

## Supplementary data

### **Functional reconstitution of the type IVa pilus assembly system from enterohaemorrhagic *Escherichia coli***

Areli Luna Rico<sup>1,2</sup>, Weili Zheng<sup>3</sup>, Nathalie Petiot<sup>1</sup>, Edward H. Egelman<sup>3</sup> and Olivera Francetic<sup>1#</sup>

Contents:

Supplementary Table S1.

Supplementary Figures S1, S2 and S3.

**Table S1. Oligonucleotides used in this study**

Name	Sequence (5'-3')
Linear vector 5'	GCGGCCGCGTGCATTATGATTCTTCTCGCTTC
Linear vector 3'	CTTAAGGTGCTTATCGATGATAAGCTGTCAAAC
PpdD operon 5'	GACAGCTTATCATCGATAAGCACCTTAAGTTCTTCGTAACGCCTCGC
PpdD operon 3'	CCTATAGCACCCATGGGGCCCGTTGAAAATACC
Hof operon5'	CAACCGGGCCCCATGGGTGCTATAGGCCATAAATCGAGC
Hof operon3'	AAGCACGGCGCGCCAAAAACATTACCGCTTTTACG
PpdA operon5'	GTAATGTTTTTGGCGCGCCGTGCTTGCCAGCTTTATCAGCGC
PpdA operon 3'	AAGCGAGAAGAATCATAATGCACGCGGCCGCGCCTGTGCCTCTCCCAGC
PpdD AfIII 5'	CACCTTAAGCAAAGTAGCACCAACCAAATCAA
HofM NcoI 5'	CACCCATGGAAGGCAAGCCAGACGCATTG
PpdA Ascl 5'	CAAGGCGCGCCTTCTCCTCGCTCCATACTGC
PpdC NotI 3'	CAAGCGGCCGCATACTCGAGCTTAACATAGCGGCTCCTGA
Lac p EcoRI 5'	CACGAATTCGCAGCTGGCAGCAGAGTT
Lac p ClaI 3'	CACATCGATGGCGTAATCATGGTC
GspO Xho 5'	CAACTCGAGGTCCTTCAGGGAGCAACA
GspO NotI 3'	CAAGCGGCCGCGCAAAATTATCTGCAAG
pulCp Eco	CCCGAATTCGGATAGTCCCTATACATGC
pulCp Cla	CAAATCGATCAGCAAAAGGTACCGCCCTAC
hofB AfIII 5	CTCCTTAAGCAACTAAGGAGCGGCAATG
hofB Nco 3	CACCCATGGCCAGAGTTGCTTACTCGCCAT
ppdB Kpn 5	CACGGTACCTTTTTCTCTGCTGGAAGTGTG
ppdB Eco 3	CACGAATTCTCACAGGTTGAATCCTGTAC
ppdA Kpn 5	CACGGTACCTTATACGCTGATTGAAACGCTG
ppdA Eco 3	CACGAATTCTTACAGGCATCCTTCTGTTTT
ygdB Kpn 5	CACGGTACCTAGCTTGCTATTACAAGGAATG
ygdB Eco 3	CACGAATTCTCAGGGAAGCTGACATAACG
ppdC Kpn 5	CACGGTACCTTTTAGCCTGCCGGAGGTAATG
ppdC Eco 3	CACGAATTCTACTGACGATTCCGACAATGC
ppdD Kpn 5	CACGAATTCATTCCACAGCTCACTGAAATG
ppdD Eco 3	CACGAATTCCATTTAGTGAGCTGTGGAAT
hofB Kpn 5	CACGGTACCGAATATTCCACAGCTCAC
hofB Eco 3	CACGAATTCCAGAGTTGCTTACTCGCCAT
hofC Kpn 5	CACGGTACCGGCGAGTAAGCAACTCTG
hofC Eco 3	CACGAATTCCACGCCAGCGTTATCCCAT
hofN Kpn 5	CACGGTACCGATGAACCCGCCAATTAAT
hofN Eco 3	CACGAATTGCAACCACAGTCAAAGA
hofO Kpn 5	CACGGTACCGATGAACATGTTCTTTG
hofO Eco 3	CACGAATTCCAACCAGCGTTTAACCCTC
hofP Bam 5	CACGGATCCGGTGATGATCTTTTGTTT
hofP Eco 3	CACGAATTCGAACGTCATCCACCATC
pUT18c Rev	CTTGTCTGTAAGCGGATGC
pKT25 Rev	CTCTTCGCTATTACGCCAGC
Sxy L For	CACAAGCTTCTGTATGGATGTACAGTAC
Sxy Rev	CACGAATTCTGAGATTGGTCGTCTGCAAC
Sxy S For	CACAAGCTTCGGGAAAACATAAATTAATC

