Self-Targeting Type IV CRISPR interference in *Pseudomonas oleovorans*

Xiaohan Guo[1†], Mariana Sanchez-Londono[1†], Vicente Gomes-Filho[1], Rogelio Hernandez-Tamayo[2], Selina Rust[1], Leah M. Immelmann[1], Pascal Schäfer[1], Julia Wiegel[1], Peter L. Graumann[2,3] and Lennart Randau*[1,3]

[1] Department of Biology, Philipps Universität Marburg, Marburg, Germany
[2] Department of Chemistry, Philipps-Universität Marburg, Marburg, Germany
[3] SYNMIKRO, LOEWE Center for Synthetic Microbiology, Marburg, Germany

* To whom correspondence should be addressed. Tel: +49 (0)6421 28 23320; Fax: +49 (0)6421-178 599; Email: lennart.randau@staff.uni-marburg.de

† Joint Authors

Extended data content

Extended data tables

Extended data figures

Extended data tables

Extended data Table 1. Identification of PAM sequences for perfect protospacer matches of indicated *P. oleovorans* CRISPR array spacers

Type I-E CRISPR array					
Spacer sequence	**contig**	**position**	**target accession number**	**target position**	**PAM**
AACAGCCAATCAATGGTGCCGTTGAACAGATA	NIUB01000004.1	171430-171461	MN366360.1	15011-15042	ACGTCAACAAGTATCTGTTCAACGGCACCATTGATTGGCTGTTCGATGCC
TTCTCGACGTGGCGGGCGACCAGCAGGAAGCG	NIUB01000004.1	157732-157763	CP046060.1	6883912-6883943	AGTCGCGCAAGCGCTTCCTGCTGGTCGCCCGCCACGTCGAGAAGGTGCCG
Type IV-A CRISPR arrays*					
Spacer sequence	**contig**	**position**	**target accession number**	**target position**	**PAM**
CCTTATCCGCCAAATGCGGCCTCAGCATGATG	NCJV01000006.1	17259-17290	CP013124.1/*Pseudomonas mendocina* S5.2	1584335-1584369	GTAAAGGAAAGCATCATGCTGAGGCCGCATTTGGCGGATAAGGGCAGCTA
GAGCCCCTTCTTCGCCAAGAAACTCGCGACCA	CP018048.1	38693-38724	CP045916.1/*Pseudomonas aeruginosa* CF39S	6770323-6770354	GGGACGATAAGGAGCCCCTTCTTCGCCAAGAAACTCGCGACCAACGGCCA
AATTTGCCGAGCGCCTTGGATCACAGGAACAC	CP018048.1	37596-37627	CP033832.1/ *Pseudomonas aeruginosa* FDAARGOS_505	1747072-1747103	TTTTGGCGAAGGTGTTCCTGTGATCCAAGGCGCTCGGCAAATTCCGGTGG
CGTATCAATAGGCACCGTGCGATCCCGACCAC	CP018048.1	38388-38419	CP013124.1/*Pseudomonas mendocina* S5.2	1584270-1584301	CGGCACGAAAGGTGGTCGGGATCGCACGGTGCCTATTGATACGCCAGAGC
TCGCGTCTCGAATCTGATGCGTAACTTGGATG	Niub01000035.1	23923-23954	CP031606.1/*pilN*	509142-509173	CCACTCAGAAGCATCCAAGTTACGCATCAGATTCGAGACGCGATTATTCG
GGTCCCTATATCCCTTAATTTTGAGAAGCTGA	Niub01000035.1	23556-23587	MH547561.1/*tnsE*	20183-20214	GGCGCGCCAAGTCAGCTTCTCAAAATTAAGGGATATAGGGACCACTTCAG
CAGCCAGTGGGTGACAGAATGAACCTGCCCCG	AP015030.1	113850-113881	CP032257.1/ *Pseudomonas aeruginosa* AR_0111	3623038-3623069	TCGGCCCTAAGCAGCCAGTGGGTGACAGAATGAACCTGCCCCGTCTAAAC

***** Type IV-A spacer sequences were collected from four highly similar systems found in *Pseudomonas* species. The position of the conserved PAM is underlined.

