Genomics Data 8 (2016) 134–138

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Data in Brief

Whole genome sequence analysis of an Alachlor and Endosulfan degrading *Pseudomonas* strain W15Feb9B isolated from Ochlockonee River, Florida

Ashvini Chauhan [a,*], Ashish Pathak [a], Ayman Y.I. Ewida [b], Zabrenna Griffiths [a], Paul Stothard [c]

[a] Environmental Biotechnology and Genomics Laboratory, School of the Environment, 1515 S. Martin Luther King Jr. Blvd., Suite 305B, FSH Science Research Center, Florida A&M University, Tallahassee, FL- 32307, USA
[b] Department of Microbiology, Central Laboratory for Environmental Quality Monitoring (CLEQM), National Water Research Center, Egypt
[c] Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G2P5, Canada

ARTICLE INFO

Article history:
Received 11 May 2016
Received in revised form 14 May 2016
Accepted 15 May 2016
Available online 16 May 2016

Keywords:
Alachlor
Endosulfan
Biodegradation
Whole Genome Sequencing (WGS)
Pseudomonas

ABSTRACT

We recently isolated a *Pseudomonas* sp. strain W15Feb9B from Ochlockonee River, Florida and demonstrated potent biodegradative activity against two commonly used pesticides - Alachlor [(2-chloro-2′,6′-diethylphenyl-N (methoxymethyl)acetanilide)] and Endosulfan [(6,7,8,9,10,10-hexachloro-1,5,5a,6,9,9a-hexahydro-6,9methano-2,3,4-benzo(e)di-oxathiepin-3-oxide], respectively. To further identify the repertoire of metabolic functions possessed by strain W15Feb9B, a draft genome sequence was obtained, assembled, annotated and analyzed. The genome sequence of strain 2385 has been deposited in GenBank under accession number JTKF00000000; BioSample number SAMN03151543. The sequences obtained from strain 2385 assembled into 192 contigs with a genome size of 6,031,588, G + C content of 60.34, and 5512 total number of putative genes. RAST annotated a total of 542 subsystems in the genome of strain W15Feb9B along with the presence of 5360 coding sequences. A genome wide survey of strain W15Feb9B indicated that it has the potential to degrade several other pollutants including atrazine, caprolactam, dioxin, PAHs (such as naphthalene) and several chloroaromatic compounds.

Specifications	
Organism	*Pseudomonas* sp.
Strain	W15Feb9B
Sequencer or array type	Illumina HiSeq 2000
Data format	Processed
Experimental factor	Strain isolated from enrichments spiked with Endosulfan and Alachlor
Experimental features	Whole genome sequence analysis, assembly and annotation
Consent	Not applicable
Sample source location	Ochlockonee River, Florida

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/bioproject/265207.

* Corresponding author at: Environmental Biotechnology and Genomics Laboratory, School of the Environment, 1515 S. MLK Blvd., Suite 305B FSHSRC, Florida A&M University, Tallahassee, FL- 32307, USA.
E-mail address: ashvini.chauhan@famu.edu (A. Chauhan).

BioSample ID: SAMN03151543.
WGS ID: JTKF00000000.

2. Text

On a global scale, both endosulfan [(6,7,8,9,10,10-hexachloro-1,5,5a,6,9,9a-hexahydro-6,9methano-2,3,4-benzo(e)di-oxathiepin-3-oxide] and alachlor [(2-chloro-2′,6′-diethylphenyl-N (methoxymethyl)acetanilide)], are two widely used pesticides that are used to enhance agriculture productivity. In fact, worlwide estimates shown that about three million metric tons of pesticides are applied each year [1]. Environmental risk assessment studies have demonstrated that exposure to Endosulfan could result in both acute and chronic risks in non-target animals, mainly birds, fish and mammals [2]. According to the World Health Organization (WHO) estimates, three million people are poisoned by pesticides every year [3].

