# Autoantibodies and anti-microbial antibodies: Homology of the protein sequences of human autoantigens and the microbes with implication of microbial etiology in autoimmune diseases

**Peilin Zhang, MD., Ph.D.** PZM Diagnostics, LLC Charleston, WV 25301

Correspondence:

Peilin Zhang, MD., Ph.D. PZM Diagnostics, LLC. 500 Donnally St., Suite 303 Charleston, WV 25301 Email: <u>pzmllc@gmail.com</u> Tel: 304 444 7505 bioRxiv preprint doi: https://doi.org/10.1101/403519; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

### Abstract

Autoimmune disease is a group of diverse clinical syndromes with defining autoantibodies within the circulation. The pathogenesis of autoantibodies in autoimmune disease is poorly understood. In this study, human autoantigens in all known autoimmune diseases were examined for the amino acid sequences in comparison to the microbial proteins including bacterial and fungal proteins by searching Genbank protein databases. Homologies between the human autoantigens and the microbial proteins were ranked high, medium, and low based on the default search parameters at the NCBI protein databases. Totally 64 human protein autoantigens important for a variety of autoimmune diseases were examined, and 26 autoantigens were ranked high, 19 ranked medium to bacterial proteins (69%) and 27 ranked high and 16 ranked medium to fungal proteins (66%) in their respective amino acid sequence homologies. There are specific autoantigens highly homologous to specific bacterial or fungal proteins, implying potential pathogenic roles of these microbes in specific autoimmune diseases. The computational examination of the primary amino acid sequences of human autoantigens in comparison to the microbial proteins suggests that the environmental exposure to the commensal or pathogenic microbes is potentially important in pathogenesis of a majority of autoimmune diseases, providing a new direction for further experimental investigation in searching for new diagnostic and therapeutic targets for autoimmune diseases.

bioRxiv preprint doi: https://doi.org/10.1101/403519; this version posted August 29, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

### Introduction

Autoimmune disease is characterized by the presence of circulating autoantibodies. Some autoantibodies, for example, anti-smith antibodies, are more specific and pathognomonic to a specific autoimmune disease such as systemic lupus erythematosus (SLE)(1, 2). Others such as anti-nuclear antibody, are non-specific and present in many clinical diseases or even normal healthy individuals (1, 2). The questions of how and why these autoantibodies are generated in patients remain largely unanswered. Microbiome study demonstrated the presence of trillions of various microbes within the body, and there is an intimate symbiotic relationship between these microbes and the human host in various aspects of human tissue and organ functions (3-5). In addition, there are numerous species of microbial DNA within the blood circulation without the culturable microbes (6). The presence of the microbial DNA in blood without culturable microbes raises two possibilities: the blood culture methods are insensitive to detect the microbes but the microbes are present within the circulation (7), or alternatively the intact microbes are destroyed by the host immune system but the microbial DNA and proteins are present in the circulation, eliciting the human immune response to these microbial DNA/proteins. The human immune responses to the microbial DNA and/or proteins may manifest as anti-microbial antibodies with potential cross-reactivity to human tissues through molecular mimicry (8, 9). In the process of identifying the infectious agents in Crohn's disease and Sjogren's syndrome, we have identified a group of microbial proteins from the commensal or pathogenic microbes reacting to patients' plasma, and we showed that there were elevated antibodies within the circulation of patients reacting to the microbial proteins (8). Furthermore, we showed that there is a cross reactivity of anti-microbial antibodies to the human tissues. It is reasonable to assume that the anti-microbial antibodies in response to the microbes, either pathogenic or commensal, can be detected as autoantibodies in autoimmune diseases, given the presence of cross-reactivity of anti-microbial antibodies against human tissues (8).

In current study, I took one step further to examine the primary amino acid sequences of all known human autoantigens in autoimmune diseases and compared these amino acid sequences to the microbial proteins including the bacterial and fungal proteins,

using information from the Genbank and BLAST search tools publically available at National Center for Biotechnology Information (NCBI)

(https://blast.ncbi.nlm.nih.gov/Blast.cgi). Our results are surprising and more than two third of the human autoantigens important for a variety of autoimmune diseases are homologous to the microbial proteins including bacterial and fungal proteins. These microbial proteins may elicit the human antibody responses with potential cross-reactivity to the human tissues, leading to specific human tissue damage and autoimmune diseases. The presence of anti-microbial antibodies in circulation in autoimmune diseases suggests a potential new mechanism of autoantibody production and autoimmune diseases. The computational analysis of the primary protein sequences for homology represents the initial step toward understanding the production of autoantibodies in autoimmune diseases.

## Methods:

All known human autoantigens associated with human autoimmune diseases are listed in Table 1. The protein (amino acid) sequences of these human autoantigens were searched from the Genbank (https://www.ncbi.nlm.nih.gov/protein/?term=), and the specific amino acid sequences and/or specific Genbank accession numbers were used for BLASTP search against the bacterial database (Bacteria (taxid:2)) and fungal database (Fungi (taxid:4751)) using the previously defined search parameters (default parameters). The homology was ranked as low, medium and high using previously defined (default) parameters, and denoted in color from NCBI (blue - low, mediumpink, red — high) (Figure 1). Screenshot photos were taken to demonstrate the homology between the human proteins with the various microbial proteins with colored graphical illustrations and the specific amino acid alignments between the human autoantigens and the perspective microbial proteins (bacterial and/or fungal). The specific homology scores from the default NCBI search algorism including compositional matrix adjustment, positive identity and gaps were not used because of the lack of expertise from the author's perspective. The low, medium and high homology definitions based on the colored coded graphics were used to estimate the similarity of

two protein sequences from two separate species. Unique specific microbes were also denoted in Table 1 for specific homologous human autoantigens. Each and all human autoantigen with their respective search data are illustrated in a book to be published later (in preparation). The current writing is the summary of all the search data with a goal to point to an entirely different direction for further experimental investigation for the mechanism of autoantibodies and autoimmune diseases.

## **Results:**

## 1. Genbank search, definition of low, medium and high homology between the human autoantigens and the microbial proteins

There are three examples of BLASTP searches for definitions of low, medium and high homology between the amino acid sequences of human autoantigens and the microbial proteins (Figure 1). Using the Genbank accession number for each individual human autoantigen (SSA52/Ro52, Genbank accession number AAA36581, and Histone H2A, Genbank accession number P20671, anti-Jo1 autoantigen, Genbank accession number P07814) and BLASTP search, three matched results were shown in Figure 1. The low, medium, and high homologies between the SSA52/R052, Histone H2A and Jo1 and the microbial proteins were illustrated by blue, pink and red colored graphics followed by the amino acid sequence alignments with the respective microbial proteins from specific bacterium or fungi. It is worth noting that antibody production requires a small stretch of amino acid (epitope) within an appropriate antigenic structure. The examination of the amino acid sequences can only show the homology and alignment between the human proteins and the microbial proteins without predicting the threedimensional structure. As a general rule, the experimental significance of this homology search is limited to the knowledge of computational prediction, and the predictive information requires vigorous experimental validation to be clinically relevant. The importance and clinical significance of these human autoantigens in the specific autoimmune diseases is beyond the scope of this study.

