Phylogenetic Analysis of *Shewanella* Strains by DNA Relatedness derived from Whole Genome Microarray DNA-DNA Hybridization and comparisons with other methods

L.Y. Wu[1,2], T.Y. Yi[1], J.D. Van Nostrand[1,2],, J. Zhou[1,2]

[1]Institute for Environmental Genomics & Department of Botany/Microbiology, University of Oklahoma, Norman, OK, [2]Virtual Institute for Microbial Stress and Survival, Berkeley, CA

ESPP 2 Environmental Stress Pathway Project

DOE GENOMICS:GTL ACCELERATING DISCOVERY FOR ENERGY AND ENVIRONMENT — OFFICE OF SCIENCE U.S. DEPARTMENT OF ENERGY

Q-2386

VIMSS Virtual Institute for Microbial Stress and Survival

OAK RIDGE NATIONAL LABORATORY Managed by UT Battelle for the Department of Energy

IEG Institute for ENVIRONMENTAL GENOMICS

MICHIGAN STATE UNIVERSITY

Berkeley University of California

http://vimss.lbl.gov/

ABSTRACT

Phylogenetic analyses were done for the Shewanella strains isolated from Baltic Sea (38 strains), US DOE Hanford Uranium bioremediation site [Hanford Reach of the Columbia River (HRCR), 11 strains], Pacific Ocean and Hawaiian sediments (8 strains), and strains from other resources (16 strains) with three out group strains, *Rhodopseudomonas palustris*, *Clostridium cellulolyticum,* and *Thermoanaerobacter ethanolicus* X514, using DNA relatedness derived from WCGA-based DNA-DNA hybridizations, sequence similarities of 16S rRNA gene and *gyr*B gene, and sequence similarities of 6 loci of Shewanella genome selected from a shared gene list of the Shewanella strains with whole genome sequenced based on the average nucleotide identity of them (ANI). The phylogenetic trees based on 16S rRNA and *gyr*B gene sequences, and DNA relatedness derived from WCGA hybridizations of the tested *Shewanella* strains share exactly the same sub-clusters with very few exceptions, in which the strains were basically grouped by species. However, the phylogenetic analysis based on DNA relatedness derived from WCGA hybridizations dramatically increased the differentiation resolution at species and strains level within *Shewanella* genus. When the tree based on DNA relatedness derived from WCGA hybridizations was compared to the tree based on the combined sequences of the selected functional genes (6 loci), we found that the resolutions of both methods are similar, but the clustering of the tree based on DNA relatedness derived from WMGA hybridizations was clearer. These results indicate that WCGA-based DNA-DNA hybridization is an idea alternative of conventional DNA-DNA hybridization methods and it is superior to the phylogenetics methods based on sequence similarities of single genes. Detailed analysis is being performed for the re-classification of the strains examined.

Keywords: Phylogenetic analysis, Shewanella, whole genome Microarray

METHODS AND STRAINS

1. Strains tested (table 1);
2. DNA isolation. The genomic DNAs were isolated from pure cultures using phenolchloroform method.
3. 16S rRNA and *gyr*B amplification and phylogenetic tree construction. 16S rRNA and *gyr*B were amplified from genomic DNAs, Sequenced at Oklahoma Medical Research Center. Phylogenetic tree were constructed using MEGA software with NJ method.
4. Microarray Construction, Labeling, and Hybridization. Genomic DNAs were printed on glass slides at concentrations of 200 ng/µL. Genomic DNA from collected *Shewanella* spp., as well as several strains not belonging to Shewanella (*Thermoanaerobacter ethanolicus* X514, *Rhodopseudomonas palustris*, and *Paracoccus denitrificans) as outside groups* were printed on the arrays. Whole genomic DNA of target strains was fluorescently labeled using the random priming method. Microarray hybridization was carried out on Tecan and performed as described by Wu, L et al (2004).
5. Microarray Scanning and Data Analysis. A ScanArray 5000 Microarray Analysis System was used for scanning microarrays at a resolution of 5 µm with 75% laser power and 65% PMT gain. Scanned image displays were analyzed using ImaGene version 6.0. A grid of individual circles defining the location of each DNA spot on the array was superimposed on the image to designate each fluorescent spot to be quantified. Mean signal intensity was determined for each spot. For decreasing the variations DNA amount from printing, the corresponding signals of Syto 61 of the same arrays stained after completely washed were used for normalization.
6. Data analysis. A distant matrix of DNA relatedness was constructed according to mean signal intensity among tested strains, and clustered using MEGA software. Phylogenetic analyses were done for sequence similarities of 16S rRNA *gyr*B gene, and multiple loci using the same methods.

