Supplementary Figures

ata Selection										
NPS Data Se	lection - Enter USI or S	Spectrum Peaks								
Spectrum USI	mzspec:GNPS:GNPS-LIBR/	ARY:accession:CCMS	SLIB000000	85687	7					
	Enter one peak per line a	as follows:								
Spectrum Peak	m/z1 intensity1 m/z2 intensity2 m/z3 intensity3 									
Spectrum Peaks Precursor m/z	m/z1 intensity1 m/z2 intensity2 m/z3 intensity3 				Charge	charge				
Spectrum Peaks Precursor m/z PM Tolerance (I	m/z1intensity1 m/z2 intensity2 m/z3 intensity3 precursor m/z	ment Tolerance (Da)	0.05		Charge Cosine Three	charge shold 0	7	Minimum Ma	atched Peaks	3

Supplementary Figure 1. microbeMASST web app

Users can access the microbeMASST web app at <u>https://masst.gnps2.org/microbemasst/</u>. They can search for MS/MS spectrum against the microbeMASST reference database by either providing a USI in `Spectrum USI` or by inputting fragment ions and their intensities and the precursor mass in `Spectrum Peak' and `Precursor m/z` respectively. Search parameters, such as parent ion and fragment ions tolerances, cosine threshold, and minimum matching peaks can be modified. Analog search can also be enabled. Finally, to submit a search query users have just to click either on `Search microbeMASST by USI` or `Search microbeMASST by Spectrum Peak` based on the information that they have provided. Search jobs can be easily shared by clicking on `Copy Link`.

