1 *Bacillus* **endospore appendages form a novel family of disulfide-linked pili**
2

3 Brajabandhu Pradhan[1,2,§], Janine Liedtke[3,§], Mike Sleutel[1,2,§], Toril Lindbäck[3], Ann-Katrin

4 Llarena[3], Ola Brynildsrud[3,4], Marina Aspholm[3*], Han Remaut[1,2*]

5 1. Structural and Molecular Microbiology, VIB-VUB Center for Structural Biology, VIB,

6   1050 Brussels, Belgium.

7 2. Department of Bioengineering Sciences, Structural Biology Brussels, Vrije Universiteit

8   Brussel, 1050 Brussels, Belgium.

9 3. Department of Paraclinical Sciences, Faculty of Veterinary Medicine, the Norwegian

10   University of Life Sciences, PO Box 369 Sentrum, N-0102, Oslo, Norway

11 4. Division of Infection Control and Environmental Health, Norwegian Institute of Public

12   Health, P.O. Box 222 Skøyen, 0213 Oslo, Norway

13

14 [§] These authors contributed equally

15 *Correspondence: han.remaut@vub.be and marina.aspholm@nmbu.no

16

17

18 **Summary**

19 *Bacillus cereus sensu lato* is a group of Gram-positive endospore-forming bacteria with high

20 ecological diversity. Their endospores are decorated with micrometer-long appendages of

21 unknown identity and function. Here we isolate endospore appendages (Enas) from the

22 food poisoning outbreak strain *B. cereus* NVH 0075-95 and find proteinaceous fibers of two

23 main morphologies. By using cryo-EM and 3D helical reconstruction we show that *Bacillus*

24 Enas form a novel class of Gram-positive pili. Enas consist of single domain subunits with

25 jellyroll topology that are laterally stacked by β-sheet augmentation. Enas are longitudinally

26 stabilized by disulfide bonding through N-terminal connector peptides that bridge the

27 helical turns. Together, this results in flexible pili that are highly resistant to heat, drought

28 and chemical damage. Phylogenomic analysis reveals the presence of defined *ena* clades

29 amongst different eco- and pathotypes. We propose Enas to represent a novel class of pili

30 specifically adapted to the harsh conditions encountered by bacterial spores.

33

34   **Introduction**

35   When faced with adverse growth conditions, bacteria belonging to the phylum *Firmicutes*

36   can differentiate into the metabolically dormant endospore state. These endospores exhibit

37   extreme resilience towards environmental stressors due to their dehydrated state and

38   unique multilayered cellular structure, and can germinate into the metabolically active and

39   replicating vegetative growth state even hundreds of years after their formation (Setlow,

40   2014). In this way, Firmicutes belonging to the classes Bacilli and Clostridia can withstand

41   long periods of drought, starvation, high oxygen or antibiotic stress.

42   Endospores typically consist of an innermost dehydrated core which contains the bacterial

43   DNA. The core is enclosed by an inner membrane surrounded by a thin layer of

44   peptidoglycan that will become the cell wall of the vegetative cell that emerges during

45   endospore germination. Then follows a thick cortex layer of modified peptidoglycan that is

46   essential for dormancy. The cortex layer is in turn surrounded by several proteinaceous coat

47   layers (Atrih and Foster, 1999). In some *Clostridium* and most *Bacillus cereus* group species,

48   the spore is enclosed by an outermost loose-fitting paracrystalline exosporium layer

49   consisting of (glyco)proteins and lipids (Stewart, 2015). The surface of *Bacillus* and

50   *Clostridium* endospores can also be decorated with multiple micrometers long and a few

51   nanometers wide filamentous appendages, which show a great structural diversity between

52   strains and species (Hachisuka and Kuno, 1976; Rode et al., 1971; Walker et al., 2007).

53   Spores of species belonging to the *B. cereus* group are often covered with appendages

54   which morphologically resemble pili of Gram-negative and Gram-positive bacteria when

55   imaged by negative strain TEM (Ankolekar and Labbe, 2010; Smirnova et al., 2013). The

56   endospore appendages, hereafter called Enas, vary in number and morphology between *B.*

57   *cereus* group strains and species, and some strains even simultaneously express Enas of

58   different morphologies (Smirnova et al., 2013). Structures resembling the Enas have not

59   been observed on the surface of the vegetative cells suggesting that they may represent

60   spore-specific fibers.

61 Although the presence of endospore appendages in species belonging to the *B. cereus* group

62 was reported already in the '60s, efforts to characterize their composition and genetic

63 identity have failed due to difficulties to solubilize and enzymatically digest the fibers

64 (DesRosier and Lara, 1981; Gerhardt and Ribi, 1964). Therefore, there is no genetic or

65 structural information and very limited functional data available for endospore appendages.

66 Here we isolate Enas from the food poisoning outbreak strain *B. cereus* NVH 0075-95 and

67 find proteinaceous fibers of two main morphologies. By using cryo-EM and 3D helical

68 reconstruction we show that the major form of Enas represents a novel class of Gram-

69 positive pili. A unique architecture of subunit stabilization by lateral β-augmentation and

70 longitudinal disulfide crosslinking gives rise to pili that combine high flexibility with high

71 resistance to heat, drought and chemical damage. The genetic identity of the S-type Enas

72 was deduced from the structural model and confirmed by analysis of mutants lacking genes

73 encoding potential Ena protein subunits. S-type Ena fibers are encoded by three associated

74 genes which are present in most species of the *B. cereus* group. Remarkably, recombinant

75 Ena subunits spontaneously self-assemble *in vitro* and *in vivo* into protein nanofibers with

76 native Ena-like properties and structure.

77

78 **Results**

79 ***Bacillus cereus* NVH 0075-95 show endospore appendages of two morphological types**

80 Negative stain EM imaging of *B. cereus* strain NVH 0075-95 showed typical endospores with

81 a dense core of ~1 μm diameter, tightly wrapped by an exosporium layer that on TEM

82 images emanates as a flat 2-3 μm long saclike structure from the endospore body (Figure

83 1A). The endospores showed an abundance of micrometer-long appendages (Enas) (Figure

84 1A). The average endospore counted 20 - 30 Enas ranging from 200 nm to 6 μm in length

85 (Figure 1E), with a median length of approximately 600 nm. The density of Enas appeared

86 highest at the pole of the spore body that lies near the exosporium. There, Enas seem to

87 emerge from the exosporium as individual fibers or as a bundle of individual fibers that

88 separates a few tens of nanometers above the endospore surface (Figures 1B and S1B).

89 Closer inspection revealed that the Enas showed two distinct morphologies (Figure 1 C, D).

90 The main or "Staggered-type" (S-type) morphology represents approximately 90% of the

3

91   observed fibers. S-type Enas have a width of ~110 Å and give a polar, staggered appearance

92   in negative stain 2D classes, with alternating scales pointing down to the spore surface. At

93   the distal end, S-type Enas terminate in multiple filamentous extensions or "ruffles" of 50 -

94   100 nm in length and ~35 Å thick (Figure 1C). The minor or "Ladder-like" (L-type) Ena

95   morphology is thinner, ~80 Å in width, and terminates in a single filamentous extension with

96   dimensions similar to ruffles seen in S-type fibers (Figure 1D). L-type Enas lack the scaled,

97   staggered appearance of the S-type Enas, instead showing a ladder of stacked disk-like units

98   of ~40 Å height. Whereas S-type Enas can be seen to traverse the exosporium and connect

99   to the spore body, L-type Enas appear to emerge from the exosporium (Figure S1A). Both

100 Ena morphologies co-exist on individual endospores (Figure S1C). Neither Ena morphology is

101 reminiscent of sortase-mediated or type IV pili previously observed in Gram-positive

102 bacteria (Mandlik et al., 2008; Melville and Craig, 2013). In an attempt to identify their

103 composition, shear force extracted and purified Enas were subjected to trypsin digestion for

104 identification by mass spectrometry. However, despite the good enrichment of both S- and

105 L-type Enas, no unambiguous candidates for Ena were identified amongst the tryptic

106 peptides, which largely contained contaminating mother cell proteins, EA1 S-layer and spore

107 coat proteins. Attempts to resolve the Ena monomers by SDS-PAGE were unsuccessful,

108 including strong reducing conditions (up to 200 mM β-mercaptoethanol), heat treatment

109 (100 °C), limited acid hydrolysis (1h 1M HCl), or incubation with chaotropes such as 8M urea

110 or 6M guanidinium chloride. Ena fibers also retained their structural properties upon

111 autoclaving, desiccation or treatment with proteinase K (Figure S1C).

112 **Cryo-EM of endospore appendages identifies their molecular identity**

113 To further study the nature of the Enas, fibers purified from *B. cereus* NVH 0075-95

114 endospores were imaged by cryogenic electron microscopy (cryo-EM) and analyzed using 3D

115 reconstruction. Isolated fibers showed a 9.4:1 ratio of S- and L-type Enas, similar to what

116 was seen on endospores. Boxes with a dimension of 300 x 300 pixels (246 x 246 Å^2) were

117 extracted along the length of the fibers, with an inter-box overlap of 21 Å, and subjected to

118 2D classification using RELION 3.0 (Zivanov et al., 2018). Power spectra of the 2D class

119 averages revealed a well-ordered helical symmetry for S-type Enas (Figure 2A, B), whereas L-

120 type Enas primarily showed translational symmetry (Figure 1D). Based on a helix radius of

121 approximately 54.5 Å, we estimated layer lines Z' and Z" in the power spectrum of S-type

122 Enas to have a Bessel order of -11 and 1, respectively (Figure 2A, B). In the 2D classes

123 holding the majority of extracted boxes the Bessel order 1 layer line was found at a distance

124 of 0.02673 Å^{-1} from the equator, corresponding to a pitch of 37.4 Å, in good agreement with

125 spacing of the apparent 'lobes' seen also by negative stain (Figures 1C, 2B and S1). The

126 correct helical parameters were derived by an empirical approach in which a systematic

127 series of starting values for subunit rise and twist were used for 3D reconstruction and real

128 space Bayesian refinement using RELION 3.0 (He and Scheres, 2017). Based on the

129 estimated Fourier – Bessel indexing, input rise and twist were varied in the range of 3.05 –

130 3.65 Å and 29 – 35 degrees, respectively, with a sampling resolution of 0.1 Å and 1 degree

131 between tested start values. This approach converged on a unique set of helical parameters

132 that resulted in 3D maps with clear secondary structure and identifiable densities for

133 subunit side chains (Figure 2C). The reconstructed map corresponds to a left-handed 1-start

134 helix with a rise and twist of 3.22937 Å and 31.0338 degrees per subunit, corresponding to a

135 helix with 11.6 units per turn (Figure 2D). After refinement and postprocessing in RELION

136 3.0, the map was found to be of resolution 3.2 Å according to the $FSC_{0.143}$ criterion.

137   The resulting map showed well defined subunits comprising an 8-stranded β-

138 sandwich domain of approximately 100 residues (Figure 2E). The side chain density was of

139 sufficient quality to manually deduce a short motif with the sequence F-C-M-V/T-I-R-Y

140 (Figure S2A). A search of the *B. cereus* NVH 0075-95 proteome (GCA_001044825.1)

141 identified two hypothetical proteins of unknown function, namely KMP91697.1 and

142 KMP91698.1, encoded by TU63_02435 and TU63_02440 respectively (Figure S2B). Further

143 inspection of the electron potential map and manual model building of the Ena subunit

144 showed this to fit well with the protein KMP91698.1. TU_63_02440 is located 15 bp

145 downstream of the TU63_02435 locus. Both genes encode hypothetical proteins of similar

146 size (117 and 126 amino acids and estimated molecular weights of 12 and 14 kDa, for

147 KMP91698.1 and KMP91697.1, respectively), with 39% pairwise amino acid sequence

148 identity, a shared domain of unknown function (DUF) 3992 and similar Cys patterns (Figure

149 S2B). Further downstream of TU_63_02440, on the minus strand, the locus TU63_0245

150 encodes a third DUF3992 containing hypothetical protein (KMP91699.1), of 160 amino acids

151 and an estimated molecular weight of 17 kDa. As such, KMP91697.1, KMP91698.1 and

152 KMP91699.1 are regarded as candidate Ena subunits, hereafter dubbed Ena1A, Ena1B and
153 Ena1C, respectively (Figure S2B,C).

