

BRENDA kinetic data analysis

Ivan Domenzain

June 2022

Contents

Summary	1
1. Loading data, scripts and packages	1
1.1 Loading packages	1
1.2 Loading R scripts for this study	1
1.3 Loading enzyme data from BRENDA database	1
2. Preprocess data	2
2.1 Extend dataset with specific activity values	2
2.2 Get enzymes information (pathways, genes) from KEGG ftp	2
3. Data analysis	5
Data composition	5
3.1.1 Data composition per enzyme classes	5
3.1.2 Data composition per organism	7
3.1.3 Data composition per metabolic groups	9
3.2 Data dispersion	11
3.3 k_{cat} distributions	14
3.3.1 k_{cat} distributions per enzymes family	14
3.3.2 k_{cat} distributions per metabolic context	17
3.3.3 k_{cat} distributions per organism Kingdoms	18
3.3.4 k_{cat} distributions per metabolic pathways groups per organism Kingdoms	21

Summary

This is a global statistical study of turnover numbers (k_{cat} parameters) for enzymatic reactions reported in the BRENDA database.

1. Loading data, scripts and packages

1.1 Loading packages

```
options(repos = c(
  yihui = 'https://yihui.r-universe.dev',
  CRAN = 'https://cloud.r-project.org'
))

install.packages('xfun')
library(tidyr)
library(dplyr)
library(plyr)
library(knitr)
library(magick)
library(gridExtra)
library(VennDiagram)
library(kableExtra)
library(ggplot2)
library(ggrepel)
library(viridis)
library(tinytex)
```

1.2 Loading R scripts for this study

```
read_chunk('loadData.R')
read_chunk('preprocessData.R')
read_chunk('analyseData.R')
read_chunk('plotData.R')
```

1.3 Loading enzyme data from BRENDA database

All available k_{cat} numbers, specific activities and molecular weights of enzymes have been retrieved from the BRENDA database, using the provided SOAP methods for python in the BRENDA website. k_{cat} numbers entries correspond to measurements that have been performed for a specific EC number, substrate and organism of origin; the same applies to specific activities, however such parameters are not reported in a substrate specific way; Molecular weights are reported for specific pairs of EC number - organism of origin.

Parameters entries for mutant and recombinant enzymes have been discarded, more details of the retrieval steps are described in the python scripts stored in the directory geckomat/brenda_parser of the GECKO toolbox.

- Loading k_{cat} values [1/s]

- Loading Specific Activity values [umol/min/mg]
- Loading MWeights values from BRENDa [mmol/mg]
- Loading enzyme data from KEGG

2. Preprocess data

2.1 Extend dataset with specific activity values

Get extra k_{cat} values from Specific activities and molecular weights data using the relationship: $k_{cat} = SA * M_{weight}$

The original k_{cat} dataset consisted of:

```
## [1] "30162 entries for 3019 different EC numbers"
```

The extended dataset consists of:

```
## [1] "38280 entries for 4130 different EC numbers"
```

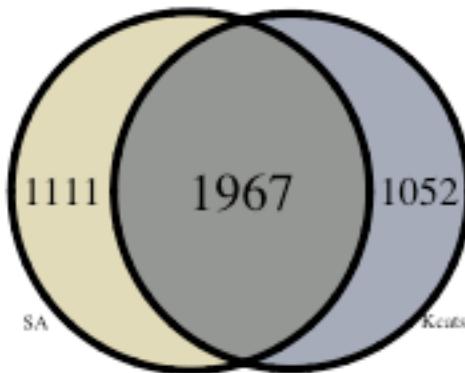


Figure S1.1: Source of unique EC numbers in the dataset

From here onwards, the term dataset will refer to the extended dataset obtained by merging k_{cat} values with those obtained from the specific activities dataset.

2.2 Get enzymes information (pathways, genes) from KEGG ftp

In this study, kinetic parameters distributions are going to be analysed for different enzyme groups, based on enzyme classes, host organism's taxonomy and metabolic context. For the latter, the KEGG pathways classification is used for identifying those enzyme entries in the dataset that have been annotated as part of the following groups:

- Carbohydrates and Energy Metabolism (CEM)

- Amino acids and Lipids Metabolism (ALM)
- Intermediate and Secondary Metabolism (ISM)
- Others

The KEGG database contains pathways and genes information for:

```
## [1] "4921 different EC numbers"
```

A venn diagram for the EC number entries available in BRENDA and KEGG databases is shown below:

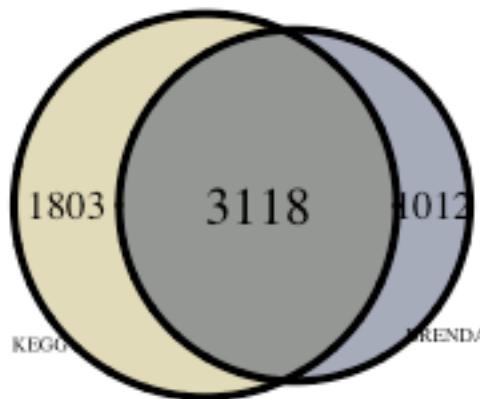


Figure S1.2: Unique EC numbers in BRENDA and KEGG.

The KEGG database provides information that links each EC number entry to all of the pathways in which it might participate. This allows to add a metabolic subgroup to each of the entries in the extended dataset used in this study. Therefore, 3,118 unique EC numbers in the dataset can be classified by metabolic context groups (metGroups).

Note: The EC numbers not present in KEGG are annotated as part of the group ‘Others’ for this study.

The classified pathways are:

pathways	group
ec00010 Glycolysis / Gluconeogenesis	CEM
ec00020 Citrate cycle (TCA cycle)	CEM
ec00030 Pentose phosphate pathway	CEM
ec00040 Pentose and glucuronate interconversions	CEM
ec00051 Fructose and mannose metabolism	CEM
ec00052 Galactose metabolism	CEM
ec00053 Ascorbate and aldarate metabolism	CEM
ec00500 Starch and sucrose metabolism	CEM
ec00520 Amino sugar and nucleotide sugar metabolism	CEM
ec00620 Pyruvate metabolism	CEM
ec00630 Glyoxylate and dicarboxylate metabolism	CEM
ec00640 Propanoate metabolism	CEM
ec00650 Butanoate metabolism	CEM
ec00660 C5-Branched dibasic acid metabolism	CEM

pathways	group
ec00562 Inositol phosphate metabolism	CEM
ec00190 Oxidative phosphorylation	CEM
ec00195 Photosynthesis	CEM
ec00196 Photosynthesis - antenna proteins	CEM
ec00710 Carbon fixation in photosynthetic organisms	CEM
ec00720 Carbon fixation pathways in prokaryotes	CEM
ec00680 Methane metabolism	CEM
ec00910 Nitrogen metabolism	CEM
ec00920 Sulfur metabolism	CEM
ec00061 Fatty acid biosynthesis	ALM
ec00062 Fatty acid elongation	ALM
ec00071 Fatty acid degradation	ALM
ec00072 Synthesis and degradation of ketone bodies	ALM
ec00073 Cutin, suberine and wax biosynthesis	ALM
ec00100 Steroid biosynthesis	ALM
ec00120 Primary bile acid biosynthesis	ALM
ec00121 Secondary bile acid biosynthesis	ALM
ec00140 Steroid hormone biosynthesis	ALM
ec00561 Glycerolipid metabolism	ALM
ec00564 Glycerophospholipid metabolism	ALM
ec00565 Ether lipid metabolism	ALM
ec00600 Sphingolipid metabolism	ALM
ec00590 Arachidonic acid metabolism	ALM
ec00591 Linoleic acid metabolism	ALM
ec00592 alpha-Linolenic acid metabolism	ALM
ec01040 Biosynthesis of unsaturated fatty acids	ALM
ec00230 Purine metabolism	ALM
ec00240 Pyrimidine metabolism	ALM
ec00250 Alanine, aspartate and glutamate metabolism	ALM
ec00260 Glycine, serine and threonine metabolism	ALM
ec00270 Cysteine and methionine metabolism	ALM
ec00280 Valine, leucine and isoleucine degradation	ALM
ec00290 Valine, leucine and isoleucine biosynthesis	ALM
ec00300 Lysine biosynthesis	ALM
ec00310 Lysine degradation	ALM
ec00220 Arginine biosynthesis	ALM
ec00330 Arginine and proline metabolism	ALM
ec00340 Histidine metabolism	ALM
ec00350 Tyrosine metabolism	ALM
ec00360 Phenylalanine metabolism	ALM
ec00380 Tryptophan metabolism	ALM
ec00400 Phenylalanine, tyrosine and tryptophan biosynthesis	ALM
ec00410 beta-Alanine metabolism	ISM
ec00430 Taurine and hypotaurine metabolism	ISM
ec00440 Phosphonate and phosphinate metabolism	ISM
ec00450 Selenocompound metabolism	ISM
ec00460 Cyanoamino acid metabolism	ISM
ec00471 D-Glutamine and D-glutamate metabolism	ISM
ec00472 D-Arginine and D-ornithine metabolism	ISM
ec00473 D-Alanine metabolism	ISM
ec00480 Glutathione metabolism	ISM
ec00510 N-Glycan biosynthesis	ISM

pathways	group
ec00513 Various types of N-glycan biosynthesis	ISM
ec00512 Mucin type O-glycan biosynthesis	ISM
ec00515 Mannose type O-glycan biosynthesis	ISM
ec00514 Other types of O-glycan biosynthesis	ISM
ec00532 Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	ISM
ec00534 Glycosaminoglycan biosynthesis - heparan sulfate / heparin	ISM
ec00533 Glycosaminoglycan biosynthesis - keratan sulfate	ISM
ec00531 Glycosaminoglycan degradation	ISM
ec00563 Glycosylphosphatidylinositol (GPI)-anchor biosynthesis	ISM
ec00601 Glycosphingolipid biosynthesis - lacto and neolacto series	ISM
ec00603 Glycosphingolipid biosynthesis - globo and isogloblo series	ISM
ec00604 Glycosphingolipid biosynthesis - ganglio series	ISM
ec00540 Lipopolysaccharide biosynthesis	ISM
ec00550 Peptidoglycan biosynthesis	ISM
ec00511 Other glycan degradation	ISM
ec00730 Thiamine metabolism	ISM
ec00740 Riboflavin metabolism	ISM
ec00750 Vitamin B6 metabolism	ISM
ec00760 Nicotinate and nicotinamide metabolism	ISM
ec00770 Pantothenate and CoA biosynthesis	ISM
ec00780 Biotin metabolism	ISM
ec00785 Lipoic acid metabolism	ISM
ec00790 Folate biosynthesis	ISM
ec00670 One carbon pool by folate	ISM
ec00830 Retinol metabolism	ISM
ec00860 Porphyrin and chlorophyll metabolism	ISM
ec00130 Ubiquinone and other terpenoid-quinone biosynthesis	ISM

3. Data analysis

Data composition

3.1.1 Data composition per enzyme classes

Not all of the enzyme classes are of the same interest for the scientific community. Counting the number of data entries available for every unique EC number allows the identification of the top studied enzyme classes.

The top ten represented EC numbers in the dataset are:

Table S1.2: Top10 enzyme classes by number of reported Kcat values

ECnumbers	Entries	% of total entries
EC3.5.2.6	581	1.52 %
EC3.1.1.1	335	0.88 %
EC1.1.1.1	316	0.83 %
EC3.2.1.21	300	0.78 %
EC3.4.21.4	270	0.71 %
EC2.3.2.5	259	0.68 %
EC3.4.21.83	208	0.54 %
EC1.1.99.18	192	0.5 %
EC1.10.3.2	168	0.44 %
EC3.4.21.35	165	0.43 %

Actually, most of the enzyme classes are poorly represented in the dataset, with at least 50% percent of the total number of unique EC numbers having less than 4 reported k_{cat} values (for different organisms and substrates).

