[Frontiers in Bioscience 13, 842-855, January 1, 2008]

Homologous housekeeping proteins in Nocardia – the NoDaMS proteomic database

Simone Koenig[1], Anna-Maria Mehlich[1], Doreen Ackermann[1], Mohammed E. Hamid[2], Nageeb S. Saeed[3], Moawia Mukhtar[4], Mogahid El Hassan[5]

[1]Integrated Functional Genomics, Interdisciplinary Center for Clinical Research, University of Muenster, Germany, [2]Department of Microbiology, College of Medicine, King Khalid University, Abha, Kingdom of Saudi Arabia, [3]General Directorate of Laboratories and Medical Research, Federal Ministry of Health, Sudan, [4]General Directorate of the Institute of Endemic Diseases, University of Khartoum, Sudan, [5]College of Medical Laboratory Sciences, Sudan University of Science and Technology, Khartoum, Sudan

TABLE OF CONTENTS

1. ABSTRACT

Nocardiosis is on the rise but hard to diagnose and the application of advanced subtyping technologies is called for. While the genomic sequence for the most virulent strain, *Nocardia farcinica* is available, proteome data are essentially non-existent. Nevertheless, they are necessary for functional studies on virulence and disease prevention. Here, comparative gel electrophoresis (PAGE)-based analyses of the five Nocardia strains SD1828, *N. africana* SD910, SD 925, *N. sp.* 1086, and *N. asteroides* N317 are discussed. The two-dimensional gel images of all strains are similar and dominated by housekeeping proteins such as chaperones and metabolic enzymes. The sequences of many proteins are highly homologous among strains and in some cases Mycobacterium sequences are closer matches to the unknown than those of *N. farcinica*. All mass spectrometry data are made available in the NoDaMS database at URL http://ifg.uni-muenster.de/ (Proteomics-Projects-NoDaMS) for re-evaluation with fresh sequencing information. Assignments, homology analyses, and peptide matches are presented. This data review comprises the first comprehensive summary of proteomic data of Nocardia.

2. INTRODUCTION

Nocardiae are Gram-positive aerobic actinomycetes, which are predominantly saprophytic (1-2), but also include species forming parasitic association with animals and plants (3-4). They are in the same family as clinically and industrially important genera such as *Mycobacterium, Streptomyces, Corynebacterium* and *Rhodococcus* and they are known to cause a variety of infections in humans and animals. According to information from the Center of Disease Control and Prevention (5) 80 percent of cases present as invasive pulmonary infection, disseminated disease, or brain abscess, and 20 percent present as cellulitis. *Nocardia asteroides* causes at least 50 percent of invasive infections. In the United States, an estimated 500-1,000 new cases of *Nocardia* infection occur annually. The number of cases has increased with the overall rise in the number of severely immunocompromised persons and diagnosis is a major challenge. *N. farcinica* frequently is resistant to antimicrobial agents, including the drug of choice trimethoprim-sulfamethoxazole, and has been demonstrated to be more virulent in an animal model. A new combination

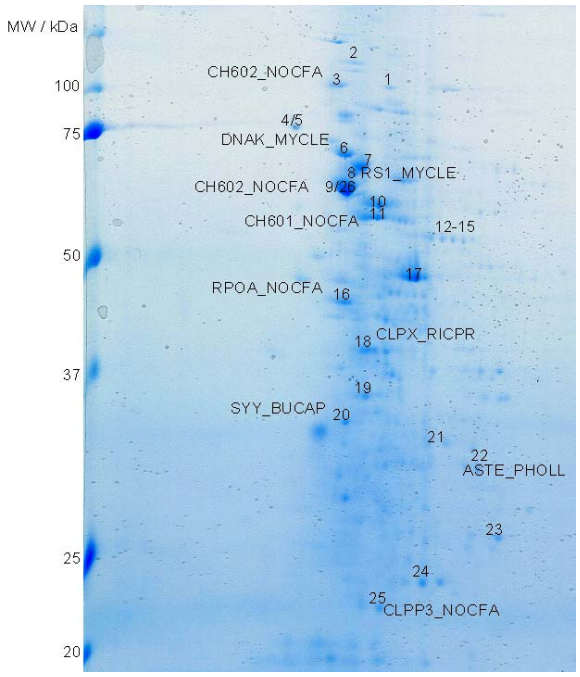

Figure 1. Representative spot assignments for *N. africana* SD925, pH 3-7 . For labels see Tables and Suppl..

drug therapy (sulfonamide, ceftriaxone, and amikacin) has shown promise for infections difficult to treat. The hope is that the application of newer molecular diagnostic and subtyping methods may assist in earlier diagnosis and outbreak investigation. (5).

The complete genomic sequence of *N. farcinica* IFM 10152 was published in 2004 (6), but investigations on the protein level are virtually non-existent as was summarized in (7). However, knowledge of the soluble proteome is crucial for functional investigations and it bears the potential for speeding up diagnosis. Therefore, several Nocardia strains of importance in Sudan were studied in a comparative way and the proteomes were landmarked in preparation for directed sub-proteome analysis.

3. EXPERIMENTAL CONSIDERATIONS

3.1. Strains

N. africana SD910*,* a new species of *Nocardia,* as well as *N. africana* SD925 were isolated from clinical specimens (sputum) obtained from patients suffering from pulmonary infection admitted at Chest Unit of Khartoum Teaching Hospital in Sudan (8). *N. farcinica* SD1828 was purified from goat milk and *Nocardia*. sp*. 1086* from cattle milk. *N. asteroides* N317 (ATTCC 19247 T) was a kind gift from M. Goodfellow (University of Newcastle).

The strains were cultivated on Glucose Yeast Extract Agar (GYEA) medium (9) for 2-3 days at 37 degree Celsius and then checked for purity. The isolates were Gram-stained and their morphology was analysed using light microscope Olympus-CH2-Japan prior to antimicrobial sensitivity testing. The single disk method (10) was used. The morphology was further studied with modified Zeihl Neelson stain (7). *Nocardia* species were harvested and inactivated in a water bath at 90 degrees Celsius for 1 h. The number of cells was approximately 15 x 10^7 (11).

3.2. Protein purification and analysis

The cells were washed with sterile phosphate-buffered saline and then centrifuged at 15000 xg for 5 min. Supernatants were discarded and the pellet was washed again by resuspending in MilliQ water and centrifuged at 10000 xg for 10 min. The supernatant was discarded. Then, 1 ml of lysis buffer (8 M urea, 2M thiourea, 4 percent Chaps, 1 percent Triton X-100, 65 mM DTT, 10 mM Tris-base) was added to each sample in addition to 100 microl protease inhibitor cocktail set III (Calbiochem), 20 microl benzonase (500 U, Novagen) and 0.5 mg zirconia/silica beads (cell suspension/beads v/v 3/1, 0.1 mm, BioSpec Products Inc.). The mixture was vortexed for 1 min. Samples were lysed by 5 cycles of freezing in liquid nitrogen, thawing at 37 degrees Celsius in a water bath, and shaking on a mixer mill (Retsch MM300, Qiagen; 30 Hz, 30 min). Samples were then spun down at 15000 x g for 5 min. The supernatant with the soluble proteins was collected and their concentration was determined (Cytoskeleton).

Protein mixtures of bacterial lysates were purified and up-concentrated using the CleanUp Kit from BioRad (Munich, Germany). Two-dimensional gel electrophoresis (2D-PAGE) was performed according to general protocols of BioRad with only slight modifications. 17 cm pH3-10 and 4-7 IPG strips were actively rehydrated applying 500 microg protein for 12 h at 50 V and 20 degree Celsius. 12 percent separating gels (1 mm x 16 cm x 20 cm) were run for the second dimension and stained with Coomassie Brilliant Blue G250. Image analysis was performed subsequently with PDQuest 6.2 (BioRad).

Prominent spots were excised and tryptically digested in the gel according to the procedure described in (12). For matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS), samples were purified using ZipTips C18 (Millipore, Bedford, MA, USA). Peptide maps were generated with MALDImicroMX instrument (Micromass/Waters, Manchester, UK). Alpha-cyano-4-hydoxycinnamic acid was used as matrix and 1 microl of analyte was applied. Database queries were carried out using Mascot (MatrixScience Ltd., London, UK) in-house with species limitation on bacteria. All data including supplementary material to this paper can be obtained from the NoDaMS web site (Nocardia Database on Mass Spectrometry analyses) on the following address: http://ifg.uni-muenster.de (Proteomics-Projects-NoDaMS).

