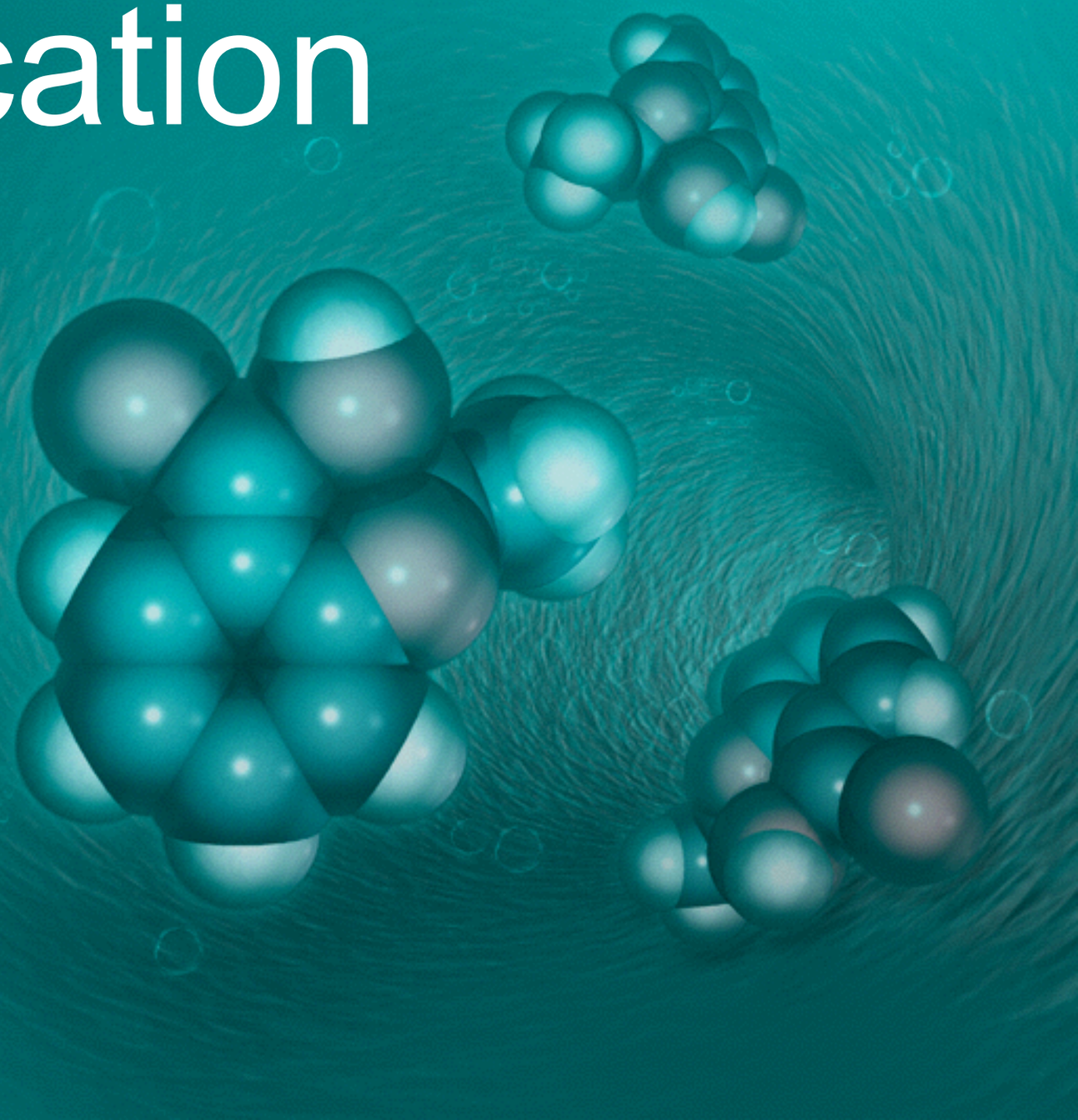
Multi-Omics Bioinformatics across Application Domains



Christoph Steinbeck

European Bioinformatics Institute
(EMBL-EBI)

The European Molecular Biology Laboratory

- 60 nationalities
- 1800 personnel

Main laboratory

Heidelberg, Gemany

Bioinformatics

Hinxton, Cambridge, UK

Structural biology

Grenoble, France

Neuroscience

Monterotondo, Rome, Italy

Structural biology

Hamburg, Germany

The

European Bioinformatics Institute

(EBI)

The

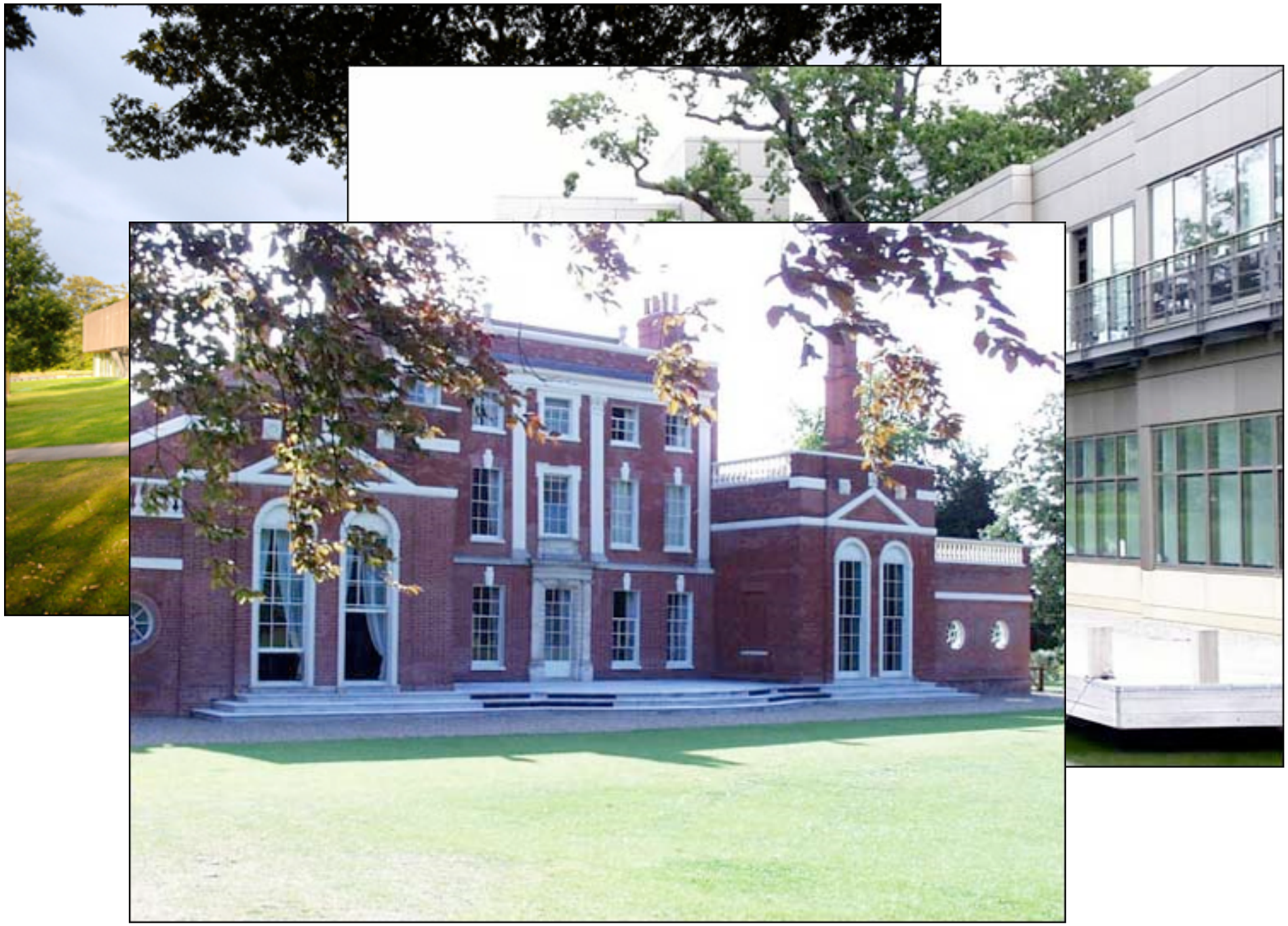
European Bioinformatics Institute

(EBI)

The

European Bioinformatics Institute

(EBI)



The

European Bioinformatics Institute

(EBI)

A word about BrExit: EMBL member states

Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland and the UK

Associate member states:
Argentina, Australia

Prospect member states:
Hungary, Poland, Slovak Republic

Our funders

- EMBL-EBI is primarily funded by EMBL member states.
- Other major funders:
  - European Commission
  - Research Councils UK
  - National Institutes of Health
  - Wellcome Trust
  - Industry Programme

BBSRC bioscience for the future

wellcome trust

NATIONAL INSTITUTES OF HEALTH

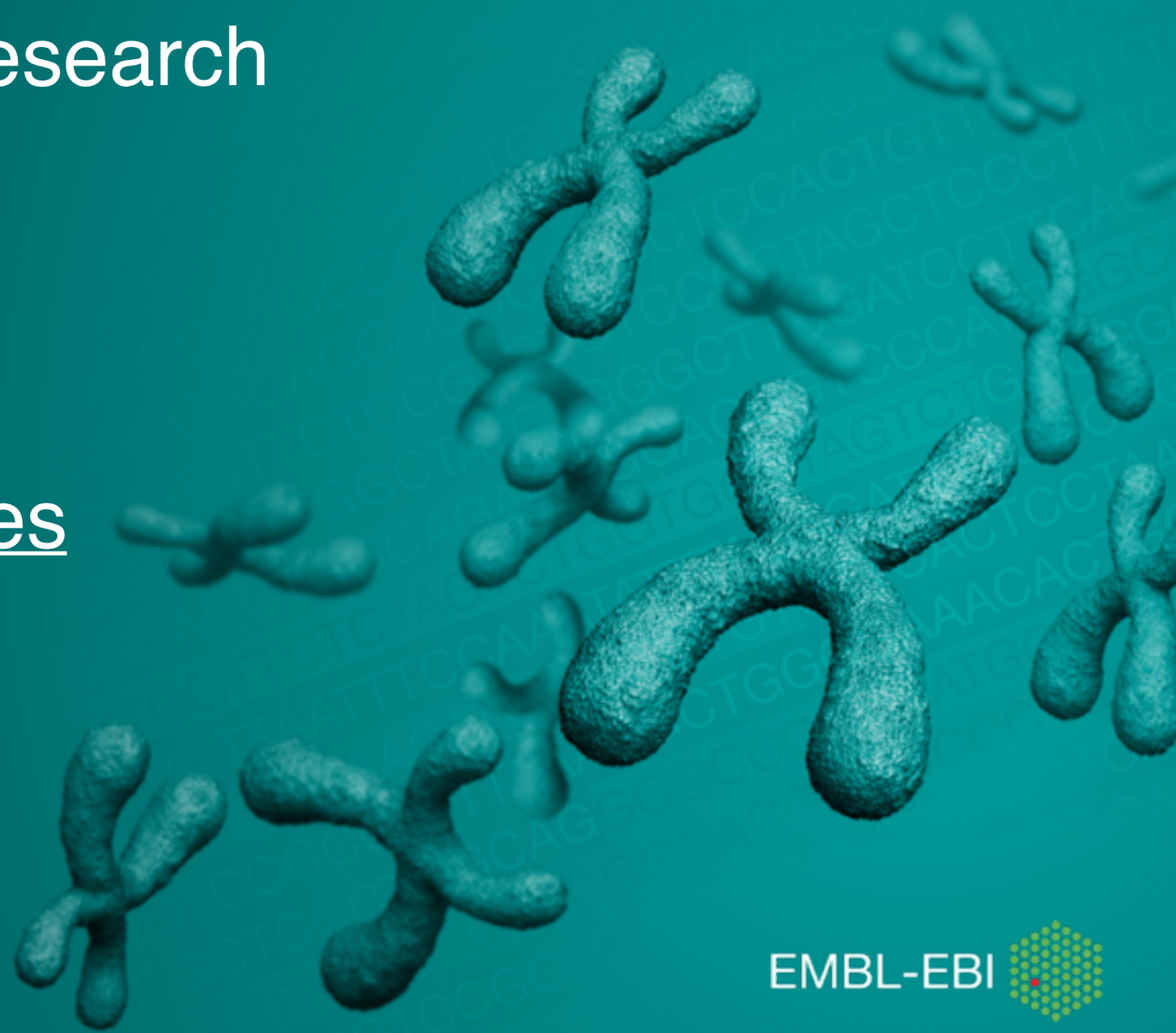
RESEARCH COUNCILS UK

MRC Medical Research Council

Bioinformatics services

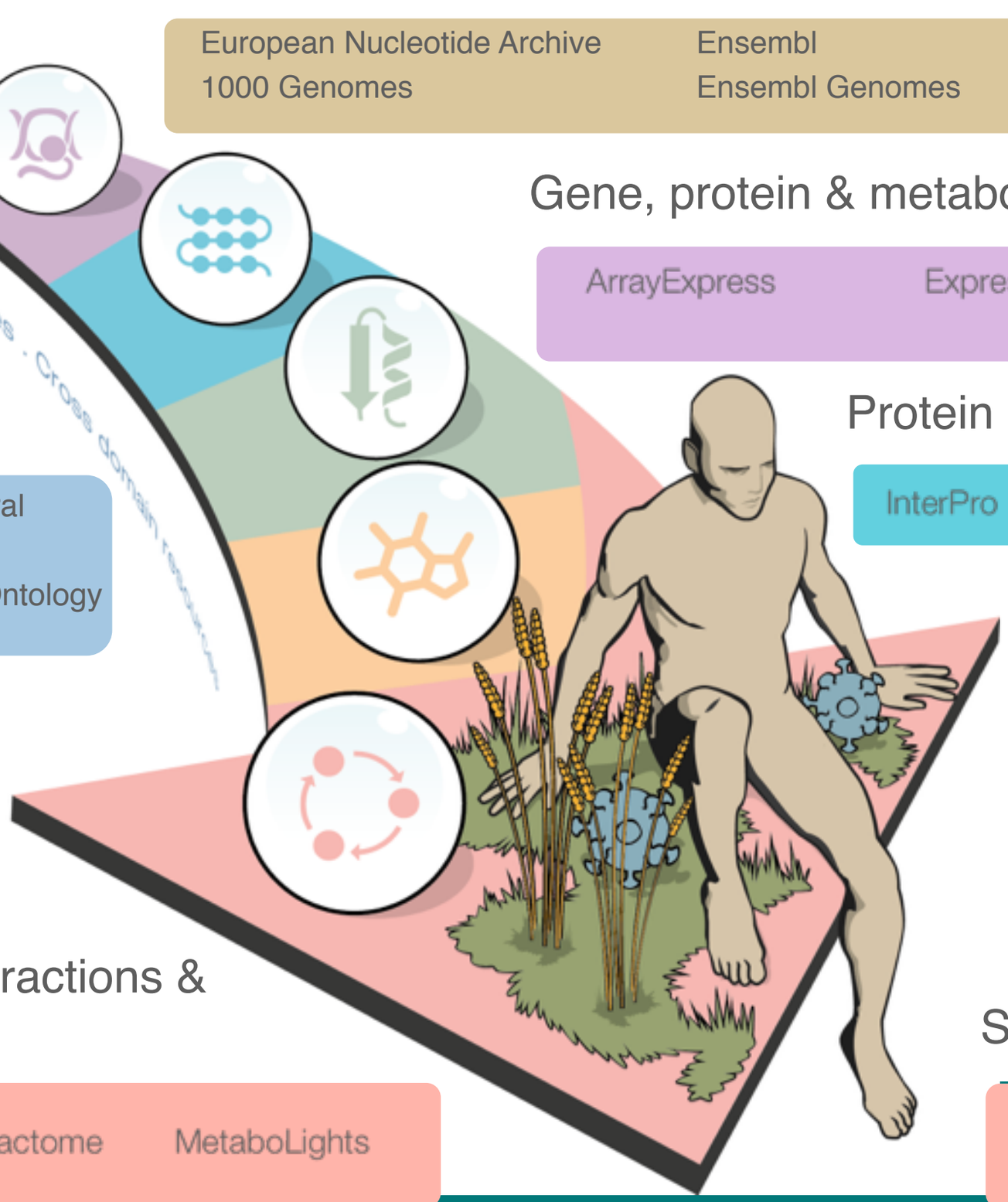
Data and tools to
support life-science research

www.ebi.ac.uk/services



EMBL-EBI

European Bioinformatics Institute (EBI)