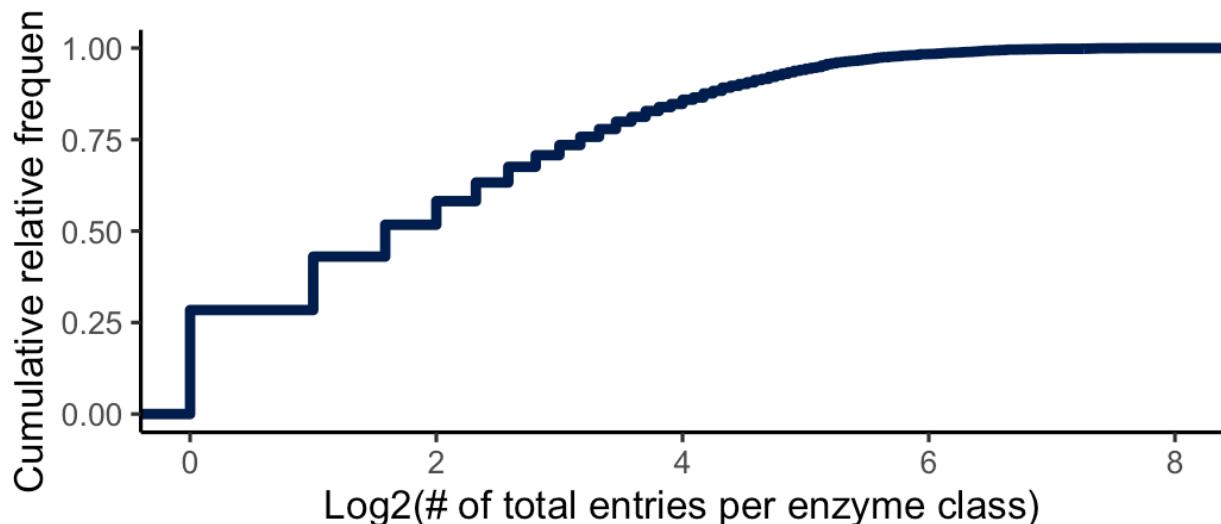


Figure S1.3: Number of entries per enzyme class.

The dataset also presents heterogeneity when it comes to the number of reported parameters for the different families of enzyme classes (top-level EC numbers), being hydrolases (EC3.x.x.x) and oxidoreductases (EC1.x.x.x) the most represented families, comprising almost 70% of the total number of entries.

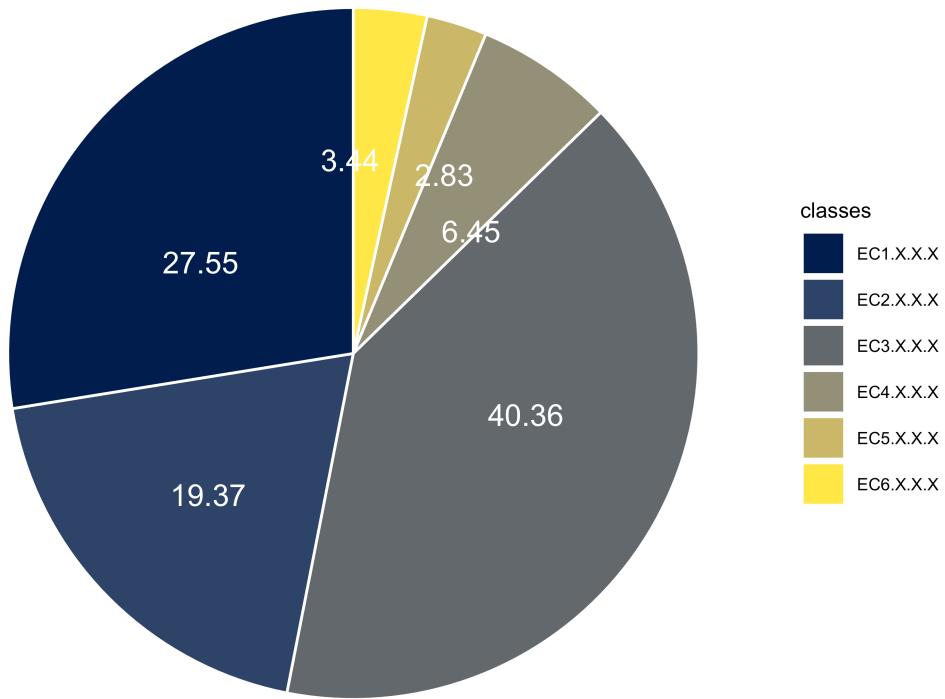


Figure S1.4: Dataset composition by enzyme classes

Table S1.3: Average number of entries per enzyme class for the different enzyme families.

Top-Level_ECnumber	Enzymes_family	Average_entries_per_ECnumber
EC1.X.X.X	Oxidoreductases	8
EC2.X.X.X	Transferases	6
EC3.X.X.X	Hydrolases	14
EC4.X.X.X	Lyases	5
EC5.X.X.X	Isomerases	6
EC6.X.X.X	Ligases	8

3.1.2 Data composition per organism

The different organisms have been grouped into five different “Kingdoms” according to the KEGG taxonomy information:

- Eukaryotes//Animals
- Eukaryotes//Plants
- Eukaryotes//Protists
- Eukaryotes//Fungi
- Prokaryotes//Bacteria
- Prokaryotes//Archaea

Number of reported k_{cat} entries per organism:

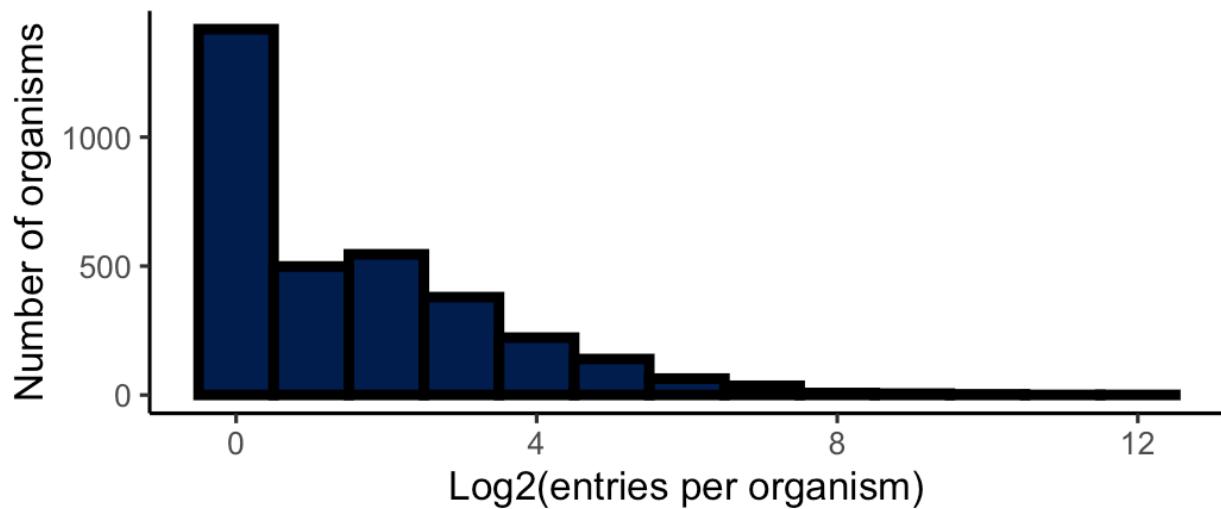


Figure S1.5: Number of entries per organism.

Table S1.4: Top10 organisms by number of reported Kcat values

Organism	Kingdom	Entries	% of total entries
Homo sapiens	Animals	4212	11 %
Escherichia coli	Bacteria	2131	5.57 %
Rattus norvegicus	Animals	1384	3.62 %
Saccharomyces cerevisiae	Fungi	734	1.92 %
Sus scrofa	Animals	733	1.91 %
Bos taurus	Animals	710	1.85 %
Mus musculus	Animals	654	1.71 %
Arabidopsis thaliana	Plants	495	1.29 %
Mycobacterium tuberculosis	Bacteria	475	1.24 %
Pseudomonas putida	Bacteria	377	0.98 %

Representation of the different organism Kingdoms in the analysed dataset

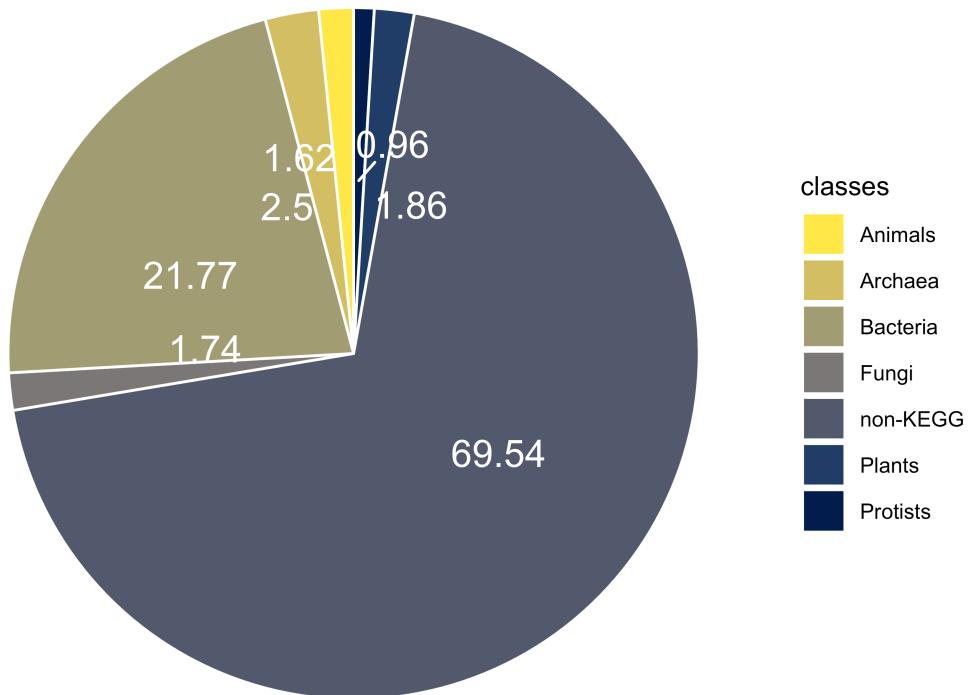


Figure S1.6: Dataset composition by organisms Kingdom

As it has been shown, the studied dataset is biased towards certain well-studied organisms, letting most of the included organisms with a very poor representation in the k_{cat} dataset.

3.1.3 Data composition per metabolic groups

Representation of the different metabolic groups in the analysed dataset.

