

Supplementary material

Table S1: Proteins with serine rich domains

Fig. S1: Distribution in functional categories of the genes with non streptococcal best BlastP hits. In blue: absolute number, in red: percentage as compared to the total number of genes in the respective functional category.

Fig. S2: Distribution in functional categories of the genes with no ortholog among the 10 sequenced species of *Streptococcus* but with a streptococcal best BlastP hit. In blue: absolute number, in red: percentage compared to the total number of genes in the respective functional category.

Fig. S3: Phylogenetic relationship of RibA and RibD protein sequences with representative sequences from Gram positive bacteria. A. RibA; B. RibD. The two trees show similar topologies. The streptococcal proteins are boxed. *S. gallolyticus* proteins cluster with *Lactobacillus reuterii* orthologs, *S. pneumoniae* with *Bifidobacterium longum* and *S. agalactiae* represents a separate branch. The scale bar represents 0.2 substitutions per site. The numbers at the branches are posterior probabilities indicating the support for the branch. GI numbers for sequences are given in parentheses.

Table S1. Proteins with serine rich domains

| Protein | Specificity or activity ¹ | Signal peptides and serine rich domains ² |
|---|--------------------------------------|---|
| Lipoproteins ABC transporter component | | |
| Gallo_0116 | Sugars | MTMTKLT TKIKLWLSVLVLL LSACVGGYVYYQM QK TILKIGVYAG |
| Gallo_0122 | | MKRNMAAIMALMLAVLT LF GCSSSKSTSDDV I HIGIL |
| Gallo_0163 | sugars | MKKGLLTIGMTALAAVTLVGCSSGSSDSDVETITFINH |
| Gallo_0192 | multiple sugars | MKWRKQLIAAGIVALATTAGLILTACSKEDDTSADGKVTIEY |
| Gallo_0324 | oligopeptide | MLLKSKT WKRI GLGAVTLVSAAVLAACGGSSSSSSSSSDEINWYTPTEIS |
| Gallo_0414 | amino acids | MKKRRLLSFGFLFLLTLAACS NQ SQSSGKTVIKVATD |
| Gallo_0754 | sugar | MKIKTLSAISLTVIPSLILANFCHFPSSQTNSTGESSIFNGELEE G AVI |
| Gallo_0874 | ferichrome | MKKKLTLLLTAMMVLVMAVFLGACSSSSSSNSTSQSSDASE |
| Gallo_1000 | amino acids | MKFKKVL MGALALV STLTLAACSSSSSKKATSTDQWD T YVK |
| Gallo_1136 | | MNKKIVGLGLAAVATLALGACSR NS SSSSSSDSSVKA I IVTDTG |
| Gallo_1136 | | MKTKQTLLYSVCALTLLSLGACSQTSSTSSSTSSSSSTVASTTSA |
| Gallo_1149 | phosphate | MKKMKKLSFLLLLITSVSIILSGCASWIDKQ S ITSVGST |
| Gallo_1234 | amino acids | MKIVKRFLAIVSLLVVVLVGCSTSKSSSSTSSSSSSSGNTAKARTL |
| Gallo_1392 | amino acids | MKIKKIFLGVLALISVLTLAACGSSSNENLQEKIVK |
| Gallo_1394 | amino acids | MKKFFRRFILLIIVGSLSACSSSTGVSQSSIQDKG 267-TYTQSSSSSDEZ |
| Gallo_1395 | amino acids | MK V SKLFGGLAVVVASLFLLSACGSSSEDT S VSDIKDK 284-AYELSVSDSSDSSDSDZ |
| Gallo_1399 | maltose | MKKN T WKKMVLGAGLTLAGSVLVACSNSSSNSSSSSSSKTIK L WVPTGAKE |
| Gallo_1412 | | MEIK N WKR V GLGAVTLLSAAVLAACGNSSSSSSSSSDEINWFIPTEISTLDISK |
| Gallo_1413 | | MEFKSTWKR V GLGVVSLASAALLAACGNSSSSSSSSNEINWYTPTEILTLDISK |
| Gallo_1760 | amino acids | MHKKIALTALTFLASIALAACSKSPDVTANATGTTIG |

| | | |
|------------|-------------|---|
| Gallo_1772 | iron | MKKFFAVLTTFLATFLLVACHNTSSTSDDETELSSMPKITGFSYE |
| Gallo_1844 | | MKLKFLFGLASVAFASVLLAACGSSSSSSSSDDTTLKVGIMTLD 293-DGADIPAWDGATTSSSSSDDZ |
| Gallo_1845 | amino acids | MKKILGVTGVALASTTFLAACSSSSSSSSSSSGKEEVVFATVGTTA |
| Gallo_1966 | phosphate | MKKWKMLMLLVFVGVGVVLTACSKSNSSESKQISVVSRE |
| Gallo_2047 | manganese | MKKITSLICLLLIICILGACATTRTSEKNKLNVVVVTNSILSDITQNIK |
| Gallo_2130 | metal ion | MKKKVLIVTGFVATILTGCGASQVATSSSTSDSNVVIKVGANITPHAEILEEA |