Genes, genomes & variation

European Nucleotide Archive
1000 Genomes
Ensembl
Ensembl Genomes
European Genome-phenome Archive
Metagenomics portal

Gene, protein & metabolite expression

ArrayExpress
Expression Atlas
Metabolights
PRIDE

Protein sequences, families & motifs

InterPro
Pfam
UniProt

Molecular structures

Protein Data Bank in Europe
Electron Microscopy Data Bank

Chemical biology

ChEMBL
ChEBI

Systems

BioModels
BioSamples
Enzyme Portal

Literature & ontologies

Europe PubMed Central
Gene Ontology
Experimental Factor Ontology

Reactions, interactions & pathways

IntAct
Reactome
MetaboLights

Big data, big demand

~18.5 million
requests to EMBL-EBI websites every day

Scientists at over
5 million
unique sites use EMBL-EBI websites

EMBL-EBI handles
9.2 million
jobs on average per month

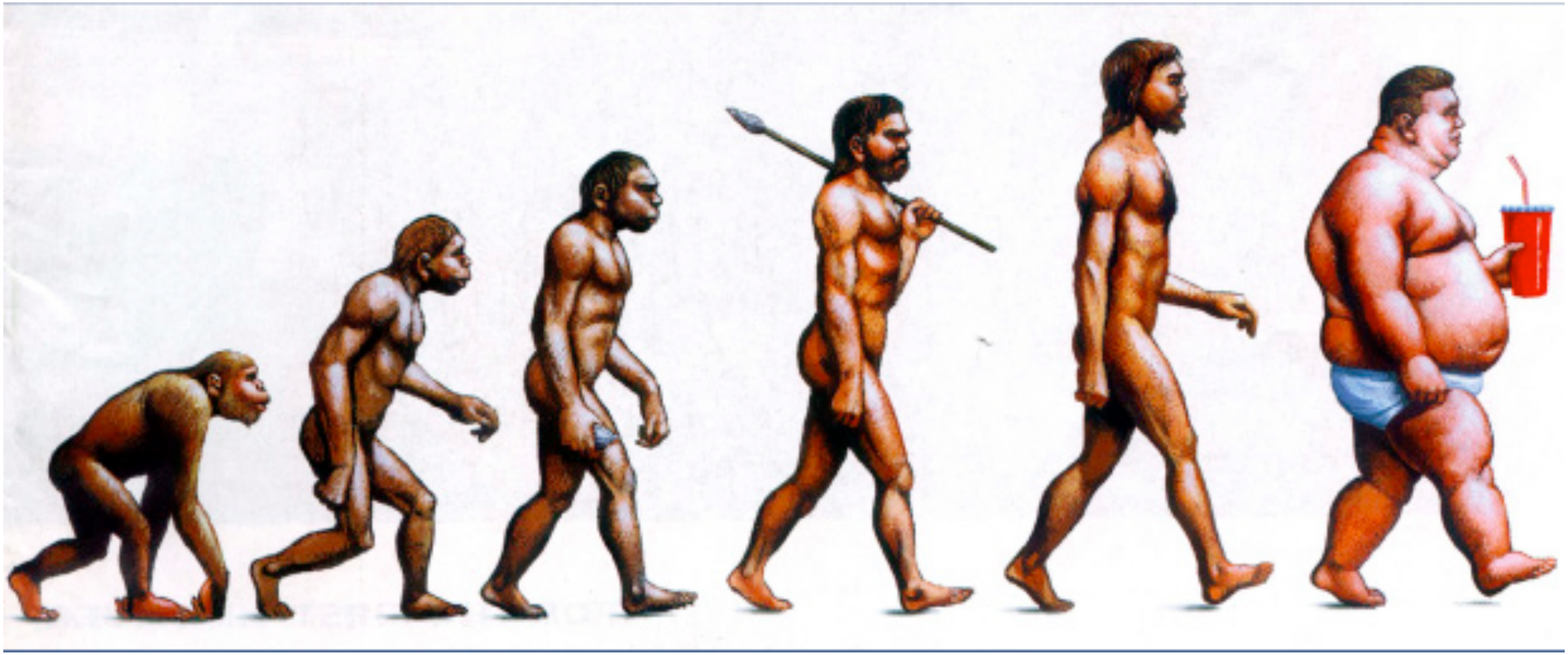
60 petabytes
of EMBL-EBI storage capacity

Post Genomic Era

[O]ur understanding of the human genome has changed in the most fundamental ways. The small number of genes -- some 30,000 -- supports the notion that we are not hard wired. We now know the notion that one gene leads to one protein, and perhaps one disease, is false.

Craig Venter, June 2001

Genes are not the full story



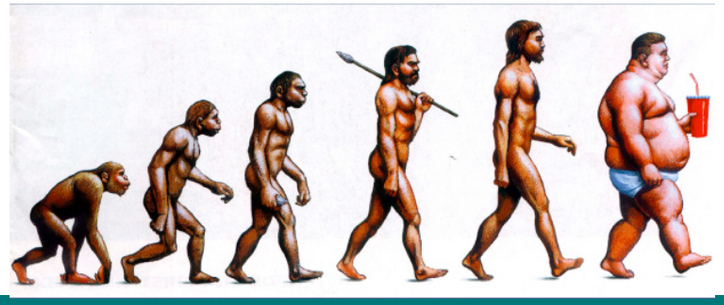
Daily Mail

'Indians are getting as fat as Americans': Obesity crisis swells among India's middle class youth as children choose Western fast food over traditional cuisine

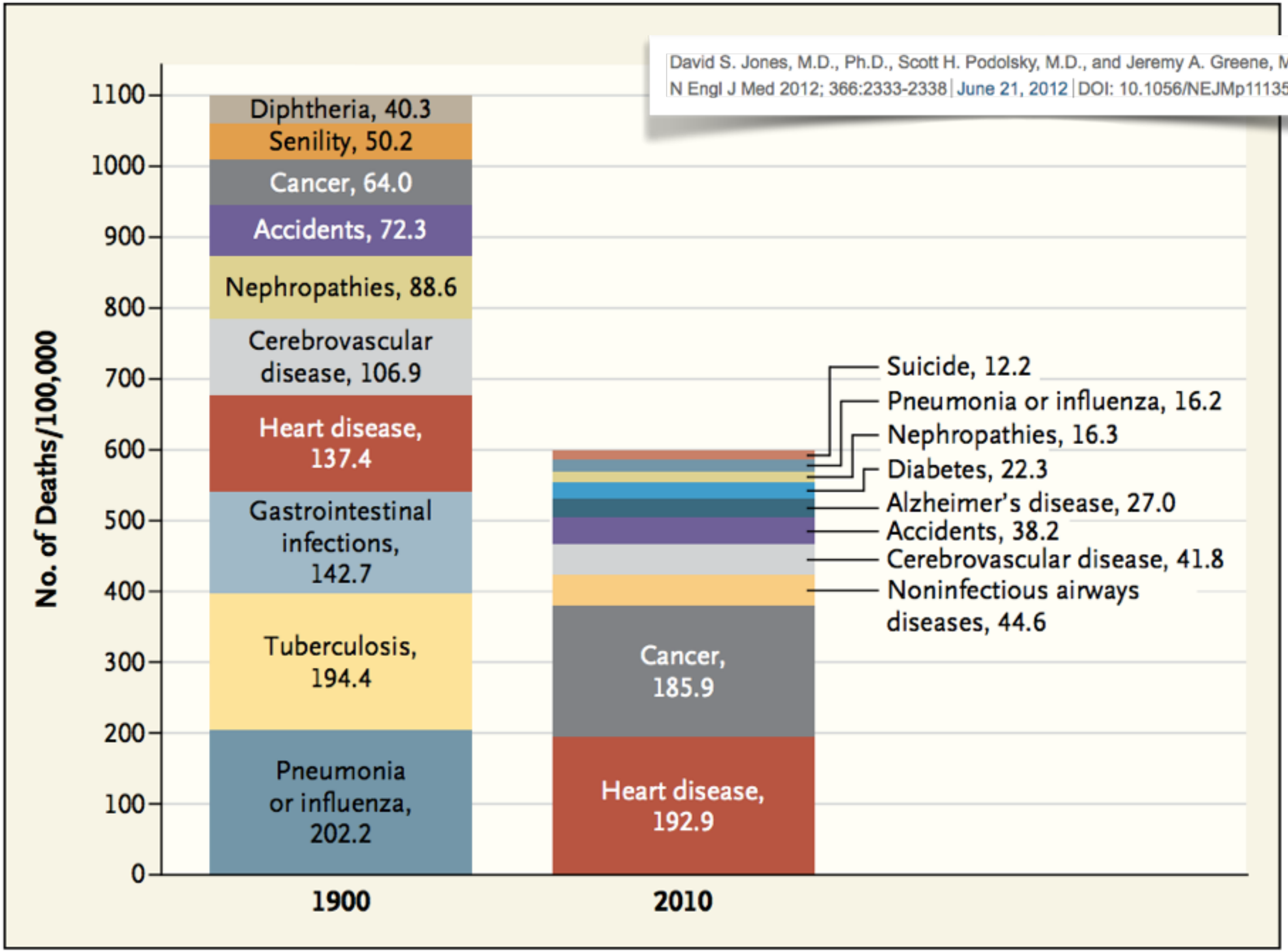
- Obesity spreading across India as a result of Western food invasion
- Indian fast food market worth £7bn - expected to double by 2016
- Doctors fitting gastric bands on children as young as 13
- As a result of the obesity epidemic the country also has diabetes epidemic
- India already has largest diabetes population in world: 50m sufferers
- Medical professionals expect it to reach 100m soon
- Indians are more genetically predisposed to developing diabetes
- Documentary India's Supersize Kids on BBC iPlayer now

By MARTHA DE LACEY

PUBLISHED: 16:58, 15 August 2013 | **UPDATED:** 17:00, 15 August 2013

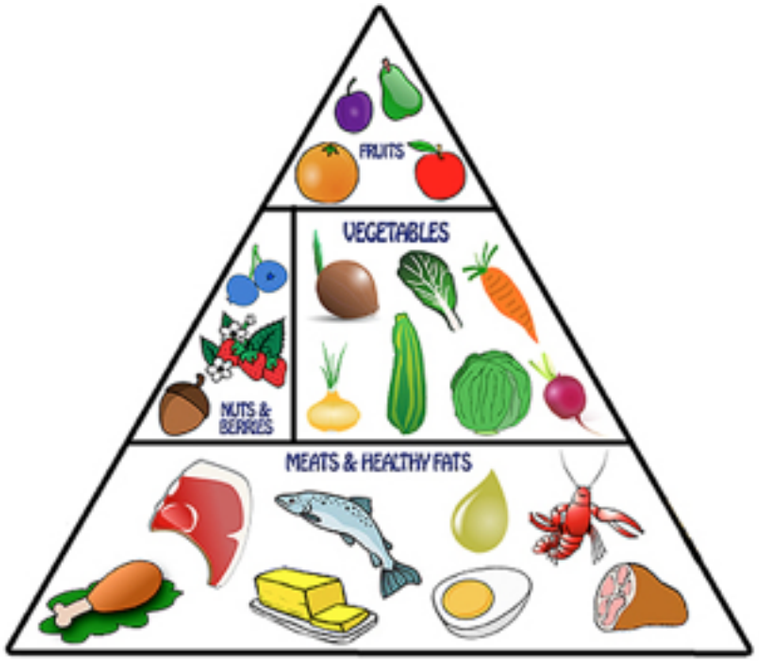


David S. Jones, M.D., Ph.D., Scott H. Podolsky, M.D., and Jeremy A. Greene, M.D., Ph.D.
N Engl J Med 2012; 366:2333-2338 | June 21, 2012 | DOI: 10.1056/NEJMp1113569



Top 10 Causes of Death: 1900 vs. 2010.

EMBL-EBI



Nutrition

Disease

Exercise

Phenome/ Exposome

Age

Environment

Drugs

The Metabolome is the most accessible and dynamically changing Molecular Phenotype

EMBL-EBI

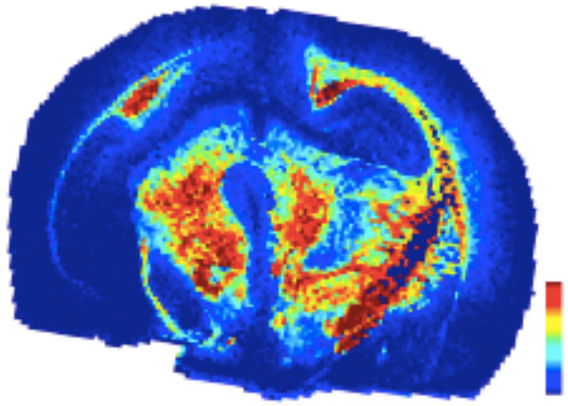
Metabolomics

Measures **occurrence** and **concentrations** of many small molecules (**metabolites**) in an organism at once.

Diagnostic fluids

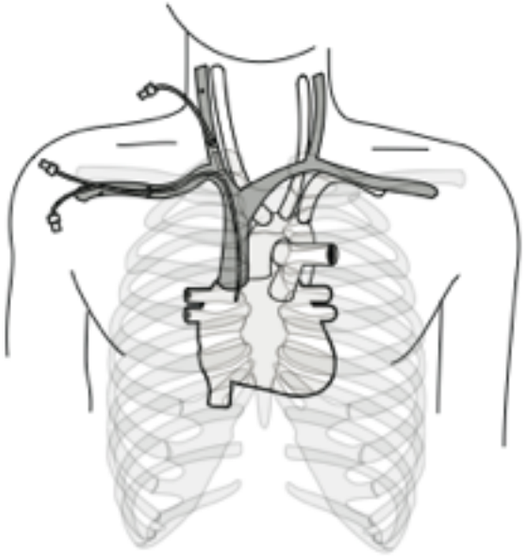
Urine
(time-averaged data)

Plasma
(snap-shot data)



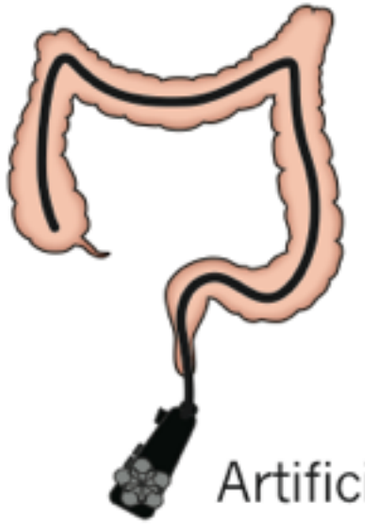
Organism Parts

Other accessible analytical compartments



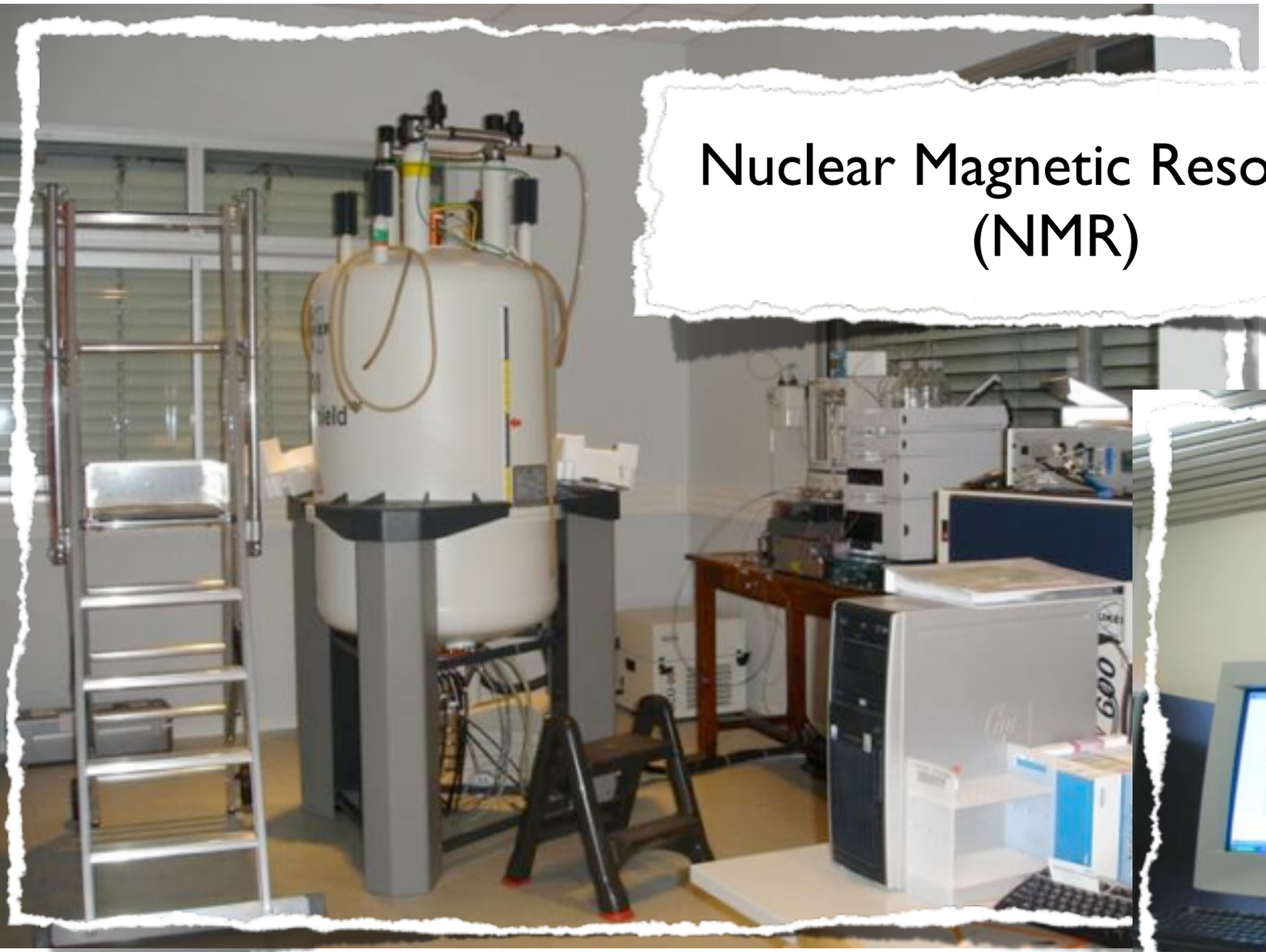
Specialized fluids and biopsies
(selected fluids)

