

RESEARCH ARTICLE

Open Access



Comparative genomics and evolution of the amylase-binding proteins of oral streptococci

Elaine M. Haase^{1*}, Yurong Kou^{1,4}, Amarpreet Sabharwal¹, Yu-Chieh Liao², Tianying Lan³, Charlotte Lindqvist³ and Frank A. Scannapieco¹

Abstract

Background: Successful commensal bacteria have evolved to maintain colonization in challenging environments. The oral viridans streptococci are pioneer colonizers of dental plaque biofilm. Some of these bacteria have adapted to life in the oral cavity by binding salivary α -amylase, which hydrolyzes dietary starch, thus providing a source of nutrition. Oral streptococcal species bind α -amylase by expressing a variety of amylase-binding proteins (ABPs). Here we determine the genotypic basis of amylase binding where proteins of diverse size and function share a common phenotype.

Results: ABPs were detected in culture supernatants of 27 of 59 strains representing 13 oral *Streptococcus* species screened using the amylase-ligand binding assay. N-terminal sequences from ABPs of diverse size were obtained from 18 strains representing six oral streptococcal species. Genome sequencing and BLAST searches using N-terminal sequences, protein size, and key words identified the gene associated with each ABP. Among the sequenced ABPs, 14 matched amylase-binding protein A (AbpA), 6 matched amylase-binding protein B (AbpB), and 11 unique ABPs were identified as peptidoglycan-binding, glutamine ABC-type transporter, hypothetical, or choline-binding proteins. Alignment and phylogenetic analyses performed to ascertain evolutionary relationships revealed that ABPs cluster into at least six distinct, unrelated families (AbpA, AbpB, and four novel ABPs) with no phylogenetic evidence that one group evolved from another, and no single ancestral gene found within each group. AbpA-like sequences can be divided into five subgroups based on the N-terminal sequences. Comparative genomics focusing on the *abpA* gene locus provides evidence of horizontal gene transfer.

Conclusion: The acquisition of an ABP by oral streptococci provides an interesting example of adaptive evolution.

Keywords: Commensal, Adaptation, Amylase, Phylogenetics, Horizontal gene transfer

Background

The importance of commensal bacteria to the health and wellbeing of the human host is now becoming more recognized. Successful commensals find ways to maintain colonization in challenging environments. In the oral cavity, bacteria are subjected to extremes in pH, the fluctuating availability of nutrients, the stress of salivary flow, and the physical shear forces mediated by chewing and swallowing. Oral bacteria have evolved numerous

mechanisms that influence colonization of this environment. One such mechanism is the production of surface proteins that facilitate adhesive interactions with host and bacterial ligands. These interactions often result in enzymatic activity and/or intra- and inter-species signaling functions that help to maintain a relatively stable collection of bacterial species comprising the oral microbiome.

Viridans streptococci are the predominant commensal bacteria colonizing the oral cavity and dental plaque biofilm. These bacteria cluster into five groups (mitis, mutans, anginosus, salivarius and bovis) based on 16S rRNA gene sequencing [1, 2]. Some species within each

* Correspondence: haase@buffalo.edu

¹Department of Oral Biology, School of Dental Medicine, University at Buffalo, State University of New York, Buffalo, NY, USA

Full list of author information is available at the end of the article



group, with the exception of bovis, have surface proteins that bind α -amylase [3–6], the predominant enzyme and one of the most abundant components in the saliva of many mammalian species [7–9].

Amylase, together with other salivary components, forms a pellicle that covers all surfaces of the oral cavity [10]. Bacteria have evolved mechanisms to adhere to the salivary pellicle by binding to various components within it, thus providing the substrate for dental plaque formation. Several species of oral streptococci produce cell wall-associated amylase-binding proteins [3, 6, 11]. In addition to mediating the adhesion of bacteria to the salivary pellicle, the binding of host amylase to the bacterial cell surface can facilitate starch metabolism and bacterial growth [12, 13].

Three types of amylase-binding proteins (ABPs) have been formally named, amylase-binding protein A (AbpA) [14], amylase-binding protein B (AbpB) [15], and amylase-binding protein C (AbpC) [16]. AbpA, essential for amylase binding to the streptococcal surface, is a unique protein with no known conserved domains. It is co-transcribed with the downstream sortase B (*srtB*) gene, shown to be essential for covalent attachment of AbpA to the bacterial cell wall [17–19]. AbpB, which is not essential for amylase binding to the bacterial cell surface [15], is homologous to and functions as a dipeptidyl-peptidase [20]. Currently, in the NCBI database, AbpA and AbpB are designated as ‘amylase-binding protein’ and ‘peptidase C69’, respectively. AbpC differs considerably from AbpA and AbpB, and its functional significance remains to be determined [16].

In addition to these three ABPs, oral streptococcal genomes encode other ABPs with molecular weights ranging from 20 to 87 kDa [3]. To date, there has been no formal analysis of the relationship of each of these proteins to each other. We postulate that some or all of these proteins are related and have been acquired by interspecies horizontal gene transfer followed by mutation or recombination. The goal of this study was to therefore investigate how ABPs of oral streptococci are evolutionarily related. Here, using a combination of approaches, we provide evidence of the amylase-binding phenotype among a diverse group of oral streptococci, and the possible acquisition of these genes through horizontal gene transfer. These results provide an interesting example of adaptive evolution.

Results

Detection of amylase-binding proteins

At the time these experiments were initiated in 2013, few ORFs were annotated as ABPs in the NCBI database. The aim of this study was to search for additional ABPs both in oral streptococcal isolates and in silico.

We began by screening 59 oral streptococcal isolates from our collection (Table 1) for the presence of ABPs using the in vitro amylase-ligand binding assay. Although AbpA is initially associated with the cell wall in mid-log phase, by stationary phase AbpA is secreted into the supernatant providing easy access to the native protein. Therefore, AbpA-like proteins and other potential ABPs were obtained from concentrated culture supernatants, resolved by SDS-PAGE, followed by staining with Coomassie blue (Fig. 1a). Proteins from a duplicate gel run at the same time were electrotransferred to polyvinylidene fluoride (PVDF) membrane and assayed for binding to salivary α -amylase in the amylase ligand-binding assay (Fig. 1b). To further characterize these ABPs, selected bands, as listed in Table 2 and outlined by a box in Fig. 1a, were cut from a corresponding Coomassie blue-stained blot (not shown) and sent for N-terminal sequencing. The N-terminal sequence of ABPs from *Streptococcus parasanguinis* MGH413 and *Streptococcus salivarius* KB005 were determined previously (unpublished).

Several streptococcal isolates expressed more than one ABP. Two ABPs were identified previously in *Streptococcus gordonii* Challis CH1, namely AbpA (20 kDa) and AbpB (82 kDa) [14, 15]. As shown in Fig. 1b, additional ABPs from different strains were identified ranging from 20 to 87 kDa. N-terminal sequencing of all ABPs was cost prohibitive, therefore, proteins were selected to represent a range of molecular weights. Of the 36 protein bands sent for sequencing, 33 bands yielded N-terminal sequences for further in silico analysis. Results from this and a previous study [3] revealed that 41 of 79 (52%) oral streptococcal strains showed the ability to bind amylase by one or more methods (Table 1). Based on these data, none (0/6) of the anginosus group (*S. anginosus*, *S. intermedius*); 64% (30/47) of the mitis group (*S. australis*, *S. cristatus*, *S. gordonii*, *S. infantis*, *S. mitis*, *S. oralis*, *S. parasanguinis*, *S. sanguinis*); none (0/15) of the mutans group (*S. mutans*, *S. sobrinus*), and all (11/11) of the salivarius group (*S. salivarius*) were able to bind salivary α -amylase. As previously noted, all *S. gordonii* and no *S. mutans* strains bound amylase. Draft genomes of *S. infantis* strains ATCC 700779 and SK970 do not carry a *abpA* or *srtB* sequence, unlike *S. infantis* strains SK1076, SK1302, UC921A, and UC6950A. Thus, it is not surprising that *S. infantis* ATCC 700779 was negative in the amylase-ligand binding assay. Many *Streptococcus* species were heterogeneous with respect to amylase binding [6]. Of the 59 strains screened, 18 were selected for genome sequencing. The ability of bacterial cells from each of these strains to bind amylase correlated with the amylase binding of their supernatant proteins (Table 1).

Table 1 Amylase binding by streptococcus strains in our collection

Streptococcus species	Strain designation	Amylase ligand-binding overlay ^a (ca. mw kDa)	Amylase-binding activity ^b	GenBank accession number	
<i>S. anginosus</i>	ATCC 33397 (NCTC 10713)	-	-		
	UC2953	-	ND		
	UC9218	-	ND		
	NCTC 10708 ^c	-	-		
	NCTC 10709 ^c	-	-		
<i>S. australis</i>	ATCC 700641	22, 35, 86	+	NZ_AFUD01000015	
<i>S. cristatus</i>	CC5A	26, 26, 84	+	NZ_JYGJ00000000 ^{dj}	
	CR3	28, 28, 84	+	NZ_JYBK00000000 ^{dj}	
	CR311 (ATCC 51100)	30, 82 ^e	+	NZ_AFUE01000002	
<i>S. gordonii</i>	ATCC 10558 (NCTC 7865)	20, 82 ^e	ND		
	Blackburn (NCTC 10231)	20, 82 ^e	ND		
	Challis CH1	20, 82 ^e	+	NC_009785	
	I141 ⁱ	20, 84	+	NZ_JYOZ00000000 ^d	
	CN2814	20, 82 ^e	ND		
	FAS4	20, 82 ^e	ND		
	GEO2	20, 82 ^e	ND		
	G9B	20, 82 ^e	+	NZ_JYGL00000000 ^{dj}	
	UB10712 ^{fi}	20, 37, 82 ^e	+	NZ_JYGN00000000 ^d	
	JF2	20, 82 ^e			
	LGR2	20, 82 ^e	ND		
	MJ2	20, 82 ^e	ND		
	M5	20, 82 ^e	ND		
	SPED3	20, 82 ^e	ND		
<i>S. infantis</i>	ATCC 700779	-	-	PRJNA158721	
	UC921A ⁱ	26, 30	+	NZ_JYGT00000000 ^d	
	UC6950A ⁱ	30	+	NZ_JYOV00000000 ^d	
<i>S. intermedius</i>	ATCC 27335 (SK54)	-	-	PRJNA197004	
<i>S. mitis</i>	N551/SD142 (ATCC 49456) (NCTC 12261)	36 ^g	+	EF989012.1	
	OT25	36	+	NZ_JYGP00000000 ^{d, j}	
	SK137	36, 50, 63	+	NZ_JYDQ00000000 ^{dj}	
	SK145	37, 48, 65	+	NZ_JYGS00000000 ^{dj}	
	UC2948	-	-		
	UC3161	-	-		
	<i>S. mutans</i>	NCTC 10449 (ATCC 25175)	-	-	
		BM71	-	ND	
		GS5	-	ND	
Ingbritt		-	-		
LT11		-	ND		
OMZ175		-	ND		
NG8		-	ND		
VT321		-	ND		
V202	-	ND			
<i>S. oralis</i>					

