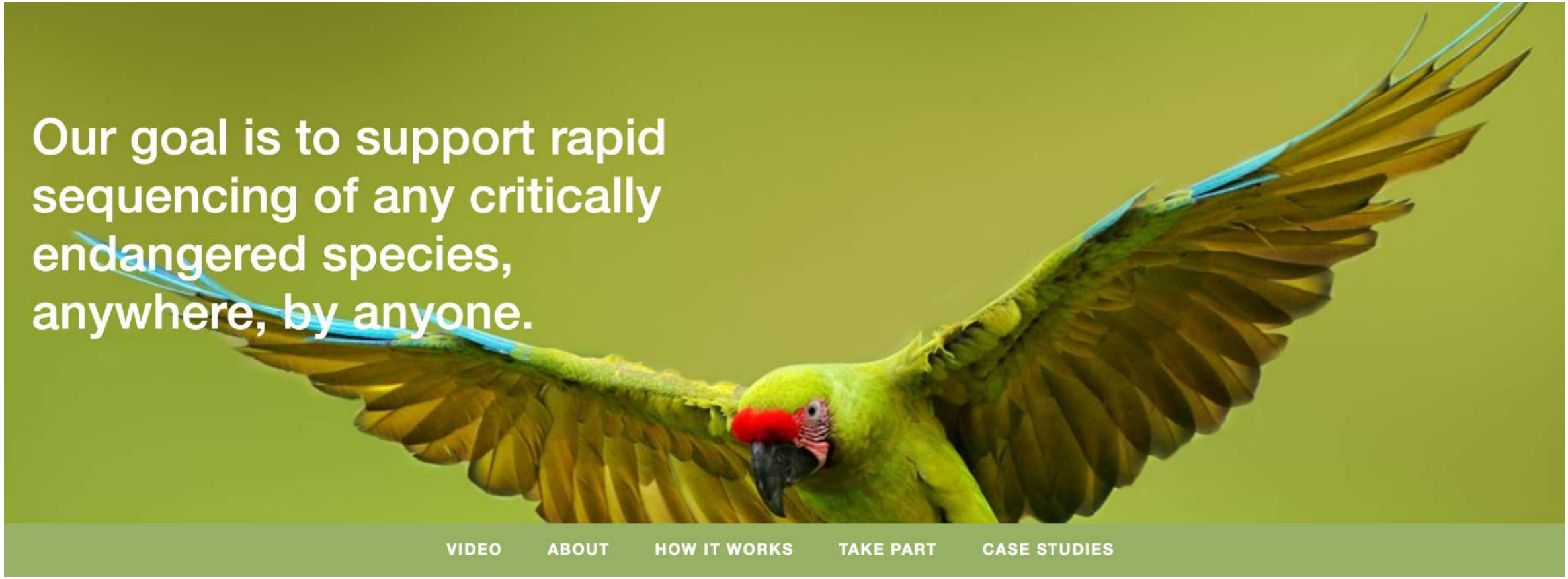
Amidst a global biodiversity crisis[1], the word ‘biodiversity’ has become indispensable for conservation and management[2]. Yet, biodiversity is often used as a buzzword in scientific literature. Resonant titles of papers claiming to have studied ‘global biodiversity’ may be used to promote research focused on a few taxonomic groups, habitats, or facets of biodiversity — taxonomic, (phylo)genetic, or functional. This usage may lead to extrapolating results outside the target systems of these studies with direct consequences for our understanding of life on Earth and its practical conservation. Here, we

We found that as many as 22% of the papers using the word ‘biodiversity’ in the title did not measure biodiversity at any level. This suggests that biodiversity is often used as a theoretical concept rather than a measurable phenomenon[2].

Across the remaining 661 papers, the proportion of biodiversity investigated by each study showed a highly skewed distribution, with most studies sampling a small proportion of biodiversity and a long tail of comparatively few studies sampling higher proportions (mean ± SE: 3.86% ± 0.15%; mode: 1.78%; range: 1.78–44.64%; Figure S2). The taxonomic scope of papers has not increased in recent years either (Figure 1A).

Mammola *et al* (2023) Current Biology

ORG.one (from Oxford Nanopore)



https://www.youtube.com/watch?v=K83GJw69fTA

(video ; around 5 minutes)

Case Example: **Biodiversity Genomics Europe**

(launched 28th September **2022**)

https://biodiversitygenomics.eu/

Biodiversity in crisis

An estimated 25% of species are threatened with extinction worldwide due to large-scale environmental change*. Addressing this global biodiversity crisis requires an understanding of the diversity of life on Earth, how that diversity functions and interacts, and how biodiversity responds to different environmental pressures.

However, after centuries of research, an estimated 80% of the world's multicellular species still await scientific discovery and description. Even for described species, telling them apart is often difficult, and knowledge of their distributions, variation, properties, inter-dependencies, and conservation status remain patchy and incomplete.

* (IPBES, 2019)

https://biodiversitygenomics.eu/

Biodiversity in crisis

An estimated 25% of species are threatened with extinction worldwide due to large-scale environmental change*. Addressing this global biodiversity crisis requires an understanding of the diversity of life on Earth, how that diversity functions and interacts, and how biodiversity responds to different environmental pressures.

However, after centuries of research, an estimated 80% of the world's multicellular species still await scientific discovery and description. Even for described species, telling them apart is often difficult, and knowledge of their distributions, variation, properties, inter-dependencies, and conservation status remain patchy and incomplete.

* (IPBES, 2019)

Key messages:

1. To scale up the quantification of biological diversity
   (communities, species, individuals)
2. To scale up the collation of metadata
   (ecology, distribution, metadata, monitoring)
3. To scale up the study of biology
4. Applications and conservations

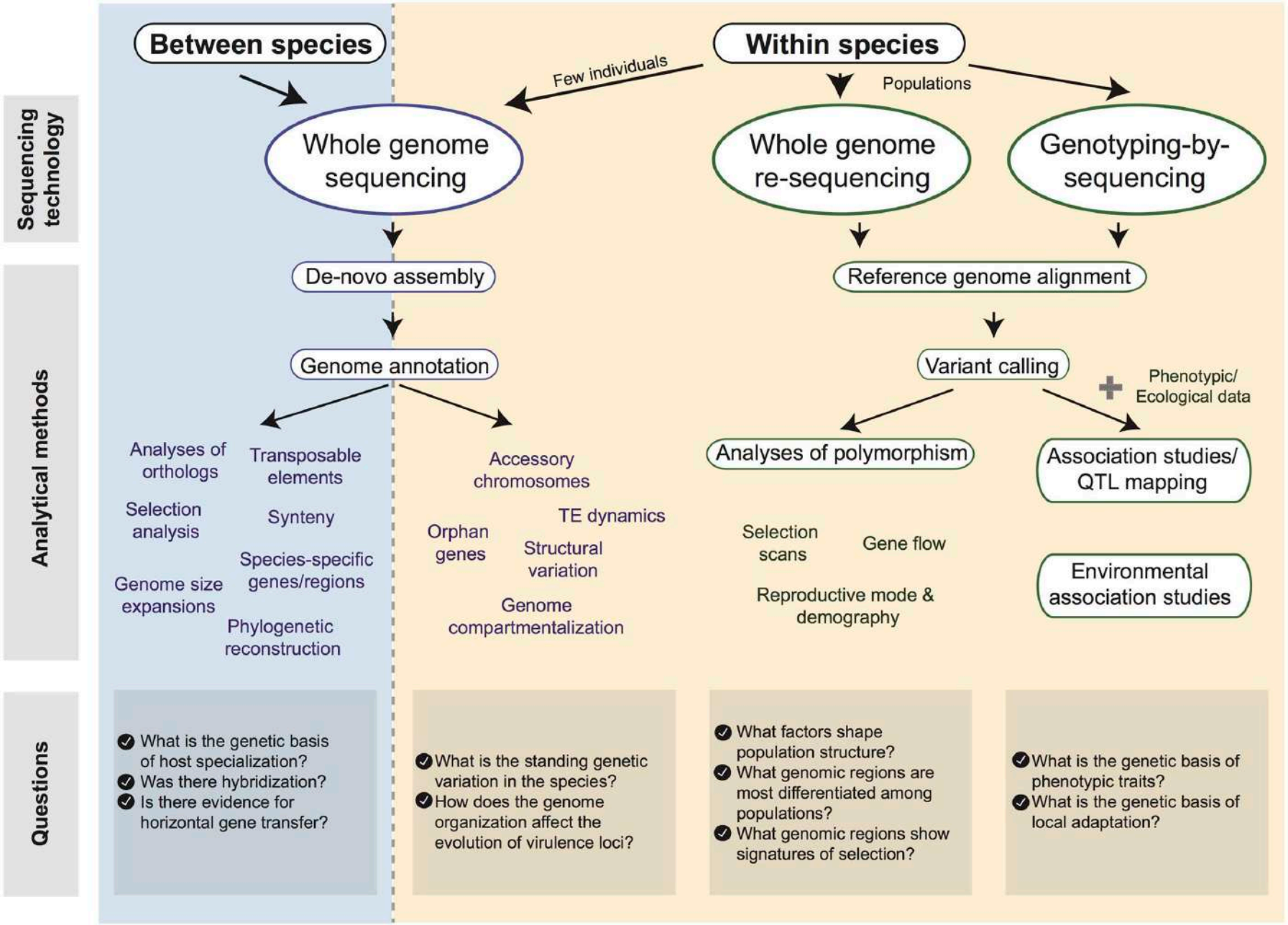
https://biodiversitygenomics.eu/

Topic one: Study biological diversity using genomes

Sequencing all genomes

Darwin TREE of LIFE

EARTH BIOGENOME PROJECT



https://www.frontiersin.org/articles/10.3389/fpls.2017.00119/full

Topic two: Study biological diversity using barcoding

DNA barcoding? Large barcoding?
Metabarcoding? Amplicons? What's the difference?

Note to self: two video talks at 20 minutes in total; + 5 mins in discussion?

Molecular identification of species – DNA barcode

THE ROYAL SOCIETY

Received 29 July 2002
Accepted 30 September 2002
Published online 8 January 2003

Biological identifications through DNA barcodes

Paul D. N. Hebert*, Alina Cywinska, Shelley L. Ball and Jeremy R. deWaard

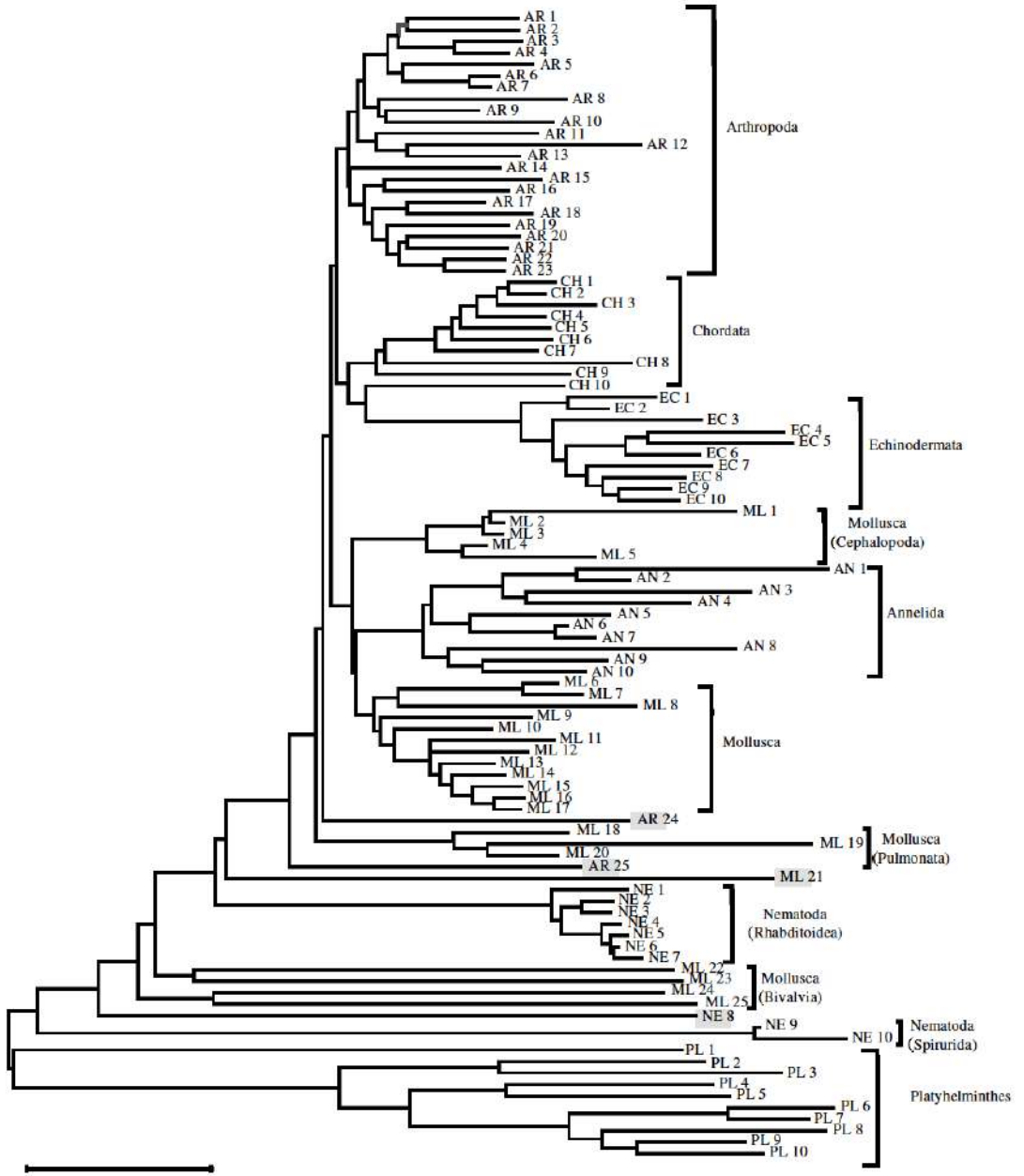
Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Although much biological research depends upon species diagnoses, taxonomic expertise is collapsing. We are convinced that the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon 'barcodes'. We establish that the mitochondrial gene cytochrome *c* oxidase I (COI) can serve as the core of a global bioidentification system for animals. First, we demonstrate that COI profiles, derived from the low-density sampling of higher taxonomic categories, ordinarily assign newly analysed taxa to the appropriate phylum or order. Second, we demonstrate that species-level assignments can be obtained by creating comprehensive COI profiles. A model COI profile, based upon the analysis of a single individual from each of 200 closely allied species of lepidopterans, was 100% successful in correctly identifying subsequent specimens. When fully developed, a COI identification system will provide a reliable, cost-effective and accessible solution to the current problem of species identification. Its assembly will also generate important new insights into the diversification of life and the rules of molecular evolution.