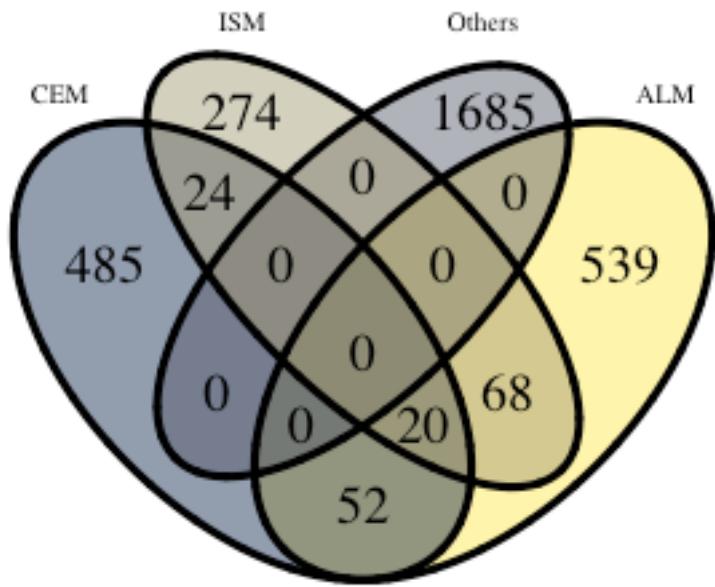
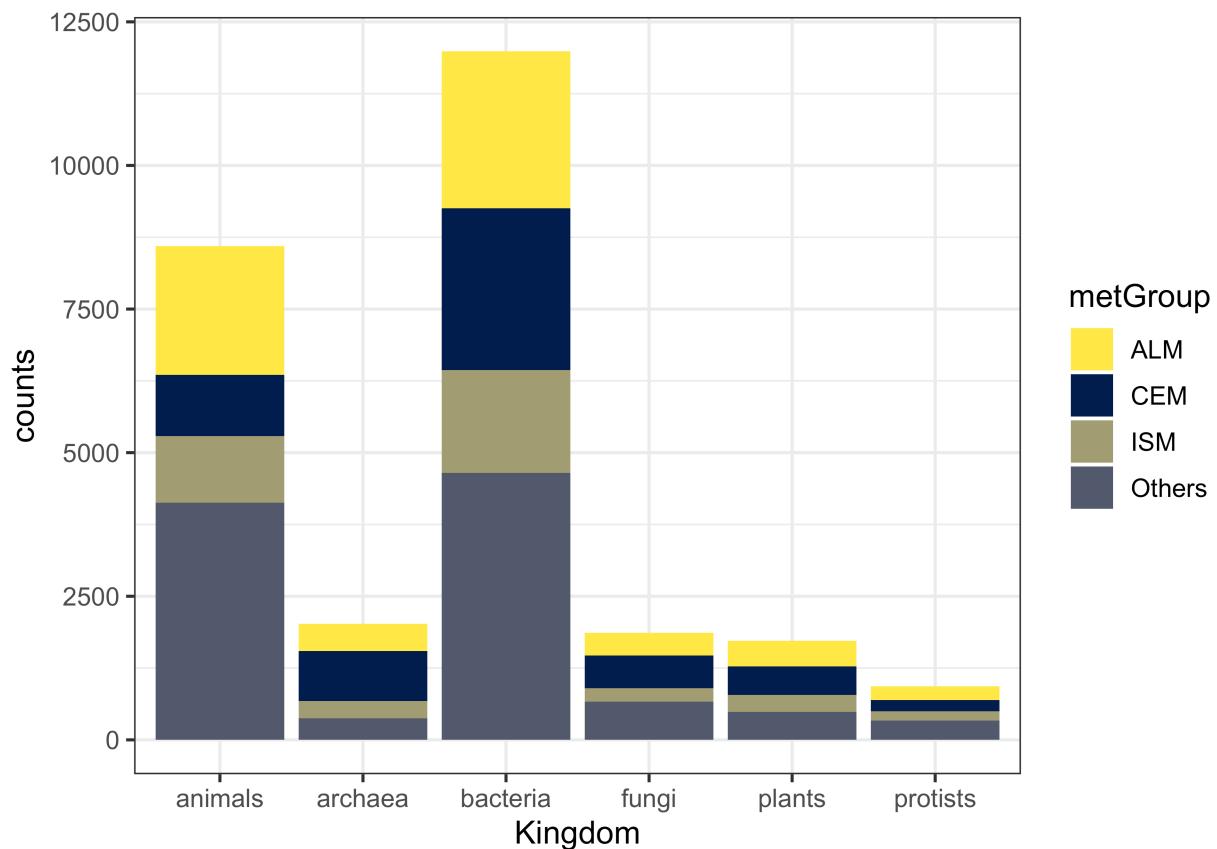


Figure S1.7: Unique EC numbers in the dataset by metabolic subgroup

Unique EC numbers per metabolic groups for the different organism Kingdoms.



The dataset is not just heterogeneous regarding its composition by organism kingdoms, but also within each of these kingdoms, different metabolic groups of enzymes have been studied in a diverse way, showing a majority of reported values for CEM and ALM enzyme classes for most of the cases.

3.2 Data dispersion

It is said that evolution has shaped kinetic parameters in a very specific way, differentiating catalytic activities by metabolic context, phylogenetic origin and substrate specificity. If such specialization exists for a big dataset as this one, then a wide spanning of reported values can be expected.

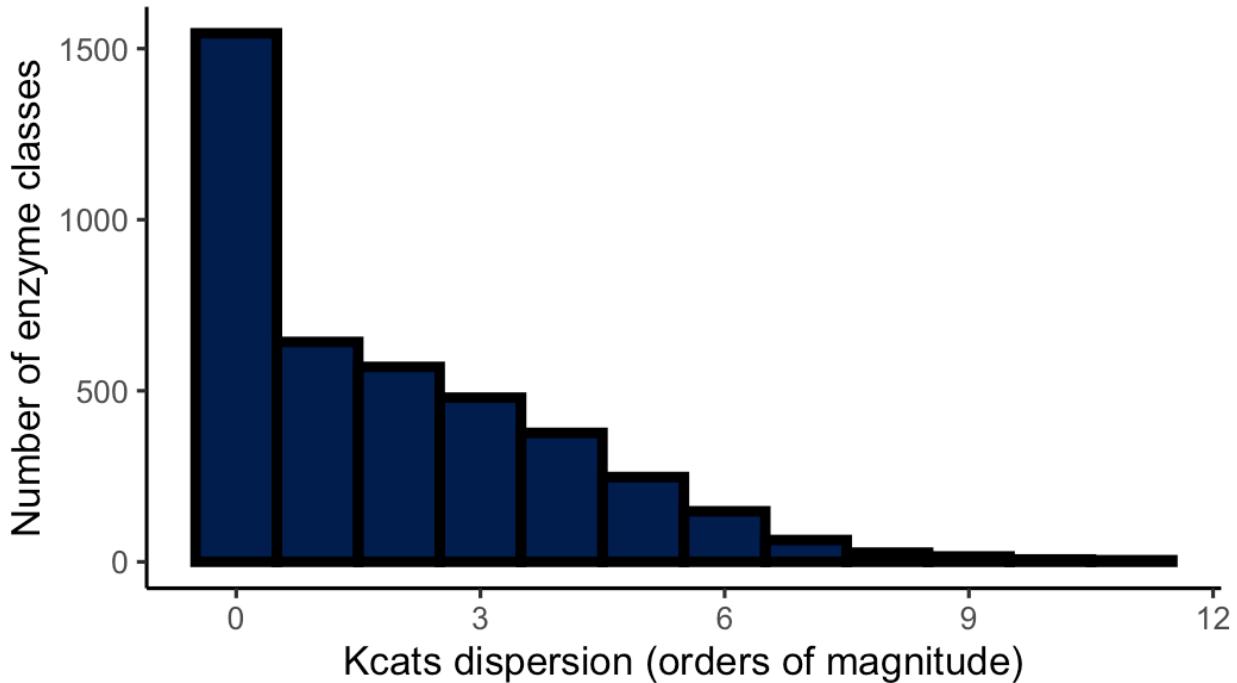


Figure S1.8: Kcats dispersion per EC number.

The previous histogram shows how most of the studied enzyme classes present a very narrow spanning of their reported k_{cat} values, however few enzyme classes (EC numbers) with a surprisingly wide spanning of catalytic activities, (**11 orders of magnitude!**) can also be found in the dataset.

However, it is also important to analyze how large is the variance of a given distribution related to its characteristic values. For this, a normalized dispersion is metric, which here will be called as “spreading” and is defined as:

$$spreading = \log_{10}(Median)/[\log_{10}(Max) - \log_{10}(Min)].$$

This metric allows to explore how wide or narrow a distribution is compared to its median value. A value lower than 1 means that the distribution spanning is higher than its median value (in terms of orders of magnitude).

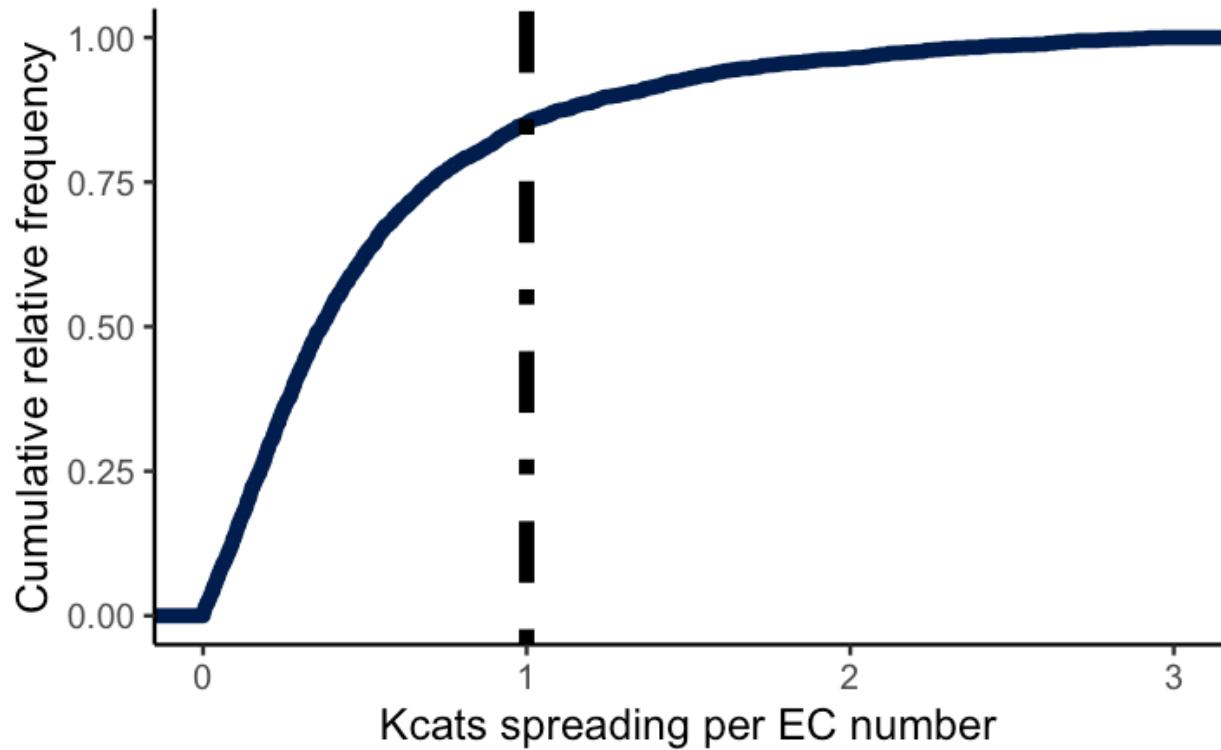


Figure S1.9: K_{cats} spreading per EC number.

This cumulative distribution shows that around 20% of the enzyme classes show very wide distributions of k_{cat} values in which a median or mean value cannot be considered as representative of the distribution due to its large spanning.

