SUPPLEMENTARY INFORMATION

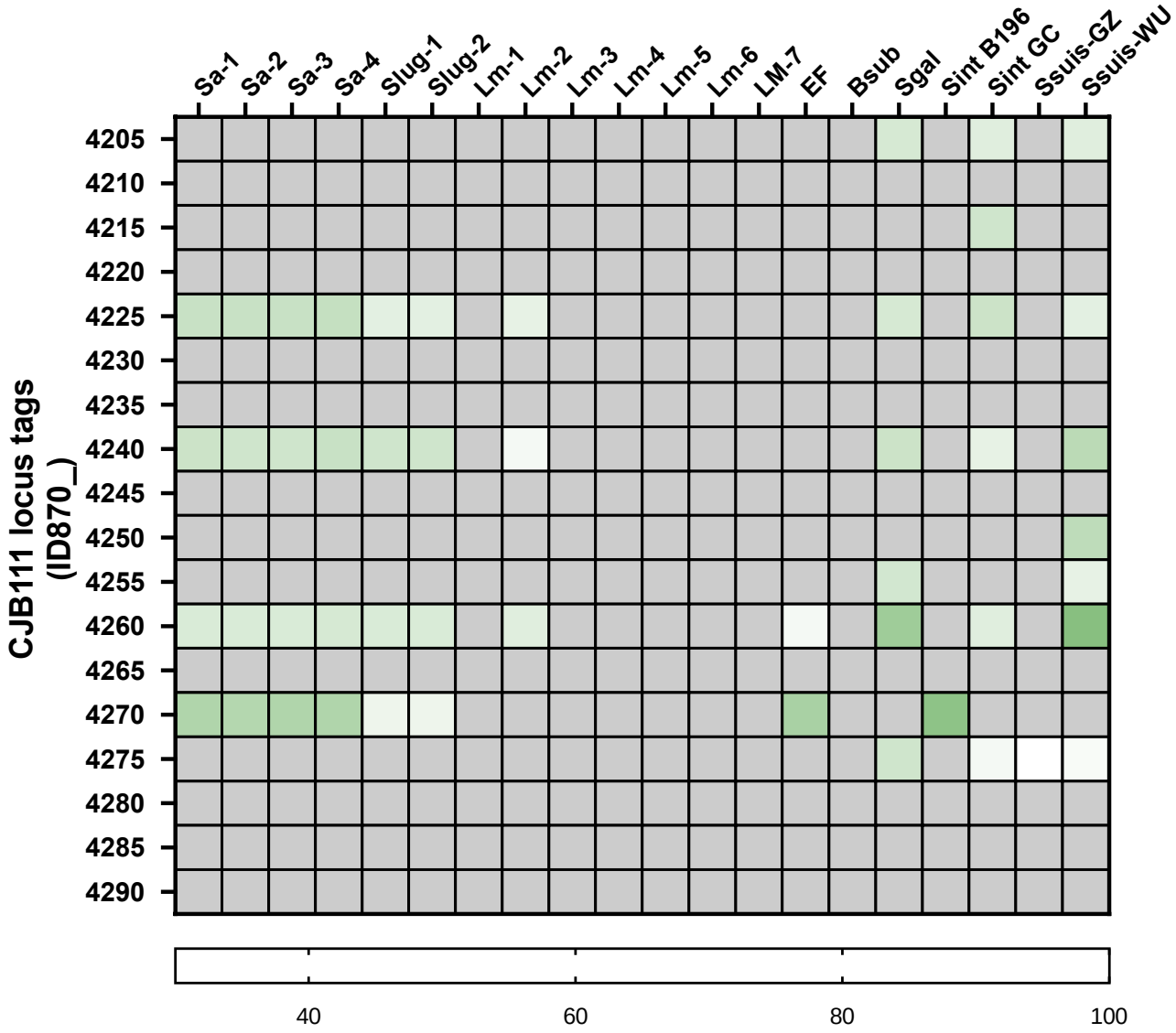
Heterogeneity of the group B streptococcal type VII secretion system and influence on colonization of the female genital tract