Once released into the environmental, pesticides can be microbially degraded given that favorable conditions are present that facilitate biodegradation. We previously isolated bacteria with the ability to degrade both endosulfan and alachlor using standard enrichment techniques

http://dx.doi.org/10.1016/j.gdata.2016.05.008

containing these pesticides as the sole source of carbon and energy [4]. This resulted in the isolation of several potent pesticide degrading strains among which, *Pseudomonas* sp. strain W15Feb9B emerged as a potent biodegradative strain; it degraded 55% and 93% of the spiked Alachlor and Endosulfan, respectively [4]. Pseudomonads are widely-distributed bacteria colonizing soils, water and even serving as the significant constituents of animal or plant host microbiome by playing important roles. Additionally, Pseudomonads are extremely important for their abilities in the biodegradation and biotransformation of biogenic and xenobiotic pollutants and thus possess immense potential for a variety of biotechnological applications. Here we report analysis of the draft genome sequence of strain W15Feb9B to further catalog the plethora of relevant traits that this strain might possess.

Briefly, genomic DNA was extracted from strain W15Feb9B and prepared for sequencing on an Illumina HiSeq2000 instrument as described previously [5]. The genome was *de novo* assembled using paired-end reads using the software package CLC Genomics Workbench V6.0 (CLCbio, Cambridge, MA). Contigs were then annotated and genes predicted by IMG ER [6], RAST [7] and the Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP), version 2.0, available from NCBI. As shown in Fig. 1, using CGView [8], a circular genomic map was obtained from the assembled 192 contigs of strain W15Feb9B exhibiting a genome size of 6,031,588 (G + C content 60.34 and N50 of 156,427 bases) and 5512 total number of putative genes, 2 copies of the 16S, 2 copies of the 23S, and 52 trRNA genes, respectively.

A genome-wide analysis of strain W15Feb9B revealed approximately 79.86% protein coding genes associating with function prediction, 29.3% genes were found connected to KEGG pathways and 72.48% genes associated with clusters of orthologous groups of proteins (COGs). We then selected the genome sequences of a total of 1210 Pseudomonads that are available in the IMG/ER to run comparative analysis of strain W15Feb9B, which revealed that it clustered closely with the *Pseudomonas fluorescence* group (data not shown). We then selected a cohort of closely related Pseudomonads, especially those with biodegradative traits, and performed a hierarchical clustering analysis based on the presence of categories of genes and genomes from KEGG (Kyoto Encyclopedia of Genes and Genomes). As shown in Fig. 2, this revealed that strain W15Feb9B was closely related to *Pseudomonas antarctica* BS2772 and *Pseudomonas fluorescence* ATCC13525. Other related groups to strain W15Feb9B belonged to the well-known

