Comparative genetics of the rdar morphotype in *Salmonella*

White, A. P.\(^1\) and M.G. Surette*\(^{1,2}\)

\(^1\) Department of Microbiology and Infectious Diseases, and \(^2\) Department of Biochemistry and Molecular Biology, University of Calgary, Calgary T2N 4N1, Alberta, Canada.

Running title: Genetics of *Salmonella* rdar morphotypes

Key Words: rdar morphotype, *agfD*, *csgD*, extracellular matrix, *Salmonella*

* corresponding author
ABSTRACT (223 words)

The *Salmonella* rdar morphotype is a distinct, rough and dry colony morphology formed by the extracellular interaction of thin aggregative fimbriae (Tafi or curli), cellulose and other polysaccharides. Cells in rdar colonies are more resistant to desiccation and exogenous stresses which is hypothesized to aid in the passage of pathogenic *Salmonella* spp. between hosts. Here we analyzed the genetic and phenotypic conservation of the rdar morphotype throughout the entire *Salmonella* genus. The rdar morphotype was conserved in 90% of 80 isolates representing all 7 *Salmonella* groups, however, the frequency was only 31% in a reference set of 16 strains (*Salmonella* reference collection C; SARC). Comparative gene expression analysis was used to separate *cis* and *trans* acting effects on promoter activity for the 16 SARC strains, focusing on the 780 bp intergenic region containing divergent promoters for the master regulator of the rdar morphotype (*agfD*) and the Tafi structural genes (*agfB*). Surprisingly, promoter functionality was conserved in most isolates and loss of the phenotype was primarily due to defects in *trans* acting regulatory factors. We hypothesize that *trans* differences have been caused by domestication whereas *cis* differences, detected for subsp. *arizonae* isolates, may reflect an evolutionary change in lifestyle. Our results demonstrate that the rdar morphotype is conserved throughout the salmonellae, but also emphasize that regulation is an important source of variability among isolates.
INTRODUCTION

Numerous bacterial species produce fimbriae, extracellular protein polymers that contribute to adherence, both to biotic (host) and abiotic surfaces. For pathogenic Salmonella spp., genome sequencing efforts have identified over 15 distinct fimbrial types (S. enterica serovars Typhimurium (43), Typhi (45), Paratyphi (42)). Pioneering work by Baumler et al. (5) and subsequent genomic comparisons (15, 46, 62) revealed that most fimbrial operons have a scattered distribution throughout the salmonellae. Early hypotheses that fimbriae were involved in adherence to host cells suggested that numerous fimbrial types would contribute to the host-specificities and tissue tropisms of different Salmonella spp. However, despite much research, the role of fimbriae in the pathogenesis of Salmonella is still not well understood. In addition, no clear links have been made connecting one fimbrial type to a particular animal host or disease process (62).

Thin aggregative fimbriae (Tafi, curli) are an exception to the general trend among Salmonella fimbriae. The divergent agfDEFG/BAC (csgDEFG/BAC) operons coding for Tafi biosynthesis have been detected in almost all Salmonella isolates tested to date (5, 19). Virtually identical operons have also been identified in E. coli and other enterobacterial species (8, 50, 70). In Salmonella, Tafi are produced together with cellulose (71), capsular and extracellular polysaccharides (18, 27, 68), and BapA (39), all of which come together to form a recalcitrant extracellular matrix that links individual cells together. Extracellular matrix production is associated with multicellular properties of Salmonella, both in the formation of the rdar morphotype (17, 53) and the formation of pellicles at the air-liquid interface in standing culture (52, 57). These phenotypes enhance the resistance of Salmonella to antimicrobial stresses (2, 56) and contribute to long-term survival (69).
The regulatory pathways controlling extracellular matrix production in *Salmonella* are outlined in Figure 1. Primary control takes place within the large intergenic region between the master regulator, *agfD*, and the Tafi structural genes, *agfBAC*. Transcription of *agfD* is dependent upon the stationary phase-inducible sigma factor, RpoS, and is maximal in the late exponential or early stationary phase of growth (26, 53, 69). Crl acts as a co-factor in this process by stimulating the binding of the RpoS-RNA polymerase complex to *agf* promoter regions (9, 49). *Trans* acting regulatory proteins required for *agfD* transcription include OmpR and MlrA. The arrangement of OmpR binding sites in the *agfD* promoter region is similar to the well-characterized *ompF* promoter (34) with a high-affinity binding site for activation under conditions of low osmolarity and low affinity binding sites that shut off transcription at higher osmolarity (24). The mechanism for MlrA activation of *agfD* transcription is unknown and no binding sites have been identified (14). Phosphorylated CpxR (CpxR-P) acts as a repressor of *agfD* transcription in high salt concentrations, through binding to sites in both *agfD* and *agfB* promoter regions (36, 47). Additional regulatory proteins affecting *agfD* transcription include HN-S and IHF (25, 36) as well as the RcsC/B (20) and TolQRA (64) systems.

Once AgfD is produced, it stimulates expression of several extracellular matrix components, either directly or indirectly (Figure 1). AgfD has an C-terminal DNA binding domain with homology to the LuxR family of transcriptional regulators (28) and N-terminal putative receiver domain (52). However, no activating signal has yet been identified. AgfD directly activates transcription of the *agfBAC* operon and *adrA*, which, in turn, activates cellulose biosynthesis (71). *Salmonella* capsular polysaccharide (27) and BapA (39) are also regulated by AgfD, but the exact mechanisms have not been determined. AgfD also positively regulates *glyA* (16) and represses transcription of several genes that inhibit biofilm formation (13). The growing
size and complexity of the AgfD regulon shows that this protein has multiple effects on cell physiology in Salmonella.

Since Tafi are one of the few conserved fimbrial types in Salmonella, it has been hypothesized that they may have a generalist function. However, the mere presence of Tafi genes does not prove that these organelles are produced by most or all isolates. To better assess the conservation of Tafi production and the rdar morphotype throughout the Salmonellae, we performed comparative genetic analysis of the important and highly variable agfD and agfB promoters from Salmonella reference collection C isolates (SARC, (10)). The SARC consists of 16 strains (SARC16) from an expanded set of 96 isolates (SARC96) from all phylogenetic lineages, including S. enterica subspecies (or group) I, II, IIIa, IIIb, IV, VI and S. bongori (group V) (10). The rdar morphotype was conserved in 80% of SARC96 isolates but in only 31% of SARC16 isolates. agfD and agfB promoter function was conserved in all SARC16 isolates, except two subsp. arizonae isolates that had clear sequence (cis) mutations resulting in inactive agfD and agfB promoters. Six of the remaining SARC16 isolates possessed upstream regulatory (trans) mutations. Three additional isolates reverted to the rdar morphotype when grown in long-term culture, via cis or trans changes that resulted in increased agfD transcription. Our results suggest that Tafi production and the associated rdar morphotype are conserved in all groups of Salmonella, except S. enterica subsp. arizonae. The predominance of regulatory mutations rather than structural gene mutations was unexpected and highlights the importance of cis and trans regulatory elements as a source of genetic and phenotypic variation.