Keywords: molecular taxonomy; mitochondrial DNA; animals; insects; sequence diversity; evolution

COI gene

Cited 13,579 times (google scholar)



Molecular identification of species – DNA barcode

DNA barcoding a useful tool for taxonomists

Sir — The Consortium for the Barcode of Life (CBOL; see www.barcoding.si.edu) is an international initiative of natural history museums, herbaria, other biodiversity research organizations, governmental organizations and private companies which wish to promote the development and use of DNA barcoding.

CBOL is in complete agreement with the major point raised by M. C. Ebach and C. Holdrege in Correspondence, that "DNA barcoding is no substitute for taxonomy" (*Nature* **434**, 697; 2005).

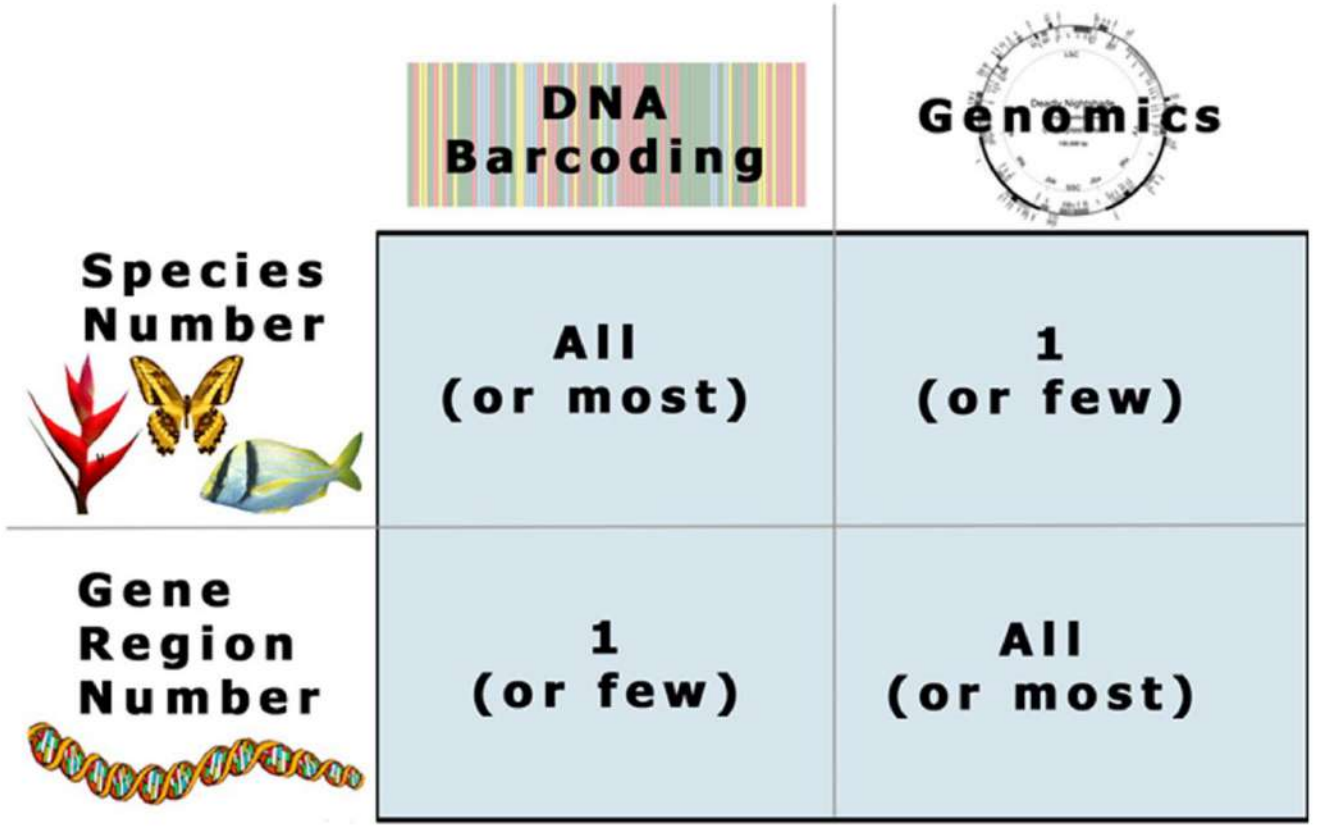
CBOL views barcoding as a useful tool for taxonomists and a cost-effective system with which non-specialists, such as border inspectors, can assign unidentified specimens to known species. In both cases, CBOL views barcoding as part of taxonomy and rejects the idea that DNA taxonomy will replace the practice of taxonomy based on diverse character sets.

Taxonomists have begun using DNA barcodes in three ways. First, barcoding can be used as a 'triage' tool for sorting new collections into units based on barcode sequences, of which some will belong to known species and others will be new to science. In CBOL's view, only expert taxonomists can resolve the relationship between new barcode-based clusters and species.

Second, DNA barcodes can also help assign specimens to known species in those cases where morphologic features are missing (in the case of immature, partial or damaged specimens) or misleading (as in sexually dimorphic species). Third, barcodes can also be used as a supplement to other taxonomic datasets in the process of delimiting species boundaries.

Ebach and Holdrege are correct in stating "DNA barcoding generates information, not knowledge". CBOL believes that this information can make systematists and the consumers of taxonomic information more knowledgeable. Therein lies its potential value.

David E. Schindel, Scott E. Miller
Consortium for the Barcode of Life, National Museum of Natural History, Smithsonian Institution, PO Box 37012, MRC-105, Washington, DC 20013-7012, USA



Schindel and Miller (2005) Nature
Kress and Erickson (2008) PNAS

International Barcode of Life

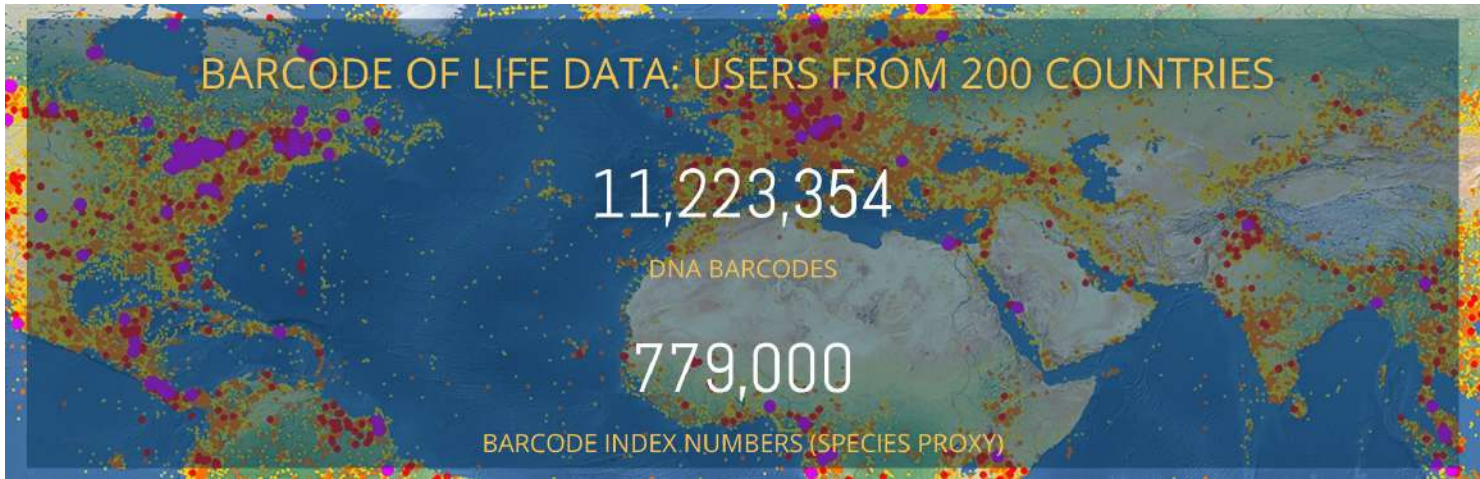
BIOSCAN: tracking biodiversity on Earth (2:00)

https://www.youtube.com/watch?v=K1AchBQHnw4

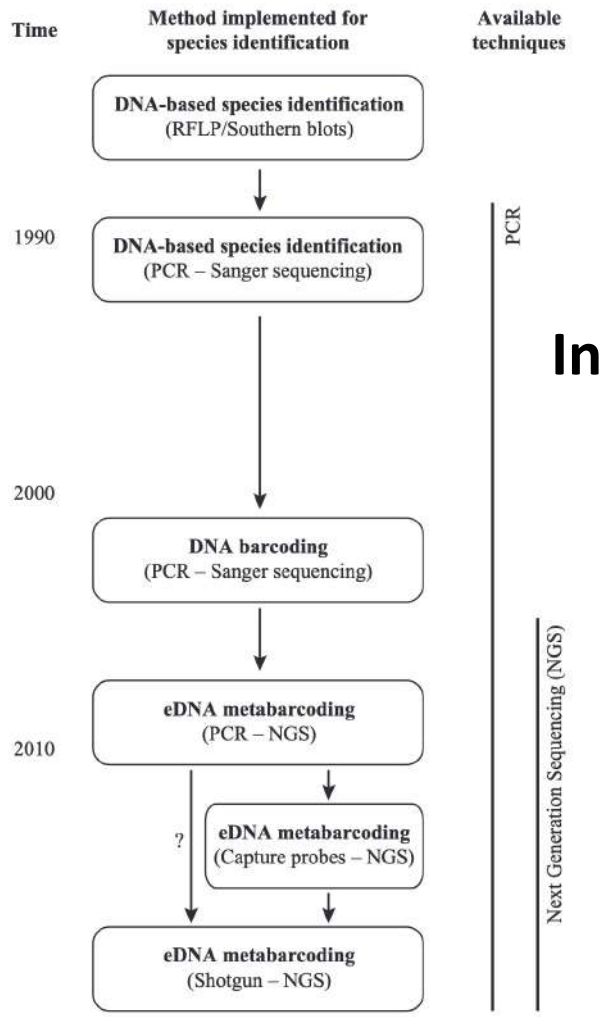
The Centre for Biodiversity Genomics: a look inside the world's leading DNA barcoding facility (6:18)

https://www.youtube.com/watch?v=SHwd0bP4zRk

Also useful reading (**discussion**): Four years of DNA barcoding: Current advances and prospects Frezel and Leblois (2008) Infection, Genetics and Evolution



Molecular identification of species in a community – metabarcoding



Individuals

DNA barcode: a small piece of the genome (marker) found in a broad range of species. The standardized barcode for most animals is a fragment of the mitochondrial *COI* gene, the standardized barcode for plants is a fragment of the plastid gene ribulose 1,5-bisphosphate carboxylase gene (*rbcL*) combined with a fragment of the maturase (*matK*) gene, whereas the barcode for fungi is the nuclear internal transcribed spacer (ITS) of the ribosomal DNA. CBOL (http://www.barcodeoflife.org/) has standardized this method of species identification, and has developed the corresponding sequence reference database for these markers [10].

DNA barcoding: the identification of species using standardized DNA fragments. The ideal DNA barcoding procedure starts with well-curated voucher specimens deposited in natural history collections and ends with a unique sequence deposited in a public reference library of species identifiers that could be used to assign unknown sequences to known species [7,43].

Metabarcoding: a rapid method of high-throughput, DNA-based identification of multiple species from a complex and possibly degraded sample of eDNA or from mass collection of specimens. The metabarcoding approach is often applied to microbial communities, but can be also applied to meiofauna or even megafauna.

Community

Tabernet et al (2012) Molecular Ecology

Cristescu et al (2014) TREE

From DNA barcoding to Metabarcoding

Need high throughput ;
Capped by sequencing technology

Criteria for evaluation	Barcoding	Metabarcoding
Size	Sizes usually longer than 500 bp	Sizes <400 bp are appropriate for degraded DNA
Specificity	At the taxon level	Specific across a divergent group of targeted taxa, but not beyond
		Broad application of single primer pair beyond targeted groups compromises depth of coverage
		Multiple primer pairs can be employed when amplification bias across divergent taxonomic groups is severe. Each targeted group can be amplified by a specific primer-pair.
Versatility	Extensive versatility beyond the taxon of interest is not essential, but can enhance projects charged with comprehensive coverage of large taxonomic groups	High versatility to amplify equally and exhaustively all target groups
Taxonomic resolution	Taxonomic resolution at the species level is desirable	Taxonomic resolution, ideally to the species level, is required; requires validation based on mock communities or similar methods
Well-understood mode of evolution	A distinct break between the intra- and interspecific levels of genetic divergence is required	Desirable for enabling good global alignments that allow valid recovery of OTUs
		Knowledge on the intra- and interspecific levels of genetic divergence across the targeted groups is required
Comprehensive taxonomic database	Building a comprehensive database is a major goal of the barcoding approach	Comprehensive taxonomic database based on verified and curated specimens is desirable; many metabarcodes used currently do not have an associated taxonomic database

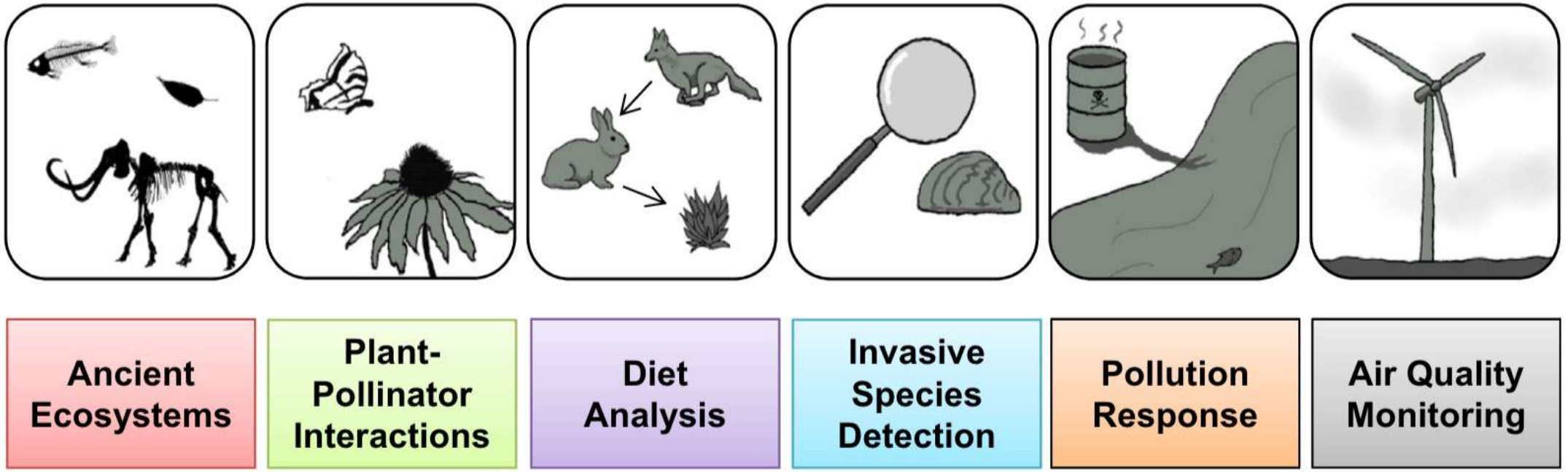
Cristescu et al (2014) TREE

Metabarcoding

How we can detect pretty much anything - Hélène Morlon and Anna Papadopoulou (5:54)
https://www.youtube.com/watch?v=bdwU_ZPk1cY