Brady L. Spencer[1], Alyx M. Job[1], Clare M. Robertson[2], Zainab A. Hameed[2], Camille Serchejian[2], Caitlin S. Wiafe-Kwakye[3], Jéssica C. Mendonça[1,4], Morgan A. Apolonio[1, 5], Prescilla E. Nagao[4], Melody N. Neely[3], Natalia Korotkova[6,7], Konstantin V. Korotkov[7], Kathryn A. Patras[2,8], Kelly S. Doran*[1]

[1]University of Colorado-Anschutz, Department of Immunology and Microbiology, Aurora, CO, USA
[2]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, USA.
[3]University of Maine, Molecular & Biomedical Sciences, Orono, ME, USA
[4]Rio de Janeiro State University, Roberto Alcântara Gomes Biology Institute, Rio de Janeiro, RJ, Brazil
[5]National Summer Undergraduate Research Program, University of Arizona, Tucson, AZ, USA.
[6]Department of Microbiology, Immunology and Molecular Genetics, University of Kentucky, Lexington, KY, USA.
[7]Department of Molecular and Cellular Biochemistry, University of Kentucky, Lexington, KY, USA.
[8]Alkek Center for Metagenomics and Microbiome Research, Baylor College of Medicine, Houston, TX, USA.

***Corresponding author**:
Kelly S. Doran
Department of Immunology and Microbiology
University of Colorado-Anschutz
Phone 303-724-4240
Email: kelly.doran@cuanschutz.edu

Running title: GBS T7SS diversity impacts colonization



Supp. Fig. 1. Comparison of GBS T7SS effectors to other Gram-positive organisms.
Heat map comparing CJB111 T7SS protein homology to a panel of eight Gram-positive organisms in which the T7SSb has been studied previously: *S. aureus* (variants 1-4), *S. lugdunensis* (variants 1, 2), *Listeria monocytogenes* (variants 1-7), *E. faecalis* (strain OG1RF), *Bacillus subtilis* (PY79), *S. gallolyticus* (strain TX20005), *S. intermedius* (strains B196 and GC1825), and *S. suis* (strains GZ5065 and WUSS351). See **Supp. Table 2** for individual strain information. Heat map color intensity is based on Geneious alignment grade, which considers query coverage and percent identity. Shading of the locus tag numbers on the x-axis corresponds to the key for gene color in **Fig 1**. Most downstream CJB111 T7SS effectors have little to no homology across species.

A

putative LXG associated α-helical protein 1 (TIGR04197 family T7SS effector):

	CNCTC10/84 (III)	NEM316 (II)	CJB111 (I)	C001 (III)
CNCTC10/84 (III)		34	41	33
NEM316 (II)	34		62	65
CJB111 (I)	41	62		75
C001 (III)	33	65	75	

B

	strain	LAP1/TIGR04197 family type VII secretion effector locus tag
subtype I	CJB111	ID870_04205
subtype II	NEM316	GBS_RS05720
subtype III	CNCTC10/84	W903_RS05450
	C001	GT95_RS05850

C

D

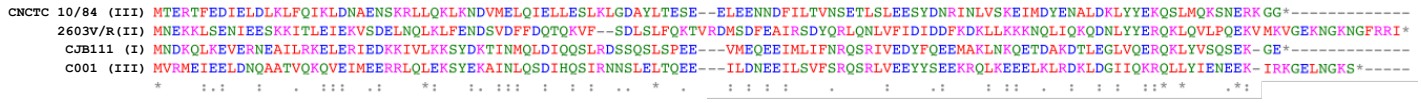
putative LXG associated α-helical protein 2

	2603V/R (II)	CNCTC10/84 (III)	CJB111 (I)	C001 (III)
2603V/R (II)		15	21	19
CNCTC10/84 (III)	15		26	24
CJB111 (I)	21	26		41
C001 (III)	19	24	41	

