

# BCP-MG: A Web Server for Predicting Bacterial Community of Metagenome

De-Guo Xia  
Hao-Ran Zheng  
Gui-Sheng Li  
Zhi-Qiang Liu  
Kai Zhao

---

## Method Article

**Keywords:** metagenome, bacterial community, predict

**Posted Date:** May 18th, 2010

**DOI:** <https://doi.org/10.1038/nprot.2010.99>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

## Introduction

The researchers have been paying more attention to culturing individual microorganisms for a long time, but the majority of microorganisms found in natural ecosystems cannot be easily cultured in the laboratory<sup>1</sup>. Metagenomics analysis of genetic material obtained directly from the environment without isolating and culturing the species in a laboratory, which is not possible with traditional methods since only a tiny fraction of all microorganisms ( $\approx 1\%$ ) can be cultivable using standard techniques<sup>2</sup>. The metagenomics analysis employs techniques that enable researchers to obtain, and sequence the genomic content of microbial communities directly, thereby bypassing the need for prior cultivation of all individual organisms present in the sample<sup>3,4</sup>. The metagenomics offers scientists the method to study the structure of the microorganism community. The current research of metagenome focuses on gene annotation<sup>5-7</sup>, reconstruction of the metabolism network<sup>8</sup>, analysis the diversity of the microorganisms community<sup>9,10</sup> and so on, which are mainly based on the known genetic fragments<sup>11-13</sup>. Since diversity in metagenomic samples is often too large to provide a high sequencing coverage of single species<sup>14-16</sup>, these methods cannot give a microscopy view of the microbial community structure from a metagenomic sample data. So, we present a tool which can predict a bacterial community from a metagenomic sample data according to the enzyme information extracted from the sample or the metabolism network reconstructed from this data. The predicted community not only covers all the known metabolic functions, but also can be used to discover unknown functions of the metagenomic samples. This tool can help researcher to form an overall understanding and find new functions in the metagenomic sample data. The webaddress for this tool is: "<http://bioinfo.ustc.edu.cn/software/bcp-mg/>":<http://bioinfo.ustc.edu.cn/software/bcp-mg/>

## Equipment

A computer with access to the internet and a web browser. **Data** Your input is a flat file which contains the enzyme information extracted from the metagenomic sample data or the reactions reconstructed from this data. Each enzyme or reaction takes one line. **Programs** In practice, BCP-MG server process query data using a computationally intensive bioinformatics protocol. (A detailed flow-chart of the protocol is provided in Figure 1.) BCP-MG is a php web server, and the core background program was written in C++ which uses an improved set-covering algorithm to predict the bacterial community.

## Procedure

1. Choose the data type you used for representing the metagenomic sample data. The data type includes enzyme and reaction. You can use the enzyme information extracted from the sample to predict the bacterial community; reaction information reconstructed from the sample can be used also.
2. upload the

data file The data file must be a text file. \! CAUTION Uploaded file must be a plain text file \ (generally using ASCII or Unicode schemes) while other rich text formats like those produced by most text editing tools e.g. Microsoft Word , cannot be processed by BCP-MG server. It may take a little time to upload this file depends on the size of the file you want to upload and the bandwidth limit. 3.Choose the metabolic database to use There are two metabolic databases that you can use, BioCyc and KEGG. Individual databases or the union can be applied. Please note that there are many differences between these two databases. One is that the organisms' number and kinds are not identical in the two databases. The other is that even the same organism in the two databases will have different metabolic network for they have used diverse pathway reconstruction algorithms and curative methods. 4.Choose the organisms selecting strategy There are two organisms selecting strategies in use: all-organisms and base-set-organisms. If all-organisms is chosen, the predict algorithm considers all organisms without predilection which may introduce some organisms with same strain. Reversely, the base-set-organisms strategy eliminates these organisms with same strain and reserves one typical organism. 5.Check for any messages. If your data were uploaded successfully to the BCP-MG server, you can see the message "upload successfully", and then a preprocess script will be employed to check whether your data is complied with the request. Depending on the success or failure of enzymes or reactions information submission different messages will be displayed. If you have not given required information, an error message, "Incorrect file content. Please make sure the content of the file uploaded are enzymes or reactions\!" will be displayed, you need to recheck and fill the required information for successful submission. 6.Obtain the results. Once your submission is successful, the data will be processed by the server and the results can be accessed by pressing the "Show Results" button which is available on the neck of the process finished. Time required for processing of your data highly depends on the size of your query and on the parameters you have chosen \ (please see the section on Time Taken to obtain an indication). 7.Interpret the results. The results are displayed using a table, see Figure 3. The table displays the size of the predicted bacterial community and the microorganisms with details \ (Domain, Phylum, Class, Order, Genus and Population according to the KEGG)

