Genomic analysis of *Chromohalobacter salexigens*: clues about its carbon metabolism and the nature of its halophilic properties

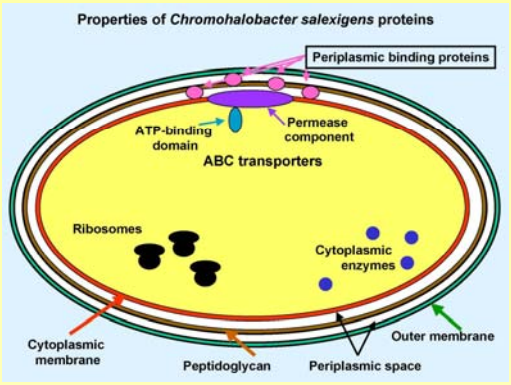
Aharon Oren[1], Laszlo N. Csonka[2], Frank Larimer[3], Paul Richardson[4], Alla Lapidus [4], Bradley W. Goodner[5] and Adam D. Ewing[5]

[1] The Hebrew University of Jerusalem, Jerusalem, Israel; [2] Purdue University, West Lafayette IN, USA; [3] Oak Ridge National Laboratory, Oak Ridge TN, USA; [4]DOE Joint Genome Institute, Walnut Creek CA, USA; [5] Hiram College, Hiram, OH, USA

***Chromohalobacter salexigens* DSM 3043 is a moderately halophilic member of the γ-Proteobacteria. It grows at salt concentrations between 0.9 and 25% with an optimum at 7.5-10%. Pulsed-field gel electrophoresis of total DNA showed that *C. salexigens* possesses a chromosome of ~3.9 Mbp (63.9% G+C) and a low copy plasmid of < 100 kbp.**

A draft sequence of the *C. salexigens* genome has been determined to an 8X coverage by the Joint Genome Institute of the US Department of Energy. Within the 3.7 Mbp unique sequences generated, 3370 predicted protein-coding genes were identified and provisionally annotated (http://genome.jgi.psf.org/microbial/index.html).

Over 60% of the ORFS show the highest similarity to orthologs from other γ-Proteobacteria. From the sequence, we find good indication that the organism has all the enzymes of glycolysis, hexose monophosphate shunt, Entner-Doudoroff pathway and the TCA cycle. We were able to rationalize the pathways of metabolism of many of the common sugars and mono- and dicarboxylic acids. We could account only partially for the metabolic pathway of benzoate, 4-hydroxybenzoate, and 3,4-hydroxybenzoate.



Compound	Enzymes in that could be predicted from the sequence	Predicted enzymes in catabolic pathway that could not be found in the sequence	Can we account for all the necessary metabolic enzymes?
D-Glucose	several potential ABC and ?ny-coupled permeases; two glucokinases; PQQ-dependent glucose oxidase; gluconolactonase		Yes
D-Gluconate	H+-coupled transporter; gluconkinase		Yes
D-Fructose	PTS enzyme I; PTS Hpr; PTS Enzyme IIBC^{Fru}; fructose-1-phosphate kinase; fructose-6-phosphate kinase		Yes
D-Mannose	hexokinase?; phosphomannose isomerase		Yes
D-Galactose	hexokinase?; UDP-glucose pyrophosphorylase; UDP-glucose 4-epimerase	galactose-1-P uridyl transferase	No
myo- (meso)-Inositol	inositol dehydrogenase, poor match; 2-keto-myo-inositol dehydrase, poor match; possible 2,3-diketo-4-d-inositol hydratase; possible 2-deoxy-5-ketogluconate-P aldolase; malonyl (methylmalonyl?) semialdehyde dehydrogenase; myo-inositol 1-P phosphatase (I-1-Pase), function unknown	2-d-5-ketogluconate kinase	No
Lactose		β-galactosidase 6-P-β-galactosidase	No
Maltose	α-glucosidase (maltase)		Yes
Trehalose	α-α-phosphotrehalase (α-glucosidase?)	periplasmic trehalase	Maybe
Mannitol	mannitol → fructose dehydrogenase		Yes
Sorbitol	sorbitol → fructose dehydrogenase		Yes
Galactitol (dulcitol)	tagatose-6-P kinase; possible tagatose (fructose?)-1,6-P$_2$ aldolase	transport or phosphorylation system; galactitol-P dehydrogenase	No
D-Glucarate (saccharate)	glucarate dehydratase; 5-keto-4-d-D-glucarate aldolase		Yes
Sucrose	α-glucosidase (maltase)		Yes
L-Arabinose	ribulokinase?	arabinose ↔ ribulose isomerase; ribulose-5-P epimerase	No