Table 1 Amylase binding by streptococcus strains in our collection (Continued)

<i>ssp oralis</i>	COL85/1862 ⁱ	26, 29	+	NZ_JYGM00000000 ^d
<i>ssp oralis</i>	OP51 ⁱ	25, 26, 30	+	NZ_JYGO00000000 ^d
<i>ssp oralis</i>	SK141 ^{h,i}	26, 26, 30	+	NZ_JYGR00000000 ^d
<i>ssp tigurinus</i>	UC5873 ⁱ	26, 30	+	NZ_JYGU00000000 ^d
	ATCC 10557 (NCTC 7864)	-	-	
	BU174	-. ^e	ND	
	C104	-	ND	
	CR834	-. ^e	ND	
	KS32AR	-	ND	
	MPD1	-. ^e	ND	
	SK92	-	-	
	34	-	ND	
	9811	-. ^e	ND	
	16532AR	-	ND	
<i>S. parasanguinis</i>	MGH413	21, 87	+	NZ_JYOW00000000 ^{d,j}
	VT517 ⁱ	22, 84	+	NZ_JYPA00000000 ^d
<i>S. salivarius</i>	KB005	29	+	NZ_JYOX00000000 ^{d,j}
	CDC013	29	ND	
	CDC017	29	ND	
	CM103	29	ND	
	MG568	29	ND	
	MG587	29, 51	ND	
	MG691	29, 49	ND	
	UC810	29	ND	
	UC3162	29, 50	+	NZ_JYOY00000000 ^d
	TOUR	29	ND	
	13,419-R	29	ND	
<i>S. sanguinis</i>	ATCC 10556 (NCTC 7863)	-	-	
	HPC1	-	ND	
	MPC1	-. ^e	ND	
	UC9433	-	-	
	804	-	ND	
<i>S. sobrinus</i>	B13	-	ND	
	SL1	-	-	
	K1R	-	ND	
	OMZ 176	-	ND	
	6715-WT13	-	ND	

ND not done

^aAmylase-ligand binding assay: detects the ability of proteins from 20x concentrated culture supernatant to bind salivary amylase in whole saliva by Far-Western blot (Methods)

^bAmylase-binding activity: measures the ability of bacterial cells to bind salivary amylase in whole saliva and reduce hydrolysis in starch agarose. (+) ABP, (-) ABP on bacterial cells (Methods)

^c*S. milleri* reclassified as *S. anginosus* in Ruoff et al. [48]

^dSabharwal et al. [22]; strains sequenced for this study

^eBrown et al. [3]

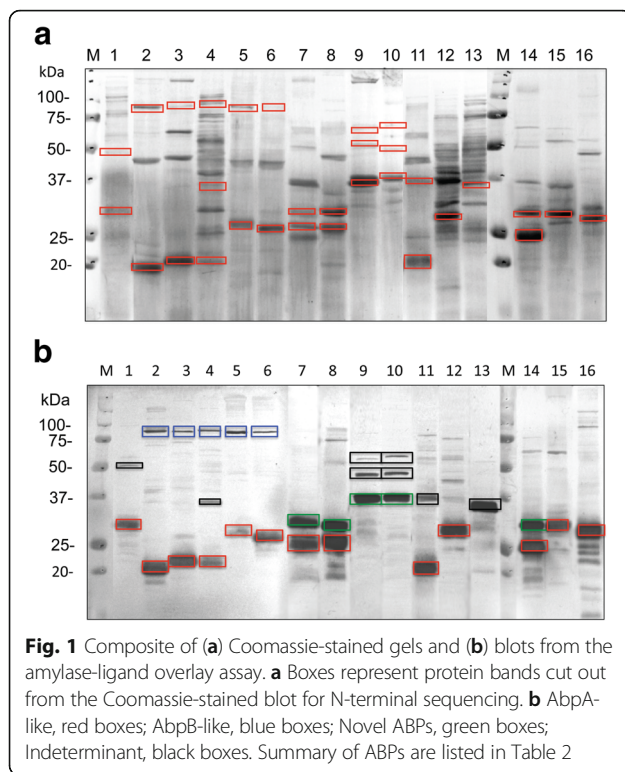
^fStrain labeled NCTC 10712 (*S. mitis*) in our collection is now corrected to *S. gordonii*, UB10712 upon whole genome sequencing

^gVorrasi et al. [16]

^hThe strain from our collection named *S. mitis* in Brown et al. [3] Submitted to GenBank as *S. mitis*; updated to *S. oralis* in 2015 as per reclassification Takenouchi-Ohkubo et al. [28] and Kilian et al. [49]

ⁱUpdated taxonomic classifications for Mitis group, Jensen et al. [23]

^jReference strains used for genome post-assembly: *S. cristatus* ATCC 51100 for *S. cristatus* CC5A and CR3; *S. gordonii* Challis substrain CH1 for *S. gordonii* G9B; *S. mitis* B6 for *S. mitis* OT25, SK145, SK137; *S. parasanguinis* ATCC 15912 for *S. parasanguinis* MGH413; *S. salivarius* CCH53 for *S. salivarius* KB005



Assembly of whole genome sequencing data

Sequenced genomes were assembled de novo

The recently developed algorithm, MyPro [21], facilitated assembly and annotation of the draft genomes. When closely related streptococcal genomes were available for reference, a post-assembly algorithm was used to reduce gaps in the de novo assemblies [21, 22]. Post-assembled genomes consisted of fewer gaps in the assembly, with a median of 6 contigs. De novo assembled genomes ranged in size from 1.79 to 2.29 Mb, with an average G + C content of 40.8% and 1916 coding sequences [22].

Taxonomic designation

There have been changes in the taxonomic assignment of several mitis group *Streptococcus* species subsequent to the availability of comparative phylogenetic analyses of core genomes, multilocus sequence analysis (MLSA), and 16S rRNA gene sequence data [23]. In the current study, revised streptococcal taxonomy showed changes in species nomenclature of 25 strains used in phylogenetic analyses. Briefly, several members of *mitis*, *oligofermentans*, *pseudopneumoniae*, and *sanguinis* species that represent ABS were revised to other *Streptococcus* species based on MLSA and whole genome core phylogeny [23]. Taxonomic reassignments are summarized in Additional file 1: Table S1, and updates have also been made to our draft genome submissions [22].

Identification of AbpA-like sequences

The AbpA protein sequence (20–26 kDa) from our newly sequenced strains and homologs obtained in silico were used to further categorize AbpA-like producing streptococcal strains into five subgroups based on the N-terminal sequence. In-depth phylogenetic analysis of AbpA for this study was confined to predominant human oral streptococcal strains and to strains where a sortase B (*srtB*) gene was immediately downstream of the putative *abpA* gene.

Comparison of complete *abpA* gene sequences revealed more heterogeneity among amylase-binding streptococci than previously observed. The *Streptococcus* species and strains within each AbpA-like subgroup along with the associated N-terminal sequences are listed in Additional file 2: Table S2. Subgroup 1 consists primarily of *S. gordonii* strains, but also includes a strain of *S. cristatus*. The well-characterized AbpA protein of *S. gordonii* Challis CH1 is a member of this subgroup [12, 14]. Subgroup 2 is predominantly *S. cristatus* strains, but also contains the Gram-positive opportunistic pathogen *Gemella haemolysans*, a member of the human flora of the oral cavity and upper respiratory tract [24, 25]. Since *G. haemolysans abpA* lacks the typical downstream *srtB* gene, it was excluded from phylogenetic analysis. *S. salivarius* dominates Subgroup 3, but also contains some *S. oralis* and *S. parasanguinis* strains. Subgroup 4 is a mix of *S. oralis*, *S. infantis*, and *S. parasanguinis*. Subgroup 5 is the most diverse group; although *S. parasanguinis* predominates, *S. australis*, *S. infantis*, and *Streptococcus vestibularis* strains are represented among others. The N-terminal sequences not only serve to subgroup AbpA-like proteins, but may contribute to the ability of AbpA to bind amylase, as in *S. gordonii* [19, 26].

The AbpA-like protein sequences from each of the subgroups were aligned using CLUSTAL Omega [27]. AbpA sequences of individual subgroup members are well conserved with the exception of subgroup 5, which is the most diverse. When the consensus sequence from each subgroup was aligned, the signal sequence, sortase B-binding motif, and C-terminal end are the most concordant (Fig. 2). These protein sequences contain 195 to 228 amino acid residues and each subgroup revealed a similar arrangement schematically represented in Fig. 3, i.e. a well-conserved signal sequence of 23 residues, followed by one of five subgroup-specific N-terminal consensus sequences (11–13 residues), followed by a variable region of low to moderate homology (152–186 residues), and ending with a putative sortase B-binding motif [LP(K/N)T(S/A/H)] and a termination sequence [A(V/A)K].