## Timing

Time taken highly depends on the size of the enzymes or reactions information of the metagenomic sample data, the metabolism database and the other parameters chosen by the user \ (more details are shown on Figure 2). For example, the GS000a Shotgun - Open

## Troubleshooting

If the server does not accept the input data for prediction, the error might be caused by one of the following reasons: 1.Input file is not a txt file 2.The content of the input file is not enzymes or reactions 3.The format of the input file is incorrect. 4.The content of the input file does not match the data type you have selected. Error information will be given if you set data type as enzyme but upload a data file with reactions information.

# Anticipated Results

The analysis of metagenome samples by BCP-MG protocol provides a quick and conservative – but reliable - prediction of bacterial community conditions. Here we use Acid Mine Drainage Biofilm (4441137.3) as an example, which has been well studied<sup>17-19</sup>. The predicted bacterial community, with 22 organisms, is smaller than the community generated by MG-RAST which includes 69 organisms. The comparison detail is shown in Table. 1 with different taxonomic ranks considered (phylum, class and order). At the rank of phylum, 17 out of 22 organisms in the predicted community correspond to the community generated by MG-RAST, while 15 at the class level and 4 at the order level.

## References

1. Amann, R.L., Ludwig, W. & Schleifer, K.H. Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation. *Microbiological Reviews* 59, 143-169 (1995).
2. Hugenholtz, P., Goebel, B.M. & Pace, N.R. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *Journal of Bacteriology* 180, 4765-4774 (1998).
3. Hugenholtz, P. Exploring prokaryotic diversity in the genomic era. *Genome Biol* 3, REVIEWS0003 (2002).
4. Rappe, M.S. & Giovannoni, S.J. The uncultured microbial majority. *Annu Rev Microbiol* 57, 369-394 (2003).
5. Meyer, F. et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *Bmc Bioinformatics* 9, - (2008).
6. Kristiansson, E., Hugenholtz, P. & Dalevi, D. ShotgunFunctionalizeR: an R-package for functional comparison of metagenomes. *Bioinformatics* 25, 2737-2738 (2009).
7. Kislyuk, A., Bhatnagar, S., Dushoff, J. & Weitz, J.S. Unsupervised statistical clustering of environmental shotgun sequences. *Bmc Bioinformatics* 10, - (2009).
8. Ye, Y.Z. & Doak, T.G. A Parsimony Approach to Biological Pathway Reconstruction/Inference for Genomes and Metagenomes. *Plos Computational Biology* 5, - (2009).
9. Tyson, G.W. et al. Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* 428, 37-43 (2004).
10. Schreiber, F., Gumrich, P., Daniel, R. & Meinicke, P. Treephyler: fast taxonomic profiling of metagenomes. *Bioinformatics* (2010).
11. Wooley, J.C., Godzik, A. & Friedberg, I. A Primer on Metagenomics. *Plos Computational Biology* 6, - (2010).
12. Schloss, P.D. & Handelsman, J. Metagenomics for studying unculturable microorganisms: cutting the Gordian knot. *Genome Biol* 6, 229 (2005).
13. Riesenfeld, C.S., Schloss, P.D. & Handelsman, J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* 38, 525-552 (2004).
14. Hoff, K.J., Lingner, T., Meinicke, P. & Tech, M. Orphelia: predicting genes in metagenomic sequencing reads. *Nucleic Acids Res.* 37, W101-W105 (2009).
15. Tringe, S.G. & Rubin, E.M. Metagenomics: DNA sequencing of environmental samples. *Nature Reviews Genetics* 6, 805-814 (2005).
16. Kunin, V., Copeland, A., Lapidus, A., Mavromatis, K. & Hugenholtz, P. A Bioinformatician's Guide to Metagenomics. *Microbiol Mol Biol R* 72, 557-578 (2008).
17. Baker, B.J. & Banfield, J.F. Microbial communities in acid mine drainage. *FEMS Microbiol Ecol* 44, 139-152 (2003).
18. Xie, X., Xiao, S. & Liu, J. Microbial communities in acid mine drainage and their interaction with pyrite surface. *Curr Microbiol* 59, 71-77 (2009).
19. Bond, P.L., Druschel, G.K. & Banfield,

