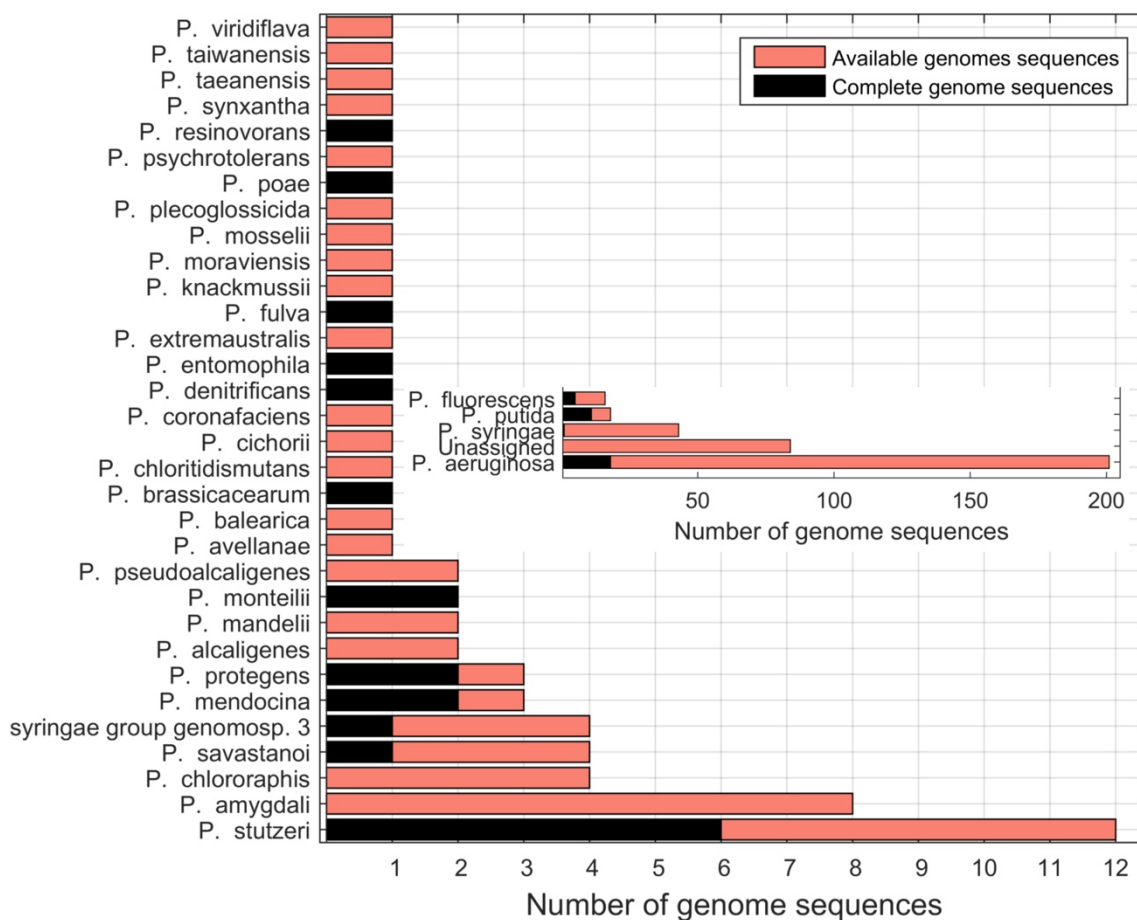


Comparison of 432 Pseudomonas strains through integration of genomic, functional, metabolic and expression data

Jasper J Koehorst^{1,+}, Jesse CJ van Dam^{1,+}, Ruben GA van Heck¹, Edoardo Saccenti¹, Vitor AP Martins dos Santos^{1,2}, Maria Suarez-Diez^{1,=*}, and Peter J Schaap^{1,=}

¹Wageningen University, Laboratory of Systems and Synthetic Biology, Wageningen, 6708 WE, The Netherlands ²LifeGlimmer GmbH, Berlin, Germany * maria.suarezdiez@wur.nl

⁺these authors contributed equally to this work ⁼these authors contributed equally to this work