Other lipoproteins

| | | |
|------------|---------------------------|---|
| Gallo_0624 | | MKKSCKITVTLCSVLLLGACSKKEKSSQVSTSHSKVTQVSKSSIK |
| Gallo_0696 | | MKKTVTYLALAATSVLFLTACSNNNQESTNTSSSSSTSSSEETTGGGFVT |
| Gallo_0751 | | MKKGFIILTIAALSFTLTACSONSQTSSSSGTSTSEVSSNKVTTY |
| Gallo_0883 | | MKKLGCFILVMMLLVLSACSSVQDNQAKQTQTTTTSKTEEEKMVT |
| TanA | tannase | MPRKKWFFTSSAVLLCSAMLTACSSSSNSSTSSSSSQNTTAST |
| Gallo_1117 | | MKTKQTLLYSVCALTLLSLGACSQTSSTSTSSSSSTVASTTSASTS |
| Gallo_1367 | | MKKKLRCYLTLLLSLGLAACSTHSQNNTNQSSSSTTVSSSVKKSTNSSET STSTGSSESEEEASSTESSSTSTHSQNNTNQSSSSTTVSSSVKKSTNSS |
| Gallo_1245 | | MKKIAIFVLSVVSFLVVGCSNSTSGGFSNLPDHLGDNKAVSNLAGAGVTYN |
| Gallo_1586 | | MKKGILFFVSVLAVLAACSNSTGKQESSKQTNESQSSYYQVTSDDISITKDA |
| Gallo_1601 | | MKLVVFLGVMLVSI FALAACSTNSSSSKSKTTTTINGTKY |
| Gallo_1710 | | MKKKLLVALLAVMSVFLLVGCSSKDDLKSGKYYGINYN |
| Gallo_1717 | Peptidyl-prolyl isomerase | MKKFWSFGLMVLCLASLGCESITRAIRGDDYVDAKIAASSSEAAASASASASSSK |
| Gallo_1778 | | MKKLVLLGTMILLSGCSKKSQSVSGSWYTNSETISSSSSTVSQTESNEE |
| Gallo_1814 | | MKKKLNRVLFSGLSLSLLFLTGCVSRDSSGNPTGTIWHILGEPMAKLIQYFANNV |
| Gallo_1907 | | MKKRKFLGLSVLLLTI LAGCGADSELTDVSDIQTEINKDYSTESEMNIIMDIIS |