3.3 k_{cat} distributions

3.3.1 k_{cat} distributions per enzymes family

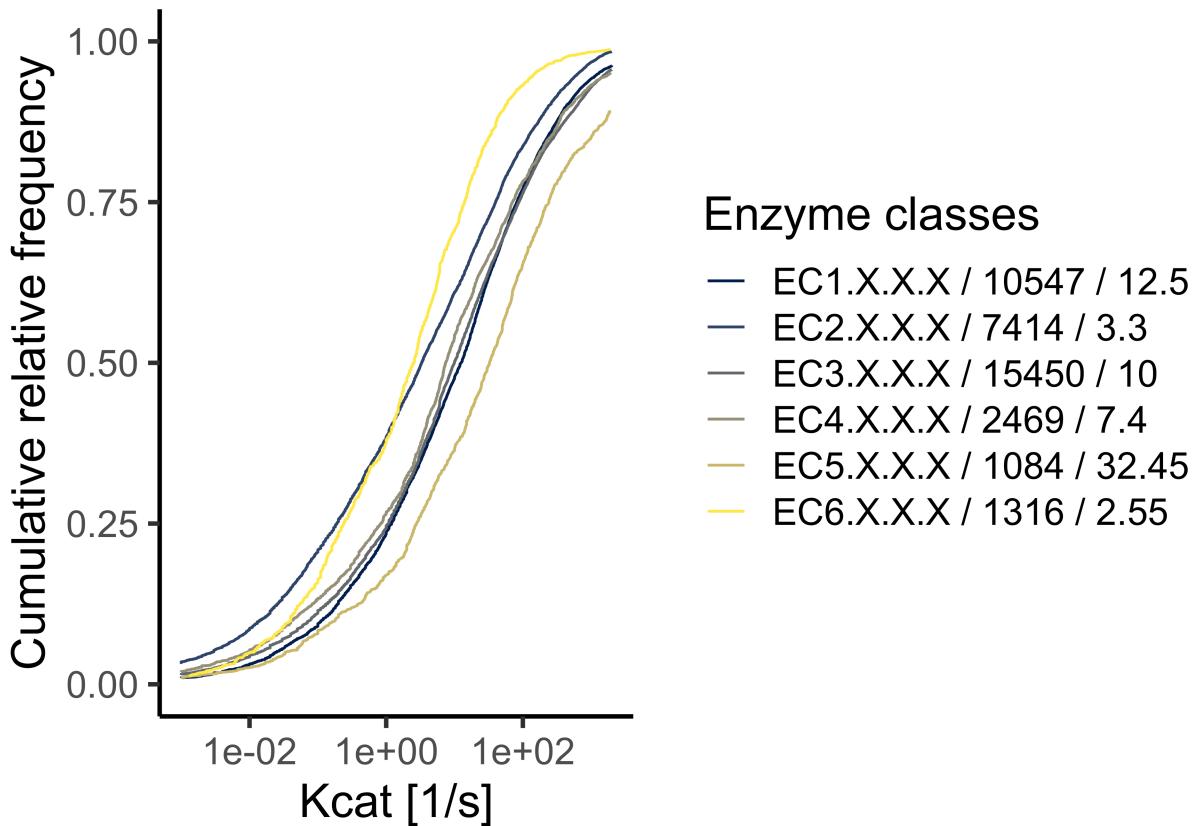


Figure S1.10: Cumulative distributions for Kcat values per enzyme family

In order to test if distributions of k_{cat} values differ significantly amongst enzyme families, a pairwise Kolmogorov-Smirnov statistical test is applied to every possible combination of distributions.

Table S1.5: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	EC1.X.X.X	EC2.X.X.X	EC3.X.X.X	EC4.X.X.X	EC5.X.X.X	EC6.X.X.X
EC1.X.X.X	1.00e+00	0	0.0000249	0.0000000	0	0
EC2.X.X.X	0.00e+00	1	0.0000000	0.0000000	0	0
EC3.X.X.X	2.49e-05	0	1.0000000	0.0007033	0	0
EC4.X.X.X	0.00e+00	0	0.0007033	1.0000000	0	0
EC5.X.X.X	0.00e+00	0	0.0000000	0.0000000	1	0
EC6.X.X.X	0.00e+00	0	0.0000000	0.0000000	0	1

As expected, enzyme catalytic activity distributions differ significantly across enzyme families. It should be noted that enzyme families represent groups of biochemical reactions which drastically differ in their mechanisms. To dig deeper into the catalytic specialization of different enzyme sub-families, wild cards can be introduced into the dataset EC numbers to obtain the subfamilies with more data entries available.

Table S1.6: Top10 represented enzyme subfamilies (2 wild-cards)

Enzyme groups	Number of entries
EC3.4.X.X	6292
EC1.1.X.X	4086
EC3.2.X.X	3189
EC3.1.X.X	2980
EC2.7.X.X	2196
EC3.5.X.X	1987
EC2.3.X.X	1467
EC2.4.X.X	1318
EC4.1.X.X	1034
EC1.14.X.X	1005

The top1 represented enzyme subfamily (top1 EC# w/1 wild-card) comprises different groups of enzymes (EC#'s w/2 wild-cards), the next plot shows how different k_{cat} distributions are for these enzyme groups.

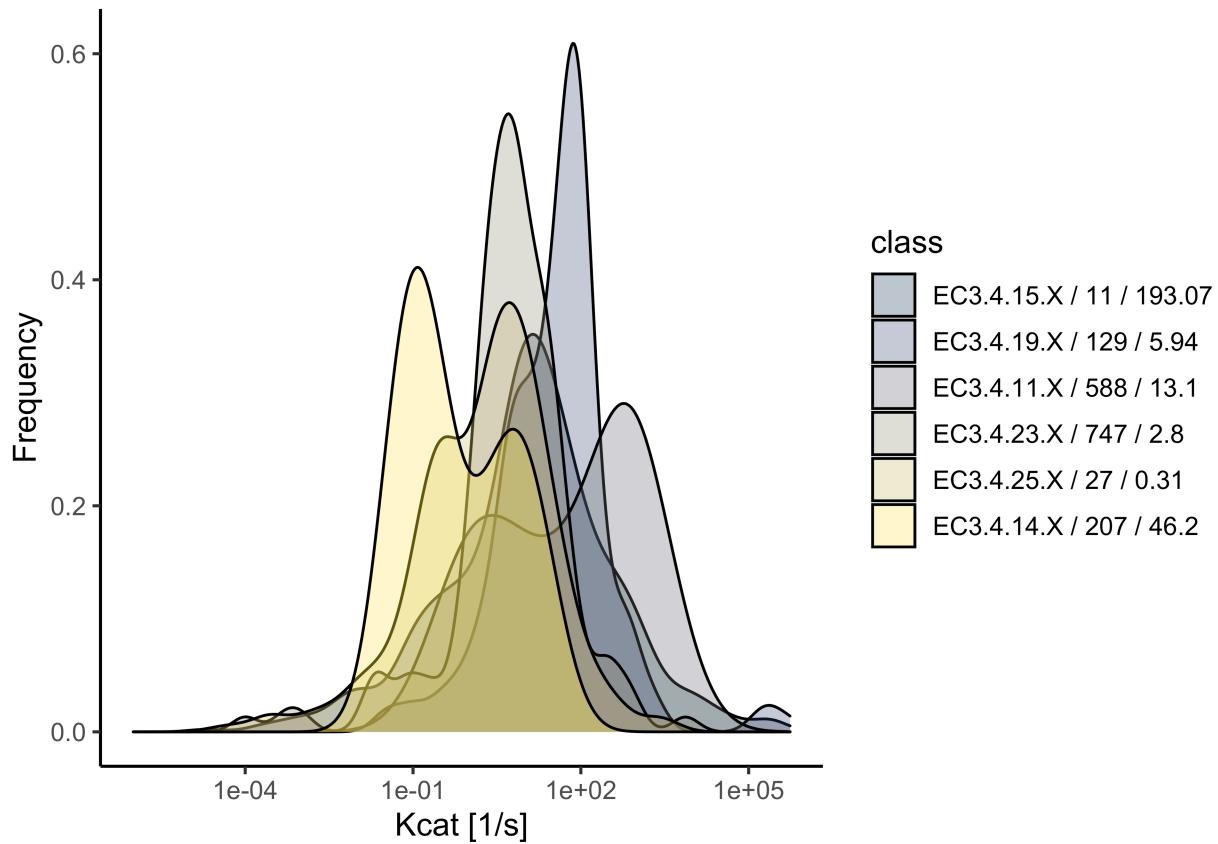


Figure S1.11: Cumulative distributions for Kcat values for different enzyme groups in the most represented enzyme subfamily (2 wild-cards)

Table S1.7: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	EC3.4.15.X	EC3.4.19.X	EC3.4.11.X	EC3.4.23.X	EC3.4.25.X	EC3.4.14.X
EC3.4.15.X	1.0000000	0.0175175	0.1215817	0.0064116	0.0105730	0.0503262
EC3.4.19.X	0.0175175	1.0000000	0.0000472	0.0000022	0.0000615	0.0000000
EC3.4.11.X	0.1215817	0.0000472	1.0000000	0.0000000	0.0000648	0.0000003
EC3.4.23.X	0.0064116	0.0000022	0.0000000	1.0000000	0.0142993	0.0000000
EC3.4.25.X	0.0105730	0.0000615	0.0000648	0.0142993	1.0000000	0.0000000
EC3.4.14.X	0.0503262	0.0000000	0.0000003	0.0000000	0.0000000	1.0000000

The same can be observed for the top-2 represented enzyme subfamily.

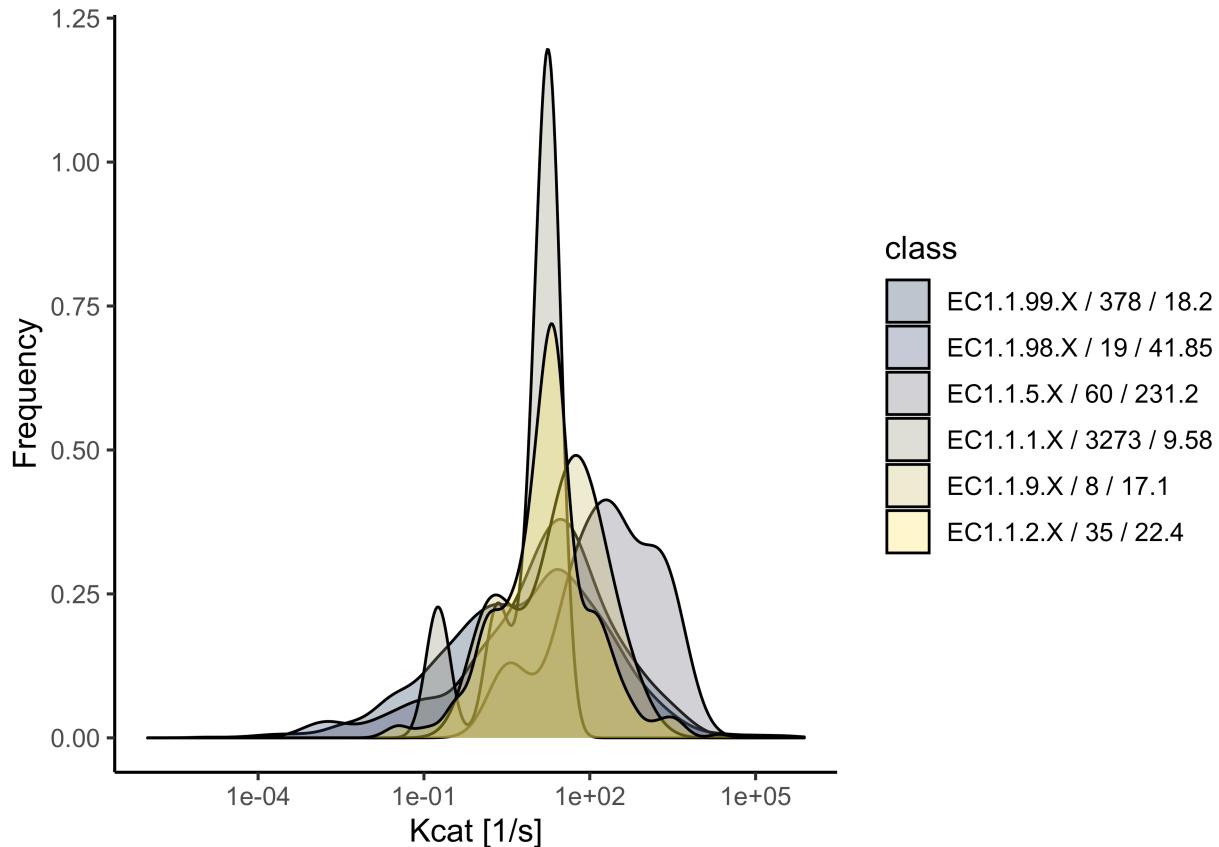


Figure S1.12: Cumulative distributions for Kcat values for different enzyme groups in the most represented enzyme subfamily (2 wild-cards)