Extended data Table 2. List of spacers used for Type IV-A CRISPR-Cas targeting

Spacer sequence	Assay	Description
ACCAGAATCGGCACAACGCCGGTAAACAGTT	Fluorescence intensity measurement/ Plasmid curing	PAM: AAG, protospacer on coding strand
CTGGTGACCACCCTGACCTATGGCGTTCAGTG	Fluorescence intensity measurement	PAM: AAG, protospacer on non-coding strand
CCACGGAACCGGCAGTTTACCGGTGGTGCAAA	Fluorescence intensity measurement/ Plasmid curing	PAM: CGG (non-functional)
CGTTGTGCCGATTCTGGTGGAACTGGATGGTG	Fluorescence intensity measurement/ Plasmid curing	non-targeting negative control
ACAGGCGGCAGTAAGGCGGTCGGGATAGTTTTG	*LacZ* targeting assay	PAM: AAG, protospacer on coding strand
CTGCCATTGTCAGACATGTATACCCCGTACGTG	*LacZ* targeting assay	PAM: AAG, protospacer on non-coding strand
TCAACTCGCACTGTGAGGGTCACATGGGCGTT	*LacZ* targeting assay/phage assay	non-targeting negative control
TCTATCTCTCACAAATTCCGGGACTGGTAAAC	Phage assay	PAM: AAG, *geneE,* coding strand
AGAAAGTCTATCTCTCACAAATTCCGGGACTG	Phage assay	PAM: CGG, *geneE* coding strand
CATCCAAGTTACGCATCAGATTCGAGACGCGA	Efficiency of transformation assays	PAM: AAG
AATTCCTGCGTTTAGCCGCTGAAGGCTACAAC	CRISPRi assay	PAM: AAG, *trpE*, protospacer on coding strand
GCTTCTTCGGCCCCACCGCACACCTGCACCAC	CRISPRi assay	PAM: AAG, *hmgA*, protospacer on coding strand

Extended data Table 3. List of differentially expressed genes with significant adjusted p-values identified from the RNA-seq data comparing *P. oleovorans* WT and delta CRISPR array strains. Genes are sorted from highest to the lowest Log2FoldChange values, all of them with significant adjusted p-value (padj)

