

## ***Supplementary Material***

### Tables

**Table S1. Well-characterized reference strains and field isolates belonging to established *Pandoraea* species.** CF, cystic fibrosis. LMG, BCCM/LMG Bacteria Collection, Laboratory of Microbiology, Ghent University, Ghent, Belgium; CCM, Czech Collection of Microorganisms, Masaryk University, Brno, Czech Republic; CCUG, Culture Collection University of Gothenburg, Department of Clinical Bacteriology, Sahlgrenska University Hospital, Gothenburg, Sweden; CDC, Centers for Disease Control, United States Public Health Service, Atlanta, USA; KACC, Korean Agricultural Culture Collection, National Institute of Agricultural Science and Technology, Suwon, Republic of Korea.

Strain	Other strains designations	Source	Depositor	Reference
<b><i>Pandoraea apista</i></b>				
LMG 16407 <sup>T</sup>	5893 <sup>T</sup> , CCUG 38412 <sup>T</sup>	CF sputum (Denmark)	N. Hoiby	(Coenye et al., 2000)
LMG 18089	CEP0633	CF sputum (United States)	D. P. Speert	(Coenye et al., 2000)
LMG 18818	AU0193	CF sputum (United States, 1997)	Own isolate	(Coenye et al., 2000)
R-10112	AU1465	CF sputum (United States, 1999)	Own isolate	
R-14454	AU2513	CF sputum (United States, 2000)	Own isolate	
R-15345	AU2160	CF sputum (United States, 2000)	Own isolate	
R-31796	54459A101	CF sputum (Denmark)	N. Hoiby	
<b><i>Pandoraea faecigallinarum</i></b>				
LMG 28171 <sup>T</sup>	CCM 2766 <sup>T</sup>	Chicken, dung	CCM	(Sahin et al., 2011)
<b><i>Pandoraea fibrosis</i></b>				
LMG 29626 <sup>T</sup>	6399 <sup>T</sup>	CF sputum (Australia, 2012)	R. Ee	(See-Too et al., 2019)
LMG 31113	AU17150	CF patient (United States, 2008)	Own isolate	
R-11943	CUL 264	Water (Lebanon, 1996)	J.-M. Meyer	
<b><i>Pandoraea norimbergensis</i></b>				
LMG 18379 <sup>T</sup>	CCUG 39188 <sup>T</sup>	Water (Germany)	CCUG	(Coenye et al., 2000)
LMG 13019	90/072	Human blood (Belgium, 1990)	S. Lauwers	(Coenye et al., 2000)
LMG 16603	CCUG 34867	Human bronchial alveolar lavage (Sweden, 1995)	CCUG	(Coenye et al., 2000)

*Pandoraea oxalativorans*

LMG 28169 <sup>T</sup>	CCM7677 <sup>T</sup>	Rhizosphere soil (Turkey)	CCM	(Sahin et al., 2011)
------------------------	----------------------	---------------------------	-----	----------------------

*Pandoraea pnomenusa*

LMG 18087 <sup>T</sup>	C1513 <sup>T</sup> , CCUG 38742 <sup>T</sup>	CF sputum (United Kingdom)	J. R. W. Govan	(Coenye et al., 2000)
LMG 18817	PC13	CF patient (United States, 1986)	Own isolate	(Coenye et al., 2000)
LMG 18820	PHA 1119	Sludge (Germany)	J. Mergaert	(Coenye et al., 2000)
LMG 24771	AI-S130	Plant (Azadirachta indica A. Juss) rhizoplane (India)	M. Madhaiyan	
LMG 31119	AU1039	CF patient (United States, 1999)	Own isolate	
R-1454	C7351	CF sputum (Canada)	D. P. Speert	(Coenye et al., 2000)
R-4805	HI2344	CF patient (United States, 1998)	Own isolate	(Coenye et al., 2000)
R-7719	K160189KP	CF patient (Belgium, 1999)	H. Franckx	
R-7902	K240586BA	CF patient (Belgium, 1999)	H. Franckx	
R-15667	CDC G8107	Human Blood (United States, 1993)	CDC	(Daneshvar et al., 2001)

*Pandoraea pulmonicola*

LMG 18106 <sup>T</sup>	FC330 <sup>T</sup> , CCUG 38759 <sup>T</sup>	CF patient (Canada)	D. P. Speert	(Coenye et al., 2000)
LMG 18107	L. Saimann41	CF patient (United States)	L. Saimann	(Coenye et al., 2000)
LMG 18108	L. Saimann42	CF patient (United States)	L. Saimann	(Coenye et al., 2000)