J.F. Comparison of acid mine drainage microbial communities in physically and geochemically distinct ecosystems. Appl Environ Microbiol 66, 4962-4971 \ (2000).

## Acknowledgements

This work has been supported by the National Key Technologies R&D Program \ (2006CB910705).

## Figures

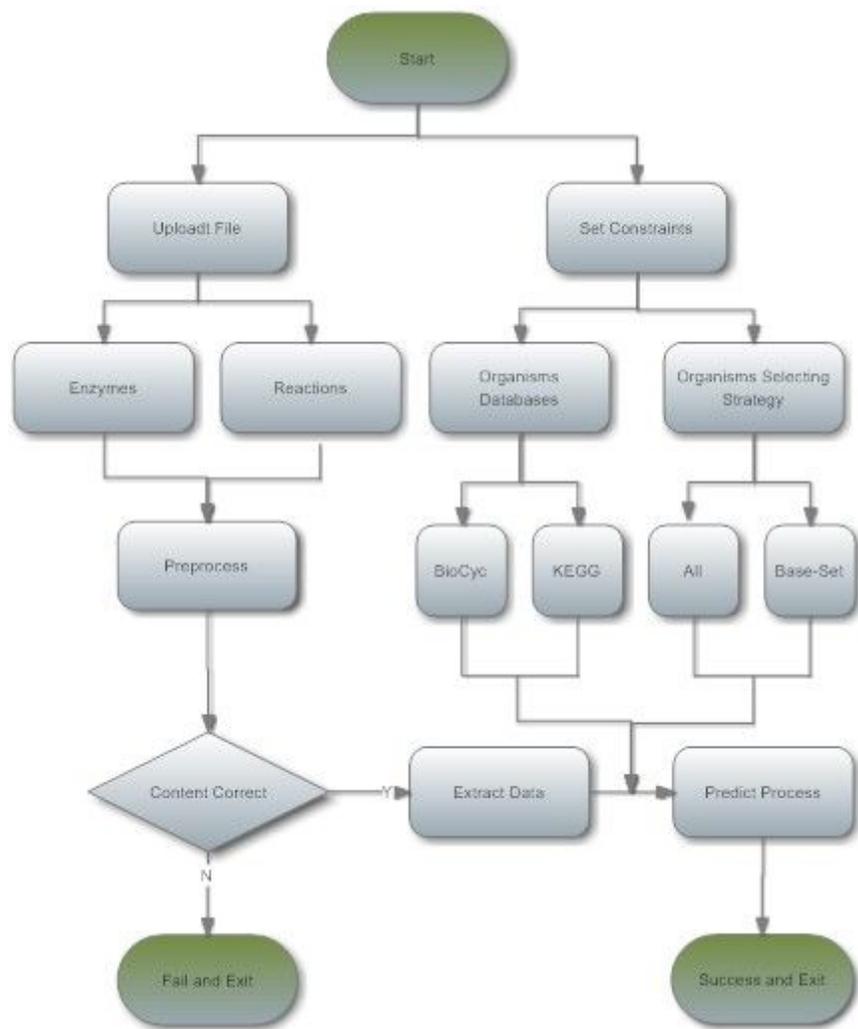


Figure 1

Flowchart illustrating the methodology

**Predict Configurations**

Select the data type  reactions list  ec list

\* The data that you upload for predict is either a list of reactions or ec numbers.

Upload your data  (Example)

Select the database  KEGG  BioCyc  Both

\* Select the database that you want to use for predict. There may be some differences in organisms number or metabolite networks of each organism between the two databases.

Organisms Limits Using  All  Base Set of the organisms

\* Predict the Bacterial Community using all or the base set of the organisms that exist in our database.

Figure 2

BCP-MG input options

**Task Result**

Number	Details (name) (Domain Phylum Class Order Genus Population)
	pchl-425-01cyc (Parachlamydia sp) (Prokaryotes Bacteria Verrucomicrobia)
	yeastcyc (S. cerevisiae S288C) (Prokaryotes Archaea Crenarchaeota)
	mme246198cyc (Mycobacterium smegmatis MC2 155) (Prokaryotes Bacteria Actinobacteria)
	avar240292cyc (Anabaena variabilis ATCC 29413) (Prokaryotes Bacteria Cyanobacteria)
	rzph272943cyc (Rhodobacter sphaeroides 2.4.1) (Prokaryotes Bacteria Proteobacteria Alphalobacteria)
	ecocyc (Escherichia coli K12) (Prokaryotes Bacteria Proteobacteria GammaEnterobacteria)
	lpla220688cyc (Lactobacillus plantarum WCFS1) (Prokaryotes Bacteria Firmicutes Lactobacillales)