а

Showing 1 to 4 of 21 entries

Library matches Dataset r	natches Taxa matches Para	meters						
Copy Excel CSV Show	4 v entries							Search:
Delta Mass GNPSLibr	iryAccession	USI		Charge	Cosine	Matching Peaks	Adduct	CompoundName
0 CCMSLIBI	0005435750	mzspec:GNPS:GNPS-LIBHAH mzspec:GNPS:GNPS-LIBHAH	Viaccession:CCMSLIB00005435750	1	1	9	M+H M+H	Yersiniabactin
	4			¥				
Latint Ubrary Spectrum Information			GNPS Metabolomes Uti					
Spectrum ID Compound Name	CCASLIBOOD Verstnadu	5435750 actin						
Pl Data Collector	Pieter Dorr Julio Gau	estein agitz	USI Data Selection	Copy Link	Data Visualization			
Pubmed			Spectrum US2 mzspec:GNPS:GNPS-LIBRAR	accession CCMSL80000		versal Spectrum Identifier 1 mcspec/6 SPLASH Identifier 1 splash10-	NPS-GNPS-LIBRARY.accessio 2002-0690000000-d363/w94	erCCMSUB00005435750
Original Submitter	Innered		Spectrum US: mztiple: GNPS-SNPS-UBRAR	accession CCMSU80000	and the second s	versal Spectrum Identifier 2 misspecifi SPLASH Identifier 2 splash0-	NPS-GNPS-LIBRARY accessio 2002-069020020-it353hr94	erCCMSLIB00005435750 Emilie IsBeetila0188a
Aust Recent Revisor	Gold Spec	la drum	Desite Desite			Download as JSON Download a	s PNG Download as SVG	
Section Inchi-15/C2	CC1252(cr)(cr)(cr)(cr)(cr)(CSC)(qr)(CSC)(qr)(cr)(cr)(cr)(cr)(qr)(qr)(qr)(qr)(qr)(qr)(qr)(qr)(qr)(q		Drawing Controls	pres.	Top: map	e:GNPS:GNPS-LIBRARY:accession:	CMSLIB00005435750	
			Figure Size 10	in X 6 in	Bottom: mzs	Bottom: mzspec:GNPS-CIBRARY: accession:CC Precurror w/r 482-1310 Charge 1		
			Mass Range mm s.	myz - mass myz		Cosine similarity = 1.0900		
			Maximum Intensity 125	*	3875 -			
			Label Precision 4	decimats		1		
			Label Rotation 90					
	10 10	T HO	Cosine travel	nd .	1	+		
Structure	+, 5	MAG	Fragment Tolerance	0220				
			6.1	and a	-			
	CH	•	Show und Yas					
			UI Adjustment		- in		#0 H28	800
			Selection Sine					
Escherichia_coli_MS_200-1 Escherichia_coli_Nissle_1917	749550 316435	strain 0.94 strain 0.92	8 <u>mzspec:MSV0000</u> 7 <u>mzspec:MSV0000</u>	82045:Aerobic_BHI : 83383:std_mix3_Fe_	356 BA4 01 22675:sc conc46 2 long:scan:19	an:1221 810	/	MSV000082045 Open MSV000083383 Open
Escherichia_coli_Nissle_1917	316435	strain 0.92	7 mzspec:MSV0000	83383:std mix3 Fe	conc46_1_long:scan:18	98	/	MSV000083383 Open
Showing 1 to 4 of 129 entries			/			/		Previous 1 2 3 4 5 3
			*			+		
SGNPS Metatologies USI				MassIVE MSV	/000083383	Not Public		
USI Data Selection	Data Visualiza	ation		GNPS - E. Coli Nissle v	wild type versus knockout	strains grown with altered zi	nc levels	
Sector (A) anne (M/S/M/S) (8947	ATTENING COMPLETION	SC-S Universal Spectrum Identifier 1 missour ONP	S ONPS-LIBRARYaccession CCMSL/800005435750	Salacrite Convert Arana	Add Researcy Add Researcy R			
		SPLASH identifier 1 splash/0-000 Universal Spectrum identifier 2 mzspec/MSV	12-069000000-c0531e96b8eed8e086088e 000083383.std_mir3_Fe_corc46_3_torg.scan:1892	Description				Annaly sector and
Spectrum USI matpec MSV000083383 etd.	nic), Fe_core4(L3_lorg)	Download as JSDN Download as P	NG Download as 0/0	Metabolomics data from E. C and enriched M-9 media. [do	oli Nissle (wild type, WT) and 3 E. G k:10.25345/C5862F] [dataset license	ki Nissle knockout (KD) strains grown in M- CCD 1.0 Universal (CCD 1.0)]	Pinedia	Total Size: 99.50 GB Spectra: 195,918
Drawing Controls	and a second sec	Top: maspec:GNPS:GNPS-LIBRARY:accession:CC	MSLIB00065435750	Keywords: Zinc ; E. Coli Niss	sie -			Subscribers: 1
Figure Size 10	m X G m	Precursor w/g: 482-1310 Charge: Bottom: mzspec:MSV000083383stid_mit/3_Fe_cone4 Precursor w/g: 482-1254 Charge:	1 46_3_long:scan:1892	Contact			_	Owner Reanalyses Identification Results Proteins (Reported): O N/A N/A
Mass Range min n	m(z - max) m/z	Cosine similarity = 0.9843		Principal Investigators: Piet (in alphabetical order) Submitting Liser: aller	ter Dorrestein, University of Californi	a San Diego, USA		Peptides: O N/A N/A Variant Peptides: O N/A N/A
Maximum Intensity 125	8	g Î						PSMs: O N/A N/A
Label Precision 4	decimals	91 						FTP Downland Link (click to copy):
Label Rotation 10	<u> </u>							(fg://massive.ucid.edu/W5/000083383/
Cosine Stand	d .							
Enagment Tolevance 0.1	mir	100						
Show Grid Yes		*						
UI Adjustment	-							
Selection Size		at 10 50 40 41						
ata Exploration								
Data Exploration								
Data Exploration	atches Taxa matches Paran	neters						Search
Library matches Dataset m Copy Excel CSV Show Name	atches Taxa matches Paran 4 v entries NCBI	neters Rank			Matches	_ Samiles		Search:
Library matches Dataset m Copy Excell CSV Show Name	atches Taxa matches Paran 4 orinies NCBI 131567	neters Rank cellular organ	iisms	3	Matches 129	. Samples 60306		Search: Fraction 0.0021
Data Exploration Ubrary matches Dataset m Copy Excel CSV Show Name Foot Prokaryote	atches Taxa matches Paran A v entries NCBI 131567 2_placeholder	neters Rank cellular organ superkingdon	nisms n_placeholder	3	Matches 129 110	 Samples 60306 44836 		Search: Fraction 0.0021 0.0025
ata Exploration Library matches Dataset m Cegy Excel CSV Show Name Prokaryote Bacteria	atches Taxa matches Paran 4 ventries NCBI 131567 2_placeholder 2	eters Bank cellular organ superkingdon superkingdon	nisms n_placeholder n	3	Matches 129 110 110	Samples 60306 44836 44836		Search: Fraction 0.0021 0.0025 0.0025

