Proteomic analysis of an extreme halophilic archaeon, 

*Halobacterium sp. NRC-1*

Young Ah Goo, Eugene C. Yi, Nitin S. Baliga, Weiguo A. Tao, Min Pan, Ruedi Aebersold, David R. Goodlett, Leroy Hood, Wailap V. Ng*

Institute for Systems Biology, 1441 N 34th St, Seattle, WA 98103

**Running title:** Proteomic analysis of *Halobacterium* NRC-1

*Corresponding Author

Phone: 206-732-1412; Fax: 206-732-1260; Email: vng@systemsbiology.org
Summary

*Halobacterium* sp. NRC-1 insoluble membrane and soluble cytoplasmic proteins were isolated by ultracentrifugation of whole cell lysate. Using a Finnigan LCQ-DECA mass spectrometer equipped with a C18 trap ESI-emitter/micro-LC column, a number of trypsin generated peptide tags from 426 unique proteins were identified. This represents approximately one-fifth of the theoretical proteome of *Halobacterium*. Of these, 232 proteins were found only in the soluble fraction, 165 were only in the insoluble membrane fraction, and 29 were in both fractions. There were 72% and 61% previously annotated proteins identified in the soluble and membrane protein fractions, respectively. Interestingly, fifty-seven of previously un-annotated proteins found only in *Halobacterium* NRC-1 were identified. Such proteins could be interesting targets for understanding unique physiology of *Halobacterium* NRC-1. A group of proteins involved in various metabolic pathways were identified among the expressed proteins suggesting these pathways were active at the time the cells were collected. This data containing a list of expressed proteins, their cellular locations and biological functions could be used in future studies to investigate the interaction of the genes and proteins in relation to genetic or environmental perturbations.
Introduction

Since the completion of the first bacterial genome, *Haemophilus influenzae* (1), more than one hundred microbial genome sequences including *Halobacterium* sp. NRC-1 (2,3) have been determined (www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html). These sequences constitute the primary digital information for global understanding of physiology, pathogenicity, and molecular machineries essential for the survival or adaptation of the organisms in different environmental conditions. The *Halobacterium* genome consists of a 2,014-kb large chromosome and two smaller chromosomes (191 kb and 365 kb) encoding approximately 2,630 putative protein genes (2,3). Among these, 41% matched to genes of known function in public databases. The predicted proteome is highly acidic with a median isoelectric point of 4.9 (4). The high negative surface charge of predicted proteins provides a major mechanism for stabilizing the proteins in nearly saturated intracellular salinity where other conventional proteins would become denatured (4,5).

The archaeon *Halobacterium* sp. NRC-1 provides a relatively simple model for understanding a complex system of how cells adjust to various environmental stimuli. *Halobacterium* flourishes in extremely saline environments (> 4M salts) and its metabolism is subject to fluctuations in sunlight, oxygen, temperature, nutrients and salinity. *Halobacterium* thrives in this harsh environment by appropriately tuning its extraordinary physiology in response to different environmental stimuli. For example, it can relocate, in search of favorable environments, using sensors that can discriminate beneficial and detrimental spectra of light (6,7,8), aerotaxis transducer (HtrVIII) (9), and buoyant gas-filled vesicles (10,11). The *Halobacterium* transducer, HtrVIII, combines subunit I core structures of eukaryotic cytochrome c oxidase and eubacterial methyl-accepting chemotaxis proteins to mediate aerotaxis (9). One of
interesting features of Halobacterium is its ability to survive aerobically as a chemoheterotroph and anaerobically using light and/or arginine as energy sources. Halobacterium sp. derives energy from light by its retinal-containing light-driven ion transporters, bacteriorhodopsin and halorhodopsin (12-16). Halobacterium can also ferment arginine via the arginine deiminase pathway to yield one mole of ATP for each mole of fermented arginine (17,18).

Its intriguing physiology together with the availability of a complete genome sequence, led us to catalogue via a simplified shotgun proteomic methodology (19,20) the proteins expressed by Halobacterium sp. NRC-1 in membrane and cytoplasmic compartments. In addition, the proteins involved in metabolic pathways in Halobacterium sp. NRC-1 under standard culture conditions were investigated. Herein we present the results of our initial investigation of the Halobacterium proteome using a simple shotgun proteomic approach that involves bulk digestion of co-purified proteins with trypsin followed by a single stage of microcapillary HPLC electrospray ionization tandem mass spectrometry (LC-ESI-MS/MS) analysis of peptides using an ion trap mass spectrometer. The data contains a list of expressed proteins derived by searching peptide tandem mass spectra against the theoretical protein database of Halobacterium sp. NRC-1 (2,3), their cellular locations (i.e. membrane, cytoplasm or both) deduced from subcellular fractionation prior to proteome analysis, and the putative biological functions of proteins.

Experimental procedure

Strain. Halobacterium sp. NRC-1 (ATCC700922) was cultured at 37 °C in basal salt medium containing 1% peptone (Oxoid, Hampshire, England) and trace metals as previously described (21).

Protein preparation. Membrane and soluble-cytoplasmic proteins were isolated using a
protocol modified from a halophiles laboratory manual and Oesterhelt (21,22). One liter of *Halobacterium sp. NRC-1* culture was grown to OD$_{600} = \sim 2.0$ and pelleted by centrifugation at 7,500 rpm at 4 °C for 10 minutes. Pellets were resuspended in 20 ml basal salt solution containing 0.5 mg each of DNaseI and RNaseA, and 1 mM of proteinase inhibitor, phenylmethylsulfonyl fluoride (PMSF). Cells were lysed by osmotic shock against a 40X excess of deionized water within a dialysis tubing bag (spectra/por® membrane MWCO: 3,500, Spectrum, CA). Cell debris was removed by centrifugation at 10,000g for 30 minutes. The remaining cell lysates were then separated into the soluble and membrane fractions by ultracentrifugation at 53,000g for 2 hours. The membrane fraction, a pellet at the bottom of the tube, and the soluble fraction, the aqueous supernatant portion, were then collected. The membrane was loaded on top of 30 % sucrose cushion and ultracentrifuged at 53,000g at 10 °C over night. The membrane fraction was collected and washed three times in 10 ml basal salt solution using a hand held electrical homogenizer (Tissue-Tearor, Fisher). Membrane proteins were then collected by centrifugation at 53,000g for 2 hours at 10 °C. The pellet was resuspended in residual basal salt solution and then transferred to a microcentrifuge tube. The residual aqueous basal salt solution was removed by a brief spin at 14,000 rpm. The soluble protein fraction was dialyzed against five changes of 100X volume of deionized water at 4 °C to reduce the salt concentration that in excess might inhibit the protease reaction and mass spectrometry analysis.

**Protease digestion.** One hundred µg of proteins were digested with 2 µg of trypsin (Promega, Madison, WI) in 50 mM sodium bicarbonate (pH 8.3) at 37 °C overnight. Soluble proteins were lyophilized after digestion. Membrane proteins were digested in the presence of 0.5% SDS to
aid solubilization. After the protease reaction, SDS was removed by precipitating proteins with 70 % acetone or by chromatography using cation exchange cartridge (OASIS MCX, Waters, MA) according to manufacturer’s procedure. The proteins were lyophilized and stored at –80 °C and resuspended in 100 µl of 0.4 % acetic solution prior to mass spectrometer analysis.

**Tandem mass spectrometry.** Trypsin digested peptides were analyzed by µLC-ESI-MS/MS using an LCQ-DECA mass spectrometer (Thermo Finnigan, San Jose, CA) equipped with a C18 trap ESI-emitter/micro-LC column. Trypsin digested peptides (2 µg) were loaded to a Hewlett Packard/Agilent 1100 Series HPLC system using a Famos Autosampler (LC Packings, San Francisco). The peptides bound to the C18 matrix were eluted by acetonitrile gradient (5 % to 35 %) by mixing acetonitrile with 0.4 % acetic acid in water. The eluted peptides were injected into the mass spectrometer by nano electrospray ionization (19,23). Mass spectra were acquired by data-dependent ion selection from a full range as well as discrete and narrow survey scan m/z ranges to increase the number of identifications. Proteins were identified from tandem mass spectra using the SEQUEST (24) database search engine to search against the *Halobacterium* NRC-1 predicted protein database (3).

**Membrane domain prediction.** *Halobacterium* putative proteins were analyzed for the presence of transmembrane domains using the TMpred (25) and TMHMM programs (26,27). The TMpred program predicts membrane spanning regions (MSRs) and their orientation based on the statistical analysis of TMbase, a database of transmembrane proteins and their helical membrane-spanning domains. The prediction is based on an algorithm using a combination of several weight-matrices for investigating the local properties of amino acid sequences. The
program TMHMM, on the other hand, takes a global approach to determine the topology of an entire protein based on Hidden Markov models. The stand-alone TMpred program was installed on a SUN Microsystem Enterprise 420R server. TMHMM (v. 2.0) was run through the web interface (www.cbs.dtu.dk/services/TMHMM).

Results

Membrane protein prediction. The *Halobacterium* NRC-1 genome encodes 2,682 putative protein-coding genes (3). Among these, 2,413 genes are unique. The TMHMM program predicted 544 membrane proteins containing 1 to 24 MSR(s), among which 163 were annotated proteins, 122 were conserved hypothetical proteins (CHP), and 259 were hypothetical proteins (HP). On the other hand, TMpred detected 929 membrane proteins containing 1 to 22 MSR(s) with total score more than 1000, among which 377 were annotated proteins, 202 were CHP, and 350 were HP. A score >500 is considered to be statistically significant in TMpred prediction (25). TMpred also detected all 544 membrane proteins predicted by the TMHMM program with a minimal TMpred total score of 1194.

Mass spectrometry peptide analysis. Two µg of trypsin digested peptide mixtures from the membrane and soluble proteins were analyzed by µLC-ESI-MS/MS with the following different m/z ranges from which ions were selected for collision-induced dissociation (CID): 1 (400 to 2000 m/z), 4 (400~800, 800~1200, 1200~1600, and 1600~2000 m/z), or 16 (400~500, 500~600, 600~700, ..., and 1900~2000 m/z). Using a different m/z range has been shown to increase the number of novel peptides selected for CID (28,29). The tandem mass spectra were analyzed using the SEQUEST database search program (24) with the *Halobacterium* NRC-1 protein...
database. Search results were processed using the INTERACT web interface, a software tool that allows internet-based data display, data filtering, and data sorting (30). Recently developed statistical modeling algorithms to compute probabilities associated with peptide (PeptideProphet\textsuperscript{TM}) (31) and protein (ProteinProphet\textsuperscript{TM}) (32) sequence assignments that distinguish correct from incorrect database search results were used to validate the search results. These tools allowed assigning probabilities to all identifications and offering standardized interpretation of results by reducing the need for manual verification. In particular, these tools enabled rapid and objective evaluation of large proteomic datasets. A detailed application of these tools has been recently published (33). More information on these applications can be found on the Proteomics pages at http://www.systemsbiology.org/ and they are open source. In this study, we report proteins with probability at least 0.5. Probability 0.5 means that according to the statistical model, the sequence match given is 50% likely to be correct. These resulted in the identification of 426 proteins with false-positive rate of 3.7% (Figure 1).

**Proteins in membrane fraction.** In the tandem mass spectrometry analysis, 165 proteins were identified only in the membrane fraction but not in the soluble protein fraction (Table 1). There were 100 (60.6%) annotated proteins, 29 (17.6%) conserved hypothetical proteins (CHP), and 36 (21.8%) hypothetical proteins (HP). The transmembrane domain prediction program TMHMM predicted 90 (54.5%) proteins containing 1 to 22 membrane domains and TMpred detected 123 (74.5%) proteins with 1 to 22 membrane domains with scores greater than 1000. No membrane domain was detected by either program in 31 (18.8%) proteins.

**Proteins in soluble fraction.** A total of 232 proteins were identified only in the soluble fraction
but not in the membrane fraction. Of these, 168 (72.4%) were annotated proteins, 45 (19.4%) were CHP, and 19 (8.2%) were HP (Table 2). TMHMM detected only three membrane proteins and each contained only a single putative membrane domain. TMpred detected 40 (17.2%) putative membrane proteins with scores greater than 1000.

**Proteins in both membrane and soluble fractions.** A total of twenty-nine proteins were identified in both soluble and membrane fractions. These included protein components involved in large complex structures such as the ribosome, flagella, and gas vesicle (Table 3). There were 27 annotated proteins and 2 HP. TMHMM predicted three of the proteins contain one membrane domain, and TMpred predicted that eight of the proteins contain one or two membrane domains with >1000 score. No membrane domain was predicted in 20 (69%) proteins.

