Supplementary Information

Genomic features of the *Helicobacter pylori* strain PMSS1 and its virulence attributes as deduced from its *in vivo* colonization patterns

Victoria Dyer[1], Holger Brüggemann[2], Meike Sörensen[1], Anja A Kühl[3], Kirstin Hoffman[1], Volker Brinkmann[1], Maria del Mar Reines[1], Stephanie Zimmerman[1], Thomas F Meyer[1,#], and Manuel Koch[1].

[1] Department of Molecular Biology, Max Planck Institute for Infection Biology, 10117 Berlin Germany

[2] Department of Biomedicine, Aarhus University, 8000 Aarhus C, Denmark

[3] Medical Department, Division of Gastroenterology, Infectiology and Rheumatology, Campus Benjamin Franklin, Charité, 12200 Berlin, Germany

Supplementary Figure 1. PMSS1 persistently colonises the gastric mucosa of C57BL/6 mice.
Quantification of the colony forming units (CFU) cultured from the gastric tissue of C57BL/6 mice, infected on two alternate days with >10^8 CFU of PMSS1 or P12 (inoculum), at two and four m.p.i. Filled circles represent individual mice; bars indicate the mean level of colonisation; n.d.: not detected.

Supplementary Figure 2. T4SS morphology is maintained although the functionality is altered.
Scanning electron microscopy of the wild type *H. pylori* strains PMSS1 and P12 in addition to the *in vivo* bacterial isolates Iso6 and Iso8 show no differences in the phenotype of the T4SS pili (white arrow) although the functionality is impaired in Iso6. Scale bars: 500 nm.

Supplementary Figure 3. Comparison of mutations in CagPAI from different isolates
(A) Sanger sequencing of *cag*W for Iso1 did not yield any mutation as observed in Iso6, thus the mutation probably occurs within another of the *cag* genes essential for the T4SS functionality. **(B)** The single nucleotide insertion was retained in all bacterial isolates from the group Iso6. **(C)** Intestinal metaplasia observed in the mice from which the Iso6 bacterial isolate was obtained.

Supplementary Figure 4. Comparison of two H. pylori PMSS1 sequences
(A) SNP analysis of two PMSS1 chromosomal sequences and the derivative strains SS1, Iso6 and Iso8. Parsnp was used to interfere phylogeny. Reference strain was our PMSS1 sequence (GenBank: AZBR00000000). 5 high-quality SNPs are detected in PMSS1 published by Draper et al. [15] and 10 SNPs were detected in Iso6 and Iso8, respectively. **(B)** Variation of cagA copy number in the two PMSS1 chromosomal sequences. A comparison using the ACT tool was carried out. The published PMSS1 sequence of Draper et al contains four cagA copies, whereas our sequence apparently only contains one copy. Differences might be also due to the different sequencing methods used in the two studies (PacBio and Illumina, respectively).
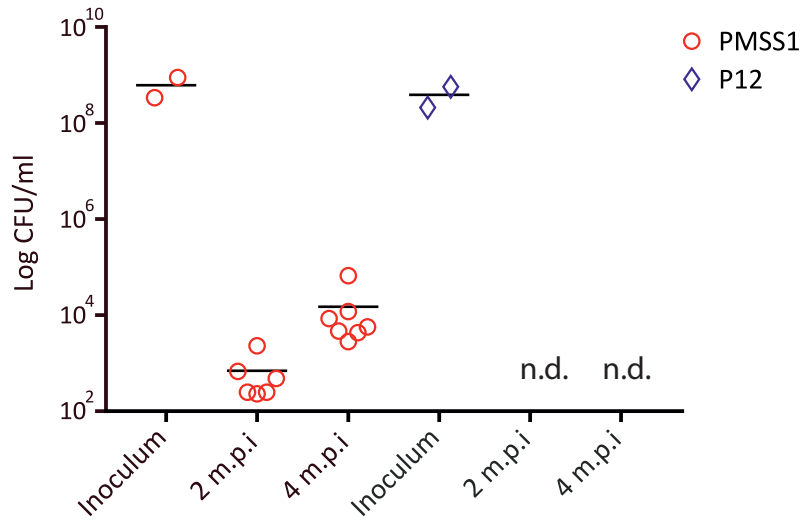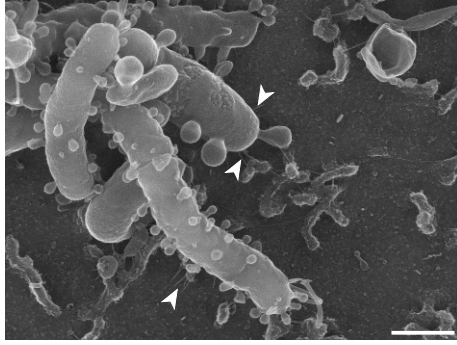
A