DNA metabarcoding for biodiversity monitoring (4:12)
https://www.youtube.com/watch?v=YiQKwpl0pq0

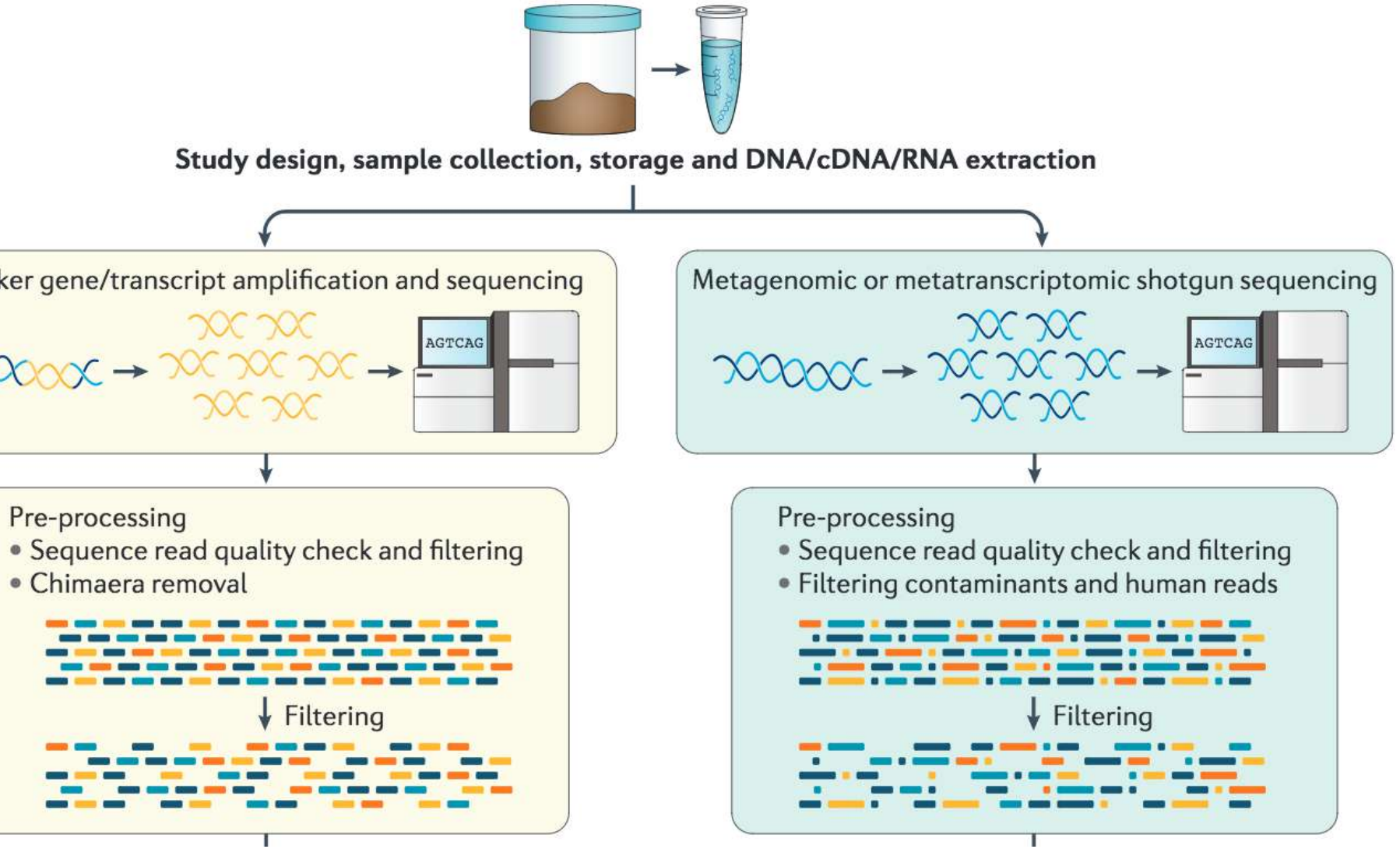
Applications of environmental DNA metabarcoding in aquatic and terrestrial ecosystem



Ruppert et al (2019) Global Ecology and Conservation

Combining the two together? – metagenomics

	DNA Barcoding	Genomics
Species Number	All (or most)	MANY
Gene Region Number	1 (or few)	All (or most)



Claesson, Clooney & O'Toole (2017) Nature Review Genetics

Resolved bacterial genomes from metagenomics samples using long reads

nature communications

Article

https://doi.org/10.1038/s41467-022-34149-0

HiFi metagenomic sequencing enables assembly of accurate and complete genomes from human gut microbiota

Received: 21 June 2022

Chan Yeong Kim[1,3,4], Junyeong Ma[1,4] & Insuk Lee[1,2]

"102 complete Metagenome-assembled genomes (cMAGs) obtained by Pacific Biosciences (PacBio) high-accuracy long-read (HiFi)metagenomic sequencing of five human fecal samples."

nature biotechnology

LETTERS

https://doi.org/10.1038/s41587-020-0422-6

Check for updates

OPEN

Complete, closed bacterial genomes from microbiomes using nanopore sequencing

Eli L. Moss[1,3], Dylan G. Maghini[1,3] and Ami S. Bhatt[1,2]

"We used our methods to analyze metagenomics data from 13 human stool samples. We assembled 20 circular genomes, including genomes of *Prevotella copri* and a candidate *Cibiobacter* sp."

We now have two paths with different emphasis

Rare species that may extinct
New species but with little basic biology known
Indicator species / Iconic species

Model species that can be experimented in lab
Model species but with extensive basic biology known
Keystone species

Complementary Omic approaches *

Conservation genomics

Ecology and evolution of species

Interested in population demography / prediction

Response to environment

Ecological genomics / **Molecular Ecology**

Ecology and evolution of species

Interested in how the genomic bases of traits/phenotypes

Response to environment

* To be taught in lecture 3-6

Topic 3: Ecological genomics / Molecular Ecology

Ecological genomics (EG) / Molecular Ecology definition

"A unique combination of disciplines is emerging — evolutionary and ecological **functional** genomics — which **focuses on the genes that affect ecological success and evolutionary fitness in natural environments and populations**

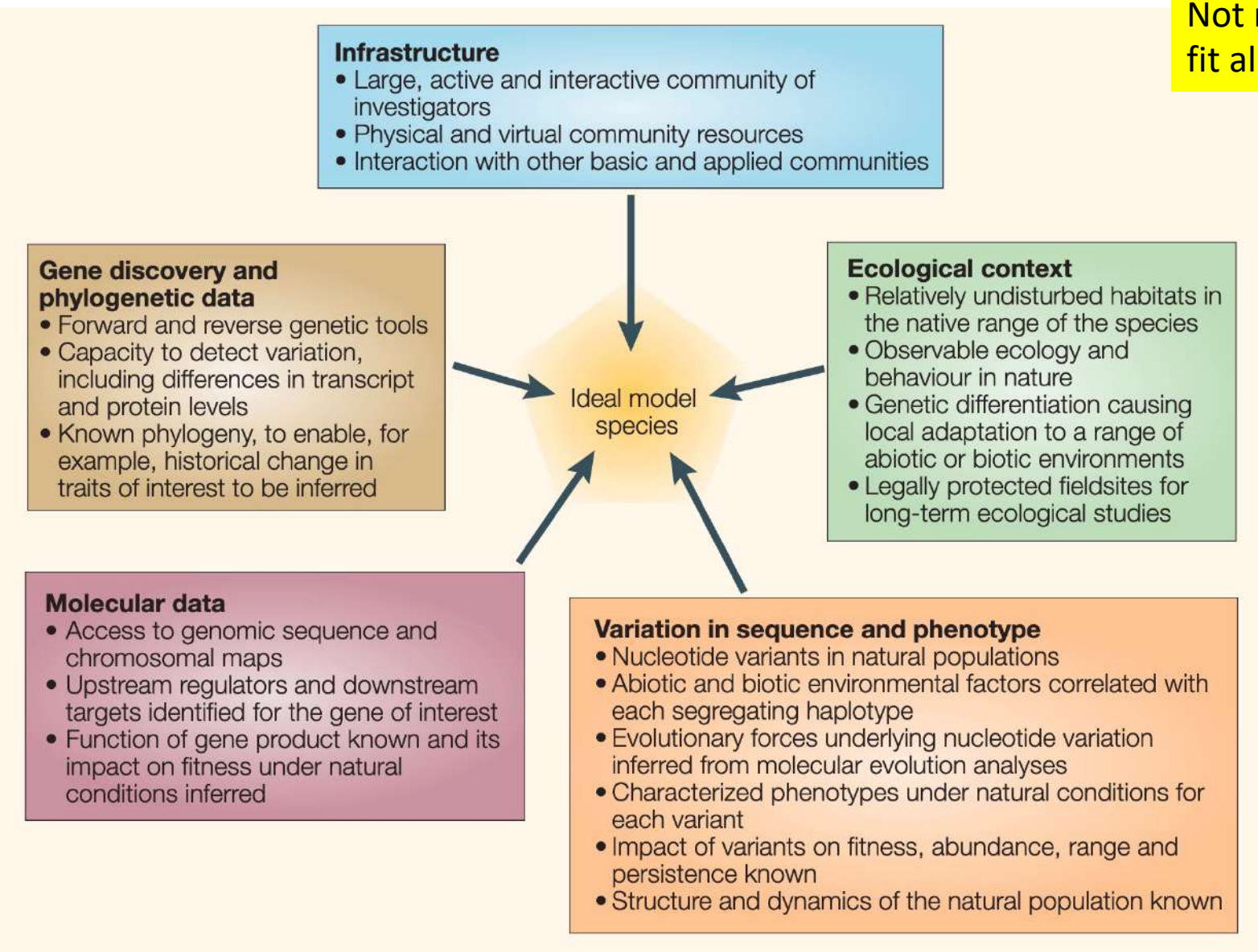
Molecular Ecology is almost synonymous with the field but usually performed on non-model/wild species.

- "the focus is on organisms that inhabit natural environments and the goal of researchers is to explain variation in DARWINIAN FITNESS in populations, and variation in size, range, longevity and diversity among populations, species and higher taxa
- Identify gene or genes of interest
- This is challenging and requires multiple disciplines (ecology, evolution, functional biology and genomics)
- **And carry out experiment to reveal its functions and molecular details**

Feder and Mitchell-Olds (2003) Nature Reviews Genetics

An ideal model organism for EG

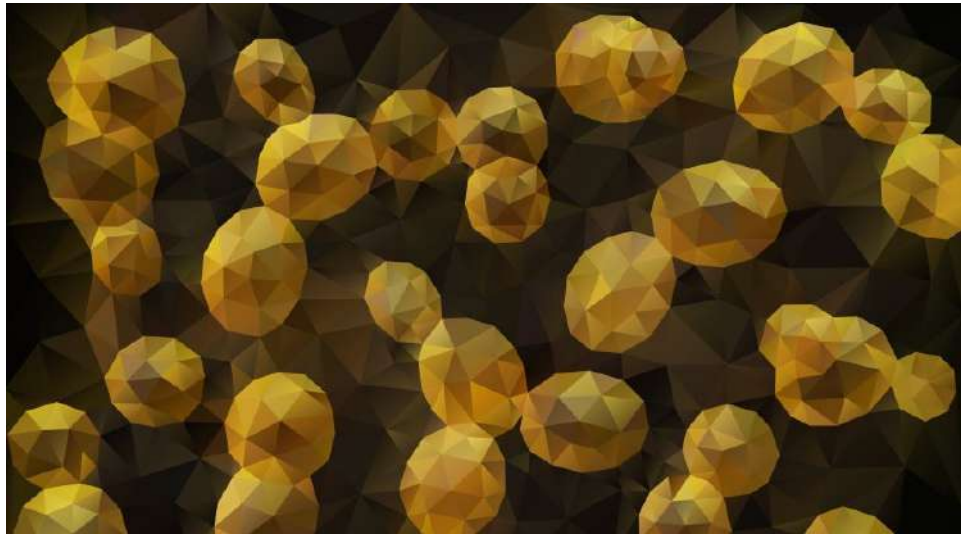
Not many organisms fit all these criteria



Feder and Mitchell-Olds (2003) Nature Reviews Genetics

Model organisms

- Easy to maintain and breed in a laboratory setting.
- Many model organisms can breed in large numbers.
- Some have a very short generation time, which is the time between being born and being able to reproduce, so several generations can be followed at once
- Mutants allow scientists to study certain characteristics or diseases.
- Easy and cheap genetic manipulation
- Some model organisms have orthologs to humans.
- Model organisms can be used to create highly detailed genetic maps.
- Or they may occupy a pivotal position in the evolutionary tree



https://www.yourgenome.org/facts/what-are-model-organisms

Research in model yeast *Saccharomyces cerevisiae*

Yeast: An Experimental Organism for 21st Century Biology

David Botstein*,[1] and Gerald R. Fink†
*Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, and †Whitehead Institute for Biomedical Research and Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

- Biotechnology
- Fermentation
- Synthetic biology
- High-throughput / Systematic
- Light sensing

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3213361/pdf/695.pdf

Ecological genomics

Genotypes

- Genotype frequencies
- Genomic variations
- Population genomics
- Comparative genomics

Phenotypes

- Phenotype frequencies
- Phenotype plasticity
- Development

Traditional model organisms

Ecology

- Abiotic
- Biotic
- Short term / long term

Ecologists

Ecological genomics / Molecular Ecology

Genotypes

- Genotype frequencies
- Genomic variations
- Population genomics
- Comparative genomics