Pathological fluids



Artificial fluids

Metabolomics uses a wide-range of analytical techniques

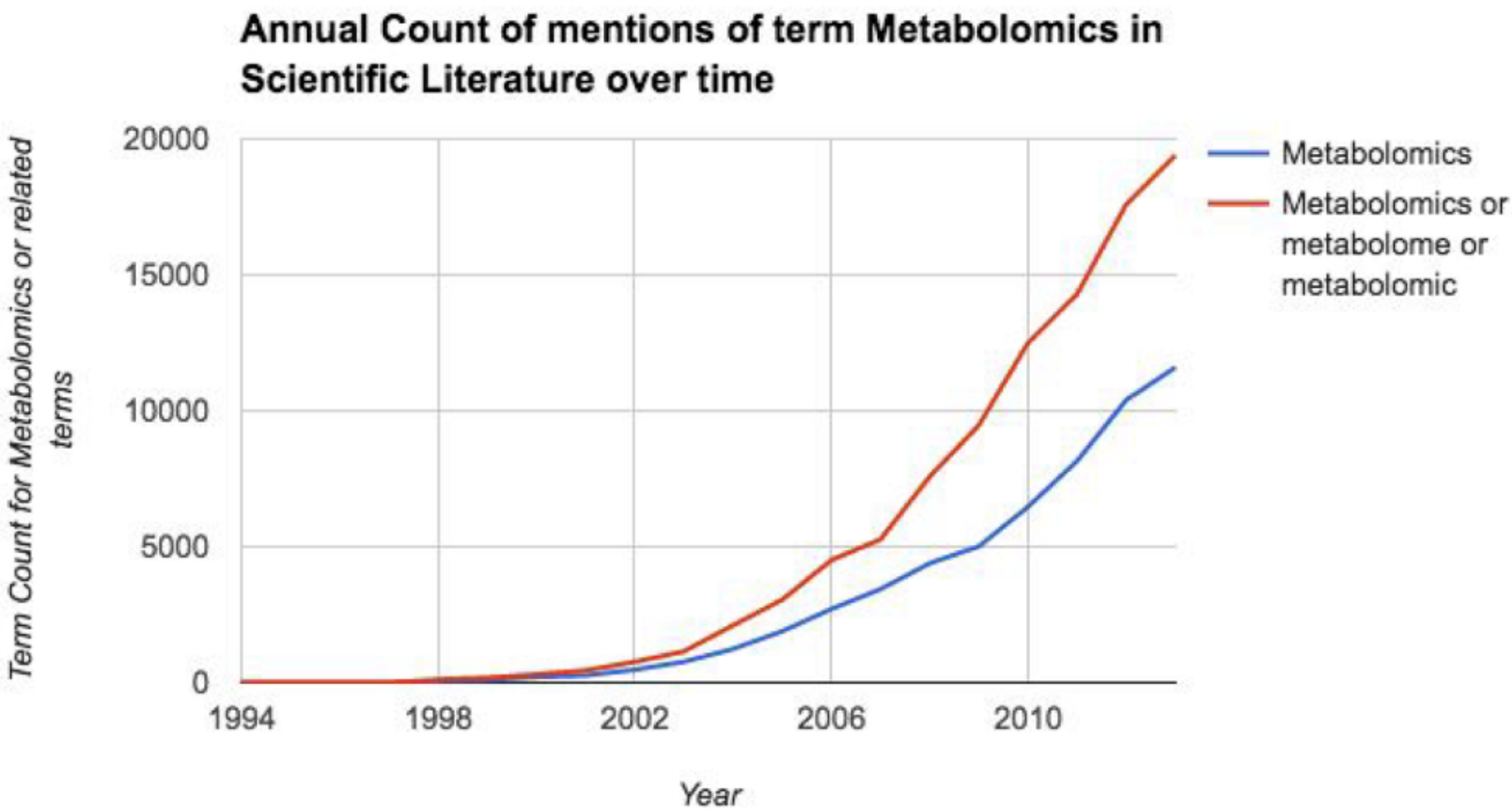


Nuclear Magnetic Resonance (NMR)

Mass Spec

EMBL-EBI

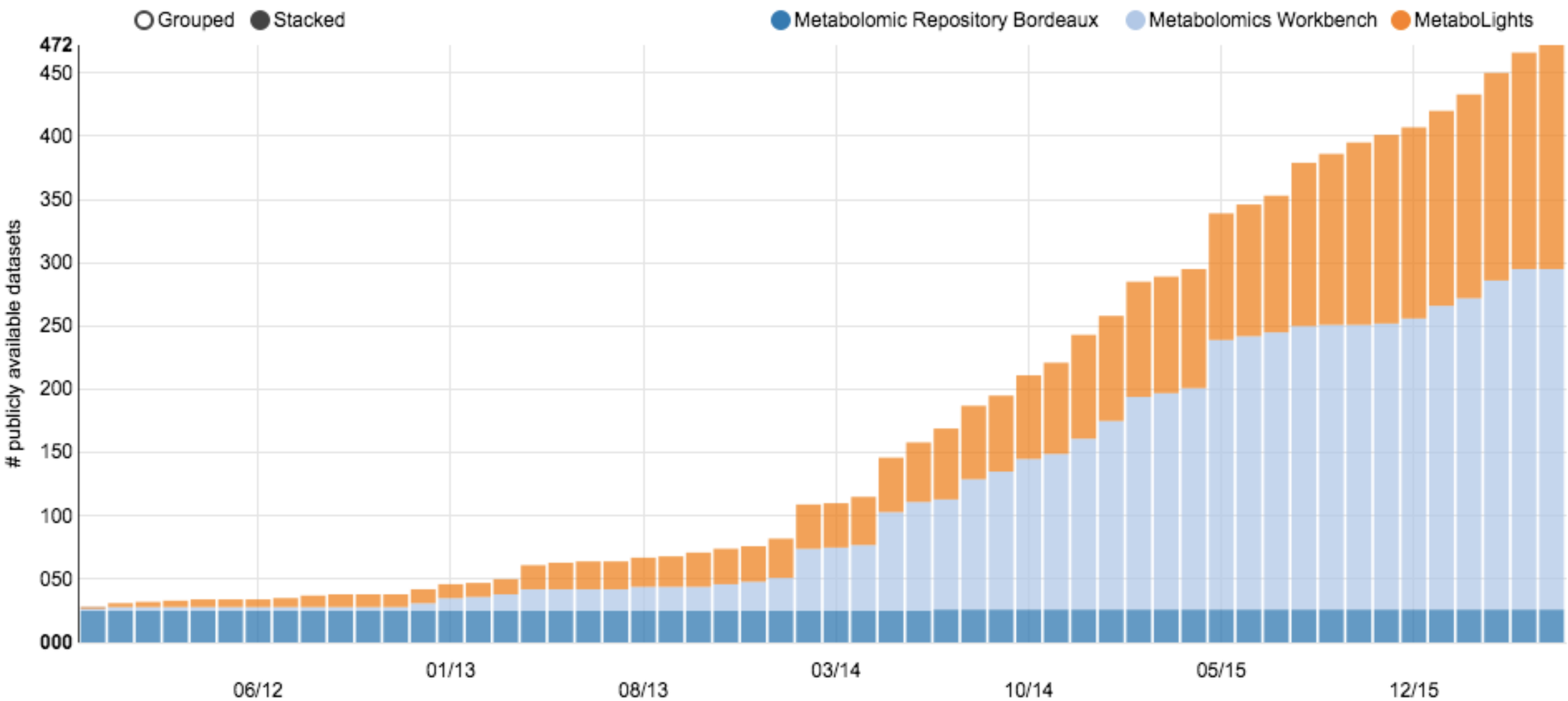
Metabolomics has taken off world-wide



MetaboloMeXchange

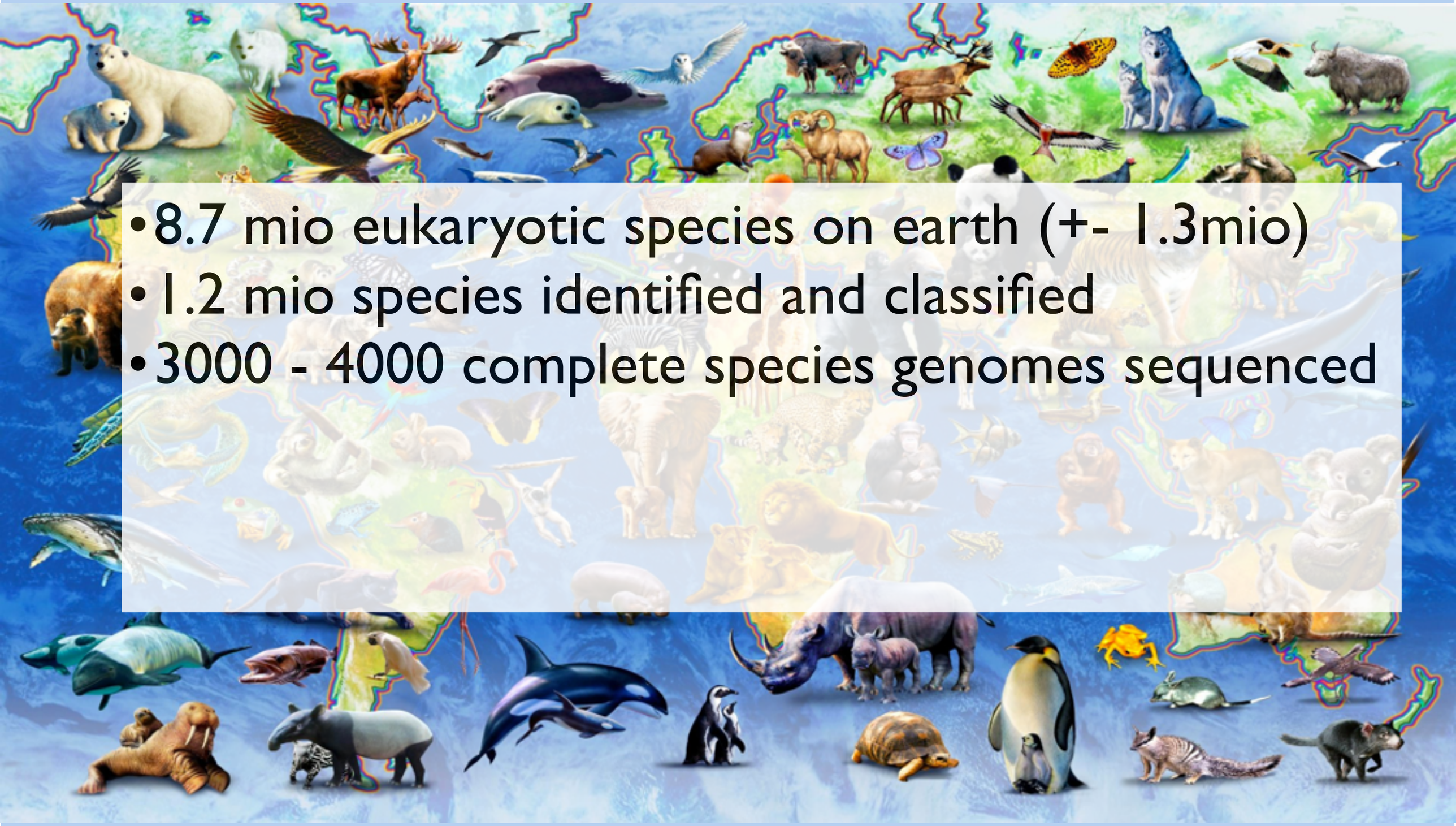
An international data aggregation and notification service for metabolomics.

Stats

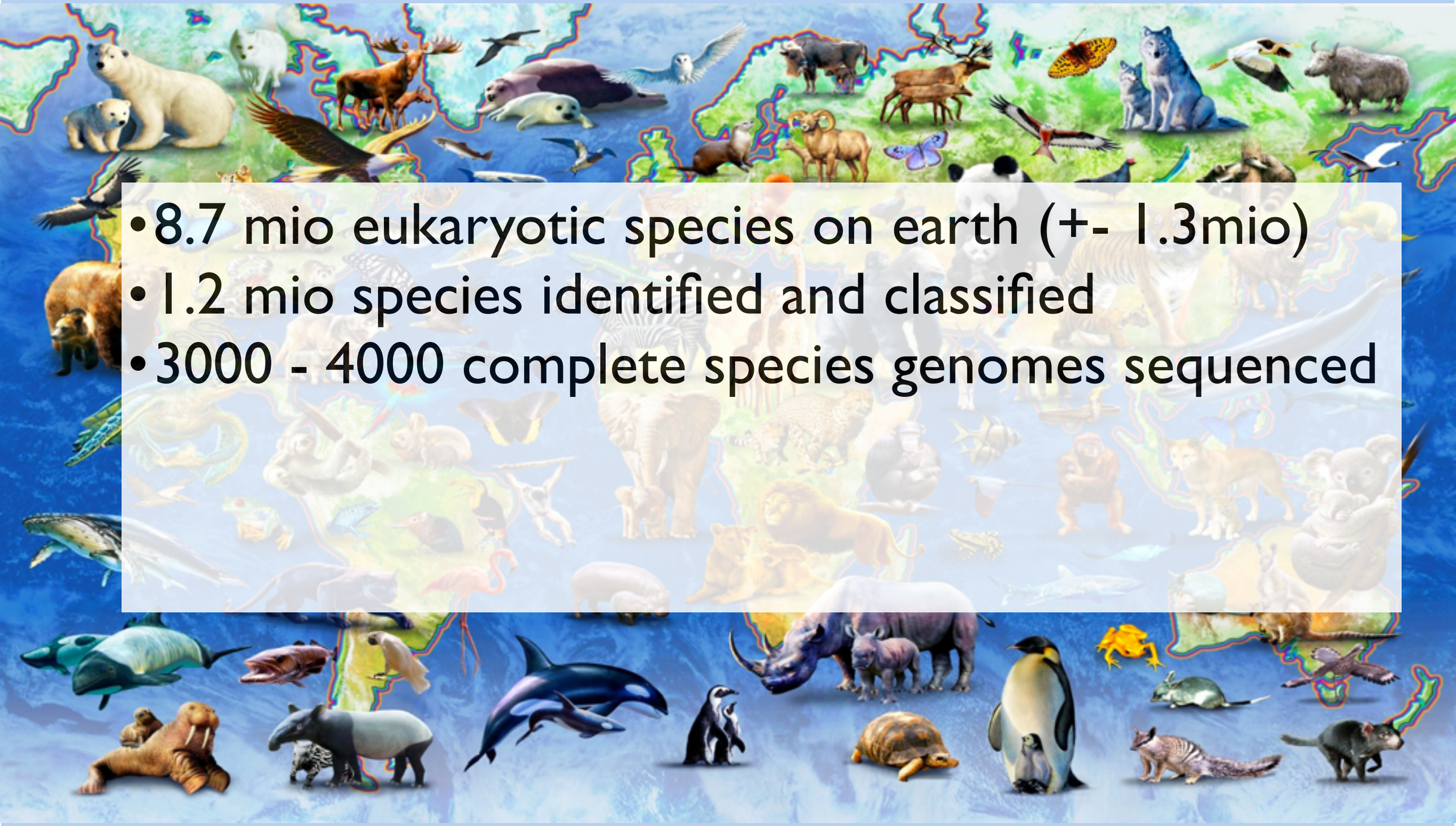


- 8.7 mio eukaryotic species on earth (+- 1.3mio)