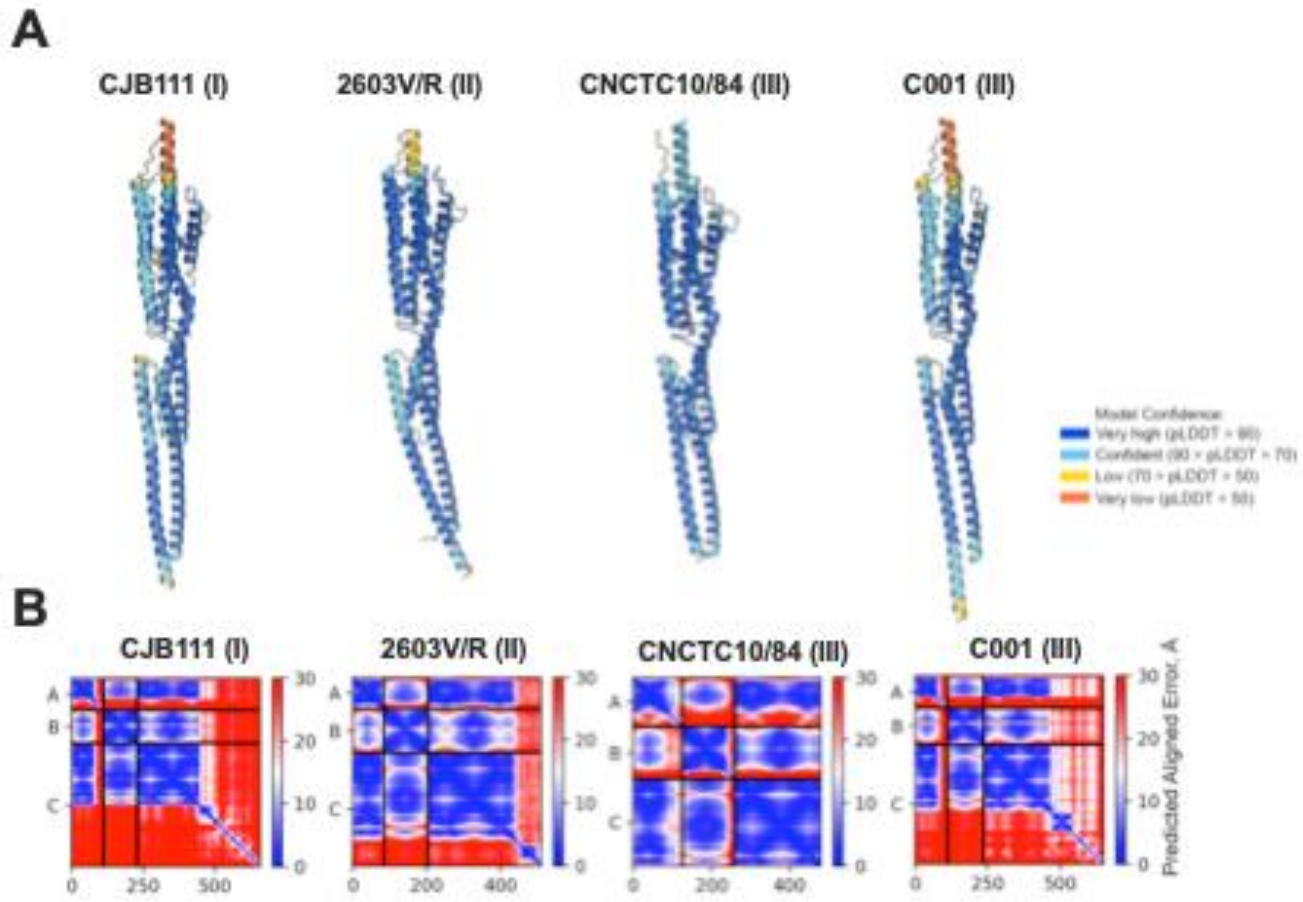
E

	strain	LAP2 locus tag
subtype I	CJB111	ID870_04210
subtype II	2603V/R	GBS_RS07885
subtype III	CNCTC10/84	W903_RS05445
	C001	GT95_RS05845