*Pandoraea sputorum*

LMG 18819 <sup>T</sup>	AU0124 <sup>T</sup> , CCUG 39682 <sup>T</sup>	CF patient (United States, 1997)	Own isolate	(Coenye et al., 2000)
LMG 18100	C4946	CF patient (Canada)	D. P. Speert	(Coenye et al., 2000)
LMG 18109	L. Saimann57, HI2111	CF patient (United States)	L. Saimann	
LMG 20601	CDC G5084	Human sinus (United States, 1990)	CDC	(Daneshvar et al., 2001)
LMG 31120	AU1359	CF patient (United States, 1999)	Own isolate	
LMG 31121	V02 34922	CF patient (Switzerland, 2007)	R. Zbinden	
R-2456	PC212	Source unknown (United States, 1987)	Own isolate	(Coenye et al., 2000)
R-11656	AU1831	CF patient (United States, 2000)	Own isolate	
R-13191	AU2075	CF patient (United States, 2000)	Own isolate	
R-13758	AU2302	CF patient (United States, 2000)	Own isolate	

R-13761	AU2321	CF patient (United States, 2000)	Own isolate	
R-14315	AU2389	CF patient (United States, 2000)	Own isolate	
<b><i>Pandoraea terraе</i></b>				
LMG 30175 <sup>T</sup>	SE-S21 <sup>T</sup> , JCM 30137 <sup>T</sup>	Forest soil (South Korea, 2014)	JCM	(Jeong et al., 2016)
<b><i>Pandoraea thiooxydans</i></b>				
LMG 24779 <sup>T</sup>	KACC 12757 <sup>T</sup>	Rhizosphere soil (Sesamum indicum L.) (South Korea)	KACC	(Anandham et al., 2010)
<b><i>Pandoraea vervacti</i></b>				
LMG 28170 <sup>T</sup>	CCM 7667 <sup>T</sup> , NS15 <sup>T</sup>	Soil (Turkey)	CCM	(Sahin et al., 2011)

**Table S2. Genomes sequenced in the present study.**

Strain	Project	Contigs	Size (bp)	N50 (bp)	Coverage (x)	%GC	CDS
<i>P. apista</i> LMG 18089	PRJEB30685	27	5,815,466	457,211	134	62.7	5,279
<i>P. fibrosis</i> LMG 31113	PRJEB30745	43	5,605,513	483,367	185	62.8	4,943
<i>P. pnomenusa</i> LMG 31119	PRJEB30696	46	5,305,298	262,716	119	64.9	4,699
<i>P. sputorum</i> LMG 20601	PRJEB30706	62	6,264,179	471,825	155	62.7	5,539
<i>P. sputorum</i> LMG 31120	PRJEB30707	17	5,956,418	751,203	114	62.7	5,301
<i>P. sputorum</i> LMG 31121	PRJEB30708	113	6,453,978	322,419	117	62.8	5,652
<i>P. terrae</i> LMG 30175 <sup>T</sup>	PRJEB30813	81	6,176,823	194,136	111	62.8	5,575
<i>P. anapnoica</i> sp. nov. LMG 31117 <sup>T</sup>	PRJEB30755	48	6,126,688	278,466	141	62.4	5,364
<i>P. anhela</i> sp. nov. LMG 31108 <sup>T</sup>	PRJEB30724	61	6,046,012	256,277	138	63.4	5,188
<i>P. aquatica</i> sp. nov. LMG 31011 <sup>T</sup>	PRJEB30756	17	5,958,127	792,622	123	62.9	5,238
<i>P. bronchicola</i> sp. nov. LMG 20603 <sup>T</sup>	PRJEB30725	34	5,351,123	323,330	158	63.0	4,753
<i>P. capi</i> sp. nov. LMG 20602 <sup>T</sup>	PRJEB30721	31	5,852,144	401,082	115	63.4	5,056
<i>P. captiosa</i> sp. nov. LMG 31118 <sup>T</sup>	PRJEB30757	36	6,139,582	614,074	134	63.3	5,340
<i>P. cepalis</i> sp. nov. LMG 31106 <sup>T</sup>	PRJEB30715	56	5,274,229	187,061	208	63.7	4,730
<i>P. cepalis</i> sp. nov. LMG 31107	PRJEB30716	32	5,159,566	300,362	131	63.5	4,626
<i>P. commovens</i> sp. nov. LMG 31010 <sup>T</sup>	PRJEB30753	26	6,036,949	540,061	122	62.6	5,308
<i>P. communis</i> sp. nov. LMG 31110 <sup>T</sup>	PRJEB30740	17	5,708,603	837,780	139	62.6	5,067
<i>P. communis</i> sp. nov. LMG 31111	PRJEB30741	55	5,566,071	278,275	150	62.5	5,064
<i>P. eparura</i> sp. nov. LMG 31012 <sup>T</sup>	PRJEB30718	35	5,205,577	435,937	135	63.7	4,621
<i>P. horticolens</i> sp. nov. LMG 31112 <sup>T</sup>	PRJEB30744	68	6,008,490	290,798	122	62.3	5,378
<i>P. iniqua</i> sp. nov. LMG 31009 <sup>T</sup>	PRJEB30748	17	6,339,129	1,113,537	117	63.1	5,521
<i>P. iniqua</i> sp. nov. LMG 31115	PRJEB30749	14	6,296,634	1,159,992	143	63.1	5,445
<i>P. morbifera</i> sp. nov. LMG 31116 <sup>T</sup>	PRJEB30750	47	5,233,298	316,192	161	64.7	4,676
<i>P. nosoerga</i> sp. nov. LMG 31109 <sup>T</sup>	PRJEB30729	41	4,862,114	229,370	205	66.1	4,266
<i>P. pneumonica</i> sp. nov. LMG 31114 <sup>T</sup>	PRJEB30747	12	5,845,078	1,392,766	181	62.5	5,202
<i>P. soli</i> sp. nov. LMG 31014 <sup>T</sup>	PRJEB30720	51	4,961,982	370,563	145	63.6	4,395