Identification of AbpB-like sequences

Several strains of streptococci produce an 80–87 kDa ABP that is consistent with AbpB (Fig. 1b, Additional

Table 2 Amylase-binding protein summary from Fig. 1b blot

Lane no.	Species	Strain	No. Bands	AbpA-like ca. mw	AbpB-like ca. mw	Novel ca. mw
1	<i>S. salivarius</i>	UC3162	2	29 ^b		50 ^c
2	<i>S. gordonii</i>	I141	2	20 ^b	84 ^e	
3	<i>S. parasanguinis</i>	VT517	2	22 ^b	84 ^e	
4	<i>S. australis</i>	ATCC 700641	3	22 ^e	86 ^e	35 ^d
5	<i>S. cristatus</i>	CR3	2	28 ^b	84 ^e	28 ^{b,f}
6	<i>S. cristatus</i>	CC5A	2	26 ^e	84 ^e	26 ^{b,f}
7	<i>S. oralis</i>	SK141	2	26 ^e		26 ^{b,g} 30 ^{b,h}
8	<i>S. infantis</i>	UC921A	2	26 ^b		30 ^{b,h}
9	<i>S. mitis</i>	SK137	3	None	None	36 ^{b,i} ,50 ^c ,63 ^d
10	<i>S. mitis</i>	SK145	3	None	None	37 ^{b,i} ,48 ^d ,65 ^c
11	<i>S. gordonii</i>	UB10712	2	20 ^b	82 ^e	37 ^d
12	<i>S. oralis</i>	COL85/1862	1	29 ^b		26 ^{e,g}
13	<i>S. mitis</i>	OT25	1	None		36 ^d
14	<i>S. oralis</i>	OP51	2	25 ^b		26 ^{e,g} ,30 ^{b,h}
15	<i>S. oralis</i>	UC5873	1	30 ^b		26 ^{e,g}
16	<i>S. infantis</i>	UC6950	1	30 ^b		
	<i>S. parasanguinis</i> ^a	MGH413	2	21 ^b	87 ^e	
	<i>S. cristatus</i> ^a	NCTC 51100/CR311	2	30 ^b	82 ^e	26 ^{e,f}
	<i>S. salivarius</i> ^a	KB005	1	29 ^b		

^aN-terminal sequences determined previously, not published

^bSequence determined based on N-terminal sequence

^cInsufficient quantity for N-terminal sequencing

^dIdentity indeterminant from N-terminal sequence

^eSequence based on whole genome sequence and/or primer walking

^fPeptidoglycan-binding protein

^gGlutamine ABC transporter

^hHypothetical protein

ⁱCholine-binding protein

file 3: Table S3). Given that the N-terminal amino acid sequences of putative AbpBs were indeterminant due to low protein content in the bands, genomic sequences were instead obtained by primer walking supplemented with NCBI database searches using key words. These data matched a highly conserved group of proteins from the peptidase C69 family (Additional file 4: Figure S1, Additional file 3: Table S3). Previously, a dipeptidyl-peptidase function was confirmed experimentally for AbpB from *S. gordonii* Challis CH1 [20].

Identification of novel ABPs

Several strains expressing ABPs between 30 and 65 kDa, with N-terminal sequences unlike AbpA or AbpB, were considered novel (Additional file 5: Table S4). Four distinct N-terminal sequences were identified that were consistent with previously described proteins including the LysM superfamily, a peptidoglycan-binding protein; glutamine ABC-type transporter, substrate-binding protein; a hypothetical protein; and a choline-binding protein, cell wall- or glucan-binding domain. Novel proteins were

identified in seven of the isolates from this study (CR3, CC5A, SK141, OP51, UC5873, COL85/1862, UC921A), as well as CR311 (ATCC 51100) sequenced earlier. The multiple alignments with other proteins from different *Streptococcus* species obtained from BLAST are presented in Additional file 6: Figure S2; alignments within each group of proteins were well conserved. The AbpC protein from *S. mitis* NS51, identified previously [16], could be aligned with the choline-binding proteins from *S. oralis* strains. Although a glutamine ABC-transporter substrate-binding protein from *Streptococcus pneumoniae* R6 and a choline-binding protein from *S. pneumoniae* SMRU2014 align with novel ABPs, to date no *S. pneumoniae* strains have been shown experimentally to bind salivary amylase.

Sequence similarity network (SSN)

Based on the least stringent alignment score, the SSN grouped all AbpAs into one cluster and all AbpBs into another cluster, while novel ABPs were grouped into four separate clusters (Fig. 4a). The distinct clustering

Subgroup 1

S.cristatus_CC5A	1	MKKVLLSSVVALSLFALAVPFAA	DEATNAARQGD	--160AA--	LPKTSAVK	203
S.gordonii_771SOLI	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARNND	--156AA--	LPKTSAVK	199
S.gordonii_IE35	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARKND	--153AA--	LPKTSAVK	196
S.gordonii_I141	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARKND	--153AA--	LPKTSAVK	196
S.gordonii_CH1	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARNND	--152AA--	LPKTSAVK	195
S.gordonii_KCOM1506	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARNND	--152AA--	LPKTSAVK	195
S.gordonii_UB10712	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARNND	--152AA--	LPKTSAVK	195
S.gordonii_2_1_36FAA	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARNND	--152AA--	LPKTSAVK	195
S.gordonii_787SOLI	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARNND	--152AA--	LPKTSAVK	195
S.gordonii_G9B	1	MKKVLLSSVVALTLFAAAAPVFS	ADEATDAARKND	--152AA--	LPKTSAVK	195

****:*****:*** *.*:*****:***: * *****

Subgroup 2

S.cristatus_CR3	1	MKKVLLSSVVALSLFAVAAPAF	ENPINGGANTPGA	--167AA--	LPKTSAVK	211
S.cristatus_CR311/ATCC51100	1	MKKVLLSSVVALSLFAVAAPAF	ENPINGGANTPGA	--167AA--	LPKTSAVK	211
S.cristatus_142SOLI	1	MKKVLLSSVVALSLFAVAAPAF	ENPINGGANTPGA	--167AA--	LPKTSAVK	211
S.cristatus_1014SOLI	1	MKKVLLSSVVALSLFAVAAPAF	ADNSINGGANTPGA	--164AA--	LPKTSAVK	208
S.cristatus_1015SOLI	1	MKKVLLSSVVALSLFAVAAPAF	ADNSINGGANTPGA	--164AA--	LPKTSAVK	208

*****:*****:*** ***** *****

Subgroup 3

S.sp._HSISS3	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_KB005	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_M18	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_CCHSS3	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_57_I	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_726SSAL	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_1270SSAL	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.salivarius_UC3162	1	MKKVLLTSAAVLALFAAAAPAV	ADINGGANTPGA	--170AA--	LPNTAAVK	212
S.oralis_ssp._oralis_COL85/1862	1	MKKVLLSSVVALAVFSAAAPAF	ADINGGANTPGA	--173AA--	LPKTHAAK	212
S.oralis_ssp._tigurinus_UC5873	1	MKKVLLSSVVALAVFSAAAPAF	ADINGGANTPGA	--173AA--	LPKTHAAK	215
S.parasanguinis_C1A	1	MKKVLLSSVAALAVFAAAAPAL	ADNNGGANTPRG	--167AA--	LPKTSAVK	209
S.parasanguinis_1287SPAR	1	MKKVLLSSVVALAVFAAAAPAL	ADNNGGANTPRG	--157AA--	LPKTSAVK	199
S.parasanguinis_392SPAR	1	MKKVLLSSVAALAVFAAAAPAL	ADNNGGANTPRG	--162AA--	LPKTSAVK	204

*****:.....*:*****:*** ***** *

Subgroup 4

S.infantis_UC6950A	1	MKKVLLSSVVALAVFSAAAPAF	ADNSYSSNALI	--186AA--	LPKTHAAK	227
S.parasanguinis_349SPAR	1	MKKVLLSSVAALAVFAAAAPAF	ADNSYSSNALI	--158AA--	LPKTSAVK	199
S.oralis_ssp._oralis_OP51	1	MKKVLLSSVAALAVFSAAAPAF	ADNSYSSNALI	--169AA--	LPKTHAAK	210
S.oralis_SK141	1	MKKVLLSSVAALAVFSAAAPAF	ADNSYSSNALI	--169AA--	LPKTHAAK	210
S.infantis_UC921A	1	MKKVLLSSVAALAVFSAAAPAF	ADNSYSSNALI	--171AA--	LPKTHAAK	212
S.infantis_SPAR10	1	MKKVLLSSVAALAVFSAAAPAF	ADNSYSSNALI	--168AA--	LPKTHAVK	209

