

Whole-Genome Sequences of 13 Endophytic Bacteria Isolated from Shrub Willow (*Salix*) Grown in Geneva, New York

Huan You Gan,^a Han Ming Gan,^a Michael A. Savka,^b Alexander J. Triassi,^b Matthew S. Wheatley,^b Lawrence B. Smart,^c Eric S. Fabio,^c André O. Hudson^b

School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, Selangor, Malaysia^a; The Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA^b; Department of Horticulture, Cornell University, New York State Agricultural Experiment Station, Geneva, New York, USA^c

Shrub willow, *Salix* spp. and hybrids, is an important bioenergy crop. Here we report the whole-genome sequences and annotation of 13 endophytic bacteria from stem tissues of *Salix purpurea* grown in nature and from commercial cultivars and *Salix viminalis* × *Salix miyabeana* grown in bioenergy fields in Geneva, New York.

Received 14 April 2014 Accepted 15 April 2014 Published 8 May 2014

Citation Gan HY, Gan HM, Savka MA, Triassi AJ, Wheatley MS, Smart LB, Fabio ES, Hudson AO. 2014. Whole-genome sequences of 13 endophytic bacteria isolated from shrub willow (*Salix*) grown in Geneva, New York. *Genome Announc.* 2(3):e00288-14. doi:10.1128/genomeA.00288-14.

Copyright © 2014 Gan et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to André O. Hudson, aohsbi@rit.edu.

Bioenergy produced from plant biomass as a feedstock has the potential to mitigate concerns regarding climate change and sustainability, among others (1). Shrub willow, *Salix* spp. and hybrids, is an attractive plant for bioenergy given its rapid growth and sustainable growth characteristics (2). There have been numerous efforts regarding the development of shrub willow cultivars suitable for the production of bioenergy (3). Endophytes have been isolated from tissues of tree willows in nature (4), but it is unknown what role endophytes might play in shrub willows growing in intensively managed bioenergy fields. To gain some insights into endophytic bacteria that associate with shrub willow, we embarked on a project which resulted in the identification of 69 unique bacteria, of which 13 were subjected to whole-genome sequencing and annotation. Plant-associated bacteria that are beneficial to *Salix* have the potential to improve crop production while reducing need for inputs. In addition, the identification of

bacteria that are detrimental to growth and development through phytopathogenesis is also of interest as a foundation for breeding for resistance and maximizing growth potential. The endophytic bacteria were initially isolated from surface-sterilized stem tissues from wild accessions of *S. purpurea* growing in nature and from commercial cultivars of *S. purpurea* and *S. viminalis* × *S. miyabeana* growing in fertilized and nonfertilized bioenergy fields in Geneva, NY. The 13 endophytes were initially identified by amplification and nucleotide sequence analysis of the variable 3 region of the 16S rRNA gene (5).

Genomic DNA was isolated from the endophytes using a GenElute bacterial genomic kit (Sigma-Aldrich, St. Louis, MO) and prepared for whole-genome sequencing using a Nextera XT library preparation kit (Illumina, San Diego, CA). Whole-genome sequencing was performed using the Illumina Miseq (150-bp paired-end reads). The reads were error corrected and assembled *de novo* using

TABLE 1 Sequencing and annotation results for the 13 endophytes isolated from *Salix*

Strain	Source ^a	SubID	BioProject no.	BioSample no.	Accession no.	Organism	Genome coverage (×)	Genome size (bp)	No. of contigs	No. of ORFs	No. of tRNAs	No. of rRNAs
RIT273	Fabius	SUB467427	PRJNA239282	SAMN02676620	JFOK00000000	<i>Pantoea agglomerans</i>	138	5,365,338	26	4,914	75	17
RIT283	Fabius	SUB467430	PRJNA239283	SAMN02676621	JFOJ00000000	<i>Staphylococcus haemolyticus</i>	289	2,527,922	81	2,445	49	14
RIT288	Fish Creek	SUB468060	PRJNA239284	SAMN02676622	JFYN00000000	<i>Pseudomonas</i> sp.	191	6,273,290	44	5,547	61	10
RIT293	Fabius	SUB468074	PRJNA239285	SAMN02676623	JFYO00000000	<i>Microbacterium oleivorans</i>	194	2,898,622	11	2,782	49	8
RIT304	Wild	SUB468076	PRJNA239286	SAMN02676624	JFYP00000000	<i>Micrococcus luteus</i>	166	2,506,829	183	2,248	53	4
RIT305	Wild	SUB468079	PRJNA239287	SAMN02676625	JFYQ00000000	<i>Micrococcus luteus</i>	200	2,612,381	110	2,350	49	6
RIT308	Fabius	SUB468082	PRJNA239288	SAMN02676626	JFYR00000000	<i>Janthinobacterium lividum</i>	189	6,212,741	44	5,431	83	19
RIT309	Fabius	SUB468084	PRJNA239289	SAMN02676627	JFYS00000000	<i>Stenotrophomonas</i> sp.	194	4,634,795	45	4,141	77	6
RIT313	Fabius	SUB468085	PRJNA239290	SAMN02676628	JFYT00000000	<i>Delftia</i> sp.	56	6,698,360	122	5,936	78	5
RIT324w	Fabius	SUB468086	PRJNA239291	SAMN02676629	JFYU00000000	<i>Micrococcus luteus</i>	459	2,635,230	118	2,381	53	7
RIT328	Fish Creek	SUB468089	PRJNA239292	SAMN02676630	JFYV00000000	<i>Sphingomonas</i> sp.	73	4,343,511	56	4,002	58	3
RIT341	Fabius	SUB468091	PRJNA239293	SAMN02676631	JFYW00000000	<i>Exiguobacterium</i> sp.	138	3,107,022	15	3,168	62	20
RIT357	Wild	SUB468093	PRJNA239294	SAMN02676632	JFYX00000000	<i>Pseudomonas</i> sp.	69	3,107,022	49	5,552	60	7