## 2. Microbial proteins reactive to human plasmas in Crohn's disease and Sjogren's syndrome are highly homologous to human protein homologues:

We have previously identified a panel of bacterial proteins reactive to human plasmas from patients with Crohn's disease and Sjogren's syndrome, and we demonstrated that there were elevated anti-microbial antibodies within the circulation of these patients (8). We also demonstrated that the specific antibodies against the microbial proteins produced in vitro can cross react to human tissues. Currently, the amino acid sequences of these microbial proteins were used to search Genbank human protein database to compare the microbial proteins with human proteins (Figure 2). This is a reverse exercise of the methods in Figure 1. The query sequences from the bacterial proteins were used to search the human protein database (Homo sapiens (taxid:9606)). The homologies between the microbial proteins EF-G, ATP5a from *Staphylococcus* aureus /pseudintermedius (10), Hsp65 from mycobacterium avium subspecies hominissuis/Mycobacterium tuberculosis, and EF-Tu from Escherichia coli to the human proteins were significantly high, and the antibody against the microbes can cross react to the human tissues as shown previously (8). The clinical significance of these microbial proteins and their respective antibodies within the circulation in Crohn's disease and Sjogren's syndrome have been previously discussed (8, 9).

## 3. Homology of all human autoantigens to microbial proteins

Using the same principle and search methods with identical search parameters, the known human autoantigens important for human autoimmune diseases were examined, and the results were listed in Table 1. Totally 64 protein autoantigens in a variety of autoimmune diseases were examined against the microbial protein databases including bacterial database (Bacteria (taxid:2)) and fungal database (Fungi (taxid:4751)). There were 25 autoantigens highly homologous to the bacterial proteins, 29 highly homologous to the fungal proteins. 19 human autoantigens showed medium homology to the bacterial proteins and 16 to fungal proteins. Two examples of high protein sequence homology between the human autoantigens and the microbial proteins

(bacterial and fungal) are shown in Figure 3. Combining the high and medium homology groups, there are 68.8% of human autoantigens showing medium to high homology to the bacterial proteins and 69% to the fungal proteins. Phospholipids can also be antigenic in autoimmune diseases such as Guillain-Barre syndrome or anti-phospholipid antibody syndromes, but the phospholipids serve as haptens in antibody response, and the haptens usually need to combine with carrier proteins to be antigenic (11). The phospholipids can derive from plasma membranes of eukaryotic or prokaryotic cells, and it is difficult to determine the sources of these phospholipids in these patient populations. It is also unclear if the patients with anti-phospholipid antibodies carry other circulating anti-microbial antibodies against other microbial proteins.

There is other important information that is medically relevant to clinical management of a variety of patients with autoimmune diseases, and this information requires experimental validation. Human autoantigen U5 ribonuclear protein in SLE is highly homologous to that of the *Candida albicans*. Human proinsulin sequence is homologous to a specific hypothetical protein of *Enterococcus faecium (12)*. Histidine tRNA ligase of anti-Jo1 antibody important for inflammatory myopathy is highly specific to that of *Clostridium*, and the antigens from anti-smooth muscle antibodies in autoimmune hepatitis were found to be specific to those of *Escherichia coli*. Specific clinical management plan can be devised if these relationships between the specific autoantigens and specific microbial proteins are experimentally validated.

Among the autoantigens highly homologous to the microbial proteins (red colored), most commonly seen were enzymes with catalytic functions in both human host cells and the microbial cells. The structural proteins in cytoskeleton are also common, and these structural proteins are well conserved across the microbial or all species with important functions in cell division and cell mobility (Table 2). Others proteins with various functions such as nucleoproteins, regulatory proteins, immune related proteins and ion channels are also identified. It is conceivable that the human immune responses to these microbial proteins lead to antibody production and these antimicrobial antibodies will cross-react to the human tissue through molecular mimicry, leading to collateral human tissue damage.

## **Discussion:**

It is well known that the B-cells and plasma cells are critical for autoimmune diseases (13). Through examination of all human autoantigens against the bacterial and fungal protein databases, approximately two third of the human autoantigens are significantly homologous to the microbial proteins including bacterial and fungal proteins in their respective primary amino acid sequences. The highly homologous protein sequences between the human host and the microbes suggest a reasonable probability that the autoantibodies in autoimmune diseases are derived from the host immunity against the microbes present in the human body, commensal or pathogenic, bacterial or fungal in origins. It is now known that human host is colonized by trillions of commensal microbes and these commensal microbes are intimate components of human development after birth (3-5). Exposure to environmental microbes including bacteria and fungi help develop normal immunity to prevent pathogenic infections, regulate human metabolic activity and various tissue functions (3-5).

It is noteworthy that no viral proteins are examined for autoantigens as the viral immunity is vastly different from the bacterial or fungal immunity. Anti-viral antibodies are protective against the subsequent viral infection, forming the basis of modern medical vaccination. However, exceptions are well-known that the viruses can evade the human immune system by hiding in the intracellular compartments and reactivated in response to the body stress conditions or immune compromised conditions. The examples of reactivation of Epstein-Bar virus (EBV) and herpes simplex virus in immune compromised patients are well-documented, although the mechanism of reactivation is yet to be proven. Specific anti-viral antibodies with cross reactivity to human tissues remain to be a possible mechanism for autoimmune diseases.

Production of antibody in vivo in response to a specific antigen is well studied, and a large scale production of antibody for pharmaceutical industry is performed routinely. However, the mechanism of the production of specific autoantibody in vivo, such as anti-CCP antibody (anti-CCP, cyclic citrullinated peptide, ACPA) in rheumatoid arthritis (RA) is still intriguing. Citrullination or deimination of protein is to convert the aiming acid arginine to amino acids citrulline. Citrulline is an unusual amino acid and it is not one of the standard 20 amino acids encoded by DNA in the genetic code. Citrullination is the result of post-translational modification by a group of enzymes arginine deiminases or peptidylarginine deiminase (PADs). Deimination of proteins will change the hydrophobicity since arginine is positively charged and citrulline is not, and the protein Citrullination leads to abnormal folding of proteins and its functions. Citrullination of proteins will also induce abnormal immune response by generating anticitrullinated protein antibodies, leading to autoimmune disease such as RA and multiple sclerosis (MS). There are multiple PADs in human with enzymatic deiminase activities. Human PADs distribute in a variety of tissues in a tissue specific fashion, and these enzymes likely play important roles in cellular functions and signal transduction. There are also many bacterial deiminases that share significant homologies with human proteins. Bacterial deiminases, from either commensal or pathogenic, can potentially citrullinate human proteins, leading to abnormal citrullinated proteins to induce human immune response and autoimmune diseases. Alternatively, bacterial deiminases can directly induce human immune response upon entering the human body, and antideiminase antibody generated against the bacterial deiminases can cross-react to the human enzymes (deiminases), leading to autoantibodies and autoimmune diseases, although neither the reports of anti-deiminase autoantibodies in human diseases are documented nor any effort is made to discover anti-deiminase antibody in autoimmune diseases. However, in mouse model, a recent study demonstrated that the immunized mice with the murine or the human PAD2 or PAD4 enzyme can generate anticitrullinated peptide anybody, and these antibodies against citrullinated peptides only occur with the enzyme (PAD) bound to the targets (14). It appears that the immunogenic epitope is the peptide-enzyme complex. Furthermore, the bacterial PAD from porphyromonase gingivalis (periodontitis) is important in periodontal disease and it

is also important for RA (15). At this point, the data to support the presence of antibacterial PAD antibodies or anti- PAD autoantibodies in autoimmune diseases are lacking, although it is plausible with a new mechanism of disease process.

The present study of the autoantigens and the microbial proteins is a computational prediction and requires vigorous experimental validation. Specific epitopes eliciting antibody productions in vivo are usually small peptides, and BLASTP analysis could miss some small areas of amino acids that are antigenic. Antigenicity of any given protein is not random. Epitope mapping of human autoantigens may help answer these questions. Combination of direct comparison of the human autoantigens and the microbial proteins with specific antigenicity mapping can help anticipate the epitopes of any given proteins and facilitate the specific epitope mapping. These computational tools are widely available in public and make it possible to study the specific autoimmune diseases for better diagnostics and therapeutics. Many microbial proteins identified by the BLASTP with homology to the human autoantigens are hypothetical proteins from the known or unknown microbes, and significant effort is required to characterize these unknown microbes (bacteria or fungi) and the functions of their hypothetical proteins so that clinical significance of these microbial proteins can be determined. Some hypothetical proteins are derived from the well-known common microorganisms such as E. coli, Enterobacters or Candida albicans, and the functions of these microbes within the body and disease state is yet to be completely understood.