## Supplementary Figures

**Figure S1. Sequence conservation of enterobacterial major pilins.** Sequence identity (in %) is shown between each pair of PpdD proteins from the 20 diverse enterobacterial species analysed in Fig. 2. The matrix is created by Clustal 2.1.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.
1.EHEC	100.00	99.32	90.34	88.11	86.21	97.26	45.39	46.76	45.89	60.42	59.59	84.62	86.21	62.50	64.14	80.69	62.76	58.33	45.00	44.29
2.E. coli K12	99.32	100.00	90.34	88.11	86.90	97.95	45.39	47.48	45.89	60.42	60.27	84.62	86.21	62.50	64.14	80.69	63.45	59.03	45.00	44.29
3.Citrobacter	90.34	90.34	100.00	85.31	87.59	90.34	46.10	46.04	45.52	60.14	57.24	81.12	86.90	61.54	61.11	79.31	59.72	55.24	46.43	46.43
4.Klebsiella	88.11	88.11	85.31	100.00	82.52	87.41	42.55	48.20	46.15	59.57	54.55	86.01	82.52	63.12	64.08	82.52	61.27	58.16	44.29	47.14
5.Salmonella	86.21	86.90	87.59	82.52	100.00	85.52	44.68	46.76	46.21	60.84	59.31	78.32	81.38	62.24	63.19	77.93	60.42	56.64	48.57	45.00
6.Shigella	97.26	97.95	90.34	87.41	85.52	100.00	46.10	46.76	46.58	59.03	58.90	83.22	86.90	62.50	62.76	82.07	64.14	58.33	45.71	45.00
7.Proteus	45.39	45.39	46.10	42.55	44.68	46.10	100.00	37.68	41.84	46.04	41.84	44.68	42.55	46.04	43.66	44.68	41.43	40.71	43.17	46.04
8.Morganella	46.76	47.48	46.04	48.20	46.76	46.76	37.68	100.00	42.45	47.45	45.32	48.92	45.32	44.20	47.83	46.04	46.76	40.58	46.04	44.60
9.Edwardsiella	45.89	45.89	45.52	46.15	46.21	46.58	41.84	42.45	100.00	51.72	47.30	51.75	45.52	50.00	45.58	47.59	46.94	45.14	42.14	42.86
10.Serratia	60.42	60.42	60.14	59.57	60.84	59.03	46.04	47.45	51.72	100.00	62.07	58.16	60.84	60.42	61.38	57.34	61.81	58.45	50.00	48.55
11.Brenneria	59.59	60.27	57.24	54.55	59.31	58.90	41.84	45.32	47.30	62.07	100.00	56.64	54.48	60.69	61.64	56.55	69.39	52.08	48.57	48.57
12.Raoultella	84.62	84.62	81.12	86.01	78.32	83.22	44.68	48.92	51.75	58.16	56.64	100.00	79.72	63.83	62.68	79.02	61.97	56.74	42.86	44.29
13.Kluyvera	86.21	86.21	86.90	82.52	81.38	86.90	42.55	45.32	45.52	60.84	54.48	79.72	100.00	62.24	61.11	81.38	59.72	55.24	46.43	45.71
14.Pantoea	62.50	62.50	61.54	63.12	62.24	62.50	46.04	44.20	50.00	60.42	60.69	63.83	62.24	100.00	71.92	62.24	60.00	54.93	48.92	48.92
15.Erwinia	64.14	64.14	61.11	64.08	63.19	62.76	43.66	47.83	45.58	61.38	61.64	62.68	61.11	71.92	100.00	59.72	56.55	51.05	46.04	45.32
16.Trabulsiella	80.69	80.69	79.31	82.52	77.93	82.07	44.68	46.04	47.59	57.34	56.55	79.02	81.38	62.24	59.72	100.00	59.72	54.55	44.29	45.71
17.Dickeya	62.76	63.45	59.72	61.27	60.42	64.14	41.43	46.76	46.94	61.81	69.39	61.97	59.72	60.00	56.55	59.72	100.00	56.64	45.00	45.71
18.Yersinia	58.33	59.03	55.24	58.16	56.64	58.33	40.71	40.58	45.14	58.45	52.08	56.74	55.24	54.93	51.05	54.55	56.64	100.00	48.20	46.76
19.Xenorhabdus	45.00	45.00	46.43	44.29	48.57	45.71	43.17	46.04	42.14	50.00	48.57	42.86	46.43	48.92	46.04	44.29	45.00	48.20	100.00	65.71
20.Photorhabdus	44.29	44.29	46.43	47.14	45.00	45.00	46.04	44.60	42.86	48.55	48.57	44.29	45.71	48.92	45.32	45.71	45.71	46.76	65.71	100.00