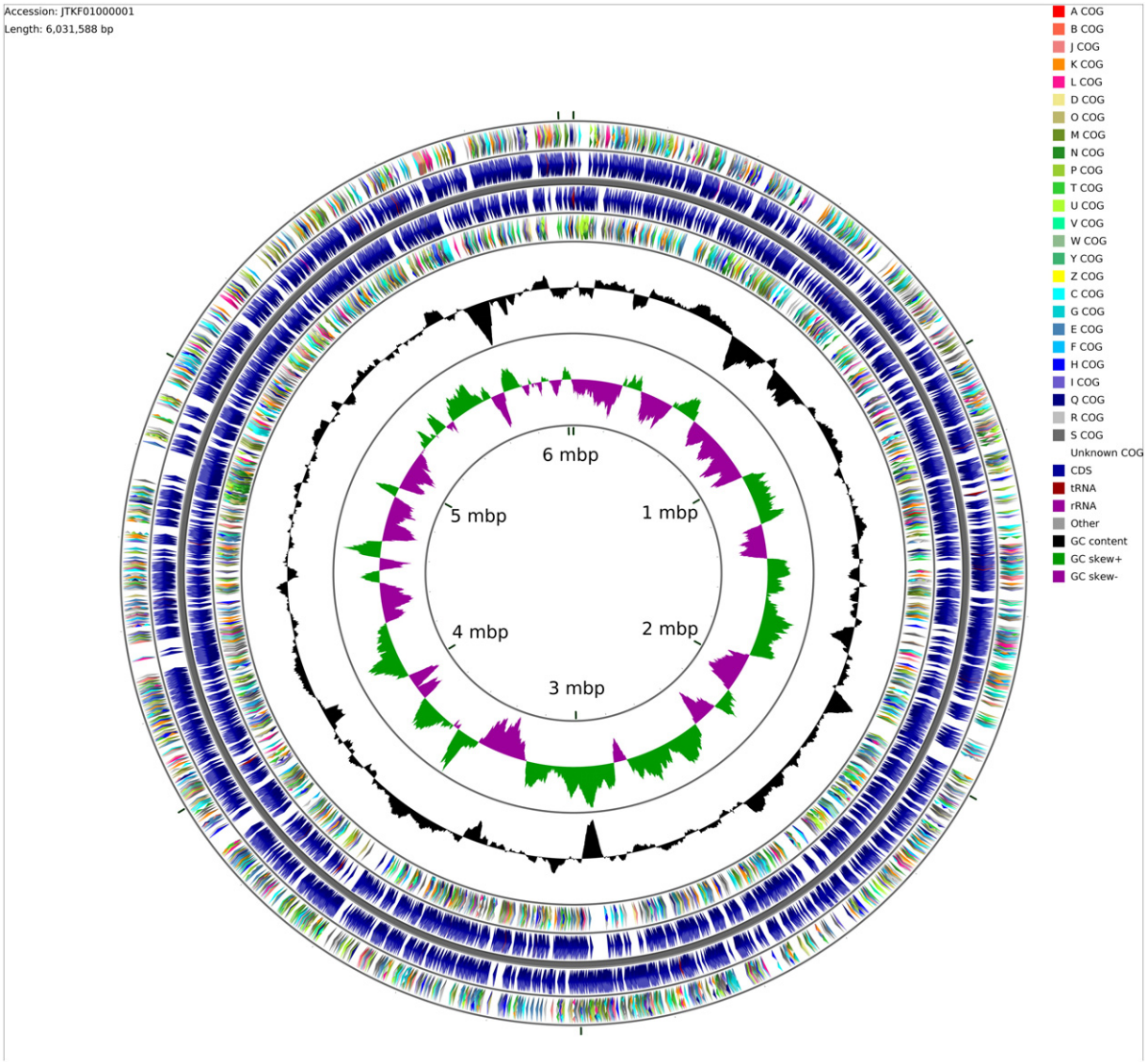

Fig. 1. Circular genome map of *Pseudomonas* sp. strain W15Feb9B with the first (outermost) and fourth rings depicting COG categories of protein coding genes on the forward and reverse strands, respectively. The second and third rings show the locations of protein coding, tRNA, and rRNA genes on the forward and reverse strands, respectively. The black plot depicts GC content with the peaks extending towards the outside of the circle representing GC content above the genome average, whereas those extending towards the center mark segments with GC content lower than the genome average. The innermost plot depicts GC skew. Both base composition plots were generated using a sliding window of 50,000 nt.