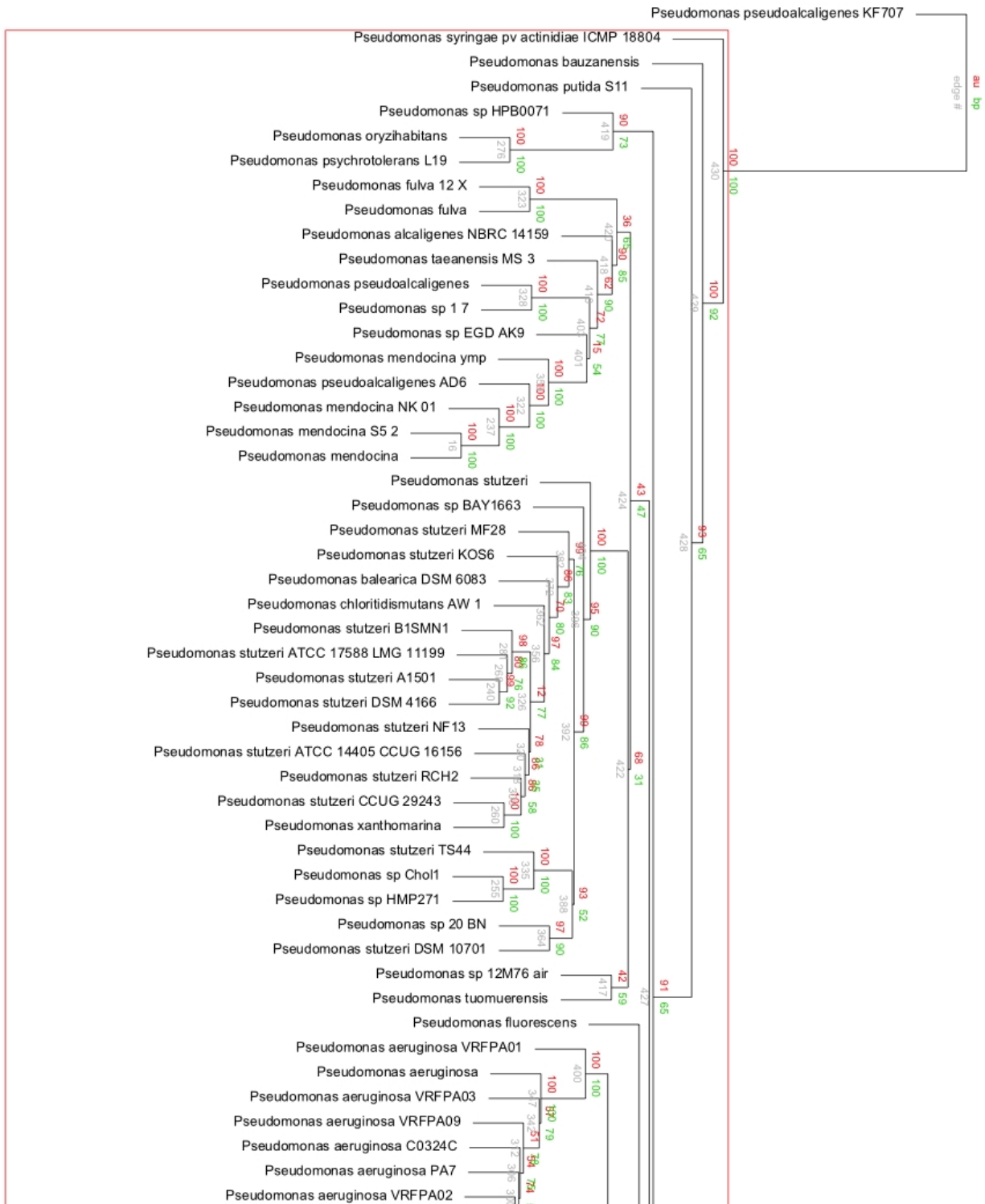
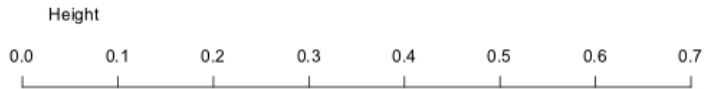
Supplementary Figure 1:
Overview of available *Pseudomonas* genome sequences analyzed in this study arranged per species.