Advances in genomics really kick off this field; rather than choose a model species we can ask virtually any questions across all organisms

Phenotypes

- Phenotype frequencies
- Phenotype plasticity
- Development

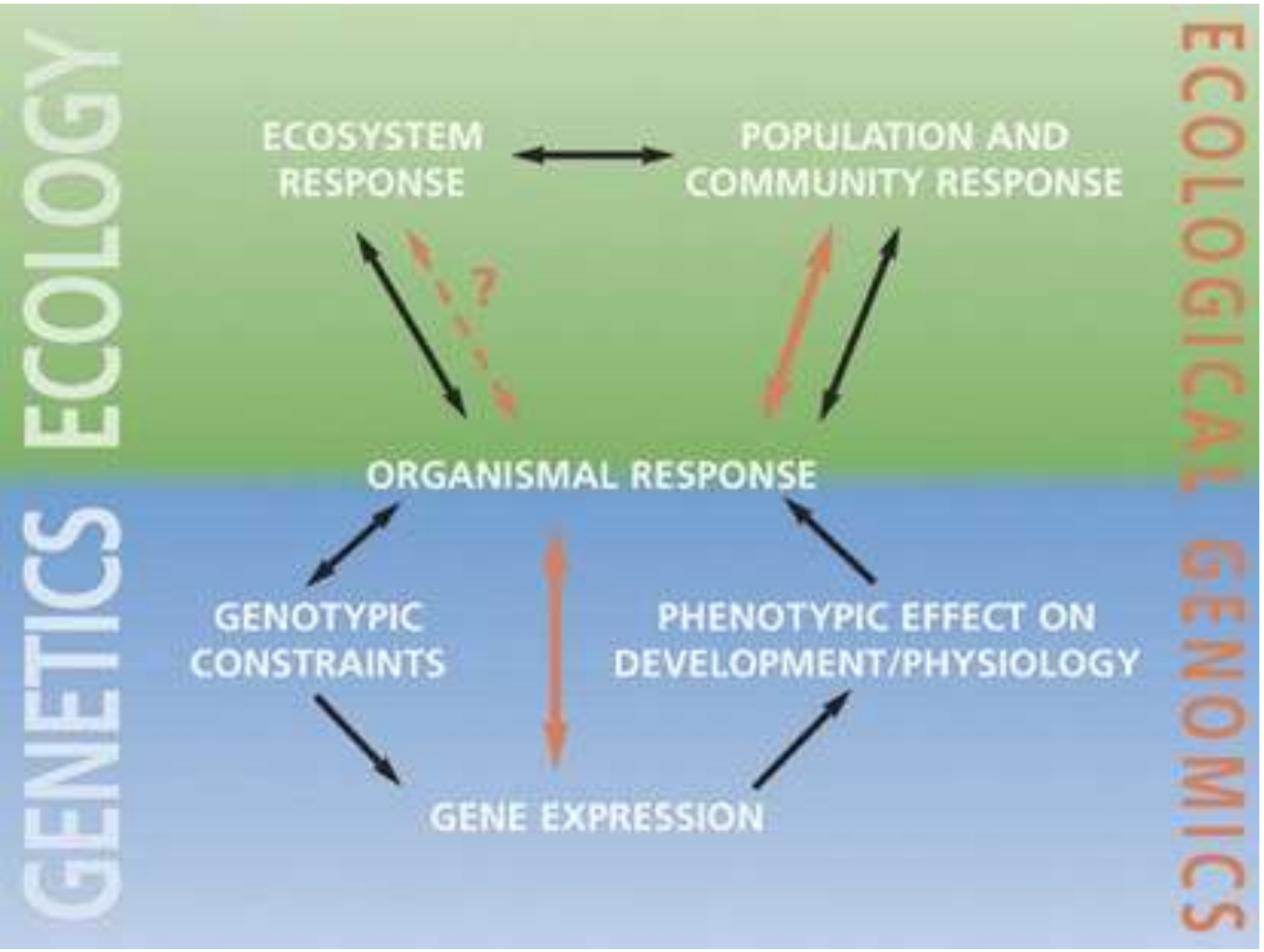
Traditional model organisms

Ecology

- Abiotic
- Biotic
- Short term / long term

Ecologists

Conceptual framework for eco genomics



Ecological interactions between the organism, the population and community levels and the ecosystem

Interactions between the levels, with organismal responses affecting and being affected by its genotype, which in turn affects what genes are expressed and at what levels, which in turn has effects on the phenotype of the organism, ultimately leading to its overall response.

Ecological genomic studies seek to integrate these disciplines (orange arrows) through the use of functional genomics approaches.

Ungerer *et al* (2007) Heredity

Experimental approaches to assess evolutionary responses to climate change using different sets of biological data.

approach		data type/method	genomic resolution	necessity of preliminary knowledge (*)	applicability to different taxa	inferable biological information	strengths	weaknesses
population genetics	reduced representation requencing	marker-based genotyping	limited	genetic markers	unlimited	genetic variation, N_e, demography, candidate genes, structural variation, signatures of selection, candidate loci/genes, historical and experimental N_e, patterns of polygenic variation	cost effective sequencing (esp. large genomes)	fraction of genomic variation, no info about phenotype
		exome capture		probes				
		RNA-Seq		no				
		RAD-Seq		no				
	whole genome sequencing	**sequencing of natural individuals or populations**	high	no	unlimited		genome-wide variation, sustainable data for different research questions, reveals real selection in natural populations	feasability (costs for large genome, material, infrastructure), indirect phenotypic inference
		evolve & resequencing	very high	no	model organisms with short life-span for large-scale EE		resolution from phenotype to genotype	artifical experiment, technical complexity
quantitative genetics	classical quantitative genetics	phenotypes and resemblance between relatives	no	phenotypic traits, variation in fitness, pedigree information	cultivable taxa for common garden experiments or taxa with known pedigrees	additive genetic co-variance (V_A) of traits, strength of selection, evolutionary response	direct phenotypic inference, estimate of evolutionary response	experiments or long-term monitoring necessary
	genomic quantitative genetics	**GRMs or genotyping (genome-wide neutral SNPs)** (**)	high - very high (**)	no (**)	unlimited		more accurate estimates of V_A compared to classical QG, idenpendent of experiments if fitness estimates can be derived from GRMs	large-scale spatial and/or temporal genotyping of populations, pending innovations to quantify fitness from GRMs

Waldvogel *et al* (2020) Evolution Letters

Topic four: Measuring biodiversity

Discussion time: How do we measure biodiversity?

- What does it mean by “High biodiversity” as oppose to “low biodiversity”?
- What is “biodiversity hotspot”?
- How do we **measure** biodiversity? Unit of biodiversity?
- And why do we need them?

Biodiversity: measurement and estimation

Preface

JOHN L. HARPER[1] AND DAVID L. HAWKSWORTH[2]

[1] *Cae Groes, Glan-y-coed Park, Dwygyfylchi, Penmaenmawr, Gwynedd LL34 6TL, U.K.*
[2] *International Mycological Institute, Bakeham Lane, Egham, Surrey TW20 9TY, U.K.*

SUMMARY

In introducing a series of 11 papers on the measurement and estimation of biodiversity, eight crucial questions are posed: What is 'biodiversity'? Is biodiversity just the number of species in an area? If biodiversity is more than the number of species how can it be measured? Are all species of equal weight? Should biodiversity measures include infraspecific genetic variance? Do some species contribute more than others to the biodiversity of an area? Are there useful indicators of areas where biodiversity is high? And can the extent of biodiversity in taxonomic groups be estimated by extrapolation? In addition, the modern concept of biological diversity is attributed to Elliot R. Norse and his colleagues.

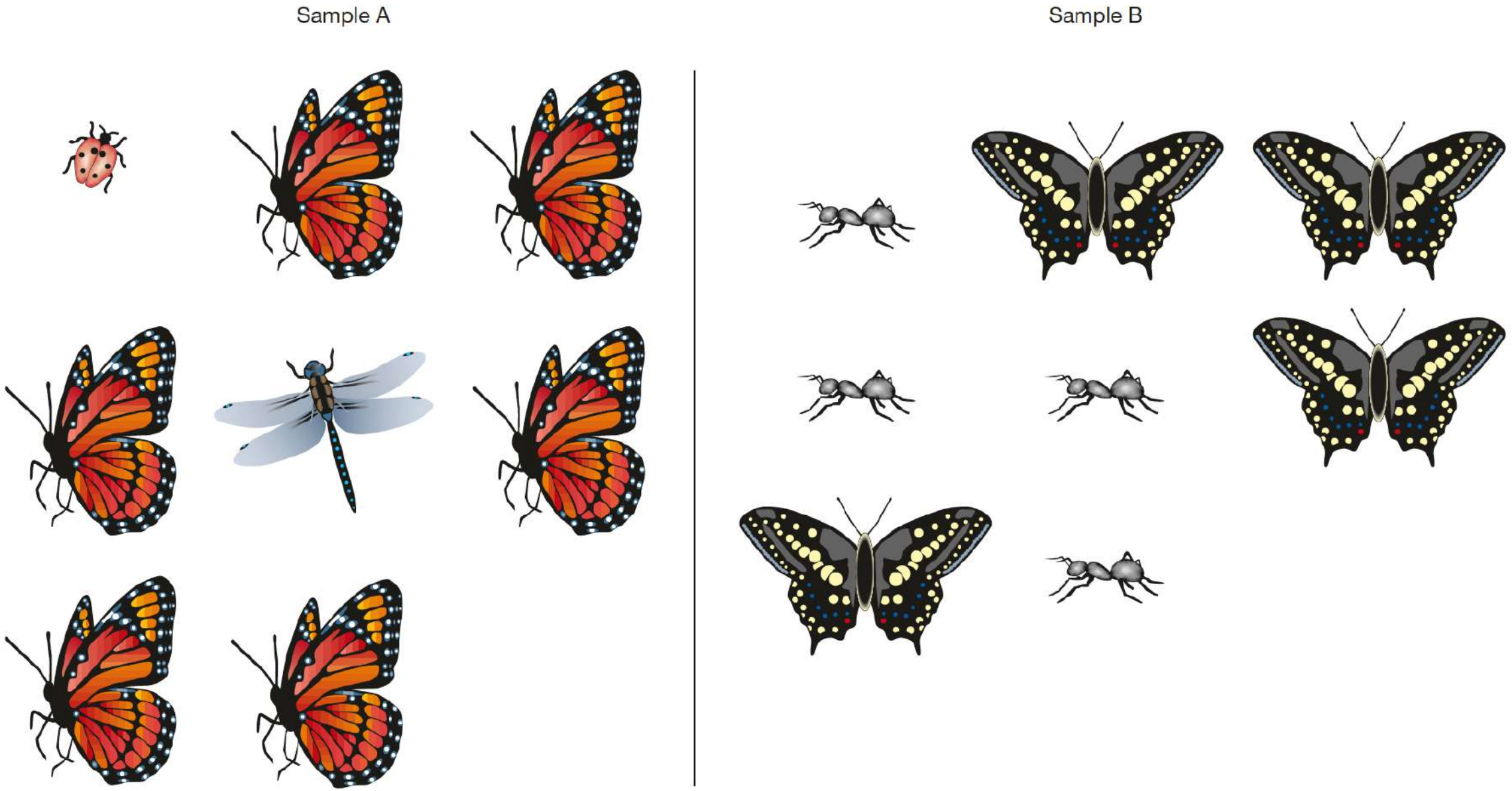
4. IF BIODIVERSITY IS MORE THAN THE NUMBER OF SPECIES HOW CAN IT BE MEASURED?

*

Harper and Hawksworth (1994) Phil Trans R Soc London B

* Also contains nice introduction on history of **biodiversity**

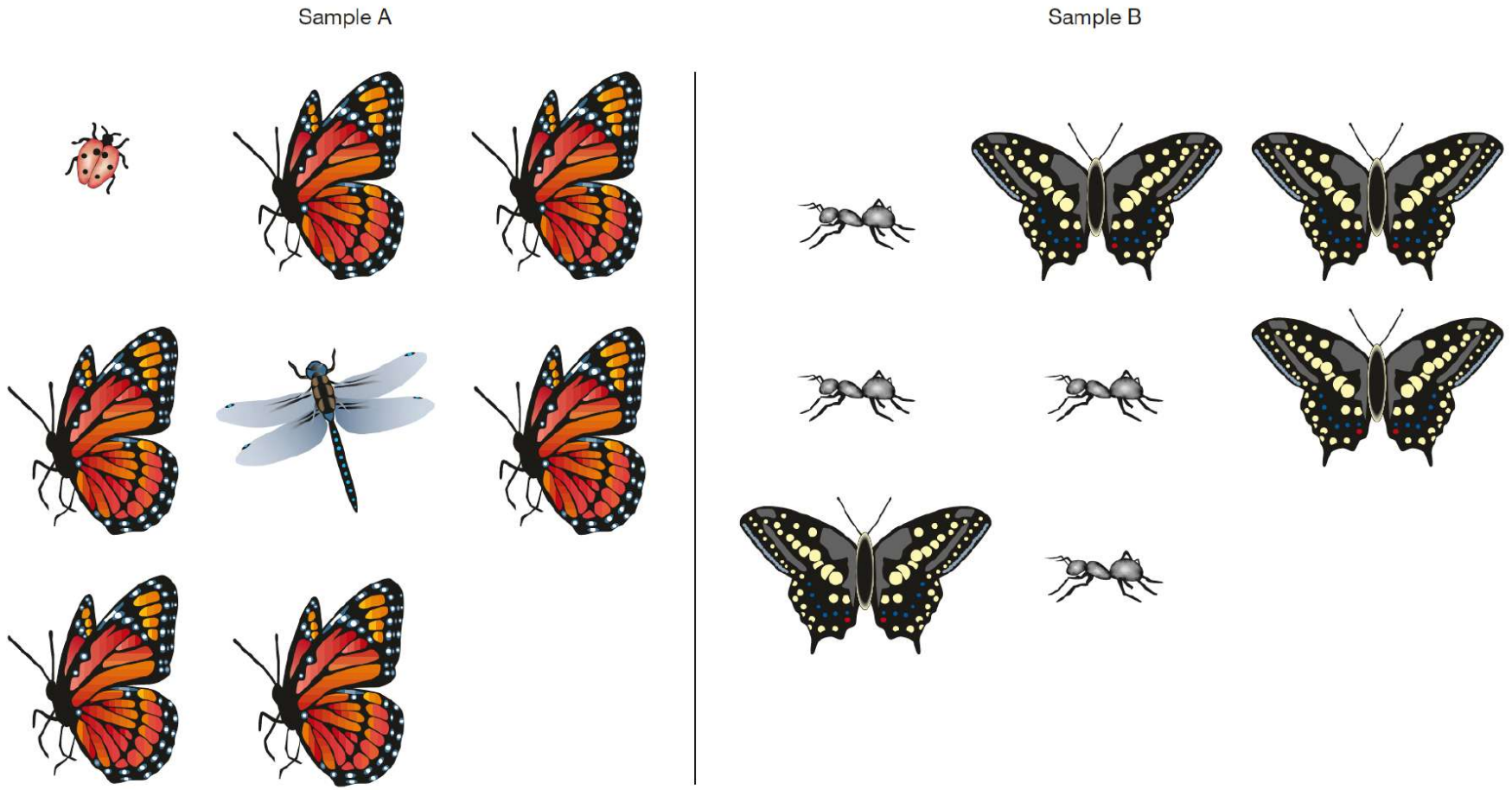
Which of the two populations do you consider to have a higher biodiversity?