154

Ena1B self-assembles into endospore appendage-like nanofibers *in vitro*

156 To confirm the subunit identity of the endospore appendages isolated from *B. cereus*
157 NVH0075-95, we cloned a synthetic gene fragment corresponding to the coding sequence of
158 Ena1B and an N-terminal TEV protease cleavable 6xHis-tag into a vector for recombinant
159 expression in the cytoplasm of *E. coli*. The recombinant protein was found to form inclusion
160 bodies, which were solubilized in 8M urea before affinity purification. Removal of the
161 chaotropic agent by rapid dilution resulted in the formation of abundant soluble crescent-
162 shaped oligomers reminiscent of a partial helical turn seen in the isolated S-type Enas
163 (Figure S2A-E), suggesting the refolded recombinant Ena1B (*rec*Ena1B) adopts the native
164 subunit-subunit β-augmentation contacts (Figure S2E). We reasoned that *rec*Ena1B self-
165 assemble into helical appendages arrested at the level of a single turn due to steric
166 hindrance by the 6xHis-tag at the subunits N-terminus. Indeed, proteolytic removal of the
167 affinity tag readily resulted in the formation of fibers of 110 Å diameter and with helical
168 parameters similar to S-type Enas, though lacking the distal ruffles seen in *ex vivo* fibers
169 (Figure S2F). CryoEM data collection and 3D helical reconstruction was performed to assess
170 whether *in vitro rec*Ena1B nanofibers were isomorphous with *ex vivo* S-type Enas. Real
171 space refinement of helical parameters using RELION 3.0 converged on a subunit rise and
172 twist of 3.43721 Å and 32.3504 degrees, respectively, approximately 0.2 Å and 1.3 degrees
173 higher than found in *ex vivo* S-type Enas, and corresponding to a left-handed helix with a
174 pitch of 38.3 Å and 11.1 subunits per turn. Apart from the minor differences in helical
175 parameters the 3D reconstruction map of *in vitro* Ena1B fibers (estimated resolution of 3.2
176 Å; Figure S3A, B) was near isomorphous to *ex vivo* S-type Enas in terms of size and
177 connectivity of the fiber subunits (Figure S3D). Closer inspection of the 3D cryoEM maps for
178 *rec*Ena1B and *ex vivo* S-type Ena showed an improved side chain fit for Ena1B residues in the
179 former (Figure S3B, C, D) and revealed regions in the *ex vivo* Ena maps that showed partial
180 side-chain character of Ena1A, particularly in loop L1, L3, L5 and L7 (Figure S2B, S3B,C).
181 Although the Ena1B character of the *ex vivo* maps is dominant, this suggested that *ex vivo* S-
182 type Enas consist of a mixed population of Ena1A and Ena1B fibers, or that S-type Enas have

6

183   a mixed composition comprising both Ena1A and Ena1B. Immunogold labelling using sera

184   generated with *rec*Ena1A or *rec*Ena1B showed subunits-specific labeling within single Enas,

185   confirming these have a mixed composition of Ena1A and Ena1B (Figure S3E). No staining of

186   S-type Enas was seen with Ena1C serum (Figure S3E). No systematic patterning or molar

187   ratio for Ena1A and Ena1B could be discerned from immunogold labelling or helical

188   reconstructions with an asymmetric unit containing more than one subunit, suggesting the

189   distribution of Ena1A and Ena1B in the fibers to be random. Apart from some side chain

190   densities with mixed Ena1A and Ena1B character, the cryoEM electron potential maps of the

191   *ex vivo* Enas showed a unique main chain conformation, indicating the Ena1A and Ena1B

192   have near isomorphous folds.

193

Enas represent a novel family of Gram-positive pili

195   Upon recognizing that native S-type Enas show a mixed Ena1A and Ena1B

196   composition, we continued with 3D cryoEM reconstruction of *rec*Ena1B for model building.

197   The Ena subunit consists of a typical jellyroll fold (Richardson, 1981) comprised of two

198   juxtaposed β-sheets consisting of strands BIDG and CHEF (Figure 2F). The jellyroll domain is

199   preceded by a flexible 15 residue N-terminal extension hereafter referred to as N-terminal

200   connector ('Ntc'). Subunits align side by side through a staggered β-sheet augmentation

201   (Remaut and Waksman, 2006), where the sheet composed of strands BIDG of a subunit i is

202   augmented with strands CHEF of the preceding subunit i-1, and strands CHEF of subunit i are

203   augmented with strands BIDG of the next subunit in row i+1 (Figure 2F, Figure S4A, B). As

204   such, the packing in the endospore appendages can be regarded as a slanted β-propeller of

205   8-stranded β-sheets, with 11.6 blades per helical turn and an axial rise of 3.2 Å per subunit

206   (Figure 2E). Subunit-subunit contacts in the β-propeller are further stabilized by two

207   complementary electrostatic patches on the Ena subunits (Figure S4C). In addition to these

208   lateral contacts, subunits across helical turns are also connected through the Ntc's. The Ntc

209   of each subunit i makes disulfide bond contacts with subunits i-9 and i-10 in the preceding

210   helical turn (Figure 2F, Figure S4B). These contacts are made through disulfide bonding of

211   Cys 10 and Cys 11 in subunit i, with Cys 109 and Cys 24 in the strands I and B of subunits i-9

212   and i-10, respectively (Figure 2F, S4B). Thus, disulfide bonding via the Ntc results in a

213   longitudinal stabilization of fibers by bridging the helical turns, as well as in a further lateral

214 stabilization in the β-propellers by covalent cross-linking of adjacent subunits. The Ntc

215 contacts lie on the luminal side of the helix, leaving a central void of approximately 1.2 nm

216 diameter (Figure S4D). Residues 12-17 form a flexible spacer region between the Ena

217 jellyroll domain and the Ntc. Strikingly, this spacer region creates a 4.5 Å longitudinal gap

218 between the Ena subunits, which are not in direct contact other than through the Ntc

219 (Figure 3C, S2B). The flexibility in the Ntc spacer and the lack of direct longitudinal protein-

220 protein contact of subunits across the helical turns create a large flexibility and elasticity in

221 the Ena fibers (Figure 3). 2D class averages of endospore-associated fibers show longitudinal

222 stretching, with a change in pitch of up to 8 Å (range: 37.1 – 44.9 Å; Figure 3D), and an axial

223 rocking of up to 10 degrees per helical turn (Figure 3A, B).

224 Thus, *B. cereus* endospore appendages represent a novel class of bacterial pili,

225 comprising a left-handed single start helix with non-covalent lateral subunit contacts formed

226 by β-sheet augmentation, and covalent longitudinal contacts between helical turns by

227 disulfide bonded N-terminal connecter peptides, resulting in an architecture that combines

228 extreme chemical stability (Figure S1) with high fiber flexibility.

229

230 **The *ena1* coding region for S-type Enas**

231 In *B. cereus* NVH 0075-95 Ena1A, Ena1B and Ena1C are encoded in a genomic region flanked

232 upstream by *dedA* (genbank protein-id: KMP91696.1) and a gene encoding a 93-residue

233 protein of unknown function (DUF1232, genbank: KMP91695.1) (Figure 4A). Downstream,

234 the *ena*-gene cluster is flanked by a gene encoding an acid phosphatase (TU63_02450).

235 Within the *ena*-gene cluster, *ena1A* and *ena1B* are found in forward, and *ena1C* in reverse

236 orientation, respectively (Figure 4A). PCR analysis of NVH 0075-95 cDNA made from mRNA

237 isolated after 4 and 16 h of culture, representative for vegetative growth and sporulating

238 cells, respectively, indicated that *ena1A* and *ena1B* are co-expressed from a bicistronic

239 transcript during sporulation but not during vegetative growth (Figure 4B). A weak

240 amplification signal was observed in vegetative cells when the forward primer was located

241 in *dedA* upstream of *ena1A* and the reverse primer was located within *ena1B* (Figure 4B,

242 lane 2) suggesting that some *enaA* and *enaB* is coexpressed with *dedA*. This was observed in

243 vegetative cells or very early in sporulation but not during later sporulation stages and may

244   represent a fraction of improperly terminated *dedA* mRNA. Reverse transcription

245   quantitative PCR (RT-qPCR) analysis showed increased expression of *ena1A*, *ena1B* and

246   *ena1C* in sporulating cells compared to vegetative cells (Figure 4B). CryoEM maps and

247   immuno-gold TEM analysis of *ex vivo* S-type Enas indicated these contain both Ena1A and

248   Ena1B (Figure S3B-D). To determine the relative contribution of Ena1 subunits to *B. cereus*

249   Enas we made individual chromosomal knockouts of *ena1A*, *ena1B*, as well as *ena1C* in

250   strain NVH 0075-95 and investigated their respective endospores by TEM. All *ena1* mutants

251   made endospores of similar dimensions to WT and with intact exosporium (Figure 5A, Figure

252   S5). Both the *ena1A* and *ena1B* mutant resulted in endospores completely lacking S-type

253   Enas, in agreement with the mixed content of *ex vivo* fibers. The *ena1C* mutant also resulted

254   in the loss of S-type Ena on the endospores (Figure 5A), even though staining with anti-

255   Ena1C serum did not identify the presence of the protein inside S-type Enas (Figure S3D). All

256   three mutants still showed the presence of L-type Enas, of similar size and number density

257   as WT endospores, although statistical analysis does not rule out L-type Enas to have a slight

258   increase in length in the *ena1B* and *ena1C* mutants (length p=0.003 and <0.0001, resp.)

259   (Figure 5B). Thus, Ena1A, Ena1B and Ena1C are mutually required for *in vivo* S-type Ena

260   assembly, but not for L-type Ena assembly. Complementation of the *ena1B* mutant with a

261   low copy plasmid (pMAD-I-*Sce*I) containing *ena1A-ena1B* restored S-type Ena expression.

262   Plasmid-based expression of these subunits resulted in an average ~2-fold increase in the

263   number of S-type Enas per spore, and a drastic increase in Ena length, now reaching several

264   microns (Figure 5A, B, Figures S5D). Thus, the number and length of S-type Enas depend on

265   the concentration of available Ena1A and Ena1B subunits. Notably, several endospores

266   overexpressing Ena1A and Ena1B appeared to lack an exosporium or showed the

267   entrapment of S-type Enas inside the exosporium (Figure S5C, D). This demonstrates that S-

268   type Enas emanate from the spore body, and that a disbalance in the concentration or

269   timing of *ena* expression can result in mis-assembly and/or mislocalization of endospore

270   surface structures. Contrary to S-type Enas, close inspection of the WT and mutant

271   endospores suggests that L-type Enas emanate from the surface of the exosporium rather

272   than the spore body. The molecular identity of the L-type Ena, or the single or multiple

273   terminal ruffles seen, respectively, in L- and S-type Enas was not determined in the present

274   study.