F

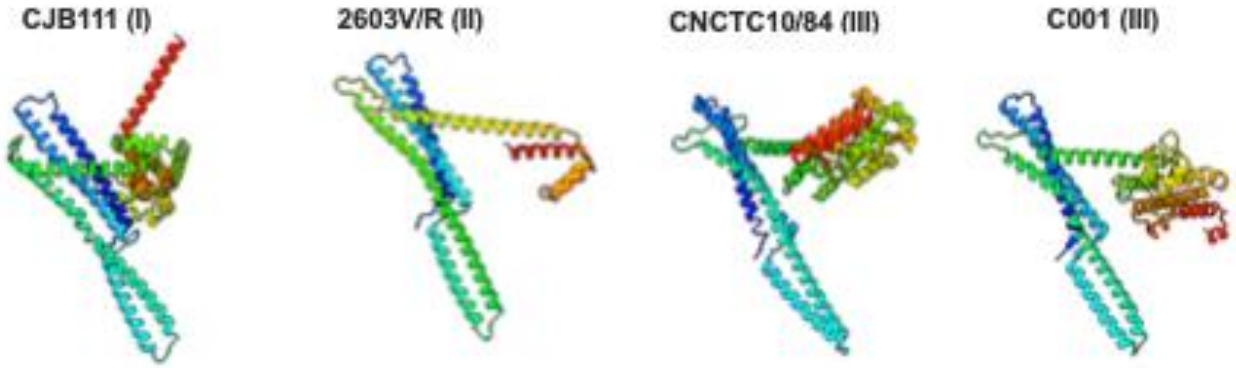


Supp. Fig. 2. Subtype-specific LXG-associated proteins are encoded by GBS T7SS.
Percent identity matrices and Clustal Omega alignments of **A-C)** LXG-associated protein 1 (Lap1) and **D-F)** LXG-associated protein 2 (Lap2) sequences from CJB111 (subtype I), 2603V/R (subtype II), CNCTC 10/84 (subtype III), and C001 (subtype III). In the above matrices, the purple shading corresponds to the level of identity between two strains (on a spectrum of 0 to 100% identity), with darker shading indicative of higher percent identity. Conserved putative T7SS-associated FXG and FxxxD/E motifs are highlighted by red boxes in **C.**



Supp. Fig. 3. Confidence scores and predicted aligned error for LXG-Lap AlphaFold models in Fig. 2A. **A)** Predicted LXG-Lap complex models shown in **Fig. 2A** but with color corresponding to per-residue confidence level (predicted local distance difference test [pLDDT] score 1-100; pLDDT > 90 are expected to be modelled to high accuracy). **B)** Predicted aligned error for each LXG-Lap complex, with colors indicating the confidence of domain positions (higher predicted error in red, lower predicted error in blue).

A



B

Transmembrane proteins downstream of LXG:

	CNCTC10/84 (III)	C001 (III)	CJB111 (I)	2603V/R (II)
CNCTC10/84 (III)		21	20	15
C001 (III)	21		22	19
CJB111 (I)	20	22		31
2603V/R (II)	15	19	31	

	strain	downstream genes locus tag
subtype I	CJB111	ID870_04220
subtype II	2603V/R	SAG_RS07870
subtype III	CNCTC10/84	W903_RS05435
	C001	GT95_RS05845

C

DUF4176-containing proteins

subtype I, CJB111	subtype II, 2603V/R	subtype III, CNCTC 10/84	subtype IV, COH1
ID870_RS04225	SAG_RS07865	W903_RS05425	RS08075
ID870_RS04240	SAG_RS13455 (pseudo)	W903_RS05405	
ID870_RS04260	SAG_RS11090	W903_RS05385	
ID870_RS01255 (pseudo)		W903_RS18490	