Genes	Annotation	Basemean	Log2FoldChange	Lfcse	Stat	pvalue	padj
PSOLE_RS11635	**PilO**	**655.947**	**60.793**	**0.8325**	**73.021**	**2.83E-13**	**1.14E-09**
PSOLE_RS11640	**PilN**	**367.056**	**59.241**	**10.144**	**58.400**	**5.22E-09**	**2.33E-06**
PSOLE_RS18395	Hypothetical	1.412.758	40.668	0.7106	57.231	1.05E-08	4.20E-06
PSOLE_RS19030	Hypothetical	1.563.214	40.355	0.6821	59.162	3.30E-09	1.80E-06
PSOLE_RS19025	Hypothetical	3.446.918	39.182	0.6219	63.007	2.96E-10	3.96E-07
PSOLE_RS14300	Hypothetical	3.562.531	38.254	0.6290	60.814	1.19E-09	9.55E-07
PSOLE_RS1163	**PilP**	**418.189**	**37.286**	**0.6942**	**53.711**	**7.82E-08**	**2.36E-05**
PSOLE_RS19020	Hypothetical	558.652	35.548	0.6980	50.929	3.53E-07	7.44E-05
PSOLE_RS14285	toprim domain-containing protein	10.710.950	34.068	0.6629	51.393	2.76E-07	6.50E-05
PSOLE_RS14310	Hypothetical	2.068.435	33.712	0.6175	54.594	4.78E-08	1.60E-05
PSOLE_RS18400	Hypothetical	1.847.183	33.670	0.6041	55.732	2.50E-08	9.12E-06
PSOLE_RS14320	DUF3150 domain-containing protein	11.652.646	31.337	0.6378	49.136	8.94E-07	0.00016
PSOLE_RS19015	Hypothetical	2.068.333	31.308	0.6667	46.961	2.65E-06	0.00041
PSOLE_RS14335		1.429.675	31.155	0.6624	47.032	2.56E-06	0.00041
PSOLE_RS19040	DNA cytosine methyltransferase	8.872.645	31.113	0.6257	49.728	6.60E-07	0.00012
PSOLE_RS18385	Hypothetical	3.457.336	30.702	0.5881	52.202	1.79E-07	4.78E-05
PSOLE_RS02515	DUF1850 domain-containing protein	2.358.322	30.688	0.7232	42.434	2.20E-05	0.00285
PSOLE_RS14325	Hypothetical	5.362.715	30.522	0.6641	45.962	4.30E-06	0.00064
PSOLE_RS14305	MoxR family ATPase	20.680.491	30.425	0.6123	49.691	6.73E-07	0.00012
PSOLE_RS14295	Hypothetical	3.409.122	30.420	0.5673	53.621	8.23E-08	2.36E-05
PSOLE_RS18365	Hypothetical	1.915.138	30.209	0.5041	59.930	2.06E-09	1.38E-06
csf2	Csf2/Cas7	5.197.382	29.179	0.5786	50.430	4.58E-07	9.19E-05

PSOLE_RS02525	TAXI family TRAP transporter	7.458.953	29.145	0.5683	51.287	2.92E-07	6.50E-05
PSOLE_RS14315	Hypothetical	31.733.390	28.202	0.5441	51.828	2.19E-07	5.48E-05
PSOLE_RS18405	Hypothetical	13.201.113	27.368	0.4438	61.662	6.99E-10	7.01E-07
PSOLE_RS14330	Hypothetical	3.925.610	27.359	0.6241	43.840	1.16E-05	0.00167
PSOLE_RS02520	TRAP transporter permease	9.514.720	25.855	0.6012	43.007	1.70E-05	0.00228
PSOLE_RS18375	Hypothetical	1.191.192	25.569	0.5858	43.651	1.27E-05	0.00176
PSOLE_RS11625	**PilQ**	**8.841.373**	**25.188**	**0.4268**	**59.022**	**3.59E-09**	**1.80E-06**
PSOLE_RS16455	Hypothetical	422.705	22.771	0.6869	33.149	0.00092	0.09935
PSOLE_RS14370	Hypothetical	1.529.682	20.279	0.5464	37.114	0.00021	0.02504
PSOLE_RS18370	Hypothetical	537.392	20.207	0.5540	36.475	0.00026	0.03034
PSOLE_RS11645	**PilM**	**1.756.765**	**18.900**	**0.4732**	**39.944**	**6.49E-05**	**0.00813**
PSOLE_RS16250	Hypothetical	9.287.882	14.686	0.4398	33.396	0.00084	0.09347
zwf	Glucose-6-phosphate dehydrogenase	11.015.997	-16.409	0.4484	-36.593	0.00025	0.02983
pgl	PGL	5.022.416	-17.786	0.5378	-33.070	0.00094	0.09951
PSOLE_RS15885	alanine:cation symporter family	3.472.504	-25.920	0.5501	-47.123	2.45E-06	0.00041
PSOLE_RS23125	16S ribosomal RNA	26.984.244	-29.933	0.4277	-69.986	2.59E-12	5.18E-09