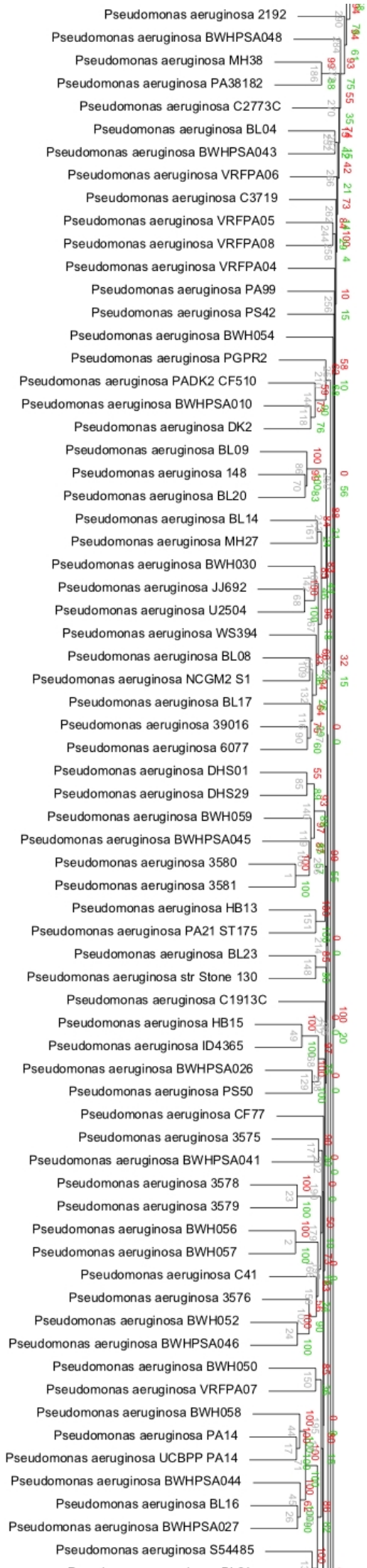
Supplementary Table 2:

Comparison between the de-novo annotation and the original annotation. *Date* refers to the deposition date of the genome sequence in the NCBI database as retrieved from <https://www.ncbi.nlm.nih.gov/>. #Genes, #Domains, #GO and #ECs indicate the difference between the total number of genes, protein domains, GO terms and EC numbers between the *de-novo* and the original annotation, respectively.