Table 1. Bacterial strans used in this study

No.	Bacteria Strains	Source of Isolation	No.	Bacteria Strains	Source of Isolation
1	Shewanella sp WAB-1	Wabash River, IN	39	Shewanella baltica OS117	Baltic Sea, Gottland Deep
2	Shewanella sp W3-6-1	marine sediments, Pacific Ocean, (997m, 9-10 cm core), Washington State	40	Shewanella baltica OS109	Baltic Sea, Gottland Deep
3	Shewanella sp W3-18-1	marine sediments, Pacific Ocean, (630m, 5-6 cm core), Washington State	41	Shewanella baltica OS107	Baltic Sea, Gottland Deep
4	Shewanella frigidimarina SW38	North Sea, UK	42	Shewanella baltica OS106	Baltic Sea, Gottland Deep
5	Shewanella putrefaciens SW2.3	temperate marine estruary, UK	43	Shewanella algae OK-1	tetrodotoxin producing red algae
6	Shewanella sp SP-7	clinical isolates, from Seattle WA	44	Shewanella sp MR-7	Black Sea, 60 m depth
7	Shewanella sp SP-32	clinical isolates, from Austria	45	Shewanella oneidensis MR-4	Black Sea, 5 m depth
8	Shewanella sp SP-3	clinal isolate, unknown source	46	Shewanella sp MR-30	freshwater sediments, Green Bay, WS
9	Shewanella sp SP-22	clinical isolates, from Seattle WA	47	Shewanella oneidensis MR-1	Lake Oneida sediments
10	Shewanella sp SP-10	clinical isolates, from Seattle WA	48	Shewanella sp LMP-1	chemocline, Lower Mystic Pond, MA
11	Shewanella amazonensis SB2B	low salinity mud flat sediment at a depth of 1 m, Cabo Cassipore, Amapa, Brazil, 1996	49	Shewanella putrefaciens IR-1	Rice Paddy, Korea
12	Shewanella sp PV-4	Naha Vents, Hawaii	50	Shewanella oneidensis DLM7	Lower Green Bay sediments, Lake Michigan
13	Shewanella baltica OS710	Baltic Sea, Gottland Deep	51	Shewanella putrefaciens CN-32	subsurface rock at Cerro Negro, NM, 1994, Yitai Liu
14	Shewanella baltica OS696	Baltic Sea, Gottland Deep	52	Shewanella sp CL245/73	human cerebrospinal fluid
15	Shewanella baltica OS681	Baltic Sea, Gottland Deep	53	Shewanella sp CE-1	Melt Pool, near Cape Evans, Ross Island, Antarctica
16	Shewanella baltica OS678	Baltic Sea, Gottland Deep	54	Shewanella sp BC-1	Karst Stream water, Blue Clay, IN
17	Shewanella baltica OS652	Baltic Sea, Gottland Deep	55	Shewanella putrefaciens ATCC 8012	surface tainted butter
18	Shewanella baltica OS650	Baltic Sea, Gottland Deep	56	Shewanella putrefaciens ATCC 19857	machine oil
19	Shewanella baltica OS645	Baltic Sea, Gottland Deep	57	Shewanella putrefaciens ATCC 12099	machine oil
20	Shewanella baltica OS641	Baltic Sea, Gottland Deep	58	Shewanella sp ANA-3	arsenic-treated wood in brackish estuary, Woods Hole, MA
21	Shewanella baltica OS631	Baltic Sea, Gottland Deep	59	Shewanella putrefaciens 200	oil pipeline, Alberta, Canada
22	Shewanella baltica OS628	Baltic Sea, Gottland Deep	60	Shewanella putrefaciens 95	surface tainted butter
23	Shewanella baltica OS625	Baltic Sea, Gottland Deep	61	Shewanella baltica 63	oil brine
24	Shewanella baltica OS288	Baltic Sea, Gottland Deep	62	Shewanella SP WS-18	
25	Shewanella baltica OS286	Baltic Sea, Gottland Deep	63	Shewanella sp 1	Hanford Reach of the Columbia River
26	Shewanella baltica OS250	Baltic Sea, Gottland Deep	64	Shewanella sp 2	Hanford Reach of the Columbia River
27	Shewanella baltica OS230	Baltic Sea, Gottland Deep	65	Shewanella sp 3	Hanford Reach of the Columbia River
28	Shewanella baltica OS225	Baltic Sea, Gottland Deep	66	Shewanella sp 4	Hanford Reach of the Columbia River
29	Shewanella baltica OS223	Baltic Sea, Gottland Deep	67	Shewanella sp 5	Hanford Reach of the Columbia River
30	Shewanella baltica OS195	Baltic Sea, Gottland Deep	68	Shewanella sp 9	Hanford Reach of the Columbia River
31	Shewanella baltica OS193	Baltic Sea, Gottland Deep	69	Shewanella sp 10	Hanford Reach of the Columbia River
32	Shewanella baltica OS190	Baltic Sea, Gottland Deep	70	Shewanella sp 11	Hanford Reach of the Columbia River
33	Shewanella baltica OS189	Baltic Sea, Gottland Deep	71	Shewanella sp 12	Hanford Reach of the Columbia River
34	Shewanella baltica OS187	Baltic Sea, Gottland Deep	72	Shewanella sp 16	Hanford Reach of the Columbia River
35	Shewanella baltica OS185	Baltic Sea, Gottland Deep	73	Shewanella sp 19	Hanford Reach of the Columbia River
36	Shewanella sp 184	surface tainted butter	74	Rhodopseudomonas palustris	
37	Shewanella baltica OS167	Baltic Sea, Gottland Deep	75	Thermoanaerobacter ethanolicus X514	
38	Shewanella baltica OS155	Baltic Sea, Gottland Deep	76	Clostridium cellulolyticum	