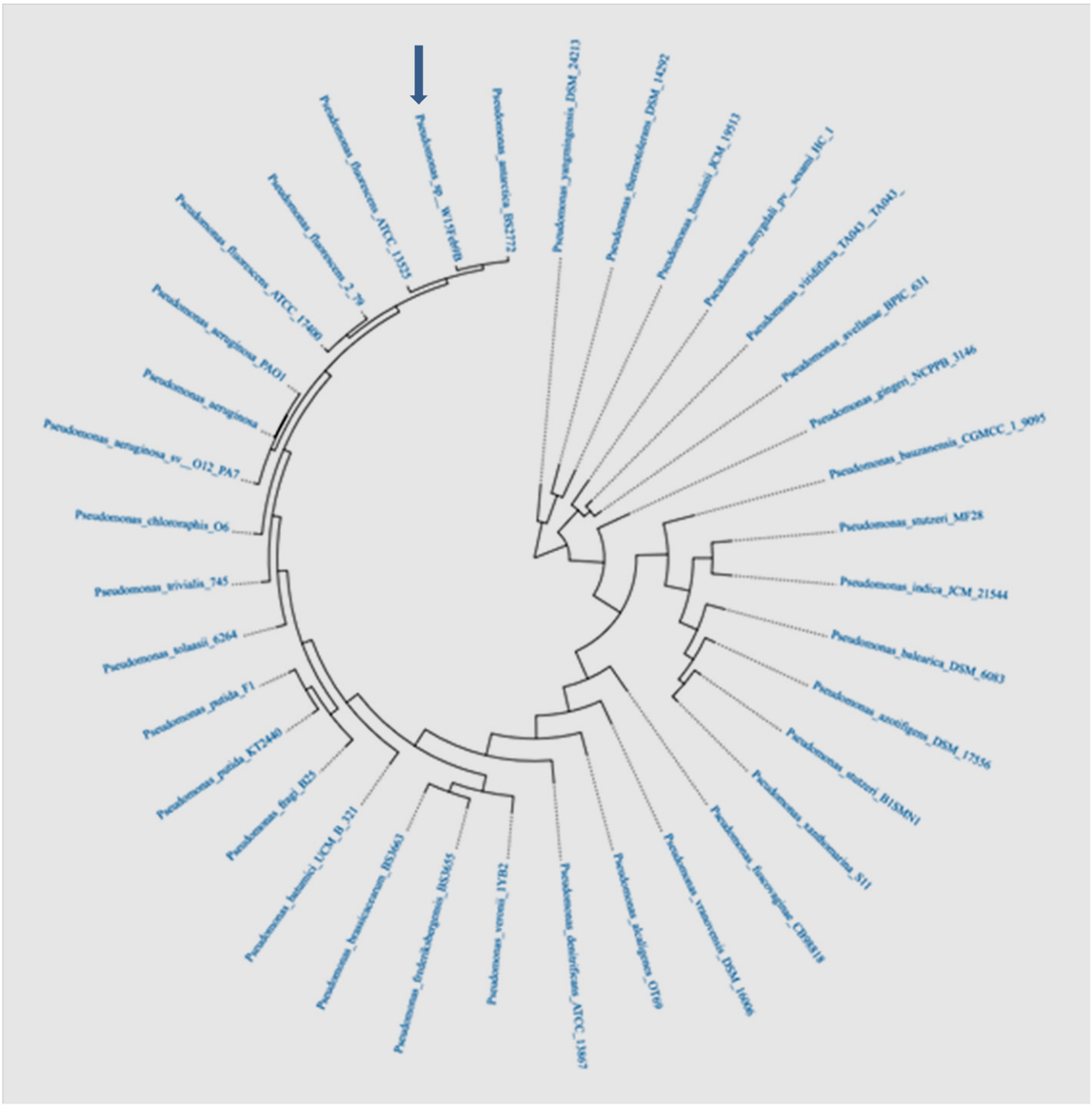

Fig. 2. Hierarchical clustering analysis based phylotree of strain W15Feb9B (identified using the arrow) along with a cohort of other Pseudomonads possessing biodegradative traits. The tree was generated based on the presence of the KEGG (Kyoto Encyclopedia of Genes and Genomes) category of genes present in the selected Pseudomonads. the selected Pseudomonads.

biodegradative strains from the *Pseudomonas aeruginosa* and *Pseudomonas putida* clades, respectively.

Fig. 3 shows a plethora of functional traits identified in strain W15Feb9B connected with KEGG pathways. Of major note are a total of 186 genes identified for energy metabolism, (5.83% of its total genome) along with 68 genes for xenobiotics biodegradation and metabolism (2.13% of its total genome). Other examples of biodegradative genes found in strain W15Feb9B include those involved in the pathways for the degradation of toluene, xylene and PAHs, chloroalkane, chloroalkene and nitrotoluene, respectively.

Additionally, a total of 3995 protein coding genes associated with COGs in strain W15Feb9B were found with further classification into 24 categories. The following 5 most abundant COG subsystems in strain W15Feb9B were related to: amino acid transport and metabolism (n = 487); general function prediction only (n = 411); transcription (n = 393); signal transduction mechanisms (n = 304); cell wall/membrane/envelope biogenesis (n = 279). The strain also contains 22 biosynthetic gene clusters containing 373 genes; a total of 6.77% of the total genome size of strain W15Feb9B. These findings suggest *Pseudomonas* sp. strain W15Feb9B to harbor various genome-enabled metabolic and catabolic processes.