MATERIALS AND METHODS

Bacterial strains, media, and growth conditions.
S. enterica serovar Typhimurium strain ATCC 14028 (ATCC 14028; American Type Culture Collection), serovar Enteritidis 27655-3b (SE 3b; (21)) and ATCC 14028 containing the SE 3b agfD promoter region (ST 3b; described below) were used as reference strains in this study. SARC isolates have been described previously (10). Strains were routinely grown for 20 h at 37 °C with agitation in 1% tryptone, pH 7.2 (T) or Miller’s Luria-Bertani broth (1.0% salt) supplemented with 50 µg/ml Kanamycin (Kn), 100 µg/ml Ampicillin (Ap), or 30 µg/ml Chloramphenicol (Cm) as required, before performing additional experiments. For growth of cells on agar, cultures were diluted to an optical density of 1.0 at 600 nm and 1 µl was spotted onto T media containing 1.5% Difco agar (T agar). To visualize the production of cellulose, cells were grown on T agar containing 200 µg/ml calcofluor white (fluorescent brightener 28; Sigma-Aldrich Canada). Colony and luminescence pictures were taken with a FluorChem 8900 camera system (Alpha Innotech).

Generation of luciferase reporters.

Template DNA from each Salmonella strain was prepared following the method of Walsh et al. (65). agfB and agfD, mlrA, or rpoS promoter-containing fragments were PCR amplified using primers agfD1 and agfD2 (69), mlrA1 (GATTAAACTCGAGCATACCCGCAA [a XhoI site is underlined]) and mlrA2 (GACGGATCCATCGTTTCACCCTTGCTC [a BamHI site is underlined]), or rpoS1 (GCCCTCGAGCAGGTCTGCACAAATTTC [a XhoI site is underlined]) and rpoS2 (GCGGGATCCGTATTCTGACTCAAAAGGTG [a BamHI site is underlined]). PCR products were purified and sequentially digested with XhoI and BamHI (Invitrogen Canada Inc.), and ligated using T4 DNA ligase (Invitrogen Canada Inc.) into pCS26-Pac (XhoI-BamHI) or pU220 (BamHI-XhoI) reporter vectors containing the luxCDABE operon from Photobacterium.
*luminescens* (7). The RpoS-responsive synthetic promoter::*lux* fusion sig38-H4 has been previously described (69). Plasmids were transformed into all *Salmonella* strains by electroporation (Gene Pulser 2.0; Bio-Rad Laboratories Inc.).

130 **Real-time gene expression/Bioluminescence assays**

Overnight cultures were diluted 1 in 600 in T broth to a final volume of 150 µl, and supplemented with antibiotics as required, in 96-well clear-bottom black plates (9520 Costar; Corning Inc.). The culture in each well was overlaid with 50 µl of mineral oil prior to starting the assays. Cultures were assayed for luminescence (0.1s; counts per second or CPS) and absorbance (620 nm, 0.1s) every 30 min during growth at 28 °C with agitation in a Wallac Victor² (Perkin-Elmer Life Sciences, Boston, Mass.). The background expression of the pCS26 and pU220 vectors is the instrument background (~150 CPS; (7)). A maximum expression from an average *Salmonella* promoter is 1000-5000 CPS and the dynamic range can reach greater than 1,000,000 CPS; *agfD* and *agfB* have very high activity. Generally, values greater than 400 CPS are considered significant.

**DNA sequencing and sequence alignments.**

Promoter::*lux* plasmids were purified from *Salmonella* (QIAspin miniprep kit, Qiagen Inc.) and DNA sequencing was performed by Macrogen (Seoul, South Korea) using primers pZE05 and pZE06 (7). DNA sequences were assembled using Contig Express (Vector NTI 7.0; Informax, Bethesda, MD). For each isolate, sequence discrepancies were resolved by sequencing additional clones. Multiple sequence alignment of the intergenic regions was performed using the Clustal W algorithm (Vector NTI 7.0; Informax, Bethesda, MD). The neighbor-joining
dendogram and bootstrap values, anchored on ATCC 14028, were calculated using MEGA version 3.0 (38).

**Preparation of ATCC 14028 genomic library and additional plasmid vectors.**

Purified ATCC 14028 genomic DNA was partially digested with Sau3a, and fragments were separated by sucrose density gradient centrifugation (55). 15 to 20 kb fragments were isolated, dialyzed against water and ligated into BamHI-digested pBR322 before transformation into Sarc1. Transformant colonies were isolated on LB Ap agar, pBR322 plasmids were purified (QIAspin miniprep kit, Qiagen) and DNA sequencing was performed (Macrogen, Seoul, South Korea) using primers pBR322seqB (AAGGAGCTGACTGGGTTGAAGG) and pBR322seqC (TCGGCACCGTCACCCCTGGA). Purified plasmids were re-transformed into Sarc1 to confirm that they conferred a switch in colony morphology.

DNA fragments containing mlrA or rpoS with native promoters were PCR amplified from ATCC 14028 using primers mlrAcloneFOR (GTCGGATCCCCAGATTTAAACTCGTACATAC [a BamHI site is underlined]) and mlrAcloneREV (GTCGGATCTCTGTGTTAAAACGCAAGG [a BamHI site is underlined]) or rpoScloneFor1 (GCCGAATTCCAGGTCTGCACAAATTC [an EcoRI site is underlined]) and rpoScloneREV (GCCAAGCTTTGACAAGGTTACTTACCTCGC [a HindIII site is underlined]). PCR products were purified and ligated into BamHI- or EcoRI- and HindIII-digested pBR322. To move rpoS into pACYC184, pBR322/rpoS was digested with EcoRI and the linearized plasmid was incubated at 68 °C for 30 min with 2.5 Units of Pfx DNA polymerase (Invitrogen Canada Inc.) to generate blunt-ended DNA. The linearized plasmid was digested with HindIII and the rpoS-containing fragment was ligated into HindIII and HincII-digested pACYC184 (Fermentas Life
Sciences). Plasmids were transformed into *Salmonella* strains by electroporation. Sarc9 was not included in these experiments because of its recalcitrance to genetic manipulation and its unique requirement for salt (>0.5%) in the growth media used.

**Generation of ATCC 14028 mutant strains**

An in-frame deletion of 949 bp in *rpoS* (encoding amino acids 7 to 323 in RpoS) was generated using overlap extension PCR (32) with primers rpoScloneFor1, rpoSRev1 (CGCTTCGATATCAGCGTATTCTGACTCA), rpoSFor2 (AATACGCTGATATCGAAGCGCTGTTCCG) and rpoSRev2 (GCCAAGCTTGTCGCAACATGACCCTGGT [a *HindIII* site is underlined]). Italicized sequences correspond to regions of identity between rpoSRev2 and rpoSFor2. PCR products were purified, digested with *Eco*RI and *Hind*III and ligated into pHSG415 (30). To generate strain ST 3b, the *agfD* promoter region from SE 3b was PCR amplified using primers agfD3b1 (AGTGAATTCGCTTCTTATCCGCTTCC [a *Eco*RI site is underlined]) and agfD3b2 (GTAAAGCTTTACTATCAAATCTAAACTTCAA [a *HindIII* site is underlined]) and cloned into pHSG415. Mutations were introduced into the chromosome of ATCC 14028 following established procedures (66, 67). *ΔrpoS* isolates were identified by smooth colony morphologies and lack of catalase activity when grown on T agar at 28 °C; chromosomal deletions in *rpoS* were confirmed by PCR. ST 3b isolates were selected by ability to form rdar colonies when grown at 37 °C on T agar; *agfD* promoter sequence mutations were confirmed by DNA sequencing.

