1   The *blp* locus of *Streptococcus pneumoniae* plays a limited role in the selection of which strains

2   can co-colonize the human nasopharynx

3

4   Carina Valente[1], Suzanne Dawid[2,3], Francisco R. Pinto[4], Jason Hinds[5,6], Alexandra S. Simões[1],

5   Katherine A. Gould[5,6], Luís A. Mendes[1], Hermínia de Lencastre[7,8], and Raquel Sá-Leão[1#]

6

7   [1]Laboratory of Molecular Microbiology of Human Pathogens, Instituto de Tecnologia Química e

8   Biológica António Xavier, Universidade Nova de Lisboa, Oeiras, Portugal; [2]Department of

9   Pediatrics and Communicable Diseases, University of Michigan, Ann Arbor, MI, USA;

10   [3]Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI, USA;

11   [4]BioISI – Biosystems & Integrative Sciences Institute, Faculdade de Ciências da Universidade de

12   Lisboa, Lisbon, Portugal; [5]Institute for Infection and Immunity, St. George's, University of

13   London, London, UK; [6]BUGS Bioscience, London Bioscience Innovation Centre, London, UK;

14   [7]Laboratory of Molecular Genetics, Instituto de Tecnologia Química e Biológica António Xavier,

15   Universidade Nova de Lisboa, Oeiras, Portugal; [8]Laboratory of Microbiology and Infectious

16   Diseases, The Rockefeller University, New York, New York, USA.

17

18   **Running title:** Impact of the *blp* locus on pneumococcal co-colonization

19

20   **# Corresponding author**

21   Raquel Sá-Leão

22   Laboratory of Molecular Microbiology of Human Pathogens,

23   Instituto de Tecnologia Quimica e Biologica Antonio Xavier,

24   Av. Republica, 2780-157 Oeiras, Portugal

25   Tel: +351214469872

26   E-mail: rsaleao@itqb.unl.pt

27

28   *ASS present affiliation: Global Health and Tropical Medicine, Instituto de Higiene e Medicina

29   Tropical, Universidade Nova de Lisboa, Lisbon, Portugal.

30   *LAM present affiliation: Department of Biology & CESAM, Universidade de Aveiro, Aveiro,

31   Portugal.

32

33   **KEY WORDS**

34   *Streptococcus pneumoniae*; multiple serotype carriage; co-colonization; competition;

35   bacteriocins; *blp* locus

36

37   **LIST OF ABBREVIATIONS**

38   CSP (competence stimulating peptide); BIR (bacteriocin/immuntity region); *ply*NCR-RFLP

39   (pneumolysin non-coding region restriction fragment length polymorphism); MLST (multilocus

40   sequence typing); CC (clonal complex); S (singleton).

41

42 **ABSTRACT**

43 Nasopharyngeal colonization is important for *Streptococcus pneumoniae* evolution, providing

44 the opportunity for horizontal gene transfer when multiple strains co-occur. Although

45 colonization with more than one strain of pneumococcus is common, the factors that influence

46 the ability of strains to co-exist are not known. A highly variable *blp* (bacteriocin-like peptide)

47 locus has been identified in all sequenced strains of *S. pneumoniae*. This locus controls the

48 regulation and secretion of bacteriocins, small peptides that target other bacteria. In this study,

49 we analyzed a series of co-colonizing isolates to evaluate the impact of the *blp* locus on human

50 colonization to determine whether competitive phenotypes of bacteriocin secretion restrict co-

51 colonization.

52

53 We identified a collection of 135 nasopharyngeal samples with two or more strains totaling 285

54 isolates. The *blp* locus of all strains was characterized genetically with regards to pheromone

55 type, bacteriocin/immunity content and potential for locus functionality. Inhibitory phenotypes

56 of bacteriocin secretion and locus activity were assessed through overlay assays. Isolates from

57 single colonization (n=298) were characterized for comparison.

58

59 Co-colonizing strains had a high diversity of *blp* cassettes; approximately one third displayed an

60 inhibitory phenotype *in vitro*. Despite in vitro evidence of competition, pneumococci co-

61 colonized individuals independently of their *blp* pheromone type (p=0.577),

62 bacteriocin/immunity content, *blp* locus activity (p=0.798) and inhibitory phenotype (p=0.716).

63 In addition, no significant differences were observed when single and co-colonizing strains were

64 compared.

65

66 Despite clear evidence of *blp*-mediated competition in experimental models, our study suggests

67 that the *blp* locus plays a limited role in restricting pneumococcal co-colonization in humans.