Furthermore, recent body of data originating from whole genome sequencing studies has demonstrated that bacterial genomes consist of a core set of genes that encode for essential metabolic functions, along with a plethora of acquired genes from the bacterium's natural environment by horizontal gene transfer (HGT) mechanisms, providing

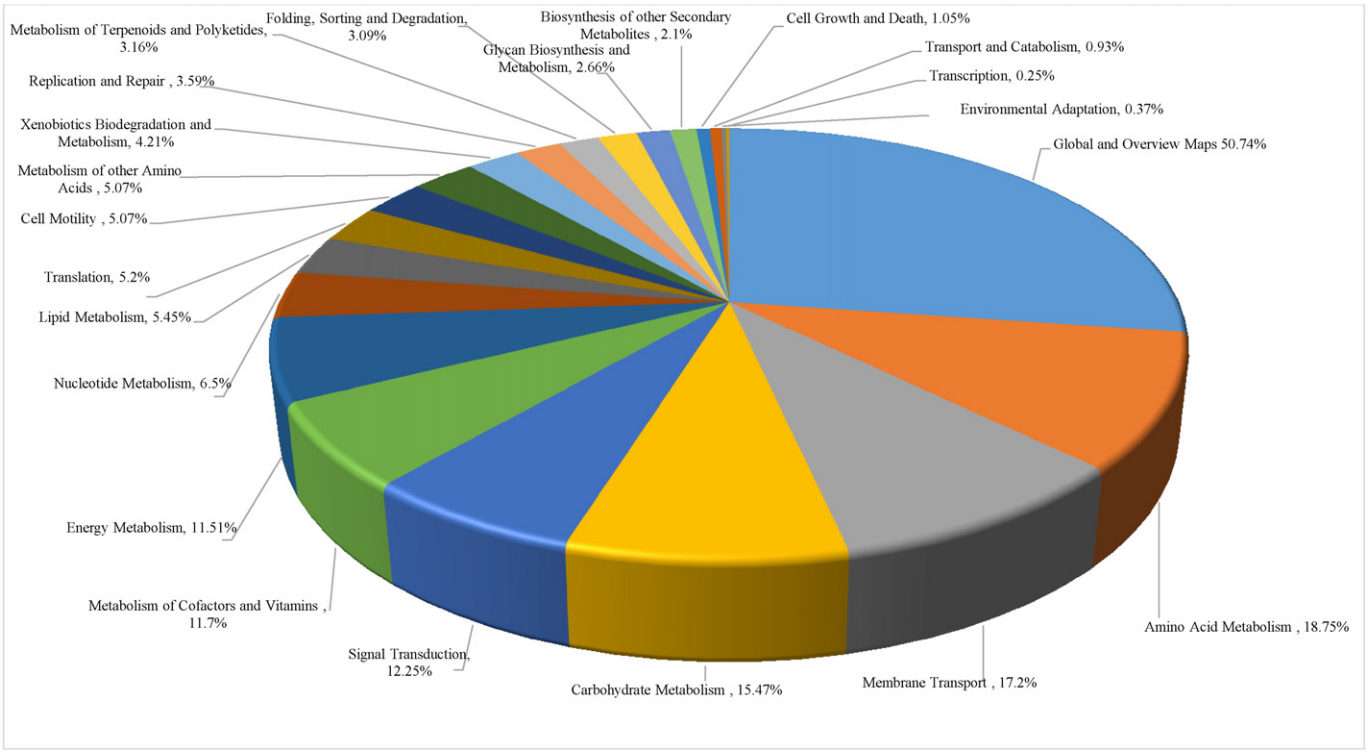

Fig. 3. Functional traits identified from the whole genome sequence of *Pseudomonas* sp. strain W15Feb9B that were found connected with KEGG pathways. The figure was prepared by analysis of the whole genome using IMG/ER and plotting the data using Microsoft Excel.