4. PROTEOMIC DATA

2D-PAGE using pH strips 3-10 (*N. africana* SD925; see "Figure 1") revealed that the soluble proteins were visible in a much smaller pI range so that all following experiments were performed with pI strips 4-7. This effect

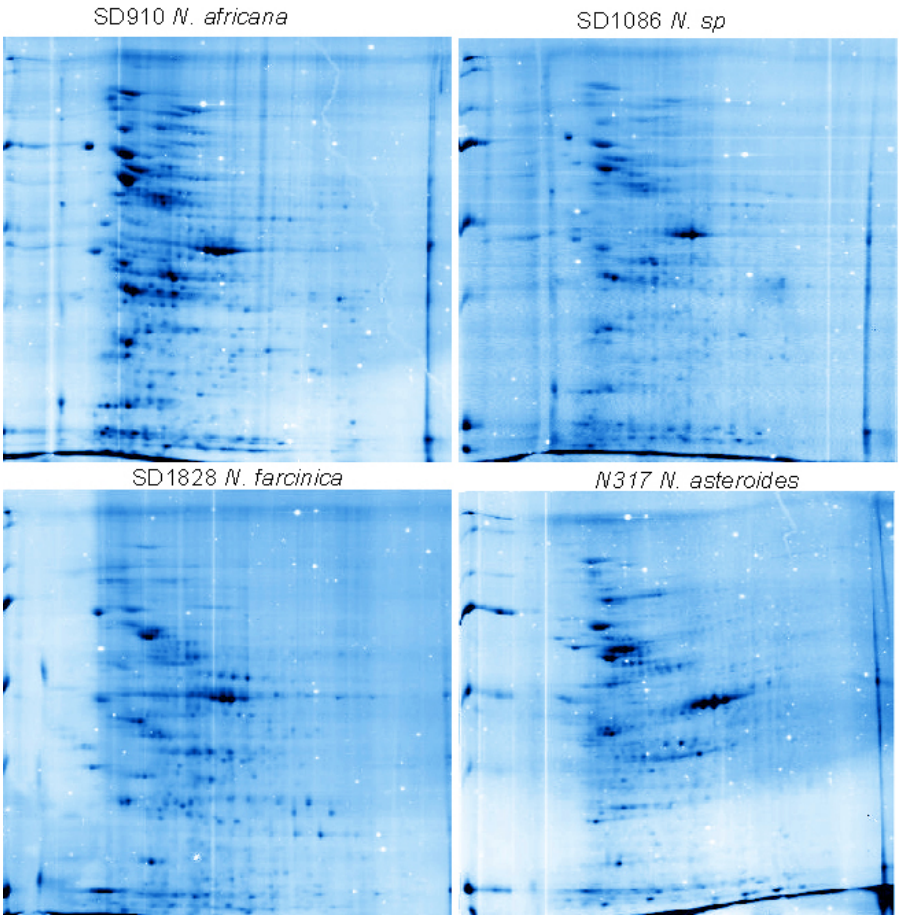

Figure 2. Gel images of *N. africana* SD910, *N. sp.* SD1086, *N. farcinica* SD1828, and *N. asteroides* N317 (pI 4-7, molecular weight MW 20-250 kDa).

had been noted before for *N. africana* SD769 (DSM44491) (7) and also strains of Mycobacterium (13). Four gel images are shown in "Figure 2". While all strains exhibited similar protein patterns, it was quite difficult to overlay them for comparative analysis. A reason for this were beside proteome specificities gel-related run variations in PAGE. The best result to that effect was obtained with the gels of *N. africana* SD910 and *N. sp*. SD1086 ("Figure 3") where regions of identical protein patterns could be located. In this study, prominent spots were excised from the gels and assigned using peptide mapping and post source decay fragmentation in an effort to evaluate the soluble proteome. All confident matches are presented in "Table 1" and "Table 2" with their SwissProt accession numbers while Supplementary Material and the original data and spectra can be downloaded from the NoDaMS database for further evaluation with the extension of the database knowledge. "Figure 1", "Figure 4" and "Figure 5" demonstrate the spot patterns with assignments for three strains. Typical spectra are exhibited in "Figure 6" and "Figure 7". As is observed in many proteomes, heat shock factors, chaperones and metabolic enzymes such as enolase dominate the visible proteome (14). Many can be found across strains as is exemplified in "Figure 8" for 60 kDa chaperonin 2 and "Figure 9" for transcriptase alpha. "Table 1" presents the peptides assigned to such common proteins in different gels and multiple spots, "Table 2" and the Supplementary Material all other hits. Often, the best match was not a *N. farcinica* protein but a sequence of Mycobacterium. A similarity analysis (Supplementary Material) underscored not unexpectedly the homology to Mycobacterium and to a lesser extent to Rhodococcus.

5. ABUNDANT NOCARDIA PROTEINS

2D-PAGE is a traditional, but also an efficient way to obtain an overview, or better a picture, from an unknown proteome. It delivers a valuable starting point for any proteome investigation within its known limits which typically concern proteins of extreme properties (very acidic, basic or hydrophobic). There are a number of evaluated 2D-PAGE protocols available in the scientific community, but we found it rather hard in the beginning, to efficiently extract the bacterial proteins. Eventually, we were satisfied with a protocol consisting of several cycles of freeze-thawing and mixer mill treatment for cell lysis. As mentioned above, the proteins extracted in this way concentrated in the acidic range of the 2D-gel so that this region was spread further using pI-strips 4-7. The spot pattern among the strains is clearly similar although no

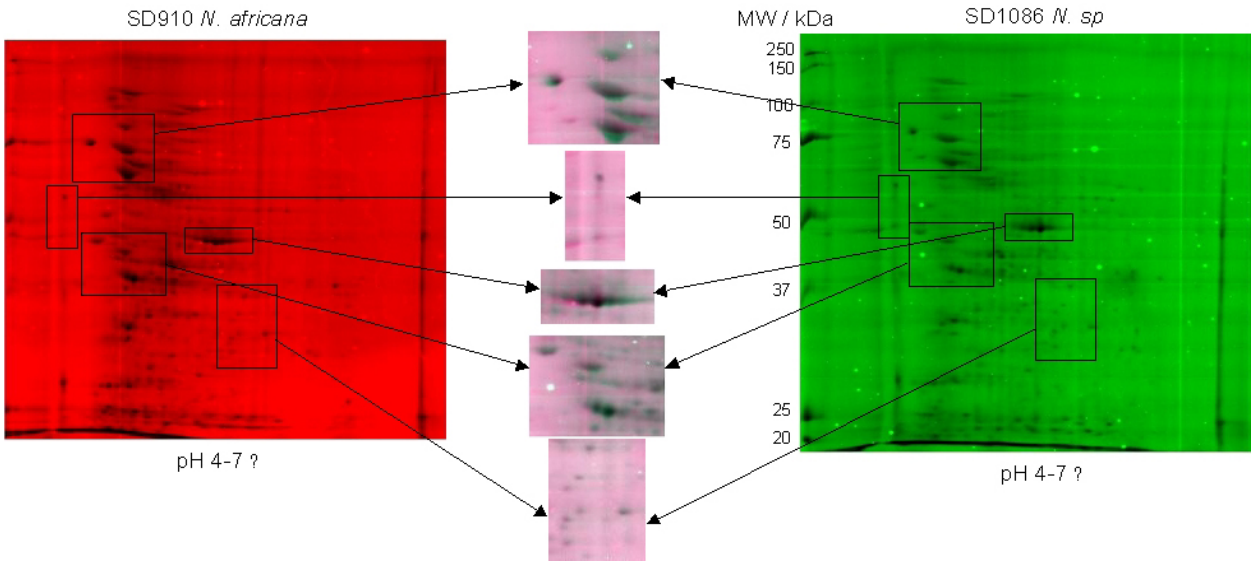

Figure 3. Overlay of gels of *N. africana* SD910 and *N. sp.* SD1086 showing the distribution of common spots.

global overlay could be achieved (apart from "Figure 3"). For such a statistical relevant differential analysis of all strains the extensive and expensive DIGE technique (15) shall be carried out in the future since it eliminates the gel-based variations. Nevertheless, the data presented here show clearly the high similarity among the Nocardia strains. This cannot be said for the comparison to Mycobacterium strains based on gel images (16-17), but no conclusion can be drawn from this fact since experimental protocols vary greatly. For a valid comparison Mycobacterium strains of interest such as *M. tuberculosis* need to be included in abovementioned DIGE study which is one of our future goals.