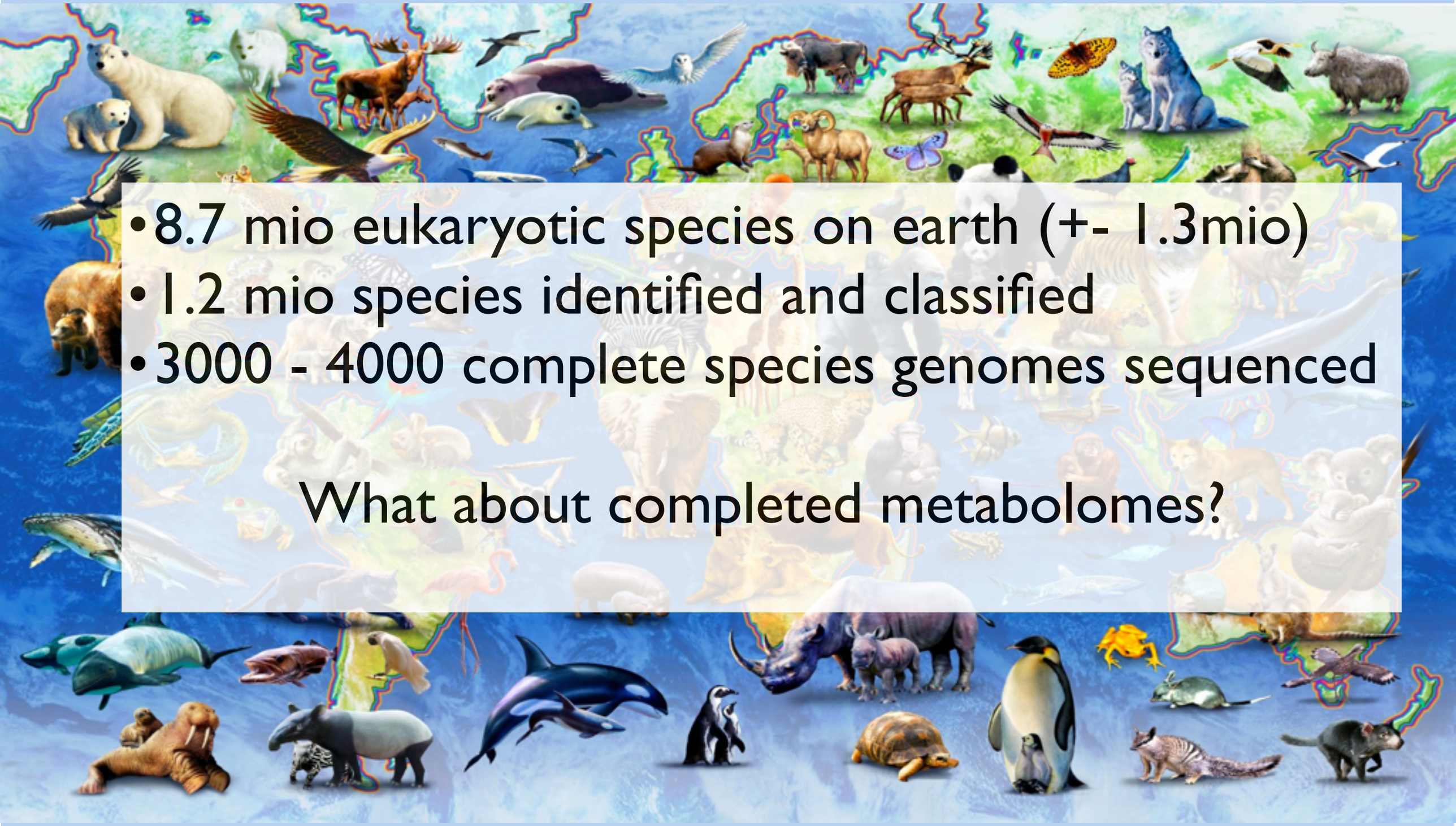
- 8.7 mio eukaryotic species on earth (+- 1.3mio)
- 1.2 mio species identified and classified



- 8.7 mio eukaryotic species on earth (+- 1.3mio)
- 1.2 mio species identified and classified
- 3000 - 4000 complete species genomes sequenced

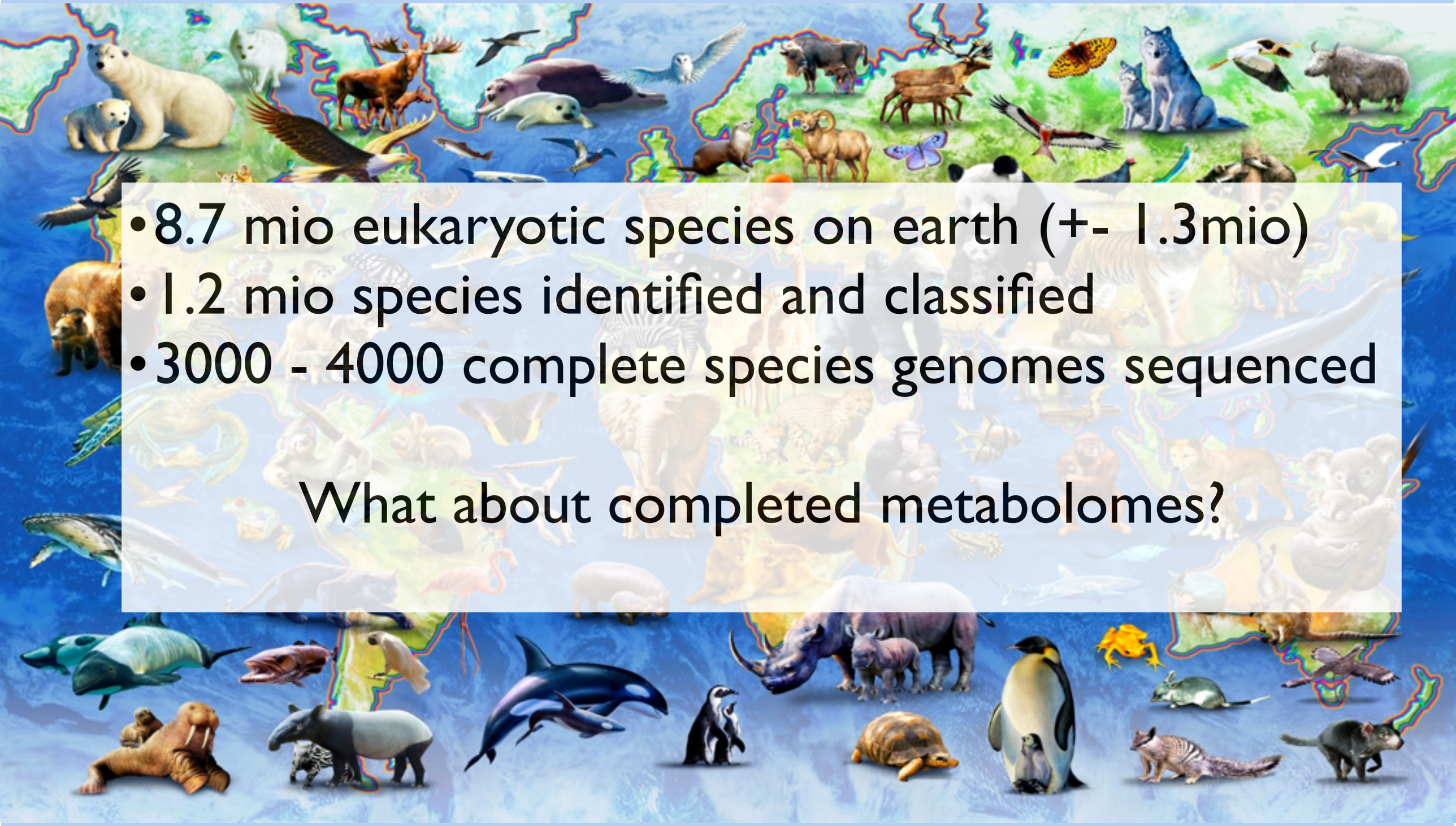


- 8.7 mio eukaryotic species on earth (+- 1.3mio)
- 1.2 mio species identified and classified
- 3000 - 4000 complete species genomes sequenced



- 8.7 mio eukaryotic species on earth (+- 1.3mio)
- 1.2 mio species identified and classified
- 3000 - 4000 complete species genomes sequenced

What about completed metabolomes?



- 8.7 mio eukaryotic species on earth (+- 1.3mio)
- 1.2 mio species identified and classified
- 3000 - 4000 complete species genomes sequenced

What about completed metabolomes?

metabolites

Metabolites. 2016 Feb 15;6(1)

MDPI

Communication

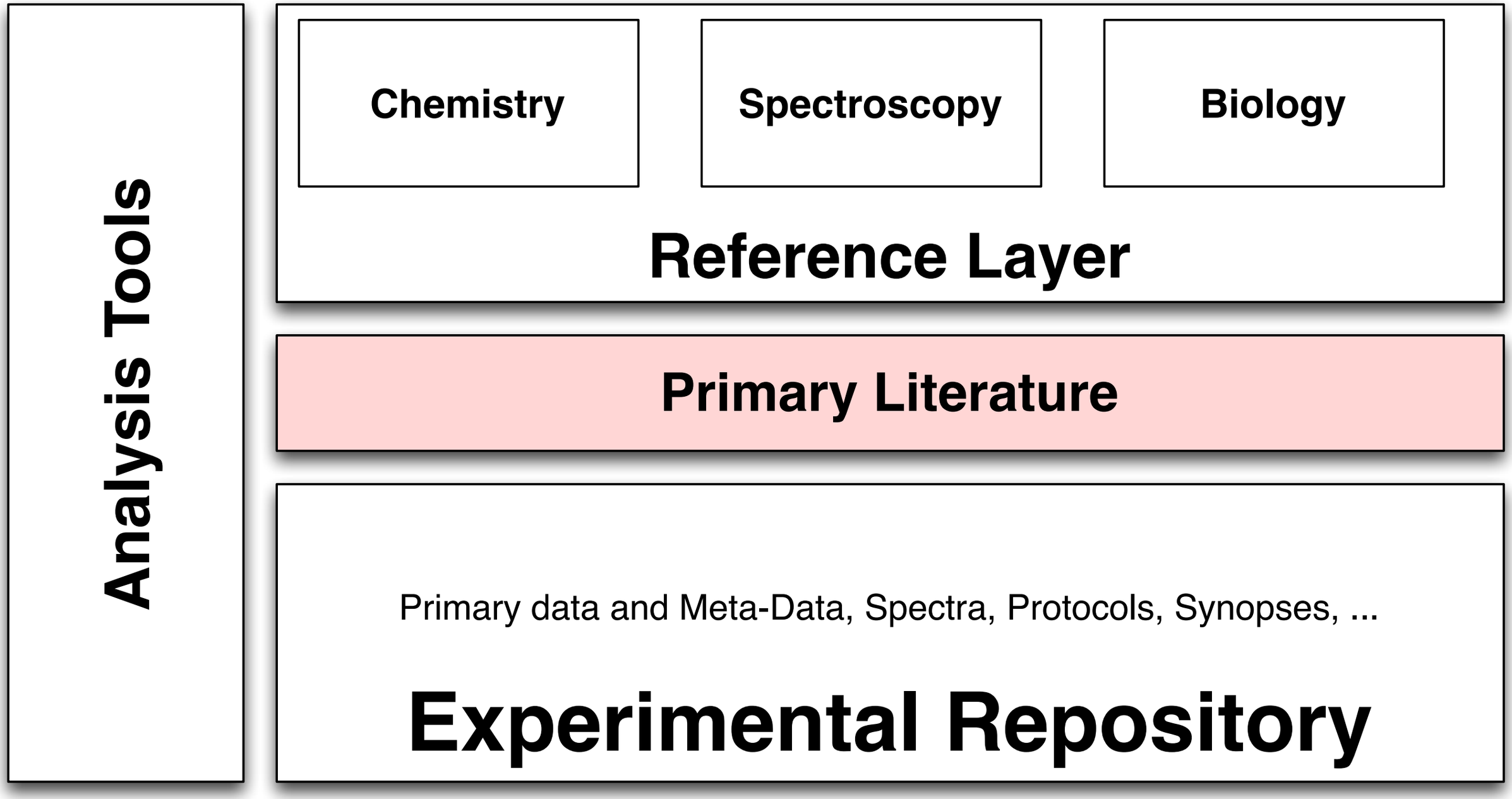
The Time Is Right to Focus on Model Organism Metabolomes

Arthur S. Edison [1], Robert D. Hall [2], Christophe Junot [3], Peter D. Karp [4], Irwin J. Kurland [5], Robert Mistrik [6], Laura K. Reed [7], Kazuki Saito [8], Reza M. Salek [9], Christoph Steinbeck [9], Lloyd W. Sumner [10] and Mark R. Viant [11,*]

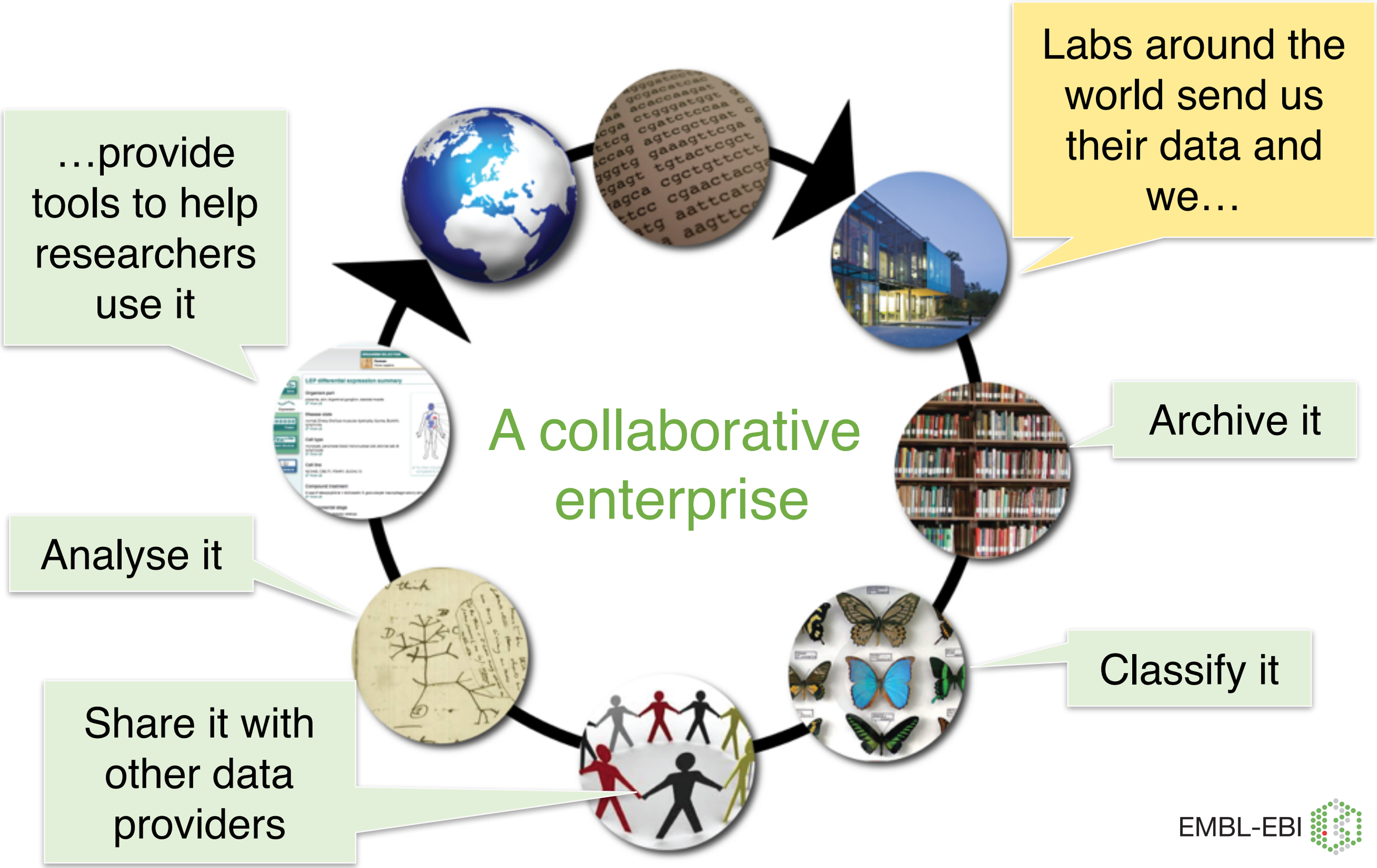
Building upon extensive genomics research, we argue that the time is now right to focus intensively on model organism metabolomes. We propose a grand challenge for metabolomics studies of model organisms: to identify and map all metabolites onto metabolic pathways, to develop quantitative metabolic models for model organisms, and to relate organism metabolic pathways within the context of evolutionary metabolomics, *i.e.*, phylometabolomics. These efforts should focus on a series of established model organisms in microbial, animal and plant research.