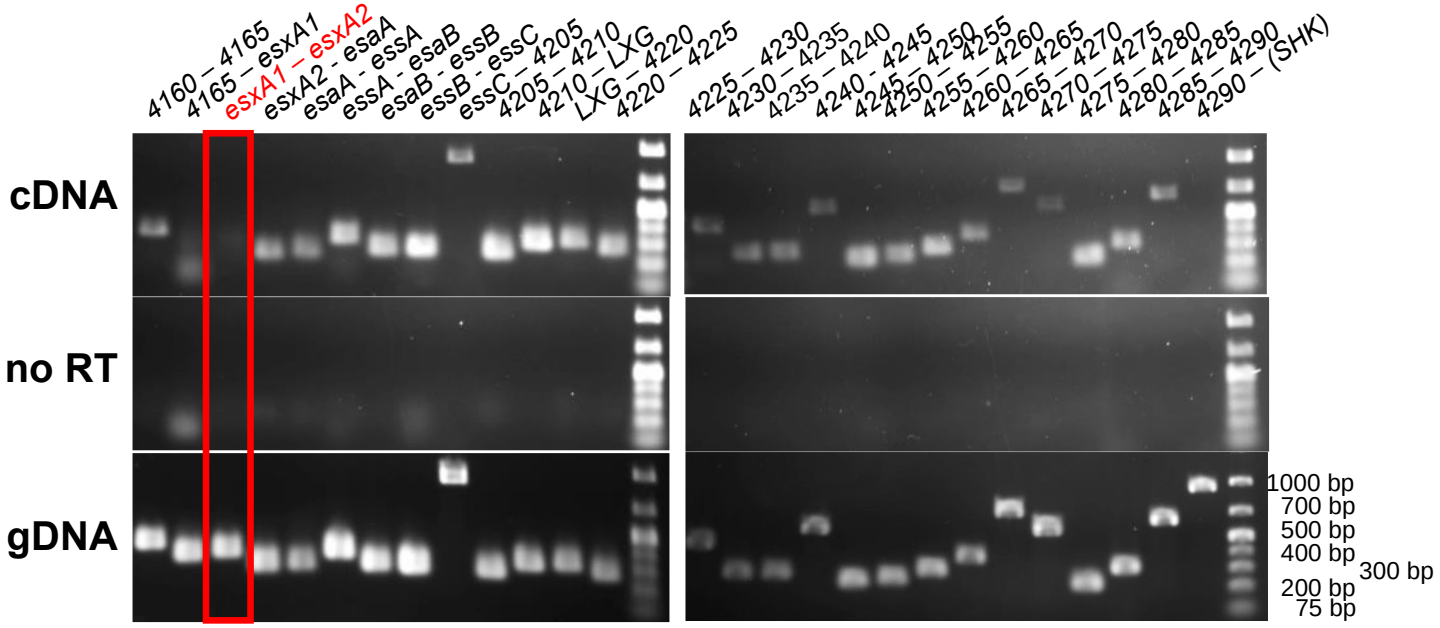
D

E

	NCTC_DUF2 (5405)	NCTC_DUF1 (5425)	CJB111_DUF1 (4225)	2603_DUF1 (7865)	CJB111_DUF3 (4260)	NCTC_DUF3 (5385)	CJB111_DUF2 (4240)	COH1_DUF$_{orphan}$ (8075)	2603_DUF$_{orphan}$ (11090)	NCTC_DUF$_{orphan}$ (18490)
NCTC_DUF2 (5405)		33	36	35	40	37	41	40	40	38
NCTC_DUF1 (5425)	33		34	33	35	32	37	36	36	31
CJB111_DUF1 (4225)	36	34		60	35	33	35	33	33	30
2603_DUF1 (7865)	35	33	60		37	48	44	44	44	44
CJB111_DUF3 (4260)	40	35	35	37		56	48	46	46	40
NCTC_DUF_3 (5385)	37	32	33	48	56		77	64	64	62
CJB111_DUF2 (4240)	41	37	35	44	48	77		85	85	81
COH1_DUF$_{orphan}$ (8075)	40	36	33	44	46	64	85		100	94
2603_DUF$_{orphan}$ (11090)	40	36	33	44	46	64	85	100		94
NCTC_DUF$_{orphan}$ (18490)	38	31	30	44	40	62	81	94	94	

Supp. Fig. 4. Modeling of full-length LXG proteins and homology of downstream putative immunity factor and DUF4176 genes