**Discussion**

In this study, we applied a simplified shotgun proteomic approach using LC-ESI-MS/MS and computational analysis to characterize the peptides in complex mixtures of trypsin digested membrane and soluble proteins. While this is a powerful technique for rapidly screening the peptide components and by inference the parent proteins in a sample, there are certain limitations to this approach which include: 1) peptide ion selection for CID during LC introduction is "top-down" and to some degree random (29), meaning that peptides which ionize well and that are from the more abundant proteins in the original mixture are the most likely to be selected; 2) for a protein to be identified, the peptide tandem mass spectrum used in the database search must be of sufficient "quality" (which in part is related to the abundance of the peptide) to match a sequence in the database; 3) the absence of a sequence in the database for
which a high quality peptide tandem mass spectrum is generated may lead to a false-positive because the software can generate a best-fit to a highly similar sequence that is present, although the probability scoring routine used minimizes this; 4) high versus low protein sequence coverage lends more weight to a protein identification and may be an indication of its relative abundance amongst proteins present in the original mixture; 5) our search results were based solely on matching predicted genome sequences (i.e., post-translational modifications of amino acids were not considered). The simplified shotgun proteomic approach was chosen instead of two-dimensional polyacrylamide gel electrophoresis – mass spectrometry methods (34,35) primarily because LC-MS/MS allows direct analysis of hydrophobic membrane proteins, and also because it serves as a relatively rapid screen of expressed proteins.

A total of 401 chromosome proteins and 25 minichromosome pNRC100 and pNRC200 proteins were identified. In order to obtain functional and physiological information regarding these expressed proteins, 295 of the expressed proteins with putative functions were searched against the KEGG Enzymes/Compounds/Genes Pathway Database (http://www.genome.ad.jp/kegg-bin/mk_point_html). A number of metabolic pathways showed more than 50% of their members present in the group of expressed proteins, and thus suggests such pathways were active at the time the cells were collected or identified proteins are constitutively expressed (Table 4).

Proteome analysis data supports the previous observation that pNRC100 and pNRC200 contain indispensable gene sequences (2,3). Among the approximately forty putative genes on the pNRC100 and pNRC200 likely to be essential or important for cell viability, ten proteins were identified in this study. These include the pNRC200 proteins, arginine deiminase (ArcA), ornithine carbamoyltransferase (ArcB), carbamate kinase (ArcC), aspartate carbamoyltransferase
catalytic subunit (PyrB), aspartate carbamoyltransferase regulatory chain (PyrI), cytochrome d oxidase chain I (CydA; also on pNRC100), sn-glycerol-1-phosphate dehydrogenase (GldA), arginine-tRNA synthetase (ArgS), Na+/H+ antiporter (NhaC3), and glycerol-3-phosphate-binding protein (UgpB).

There were twenty-nine proteins with tryptic peptides identified from both membrane and soluble protein fractions. These included cell surface glycoprotein (Csg), fifteen ribosomal proteins, four flagella proteins, a gas vesicle protein, GvpC, and eight other proteins. Among these proteins, only Csg, FlaA1a, and FlaB1 contained both TMHMM and TMpred membrane domains and DppA, FlaA2, FlaB3, SdhA, and NrdB2 contained TMpred membrane domains (scores > 1000) but no TMHMM predicted membrane domain. It is unusual to find flagellar proteins in the soluble fraction. They may be present in the soluble fraction as flagella fragments that have been detached from the basal body (36) where they are connected to the cytoplasmic membrane, or alternatively, as unassembled precursor proteins. In *Halobacterium*, the formation of gas vesicles is induced under low oxygen conditions, enabling cells to float to the surface and grow phototrophically (10,11). Interestingly, the major gas vesicle protein, GvpA, was identified only in the membrane fraction. The GvpC peptides detected in the soluble protein mixture might have come from GvpC molecules detached from gas vesicle surfaces.

Ribosomal proteins were found in soluble (thirty-two), membrane (four) or both fractions (fifteen). Since ribosomal protein complexes are large structures, some ribosomes might have co-purified with the membrane during ultracentrifugation while most of them remained in the soluble fraction. The cell surface glycoprotein (Csg) was one of the most frequently identified proteins in our analysis. Of the approximately 800 tandem mass spectra matched to Csg, 43% were from the membrane fractions. Csg may be observed in the soluble fraction if it detached
from the cell membrane surface, or possibly as very small membrane fragments suspended in the supernatant after ultracentrifugation.

We next evaluated the extent to which our biochemical fractionation was successful in segregating membrane from soluble proteins. To do this we examined the peptides from the top three membrane and soluble proteins separately. Top protein candidates were those proteins with the highest number of matched tandem mass spectra and for which all tandem mass spectra had a \( \geq 0.9 \) probability of being correctly matched. The top three soluble proteins were CctB, CctA and ArcB. Of these three proteins, peptides from CctB were selected for CID 715 times in the soluble fraction but only 5 in the membrane fraction indicating that very little of this soluble protein was found in the membrane fraction. By this same measure, CctA and ArcB were also well segregated to the soluble fraction as demonstrated by finding that 638 out of 646 tandem mass spectra were selected only in the soluble fraction for CctA and 436 out of 444 for ArcB. Likewise, the top three membrane proteins, YqgG, DppD, and Vng1802H also appeared to be well segregated to the membrane fraction because 605 out of 681, 224 out of 244 and 110 out of 124 of their tandem mass spectral matches were found only in the membrane fraction. This test suggested that the ultracentrifuge purification procedure effectively partitioned most of the proteins to either soluble or membrane fraction, reflecting their original cellular location.

There are several sets of proteins of interest in *Halobacterium sp.* NRC-1 important to its physiology. They include the six TATA-box binding protein (TBP), seven transcription factor B (TFB) proteins, seventeen Htr signal transducer family, and gas vesicle proteins (2,3). In this study, only the expression of TbpE and TfbG were detected among the multiple putative transcription factor proteins. It raises the question of whether the other transcription factors would be expressed under different physiological conditions or growth phases. This observation
may also be due to the protein being expressed at levels too low for detection. Interestingly, TbpE is located on the chromosome, while all the other Tbps are found on the pNRC100 or pNRC200 minichromosomes. Another possibility is that TbpE is the major TBP controlling transcription with TfbG. At least eleven (Htr1, Htr2, Htr3, Htr4, Htr5, Htr6, Htr8, Htr13, Htr14, Htr15, and Htr16) of the seventeen signal transducers were detected. Expression of multiple transducer proteins suggests dynamic cellular functions in response to rapidly changing environmental conditions.

The gas vesicle gene clusters, gvpACNO and gvpDEFGHIJKLM, on pNRC100 had been extensively studied by genetic approaches to identify the essential genes required for gas vesicle biogenesis (11,37-42). Through analyses of spontaneous gas vesicle deficient mutants caused by transposition of insertion sequence (ISH) elements, site-specific linker insertion, and deletion mutants, at least ten of the fourteen gvp genes were determined to be necessary for gas vesicle synthesis. Our tandem mass spectrometry analysis identified the GvpA, GvpH, and GvpN in the membrane fraction, GvpO in the soluble fraction and GvpC in both fractions. Interestingly, GvpH, and GvpO were not identified from purified gas vesicles by μLC-ESI-MS/MS analysis in our recent study (unpublished data). This suggests GvpH and GvpO may not be gas vesicle structural proteins.

Four rhodopsins involved in photobiology (6-8,12-16,43); bacteriorhodopsin (Bop), halorodopsin (Hop), sensory rhodopsin I (Sop1), and sensory rhodopsin II (Sop2) were all identified in the membrane fraction.

Our study compared the results of two transmembrane helix prediction programs used to predict the presence of membrane domains in the expressed proteins. In most cases, TMpred predicted a larger number of MSRs than TMHMM. A previous comparison of fourteen
membrane protein prediction programs on 883 defined MSRs of 188 well-characterized proteins indicated the TMHMM is currently the best performing transmembrane prediction program (44). Thus, in our study, TMHMM was primarily used to predict the membrane domains while TMpred was used as a supplemental program to support the results generated by TMHMM.

The computationally predicted membrane domains were somewhat in agreement with the results of the mass spectrometry. More than 99% of the soluble fraction proteins do not contain TMHMM predicted MSRs, and 82.8% do not contain TMpred predicted MSRs with total score greater than 1000. Accordingly, 74.5% of the membrane fraction proteins contained one or more TMpred predicted MSR(s) with scores greater than 1000, and 54.5% contained at least one TMHMM predicted MSR(s). The low percentage may due to the presence of non-integral or peripheral membrane proteins without MSRs that are unlikely to be detected by either prediction program.

Analysis of a genome sequence can provide a list of all predicted genes, yet the information regarding the expression of transcriptome or proteins is difficult to measure. It is unlikely that all of the predicted putative proteins are expressed. Among the expressed proteins, post-translational or chemical modifications could lead to the formation of new or functionally different proteins. Our study was based on matching predicted proteins of genome sequences and post-translational modifications were not considered. The failure to detect a protein in this study thus does not mean that it is absent. Sophisticated developments in mass spectrometers and enhanced sample preparation protocols will enable more expressed proteins to be identified in the future.

The integrated computational and mass spectrometry data analysis in this study gives a better understanding of the expressed proteins and cellular locations of membrane and cytoplasmic proteins in *Halobacterium sp.* NRC-1. The information obtained from this study will be useful
for subsequent work, such as expression of proteins of interest to investigate certain aspects of *Halobacterium* biology. It can also be used with integrated microarray and ICAT data (45) for systems biological approaches (46) to study novel biological processes in halophiles.
References

1. Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., et al. (1995) Science 269, 496-512

2. Ng, W. V., Ciufo, S. A., Smith, T. M., Bumgarner, R. E., Baskin, D., Faust, J., Hall, B., Loretz, C., Seto, J., Slagel, J., et al. (1998) Genome Res. 8, 1131-1141

3. Ng, W. V., Kennedy, S. P., Mahairas, G. G., Berquist, B., Pan, M., Shukla, H. D., Lasky, S. R., Baliga, N. S., Thorsson, V., Sbrogna, J., et al. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 12176-12181

4. Kennedy, S. P., Ng, W. V., Salzberg, S. L., Hood, L., and DasSarma, S. (2001) Genome Res. 11, 1641-1650

5. Lanyi, J. K. (1974) Bacteriol. Rev. 38, 272-274

6. Bogomolni, R. A., and Spudich, J. L. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 6250-6254

7. Spudich, J. L., and Bogomolni, R. A. (1984) Nature 312, 509-513

8. Spudich, J. L. (1993) J. Bacteriol. 175, 7755-7761

9. Brooun, A., Bell, J., Freitas, T., Larsen, R. W., and Alam, M. (1998) J. Bacteriol. 180, 1642-1646

10. DasSarma, S., Damerval, T., Jones, J. G., and Tandeau de Marsac N. (1987) Mol. Microbiol. 1, 365-370

11. DasSarma, S., Arora, P., Lin, F., Molinari, E. and Yin, L. R. (1994) J. Bacteriol. 176, 7646-7652

12. Oesterhelt, D., Meentzen, M., and Schuhmann, L. (1973) Eur. J. Biochem. 40, 453-463
13. Dunn, R., McCoy, J., Simsek, M., Majumdar, A., Chang, S. H., Rajbhandary, U. L., and Khorana, H.G. (1981) Proc. Natl. Acad. Sci. U. S. A. 78, 6744-6748

14. Schobert, B., and Lanyi, J. K. (1982) J. Biol. Chem. 257, 10306-10313

15. Krebs, M. P., and Khorana, H.G. (1993) J. Bacteriol. 175, 1555-1560

16. Kolbe, M., Besir, H., Essen, L. O, and Oesterhelt, D. (2000) Science 288, 1390-1396

17. Hartmann, R., Sickinger, H. D., and Oesterhelt, D. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 3821-3825

18. Ruepp, A., and Soppa, J. (1996) J. Bacteriol. 178, 4942-4947

19. Figeys, D., Ducret, A., Yates, J. R. 3rd, and Aebersold, R. (1996) Nat. Biotechnol. 14, 1579-1583

20. Link, A. J., Eng, J., Schieltz, D. M., Carmack, E., Mize, G. J., Morris, D. R., Garvik, B. M., Yates, J. R. 3rd. (1999) Nat. Biotechnol. 17, 676-682

21. Oesterhelt, D., and Stoeckenius, W. (1974) Methods Enzymol. 31, 667-678

22. DasSarma, S., Robb, F. T., Place, A. R., Sowers, K. R., Schreier, H. J., and Fleischmann, E. M. (1995) Archaea: a laboratory manual: halophiles, Cold Spring Harbor Laboratory Press, Plainview, NY