Purvis and Hector (2000) Nature

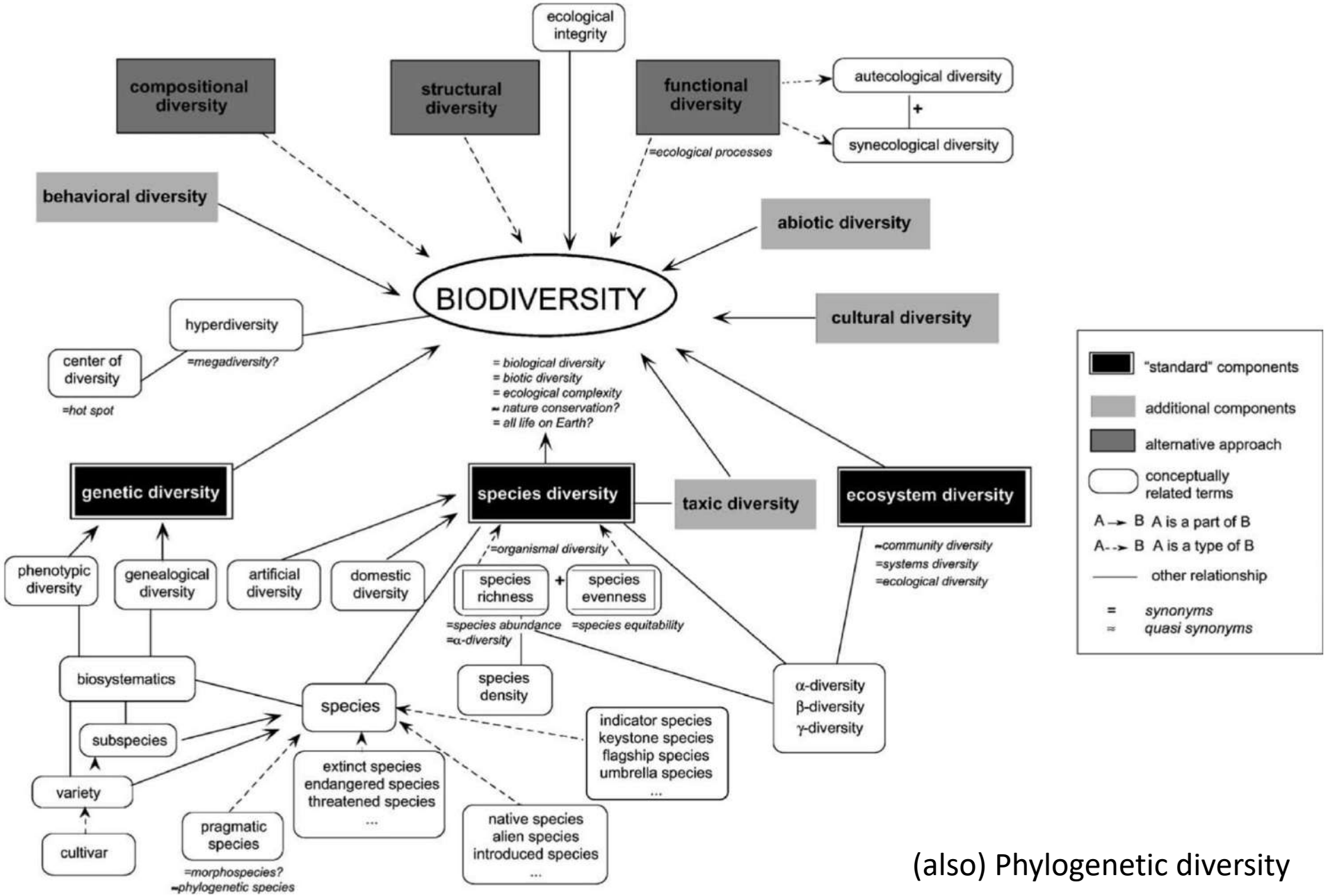
”To proceed very far with the study of biodiversity, we need to pin the concept down. We cannot even begin to look at how biodiversity is distributed, or how fast it is disappearing, unless we can put units on it.

However, any attempt to measure biodiversity quickly runs into the problem that it is a fundamentally multidimensional concept: it cannot be reduced sensibly to a single number”



Sample A could be described as being the more diverse as it contains three species to sample B’s two (**richness A>B**). But there is less chance in sample B than in sample A that two randomly chosen individuals will be of the same species (**evenness B>A**)

Purvis and Hector (2000) Nature



(also) Phylogenetic diversity

Duelli and Obrist (2003) Agriculture & Ecosystems & Environment

Three type of measures*

- Richness (Numbers of…)
  - Species?
  - OTUs?
  - Higher taxonomic level?
- Evenness
  - How are they distributed?
  - Genetic analogues? (heterozygosity)
- Differences
  - Genetic variability?
  - Morphological variability?

Association

ecosystem function

ecosystem resilience

conservation

biological control

Purvis and Hector (2000) Nature

* To be taught in the amplicon lecture

Additional notes

Scaling up!

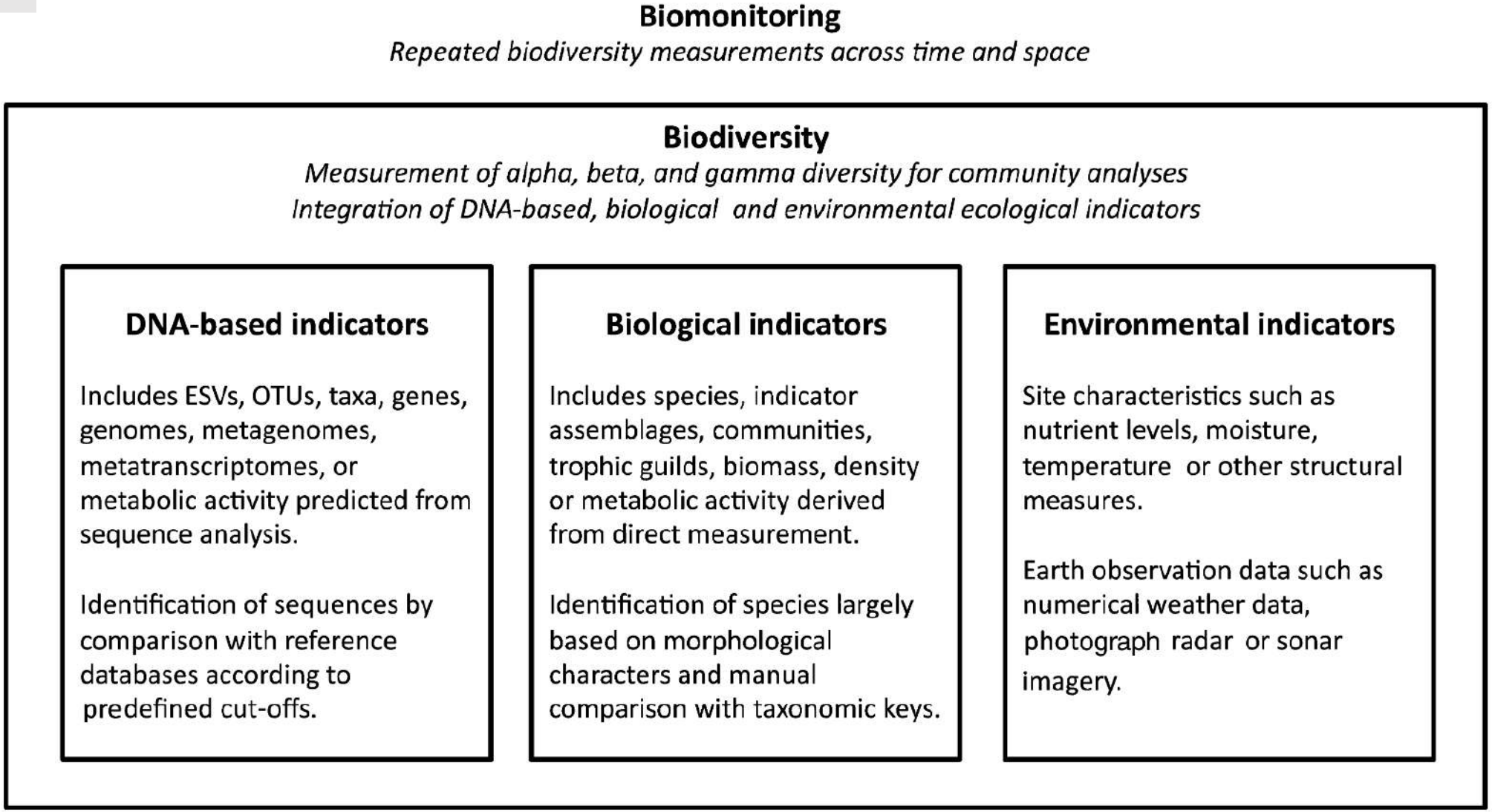
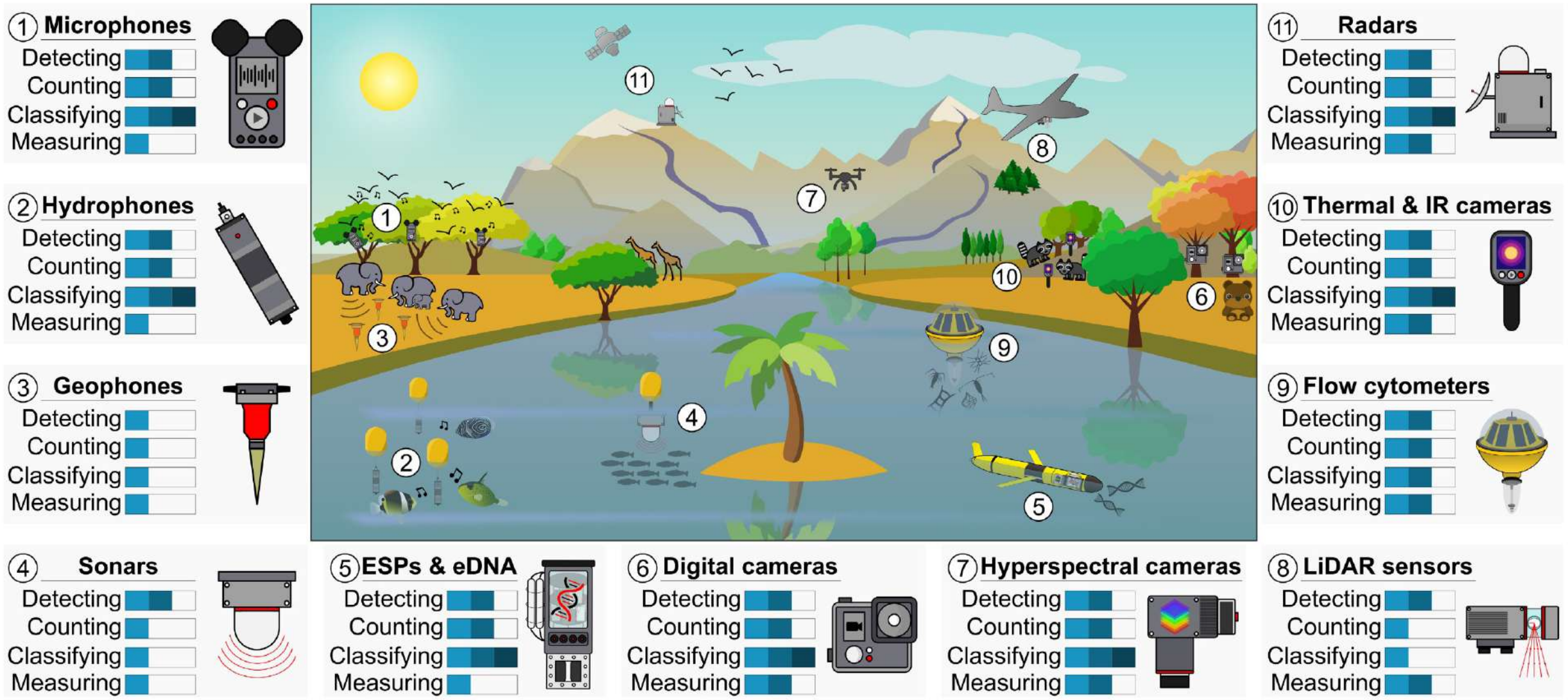


FIGURE 1 Integration of data types in biodiversity genomics. Boxes outline the various ways biodiversity can be sampled using DNA-based or traditional methods that use biological and environmental ecological indicators