As is demonstrated in "Figure 1", "Figure 4", "Figure 5" and the Tables well known housekeeping proteins populate the visible proteome. No luxury proteins were found in this study which is expected in such global approach limited by protein solubility and the dynamic range of the PAGE technique (14). The proteins presented in the Tables are all involved in synthesis and processing of DNA, RNA and protein or the major metabolic pathways, that is in the basic functioning of the cell, and they are well represented among the predicted highly expressed genes in *N. farcinica* (13). Particularly chaperones such as HSP60 and 70 and dnaJ are abundant. These molecules are transiently involved in protein and RNA folding and refolding processes after stress (i.e. heat shock). The trigger factor isolated from *N. farcinica* ("Figure 6") participates in protein export and it acts as a chaperone by maintaining the newly synthesized protein in an open conformation (see Q5Z064). Components of ATP-dependent Clp protease were found in *N. africana* SD925, a protein with chymotrypsin-like activity which hydrolyzes proteins to small peptides. They perform chaperone function and play a major role in the degradation of misfolded proteins (Q5Z062, Q9ZCN1).

Other proteins are part of transcriptional or translational processes. DNA-dependent RNA polymerase ("Figure 9") catalyzes the transcription of DNA into RNA using the four ribonucleoside triphosphates as substrates (Q5Z1K9). Elongation factors associate with ribosomes cyclically during the elongation phase of protein synthesis and catalyze the formation of the acyl bond between the incoming amino acid residue and the peptide chain (Q5YPG3). EF-Tu promotes the GTP-dependent binding of aminoacyl-tRNA to the A-site of ribosomes and EF-G the GTP-dependent translocation of the nascent protein chain from the A-site to the P-site of the ribosome (P30768). The identification of regulated EF-Tu in a study of stationary-phase cultures of *M. smegmatis* (17) was assigned significance due to the fact that ribosomal components are known to be expressed in a growth-rate-dependent way. Another protein, methionyl-tRNA formyltransferase modifies the free amino group of the aminoacyl moiety of methionyl-tRNA. The formyl group appears to play a dual role in the initiator identity of N-formylmethionyl-tRNA by promoting its recognition by IF2 and impairing its binding to EF-Tu-GTP (Q819U1). The 30S ribosomal protein S1 picked from the *N. africana* SD925 gel binds mRNA thus facilitating recognition of the initiation point (P46836). An essential enzyme of protein biosynthesis is tyrosyl-tRNA synthetase. It catalyzes the attachment of tyrosine to tRNA(Tyr) in a two-step reaction: tyrosine is first activated by ATP to form Tyr-AMP and then transferred to the acceptor end of tRNA(Tyr) (Q8KA14). Threonyl-tRNA synthetase is an aminoacyl-tRNA synthetase which catalizes the reaction ATP + L-threonine + tRNA(Thr) = AMP + diphosphate + L-threonyl-tRNA(Thr) (Q6FZZ7). Enzymes involved in nucleotide biosynthesis are phosphoribosylaminoimidazole-succinocarboxamide synthase 2 (Q98I23) and dihydroorotate dehydrogenase (Q8ZG91) while ornithine aminotransferase controls arginine catabolism in the amino acid synthesis pathway (Q9K5Z2). Succinylglutamate desuccinylase transforms 2-

Table 1. Common proteins among Nocardia strains as determined by MALDI-TOF peptide mapping with their SwissProt accession number

Elongation factors observed in three different strains

SD910 *N. africana* spot 5 assigned to elongation factor G (EF-G; Q5YPG3; EFG_NOCFA, 76971 Da)

MW	Start		End	Peptide
1014.57	320	–	328	K.IAVHPFFGK.L
1090.54	124	–	132	R.QADKYDVPR.I
1216.65	32	–	41	R.ILFYTGVNYK.I
1429.77	475	–	487	K.VEANVGKPQVAYR.E
1442.68	112	–	123	K.EGVEPQSEQVWR.Q
1772.91	142	–	156	K.LGADFYFTVQTIKDR.L
1779.92	651	–	666	K.ALVPLSEMFGYIGDLR.S
1795.92	651	–	666	K.ALVPLSEMFGYIGDLR.S + Oxidation (M)
1833.97	472	–	487	R.EFKVEANVGKPQVAYR.E
1893.98	233	–	249	K.FFGGEELTIDEIKGAIR.K
2022.08	233	–	250	K.FFGGEELTIDEIKGAIRK.M
2614.30	100	–	123	R.VLDGAVAVFDGKEGVEPQSEQVWR.Q

SD1828 *N. farcinica* spot 8 and *Nocardia sp* N1086 spot 7 assigned to elongation factor Tu.

MW	Start		End	Peptide	SD1828 *N. farcinica* spot 8 EFTU_MYCLE P30768; 43641 Da	*N. sp* N1086 spot 7 EFTU_MYCBO P0A559; 43566 Da
836.49	120	–	126	R.EHVLLAR.Q	X	
904.49	377	–	384	R.FAIREGGR.T	X	
1072.60	226	–	235	R.GTVVTGRVER.G	X	
1100.59	283	–	291	R.GIKREDVER.G	X	
1340.71	365	–	376	K.LIQPVAMDEGLR.F		X
1356.71	365	–	376	K.LIQPVAMDEGLR.F + Oxidation (M)		X
1403.63	48	–	59	K.AFDQIDNAPEER.Q		X
1412.84	127	–	139	R.QVGVPYILVALNK.S	X	X
1680.92	267	–	282	K.LLDQGQAGDNVGLLLR.G	X	X
1692.80	78	–	92	R.HYAHVDAPGHADYIK.N	X	
1730.88	62	–	76	R.GITINISHVEYQTEK.R	X	
1800.89	158	–	173	R.ELLAAQEFDEDAPVVR.V		X
1808.96	9	–	25	R.TKPHVNIGTIGHVDHGK.T	X	
1886.98	62	–	77	R.GITINISHVEYQTEKR.H	X	
2032.97	322	–	336	R.HTPFFNNYRPQFYFR.T	X	X
2880.47	207	–	232	R.ETDKPFLMPVEDVFTITGRGTVVTGR.V + Oxidation (M)	X	X

Chaperone protein dnaK (heat shock protein 70, HSP70, 70 kDa) of two different Mycobacteria strains assigned to spots in four different Nocardia strains

MW	Start		End	Peptide	SD910 *N. africana* spot 8	N317 *N. asteroidis* spot 3	SD925 *N. africana* spot 6	*N. sp* N1086 spot 2
					DNAK_MYCPA Q00488, 66347 Da		DNA_MYCLE P19993, 66507 Da	
931.51	290	–	297	R.QPFQSVVK.D			X	
972.52	280	–	287	R.ITQDLLDR.T			X	
1037.51	85	–	93	K.YTAQEISAR.V	X			X
1145.63	552	–	563	K.TALGGTDISAIK.S			X	X
1188.66	288	–	297	R.TRQPFQSVVK.D				X

1239.69	131	–	142	K.EAGQIAGLNVLR.I	X	X	X	X
1344.70	264	–	274	K.NPLFLDEQLTR.A	X	X		
1566.88	426	–	440	K.LLGSFELTGIPPAPR.G	X	X	X	X
1667.93	127	–	142	R.QATKEAGQIAGLNVLR.I				X
1825.04	361	–	377	K.DVLLLDVTPLSLGIETK.G	X		X	
1906.05	337	–	356	K.GVNPDEVVAVGAALQAGVLK.G			X	
2108.10	44	–	63	R.NGEVLVGQPAKNQAVTNVDR.T		X		X
2322.22	441	–	462	R.GVPQIEVTFDIDANGIVHVTAK.D	X		X	
2346.27	419	–	440	R.EIASHNKLLGSFELTGIPPAPR.G				X
3375.60	396	–	425	R.SETFTTADDNQPSVQIQVYQGEREIAAHNK.L		X		

60 kDa chaperonin 1 (protein Cpn60 1; groEL protein 1; Q5Z1F9; CH601_NOCFA) assigned to spot 11 in SD925 N. africana and spot 5 in N. sp N1086