*****:.....*:*****:*** ***** *

Subgroup 5

S.vestibularis_F0396	1	MKKVLLTSAAVLAVFAAGTQVFA	QGENPKNSNQLT	--174AA--	LPNTAAVK	216
S.vestibularis_ATCC49124	1	MKKVLLTSAAVLAVFAAGTQVFA	QGENPKNSNQLT	--174AA--	LPNTAAVK	216
S.vestibularis_1004SSPC	1	MKKVLLTSAAVLAVFAAGTQVFA	QGENPKNSNQLT	--174AA--	LPNTAAVK	216
S.vestibularis_1005STHE	1	MKKVLLTSAAVLAVFAAGTQVFA	QGENPKNSNQLT	--174AA--	LPNTAAVK	216
S.oralis_ssp._dentisani_bv.2.str.F0932	1	MKKVLLSSVAALAVFSAAAPVFA	QGENPNSSNQLI	--169AA--	LPKTHAAK	211
S.sp._1116SGOR	1	MKKVLLSSVVALTLFAAAAPVFA	QGENAESSNQLI	--152AA--	LPKTSAVK	194
S.infantis_SK1076	1	MKKVLLSSVAALAVFSAAAPVFA	QGENPASSNQLI	--177AA--	LPKTHAVK	219
S.infantis_900SORA	1	MKKVLLASVAALAVFSAAAPVFA	QGENPASSNQLI	--171AA--	LPKTHAAK	213
S.infantis_1172SPSE	1	MKKVLLASVAALAVFSAAAPVFA	QGENPASSNQLI	--186AA--	LPKTHAAK	228
S.infantis_SK1302	1	MKKVLLASVAALAVFSAAAPVFA	QGENPASSNQLI	--179AA--	LPKTHAAK	221
S.infantis_434SPSE	1	MKKVLLASVAALAVFSAAAPVFA	QGENPASSNQLI	--170AA--	LPKTHAAK	212
S.infantis_SK140	1	MKKVLLASVAALAVFSAAAPVFA	QGENPASSNQLI	--172AA--	LPKTHAAK	214
S.infantis_469SPSE	1	MKKVLLASVAALAVFSAAAPVFA	QGENPASSNQLI	--171AA--	LPKTHAAK	213
S.parasanguinis_MGH413	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSSSNQLI	--166AA--	LPKTSAVK	208
S.parasanguinis_65SPAR	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPNASNQLI	--159AA--	LPKTSAVK	201
S.parasanguinis_FW213	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSASNQLI	--165AA--	LPKTSAVK	207
S.parasanguinis_ATCC15912	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSASNQLI	--165AA--	LPKTSAVK	207
S.parasanguinis_VT517	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSASNQLI	--165AA--	LPKTSAVK	207
S.parasanguinis_344SPAR	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSASNQLI	--172AA--	LPKTSAVK	207
S.parasanguinis_889SPAR	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSASNQLI	--172AA--	LPKTSAVK	207
S.parasanguinis_766SPAR	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPSASNQLI	--172AA--	LPKTSAVK	207
S.infantis_342SPSE	1	MKKVLLSSVAALAVFSAAAPVFA	QGENPNASNQLI	--168AA--	LPKTHAAK	210
S.parasanguinis_F0449	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPNASNQLI	--181AA--	LPKTSAVK	223
S.parasanguinis_451SPAR	1	MKKVLLSSVVALAVFAAAAPVFA	QGENPSSSNPLI	--172AA--	LPKTSAVK	207
S.australis_ATCC700641	1	MKKVLLSSVAALAVFAAAAPVFA	QGENPNASNQLI	--170AA--	LPKTSAVK	205
S.australis_I-G2	1	MKKVLLSSVAALSLFAAAAPVFA	QGENPNASNQLI	--173AA--	LPKTSAVK	208

*****:.....*:*****:*** ***** ** *

Fig. 2 CLUSTAL Omega alignment of functional regions from AbpA-like protein sequences. Similarity in signal sequence (blue), sortase-binding motif (violet) and termination codon (orange). Subgroup (1 to 5) differentiation is based on the N-terminal sequence (red)

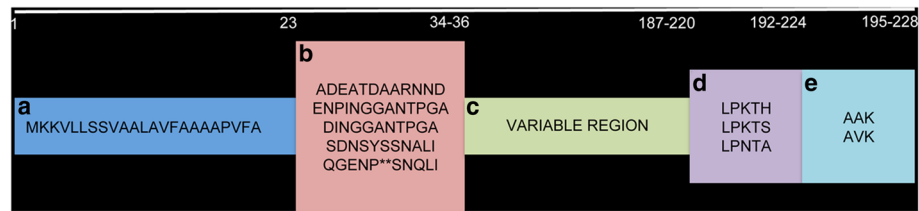


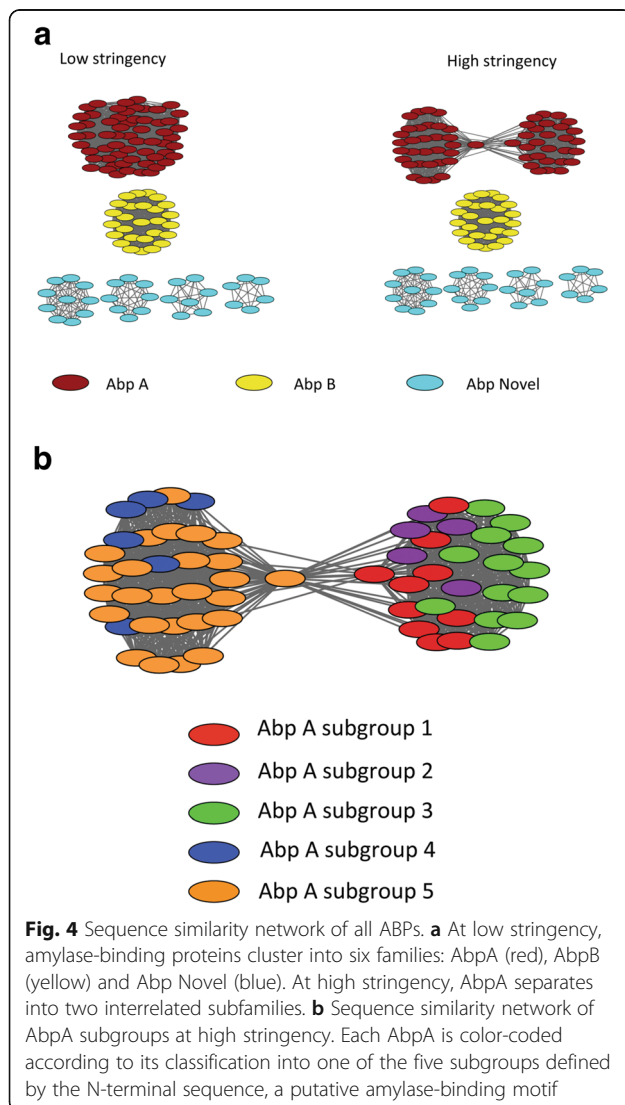
Fig. 3 Schematic of AbpA-like proteins. **a** Signal sequence; **b** N-terminal regions for subgroups 1–5; **c** variable region; **d** sortase B-binding motif; and **e** termination codon

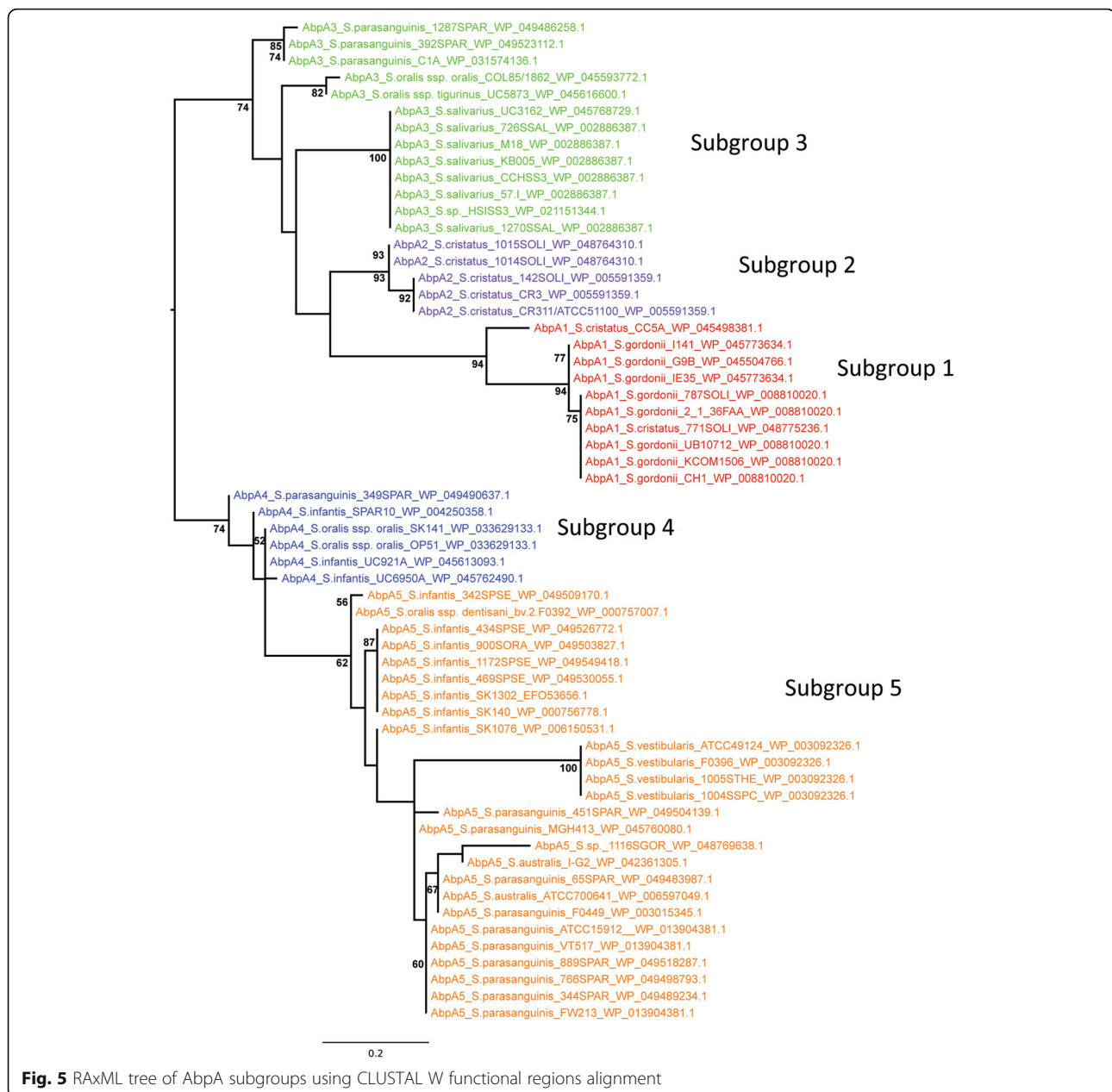
among these novel ABP groups suggests that individual clusters represent different gene families. In the high stringency SSN, all AbpAs formed a dumbbell-shaped cluster with a high degree of interconnectivity between the two bells. As shown in Fig. 4b, each AbpA was color-coded according to its classification into one of the five subgroups defined by the N-terminal

sequence, a putative amylase-binding motif. The clustering of AbpA subgroups follows the RAxML phylogeny (Fig. 5).