Porter and Hajibabaei (2017) Molecular Ecology

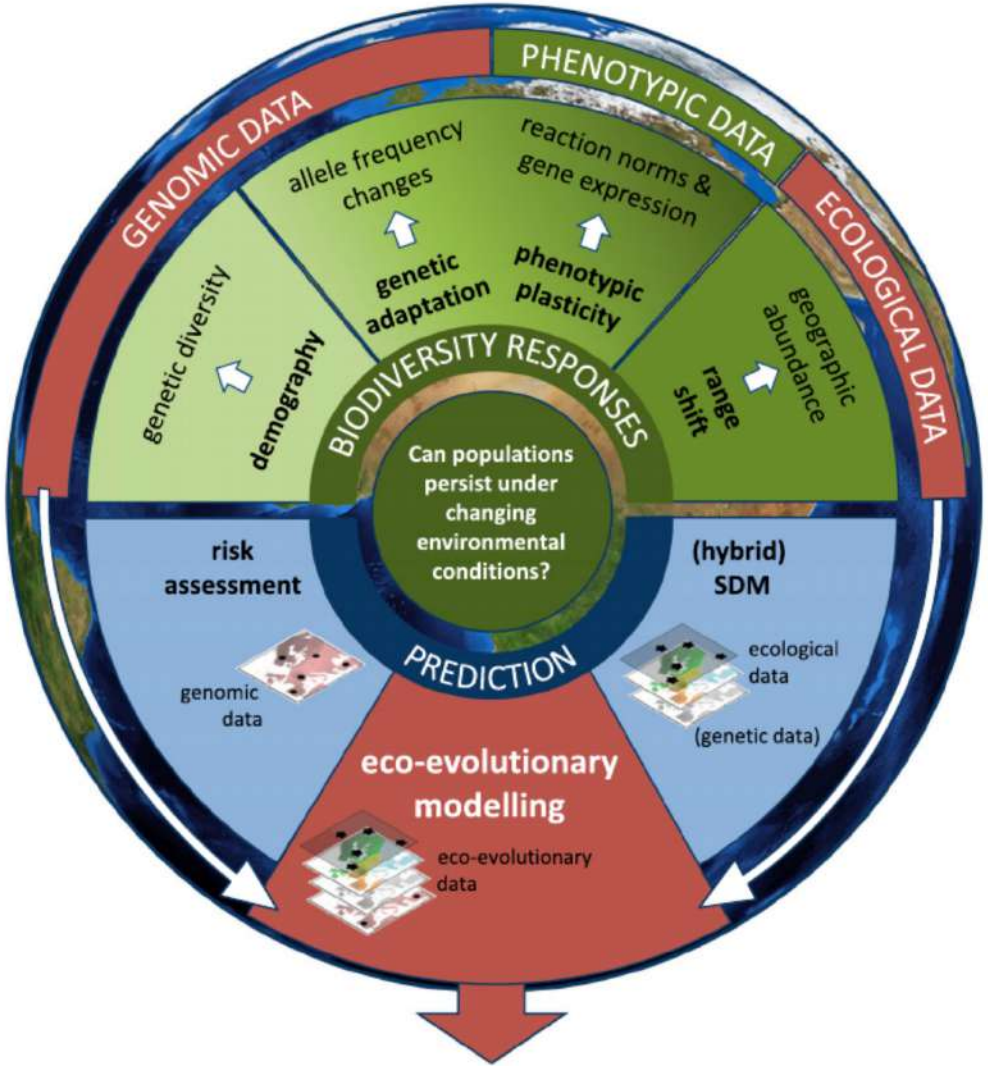
“Towards the fully automated monitoring of ecological communities”



*

Besson et al (2022) Ecology Letters

Predicting species' responses to climate change

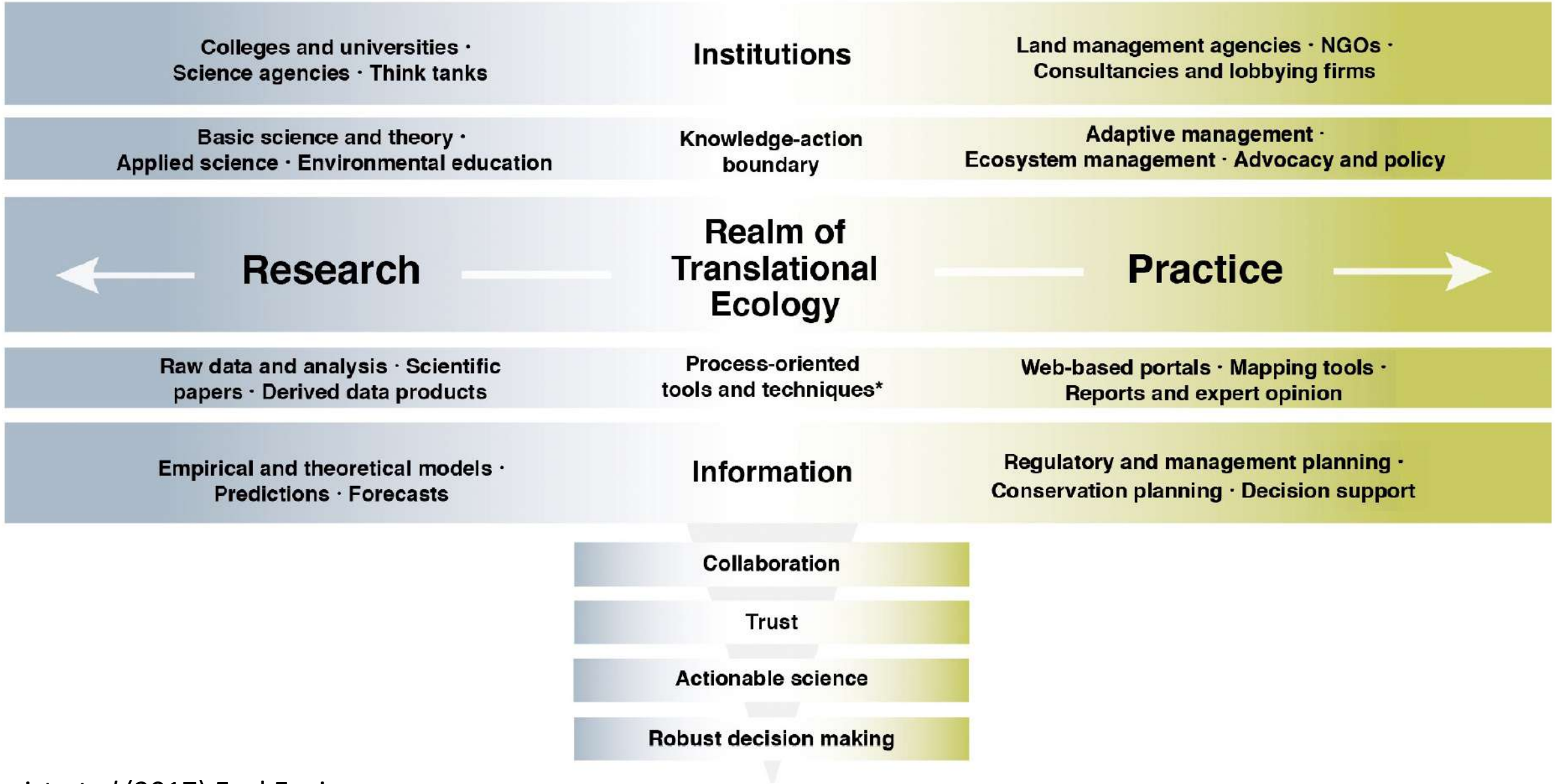


promising strategy to use genomics for predictions:
fast enough and for a broad taxon spectrum

ecosystem management & conservation strategies

Waldvogel *et al* (2020) Evolution Letters

Translational Ecology

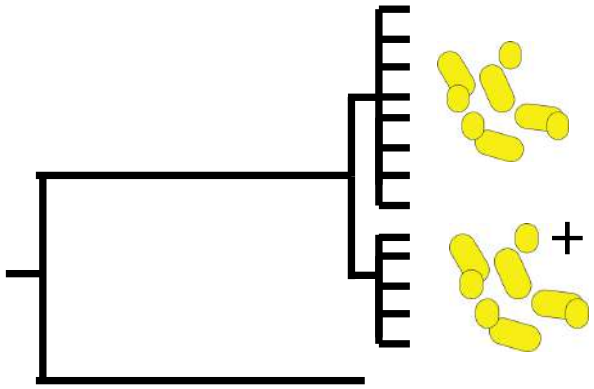


Enquist *et al* (2017) Ecol Environ

Summary: quantifying diversity (基因多樣性) at different timescales

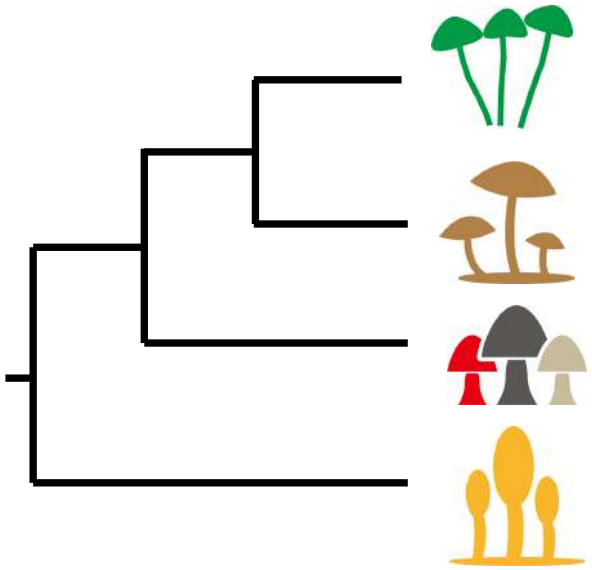
million years ago

Population genomics
(**PhD**)



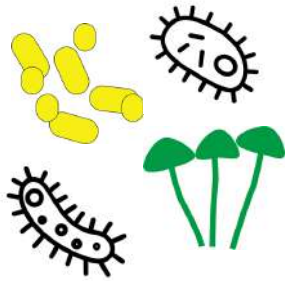
群體遺傳學

Comparative
genomics/transcriptomics
(**Postdoc**)



比較基因體學

Metagenomics/
metatranscriptomics



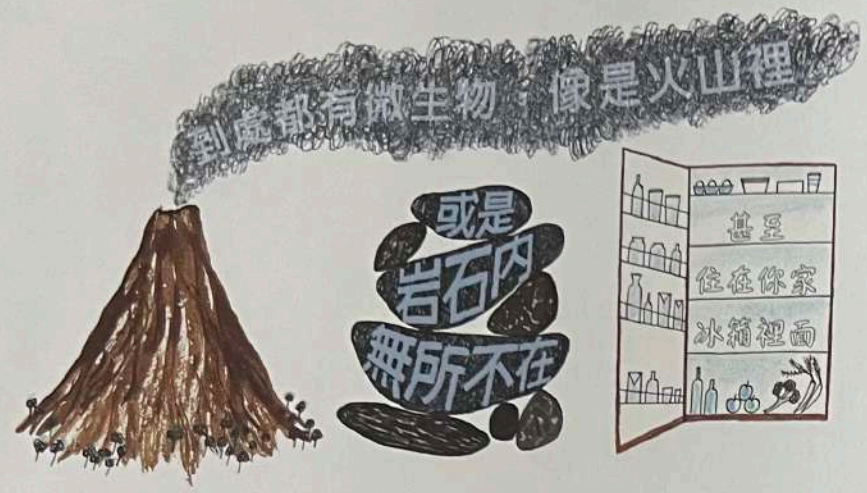
總體基因體學

Main question 1: Characterise and understand microbial **biodiversity**

Main question 2: **Pathogen** evolution

為什麼要研究微生物？

用很棒的一本書跟~~女兒~~大家解釋



到處都有微生物，無論是在海洋裡，在陸地上，在土壤裡，還是在空氣中。它們住在其他生物無法生存的地方，像是火山裡，岩石內，或是你家冰箱裡面，

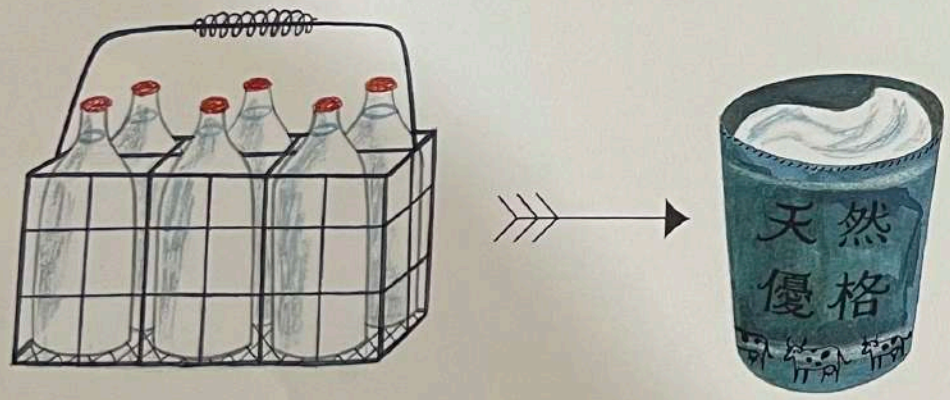
也會出現在動物、植物的表面和體內。

此時此刻，你皮膚上的微生物，
就比全世界人口還要多，而你肚子裡的微生物，
是全世界人口的10倍，甚至100倍。

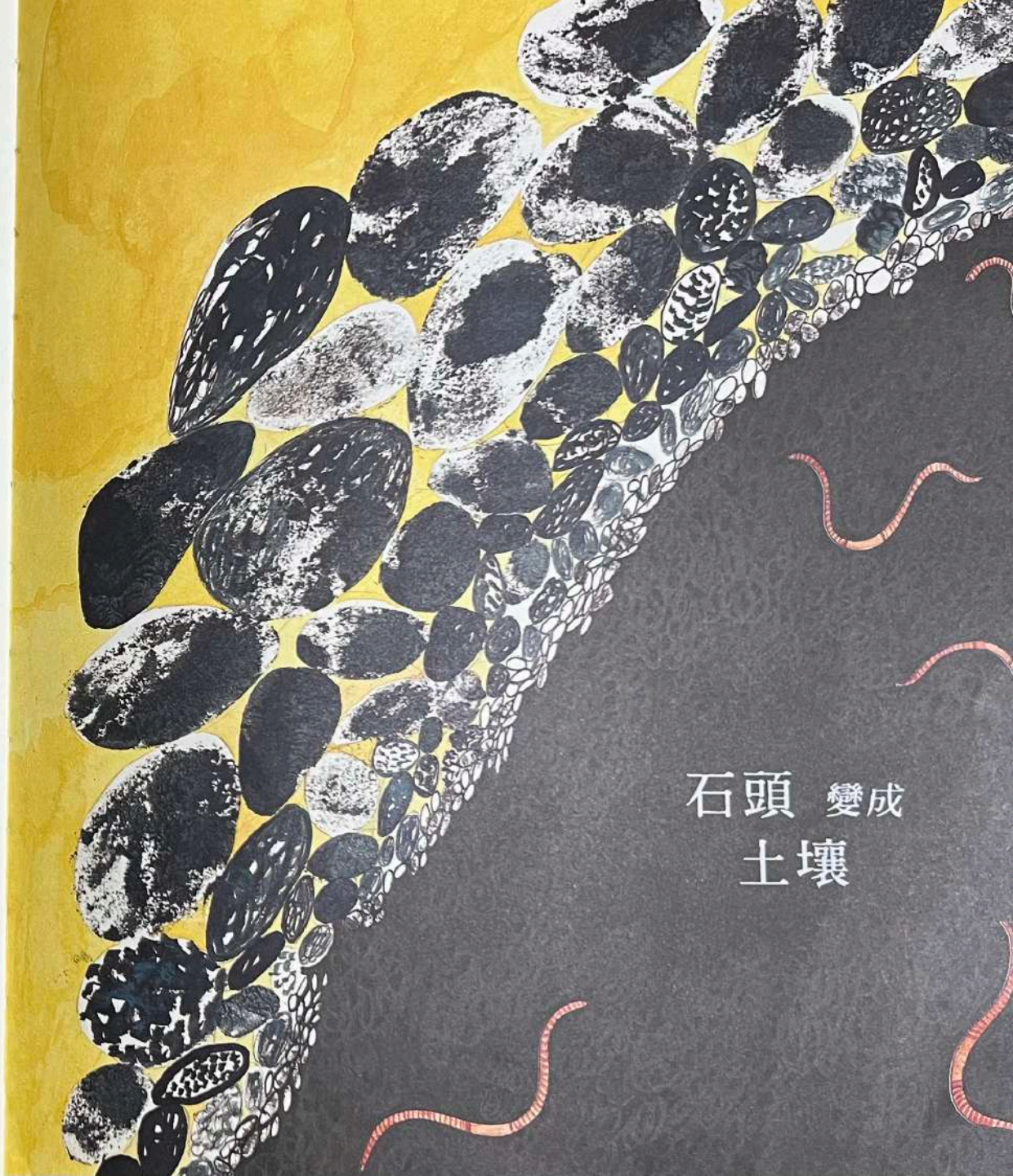
（別擔心，雖然有些微生物會害你生病，但那些長住在你身體表面和體內的微生物，會讓你保持健康。）