## **Conclusion:**

More than two third of the autoantigens important for human autoimmune diseases are homologous to the microbial proteins, suggesting a majority of the autoantibody production are against the microbial protein origin. Further experimental validation is required for better understanding of the autoantibodies and the autoimmune diseases.

## Figure legends:

**Figure 1.** Definition of protein sequence homology with colored graphic illustration and the sequence alignment between the human autoantigens and the microbial proteins. The BLASTP search parameters were of the default setting, and the colored graphical illustrations were from the NCBI.

**Figure 2.** Highly homologous bacterial proteins identified in Crohn's and Sjogren's patients to human proteins by the default BLASTP search with graphic illustrations and protein sequence alignments. Individual microbial proteins and their Genbank accession numbers were listed on top of the graphics.

**Figure 3.** Representative high homology alignments between the human autoantigens and the microbial proteins (SSA60 and mitochondrial protein M2 complex E2 component with their Genbank accession numbers).

**Table 1.** Human autoantigens in autoimmune diseases and homology to the microbial proteins by BLASTP analysis. All common human autoantigens were listed with their respective search data against the bacterial and fungal protein database. The unique microbes (bacterial or fungi) were also denoted.

**Table 2.** High homology between the human autoantigens and the microbial proteins in categories based on the protein structures and functions.

## **Conflict of interest**

PZM Diagnostics, LLC is a private clinical laboratory registered in the State of West

Virginia, USA. The author is the co-founder and the stake owner of the laboratory.

## References

1. Kumar V, Abbas, AK., Aster, JC. Robbins and Cotran Pathologic Basis of Disease - 9th edition. 9th Ed. ed.: Elsevier; 2015.

2. Kasper D, Fauci, AS., Hauser, SL., Longo, DL., Jameson, JL., Loscalzo, J. Harrison's Principles of Internal Medicine. 19th Ed. ed.: McGraw Hill Education; 2015.

3. Gilbert JA, Blaser MJ, Caporaso JG, Jansson JK, Lynch SV, Knight R. Current understanding of the human microbiome. Nat Med. 2018 Apr 10;24(4):392-400.

4. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, et al. Microbiome-wide association studies link dynamic microbial consortia to disease. Nature. 2016 Jul 7;535(7610):94-103.

5. Young VB. The role of the microbiome in human health and disease: an introduction for clinicians. Bmj. 2017 Mar 15;356:j831.

6. Kowarsky M, Camunas-Soler J, Kertesz M, De Vlaminck I, Koh W, Pan W, et al. Numerous uncharacterized and highly divergent microbes which colonize humans are revealed by circulating cell-free DNA. Proc Natl Acad Sci U S A. 2017 Sep 05;114(36):9623-8.

7. Zhang P, Minardi LM, Kuenstner JT, Zekan S, Kruzelock R. Extracellular Components in Culture Media of Mycobacterium Avium Subspecies and Staphylococci with Implications for Clinical Microbiology and Blood Culture. American Journal of Infectious Diseases and Microbiology. 2016;4(6):112-7.

8. Zhang P, Minardi, LM., Kuenstner, JT., Zekan, SM., Zhu, F., Hu, YL., Kruzelock, R. Cross–reactivity of antibodies against microbial proteins to human tissues as basis of Crohn's disease and Sjogren's syndrome. <u>http://biorxivorg/content/early/2017/03/13/116574</u>. 2017.

9. Zhang P, Minardi LM, Kuenstner JT, Zekan SM, Kruzelock R. Anti-microbial Antibodies, Host Immunity, and Autoimmune Disease. Front Med (Lausanne). 2018;5:153.

10. Somayaji R, Priyantha MA, Rubin JE, Church D. Human infections due to Staphylococcus pseudintermedius, an emerging zoonosis of canine origin: report of 24 cases. Diagn Microbiol Infect Dis. 2016 Aug;85(4):471-6.

11. Kochibe N, Nicolotti RA, Davie JM, Kinsky SC. Stimulation and inhibition of anti-hapten responses in guinea pigs immunized with hybrid liposomes. Proc Natl Acad Sci U S A. 1975 Nov;72(11):4582-6.

12. Niegowska M, Paccagnini D, Burrai C, Palermo M, Sechi LA. Antibodies against Proinsulin and Homologous MAP Epitopes Are Detectable in Hashimoto's Thyroiditis Sardinian Patients, an Additional Link of Association. PLoS One. 2015;10(7):e0133497.

13. Aschermann S, Lehmann CH, Mihai S, Schett G, Dudziak D, Nimmerjahn F. B cells are critical for autoimmune pathology in Scurfy mice. Proc Natl Acad Sci U S A. 2013 Nov 19;110(47):19042-7.

14. Arnoux F, Mariot C, Peen E, Lambert NC, Balandraud N, Roudier J, et al. Peptidyl arginine deiminase immunization induces anticitrullinated protein antibodies in mice with particular MHC types. Proc Natl Acad Sci U S A. 2017 Nov 21;114(47):E10169-E77.

15. Gabarrini G, de Smit M, Westra J, Brouwer E, Vissink A, Zhou K, et al. The peptidylarginine deiminase gene is a conserved feature of Porphyromonas gingivalis. Sci Rep. 2015 Sep 25;5:13936.

	•	-	•				
Autoimmune disease	Autoantibody	Autoantigen	Bacterial homology (high, medium, low)	Fungal homology (high, medium, low)	Unique to bacteria	Unique to fungi	Unique microbes
	Anti-SSA/Ro	Ribonucleoproteins	SSA60 High	SSA60 High			
Systemic lupus erythematosus			SSA52 Low	SSA52 Low			
	Anti-ds-DNA	Double-stranded DNA					
	Anti-Smith	snRNP core proteins	U5 snRNP 100K High	U4/U6-U5 complex High		U5	Candida
	Anti-histone	Histone	Medium	Medium			Gandida
	Anti-thrombin/lupus anti-	Thrombin		Medium			
	coagulant Anti-ubiquitin	Ubiquitin	Medium	High			Acinetorbacter,
		e and annual	High				streptomyces
Sjogren's syndrome	Anti-SSA/Ro	Ribonucleoproteins	<b>g</b>				
	Anti-SSB/La		Medium	Medium			
CREST syndrome	Anti-Centromere	Centromere	CENP-A, B, C	Medium			
Inflammatory myopathy	Anti-Jo1	Histidine-tRNA ligase	Medium <mark>High</mark>	High			Clostridium
Primary biliary cirrhosis	Anti-p62	Nucleoporin 62	No bacterial homology	Medium		p62	
	Anti-Sp100	Sp100 nuclear	No bacterial homology	Low			
	Anti-gp210	antigen Nucleoporin 210	No bacterial homology	No fungal homology			
	Anti-mitochondria	Mitochondria	M2, High, p52 High	High, High			
	ATP-dependent DNA helicase 2 autoantigen p86, p70)	subunit 2 (Lupus Ku	No bacterial homology	High		Ku	
Systemic sclerosis	Anti-topoisomerase	Type I topoisomerase	High	High			
Celiac disease	Anti-tTG	(Scl-70) Transglutaminase	Medium	Low	tTG		
	Anti-actin	Actin	High	High			
Dermatitis herpetiformis	Anti-eTG	Transglutaminase	Medium	Low	eTG		
Rheumatoid arthritis	RF	lgG	High	High			
	Anti-CCP	Cyclic citrulinated	PAD1, High	Medium	CCP		
Autoimmune hepatitis	Liver kidney microsomal type 1	P450 2D6	High	Medium	Liver kidney microsome 1 (P450 2D6)		
	Ant-soluble liver/liver-pancreas	UGA suppressor	High	High	(		Chlamydia
	antibody Anti-smooth muscle	tRNA protein Smooth muscle	SMA, myosin, tropomyosin, F-actin High	High			Escherichia coli
Granulomatosis with polyangiitis	C-ANCA	Leukocyte Protease 3	Medium	High			
poryangino		Autoantigen 56kD (CAP-50, P50559)	High	High			