**Figure S2. Full induction of PpdD synthesis requires CRP-S sequence upstream of the *sxy* gene.** The *E. coli* K-12 *sxy* gene was amplified by PCR using primers Sxy For L (for pCHAP8746) or Sxy For S (for pCHAP8744) and Sxy Rev. The fragments were digested with EcoRI and HindIII and cloned in the pSU19 vector under control of *placZ* promoter. The resulting clones contained DNA sequences shown on the left pCHAP8746 containing the *sxy* gene fragment with the CRP-S sequence shown in red and pCHAP8744 containing the *sxy* gene with its promoter (in blue) without the upstream CRP-S site. Right, immunodetection of PpdD in bacteria of strain PAP7460 carrying plasmid pMS10 and either pCHAP8744 or pCHAP8746. Bacteria were grown for 48 hours on LB plates containing Ap, Cm and IPTG. Migration of Mw markers is indicated on the left with their mass shown in kDa; PpdD migration on SDS-PAGE is indicated on the right.



**Figure S3. Sequence alignments of selected T4PS and T2SS components. A.** Sequence alignment of the minor pseudopilin PulH of the *K. oxytoca* T2SS and the minor T4a pilin PpdB from EHEC. **B.** Alignment of the major pseudopilin PulG and the major pilin PpdD sequences. **C** Sequence alignment of the T2SS AP component PulM with T4aP protein HofN from EHEC. **D,** Sequence alignment of the T2SS AP component PulM with T4aP protein HofO from EHEC. The transmembrane segments of these proteins are highlighted in blue. Percentage if identical and similar residues are indicated below each comparison.

**A**

```

Method:          Diagonals (BLOSUM62)
Layout:          Standard
Mismatch penalty: Smaller (1)
Gap penalty:    Medium (2)
A A Weighting:  BLOSUM62

          20      40      60      80      100      120
PulH 1  FTLLEMLLILLMGVSRGHVLLRFFASRDDSAQT--LRAF-ERQLR-FVQQRGLQT-GQFFGVSVHPDRHQFLVLEARDGADPAPADDGHSVFRHLPLRAGVATSGS IAGKLNLRFRQGERW 120
PpdB 1  F+LLE+++ + + V          PA + + S T LA E LR F + LQ G 6+ + + A ++ W          + + + + + + + + + + + + + + + + + + + + + + + +
          20      40      60      80      100      120
          121 TPSDNPDVLIFFSGEMTPFALTGEPG IAFNRAGEGLPEPEQAQ*-----
          126 DKVTNFDRI I IDTFVQRQVDSGFSVLTVMNRASKSEPTQTVVDASVSVTGFNL*
          140      160      180
% Identity = 19.2 (35/182)   % Homology = 12.6 (23/182)   % Total = 31.9 (58/182)