這就是為什麼微生物吃過的東西
並不會一口一口消失不見。
這些東西會慢慢的變成另一種東西……

食物　變成　堆肥



牛奶　變成　優格



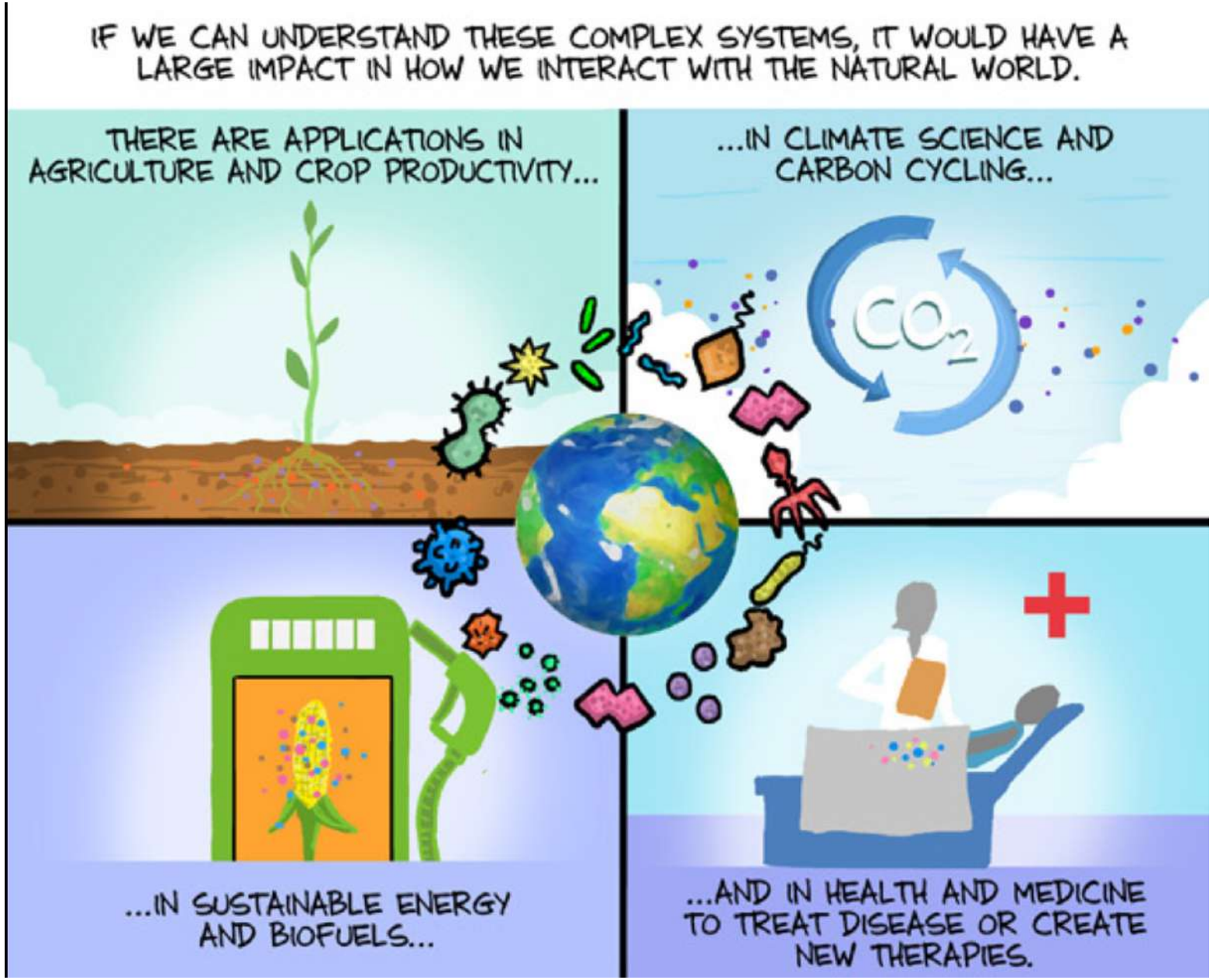
石頭 變成
土壤

而且，因為微生物很擅長
製造出更多微生物，
有時候它們能做出
非常了不起的事情。

它們可以讓高山崩塌，
形成峭壁。
它們可以把海水染紅，
讓天空的雲變多，
還可以變出很多雪花。

Why should we care (loss of) microbial biodiversity?

- Ethics 道德
- Esthetic 美學
- **Harnessing 利用***
- **Ecosystem functioning 生態系統功能***



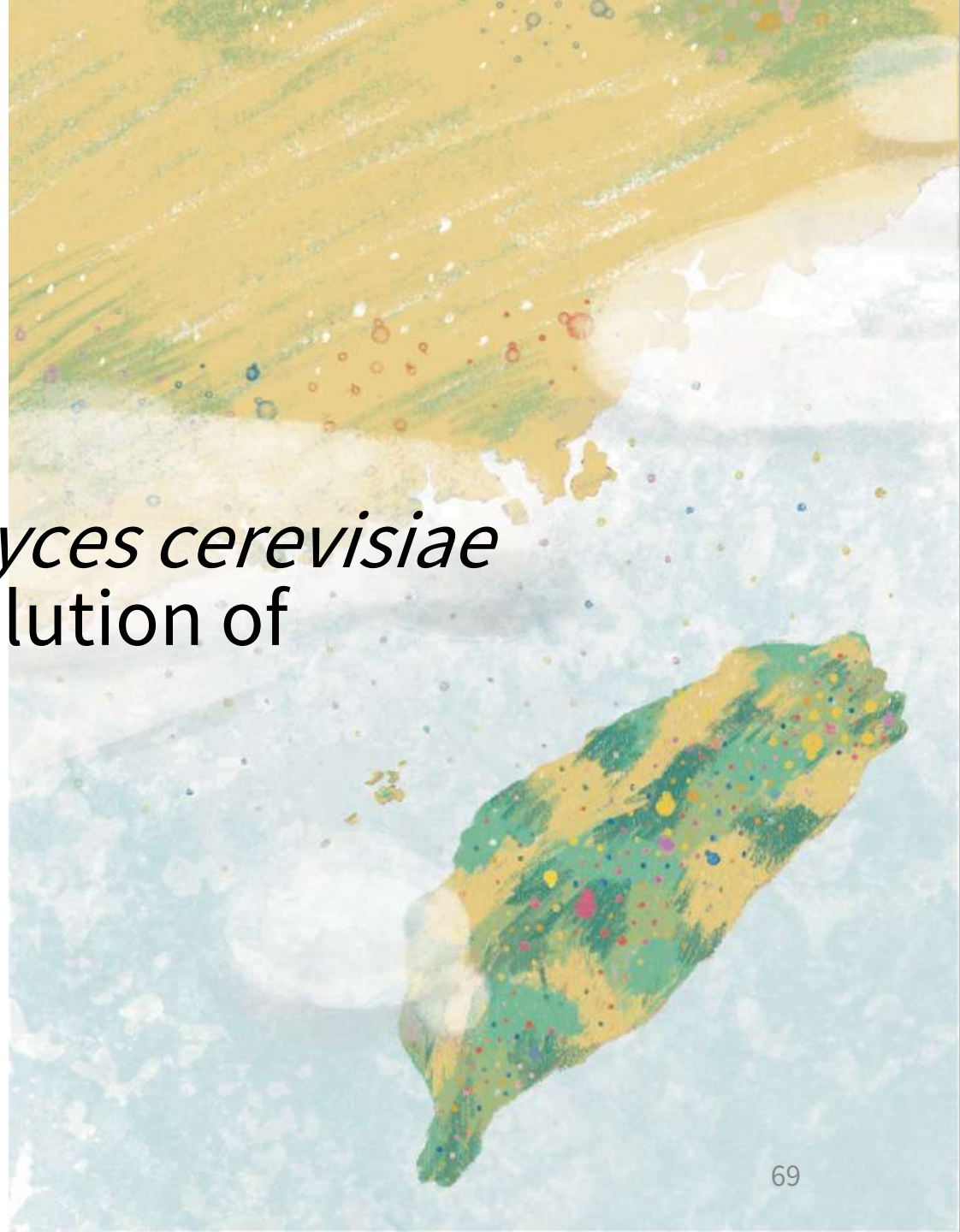
Phd comics

PR Ehrlich and EO Wilson (1991) Science

案例研究分享

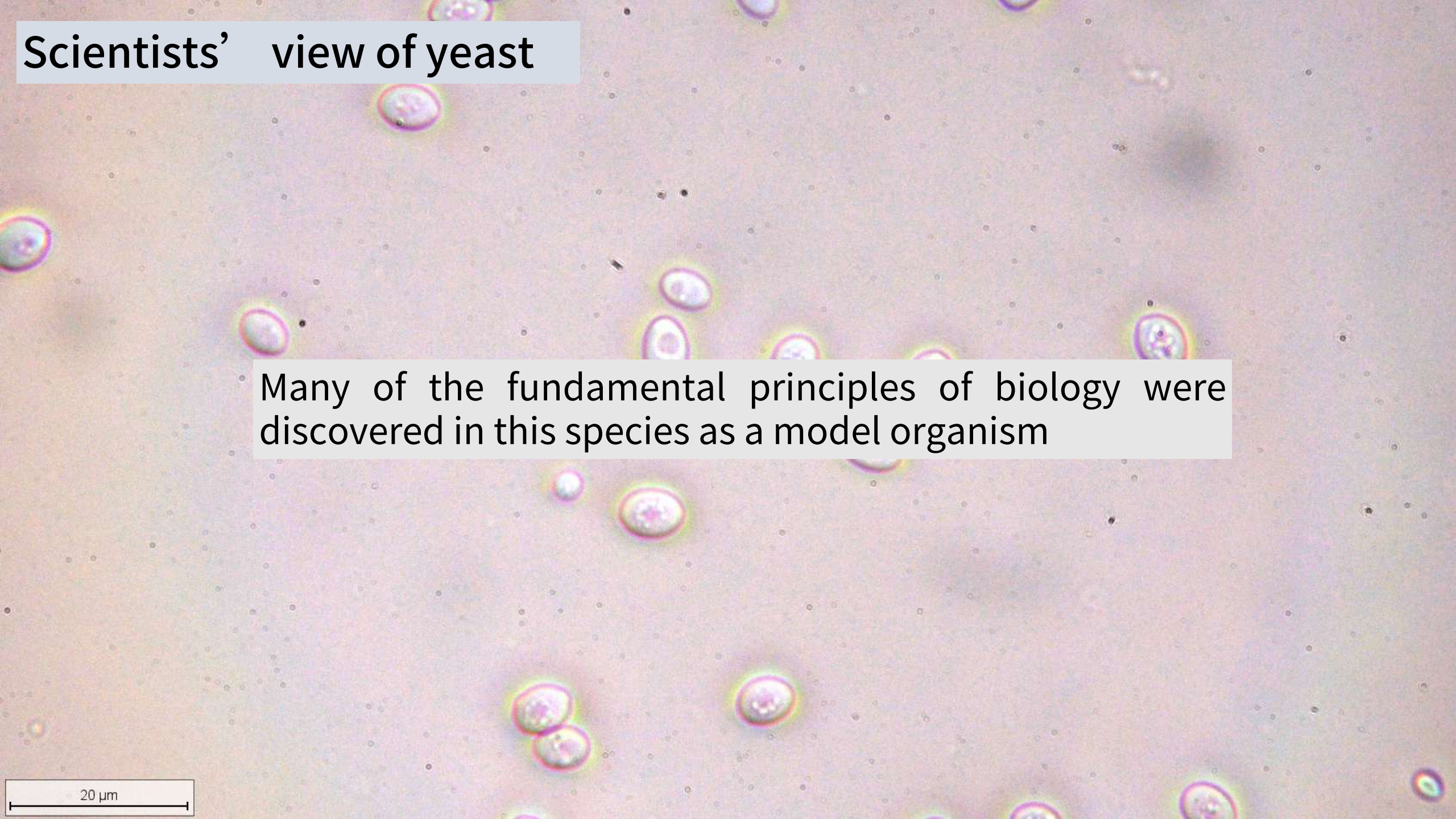
Extensive sampling of *Saccharomyces cerevisiae* in Taiwan reveals ecology and evolution of predomesticated lineages

尋覓臺灣野生的釀酒酵母，低調卻無所不在



Scientists’ view of yeast

Many of the fundamental principles of biology were discovered in this species as a model organism