## Table 1: Human autoantigens and homologies to microbial proteins in autoimmune diseases

	autoantigens and no	mologics to mic			iscuses		
Microscopic polyangiitis, eosinophilic granulomatosis,	p-ANCA	Myeloperoxidase	High	High			
systemic vasculitides Hashimoto's thyroiditis	Anti-TPO	Thyroid peroxidase (microsomal)	High	High			
	Anti-TG	Thyroglobulin	Medium	Medium			
		Graves's soluble carrier protein, P16260	Medium	High		Graves' P16260	
Graves' disease	Anti-TSH	TSH receptor	Low	Low		110200	
Myasthenia gravis	Anti-AchR	Nicotine acetylcholine receptor	Low	No fungal homology			
	Anti-MUSK	Muscle specific kinase	Medium	Medium			Salmonella
Lambert-Eaton myasthenic syndrome	Anti-VGCC	Voltage-gated calcium channel (P/Q type)	Medium	High		Calcium channel	
Limbic encephalitis, Isaac's syndrome (autoimmune neuromyotonia)	Anti-VGKC	Voltage-gated potassium channel	Medium	Medium			
Polymyositis	Anti-SRP	Signal recognition particle	High	High			
Scleromyositis		Exosome complex					
Limbic encephalitis, encephalomyelitis, subacute sensory neuronopathy	Anti-Hu (ANNA-1)	Neuronal nuclear proteins	Hu-D, Medium	Medium			
	Anti-Ma		No bacterial homology	No fungal homology			
Paraneoplastic cerebellar degeneration	Anti-Yo	Cerebellar Purkinje cells	No bacterial homology	Low			
	Anti-Tr	Glutamate receptor	No bacterial homology	No fungal homology			
Opsocionus myoclonus syndrome	Anti-Ri (ANNA-2)	Neuronal nuclear proteins					
Stiff person syndrome, Diabetes type I	Anti-GAD	Glutamate decarboxylase	High	High			
	Anti-amphiphysin	amphiphysin	No bacterial homology	Medium		amphiphysin	
Optic neuropathy, chorea	Anti-CRMP-5	Collapsin response mediator protein 5	High	High			
Sydenham's chorea, PANDAS		Basal ganglia neuron	Dopamine D2 receptor, Medium	Low			
Anti-NMDA receptor encephalitis	Anti-NMDA receptor	N-methyl-D-aspartate receptor (NMDA)	Medium	No fungal homology			
Neuromyelitis optica (Devic's syndrome)	Anti-NMO	Aquaporin-4	Medium	Medium			
		Human interferon	No bacterial homology	Medium		IFN-G	
		gamma BP230 (BPAG1)	No bacterial homology	High (first 400 aa)		BP230	
		BP180 (BPAG2)	Low (60 to 138 aa identical to E. coli	Low	BP180		E. coli

## Table 1: Human autoantigens and homologies to microbial proteins in autoimmune diseases

## Table 1: Human autoantigens and homologies to microbial proteins in autoimmune diseases

	•	e	•				
			hypothetical protein)				
Nephrotic and nephritis		Factor H	Low (40-50 aa identical to E. coli hypothelical protein	No fungal homology			E. coli
		Complement C3	High	No fungal homology	C3		Klebsiella
		C1q receptor/Calreticulin	High	High			Kangiella spongicola
		ACE (angiotensin converting enzyme)	High	High			
Neurodegenerative diseases		Presenillin 1	Low (50 aa identical to E. coli hypothetical protein	High		Presenilin	
		Nacastrin	Low	Medium			
		MMP1	Low	Medium			
Type 1 diabetes		Insulin	Medium (44 aa identical to Enterobacter faecium)	Low	Insulin		Enterococcus faecium
Mixed connective tissue	Anti-RNP	Prohormone convertase Ribonucleoprotein	Low	High		Prohornmone C	
Antiphospholipid syndrome	Anti-phsopholipid	Phospholipid					
Miller-Fisher syndrome	Anti-ganglioside	Gangliosides GQ1B					
Acute motor axonal neuropathy		Ganglioside GD3					
Multifocal motor neuropathy with conduction block (MMN)		Ganglioside GM1					
			Totally 64 protein antigens <mark>25 high</mark>	29 high			

68.8%

69.20%

## Table 2: Protein classification of the autoantigens and microbial proteins

	Autoantibody	Autoantigen (Genbank accession #)	Bacterial homology (high)	Fungal homology (high)
Enzymes	Anti-Jo1	Histidine-tRNA ligase (P07814)	High	High
	Anti-mitochondria	Mitochondria (M2 and p52) (P10515, P11182)	High	High
	Anti-topoisomerase	Type I topoisomerase (Scl-70) (AAL05624)	High	High
	Liver kidney microsomal type 1	Liver kidney microsome 1 (P450 2D6) (AIA09571)	High	Medium
	C-ANCA	Leukocyte Protease 3 (P24158)	Medium	High
		Autoantigen 56kD (CAP-50, P50559)	High	High
	p-ANCA	Myeloperoxidase (AAA59863)	High	High
	Anti-TPO	Thyroid peroxidase (microsomal) (AAA61217)	High	High
		ACE (angiotensin converting enzyme) (AAR03504)	High	High
		Presenillin 1 (AAB46371)	Low (50 aa identical to E. coli hypothetical protein	High
		Prohormone convertase (insulin) (AAQ89322)	Low	High
	Anti-GAD	Glutamate decarboxylase (CAA49554)	High	High
Structural proteins	Anti-smooth muscle	Smooth muscle (SMA, myosin, tropomyosin, F- actin)(P62736, EAW66154, AAB59509, 5TBY-F)	High	High
	Anti-ubiquitin	Ubiquitin (CAA28495)	High	High
	Anti-CRMP-5	Collapsing response mediator protein 5 (Q9BPU6) BP230 (BPAG1) (Q03001)	High No bacterial homology	High High (first 400 aa)
Immune protein		Complement C3b (NP_000055)	High	No fungal homology
		C1q receptor/Calreticulin (P19474)	High	High
Nucleoproteins	Anti-SSA/Ro (SSA60)	Ribonucleoproteins (NP_001035828)	High	High
	Anti-Smith (U5 snRNP 100K, U4/U6- U5 complex)	snRNP core proteins (AAB87902, SC5314)	High	High
	. ,	Graves's soluble carrier protein, P16260	Medium	High
lon channel	Anti-VGCC	Voltage-gated calcium channel (P/Q type) (O00555)	Medium	High
	Anti-SRP	Signal recognition particle (NP_003127)	High	High

#### SSA52 --- Low

	INC search	rimila	-		SPRY super	
Dis	tribution of the to Mouse over to		on 26 subject se k to show alignm			
■<40	Color 40-50	key for align	ment scores	0 .	=200	1
1	90	Que 180	270	360	450	
-		_				
	_	_	_			
-						

hypothetical protein A6F71\_09265 [Cycloclasticus sp. symbiont of Poecilosclerida sp. M]