Phylogenetic analysis of AbpA-like proteins

As shown in Figs. 2 and Additional file 7: Figure S3, the signal sequence, sortase binding motif, and C-terminal regions of the AbpA-like proteins aligned with high reliability, while the N-terminal amylase-binding region was less reliably aligned due to a high degree of sequence divergence. A RAxML analysis based on these functional regions produced a phylogenetic tree that grouped the AbpA subgroups into two major clades receiving moderate bootstrap support (74%), with subgroups 1, 2, and 3 grouping into one clade, and subgroup 4 and 5 grouped into another clade (Fig. 5). However, the interrelationships among subgroups within these two clades are poorly supported, and while subgroups 1, 2, and 5 each receive moderate to strong support, members of subgroup 3 and 4 are non-monophyletic. Subgroup 3 members are split into three well-supported clades following taxonomy (*S. parasanguinis*, *S. oralis*, and *S. salivarius*, respectively). Different PRANK alignments corroborate grouping the AbpA subgroups into the same two major clades (only trees based on complete protein sequences and threshold 7 are shown here; Additional files 8 and 9: Figures S4 and S5). However, a number of incongruent relationships among subgroup members within these two major clades are found. For example, in all trees based on the PRANK alignments, subgroup 3 members now form two clades, and subgroup 4 members are nested within subgroup 5 members, rendering subgroup 5 non-monophyletic. In the analysis based on complete protein sequences, subgroups 1 and 2 are strongly supported as sister groups (Additional file 8: Figure S4), which is also supported by their very similar N-terminal sequences (Figs. 2 and 5). However, in the analyses of trimmed alignments based on PRANK reliability score thresholds, subgroup 2 is either nested within subgroup 3 (Additional file 9: Figure S5) or is sister to subgroup 3 (not shown). Since these phylogenetic trees are unrooted, i.e. reconstructed without outgroup for comparison,





conclusions regarding common ancestry are limited. Yet the high degree of overall sequence divergence and relative lack of interrelationships among most subgroups suggest that AbpA-like proteins have evolved independently.

Evidence of horizontal gene transfer

Complete and nearly complete genomes containing the *abpA-srtB* gene locus were analyzed for evidence of horizontal gene transfer. The G + C content of *abpA* ranged from 38.91 to 42% from genomes with a G + C content ranging from 39.9 to 42.1%. This minimal deviation in G + C content is consistent with ancient genes in closely related species such as the immunoglobulin

A1 protease (*iga*) gene from *Streptococcus* species [28]. Synteny maps were constructed (Fig. 6) to examine the immediate gene neighborhood around the *abpA-srtB* gene locus among genomes representative of different AbpA subgroups. There is lack of synteny conservation in genes flanking *abpA-srtB* in *Streptococcus* species experimentally found to bind amylase, in species with in silico evidence of the gene locus, and around the gene encoding the AbpA ortholog in *G. haemolysans* M341. Comparison of *S. cristatus* ATCC 51100 to *S. cristatus* AS 1.3089 demonstrates the presence and absence of *abpA-srtB*, respectively, in closely related strains while preserving the immediate synteny neighborhood. Overall, among ABS species and strains in

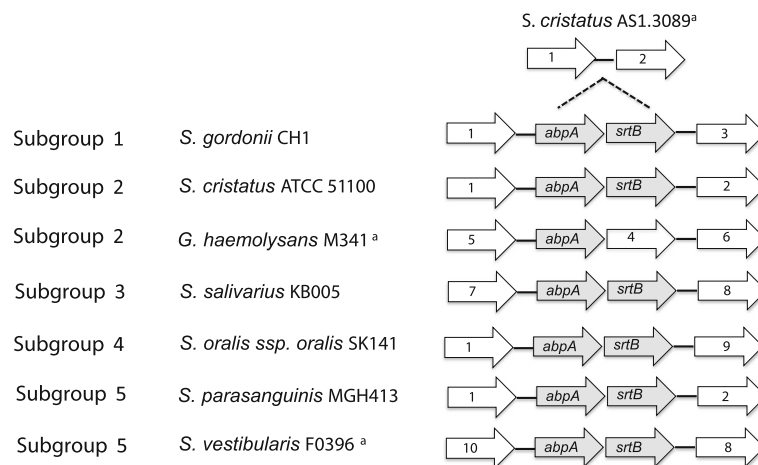


Fig. 6 Synteny mapping of *abpA-srtB*. *S. cristatus* AS1.3089 did not carry the *abpA-srtB* locus. ORFs flanking the *abpA-srtB* locus were deduced from genome sequences available at NCBI. All strains except *S. cristatus* AS1.3089, *S. vestibularis* F0396, and *G. haemolyans* M341 have been tested in vitro for the ability to bind amylase. Flanking genes include: (1) ribose-phosphate pyrophosphokinase; (2) aminotransferase; (3) haloacid dehalogenase; (4) MFS transporter; (5) peptidase M42; (6) peptidase; (7) DNA-binding transcriptional regulator, XRE-family; (8) metallophosphatase; (9) CoA-binding protein; and (10) IS200/IS605 family transposase. ^a*in silico* evidence only

this study, the genes flanking *abpA-srtB* show conservation within species, but not between species. In addition, transposon-mediated mutagenesis with Tn916 used to inactivate *abpA* demonstrated a transposon integration hotspot immediately upstream of *abpA* [14]. Together, these observations suggest the co-transcribed *abpA-srtB* locus is a genetic island that may have been acquired by horizontal transfer.

Molecular modeling of AbpA subgroups

In silico modeling with I-TASSER revealed similar tertiary structures for representative AbpA proteins from each subgroup. The molecular modeling performed here is limited by the availability of NMR data on AbpA structure, which until recently was not reported [29]. Overall, helical arrangement predominated the N-terminal half of the molecule and C-terminal half showed a coiled structure (Fig. 7).

Discussion

In this study, we expand upon previous studies to better understand how oral streptococci use different proteins to bind amylase. This interaction allows the bacteria to adapt to the oral environment by adhering to the salivary pellicle and initiating colonization of the tooth surface, as well as to efficiently metabolize dietary starch as an energy source. An in vitro assay to assess amylase binding to the cell surface of oral *Streptococcus* species has also been used previously in an attempt to classify viridans streptococci [6, 11, 30]. Here, this method was used to confirm the results of amylase binding to denatured proteins in the culture supernatant. Most *Streptococcus* strains examined produced an AbpA-like protein, the best-studied and most predominant

protein that mediates the binding of amylase to the bacterial surface. But bacterial cells from four strains, *S. mitis* OT25, SK137, SK145, and NS51 bound amylase without AbpA. These strains produced novel proteins, in particular the AbpC of *S. mitis* NS51. The actual function, cellular location, exposure to the extracellular environment, and relevance of these novel ABPs to the fitness of these strains, remains to be determined.

At least six separate families of ABPs comprising AbpA-like, AbpB-like, and novel proteins were identified in *Streptococcus* strains sequenced in this study in combination with homologs obtained from NCBI RefSeq <https://www.ncbi.nlm.nih.gov/refseq/>. There was no significant similarity between the families based on

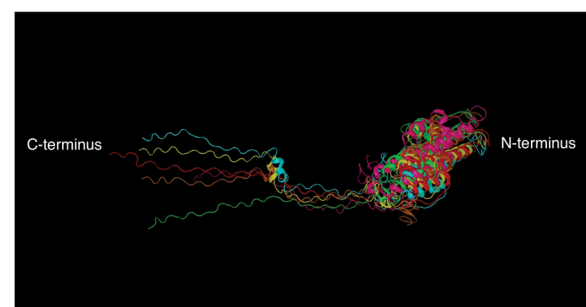


Fig. 7 Tertiary protein structure of representative AbpA sequences. N-terminal half of AbpA is predominated by helical spatial arrangement and the C-terminal half is largely coiled. NMR structure of *S. gordonii* CH1 is in red and the predicted structures for representative sequences are coded in various colors (Green-Subgroup1-WP_008810020.1; Yellow-Subgroup2-WP_005591359.1; Purple-Subgroup3-WP_002886387.1; Blue-Subgroup4-WP_033629133.1; Orange-Subgroup5-WP_003092326.1)

protein sequence alignment. No readily identifiable amylase-binding or sortase-binding motifs were present except within the AbpA-like family. Thus, binding of amylase by each of these protein families is likely due to secondary structural interaction rather than by primary amino acid sequences. This was especially evident for AbpA-like proteins having moderate amino acid similarity and identity, but with similar predicted secondary structures as presented in this study and by Liang et al. [19].

General functions of ABPs may be in promoting oral colonization and fitness. When ABPs bind salivary amylase, they enable close contact of bacteria with oral surfaces coated with saliva and promote subsequent colonization. When first described, AbpA was a novel protein [14]. Homologs are now found in many oral streptococcal species of humans and animals whose saliva contains amylase activity [32, 33]. In addition to playing a role in adhesion, biofilm formation, providing a convenient source of saccharolytic nutrition through the hydrolysis of dietary starch, and possibly providing relief from oxidative stress, there may be as yet undescribed functions for AbpA [12, 17]. AbpB protein sequences are conserved (63–100% identical), suggesting they have a primary function as dipeptidases, which may have a role in nutrient acquisition pathways [20, 34]. The novel ABPs have been annotated with a variety of cell wall-associated functions. Association of AbpA with the bacterial cell wall, especially at the site of nascent cell wall synthesis, was previously demonstrated [32], so the finding that other unrelated ABPs may also associate with the cell wall is supportive of the hypothesis that amylase binding is beneficial to the growth of the organism [12, 34]. Alternatively, the ability to bind amylase may just be a coincidental function for AbpB and the novel ABPs. How the acquisition of the *abpA-srtB* locus or any of the other genes encoding potentially novel ABPs enhances the survival and role of the commensal streptococcal strains as primary colonizers of the oral cavity requires further study.