RESULTS: Multiple Locus Sequence Typing (MLST)

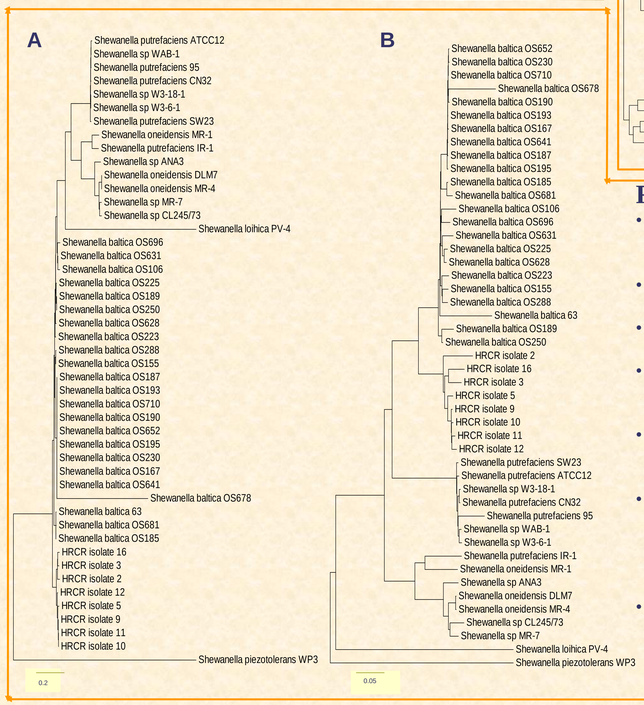

Figure 1. Trees based on MLST

- Genes selected: Three genes coding permease, phosphoglycolate phosphatase, and thymidylate kinase; one gene coding OsmC/Ohr family protein; an unknown periplasmic protein, and another unknown protein.
- Sequences of these genes were collected for about 20 sequenced *shewanella* strains.
- For those strains the whole genome sequences not available, these genes were PCA amplified and sequenced.
- Phylogenetic trees were generated based on sequences of each single functional gene and the combined sequences of all selected functional genes using MEGA 4.0.
- The trees based on sequences of a **single gene** coding an unknown protein (Fig. 1 A) and the **combined sequences** of all selected functional genes (Fig. 1 B) are shown.
- The differentiation resolution of phylogenetic analysis based on a single functional gene is very low and the classification is not consistent with other methods; however, while based on the combined sequences it is higher and the clustering of the tree is more clear and consistent with other methods.
- When the tree based on DNA relatedness derived from WCGA hybridizations was compared to the tree based on the combined sequences of the selected functional genes, we found that the resolutions of both methods were similar, but the clustering of the tree based on DAN relatedness is clearer.

RESULTS: Phylogenetic Analyses based on DNA Relatedness, Sequence similarities of 16S RNA and *gyr*B genes

