1   Comparative genome analysis of *Pseudomonas* genomes including *Populus-*

2   associated isolates

3

4   Se-Ran Jun[1], Trudy M. Wassenaar[3], Intawat Nookaew[2], Loren Hauser[2], Visanu Wanchai[2], Miriam

5   Land[2], Collin Timm[2], Tse-Yuan S. Lu[2], Christopher W. Schadt[2], Mitchel J. Doktycz[2], Dale A.

6   Pelletier[2], David W. Ussery[2,#]

7

8   Running Head: Comparative *Pseudomonas* genome analysis

9   [1] Joint Institute for Computational Sciences, University of Tennessee, Knoxville, TN, USA

10   [2] Biosciences Division, Oak Ridge National Laboratories, Oak Ridge, TN, USA

11   [3] Molecular Microbiology and Genomics Consultants, Zotzenheim, Germany

12

13

14   #Address correspondence to David W. Ussery
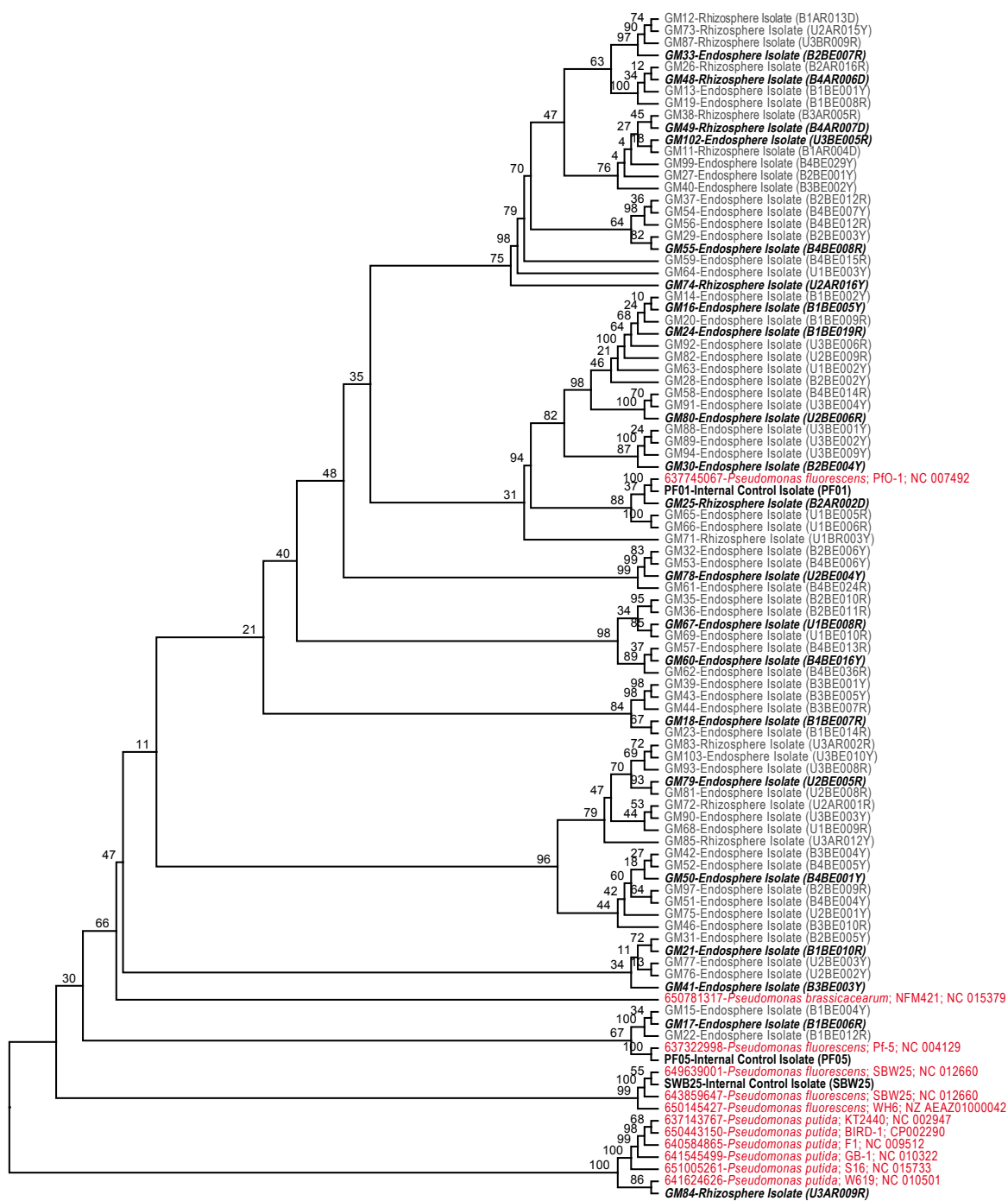
15   Oak Ridge National Laboratory