Table S1.8: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	EC1.1.99.X	EC1.1.98.X	EC1.1.5.X	EC1.1.1.X	EC1.1.9.X	EC1.1.2.X
EC1.1.99.X	1.0000000	0.0466039	0.0000000	0.0000000	0.0644177	0.3665778
EC1.1.98.X	0.0466039	1.0000000	0.0017105	0.0724222	0.0224137	0.7081543
EC1.1.5.X	0.0000000	0.0017105	1.0000000	0.0000000	0.0000329	0.0000097
EC1.1.1.X	0.0000000	0.0724222	0.0000000	1.0000000	0.1377232	0.2677261
EC1.1.9.X	0.0644177	0.0224137	0.0000329	0.1377232	1.0000000	0.0638382
EC1.1.2.X	0.3665778	0.7081543	0.0000097	0.2677261	0.0638382	1.0000000

It can be seen that kinetic parameters not only differ significantly across general enzyme families (top-level EC numbers), but also across different enzyme subgroups (EC#'s w/ 2 wild-cards) within these families.

3.3.2 k_{cat} distributions per metabolic context

As all k_{cat} entries were assigned a metabolic subgroup, according to the KEGG pathways classification, the effect of metabolic context on k_{cat} values differentiation was also assessed. *Fig.S1.13* shows that, in general, enzymes in the central carbon and energy metabolism pathways tend to be significantly faster than those involved in amino acids, lipids and secondary metabolism.

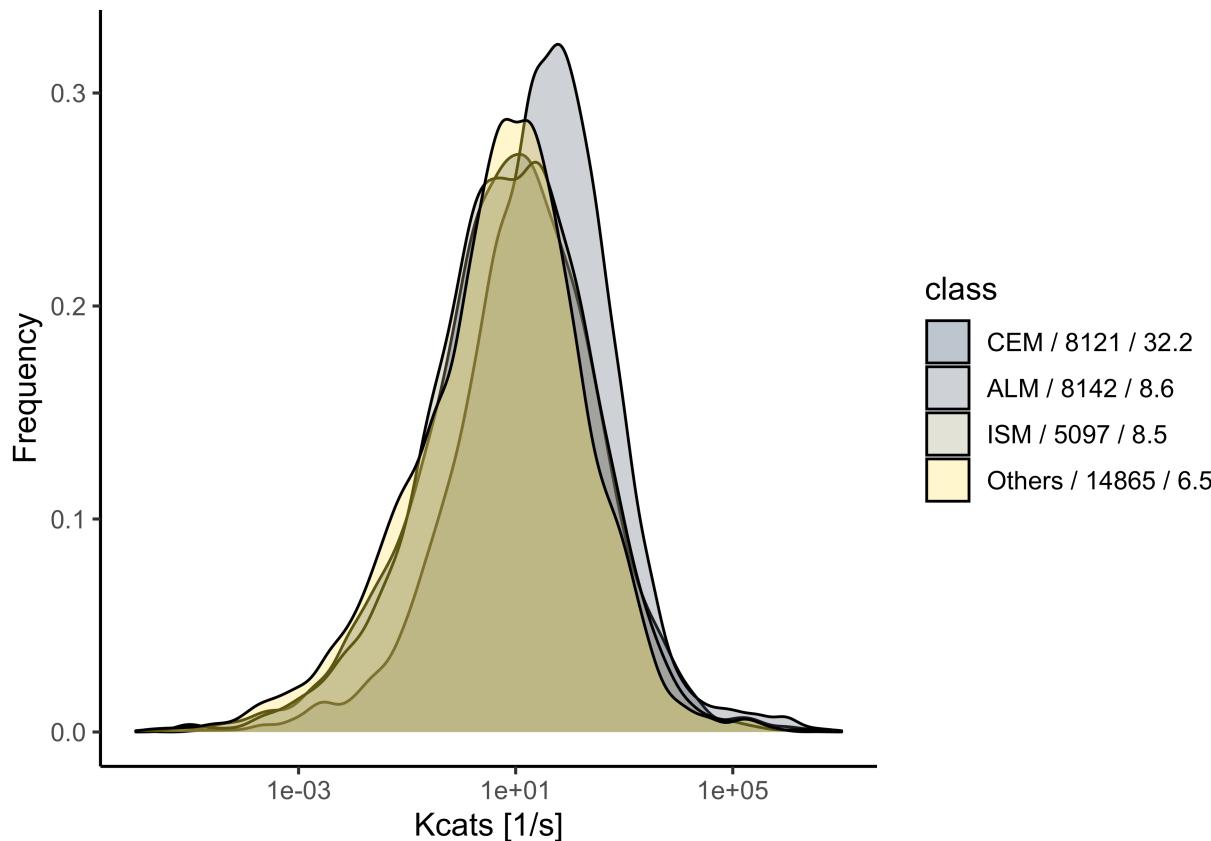


Figure S1.13: Kcat distributions sorted by metabolic context of enzymes

3.3.3 k_{cat} distributions per organism Kingdoms

Each entry of the retrieved k_{cat} values contains information regarding substrate and organism, it is then possible to assign a phylogenetic classification to each entry by retrieving the classification of the organism of origin from the KEGG organisms database. Therefore, comparison of k_{cat} distributions by different taxonomy levels is possible. Fig.S1.14 shows k_{cat} cumulative distributions when sorted by organism kingdom, where it is evident that activity values for enzymes in microbial organisms (fungi, bacteria and archaea) tend to be significantly higher than those for other organisms such as animals and plants.

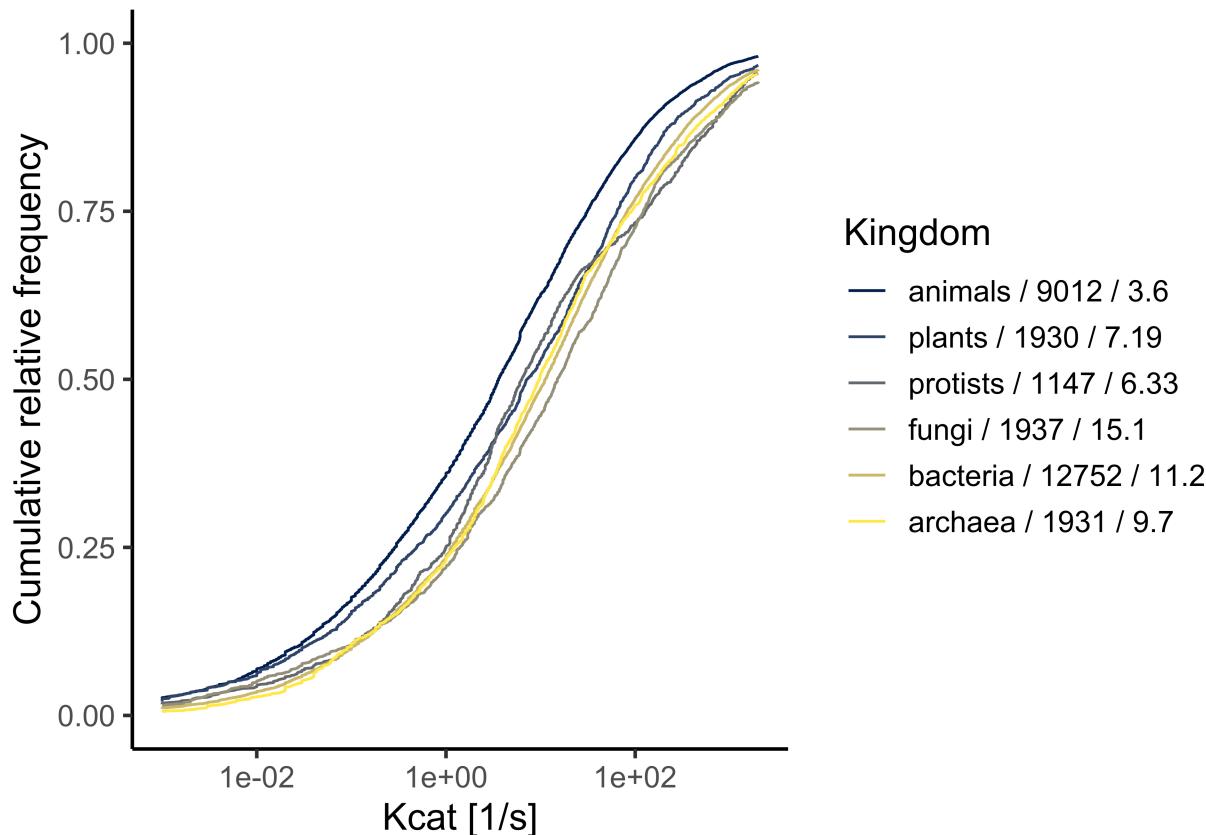


Figure S1.14: Cumulative distributions for Kcat values per organism kingdoms

Table S1.9: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	animals	plants	protists	fungi	bacteria	archaea
animals	1	0.00e+00	0.0000000	0.00e+00	0.0000000	0.0000000
plants	0	1.00e+00	0.0000510	4.00e-07	0.0000004	0.0000479
protists	0	5.10e-05	1.0000000	1.00e-07	0.0000606	0.0129826
fungi	0	4.00e-07	0.0000001	1.00e+00	0.0000341	0.0000252
bacteria	0	4.00e-07	0.0000606	3.41e-05	1.0000000	0.1634046
archaea	0	4.79e-05	0.0129826	2.52e-05	0.1634046	1.0000000

However, due to the impact of metabolic context on k_{cat} values differentiation (Fig.S1.13), this should also be taken into account when studying the role of phylogeny. Figs.S1.15 – 20 show that enzymes in Fungi organisms tend to be significantly faster than those for other organisms across **all** metabolic contexts.

Additionally, it was found that enzymes for higher organisms, such as animals, tend to display activity values that are lower, in general, than those for microbial organisms across **all** studied metabolic groups.