那釀酒酵母的自然環境在哪呢？

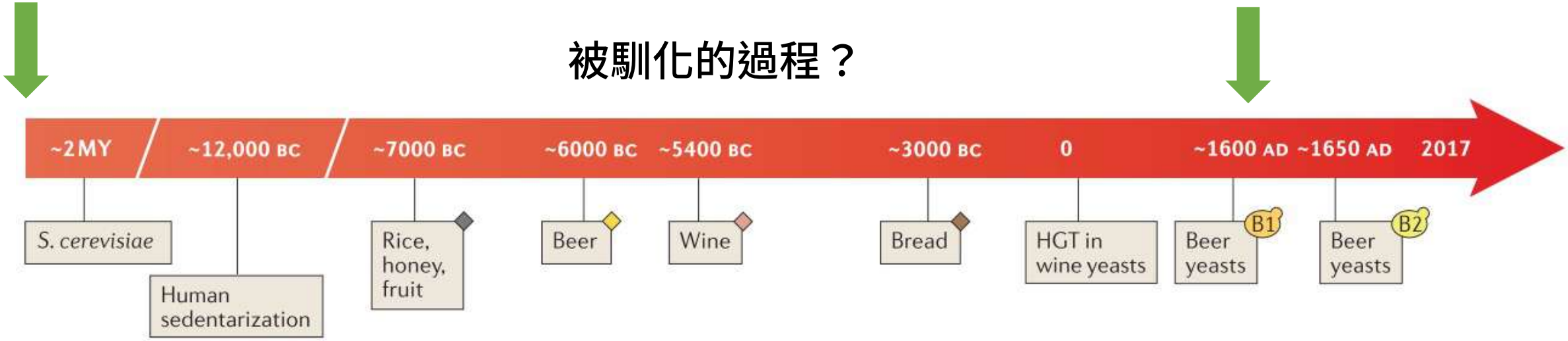
Evolution of *S. cerevisiae*

如果說釀酒酵母是透過動植物、尤其是人類而傳播，
那它最原本的發源地是在哪裡呢？

馴化前的過程？＊＊＊

第一個釀酒菌株？

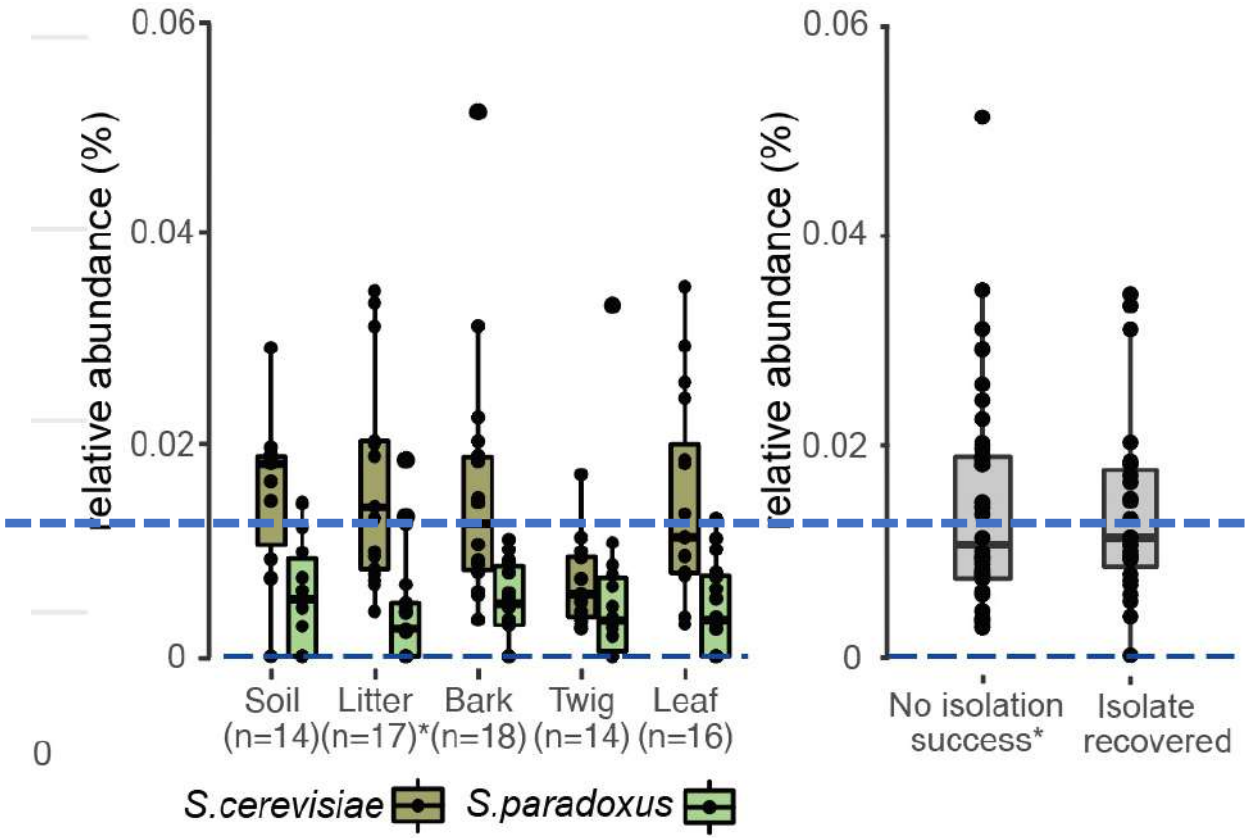
被馴化的過程？



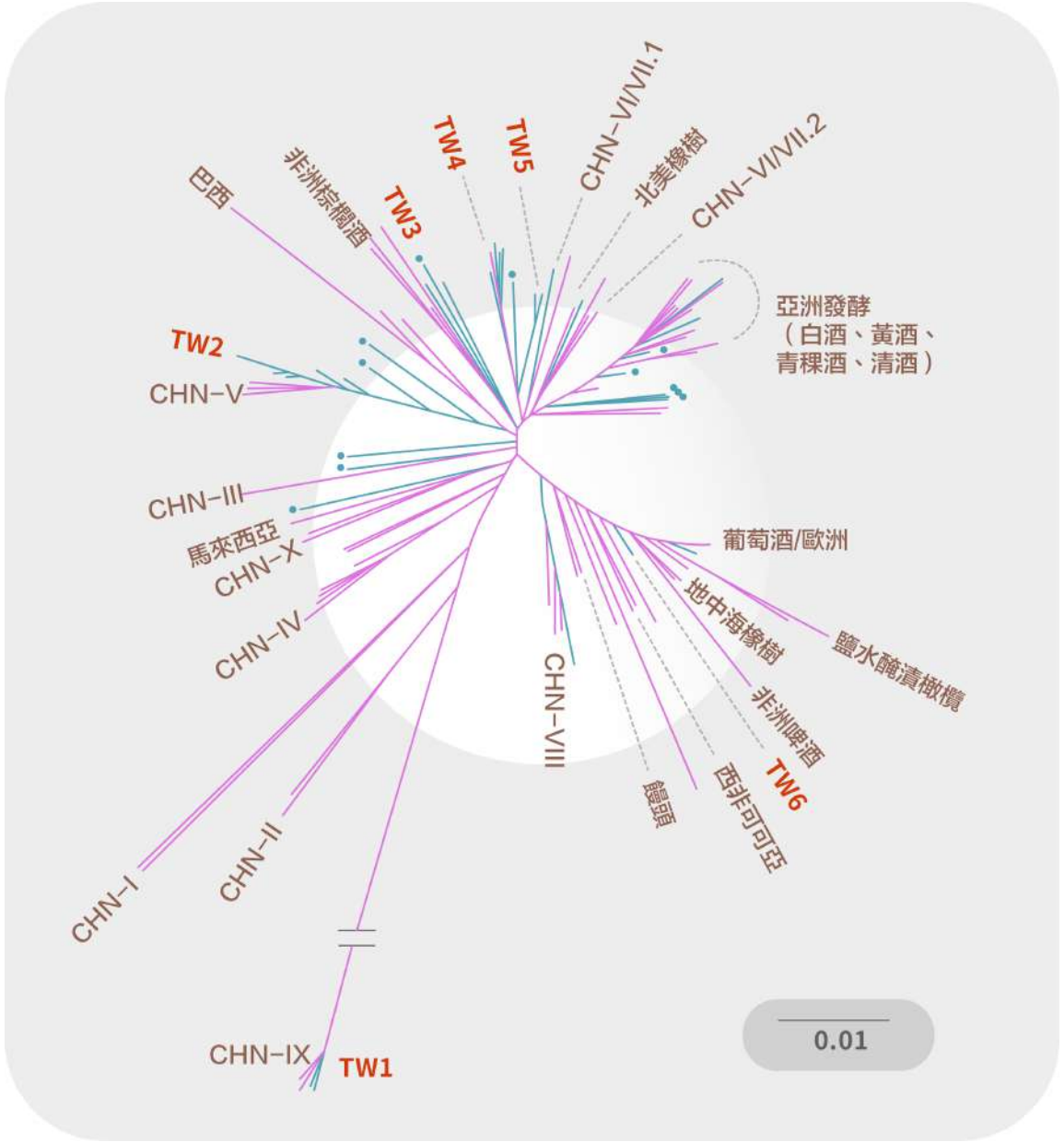
> **蔡怡陞歸納出的模式是：臺灣野外森林中，釀酒酵母普遍存在，但是比例非常低，可謂低調卻無所不在。**

如何歸納出以上結論呢？這要利用如今基因體學的新工具：總體基因體學（metagenomic）。原理是取得環境樣本後，直接定序其中所有 DNA 片段，或是所有物種都有的擴增子（amplicon），再與資料庫對照；如此一來，便能估計目標佔整體的比例，蔡怡陞團隊就是去估算釀酒酵母佔其生長環境中的比例。

從環境採樣培養出釀酒酵母以後，由中研院定序核心實驗室的呂美曄，回頭定序該樣本的擴增子，接著由蔡怡陞實驗室的林渝非分析。野外採集的樣本中，絕大部分是細菌，通常高達至少 99% 之多；剩下多半為真菌（和原生生物等等），其中只有極低比例是釀酒酵母，最多也只佔 0.012%。因此**同樣是細菌、真菌等微生物，釀酒酵母的存在感是低於 1% 中的 0.012% 以下，換句話說，不超過百萬分之 12！**



研之有物



將臺灣的一百多個菌株擺進演化樹，驚奇的事發生了！臺灣存在的釀酒酵母們，竟然也被歸類進各大譜系，並有新的譜系，這表示臺灣的釀酒酵母多樣性，和中國一樣高。

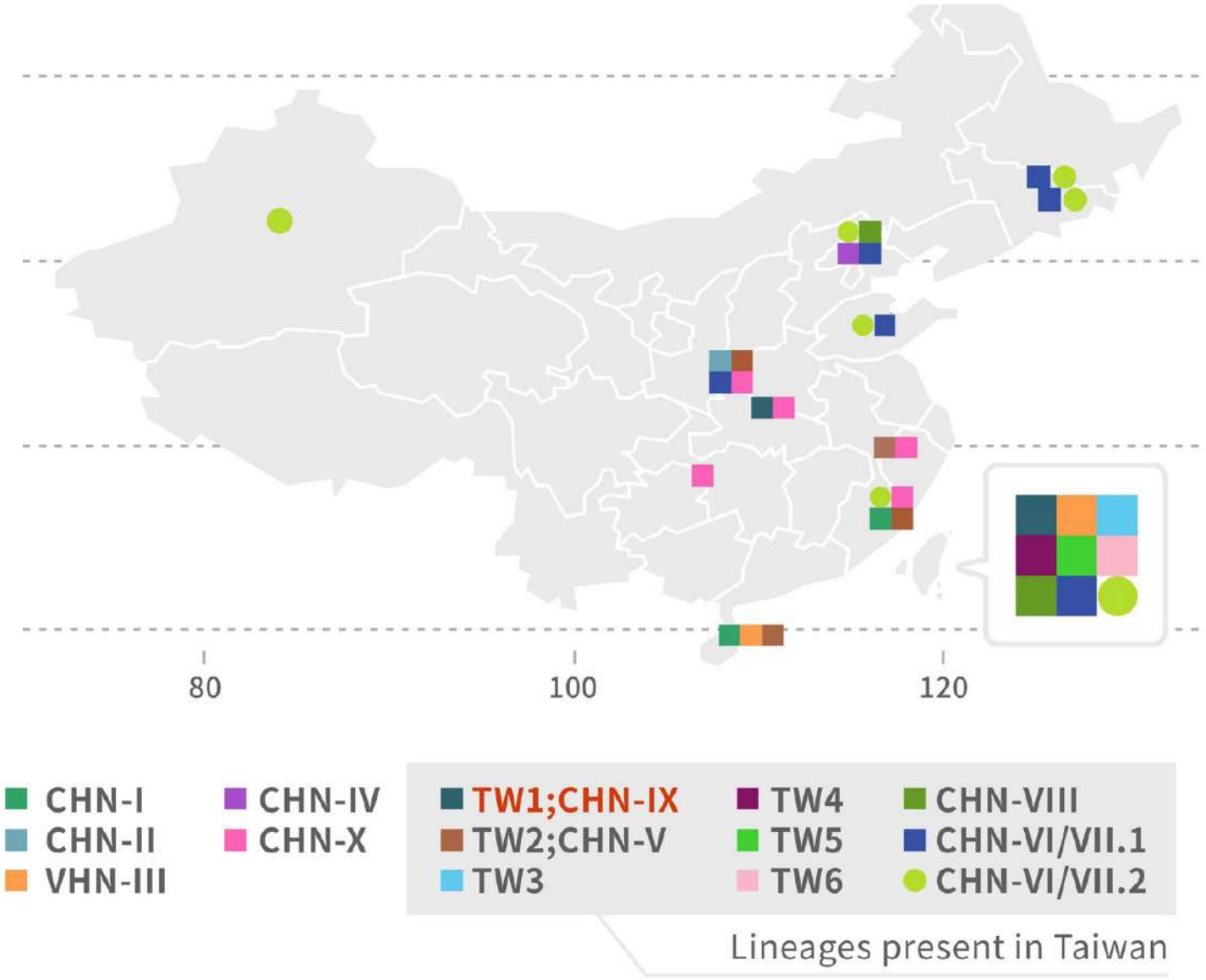
而且還有一款進入之前於中國採集到，與同類最早分家的那一群。

採集到的臺灣野生釀酒酵母譜系中，發現有一款和先前中國採集樣本都是最早分家的一群（黃框處），地理傳播也交織在一起。

其中 TW1 和 CHN-IX 皆為最早分家的一群，證明了台灣是發跡地之一。

這表示臺灣的釀酒酵母多樣性，也和中國一樣非常高。

Biogeography of *S. cerevisiae*



野生釀酒酵母在中國與臺灣的實際採樣分布，發現臺灣譜系的數量是全世界同尺度地區中最高的。

小小的臺灣擁有如此高的多樣性，就是讓人驚奇之處。