Gallo_2214

MKKFATSIIILISLV**LVGCTH**STDSNSNSTKTSEAGQSQTIGEILNKDKRTIWYST

Other proteins

| | | |
|------------|---------------------------------|---|
| Gallo_0019 | Glucan binding | 309-AATV TPTATTTTSSSSSSSSASSSSSSSSASTSSTASTSTSSSSSSSSSSSVNTY PVG-358 |
| PbpF - | Penicillin-binding protein 2C | 720-DLVY STDDSTDTSSSSSSSSSDS ILD-745 |
| Gallo_0239 | D-ala d-ala carboxypermase | 46- SSTDSSSLVSESSTSTASATDLPDVSSSD -76 |
| Gallo_0388 | | 1-MKTSRSEVTYLILSIILIAFGAGIYLIFGNNYASTSSSQTTTAV SSSSSTSEEDLEANA |
| PonA - | Penicillin-binding protein 1A | 680-SG SSNNYTYSSSTSSSVYSNIYGSSSSSSSEESTAESSSEDSSTESSNEADNNSDNSNTES SDDZ |
| Gallo_0596 | | 111-P STTATSTSTSSSSSSSSTSETKTSST -137 |
| Gallo_0673 | Peptidylprolyl isomerase (2 TM) | 329-SILSQY ATSSSSSSSSSSSSSSSSSSSSDDSSSSSSTEESSSTTES SDEZ |
| Gallo_0878 | <i>C. difficile</i> tcdC like | 52- SSSDSDDEEDYDYDDDDYSSSSSSSYSSSSSSSSSSSYSSSSSE -97 |
| Gallo_1348 | | 1-MSQKYSRRQK SSSKNDKSAAPT KHIKTGFSALQKTVALIGSILSIIVASITISNALKGSNSKT TDKSTTTTTV IKESDSDSKTDT SASTYSSASANN NYNSGNNNTYSSSSD TTYSSSSATSQSSSTPSSD TTSD TTTTSSNTETTEZ |
| Gallo_1358 | Polyglycerol P synthase | 717-LFKAPSYMQLH SSSSSSSSSSSSSSSSDGSZ |
| Gallo_1363 | | 75- SSSSVSSSEASSSETTAETTTDESSTDASSDDTTESSSTGSGDTIT -123 |
| Gallo_1899 | | 50- GGSSGGFSSGFSSGSSSSSSTSTSSGDGD -78 |
| Gallo_1903 | C-term domain of protein kinase | Several motifs in the C-terminal domain |
| Gallo_2249 | One TM | 186-SDSNNASTSYSVVH SSSSTTTDSTSSSESSSTSESSSEED -226 |
| prkC | Protein kinase (1 TM) | 597- SSSSSESSEKESSSSSSSTDDSSSSTETSDEZ |

1 Other characteristics are indicated like the prediction of transmembrane domains (TM)

2 The putative lipo-box is indicated in blue and the serine rich regions in red. Numbers indicate the position in the protein and Z the C-terminal end.