MW	**Start**		**End**	**Peptide**	***N. africana* spot 11**	***N. sp* N1086 spot 5**
888.46	394	–	402	R.VEDAVSAAK.A		X
964.49	276	–	283	K.APFFGDRR.K	X	X
1244.67	379	–	390	K.VGAATETALKER.K	X	X
1333.70	2	–	12	M.AKQIEFDEKAR.R	X	
1335.56	350	–	360	R.EIEATDSDWDR.E	X	
1451.86	22	–	35	K.LADAVKVTLGPRGR.H	X	
1464.74	1	–	12	-.MAKQIEFDEKAR.R	X	X
1574.81	42	–	57	K.AFGGPTVTNDGVTIAR.D	X	X
1817.86	326	–	343	K.DETTIIDGAGTAEDIAAR.A	X	X
1914.96	58	–	74	R.DIDLEDPFENLGAQLVK.S	X	X
2455.35	444	–	467	R.KALEAPLFWIASNAGLDGAVVVSK.V	X	X
2736.40	473	–	498	K.EGFNAATLSYGDLLTDGVVDPVKVTR.S	X	X

60 kDa chaperonin 2 (protein Cpn60 2; groEL protein 2; heat shock protein 60; Q9AFA6; CH602_NOCFA, 56373 Da). In the case of N317 N. asteroidis the match to the 65 kDa antigen of M. paratuberculosis (P42384; CH602_MYCPA; 56477 Da) was better rated than the hit for the to Nocardia protein

MW	**Start**		**End**	**Peptide**	**SD910 spot 10**	**SD925 spot 3**	**SD925 spot 26**	**SD1828 *N. farcinica* spot 4**	**N317 spot 4 MYCPA**	**N317 spot 4 NOCFA**	**N317 spot 5**	***N. sp* N1086 spot 4**
					N. africana				***N. asteroidis***			
939.56	230	–	237	K.DLLPLLEK.V								X
986.54	18	–	27	R.GLNSLADAVK.V	X	X						X
1066.49	4	–	12	K.TIAYDEEAR.R		X	X	X			X	X
1122.63	391	–	399	R.KHRIEDAVR.N	X			X				
1202.64	271	–	282	K.SVAVKAPGFGDR.R				X				
1222.59	4	–	13	K.TIAYDEEARR.G	X	X	X	X		X	X	X
1222.59	3	–	12	K.TIAYDEEARR.G					X			
1265.63	2	–	12	M.AKTIAYDEEAR.R		X		X				
1272.67	379	–	390	K.AGAATEVELKER.K	X	X	X	X			X	X
1279.74	105	–	117	R.NVAAGANPLGLKR.G	X	X	X	X		X	X	X
1279.74	104	–	116	R.NVAAGANPLGLKR.G					X			
1400.76	379	–	391	K.AGAATEVELKERK.H				X				
1421.73	2	–	13	M.AKTIAYDEEARR.G	X			X				

1453.88	225	–	237	K.VSTVKDLLPLLEK.V	X			X				
1476.69	195	–	207	K.GYISGYFVTDAER.Q					X			
1502.70	196	–	208	K.GYISGYFVTDPER.Q				X		X	X	X
1528.79	43	–	57	K.WGAPTITNDGVSIAK.E	X	X	X	X		X	X	X
1552.77	1	–	13	-.MAKTIAYDEEARR.G				X				X
1578.88	101	–	116	R.EGLRNVAAGANPLGLK.R				X				
1641.87	450	–	465	K.QIAFNAGLEPGVVAEK.V	X	X	X	X			X	X
1656.88	42	–	57	K.KWGAPTITNDGVSIAK.E							X	
1677.84	4	–	17	K.TIAYDEEARRGLER.G	X			X				
1699.72	349	–	362	R.TEIENSDSDYDREK.L								X
1772.96	209	–	224	R.QEAVLEDPYILLVGSK.V	X	X	X	X				X
1866.88	192	–	207	R.FDKGYISGYFVTDAER.Q					X			
1892.89	193	–	208	R.FDKGYISGYFVTDPER.Q			X	X		X	X	
1974.02	58	–	74	K.EIELEDPYEKIGAELVK.E	X	X	X			X		X
1974.02	57	–	73	K.EIELEDPYEKIGAELVK.E					X			
2074.05	80	–	100	K.TDDVAGDGTTTATVLAQALVR.E	X	X	X			X	X	X
2074.05	79	–	99	K.TDDVAGDGTTTATVLAQALVR.E					X			
2202.15	79	–	100	K.KTDDVAGDGTTTATVLAQALVR.E	X	X	X	X		X	X	X
2202.15	78	–	99	K.KTDDVAGDGTTTATVLAQALVR.E					X			
2226.00	349	–	366	R.TEIENSDSDYDREKLQER.L	X		X	X			X	
2272.22	320	–	341	R.KVVVTKDETTIVEGAGDAEAIK.G			X					
2357.24	321	–	343	K.VVVTKDETTIVEGAGDAEAIKGR.V				X				
2421.36	442	–	465	R.VALSAPLKQIAFNAGLEPGVVAEK.V				X				
2430.26	319	–	342	R.KVVVTKDETTIVEGAGDSDAIAGR.V					X			
2449.36	441	–	464	R.VALEAPLKQIAFNGGLEPGVVAEK.V					X			
2547.24	141	–	166	K.EQIAATAGISAGDASIGELIAEAMDK.V + Oxidation (M)							X	
2567.24	170	–	192	K.EGVITVEESNTFGLQLELTEGMR.F + Oxidation (M)								X
2629.39	75	–	100	K.EVAKKTDDVAGDGTTTATVLAQALVR.E				X				
2657.43	499	–	524	R.SALQNAASIAALFLTTEAVVADKPEK.A	X	X	X			X	X	X
2657.40	78	–	103	K.KTDDVAGDGTTTATVLAQALVREGLR.N					X			
2818.61	238	–	264	K.VIQAGKPLLIIAEDVEGEALSTLVVNK.I								X
2831.42	141	–	169	K.EQIAATAGISAGDASIGELIAEAMDKVGK.E + Oxidation (M)							X	
3010.45	466	–	495	K.VSNLEAGHGLNADSGEYEDLLAAGVADPVK.V				X				
3366.67	466	–	498	K.VSNLEAGHGLNADSGEYEDLLAAGVADPVKVTR.S				X				

DNA-directed RNA polymerase alpha chain (EC 2.7.7.6; RNAP alpha subunit; transcriptase alpha; RPOA_NOCFA, Q5Z1K9; 37963 Da)

MW	Start		End	Peptide	SD910 spot 24	SD925 spot 16	*SD1828 N.* spot 9	*farcinica* spot 10	N317 *N. asteroidis* spot 9
					N. africana				
1060.54	279	–	287	R.TESDLLDIR.N	X	X			X
1065.58	174	–	182	K.VTYKVEATR.V			X	X	

1098.63	134	–	142	K.LEIELVVER.G		X			
1117.65	214	–	223	K.TLVELFGLAR.E	X	X	X	X	X
1187.64	143	–	153	R.GRGYVPAVQNK.A			X	X	
1265.67	267	–	278	R.EGVHTVGELVAR.T	X	X	X	X	X
1271.70	71	–	81	K.EDVTDIILNLK.G		X			
1283.74	132	–	142	K.GKLEIELVVER.G		X			X
1329.75	162	–	173	R.IPVDSIYSPVLK.V	X	X	X		X
1421.77	266	–	278	K.REGVHTVGELVAR.T		X			
1484.86	41	–	55	R.TLLSSIPGAAVTSIR.I	X	X	X	X	X
1563.81	187	–	199	R.TDFDRLILDVETK.N		X	X	X	X
1577.85	174	–	186	K.VTYKVEATRVEQR.T			X	X	
1610.87	56	–	70	R.IDGVLHEFTTVPGVK.E	X	X			X
1634.83	279	–	292	R.TESDLLDIRNFGQK.S			X	X	
1908.91	82	–	98	K.GLVVSSEEDEPVTMYVR.K	X	X	X	X	X
1924.91	82	–	98	K.GLVVSSEEDEPVTMYVR.K + Oxidation (M)			X	X	X
2071.13	154	–	173	K.ASGAEIGRIPVDSIYSPVLK.V			X	X	X
2099.10	1	–	18	-.MLISQRPTLTEEVIAENR.S			X	X	
2110.07	21	–	39	K.FTIEPLEPGFGYTLGNSLR.R	X	X			
2115.10	1	–	18	-.MLISQRPTLTEEVIAENR.S + Oxidation (M)			X	X	
2266.17	21	–	40	K.FTIEPLEPGFGYTLGNSLRR.T	X				
2325.20	19	–	39	R.SKFTIEPLEPGFGYTLGNSLR.R	X	X	X	X	X

Enolase (EC 4.2.1.11; 2-phosphoglycerate dehydratase, 2-phospho-D-glycerate hydrolyase; Q5YQ30; ENO_NOCFA; 44966 Da)