Supplementary Figure 2. Complementary output of microbeMASST

a) The queried MS/MS spectrum is searched against the GNPS libraries and, if matches are found, possible annotations are returned. Information on cosine similarity and number of matching peaks is provided and users can explore the associated GNPS Library Spectrum page and inspect mirror plots. b) Information of matching scans in the sample from the different taxa is provided. In addition to cosine similarity score, matching fragments, and the possibility to inspect mirror plots, users can retrieve the MassIVE ascension number of the project together with contact information on who deposited the data. c) Number of matches for each taxonomic level are returned. Users can observe how many samples for that specific taxon are part of microbeMASST and see how broadly the molecule is distributed within it (fraction).

Previous 1 2 3 4 5 6 Next



Supplementary Figure 3. Examples of microbeMASST outputs

Additional examples of results obtained from searching from <u>mevastatin</u>, <u>arylomycin A4</u>, <u>yersiniabactin</u>, <u>promicroferrioxamine</u>, <u>glutamate-cholic acid</u>, and <u>glutamate deoxycholic acid</u>. In all cases the molecules were found in the monocultures of known producers and not in human cell lines, confirming that microbeMASST can be used to search for microbial-derived molecules.



Supplementary Figure 4. Distribution putative microbial metabolites across phyla

a) Of the 2,425 MS/MS spectra that had a match exclusively to microbial monocultures, 1,673 were found only in bacteria, 95 in fungi, and 657 in both. b) Chemical classes were then predicted using SIRIUS and CANOPUS and their distribution across the different phyla was visualized using an Upset plot.



Supplementary Figure 5. Contextualization of the molecular network

The selected 512 MS/MS spectra were mapped back to the full molecular network generated from the data acquired from GF and SPF mice (MSV000079949). Only molecular families containing at least one of the MS/MS spectra of interest were retained and used to generate putative annotations of the unannotated spectra of interest. Each node represents a spectrum (ion), light blue indicates MS/MS of interest while green indicates spectra that were not retained in the performed downstream analysis. Within each node the m/z of the precursor ion is indicated, while on each edge the m/z difference between two connected nodes is reported. a) lons with precursor mass 318.181 and directly connected to a spectrum annotated as the dipeptide Trp-Phe (m/z 352.144) can be tentatively annotated as Trp-Iso/Leu. b) lons with precursor mass 423.362 connected to Soyasapogenol C, which can represent a soyasapogenol molecule with a loss of water (delta mass 18.01). Abbreviations: Trp, tryptophan; Phe, phenylalanine; Iso, isoleucine; Leu, leucine.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryFigure1.ai
- SupplementaryFigure2.ai
- SupplementaryFigure3.ai
- SupplementaryFigure4.ai
- SupplementaryFigure5.ai
- SupplementaryTable1.csv
- SupplementaryTable2.csv
- SupplementaryTable3.csv