68

69 **IMPORTANCE**

70 Nasopharyngeal colonization with *Streptococcus pneumoniae* (pneumococcus) is important for

71 pneumococcal evolution as it represents the major site for horizontal gene transfer when

72 multiple strains co-occur, a phenomenon known as co-colonization. Understanding how

73 pneumococcal strains interact within the competitive environment of the nasopharynx is of

74 chief importance in the context of pneumococcal ecology. In this study we used an unbiased

75 collection of naturally co-occurring pneumococcal strains and showed that a biological process

76 frequently used by bacteria for competition – bacteriocin production - is not decisive in the co-

77 existence of pneumococci in the host, contrary to what has been shown in experimental

78 models.

79

80 **INTRODUCTION**

81 *Streptococcus pneumoniae* is an important bacterial pathogen associated with high morbidity

82 and mortality worldwide (1-3). Notwithstanding, disease is a rare event compared with the

83 frequency of asymptomatic nasopharyngeal colonization (4).

4

84

85   Nasopharyngeal colonization is particularly frequent among young children (5) and multiple

86   strains can be simultaneously detected in the same sample (6-8). As pneumococcus is naturally

87   competent and known to evolve mainly through genetic recombination (9), co-colonization is

88   fundamental for its evolution (10, 11). Donker *et al* correlated the extensive recombination

89   among pediatric colonization strains with a high prevalence of co-colonization in that

90   population, highlighting the importance of having a better understanding of this phenomenon.

91

92   Little is known about how pneumococcal strains interact with each other within the

93   nasopharynx. Data from longitudinal studies have shown that the pneumococcal population

94   colonizing children is in constant turnover (5, 12). Although over 95 serotypes have been

95   described to date in the pneumococcal population (13, 14), only a small subset are dominant in

96   carriage suggesting a higher competitive ability of these serotypes (15, 16). The fact that the

97   pneumococcus' natural niche is a dense polymicrobial environment with limited resources also

98   suggests that competitive interactions are important in its lifestyle (17).

99

100 To date, two main molecular mechanisms have been implicated in pneumococcal intra-species

101 competition: competence-mediated fratricide and bacteriocin production (18-21). These

102 systems were historically thought of as independent. However, recent work has demonstrated

103 that they are coordinately regulated suggesting that, under some circumstances, fratricide and

104 bacteriocin production may work in concert to target neighboring cells (22, 23).

105

106   Fratricide is used by competent cells not only to eliminate direct competitors, but also to obtain

107   nutrients and DNA released by non-competent lysed cells (19). Because pneumococcal

108   populations are divided into two pherotypes that express and respond to distinct competence

109   stimulating peptide (CSP) types, it was originally hypothesized that CSP might influence co-

110   colonization as a result of fratricide-mediated exclusion. Two studies that evaluated the impact

111   of CSP pherotype on the co-existence of pneumococci in humans could not confirm this

112   hypothesis (24, 25).

113

114   In *S. pneumoniae*, the *blp* locus (associated with bacteriocin production) has been shown to play

115   a role in intraspecies competition in a murine model of co-colonization (20). The *blp* locus

116   encodes the factors required for the regulation, production and secretion of bacteriocins and

117   their associated immunity proteins. The locus is controlled by a two component regulatory

118   system consisting of a histidine kinase (BlpH) and a response regulator (BlpR). The system is

119   activated when the signalling peptide, BlpC, binds to its cognate BlpH receptor (20, 21) resulting

120   in phosphorylation and activation of the regulator, BlpR. BlpR activation results in upregulation

121   of the entire *blp* locus (26) including the bacteriocin/immunity region (BIR) that encodes the

122   bacteriocins and their cognate co-transcribed immunity proteins (27, 28). Bacteriocins (and

123   BlpC) are primarily secreted by the ABC transporter BlpAB (20-23, 26).

124

6

125   The *blp* locus is highly diverse. There is significant variability in the BIR, at least sixteen

126   bacteriocin peptides have been described, several with allelic variability and strains can vary

127   with respect to the number of encoded bacteriocins (27, 28). Also, at least five types of the

128   peptide pheromone BlpC have been found (21, 26-28). In addition, a significant fraction of

129   pneumococcal strains carry a disruption in the genes encoding the BlpAB transporter (27). These

130   strains can respond to exogenous pheromone with the production of immunity proteins but

131   cannot secrete bacteriocins. The prevalence of these strains suggests that strategies that limit

132   locus stimulation and peptide secretion may provide an energetic advantage. It was recently

133   shown that *blpAB* disrupted strains can secrete a fraction of the *blp* pheromone, BlpC, through

134   the homologous competence transporter, ComAB, allowing for self-stimulation of the locus (22,

135   23). This secretion only occurs during the competent state when sufficient ComAB is produced

136   to support BlpC secretion and is enhanced by cross stimulation of some of the *blp* promoters by

137   the competence response regulator, ComE. The universal presence of the *blp* locus in

138   pneumococci and the significant diversity of its content suggest that this locus provides a

139   competitive advantage either during colonization or transmission (28). No studies to date have

140   used colonizing isolates from humans to examine whether *blp* locus activity is sufficiently

141   widespread to restrict the strains that can co-exist within the nasopharynx.