275

Phylogenetic distribution of the *ena1A-C* genes

277   To investigate the occurrence of *ena1A-C* within the *B. cereus s.l.* group and other relevant

278   species of the genus *Bacillus*, pairwise tBLASTn searches for homologs of Ena1A-C were

279   performed on a database containing all available closed, curated *Bacillus* spp. genomes,

280   with the addition of scaffolds for species for which closed genomes were lacking (*n*=735,

281   Table S3). Homologs and orthologs with high coverage (>90%) and high amino acid

282   sequence identity (>80%) to Ena1A or Ena1B of *B. cereus* NVH 0075-95 were found in 48

283   strains including 11 of 85 *B. cereus* strains, 13 of 119 *B. wiedmannii* strains, 14 of 14 *B.*

284   *cytotoxicus* strains, one of one *B. luti* (100%) strain, three of six *B. mobilis* strains, three of 33

285   *B. mycoides* strains, one of one *B. tropics* strain and both *B. paranthracis* strains analyzed. Of

286   these strains, only 31 also carried a gene encoding a homolog with high sequence identity

287   and coverage to Ena1C of *B. cereus* NVH 0075-95 (Figure 6). All investigated *B. cytotoxicus*

288   genomes (14/14) encoded hypothetical Ena1A and Ena1B proteins, but only 12/14 encoded

289   an Ena1C ortholog, which showed only a moderate amino acid conservation compared to

290   the Ena1C of *B. cereus* NVH 0075-95 (mean 63.9% amino acid sequence identity) (Figure 6,

291   Figure S5).

292

293   Upon searching for Ena1A-C homologs in *B. cereus* group genomes, a candidate orthologous

294   gene cluster encoding hypothetical EnaA-C proteins was discovered. These three proteins

295   had, respectively, an average of 59.3±0.9%, 43.3±1.6% and 53.9±2.2% amino acid sequence

296   identity with Ena1A, Ena1B and Ena1C of *B. cereus* NVH0075-95, and shared gene synteny

297   (Figure 6b). The orthologous *ena* gene cluster was named *ena2A-C*. Except for *B. subtilis*

298   (*n*=127) and *B. pseudomycoides* (*n*=8), all genomes analyzed (*n*=735) carried either *ena1*

299   (*n*=48) or the *ena2* (*n*=476) gene cluster. *Ena1A-C* or the *ena2A-C* were never present

300   simultaneously and no chimeric *ena1A-C/2A-C* clusters were discovered among the

301   genomes analyzed (Figure 6). In addition to the main split between Ena1A-C and Ena2A-C in

302   the protein trees, distinct sub-clusters were seen among Ena1A, Ena1B and, especially,

303   Ena1C sequences (Figure S5). The Ena1A sequences separated into two main sub-clusters:

304   one present in the majority of *B. cytotoxicus* strains and another found in *B. wiedmanni* and

305   *B. cereus* strains (Figure S5A). More variation was evident for EnaB proteins: Ena1B

306   sequences formed two clusters; one containing *B. cereus* and *B. wiedmannii* isolates, and

307   the other with *B. cytotoxicus* (Figure S5B). Also, a separate sub-cluster of Ena2B proteins

308   was seen (Figure S5B), containing isolates of *B. mycoides, B. cereus, B. thuringiensis, B.*

309   *pacificus,* and *B. wiedmannii* that shared around ~78% and ~48% sequence identity with the

310   remainder of Ena2B and Ena1B, respectively. EnaC was the most variable of the three

311   proteins: Ena1C formed a monophyletic clade containing isolates of *B. wiedmanni, B. cereus,*

312   *B. anthracis, B. paranthracis, B. mobilis, B. tropicus,* and *B. luti,* but had considerable

313   sequence variation in species and strains carrying *Ena2AB* as well as in subset of strains

314   carrying *Ena1AB.*

315

316   The *ena2A-C* homo- or orthologs were much more common among *B. cereus* group strains

317   than the *ena1A-C* genes; all investigated *B. toyonensis* (*n=204*), *B. albus* (*n=1*), *B.*

318   *bombysepticus* (*n=1*), *B. nitratireducens* (*n=6*), *B. thuringiensis* (*n=50*) genomes and in the

319   majority of *B. cereus* (87%, 74/85), *B. wiedmannii* (105/119, 89.3%), *B. tropicus* (71%, 5/7,)

320   and *B. mycoides* (91%, 30/33) had the Ena2A-C form of the protein (Figure 6). No *ena*

321   orthologs were found in *B. subtilis* (*n=127*) or *B. pseudomycoides* (n=8) genomes or in any

322   other genomes outside the *B. cereus* group except for three misclassified *Streptococcus*

323   *pneumoniae* genomes (GCA_001161325, GCA_001170885, GCA_001338635) and one

324   misclassified *B. subtilis* genome (GCA_004328845). These genomes and the *B. subtilis* were

325   re-classified as *B. cereus* when re-analyzed with three different methods for taxonomic

326   classification (Masthree, 7-lociMLST and Kraken, see Methods). The genomes of a few

327   *Peanibacillus* spp. strains had genes encoding hypothetical proteins with a low level of

328   amino acid sequence similarity to Ena1A-C, and genes encoding hypothetical proteins with

329   some similarity to Ena1A and Ena1B were also found in the genome of a *Cohnella abietis*

330   strain (GCF_004295585.1). These hits outside of *Bacillus* genus was in the DUF3992 domain

331   of these genes, which is found in *Anaeromicrobium, Cochnella,* and of the order Bacillales.

332

333   A few genomes had deviations in the *ena*-gene clusters compared to other strains of their

334   species. Two of three *B. mycoides* strains (GCF_007673655 and GCF_007677835.1) lacked

335   the *ena1C* allele downstream of the *ena1A-B* operon (data not shown). However, potential

336   *ena1c* orthologs encoding hypothetical proteins with 50% identity to Ena1C of *B. cereus* NVH

337   0075-95 were found elsewhere in their genomes. One genome annotated as *B. cereus*

338   (strain Rock3-44 Assembly: GCA_000161255.1) grouped with these strains of *B. mycoides*

339   (Figure 6) and shared their *ena1A-C* distribution pattern with. *B. thuringiensis* usually carries

340   *ena2* gene, but a genome annotated as *B. thuringiensis* (strain LM1212, GCF_003546665)

341   lacked all *ena* genes. This strain was nearly identical to the reference strain of *B. tropicus*,

342   which also lacked both the *ena* gene clusters. One *B. toyonensis* strain AFS086269

343   GCF_002568845 also lacked all three genes, while the remainder of the 204 strains of *B.*

344   *toyonensis* all had *ena2A-C*.

345

346

347   **Discussion**

348   Endospores formed by *Bacillus* and *Clostridium* species frequently carry surface-attached

349   ribbon- or pilus-like appendages (Driks, 2007), the role of which has remained largely

350   enigmatic due to the lack of molecular annotation of the pathways involved in their

351   assembly. Half a century following their first observation (Hachisuka and Kuno, 1976;

352   Hodgikiss, 1971), we employ high resolution *de novo* structure determination by cryoEM to

353   structurally and genetically characterize the appendages found on *B. cereus* spores. We

354   found that *B. cereus* Enas come in two main morphologies: 1) staggered or S-type Enas that

355   are several micrometers long and emerge from the spore body and traverses the

356   exosporium, and 2) smaller, less abundant ladder- or L-type Enas that appears to directly

357   emerge from the exosporium surface. Our phylogenetic analyses of S-type fibers reveal Ena

358   subunits belonging to a conserved family of proteins encompassing the domain of unknown

359   function DUF3992.

360   Covalent bonding, and the highly compact jellyroll fold result in a high chemical and physical

361   stability of the Ena fibers, withstanding desiccation, high temperature treatment, and

362   exposure to proteases. The formation of linear filaments of multiple hundreds of subunits

363   requires stable, long-lived subunit-subunit interactions with high flexibility to avoid that a

364   dissociation of subunit-subunit complexes results in pilus breakage. This high stability and

365   flexibility are likely to be adaptations to the extreme conditions that can be met by

366   endospores in the environment or during the infectious cycle. Two molecular pathways are

367   known to form surface fibers or "pili" in Gram-positive bacteria: 1) sortase-mediated pilus

368   assembly, which encompasses the covalent linkage of pilus subunits by means of a

369   transpeptidation reaction catalyzed by sortases (Ton-That and Schneewind, 2004), and 2)

370   Type IV pilus assembly, encompassing the non-covalent assembly of subunits through a

371   coiled-coil interaction of a hydrophobic N-terminal helix (Melville and Craig, 2013). Sortase-

372   mediated pili and Type IV pili are formed on vegetative cells, however, and to date, no

373   evidence is available to suggest that these pathways are also responsible for the assembly of

374   endospore appendages.

375   Until the present study, the only species for which the genetic identity and protein

376   composition of spore appendages has been known, is the non-toxigenic environmental

377   species *Clostridium taeniosporum,* which carry large (4.5 μm long, 0.5 μm wide and 30 nm

378   thick) ribbon-like appendages, which are structurally distinct from those found in most

379   other *Clostridium* and *Bacillus* species. *C. taeniosporum* lacks the exosporium layer and the

380   appendages seem to be attached to another layer, of unknown composition, outside the

381   coat (Walker et al., 2007). The *C. taeniosporum* endospore appendages consist of four major

382   components, three of which have no known homologs in other species and an orthologs of

383   the *B. subtilis* spore membrane protein SpoVM (Walker et al., 2007). The appendages on the

384   surface of *C. taeniosporum* endospores, therefore, represent distinct type of fibers than

385   those found on the surface of spores of species belonging to the *B. cereus* group.

386   Our structural studies uncover a novel class of pili, where subunits are organized into

387   helically wound fibers, held together by lateral β-sheet augmentation inside the helical

388   turns, and longitudinal disulfide cross-linking across helical turns. Covalent cross-linking in

389   pilus assembly is known for sortase-mediated isopeptide bond formation seen in Gram-

390   positive pili (Ton-That and Schneewind, 2004). In Enas, the cross-linking occurs through

391   disulfide bonding of a conserved Cys-Cys motif in the N-terminal connector of a subunit i, to

392   two single Cys residues in the core domain of the Ena subunits located at position i-9 and i-

393   10 in the helical structure. As such, the N-terminal connectors form a covalent bridge across

394   helical turns, as well as a branching interaction with two adjacent subunits in the preceding

395   helical turn (i.e. i-9 and i-10). The use of N-terminal connectors or extensions is also seen in

396   chaperone-usher pili and *Bacteroides* Type V pili, but these system employ a non-covalent

397   fold complementation mechanism to attain long-lived subunit-subunit contacts, and lack a

398   covalent stabilization (Sauer et al., 1999; Xu et al., 2016). Because in Ena the N-terminal

399   connectors are attached to the Ena core domain via a flexible linker, the helical turns in Ena

13

400 fibers have a large pivoting freedom and ability to undergo longitudinal stretching. These

401 interactions result in highly chemically stable fibers, yet with a large degree of flexibility.

402 Whether the stretchiness and flexibility of Enas carry a functional importance remains

403 unclear. Of note, in several chaperone-usher pili, a reversible spring-like stretching provided

404 by helical unwinding and rewinding of the pili has been found important to withstand shear

405 and pulling stresses exerted on adherent bacteria (Fallman et al., 2005; Miller et al., 2006).

406 Possibly, the longitudinal stretching seen in Ena may serve a similar role.

407 Typical Ena filaments have, to the best of our knowledge, never been observed on the

408 surface of vegetative *B. cereus* cells indicating that they are endospore-specific structures. In

409 support of that assumption, RT-qPCR analysis NVH 0075-95 demonstrated increased *ena1A-*

410 *C* transcript during sporulation, compared to vegetative cells. A transcriptional analysis of *B.*

411 *thuringiensis* serovar *chinensis* CT-43 at 7 h, 9 h, 13 h (30 % of cells undergoing sporulation)

412 and 22 hours of growth has previously been performed (Wang et al., 2013). It is difficult to

413 directly compare expression levels of *ena1A, B* and *C* in *B. cereus* NVH 0075-95 with the

414 expression level of *ena2A-C* in *B. thuringiensis* serovar *chinensis CT-43* (CT43_CH0783-785)

415 since the expression of the latter strain was normalized by converting the number of reads

416 per gene into RPKM (Reads Per Kilo bases per Million reads) and analyzed by DEGseq

417 software package, while the present study determines the expression level of the *ena* genes

418 relative to the house keeping gene *rpoB*. However, both studies indicate that *enaA* and *enaB*

419 are only transcribed during sporulation. By searching a separate set of published

420 transcriptomic profiling data we found that *ena2A-C* also are expressed in *B. anthracis*

421 during sporulation (Bergman et al., 2006), although Enas have not previously been reported

422 from *B. anthracis* spores.