MW	**Start**		**End**	**Peptide**	**SD910 *N. africana* spot 25**	***N. sp* N1086 spot 8**	***N. sp* N1086 spot 9**
955.55	2	–	10	M.AIIEQVGAR.E	X	X	X
994.49	415	–	423	R.YAGDVAFPR.F	X	X	X
1045.50	405	–	414	R.IEDALGDSAR.Y		X	X
1172.60	180	–	189	R.WGAEVYHALK.A	X		X
1249.63	120	–	131	R.AAAESSGLELFR.Y	X	X	X
1423.85	105	–	119	R.LGANALLGVSLAVAR.A	X	X	X
1627.83	88	–	102	R.TVDQVLLDLDGTPDK.S		X	
1714.89	305	–	319	K.IQLVGDDLFVTNPER.L	X	X	X
1779.88	35	–	52	R.AAVPSGASTGEHEAVELR.D	X	X	X
1870.96	88	–	104	R.TVDQVLLDLDGTPDKSR.L	X	X	X
1898.96	17	–	34	R.GNPTVEVEIALDDGTLTR.A	X	X	X
2079.10	336	–	354	K.VNQIGTLTETLDAVELAHR.N	X	X	X
2161.07	365	–	386	R.SGETEDTTIADLAVAVGSGQIK.T		X	X
2263.21	66	–	87	K.AVEGVLDEIAPAVIGLDAVEQR.T	X	X	X
2280.07	35	–	57	R.AAVPSGASTGEHEAVELRDGGDR.Y		X	
2441.28	305	–	326	K.IQLVGDDLFVTNPERLEDGIAK.G	X	X	X

Table 2. Proteins assigned in the five strains studied

SD910 *N. africana*					
Spot #	**Protein name**	**Abbreviation**	**SwissProt accession #**	**MW [Da]**	**Confidence**
5	Elongation factor G (EF-G)	EFG_NOCFA	Q5YPG3	76971	+++
8	Chaperone protein dnaK (heat shock 70 kDa protein, HSP70)	DNAK_MYCPA	Q00488	66347	++
10	60 kDa chaperonin 2 (protein Cpn60 2, groEL protein 2, heat shock protein 60)	CH602_NOCFA	Q9AFA6	56373	+++
12	Putative phosphoribosylaminoimidazole-succinocarboxamide synthase 2 (EC 6.3.2.6)	PUR72_RHILO	Q98I23	35069	++
22	Dihydroorotate dehydrogenase (EC 1.3.3.1, dihydroorotate oxidase, DHOdehase, DHODase	PYRD_YERPE	Q8ZG91	36808	++
24	DNA-directed RNA polymerase alpha chain, EC 2.7.7.6, RNAP alpha subunit, transcriptase alpha chain)	RPOA_NOCFA	Q5Z1K9	37963	+++
25	Enolase (EC 4.2.1.11, 2-phosphoglycerate dehydratase, 2-phospho-D-glycerate hydro-lyase)	ENO_NOCFA	Q5YQ30	44966	++
28, 41	Human keratin				+++
SD925 *N. africana*					
Spot #	**Protein name**	**Abbreviation**	**SwissProt accession #**	**MW [Da]**	**Confidence**
3	60 kDa chaperonin 2 (Protein Cpn60 2, groEL protein 2, heat shock protein 60)	CH602_NOCFA	Q9AFA6	56373	++
6	Chaperone protein dnaK, heat shock protein 70, heat shock 70 kDa protein, HSP70)	DNAK_MYCLE	P19993	66507	++
8	30S ribosomal protein S1	RS1_MYCLE	P46836	53246	++
11	60 kDa chaperonin 1 (Protein Cpn60 1) (groEL protein 1)	CH601_NOCFA	Q5Z1F9	55769	++
16	DNA-directed RNA polymerase alpha chain (EC 2.7.7.6, RNAP alpha subunit)	RPOA_NOCFA	Q5Z1K9	37963	+++
18	ATP-dependent Clp protease ATP-binding subunit clpX	CLPX_RICPR	Q9ZCN1	46878	+
20	Tyrosyl-tRNA synthetase (EC 6.1.1.1, tyrosine--tRNA ligase, TyrRS)	SYY_BUCAP	Q8KA14	48954	+
22	Succinylglutamate desuccinylase (EC 3.1.-.-)	ADTE_PHOLL	Q7N2H1	36866	+
25	ATP-dependent Clp protease proteolytic subunit 3 (EC 3.4.21.92, endopeptidase Clp 3)	CLPP3_NOCFA	Q5Z062	24131	+++
26	60 kDa chaperonin 2, protein Cpn60 2, groEL protein 2, heat shock protein 60)	CH602_NOCFA	Q9AFA6	56373	+++
SD1828 *N. farcinica*					
Spot #	**Protein name**	**Abbreviation**	**SwissProt accession #**	**MW [Da]**	**Confidence**
1	Trigger factor (TF)	TIC_NOCFA	Q5Z064	50244	+++
4	60 kDa chaperonin 2 (Protein Cpn60 2, groEL protein 2, heat shock protein 60)	CH602_NOCFA	Q9AFA6	56373	+++
6	Transaldolase A (EC 2.2.1.2)	TALA_ECO57	P0A869	35636	++
8	Elongation factor Tu (EF-Tu)	EFTU_MYCLE	P30768	43641	++
9	DNA-directed RNA polymerase alpha chain (EC 2.7.7.6, RNAP alpha subunit, transcriptase alpha chain)	RPOA_NOCFA	Q5Z1K9	37963	++
10	DNA-directed RNA polymerase alpha chain (EC 2.7.7.6, RNAP alpha subunit, transcriptase alpha chain)	RPOA_NOCFA	Q5Z1K9	37963	+++
14	Phosphate import ATP-binding protein pstB (EC 3.6.3.27, phosphate-transporting ATPase)	PSTB_BACTN	Q8A853	28403	+
N317 *N. asteroidis*					
Spot #	**Protein name**	**Abbreviation**	**SwissProt accession #**	**MW [Da]**	**Confidence**
3	Chaperone protein dnaK (heat shock protein 70, heat shock 70 kDa protein, HSP70)	DNAK_MYCPA	Q00488	66347	++
4	60 kDa chaperonin 2 (protein Cpn60 2, groEL protein 2)	CH602_MYCPA CH602_NOCFA	P42384 Q9AFA6	56477 56373	++
5	60 kDa chaperonin 2 (protein Cpn60 2, groEL protein 2)	CH602_NOCFA	Q9AFA6	56373	+++
7	1-deoxy-D-xylulose 5-phosphate reductoisomerase (EC 1.1.1.267, DXP reductoisomerase)	DXR_SALCH	Q57T35	43311	+
9	DNA-directed RNA polymerase alpha chain (EC 2.7.7.6, RNAP alpha subunit)	RPOA_NOCFA	Q5Z1K9	37963	+++
10	Chaperone protein dnaJ	DNAJ_CHLTE	Q8KCD8	43224	++
***N. sp* SDN1086**					
Spot #	**Protein name**	**Abbreviation**	**SwissProt accession #**	**MW [Da]**	**Confidence**
7	Isocitrate dehydrogenase kinase/phosphatase (EC 2.7.1.116, EC 3.1.3.-)	ACEK_PHOLL	Q7MZA0	68747	++
2	Chaperone protein dnaK (heat shock protein 70, heat shock 70 kDa protein, HSP70)	DNAK_MYCLE	P19993	66507	++
3	Threonyl-tRNA synthetase (EC 6.1.1.3, threonine--tRNA ligase, ThrRS)	SYT_BARQU	Q6FZZ7	75656	++
4	60 kDa chaperonin 2 (protein Cpn60 2, groEL protein 2, heat shock protein 60)	CH602_NOCFA	Q9AFA6	56373	+++
5	60 kDa chaperonin 1 (protein Cpn60 1, groEL protein 1)	CH601_NOCFA	Q5Z1F9	55769	++
7	Elongation factor Tu (EF-Tu)	EFTU_MYCBO	P0A559	43566	+
8	Enolase (EC 4.2.1.11, 2-phosphoglycerate dehydratase, 2-phospho-D-glycerate hydro-lyase)	ENO_NOCFA	Q5YQ30	44966	+++
9	Enolase (EC 4.2.1.11, 2-phosphoglycerate dehydratase, 2-phospho-D-glycerate hydro-lyase)	ENO_NOCFA	Q5YQ30	44966	+++
11	Methionyl-tRNA formyltransferase (EC 2.1.2.9)	FMT_BACCR	Q819U1	34755	++
12	Cytidylate kinase (EC 2.7.4.14, CK, cytidine monophosphate kinase, CMP kinase)	KCY_MYCPN	P75308	24552	+
13	Ornithine aminotransferase (EC 2.6.1.13, ornithine--oxo-acid aminotransferase)	OAT_BACHD	Q9K5Z2	44298	+

For supplementary Tables 3 and 4 see http://ifg-izkf.uni-muenster.de/ (Proteomics-Projects-NoDaMS). Subtle differences in confidence in the match are expressed with +++ (very), ++ (high confidence), + (good peak assignment). For all other search results see the NoDaMS database

N-succinylglutamate into succinate and glutamate in amino acid degradation 8Q7N2H1).