**Long-term standing culture experiments.**
1 A$_{600}$ (~10$^9$ cells) of ATCC 14028, Sarc1, Sarc2, Sarc4, Sarc8, Sarc11, Sarc14 and Sarc16 were inoculated into 5 ml of 1% tryptone broth or Miller’s LB (1% NaCl) and grown in loosely-capped 16 x 125 mm borosilicate culture tubes at room temperature for up to 21 days. After pellicle formation had occurred, material was carefully removed from the air-liquid interface, resuspended in 1 ml of sterile PBS and broken up using a tissue homogenizer until uniform turbidity was reached (~ 20 s). Alternatively, cultures were vortexed to resuspend the pellicles. Homogenized pellicle mixtures were serially diluted onto T agar and incubated at RT to isolate individual colonies.

RESULTS

Conservation of the rdar morphotype throughout the salmonellae.

When SARC96 isolates were grown on T agar at RT for up to 14 days, three distinct phenotypic classes were observed. 79% of isolates formed rdar colonies with complete surface patterns that could be lifted off the agar surface intact, 4% formed colonies with incomplete patterns and 17% formed smooth, non-aggregative colonies without surface patterns (Table 1). Cellulose production was tested by growing isolates on agar containing calcofluor (57); all rdar$^+$ isolates were cellulose$^+$ (Table 1). In general, most incomplete pattern and smooth isolates were negative for cellulose production, with some exceptions (Table 1). S. enterica subspecies arizonae (group IIIa) was unique in that all isolates were smooth and did not produce cellulose (Table 1). In general, multicellular behaviour (rdar morphotype) was conserved in 6 of 7 Salmonella groups.

Smooth isolates were evenly distributed within the different Salmonella groups, but many belonged to the conventional SARC16 set, consisting of at least two isolates from each Salmonella subgroup (10). Four SARC16 isolates were rdar$^+$, one had an intermediate
phenotype, and the remaining 11 isolates were smooth (Table 1). For all *Salmonella* groups, except IIIa, the smooth colony morphologies of the SARC16 isolates were atypical compared to most isolates in the SARC96 set (Table 1).

**Conservation of agfD and agfB promoter activity in SARC16 isolates.**

To analyze the genetic conservation of Tafi production throughout the Salmonellae, the *agfDEFG-agfBAC* intergenic region was amplified from each SARC16 isolate and used to generate *agfD* and *agfB* promoter luciferase fusions (7). This DNA region included the 521 bp inter-transcript region and 5’ untranslated regions of *agfB* and *agfD* (24). We were unable to amplify *agfD-agfB* fragments from subsp. *arizonae* isolates Sarc5 and Sarc6; therefore, Sarc35 and Sarc37 isolates from the SARC96 collection were used in their place for all subsequent experiments (Table 1). Three subspecies I reference strains were included as controls in these experiments: serovar Typhimurium ATCC 14028 (ATCC 14028), serovar Enteritidis 27655-3b (SE 3b) and ATCC 14028 containing the *agfD* promoter region from SE 3b (ST 3b).

To determine if each set of promoters were functional and to compare their activities directly (i.e., independent of variation in levels of transcription factors), expression was measured in the same strain background, ATCC 14028. Surprisingly, *agfD* and *agfB* promoters from 17 of the 19 strains were active (Fig. 2A). The variation in magnitude for active *agfD* promoters varied from 96,000 to 520,000 luminescence counts per second (CPS), whereas for *agfB* the expression levels varied from 430,000 to 740,000 CPS (Fig. 2A). Some *cis* (sequence) variability was reflected in the variation in promoter strength. However, since the temporal pattern of expression remained conserved (Fig. 2B and C), it appears that important regulatory sites in the *agf* promoters are functionally conserved throughout the salmonellae. Peak
expression for \textit{agfD} promoters lasted approximately five to six hours (Fig. 2B), whereas for \textit{agfB} promoters the maximum expression lasted only 30 to 60 minutes (Fig 2C). The only inactive \textit{agf} promoters were obtained from subsp. \textit{arizonae} isolates Sarc35 and Sarc37; these promoters were inactive when ATCC 14028 was grown under a variety of additional conditions (data not shown).

**Sequence conservation of the \textit{agfBAC-agfDEFG} intergenic region between different \textit{Salmonella} isolates.**

Alignment of the \textit{agfB} and \textit{agfD} promoter sequences showed a high amount of divergence among the SARC16 isolates and three subspecies I reference strains (Fig. 3). The overall sequence identity for the region between \textit{agfB} and \textit{agfD} was only 67\%. Most sequence changes were group specific and clustered into regions not immediately adjacent to the \textit{agf} promoters (Fig. 3, black bars). Consistent with this, each \textit{Salmonella} subgroup formed a separate node in the alignment (Fig. 3).

Unique \textit{cis} changes were identified for the inactive \textit{agf} promoters from the subsp. \textit{arizonae} isolates, Sarc35 and Sarc37. For \textit{PagfB}, both isolates had a unique G to T sequence change in a recently predicted AgfD binding site (GGGTGAGTTA; (13)) near the -35 region (Fig. 3, upper circled region). For \textit{PagfD}, the \textit{arizonae} isolates both possessed three unique changes in the activating OmpR binding site required for \textit{agfD} transcription (24, 53) (Fig. 3, lower circled region). Two of the changes were in the most highly conserved nucleotides in the OmpR consensus binding sequence (ACNTTTNGNTACANNTAT; (24, 34).

SE 3b and ST 3b strains possessed a G to T transversion in the activating OmpR site adjacent to the -35 region of the \textit{agfD} promoter (Fig. 3). This specific promoter change relieves
RpoS-dependency and allows for \textit{agfD} transcription by RpoD ($\sigma^{70}$) (53). As a result, expression is more constitutive and increased in magnitude, allowing for Tafi and cellulose production at 37 °C. Sarc15 was also capable of producing Tafi and cellulose at 37 °C (Table 1), but no changes were detected in the activating OmpR binding site. One unique change was identified in the region containing putative OmpR binding sites D3-6 (24) further upstream of the Sarc15 \textit{agfD} promoter. This binding region has been linked to the repression of \textit{agfD} transcription by high levels of phosphorylated OmpR (24), and it is possible that the change identified may relieve this repression. The -10 and -35 promoter regions for both \textit{agfD} and \textit{agfB} were almost absolutely conserved in all \textit{Salmonella} isolates (Fig. 3, boxed regions). Sarc16 was the only strain where a change was identified, but this did not prevent transcription (Fig. 2A). Altogether, these results demonstrated that \textit{agfD} and \textit{agfB} promoter function was conserved for six of seven \textit{Salmonella} subgroups, despite sequence differences.