D-Ribose	ribokinase		Yes
D-xylose	xylulose kinase	xylose ↔ xylulose isomerase	No
D-Erythritol	kinase?	erythritol-1-P dehydrogenase; erythrulose-1-P dehydrogenase	No
TCA cycle intermediates: citrate, ketoglutarate, succinate, fumarate, malate	several potential Na$^+$ and H$^+$ linked tri- and dicarboxylic acid transporters, specificity can't be inferred; subsequent metabolism via TCA cycle		Yes
Acetate	acetyl CoA synthetase; isocitrate lyase; malate synthase		Yes
Ethanol	alcohol dehydrogenases (uncertain specificity); aldehyde dehydrogenases (uncertain specificity)		Maybe
Glycerol	glycerol kinase; glycerol-3-phosphate dehydrogenase		Yes
D-Tartrate	tartrate dehydratase α, β subunits		Yes
D,L-Glycerate	D-glycerate kinase		Yes on D-glycerate
Propionate	propionyl CoA synthetase; 2-methylcitrate synthase; 2-methylcitrate dehydrates; 2-methylisocitrate lyase		Yes
Malonate	malonyl-CoA. ACP-SH transferase malonate decarboxylase (β, γ subunits, poor match	malonate decarboxylase α, γ subunits 2-(5'-triphosphoribosyl)-3'-dephospho-CoA synthase; phosphoribosyl-dephospho-CoA transferase	No

Compound	Enzymes in that could be predicted from the sequence	Predicted enzymes in catabolic pathway that could not be found in the sequence	Can we account for all the necessary metabolic enzymes?
Benzoate	See figure		Maybe
Protocatechuate (3, 4-dihydroxy-benzoate)	protocatechuate 3,4-dioxygenase α, β subunits; 3-carboxy-cis,cis-muconate cycloisomerase; 4-carboxymuconolactone decarboxylase		Yes, if 3-oxoadipate enol-lactone can be metabolized
4-Hydroxy-benzoate	4-hydroxybenzoate 3-monooxygenase		Yes
Toluene	toluene 2,3-dioxygenase α subunit toluene cis-dihydrodiol dehydrogenase catechol 2,3-dioxygenase I, II 2-hydroxy-6-oxohepta-2,4-dienoate hydrolase; poor match to 4-hydroxy-2-ovovalerate aldolase	toluene 2,3-dioxygenase β subunit, 2-oxopent-4-enoate hydratase; toluene 2-monooxygenase; toluene 3-monooxygenase; toluene 4-monooxygenase; toluene / o-xylene monooxygenase; 4-cresol dehydrogenase; p-hydroxybenzaldehyde dehydrogenase	No

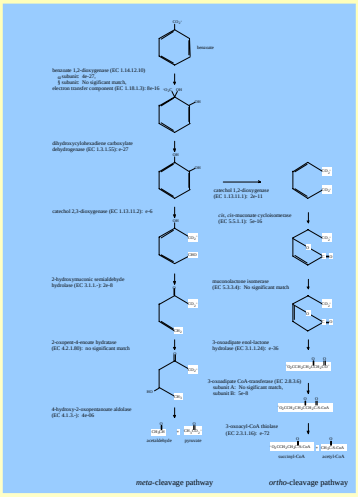
Comparison of the amino acid composition of different categories of proteins of *C. salexigens* and non-halophilic γ-Proteobacteria (*E. coli*, *P. aeruginosa*, *V. cholerae*) showed only a slight excess of acidic residues in the cytoplasmic proteins, and no significant differences were found in the acidity of membrane-bound proteins. In contrast, the periplasmic binding proteins of the ABC transport systems of *C. salexigens* have a pronouncedly lower mean pI value than the non-halophiles. *V. cholerae*, adapted to brackish water, shows intermediate values.