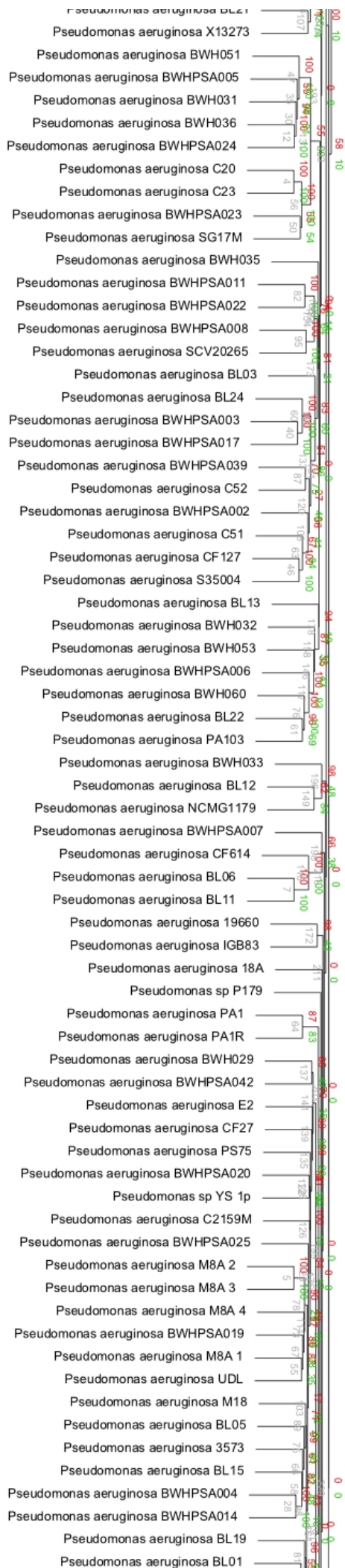
Organism	Date	#Genes	#Domains	#Go	#ECs
<i>Pseudomonas_aeruginosa_PA01</i>	16/05/00	109	4	17	2
<i>Pseudomonas_putida_KT2440</i>	08/04/02	244	14	94	4
<i>Pseudomonas_syringae_pv__tomato_str__D C3000</i>	25/03/03	245	2	38	1
<i>Pseudomonas_putida_ND6</i>	20/05/03	-682	-2	31	1
<i>Pseudomonas_syringae_pv__phaseolicola_1 448A</i>	04/03/04	407	4	50	2
<i>Pseudomonas_syringae_pv__syringae_B728 a</i>	04/05/05	118	-1	-2	1
<i>Pseudomonas_protegens_Pf_5</i>	13/05/05	168	24	1	0
<i>Pseudomonas_fluorescens_Pf0_1</i>	08/08/05	17	1	-25	0
<i>Pseudomonas_fluorescens_SBW25</i>	09/01/06	109	2	20	0
<i>Pseudomonas_entomophila_L48</i>	27/01/06	-61	6	-7	2
<i>Pseudomonas_stutzeri_A1501</i>	03/04/06	49	5	41	-1
<i>Pseudomonas_aeruginosa_UCBPP_PA14</i>	18/08/06	16	-3	0	0
<i>Pseudomonas_aeruginosa_PA7</i>	06/09/06	-255	-1	0	0
<i>Pseudomonas_mendocina_ymp</i>	17/04/07	39	-1	11	0
<i>Pseudomonas_putida_F1</i>	25/05/07	45	-1	-7	1
<i>Pseudomonas_putida_GB_1</i>	24/01/08	37	1	-6	0
<i>Pseudomonas_putida_W619</i>	21/02/08	55	4	-7	0
<i>Pseudomonas_aeruginosa_LESB58</i>	02/09/08	134	4	34	0
<i>Pseudomonas_putida_BIRD_1</i>	18/10/10	191	5	46	0
<i>Pseudomonas_mendocina_NK_01</i>	01/01/11	50	2	28	2
<i>Pseudomonas_aeruginosa_M18</i>	18/01/11	87	4	-6	2
<i>Pseudomonas_brassicacearum_subsp...NFM 421</i>	11/03/11	-39	3	-3	0
<i>Pseudomonas_stutzeri_DSM_4166</i>	29/03/11	33	3	28	2
<i>Pseudomonas_fulva_12_X</i>	29/04/11	2	-13	0	0
<i>Pseudomonas_stutzeri_ATCC_17588__LMG_11199</i>	22/06/11	-20	2	28	0
<i>Pseudomonas_putida_S16</i>	28/06/11	249	0	41	0
<i>Pseudomonas_fluorescens_A506</i>	09/09/11	102	2	11	0
<i>Pseudomonas_stutzeri_RCH2</i>	05/10/11	39	0	6	0
<i>Pseudomonas_aeruginosa_DK2</i>	22/11/11	-12	2	-2	2

Pseudomonas fluorescens_F113	22/11/11	183	4	12	0
Pseudomonas protegens_CHA0	08/12/11	26	7	37	1
Pseudomonas fluorescens_R124	05/03/12	0	0	0	0
Pseudomonas aeruginosa_NCGM2_S1	10/04/12	-42	3	50	1
Pseudomonas stutzeri_CCUG_29243	11/06/12	56	0	-9	0
Pseudomonas putida_DOT_T1E	04/07/12	29	47	0	0
Pseudomonas stutzeri_DSM_10701	10/07/12	23	0	13	0
Pseudomonas putida_HB3267	23/07/12	39	0	-5	1
Pseudomonas sp_UW4	17/09/12	70	-5	0	0
Pseudomonas sp_VLB120	26/11/12	53	4	1	0
Pseudomonas aeruginosa_PA1	25/12/12	182	5	106	0
Pseudomonas aeruginosa_PA1R	25/12/12	179	5	107	0
Pseudomonas poae_RE_1_1_14	11/01/13	2	0	-22	0
Pseudomonas denitrificans_ATCC_13867	15/02/13	5	3	-2	2
Pseudomonas aeruginosa_B136_33	09/04/13	-10	-2	-19	0
Pseudomonas putida_H8234	06/05/13	33	5	31	0
Pseudomonas resinovorans_NBRC_106553	04/06/13	-105	1	-22	0
Pseudomonas putida_NBRC_14164	14/06/13	-14	0	-1	0
Pseudomonas aeruginosa_RP73	17/06/13	42	1	16	1
Pseudomonas chloritidismutans_AW_1	22/07/13	0	4	31	1
Pseudomonas aeruginosa_PAO581	14/08/13	36	2	2	1
Pseudomonas aeruginosa_c7447m	05/09/13	32	-2	3	0
Pseudomonas aeruginosa_PAO1_VE13	22/10/13	11	1	4	0
Pseudomonas sp_TKP	14/11/13	-50	0	-23	0
Pseudomonas aeruginosa_MTB_1	27/11/13	-1	-21	1	0
Pseudomonas aeruginosa_SCV20265	02/12/13	-109	3	39	0
Pseudomonas aeruginosa_LES431	03/12/13	42	4	5	0
Pseudomonas monteilli_SB3078	11/12/13	23	3	12	0
Pseudomonas monteilli_SB3101	11/12/13	24	2	11	1