\textbf{Native \textit{agfD} and \textit{agfB} promoter expression in SARC16 isolates.}

To analyze trans regulatory differences between smooth and rdar+ SARC16 isolates, expression of each set of functional \textit{agfD} and \textit{agfB} promoters was tested in their native strain backgrounds (Fig. 4). Sarc35 and Sarc37 isolates were not included in this analysis since both isolates possessed non-functional \textit{agf} promoters. In general, \textit{agf} expression levels in all SARC16 isolates were reduced compared to the three subspecies I reference strains (Fig. 4). Differences were observed between the rdar+ and smooth isolates, however. All rdar+ SARC isolates had \textit{agfD} and \textit{agfB} expression levels above 10,000 CPS (Fig. 4; Sarc3, 7, 13 and 15). Sarc10, which had an intermediate rdar phenotype, also had expression of both promoters above 10,000 CPS (Fig. 4). All smooth isolates had \textit{agfB} expression below 10,000 CPS (Fig. 4) and only three strains had
agfD expression above 10,000 CPS (Fig. 4, Sarc1, 8 and 12). On average, expression in the smooth strain backgrounds was reduced 20-fold for agfD and > 1000 fold for agfB compared to expression of the same promoters in the ATCC 14028 background (Fig. 2A). In contrast, expression levels in rdar+ isolates were only reduced 4-fold for agfD and 20-fold for agfB.

RpoS activity in rdar+ and smooth SARC strains.

Since RpoS is one of the central regulators of agfD transcription and its activity can vary among Salmonella isolates (35, 48), we investigated if reduced agfD expression in the smooth SARC isolates could be correlated with a reduction in RpoS activity. We measured the expression of a synthetic RpoS-dependent promoter::lux fusion (sig38-H4; (69) in each strain background (Fig. 5). ATCC 14028, known to be RpoS+, had expression levels of 48,000 CPS. In contrast, known RpoS-deficient strains, ATCC 14028 ΔrpoS and SE 3b (1), had expression levels below 10,000 CPS (Fig. 5; dotted line). Most SARC16 isolates had expression above 30,000 CPS and were assumed to have functional RpoS. Four isolates were identified as putative rpoS mutants (Fig. 5; Sarc4, Sarc12, Sarc15 and Sarc16) and all, except Sarc15, had smooth colony morphologies (Table 1). Reduced RpoS activity was confirmed in three of four isolates by an absence of catalase activity when colonies were treated with hydrogen peroxide (data not shown).

Restoration of the rdar morphotype in Sarc1.

To determine the genetic defect(s) in Sarc1 (serovar Typhimurium), we transformed this isolate with an ATCC 14028 genomic DNA library in pBR322. Approximately 10,000 transformants were screened and three rdar+ isolates were obtained. Plasmids purified from all three isolates contained mlrA (yehV), a known positive regulator of agfD transcription (14, 23). When Sarc1
was transformed with pBR322 containing only *mlrA*, this was sufficient to restore the rdar morphotype (Fig. 6A). Overexpression of *mlrA* resulted in a 50-fold increase in *agfD* expression and a 200-fold increase in *agfB* expression (Fig. 6B). Sarc1 + pBR322/*mlrA* colony morphology (Fig. 6A, right panel) was similar to SE 3b, which has more constitutive expression of extracellular components (53). Consistent with this, overexpression of *mlrA* caused an increase in calcofluor binding, indicating that cellulose was produced (data not shown). We also detected a near two-fold increase in RpoS activity in Sarc1 + pBR333/*mlrA* (Fig. 6B). The significance of increased RpoS activity is not known, but may reflect changes in cell physiology and growth conditions due to over production of Tafi and cellulose.

**Restoring the rdar morphotype in smooth SARC16 strains.**

When pBR322/*mlrA* was transformed into additional smooth SARC16 isolates, Sarc2, Sarc8, and Sarc14 were restored to the rdar morphotype (Fig. 7, + *mlrA*). For each isolate, the change in colony morphology (Fig. 7; images on left) was correlated with an increase in luminescence from a native *agfB* promoter::*lux* reporter (Fig. 7; images on right). The rdar morphotype was not restored in RpoS-deficient isolates Sarc4 and Sarc16. When RpoS activity was restored by transformation with pACYC/*rpoS*, Sarc16 displayed a switch to rdar morphology (Fig. 7, + *rpoS*). For Sarc4, transformation with both *mlrA* and *rpoS* plasmids resulted in partial complementation of the rdar morphotype (Fig. 7, *mlrA* + *rpoS*). The incomplete pattern observed could reflect the presence of additional regulatory defects that were not fully complemented.

*m*lr*A* and *rpoS* expression in smooth SARC16 strains
Garcia et al. (23) recently described a smooth *S. enterica* serovar Typhimurium isolate with a deficiency in *mlrA* transcription, resulting in a lack of Tafi and cellulose production. Following from these results, we hypothesized that smooth SARC16 isolates might also have defects in native *mlrA* expression. However, expression of an ATCC 14028 *mlrA* promoter::*lux* reporter was not impaired in Sarc1, Sarc2, Sarc8 and Sarc14. This showed that regulatory networks required for *mlrA* transcription were intact in each isolate. In addition, *PmlrA::lux* reporters generated from each isolate had wild-type expression levels in ATCC 14028 (data not shown), indicating that the *mlrA* promoters were functional. Lastly, the predicted amino acid sequence of MlrA from each isolate was identical to ATCC 14028 (data not shown). These three experiments gave strong evidence that Sarc1, Sarc2, Sarc8 and Sarc14 were not impaired in *mlrA* expression and therefore, presumably had defects in other genes required for *agfD* transcription.

In contrast, reduced *mlrA* expression was detected in Sarc4 and Sarc16 (Table 2). This can be explained since Sarc4 and Sarc16 were predicted to be RpoS-deficient (Fig. 4) and *mlrA* transcription is dependent upon RpoS (14). Expression of an ATCC 14028 *PrpoS::lux* fusion was high in Sarc4 but reduced to background levels in Sarc16 (Table 2). This indicated that Sarc16 had defects in the upstream regulatory network required for *rpoS* transcription, whereas Sarc4 was primarily intact. Sarc16 also had a deletion of a single T in codon 55 of *rpoS*, predicted to result in a truncated protein. It was assumed that Sarc4 also possessed sequence alterations, but we were unable to PCR amplify *rpoS*-containing fragments from this isolate.

**Pellicle formation by smooth SARC strains.**

*Salmonella* spp. normally form pellicles at the air-liquid interface in standing liquid cultures; in rich media, this requires production of Tafi and cellulose (52, 57, 71). The air-liquid interface
represents a favourable niche that provides bacterial cells with increased access to oxygen. In other bacterial species (i.e., *Pseudomonas fluorescens*), reversion to pellicle formation, or adaptive divergence, can occur through mutations that increase the production of extracellular polymers (59, 60). Therefore, we investigated if smooth SARC16 isolates would form pellicles if grown in standing liquid culture for long time periods.