16   Biosciences Division

17   Oak Ridge, TN 37831

18   Phone: (865) 574-8201

19   Fax: (865) 574-3555

20   Email: usserydw@ornl.gov

21

22

Figure S1. Phylogenetic analysis of partial *rpoD* genes found in isolates of *Pseudomonas* from

25   *Populus deltoides* roots and selected references taxa (Maximum likelihood, bootstraps = 100).

26   Genes from reference isolates are listed by their GI number. Strains from the current study are

27   labeled to denote the compartment (endosphere or rhizosphere) and tree from which the

28   isolate originated. Isolates representing diverse and well-supported groups were selected for

29   complete genome sequence analysis, indicated in bold. Reference complete genomes are

30   shown in red.
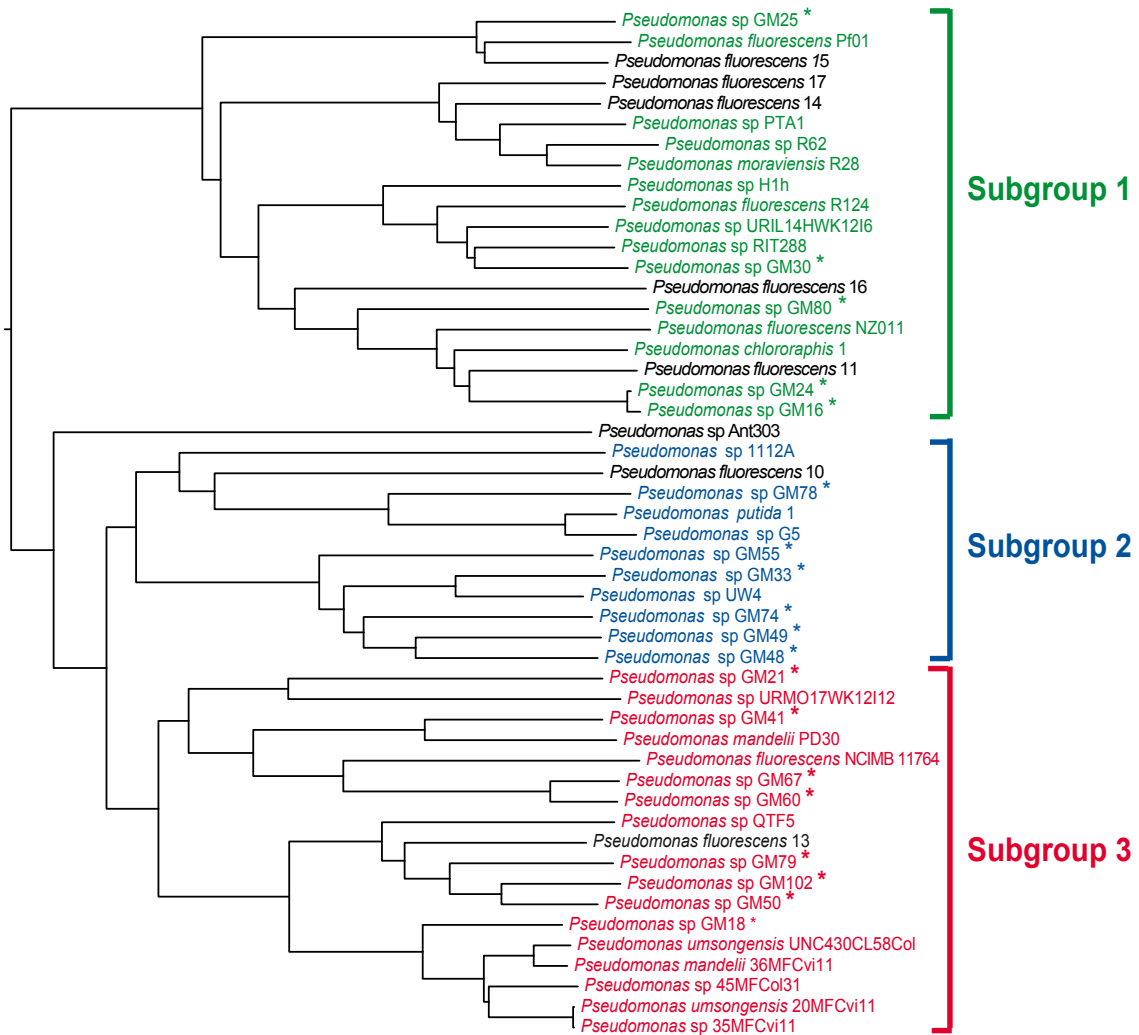
31

32

33

34

35

36