76.41



416

100 100

425

99 54

100 10

58 10

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

100 100

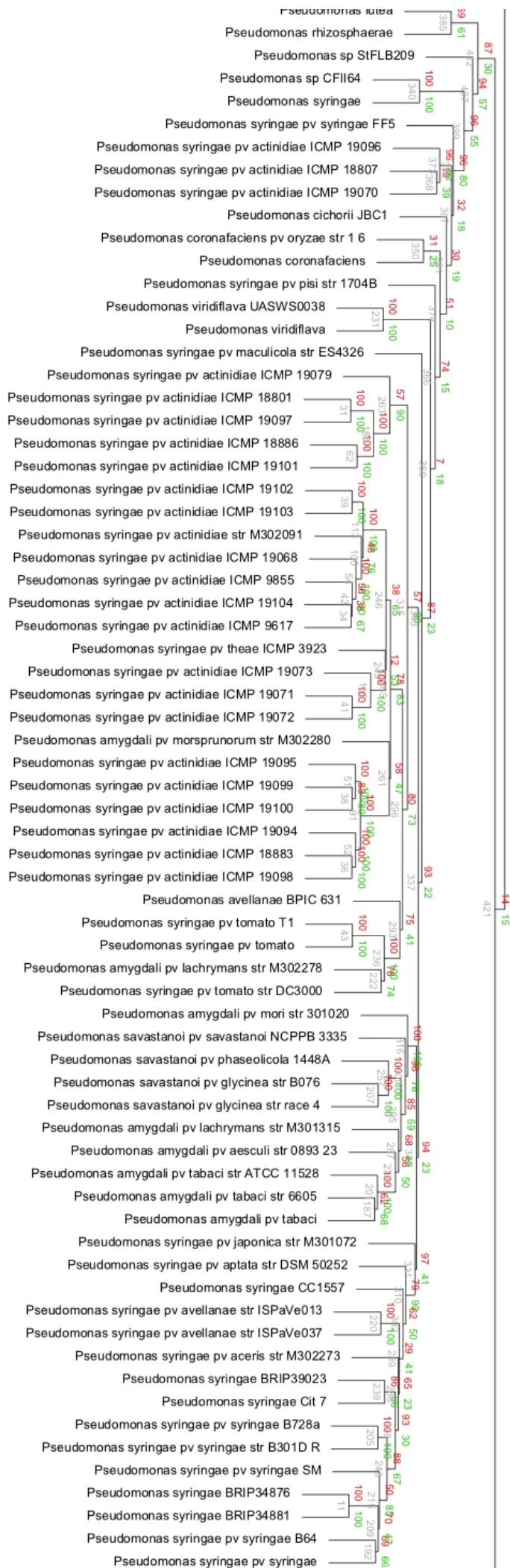