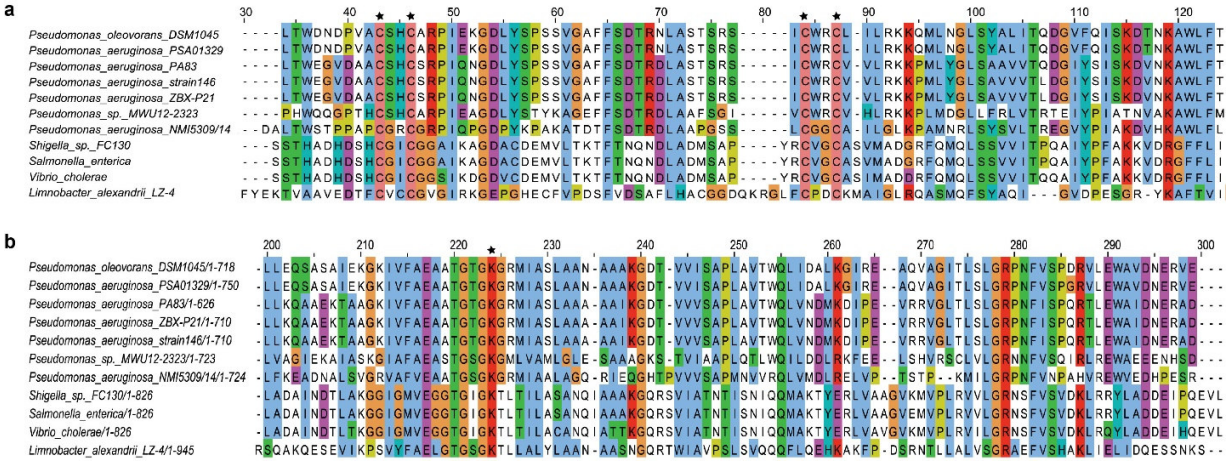
Extended data Table 4. Plasmids used for genetic manipulation of *P. oleovorans*

Plasmid	Features	Source
pEMG	Suicide vector used for deletions and insertions in Gram-negative bacteria; *oriT, traJ, lacZα, oriV(R6K);* Km^R	[40]
pMSL13	Derivative of pEMG vector carrying the HR for insertion of mNeonGreen FP in the C-terminal end of *Csf5*/Cas6 in *P. oleovorans*.	This work
pMSL15	Derivative of pEMG vector carrying the HR for deletion of the CRISPR array of the type IV CRISPR-Cas system in *P.oleovorans*	This work
pKB223	Encodes LacI-controlled expression of surface-displayed C18G and PvirK-mNeonGreen reporter	[41]/Addgene ID #170040
pSEVA6213S	Helper plasmid; *oriV* (RK2), $P_{EM7} \rightarrow$ *I-SceI;* Gm^R	[42]
pHERD30T	pUCP30T derivative plasmid; *ori* (pBR322); araC-PBAD cassette; Gm^R	[43]
pUCP18	pUC derivative plasmid; *ori*(pMB1); Amp^R	[44]