Species Metabolomes are being assembled on the fly right now through data sharing in Metabolomics

MetaboLights Database at the EBI



MetaboLights Database at the EBI



EMBL-EBI

MetaboLights Database at the EBI

Labs around the
rld send us
eir data and
we…

…provide
tools to hel
researcher
use it

Data at the EBI
can be
used freely
by anyone for any
purpose

Archive it

Analyse it

Classify it

Share it with
other data
providers

EMBL-EBI

MetaboLights Repository at the EBI

Labs around the world send us their data and we…

…provide tools to hel researcher use it

EBI databases are supported over decades

Archive it

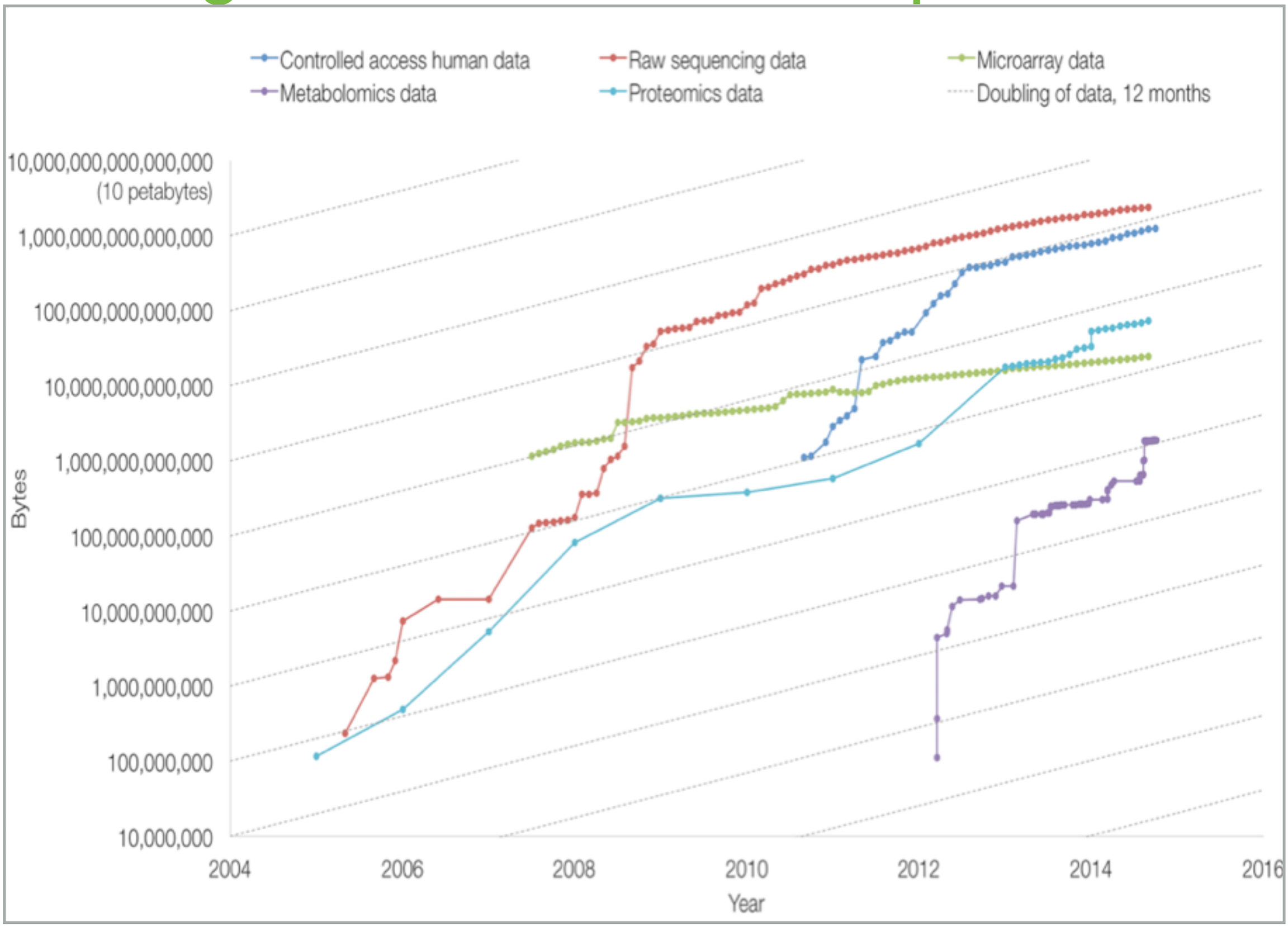
Analyse it

Classify it

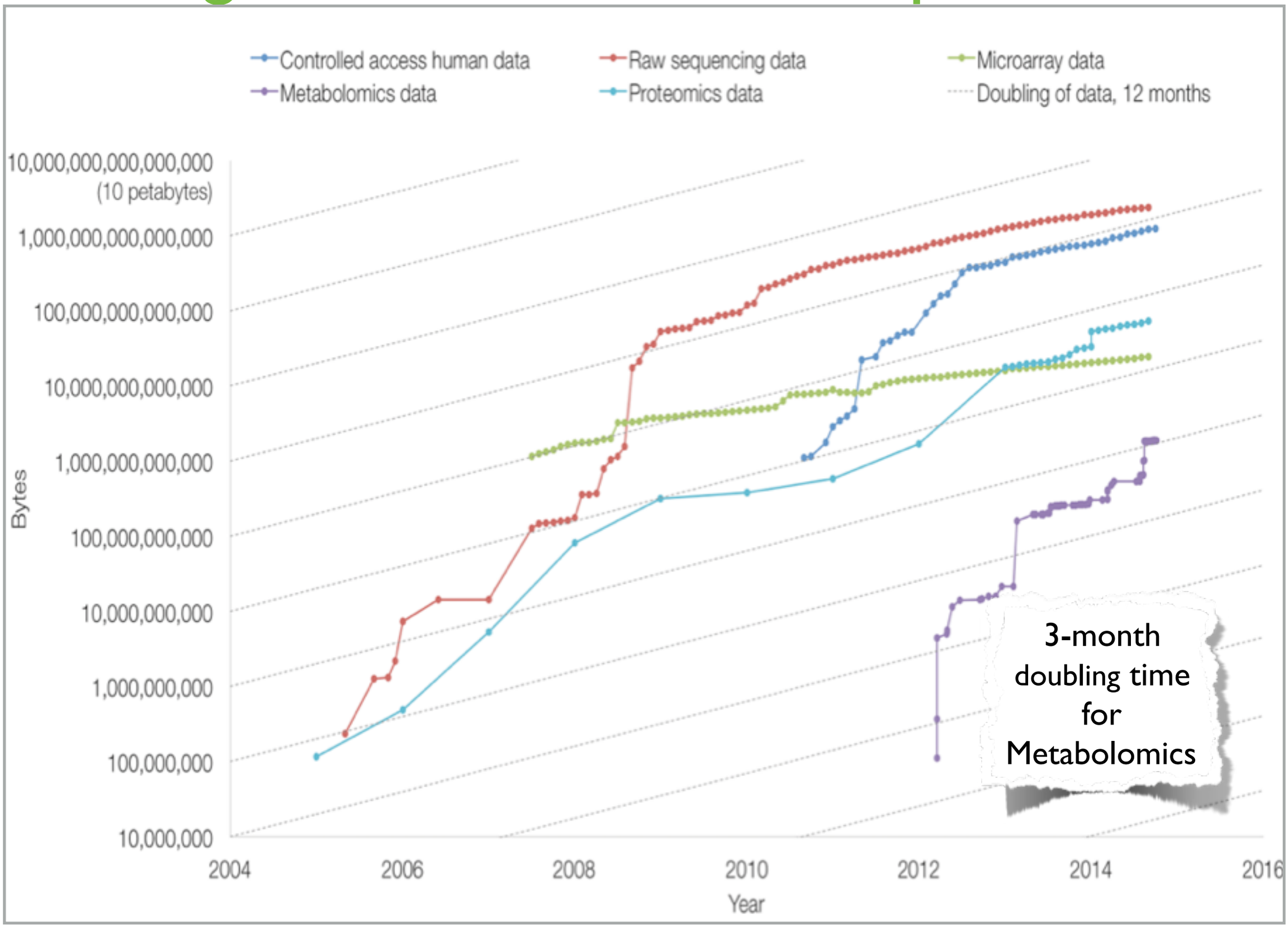
Share it with other data providers

EMBL-EBI

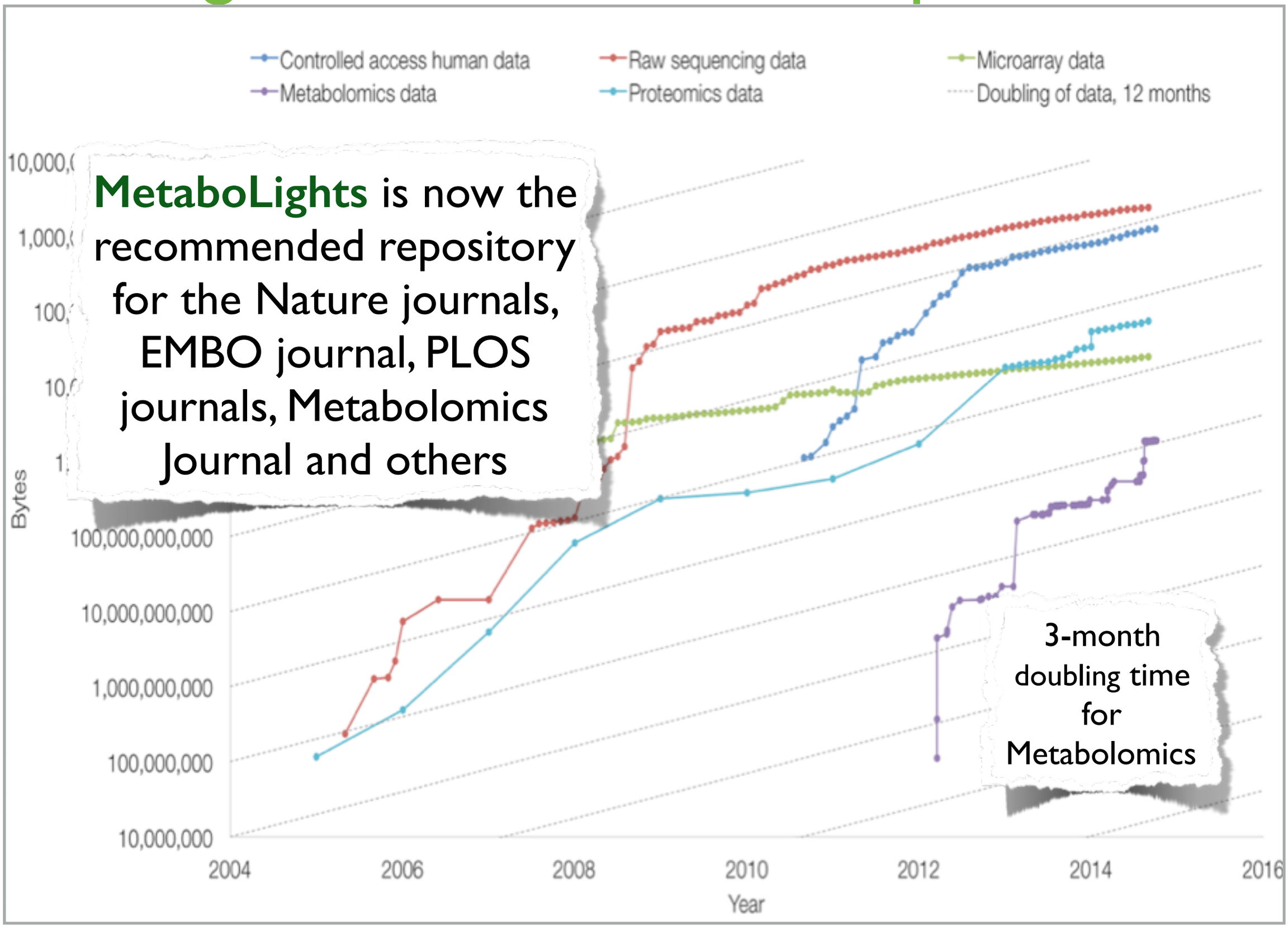
Data growth in EBI data repositories

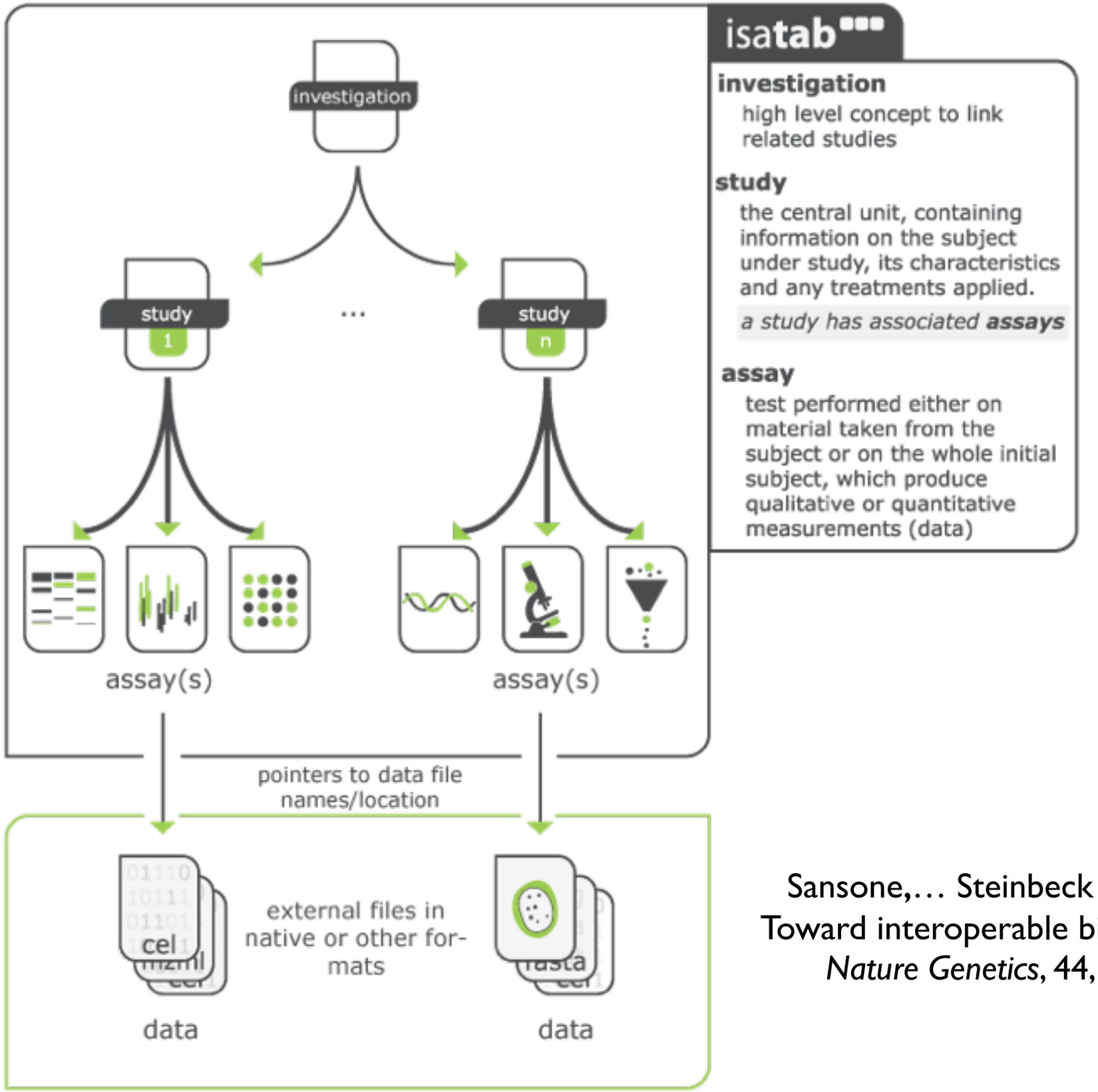


Data growth in EBI data repositories



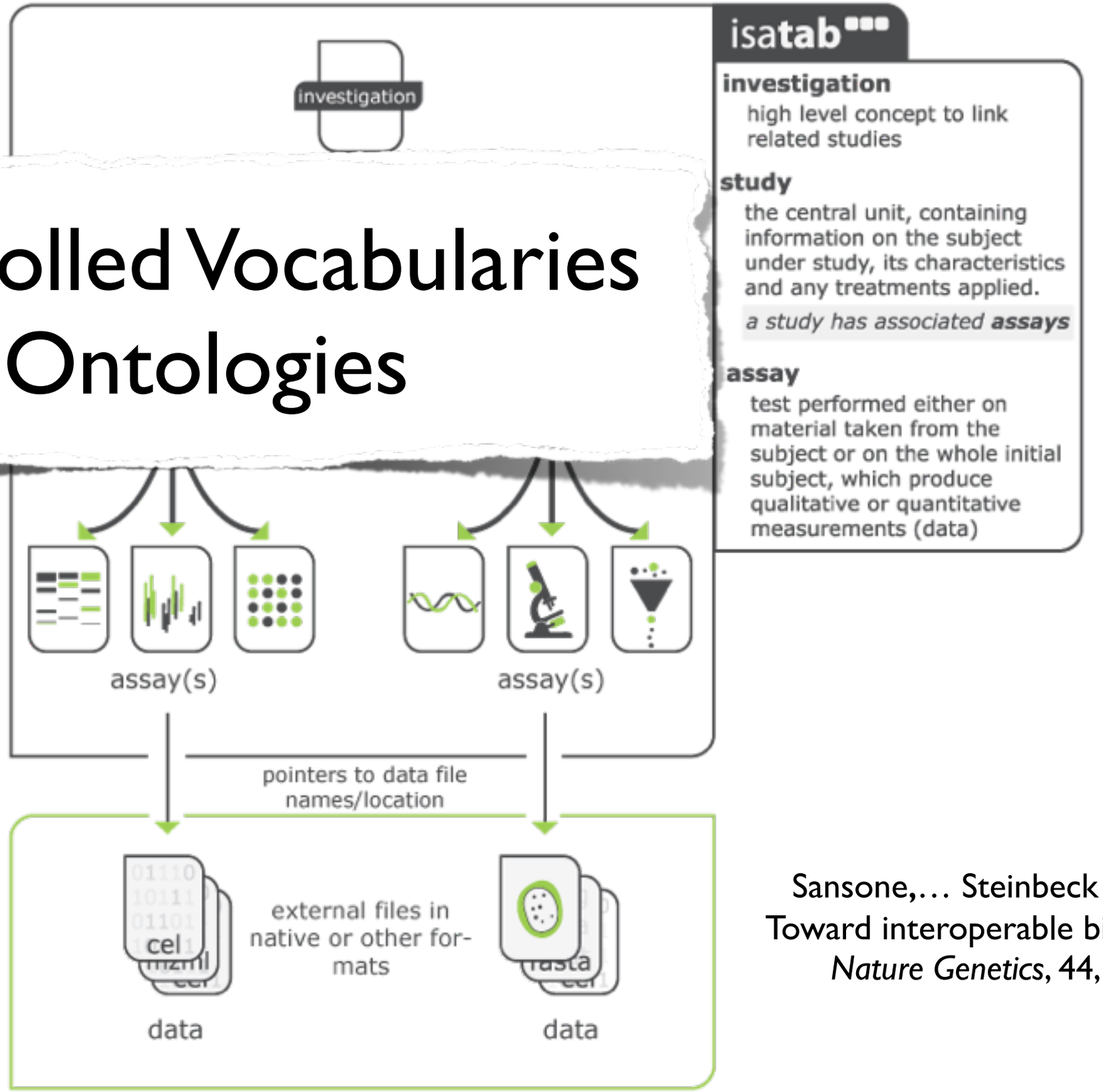
Data growth in EBI data repositories





Sansone,… Steinbeck et al. (2012)
Toward interoperable bioscience data.
Nature Genetics, 44, 121–126.

Controlled Vocabularies
Ontologies



isatab

investigation
high level concept to link related studies

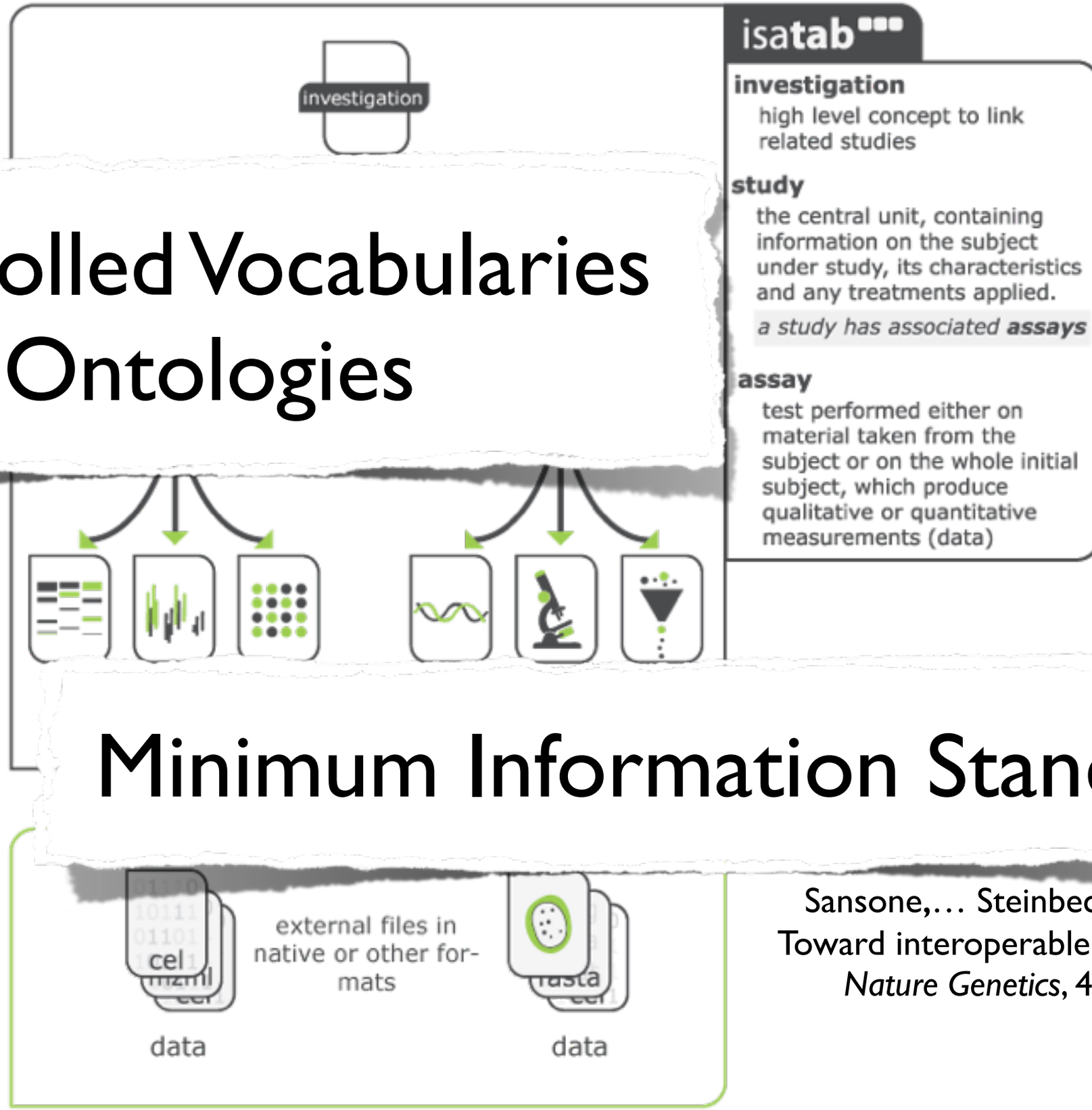
study
the central unit, containing information on the subject under study, its characteristics and any treatments applied.

a study has associated ***assays***

assay
test performed either on material taken from the subject or on the whole initial subject, which produce qualitative or quantitative measurements (data)

Sansone,… Steinbeck et al. (2012)
Toward interoperable bioscience data.
Nature Genetics, 44, 121–126.

Controlled Vocabularies
Ontologies



Minimum Information Standards

Sansone,… Steinbeck et al. (2012)
Toward interoperable bioscience data.
Nature Genetics, 44, 121–126.

Repository Entry

EMBL-EBI

Services | Research | Training | About us

MetaboLights

Search

Examples: alanine, Homo sapiens, urine, MTBLS1

Home | Browse Studies | Browse Compounds | Browse Species | Analysis | Download | Help | Give us feedback | About | Submit Study | Login

MetaboLights > Study

Submitted — In curation — In review — Public

MTBLS288: Identification of Conserved and Diverse Metabolic Shifts during Rice Grain Development (GC-MS assay)

Authors: **Shi Jianxin , Hu Chaoyang**

Share Study

Submitted: **30-Nov-2015** , Release date: **03-May-2016** , Update date: **03-May-2016**

Submitted by: Chaoyang Hu | Study status: Public

Study Description

View Metabolites Assay | Download Study files

Here we investigated the dynamic metabolic changes along the rice grain development of two japonica and two indica cultivars using non-targeted metabolomics approach, in which we successfully identified 214 metabolites. Principal component and clustering analysis revealed both cultivar and developmental stage dependent metabolic changes in rice grains. Generally, the stage specific metabolic kinetics corresponded well to the physiological status of the developing grains, and some of metabolic changes in developing grain of monocot rice are common with those of dicot Arabidopsis and tomato, while others show rice specific patterns.