100 100

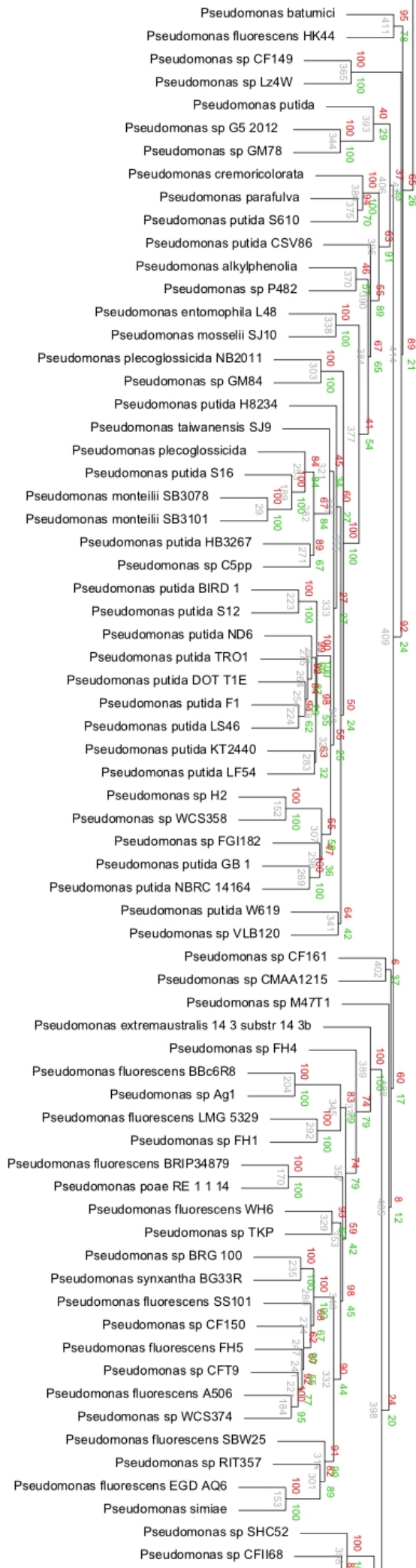
100 100

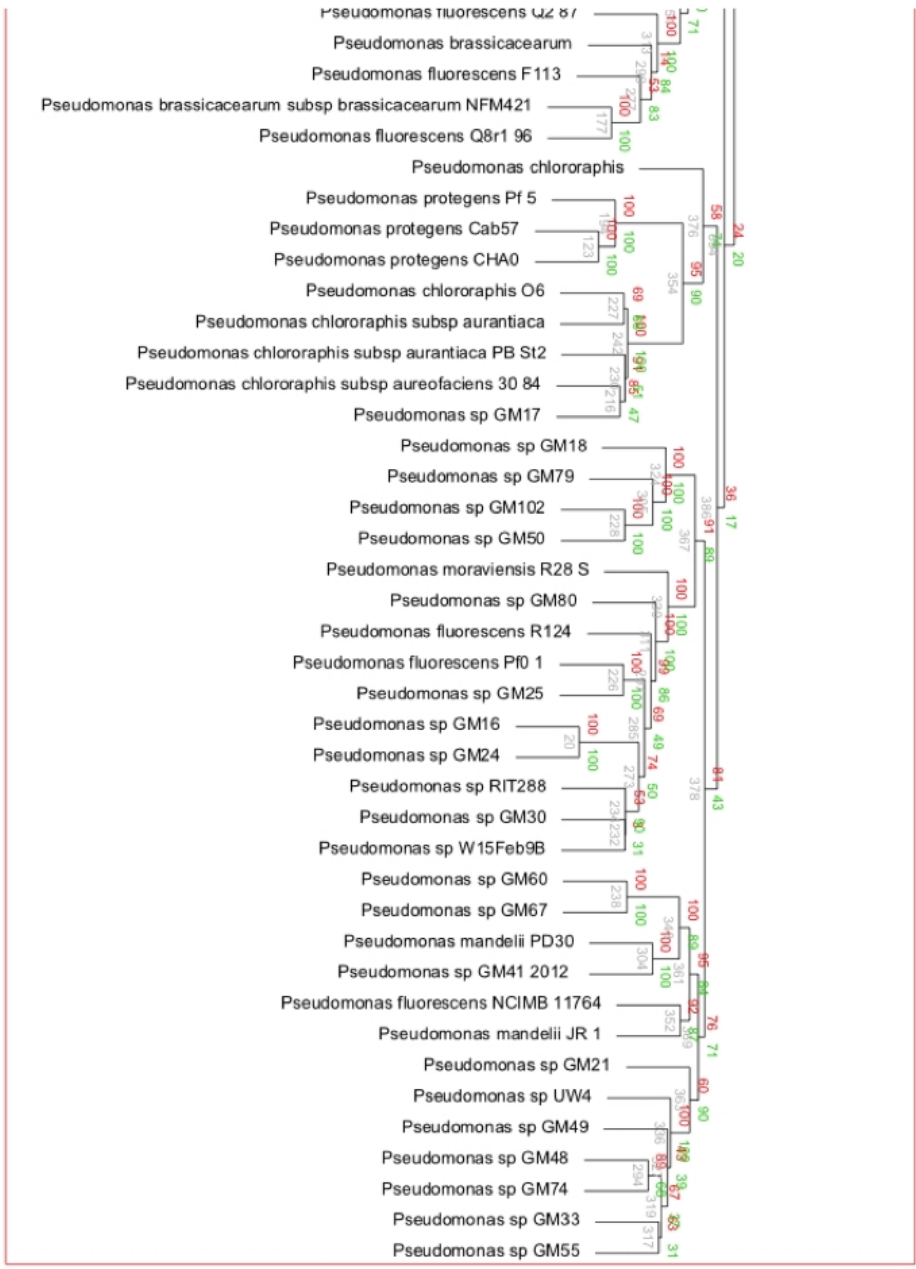
Cluster dendrogram with AU/BJP values (%)