23. Ducret, A., Van Oostveen, I., Eng, J. K., Yates, J. R. 3rd and Aebersold, R. (1998) Protein Sci. 7, 706-719

24. Eng J. K., McCormack, A. L., Yates, J. R. 3rd. (1994) J. Am. Soc. Mass Spectrom. 5, 976-989

25. Hofmann, K., and Stoffel, W. (1993) Biol. Chem. Hoppe-Seyler 374, 166

26. Sonnhammer, E. L., von Heijne, G., and Krogh, A. (1998) Nucleic Acids Res. 26, 320-322
27. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. (2001) J. Mol. Biol. 305, 567-580
28. Spahr, C. S., Davis, M. T., McGinley, M. D., Robinson, J. H., Bures, E. J., Beierle, J., Mort, J., Courchesne, P. L., Chen, K., Wahl, R. C., Yu, W., Luethy, R., and Patterson, S. D. (2001) Proteomics 1, 93-107
29. Yi, E. C., Marelli, M., Lee, H., Purvine, S. O., Aebersold, R., Aitchison, J. D., and Goodlett, D. R. (2002) Electrophoresis 23, 3205-3216
30. Han, D. k., Eng, J., Zhou, H., and Aebersold, R. (2001) Nat. Biotechnol. 19, 946-951
31. Nesvizhskii, A. L., Keller, A., Kolker, E., and Aebersold, R. (2003) Anal. Chem. In press
32. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Anal. Chem. 74, 5383-5392
33. von Haller, P. D., Yi, E. C., Donohoe, S., Vaughn, K., Keller, A., Nesvizhskii, A. I., Eng, J., Li, X., Goodlett, D. R., Aebersold, R., and Watts, J. D. (2003) Mol. Cell. Proteomics In press
34. Garrels, J. I., McLaughlin, C. S., Warner, J. R., Futzer, B., Latter, G. I., Kobayashi, R., Schwender, B., Volpe, T., Anderson, D. S., Mesquita-Fuentes, R., and Payne, W. E. (1997) Electrophoresis 18, 1347-1360
35. Gygi, S. P., Corthals, G. L., Zhang, Y., Rochon, Y., and Aebersold, R. (2000) Proc. Natl. Acad. Sci. U. S. A. 97, 9390-9395
36. Tarasov, V. Y., Pyatibratov, M. G., Tang, S. L., Dyall-Smith, M., and Fedorov, O. V. (2000) Mol. Microbiol. 35, 69-78
37. Jones, J. G., Hackett, N. R., Halladay, J. T., Scothorn, D. J., Yang, C. F., Ng, W. L., and
DasSarma, S. (1989) Nucleic Acids Res. 17, 7785-7793

38. Jones, J. G., Young, D. C., and DasSarma, S. (1991) Gene 102, 117-122

39. Ng, W. L., Kothakota, S., and DasSarma, S. (1991) J. Bacteriol. 173, 1958-1964

40. Halladay, J. T., Ng, W. L., and DasSarma, S. (1992) Gene 119, 131-136

41. Halladay, J. T., Jones, J. G., Lin, F., MacDonald, A. B., and DasSarma, S. (1993) J. Bacteriol. 175, 684-692

42. Pfeifer, F., Gregor, D., Hofacker, A., Plosser, P., and Zimmermann, P. (2002) J. Mol. Microbiol. Biotechnol. 4, 175-181

43. Haupts, U., Haupts, C., and Oesterhelt, D. (1995) Proc. Natl. Acad. Sci. U. S. A. 92, 3834-3838

44. Moller, S., Croning, M. D. and Apweiler, R. (2001) Bioinformatics 17, 646-653

45. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) Nat. Biotechnol. 17, 994-999

46. Baliga, N. S., Pan, M., Goo, Y. A., Yi, E. C., Goodlett, D.R., Dimitrov, K., Shannon, P., Aebersold, R., Ng, W. V., and Hood, L. (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 14913-14918
Figure Legends

Figure 1. Identification of membrane and soluble proteins from *Halobacterium sp.* NRC-1. ProteinProphet™ was used to identify 426 proteins with a probability ≥ 0.5. Of these 232 proteins were found only in the soluble fraction, 165 were only in the insoluble membrane fraction, and 29 were in both fractions.
Table 1. Proteins identified in membrane fraction.

The functional categories of proteins, the gene ID numbers, the probability calculated by empirical statistical model, the abbreviated name of protein, putative function, and the predicted number of transmembrane domains by TMHMM and TMpred are listed. Conserved hypothetical proteins (CHP), and hypothetical proteins (HP) are listed under the functional categories.

| Functional Categories | Gene ID       | Probability | Protein | Putative Function                  | TMHMM | TMpred |
|-----------------------|---------------|-------------|---------|------------------------------------|-------|--------|
| Amino acid metabolism | VNG0104G      | 1           | SerA3   | phosphoglycerate dehydrogenase     | 0     | 2      |
|                       | VNG1814G      | 1           | CarB    | carbamoyl-phosphate synthase large subunit | 0     | 0      |
|                       | VNG2120G      | 1           | YusM    | proline dehydrogenase             | 0     | 0      |
|                       | VNG2418G      | 1           | AspC1   | aspartate aminotransferase         | 0     | 1*     |
|                       | VNG2421G      | 1           | Hal     | O-acetyl homoserine                | 0     | 2      |
| Cell envelope component | VNG0180G      | 0.98        | Hop     | halorhodopsin                      | 7     | 7      |
|                       | VNG0321G      | 1           | Ids     | bifunctional short chain isoprenyl diphosphate synthase | 1     | 1      |
|                       | VNG1187G      | 1           | Pan1    | membrane protein                   | 0     | 1      |
|                       | VNG1467G      | 1           | Bop     | bacteriorhodopsin                  | 6     | 7      |
|                       | VNG1660G      | 1           | Sop1    | sensory rhodopsin I                | 7     | 7      |
|                       | VNG1764G      | 0.98        | Sop2    | sensory rhodopsin II               | 7     | 7      |
|                       | VNG6265G      | 1           | YcdH    | adhesion protein                   | 0     | 1      |
| Cellular process      | VNG0129G      | 1           | Hsp4    | heat shock protease protein        | 4     | 4      |
|                       | VNG0355G      | 1           | Htr14   | Htr14 transducer                   | 1     | 2      |
|                       | VNG0375G      | 1           | SecE    | protein translocase                | 1     | 1      |
|                       | VNG0614G      | 0.79        | Htr16   | Htr16 transducer                   | 1     | 1      |
|                       | VNG0793G      | 1           | Htr6    | Htr6 transducer                    | 2     | 2      |
|                       | VNG0806G      | 1           | Htr4    | Htr4 transducer                    | 2     | 3      |
|                       | VNG0971G      | 0.99        | CheA    | chemotaxis protein                 | 0     | 2      |
|                       | VNG0976G      | 1           | cheW1   | chemotaxis protein                 | 0     | 0      |
|                       | VNG1523G      | 1           | Htr8    | Htr8 transducer                    | 5     | 5      |
|                       | VNG1529G      | 1           | Htr1    | Htr1 transducer                    | 1     | 2      |
|                       | VNG1760G      | 1           | Htr5    | Htr5 transducer                    | 2     | 2      |
|                       | VNG1765G      | 1           | Htr2    | Htr2 transducer                    | 1     | 3      |
|                       | VNG1801G      | 1           | Hsp1    | small heat shock protein           | 0     | 0      |
|                       | VNG1856G      | 1           | Htr3    | Htr3 transducer                    | 1     | 1      |
|                       | VNG1987G      | 1           | SecD    | protein-export membrane protein    | 6     | 5      |
|                       | VNG1988G      | 1           | SecF    | protein-export membrane protein    | 6     | 5      |
|                       | VNG5025G      | 1           | GvpH1   | GvpH protein, cluster A            | 0     | 0      |
|                       | VNG6024G      | 1           | GvpN1   | GvpN protein, cluster A            | 0     | 0      |
|                       | VNG5030G      | 1           | GvpA1   | GvpA protein, cluster A            | 0     | 0      |
|                       | VNG6029G      | 1           | GvpA2   | GvpA protein, cluster A            | 0     | 0      |
|                       | VNG5033G      | 1           | GvpN1   | GvpN protein, cluster A            | 0     | 0      |
|                       | VNG6032G      | 1           | GvpN2   | GvpN protein, cluster A            | 0     | 0      |
| Cofactor Metabolism | VNG1635G | 0.98 | CbiM | cobalamin biosynthesis protein | 6 | 6 |
|---------------------|----------|------|------|--------------------------------|---|---|
|                     | VNG1776G | 1    | NirH | heme biosynthesis protein      | 0 | 0 |
| DNA Replication,    | VNG0884G | 1    | Top6A| DNA topoisomerase VI subunit A | 0 | 0 |
| Repair, and         | VNG2372G | 1    | Rad24c| DNA repair protein            | 0 | 0 |
| Recombination       | VNG2473G | 1    | RadA1| DNA repair protein            | 0 | 1 |
| Energy Metabolism   | VNG0412G | 1    | FolP | dihydropteroate synthase      | 0 | 0 |
|                     | VNG0637G | 1    | NdhG5| NADH dehydrogenase/oxidoreductase | 0 | 1 |
|                     | VNG0639G | 1    | NdhG4| NADH dehydrogenase/oxidoreductase | 6 | 7 |
|                     | VNG0646G | 1    | NuoL | F420H2:quinone oxidoreductase | 17 | 15 |
|                     | VNG0648G | 1    | NdhG3| NADH dehydrogenase/oxidoreductase | 11 | 10 |
|                     | VNG0657G | 1    | CoxA2| cytochrome c oxidase subunit I | 13 | 14 |
|                     | VNG0665G | 1    | CoxB1| cytochrome c oxidase subunit II | 3 | 3 |
|                     | VNG0891G | 1    | YjI  | NADH dehydrogenase            | 0 | 1 |
|                     | VNG1308G | 1    | SdhB | succinate dehydrogenase subunit B | 0 | 1 |
|                     | VNG1498G | 1    | CelM | endoglucanase                  | 0 | 0 |
|                     | VNG2141G | 1    | AtpC | H^+-transporting ATP synthase subunit C | 0 | 1* |
|                     | VNG2143G | 1    | AtpK | H^+-transporting ATP synthase subunit K | 2 | 2 |
|                     | VNG2193G | 1    | CoxA1| cytochrome c oxidase subunit I | 10 | 11 |
|                     | VNG2195G | 1    | CoxB2| cytochrome c oxidase subunit II | 0 | 0 |
|                     | VNG5055G, | 1    | CydA | cytochrome d oxidase chain I | 9 | 11 |
|                     | VNG5242G, |      |      |                                 |   |   |
|                     | VNG6055G, |      |      |                                 |   |   |
|                     | VNG6473G |       |      |                                 |   |   |
| Miscellaneous       | VNG0249G | 1    | Fbr  | cytochrome-like protein         | 0 | 2 |
|                     | VNG0303G | 1    | Lon  | ATP-dependent proteinase homolog | 1 | 2 |
|                     | VNG0459G | 1    | NodP | nodulation protein             | 0 | 0 |
|                     | VNG0540G | 1    | Imp  | immunogenic protein            | 0 | 1* |
|                     | VNG0620G | 1    | Edp  | proteinase IV homolog          | 1 | 3 |
|                     | VNG0635G | 1    | NoIB | NADH dehydrogenase/oxidoreductase-like protein | 3 | 3 |
|                     | VNG0640G | 0.98 | NoID | NADH dehydrogenase/oxidoreductase-like protein | 0 | 0 |
|                     | VNG0795G | 1    | HcpC | halocyanin precursor-like      | 0 | 2 |
|                     | VNG1428G | 1    | HtA  | Htr-like protein               | 7 | 7 |
|                     | VNG1932G | 1    | NolA | NADH dehydrogenase/oxidoreductase-like protein | 0 | 2 |
|                     | VNG2086G | 1    | Hpb  | possible phosphate binding protein | 0 | 1 |
|                     | VNG2196G | 1    | HcpB | halocyanin precursor-like      | 3 | 3 |
|                     | VNG2308G | 1    | Hlp  | hemolysin protein              | 2 | 3 |
|                     | VNG2320G | 1    | HdrD | heterodisulfid reductase       | 5 | 6 |
|                     | VNG2358G | 1    | AppA | oligopeptide binding protein   | 0 | 1 |
|                     | VNG6301G | 1    | Aph  | alkaline phosphatase           | 0 | 1 |
| Nucleotide Metabolism | VNG0632G | 1    | PurK | phosphoribosylaminomimidazole carboxylase ATP binding subunit | 0 | 0 |
|                     | VNG1408G | 1    | Ush  | UDP-sugar hydrolase            | 2 | 2 |
| Gene       | Accession | Predicted Protein Function                                                                 |
|------------|-----------|-------------------------------------------------------------------------------------------|
| VNG2507G   | 0.99      | PyrD dehydroorotate dehydrogenase                                                         |
| VNG0177G   | 1         | Rpl15e 50S ribosomal protein L15E                                                          |
| VNG0551G   | 1         | Rpl44e 50S ribosomal protein L44E                                                          |
| VNG1433G   | 1         | Rps17e 30S ribosomal protein S17E                                                           |
| VNG1866G   | 0.98      | Map methionyl aminopeptidase                                                               |
| VNG2469G   | 1         | Rpl39e 50S ribosomal proteins L39E                                                          |
| VNG0002G   | 1         | YvrO amino acid ABC transporter, ATP-binding protein                                        |
| VNG0174G   | 1         | Cat1 cationic amino acid transporter                                                       |
| VNG0453G   | 1         | PstA2 phosphate ABC transporter permease                                                    |
| VNG0455G   | 1         | PstC2 phosphate ABC transporter permease                                                    |
| VNG0457G   | 1         | PhoX phosphate ABC transporter periplasmic phosphate-binding                               |
| VNG0794G   | 1         | YufN ABC transporter (lipoprotein)                                                          |
| VNG0924G   | 1         | Ibp iron-binding protein                                                                     |
| VNG1634G   | 0.97      | CbiN cobalt transport protein                                                               |
| VNG1762G   | 0.99      | ProX putative ABC transporter                                                               |
| VNG2343G   | 1         | YkID oligopeptide ABC transporter ATP-binding                                              |
| VNG2346G   | 1         | DppC2 dipeptide ABC transporter permease                                                    |
| VNG2347G   | 1         | DppB1 dipeptide ABC transporter permease                                                    |
| VNG2359G   | 0.99      | AppB oligopeptide ABC permease                                                             |
| VNG2378G   | 1         | NosF1 copper transport ATP-binding protein                                                  |
| VNG2483G   | 1         | PstA1 phosphate ABC transporter permease                                                    |
| VNG2486G   | 1         | YutG phosphate ABC transporter binding                                                      |
| VNG2527G   | 1         | DppD dipeptide ABC transporter ATP-binding                                                  |
| VNG2529G   | 1         | DppB2 dipeptide ABC transporter permease                                                    |
| VNG6277G   | 1         | UgpB glycerol-3-phosphate-binding protein precursor                                          |
| VNG6313G   | 1         | NhaC3 Na+/H+ antiporter                                                                    |
| VNG0593G   | 0.99      | Dmd diphosphate decarboxylase                                                              |
| VNG0748G   | 1         | PrkA kinase anchor protein                                                                  |
| VNG1068G   | 1         | Tot transmembrane oligosaccharyl transferase                                                |
| VNG2106G   | 0.99      | Sdh succinate dehydrogenase subunit                                                        |
| VNG2422G   | 0.99      | GlcD glycolate oxidase subunit                                                              |
| VNG0361C   | 0.99      | Conserver hypothetical protein                                                              |
| VNG0406C   | 0.98      | Conserver hypothetical protein                                                              |
| VNG0439C   | 0.99      | Conserver hypothetical protein                                                              |
| VNG0498C   | 1         | Conserver hypothetical protein                                                              |
| VNG0537C   | 1         | Conserver hypothetical protein                                                              |
| VNG0560C   | 1         | Conserver hypothetical protein                                                              |
| VNG0573C   | 0.96      | Conserver hypothetical protein                                                              |
| VNG0582C   | 1         | Conserver hypothetical protein                                                              |
| VNG0586C   | 1         | Conserver hypothetical protein                                                              |
| VNG0675C   | 1         | Conserver hypothetical protein                                                              |
| VNG0727C   | 1         | Conserver hypothetical protein                                                              |
| VNG0754C   | 1         | Conserver hypothetical protein                                                              |
| VNG0801C   | 1         | Conserver hypothetical protein                                                              |
| VNG0903C   | 1         | Conserver hypothetical protein                                                              |
| VNG1021C   | 0.98      | Conserver hypothetical protein                                                              |
| VNG1323C   | 0.98      | Conserver hypothetical protein                                                              |
| VNG1475C   | 1         | Conserver hypothetical protein                                                              |
| Accession     | TMpred Score | Description                  | TMpred Score | Description                  |
|--------------|--------------|------------------------------|--------------|------------------------------|
| VNG1746C     | 1            | Conserver hypothetical protein | 0            | 0                            |
| VNG2121C     | 0.86         | Conserver hypothetical protein | 2            | 2                            |
| VNG2282C     | 1            | Conserver hypothetical protein | 0            | 2                            |
| VNG2285C     | 1            | Conserver hypothetical protein | 2            | 2                            |
| VNG2329C     | 0.51         | Conserver hypothetical protein | 4            | 4                            |
| VNG2395C     | 1            | Conserver hypothetical protein | 0            | 1                            |
| VNG2549C     | 1            | Conserver hypothetical protein | 0            | 1                            |
| VNG2555C     | 1            | Conserver hypothetical protein | 1            | 1                            |
| VNG2587C     | 0.98         | Conserver hypothetical protein | 0            | 0                            |
| VNG5143C     | 1            | Conserver hypothetical protein | 10           | 11                           |
| VNG6268C     | 1            | Conserver hypothetical protein | 15           | 16                           |
| VNG6296C     | 1            | Conserver hypothetical protein | 0            | 2                            |