72.4 bits(176) 6e-10 Compositional matrix adjust. 63/247(26%) 114/247(46%) 26/247(10%) NASAARLTNNWEEVTCPICLDPFVEPVSIECGHSFCQECISQVG-----KGGGSV-CPVCR 55 NA A L ++ E++ C +CLD + +P ++C H FC++C+ +G +G S+ CP CR NAEEA-LKIIEEQLKCSVCLDIYTDPKLLQCFHVFCRKCLVPLGVRDQQQLSLTCPTCR 59

Identities

-----RFLLKN-LRPNRQLANMVNNLKEISQEAREGTQGERCAVH-GERLHLF 102

+ FL+ N L +++L N + + Q RC+ H GE L+ + KVTPRTVAGLOSAFLINNLLEAHKKLLNPLGKGNTTPPTSASAVO--RCSEHAGEDLNFY 117

 Query
 103
 CEKOGKALCWVCAQSRKHRDHANVPLEEAAQEYQEKLQVALGELRRKQELAEKLEVEIAI
 162

 C+
 K+C
 CA
 H
 ++A + V
 E++ ++L
 H
 E+

 Sbjct
 118
 COTCOKLICHCHABNGO-NNETGYCADEXKTEMENGEDVYTKALHELKK
 176
 177

Sequence ID: ORU94389.1 Length: 689 Number of Matches: 1

**Expect Method** 

Range 1: 1 to 243 GenPept Graphics

Score

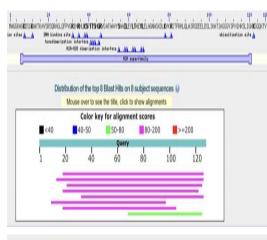
Query Sbjct

Query 56

Sbjct 60

Query 219 AQQSQAL 225 Sbjet 237 AQLKSCL 243

#### Histone --- Medium



#### Bownload v GenPept Graphics

V Next Match & Previous Match

Gaps

Positives

hypothetical protein A7M48\_18550 [Acinetobacter baumannii] Sequence ID: OIC85867.1 Length: 124 Number of Matches: 1

Score		Expect	Method	Identities	Positives	Gaps
130 bit	s(328)	7e-37	Compositional matrix adjust.	69/109(63%)	82/109(75%)	2/109(1%)
Query			LQFPVGRIHRHLKSRTTSHGRVGAT LOFPVGRIHR L+ + RVGA			
Sbjct			LQFPVGRIHRLLRKGNYAE-RVGAG			
Query	79		HLQLAIRGDEELDSLIK-ATIAGGG			
Sbjet	75		HLQLAIRNDEELNKLLSGVTIAQGG			

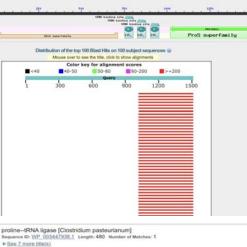
#### Bownload - GenPept Graphics

#### hypothetical protein [Pseudooceanicola sp. 157] Sequence ID: WP\_100165103.1 Length: 159 Number of Matches: 1

#### ► See 1 more title(s)

Range 1: 25 to 137 GenPept Graphics Thest Match & Previous					
Score	Expect	Method	Identities	Positives	Gaps
120 bits(301)	2e-32	Compositional matrix adjust.	67/114(59%)	83/114(72%)	2/114(1%)

## Jo1 (aminoacyl-tRNA synthetase) -- High



Score		Expect	Method		Identities	Positives	Gaps
461 bit	ts(1185)	9e-145	Compos	itional matrix adjust.	238/494(48%)	324/494(65%)	32/494(6%
Query				EMIEYHDISGCYILRPWA			
Sbjct.				E***Y * GC ILRP+A ELVDYSSVKGCVILRPYA			
Query	1081			DFAPEVAMVTRSGETELA FAPEVAMVT+ G EL		YAK VOS++DL	
Sbjet	77			GPAPEVAWVTRGGNDELT			
Query	1141			KEPOPFLATREFLWOEGH X +PFLAT EFLWOEGH			1200
Sbjct	137			KTTRPFLRT EFLWQEGH			195
Query				FAGGDYTTTIEAFISASG FAG + T TIE+ + G	RAIQGGTSHHLGQN +A+0 GTSH GON		1260
Sbjet				FAGAEATTTIESLMH-DO			252
Query	1261	IPGEK-QF	AYONSHG	TTR IG + MVHGD+ 0	LVLPPRVACVQVV	IIPCGITNALSEED	1319
Sbjet	253			MTTRIIGAIINVHGDDEG			304
Query				V RV+ D+ D +PG			1379
Sbjet				AKV-ARVENDISDE-TPO			362
Query				NEAETKLQAILEDIQVTL +E ETK+ +LEDI ++		ANTHEDFORILDS	
Sbjct				DELETKIPELLEDINKSN			
Query	1440			EDWIKKTTARDQOLEPGA ED IS+ T	PSMGARSLCIPFKI GA S CIPF+	PLCELQPGARCVCG	1498
Sbjct	423			EDKIKEDT			466
Query	1499	KNPARYYT AX+	LFGRSY +GR+Y	1512			
Sbjct	467	GKRAKHNV		480			

