Supplementary figures

Transcriptomics in serum and culture medium reveal shared and differential gene regulation in pathogenic and commensal *Streptococcus* suis



Figure S1: Growth curves. The strains were grown in 96-well plates with 200 μ L total volume per well and incubated at 37 °C with 5% CO₂. Each point represents the mean value of the strain sampled in minimum 3 separate plates with 2 technical replicates per plate.



Figure S2: High ADS expression in S11 grown in THY. **A)** PCA on the core genome transcriptome of all strains shows sample S11T (strain S11 grown in THY) as an outlier. Each label represents a single sample, eclipses represent 75% confidence level, arrows show genes driving the separation of samples; samples in the direction the arrow is pointing have higher expression of the gene. **B)** Strain S11 also had the largest transcriptome difference between THY and APS cultures of the same strain (core genome transcriptome Bray-Curtis dissimilarity). **C)** Boxplot of expression (shown as log10(TPM) of *arcABC* and adjacent genes.



Figure S3: Opacification factor of serum (*ofs*). **A**) *ofs* occurs in different (suggested pseudogene) variants, but in similar genomic context. The 3 strains with high opaficification activity in serum (red box) share a similar variant (although there is substantial nucleotide level variation between these 3 strains). **B**) The 3 commensals with *ofs* had high expression of the gene. The other commensal strains did not have any detected *ofs* homologue.



Figure S4: Boxplots showing expression of the overall most expressed GO terms in APS vs THY.



Figure S5: Genes strongly regulated between media in sequence type 1 strains. A) RDA on log2fold gene expression change constrained by ST1 clade strains vs all other strains. Each label represents a single sample, eclipses represent 75% confidence level, arrows show genes driving the separation of samples; samples in the direction the arrow is pointing have higher expression of the gene. B) Boxplot of the log2fold change of the 10 strongest drivers (genes) of the RDA1 axis. C) Boxplot of the TPM expression of the same 10 genes in APS and THY.



Figure S6: Maximum likelihood trees of *cps* genes (DNA). **A)** SSU_RS02755, *cpsR/gntR/orf2Y* **B)** SSU_RS02765, *cpsA* **C)** SSU_RS02770, *cpsB* **D)** SSU_RS02775, cpsC **E)** SSU_RS02780, *cpsD*. Tip labels: [Strain]_[gene number]. Except from outliers D13 and D20, *cpsR* is more conserved than *cpsABCD*. Diverged *cpsR* copies in strains D43 and D49 may underlie different *cps* regulation compared to other serotype 2 strains.



Figure S7: Multiple sequence alignment of *cpsR* (SSU_RS02755) protein sequences. Strains D43 and D49 do not share sequence similarity with other serotype 2 strain, except from a R/S SNP in position 17.



Figure S8: Divergent *lac* gene cluster expression in strain D13. **A)** Several variants of the *lac* gene cluster are carried by the strains, and these may occur alone or two (at different locations) in the same genome. D13 has 2 *lac* gene clusters, which both feature inverted or pseudogene *lacR* copies (although other strains sharing the inverted variant had exopressino as expected). Genomes with several copies highlighted in color. Annotation based on Afzal et. al. 2014 (https://doi.org/10.1128/AEM.01370-14) and *S. pneumoniae* D39 assembly CP000410.2. **B)** *lacR* is highly expressed in D13, but *lacA* and *lacB* (and *lacT*) are nonetheless highly expressed in both media. P1/7 has unexpectly low *lac* expression compared to other ST1 strains.



Figure S9: Gene expression of the *tro/mnt* gene cluster. A) ML tree of *perR* with *tro* expression heatmap. B) ML tree of *troR* with *tro* expression heatmap. C) Raw RNA-seq coverage of the *tro* gene cluster in strain D13 (visualized in IGV, https://igv.org). D) Raw RNA-seq coverage of the *tro* gene cluster in strain P1/7.