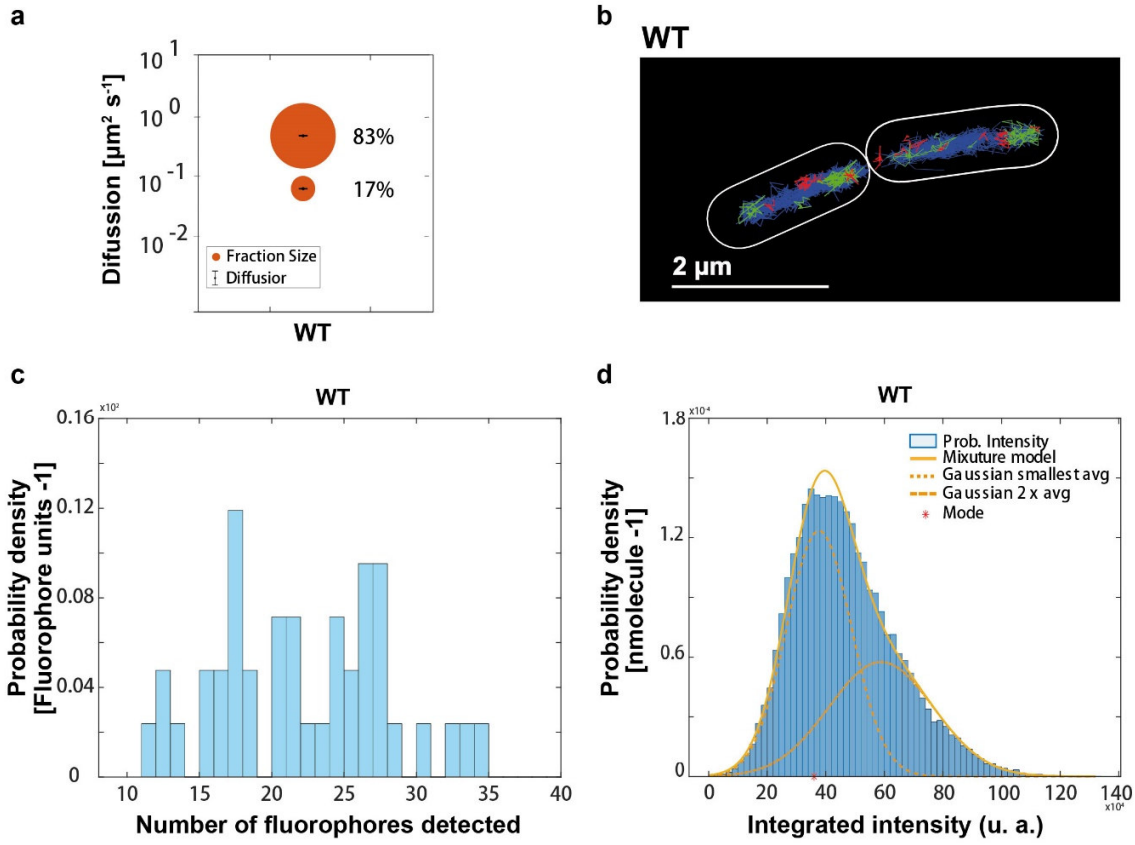
Table S5. Primers used for RT-qPCR

Primer	Sequence	Efficiency	R^2
recA fw	GTGCATGTCACCCATTTCGCCTT	110%	0.99
recA rv	CAACCGGACACCGGTGAACAG		
pilN fw	GTGCGGCAATCAGTAACCAG	110%	0.99
pilN rv	AAATACACACCGTCCGGCAG		

Extended data Figures



Extended data Fig. 1. Multiple sequence alignments of conserved Csf1 and DinG sections. The Clustal X default colour scheme is applied[4], positions of point mutations investigated in this study are labelled with an asterisk. **a**. Multiple sequence alignment of Type IV-A Csf1, cysteine residues at position 43, 46, 84 and 87 are conserved. **b**. Multiple sequence alignment of Type IV-A associated DinG proteins; a variant walker-A motif with consensus sequence TGXGK is identified.



Extended data Fig.2. Single Molecule Tracking of Cas6-mNeonGreen in *P. oleovorans*. a. Bubble plot showing fraction sizes (size of the bubble) and diffusion constants (y-axis) for Cas6-mNeonGreen in the *P. oleovorans* strain containing a wild type Type IV-A CRISPR array. Step size distributions reveal two populations, a mobile (upper circle) and a static (lower circle) fraction. **b.** Representative 2µm cell from wild type CRISPR array strains showing confinement (red), transition (green) and freely diffusive (blue) molecules. **c.** Distribution density function of the number of detected fluorophores in all cells. **d.** Distribution density function of 57329 integrated spot intensities. In the best estimation, there are two populations of average integrated intensity I1∼=29341 and I2∼=27901 u.a., with a proportion of 57%/43%. Therefore, the estimation of the number of fluorophores after accounting for crRNP formation and simulation corrections is 26 per cell.