Of the metabolic enzymes enolase is one of the most abundant ones and found in the soluble proteome across species. It is active in carbohydrate degradation catalyzing the reaction 2-phospho-D-glycerate = phosphoenolpyruvate + water (Q5YQ30). Transaldolase is important for the balance of metabolites in the pentose-phosphate pathway (P0A869). Isocitrate dehydrogenase

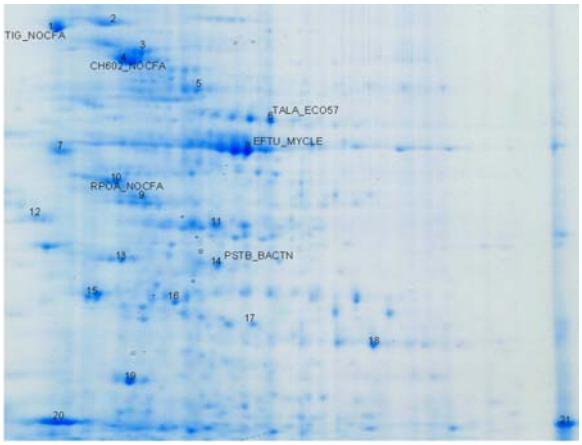

Figure 4. Representative spot assignments for *N. farcinica* SD1828, pH 4-7, MW 25-75 kDa. For labels see Tables and Suppl..

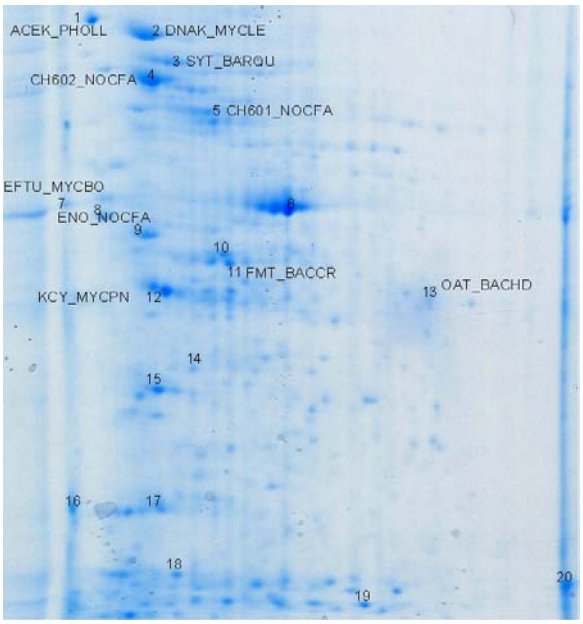

Figure 5. Representative spot assignments for *N. sp.* SD1086, pH 4-7, MW 25-75 kDa. For labels see Tables and Suppl..

kinase/phosphatase is a bifunctional enzyme which can phosphorylate or dephosphorylate isocitrate dehydrogenase Q7MZA0). In this regulatory mechanism bacteria can bypass the Krebs cycle in response to the source of carbon. 1-deoxy-D-xylulose 5-phosphate reductoisomerase is part of isoprenoid biosynthesis and catalyzes the NADP-dependent rearrangement and reduction of 1-deoxy-D-xylulose-5-phosphate to 2-C-methyl-D-erythritol 4-phosphate (Q57T35). Typically, also proteins participating in phosphate transfer are detectable on gels. The phosphate import ATP-binding protein pstB is part of the ABC transporter complex pstSACB and responsible for energy coupling to the transport system (Q8A853). Cytidylate kinase is a member of the adenylate kinase superfamily and an enzyme for the reaction ATP + (d)CMP = ADP + (d)CDP.

MALDI-MS is a technique which relies on protein identification based on the search of peptide mass data against *in silico* digested protein databases. Due to the statistical nature of this approach, false-positive answers cannot be entirely excluded. The experienced user limits this effect by sensibly adjusting the search space and monitoring the results. Interestingly, a high search score does not necessarily reflect a correct match and a correct hit might exhibit a low score. Therefore, great care and experience has produced the data presented. Nevertheless, there are some matches labelled by a lower confidence level in "Table 2" which are worrysome despite good peak assignment. Therefore, we would like to stress that MALDI-MS results, in principle, always require validation by other methods such as fragmentation analysis or Western blotting. This is not made clear in most publications based on peptide mass fingerprinting. In particular with proteomes not completely sequenced or even unknown, the danger of misassignments increases in addition to the usual difficulties with modifications, isoforms and unspecific cleavage. It is reasons like this which are responsible for the fact that answers for gel-separated proteins of the *N. farcinia* proteome contain sequences of other bacteria. In the case of EFTU_MYCLE it is a simple mathematical imbalance which made the very homologues Mycobacterium sequence (see BLAST comparison in the Supplementary Material) the number one hit in the list. In other cases such as PSTB_BACTN where the peak assignment is good but the species not correct the user is warned that this search answer needs further validation efforts. In addition, the susceptibility of the individual proteins to the digestion enzyme, usually trypsin, is a factor influencing the quality of the peptide map and therefore identification. Particularly smaller proteins may only generate few peptides or exhibit spectra with one or two major peaks ("Figure 7"). Multiple passes changing the enzyme might help in difficult cases. However, the fine-tuning is step three. Global overviews are mostly satisfied with LC-MS/MS and those experiments are in progress. All data obtained so far are made available for free re-evaluation with fresh sequence data from the NoDaMS website in an effort to progress proteomic analysis of Nocardia.

6. CONCLUSION

Proteomic data of Nocardia are sparse despite the organisms importance. A comparative proteomic study of five Nocardia strains was conducted for the first time. Gel electrophoresis served to produce patterns of the soluble proteins which are concentrated in the acidic range of the gel. It confirmed the expected similarity of the strains. The next step is a statistical relevant comparison based on DIGE where effects due to gel-to-gel variations are minimized in order to filter out the few differences possibly relevant for virulence. Landmarks include a number of housekeeping proteins such as chaperones and metabolic enzymes and they show great homology among strains. A large body of MS data is made available in the NoDaMS

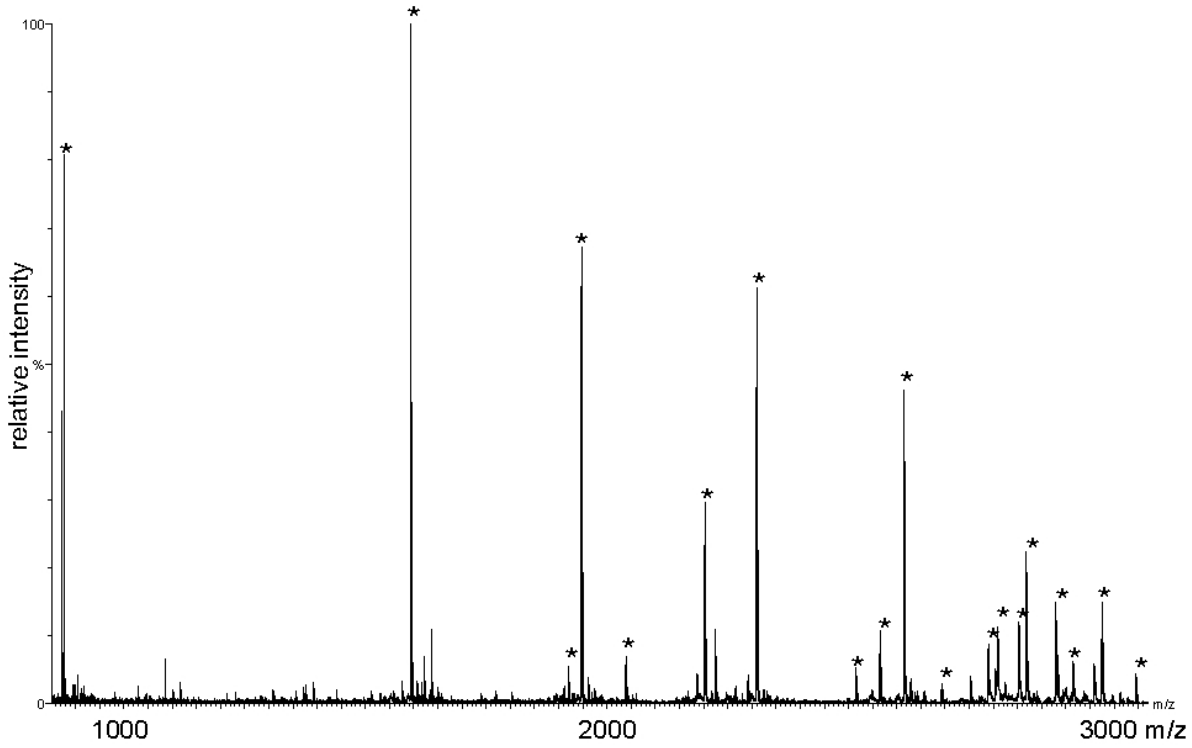

Figure 6. Representative peptide map for *N. farcinica* SD1828, spot 1 – TIG_NOCFA.