Supplementary Figure 1
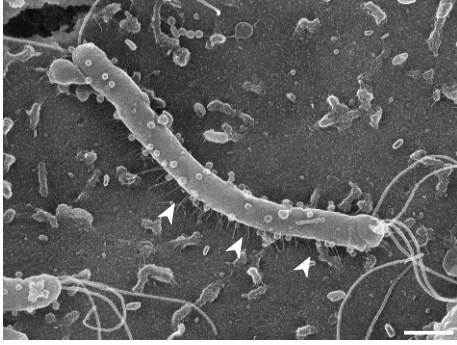
A

P12   PMSS1

Iso6   Iso8

Supplementary Figure 2

A

```
Iso6      GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
Iso_1.1   GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
Iso_1.2   GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
Iso_1.3   GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
Iso_1.4   GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
Iso_1.5   GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
Iso_1.6   GGGATTTTTGTTTTCATTTCTTTTATGAAATTGAGGGATATTTCAAATGGTGTTCAGATT    840
          ************************************************************

Iso6      AACATAGGTTTTGAATACATGCGATTTATTTGGGGGGACATTATTCAAAATGGCGATGGT    900
Iso_1.1   AACATAGGTTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGGT    899
Iso_1.2   AACATAGGTTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGGT    899
Iso_1.3   AACATAGGTTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGGT    899
Iso_1.4   AACATAGGTTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGGT    899
Iso_1.5   AACATAGGTTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGGT    899
Iso_1.6   AACATAGGTTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGGT    899
          *****************************  ****************************