142

143   In this study we explored the role of the *blp* locus and bacteriocin secretion on the co-existence

144   of *S. pneumoniae* in human nasopharyngeal samples.

145

146 **METHODS**

147 **Sample selection and isolation of pneumococci**

148 Nasopharyngeal samples collected from healthy children attending day-care centers in Oeiras

149 and Montemor-o-Novo, Portugal, were retrospectively selected from cross-sectional studies

150 conducted in 2001, 2006, 2007, 2009, and 2010 (29-32). In those studies, samples were

151 routinely plated in selective media for *S. pneumoniae* (5% blood trypticase soy agar containing

152 5mg/L gentamicin) and incubated overnight at 37°C under anaerobic conditions with an

153 optochin disk. On the second day, pneumococcal colonies with different morphologies were

154 picked and plated separately, one colony per morphology. In the majority of the cases (>90%)

155 only one morphology was detected with naked eye. The bacterial lawn of the primary selective

156 plate, herein called "primary sample", was collected and frozen at -80°C in 1mL Mueller-Hinton

157 broth with 30% glycerol. On the third day, pneumococcal pure cultures were also frozen.

158

159 Pneumococcal strains were serotyped by multiplex PCR as described previously (33, 34),

160 [http://www.cdc.gov/streplab/pcr.html], or by the Quellung reaction, as appropriate

161 (35)(Statens Serum Institute, Copenhagen, Denmark).

162

163 To maximize the likelihood of detecting significant associations between serotypes in multiple

164 carriage events, and given the high serotype diversity of pneumococci, we first identified all

165 samples containing pneumococci. Then, based on our previous observation that the serotypes

166 more frequently found in single and multiple carriage events are essentially the same (36), we

167 identified the serotypes with highest cumulative prevalence in the cross-sectional studies

168 described above. These were: 3, 6A, 6C, 11A, 15A, 15B/C, 19A, 19F, 21, 23A, 23B and non-

169 encapsulated pneumococci (NT). In total, 1,415 samples were identified and selected for

170 molecular detection of co-colonization. Of these, 66.2% (936/1,415) were collected from

171 children vaccinated with at least one dose of either the 7- or the 10-valent pneumococcal

172 conjugate vaccine, 30.7% (435/1,415) were collected from non-vaccinated children, and 3.1%

173 (44/1,415) were collected from children from whom the vaccination status was unknown.

174

175 **DNA isolation**

176 Total DNA was isolated from 200µL of the primary sample frozen stock using the High Pure PCR

177 Template Preparation Kit (Roche Diagnostics, Mannheim, Germany) according to the

178 manufacturer's instructions.

179

180 **Detection of co-colonization and capsular typing of the co-colonizing strains**

181 Detection of co-colonization was done using the plyNCR-RFLP (37) and the BuG@S SP-CPSv1.4.0

182 microarray for molecular serotyping as previously described (36).

183

184 All serotypes identified by the microarray were confirmed by PCR using as template purified

185 DNA of the primary sample.

186

9

187   For the interpretation of the microarray results, a serotype was considered a major serotype

188   whenever its relative abundance was ≥70%. Accordingly, a serotype was considered a minor

189   serotype whenever its relative abundance was ≤30%. Serotypes with relative abundances > 30%

190   and <70% were considered co-dominant.

191

192   **Isolation of pneumococcal strains from the co-colonized samples**

193   Aliquots of the primary sample frozen stock were prepared, serially diluted and plated in 5%

194   blood trypticase soy agar containing 5mg/L gentamicin plates to obtain isolated colonies.

195   Individual colonies were picked for amplification of serotype-specific capsular genes. The

196   number of colonies picked was calculated as described by Huebner *et al*. (38). One colony of

197   each capsular type was subcultured and frozen. For isolation and correct identification of non-

198   encapsulated pneumococci the method described by Simões *et al.* was used (39).

199

200   **MLST**

201   Multilocus sequence typing was done as described previously (40). Novel STs and the

202   corresponding strain information were deposited at the public MLST database for *S.*

203   *pneumoniae* (http://pubmlst.org/spneumoniae/).

204

205   **Assignment of pherotype CSP1 or CSP2**

206   Pherotype assignment was done by multiplex PCR amplification of specific *comC* gene

207   fragments, as described previously (41).

10

208

PCR assignment of *blpC* type

Assignment of the five *blpC* types described up to now was done by amplification of specific

blpC gene fragments, using primers designed for this study (Table S1). The PCR reactions varied

according to primer pair, as described in Table S1.

Overlay assays

Inhibition and activity overlay assays were performed as described elsewhere (27) and using the

reporter strains listed in Table 1.