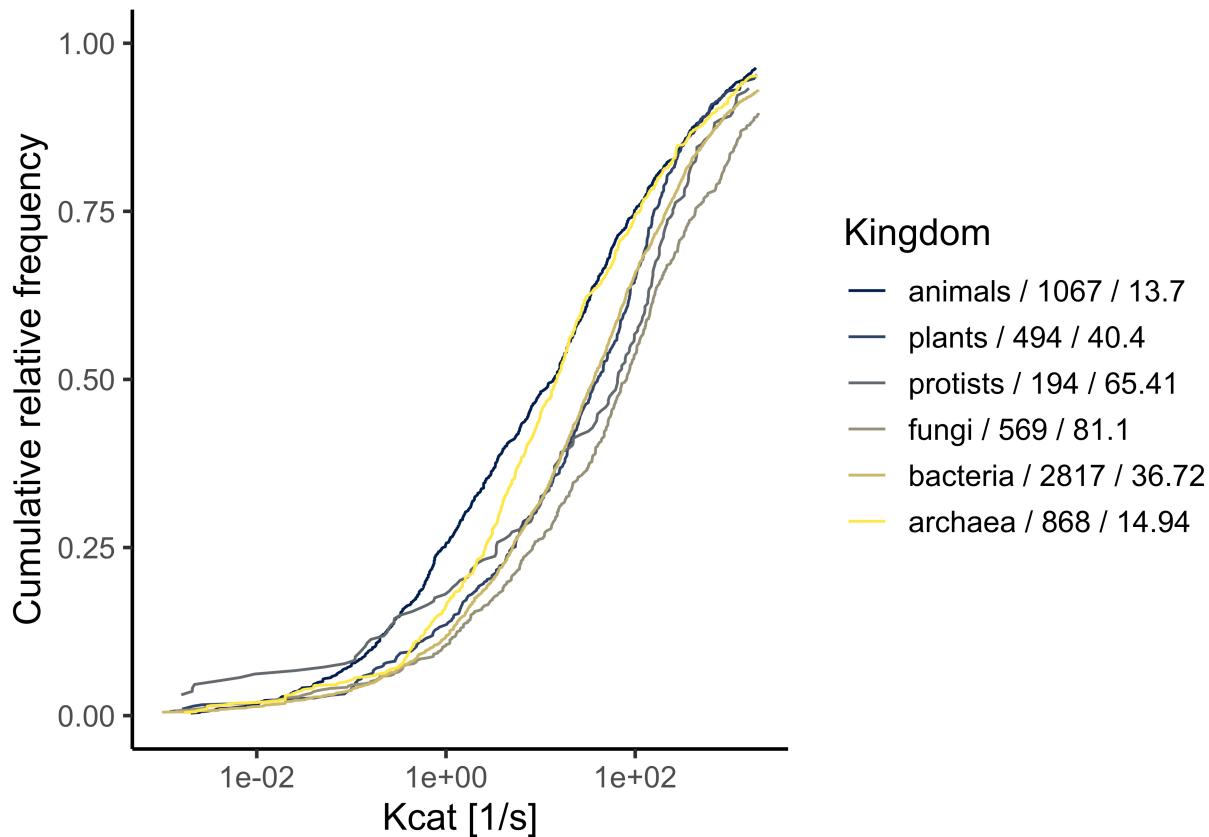


Figure S1.15: Cumulative distributions for Kcat values in CEM per organism kingdoms

Table S1.10: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	animals	plants	protists	fungi	bacteria	archaea
animals	1.00e+00	0.0000000	0.0000001	0.0000000	0.0000000	3.41e-05
plants	0.00e+00	1.0000000	0.1084608	0.0000540	0.2147233	1.00e-07
protists	1.00e-07	0.1084608	1.0000000	0.1037212	0.0420287	1.80e-06
fungi	0.00e+00	0.0000540	0.1037212	1.0000000	0.0000002	0.00e+00
bacteria	0.00e+00	0.2147233	0.0420287	0.0000002	1.0000000	0.00e+00
archaea	3.41e-05	0.0000001	0.0000018	0.0000000	0.0000000	1.00e+00

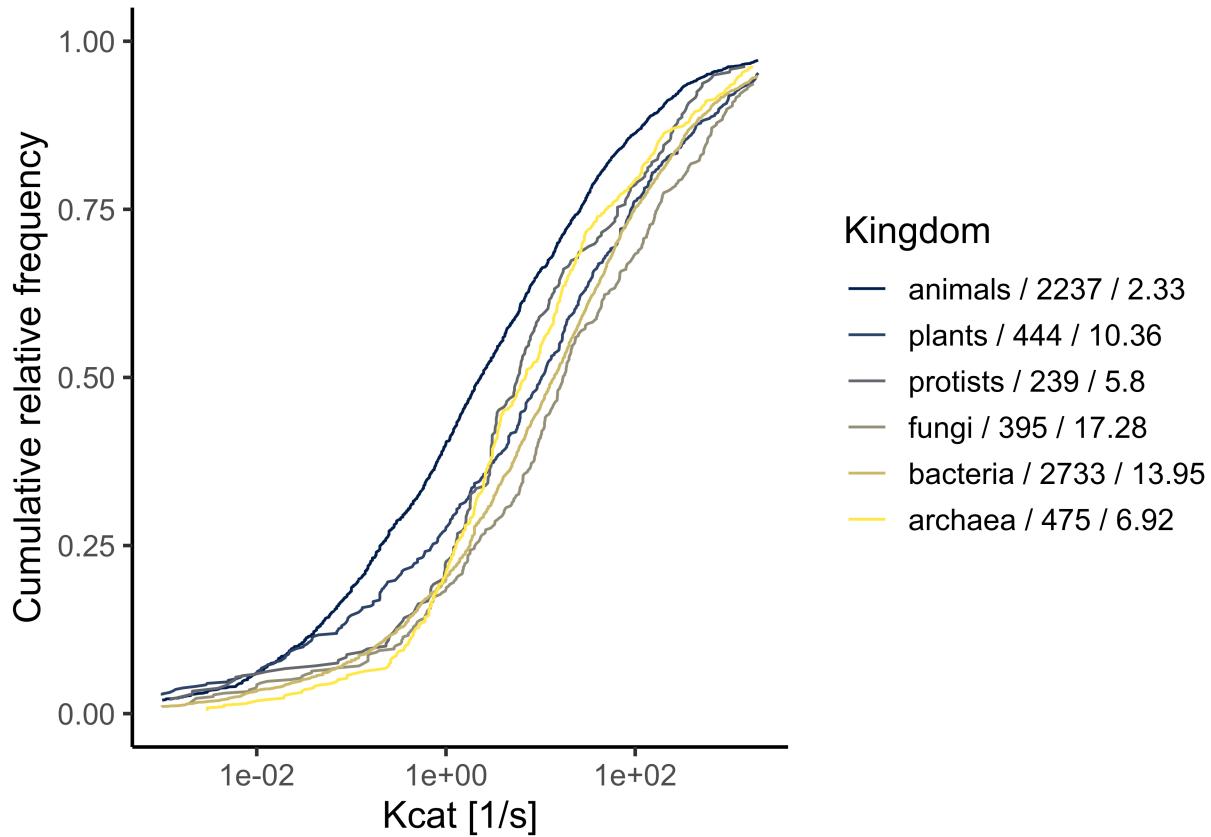


Figure S1.16: Cumulative distributions for Kcat values in ALM per organism kingdoms

Table S1.11: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	animals	plants	protists	fungi	bacteria	archaea
animals	1e+00	0.0000000	0.0000001	0.0000000	0.0000000	0.0000000
plants	0e+00	1.0000000	0.0795835	0.0110713	0.0094757	0.0012962
protists	1e-07	0.0795835	1.0000000	0.0000212	0.0005014	0.6270489
fungi	0e+00	0.0110713	0.0000212	1.0000000	0.0476123	0.0000556
bacteria	0e+00	0.0094757	0.0005014	0.0476123	1.0000000	0.0000591
archaea	0e+00	0.0012962	0.6270489	0.0000556	0.0000591	1.0000000

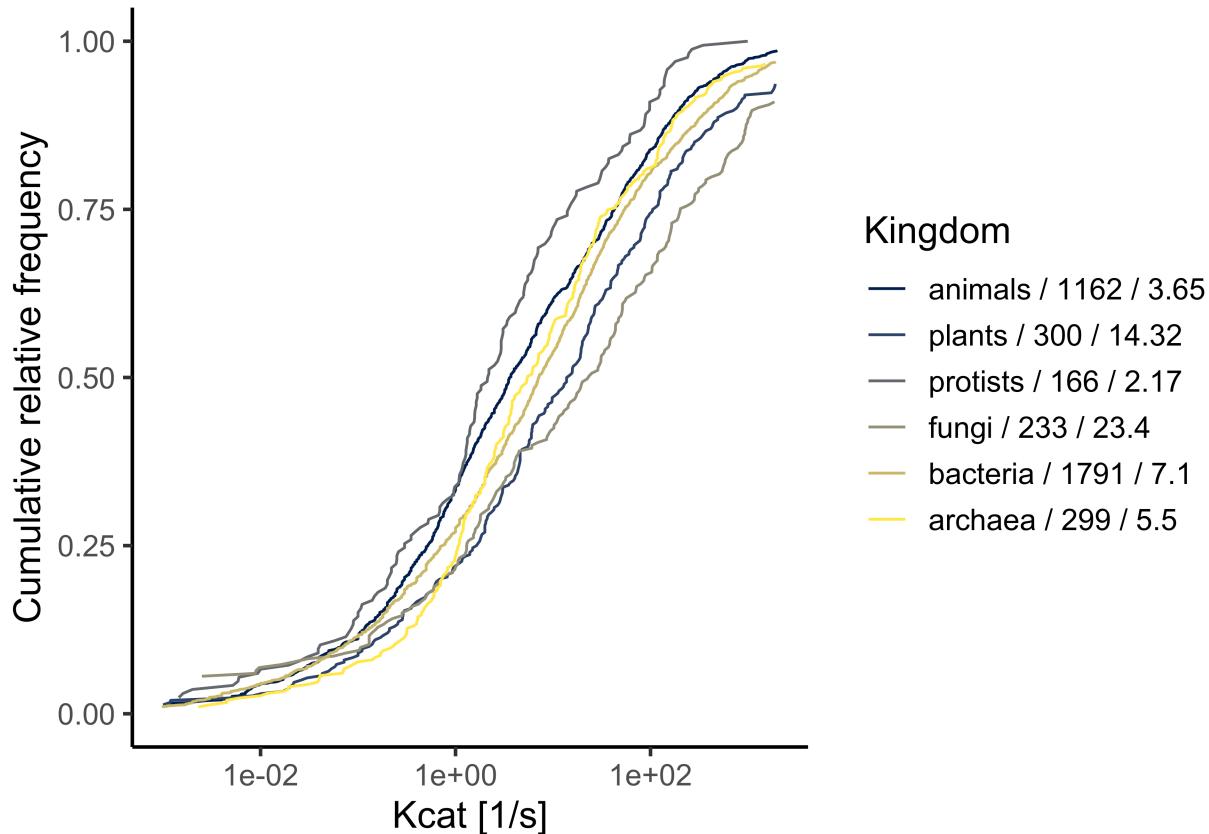


Figure S1.17: Cumulative distributions for Kcat values in ISM per organism kingdoms

Table S1.12: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	animals	plants	protists	fungi	bacteria	archaea
animals	1.0000000	0.0000038	0.0420024	0.0000005	0.0000303	0.0102206
plants	0.0000038	1.0000000	0.0000009	0.1433978	0.0197837	0.0138157
protists	0.0420024	0.0000009	1.0000000	0.0000000	0.0000186	0.0049434
fungi	0.0000005	0.1433978	0.0000000	1.0000000	0.0000094	0.0000057
bacteria	0.0000303	0.0197837	0.0000186	0.0000094	1.0000000	0.1498901
archaea	0.0102206	0.0138157	0.0049434	0.0000057	0.1498901	1.0000000

3.3.4 k_{cat} distributions per metabolic pathways groups per organism Kingdoms

It has been shown that, overall, enzymes in central carbon and energy metabolism have been reported to have higher activity values than those in other metabolic contexts (*Fig.S1.13*). Nonetheless, enzyme activity also seems to be related to the phylogeny of their organism of origin. In order to assess if CEM related enzymes are faster than others, distributions of k_{cat} values for each studied metabolic context were statistically compared for each of the KEGG kingdoms of life.