	ecoo157cyc (Escherichia coli O157:H7 EDL933) (Prokaryotes Bacteria Proteobacteria GammaEnterobacteria)
	snub309607cyc (Salinibacter ruber DSM 13855) (Prokaryotes Bacteria Bacteroides)
	bhp36773cyc (Burkholderia sp. 383) (Prokaryotes Bacteria Proteobacteria Beta)
	ecol354106cyc (Escherichia coli UT89) (Prokaryotes Bacteria Proteobacteria GammaEnterobacteria)
	brme358391cyc (Brucella melitensis biovar Abortus) (Prokaryotes Bacteria Proteobacteria Alphaproteobacteria)
Predicted Bacterial Community	25
	agroccyc (Agrobacterium tumefaciens C58) (Prokaryotes Bacteria Proteobacteria Alphaproteobacteria)
	rub269796cyc (Rhodospirillum rubrum ATCC 11170) (Prokaryotes Bacteria Proteobacteria Alphaproteobacteria)
	bcer228900cyc (Bacillus cereus ATCC 14579) (Prokaryotes Bacteria Firmicutes Bacillales)
	cvio243365cyc (Chromobacterium violaceum ATCC 12472) (Prokaryotes Bacteria Proteobacteria Beta)
	mhun323259cyc (Methanospirillum hungatei JF-1) (Prokaryotes Archaea Euryarchaeota)
	rden375451cyc (Roseobacter denitrificans OCh 114) (Prokaryotes Bacteria Proteobacteria Alphaproteobacteria)
	lnt-130-01cyc (Leptospira interrogans L1-130) (Prokaryotes Bacteria Spirocheta)
	asp76114cyc (Azotarcus sp. EbN1) (Prokaryotes Bacteria Proteobacteria Beta)
	mmaz192952cyc (Methanosarcina mazei Go1) (Prokaryotes Archaea Euryarchaeota)
	bxen266265cyc (Burkholderia xenovorans LB400) (Prokaryotes Bacteria Proteobacteria Beta)
	frantcyc (Francisella tularensis tularensis SCHU S4) (Prokaryotes Bacteria Proteobacteria GammaAlphaproteobacteria)
	vohocyc (Vibrio cholerae O1 biovar eltor str. N16961) (Prokaryotes Bacteria Proteobacteria GammaAlphaproteobacteria)
	hsom205914cyc (Haemophilus somnus 125PT) (Prokaryotes Bacteria Proteobacteria GammaAlphaproteobacteria)