Pellicle formation by ATCC 14028 occurred within three days growth at RT in both LB and 1% tryptone (data not shown). Two smooth isolates, Sarc8 and Sarc14, formed thick pellicles in both media between 7-14 days of growth at RT (data not shown). In contrast, Sarc1 and Sarc2 did not form pellicles for the duration of the experiment. For Sarc4 and Sarc16, pellicle formation was observed between 14 and 21 days, but only in one culture medium (1% tryptone for Sarc4, LB for Sarc16). When pellicle material from each strain was homogenized and individual cells grown out on T agar, two colony types were observed, smooth (cellulose-') and rdar-like (cellulose+). When re-inoculated into standing liquid culture, the Sarc4 and Sarc16 rdar-like (revertant) isolates formed pellicles within 3-4 days of growth, whereas the smooth isolates did not (Fig. 8A, B). DNA sequencing proved that the revertant isolates possessed single base pair *cis* mutations in the *agfD* promoter region (Fig. 8C). The change for the Sarc4 revertant was within the activating OmpR binding site, whereas the Sarc16 revertant possessed a change in the -35 promoter region (Fig. 8C). Expression levels of the Sarc4 and Sarc16 revertant promoters were increased 3.5 fold and ~18 fold compared to native promoters (Fig. 8C). SE 3b was included for comparison purposes; the *agfD* promoter change in SE 3b, which is at the same position as the Sarc4 revertant, caused an 8.6 fold increase in expression levels (Fig. 8C).

The *S. bongori* isolate Sarc11 was also included in the pellicle-formation experiment. Between 14 and 21 days of growth, a thin pellicle began to form at the air-liquid interface in the
1% tryptone culture. Two different colony types were isolated from the pellicle material, non-aggregative and smooth (82%) and rdar (18%) (Fig. 9B). When re-inoculated into 1% tryptone, rdar isolates formed a thin pellicle at the air-liquid interface, whereas smooth isolates did not (Fig. 9A). The difference in pellicle-forming ability was not attributed to agfD promoter (cis) mutations, since both native and revertant isolates had identical sequences. To determine if native Sarc11 and the revertant isolate differed in their levels of trans regulators, we tested the expression of the same agfD and agfB promoter fusions in each strain background (Fig. 9C). In the revertant isolate, agfD expression was increased 16-fold and agfB expression was increased 2000-fold above native levels, whereas RpoS activity remained the same (Fig. 9C). This suggested that change(s) in trans regulatory factors in Sarc11 had occurred upstream of agfD rather than reversion within the agfDEFG or agfBAC operons.

DISCUSSION

The primary aim of this study was to determine the conservation of Tafi production and rdar morphotype formation throughout the salmonellae. In total, 80% of 96 isolates from all phylogenetic lineages of Salmonella were capable of thin aggregative fimbriae (Tafi or curli) and cellulose production and were rdar+. This extends previous reports on conservation of the rdar morphotype within S. enterica subgroup I: >90% of 800 serovar Typhimurium and Enteritidis isolates (51), 70% of 204 serovar Enteritidis isolates (57), and 72% of 71 isolates from 28 different serovars (58) from clinical, food, animal, and environmental sources. S. enterica subgroup I isolates primarily infect warm blooded hosts and are responsible for most human disease cases, whereas the other subgroups, including S. bongori, are primarily associated with cold-blooded hosts (11). Conservation of Tafi production and rdar morphotype formation in
these other subgroups suggests an important role in the lifecycles of diverse *Salmonella* isolates. Due to characteristic growth conditions (growth under 30 °C, nutrient limitation, low osmolarity; (17, 26) and increased resistance to dessication (27, 69) and anti-microbial agents (2, 56, 57, 69), we hypothesize that the rdar morphotype has a fundamental role for *Salmonella* survival outside host environments. This may contribute to contamination of different food products (4, 31) and passage between susceptible hosts.

The SARC16 isolates were analyzed in more detail to assess the genetic conservation of the *agfDEFG* and *agfBAC* operons within the entire *Salmonella* genus. Despite a large amount of sequence divergence, *agfD* and *agfB* promoter function was conserved in 14 isolates from six of seven *Salmonella* groups. In addition, the magnitude and profile of expression did not vary greatly between promoters from diverse isolates. The estimated evolutionary distances calculated from alignment of the entire 780 bp region between *agfDEFG* and *agfBAC* operons closely matched SARC strain comparisons based on over 12,000 bases of coding sequence (10). We interpret these results to indicate that most of the changes in the *agf* intergenic region that exist between isolates are “neutral” and representative of genetic drift, while essential regulatory regions required for *agfD* and *agfB* expression have been conserved. The divergence observed agrees with recent sequence comparisons by Peter Reeves and colleagues (33) showing that intergenic regions have higher mutation rates than coding sequences within serovar Typhimurium.

The most divergent sequences were obtained from *S. enterica* subsp. *arizonae* (group IIIa) and *S. bongori* (group V) isolates. This is consistent with previous DNA microarray studies (15, 46). However, only subsp. *arizonae* isolates possessed inactive *agfD* and *agfB* promoters. Clear sequence changes were identified in both Sarc35 and Sarc37: mutation of the activating
OmpR binding site for PagfD, and mutation of a predicted AgfD binding site (GGGTGAGTTA; (13) for PagfB (Table 3). These sequence changes likely prevent AgfD and OmpR binding and or subsequent activation of transcription. We could not amplify promoter regions from five additional subsp. arizonae isolates, including Sarc5 and Sarc6, possibly due to larger sequence differences. We believe that these cis changes are strong evidence of selection against Tafi production and the rdar morphotype in subsp. arizonae. One of the major differences from other

Salmonella groups is that subsp. arizonae isolates are common gut inhabitants of reptiles and snakes and could be part of the commensal microflora in these animals (41). Thus, subsp. arizonae isolates may not need to survive outside of their hosts for long time periods. In addition, subsp. arizonae isolates are known to cause disease primarily in patients who are immunocompromised (41). Whether the difference in agf promoter function in subsp. arizonae can be related to decreased infectivity has yet to be determined.

Even though most SARC16 isolates possessed functional agf promoters, 11 of 16 isolates did not produce Tafi and cellulose when grown on T agar. The rdar morphotype was restored in six smooth isolates by increased copy number of known trans regulators of agfD transcription (Table 3). Each of the four isolates where rdar was restored by increased copy number of mlrA had normal levels of native mlrA expression and the precise mutations were not identified. To explain our results, we concluded that over-expression of mlrA compensated for other trans regulatory defects, resulting in increased expression of agfD and production of downstream extracellular components. We assume that the trans defects are in the same upstream pathway required for agfD transcription, but do not effect mlrA transcription. The other isolates, Sarc4 and Sarc16, were identified as being RpoS-deficient. Mutations in rpoS have been detected in
other *Salmonella* isolates and are thought to reflect either natural variation or laboratory-induced changes (35, 48, 61).