See study MTBLS286 and MYTBLS287 for the LC-MS pos and neg assays respectively for this study.

Study Design Description | Protocols | Samples | Assay | Study Files | Study Validation | Pathways

Repository Entry

Metabolites

Show 10 entries

Search:

Metabolite identification	Chemical formula	Species	Smiles	InChI
5-Oxoproline (Pyroglutamate) (CHEBI:16010)	C5H7NO3	Oryza sativa	OC(=O)C1CCC(=O)N1	InChI=1S/C5H7NO3/c7-4-2-1-3(6-4)5(8)9/h3H,1-2H2,(H,6,7)(H,8,9)
Adenine (CHEBI:16708)	C5H5N5	Oryza sativa	C1=NC2=C(N1)C(=NC=N2)N	InChI=1S/C5H5N5/c6-4-3-5(9-1-7-3)10-2-8-4/h1-2H,(H3,6,7,8,9,10)
Alanine (CHEBI:16449)	C3H7NO2	Oryza sativa	CCOC(=O)N	InChI=1S/C3H7NO2/c1-2-6-3(4)5/h2H2,1H3,(H2,4,5)
Arginine (CHEBI:29016)	C6H14N4O2	Oryza sativa	C(CC(C(=O)O)N)CN=C(N)N	InChI=1S/C6H14N4O2/c7-4(5(11)12)2-1-3-10-6(8)9/h4H,1-3,7H2,(H,11,12)(H4,8,9,10)
Benzoate (CHEBI:30746)	C7H6O2	Oryza sativa	OC(=O)c1ccccc1	InChI=1S/C7H6O2/c8-7(9)6-4-2-1-3-5-6/h1-5H,(H,8,9)
Beta-Alanine (CHEBI:16958)	C3H7NO2	Oryza sativa	C(CN)C(=O)O	InChI=1S/C3H7NO2/c4-2-1-3(5)6/h1-2,4H2,(H,5,6)
Dehydroascorbate (CHEBI:17242)	C6H6O6	Oryza sativa	OCC(O)C1OC(=O)C(=O)C1=O	InChI=1S/C6H6O6/c7-1-2(8)5-3(9)4(10)6(11)12-5/h2,5,7-8H,1H2
Fructose (CHEBI:28757)	C6H12O6	Oryza sativa		
Fumarate	C4H4O4	Oryza sativa	OC(=O)C=CC(O)=O	InChI=1S/C4H4O4/c5-3(6)1-2-4(7)8/h1-2H,(H,5,6)

Reference Layer

EMBL-EBI

Services | Research | Training | About us

MetaboLights

Search

Examples: alanine, human, urine, MTBLS1

Home | Browse Studies | Browse Compounds | Browse Species | Download | Help | Give us feedback | About | Submit Study | Login

MetaboLights > Species search

Species selection page

See below for some direct links to some common model organisms and a wider list of all the organisms we have information about.

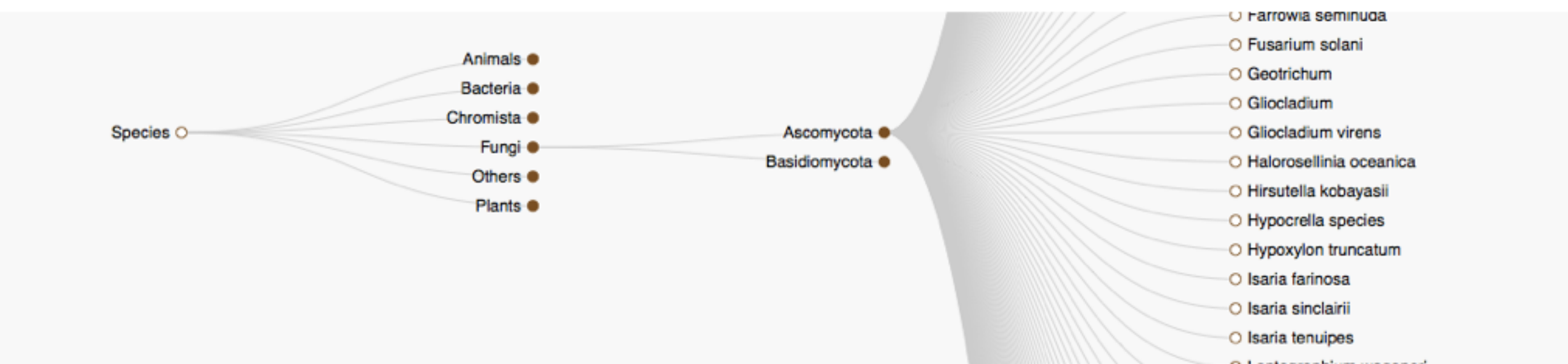
Model organisms

- Homo sapiens (Human)
- Mus musculus (Mouse)
- Arabidopsis thaliana (thale cress)
- E. coli
- Saccharomyces cerevisiae (Baker's yeast)
- Caenorhabditis elegans

Taxonomy Search

Start typing the first 3 letters of the species name

Taxonomy Browser (1593 species)



Filter your results

Type

- [x] compound

Compound features

- [ ] Species
- [ ] Pathways
- [ ] Reactions
- [x] NMR
- [ ] MS

Technology

- [ ] mass spectrometry
- [ ] NMR spectroscopy

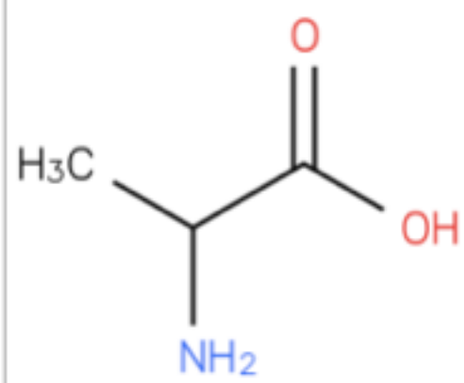
Organism

Find your Organism

- [x] Caenorhabditis elegans
- [ ] Homo sapiens
- [ ] Mus musculus
- [ ] Saccharomyces cerevisiae
- [ ] Escherichia coli
- [ ] Vitis vinifera
- [ ] reference compound
- [ ] Arabidopsis thaliana
- [ ] Daphnia magna
- [ ] Chaetomium globosum
- [ ] Streptomyces species
- [ ] Mycoplasma genitalium
- [ ] Salmonella typhimurium
- [ ] Rubia yunnanensis
- [ ] Centella asiatica
- [ ] Cordyceps sinensis

15 results, showing 1 to 10

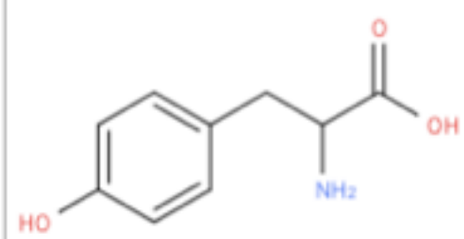
« ‹ Page 1 of 2 › »



alanine (MTBLC16449)

An alpha-amino acid that consists of propionic acid bearing an amino substituent at position 2.