Figure 3

BCP-MG result

		BCP-HG	MG-RAST	
Actinobacteria		Mycobacterium smegmatis	Acidimicrobium	
			Brevibacterium	
			unclassified_Microbacteriales	
			Arthrobacter	
			Rhodococcus	
			Amycolatopsis	
			unclassified_Actinomycetales	
			unclassified_Actinobacteria	
Firmicutes	Cyanobacteria	Anabaena variabilis	Oscillatoria	
		Cyanospora sp. PCC 7424	unclassified_Family 4.1	
		Synechococcus sp. WH7803	unclassified_Deferribacteriales	
	Bacilli	Bacillus amyloquelificans	Bacillus	
			unclassified_Bacillales	
			Paenibacillus	
	Mollicutes		unclassified_Bacillales	
			Acholeplasma	
			Ureaplasma	
Clostridia		unclassified_Mollicutes		
		Caloramprobacter		
		Clostridium difficile		
		Haloferax volcanium		
		unclassified_Lachnospiraceae		
		Dehalobacter		
		Pelotomaculum		
		unclassified_Syntrophomonadaceae		
		unclassified_Clostridiales		
		unclassified_Clostridia		
Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobium japonicum	
			unclassified_Methylocystaceae	
		Rhodospirillales	Magnetospirillum magnetium	
			Rhizobium	
	Sphingomonadales	Beta	Acidovorax avenae	Erythrobacter
			Azoarcus sp. BH72	unclassified_Burkholderiales
			Aromatobacterium aromaticum	Dechloromonas
			Bordetella pertussis	
			Desulfobacterium autotrophicum	Desulfosphaerulum
			Desulfosphaerulum psychrophilum	unclassified_Deltaproteobacteria
	Gammaproteobacteria	enterobacteria	Candidatus_Biochlamydia pennsylvanicus	unclassified_Enterobacteriaceae
			Yersinia enterocolitica	
		Atheromonadales		Pseudoalteromonas
				Shewanella
				Thalassomonas
				unclassified_Atheromonadales
				Methylobacter
				Oceanospirillales
				unclassified_Oceanospirillales
				Pseudomonadales
		Pseudomonas		
		unclassified_Xanthomonadales		
		unclassified_Gammaproteobacteria		
Basidiomycetes	Cryptosporidium parvum			
Chlorobacteria	Chlorobium chlorochromatii			
Aquificae		Sulfurohydrogenibium		
Bacteroidetes			Prevotella	
			Psychroserpens	
			Tenacibaculum	
			unclassified_Flavobacteriales	
			Chlorohaga	
			Hymenobacter	
			unclassified_Flavobacteriales	
			Halicomonas	
			unclassified_Bacteroidetes	
			Thermus	
Denitrospira-Thermus				
Genera_Incertae_sedis_OPF10		OPF10		
Nitrospira		Leptospirillum		
		Thermodesulfobacterium		
Fluctococcoides		Prellis		
Spirillozoetes		Thiospira		
Thermotoga		Thermotoga		
unclassified_Bacteria		unclassified_Bacteria		