Adaptive divergence (i.e. pellicle formation) was observed for three smooth SARC16 isolates previously unable to colonize the air-liquid interface of standing cultures. Each isolate acquired mutations that elevated *agfD* expression, either directly through *cis* mutations in the *PagfD* region (Sarc4 and Sarc16) or through *trans* regulatory changes upstream of *agfD* (Sarc11) (Table 3). The nucleotide change in Sarc4 was at the same position as the *PagfD* change in SE 3b (53) and may represent a mutational hotspot within the OmpR binding region (ACNTTTNGNTACANTAT; (24)). For Sarc16, the change was in the -35 region and shifted the *PagfD* sequence closer to the $\sigma^{70}$ consensus (40). In *E. coli*, enhanced Tafi (curli) production has been linked to *PagfD* (Pcs*D*) mutations in the -10 region that move the sequence closer to the $\sigma^{70}$ consensus (63). The changes that occurred in Sarc4 and Sarc16 were also predicted to allow for $\sigma^{70}$-based transcription of *agfD*. Sarc15 had similar properties to the revertant isolates and was also predicted to be RpoS-deficient. Our results suggest that *Salmonella* isolates lacking native RpoS activity can easily revert to rdar formation by acquiring *cis* promoter mutations that activate *agfD* transcription. For Sarc11, the precise regulatory change was not identified and may have occurred through several different pathways, although the defect was determined not to be in *mlrA* (data not shown).

The high prevalence of *trans* regulatory mutations in the SARC16 isolates may be the result of domestication. Bacille Calmette-Guérin (BCG) is one well-documented example of this phenomenon (6), which has also occurred in several commonly studied bacterial pathogens (22). The initial descriptions of curli (Tafi) (44) and positive regulators *crl* (3) and *mlrA* (14) were in *E. coli* HB101, a commonly used strain with *trans* regulatory mutations. Garcia et al. (23) have
also recently described a variant of commonly used *S. enterica* serovar Typhimurium SL1344 with a *trans* regulatory mutation. Furthermore, repeated subculturing of *E. coli* OH157:H7 (63) and *S. enterica* serovar Typhimurium ((53); C. D. Davidson, A. P. White and M. G. Surette, unpublished data) is known to induce phenotype switching with respect to the rdar morphotype. The discovery of *trans* mutations was surprising since the 4.4 kb *agfDEFG/BAC* (Tafi) and 14.2 kb *bcsABZC/EFG* (cellulose) operons represent large regions where inactivating mutations could accumulate. It is possible that *cis* mutations and/or structural gene mutations are more indicative of a change in lifestyle, such as what has been observed in *Shigella spp.* (54). Collectively, our experiments show that SARC16 phenotypes are not representative of the salmonellae and the rdar morphotype may be more prevalent than we have measured.

In the modular description of cellular organization (29), the whole Tafi, cellulose and extracellular matrix network could exist as a single “survival” module under the control of AgfD (CsgD). With the discovery of BapA (39) and an O-antigen capsule (27), the network of AgfD-regulated extracellular components is growing. Our results demonstrate that loss of the rdar morphotype in *Salmonella* primarily results from regulatory mutations affecting AgfD expression and not from mutation in genes for Tafi or cellulose biosynthesis. Thus, AgfD is the point of integration of multiple physiological and environmental inputs. Detailed promoter sequence and function comparisons allowed us to separate *cis* and *trans* effects on regulation. Overall, diverse isolates within the *Salmonella* genus have retained the genetic capacity and phenotypic ability to produce the extracellular matrix, which may contribute to the world-wide persistence of these important pathogens.

REFERENCES


Table 1. Prevalence of multicellular pattern formation (rdar morphotype) in the salmonellae (*Salmonella* reference collection C; (10)).

<table>
<thead>
<tr>
<th><em>Salmonella</em> strains&lt;sup&gt;a&lt;/sup&gt;</th>
<th>No. strains tested</th>
<th>Pattern (rdar)</th>
<th>Incomplete pattern</th>
<th>No pattern</th>
<th>Calcofluor binding&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subgroup I</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 1-11</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>SARC16-1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-2</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 12-33</td>
<td>22</td>
<td>18*</td>
<td>2</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>SARC16-3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup IIIa</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-5</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-6</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96-35</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96-37</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup IIIb</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 38-42</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>SARC16-7</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-8</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup IV</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 43-70</td>
<td>28</td>
<td>24</td>
<td>2</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>SARC16-9</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup V</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 71-83</td>
<td>13</td>
<td>10*</td>
<td>0</td>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>SARC16-11</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>+/-</td>
</tr>
<tr>
<td>SARC16-12</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup VI</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 84-92</td>
<td>9</td>
<td>8*</td>
<td>0</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>SARC16-13</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-14</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Subgroup IV&lt;sup&gt;d&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96 93-96</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>SARC16-15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC16-16</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Totals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SARC96</td>
<td>96</td>
<td>76 (79%)</td>
<td>4 (4%)</td>
<td>16 (17%)</td>
<td>80 (83%)</td>
</tr>
<tr>
<td>without SARC16</td>
<td>80</td>
<td>72 (90%)</td>
<td>3 (4%)</td>
<td>5 (6%)</td>
<td>74 (93%)</td>
</tr>
<tr>
<td>SARC16</td>
<td>16</td>
<td>4 (25%)</td>
<td>1 (6%)</td>
<td>11 (69%)</td>
<td>6 (38%)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Strain subgroups were determined by the patterns of the associated *Salmonella* reference collection C.

<sup>b</sup> No. isolates with indicated phenotype.

<sup>c</sup> No. strains tested.

<sup>d</sup> Subgroup I examples.
a SARC16 isolates represent #2, 9, 12, 20, 34, 36, 41, 42, 47, 59, 73, 76, 89, 91, 93 and 94 within SARC96 and are referred to as Sarc1-16 in the main text. SARC96 #35 and 37 are referred to as Sarc35 and Sarc37 in the main text. Group V has been designated *S. bongori*; all other groups are part of *S. enterica*.

b Colony morphology and associated phenotypes were recorded after growth on T agar at 28 °C for up to seven days; pattern (rdar) - colonies had complete surface patterns and could be lifted intact from the agar surface; Incomplete pattern - colonies could not be lifted intact from the agar surface; No pattern - colonies were non-aggregative and smooth. SARC96 #15, 74, 86 and 93 isolates also formed rdar colonies at 37 °C (*).

c Fluorescence of colonies grown on T agar supplemented with 200 µg/ml calcofluor was observed under a 366 nm UV light source; +/-, strains with fluorescence intermediate between ATCC 14028 positive control and Δ*bcsA* negative control strains.

d These four strains were originally designated as a separate subgroup (VII; (10) related to subgroup IV, but this classification has not been widely adopted (12). Therefore, the four isolates are considered members of *S. enterica* subgroup IV (subspecies *houtenae*).
Table 2. Variability of P\textit{mlrA} and P\textit{rpoS} expression in smooth SARC16 strains.

<table>
<thead>
<tr>
<th>Strain\textsuperscript{a}</th>
<th>Promoter expression\textsuperscript{b}</th>
<th>\textit{mlrA}</th>
<th>\textit{rpoS}</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATCC 14028</td>
<td></td>
<td>5200 ± 580</td>
<td>45800 ± 940</td>
</tr>
<tr>
<td>Sarc1</td>
<td></td>
<td>6600 ± 470</td>
<td></td>
</tr>
<tr>
<td>Sarc2</td>
<td></td>
<td>6600 ± 90</td>
<td></td>
</tr>
<tr>
<td>Sarc4</td>
<td></td>
<td>160 ± 80</td>
<td>35400 ± 5300</td>
</tr>
<tr>
<td>Sarc8</td>
<td></td>
<td>14400 ± 410</td>
<td></td>
</tr>
<tr>
<td>Sarc14</td>
<td></td>
<td>5700 ± 370</td>
<td></td>
</tr>
<tr>
<td>Sarc16</td>
<td></td>
<td>130 ± 30</td>
<td>700 ± 260</td>
</tr>
</tbody>
</table>

\textsuperscript{a} \textit{Salmonella} strains as described in Figure 2.