**Table S7. The frequency of orthologous versus non-orthologous CDS varies among species.**

Pearson chi-square analysis testing the independence of gene conservation (orthologous vs. non-orthologous CDS) and species ( $\chi^2(29) = 5863$ ,  $p < 0.001$ ). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or - if the standardized residual is  $> 1.96$  or  $< -1.96$ , respectively, and significant at  $p < 0.05$ .

	Orthologous CDS		Non-orthologous CDS
<i>Ca. P. novymondanis</i>	920 (-1.058)	+	48 (8.344)
<i>P. anapnoica</i>	5307 (0.385)	-	57 (-3.032)
<i>P. anhela</i>	5012 (-1.313)	+	176 (10.352)
<i>P. apista</i>	+ 97304 (4.003)	-	305 (-31.560)
<i>P. aquatica</i>	5118 (-0.516)	+	120 (4.067)
<i>P. bronchicola</i>	4649 (-0.420)	+	104 (3.311)
<i>P. capi</i>	10134 (-1.117)	+	278 (8.808)
<i>P. captiosa</i>	5221 (-0.475)	+	119 (3.744)
<i>P. cepalis</i>	13600 (-0.901)	+	326 (7.101)
<i>P. commovens</i>	5226 (0.029)		82 (-0.225)
<i>P. communis</i>	15005 (-0.261)	+	274 (2.059)
<i>P. eparura</i>	4510 (-0.561)	+	111 (4.421)
<i>P. faecigallinarum</i>	4719 (-0.898)	+	139 (7.076)
<i>P. fibrosis</i>	13584 (1.193)	-	78 (-9.407)
<i>P. horticolens</i>	5168 (-1.716)	+	210 (13.526)
<i>P. iniqua</i>	10821 (0.276)	-	145 (-2.175)
<i>P. morbifera</i>	4582 (-0.294)	+	94 (2.318)
<i>P. norimbergensis</i>	5177 (0.320)	-	60 (-2.519)
<i>P. nosoerga</i>	4188 (-0.161)		78 (1.270)
<i>P. oxalativorans</i>	5237 (-1.063)	+	163 (8.378)
<i>P. pneumonica</i>	5096 (-0.330)	+	106 (2.602)
<i>P. pnomenusa</i>	32135 (0.920)	-	350 (-7.251)
<i>P. pulmonicola</i>	4719 (-1.039)	+	149 (8.189)
<i>P. soli</i>	4346 (0.313)	-	49 (-2.470)
<i>P. sputorum</i>	21040 (0.017)		336 (-0.138)
<i>P. terrae</i>	- 4943 (-7.340)	+	632 (57.864)
<i>P. terrigena</i>	4758 (-0.617)	+	120 (4.863)
<i>P. thiooxydans</i>	- 8010 (-2.688)	+	377 (21.187)
<i>P. vervacti</i>	4758 (0.337)	-	53 (-2.657)
<i>Pandoraea sp.</i>	10592 (0.628)	-	105 (-4.949)

**Table S8. Orthogroup specificity varies among COG categories.**

Pearson chi-square analysis testing the independence of orthogroup specificity and COG category ( $\chi^2(66) = 522$ ,  $p < 0.001$ ). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or - if the standardized residual is  $>1.96$  or  $<-1.96$ , respectively, and significant at  $p < 0.05$ .