| HP           | VNG0001H     | Hypothetical protein         | 2            | 2                            |
|--------------|--------------|------------------------------|--------------|------------------------------|
|              | VNG0005H     | Hypothetical protein         | 11           | 14                           |
|              | VNG0058H     | Hypothetical protein         | 9            | 9                            |
|              | VNG0322H     | Hypothetical protein         | 3            | 3                            |
|              | VNG0331H     | Hypothetical protein         | 2            | 2                            |
|              | VNG0352H     | Hypothetical protein         | 1            | 1                            |
|              | VNG0509H     | Hypothetical protein         | 1            | 1                            |
|              | VNG0516H     | Hypothetical protein         | 2            | 2                            |
|              | VNG0578H     | Hypothetical protein         | 4            | 4                            |
|              | VNG0584H     | Hypothetical protein         | 0            | 1                            |
|              | VNG0585H     | Hypothetical protein         | 2            | 2                            |
|              | VNG0590H     | Hypothetical protein         | 4            | 4                            |
|              | VNG0617H     | Hypothetical protein         | 0            | 0                            |
|              | VNG0743H     | Hypothetical protein         | 0            | 1*                           |
|              | VNG0759H     | Hypothetical protein         | 2            | 3                            |
|              | VNG0767H     | Hypothetical protein         | 2            | 2                            |
|              | VNG0913H     | Hypothetical protein         | 11           | 12                           |
|              | VNG1002H     | Hypothetical protein         | 5            | 7                            |
|              | VNG1034H     | Hypothetical protein         | 4            | 4                            |
|              | VNG1120H     | Hypothetical protein         | 0            | 1                            |
|              | VNG1250H     | Hypothetical protein         | 1            | 1                            |
|              | VNG1315H     | Hypothetical protein         | 0            | 1*                           |
|              | VNG1440H     | Hypothetical protein         | 2            | 2                            |
|              | VNG1538H     | Hypothetical protein         | 1            | 1                            |
|              | VNG1619H     | Hypothetical protein         | 5            | 5                            |
|              | VNG1802H     | Hypothetical protein         | 0            | 0                            |
|              | VNG1934H     | Hypothetical protein         | 2            | 2                            |
|              | VNG1965H     | Hypothetical protein         | 8            | 8                            |
|              | VNG2328H     | Hypothetical protein         | 3            | 3                            |
|              | VNG2380H     | Hypothetical protein         | 5            | 4                            |
|              | VNG2403H     | Hypothetical protein         | 0            | 1                            |
|              | VNG2413H     | Hypothetical protein         | 4            | 3                            |
|              | VNG2562H     | Hypothetical protein         | 0            | 2                            |
|              | VNG2599H     | Hypothetical protein         | 1            | 1                            |
|              | VNG2678H     | Hypothetical protein         | 2            | 2                            |
|              | VNG6251H     | Hypothetical protein         | 1            | 1                            |

* TMpred total score < 1000
Table 2. Proteins identified in soluble fraction.

The functional categories of proteins, the gene ID numbers, the probability calculated by empirical statistical model, the abbreviated name of protein, putative function, and the predicted number of transmembrane domain by TMHMM and TMpred are listed. Conserved hypothetical proteins (CHP), and hypothetical proteins (HP) are listed under the functional categories.

| Functional Categories              | Gene ID          | Probability | Protein | Putative Function                                                                 | TMHMM | TMpred |
|-----------------------------------|------------------|-------------|---------|-----------------------------------------------------------------------------------|-------|--------|
| **Amino acid metabolism**         | VNG0161G         | 1           | GdhB    | glutamate dehydrogenase                                                          | 0     | 2      |
|                                   | VNG0606G         | 1           | YrhA    | cysteine synthase                                                                | 0     | 2      |
|                                   | VNG0628G         | 1           | GdhA1   | glutamate dehydrogenase                                                          | 0     | 1*     |
|                                   | VNG0629G         | 1           | AspB2   | aspartate aminotransferase                                                        | 0     | 1*     |
|                                   | VNG0796G         | 1           | Cgs     | cystathionine gamma-synthase                                                      | 0     | 2      |
|                                   | VNG1172G         | 1           | MetB    | cystathionine alpha synthase                                                     | 0     | 1      |
|                                   | VNG1414G         | 1           | GlyA    | glycine hydroxymethyltransferase                                                 | 0     | 0      |
|                                   | VNG1912G         | 1           | TrpD2   | phosphoribosyl transferase                                                       | 0     | 1      |
|                                   | VNG2093G         | 1           | GlnA    | glutamine synthetase                                                             | 0     | 0      |
|                                   | VNG2100G         | 1           | IleA    | threonine dehydratase                                                           | 0     | 2      |
|                                   | VNG2224G         | 1           | Ocd1    | ornithine cyclodeaminase                                                         | 0     | 1*     |
|                                   | VNG2251G         | 1           | AchY    | adenosylhomocysteinase                                                           | 0     | 0      |
|                                   | VNG2294G         | 0.98        | HisA    | phosphoribosylformimino-5-aminomimidazole carboxamide ribotide isomerase         | 0     | 1*     |
|                                   | VNG2424G         | 1           | SerA1   | phosphoglycerate dehydrogenase                                                   | 0     | 1      |
|                                   | VNG6315G         | 1           | ArcB    | ornithine carbamoyltransferase                                                   | 0     | 0      |
|                                   | VNG6316G         | 1           | ArcC    | carbamate kinase                                                                 | 0     | 0      |
|                                   | VNG6317G         | 1           | ArcA    | arginine deiminase                                                               | 0     | 0      |
| **Cell envelope component**       | VNG0428G         | 1           | Fad2    | enoyl-CoA hydratase                                                              | 0     | 2      |
|                                   | VNG0679G         | 1           | Acd4    | acyl-CoA dehydrogenase                                                           | 0     | 1*     |
|                                   | VNG0681G         | 1           | Hbd1    | 3-hydroxyacyl-CoA dehydrogenase                                                  | 1     | 1      |
|                                   | VNG0771G         | 1           | AldY2   | aldehyde dehydrogenase (retinol)                                                 | 0     | 1      |
|                                   | VNG0775G         | 1           | Acd2    | acyl-CoA dehydrogenase                                                           | 0     | 1*     |
|                                   | VNG1313G         | 1           | Hbd2    | 3-hydroxyacyl-CoA dehydrogenase                                                  | 0     | 1*     |
|                                   | VNG1482G         | 1           | Acd5    | acyl-CoA dehydrogenase                                                           | 0     | 2      |
|                                   | VNG1532G         | 1           | Acc     | biotin carboxylase                                                                | 0     | 0      |
|                                   | VNG2122G         | 1           | IlvE2   | branched-chain amino-acid aminotransfer                                          | 0     | 0      |
|                                   | VNG2513G         | 1           | AldY1   | aldehyde dehydrogenase (retinol)                                                 | 0     | 1*     |
|                                   | VNG5084G,         |              | Crt     | carotenoid biosynthetic protein                                                  | 0     | 0      |
|                                   | VNG5213G, VNG6081G, |              |         |                                                                                  |       |        |
|                                   | VNG6445G         |              |         |                                                                                  |       |        |
| **Cellular process**              | VNG0101G         | 1           | CspD1   | cold shock protein                                                               | 0     | 0      |
|                                   | VNG0166G         | 1           | PsmB    | proteasome, subunit beta                                                         | 0     | 0      |
| Gene ID     | Gene Name                      | Description                                      | Value 1 | Value 2 |
|------------|-------------------------------|--------------------------------------------------|---------|---------|
| VNG0491G   | DnaK                          | heat shock protein                               | 0       | 1*      |
| VNG0494G   | GrpE                          | heat shock protein                               | 0.98    | 0       |
| VNG0880G   | PsmA                          | proteasome, subunit alpha                         | 1       | 2       |
| VNG0942G   | CheW2                         | chemotaxis protein                               | 0.76    | 0       |
| VNG0958G   | Htr15                         | Htr15 transducer                                 | 0       | 1*      |
| VNG1013G   | Htr13                         | Htr13 transducer                                 | 0.76    | 1*      |
| VNG1147G   | Cdc48b                        | cell division cycle protein                       | 1       | 1*      |
| VNG1667G   | Cdc48c                        | cell division cycle protein                       | 0       | 0       |
| VNG1836G   | CspD2                         | cold shock protein                                | 0       | 0       |
| VNG1933G   | FtsZ3                         | cell division protein                             | 0       | 2       |
| VNG2096G   | CctB                          | thermosome subunit beta                           | 0       | 3       |
| VNG2226G   | CctA                          | thermosome subunit alpha                          | 0       | 1       |
| VNG2443G   | DpsA                          | starvation induced DNA binding protein            | 0       | 0       |
| VNG5034G,  | GvpO1                         | GvpO protein, cluster A                           | 0       | 0       |
| VNG5089G,  | SojC1                         | Spo0A activation inhibitor                        | 0       | 0       |
| VNG5208G,  |                              |                                                  |         |         |
| VNG6086G   | GvpO2                         | GvpO protein, cluster B                           | 0       | 0       |
| VNG6246G   |                              | peroxidase / catalase                             | 0       | 1*      |
| VNG1555G   | CobH                          | cobalamin biosynthesis                            | 0       | 3       |
| VNG1566G   | CobN                          | cobalamin biosynthesis protein                   | 0       | 1*      |
| VNG2604G   | Thi1                          | thiamine biosynthetic enzyme                      | 0       | 1*      |
| VNG2606G   | ThiD                          | hydroxymethylpyrimidine phosphate kinase          | 0       | 0       |
| VNG1190G   | Sod1                          | superoxide dismutase                              | 0       | 1*      |
| VNG2256G   | Pcn                           | proliferating-cell nuclear antigen               | 0       | 0       |
| VNG2280G   | RfaC                          | replication factor C small subunit               | 0       | 0       |
| VNG2620G   | UvrD                          | repair helicase                                   | 0       | 0       |
| VNG0095G   | GapB                          | glyceraldehyde 3-phosphate dehydrogenase         | 0       | 0       |
| VNG0259G   | Ipp                           | inorganic pyrophosphatase                         | 0       | 0       |
| VNG0330G   | PpsA                          | phosphoenolpyruvate synthase                     | 0       | 0       |
| VNG0467G   | YafB                          | aldehyde reductase                                | 0       | 0       |
| VNG0474G   | PorA                          | pyruvate ferrodoxin oxidoreductase, subunit alpha | 0       | 2       |
| VNG0684G   | Fbp                           | fructose-bisphosphatase                           | 0       | 2       |
| VNG0940G   | Acs3                          | acetyl-CoA synthetase                             | 0       | 1*      |
| VNG0997G   | Acs2                          | acetyl-CoA synthetase                             | 0       | 3       |
| VNG1128G   | KorA                          | putative 2-ketoglutarate ferrodoxin oxidoreductase (alpha) | 0       | 1*      |
| VNG1142G   | Eno                           | phosphopyruvate hydratase                         | 0       | 1*      |
| VNG1356G   | FumC                          | fumarate hydratase                                | 0       | 0       |
| VNG1541G   | SucC                          | succinyl-CoA synthetase beta chain               | 0       | 1*      |
| VNG1542G   | SucD                          | succinyl-CoA synthetase alpha chain              | 0       | 0       |
| VNG1873G   | Icd                           | isocitrate dehydrogenase, NADP                   | 0       | 1*      |
| VNG1914G   | PpiA                          | peptidyl-prolyl isomerase                         | 0       | 0       |
| VNG2102G   | CitZ                          | citrate synthase                                 | 0       | 1*      |
| VNG2138G   | AtpB                          | H+-transporting ATP synthase subunit B           | 0       | 0       |
| VNG2139G   | AtpA                          | H+-transporting ATP synthase subunit A           | 0       | 0       |
| VNG2142G   | AtpE                          | H+-transporting ATP synthase subunit E           | 0       | 0       |
| VNG2151G   | EtfA                          | electron transfer flavoprotein subunit alpha     | 0       | 1*      |
| Gene Name      | Protein Name               | Description                                         | ORF | Notes  |
|---------------|---------------------------|-----------------------------------------------------|-----|--------|
| VNG2217G      | PdhA2                     | pyruvate dehydrogenase alpha subunit                | 1   | *
| VNG2218G      | PdhB                      | pyruvate dehydrogenase beta subunit                 | 1   | *
| VNG2219G      | Dsa                       | dihydrolipoamide S-acetyltransferase                | 1   | *
| VNG2220G      | LpdA                      | dihydrolipoamide dehydrogenase                     | 2   | 0      |
| VNG2229G      | FerA2                     | ferredoxin                                          | 0   | 0      |
| VNG2367G      | MdhA                      | L-malate dehydrogenase                              | 0   | 1      |
| VNG2496G      | FerB                      | ferredoxin                                          | 0   | 3      |
| VNG2574G      | Can                       | aconitase                                           | 0   | 1      |
| VNG2617G      | Adh2                      | alcohol dehydrogenase                               | 0   | 1      |
| VNG6270G      | GldA                      | sn-glycerol-1-phosphate dehydrogenase              | 0   | 0      |
| VNG0186G      | PepB1                     | aminopeptidase homolog                              | 0.56| 0      |
| VNG0523G      | Inb                       | oxidoreductase homolog                              | 1   | 0      |
| VNG1416G      | FoD                       | methylenetetrahydrofolate dehydrogenase            | 0.99| 0      |
| VNG0414G      | PurH                      | phosphoribosylaminomimidazole-succinocarboxamide formyltransferase | 1   | 2      |
| VNG0448G      | PyrE1                     | orotate phosphoribosyl transferase                 | 0.99| 0      |
| VNG0559G      | Apt                       | adenine phosphoribosyltransferase                  | 0   | 1      |
| VNG0890G      | Imd2                      | inosine-5'-monophosphate dehydrogenase             | 0   | 0      |
| VNG0893G      | Udp2                      | uridine phosphorylase                               | 0.99| 0      |
| VNG1001G      | GuaB                      | inosine monophosphate dehydrogenase                | 0   | 1      |
| VNG1048G      | Udg1                      | UDP-glucose dehydrogenase                           | 0   | 1      |
| VNG1089G      | PurA                      | adenylosuccinate synthase                           | 0   | 0      |
| VNG1160G      | Ndk                       | nucleoside diphosphate kinase                       | 0   | 1      |
| VNG1305G      | PurD                      | phosphoribosylglycinamidase synthetase              | 0   | 0      |
| VNG1724G      | Adk                       | adenylate kinase                                    | 0   | 0      |
| VNG1830G      | PyrG                      | CTP synthase                                        | 0.98| 1      |
| VNG2118G      | PyrE2                     | orotate phosphoribosyl transferase                 | 0   | 2      |
| VNG2600G      | TrxA2                     | thioredoxin                                         | 0   | 0      |
| VNG6309G      | PyrB                      | aspartate carbamoyltransferase catalytic subunit    | 0   | 0      |
| VNG6311G      | PyrI                      | aspartate carbamoyltransferase regulatory chain     | 0   | 0      |
| VNG0550G      | Rps27e                    | 30S ribosomal protein S27E                         | 0.95| 0      |
| VNG0787G      | Rps3e                     | 30S ribosomal protein S3E                          | 0   | 0      |
| VNG0790G      | Rps15p                    | 30S ribosomal protein S15P                         | 0   | 0      |