423 Without knowledge on the function of Enas, we can only speculate about their

424 biological role. The Enas of *B. cereus* group species resemble pili, which in Gram-negative

425 and Gram-positive vegetative bacteria play roles in adherence to living surfaces (including

426 other bacteria) and non-living surfaces, twitching motility, biofilm formation, DNA uptake

427 (natural competence) and exchange (conjugation), secretion of exoproteins, electron

428 transfer (Geobacter) and bacteriophage susceptibility (Lukaszczyk et al., 2019; Proft and

429 Baker, 2009). Some bacteria express multiple types of pili that perform different functions.

430 The most common function of pili-fibers is adherence to a diverse range of surfaces from

14

431   metal, glass, plastics rocks to tissues of plants, animals or humans. In pathogenic bacteria,

432   pili often play a pivotal role in colonization of host tissues and function as important

433   virulence determinants. Similarly, it has been shown that appendages, expressed on the

434   surface of *C. sporogenes* endospores, facilitate their attachment to cultured fibroblast cells

435   (Panessa-Warren et al., 2007). The Enas are, however, not likely to be involved in active

436   motility or uptake/transport of DNA or proteins as they are energy demanding processes

437   that are not likely to occur in the endospore´s metabolically dormant state. Enas appear to

438   be a widespread feature among spores of strains belonging to the *B. cereus* group (Figure 6),

439   a group of closely related *Bacillus* species with a strong pathogenic potential (Ehling-Schulz

440   et al., 2019). For most *B. cereus* group species, the ingestion, inhalation or the

441   contamination of wounds with endospores forms a primary route of infection and disease

442   onset. Enas cover much of the cell surface so that they can be reasonably expected to form

443   an important contact region with the endospore environment and may play a role in the

444   dissemination and virulence of *B. cereus* species. Our phylogenetic analysis shows a

445   widespread occurrence of Enas in pathogenic *Bacilli*, and a striking absence in non-

446   pathogenic species such as *B. subtilis,* a soil-dwelling species and gastrointestinal

447   commensal that has functioned as the primary model system for studying endospores.

448   Ankolekar *et al.*, showed that all of 47 food isolates of *B. cereus* produced endospores with

449   appendages (Ankolekar and Labbe, 2010). Appendages were also found on spores of ten out

450   of twelve food-borne, enterotoxigenic isolates of *B. thuringiensis,* which is closely related to

451   *B. cereus,* and best known for its insecticidal activity (Ankolekar and Labbe, 2010).

452   The cryo-EM images of *ex vivo* fibers showed 2-3 nm wide fibers (ruffles) at the terminus of

453   S- and L-type Enas. The ruffles resemble tip fibrilla of P-pili and type 1 seen in many Gram-

454   negatives bacteria of the family Enterobacteriaceae (Proft and Baker, 2009). In Gram-

455   negative pilus filaments, the tip fibrilla provides adhesion proteins with a flexible location to

456   enhance the interaction with receptors on mucosal surfaces (Mulvey et al., 1998). No ruffles

457   were observed on the *in vitro* assembled fibers suggesting that their formation require

458   additional components than the Ena1B subunits.

459   We present the molecular identification of a novel class of spore-associated appendages or

460   pili widespread in pathogenic *Bacilli*. Future molecular and infection studies will need to

461   determine if and how Enas play a role in the virulence of spore-borne pathogenic *Bacilli*. The

462   advances in uncovering the genetic identity and the structural aspects of the Enas presented

463   in this work now enable *in vitro* and *in vivo* molecular studies to tease out their biological

464   role(s), and to gain insights into the basis for Ena heterogeneity amongst different *Bacillus*

465   species.

466

467   **Acknowledgements**

476   **Author contributions**

477   B.P and M.S. performed TEM imaging, structural studies, and recombinant ena1B

478   production and analysis. J.L. and T.L. produced endospores, performed TEM imaging,

479   isolated Enas and conducted genetic studies. A-K.L. and O.B. conducted the phylogenetic

480   analysis. H.R. and M.A. designed and supervised experiments, and wrote the paper, with

481   contributions from all authors.

482

483   **Figure legends**

484   **Figure 1. *B. cereus* endospores carry S and L-type Enas. (A, B)** negative stain TEM image of

485   *B. cereus* NVH 0075-95 endospore, showing spore body (SB), exosporium (E), and endospore

486   appendages (Ena), which emerge from the endospore individually or as fiber clusters

487   (boxed). At the distal end, Enas terminate in a single or multiple thin ruffles (R). **(C, D)** Single

488   fiber cryoTEM images and negative stain 2D class averages of S-type (C) and L-type Enas (D).

489   **(E)** Length distribution of S- and L-type Enas and number of Enas per endospore (inset),

490   (*n*=1023, from 150 endospores, from 5 batches). See also Figure S1.

491

492   **Figure 2. CryoTEM structure of S-type Enas. (A, B)** Representative 2D class average (A) and

493   corresponding power spectrum **(B)** of *B. cereus* NVH 0075-95 S-type Enas viewed by

494   cryoTEM. Bessel orders used to derive helical symmetry are indicated. **(C)** Reconstituted

495   cryoEM electron potential map of *ex vivo* S-type Ena (3.2 Å resolution). **(D)** Side and top

496   view of a single helical turn of the *de novo*-built 3D model of S-type Ena shown in ribbon

497   representation and molecular surface. Ena subunits are labelled i to i-10. **(F)** Ribbon

498   representation and topology diagram of the S-type Ena1B subunit (blue to red rainbow from

499   N- to C-terminus), and its interaction with subunits i-9 (sand) and i-10 (green) through

500   disulfide crosslinking.

501

502   **Figure 3. Ntc linkers give high flexibility and elasticity to S-type Enas. (A)** CryoTEM image of

503   an isolated S-type Ena making a U-turn comprising just 19 helical turns (shown schematically

504   in orange). **(B, C)** Cross-section and 3D cryoTEM electron potential map of the S-type Ena

505   model, highlighting the longitudinal spacing between Ena1B jellyroll domains as a result of

506   the Ntc linker (residues 12-17). **(D)** Negative stain 2D class averages of endospore-

507   associated S-type Enas show variation in pitch and axial curvature.

508

509   **Figure 4. *ena* is bicistronic and expressed during sporulation. (A)** Chromosomal

510   organization of the *ena* genes and primers used for transcript analysis (arrows). **(B)** Agarose

511   gel electrophoresis (1%) analysis of PCR products using indicated primer pairs and cDNA

512   made of mRNA isolated from NVH 0075-95 after 8 and 16 hours growth in liquid cultures or

513   genomic DNA as control. Of note, the transcription of *ena*1C was surprisingly higher than

514   *ena1A* and *ena1B*, which are components of the major appendages. **(C)** Transcription level

515   of *ena1A* (x), *ena1B* (▲), *ena1C* (▯) and *dedA* (▯) relative to *rpoB* determined by RT-qPCR

516   during 16▯hours of growth of *B. cereus* strain NVH 0075-95. The dotted line represents the

517   bacterial growth measured by increase in OD_{600}. Whiskers represent standard deviation of

518   three independent experiments.

519

520   **Figure 5. Composition of S- and L-type Ena. (A)** Representative negative stain images of

521   endospores of NVH 0075-95 mutants lacking *ena1A, ena1B, ena1A* and *B* or *ena1C*, as well

522   as the *ena1B* mutant complemented with *ena1A-ena1B* from plasmid (pAB). Inset are 2D

523   class averages of Enas observed on the respective mutants. **(B)** Length distribution and

524   number of Enas found on WT and mutant NVH 0075-95 endospores. Statistics: pair-wise

525   Mann-Whitney U tests against WT (*n*: ≥18 spores; *n*: ≥50 Enas; ns: not significant, * p<0.05,

526   ** p<0.01, *** p<0.001 and **** p<0.0001. ---: mean ± s.d.)

527

528   **Figure 6. Ena is widespread in pathogenic Bacilli. (A)** Ena1 and Ena2 loci with average

529   amino acid sequence identity indicated between the population of EnaA-C ortho- and

530   homologs. Ena1C shows considerably more variation and is in *B. cytotoxicus* different from

531   both Ena1C and Ena2C (see Figure S5C), while other genomes have *enaC* present at different

532   loci (applies to two isolates of *B. mycoides*). **(B)** Distribution of *ena1/2A-C* among *Bacillus*

533   species. Whole genome clustering of the *B. cereus s.l.* group and *B. subtilis* created by

534   Mashtree (Katz et al., 2019; Ondov et al., 2016) and visualized in Microreact (Argimon et al.,

535   2016). Rooted on *B. subtilis*. Traits for species (colored nodes), Bazinet clades and presence

536   of *ena* are indicated on surrounding four rings in the following order from inner to outer:

537   clade according to Bazinet 2017 (when available) (Bazinet, 2017) (see legend), and presence

538   of *enaA, enaB* and *enaC* (for all three, *ena1*: teal, *ena2*: orange, different locus: cyan).

539   When no homo- or ortholog was found, the ring is grey. Ena1A-C and Ena2A-C are defined

540   as present when a homologous protein of the corresponding genome has high coverage

541   (>90%) and >80% and 50-65% sequence identity, respectively, with Ena1A-C of the NMH

542   0095/75   strain   (see   Table   S4).   Interactive   tree   accessible   at

543   https://microreact.org/project/vn2oWw7zM3cwejEFNoRGWA/0024f86c

544

545   **METHODS**

546   **Culture of *B. cereus* and appendages extraction**

18

547   For extraction of Enas the *B. cereus* strain NVH 0075-95 was plated on blood agar plates and
548   incubated at 37 °C for 3 months. Upon maturation, the spores were resuspended and
549   washed in milli-Q water three times (centrifugation 2,400 g at 4 °C). To get rid of various
550   organic and inorganic debris, the pellet was then resuspended in 20% Nycodenz (Axis-
551   Shield) and subjected to Nycodenz density gradient centrifugation where the gradient was
552   composed of a mixture of 45% and 47% (w/v) Nycodenz in 1:1 v/v ratio. The pellet
553   consisting only of the spore cells was then washed with 1 M NaCl and TE buffer (50 mM Tris-
554   HCl; 0.5 mM EDTA) containing 0.1% SDS respectively. To detach the appendages, the
555   washed spores were sonicated at 20k Hz ± 50 Hz and 50 watts (Vibra Cell VC50T; Sonic &
556   Materials Inc.; U.S.) for 30 seconds on ice followed by centrifugation at 4500 g and
557   appendages were collected in the supernatant. To further get rid of the residual
558   components of spore and vegetative mother cells n-hexane was added and vigorously mixed
559   with the supernatant in 1:2 v/v ratio. The mixture was then left to settle to allow phase
560   separation of water and hexane. The hexane fraction containing the appendages was then
561   collected and kept at 55 °C under pressured air for 1.5 hours to evaporate the hexane. The
562   appendages were finally resuspended in mill-Q water for further cryo-EM sample
563   preparation.