Assessment of *blpA* integrity

Detection of the 4 bp repeat insertion was done by PCR amplification of a *blpA* gene fragment

containing the region of repeat insertion using forward primer 21 described in (27) and reverse

primer with sequence AGCCGCTGATGAAATGGGC, followed by digestion with Cac8I (New

England Biolabs, Ipswich, MA, USA), according to the manufacturer's instructions. Restriction

occurs only when the 4 bp repeat is present.

Samples in which we failed to amplify the region containing the repeat insertion were amplified

with primers 18 and 19 described in (27) for amplification of the total *blpA* gene to confirm

presence of possible deletions by comparison of the size of the PCR product with that of a

control strain with an intact *blpA* gene.

11

229

230 **RFLP analysis of the bacteriocin/immunity region (BIR) and bacteriocin content prediction**

231 RFLP profiles of all isolates were determined as previously described (27). Briefly, PCR products

232 were obtained with primers 1 and 2, purified and digested with AseI (New England Biolabs,

233 Ipswich, MA, USA), according to the manufacturer's instructions. Digestion products were

234 analyzed by capillary electrophoresis and isolates with identical restriction patterns were

235 assigned to the same group.

236

237 BIR content was predicted by comparison of RFLP profiles with those of sequenced strains

238 available at GenBank, using either VectorNTI or CLC Genomics softwares. Overlay assays with

239 reporter strains of known BIR contents (Table 1) were performed as a control for *in silico*

240 predictions.

241

242 **Data analysis**

243 The frequencies of several trait types (serotypes, genotype, CSP, BlpC, BlpA and inhibitory

244 activity) were compared between co-colonization and single colonization samples (sample

245 types). To detect statistically significant differences, 5000 random datasets were generated by

246 randomly permuting the strain's trait and sample types. Two tailed p-values were computed by

247 comparing individual frequencies of each trait type in each sample type to permutation-derived

248 frequencies. The deviation of the complete trait/sample types distribution from the null

249 hypothesis was evaluated through a Fisher's Exact Test. A similar analysis was performed to

250   compare strains classified as minor, co-dominant or major, according to their relative

251   abundance.

252

253   The frequency with which pairs of serotypes (or of genotypes, CSP, BlpC and BlpA types) appear

254   together in co-colonizing samples was also compared with a null hypothesis where strains mix

255   randomly. The null distribution was estimated through 5000 random assignments of strains to

256   the samples with co-colonization events. Each sample had to contain at least one strain from

257   the set of serotypes used to select the samples included in this study (3, 6A, 6C, 11A/D, 15A,

258   15B/C, 19A, 19F, 21, 23A, 23B, NT). A similar analysis was performed to compare the pair type

259   frequency among samples with major/minor versus co-dominant co-colonization events.

260

261   Frequency deviations were considered significant when p-values<0.05.

262

263   **RESULTS**

264   *S. pneumoniae* **strains co-colonizing individuals are highly diverse and co-exist independently**

265   **of the capsular type and genotype**

266   Out of the 1,415 samples included in the study to screen for pneumococcal co-colonization, 285

267   pneumococcal strains were isolated from 135 co-colonized samples (Figure 1). Strains were

268   found to be highly diverse: 36 capsular types were found and MLST genotyping clustered strains

269   in 19 clonal complexes (CC, CC1-CC19) and 36 singletons (S, S1-S36) (Figure 2). MLST analysis

270   revealed 24 novel STs, 10 of which were originated by the presence of new alleles (STs 9146,

13

271   9147, 9148, 9149, 9150, 9151, 9722, 9724, 9725, and 9726) and 14 resulted from novel allele

272   combinations (STs 9152, 9153, 9154, 9155, 9156, 9157, 9158, 9159, 9160, 9161, 9162, 9163,

273   9164, and 9720). Among the latter, for three STs (ST9160, ST9162 and ST9164) the allele

274   separating the novel ST from its SLV was shared with its co-colonizing strain, suggesting that

275   within host recombination might have occurred with the co-colonizing strain (Table 2).

276

277   Positive associations were found between some serotypes and sequence types in this collection,

278   although these correlations are of unclear significance. Statistically significant correlations are

279   listed in Table S2.

280

281   CSP assignment of all strains showed that 35.2% of the interactions occurring in the 135 co-

282   colonized samples were between two strains of CSP1, 10.9% were between two strains of CSP2,

283   and 51.5% were between strains of different pherotypes (CSP1, CSP2, and other type present in

284   strains from serotype 38). When the frequency in which CSP types appear together in co-

285   colonization was compared with the frequency estimated under the hypothesis that strains mix

286   randomly, the results suggested a tendency for co-colonization with strains of different CSP

287   types, although with low robustness (p=0.048).

288

289   Despite some positive associations found at the serotype and genotype levels, the large

290   diversity of our collection seems to suggest that pneumococcal strains co-colonize the human

291   nasopharynx independently of their capsule, genetic background and CSP type.