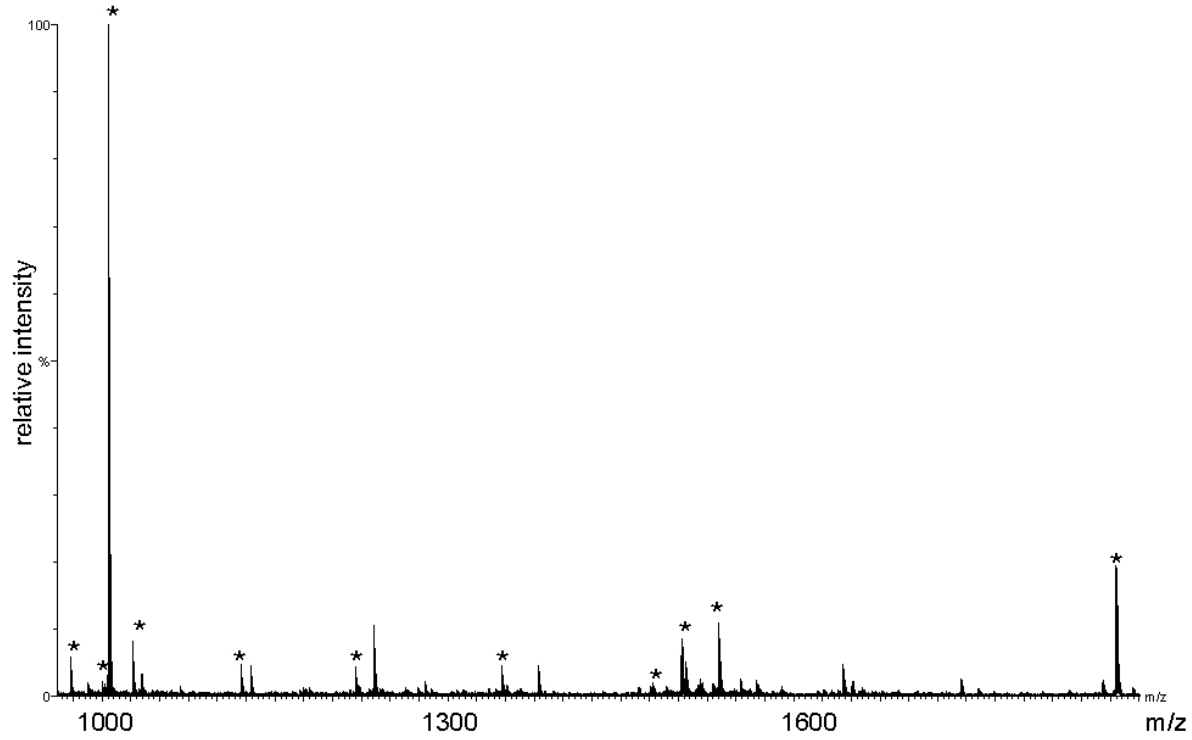

Figure 7. Representative peptide map for a digest showing one prominent peak *N. africana* SD925, spot 8.

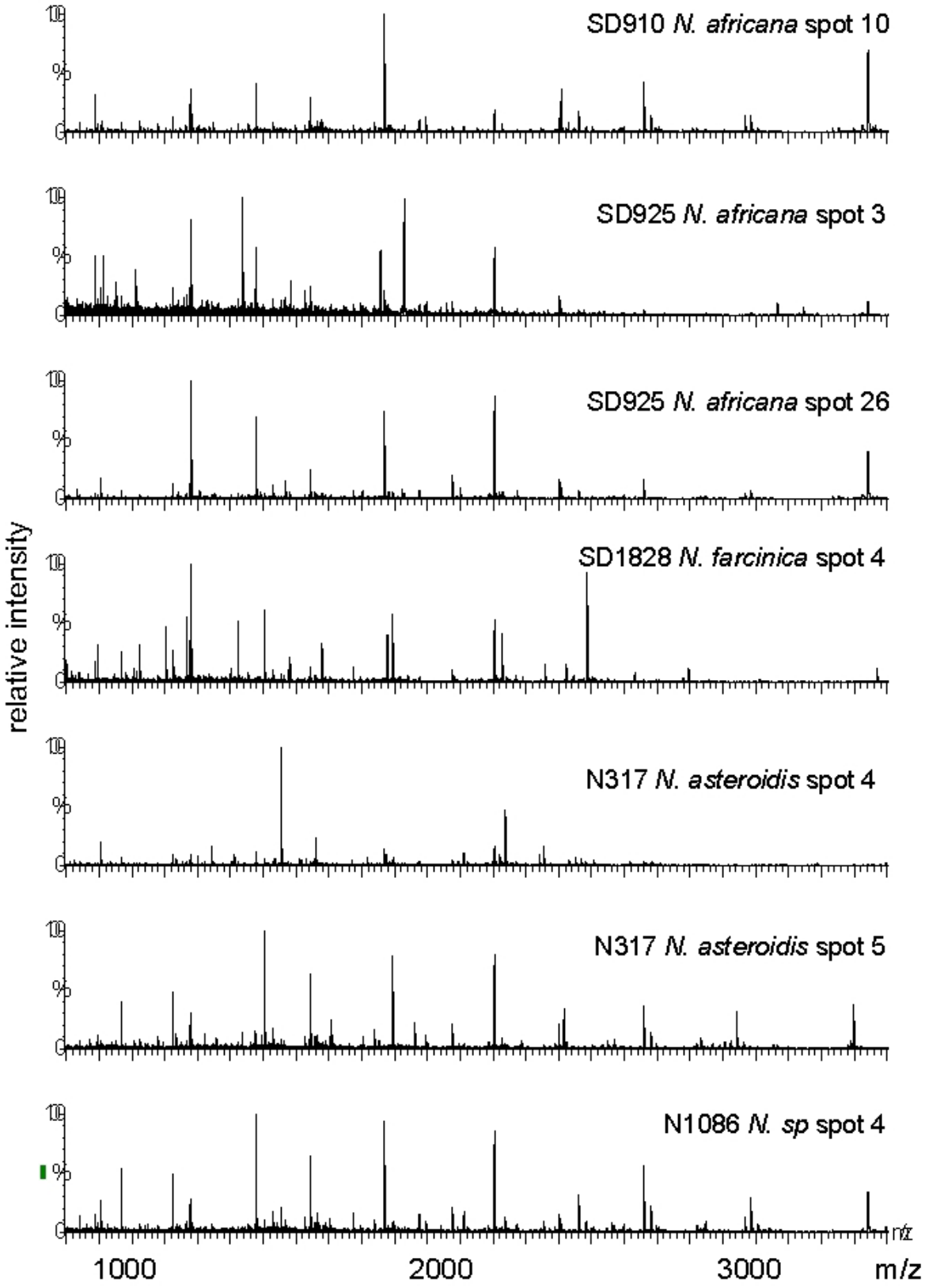

Figure 8. Comparison of peptide maps of 60 kDa chaperonin 2 in different strains and spots.