## Figure 2

#### Hsp65 (P9WPE6) ATP5a (KZK18041) EF-Tu (AAA50993) EF-G (WP\_014614745) 10 JAM A MA Fill superfamily ion of the top 111 Blast Hits on 100 subject sequences 4 Distribution of the top 100 Blast Hits on 100 subject see chaperonin\_like superfamily o see the tille, click to show alig Distribution of the top 32 Blast Hits on 32 subject sequences 📦 Color key for alignment scores 40-50 50-80 50-200 Color key for aligne =>=200 Distribution of the top 09 Blast Hits on 69 subject sequences @ 40.50 80,200 B >= 200 Acuse over to see the title, click to show alignments e the title, click to show aligne 140 210 280 350 200 300 +00 500 Color key for alignment scores Color key for alignment scores ■ 50-80 ■ 80-200 ■ >=200 40-50 40-50 80-200 300 500 200 400 elongation factor G [Homo sapiens] 60 kDa heat shock protein, mitochondrial [Homo sapiens] Bequence ID: NP 002147.2 Length: 573 Number of Matchesi 1 ATP synthase subunit alpha, mitochondrial isoform a precursor [Homo sapiens] elongation factor Tu, mitochondrial precursor [Homo sapiens] nce ID: NP 004037.1 Length: 553 Number of Matches: ce ID: AAK58877.1 Length: 751 Number of Matches: 1 Sequence ID: NP 003312.3 Length: 455 Number of Matches: 1 > See 13 more title(s) > See 14 more title(s) ▶ See 4 more title(s) Range 1: 44 to 734 Genfrot Graphics Range 1: 27 to 554 GenPeot Graphics Identities Positives Range 1: 50 to 553 Genheat Graphics Range 1: 50 to 444 Genfrot Graphics Method 540 bits(1390) 0.0 Compositional matrix adjust. 292/693(42%) 424/693(61%) 13/693(1%) 469 bits(1206) 1e-159 Compositional matrix adjust. 248/528(47%) 364/528(68%) 3/528(0 600 bits(1546) 0.0 Compositional matrix adjust. 297/504(59%) 377/504(74%) 10/504(1%) Expect Method Identities Positives Gaos Overy 8 FRNIGIMARIDAGETTTTERILYYTGRIHEIGETH--EG-ASOMDWBEGEODRGITIT 64 RNIGI AHID-GETT TER-LYYTGRI X+ E +G + MD ME E+ RGITI 469 bits(1207) 1e-163 Compositional matrix adjust. 222/395(56%) 299/395(75%) 2/395(0%) Query 2 Query Shjot 44 ++ + + + G V+ IGDGIA VHGL +V A E++EF +G+ G+ Shirt 27 ERERFERTERIVENVGTIGHVDHGKTTLTAAITTVLAKTYGGAARAFDQIDNAPEEKARGI 61 +K+ + R KPHVNVGTIGHVDHGKTTLTAAIT+LA+ G + ++IDNAPEE+ARGI AKKTYVRDKPHVNVGTIGHVDHGKTLTAAITKILARGGGARFKKYEEIDNAPEERARGI 109 16 Sbjct 50 100 Query 2 Query 65 124 NOAT 124 Query 62 Query 67 124 Sbjet 50 AAT N+ +NIIDTPGHVDFT+EVER+LRVLDGAV VL A GV+ QT TV RQ \*\* GANP\* \*\*RG VA T++ AGDG Sbjet 104 Sbjct 87 PVEINRGVML 146 Shict 110 149 Query 62 TINTSHVEYDTPTRHYAHVDCPGHADYVKNNITGAAQNDGAILVVAATDGPMPQTRENIL 121 Overy 125 GVPRIVFVNKHDELGANFEYSVSTLHDELGANAAPIGLPIGAEDEFEAIIDLVEHKC / VF + F+NK+D++G+N ++ + +L NAA +G+P+G E F+ 1+DL+E + Query 122 ETLL&GAKEVETKEQIAATAAISA-GDQ&IGDLIAEAMDKVGHEGVITVEESHT 180 L K +K V T E+IA A IEA GD+ IG++I++AM EVG +GVITV++ T Query 127 184 IN HEVEY T RHYAH DCPGHADYVKNNITG A +DG ILVVAA DGPHPOTREH+ KA G++ B SV EP+OTGIES fibict 110 TINAAHVEYSTAARHYAHTDCPGHADYVKNHITGTAPLDGCILVVAJ ANDGPHPQTREHLL 169 Sbjct 147 206 Sbjet 164 Sbjct 170 229 Query 122 LGRQVQVPYIIVFLNKCDNVDOEELLELVEMEVRELLSQVDFPGOOTPIVRGSALKALEG 181 L R0+GV +++V+HK D V D E++ELVE\*E+RELL+++ G++TP++ GSAL ALEG Bbjot 170 LARQIGVEHVVYYNKADAVQOSEMVELVELEIRELJTEFGYKGEETPVIVGSALCALEG 229 Query 181 K+S+VKDLLPLLEKVIG 240 K+S+++++P LE Query 187 238 Query 185 EF TNDLGTEIEEIEIPEDHLDRAEEARASLIEAVAETSDELHEKYLGDEEISVSELKEA D G + EIP + A + R LIE VA + \*\*L E \*L +\* S\*S\*LK AJ D G D + EIP + A + R LIE VA + \*\*L E \*L +\* S\*S\*LK AJ 289 Sbjot 207 LN ALEIANA 266 Sbjct 230 Sbjet. 224 Query 241 AGEPLLIIAE QUEYY 182 -OAEWEAK-ILELAGFLDSYJPEPERAIDKPPLLPIEDVFSISGRGTVVTGRVERGIIKV 239 D E K + +L +OHYIP P R ++KPFLD+E V+S+ GRGTVVTG +ERGI+K Sbjet 230 RDPELGLKSVQKLLGAVDTYIPVPARDLEKPFLLPVERVYSVPGRGTVVTGTLERGILKK 289 EGEALSTLVVNKIRGTFESVAVEAPG 300 Query 239 298 Query 245 QATTNVEFYPVLCGTAFKNKGVOLHLDAVIDYLPSPLDVK--PIIGHR \*AT F PV G+A KNKGVQ \*LDAV+YLP\*P +V+ I+ 11AEDV+GEALSTLV+S+++ + VAVEADGCD IX L-UMAI IGG V 11AEDV-DEALSTLV-SH+++ + VAVEADGCD IX L-UMAI IGG V Y\*G \*M E F NGKH LI\*YDOL\*EQA AYR \* E\* I Sbjot 290 349 Sbjet 267 Sbjet. Query 299 Query 301 359 LGK +V+VTEDETTIVEGAGDTDAIAGEVA LGK +V+VTED+ +++G GD I R+ + Query 240 GEEVEIVGIKETQKSTCTGVENFRKLLDEGRAGENVGVLLRGIKREEIERGQVLAKEGTI 299 GFE E+tG + ++ TG+BMF K L+ Ac+N+G L+RG+KRE++ RG V+ KPG+I Sbjet 290 GDECELLGENKIRTVYTGIENFRSLERAKIDAVKPKGLERGLIKVEVFORGI 349 Query 301 Sbjet 327 EE Sbjet 350 40.9 D+S F LAFE+ + G+LT+ R Y G + G + N+ E+ R+ RL +NE 386 Shiet 344 Query 360 419 Query 159 SLSVSRVGGSAQIKANKKVAGTLRLDLASYRELESPAQFGSDLDEATASKLERGKRTVE 418 SLSVSRVG +AQ +ANK+VAGT++L+LA YRE+ +PAQFGSDLD AT L RG R E \*AAVEEGIVAGGGVTLLQ \*AAVEEGIV GGG LL\* TRAAVEEGIVLGOGCALLB EKLQERLAKLAGGVAVIKAGAATEVELKERKHRIEDAV EKL ERLAKL+ GVAV+K G ++VE+ E+K R+ DA+ Query 300 KPHIKPEERVVILSKOEGGHHPFFKGYRPGPFWTFOVFGTELEEGVENVHPGONIKH 359 KPH K E+VYLLSK-EGGRH FF T D+ I\_EP E+HPG++K Ebjet 350 KPHGKVEAQVYLLSKEEGGRHKPFVSHFMUVHFSLTHEMACRILLPEKELANDGEDLKF 409 Query HQEIDTVYSGDIAAAVGLRDTGTGDTLCGERND-IILESMEFFEPVIHLSVEPKSK/ ++++ VY+GDI & G+ D +GDT + N + +ES+ P+PVI ++++P +K Sbjct 410 100 Sbjet 38 446 Shiet 403 46 Query 419 VLRQDQNRFLFVEHQVLIIYALTRGVLD0IFVVDITRFEAEIIEMTKSNANEIFTEIRET 478 \*LKQ Q F+ 4E QV \*IYA RGVLD + IT+FE + 5 IR LLKQGQTFPALIEQVAVIJAQUVGFLLAEPTKITKFUNFLSHVV3Q0GALLGTIRAD 529 QUERY 420 AAPTLDELELEG-DEATGANIVKVALEAPLKO EXVENLEAGEDINAD 478 + G +A 506 P LD L D+ G I+E L+ P IA N+G+ ++ EX-Sbjet 447 CIPALDSLTPAMEDOKIGIEIIKRTLKIPAMTIAKNAOUNCETTUET Sbjct 470 Query 360 VVTLIHPIAMDCGLRFAIREGGRTVGAGVVAKVLS 394 + L P+ ++ G RF +R+G RT+G G+V L+ Bbjet 410 NLILRQPNILMKGQRFTLRDGNRTIGTGLYPNTLA 444 DENTEMATURE EDPTFANTORTOGY I GUNGELALDI LVDDMGLATTELTURE \*E ++ + EDPTF + D E ++ I GMGELALFI E++E+ C G P EFFSKIGFTFROTTEVYFOTENKETVISGMGELALE IYAQULERETGOPCITOKP 521 Query 479 GGLP--ADEXFEGAINEFFEGFEA 500 G + +D K + F GP+A Sbjet 467 DL Query 479 TGVYEDLLAAGVADPVKVTRSALQNAASIAGLFLTTEAVVADKPEREF G + \*\*+ G+ DP KV R+AL +AA +A L T E VV + P++EB Query 479 MVSYMETFRESAQVQGKFSRQSGGRQYGDVNIEFTP--HETGAGFEFENAIVGGVVPRE 536 V++RET + \*QSGG GQYG V P E EF + G +P++