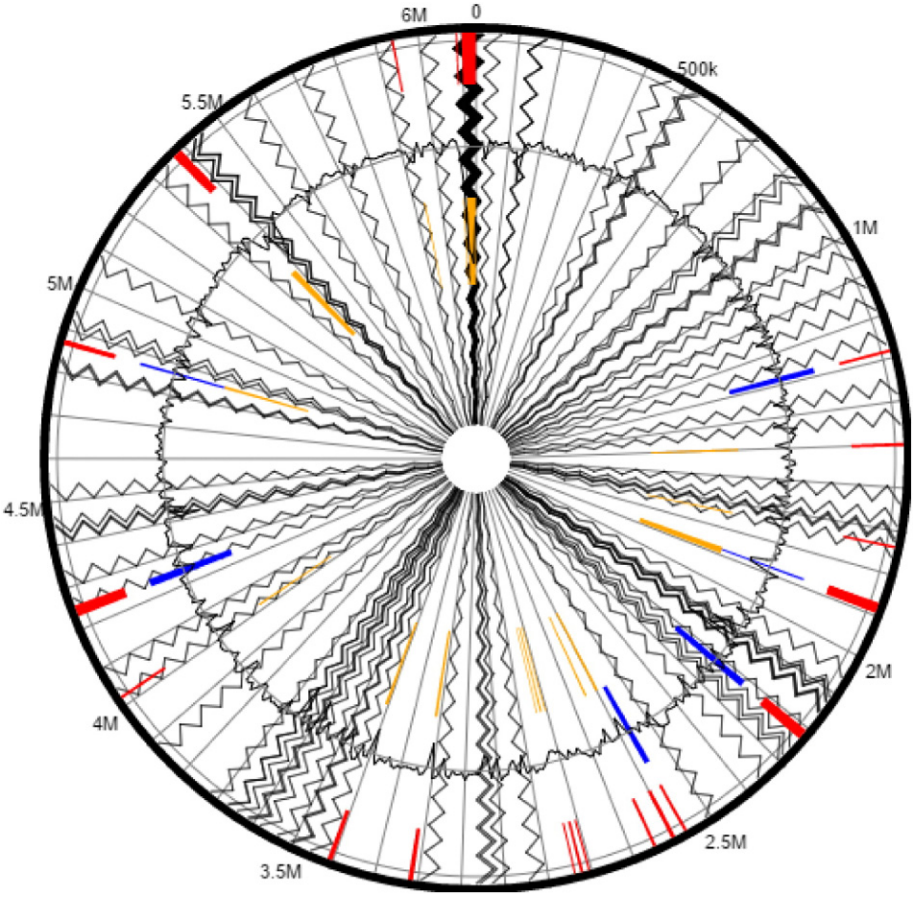

Fig. 4. Putative genomic islands (GEIs) predicted within the genome of *Pseudomonas* sp. strain W15Feb9B when aligned against the complete genome of *Pseudomonas fluorescens* A506 as the reference. The outer black circle represents the scale line in Mbps and the black zig-zag line plot delineates each of the 192 contigs identified from strain W15Feb9B. GEIs obtained from each of the following methods are shown in color: SIGI-HMM (orange), IslandPath-DIMOB (blue), and integrated detection (red), respectively.

the host with evolutionary adaptive traits and genome plasticity. Many such HGT-acquired accessory genes occur as orthologous blocks referred to as genomic islands (GEIs) [9]. GEIs have been mainly attributed to render virulence or antibiotic resistance to the host bacteria, but genome sequencing studies have also identified other adaptive functional traits brought about by GEIs to the host bacterium. As a result, GEIs are now broadly classified into 4 categories based on their associated functions - pathogenicity islands (PAIs), harboring virulence genes; metabolic islands (MIs), genes for the biosynthesis of secondary metabolites; resistance islands (RIs), resistance genes, for example, against antibiotics; and symbiotic islands (SIs), that facilitate symbiotic associations of the host with other micro- and macroorganisms.