```

**B**

```

Method:          Diagonals (BLOSUM62)
Layout:          Standard
Mismatch penalty: Smaller (1)
Gap penalty:    Medium (2)
A A Weighting:  BLOSUM62

          20      40      60      80      100      120
PulG 1  FTLLEIMVYVILGLVLRSLVYFPHLMGNKEKADARQKVVSDLVVALEGLDMYKLDNSRVPITTEQGLQALVSRAPREPHARNYPEGGY IARLPQDPHGSVQVLLSPGQHGQVD I FSLGPDGVPESSND 125
PpdD 1  FTLLEIMVYV I I I I L S A I G I I A V Q N Y L R K A A L T D M L Q T F V P Y R T A V E L C A L E H G L D T C D G G S N G I P S P T T T R V S A M S V A K G V V S L T G Q E S L N G L S V M P T P G M N A N G V T G H A R N C I Q S D S A 125
          20      40      60      80      100      120
          126 IGHWTIOKK-----
          126 LQQACEDVFRFDDRI
          140
% Identity = 17.1 (24/140)   % Homology = 20.7 (29/140)   % Total = 37.9 (53/140)

```

**C**

```

Method:          Diagonals (BLOSUM62)
Layout:          Standard
Mismatch penalty: Smaller (1)
Gap penalty:    Medium (2)
A A Weighting:  BLOSUM62

          20      40      60      80      100      120
PulM 1  MHNLLA-L-N-QQRTAERCLLLGHAVVLL IGLVYVTLHQPQNR-ERQWRQLTAR-EQ--A-SLQNM--RQQTPL-IARLNKPKPTAPEEPSTV I MREARRHGLT I VR-LQP---QGRSLT 110
HofN 1  M+ + L W QQR          L M V L + V TL          + ER+ L + EQ A SLQ R A IR+ R Q+          IR+ R Q+          + + + + + + + + + + + + + + + + + + + + + + + +
          20      40      60      80      100      120
          111 VQPADFOALMHNLDALGQ-AGHTTATLAVTRVAQPGWVTNTLVLEASDEK---
          126 GLTTS I T A L N A E T S L R Q D S F H L N R G A T Q Q D A G R W Q F E V Q L T R K V S D E H V L *
          140      160      180
% Identity = 25.6 (46/180)   % Homology = 6.1 (11/180)   % Total = 31.7 (57/180)

```

**D**

```

Method:          Diagonals (BLOSUM62)
Layout:          Standard
Mismatch penalty: Smaller (1)
Gap penalty:    Medium (2)
A A Weighting:  BLOSUM62

          20      40      60      80      100      120
PulM 1  MHNLLALQQRTAERCLLLGHAVVLL IGLVYVTLHQPQNRERQWRQLTAR-EQ--A-SLQNM--RQQTPL IARLNKPKPTAPEEPSTV I MREARRHGLT I VALQPQGRSLT VQPADFOALMHNLDRA 125
HofO 1  M-NMFFDHFATSP=RLRQFCUAFULLM=LV=TL I FLSSTHHEE=RDAL I ALRASHHQA=TLVRLVDTTFPSEEKTLFPSPDFQLPQAQLVFHPSAQGGELAKLTLHEAVPSAFRTL=A 116
          20      40      60      80      100      120
          126 LQQDHTTATLAVTRVAQPGWVTNTLVLEASDEK
          117 ERNYSVSRFSLSVEGDOLLFTLQLETPHE6*=====
          140
% Identity = 18.6 (30/161)   % Homology = 8.7 (14/161)   % Total = 27.3 (44/161)

```