## Figure 3

## SSA60/Ro60 (NP\_001035828)

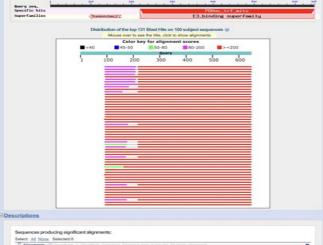
inperfamilies			superfamily				AFR S	SIDER	Family
							_		
	Distri		100 Blast Hits on 100 s see the blas, click to sho		nces 🗟				
		Color	key for alignment s	scores				-	
	<b>a</b> <40	40-50	50-80	80-200		200	_	- 1	
	5	100	200 300	4	0	5	00	- 1	
							_	- 1	
	_						-	- 1	
	_						-	- 1	
								- 1	
	-						-	- 1	
								- 1	
								- 1	
								- 1	
	-							- 1	
							-	- 1	
								- 1	
							-	- 1	
	_						-	- 1	
							=	- 1	
							E		
							Ē		
criptions									
criptions									
criptions.	ligsthart stigm	ents:							
Sequences producing s Select Al Nazy Select		weite:							
Sequences producing s		ints:	a the diractly William						
Sequences producing s Select Al Nazy Select	ed.0	erfa: Descripton		Max	Total	Duery	E	iterat	Accession
Sequences producing t Select: <u>All Norm</u> Delect 3 Algements 3	edo	Description	a ba dan da Wasa				value		Accession
Sequences producing s Select: <u>All Name</u> Select <u>Alignments</u>	ed 0	Description		90079 441	441	001V887 (97%)	value 6e-147	45%	Accession WP 034341831
Sequences producing s lefect. All Name Select Alignments	ed 0	Description Social Theorem		441 440	441 440	97% 97%	value 6e-147 2e-145	45%	Accession WP 004041801 WP 000719490
Sequences producing t Select All Song Select Algements DEC/E Senae cost DEC/E Senae cost	ed 0 Internet Construction Internet Internet Construction Internet Internet Internet	Description coccus intervents coccus Entervents cohema camulat		441 440 432	441 440 432	97% 97% 97% 90%	value 5e-147 2e-145 5e-143	45% 42% 42%	Accession WP 004041801 WP 000710400, WP 007410407,
Sequences producing a Select al Stand Select Algements 1907/E dense cets 1907/E dense cets 1907/E dense cets 1907/E dense cets 1907/E dense cets	ed 0 anital a Construct anital anital (Dens anital anital (Dens anital anital (Dens anital Dens anital (Dens	Description coccus missemals conse E-tothernell sphere consult return 25	k.addudeestil	441 440 432 417	441 440 432 417	97% 97% 90% 98%	value 6e-147 2e-146 5e-143 1e-137	45% 42% 42%	Accession WP 004041801 WP 000719680 WP 000719680 EDMINAGE1
Sequences producing Select 21 Nam Delect Augments 1920/Edman.com 1920/Edman.com 1920/Edman.com 1920/Edman.com 1920/Edman.com 1920/Edman.com	edd arcine anoten (Deiro arcine anoten (Deiro arcine anoten (Deiro rocheanoten (Deiro rocheanoten (Deiro Carobiatus Surate	Description Description sciolus missemals idatus Entatheoret spheres samulat istes sci.PSI sciente Sacterium Ri	PCOPHERCE II FULL	441 440 412 417 11.22 408	441 440 432 417 408	97% 97% 90% 98% 98%	value 5e-147 2e-145 5e-143 1e-137 3e-134	45% 42% 42% 47% 41%	Accession WP 004041001 WP 000710400 WP 000710400 WP 000740807 EDMD446.1 D02298546.1
Sequences producing t Select Al Nam Select Adjuments 1902 Edman-colt 1902 Edma	ed 0 Interne Inches (Deno Inches Inches (Deno Inches Inches (Deno Inches Inches (Deno Inches Inches (Deno Inches Inches (Deno	Constant Description Description Soluture Entrativenet opherna camulati initian and PDE edenia bacterium Ro	PCOPHERCE II FULL	441 440 412 417 417 416 405	441 440 432 417 408 407	97% 97% 98% 98% 98% 98%	value 5e-147 2e-145 5e-143 1e-137 3e-134 7e-134	40% 42% 42% 47% 47% 42%	Accession WP GALARIES WP 000718000 WP 007118000 WP 007118000 EDNET408.1 CO2290548.1 GAUGEDS 1
Beguences producing to Belefit Al Store Telecit Algomenta TROVE annan-sett TROVE annan-sett Bisto Schemato-sett Bisto Schemato-sett Trove annan-sett Trove annan-sett Trove annan-sett Trove annan-sett Trove annan-sett	ed 0 anima, protein (Deno anima, protein (Deno anima, protein (Deno anima, protein (Deno Constitution, Streach KGO11, 08640 (Anima anima, protein (Deno	Description control missional pharms provide relation activity activity bacterian (2) activity bacterian (2) activity bacterian (2)	PCOPHERCE II FULL	441 440 412 417 11_22 408 407 402	441 440 432 417 408 457 402	97% 97% 98% 98% 98% 98%	value 5e-147 2e-145 5e-143 1e-137 3e-134 7e-134 1e-131		Accession VMP 034041801 VMP 030718680 VMP 0307136880 EDMITMENT COLUMNED 0405030.1 VMP 01284208
Seguencies producing t Select. 20 Xiona Select 1907/E. deman-cent 1907/E. deman-cent 50 XIO/E. deman-cent 50 XIO/E. deman-cent 50 XIO/E. deman-cent 1907/E. deman-cent 1907/E. deman-cent	ed 0 Initial Inc. Construction Interna Interferen (Chernol Interna Interferen (Chernol Interferen Interferen (Chernol Constitution, Survator Interferen Interferen (Chernol Interna Interferen (Chernol Interna Interferen (Chernol Interna Interferen (Chernol	Description Description Solution Entertained Solution Entertained Solution and Phil Interna Encoderant (E) and an engenerated Antilancier antij	komenti FCSPHORO2 II FULL Intel8-18	441 440 412 417 417 406 407 402 400	441 440 432 417 408 427 402 400	27% 27% 26% 26% 26% 26% 26%	value 5e-147 2e-145 5e-143 1e-137 3e-134 7e-134 1e-131 6e-131	455 475 475 475 475 475 475	Accession WP 034341531 WP 03741868 WP 08741868 E20674481 C02285481 C02285481 047 01244283 WP 05320868
Seguricos producing s Select Al Sura Select Agrinenta TROVE domas care TROVE domas care Statu al Africato Statu al Africa	edd aming anthen (Serb aming anthen (Serb softwarthen Perlo softwarthen Brod (Sentitien Sorta (Sentitien Sorta (Sentitien Sorta (Sentitien Sorta) (Sentitien Sorta) aming anthen (Serb aming anthen (Serb aming anthen Sorta)	Description Description Colour, Hosenaria Solar, 2 Excellent Solaris Sactoria, 25 Solaris Sactoria, 25 Solaris Sactoria, 25 Solaris Sactoria, 25 Solaris Sactoria, 20 Solaris Sac	komenti FCSPHORO2 II FULL Intel8-18	441 440 412 417 417 417 408 400 400 400	80000 441 440 432 417 408 427 400 400 400	07% 97% 98% 98% 98% 98% 98% 98% 98% 98%	value 5e-147 2e-146 5e-143 1e-137 3e-134 7e-134 1e-131 5e-131	455 455 455 455 455 455 455	Accession VMP_004341801 VMP_002118080 VMP_002412802 EDME7468_1 CARMETARE_1 CARMETARE_1 VMP_012802885 VMP_0128028852 VMP_002221588
Segurations producing to interest AS Mura Select 20 Auguments 20 1900/C. Strategic and 1900/C. Strategic and 1	+10 anima andren (Dero anima andren (Dero	Control Control Description cocide Historical address Statement address Statement address Statement address Statement Address address Address and Address and Addr	ecological di Pala, nice 18:13 ni	441 440 412 417 407 407 400 400 307	441 440 432 417 405 407 400 400 297	97% 97% 98% 98% 98% 98% 98% 98% 98% 98% 98% 98	value 6e-147 2s-146 5e-143 1e-137 3e-134 1e-131 6e-131 5e-131 1e-129	455 455 455 455 455 455	Accession VPF 02424/VE1 VPF 02421/VE02 VPF 02424/022 C22962448.1 C22962448.1 C22962448.1 VPF 0212452888. VPF 0512212888. VPF 0512212888. VPF 051221288.
Segurations producing to interest AS Mura Select 20 Auguments 20 1900/C. Strategic and 1900/C. Strategic and 1	+10 anima andren (Dero anima andren (Dero	Control Control Description cocide Historical address Statement address Address Rit deta Statement address Address Rit deta descriptions Address and Address and A	komenti FCSPHORO2 II FULL Intel8-18	441 440 412 417 417 417 408 400 400 400	80000 441 440 432 417 408 427 400 400 400	07% 97% 98% 98% 98% 98% 98% 98% 98% 98%	value 6e-147 2s-146 5e-143 1e-137 3e-134 1e-131 6e-131 5e-131 1e-129	455 455 455 455 455 455	Accession VMP_004341801 VMP_002118080 VMP_002412802 EDME7468.1 CARMETINE.1 VMP_01240480 VMP_0124049880 VMP_002221568