**Miscellaneous**

| Gene Name      | Protein Name               | Description                                         | ORF | Notes  |
|---------------|---------------------------|-----------------------------------------------------|-----|--------|
| VNG2616G      | PepB2                     | aminopeptidase homolog                              | 0   | 0      |
| VNG1136G      | Rpb3                      | RNA-directed RNA polymerase II                      | 0   | 0      |
| VNG11160G     | Ndk                       | nucleoside diphosphate kinase                       | 0   | 0      |
| VNG1922G      | Trh5                      | transcription regulator                              | 0   | 0      |
| VNG2036G      | Hix1                      | HoxA-like transcriptional regulator                 | 0   | 0      |
| VNG2094G      | Trh4                      | transcription regulator                              | 0   | 0      |

**Nucleotide Metabolism**

| Gene Name      | Protein Name               | Description                                         | ORF | Notes  |
|---------------|---------------------------|-----------------------------------------------------|-----|--------|
| VNG0451G      | PhoU                      | transcriptional regulator                            | 0.99| 0      |
| VNG0536G      | SirR                      | transcription repressor                              | 0   | 0      |
| VNG1922G      | TrhS                      | transcription regulator                              | 0   | 0      |
| VNG2036G      | Hix1                      | HoxA-like transcriptional regulator                 | 0.98| 0      |
| VNG2094G      | Trh4                      | transcription regulator                              | 0.98| 0      |

**Transcription**

| Gene Name      | Protein Name               | Description                                         | ORF | Notes  |
|---------------|---------------------------|-----------------------------------------------------|-----|--------|
| VNG0401G      | Epf2                      | mRNA 3'-end processing factor homolog               | 1   | 1      |
| VNG1136G      | Rpb3                      | DNA-directed RNA polymerase II                      | 1   | 0      |
| VNG2243G      | TbpE                      | transcription initiation factor IID                 | 1   | 0      |
| VNG2662G      | RpoC                      | DNA-directed RNA polymerase subunit C               | 1   | 0      |
| VNG2664G      | RpoA                      | DNA-directed RNA polymerase subunit A               | 1   | 0      |
| VNG2665G      | RpoB'                     | DNA-directed RNA polymerase subunit B'              | 1   | 0      |
| VNG2666B      | RpoB''                    | DNA-directed RNA polymerase subunit B''             | 1   | 0      |

**Translation**

| Gene Name      | Protein Name               | Description                                         | ORF | Notes  |
|---------------|---------------------------|-----------------------------------------------------|-----|--------|
| VNG0461G      | AspS                      | aspartyl-tRNA synthetase                             | 0   | 0      |
| VNG0550G      | Rps27e                    | 30S ribosomal protein S27E                          | 0.95| 0      |
| VNG0787G      | Rps3e                     | 30S ribosomal protein S3E                          | 0   | 0      |
| VNG0790G      | Rps15p                    | 30S ribosomal protein S15P                          | 0   | 0      |
| Accession Number  | Gene ID | Description                        | Rpl10p | 50S ribosomal protein L10P | 0  | 0 |
|-------------------|---------|------------------------------------|--------|---------------------------|----|----|
| VNG1104G          | 1       | Rpl10p                             |        |                           | 0  | 0 |
| VNG1105G          | 1       | Rpl1p                              |        | 50S ribosomal protein L1P  | 0  | 0 |
| VNG1108G          | 1       | Rpl1lp                             |        | 50S ribosomal protein L11P | 0  | 0 |
| VNG1112G          | 1       | Rps13p                             |        | 30S ribosomal protein S13P | 0  | 0 |
| VNG1133G          | 1       | Rps4p                              |        | 30S ribosomal protein S4P  | 0  | 0 |
| VNG1138G          | 1       | Rpl13p                             |        | 50S ribosomal protein L13P | 0  | 0 |
| VNG1143G          | 1       | Rps2p                              |        | 30S ribosomal protein S2P  | 0  | 0 |
| VNG1157G          | 1       | Rps6p                              |        | 30S ribosomal protein HS6   | 0  | 1*|
| VNG1158G          | 1       | Rps28e                             |        | 30S ribosomal protein S28E | 0  | 0 |
| VNG1170G          | 1       | Rpl21e                             |        | 50S ribosomal protein L21E | 0  | 0 |
| VNG1190G          | 1       | Rpl4e                              |        | 50S ribosomal protein L4E  | 0  | 0 |
| VNG1191G          | 1       | Rpl23p                             |        | 50S ribosomal protein L23P | 0  | 0 |
| VNG1193G          | 1       | Rps19p                             |        | 30S ribosomal protein S19P | 0  | 0 |
| VNG1195G          | 1       | Rpl22p                             |        | 50S ribosomal protein L22P | 0  | 0 |
| VNG1197G          | 1       | Rps3p                              |        | 30S ribosomal protein S3P  | 0  | 0 |
| VNG1198G          | 1       | Rpl29p                             |        | 50S ribosomal protein L29P | 0  | 0 |
| VNG1700G          | 1       | Rps17p                             |        | 30S ribosomal protein S17P | 0  | 0 |
| VNG1702G          | 1       | Rpl24p                             |        | 50S ribosomal protein L24P | 0  | 0 |
| VNG1703G          | 1       | Rps4e                              |        | 30S ribosomal protein S4E  | 0  | 0 |
| VNG1705G          | 1       | Rpl5p                              |        | 50S ribosomal protein L5P  | 0  | 0 |
| VNG1709G          | 1       | Rpl6p                              |        | 50S ribosomal protein L6P  | 0  | 0 |
| VNG1715G          | 1       | Rps5p                              |        | 30S ribosomal protein S5P  | 0  | 0 |
| VNG1716G          | 1       | Rpl30p                             |        | 50S ribosomal protein L30P | 0  | 0 |
| VNG1997G          | 1       | InfB                               |        | bacterial-like IF2         | 0  | 0 |
| VNG2005G          | 1       | HisS                               |        | histidyl-tRNA synthetase   | 0  | 1*|
| VNG2010G          | 1       | Rps19e                             |        | 30S ribosomal protein S19E | 0  | 0 |
| VNG2048G          | 1       | Rps24e                             |        | 30S ribosomal protein S24E | 0  | 0 |
| VNG2072G          | 1       | SerS                               |        | serine-tRNA synthetase     | 0  | 1*|
| VNG2190G          | 1       | IleS                               |        | isoleucyl-tRNA synthetase  | 0  | 1*|
| VNG2352G          | 1       | GlyS                               |        | glycine-tRNA synthetase    | 0  | 1*|
| VNG2467G          | 1       | Rpl31e                             |        | 50S ribosomal protein L31E | 0  | 0 |
| VNG2514G          | 1       | Rps6e                              |        | 30S ribosomal protein S6E  | 0  | 0 |
| VNG2649G          | 1       | Eef1a                              |        | translation elongation factor eEF-1A subunit alpha | 0 | 0 |
| VNG2654G          | 1       | Eef2                               |        | translation elongation factor eEF-2 | 0 | 0 |
| VNG6312G          | 1       | ArgS                               |        | arginine-tRNA synthetase   | 0  | 1*|
| Transport         |         |                                    |        |                            |    |    |
| VNG0115G          | 1       | YusZ1                              |        | oxidoreductase             | 0  | 0 |
| VNG0524G          | 1       | YurY                               |        | ABC transporter, ATP-binding protein | 0 | 0 |
| VNG1924G          | 1       | TrkA6                              |        | TRK potassium uptake system protein | 0 | 0 |
| Uncharacterized   |         |                                    |        |                            |    |    |
| VNG0937G          | 1       | Gap                                |        | glyceraldehyde-3-phosphate dehydrogenase | 0 | 1*|
| VNG0998G          | 1       | YajO2                              |        | probable oxidoreductase    | 0  | 0 |
| VNG1294G          | 1       | SlyD                               |        | peptidyl-prolyl cis-trans isomerase | 0 | 0 |
| VNG2001G          | 1       | yjbG                               |        | oligopeptidase             | 0  | 2 |
| VNG2593G          | 1       | TssA                               |        | probable thiosulfate sulfurtransferase | 0 | 0 |
| VNG2639G          | 1       | Uae                                |        | UDP-N-acetyllucosamine 2-epimerase | 0 | 0 |
| CHP               |         |                                    |        |                            |    |    |
| VNG0096C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 1*|
| VNG0153C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 1*|
| VNG0234C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 0 |
| VNG0239C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 0 |
| VNG0309C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 0 |
| VNG0310C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 0 |
| VNG0394C          | 1       | Conserved hypothetical protein      |        | Conserved hypothetical protein | 0 | 0 |
| VNG0424C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG0525C | 1 | Conserved hypothetical protein | 0 | 1 |
| VNG0527C | 1 | Conserved hypothetical protein | 0 | 1 |
| VNG0546C | 1 | Conserved hypothetical protein | 0 | 1 |
| VNG0654C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG0683C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG0758C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG0849C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG0941C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1038C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1088C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1093C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1149C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1168C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1169C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1276C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1297C | 1 | Conserved hypothetical protein | 0 | 1* |
| VNG1325C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1339C | 1 | Conserved hypothetical protein | 0 | 2 |
| VNG1524C | 1 | Conserved hypothetical protein | 0 | 1* |
| VNG1663C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1748C | 0.98 | Conserved hypothetical protein | 0 | 0 |
| VNG1752C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG1793C | 1 | Conserved hypothetical protein | 0 | 1* |
| VNG1898C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2021C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2099C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2117C | 0.91 | Conserved hypothetical protein | 0 | 2 |
| VNG2160C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2162C | 0.99 | Conserved hypothetical protein | 0 | 0 |
| VNG2259C | 1 | Conserved hypothetical protein | 0 | 1* |
| VNG2351C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2371C | 0.51 | Conserved hypothetical protein | 0 | 1* |
| VNG2465C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2508C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2584C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG2615C | 1 | Conserved hypothetical protein | 0 | 0 |
| VNG5069C, VNG5228C, VNG6066C, VNG64609C | 1 | Conserved hypothetical protein | 0 | 1* |
| HP | VNG0207H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG0557H | 1 | Hypothetical protein | 0 | 1* |
| HP | VNG0597H | 1 | Hypothetical protein | 0 | 1* |
| HP | VNG0782H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG1257H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG1289H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG1291H | 0.66 | Hypothetical protein | 0 | 0 |
| HP | VNG1412H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG1562H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG1925H | 1 | Hypothetical protein | 0 | 0 |
| HP | VNG2008H | 0.98 | Hypothetical protein | 0 | 0 |
| HP | VNG2146H | 1 | Hypothetical protein | 0 | 0 |
| Protein ID   | Multiplicity | Description       | TMpred Score | TMD Score |
|-------------|--------------|-------------------|--------------|-----------|
| VNG2273H    | 1            | Hypothetical protein | 0            | 0         |
| VNG2392H    | 1            | Hypothetical protein | 0            | 0         |
| VNG2509H    | 1            | Hypothetical protein | 1            | 1         |
| VNG2603H    | 1            | Hypothetical protein | 0            | 0         |
| VNG2643H    | 1            | Hypothetical protein | 0            | 0         |
| VNG5049H,   | 1            | Hypothetical protein | 0            | 0         |
| VNG5248H,   |              |                   |              |           |
| VNG6047H    |              |                   |              |           |
| VNG6479H    |              |                   |              |           |
| VNG5145H    | 1            | Hypothetical protein | 0            | 0         |