Distance: binary
 Cluster method: average



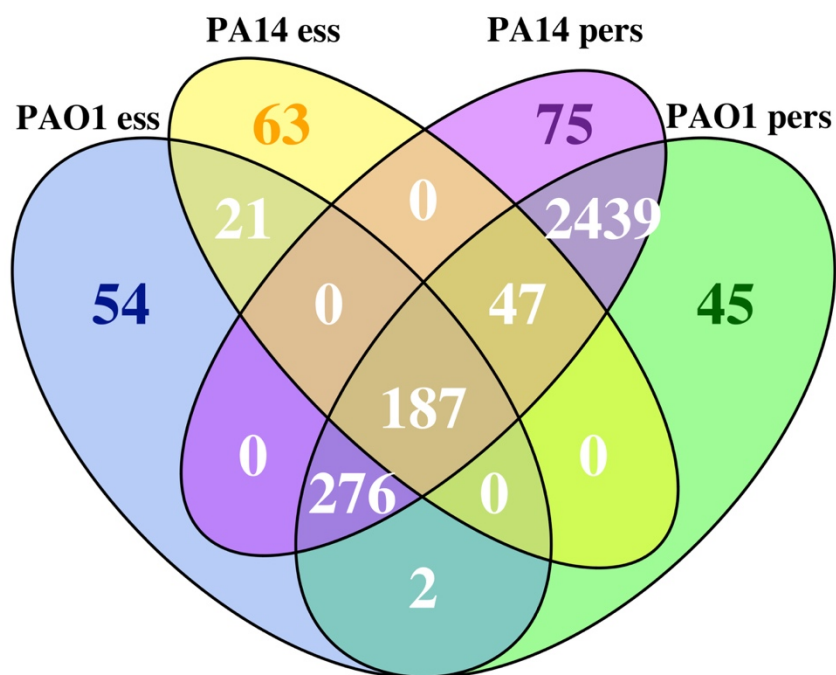




Supplementary Figure 3:

Pan-genome tree for 432 *Pseudomonas* genomes based on protein domains.

A binarized matrix indicating presence-absence of protein domains for each considered genome was obtained. The hierarchical tree was constructed using as input a Manhattan distance matrix derived from this matrix



Supplementary Figure 4:

A comparison of gene persistence and essentiality for *P. aeruginosa* PAO1 and *P. aeruginosa* PA14. PA14 ess: refers to genes found to be essential for growth of *P. aeruginosa* PA14 on LB (see Liberati *et al* 2006 PNAS); **PAO1 ess:** genes found to be essential for growth of *P. aeruginosa* on either LB, minimal medium with pyruvate addition or sputum agar (see Lee *et al* 2015 PNAS). **PA14 pers** and **PAO1 pers** indicate genes that are present *P. aeruginosa* PA14 and PAO1 respectively and have a high (>0.95) persistence in the *Pseudomonas* genus.