Figure 4

Table 1 The predicted bacterial community contrast with MG-RAST data

**Task Result**

	Number	Details (name) (Domain Phylum Class Order Genus Population)
Predicted Bacterial Community	25	pchl-e25-01cyc (Parachlamydia sp) (Prokaryotes Bacteria Verrucomicrobia)
		yeastcyc (S. cerevisiae S288C) (Prokaryotes Archaea Crenarchaeota)
		msme246196cyc (Mycobacterium smegmatis MC2 155) (Prokaryotes Bacteria Actinobacteria)
		avar240292cyc (Anabaena variabilis ATCC 29413) (Prokaryotes Bacteria Cyanobacteria)
		rsph272943cyc (Rhodobacter sphaeroides 2.4.1) (Prokaryotes Bacteria Proteobacteria Alpha/others)
		ecocyc (Escherichia coli K12) (Prokaryotes Bacteria Proteobacteria Gamma/enterobacteria)
		lpla220668cyc (Lactobacillus plantarum WCFS1) (Prokaryotes Bacteria Firmicutes Lactobacillales)
		ecoo157cyc (Escherichia coli O157:H7 EDL933) (Prokaryotes Bacteria Proteobacteria Gamma/enterobacteria)
		srub309807cyc (Salinibacter ruber DSM 13855) (Prokaryotes Bacteria Bacteroides)
		bsp36773cyc (Burkholderia sp. 383) (Prokaryotes Bacteria Proteobacteria Beta)
		ecol364106cyc (Escherichia coli UTI89) (Prokaryotes Bacteria Proteobacteria Gamma/enterobacteria)
		bmel359391cyc (Brucella melitensis biovar. Abortus) (Prokaryotes Bacteria Proteobacteria Alpha/rhizobacteria)
		agrocyc (Agrobacterium tumefaciens C58) (Prokaryotes Bacteria Proteobacteria Alpha/rhizobacteria)
		rrub269796cyc (Rhodospirillum rubrum ATCC 11170) (Prokaryotes Bacteria Proteobacteria Alpha/others)
		bcer226900cyc (Bacillus cereus ATCC 14579) (Prokaryotes Bacteria Firmicutes Bacillales)
		cvio243365cyc (Chromobacterium violaceum ATCC 12472) (Prokaryotes Bacteria Proteobacteria Beta)
		mhun323259cyc (Methanospirillum hungatei JF-1) (Prokaryotes Archaea Euryarchaeota)
		rden375451cyc (Roseobacter denitrificans OCh 114) (Prokaryotes Bacteria Proteobacteria Alpha/others)
		lint-130-01cyc (Leptospira interrogans L1-130) (Prokaryotes Bacteria Spirochete)
		asp76114cyc (Azoarcus sp. EbN1) (Prokaryotes Bacteria Proteobacteria Beta)
		mmaz192952cyc (Methanosarcina mazei Go1) (Prokaryotes Archaea Euryarchaeota)
		bxen266265cyc (Burkholderia xenovorans LB400) (Prokaryotes Bacteria Proteobacteria Beta)
		frantcyc (Francisella tularensis tularensis SCHU S4) (Prokaryotes Bacteria Proteobacteria Gamma/others)
		vchocyc (Vibrio cholerae O1 biovar eltor str. N16961) (Prokaryotes Bacteria Proteobacteria Gamma/others)
		hsom205914cyc (Haemophilus somnus 129PT) (Prokaryotes Bacteria Proteobacteria Gamma/others)