564   **Recombinant expression, purification and *in vitro* assembly of Ena1B appendages**

565   A synthetic open reading frame encoding Ena1B was codon optimized for recombinant
566   expression in *E. coli*, synthesized and cloned into pET28a expression vector at Twist
567   biosciences (Table S2). The insert was designed to have a N-terminal 6X histidine tag on
568   Ena1B along with a TEV protease cleavage site (ENLYFQG) in between. Large scale
569   recombinant expression was carried out in the T7 Express lysY/Iq *E. coli* strain from NEB. A
570   single colony was inoculated into 20 mL of LB and grown at 37 °C with shaking at 150 rpm
571   overnight for primary culture. Next morning 6 L of LB was inoculated with 20 mL/L of
572   primary culture and grown at 37 °C with shaking until the OD_{600} reached 0.8 after which
573   protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG).
574   The culture was incubated for a further 3 hours at 37 °C and harvested by centrifugation at
575   5,000 rpm. The whole-cell pellet was resuspended in lysis buffer (20 mM potassium
576   phosphate, 500 mM NaCl, 10 mM β -mercaptoethanol, 20 mM imidazole, pH 7.5) and
577   sonicated on ice for lysis. The lysate was centrifuged to separate the soluble and insoluble

578   fractions by centrifugation at 18,000 rpm for 45 min in a JA-20 rotor from Beckman Coulter.

579   The pellet was further dissolved in denaturing lysis buffer consisting 8M urea in lysis buffer.

580   The dissolved pellet was then passed over a HisTrap HP column (GE Healthcare) and

581   equilibrated with denaturing lysis buffer. The bound protein was eluted from the column

582   with elution buffer (20 mM potassium phosphate, pH 7.5, 8 M urea, 250 mM imidazole) in a

583   gradient mode (20-250 mM imidazole) at room temperature.

584   For *in vitro* Ena1B assembly, purified His-Ena1B in denaturing conditions was first

585   dialyzed against a buffer containing 20 mM Hepes, pH 7.0, 50 mM NaCl overnight at 4 °C. To

586   facilitate Ena1B self-assembly into S-type Ena filaments the 6xHis-tag was cleaved off by TEV

587   protease. TEV protease along with 100 mM β-mercaptoethanol was then added in

588   equimolar ratio and incubated for 2 hours at 37 °C. Removal of the 6xHis-tag led to the

589   assembly of the *rec*Ena1B into long Ena-like filaments Figure S2F.

590

591

592   **Ena treatment experiments to test its robustness**

593   *Ex vivo* Enas extracted from *B. cereus* strain NVH 0075-95 (see above) were resuspended in

594   deionized water, autoclaved at 121 °C for 20 minutes to ensure inactivation of residual

595   bacteria or spores, and subjected to treatment with buffer or as indicated below and shown

596   in Figure S1. To determine Ena integrity upon the various treatments, samples were imaged

597   using negative stain TEM and Enas were boxed and subjected to 2D classification as

598   described below. To test protease resistance, *ex vivo* Ena were subjected to 1 mg/mL Ready-

599   to-use Proteinase K digestion (Thermo Scientific) for 4 hours at 37 °C and imaged by TEM. To

600   study the effects of desiccation on the appendages, *ex vivo* Ena were vacuum dried at 43 °C

601   using Savant DNA120 Speedvac Concentrator (Thermo scientific) run for 2 hours at a speed

602   of 2k RPM.

603

604   **Negative-Stain Transmission Electron Microscopy (TEM)**

605   For visualization of spores and recombinantly expressed appendages by negative stain TEM,

606   formvar/carbon coated copper grids with 400-hole mesh (Electron Microscopy Sciences)

607   were glow discharged (ELMO; Agar Scientific) with a plasma current of 4 mA at vacuum for

608   45 seconds. 3 μL of sample was applied on the grids and allowed to bind to the support film

609   for 1 minute after which the excess liquid was blotted away with Whatman grade 1 filter

610   paper. The grids were then washed three times using three 15 μL drops of milli-Q followed

611   by blotting of excess liquid. The washed grids were held in 15 μL drops of 2% uranyl acetate

612   three times for, respectively, 10 seconds, 2 seconds and 1 minute durations, with a blotting

613   step in between each dip. Finally, the uranyl acetate coated grids were blotted until dry. The

614   grids were then imaged using a 120 kV JEOL 1400 microscope equipped with LaB6 filament

615   and TVIPS F416 CCD camera. 2D classes of the appendages were generated in RELION 3.0

616   (Zivanov et al., 2018) as described below.

617

618   **Preparation of cryo-TEM grids and cryo-EM data collection.**

619   QUANTIFOIL® holey Cu 400 mesh grids with 2 μm holes and 1 μm spacing were first glow

620   discharged in vacuum using plasma current of 5 mA for 1 minute (ELMO; Agar Scientific). 3

621   μL of 0.6 mg /mL graphene oxide (GO) solution was applied onto the grid and incubated 1

622   minute for absorption at room temperature. Extra GO was then blotted out and left for

623   drying using a Whatman grade 1 filter paper. For cryo-plunging, 3 μL of protein sample was

624   applied on the GO coated grids at 100% humidity and room temperature in a Gatan CP3

625   cryo-plunger. After 1 minute of absorption it was machine-blotted with Whatman grade 2

626   filter paper for 5 seconds from both sides and plunge frozen into liquid ethane at 180 °K.

627   Grids were then stored in liquid nitrogen until data collection. Two datasets were collected

628   for *ex vivo* and *rec*Ena1B appendages with slight changes in the collection parameters. High

629   resolution cryo-EM 2D micrograph movies were recorded on a JEOL Cryoarm300 microscope

630   automated with energy filter and a K2 or K3 direct electron detector run in counting mode.

631   For the *ex vivo* Ena, the microscope was equipped with a K2 summit detector and had the

632   following settings: 300 keV, 100 mm aperture, 30 frames / image, 62.5 e$^-$/Å^2, 2.315 seconds

633   exposure, and 0.82 Å/pxl. For the *rec*Ena1B dataset was recorded on a K3 detector, at a

634   pixel size of 0.782 Å/pxl, and an exposure of 64.66 e$^-$/Å^2 taken over 61 frames / image.

635   **Image processing**

21

636   MOTIONCORR2 (Zheng et al., 2017) implemented in RELION 3.0 (Zivanov et al., 2018) was used to correct for beam-induced image motion and averaged 2D micrographs were generated. The motion-corrected micrographs were used to estimate the CTF parameters using CTFFIND4.2 (Rohou and Grigorieff, 2015) integrated in RELION 3.0. Subsequent processing used RELION 3.0. and SPRING (Desfosses et al., 2014). For both the datasets, the coordinates of the appendages were boxed manually using *e2helixboxer* from the EMAN2 package (Tang et al., 2007). Special care was taken to select micrographs with good ice and straight stretches of Ena filaments. The filaments were segmented into overlapping single-particle boxes of dimension 300 x 300 pxl with an inter-box distance of 21 Å. For the *ex vivo* Enas a total of 53,501 helical fragments was extracted from 580 micrographs with an average of 2 - 3 long filaments per micrograph. For the *rec*Ena1B filaments, 100,495 helical fragments were extracted from 3,000 micrographs with an average of 4 - 5 filaments per micrograph. To filter out bad particles multiples rounds of 2D classification were run in RELION 3.0. After several rounds of filtering, a dataset of 42,822 and 65,466 good particles of the *ex vivo* and *rec*Ena1B appendages were selected, respectively.

        After running ~50 iterations of 2D classification well-resolved 2D class averages could be obtained. *segclassexam* of the SPRING package (Desfosses et al., 2014) was used to generate B-factor enhanced power spectrum of the 2D class averages . The generated power spectrum had an amplified signal-to-noise ratio with well resolved layer lines (Figure 2B). To estimate crude helical parameters, coordinates and phases of the peaks in the layer lines were measured using the *segclasslayer* option in SPRING. Based on the measured distances and phases possible sets of Bessel orders were deduced, after which the calculated helical parameters were used in a helical reconstruction procedure in RELION (He and Scheres, 2017). A featureless cylinder of 110 Å diameter generated using *relion_helix_toolbox* was used as an initial model for 3D classification. Input rise and twist deduced from Fourier – Bessel indexing were varied in the range of 3.05 – 3.65 Å and 29 – 35 degrees, respectively, with a sampling resolution of 0.1 Å and 1 degree between tested start values. So doing, several rounds of 3D classification were run until electron potential maps with good connectivity and recognizable secondary structure were obtained. The output translational information from the 3D classification was used to re-extract particles and 3D refinement was done taking a 25 Å low pass filtered map generated from the 3D

667 classification run. To improve the resolution of the EM maps multiple rounds of 3D
668 refinement were run. To further improve the resolution Bayesian polishing was performed
669 in RELION. Finally, a solvent mask covering the central 50% of the helix *z*-axis was generated
670 in *maskcreate* and used for postprocessing and calculating the solvent-flattened Fourier
671 shell correlation (FSC) curve in RELION. After two rounds of polishing, maps of 3.2 Å
672 resolution according to the $FSC_{0.143}$ gold-standard criterion as well as local resolution
673 calculated in RELION were obtained (Figure S3A).

674

675 **Model building**

676 Prior to model building unfiltered maps for recEna1B calculated by Relion were masked
677 down to three helical turns and used for cryo-EM density modification as implemented in
678 ResolveCryoEM (Terwilliger et al., 2019) from the PHENIX package (Afonine et al., 2018),
679 resulting in a map of 3.05 Å final resolution ($FSC_{0.143}$ criterium) for *rec*Ena1B. At first the
680 primary skeleton for a single asymmetric subunit from the density modified map was
681 generated in Coot (Emsley et al., 2010). Primary sequence of Ena1B was manually threaded
682 onto the asymmetric unit and fitted into the map taking into consideration the chemical
683 properties of the residues. The SSM superpose tool in coot was used to place the additional
684 protomers of the S-type Ena from a single subunit. The built model was then subjected to
685 multiple rounds real space structural refinement in PHENIX, each residue was manually
686 inspected after every round of refinement. Model validation was done in Molprobity (Davis
687 et al., 2007) implemented in Phenix. All the visualizations and images for figures were
688 generated in ChimeraX (Goddard et al., 2018), Chimera (Pettersen et al., 2004), or Pymol.

689

690 **Immuno-labelling of the Enas**
691 For antibody generation, *rec*Ena1A and *rec*Ena1C were cloned, expressed and purified according to
692 the method described above for *rec*Ena1B. Aliquots of purified *rec*Ena1A, *rec*Ena1B and *rec*Ena1C
693 were concentrated to 1 mg/mL in PBS for rabbit immunization (Davids Biotechnologie GmbH). For
694 immunostaining EM imaging, 3 μL aliquots of purified *ex vivo* Enas were deposited on
695 Formvr/Carbon grids (400 Mesh, Cu; Electron Microscopy Sciences), washed with 20 μL 1x PBS, and

696   incubated for 1 hour with 0.5% (w/v) BSA in 1x PBS. After additional washing with 1x PBS, separate

697   grids were individually incubated for 2 hours at 37 °C with 1:1000 dilutions in PBS of anti-Ena1A,

698   anti-Ena1B or anti-Ena1C sera, respectively. Following washing with 1x PBS, grids were incubated

699   for 1 hour 37 °C with a 1:2000 dilution of a 10 nm gold-labelled goat anti-rabbit IgG (G7277; Sigma

700   Aldrich), washed with 1x PBS, and negative stained and imaged on a 120 kV JEOL 1400 microscope

701   as described above.

702

703   **Quantitative RT-PCR**

704   RT-qPCR experiments were performed on isolated mRNA from *B. cereus* cultures harvested

705   from three independent Bacto media cultures (37 °C, 150 rpm) at 4, 8, 12 and 16hours post-

706   inoculation. RNA extraction, cDNA synthesis and RT-qPCR analysis was performed essentially

707   as described before (Madslien et al., 2014), with the following changes: pre-heated (65 °C)

708   TRIzol Reagent (Invitrogen) and bead beating 4 times for 2 minutes in a Mini-BeadBeater-8

709   (BioSpec) with cooling on ice in between.