\textsuperscript{b} Expression levels of P\textit{mlrA} and P\textit{rpoS::lux} fusions generated from ATCC 14028 were measured in each strain background. Values represent the mean and standard deviations of maximum luminescence (counts per second) from at least three independent experiments.
Table 3. Summary of *cis* and *trans* mutations contributing to the smooth phenotypes of select SARC isolates.

<table>
<thead>
<tr>
<th>SubSpecies (Group)</th>
<th>Isolate</th>
<th>Complementation&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Reversion&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Defect identified</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>enterica</em> (I)</td>
<td>Sarc1</td>
<td><em>mlrA</em></td>
<td></td>
<td><em>Trans</em> regulatory changes</td>
</tr>
<tr>
<td></td>
<td>Sarc2</td>
<td><em>mlrA</em></td>
<td></td>
<td><em>Trans</em> regulatory changes</td>
</tr>
<tr>
<td><em>salamae</em> (II)</td>
<td>Sarc4</td>
<td><em>mlrA</em> + <em>rpoS</em></td>
<td><em>PagfD cis</em> mutation (OmpR binding site)</td>
<td>Reduced RpoS activity; <em>Trans</em> regulatory changes</td>
</tr>
<tr>
<td><em>arizonae</em> (IIIa)</td>
<td>Sarc35</td>
<td>X</td>
<td></td>
<td><em>PagfD and PagfB cis</em> mutations</td>
</tr>
<tr>
<td></td>
<td>Sarc37</td>
<td>X</td>
<td></td>
<td><em>PagfD and PagfB cis</em> mutations</td>
</tr>
<tr>
<td><em>houtenae</em> (IV)</td>
<td>Sarc8</td>
<td><em>mlrA</em></td>
<td><em>PagfD cis</em> mutation (OmpR binding site)</td>
<td><em>Trans</em> regulatory changes</td>
</tr>
<tr>
<td></td>
<td>Sarc16</td>
<td><em>rpoS</em></td>
<td><em>PagfD cis</em> mutation (-35 region)</td>
<td>Reduced PrpoS activity; truncated RpoS</td>
</tr>
<tr>
<td><em>S. bongori</em> (V)</td>
<td>Sarc11</td>
<td>X</td>
<td><em>Trans</em> regulatory changes upstream of agfD</td>
<td><em>Trans</em> regulatory changes</td>
</tr>
<tr>
<td></td>
<td>Sarc12</td>
<td>X</td>
<td></td>
<td>Reduced RpoS activity; <em>Trans</em> regulatory changes</td>
</tr>
<tr>
<td><em>indica</em> (VI)</td>
<td>Sarc14</td>
<td><em>mlrA</em></td>
<td></td>
<td><em>Trans</em> regulatory changes</td>
</tr>
</tbody>
</table>

<sup>a</sup> Isolates formed colonies of the rdar morphotype on T or LB agar at 28°C when transformed with genes on multi-copy plasmids pBR322 or pACYC184.

<sup>b</sup> Pellicle-forming revertants were identified when SARC isolates were grown in 1% tryptone or LB standing culture at RT for 14-21 days.
Figure Legends.

Figure 1. Complex regulation of *Salmonella* extracellular matrix production.

Components regulating *agfD* transcription are indicated in boxes above the *agfDEFG* (*csgDEFG*) and *agfBAC* (*csgBAC*) operons. HN-S and IHF binding proteins and RcsBCD and TolQRA regulatory systems were not included due to the complex nature of their actions on *agfD* transcription. Arrows leading from AgfD show extracellular components that are transcriptionally regulated by AgfD. AdrA generates cyclic-di-GMP (c-di-GMP), a signalling molecule that activates cellulose biosynthesis by integral membrane Bcs (bacterial cellulose synthesis) enzymes (37). Since BapA (39) and capsular polysaccharides (CPS; 27) were only recently discovered, their precise roles in multicellular aggregation are not well understood.

Figure 2. Cis variability of *agfD* and *agfB* promoters in *Salmonella*.

Expression of *agfD* and *agfB* promoter::*lux* fusions from 19 isolates representing seven different *Salmonella* groups was measured in ATCC 14028. (A) Maximum luminescence (counts per second; CPS) values for each reporter strain are represented as the mean and standard deviation from at least three independent experiments. *Salmonella* isolates corresponding to each promoter are listed below the x-axis: SARC isolates are numbered as described in Table 1; 14028, serovar Typhimurium ATCC 14028; ST 3b, ATCC 14028-3b; SE 3b, serovar Enteritidis 27655-3b. *Salmonella* subgroup numbers are listed below each set of strains. Light production of *agfD::*lux (B) or *agfB::*lux (C) reporter strains as a function of time, normalized by the maximum CPS of each strain. The dashed line represents the level of maximum luminescence. Results from one representative experiment are shown.
Figure 3. Multiple sequence alignment of the \textit{agfBAC}-\textit{agfDEFG} intergenic region from representative isolates of each \textit{Salmonella} subgroup.

Conserved bases are shown in grey, base pair differences are shown in black, and gaps in the alignment are represented by white spaces. Transcriptional start sites (arrows) and -10 and -35 promoter regions (boxes) for \textit{agfB} and \textit{agfD} are displayed as annotated in the serovar Typhimurium LT2 genome (43). Important regulatory regions or binding sites for OmpR (hatched boxes), CpxR (black boxes), H-NS (grey box), and IHF (white box) are displayed above the alignment (24, 36, 47). The circled regions highlight changes in a putative AgfD binding site (13) and the activating OmpR site adjacent to the \textit{agfD} promoter (24). Neighbor-joining dendograms are shown on the left of the alignment and bootstrap values based on 1,000 computer-generated trees are indicated at the nodes. Strains and groups are listed as described in Figure 2. Sequence and alignments are available from the author.

Figure 4. Trans variability of \textit{agfD} and \textit{agfB} promoter expression in SARC isolates.

Expression of native \textit{agfD} and \textit{agfB} promoter::\textit{lux} reporters in SARC16 and \textit{Salmonella} subgroup I reference strains. Maximum luminescence (counts per second; CPS) values for each reporter strain are represented as the mean and standard deviation from at least three independent experiments. The dotted line marks the threshold level of \textit{agfD} and \textit{agfB} expression that correlates with rdar colony morphology. \textit{Salmonella} strains and group numbers listed below the x-axis were described in Figure 2.

Figure 5. RpoS activity in SARC isolates and \textit{Salmonella} subgroup I reference strains.
Luminescence (counts per second; CPS) from a synthetic RpoS-responsive promoter::lux fusion (sig38-H4) was measured in each strain during growth in 1% tryptone for 48 h at 28 °C. Bars represent the mean and standard deviation of maximum CPS values from at least three independent experiments. The dotted line marks the threshold level of RpoS activity; strains with expression at or below the line are known or assumed to be RpoS-deficient. *Salmonella* strains and group numbers listed below the x-axis were described in Fig. 2. An isogenic rpoS mutant strain of ATCC 14028 (∆rpoS) and pACYC/rpoS complemented strain (∆rpoS +) were included as controls for RpoS activity.