* TMpred total score < 1000
Table 3. Proteins identified in both membrane and soluble fraction.

The functional categories of proteins, the gene ID numbers, the probability calculated by empirical statistical model, the name of protein, putative function, and the predicted number of transmembrane domain by TMHMM and TMpred are listed. Hypothetical proteins (HP) are listed under the functional categories.

| Functional Categories        | Gene ID            | Probability | Protein | Putative Function                        | TMHMM | TMpred |
|------------------------------|--------------------|-------------|---------|-----------------------------------------|-------|--------|
| Cell envelope component      | VNG2679G           | 1           | Csg     | cell surface glycoprotein               | 1     | 2      |
| Cellular process             | VNG0960G           | 1           | FlaB1   | flagellin B1 precursor                 | 1     | 1      |
|                              | VNG0962G           | 1           | FlaB3   | flagellin B3 precursor                 | 0     | 1      |
|                              | VNG1008G           | 1           | FlaA1a  | flagellin A1 precursor                 | 1     | 1      |
|                              | VNG1009G           | 1           | FlaA2   | flagellin A2 precursor                 | 0     | 1      |
|                              | VNG5032G, VNG6031G | 1           | GvpC1   | GvpC protein, cluster A                | 0     | 0      |
| Cofactor Metabolism          | VNG1550G           | 1           | CbiT    | cobalamin biosynthesis glutamate-1-semialdehyde aminotransferase | 0     | 0      |
|                              | VNG2326G           | 0.99        | HemL    |                                          | 0     | 0      |
| Energy Metabolism            | VNG1125G           | 1           | KorB    | putative 2-ketoglutarate ferredoxin oxidoreductase (beta) succinate dehydrogenase subunit A | 0     | 0      |
|                              | VNG1306G           | 1           | SdhA    |                                          | 0     | 2      |
| Nucleotide Metabolism        | VNG1644G           | 1           | NrdB2   | ribonucleoside reductase large chain    | 0     | 2      |
| Translation                  | VNG0099G           | 1           | Rpl10e  | 50S ribosomal protein L10E              | 0     | 0      |
|                              | VNG1134G           | 1           | Rps11p  | 30S ribosomal protein S11P              | 0     | 0      |
|                              | VNG1137G           | 1           | Rpl18e  | 50S ribosomal protein L18E              | 0     | 0      |
|                              | VNG1139G           | 1           | Rps9p   | 30S ribosomal protein S9P               | 0     | 0      |
|                              | VNG1668G           | 1           | Rps8e   | 30S ribosomal protein S8E               | 0     | 0      |
|                              | VNG1689G           | 1           | Rpl3p   | 50S ribosomal protein L13P              | 0     | 0      |
|                              | VNG1692G           | 1           | Rpl2p   | 50S ribosomal protein L2P               | 0     | 0      |
|                              | VNG1701G           | 1           | Rpl14p  | 50S ribosomal protein L14P              | 0     | 0      |
|                              | VNG1707G           | 1           | Rps8p   | 30S ribosomal protein S8P               | 0     | 0      |
|                              | VNG1711G           | 1           | Rpl32e  | 50S ribosomal protein L32E              | 0     | 0      |
|                              | VNG1713G           | 1           | Rpl19e  | 50S ribosomal protein L19E              | 0     | 0      |
|                              | VNG1714G           | 1           | Rpl18p  | 50S ribosomal protein L18P              | 0     | 1      |
|                              | VNG1718G           | 1           | Rpl15p  | 50S ribosomal protein L15P              | 0     | 0      |
|                              | VNG2657G           | 1           | Rps7p   | 30S ribosomal protein S7P               | 0     | 0      |
|                              | VNG2658G           | 1           | Rps12p  | 30S ribosomal protein S12P              | 0     | 0      |
| Transport                    | VNG2349G           | 1           | DppA    | dipeptide ABC transporter dipeptide-binding | 0     | 2      |
| HP                           | VNG1220H           | 0.99        |         | Hypothetical protein                    | 0     | 0      |
|                              | VNG2400H           | 0.99        |         | Hypothetical protein                    | 0     | 0      |

* TMpred total score < 1000
Table 4. Expressed proteins in metabolic pathways.

Expressed proteins were analyzed for pathway involvement by searching the KEGG Enzymes/Compounds/Genes pathway database. Pathway, known number of proteins involved in a pathway in *Halobacterium sp. NRC-1*, number of protein identified in this study, gene identification number, abbreviated protein name, and putative function are listed. Some proteins are involved in more than one pathway.