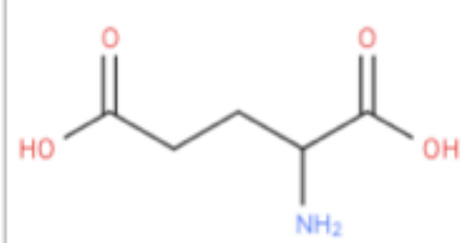
Identified in CHEBI:16449 , MTBLS20 , MTBLS8 , MTBLS3 , MTBLS1



tyrosine (MTBLC18186)

An alpha-amino acid that is phenylalanine bearing a hydroxy substituent at position 4 on the phenyl ring.

Identified in MTBLS3 , CHEBI:18186 , MTBLS55 , MTBLS1 , MTBLS8



glutamic acid (MTBLC18237)

An alpha-amino acid that is glutaric acid bearing a single amino substituent at position 2.

Identified in MTBLS8 , MTBLS3 , MTBLS29 , MTBLS36 , MTBLS1 , CHEBI:18237

EMBL-EBI

EMBL-EBI

Services | Research | Training | About us

MetaboLights

Search

Examples: alanine, Homo sapiens, urine, MTBLS1

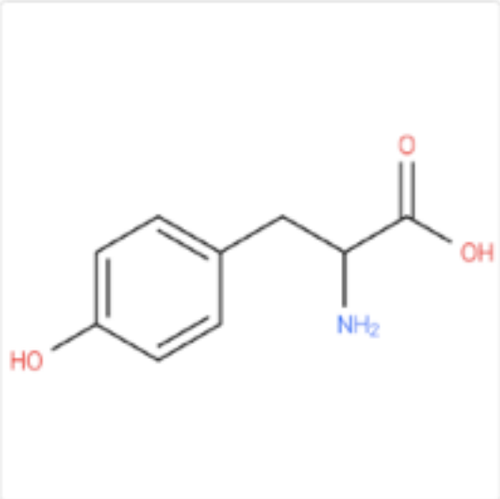
Home | Browse Studies | Browse Compounds | Browse Species | Analysis | Download | Help | Give us feedback | About | Submit Study | Login

MetaboLights > Compound page

Tyrosine

MTBLC18186

2D | 3D



tyrosine - (CHEBI:18186)

BETA

Upload Reference Spectra

Chemistry | Biology | NMR spectra | MS spectra | Literature

Definition

An alpha-amino acid that is phenylalanine bearing a hydroxy substituent at position 4 on the phenyl ring.

Chemical Properties

Synonyms

External links

Chemical Properties

Property	Value
InChIKey	OUYCCCASQSFEME-UHFFFAOYSA-N
InChI	InChI=1S/C9H11NO3/c10-8(9(12)13)5-6-1-3-7(11)4-2-6/h1-4,8,11H,5,10H2,(H,12,13)
Formula	C9H11NO3
Molecular Weight	181.18889
Exact Mass	181.07389

EMBL-EBI

MetaboLights

Search

Examples: alanine, Homo sapiens, urine, MTBLS1

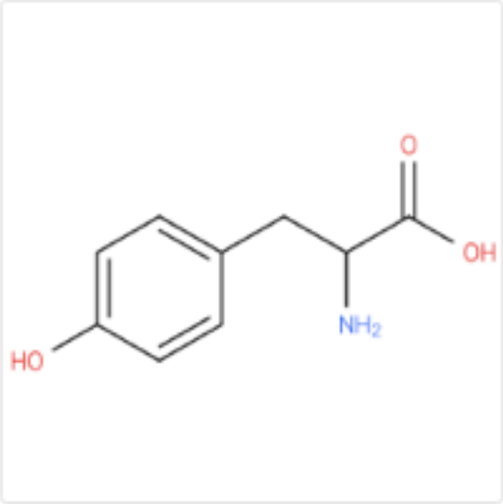
Home | Browse Studies | Browse Compounds | Browse Species | Analysis | Download | Help | Give us feedback | About | Submit Study | Login

MetaboLights > Compound page

Tyrosine

MTBLC18186

2D 3D



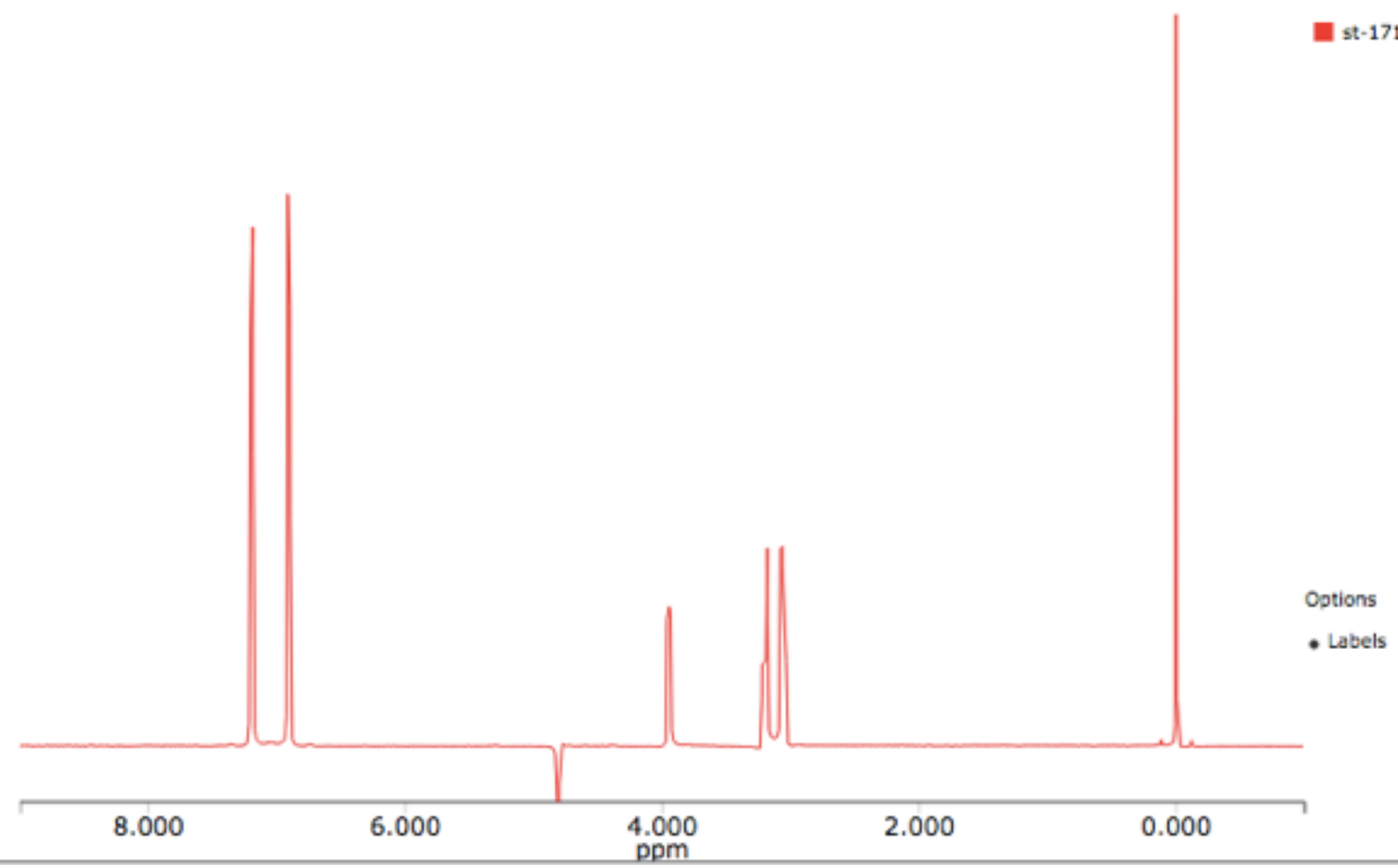
tyrosine - (CHEBI:18186)

BETA

Upload Reference Spectra

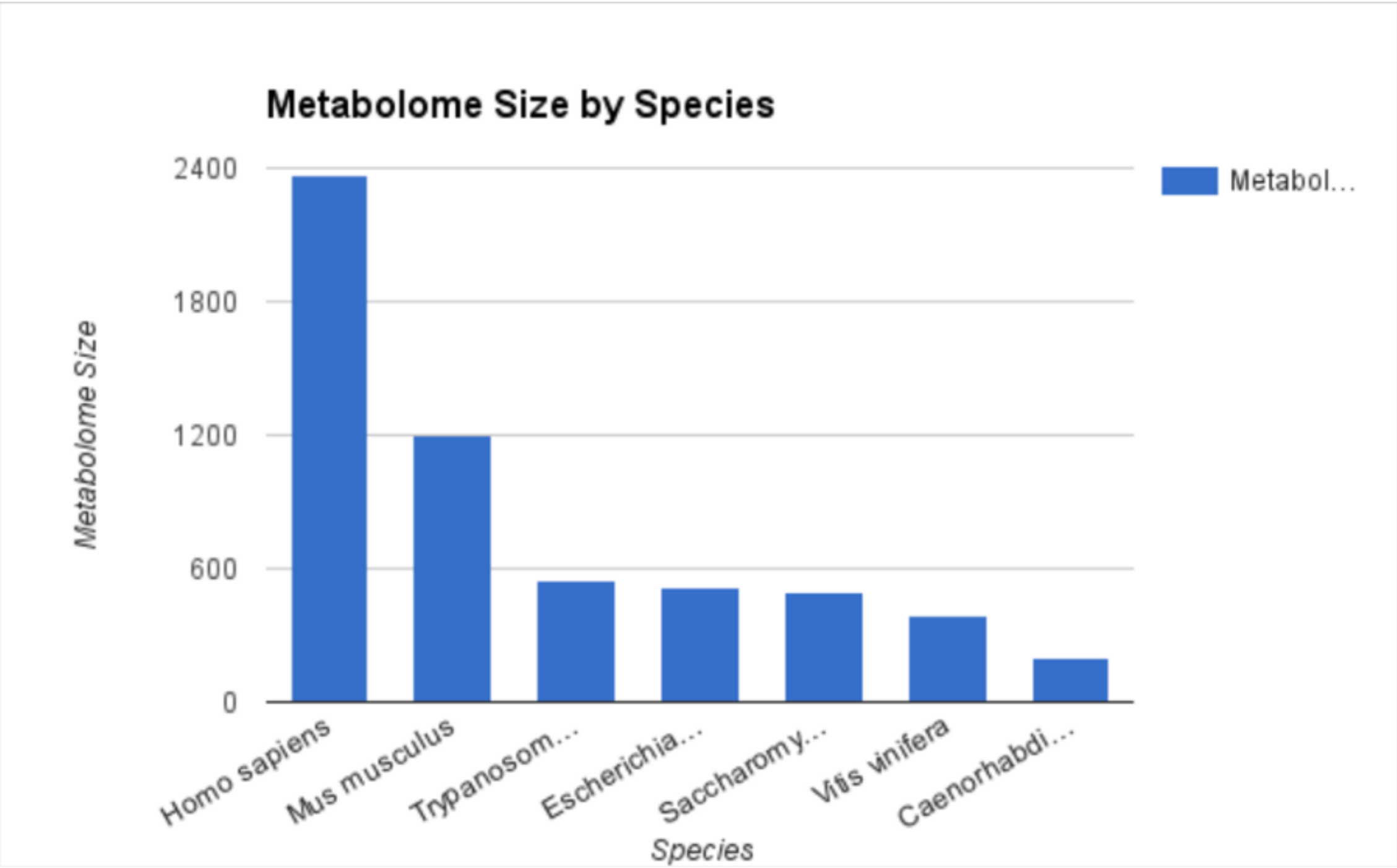
Chemistry | Biology | NMR spectra | MS spectra | Literature

BML (ID:3316) - 1H 500 MHz, WS(NOESY Presat), pH:7.0, PD=10.0s, EA=30 degree
BML (ID:3320) - 1H 500 MHz, WS(Excitation Sculpting), pH:7.0, PD=0.75s, EA=90 degree
BML (ID:3327) - 1H 500 MHz, WS(Excitation Sculpting), pH:7.4, PD=3.0s, EA=60 degree
BML (ID:3313) - 1H 500 MHz, WS(Excitation Sculpting), pH:7.0, PD=10.0s, EA=30 degree