710   Each RT-qPCR of the RNA samples was performed in triplicate, no template was added in

711   negative controls, and *rpoB* was used as internal control. Slopes of the standard curves and

712   PCR efficiency (E) for each primer pair were estimated by amplifying serial dilutions of the

713   cDNA template. For quantification of mRNA transcript levels, Ct (threshold cycle) values of

714   the target genes and the internal control gene (*rpoB*) derived from the same sample in each

715   RT-qPCR reaction were first transformed using the term E^{-Ct}. The expression levels of target

716   genes were then normalized by dividing their transformed Ct-values by the corresponding

717   values obtained for the internal control gene (Duodu et al., 2010; Madslien et al., 2014;

718   Pfaffl, 2001). The amplification was conducted by using StepOne PCR software V.2.0

719   (Applied Biosystems) with the following conditions: 50 °C for 2 minutes, 95 °C for 2 minutes,

720   40 cycles of 15 seconds at 95 °C, 1 minute at 60 °C and 15 seconds at 95 °C. All primers used

721   for RT-qPCR analyses are listed in Table S2.

722   Regular PCR reactions were performed on cDNA to confirm that *ena1A* and *ena1B* were

723   expressed as an operon using the primers 2180/2177 and 2176/2175 and DreamTaq DNA

724   polymerase (Thermo Fisher) amplified in an Eppendorf Mastercycler using the following

725   program: 95°C for 2 minutes, 30 cycles of 95°C for 30 seconds, 54°C for 30 seconds, and
726   72°C for 1 minute.

Construction of deletion mutants

728   The *B. cereus* strain NVH 0075-95 was used as background for gene deletion mutants. The
729   e*na1B* gene was deleted in-frame by replacing the reading frames with ATGTAA (5´–3´)
730   using a markerless gene replacement method (Janes and Stibitz, 2006) with minor
731   modifications. To create the deletion mutants the regions upstream (primer A and B, Table
732   S2) and downstream (primer C and D, Table S2) of the target *ena* genes were amplified by
733   PCR. To allow assembly of the PCR fragments, primers B and C contained complementary
734   overlapping sequences. An additional PCR step was then performed, using the upstream and
735   downstream PCR fragments as template and the A and D primer pair (Table S2). All PCR
736   reactions were conducted using an Eppendorf Mastercycler gradient and high fidelity
737   AccuPrime Taq DNA Polymerase (ThermoFisher Scientific) according to the manufacturer's
738   instructions. The final amplicons were cloned into the thermosensitive shuttle vector pMAD
739   (Arnaud et al., 2004) containing an additional I-SceI site as previously described (Lindback et
740   al., 2012). The pMAD-I-*Sce*I plasmid constructs were passed through One Shot™ INV110 *E.*
741   *coli* (ThermoFisher Scientific) to achieve unmethylated DNA to enhance the transformation
742   efficiency in *B. cereus*. The unmethylated plasmid were introduced into *B. cereus* NVH 0075-
743   95 by electroporation (Mahillon et al., 1989). After verification of transformants by PCR, the
744   plasmid pBKJ233 (unmethylated), containing the gene for the I-*Sce*I enzyme, was introduced
745   into the transformant strains by electroporation. The I-*Sce*I enzyme makes a double-
746   stranded DNA break in the chromosomally integrated plasmid. Subsequently, homologous
747   recombination events lead to excision of the integrated plasmid resulting in the desired
748   genetic replacement. The gene deletions were verified by PCR amplification using primers A
749   and D (Table S2) and DNA sequencing (Eurofins Genomics).

Search for orthologs and homologs of Ena1

751   Publicly available genomes of species belonging to the *Bacillus s.l.* group (Table S3 and S4)
752   was   downloaded   from   NCBI   RefSeq   database   (*n=735*,   NCB
753   (https://www.ncbi.nlm.nih.gov/refseq/, Table S1). Except for strains of particular interest

25

754　due to phenotypic characteristics (GCA_000171035.2_ASM17103v2,

755　GCA_002952815.1_ASM295281v1, GCF_000290995.1_Baci_cere_AND1407_G13175) and

756　species of which closed genomes were nonexistent or very scarce (*n=158*), all assemblies

757　included were closed and publicly available genomes from the curated database of NCBI

758　RefSeq database. Assemblies were quality checked using QUAST (Gurevich et al., 2013), and

759　only genomes of correct size (~4.9-6Mb) and a GC content of ~35% were included in the

760　downstream analysis. Pairwise tBLASTn searches were performed (e-value 1e-10, max_hspr

761　1, default settings) to search for homo- and orthologs of the following query-protein

762　sequences from strain NVH 0075-95: Ena1A, Ena1B, Ena1C (Table S3). The Ena1B protein

763　sequences used as query originated from an inhouse amplicon sequenced product, while

764　the Ena1A and Ena1C protein sequence queries originated from the assembly for strain NVH

765　0075-95 (Accession number GCF_001044825.1, protein KMP91698.1 and KMP91699.1,

766　Table S5). We considered proteins orthologs or homologs when a subject protein matched

767　the query protein with high coverage (>70%) and moderate sequence identity (>30%).

768　**Comparative genomics of the ena-genes and proteins**

769　Phylogenetic trees of the aligned Ena1A-C proteins were constructed using approximately

770　maximum likelihood by FastTree (Price et al., 2010) (default settings) for all hits resulting

771　from the tBLASTn search. The amino acid sequences were aligned using mafft v.7.310 (Katoh

772　et al., 2019), and approximately-maximum-likelihood phylogenetic trees of protein

773　alignments were made using FastTree, using the JTT+CAT model (Price et al., 2010). All Trees

774　were visualized in Microreact (Argimon et al., 2016) and the metadata of species, and

775　presence and absence for Ena1A-C and Ena2A-C overlaid the figures.

776　**References**

777　Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and

778　Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. Acta

779　Crystallogr D Struct Biol *74*, 531-544.

780　Ankolekar, C., and Labbe, R.G. (2010). Physical characteristics of spores of food-associated

781　isolates of the Bacillus cereus group. Appl Environ Microbiol *76*, 982-984.

782   Argimon, S., Abudahab, K., Goater, R.J.E., Fedosejev, A., Bhai, J., Glasner, C., Feil, E.J.,
783   Holden, M.T.G., Yeats, C.A., Grundmann, H., *et al.* (2016). Microreact: visualizing and sharing
784   data for genomic epidemiology and phylogeography. Microb Genom *2*, e000093.
785   Arnaud, M., Chastanet, A., and Debarbouille, M. (2004). New vector for efficient allelic
786   replacement in naturally nontransformable, low-GC-content, gram-positive bacteria. Appl
787   Environ Microbiol *70*, 6887-6891.
788   Atrih, A., and Foster, S.J. (1999). The role of peptidoglycan structure and structural dynamics
789   during endospore dormancy and germination. Antonie Van Leeuwenhoek *75*, 299-307.
790   Bazinet, A.L. (2017). Pan-genome and phylogeny of Bacillus cereus sensu lato. BMC Evol Biol
791   *17*, 176.
792   Bergman, N.H., Anderson, E.C., Swenson, E.E., Niemeyer, M.M., Miyoshi, A.D., and Hanna,
793   P.C. (2006). Transcriptional profiling of the Bacillus anthracis life cycle in vitro and an implied
794   model for regulation of spore formation. J Bacteriol *188*, 6092-6100.
795   Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W.,
796   Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S., *et al.* (2007). MolProbity: all-atom contacts
797   and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375-383.
798   Desfosses, A., Ciuffa, R., Gutsche, I., and Sachse, C. (2014). SPRING - an image processing
799   package for single-particle based helical reconstruction from electron cryomicrographs. J
800   Struct Biol *185*, 15-26.
801   DesRosier, J.P., and Lara, J.C. (1981). Isolation and properties of pili from spores of Bacillus
802   cereus. J Bacteriol *145*, 613-619.
803   Driks, A. (2007). Surface appendages of bacterial spores. Mol Microbiol *63*, 623-625.
804   Duodu, S., Holst-Jensen, A., Skjerdal, T., Cappelier, J.M., Pilet, M.F., and Loncarevic, S.
805   (2010). Influence of storage temperature on gene expression and virulence potential of
806   Listeria monocytogenes strains grown in a salmon matrix. Food Microbiol *27*, 795-801.
807   Ehling-Schulz, M., Lereclus, D., and Koehler, T.M. (2019). The Bacillus cereus Group: Bacillus
808   Species with Pathogenic Potential. Microbiol Spectr *7*.
809   Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of
810   Coot. Acta crystallographica Section D, Biological crystallography *66*, 486-501.
811   Fallman, E., Schedin, S., Jass, J., Uhlin, B.E., and Axner, O. (2005). The unfolding of the P pili
812   quaternary structure by stretching is reversible, not plastic. EMBO Rep *6*, 52-56.

813   Gerhardt, P., and Ribi, E. (1964). Ultrastructure of the Exosporium Enveloping Spores of
814   Bacillus Cereus. Journal of bacteriology *88*, 1774-1789.
815   Goddard, T.D., Huang, C.C., Meng, E.C., Pettersen, E.F., Couch, G.S., Morris, J.H., and Ferrin,
816   T.E. (2018). UCSF ChimeraX: Meeting modern challenges in visualization and analysis.
817   Protein Sci *27*, 14-25.
818   Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool
819   for genome assemblies. Bioinformatics *29*, 1072-1075.
820   Hachisuka, Y., and Kuno, T. (1976). Filamentous appendages of Bacillus cereus spores. Jpn J
821   Microbiol *20*, 555-558.
822   He, S., and Scheres, S.H.W. (2017). Helical reconstruction in RELION. J Struct Biol *198*, 163-
823   176.
824   Hodgikiss, W. (1971). Filamentous appendages on the spores and exosporium of certain
825   Bacillus species. In Spore research, A.N. Barker, G.W. Gould, and J. Wolf, eds. (London and
826   New York: Academic Press), pp. 211-218.
827   Janes, B.K., and Stibitz, S. (2006). Routine markerless gene replacement in Bacillus anthracis.
828   Infect Immun *74*, 1949-1953.
829   Katoh, K., Rozewicki, J., and Yamada, K.D. (2019). MAFFT online service: multiple sequence
830   alignment, interactive sequence choice and visualization. Brief Bioinform *20*, 1160-1166.
831   Katz, L.S., Griswold, T., Morrison, S.S., Caravas, J.A., Zhang, S., C., d.B.H., Deng, X., and
832   Carleton, A. (2019). Mashtree: a rapid comparison of whole genome sequence

833   files. Journal of Open Source Software *4*.
834   Lindback, T., Mols, M., Basset, C., Granum, P.E., Kuipers, O.P., and Kovacs, A.T. (2012). CodY,
835   a pleiotropic regulator, influences multicellular behaviour and efficient production of
836   virulence factors in Bacillus cereus. Environ Microbiol *14*, 2233-2246.
837   Lukaszczyk, M., Pradhan, B., and Remaut, H. (2019). The Biosynthesis and Structures of
838   Bacterial Pili. Subcell Biochem *92*, 369-413.
839   Madslien, E.H., Granum, P.E., Blatny, J.M., and Lindback, T. (2014). L-alanine-induced
840   germination in Bacillus licheniformis -the impact of native gerA sequences. BMC Microbiol
841   *14*, 101.

842   Mahillon, J., Chungjatupornchai, W., Decock, J., Dierickx, S., Michiels, F., Peferoen, M., and
843   Joos, H. (1989). Transformation of Bacillus thuringiensis by electroporation. FEMS
844   Microbiology Letters *60*, 205-210.

845   Mandlik, A., Swierczynski, A., Das, A., and Ton-That, H. (2008). Pili in Gram-positive bacteria:
846   assembly, involvement in colonization and biofilm development. Trends Microbiol *16*, 33-
847   40.

848   Melville, S., and Craig, L. (2013). Type IV pili in Gram-positive bacteria. Microbiol Mol Biol
849   Rev *77*, 323-341.

850   Miller, E., Garcia, T., Hultgren, S., and Oberhauser, A.F. (2006). The mechanical properties of
851   E. coli type 1 pili measured by atomic force microscopy techniques. Biophys J *91*, 3848-3856.