		<b>Core</b>	<b>Multiple species</b>	<b>Single species</b>	<b>Single isolate</b>
<b>Information storage and processing</b>					
J	Translation, ribosomal structure and biogenesis	+ 159 (11.999)	- 114 (-6.199)	12 (-1.625)	0 (-0.573)
K	Transcription	- 97 (-5.399)	+ 598 (2.897)	54 (0.535)	0 (-0.929)
L	Replication, recombination and repair	+ 63 (2.042)	144 (-0.909)	12 (-0.699)	0 (-0.503)
B	Chromatin structure and dynamics	1 (0.116)	2 (-0.497)	1 (1.413)	0 (-0.068)
<b>Cellular processes and signaling</b>					
D	Cell cycle control, cell division, chromosome partitioning	+ 26 (3.032)	39 (-1.046)	- 0 (-2.087)	0 (-0.274)
V	Defense mechanisms	- 22 (-2.371)	+ 137 (1.990)	- 4 (-2.095)	0 (-0.434)
T	Signal transduction mechanisms	67 (-0.346)	226 (0.221)	21 (-0.010)	0 (-0.602)
M	Cell wall/membrane/envelope biogenesis	105 (0.214)	322 (-0.313)	34 (0.545)	1 (0.640)
N	Cell motility	43 (-0.882)	166 (0.739)	12 (-0.731)	0 (-0.505)
W	Extracellular structures	0 (-1.334)	8 (0.977)	0 (-0.732)	0 (-0.096)
U	Intracellular trafficking, secretion, and vesicular transport	26 (-0.738)	107 (1.150)	- 2 (-2.343)	0 (-0.395)
O	Posttranslational modification, protein turnover, chaperones	+ 85 (4.521)	- 130 (-2.841)	20 (1.070)	0 (-0.521)
<b>Metabolism</b>					
C	Energy production and conversion	+ 125 (2.065)	301 (-1.660)	41 (1.733)	0 (-0.734)
G	Carbohydrate transport and metabolism	- 63 (-3.095)	318 (1.294)	36 (1.522)	0 (-0.693)
E	Amino acid transport and metabolism	168 (0.594)	498 (-0.591)	54 (0.816)	1 (0.185)
F	Nucleotide transport and metabolism	+ 55 (5.409)	- 54 (-3.435)	12 (1.366)	0 (-0.374)
H	Coenzyme transport and metabolism	+ 96 (3.950)	181 (-1.674)	12 (-1.675)	0 (-0.577)
I	Lipid transport and metabolism	99 (-0.005)	322 (0.360)	24 (-1.067)	0 (-0.716)
P	Inorganic ion transport and metabolism	- 67 (-2.579)	315 (1.333)	30 (0.453)	0 (-0.689)
Q	Secondary metabolites biosynthesis, transport and catabolism	29 (-1.175)	122 (0.663)	10 (-0.261)	1 (1.881)
<b>Poorly characterized</b>					
R	General function prediction only	- 81 (-2.113)	341 (0.817)	37 (1.110)	1 (0.645)
S	Function unknown	- 66 (-2.724)	315 (1.247)	32 (0.807)	1 (0.756)
<b>Mobile elements</b>					
X	Mobilome: prophages, transposons	- 1 (-5.970)	+ 160 (3.666)	5 (-1.880)	+ 3 (6.354)

**Table S9. Orthogroup specificity varies among KEGG categories.**

Pearson chi-square analysis testing the independence of orthogroup specificity and KEGG category ( $\chi^2(10) = 130$ ,  $p < 0.001$ ). Each cell in the contingency represents the observed frequency and standardized residual (in between brackets) and is preceded by + or - if the standardized residual is  $>1.96$  or  $<-1.96$ , respectively, and significant at  $p < 0.05$ .

		Core	Multiple species	Single species
09100 Metabolism		511 (-0.192)	729 (0.118)	51 (0.166)
09120 Genetic Information Processing	+	136 (7.432)	- 41 (-6.062)	5 (-0.764)
09130 Environmental Information Processing	-	83 (-4.160)	+ 227 (3.183)	17 (1.232)
09140 Cellular Processes		44 (-1.044)	84 (1.347)	1 (-1.783)
09150 Organismal Systems		3 (0.123)	4 (0.033)	0 (-0.520)
09160 Human Diseases		9 (-1.150)	22 (0.800)	2 (0.644)

## Figure legends

**Figure S1. Phylogenetic tree based on partial *recA* sequences of all *Pandoraea* strains examined.** Sequences (621-663 bp) were aligned based on their amino acid sequences and phylogeny was inferred using the Maximum Likelihood method and GTRCAT substitution model in RAxML. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches if greater than 50%. *Burkholderia cenocepacia* J2315<sup>T</sup> was used as outgroup. The scale bar indicates the number of substitutions per site. Isolates selected for whole-genome sequencing are shown in bold character type.

**Figure S2. Phylogenetic tree based on partial *gltB* sequences of all *Pandoraea* strains examined.** Sequences (600-609 bp) were aligned based on their amino acid sequences and phylogeny was inferred using the Maximum Likelihood method and GTRCAT substitution model in RAxML. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches if greater than 50%. *Burkholderia cenocepacia* J2315<sup>T</sup> was used as outgroup. The scale bar indicates the number of substitutions per site. Isolates selected for whole-genome sequencing are shown in bold character type.