Figs.S1.18 – 23 show cumulative distributions for enzymes sorted by different metabolic groups for each of the KEGG kingdoms of life. Notably, central carbon and energy metabolism enzymes present higher values, on average, than those in other metabolic contexts. Notably, when focusing on enzymes for Fungi

and Bacteria organisms (two well-studied phylogenetic groups) it turns out that all k_c at value distributions for the studied metabolic contexts, differ significantly (*Tables S1.16 – 17*).

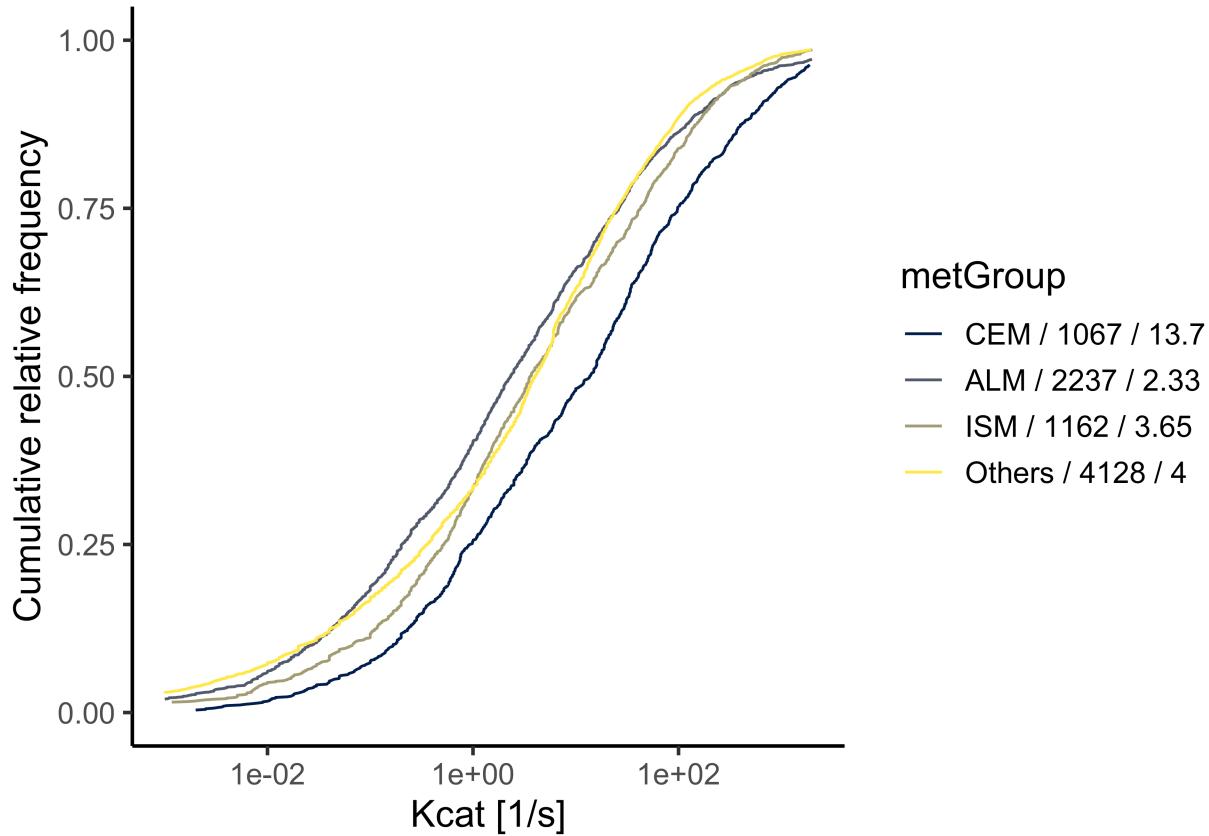


Figure S1.18: Cumulative distributions for Kcat values by metabolic subgroup for Animals

Table S1.13: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	CEM	ALM	ISM	Others
CEM	1	0.00e+00	0.0000000	0.0000000
ALM	0	1.00e+00	0.0000194	0.0000000
ISM	0	1.94e-05	1.0000000	0.0021873
Others	0	0.00e+00	0.0021873	1.0000000

Animals

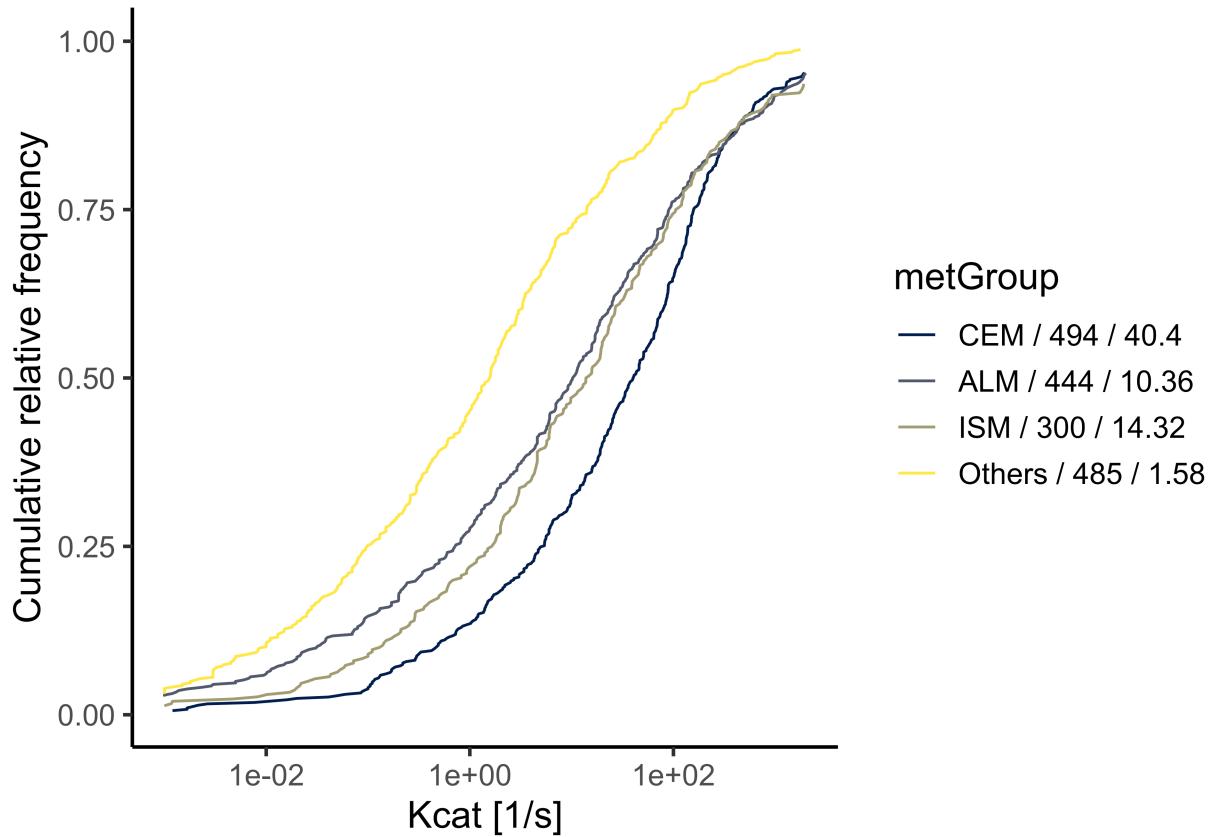


Figure S1.19: Cumulative distributions for Kcat values by metabolic subgroup for Plants

Table S1.14: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	CEM	ALM	ISM	Others
CEM	1.0000000	0.0000000	0.0001715	0
ALM	0.0000000	1.0000000	0.2961840	0
ISM	0.0001715	0.296184	1.0000000	0
Others	0.0000000	0.0000000	0.0000000	1

Plants

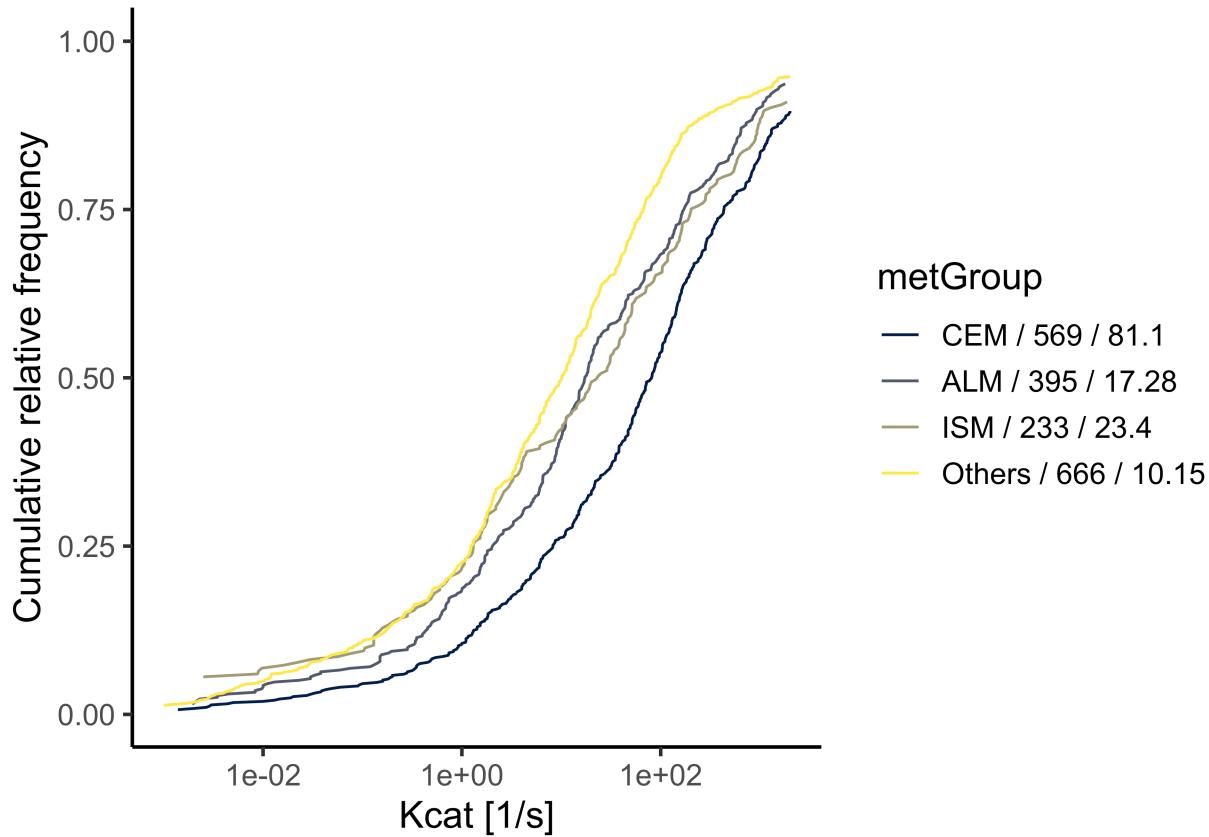


Figure S1.20: Cumulative distributions for Kcat values by metabolic subgroup for Fungi