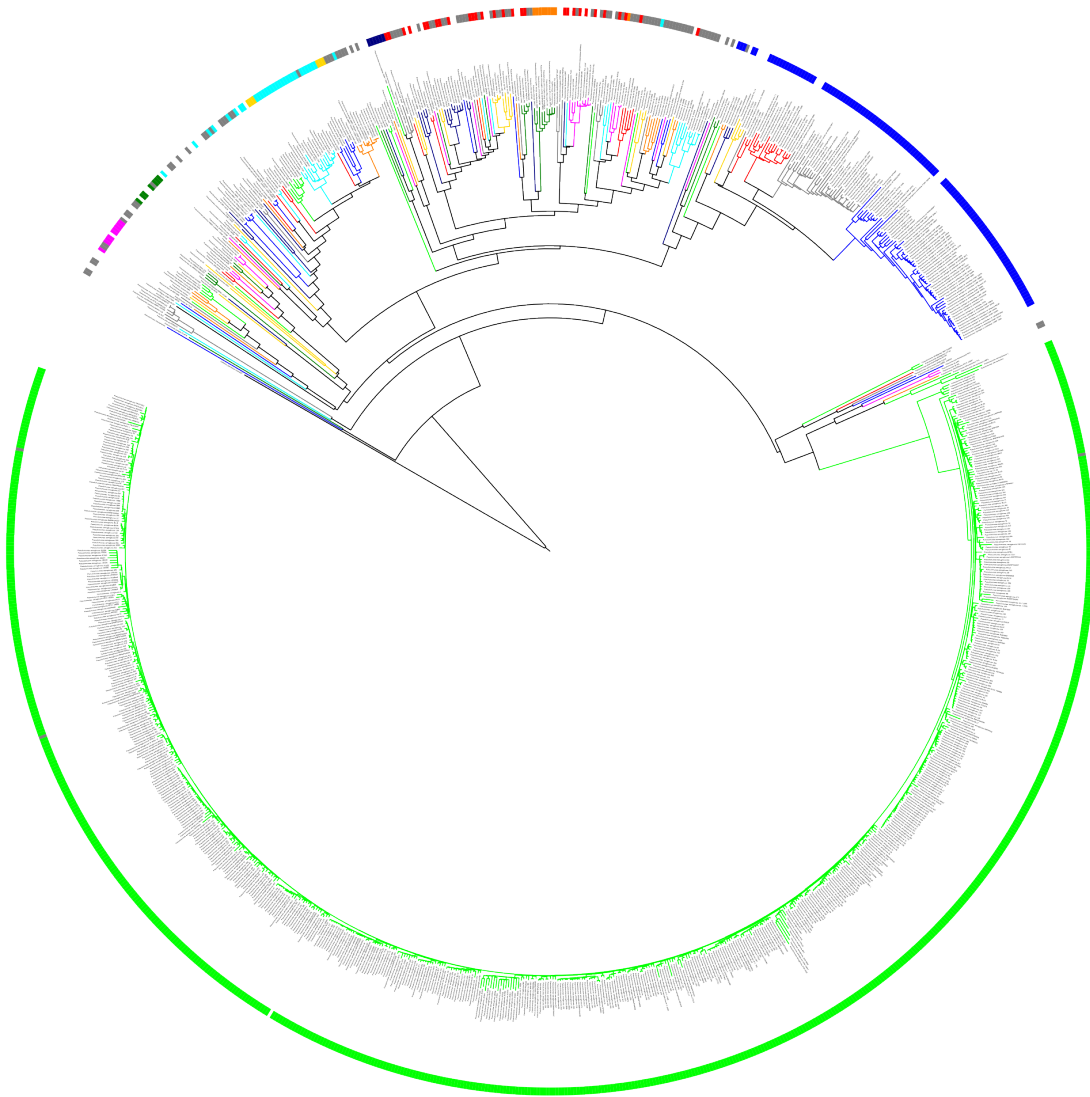37

38

39

40

41

42

43

44

45

46

Figure S2. AAI-based tree of 19 *Populus-associated Pseudomonas* isolates and related *Pseudomonas* species with newly sequenced *P. fluorescens* in black.
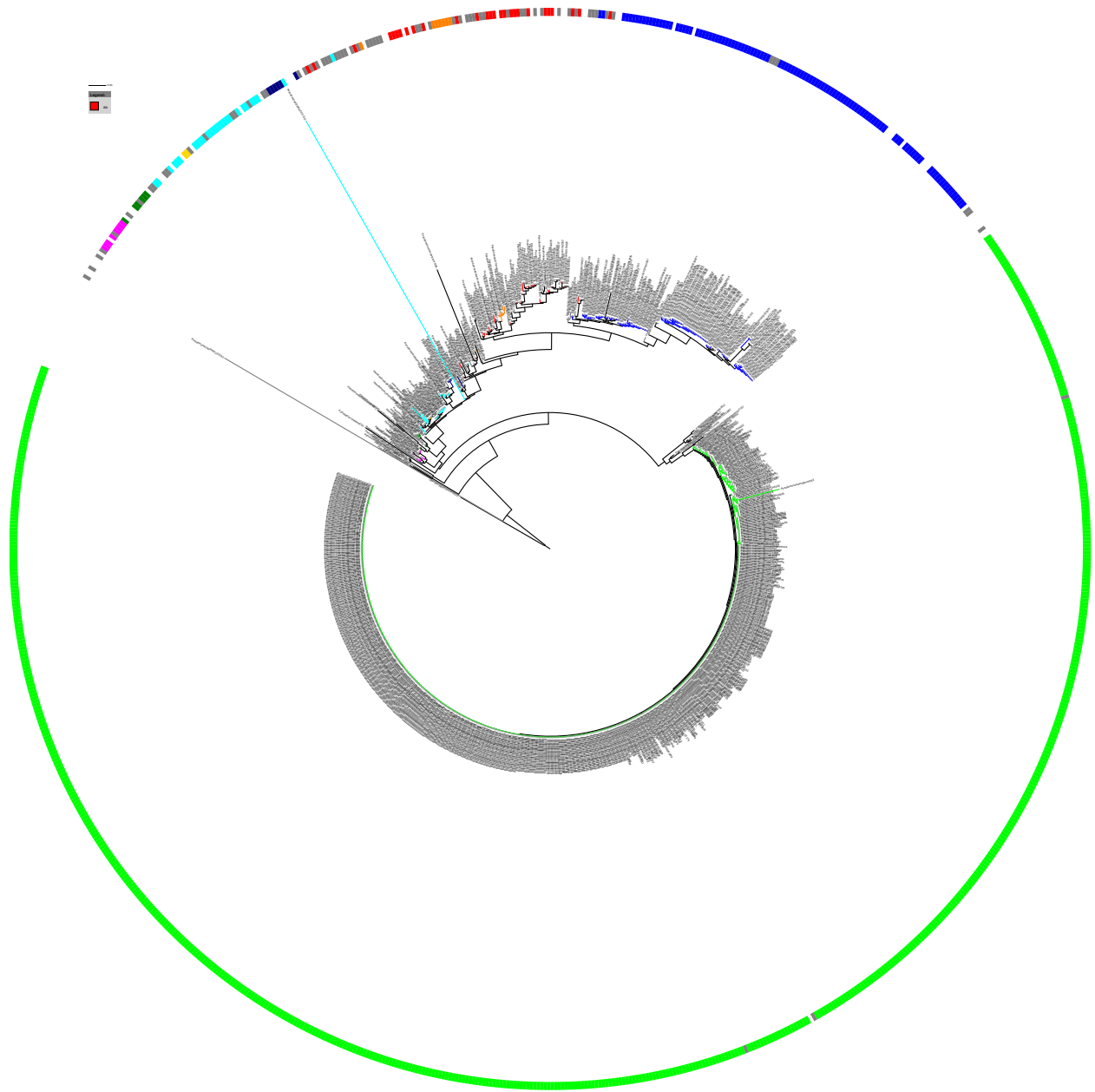
Figure S3. AAI-based tree of 1073 *Pseudomonas* genomes. The outer ring and branches are colored by *Pseudomonas* species groups defined by NCBI taxonomy.