The ability of a bacterial species to bind salivary amylase allowing it to adapt to the oral environment is not confined to cell wall components of *Streptococcus* species. Fimbria-associated proteins from *S. sanguinis* SK36 [33] and *S. mutans* CS2 [35], a *Streptococcus* species not normally known to bind amylase have been shown to bind amylase. Amylase-binding components have been described in a few other non-streptococcal oral species including a 110-kDa protein of *Fusobacterium nucleatum* subsp. *polymorphum* [36], and the outer membrane lipopolysaccharide of *Aggregatibacter actinomycetemcomitans* [37] and *Porphyromonas gingivalis* [38]. In *A. actinomycetemcomitans*, amylase interferes with bacterial adherence and biofilm formation. Amylases from

several sources are able to inhibit the growth of *P. gingivalis* [39]. How the binding of amylase in vivo in a mixed species biofilm enables or hinders interspecies interactions is not known.

Comparative pan-genomic studies provide evidence of genetic exchange with related species sharing the same ecological niche [40, 41]. Our analyses of complete and nearly complete draft genomes of AbpA-encoding oral *Streptococcus* strains revealed that all contained the *abpA-srtB* gene locus. Sortase B is essential for initial covalent attachment of AbpA to the bacterial cell wall [18, 19]. In fact this gene locus was found in 15/18 strains sequenced in this study, supporting the role of AbpA as the predominate ABP. The *abpA* gene may be an ancient gene, not unlike another gene in the mitis group of streptococci, *iga*, which encodes IgA protease [28] and is also found in several closely related *Streptococcus* species. AbpA is located in a region of conserved gene order indicating relatively recent subgroup divergence. AbpA subgroups generally correlate with species, although there is evidence of interspecies divergence. For example, within the mitis group of *Streptococcus* species, the *S. oralis* clade is more distantly related to the *S. parasanguinis* clade [23], and both contain strains that carry AbpA-like proteins from a variety of subgroups. This apparent random distribution of the co-transcribed *abpA-srtB* genes in more distantly related *Streptococcus* species, the lack of synteny conservation in closely related species, and the presence of an AbpA ortholog in a different genus (*Gemella*) suggests that these genes are part of a genomic island acquired by horizontal gene transfer.

Many oral streptococci are known to be naturally competent [42], a benefit to organisms living in a biofilm such as dental plaque. While screening our collection of *Streptococcus* strains, it became apparent that not all strains of a given species possess an ABP. The transfer of DNA encoding a beneficial trait from one member of the community to another enhances the survival of the community. It is therefore possible that the acquisition of the amylase-binding function by some oral streptococcal strains, mediated by different proteins enables these bacteria to take advantage of their ecological niche within the oral cavity.

This study demonstrates that ABPs are a diverse group of proteins that sort into at least six distinct groups with a variety of putative functions. Sequence divergence of the ABPs, especially within the AbpA group, suggests that the amylase-binding function evolved independently in separate protein families. The ability to bind amylase may benefit the host by serving to remove bacteria in saliva. The high prevalence of amylase binding among primary colonizing streptococci in dental plaque may demonstrate how adaptive evolution can enhance bacterial colonization and

survival in the oral cavity. Amylase binding may not only facilitate the nutrition of bacteria in the immediate environment, but it may also serve another unrecognized function to benefit the whole microbial community.

Conclusions

In this study, we explored the genomic and evolutionary aspects of ABPs. We demonstrated that ABPs (AbpA, AbpB and novel ABPs) cluster into six distinct but unrelated families. Further, the AbpA protein family could be subgrouped based on N-terminal sequences. While evidence for evolution of these protein families was not conclusive, comparative genomics of *abpA* gene provided evidence of horizontal gene transfer. Importantly, the acquisition of ABP by oral streptococci provided an interesting example of how bacteria may evolve to adapt to the human host.

Methods

Bacterial strains and culture conditions

A total of 79 oral streptococcal strains representing 13 species were screened for ABPs, which included 20 strains studied previously and 59 strains examined here (Table 1). All strains were obtained from our culture collection, with the exception of *S. australis* ATCC 700641 and *S. infantis* ATCC 700779, which were obtained from the American Type Culture Collection (ATCC, Manassas, VA). Bacteria were cultured from frozen stocks to agar plates containing tryptic soy broth supplemented with 0.5% (*w/v*) yeast extract (TSBY) and 1.5% Bacto agar (Becton Dickinson and Co., Sparks, MD) in a candle jar at 37 °C for 2 days. Isolated colonies were subcultured into TSBY broth and incubated in a candle jar at 37 °C for 14 to 16 h.

Human saliva collection

The University at Buffalo Human Subjects Institutional Review Board approved the saliva collection protocol. Briefly, unstimulated whole saliva, as the source of human salivary amylase, was collected from several donors by expectoration into ice-chilled tubes, clarified by centrifugation at 12,800 \times *g* for 10 min at 4 °C, aliquoted, and stored at -20 °C until use.

Detection of amylase-binding proteins (ABPs) by the amylase ligand-binding assay

Streptococcal culture supernatants from overnight cultures were collected by centrifugation at 5000 \times *g* for 10 min followed by 20-fold concentration using Amicon Ultra-4 centrifugal filter devices (Merck Millipore Ltd., Tullagreen, Carrigtwohill, Co. Cork, IRL). The concentrated supernatant proteins were separated on a 12.5% gel by SDS-PAGE. Proteins were either stained with Coomassie blue or transblotted to PVDF membrane.

The amylase-ligand binding assay was performed as previously described [14, 20]. Whole human saliva was used as the source of amylase.

N-terminal amino acid sequencing

Concentrated supernatants containing ABPs identified by the aforementioned amylase ligand-binding assay were subjected to 12.5% SDS-PAGE and separated proteins were electrotransferred from gels to a small pore PVDF ProBlott membrane (Applied Biosystems Inc., Foster City, CA) and stained with 0.1% (*w/v*) Coomassie brilliant blue R-250. The protein bands corresponding to the identified ABPs were cut from the blot and N-terminal sequenced by standard methods (ProSeq Inc., Oxford, MA). The N-terminal amino acid sequences of ABPs were used to search for homologous proteins in the National Center for Biotechnology Information database (NCBI, <http://www.ncbi.nlm.nih.gov>).

Purification and sequencing of streptococcal DNA

Genomic DNA isolated from streptococcal strains using a previously described method [20] was used as the template in PCR reactions. For high-throughput sequencing, the genomic DNA was first treated with RNase A/T1 Mix (Thermo Fisher Scientific Inc., Pittsburgh, PA) according to the manufacturers' instructions. The QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) was then used to further purify the genomic DNA. The quality was assessed by ethidium bromide stained agarose gels. DNA was quantified at A_{260} and A_{280} using the Nanodrop spectrophotometer and the Quant-iT dsDNA kit (Invitrogen, Carlsbad, CA). Samples were stored at -80 °C prior to library construction for whole genome sequencing (Center for Excellence in Bioinformatics at the University at Buffalo, Buffalo, NY).

Sequencing, quality control, de novo assembly, and annotation of streptococcal genomes

Illumina HiSeq 2500 Next-Generation Sequencing was used in rapid 150-cycle paired-end mode to perform genome sequencing of the appropriately prepared streptococcal strains. The paired-end sequencing reads of 150-bp read length and over 100X coverage were processed using MyPro (<http://sourceforge.net/projects/sb2nhri/files/MyPro>) [21], a customized software pipeline designed for prokaryotic genome assembly and annotation. Quality control tests, de novo assembly, and annotation were applied to genomes of 18 streptococcal strains. When closely related reference genomes were available, post-assembly function of MyPro was used to align and order contigs and reduce gaps in the de novo genome assemblies [22]. All draft genome sequences were deposited in GenBank and corresponding accession numbers were published prior to data mining [22].

Identification of amylase-binding protein A sequences

Complete *abpA* were manually obtained from 18 streptococcal genomes by using corresponding N-terminal sequence as a query. Thereafter, complete *abpA* sequences were used as multifasta query to BLAST against the non-redundant protein sequence (nr) database restricted to Streptococcus (taxid:1301) using an evalue of 1e-5. Results were curated for presence of N-terminal - like sequences as a measure of confirmation.

Identification of amylase-binding protein B sequences

To obtain genomic sequences previously annotated as AbpB, the NCBI RefSeq database was searched for complete streptococcal AbpB sequences by using the keywords 'amylase-binding protein' or 'amylase-binding protein B', and filters 'bacteria' and 'sequence length more than 300'. The results were un-collapsed and manually sorted to include all non-redundant streptococcal strains. To identify any additional AbpB-like sequences that may be annotated differently (e. g. peptidase C69), the database was also searched with sequences obtained from primer walking of AbpB.

Primer walking of *abpB*

To determine the *abpB* sequence from each of the strains expressing AbpB in the amylase-ligand binding assay, five sets of degenerate primers were designed according to the alignment of nucleotide sequences of known AbpBs. Primers were designed to conserved regions where possible. The primer sequences (Invitrogen) listed in Additional file 10: Table S5 cover approximately 84% of *abpB*. Conventional PCR was initiated with 90 s at 95 °C, followed by 34 cycles of 30 s at 95 °C, 30 s at 50–52 °C according to the annealing temperatures, and 2 min at 72 °C, then 10 min at 72 °C using a T100 thermo cycler (Bio-Rad, CA). Amplified products were separated in a 1.5% agarose gel, stained with ethidium bromide, and visualized under UV light. Genomic DNA of *S. gordonii* Challis CH1 and a no template control were used as the positive and negative controls, respectively. PCR amplicons purified by QIAquick PCR Purification kit or QIAquick Gel Extraction kit (Qiagen, Hilden, Germany) were sent for nucleic acid sequencing (Roswell Park Cancer Institute, Buffalo, NY). Sequences obtained from each strain were assembled using SeqMan software (www.dnastar.com) into one sequence containing the nearly complete *abpB* gene.

Identification of novel ABP sequences

Proteins that bound amylase in the amylase ligand-binding assay with molecular weights ranging from 30 to 65 kDa and whose N-terminal sequences were not similar to one of the five AbpA, or AbpB N-terminal sequences were considered novel ABPs. The NCBI RefSeq

database was searched with the N-terminal sequences of these strains to identify novel ABP sequences.