| Pathway                                      | Total number of proteins | Identified proteins | Gene ID          | Protein and Putative function                                                                 |
|----------------------------------------------|--------------------------|--------------------|------------------|------------------------------------------------------------------------------------------------|
| Glycolysis / Gluconeogenesis                 | 17                       | 8                  | VNG0771G AldY2; aldehyde dehydrogenase (retinol) |                                                                                                  |
|                                              |                          |                    | VNG0997G Acs2; acetyl-CoA synthetase              |                                                                                                  |
|                                              |                          |                    | VNG1142G Eno; EnOlaSe                             |                                                                                                  |
|                                              |                          |                    | VNG2217G PdhA2; pyruvate dehydrogenase alpha subunit |                                                                                                  |
|                                              |                          |                    | VNG2218G PdhB; pyruvate dehydrogenase beta subunit |                                                                                                  |
|                                              |                          |                    | VNG2219G Dsa; dihydrolipoamide S-acetyltransferase |                                                                                                  |
|                                              |                          |                    | VNG2220G LpdA; dihydrolipoamide dehydrogenase      |                                                                                                  |
|                                              |                          |                    | VNG2513G AldY1; aldehyde dehydrogenase (retinol)   |                                                                                                  |
| Citrate cycle (TCA cycle)                    | 18                       | 14                 | VNG0474G PorA; pyruvate ferredoxin oxidoreductase, subunit alpha |                                                                                                  |
|                                              |                          |                    | VNG1125G KorB; 2-oxoglutarate ferredoxin oxidoreductase, beta subunit |                                                                                                  |
|                                              |                          |                    | VNG1128G KorA; 2-oxoglutarate ferredoxin oxidoreductase, alpha subunit |                                                                                                  |
|                                              |                          |                    | VNG1306G SdhA; SucCinate dehydrogenase subunit A   |                                                                                                  |
|                                              |                          |                    | VNG1308G SdhB; SucCinate dehydrogenase subunit B   |                                                                                                  |
|                                              |                          |                    | VNG1356G FumC; fumarate hydratase                  |                                                                                                  |
|                                              |                          |                    | VNG1532G Acc; pyruvate Carboxylase subunit A       |                                                                                                  |
|                                              |                          |                    | VNG1541G SucC; SucCinky-CoA synthetase beta chain   |                                                                                                  |
|                                              |                          |                    | VNG1542G SucD; SucCinky-CoA synthetase alpha chain  |                                                                                                  |
|                                              |                          |                    | VNG1873G Icd; isocitrate dehydrogenase, NADP        |                                                                                                  |
|                                              |                          |                    | VNG2102G CitZ; citrate synthase                    |                                                                                                  |
|                                              |                          |                    | VNG2220G LpdA; dihydrolipoamide dehydrogenase      |                                                                                                  |
|                                              |                          |                    | VNG2367G MdhA; L-malate dehydrogenase              |                                                                                                  |
|                                              |                          |                    | VNG2574G Can; aconitase                           |                                                                                                  |
| Ascorbate and aldarate metabolism            | 3                        | 2                  | VNG0771G AldY2; aldehyde dehydrogenase (retinol)   |                                                                                                  |
|                                              |                          |                    | VNG2513G AldY1; aldehyde dehydrogenase (retinol)   |                                                                                                  |
| Fatty acid biosynthesis (path 2)             | 7                        | 4                  | VNG0428G Fad2; Enoyl-CoA hydratase                |                                                                                                  |
|                                              |                          |                    | VNG0681G Hbd1; 3-hydroxyacyl-CoA dehydrogenase      |                                                                                                  |
|                                              |                          |                    | VNG1313G Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase |                                                                                                  |
|                                              |                          |                    | VNG2063G Aca; probable acetyl-coa acetyltransferase|                                                                                                  |
| Pathway                        | Count 1 | Count 2 | Gene ID  | Description                                                                 |
|-------------------------------|---------|---------|----------|-----------------------------------------------------------------------------|
| **Fatty acid metabolism**     | 13      | 6       | VNG0428G | Fad2; Enoyl-CoA hydratase                                                  |
|                               |         |         | VNG0681G | Hbd1; 3-hydroxyacyl-CoA dehydrogenase                                      |
|                               |         |         | VNG0771G | AldY2; aldehyde dehydrogenase (retinol)                                    |
|                               |         |         | VNG1313G | Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase                |
|                               |         |         | VNG2063G | Acac; probable acetyl-coa acetyltransferase                                |
|                               |         |         | VNG2513G | AldY1; aldehyde dehydrogenase (retinol)                                    |
| **Sterol biosynthesis**       | 7       | 1       | VNG0593G | Dmd; diphosphomevalonate deCarboxylase                                     |
| **Bile acid biosynthesis**    | 11      | 6       | VNG0679G | Acd4; acyl-CoA dehydrogenase                                               |
|                               |         |         | VNG0771G | AldY2; aldehyde dehydrogenase (retinol)                                    |
|                               |         |         | VNG0775G | Acd2; acyl-CoA dehydrogenase                                               |
|                               |         |         | VNG1482G | Acd5; acyl-CoA dehydrogenase                                               |
|                               |         |         | VNG2063G | Acac; probable acetyl-coa acetyltransferase                                |
|                               |         |         | VNG2513G | AldY1; aldehyde dehydrogenase (retinol)                                    |
| **Ubiquinone biosynthesis**   | 18      | 7       | VNG0635G | NolB; NADH dehydrogenase I chain A                                        |
|                               |         |         | VNG0637G | NdhG5; NADH dehydrogenase I chain C / D                                    |
|                               |         |         | VNG0639G | NdhG4; NADH dehydrogenase I chain H                                        |
|                               |         |         | VNG0640G | NolD; NADH dehydrogenase I chain I                                        |
|                               |         |         | VNG0646G | NuoL; NADH dehydrogenase I chain L                                        |
|                               |         |         | VNG0648G | NdhG3; NADH dehydrogenase I chain N                                        |
|                               |         |         | VNG1932G | NolA; NADH dehydrogenase/oxidoreductase-like protein                        |
| **Oxidative phosphorylation** | 30      | 20      | VNG0259G | Ipp; inorganic pyrophosphatase                                             |
|                               |         |         | VNG0635G | NolB; NADH dehydrogenase I chain A                                        |
|                               |         |         | VNG0637G | NdhG5; NADH dehydrogenase I chain C / D                                    |
|                               |         |         | VNG0639G | NdhG4; NADH dehydrogenase I chain H                                        |
|                               |         |         | VNG0640G | NolD; NADH dehydrogenase I chain I                                        |
|                               |         |         | VNG0646G | NuoL; NADH dehydrogenase I chain L                                        |
|                               |         |         | VNG0648G | NdhG3; NADH dehydrogenase I chain N                                        |
|                               |         |         | VNG0657G | CoxA2; cytochrome c oxidase subunit I                                      |
|                               |         |         | VNG0665G | CoxB1; cytochrome c oxidase subunit II                                     |
|                               |         |         | VNG1306G | SdhA; SucCinate dehydrogenase subunit A                                    |
|                               |         |         | VNG1308G | SdhB; SucCinate dehydrogenase subunit B                                    |
|                               |         |         | VNG1932G | NolA; NADH dehydrogenase/oxidoreductase-like protein                        |
|                               |         |         | VNG2138G | AtpB; H+ transporting ATP synthase subunit B                               |
|                               |         |         | VNG2139G | AtpA; H+ transporting ATP synthase subunit A                               |
|                               |         |         | VNG2141G | AtpC; H+ transporting ATP synthase subunit C                               |
|                               |         |         | VNG2142G | AtpE; H+ transporting ATP synthase subunit E                               |
|                               |         |         | VNG2143G | AtpK; H+ transporting ATP synthase subunit K                               |
|                               |         |         | VNG2144G | AtpI; H+ transporting ATP synthase subunit I                               |
|                               |         |         | VNG2193G | CoxA1; cytochrome c oxidase subunit I                                      |
|                               |         |         | VNG2195G | CoxB2; cytochrome c oxidase subunit II                                     |
| **ATP synthesis**             | 8       | 6       | VNG2138G | AtpB; H+ transporting ATP synthase subunit B                               |
|                               |         |         | VNG2139G | AtpA; H+ transporting ATP synthase subunit A                               |
|                               |         |         | VNG2141G | AtpC; H+ transporting ATP synthase subunit C                               |
|                               |         |         | VNG2142G | AtpE; H+ transporting ATP synthase subunit E                               |
|                               |         |         | VNG2143G | AtpK; H+ transporting ATP synthase subunit K                               |
|                               |         |         | VNG2144G | AtpI; H+ transporting ATP synthase subunit I                               |
| Metabolic Pathway                  | Code   | Genes                                                                                           |
|-----------------------------------|--------|-------------------------------------------------------------------------------------------------|
| **Urea cycle and metabolism of amino groups** | 6      | VNG0161G: GdhB; glutamate dehydrogenase                                                    |
|                                   |        | VNG0628G: GdhA1; glutamate dehydrogenase                                                       |
|                                   |        | VNG6315G: ArcB; ornithine CarBamoyltransferase                                                 |
| **Purine metabolism**             | 26     | VNG0414G: PurH; phosphoribosylaminomimidazoleCarboxamide formyltransferase / IMP cyclohydrodase |
|                                   |        | VNG0559G: Apt; adenine phosphoribosyltransferase                                               |
|                                   |        | VNG0632G: PurK; phosphoribosylaminomimidazole Carboxylase ATP binding subunit                |
|                                   |        | VNG1001G: GuaB; inosine monophosphate dehydrogenase                                             |
|                                   |        | VNG1089G: PurA; adenyloSuccinate synthase                                                      |
|                                   |        | VNG1160G: Ndk; nucleoside diphosphate kinase                                                   |
|                                   |        | VNG1305G: PurD; phosphoribosylglycinamide synthetase                                            |
|                                   |        | VNG1644G: NrdB2; ribonucleoside-diphosphate reductase alpha chain                             |
|                                   |        | VNG1724G: Adk; adenylate kinase                                                                |
|                                   |        | VNG6316G: ArcC; CarBamoyl kinase                                                              |
| **Pyrimidine metabolism**         | 26     | VNG0448G: PyrE1; orotate phosphoribosyltransferase                                            |
|                                   |        | VNG0893G: Udp2; uridine phosphorylase                                                          |
|                                   |        | VNG1160G: Ndk; nucleoside diphosphate kinase                                                   |
|                                   |        | VNG1644G: NrdB2; ribonucleoside-diphosphate reductase alpha chain                             |
|                                   |        | VNG1814G: CarB; CarBamoyl-phosphate synthase large subunit                                    |
|                                   |        | VNG1830G: PyrG; CTP synthase                                                                  |
|                                   |        | VNG2118G: PyrE2; orotate phosphoribosyl transferase                                            |
|                                   |        | VNG2507G: PyrD; dihydroorotate dehydrogenase                                                   |
|                                   |        | VNG6309G: PyrB; aspartate CarBamoyltransferase catalytic subunit                              |
|                                   |        | VNG6311G: PyrI; aspartate CarBamoyltransferase regulatory chain                               |
| **Glutamate metabolism**          | 18     | VNG0161G: GdhB; glutamate dehydrogenase                                                      |
|                                   |        | VNG0628G: GdhA1; glutamate dehydrogenase                                                      |
|                                   |        | VNG0629G: AspB2; aspartate aminotransferase                                                    |
|                                   |        | VNG1814G: CarB; CarBamoyl-phosphate synthase large subunit                                    |
|                                   |        | VNG2093G: GlnA; glutamine synthetase                                                          |
|                                   |        | VNG6316G: ArcC; CarBamoyl kinase                                                              |
| **Alanine and aspartate metabolism** | 16     | VNG0461G: AspS; aspartyl-tRNA synthetase                                                       |
|                                   |        | VNG0629G: AspB2; aspartate aminotransferase                                                    |
|                                   |        | VNG1089G: PurA; adenyloSuccinate synthase                                                      |
|                                   |        | VNG1532G: Acp; pyruvate Carboxylase subunit A                                                  |
|                                   |        | VNG6309G: PyrB; aspartate CarBamoyltransferase catalytic subunit                              |
|                                   |        | VNG6311G: PyrI; aspartate CarBamoyltransferase regulatory chain                               |
| **Glycine, serine and threonine metabolism** | 22     | VNG1172G: MetB; cystathionine alpha synthase                                                   |
|                                   |        | VNG1414G: GlyA; glycine hydroxymethyltransferase                                               |
|                                   |        | VNG2072G: SerS; serine-tRNA synthetase                                                         |
|                                   |        | VNG2100G: IleA; threonine dehydratase                                                          |
|                                   |        | VNG2220G: LpdA; dihydrolipoamide dehydrogenase                                                  |
|                                   |        | VNG2352G: GlyS; glycine-tRNA synthetase                                                        |
|                                   |        | VNG2424G: SerA1; phosphoglycerate dehydrogenase                                                 |
| **Methionine metabolism**         | 6      | VNG0796G: Cgs; cystathionine gamma-synthase                                                    |
| Metabolism                              | Number of Genes | Genes                                                                 |
|----------------------------------------|----------------|----------------------------------------------------------------------|
| Cysteine metabolism                    | 10             | MetB; cystathionine alpha synthase                                      |
|                                        |                | VNG1172G                                                             |
|                                        |                | YrhA; cysteine synthase                                              |
|                                        |                | VNG2251G                                                             |
|                                        |                | AchY; adEnosylhomocysteinate                                          |
|                                        |                | VNG2421G                                                             |
|                                        |                | Hal; O-acetyl homoserine                                              |
|                                        | 5              | VNG0606G                                                             |
|                                        |                | VNG0629G                                                             |
|                                        |                | AspB2; aspartate aminotransfer                                        |
|                                        |                | VNG0796G                                                             |
|                                        |                | Cgs; cystathionine gamma-synthase                                      |
|                                        |                | VNG1172G                                                             |
|                                        |                | MetB; cystathionine alpha synthase                                      |
|                                        |                | VNG2421G                                                             |
|                                        |                | Hal; O-acetyl homoserine                                              |
| Valine, leucine and isoleucine         | 15             | Fad2; Enoyl-CoA hydratase                                              |
| degradation                            | 7              | VNG0428G                                                             |
|                                        |                | VNG0681G                                                             |
|                                        |                | Hbd1; 3-hydroxyacyl-CoA dehydrogenase                                 |
|                                        |                | VNG0771G                                                             |
|                                        |                | AldY2; aldehyde dehydrogenase (retinol)                              |
|                                        |                | VNG1313G                                                             |
|                                        |                | Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase          |
|                                        |                | VNG2063G                                                             |
|                                        |                | Aca; probable acetyl-coa acetyltransferase                           |
|                                        |                | VNG2122G                                                             |
|                                        |                | IlvE2; branched-chain amino acid aminotransfer                       |
|                                        |                | VNG2513G                                                             |
|                                        |                | AldV1; aldehyde dehydrogenase (retinol)                              |
| Valine, leucine and isoleucine         | 8              | IlvE2; branched-chain amino acid aminotransfer                       |
| biosynthesis                           | 4              | VNG2122G                                                             |
|                                        |                | VNG2190G                                                             |
|                                        |                | IleS; isoleucyl-tRNA synthetase                                      |
|                                        |                | VNG2217G                                                             |
|                                        |                | PdhA2; pyruvate dehydrogenase alpha subunit                          |
|                                        |                | VNG2218G                                                             |
|                                        |                | PdhB; pyruvate dehydrogenase beta subunit                            |
| Lysine biosynthesis                    | 6              | AspC1; putative aminotransfer                                        |
|                                        | 1              | VNG2418G                                                             |
| Lysine degradation                     | 9              | Fad2; Enoyl-CoA hydratase                                              |
|                                        | 6              | VNG0428G                                                             |
|                                        |                | VNG0681G                                                             |
|                                        |                | Hbd1; 3-hydroxyacyl-CoA dehydrogenase                                 |
|                                        |                | VNG0771G                                                             |
|                                        |                | AldY2; aldehyde dehydrogenase (retinol)                              |
|                                        |                | VNG1313G                                                             |
|                                        |                | Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase          |
|                                        |                | VNG1414G                                                             |
|                                        |                | GlyA; glycine hydroxyethyltransferase                                |
|                                        |                | VNG2224G                                                             |
|                                        |                | Ocd1; ornithine cyclodeaminase                                       |
|                                        |                | VNG2513G                                                             |
|                                        |                | AldV1; aldehyde dehydrogenase (retinol)                              |
| Arginine and proline metabolism       | 17             | GdhB; glutamate dehydrogenase                                        |
|                                        | 9              | VNG0161G                                                             |
|                                        |                | VNG0628G                                                             |
|                                        |                | GdhA1; glutamate dehydrogenase                                       |
|                                        |                | VNG0629G                                                             |
|                                        |                | AspB2; aspartate aminotransfer                                       |
|                                        |                | VNG0771G                                                             |
|                                        |                | AldY2; aldehyde dehydrogenase (retinol)                              |
|                                        |                | VNG2224G                                                             |
|                                        |                | Ocd1; ornithine cyclodeaminase                                       |
|                                        |                | VNG2513G                                                             |
|                                        |                | AldV1; aldehyde dehydrogenase (retinol)                              |
|                                        |                | VNG6312G                                                             |
|                                        |                | ArgS; arginine-tRNA synthetase                                       |
|                                        |                | VNG6315G                                                             |
|                                        |                | ArcB; ornithine CarBamoyltransferase                                 |
|                                        |                | VNG6316G                                                             |
|                                        |                | ArcC; CarBamate kinase                                               |
| Histidine metabolism                  | 17             | AldY2; aldehyde dehydrogenase (retinol)                              |
|                                        | 4              | VNG0771G                                                             |
|                                        |                | HisS; histidyl-tRNA synthetase                                       |
|                                        |                | VNG2005G                                                             |
|                                        |                | HisA; phosphoribosylformimin-5-aminoimidazole                         |
|                                        |                | VNG2224G                                                             |
|                                        |                | CarBoxamido ribotide isomerase                                      |
|                                        |                | VNG2513G                                                             |
|                                        |                | AldV1; aldehyde dehydrogenase (retinol)                              |
| Tyrosine metabolism                   | 9              | AspB2; aspartate aminotransfer                                       |
|                                        | 1              | VNG0629G                                                             |