58

Figure S4. 16S-rRNA based tree. We considered only 16S rRNA genes (extracted by RNAmmer

v1.2 (1)) whose length is larger than 1400 nt-long, and chose one 16S rRNA gene whose length

is closest to 1500 when the multiple copies of 16S rRNA genes are present in the genomes,

resulting in 116 *Pseudomonas* without full-length 16S rRNA genes out of 1073 *Pseudomonas*

isolates. We aligned 16S rRNA genes using MAFFT (2), and applied FastTree (3) using a

combined model of generalized time-reversible and a single rate for each site (GTR+CAT model)

65   to approximate a 16S rRNA-based maximum likelihood tree. The outer ring and branches are

66   colored by *Pseudomonas* species groups defined by NCBI taxonomy.
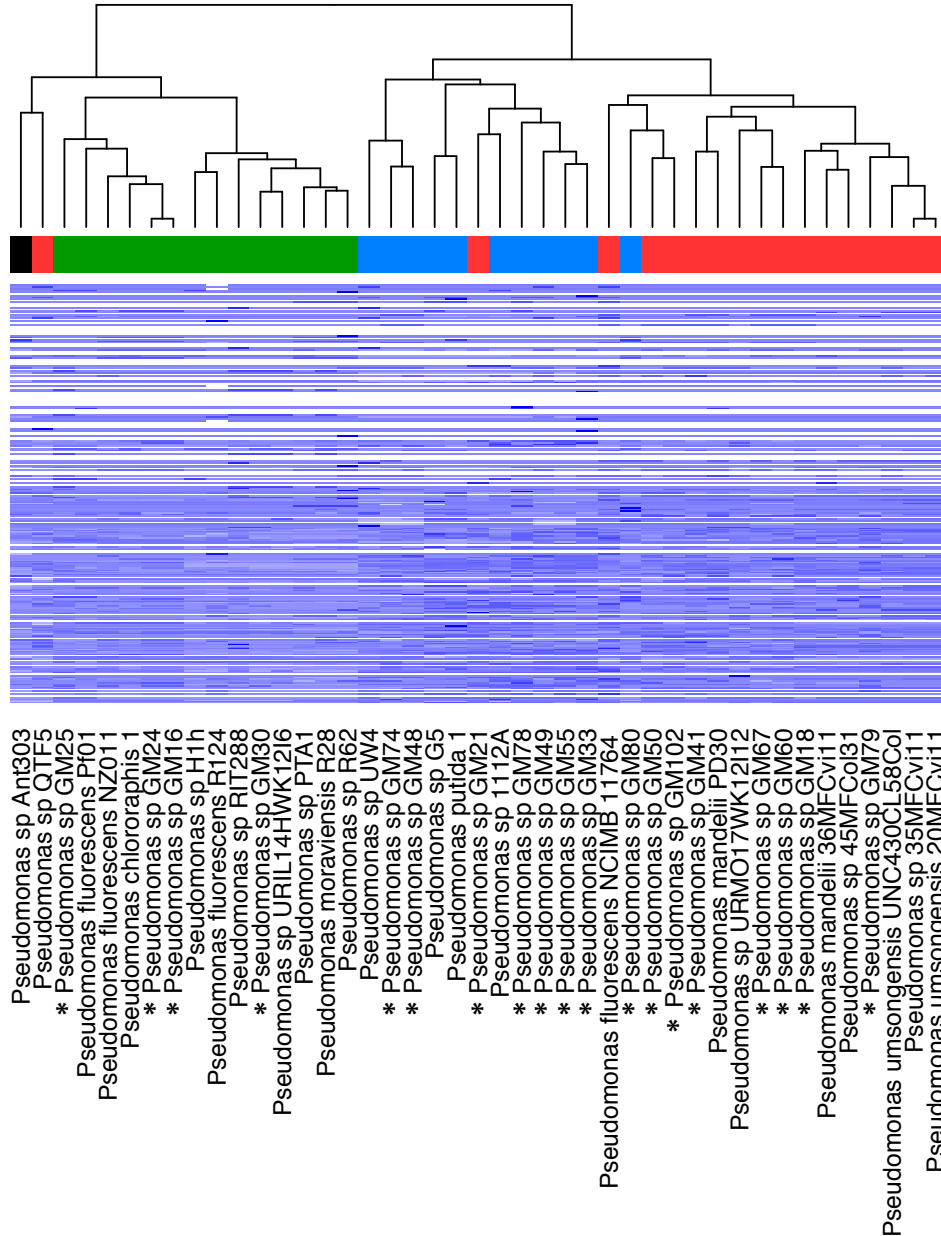
67

68

69

70

71

Figure S5. Pathway profiles analysis of PMI *Pseudomonas* isolates. The asterisk indicates PMI *Pseudomonas* isolates. The color code at the X-axis corresponds to three distinct subgroups identified in Figure 2.

79   **Table S1.** *Pseudomonas* dataset.

80

81   **Table S2.** Genomic clusters with at least five members.