#### TROVE domain-containing protein [Deinococcus misasensis]

Sequence ID: WP\_034341831.1 Length: 525 Number of Matches: 1

Score		Expect	Method		Identities	Positives	Gaps
441 bits(1	133)	6e-147	Compositional r	matrix adjust.	231/513(45%)	311/513(60%)	24/513(4%)
Query 7		OPL-NEP	QIANSQDGYVWQV O AN GY W+V		GSEGGTYYIKEQKI GSEGGT+Y+ EOKI		65
Sbjct 14			QKANHAGGYAWEV				73
Query 66	EI		QEIKSFSQEGRTT + I S+ GR				125
Sbjct 74			VERIAEISESGRAP				132
Query 12			LKESMKCGMWGRA	LRKAIADWYNEK LR+AIADWYN+K		ORNGWSHKDLLRL	185
Sbjct 13	3 FI		FRGWGRG	LRRAIADWYNQK	DLKALALQAVKYR	RDGWTHRDALRL	186
Query 18	+1	SEC	LAIVTKYITKGWK		E L+ +EA	++V++ E +	245
Sbjct 18	7 AI	IPVAPSEC	HQKLYRWITRD	Q	FEPATGLELIEAF	KQVQQASVP-EAV	232
Query 24			REHLLTNHLKSKE			/LEPGNSEVSLVC	305
Sbjct 23	3 KI	IKRHRLI	REALPTELLVRPE	VWEALLPHMPLE	AMVRNLANMTRVG	LTPMSDASRLVQ	292
Query 30			KKARIHPFHILIA KARIHP IL A				365
Sbjct 29			LKARIHPIKILAA				352
Query 36			VDVSASMNQRVLG		AMCMVVTRTEKDS A+ +V TE	VVAFSDEMVPCP V+ FS ++VP	423
Sbjct 35	3 P1	GKRIML	LDVSGSMEWGHIA	GIPGLTPRVASA	ALALVTASTEAQH	<b>WMGFSHQLVPIG</b>	412
Query 42	4 V1		VLMAMSQIPAGGT				483
Sbjct 41	3 IS		VLNTVGRVPMGGT				472
Query 48			PAKLIVCGMTSNG AKLIV GM SN		516		
Sbjct 47			AAKLIVVGMQSNR		505		

## E2 component p70 (P10515) (M2)



57	Algomenta (200-04-04) - DepEert, Graefie's Robbaschox at one-th Website at an						C
	Description			Query cover	E	1dent	Accession
	pyruvate dehydroperates complex dihvitrolipoemide acetyltransferates Michaeroleobacieria bac	399	496	54%	2e-130	49%	OFWERZ73.1
a	peruvale dehutiosenese complex dityrioliscenside acetylinanskease (Alshaprotechacteria bac	398	492	84%	6e-130	48%	P9832508.1
5	pyruvate dehydrogenase concilex dihydrolocemide acetytransferase (Phodospitilaceae bacte	397	489	84%	9e-130	50%	PC/60435.1
a,	pyruvate dehydrogenaae complex ditycholocenide acetyltransferaes (Alphaprotechacteria bac	395	490	84%	2e-129	48%	OFW75937.1
a,	perurate detudopartese conclex directolocamide acetyltansferase Tenacitaculum mesorie	395	395	85%	2e-127	40%	WP 072183424.1
ŝ	pyrusate dehydrogenase complex dhydrolicoemide adetylraeshease (Alphacoteobacteria bac	388	477	84%	5e-126	46%	OFWETROK1

pyruvate dehydrogenase complex dihydrolipoamide acetyltransferase [Alphaproteobacteria ba Sequence ID: <u>OFW69273.1</u> Length: 407 Number of Matches: 2

#### ▶ See 2 more title(s)

Score		Expect	Method	Identities	Positives	Gaps
399 bi	its(1024	4) 2e-130	Compositional matrix adjust.	209/430(49%)	284/430(66%)	29/430(6%)
Query	219		PTMTMGTVQRWEKKVGEKLSEGDLL PTMT G + +W KK G+K+S GD++			278
Sbjct	3	IEILMPALS	PTMTEGNLVKWHKKEGDKVSAGDVV	AEIETDKATMEVE	VDEGTLGKILIP	62
Query	279		TPLCIIVEKEADISAFADYRPTEVT +P+ +I+EK D +Y+	DLKPQVPPPTPPPV PTP	AAVPPTPQPLAP P P+ P	338
Sbjct	63	EGTENVKVN	SPIALILEKGEDKKVLENYKA	PTPAA/	QKEEPVSAPV-P	109
Query	339	TPSAPCPAT T ++P P 1	PAGPKGRVFVSPLAKKLAVEKGIDL GR+ SPLAK+LA EK IDL			398
Sbjct	110	TMASPAP-1	IVLSTGRIVASPLAKRLATEKNIDL	RQVQGSGPHGRIV	QDIDTFVPGSAA	168
Query	399	PAPAAVVPI A	TGPGMAPVPTG-VFTDIPISNIRRV P TG ++ D P+SN+RRV			457
Sbjct	169	RGHAL	PTHTGPLYQDKPVSNMRRV	IAQRLTESKQTVPH	FYLTLDCEIDAL	218
Query	458	LLVRKELNE L R+ +N	ILEGRSKISVNDFIIKASALACLKV K++VNDF++KA ALA V		NHVVDVSVAVST D+SVAV+	517
Sbjct	219	LAARQSINS	HFNVKVTVNDFVLKAVALALQDV	PDANASWRGETIRY	YTTSDISVAVAI	276
Query	518		FNAHIKGVETIANDVVSLATKAREG A+ K + TI+ +V +L KA+EG			577
Sbjct	277	EDGLITPIN	KMANFKSLLTISEEVKTLVQKAKEG	RLKPEEFQGGSFSV	/SNLGMFGIKQFE	336
Query	578		ILAIGASEDKLVPADNEKGFDVASM ILA+GA E + P E G VA++			637
Sbjct	337		ILAVGAGEKRPVVKEGGLAVATV			394
Query	638	YLEKPITMI Y+E P +I				
Sbjct	395	YIENPALLI				