852   Mulvey, M.A., Lopez-Boado, Y.S., Wilson, C.L., Roth, R., Parks, W.C., Heuser, J., and Hultgren,
853   S.J. (1998). Induction and evasion of host defenses by type 1-piliated uropathogenic
854   Escherichia coli. Science *282*, 1494-1497.

855   Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., and
856   Phillippy, A.M. (2016). Mash: fast genome and metagenome distance estimation using
857   MinHash. Genome Biol *17*, 132.

858   Panessa-Warren, B.J., Tortora, G.T., and Warren, J.B. (2007). High resolution FESEM and
859   TEM reveal bacterial spore attachment. Microsc Microanal *13*, 251-266.

860   Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and
861   Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and
862   analysis. J Comput Chem *25*, 1605-1612.

863   Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-
864   PCR. Nucleic Acids Res *29*, e45.

865   Price, M.N., Dehal, P.S., and Arkin, A.P. (2010). FastTree 2--approximately maximum-
866   likelihood trees for large alignments. PLoS One *5*, e9490.

867   Proft, T., and Baker, E.N. (2009). Pili in Gram-negative and Gram-positive bacteria -
868   structure, assembly and their role in disease. Cell Mol Life Sci *66*, 613-635.

869   Remaut, H., and Waksman, G. (2006). Protein-protein interaction through beta-strand
870   addition. Trends Biochem Sci *31*, 436-444.

871   Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. Adv Protein Chem
872   *34*, 167-339.

873 Rode, L.J., Pope, L., Filip, C., and Smith, L.D. (1971). Spore appendages and taxonomy of
874 Clostridium sordellii. Journal of bacteriology *108*, 1384-1389.

875 Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from
876 electron micrographs. J Struct Biol *192*, 216-221.

877 Sauer, F.G., Futterer, K., Pinkner, J.S., Dodson, K.W., Hultgren, S.J., and Waksman, G. (1999).
878 Structural basis of chaperone function and pilus biogenesis. Science *285*, 1058-1061.

879 Setlow, P. (2014). Germination of spores of Bacillus species: what we know and do not
880 know. Journal of bacteriology *196*, 1297-1305.

881 Smirnova, T.A., Zubasheva, M.V., Shevliagina, N.V., Nikolaenko, M.A., and Azizbekian, R.R.
882 (2013). [Electron microscopy of the surfaces of bacillary spores]. Mikrobiologiia *82*, 698-706.

883 Stewart, G.C. (2015). The Exosporium Layer of Bacterial Spores: a Connection to the
884 Environment and the Infected Host. Microbiol Mol Biol Rev *79*, 437-457.

885 Tang, G., Peng, L., Baldwin, P.R., Mann, D.S., Jiang, W., Rees, I., and Ludtke, S.J. (2007).
886 EMAN2: an extensible image processing suite for electron microscopy. J Struct Biol *157*, 38-
887 46.

888 Terwilliger, T.C., Ludtke, S.J., Read, R.J., Adams, P.D., and Afonine, P.V. (2019). Improvement
889 of cryo-EM maps by density modification. bioRxiv.

890 Ton-That, H., and Schneewind, O. (2004). Assembly of pili in Gram-positive bacteria. Trends
891 Microbiol *12*, 228-234.

892 Walker, J.R., Gnanam, A.J., Blinkova, A.L., Hermandson, M.J., Karymov, M.A., Lyubchenko,
893 Y.L., Graves, P.R., Haystead, T.A., and Linse, K.D. (2007). Clostridium taeniosporum spore
894 ribbon-like appendage structure, composition and genes. Mol Microbiol *63*, 629-643.

895 Wang, J., Mei, H., Zheng, C., Qian, H., Cui, C., Fu, Y., Su, J., Liu, Z., Yu, Z., and He, J. (2013).
896 The metabolic regulation of sporulation and parasporal crystal formation in Bacillus
897 thuringiensis revealed by transcriptomics and proteomics. Mol Cell Proteomics *12*, 1363-
898 1376.

899 Xu, Q., Shoji, M., Shibata, S., Naito, M., Sato, K., Elsliger, M.A., Grant, J.C., Axelrod, H.L., Chiu,
900 H.J., Farr, C.L., *et al.* (2016). A Distinct Type of Pilus from the Human Microbiome. Cell *165*,
901 690-703.

902 Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017).
903 MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron
904 microscopy. Nat Methods *14*, 331-332.

905   Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres,

906   S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in
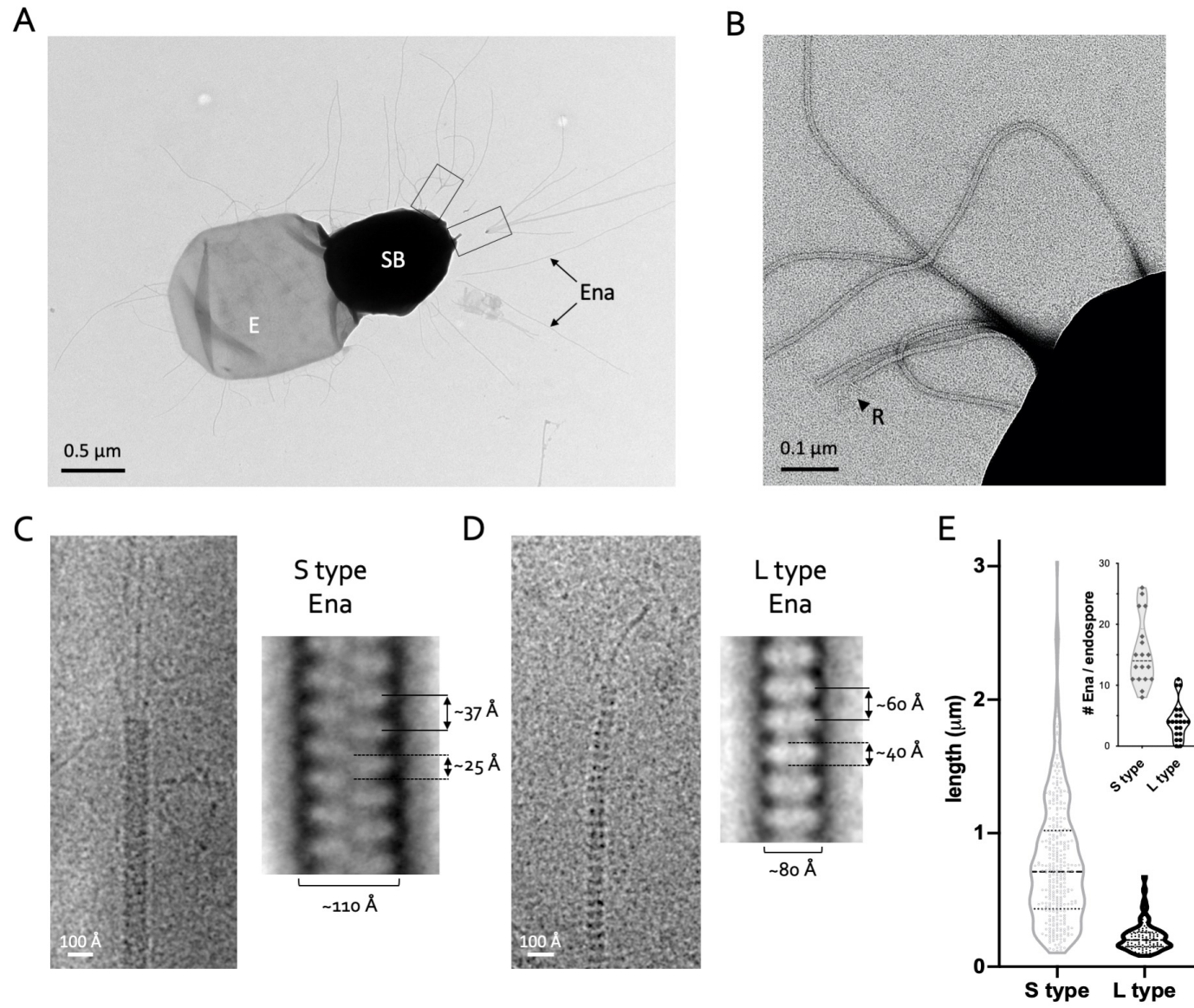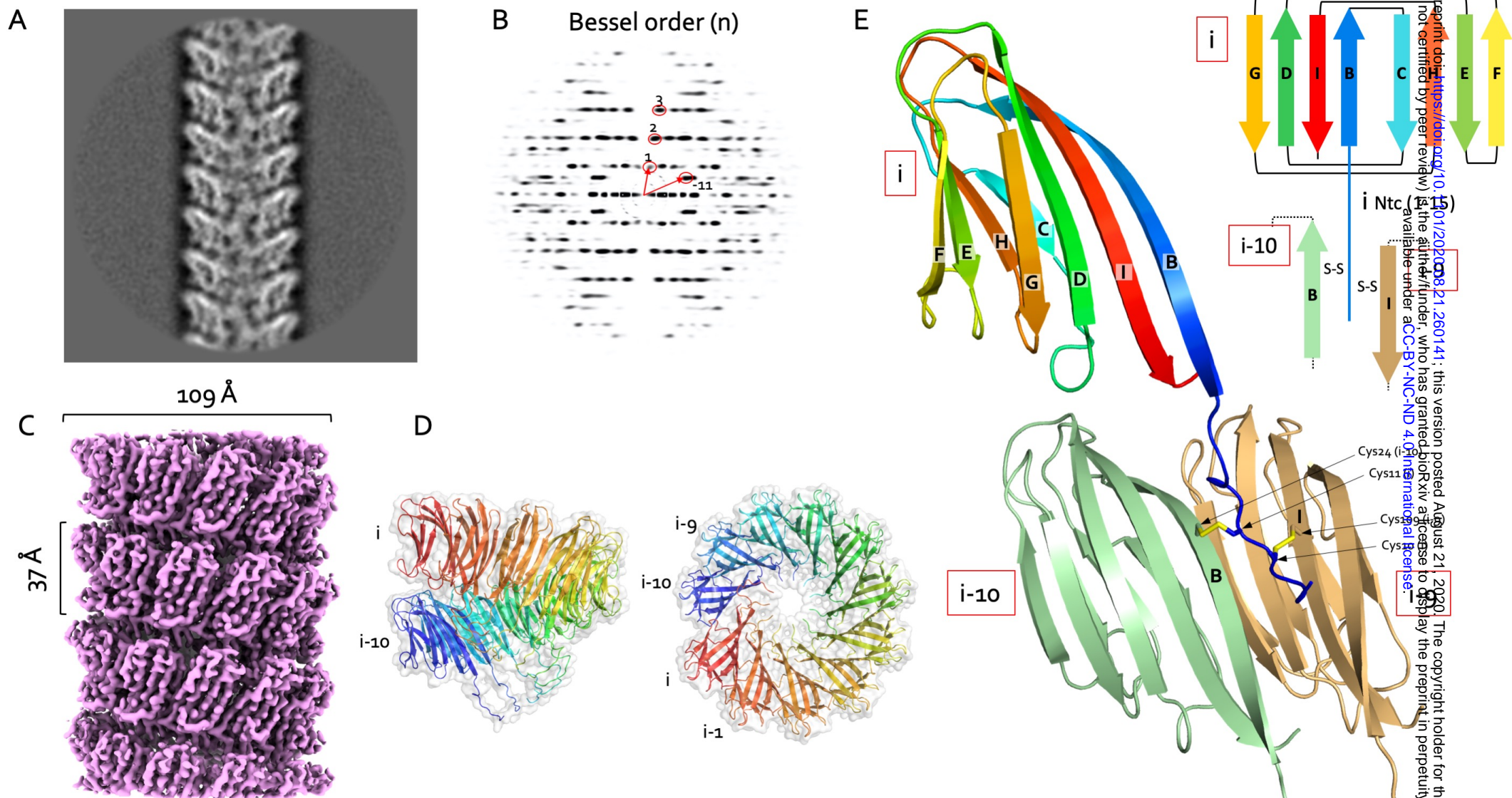
907   RELION-3. Elife 7.