35
| Pathway                                      | Gene Count | Gene Count | Gene Name                                      | Function                                               |
|----------------------------------------------|------------|------------|-----------------------------------------------|--------------------------------------------------------|
| Phenylalanine metabolism                     | 5          | 1          | VNG0629G AspB2; aspartate aminotransferase     |                                                        |
| Benzoate degradation                         | 5          | 1          | VNG2063G Aca; probable acetyl-coa acetyltransferase |                                                        |
| Tryptophan metabolism                        | 12         | 6          | VNG0428G Fad2; Enoyl-CoA hydratase             |                                                        |
|                                              |            |            | VNG0681G Hbd1; 3-hydroxyacyl-CoA dehydrogenase |                                                        |
|                                              |            |            | VNG0771G AldY2; aldehyde dehydrogenase (retinol) |                                                        |
|                                              |            |            | VNG1313G Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase |                                                        |
|                                              |            |            | VNG2513G AldY1; aldehyde dehydrogenase (retinol) |                                                        |
|                                              |            |            | VNG6294G PerA; peroxidase / catalase           |                                                        |
| Phenylalanine, tyrosine and tryptophan biosynthesis | 21        | 2          | VNG0629G AspB2; aspartate aminotransferase     |                                                        |
|                                              |            |            | VNG1142G Eno; ENolAse                          |                                                        |
| beta-Alanine metabolism                      | 6          | 4          | VNG0428G Fad2; Enoyl-CoA hydratase             |                                                        |
|                                              |            |            | VNG0771G AldY2; aldehyde dehydrogenase (retinol) |                                                        |
|                                              |            |            | VNG1313G Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase |                                                        |
|                                              |            |            | VNG2513G AldY1; aldehyde dehydrogenase (retinol) |                                                        |
| Aminophosphonate metabolism                  | 2          | 1          | VNG2418G AspC1; putative aminotransferase      |                                                        |
| Selenoamino acid metabolism                  | 6          | 4          | VNG0606G YrhA; cysteine synthase               |                                                        |
|                                              |            |            | VNG0796G Cgs; cystathionine gamma-synthase     |                                                        |
|                                              |            |            | VNG1172G MetB; cystathionine alpha synthase    |                                                        |
|                                              |            |            | VNG2251G AchY; adEnosylhomocysteinase          |                                                        |
| Cyanoamino acid metabolism                   | 2          | 1          | VNG1414G GlyA; glycine hydroxymethyltransferase |                                                        |
| D-Glutamine and D-glutamate metabolism       | 3          | 2          | VNG0161G GdhB; glutamate dehydrogenase         |                                                        |
|                                              |            |            | VNG0628G GdhA1; glutamate dehydrogenase       |                                                        |
| Glutathione metabolism                       | 1          | 1          | VNG1873G Icd; isocitrate dehydrogenase, NADP  |                                                        |
| Starch and sucrose metabolism                | 8          | 1          | VNG1498G CelM; endogluCanase                   |                                                        |
| Peptidoglycan biosynthesis                   | 1          | 1          | VNG2093G GlnA; glutamine synthetase            |                                                        |
| Glycerolipid metabolism                      | 13         | 3          | VNG0771G AldY2; aldehyde dehydrogenase (retinol) |                                                        |
|                                              |            |            | VNG2513G AldY1; aldehyde dehydrogenase (retinol) |                                                        |
|                                              |            |            | VNG6270G GldA; sn-glycerol-1-phosphate dehydrogenase |                                                        |
| Sphingoglycolipid metabolism                 | 6          | 3          | VNG0679G Acd4; acyl-CoA dehydrogenase          |                                                        |
|                                              |            |            | VNG0775G Acd2; acyl-CoA dehydrogenase          |                                                        |
|                                              |            |            | VNG1482G Acd5; acyl-CoA dehydrogenase          |                                                        |
| Metabolism Type                        | Gene ID | Gene Product | Description                                                                 |
|---------------------------------------|---------|--------------|-----------------------------------------------------------------------------|
| Pyruvate metabolism                   | 15      | VNG0330G     | PpsA; phosphoEnolpyruvate synthase                                          |
|                                       |         | VNG0771G     | AldY2; aldehyde dehydrogenase (retinol)                                     |
|                                       |         | VNG0997G     | Acs2; acetyl-CoA synthetase                                                 |
|                                       |         | VNG1532G     | Acc; pyruvate Carboxylyase subunit A                                        |
|                                       |         | VNG2217G     | PdhA2; pyruvate dehydrogenase alpha subunit                                 |
|                                       |         | VNG2218G     | PdhB; pyruvate dehydrogenase beta subunit                                   |
|                                       |         | VNG2219G     | Dsa; dihydrolipoamide S-acetyltransferase                                    |
|                                       |         | VNG2220G     | LpdA; dihydrolipoamide dehydrogenase                                        |
|                                       |         | VNG2367G     | MdhA; L-malate dehydrogenase                                                |
|                                       |         | VNG2513G     | AldY1; aldehyde dehydrogenase (retinol)                                     |
| Glyoxylate and dicarboxylate metabolism | 5      | VNG1416G     | FdD; methylenetetrahydrofolate dehydrogenase                               |
|                                       |         | VNG2102G     | CitZ; citrate synthase                                                      |
|                                       |         | VNG2367G     | MdhA; L-malate dehydrogenase                                                |
|                                       |         | VNG2574G     | Can; aconitase                                                              |
| 1,2-Dichloroethane degradation        | 2      | VNG0771G     | AldY2; aldehyde dehydrogenase (retinol)                                     |
|                                       |         | VNG2513G     | AldY1; aldehyde dehydrogenase (retinol)                                     |
| Benzoate degradation via CoA ligation | 9      | VNG0428G     | Fad2; Enoyl-CoA hydratase                                                   |
|                                       |         | VNG0681G     | Hbd1; 3-hydroxyacyl-CoA dehydrogenase                                       |
|                                       |         | VNG1306G     | SdhA; SucCinate dehydrogenase subunit A                                     |
|                                       |         | VNG1308G     | SdhB; SucCinate dehydrogenase subunit B                                     |
|                                       |         | VNG1313G     | Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase                 |
| Propanoate metabolism                 | 13     | VNG0428G     | Fad2; Enoyl-CoA hydratase                                                   |
|                                       |         | VNG0771G     | AldY2; aldehyde dehydrogenase (retinol)                                     |
|                                       |         | VNG0997G     | Acs2; acetyl-CoA synthetase                                                 |
|                                       |         | VNG1313G     | Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase                 |
|                                       |         | VNG1541G     | SucC; SucCinyl-CoA synthetase beta chain                                     |
|                                       |         | VNG1542G     | SucD; SucCinyl-CoA synthetase alpha chain                                    |
|                                       |         | VNG2513G     | AldY1; aldehyde dehydrogenase (retinol)                                     |
| Butanoate metabolism                  | 16     | VNG0428G     | Fad2; Enoyl-CoA hydratase                                                   |
|                                       |         | VNG0681G     | Hbd1; 3-hydroxyacyl-CoA dehydrogenase                                       |
|                                       |         | VNG0771G     | AldY2; aldehyde dehydrogenase (retinol)                                     |
|                                       |         | VNG1306G     | SdhA; SucCinate dehydrogenase subunit A                                     |
|                                       |         | VNG1308G     | SdhB; SucCinate dehydrogenase subunit B                                     |
|                                       |         | VNG1313G     | Hbd2; 3-hydroxyacyl-CoA dehydrogenase / Enoyl-CoA hydratase                 |
|                                       |         | VNG2217G     | PdhA2; pyruvate dehydrogenase alpha subunit                                 |
|                                       |         | VNG2218G     | PdhB; pyruvate dehydrogenase beta subunit                                   |
|                                       |         | VNG2513G     | AldY1; aldehyde dehydrogenase (retinol)                                     |
| C5-Branch dibasic acid metabolism     | 3      | VNG1541G     | SucC; SucCinyl-CoA synthetase beta chain                                     |
|                                       |         | VNG1542G     | SucD; SucCinyl-CoA synthetase alpha chain                                    |
| One carbon pool by folate             | 7      | VNG0414G     | PurH; phosphoribosylaminomimidazoleCarboxamide formyltransferase / IMP cyclohydrolase |
| Pathway                                      | Gene Symbols | Genes                                                                 |
|----------------------------------------------|--------------|----------------------------------------------------------------------|
| Methane metabolism                          | VNG1414G     | GlyA; glycine hydroxymethyltransferase                                |
|                                              | VNG1416G     | FolD; methylenetetrahydrofolate dehydrogenase                        |
|                                              | VNG6294G     | PerA; peroxidase / catalase                                           |
| Carbon fixation                              | VNG0629G     | AspB2; aspartate aminotransfer                                       |
|                                              | VNG2367G     | MdhA; L-malate dehydrogenase                                         |
| Reductive carboxylate cycle (CO2 fixation)   | VNG0330G     | PpsA; phosphoEnolpyruvate synthase                                   |
|                                              | VNG0474G     | PorA; pyruvate ferredoxin oxidoreductase, subunit alpha              |
|                                              | VNG0997G     | Acs2; acetyl-CoA synthetase                                          |
|                                              | VNG1125G     | KorB; 2-oxoglutarate ferredoxin oxidoreductase, beta subunit         |
|                                              | VNG1128G     | KorA; 2-oxoglutarate ferredoxin oxidoreductase, alpha subunit        |
|                                              | VNG1306G     | SdhA; SucCinate dehydrogenase subunit A                               |
|                                              | VNG1308G     | SdhB; SucCinate dehydrogenase subunit B                              |
|                                              | VNG1356G     | FumC; fumarate hydratase                                             |
|                                              | VNG1541G     | SucC; SucCyl-CoA synthetase beta chain                                |
|                                              | VNG1542G     | SucD; SucCyl-CoA synthetase alpha chain                               |
|                                              | VNG1873G     | lcd; isocitrate dehydrogenase, NADP                                  |
|                                              | VNG2367G     | MdhA; L-malate dehydrogenase                                         |
|                                              | VNG2574G     | Can; aconitase                                                       |
| Thiamine metabolism                          | VNG2606G     | ThiD; hydroxymethylPyrImidine phosphate kinase                        |
| Pantothenate and CoA biosynthesis            | VNG2122G     | IlVE2; branched-chain amino acid aminotransfer                       |
| Folate biosynthesis                          | VNG0412G     | FolP; folypeolylglutamate synthase / dihydrofolate synthase / dihydroproteoate synthase |
| Porphyrin and chlorophyll metabolism         | VNG1555G     | CobH; precorrin-3B C17-methyltransferase                             |
|                                              | VNG2326G     | HemL; glutamate-1-semialdehyde aminotransfer                         |
| Nitrogen metabolism                          | VNG0161G     | GdhB; glutamate dehydrogenase                                        |
|                                              | VNG0628G     | GdhA1; glutamate dehydrogenase                                       |
|                                              | VNG0796G     | Cgs; cystathionine gamma-synthase                                    |
|                                              | VNG2093G     | GlnA; glutamine synthetase                                           |
|                                              | VNG6316G     | ArcC; CarBamate kinase                                               |
| Sulfur metabolism                            | VNG0606G     | YrhA; cysteine synthase                                              |
|                                              | VNG1172G     | MetB; cystathionine alpha synthase                                   |
| Caprolactam degradation                      | VNG2418G     | AspC1; putative aminotransfer                                        |
| Alkaloid biosynthesis I                       | VNG0629G     | AspB2; aspartate aminotransfer                                       |
| Aminoacyl-tRNA biosynthesis                  | VNG0461G     | AspS; aspartyl-tRNA synthetase                                       |
|                                              | VNG2005G     | HisS; histidyl-tRNA synthetase                                       |
|                                              | VNG2072G     | SerS; serine-tRNA synthetase                                         |
|                                              | VNG2190G     | IleS; isoleucyl-tRNA synthetase                                      |
|                                              | VNG2352G     | GlyS; glycine-tRNA synthetase                                        |
|                                              | VNG6312G     | ArgS; arginine-tRNA synthetase                                       |
| RNA polymerase                               | VNG1336G     | Rpb3; DNA-directed RNA polymerase subunit D                           |
|                                              | VNG2662G     | RpoC; DNA-directed RNA polymerase subunit A”                          |
|                      |     |     |                      |                      |                      |
|----------------------|-----|-----|----------------------|----------------------|----------------------|
|                      |     |     | VNG2664G RpoA; DNA-directed RNA polymerase subunit A' |                      |                      |
| Transcription        | 13  | 2   | VNG2243G TbpE; transcription initiation factor IID |                      |                      |
| factors              |     |     |                      |                      |                      |
| Protein export       | 9   | 2   | VNG1987G SecD; protein-export membrane protein SecD | VNG1988G SecF; protein-export membrane protein SecF |                      |
Figure 1

- Soluble: 232
- Membrane: 165
- Two overlapping sets: 29

Total number of unique proteins: 426