**Figure 6. Restoration of the rdar morphotype in Sarc1.**

(A) Colony morphology of Sarc1 +/- pBR322/mlrA grown on T agar at 28 °C for 48 h. (B) Luminescence (counts per second; CPS) of key promoters in Sarc1 (black bars) or Sarc1 + pBR322/mlrA (grey bars) was measured during growth in 1% tryptone for 48 h at 28 °C. Bars represent average and standard deviations of maximum CPS values from at least three independent experiments.

**Figure 7. Restoration of agfB expression and the rdar morphotype in smooth SARC16 isolates.**

ATCC 14028, Sarc2, Sarc4, Sarc8, Sarc14 and Sarc16 strains containing native agfB::lux reporter plasmids were transformed with pBR322/mlrA, pACYC/rpoS or both plasmids together. Cells were inoculated onto T agar or LB agar (Sarc4) and grown at 28 °C for 48 h. Colony morphology on the left was recorded with a reflective light source; luminescence from each
colony is represented by images on the right. Increased luminescence is predicted to be representative of Tafi production.

**Figure 8. Pellicle formation by smooth SARC16 strains.**

Revertant isolates capable of pellicle formation were isolated from the air-liquid interface of standing liquid cultures of Sarc4 and Sarc16. Pellicle-forming ability was tested for native isolates and revertant isolates during growth in 1% tryptone (A) or LB (B) for 3 days at RT. (C) agfD promoter sequences were compared between ATCC 14028, SE 3b, Sarc4, Sarc16 and the two revertant (rev) isolates. The activating OmpR binding site is underlined and the -35 promoter sequence is in bold. (*) The fold-increase in agfD transcription represents the relative expression levels of promoter::lux fusions from each isolate measured in SE 3b ΔagfD (27), Sarc4 or Sarc16 strain backgrounds. Values represent the ratios of maximum luminescence of SE 3b divided by ATCC 14028, Sarc4rev divided by Sarc4, and Sarc16rev divided by Sarc16, plus or minus the standard deviation from at least three independent experiments.

**Figure 9. Reversion of multicellular aggregation in Sarc11 (S. bongori).**

Native Sarc11 and a pellicle-forming revertant isolate were tested for (A) pellicle-forming ability when grown in 1% tryptone at RT for 7 days, (B) colony morphology when grown on T agar at 28 °C for 48 h and (C) levels of gene expression. (C) Luminescence (counts per second; CPS) of native agfD and agfB promoter::lux fusions and the sig38-H4 RpoS-responsive reporter was measured in each strain background during growth in 1% tryptone at 28 °C for 48 h. Bars represent the mean and standard deviations of maximum CPS values from at least three independent experiments.
ACKNOWLEDGEMENTS.

We thank Ken Sanderson for providing access to SARC strains at the *Salmonella* Genetic Stock Centre (University of Calgary), and K.S., M. Elowitz, B. Bassler, W. Kay and E. Crump for critical reading of the manuscript. This work was supported by grants from the Canadian Institutes of Health Research to MGS and through Genome Prairie, Genome BC and Inimex Pharmaceuticals through the “Functional Pathogenomics of Mucosal Immunity” project. MGS is supported as an Alberta Heritage Foundation for Medical Research (AHFMR) Senior Scholar and Canada Research Chair in Microbial Gene Expression. APW is supported by a Postdoctoral Fellowship from AHFMR.
RpoS activity

Maximum CPS

10^6

10^5

10^4

10^3

10^2

10^1

14028  ST 3b  SE 3b  Sarc1  Sarc2  Sarc3  Sarc4  Sarc35  Sarc37  Sarc7  Sarc8  Sarc10  Sarc15  Sarc16  Sarc11  Sarc12  Sarc13  Sarc14  ΔrpoS  ΔrpoS +

I  II  IIIa  IIIb  IV  V  VI
A

Sarc1

Sarc1 + mlrA

B

Maximum CPS

\[
\begin{align*}
agfD::lux & : 10^4 \\
agfB::lux & : 10^5 \\
RpoS activity & : 10^5
\end{align*}
\]

\[
\begin{align*}
Sarc1 & \quad \text{Sarc1} \\
+mlrA & \quad +mlrA
\end{align*}
\]
<table>
<thead>
<tr>
<th></th>
<th>14028</th>
<th>Sarc2</th>
<th>Sarc4</th>
<th>Sarc8</th>
<th>Sarc14</th>
<th>Sarc16</th>
</tr>
</thead>
<tbody>
<tr>
<td>agfB::lux</td>
<td><img src="" alt="Image" /></td>
<td><img src="Sarc2.png" alt="Image" /></td>
<td><img src="Sarc4.png" alt="Image" /></td>
<td><img src="Sarc8.png" alt="Image" /></td>
<td><img src="Sarc14.png" alt="Image" /></td>
<td><img src="Sarc16.png" alt="Image" /></td>
</tr>
<tr>
<td>+ mlrA</td>
<td><img src="" alt="Image" /></td>
<td><img src="Sarc2+mlrA.png" alt="Image" /></td>
<td><img src="Sarc4+mlrA.png" alt="Image" /></td>
<td><img src="Sarc8+mlrA.png" alt="Image" /></td>
<td><img src="Sarc14+mlrA.png" alt="Image" /></td>
<td><img src="Sarc16+mlrA.png" alt="Image" /></td>
</tr>
<tr>
<td>+ rpoS</td>
<td><img src="" alt="Image" /></td>
<td><img src="Sarc2+rpoS.png" alt="Image" /></td>
<td><img src="Sarc4+rpoS.png" alt="Image" /></td>
<td><img src="Sarc8+rpoS.png" alt="Image" /></td>
<td><img src="Sarc14+rpoS.png" alt="Image" /></td>
<td><img src="Sarc16+rpoS.png" alt="Image" /></td>
</tr>
<tr>
<td>mlrA + rpoS</td>
<td><img src="" alt="Image" /></td>
<td><img src="Sarc2+mlrA+rpoS.png" alt="Image" /></td>
<td><img src="Sarc4+mlrA+rpoS.png" alt="Image" /></td>
<td><img src="Sarc8+mlrA+rpoS.png" alt="Image" /></td>
<td><img src="Sarc14+mlrA+rpoS.png" alt="Image" /></td>
<td><img src="Sarc16+mlrA+rpoS.png" alt="Image" /></td>
</tr>
</tbody>
</table>
### Table C

<table>
<thead>
<tr>
<th>Strain</th>
<th>14028</th>
<th>SE 3b</th>
<th>Sarc4</th>
<th>Sarc4rev</th>
<th>Sarc16</th>
<th>Sarc16rev</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TTACATTTGGTTACAAGTTTAAACAC<strong>TTGCTTT</strong>TA</td>
<td>------------</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>AC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fold-increase in agfD transcription</td>
<td>8.64 ± 0.47</td>
<td></td>
<td></td>
<td>3.51 ± 0.12</td>
<td>17.7 ± 3.2</td>
<td></td>
</tr>
</tbody>
</table>