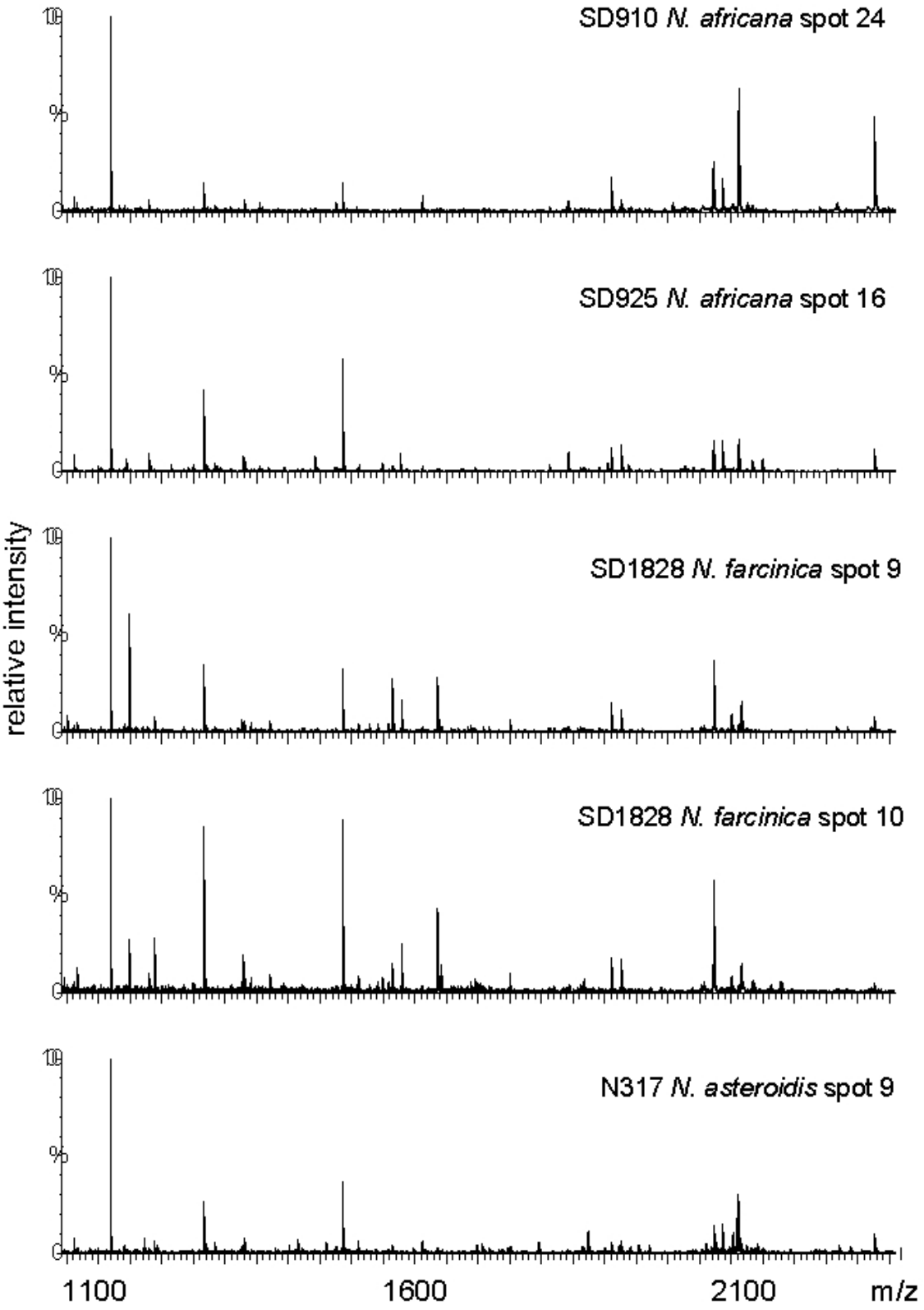

Figure 9. Comparison of peptide maps of transcriptase alpha in different strains and spots.

database which was created to share peak lists for future investigations using advanced protein databases. This data review comprises the first comprehensive summary of proteomic data of Nocardia.

7. ACKNOWLEDGEMENT

The authors are grateful to A. Mahgoub, M. M. Ahmed and Dr. Amanwel for their help during the preparation of this project. M. Poetter (Microbiology, University of Muenster) deserves our thanks for sharing his knowledge on bacterial lysis. We gratefully acknowledge Prof. Goodfellow (Newcastle) for sharing *N. asteroides* N317 (ATTCC 19247 T). The study was supported by the University of Sudan for Science and Technology, German Academic Exchange Service, ICGEB, and DFID Higher Education.

8. REFERENCES

1. Valerie A Orchard: The ecology of Nocardia and related taxa. *Zentralbl Bakteriol Suppl* 11, 167-180 (1981)

2. Michael Goodfellow and ST Williams: Ecology of actinomycetes. *Ann Rev Microbiol* 37, 189-216 (1983)

3. Michael Goodfellow: The family Nocardiaceae. In: The prokaryotes, Vol 2., 2nd Ed., Eds: Balows, A, Truper, HG, *Springer*, NY, 1188-213 (1992)

4. Blaine L Beaman and LoVelle Beaman: Differences in the interactions of *Nocardia asteroides* with macrophage, endothelial, and astrocytoma cell lines. *Clin Microbiol Rev* 7, 203-264 (1994)

5. http://www.cdc.gov/ncidod/dbmd/diseaseinfo/nocardiosis_t.htm

6. Jun Ishikawa, Atsushi Yamashita, Yuzuru Mikami, Yasutaka Hoshino, Haruyo Kurita, Kunimoto Hotta, Tadayoshi Shiba and Masahiara Hattori: The complete genomic sequence of *Nocardia farcinica* IFM 10152. *PNAS* 101(41), 14925-14930 (2004)

7. Mogahid E. ElHassan, Kanury VS Rao, Zaved Siddique, Dinesh S Kumar, Rashmi Tickoo, Nageeb S Saeed, Moawia M Mukhtar and Mohamed E Hamid: Proteomics of *Nocardia africana* (SD769) recently isolated from patients with pulmonary infection in Sudan. *Biomacromol Mass Spectrom* 1(3) (2007), in print

8. Mohamed E Hamid, Luis Maldonado, Ghada S Sharaf Eldin, Maha F Mohammed, Nageeb S Saeed and Michael Goodfellow: *Nocardia africana* sp. nov., a new pathogen isolated from patients with pulmonary infections. *Clin Microbiol* 39(2), 625-30 (2001)

9. Mohamed E Hamid and Michael Goodfellow: *In-vitro* antimicrobial susceptibility profile of soil isolates of bovine farcy organisms. *Revue Elev Med Vet Pays Trop* 50, 5-9 (1997)

10. AW Bauer, WM Kirby, JC Sherris and M Turck: Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* 45, 493-496 (1966)

11. Ashley A Miles and SS Misra: The estimation of the bacterial power of the blood. *J Hygiene* 38, 732-749 (1938)

12. Max Koltzscher, Claudia Neumann, Simone Koenig and Volker Gerke: Ca^{2+}-dependent binding and activation of dormant ezrin by dimeric S100P. *Mol Biol of the Cell* 14(6), 2372-2384 (2003)

13. Gang Wu, Lei Nie and Weiwen Zhang: Predicted highly expressed genes in *Nocardia farcinica* and the implication for its primary metabolism and nocardial virulence. *Antonie van Leeuwenhock* 89(1), 135-146 (2005)

14. Simone Koenig: The crux of proteome analysis. *Biomacromol Mass Spectrom* 1(2), 135-141 (2007)

15. Andrew Alban, Stephen O. David, Lennart Bjorkesten, Christian Andersson, Erik Sloge, Steve Lewis and Ian Curie: A novel experimental design for comparative two-dimensional gel analysis: 2D-difference gel electrophoresis incorporating a pooled standard. *Proteomics* 3(1), 36-44 (2003)

16. Xiu-Yun He, Yu-Hui Zhuang, Xiao-Gang Zhang and Guo-Li Li: Comparative proteome analysis of culture supernatant proteins of *Mycobacterium tuberculosis* H37Rv and H37Ra. *Microbes and Infection* 5, 851-856 (2003)

17. Marian C. J. Blokpoel, Marjan J. Smeulders, Julia A. M. Hubbard, Jacquie Keer and Huw D. Williams: Global analysis of proteins synthesized by *Mycobacterium smegmatis* provides direct evidence for physiological heterogeneity in stationary-phase cultures. *J Bacteriol* 187(19), 6691-6700 (2005)

Key Words: Mass spectrometry, MALDI, Nocardia, Proteomics, Chaperones, Trigger factor, Heat-shock factor, Enolase, Database

Send correspondence to: Dr. Simone Koenig, Integrated Functional Genomics, Röntgenstr. 21, 48149 Münster, Germany, Tel: 0049-251-8357164, Fax: 0049-251-8357255, E-mail: koenigs@uni-muenster.de

http://www.bioscience.org/current/vol13.htm