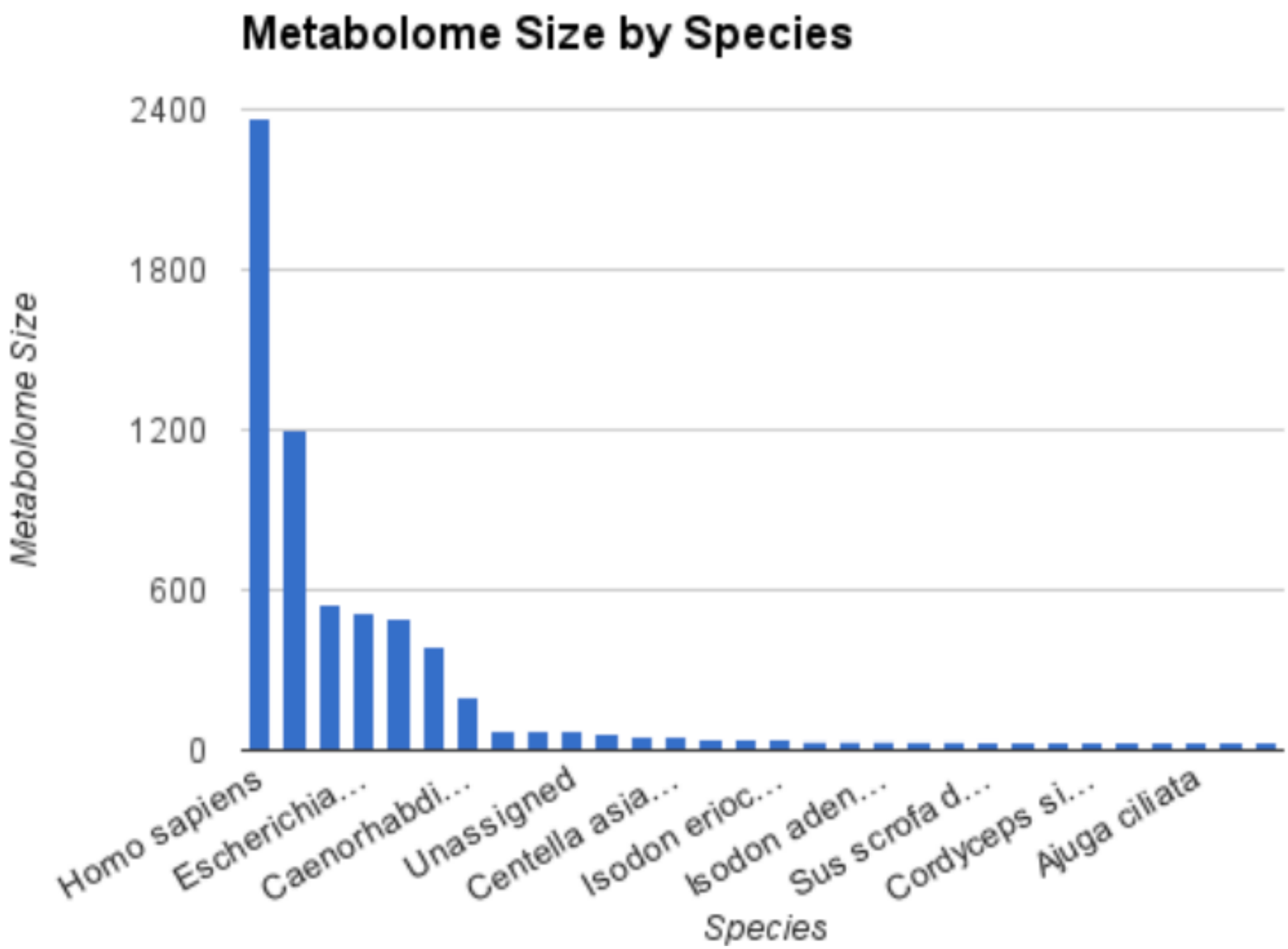


EMBL-EBI

7 most annotated metabolomes in MetaboLights

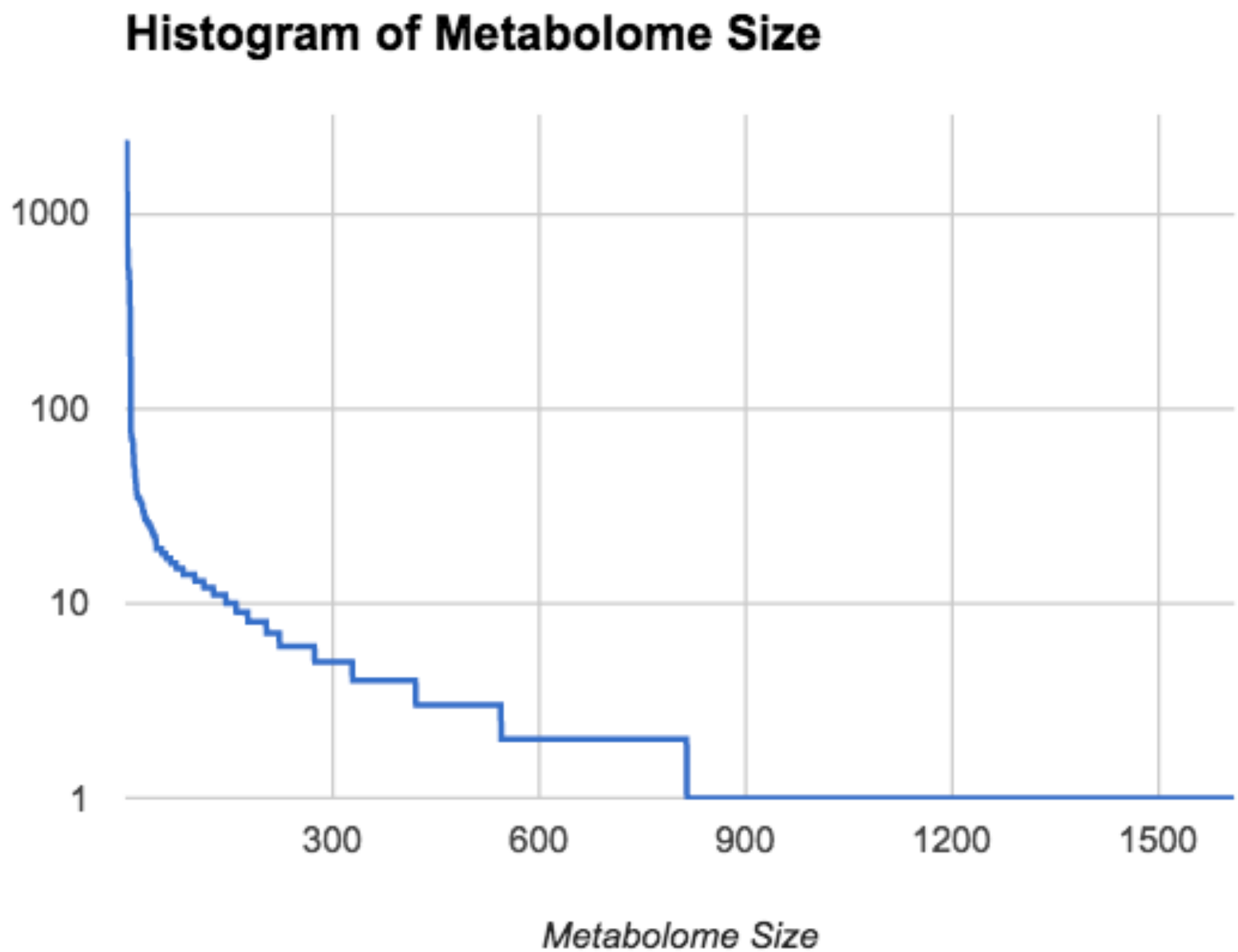


30 most annotated metabolomes in MetaboLights



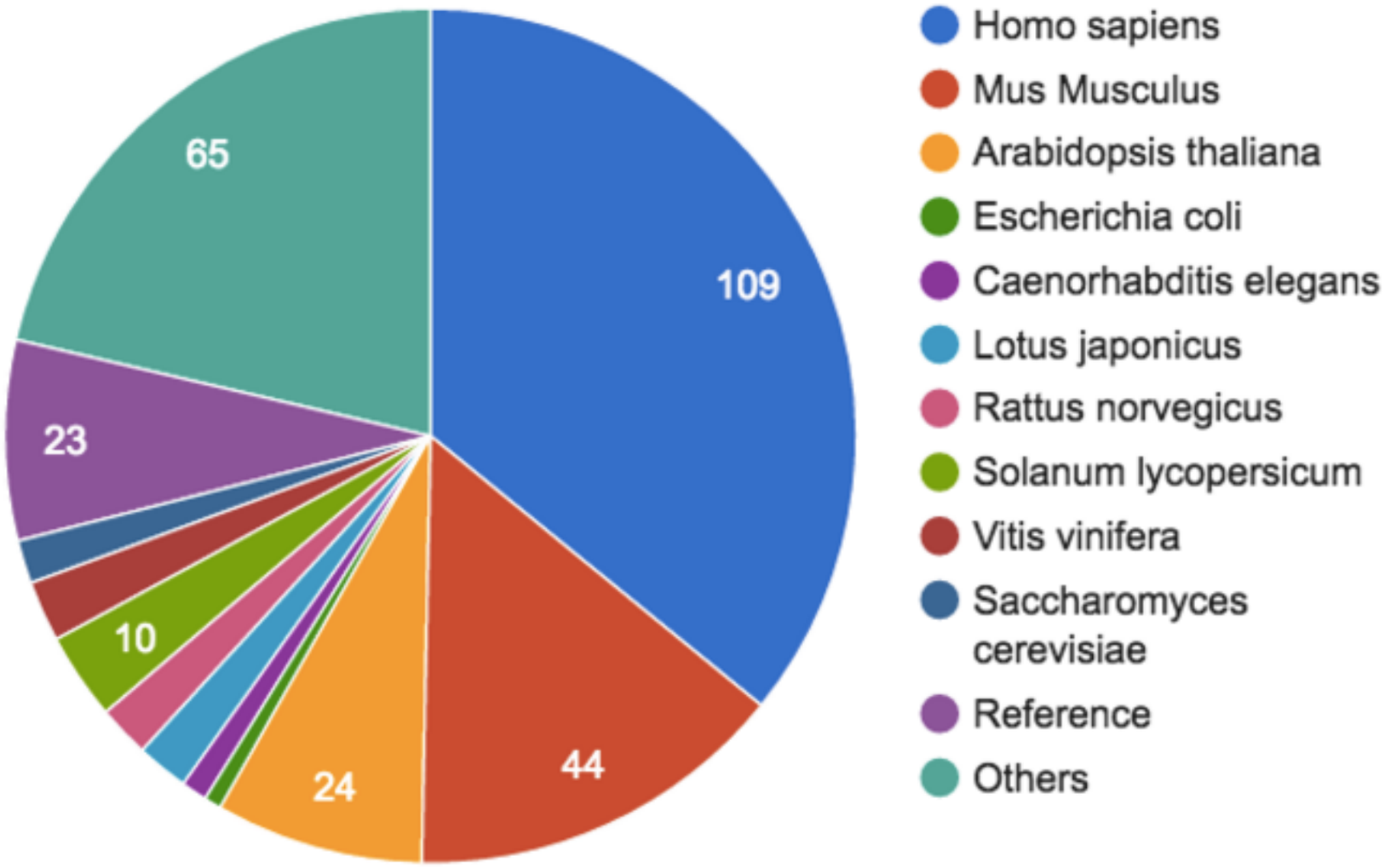
EMBL-EBI

1600 metabolome sizes in MetaboLights on a log scale



EMBL-EBI

Number of Studies in MetaboLights per Species



EMBL-EBI

SCIENTIFIC DATA | DATA DESCRIPTOR OPEN

Metabolic differences in ripening of *Solanum lycopersicum* 'Ailsa Craig' and three monogenic mutants

Stephan Beisken, Mark Earll, Charles Baxter, David Portwood, Zsuzsanna Ament, Aniko Kende, Charlie Hodgman, Graham Seymour, Rebecca Smith, Paul Fraser, Mark Seymour, Reza M. Salek & Christoph Steinbeck

Affiliations | **Contributions** | **Corresponding authors**

Scientific Data **1**, Article number: 140029 | doi:10.1038/sdata.2014.29

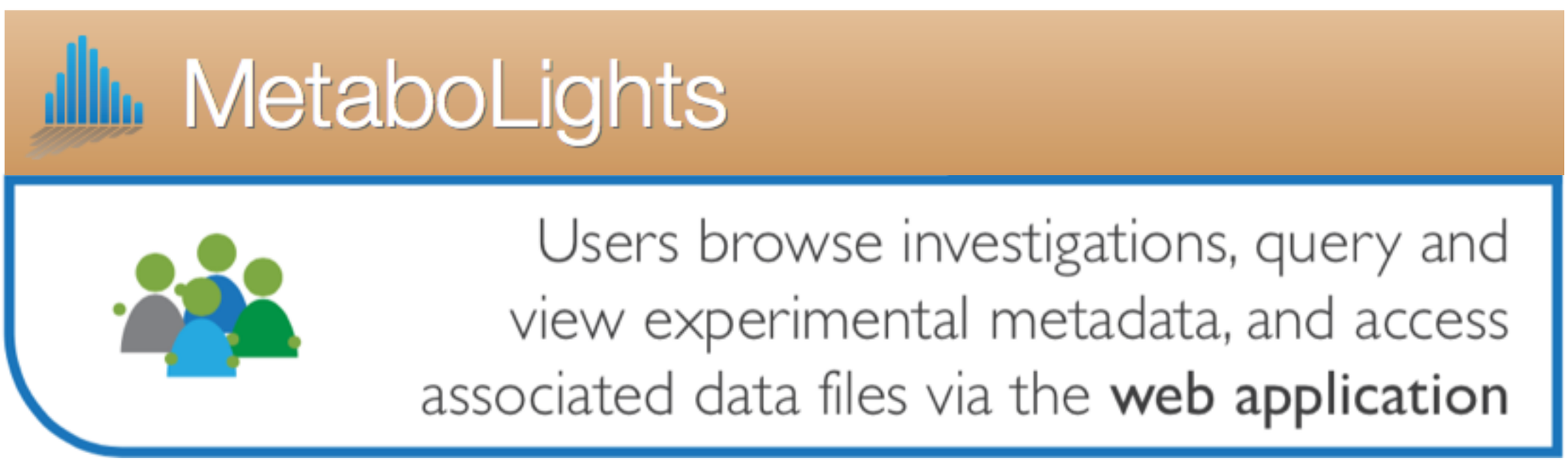
Received 10 April 2014 | Accepted 06 August 2014 | Published online 16 September 2014

Abstract

Application of mass spectrometry enables the detection of metabolic differences between groups of related organisms. Differences in the metabolic fingerprints of wild-type *Solanum lycopersicum* and three monogenic mutants, *ripening inhibitor (rin)*, *non-ripening (nor)* and *Colourless non-ripening*

Biocrates XML with raw data is annotated and uploaded

Share private prepublication studies with reviewers and other trusted parties.



Multi-Omics Data Handling

EMBL-EBI

BioSamples

BioSamples - database of sample descriptions

The BioSamples database aggregates sample information for reference samples (e.g. Coriell Cell lines) and samples for which data exist in one of the EBI's assay databases such as ArrayExpress, the European Nucleotide Archive or PRIDE. It provides links to assays an specific samples, and accepts direct submissions of sample information.

Info

- Help pages about how to search BioSamples, how to submit data, and FAQ.
- Programmatic access to query and download data using web services.
- Latest news about BioSamples database.
- The BioSamples database now offers access to RDF, and a supporting SPARQL endpoint as part of the EBI's RDF platform.
- Contact us by emailing biosamples@ebi.ac.uk

External links

- Experimental Factor Ontology
- Human Induced Pluripotent Stem Cells Initiative (HipSci)
- 1000 Genomes
- Encyclopedia of DNA Elements (ENCODE)
- Catalogue Of Somatic Mutations In Cancer (COSMIC)

Data Content

- 4,733,993 Samples
- 57,274 Groups

HipSci

EMBL-EBI

Services | Research | Training | About us

BioStudies

Search

Examples: hyperplasia, PMC516016

Home | Browse | Submit a study | Help | About BioStudies | Feedback | Login

BioStudies – database of biological studies

The BioStudies database holds descriptions of biological studies, links to data from these studies in other databases at EMBL-EBI or outside, as well as data that do not fit in the structured archives at EMBL-EBI. The database can accept a wide range of types of studies described via a simple format. It also enables manuscript authors to submit supplementary information and link to it from the publication.

Browse

601975 submissions

4 projects

601971 studies

Latest

- Drosophila lilliputian is required for proneural gene expression in retinal development [S-EPMC4946344]
- Early pediatric atopic dermatitis shows only a cutaneous lymphocyte antigen (CLA)(+) TH2/TH1 cell imbalance, whereas adults acquire CLA(+) TH22/TC22 cell subsets [S-EPMC4946641]
- A population-based study of hospital care costs during 5 years after transient ischemic attack and stroke [S-EPMC4946629]
- Population-based study of disability and institutionalization after transient ischemic attack and stroke: 10-year results of the Oxford Vascular Study [S-EPMC4946627]
- Proteome wide purification and identification of O-GlcNAc-modified proteins using click chemistry and mass

A Case for Deep Metabolome Annotation

Kingdom	Latin Name	Common Name
Bacteria	*Escherichia coli*	-
Fungi	*Saccharomyces cerevisiae*	yeast
Animal (invertebrate)	*Caenorhabditis elegans*	nematode
	Daphnia magna	water flea
	Drosophila melanogaster *	fruit fly
Animal (vertebrate)	*Danio rerio*	zebrafish
	Mus musculus	mouse
Plant	*Arabidopsis thaliana* **	thale cress
	Medicago truncatula	barrel medic, model legume
	Oryza sativa	rice
	Solanum lycopersicum	tomato

* International Drosophila Metabolomics Curation Consortium [30]; ** Metabolomics subcommittee (chaired by Kazuki Saito) within the Multinational Arabidopsis Steering Committee [31].

A Case for Deep Metabolome Annotation

Help building species metabolomes

- Submit your metabolomics study to MetaboLights
- Submit data publications (e.g. to NATURE Scientific Data)
- Be highly cited :)

Solanum lycopersicum tomato

* International Drosophila Metabolomics Curation Consortium [30]; ** Metabolomics subcommittee (chaired by Kazuki Saito) within the Multinational Arabidopsis Steering Committee [31].

Slides at

https://www.slideshare.net/csteinbeck

Funding

BBSRC bioscience for the future

EMBL-EBI

European Commission

MRC Medical Research Council

Thanks for your attention

EMBL-EBI

Thank you!

Metabolights-help@ebi.ac.uk

EMBL-EBI