14

292

293 **Genetic characterization of the *blp* locus of co-colonizing pneumococci reveals high diversity**

294 **of *blp* cassettes**

295 The genetic characterization of the *blp* locus of the co-colonizing strains was focused on the

296 RFLP analysis of the BIR, on the assignment of the *blpC* type and on the assessment of *blpA*

297 integrity. The results are summarized in Figure 2.

298

299 Thirty-nine BIR RFLP profiles were identified, from which 16 could be assigned to known BIR

300 sequences available in GenBank and accounted for 70.9% (n=202/285) of the strains. The

301 remaining 23 profiles corresponded to not yet described BIR regions and accounted for 25.6%

302 (n=73/285) of the strains. In ten strains (3.5%, 10/285) we were not able to amplify the BIR

303 region, despite several attempts. Of note, some BIR profiles, such as profiles 7 and 11, were

304 detected in several clonal complexes (CC). Other profiles, highly prevalent as well, were

305 restricted to a single CC, such as profiles 10 and 8, associated with CC5 and CC2, respectively.

306 Our BIR predictions indicated that the number of bacteriocin peptides present in the co-

307 colonizing strains varied between 2 and 6. Interestingly, profiles predicted to contain the highest

308 number of bacteriocin peptides (profiles 7 and 11) were associated mostly with strains with a

309 disrupted *blpA* (described below).

310

311 **High prevalence of strains with disrupted *blpA* gene co-colonizing in nature**

15

312   Disruption of the *blpA* transporter gene has been previously associated with a *cheater*

313   phenotype, i.e., non-inhibitory immune-only (27). Analysis of the *blpA* gene showed a disruption

314   in the gene in 155 out of 285 strains (54.4%). Of these, 151 strains contained the 4 bp repeat

315   insertion described by Son *et al.* (27) and four strains had a larger deletion of approximately 840

316   bp.

317

318   In the sets of co-colonized samples, 27.9% of dual interactions occurred between two *cheater*

319   strains, 20.6% between strains with intact *blpA* genes (non-*cheaters*), and 51.5% occurred

320   between a *cheater* and a non-*cheater* strain. Comparison of the estimated and observed

321   proportions of events involving strains with same or different *blpA* status (intact or disrupted)

322   did not show a significant difference, suggesting that the *cheater* phenotype alone does not

323   restrict co-colonization of pneumococci (p=0.713).

324

325   **Pheromone peptides BlpC are not equally distributed in the population and do not restrict co-**

326   **colonization**

327   Co-existence of strains secreting different BlpC peptides implies that the strain at higher cell

328   density would activate its *blp* locus earlier and be at a competitive advantage. To assess if the

329   type of BlpC could prevent or facilitate co-existence of pneumococci in the host we determined

330   the *blpC* allelle of the 285 co-colonizing strains. The results showed that 36.8% of the strains

331   were of type $blpC_{T4}$, 22.5% of type $blpC_{P164}$, 22.1% of type $blpC_{6A}$, 16.5% of type $blpC_{R6}$, and 2.1%

16

332   of type *blpC*$_{P155}$. Of note, this latter *blpC* type was associated exclusively with non-encapsulated

333   pneumococci.

334

335   To evaluate the ability of the strains to activate their own locus, signaling overlay assays were

336   performed for all strains using BlpC-specific reporter strains for each BlpC of the four major

337   types (Table 1). The BlpC type was confirmed in all strains in which it was possible to observe

338   locus activity, as indicated by the β-galactosidase activity in the assays (n=59). However, for

339   most strains (79%, n=226/285) we were not able to confirm phenotypically the secretion of the

340   encoded BlpC type. This could be largely explained by the fact that 69% (n=155/226) of the non-

341   inducible strains were in fact *cheaters*. *blpA* disruption was mainly associated with strains of

342   type *blpC*$_{R6}$ and *blpC*$_{T4}$ (p=0.0001 for both, Fisher's exact test) (Figure 3). In the remaining 71

343   strains a *blpA* disruption was not detected. A likely explanation for the lack of BlpC secretion in

344   these strains would be the presence of non-repeat mutations in *blpA*, or mutations in *blpB, blpH*

345   or *blpR*, a *blpC/blpH* mismatch, or mutations in the promotor regions of *blpABC* or *blpRH*, all of

346   which have been described in other collections (22, 27, 45).

347

348   Looking at the sets of co-colonization samples, 40.6% of the dual interactions were between

349   strains of the same BlpC type, while 59.4% were between strains of different BlpC types. When

350   the frequency in which BlpC types appear together in co-colonization was compared with the

351   frequency estimated under the hypothesis that strains mix randomly no significant difference

352   was found (p=0.577). The same analysis was performed considering only pairs of strains in

17

353 which secretion of BlpC was confirmed phenotypically through overlay assays in one of the

354 strains; no significant difference was found as well (p=0.798).