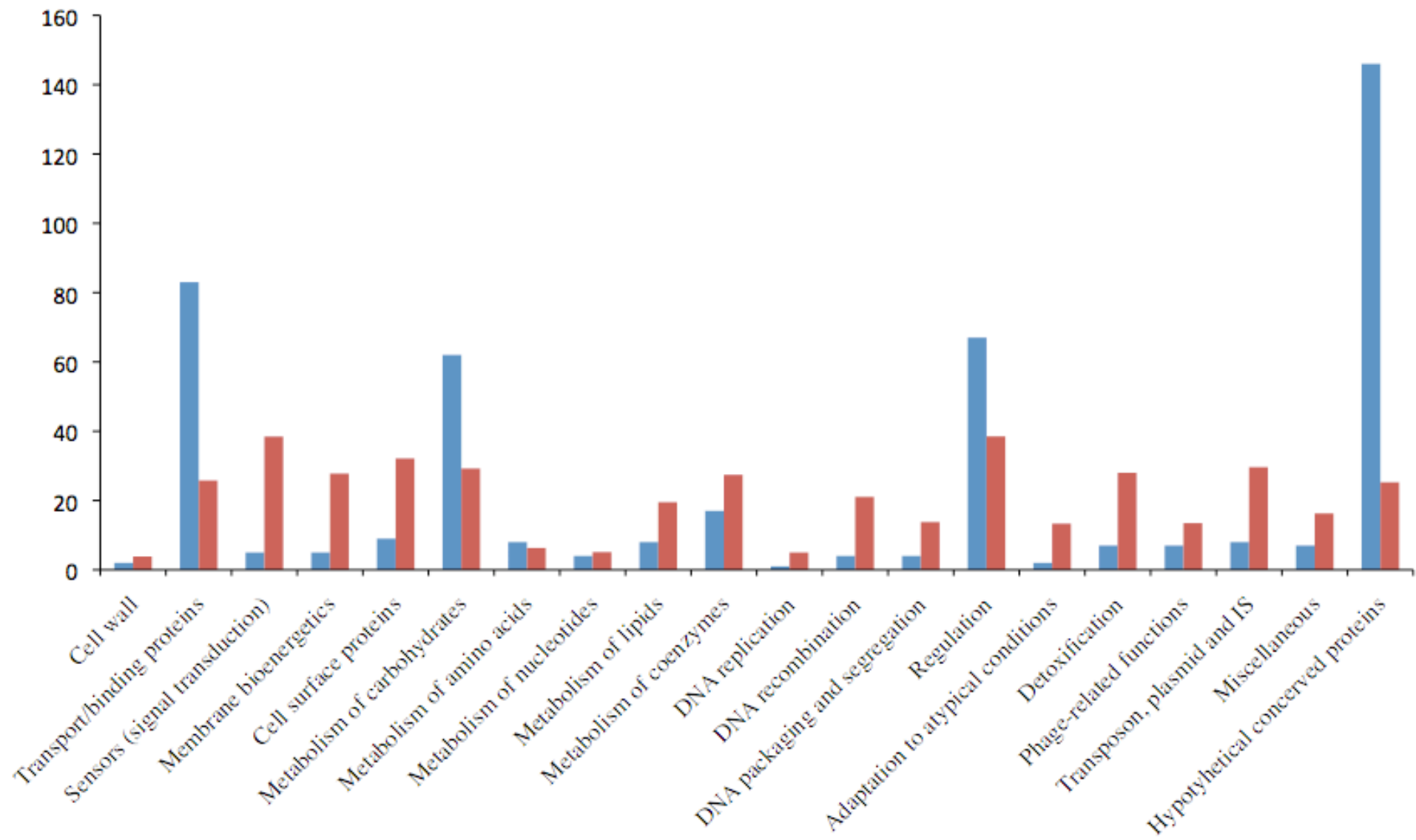


Figure S1:

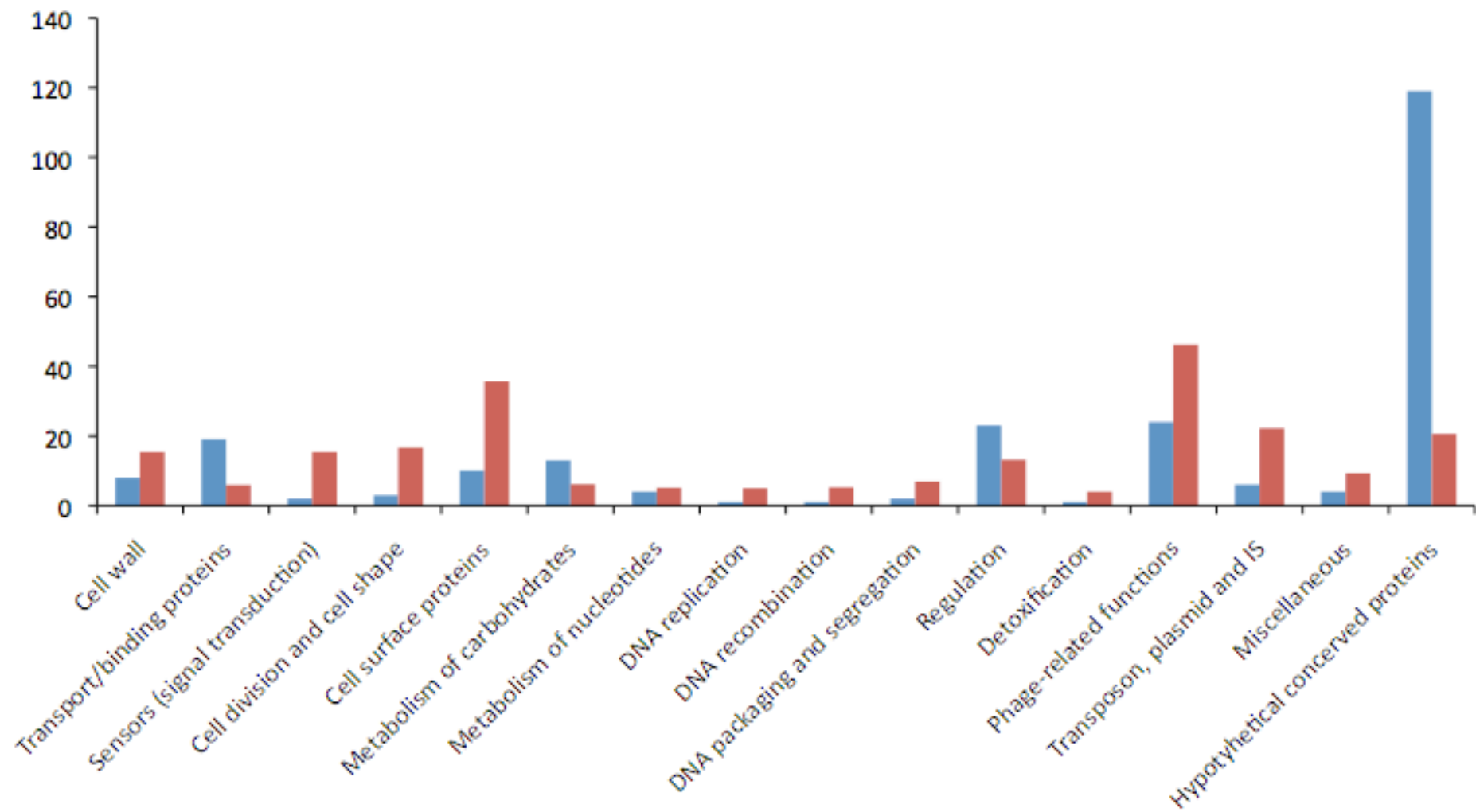


Figure S2:

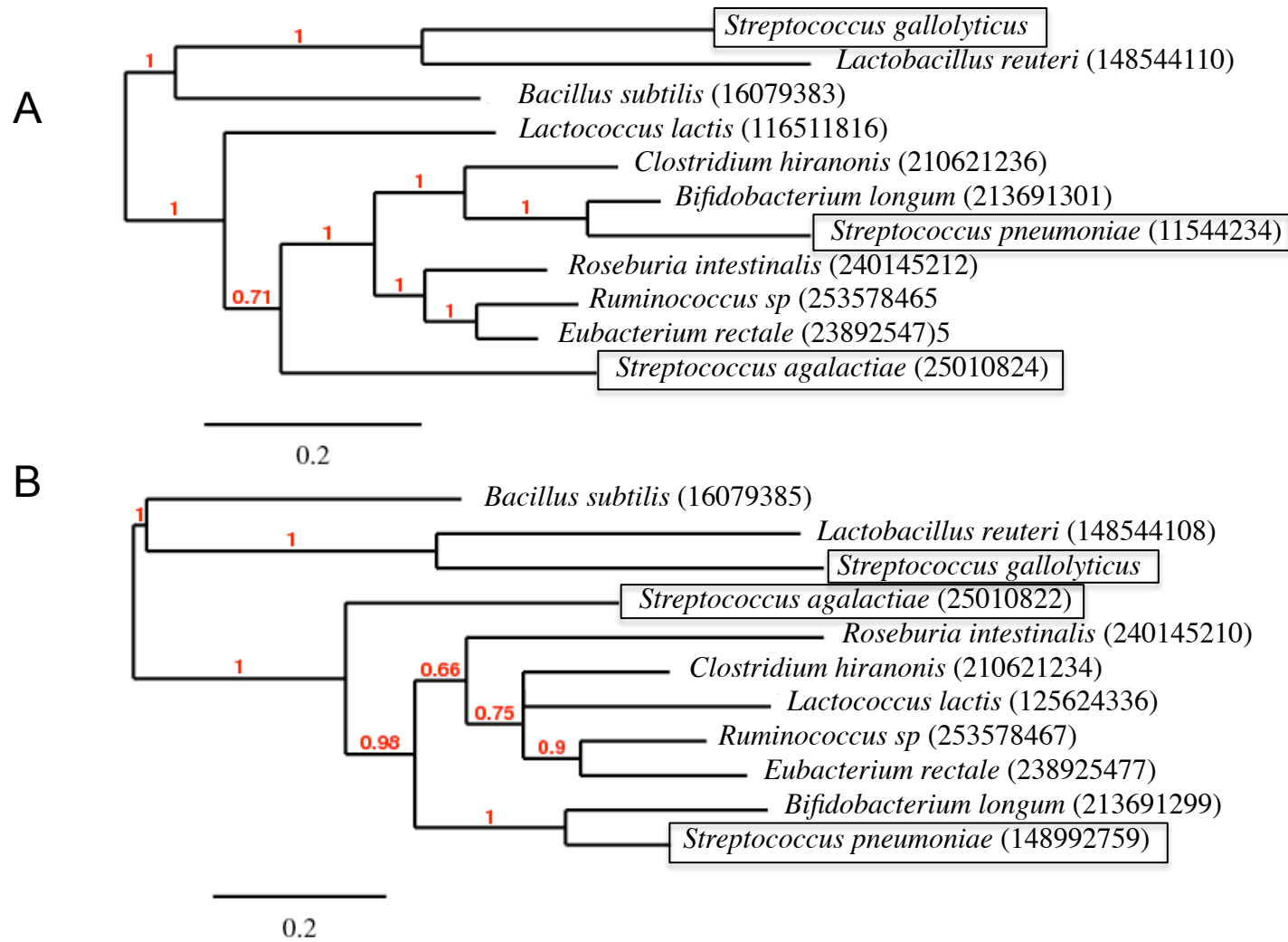


Figure S3: