# Rapid Microorganism Identification Database (www.RMIDb.org)

ASMS 20<u>06</u>

# Nathan J. Edwards<sup>1</sup>, Fernando Pineda<sup>2</sup>

<sup>1</sup>Center for Bioinformatics & Computational Biology, University of Maryland, College Park; <sup>2</sup>Molecular Microbiology & Immunology, Johns Hopkins Bloomberg School of Public Health.

#### Overview

# Public Web Interface

#### www.RMIDb.org

- Comprehensive bioinformatics platform for microorganism identification by mass spectrometry of proteins and peptides.
- Supports intact protein and trypsin or acid digest product protocols.
- Facilitates exploratory protocol development.
- Private guest and registered user workspaces.
- Sound statistical significance estimation.

## Introduction

Statistically significant rapid microorganism identification using mass spectrometry and protein sequence databases has been demonstrated for various protocols, such as intact proteins from bacterial cells (1) and tryptic (2) and acid (3) digest products from bacterial spores and viruses (4).

Protocol development for rapid microorganism identification requires the determination of preferentially observed proteins or peptides (*"identification model*") and an appropriate method for estimating the *statistical significance* of mass matches to the model (5).

The Rapid Microorganism Identification Database (<u>www.RMIDb.org</u>) provides each of these informatics needs in an easy to use web interface.

	Chi Rapid Mitrae			interteen 1	Accella l'Inch								- 1 <b>1</b>	New Model				
	* ** ** *												0	New Model				(c)
	4-50	1 D M	Seat. a	yi yi kededa	d.						Do G		-	Step 1.	Tatometer	474		
RMIDb: Rapid Microorganism Identification Database (a)												Select the properties of model premises at right Dep 2. Select a class property to group and	1.44ammalian type C-inus golug (*1836) 19354 (Peanut WB group) (#621) 19358 (Pispen pas witches' forcer grou 19357 (Pispen pas witches' forcer grou 19357 (En ywlaws group) (#627)					
Phone    Macant Selectors    Fersate   Organisme: 27922   Species: 15660											partness model proteter and ther	Species chronic property proces		-				
Hun	m    Model Se	lection 31 Fe	unte					Searces	Gebak	Atlins Co	27922 (Spec a) Post Tol (** 183441) (Second) (	8. Tri	10673	denned humarkers Proceed andy Jac General-Missing Cryptone Angel Sumarkers	Decilies as 158-2 (7558) Decilies as 158-2 (7558) Decilies as 158-2 (7557) Decilies as 158-2 (7257) Decilies advances (167) (755556 advances (167)		1	
	vailable Models										Epeciet Teronomy	Organiza						
		ners date_1_lder	213	Non)								New	vModel	Day 4. Land the mass range of humanitory	Fis-obscherunt lutercene (25) Frunst emeriese' phytoplasme (269) 2.40 degrading backmum TFD21 (034) 2101252 view (17785)		8	
relect	Interest	dan def	MRC	Mahrin	ManMen	MelTM	Protocol	-	dames	To distant	dertinet .		all a second second	Bron 4000 to 15000 Da	28/263t veus (172880) GenerOutsbary	sthi		
	Failles rubble Fromas	GeorGanilig		4000.0	0 15000.0	Lavy	Ho Digit	2006-05-25	400	47%	2745	5	5851	Dep 1. Manue the model	the neuron MP booystheses (SO 2008118) the neuron potentine base boourheses in			
	Bubcoomal Promine [*]	Species	21	4000.0	0 15090.0	Larty	He Digest		307	201	10764		45879	Encilies sublic Pro	1-ekyl 2-ecetylytarophosphosholine i Sphosphatetylarophili-Hybrophate 1-in 2-2-ocio-spheritety 2-shoophodester	10		
	TATE	Specara		400.0	0 3000.0	Larty	Trypoin Digest		39	12%	1212	2	15330	296	Protein Funde:	11	-	
	Tupsdar (*) defaed no kie						(lip to 1 barred cleanage)				(t	<b>c)</b>		Subset the query County Model	Shrinna' (CHRinamatin Departmation MCo Cold-shook' (Drivineding domesn (PPI)) Plained box' domain (PPI)(CSQ) ((Physhesphor) Availatione synthesis (C Linkson, Orcholate Article and well with	212) mAi (PP0007%)	8	

**Figure 1:** (a) <u>Comprehensive</u> bacteria & virus protein database; (b) System and <u>user defined</u> identification models; (c) New model interface.

Medel Name: EASP Proteins Jaco Definition: Species Samber of Classes: 39				(8	Figure 2: (a) <u>Sequence unique</u> SASP							
Details						peptides; (b) Significant GO categorie						
elect class			irtigert haven in Korey all hirese			peptides, (b) <u>sign</u>		J	$\pi$		alcyon	
Bacilus bal-darans (2005)	89%	74	10	309		for B. subtilis prote		_				
Alkala bibu metaline figment (200226)		67	78	156		TOP B. SUDTINS PROTE	eır	าระ				
3 3 acidos megatemas (1404)	34%	42	74	129			••••	,				
[] Bacilus robilis (1422)	34%			2012	-		<b>~</b> (	$\sim$	Cir	anifi	cont	
E Facilias corres (1795)	33%	Inde	ntification Results				1	C)	310	יו ווו ון	cant	
Controllous actroloutyleous (1422)	305		Nume Barilla editle Insta			(b)	1.1	11				
Crobeche keutophie (1452)	34%		Definition: DeseCeptings			(U)	110	dei	ntit	ticat	ion of	
Carborp Jefferman hy kegres-fermane (129950)			e of Classer 480			• • •	1 -	uu	1.01	icut		
D Themasawolacter Ingringetau (112122)	57%						Шr	<b>.</b> .		A1114	date of the second s	
2mbol.acternas Berts-philes (238)	100%	Dete	8					5. S	suc	ıbtilis by		
Denilli-bacterium hallience (45232)	100%	No. of Concession, Name										
Comises termovilus (212)	. 92%	select				hismarkers peaks matches e-value e-value	l r	ribosomal protei				
Controllous perfingent (1502)	100%		rd-inteler (00.0005840)			54 24 11 0.0136 0.00906		IDC	720	illai	protein	
D. Rentes dans (1999)	- 896		structural construent of silicous	(200007725)		54 24 11 0.0136 0.00906					· · · · · · · · · · · · · · · · · · ·	
			organzle (00/0043226)			55 24 11 0.0148 0.0115						
			rd-maile operations compilers (28			Indentification Results						
			envellier wa neutron-to	and organile (120,004322)	ð						2 X	
			atraceblar organile ((20.004			Madel Nume: Educational Proteins					(c)	
			nie mestene bond opnel			Class Definition: Species Number of Classer: 307					(0)	
			protein Toi conthe la (22,00%	61120		Number of Classer 307						
			cytoplass (30-0015737)			Densis						
			structural molecule activity (22)				A sea in	-	Arrest starts	duradus same	And a local diversity of	
			macrosulierale biosynthesis 🖸	k2.00220180		E Bacha salata (1623)	54	24	11	0.00047	133+06	
			rellalær protess metabolisen (28	2,00442420		C Bacillar International (1402)	- 0	1		13.5	0.649	
			section aretabulkers (OO 1002)?	53(0)		E Fachettra typis (285)	30	24		21.3	304	
		Dana		111		Oevhatiba stravitlemophila (1422)	32		-	255	3.97	
						<ul> <li>Vehatia information (1902)</li> <li>Wehatia information of Decould annumer (2005)</li> </ul>	32	28	-	251	3.97	
						Xethonias synar (HD)	n	24		27.8	4.43	
An Arrest of Arrest and Arrest of Ar					a second second	D Dervedan 274 02250	- 24	34		40.5	10.0	
dats mass sequence + start e	ad estemps	accession	description	ergenitte	species	Acartolator m AS91 (82970)	41	24	-	46.1	8.13	
Bacillus robsilas (1422) 691 82 001.83 43	40 5	+ with TTU 18:000044	mail acid-mithle spore	Dacibu rebile rebep.	Decilia ref.	Accentitian en ACV1 (02770)     Tectobacterian atmospheram (09770)	40	24		46.1	9.71	
			protes (aphabeta-type	ruballe ex: 163 (224)(1	0		43	54		55.0	9.71	
			TATP)-related protein			Comises perfagess (1202)		28				
752.27 VP201.76 18	25 67	4 10/10/10/20054	reop stday-bea ham	Dardho rabile nabep		Hattpldes tones (21)	48			71.0	14.2	
			prome (alpha/teta-type SASP)-eslated prome	ndelle m: 162 (22400)	9	D Torposense palleken (163)	32	- 24	3	72.0	21.5	
801.89 TVTTR 27	12	A NOT TRADUCT	mal and robble more	Tacille robile robes	and the second	Bddoebeic bacterioweau (22)	33	24	3	76.3	217	
101 (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c		· BOLIVIALOVICS	proton (alpha/seta-type EASP)-related protein	ndsla m 163 (22430)			34	54		85.4		
S31 BE ADDRESS 1	1 10	MERCANCES	Small, and exhibit sport protein C (SASP)	Bacillor extends (1423)	Decides rel							
		#16029053 utp:0:309076.1	mall acid-midde sport protein (alghafteta-type SAS		Þ							
		#1522415	ukares	Bacalhor militale (1422)		Figure 2. Motobo	4	/	nic			
		101701332345	renall, acid ordable spore protein 2 ordable protein	Recibio ministe minep- ministe en 143 (22430)	9	Figure 3: Matche				•		
		WHITE ALE STOPP	mall and soldale spore protein (alphabeta type SA)		Description of	biomarkers. Sort k						
200.91 Aministrice 1		NETTIN NETTIN NETTIN	Small, acid-soluble spore protein gamma type (SASP) Small acid-soluble protein	Facilies extelles (1422) Facilies extelles (1422)		columns may be r cells indicate repe						

### Statistical Significance

The statistical significance of microorganism identifications are commonly estimated using the method of (5). This model based approach permits an explicit *p*-value computation, but requires problematic assumptions:

- Protein masses are uniformly distributed,
- Input masses match independently.

We instead randomly sample appropriately sized sets of (sequence distinct) biomarkers without regard to the class labels, and count the number of samples that match at least as many peaks as the class of interest. We draw samples until the *p*-value estimate is within 10% of its true value, with 99% confidence. Finally, we report *e*-values to correct for multiple-testing.

#### Conclusions

The Rapid Microorganism Identification Database (<u>www.RMIDb.org</u>) provides a flexible, easy-to-use bioinformatics platform for microorganism identification. The RMIDb facilitates protocol development by supporting user-defined identification models. A new statistical significance model eliminates problematic assumptions of previous methods.

#### References

- 1. Pineda, Antoine, Demirev, Feldman, Longenecker, Lin. Anal. Chem. 2003.
- 2. English, Warscheid, Fenselau, Cotter. Anal Chem. 2003.
- 3. Yao, Demirev, Fenselau. Anal. Chem. 2002.
- 4. Swatkoski, Russell, Edwards, Fenselau. Anal. Chem. 2006.
- 5. Pineda, Lin, Fenselau, Demirev. Anal. Chem. 2000.