355

356 **Phenotypes of bacteriocin secretion do not restrict co-colonization**

357 To determine the phenotype of bacteriocin secretion, *i.e.*, the inhibitory activity of the

358 pneumococcal strains, overlay assays with a susceptible reporter pneumococcal strain P537

359 (Δ*blp*) were performed.

360

361 Eighty-four out of 285 (29.5%) co-colonizing strains displayed inhibitory activity against P537. Of

362 these, 21 strains (25%) had a disrupted *blpA* gene and no evidence of pheromone secretion

363 using reporter strains (suggesting a non-*blp* mediated inhibition) and 63 strains had an intact

364 *blpA* gene. Among the latter, 53 strains were in the group in which BlpC pheromone secretion

365 was observed in the signaling overlays. For the other 10, the signaling overlay assay had been

366 negative.

367

368 Among the strains that did not display inhibitory activity (n=201), 67% were *cheaters* and 30%

369 did not secrete pheromone in the signaling overlays, despite an apparently intact *blpA* gene.

370

371 In the sets of co-colonized samples, co-existence of strains that displayed an inhibitory

372 phenotype was not different from what would be expected by chance, suggesting that an

373 inhibitory phenotype alone does not prevent co-colonization (p=0.715).

18

374

375   Considering this result, we performed an *in silico* prediction of the outcome of the interaction of

376   each pair of strains in every sample, taking into account the genetic content of the BIR, the

377   integrity of the *blpA* transporter gene and the type of signaling peptide BlpC secreted. Our

378   prediction resulted in 50 outcomes of inhibition (30.3%) and 115 outcomes of no inhibition

379   (69.7%). Comparison of the estimated and predicted proportions of outcomes of inhibition and

380   no inhibition did not show a significant difference (p=0.274).

381

382   The results obtained with the genotypic and phenotypic approaches suggest that the *blp* locus

383   alone does not seem to prevent the co-existence of pneumococci in the host.

384

385   **Phenotypes of bacteriocin secretion are the same in strains isolated from single and co-**

386   **colonization events**

387   In order to assess whether strains found in single and co-colonization events would display

388   different genotypic and/or phenotypic *blp* characteristics, we selected a subset (n=298) of the

389   initial 1,053 nasopharyngeal samples in which only one strain was detected, for comparison

390   with the collection of co-colonized samples. This selection was performed randomly but

391   maintaining a matched number of samples from each year between both collections. The

392   characteristics of this collection are summarized in Table S3.

393

394   Overall, the prevalence of each *blpC* type was similar in the two collections, as well as the

395   distribution of inducible, non-inducible and *cheater* strains within each *blpC* type (Figures 3 and

396   4A). Similarly, no significant differences were observed in the two collections regarding the

397   proportion of strains with an inhibitory phenotype, *cheaters*, and an active locus (Figure 4B).

398

399   **DISCUSSION**

400   Bacteriocin production is a common trait in organisms that reside in polymicrobial communities.

401   Bacteriocin mediated competition has been shown to alter the composition of the microbiota in

402   environmental communities and on human surfaces (46).

403

404   Animal models of pneumococcal colonization have demonstrated that bacteriocin production

405   provides a competitive advantage during establishment on the mucosal surface (20, 27). These

406   experiments were performed by inoculating mixtures of competing strains at a 1:1 ratio

407   simultaneously into the nasopharynx of a mouse. Colonization dynamics in humans is likely to

408   be far more complex than the experimental conditions used in these studies. Although co-

409   colonization with distinct pneumococcal strains is quite common, the relative proximity of these

410   strains within the human host is not known. What is clear, based on extensive genome studies,

411   is that organisms are in close enough proximity to support genetic exchange (9, 47). In this

412   study, we used a collection of co-colonizing and single colonizing isolates that reflect true

413   colonization patterns in the human host to better understand the contribution of bacteriocin

414   production to global colonization dynamics.

415

416   Characterization of the co-colonization strains at the level of serotype and genotype enabled us

417   to investigate positive associations between specific capsular types and genotypes, although no

418   straightforward observations could be found to explain those associations. At the capsular level,

419   we explored properties shown to be dependent on the capsular type, such as the

420   polysaccharide structure, and, by association, the fitness cost of capsule production (48, 49),

421   and the surface charge of the capsular type (50). All these properties have been shown to

422   predict the prevalence of the serotypes (48, 49). Interestingly, Trzciński *et al.* (51) have shown a

423   reproducible hierarchy of capsular types in a mouse model of multiple serotype carriage, which

424   correlated with the metabolic cost and the surface charge measured *in vitro.* Our prevalence

425   results are in agreement with the proposed hypothesis but none of the explored properties

426   could explain the serotype associations identified in this study. This may have been due to the

427   fact that interactions in the human host are likely to be far more complex than in experimental

428   systems.