Table S1.15: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	CEM	ALM	ISM	Others
CEM	1.00e+00	0.0000000	0.0000102	0.0000000
ALM	0.00e+00	1.0000000	0.2495139	0.0004581
ISM	1.02e-05	0.2495139	1.0000000	0.0006279
Others	0.00e+00	0.0004581	0.0006279	1.0000000

Fungi

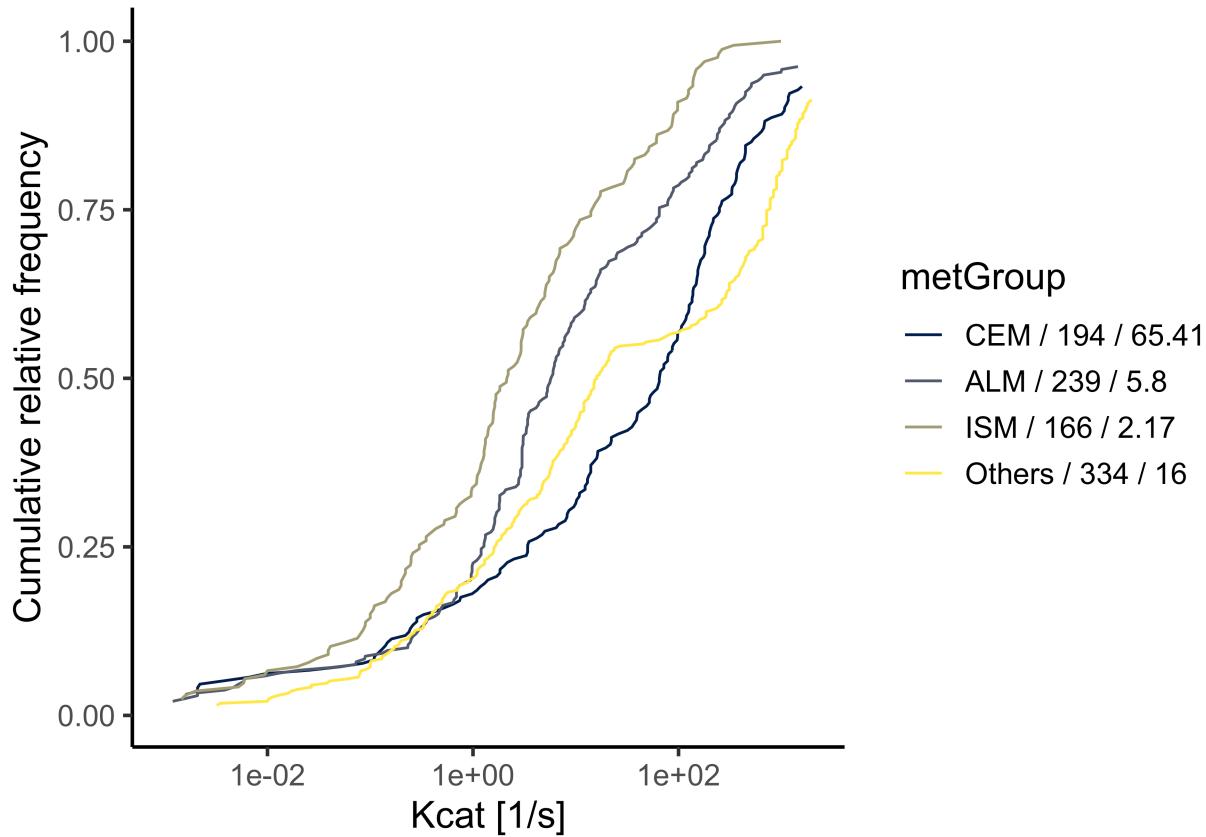


Figure S1.21: Cumulative distributions for Kcat values by metabolic subgroup for Protists

Table S1.16: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	CEM	ALM	ISM	Others
CEM	1.0000000	0.0000001	0.0000000	0.0014397
ALM	0.0000001	1.0000000	0.0016707	0.0000000
ISM	0.0000000	0.0016707	1.0000000	0.0000000
Others	0.0014397	0.0000000	0.0000000	1.0000000

Protists

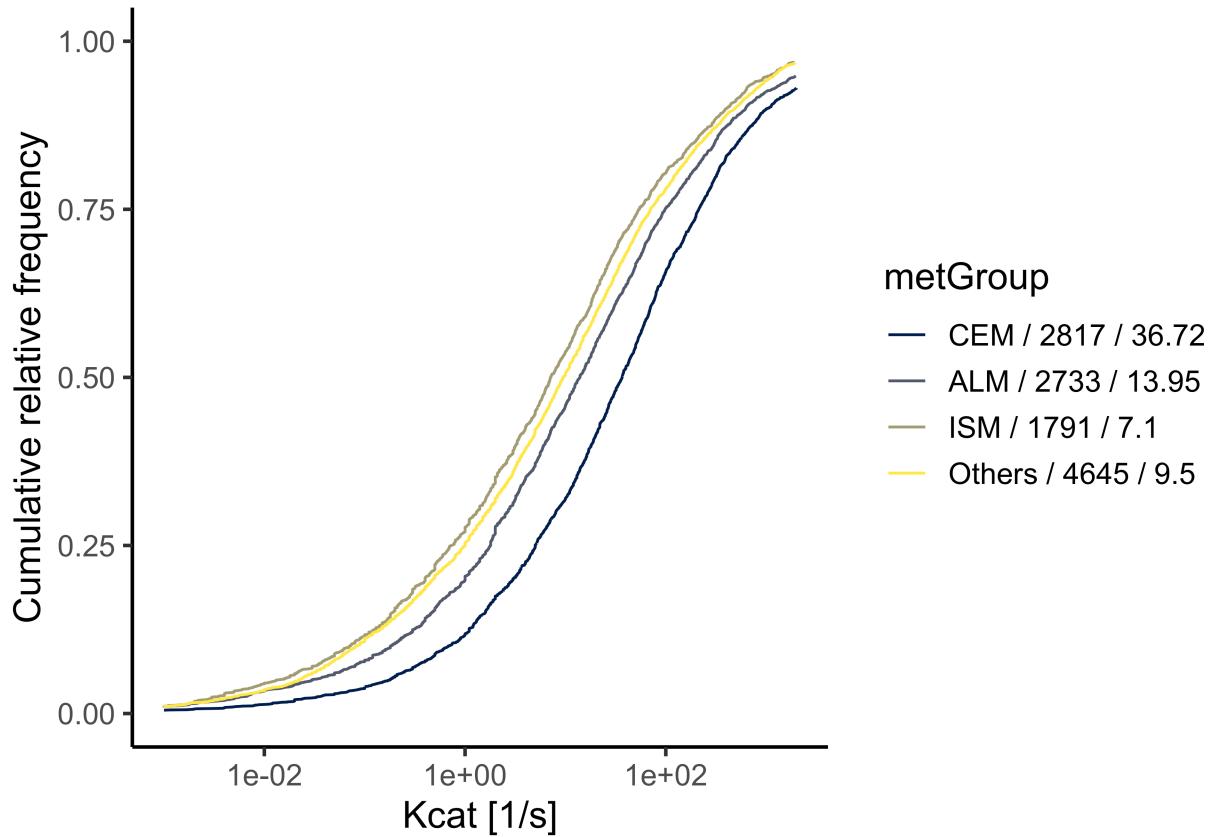


Figure S1.22: Cumulative distributions for Kcat values by metabolic subgroup for Bacteria

Table S1.17: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	CEM	ALM	ISM	Others
CEM	1	0.00e+00	0.0000000	0.0000000
ALM	0	1.00e+00	0.0000004	0.0000188
ISM	0	4.00e-07	1.0000000	0.0280320
Others	0	1.88e-05	0.0280320	1.0000000

Bacteria

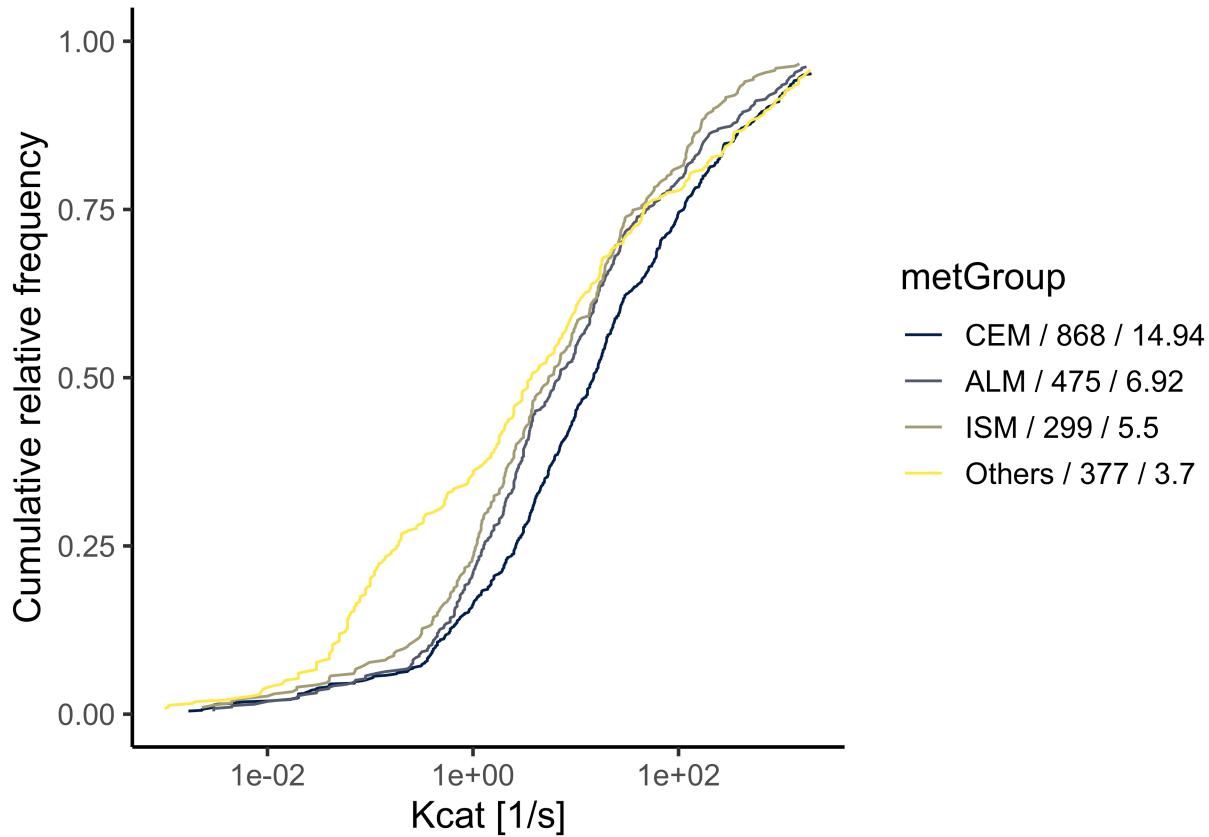


Figure S1.23: Cumulative distributions for Kcat values by metabolic subgroup for Archaea

Table S1.18: p-values under the pairwise Kolmogorov-Smirnov two-tailed statistical test

	CEM	ALM	ISM	Others
CEM	1.00e+00	0.0000211	0.0000333	0e+00
ALM	2.11e-05	1.0000000	0.8120868	0e+00
ISM	3.33e-05	0.8120868	1.0000000	8e-05
Others	0.00e+00	0.0000000	0.0000800	1e+00

Archaea