A) Robetta predicted structures for full-length putative GBS LXG toxins across subtypes I, II, and III. **B)** Percent identity matrix of putative immunity factors (transmembrane domain containing proteins) encoded for downstream of LXG genes in CJB111 (subtype I), 2603V/R (subtype II), CNCTC 10/84 (subtype III), and C001 (subtype III). **C-E)** Percent identity matrices and Clustal Omega alignments of DUF4176 protein sequences from CJB111 (subtype I), 2603V/R (subtype II), CNCTC 10/84 (subtype III), and COH1 (subtype IV). Green highlighting in **C-E** indicate orphaned DUF4176 proteins. A conserved central FXG motif within GBS DUF4176 proteins is highlighted by a red box in **D.** In all of the above percent identity matrices, the purple shading corresponds to the level of identity between two strains (on a spectrum of 0 to 100% identity), with darker shading indicative of higher percent identity.



Supp. Fig. 5. Transcriptional landscape of the CJB111 T7SS locus
RT-PCR was performed to evaluate transcriptional organization of the CJB111 T7SS locus. Using cDNA as template (and no RT-cDNA and genomic DNA as negative and positive controls, respectively), PCRs were performed using primer pairs spanning every gene junction in the putative T7SS locus. Primers used for these experiments can be found in **Supp. Table 5.** Agarose gels shown are representative of three independent experiments.

A

GBS WXG100 genes

	subtype I		subtype II			subtype III	subtype IV	
	CJB111 (CP063198.2)	**A909** (NC_007432.1)	**2603V/R** (NC_004116.1)	**NEM316** (NC_004368.1)	**515** (NZ_CP051004.1)	**CNCTC 10/84** (NZ_CP006910.1)	**COH1** (NZ_HG939456.1)	**BM110** (NZ_LT714196.1)
esxA1	ID870_04170	SAK_RS05640				W903_RS05485	none	
esxA2	ID870_04175	SAK_RS05645	SAG_RS07925	GBS_RS05755	GRB95_RS05270	W903_RS05490		
esxA3	ID870_08245	SAK_RS01430	SAG_RS03860	GBS_RS01395	GRB95_RS01415	W903_RS01365		BQ8897_RS01705
esxA4	ID870_10565	SAK_RS09855		GBS_RS10285		W903_RS10725*		