^a Wild, wild *Salix purpurea*; Fish Creek, *Salix purpurea* Fish Creek cultivar; Fabius, *Salix viminalis* × *S. miyabeana* Fabius cultivar.

Spades 2.5 (6). Scaffolding of the contigs and *in silico* gap-closing of the resulting scaffolds were performed with SSPACE and GapFiller, respectively (7, 8). Genome annotation was performed using the Prokka annotation pipeline, which incorporated Prodigal 2.60, Aragorn, and RNAmmer 1.2 for the prediction of open reading frames (ORFs), tRNAs, and rRNAs, respectively (9–11). Additional annotation of the predicted protein sequences was done using InterProScan5 (12). The key attributes for the genome sequences and annotation are summarized in Table 1. An in-depth analysis of the genes associated with plant-microbe symbiosis is under way and will be published in a subsequent report.

Nucleotide sequence accession numbers. The nucleotide sequences have been deposited at DDBJ/EMBL/GenBank under the accession numbers provided in Table 1.

ACKNOWLEDGMENTS

A.O.H., A.J.T., M.S.W., and M.A.S. acknowledge the College of Science (COS) at the Rochester Institute of Technology (RIT) for ongoing support. M.A.S. also acknowledges a Dean's Research Initiation Grant (DRIG) from the COS at RIT. This work was supported in part by a United States National Science Foundation (NSF) award to A.O.H. (MCB-1120541) in addition to support from the Monash University Malaysia Tropical Medicine and Biology Multidisciplinary Platform.

REFERENCES

- Karp A, Shield I. 2008. Bioenergy from plants and the sustainable yield challenge. *New Phytol.* 179:15–31. <http://dx.doi.org/10.1111/j.1469-8137.2008.02432.x>.
- Heller MC, Keoleian GA, Volk TA. 2003. Life cycle assessment of a willow bioenergy cropping system. *Biomass Bioenerg.* 25:147–165. [http://dx.doi.org/10.1016/S0961-9534\(02\)00190-3](http://dx.doi.org/10.1016/S0961-9534(02)00190-3).
- Smart LB, Cameron KD. 2012. Shrub willow, p 687–708. *In* Kole C, Joshi S, Shonnard D (ed), *Handbook of bioenergy crop plants*. Taylor and Francis, Boca Raton, FL.
- Doty SL, Oakley B, Xin G, Kang JW, Singleton G, Khan Z, Vajzovic A, Staley JT. 2009. Diazotrophic endophytes of native black cottonwood and willow. *Symbiosis* 47:23–33. <http://dx.doi.org/10.1007/BF03179967>.
- Hudson AO, Ahmad NH, Van Buren R, Savka MA. 2010. Sugarcane and grapevine endophytic bacteria: isolation, detection of quorum sensing signals and identification by 16S v3 rDNA sequence analysis, p 801–806. *In* Mendez-Vilas A (ed), *Current research, technology and education topics in applied microbiology and microbial biotechnology*, vol 2. Formatex Research Center, Badajoz, Spain.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single cell sequencing. *J. Comput. Biol.* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
- Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding preassembled contigs using SSPACE. *Bioinformatics* 27: 578–579. <http://dx.doi.org/10.1093/bioinformatics/btq683>.
- Boetzer M, Pirovano W. 2012. Toward almost closed genomes with GapFiller. *Genome Biol.* 13:2012–2013. <http://dx.doi.org/10.1186/gb-2012-13-6-r56>.
- Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–16. <http://dx.doi.org/10.1093/nar/gkh152>.
- Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108. <http://dx.doi.org/10.1093/nar/gkm160>.
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 29 January 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*. <http://dx.doi.org/10.1093/bioinformatics/btu031>.