Sequence similarity network

A sequence similarity network (SSN) of all the ABP protein sequences was generated using EFI-EST [43]. The relatedness is described by sequence similarity based on pairwise alignment scores calculated from an all-by-all BLAST with a default E-value 1e-05. Networks were generated based on the minimum alignment score 6 (low stringency) and the maximum alignment score 16 (high stringency), respectively. Each node, which represents one protein sequence, was color-coded according to its classification in the three ABP groups: AbpA, AbpB, and Abp Novel.

Phylogenetic analysis of ABPs

Multiple sequence alignment was performed using CLUSTAL W to align the functional regions of AbpA, including the signal sequence, the N-terminal amylase-binding region, the sortase B-binding motif, and the C-terminus. Maximum Likelihood (ML) analysis was performed on the resulting alignment using the RAxML (Randomized Accelerated Maximum Likelihood) algorithm [31]. Since the AbpA-like protein sequences are highly divergent, we also applied PRANK (Probabilistic Alignment Kit) [44] to align the complete protein sequences of AbpAs, and the reliability score of each aligned site was then annotated. The scale of the reliability score is defined as 0 to 9, indicating less reliably aligned site to more reliably aligned site. The alignment was then trimmed based on different reliability score thresholds: 1, 4, 5, 6, 7 and 9. ML analysis using RAxML was performed for each of these trimmed alignments, as well as the complete protein sequences.

Bacterial amylase-binding assay

The starch agarose assay was performed as previously [6, 11] to access the ability of bacterial cells to bind amylase in clarified whole human saliva. Control samples included saliva alone, and saliva pre-incubated with streptococcal strains known to bind amylase or unable to bind amylase.

Signatures of horizontal gene transfer surrounding *abpA*

GenBank files for ABS were compared for synteny (one gene upstream and downstream) flanking the *abpA-srtB* locus. Representative genomes for each AbpA subgroup were selected based on experimental evidence of AbpA to illustrate the variation in synteny neighborhood of *abpA-srtB* locus. The closely related genome *S. cristatus* ATCC 51100 was selected based on the lack of the *abpA-srtB* locus, and the *G. haemolysans* M341 genome was selected to illustrate an *abpA* ortholog without the typical *srtB*.

Molecular modeling of AbpA

Multispecies sequences (representing 20 sequences) from all AbpA subgroups were submitted for molecular modeling using the I-TASSER (Iterative Threading ASSEMBLY Refinement) web service, which uses a hierarchical approach for protein structure and function prediction [45]. We included the AbpA sequence from *S. gordonii* Challis CH1 as a positive control. The NMR structure of AbpA from *S. gordonii* G9B has been determined [29], and the protein sequence of *S. gordonii* Challis CH1 differs from G9B only at residue 33 containing Asn and Lys, respectively. I-TASSER predicted secondary structures and tertiary structure models (.pdb format) were obtained for all sequences along with their confidence scores. Model with highest confidence score for each subgroup were used as input for Protean3D (<http://www.dnastar.com/t-protean-3D.aspx>) for structural comparison and output in graphical format.

Streptococcal taxonomy

Concatemers of seven housekeeping genes for viridans group Streptococci in multifasta format were obtained from the authors of a previous publication [46]. These sequences were curated to represent the amylase-binding streptococcus (ABS) species used in this study. Curated sequences were used as query to BLAST against a local database comprised of whole genome sequences of ABS with an Expected (E) value of $1e-5$ [47]. The results were concatenated in-frame, in the order (*map-pfl-ppaC-pyk-rpoB-sodA-tuf*) and used for phylogenetic analysis. ABS concatemers were aligned using CLUSTAL W of the MEGA 7.0.18 package and the aligned sequences were used to construct a minimum evolution tree.

Additional files

Additional file 1: Table S1. MLSA based revised streptococcal taxonomy. Includes the genome accession number, strain identifier, old taxonomic name, and new taxonomic name for each strain. (DOCX 91 kb)

Additional file 2: Table S2. Amylase-binding protein A-like subgroup comparison. Includes NCBI protein identifiers, molecular weights, and N-terminal sequences. (DOCX 107 kb)

Additional file 3: Table S3. Amylase-binding protein B-like comparison. Includes NCBI protein identifiers, molecular weights, and N-terminal sequences. (DOCX 89 kb)

Additional file 4: Figure S1. CLUSTAL alignment of AbpB-like protein sequences. (DOCX 167 kb)

Additional file 5: Table S4. Novel amylase-binding protein comparison. Includes NCBI protein identifiers, molecular weights, and N-terminal sequences. (DOCX 82 kb)

Additional file 6: Figure S2. CLUSTAL alignment of novel amylase-binding proteins. (DOCX 145 kb)

Additional file 7: Figure S3. PRANK reliability annotation of the alignment of functional regions (signal sequence and N-terminal sequence) from AbpA-like proteins. (TIFF 11836 kb)

Additional file 8: Figure S4. PRANK tree of AbpA subgroups using the entire gene sequence. (TIFF 14602 kb)

Additional file 9: Figure S5. PRANK tree of AbpA subgroups using the entire gene sequence, reliability score 7. (TIFF 14758 kb)

Additional file 10: Table S5. Degenerate PCR primers for genes encoding AbpB-like proteins. (DOCX 73 kb)

Abbreviations

AbpA: Amylase-binding protein A; AbpB: Amylase-binding protein B; AbpC: Amylase-binding protein C; ABPs: Amylase-Binding Proteins; ATCC: American Type Culture Collection; BLAST: Basic Local Alignment Search Tool; CLUSTAL W: CLUSTER ALIGNMENT W; G + C: Guanine and Cytosine; IgA: Immunoglobulin A1 protease; *iga*: Immunoglobulin A1 protease gene; I-TASSER: Iterative Threading ASSEMBLY Refinement; kDa: kilo Dalton; MLSA: MultiLocus Sequence Analysis; NCBI: National Center for Biotechnology Information; NMR: Nuclear Magnetic Resonance; PRANK: PRobabilistic ALignMent Kit; PVDF: PolyVinylidene Fluoride; RAXML: Randomized Accelerated Maximum Likelihood; SDS-PAGE: Sodium Dodecyl Sulfate- PolyAcrylamide Gel Electrophoresis; *srtB*: Sortase B gene; SrtB: Sortase B protein; SSN: Sequence Similarity Network; TSBY: Tryptic Soy Broth supplemented with 0.5% (w/v) Yeast extract

Acknowledgements

The National Institute of Dental and Craniofacial Research Grant R01DE022673 supported this work. We thank Dr. Victor Albert and Dr. Gerald Kuldelka in the Department of Biology, and Dr. M. Margaret Vickerman in the Department of Oral Biology at the University at Buffalo for their helpful discussions. We also thank Steven Hintermeier for technical laboratory assistance.

Funding

This work was supported by the National Institute of Dental and Craniofacial Research Grant R01DE022673.

Availability of data and materials

The datasets generated and/or analyzed during the current study are available at the NCBI in GenBank at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA274768>. Draft genomes are described in [22].

Authors' contributions

EMH contributed to the design of all experiments, analyzed Western blot data, performed amylase activity assays, collated strain data and results, participated in bioinformatic data analysis, co-drafted and co-wrote the manuscript. YK performed sample preparation for Western blot and N-terminal sequencing, amylase ligand binding assay and analysis, DNA isolation for genomic sequencing, primer design and primer walking of AbpB, and co-wrote a draft of the manuscript. AS collaborated with YL in genome analysis, performed CLUSTAL alignments, horizontal gene transfer analysis, and structural modeling of AbpA, and co-wrote part of the manuscript. TL performed all phylogenetic analyses. CL consulted on types of phylogenetic analyses, and together with TL co-wrote part of the manuscript. FAS conceived of the project and advised on all aspects of experimental design and analyses, and co-wrote this manuscript. All authors have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

The University at Buffalo Human Subjects Institutional Review Board approved the human saliva collection protocol. The protocol includes a statement that the saliva will be used for research purposes, and the consent was signed by all donors.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

¹Department of Oral Biology, School of Dental Medicine, University at Buffalo, State University of New York, Buffalo, NY, USA. ²Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan. ³Department of Biological Sciences, University at Buffalo, State University of New York, Buffalo, NY, USA. ⁴Department of Oral Biology, School of Stomatology, China Medical University, Shenyang, People's Republic of China.