Genomic cluster Index	Cluster size	Species by NCBI Taxonomy
Cluster_53	684	*Pseudomonas aeruginosa* (680) *Pseudomonas otitidis* (1) *Pseudomonas sp* (3)
Cluster_25	64	*Pseudomonas syringae* (61) *Pseudomonas avellanae* (3)
Cluster_23	33	*Pseudomonas syringae* (33)
Cluster_33	16	*Pseudomonas putida* (15) *Pseudomonas monteilii* (1)
Cluster_24	16	*Pseudomonas syringae* (13) *Pseudomonas savastanoi* (3)
Cluster_32	10	*Pseudomonas monteilii* (3) *Pseudomonas putida* (1) *Pseudomonas sp* (6)
Cluster_10	9	*Pseudomonas fluorescens* (4) *Pseudomonas brassicacearum* (3) *Pseudomonas sp* (2)
Cluster_13	9	*Pseudomonas chlororaphis* (8) *Pseudomonas sp* (1)
Cluster_16	8	*Pseudomonas fluorescens* (4) *Pseudomonas sp* (4)
Cluster_14	8	*Pseudomonas fluorescens* (3) *Pseudomonas simiae* (2) *Pseudomonas sp* (3)
Cluster_34	8	*Pseudomonas putida* (3) *Pseudomonas monteilii* (2) *Pseudomonas plecoglossicida* (1) *Pseudomonas taiwanensis* (1) *Pseudomonas sp* (1)
Cluster_31	7	*Pseudomonas putida* (6) *Pseudomonas sp* (1)
Cluster_54	6	*Pseudomonas psychrotolerans* (1) *Pseudomonas oryzihabitans* (1) *Pseudomonas oleovorans* (1) *Pseudomonas sp* (3)
Cluster_2	6	*Pseudomonas umsongensis* (2) *Pseudomonas mandelii* (1)

		Pseudomonas sp (3)
Cluster_7	5	*Pseudomonas fluorescens* (1)
		Pseudomonas sp (4)
Cluster_4	5	*Pseudomonas sp* (5)

82

83 **Table S3. Membership of PMI Pseudomonas isolates into genomic clusters.**

PMI *Pseudomonas*	Genomic cluster index	Cluster size	*Pseudomonas* species by NCBI taxonomy
GM60	Cluster_0	2	*Pseudomonas sp (*2)
GM67	Cluster_0	2	*Pseudomonas sp* (2)
GM41	Cluster_1	2	*Pseudomonas mandelii* (1) *Pseudomonas sp* (1)
GM18	Cluster_2	6	*Pseudomonas umsongensis* (2) *Pseudomonas mandelii* (1) *Pseudomonas sp* (3)
GM50	Cluster_3	4	*Pseudomonas sp* (4)
GM79	Cluster_3	4	*Pseudomonas sp* (4)
GM102	Cluster_3	4	*Pseudomonas sp* (4)
GM33	Cluster_4	3	*Pseudomonas sp* (3)
GM48	Cluster_4	5	*Pseudomonas sp* (5)
GM49	Cluster_4	5	*Pseudomonas sp* (5)
GM74	Cluster_4	5	*Pseudomonas sp* (5)
GM78	Cluster_5	3	*Pseudomonas putida* (1) *Pseudomonas sp* (2)
GM16	Cluster_6	4	*Pseudomonas chlororaphis* (1) *Pseudomonas fluorescens* (1) *Pseudomonas sp* (2)
GM24	Cluster_6	4	*Pseudomonas chlororaphis* (1) *Pseudomonas fluorescens* (1) *Pseudomonas sp* (2)
GM30	Cluster_7	5	*Pseudomonas sp* (4) *Pseudomonas fluorescens* (1)
GM25	Cluster_8	2	*Pseudomonas fluorescens* (1) *Pseudomonas sp* (1)
GM17	Cluster_13	9	*Pseudomonas chlororaphis* (8) *Pseudomonas sp* (1)
GM55	Singleton	1	*Pseudomonas sp* (1)
GM84	Singleton	1	*Pseudomonas sp* (1)
GM80	Singleton	1	*Pseudomonas sp* (1)
GM21	Singleton	1	*Pseudomonas sp* (1)

84

85   **Table S4.** Subgroup-specific genes with annotation.

86

87   **REFERENCES**

88

89   1.   **Lagesen K, Hallin P, Rodland EA, Staerfeldt H-H, Rognes T, Ussery DW.** 2007.

90   RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic acids

91   research **35:**3100-3108.

92   2.   **Katoh K, Standley DM.** 2013. MAFFT multiple sequence alignment software version 7:

93   improvements in performance and usability. Molecular biology and evolution **30:**772-

94   780.

95   3.   **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2--approximately maximum-likelihood

96   trees for large alignments. PloS one **5:**e9490.

97