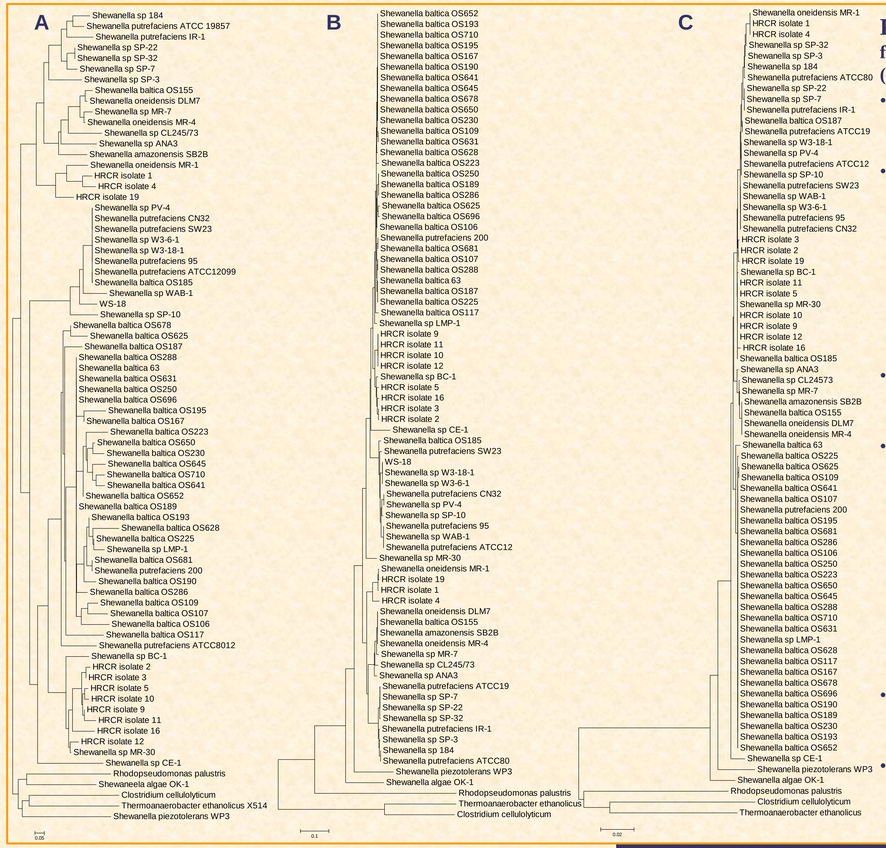

Figure 2. Trees based on DNA relatedness from WCGA hybridization (A), *gyr*B sequence (B), and 16S sequences (C).

- The trees were generated using Molecular Evolutionary Genetics Analysis software version 4.0 and the neighbor-joining algorithms.
- The threes built from different phylogenetic analyses shared a complete same sub cluster (DLM7, OS155, SB2B, MR-4, MR-7, CL245/73, ANA3), four high similar sub clusters (SP-3, SP-7, SP-32, SP-22, IR-1, ATCC19857, 184), (SP-10, WS-18, WAB-1, OS185, ATCC12099, W3-18-1, W3-6-1, CN32,SW23,PV-4), (BC-1, HRCR isolate 2, 3, 5, 9, 10, 11, 12, 16), and (MR-1, HRCR isolate 1, 4, 19).
- Almost all *Shewanella baltica* and *S. putrefaciens* strains were clustered in the same groups in three phylogenetic trees.
- In the phylogenetic tree based on whole genome array hybridization, two strains, OK-1 and WP3 were not grouped with *Shewanella* strains, and the three outgroup strains (*Rhodopseudomonas palustris*, *Thermoanaerobacter ethanolicus* and *Clostridium cellulolyticum*) were not separated perfectly as their classification position . This might be due to the common reason of DND-DNA hybridization that the hybridization signal intensity for distantly related strains are too low to give enough resolution.
- The current results are promising and indicating that MWGA is a powerful and convenient method for bacteria classification.
- Different hybridization temperature, or stringency will be evaluated for increase classification resolution at both strain level and higher hierarchical levels.

CONCLUSIONS

- The accuracy of the classification by WCGA hybridizations was confirmed by the phylogenetic analyses using 16S RNA and *gyr*B genes.
- The differentiation resolution of the WCGA hybridizations method is higher than those using sequence similarities of 16S RNA and *gyr*B genes, and the multiple locus shared by different genomes.
- Using WCGA hybridizations, the 11 HRCR strains can be clustered into two groups: one group contains strains 2, 3, 5, 9, 10, 11, 12, and 16, which are closely related to *Shewanella* sp. BC-1; another group contains strains 1, 4, and 19, which are similar to *Shewanella oneidensis* MR-1. This classification is consistent with that based on *gyr*B gene sequence similarities, and is also consistent with that based on the 16S RNA gene similarities, with one exception that strain 19 is not with the group similar to MR-1, but with the big group.
- Using WCGA hybridization method, most of the Baltic Sea strains are classified together, with only two exceptions that strain OS185 is classified with *Shewanella putrefaciens* strains, and strain OS155 is classified with *Shewanella oneindensis* strains. The classifications of these strains are similar based on 16S RNA and *gyr*B gene sequence similarities, but with a few different exceptions. WCGA hybridization can cluster the major group of Baltic Sea strains further into several sub-clusters, which is different from the other two methods.

ACKNOWLEDGEMENT

This research was supported by the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov) supported by the U. S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL program through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the U. S. Department of Energy, the Environmental Remediation Science Program (ERSP), and was supported by Oklahoma Center for the Advancement of Science and Technology under Oklahoma Applied Research Support Program.