orphaned genes in green

*frameshifted/truncated

B

Module 1 WXG100 (EsxA3) protein alignment

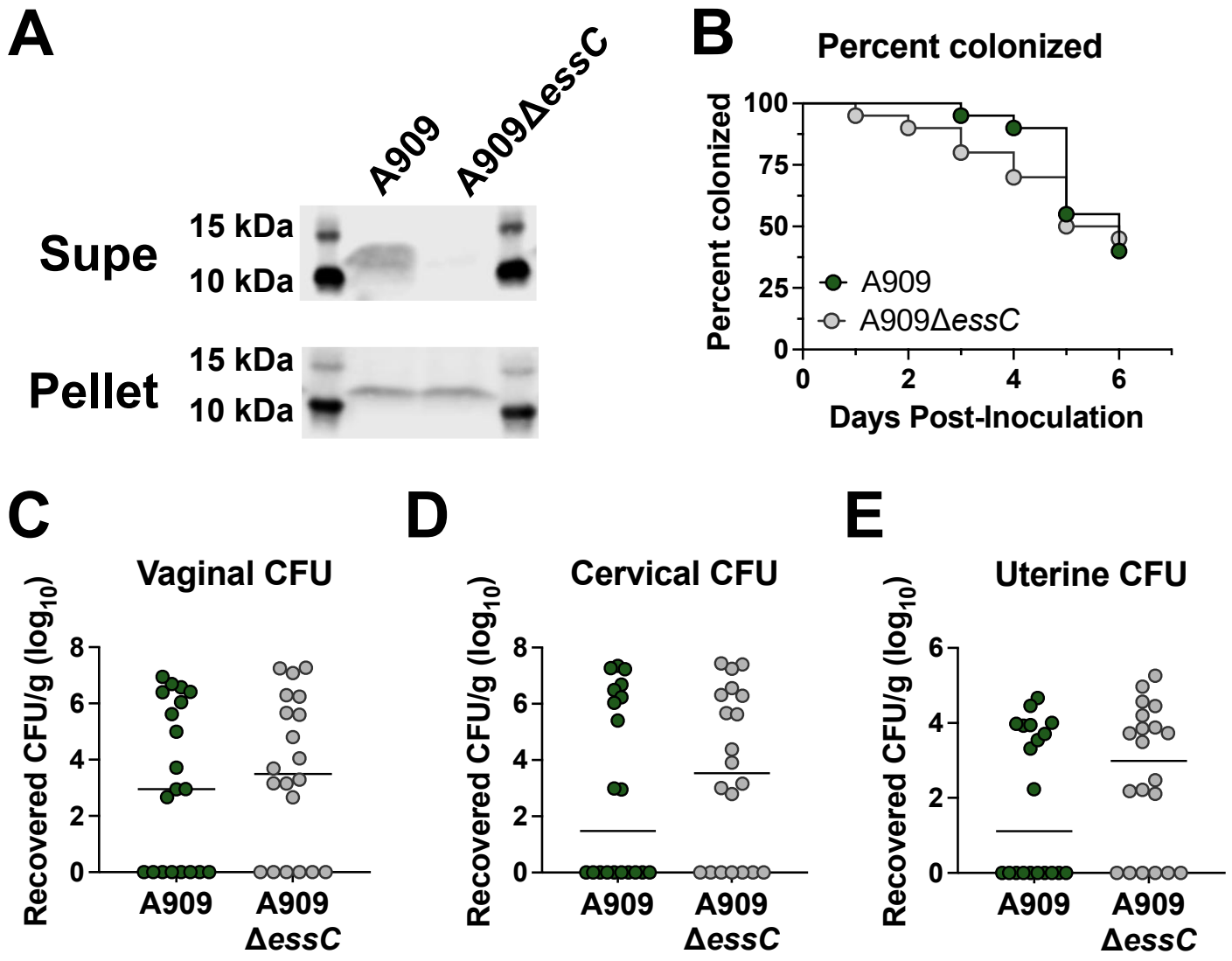
	CJB111	A909	2603V/R	NEM316	515	CNCTC 10/84	BM110
CJB111		100	100	99	99	98	100
A909	100		100	99	99	98	100
2603V/R	100	100		99	99	98	100
NEM316	99	99	99		100	97	99
515	99	99	99	100		97	99
CNCTC 10/84	98	98	98	97	97		98
BM110	100	100	100	99	99	98	

C

Module 2 WXG100 (EsxA4) protein alignment

	CJB111	A909	NEM316	CNCTC 10/84
CJB111		97	99	98
A909	97		98	95
NEM316	99	98		97
CNCTC 10/84	98	95	97	

Supp. Fig. 6. Homology of orphaned GBS WXG100 proteins.
A) Locus tag tables and percent identity matrices of **B)** Orphaned Module 1 WXG100 protein EsxA3 and **C)** Orphaned Module 2 WXG100 protein EsxA4 sequences across a panel of GBS isolates representing T7SS subtypes I-IV. Green highlighting in **A** indicates orphaned WXG100 proteins. In the above matrices **(B-C)**, the purple shading corresponds to the level of identity between two strains (on a spectrum of 0 to 100% identity), with darker shading indicative of higher percent identity.



Supp. Fig 7. Impact of EssC deficiency on GBS vaginal colonization by T7SS subtype I strain A909. A) Western blot showing EssC-dependent secretion of EsxA from subtype I strain A909, *in vitro*. **B)** Percent colonization curve of 8-week-old CD1 female mice vaginally inoculated with subtype I strain A909 or A909Δ*essC*. Statistics reflect the Log rank (Mantel-Cox) test. Recovered CFU counts from the **C)** vaginal **D)** cervical, and **E)** uterine tissue of colonized mice. In panels **C-E**, each dot represents one mouse and two independent experiments' data are combined in these figures (n = 20/group total). Plots show the median and statistics represent the Mann Whitney U test.