Gene category	*C. salexigens*	*E. coli*	*P. aeruginosa*	*V. cholerae*	*Halobacterium*
Central metabolism	5.10 ± 0.34 (12)	5.66 ± 0.49 (12)	5.80 ± 0.55 (12)	5.54 ± 0.44 (12)	4.22±0.14 (6)
Ribosomes	10.10 ± 2.15 (53)	10.41 ± 1.67 (53)	10.37 ± 1.81 (53)	10.28 ± 1.88 (53)	5.81 ± 2.56 (55)
ATP binding components	6.67 ± 1.47 (50)	6.93 ± 1.63 (73)	7.06 ± 1.62 (62)	6.90 ± 1.41 (54)	4.42 ± 1.10 (26)
Permease components	9.19 ± 1.56 (68)	9.18 ± 1.44 (77)	9.24 ± 1.49 (60)	8.65 ± 1.75 (53)	6.68 ± 2.46 (27)
Periplasmic binding components	4.54 ± 1.13 (55)	6.81 ± 1.56 (59)	7.28 ± 1.33 (44)	5.68 ± 1.03 (39)	4.11 ± 0.14 (6)

Mean pI values of different categories of proteins of *Chromohalobacter salexigens*, as compared with the orthologs form the non-halophilic *Escherichia coli* K-12, *Pseudomonas aeruginosa* PA01, *Vibrio cholerae* O1 El Tor N16961, and the extremely halophilic archaeon *Halobacterium* sp. NRC-1.

Ribosomes	Acidic	Basic	Acidic/Basic	Ser + Thr	Hydrophobic
C. salexigens	11.3	17.6	0.64	10.0	35.1
E. coli	10.9	18.1	0.60	9.6	38.1
P. aeruginosa	10.9	18.1	0.61	9.8	36.2
V. cholerae	10.7	17.9	0.60	9.5	37.8
Halobacterium	19.0	12.1	1.57	11.2	31.9
Substrate binding proteins					
C. salexigens	14.8	7.8	1.91	11.6	37.6
E. coli	11.2	10.7	1.05	11.4	37.8
P. aeruginosa	11.6	11.5	1.01	10.1	38.4
V. cholerae	11.4	9.7	1.17	12.3	37.9
Enzymes of the central metabolism					
C. salexigens	13.9	10.3	1.35	9.9	37.2
E. coli	12.8	10.7	1.19	10.6	37.5
P. aeruginosa	12.8	10.9	1.17	9.8	38.4
V. cholerae	12.4	10.4	1.20	10.2	38.2

'Halophilic' signatures of periplasmic binding proteins of ABC transporters, ribosomal proteins, and selected enzymes of the central metabolic pathways of *C. salexigens* as compared to *E. coli*, *P. aeruginosa*, *V. cholerae* and *Halobacterium* NRC-1. The values are given in mole-percent of the total number of amino acid residues or as ratios, as appropriate. Boldface: special features of the *C. salexigens* and the *Halobacterium* proteins.



Benzoate metabolic pathways. The expect values indicate the tblastn similarity scores of *C. salexigens* ORFS against queries from genes from the *P. putida* TOL plasmid (for the reactions from benzoate to catechol and for the *meta* pathway) and from *Acinetobacter* (for the *ortho* pathway).

Conclusions

1. The sequence information shows that *C. salexigens* is a versatile heterotroph, and can at least partially metabolize a number of aromatic and xenobiotic compounds. This opens the possibility that the organism might be exploited for biological cleanup of highly saline polluted environments.
2. The acidic nature of the *C. salexigens* periplasmic substrate binding proteins is indicative of salt adaptation and possibly salt dependence of these proteins, and indicates that salt requirement of proteins located external to the cytoplasmic membrane may determine salt requirement of many prokaryotes.

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under Contract No. W-7405-ENG-36.

LBNL-57689 Poster

References

Csonka, L.N., O'Connor, K., Larimer, F., Richardson, P., Lapidus, A., Ewing, A.D., Goodner, B.W., and Oren, A. 2005. What we can deduce about metabolism in the moderate halophile *Chromohalobacter salexigens* from its genomic sequence. In: Gunde-Cimerman, N., Oren, A., and Plemenitaš, A. (eds.). Adaptation to life at high salt concentrations in Archaea, Bacteria, and Eukarya. Springer, Dordrecht, in press.

Oren, A., Larimer, F., Richardson, P., Lapidus, A., and Csonka, L.N. 2005. How to be moderately halophilic with a broad salt tolerance: clues from the genome of *Chromohalobacter salexigens*. Extremophiles, in press