Iso6      CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    960
Iso_1.1   CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    959
Iso_1.2   CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    959
Iso_1.3   CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    959
Iso_1.4   CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    959
Iso_1.5   CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    959
Iso_1.6   CTCTTTTATCGCCTATGCAGGTTTTGGTTATCTTTATAAAATCTCTTATTCCATTTATTT    959
          ************************************************************
```
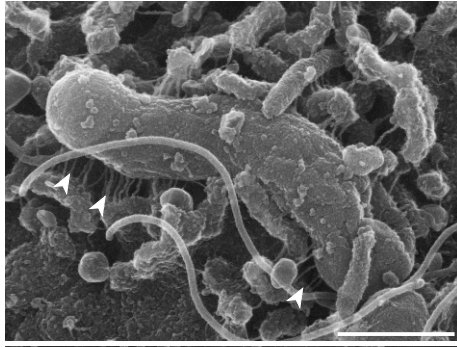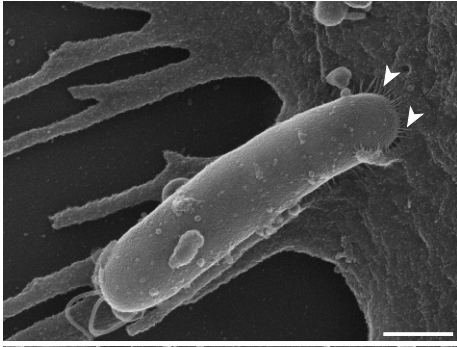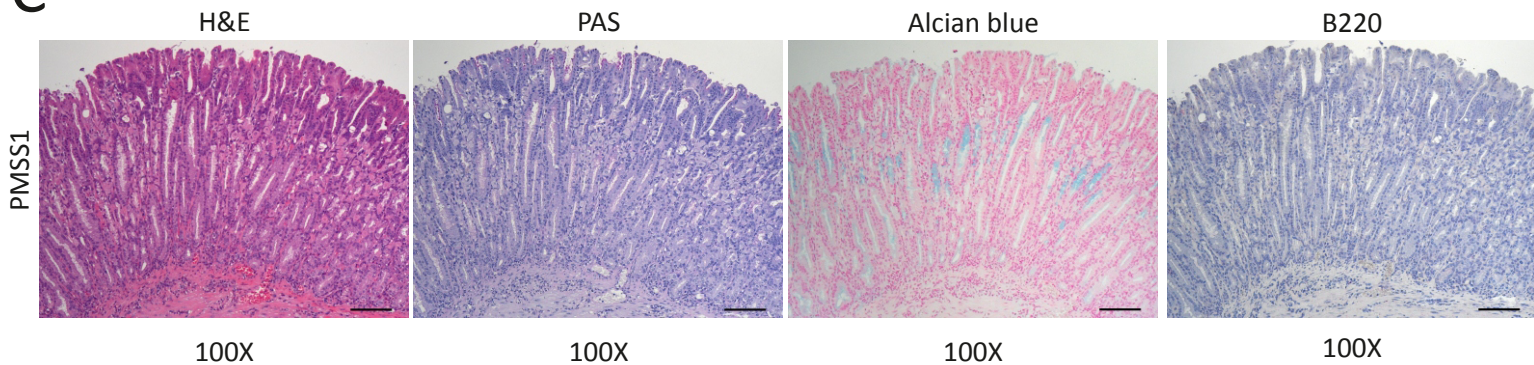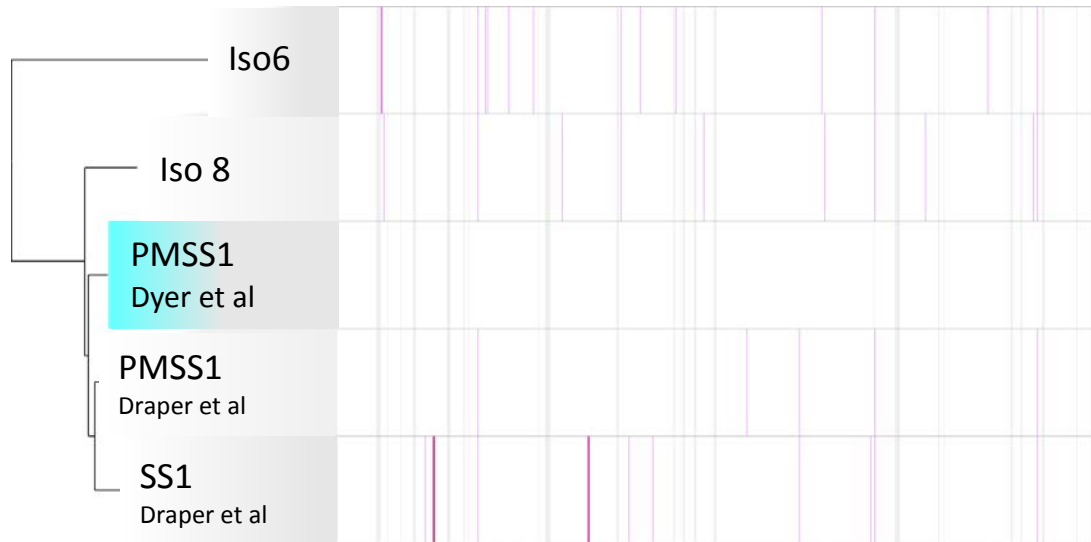
B

```
PMSS1     AACATAGG-TTTTGAATACATGCGATTTATT-GGGGGGACATTATTCAAAATGGCGATGG    898
Iso_6     AACATAGG-TTTTGAATACATGCGATTTATTTGGGGGGACATTATTCAAAATGGCGATGG    899
Iso_6.1   AACATACGGTTTTGAATACATGCGATTTATTTGGGGGGACATTATTCAAAATGGCGATGG    900
Iso_6.2   AACATACG-TTTTGAATACATGCGATTTATTTGGGGGGACATTATTCAAAATGGCGATGG    899
Iso_6.3   AACATACG-TTTTGAATACATGCGATTTATTTGGGGGGACATTATTCAAAATGGCGATGG    899
          ******  * ********************* ***************************
```

C

Supplementary Figure 3

A

B

Supplementary Figure 4