Received: 23 December 2016 Accepted: 8 April 2017

Published online: 20 April 2017

References

- Bentley RW, Leigh JA, Collins MD. Intrageneric structure of *Streptococcus* based on comparative analysis of small-subunit rRNA sequences. *Int J Syst Bacteriol.* 1991;41(4):487–94. doi:10.1099/00207713-41-4-487.
- Kawamura Y, Hou XG, Sultana F, Miura H, Ezaki T. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. *Int J Syst Bacteriol.* 1995;45(2):406–8. doi:10.1099/00207713-45-2-406.
- Brown AE, Rogers JD, Haase EM, Zelasko PM, Scannapieco FA. Prevalence of the amylase-binding protein A gene (*abpA*) in oral streptococci. *J Clin Microbiol.* 1999;37(12):4081–5.
- Douglas CW, Heath J, Hampton KK, Preston FE. Identity of viridans streptococci isolated from cases of infective endocarditis. *J Med Microbiol.* 1993;39(3):179–82. doi:10.1099/00222615-39-3-179.
- Gwynn JP, Douglas CW. Comparison of amylase-binding proteins in oral streptococci. *FEMS Microbiol Lett.* 1994;124(3):373–9.
- Kilian M, Nyvad B. Ability to bind salivary alpha-amylase discriminates certain viridans group streptococcal species. *J Clin Microbiol.* 1990;28(11):2576–7.
- Karn RC, Shulkin JD, Merritt AD, Newell RC. Evidence for post-transcriptional modification of human salivary amylase (amyl) isozymes. *Biochem Genet.* 1973;10(4):341–50.
- Boehlke C, Zierau O, Hannig C. Salivary amylase - The enzyme of unspecialized euryphagous animals. *Arch Oral Biol.* 2015;60(8):1162–76. doi:10.1016/j.archoralbio.2015.05.008.
- Aguirre A, Levine MJ, Cohen RE, Tabak LA. Immunochemical quantitation of alpha-amylase and secretory IgA in parotid saliva from people of various ages. *Arch Oral Biol.* 1987;32(4):297–301.
- Scannapieco FA, Torres G, Levine MJ. Salivary alpha-amylase: role in dental plaque and caries formation. *Crit Rev Oral Biol Med.* 1993;4(3–4):301–7.
- Douglas CW. The binding of human salivary alpha-amylase by oral strains of streptococcal bacteria. *Arch Oral Biol.* 1983;28(7):567–73.
- Rogers JD, Palmer Jr RJ, Kolenbrander PE, Scannapieco FA. Role of *Streptococcus gordonii* amylase-binding protein A in adhesion to hydroxyapatite, starch metabolism, and biofilm formation. *Infect Immun.* 2001;69(11):7046–56.
- Nikitkova AE, Haase EM, Vickerman MM, Gill SR, Scannapieco FA. Response of fatty acid synthesis genes to the binding of human salivary amylase by *Streptococcus gordonii*. *Appl Environ Microbiol.* 2012;78(6):1865–75. doi:10.1128/AEM.07071-11.
- Rogers JD, Haase EM, Brown AE, Douglas CW, Gwynn JP, Scannapieco FA. Identification and analysis of a gene (*abpA*) encoding a major amylase-binding protein in *Streptococcus gordonii*. *Microbiology.* 1998;144(Pt 5):1223–33.
- Li L, Tanzer JM, Scannapieco FA. Identification and analysis of the amylase-binding protein B (*AbpB*) and gene (*abpB*) from *Streptococcus gordonii*. *FEMS Microbiol Lett.* 2002;212(2):151–7.
- Vorrasi J, Chaudhuri B, Haase EM, Scannapieco FA. Identification and characterization of amylase-binding protein C from *Streptococcus mitis* NS51. *Mol Oral Microbiol.* 2010;25(2):150–6. doi:10.1111/j.2041-1014.2009.00554.x.
- Haase EM, Feng X, Pan J, Miecznikowski JC, Scannapieco FA. Dynamics of the *Streptococcus gordonii* transcriptome in response to media composition, salivary alpha-amylase and starch. *Appl Environ Microbiol.* 2015. doi:10.1128/AEM.01221-15.
- Liang X, Chen Y, Wu H. Sortase B assemble amylase-binding protein A to streptococcal cell surface. 89th Gen. Sess. of Internat. Assoc. Dent. Res. San Diego: IADR; 2011. p. 72.
- Liang X, Liu B, Zhu F, Scannapieco FA, Haase EM, Matthews S, et al. A distinct sortase SrtB anchors and processes a streptococcal adhesin AbpA with a novel structural property. *Sci Rep.* 2016;6:30966. doi:10.1038/srep30966.
- Chaudhuri B, Paju S, Haase EM, Vickerman MM, Tanzer JM, Scannapieco FA. Amylase-binding protein B of *Streptococcus gordonii* is an extracellular dipeptidyl-peptidase. *Infect Immun.* 2008;76(10):4530–7. doi:10.1128/IAI.00186-08.
- Ruoff KL. *Streptococcus anginosus* ("Streptococcus milleri"): the unrecognized pathogen. *Clin Microbiol Rev.* 1988;1(1):102–8.
- Sabharwal A, Liao YC, Lin HH, Haase EM, Scannapieco FA. Draft genome sequences of 18 oral streptococcus strains that encode amylase-binding proteins. *Genome Announc.* 2015;3(3):e00510–5. doi:10.1128/genomeA.00510-15.
- Takenouchi-Ohkubo N, Mortensen LM, Drasbek KR, Kilian M, Poulsen K. Horizontal transfer of the immunoglobulin A1 protease gene (*iga*) from *Streptococcus* to *Gemella haemolysans*. *Microbiology.* 2006;152(Pt 7):2171–80. doi:10.1099/mic.0.28801-0.
- Kilian M, Poulsen K, Blomqvist T, Havarstein LS, Bek-Thomsen M, Tettelin H, et al. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. *PLoS One.* 2008;3(7):e2683. doi:10.1371/journal.pone.0002683.
- Jensen A, Scholz CF, Kilian M. Re-evaluation of the taxonomy of the Mitis group of the genus *Streptococcus* based on whole genome phylogenetic analyses, and proposed reclassification of *Streptococcus dentisani* as *Streptococcus oralis* subsp. *dentisani* comb. nov., *Streptococcus tigurinus* as *Streptococcus oralis* subsp. *tigurinus* comb. nov., and *Streptococcus oligofermentans* as a later synonym of *Streptococcus cristatus*. *Int J Syst Evol Microbiol.* 2016;66(11):4803–20. doi:10.1099/ijsem.0.001433.
- Liao YC, Lin HH, Sabharwal A, Haase EM, Scannapieco FA. MyPro: A seamless pipeline for automated prokaryotic genome assembly and annotation. *J Microbiol Methods.* 2015;113:72–4. doi:10.1016/j.jmimet.2015.04.006.
- Facklam R, Elliott JA. Identification, classification, and clinical relevance of catalase-negative, gram-positive cocci, excluding the streptococci and enterococci. *Clin Microbiol Rev.* 1995;8(4):479–95.
- Collins M. The Genus *Gemella*. In: Dworkin M, Falkow S, Rosenberg E, Schleifer H, Stackebrandt E, editors. *The Prokaryotes*. New York: Springer; 2006. p. 511–8.
- Gopal P, Ragnanath C, Vyas V, Shanmugam M, Ramasubbu N. Probing the interaction of human salivary alpha-amylase and amylase binding protein A (*AbpA*) of *Streptococcus gordonii*. *Mol Biol.* 2013;2:111. doi:10.4172/2168-9547.100011.
- Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using CLUSTAL Omega. *Mol Syst Biol.* 2011;7:539. doi:10.1038/msb.2011.75.
- Sethi A, Mohanty B, Ramasubbu N, Gooley PR. Structure of amylase-binding protein A of *Streptococcus gordonii*: a potential receptor for human salivary alpha-amylase enzyme. *Protein Sci.* 2015;24(6):1013–8. doi:10.1002/pro.2671.
- Douglas CW, Pease AA, Whiley RA. Amylase-binding as a discriminator among oral streptococci. *FEMS Microbiol Lett.* 1990;54(1–3):193–7.
- Nikitkova AE, Haase EM, Scannapieco FA. Taking the starch out of oral biofilm formation: molecular basis and functional significance of salivary alpha-amylase binding to oral streptococci. *Appl Environ Microbiol.* 2013;79(2):416–23. doi:10.1128/AEM.02581-12.
- Scannapieco FA, Haraszthy GG, Cho MI, Levine MJ. Characterization of an amylase-binding component of *Streptococcus gordonii* G9B. *Infect Immun.* 1992;60(11):4726–33.
- Okahashi N, Nakata M, Terao Y, Isoda R, Sakurai A, Sumitomo T, et al. Pili of oral *Streptococcus sanguinis* bind to salivary amylase and promote the biofilm formation. *Microb Pathog.* 2011;50(3–4):148–54. doi:10.1016/j.micpath.2011.01.005.
- Perrone M, Gfell LE, Fontana M, Gregory RL. Antigenic characterization of fimbria preparations from *Streptococcus mutans* isolates from caries-free and caries-susceptible subjects. *Clin Diagn Lab Immunol.* 1997;4(3):291–6.
- Zulfiqar M, Yamaguchi T, Sato S, Oho T. Oral *Fusobacterium nucleatum* subsp. *polymorphum* binds to human salivary alpha-amylase. *Mol Oral Microbiol.* 2013;28(6):425–34. doi:10.1111/omi.12036.
- Baik JE, Hong SW, Choi S, Jeon JH, Park OJ, Cho K, et al. Alpha-amylase is a human salivary protein with affinity to lipopolysaccharide of *Aggregatibacter actinomycetemcomitans*. *Mol Oral Microbiol.* 2013;28(2):142–53. doi:10.1111/omi.12011.
- Choi S, Baik JE, Jeon JH, Cho K, Seo DG, Kum KY, et al. Identification of *Porphyromonas gingivalis* lipopolysaccharide-binding proteins in human saliva. *Mol Immunol.* 2011;48(15–16):2207–13. doi:10.1016/j.molimm.2011.06.434.
- Ochiai A, Harada K, Hashimoto K, Shibata K, Ishiyama Y, Mitsui T, et al. alpha-Amylase is a potential growth inhibitor of *Porphyromonas gingivalis*, a

- periodontal pathogenic bacterium. *J Periodontol Res.* 2014;49(1):62–8. doi:10.1111/jre.12079.
41. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, Angiuoli SV, et al. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. *Genome Biol.* 2010;11(10):R107. doi:10.1186/gb-2010-11-10-r107.
 42. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, et al. Phylogenomics and the dynamic genome evolution of the genus *Streptococcus*. *Genome Biol Evol.* 2014;6(4):741–53. doi:10.1093/gbe/evu048.
 43. Cvitkovitch DG. Genetic competence and transformation in oral streptococci. *Crit Rev Oral Biol Med.* 2001;12(3):217–43.
 44. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using sequence similarity networks for visualization of relationships across diverse protein superfamilies. *PLoS One.* 2009;4(2):e4345. doi:10.1371/journal.pone.0004345.
 45. Stamatakis A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312–3. doi:10.1093/bioinformatics/btu033.
 46. Loytynoja A, Goldman N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics.* 2010;11:579. doi:10.1186/1471-2105-11-579.
 47. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics.* 2008;9:40. doi:10.1186/1471-2105-9-40.
 48. Bishop CJ, Aanensen DM, Jordan GE, Kilian M, Hanage WP, Spratt BG. Assigning strains to bacterial species via the internet. *BMC Biol.* 2009;7:3. doi:10.1186/1741-7007-7-3.
 49. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;215(3):403–10. doi:10.1016/S0022-2836(05)80360-2.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