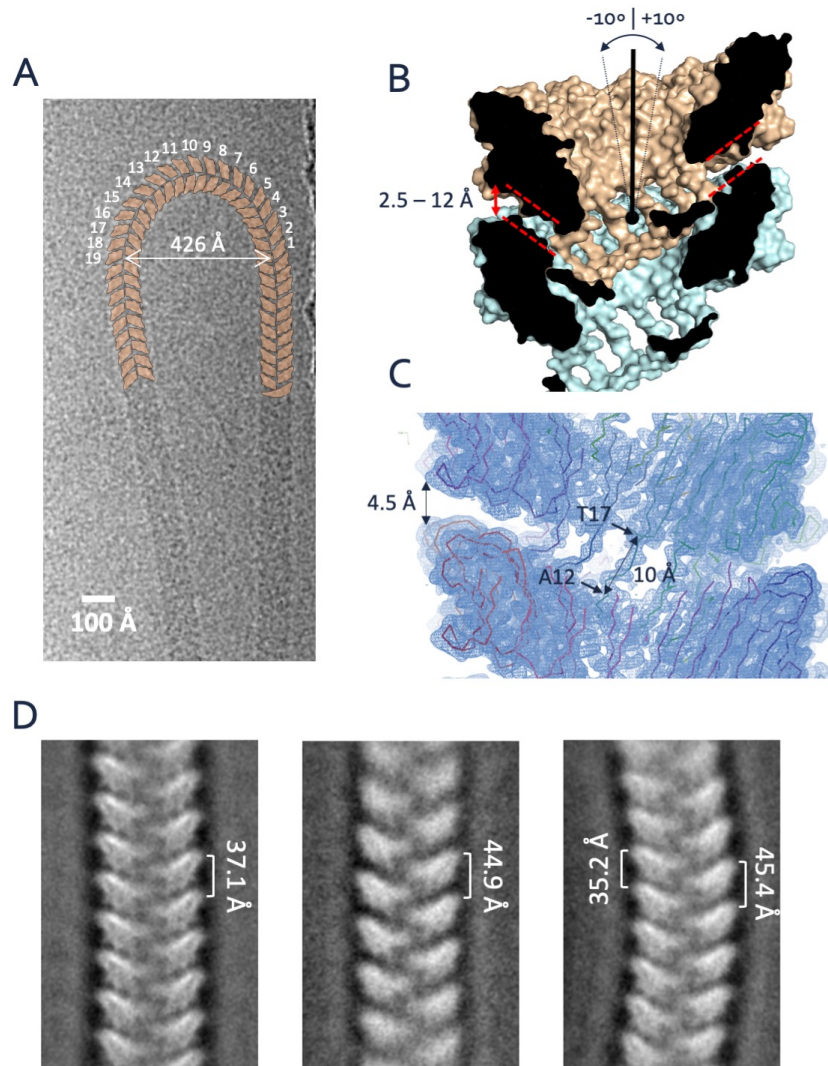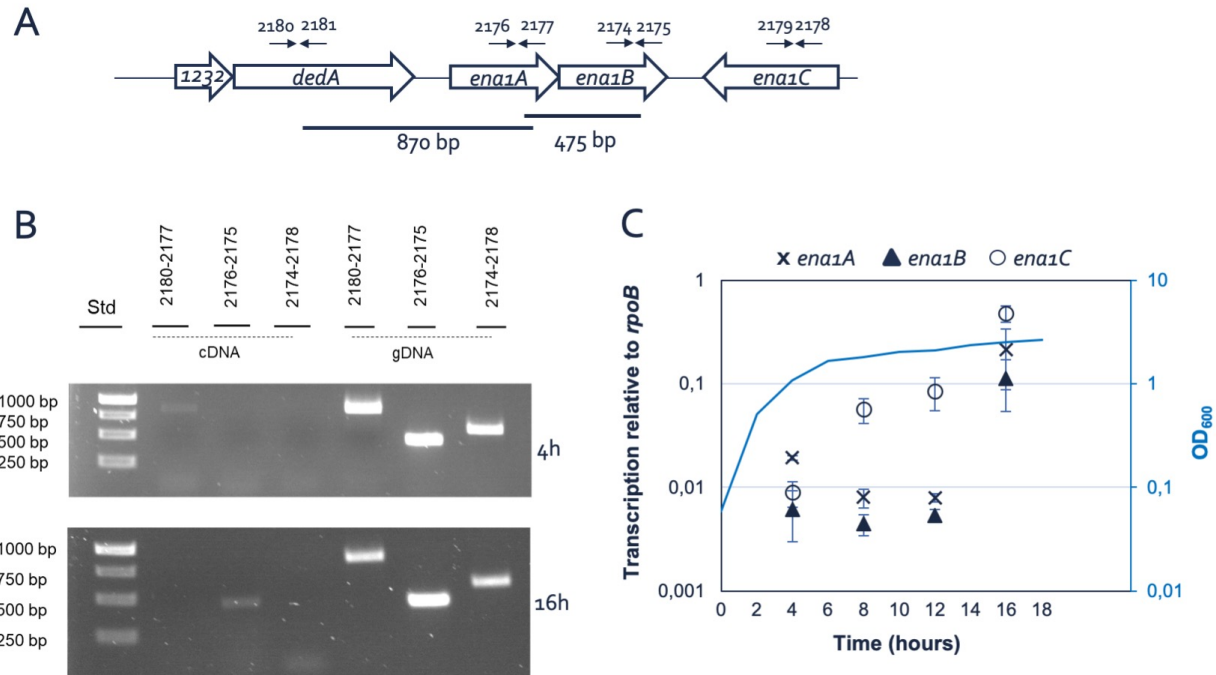908

Figure 1

Figure 2

Figure 3

Figure 4

A

B

C
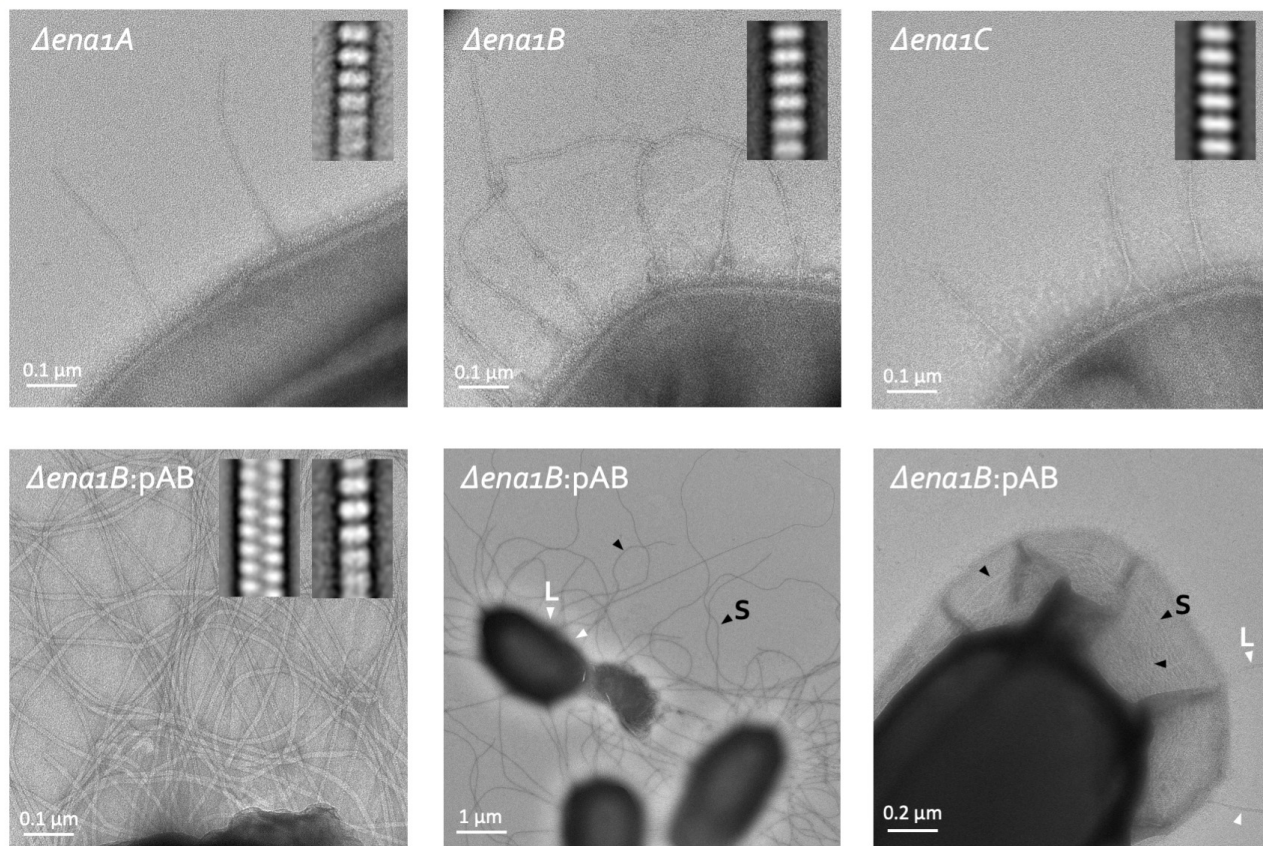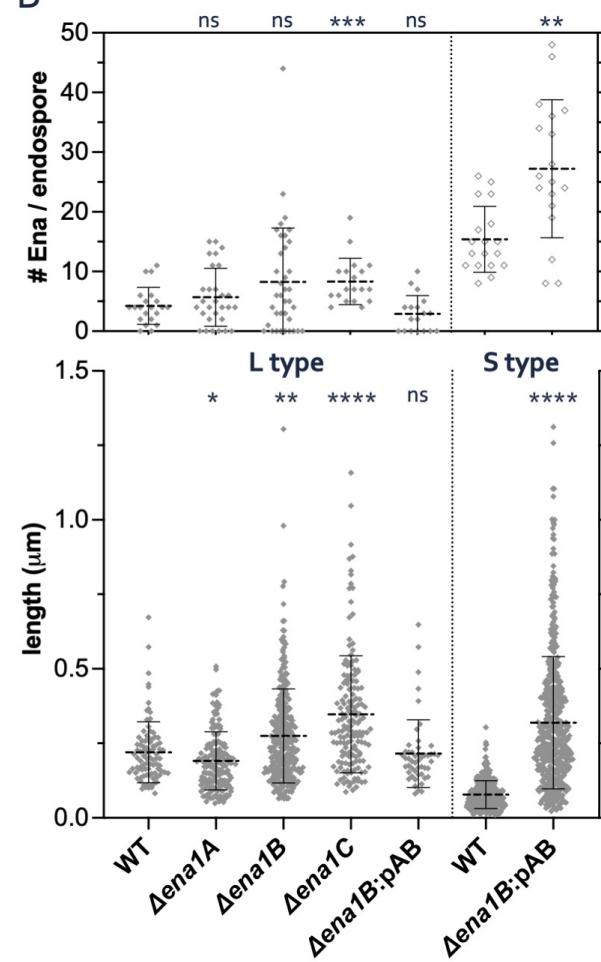
Figure 5

Figure 6

A

B

Species

- B. albus
- B. anthracis
- B. bombysepticus
- B. cereus
- B. cytotoxicus
- B. gaemokensis
- B. luti
- B. manliponensis
- B. mobilis
- B. mycoides
- B. nitratireducens
- B. pacificus
- B. paramycoides
- B. paranthracis
- B. proteolyticus
- B. pseudomycoides
- B. subtilis
- B. thuringiensis
- B. toyonensis
- B. tropicus
- B. wiedmannii

Clade

- –
- 1
- 2
- 3

Ena isoform

- Ena1
- Ena2
- absent

0.055