429

430   The genetic characterization of the *blp locus* of the co-colonizing strains allowed us to show a

431   very large diversity in this locus, originated from the diversity in the *blpC, blpA* and BIR region,

432   supporting observations from other studies (20, 26-28, 52). Notwithstanding, MLST genotyping

433   of the strains enabled us to show a fairly high level of conservation of the *blp locus* among

434   closely-related strains, suggesting that this extensive genetic diversification is occurring at a rate

435   sufficiently low for some clonality to be sustained.

436

437   Overall our genotypic and phenotypic characterization of the *blp locus* of co-colonizing

438   pneumococci seems to suggest that the phenotypes of bacteriocin secretion do not have an

439   impact in the co-existence of pneumococci in the nasopharynx, an observation that was

440   supported by the lack of differences in the results obtained in the characterization of the single

441   carriage isolates. Given the large diversity in this locus, it is not surprising that the effect of *blp*-

442   mediated competition on co-colonization is not as straightforward as might be expected and

443   several aspects must be taken into account.

444

445   First, the high diversity of BlpH receptors due to the existence of naturally occurring chimeras

446   for the *blpH$_{6A}$* gene, as described by Pinchas *et al.* (43), may affect the likelihood of cross

447   stimulation between co-colonizing pairs. These authors have shown that these BlpH variants

448   have different specificities for cognate and non-cognate BlpC peptides, which can affect the

449   outcome of neighboring *blp* locus activation.

450

451   Second, the large array of bacteriocins and immunity proteins that can be present in a strain,

452   and the fact that the pneumococcal natural niche is a polymicrobial environment, raises the

453   hypothesis that this competition mechanism might be used for mediating the interaction not

454   only with bacteria from the same species, but also with other inhabitants of the nasopharynx. In

455   fact, Lux *et al.* (52) have demonstrated the inhibitory activity of pneumococci against bacteria of

456   different species.

457

458   Finally, the natural and highly frequent occurrence of strains that display a *cheater* phenotype

459   also contributes to the complexity of this competition mechanism. The reason why these

460   *cheater* strains are so highly prevalent remains to be addressed and the advantage of displaying

461   this phenotype is not completely clear. On one hand, these strains avoid the fitness-cost of

462   bacteriocin secretion while they are still able to express immunity proteins. On the other hand,

463   they become at risk of elimination upon the encounter of a strain secreting a different BlpC

464   type. Two recent publications (22, 23) have demonstrated a regulatory connection between the

465   *com* system, controlling competence and the *blp* system. These studies demonstrated that BlpC

466   could be secreted by the competence regulated transporter, ComAB under competence

467   inducing conditions. ComAB secreted BlpC could then stimulate the *blp* locus through the BlpHR

468   regulatory mechanism. None of these studies has, however, shown the ability of these strains to

469   use ComAB as a bacteriocin secretion system, suggesting that the previously named "*cheater*"

470   strains may be able to self-activate the locus during competence but have the capability to

471   display only an immune phenotype. We have performed signalling overlays in close to 600

472   strains, from which over half had a disrupted *blpA* gene, and we were never able to see locus

473   activation in a strain with *blpA* disruption. Nevertheless, as the conditions of our assays do not

474   control for competence activation, we have repeated the analysis performed in this study

475   considering that all strains would be able to activate the *blp* locus and express immunity and the

476   conclusion that bacteriocin secretion does not restrict co-colonization was maintained:

477   predictions resulted in 95 outcomes of inhibition, 51 outcomes of no inhibition and 19 events in

478   which it was not possible to predict an outcome due to unknown bacteriocin/immunity contents

479   in the strains (p=0.628, Fisher's exact test).

480

481   Our study has some limitations. First, we did not sequence the entire BIR locus, which might

482   have helped in explaining the lack of BlpC secretion in the 71 strains with an apparently intact

483   *blpA.* Lack of BlpC secretion may have been due to mutations in *blpA, blpB, blpH* or *blpR* (22, 27,

484   45). In fact, looking at over 4,000 pneumococcal genomes, Kjos *et al.* (22) have shown that only

485   23% of the genomes had intact *blpAB* genes, which is in agreement with the 21% proportion of

486   secretor strains that we have obtained with the signaling overlays. Yet other alternative

487   explanations could be a *blpC/blpH* mismatch in these strains (43), lack of sensitivity of the assay,

488   or the requirement of particular conditions for locus activation in those strains. In addition, the

489   few non-inhibitory strains with intact pheromone secretion may have harbored mutations in

490   their bacteriocin genes that would not be detected with the RFLP analysis. Second, we did not

491   determine the opaque/transparent phenotype of our strains, which might have helped to

492   explain the lack of inhibitory activity in the 69 intact strains that failed to inhibit the susceptible

493   strain. Dawid *et al* (53) have shown that opaque and transparent variants of a strain with an

494   intact *blpA* gene had different inhibitory profiles, despite the fact that the amount of *blpMNPO*

495   transcripts was the same. These differences were attributed to different expression levels of the