**Figure 5**

**Figure 3**



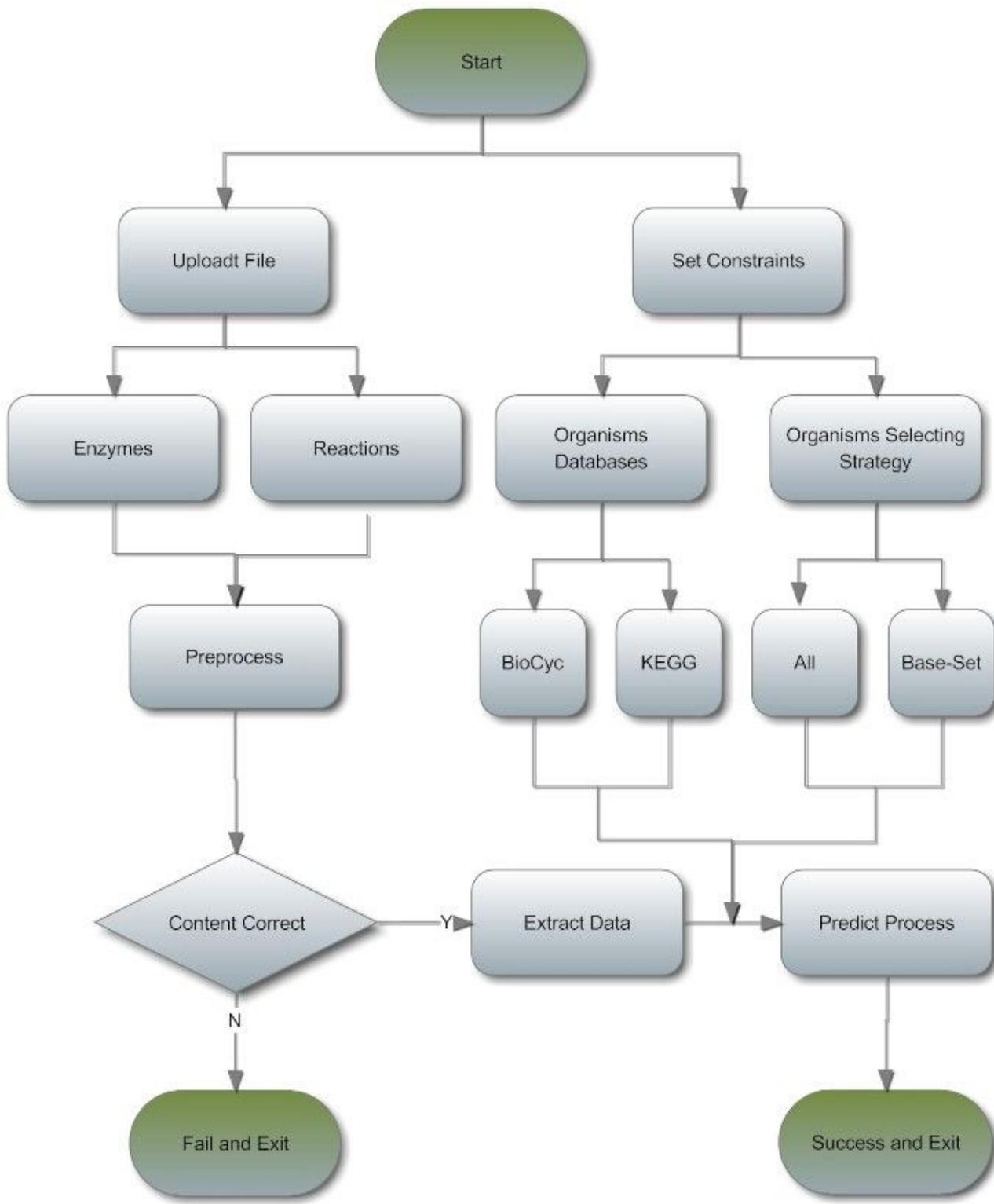


Figure 7

Figure 1

## Predict Configurations

**Select the data type**  reactions list  ec list

\* The data that you upload for predict is either a list of reactions or ec numbers.

**Upload your data**   (Example)

**Select the database**  KEGG  BioCyc  Both

\* Select the database that you want to use for predict. There may be some differences in organisms number or metabolite networks of each organism between the two databases.

**Organisms Limits** Using  All  Base Set of the organisms

\* Predict the Bacterial Community using all or the base set of the organisms that exist in our database.

Figure 8

Figure 2