We used Island Viewer to identify genomic islands (GEIs) (http://www.pathogenomics.sfu.ca/islandviewer/) [10] in strain W15Feb9B. Island Viewer predicts GEIs integrating two widely used sequence composition based GEI prediction methods - SIGI-HMM and IslandPath-DIMOB along with a comparative GEI prediction method - IslandPick. Specifically, SIGI-HMM is a sequence composition based GEI prediction method which uses Hidden Markov Model (HMM) and measures codon usage to identify possible GEIs and IslandPath-DIMOB visualizes several common characteristics of GEIs that include abnormal sequence composition or the presence of genes known to correspond to mobile genomic elements. Interestingly, a total of 44 genomic islands were identified from within the genome of *Pseudomonas* sp. strain W15Feb9B when aligned against the complete genome sequence of *Pseudomonas fluorescens* A506 as the reference (Fig. 4). When GEI-encoded gene homologues were studied by BLAST, several of them closely affiliated with genes previously shown active in catabolic processes, providing a strong evidence that the GEIs were horizontally acquired by strain W15Feb9B from other biodegradative bacteria to facilitate survival in its native environment.

3. Nucleotide sequence accession number

The draft genome sequence of *Pseudomonas* sp. strain W15Feb9B obtained in this study is available in GenBank under the accession number of JTKF00000000.

Acknowledgements

This work was jointly supported by the following grants: USAID-Egypt grant (004371) and the US Department of Defense (DoD) grants - W911NF-10-1-0146 and W911NF-10-R-0006. We also acknowledge funding provided by the Savannah River Nuclear Solutions under subcontract #0000217620 and the support provided by Dr. Stefan Green, University of Illinois at Chicago.

References

[1] D. Pimentel, Environmental and Economic Costs of the Application of Pesticides Primarily in the United States. in: R. Peshin, A.K. Dhawan (Eds.),Integrated Pest Management: Innovation-Development Process, 1, 2009.
[2] R.I.M. Sunderam, D.M.H. Cheng, G.B. Thompson, Toxicity of endosulfan to native and introduced fish in Australia. Environ. Toxicol. Chem. 11 (1992) 1469–1476, http://dx.doi.org/10.1002/etc.5620111012.
[3] Public Health Impact of Pesticides Used in Agriculture. WHO in collaboration with UN Environment Program, WHO, Geneva (1990) (accessed 05.01.15) http://whqlibdoc.who.int/publications/1990/9241561394.pdf.
[4] A.Y.I. Ewida, Biodegradation of alachlor and endosulfan using environmental bacterial strains. World Appl. Sci. J. 32 (4) (2014) 540.
[5] A. Chauhan, S. Green, A. Pathak, J. Thomas, R. Venkatramanan, Whole-genome sequences of five oyster-associated bacteria show potential for crude oil hydrocarbon degradation. Genome Announc. 1 (5) (2013) (3).
[6] V.M. Markowitz, M. Mavromatis, N.N. Ivanova, I.-M.A. Chen, K. Chu, N.C. Kyrpides, IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics 25 (2009) 2271–2278.
[7] R. Aziz, D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, K. Formsma, S. Gerdes, E. Glass, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Osterman, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology. BMC Genomics 9 (2008) 75.
[8] J.R. Grant, A.S. Arantes, P. Stothard, Comparing thousands of circular genomes using the CGView Comparison Tool. BMC Genomics 13 (2012) 202.
[9] M. Juhas, J.R. van der Meer, M. Gaillard, R.M. Harding, D.W. Hood, D.W. Crook, Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol. Rev. 33 (2) (2009) 376–393.
[10] B.K. Dhillon, M.R. Laird, J.A. Shay, G.L. Winsor, R. Lo, F. Nizam, S.K. Pereira, N. Waglechner, A.G. McArthur, M.G. Langille, F.S. Brinkman, IslandViewer 3: more flexible, interactive genomic island discovery, visualization and analysis. Nucleic Acids Res. 43 (W1) (2015) W104–W108, http://dx.doi.org/10.1093/nar/gkv401.