Variability of gene expression profiles and its association to essentiality and persistence for *P. aeruginosa* PA14 using RNAseq data

An RNAseq dataset corresponding to *P. aeruginosa* PA14 grown under 14 different environmental conditions (different growth temperatures, stages in the growth curve, osmolarity, phosphate and iron concentrations, oxygen availability, surface attachment or within biofilms) was used to assess gene expression variability in this strain (Dötsch *et al.* 2015). On the basis of differential expression analysis each gene was assigned a score, *FCS*, for transcriptional variation. Persistent genes tend to show significantly lower degree of variation in expression level than those with lower persistence (p -value <0.01); this holds true also for essential genes (Fig. S1)

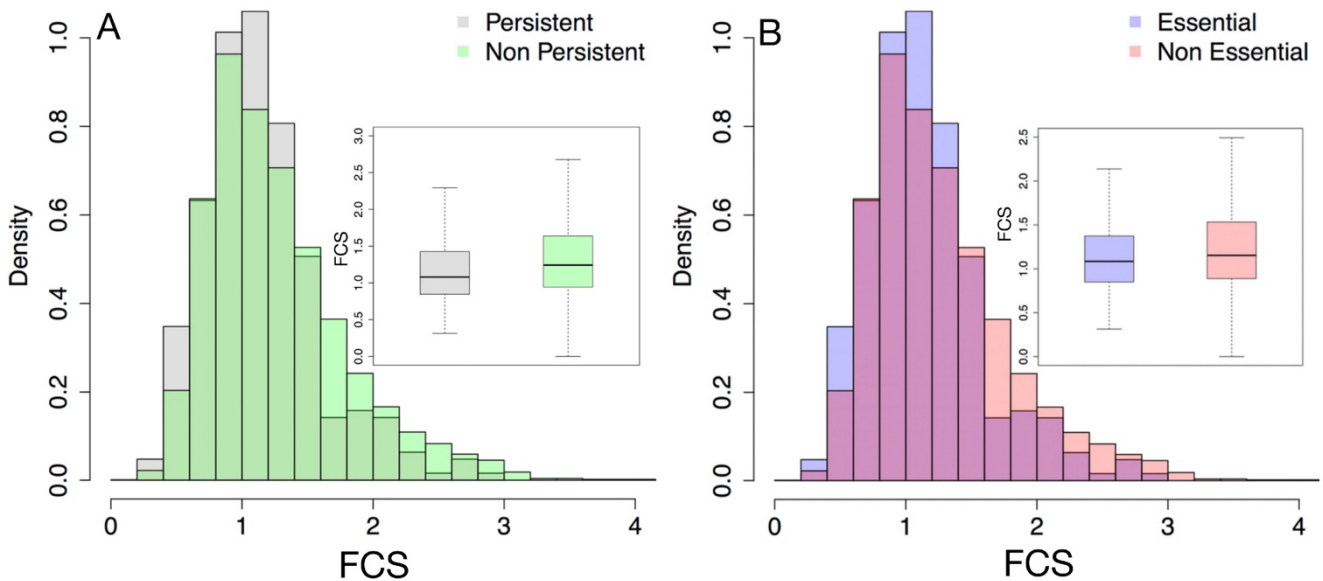


Figure S1: Variability of gene expression levels and its association with persistence and essentiality for *P. aeruginosa* PA14 using RNASeq A) Distribution of FCS index for (non) persistent genes (genes with persistence lower or higher than 0.95, respectively). The box blot gives the average value of FCS for the two groups. The difference between the mean values is significant (p -val <0.01). B) Distribution of FCS index for essential and non-essential genes with gene essentiality derived experimentally (Liberati *et al.* 2006). The box blot gives the average value of FCS for the two groups. The difference between the mean values is significant (p -val <0.01)

FCS Computation

Number of RNA-seq reads mapped to the genome of *P. aeruginosa* PA14 grown under 14 different conditions were retrieved from the GEO database (accession number: GSE55197 (Dötsch *et al.* 2015)). The data corresponds to the environmental growth conditions within biofilms, at various temperatures, osmolarities and phosphate concentrations, under anaerobic conditions, attached to a surface and conditions encountered within the eukaryotic host. For every gene we calculated the \log_2 fold change of its expression values when considering all possible pairwise comparison of conditions. These values were aggregated in an overall fold change score, *FCS*, by averaging their absolute values. Missing values for fold change in individual comparisons arising from genes with very low counts in some conditions were imputed to 0 or ± 4 , according to the significance of the differential expression ($\text{fdr}<0.05$ selected as a threshold).

Supplementary Data 6: Models and the code used to analyze the six metabolic models.