496   HtrA protease. In particular, in opaque variants expression of HtrA was higher resulting in

497   degradation of BlpC and hence in decreased bacteriocin secretion (42). Still, this link between

498   opacity variants and HtrA expression was shown for a single 6A strain only and it is unknown

24

499 whether this is a general property of pneumococcal strains. Third, the cross-sectional nature of

500 our study prevented us from measuring the effect of *blp*-mediated competition on the duration

501 of carriage. Also, we did not measure the pneumococcal load in the samples and it is possible

502 that the effects of competition might be reflected at the level of carriage density. Even so, we

503 measured the relative proportion of each strain in the samples and could not establish a

504 correlation between outcomes of bacteriocin secretion and strain density (Figure S1).

505

506 Our study has also significant strengths. We have used a very well characterized, natural

507 collection of co- and singly colonized samples identified through the use of highly sensitive

508 molecular methods. This enabled us to obtain an unbiased and highly diverse collection of both

509 single and co-colonizing pneumococcal strains to measure the impact of *blp*-bacteriocin

510 production on competition in the nasopharynx.

511

512 Our results demonstrate the importance of using human samples to support conclusions drawn

513 using idealized animal models and are likely to have implications for other bacteriocin systems

514 that have been exclusively studied in animal models. For example, similar to the case with the

515 *blp* bacteriocins, studies performed in animal models have shown that bacteriocin secretion is

516 important in *Streptococcus mutans* competition in early dental biofilms (54) and in the ability of

517 *Enterococcus faecalis* to colonize of the mammalian gut (55). In light of our study, it would be of

518 interest to determine if the expected correlations hold true when natural samples from the

519 human host are studied.

520

521   Although we have not seen an effect of *blp*-mediated competition on co-colonization, the fact

522   that this locus is present in all pneumococci and has been maintained by evolution (28) suggests

523   an important function, which is supported by the results obtained in competition experiments

524   showing that the locus is active *in vivo* (20). The link between the *blp* and *com* systems suggests

525   that the *blp* locus could be important to increase the DNA pool for transformation (22, 23). Also,

526   it could be important under specific conditions (*e.g.*, nutrient limitation) or during the

527   establishment of a strain during colonization by creating an isolated niche. Moreover, studies

528   with a longitudinal design would help to determine whether *blp*-mediated competition might be

529   acting at the level of carriage duration or density.

530

531   **AUTHOR CONTRIBUTIONS**

532   The study was designed by RSL, CV and SD. Data acquisition, analysis and interpretation were

533   performed by CV, SD, FP, JH, ASS, KG, LAM, and RSL. RSL, SD, JH and HML contributed reagents

534   or materials. The manuscript was drafted by CV and RSL and critically revised by all authors. All

535   authors read and approved the final version of the manuscript.

536

541   collection, analysis, and interpretation of data, writing of the manuscript or in the decision to

542   submit the manuscript for publication.

543

544   **REFERENCES**

545   1.   **CDC.** 2012. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal
546   polysaccharide vaccine for adults with immunocompromising conditions: recommendations of
547   the Advisory Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly
548   Report (MMWR) **61:**816-819.
549   2.   **CDC.** 2014. Use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal
550   polysaccharide vaccine among adults aged ≥65 years: recommendations of the Advisory
551   Committee on Immunization Practices (ACIP). Morbidity and Mortality Weekly Report (MMWR)
552   **63:**822-825.
553   3.   **O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, Lee E, Mulholland K,**
554   **Levine OS, Cherian T.** 2009. Burden of disease caused by *Streptococcus pneumoniae* in children
555   younger than 5 years: global estimates. Lancet **374:**893-902.
556   4.   **Bogaert D, De Groot R, Hermans PW.** 2004. *Streptococcus pneumoniae* colonisation: the key to
557   pneumococcal disease. Lancet Infect Dis **4:**144-154.
558   5.   **Gray BM, Converse GM, 3rd, Dillon HC, Jr.** 1980. Epidemiologic studies of *Streptococcus*
559   *pneumoniae* in infants: acquisition, carriage, and infection during the first 24 months of life. J
560   Infect Dis **142:**923-933.
561   6.   **Hodges RG, Mac LC, Bernhard WG.** 1946. Epidemic pneumococcal pneumonia; pneumococcal
562   carrier studies. Am J Hyg **44:**207-230.
563   7.   **Hansman D, Morris S.** 1988. Pneumococcal carriage amongst children in Adelaide, South
564   Australia. Epidemiol Infect **101:**411-417.
565   8.   **Sá-Leão R, Tomasz A, Santos Sanches I, de Lencastre H.** 2002. Pilot study of the genetic diversity
566   of the pneumococcal nasopharyngeal flora among children attending day care centers. J Clin
567   Microbiol **40:**3577-3585.
568   9.   **Chewapreecha C, Harris SR, Croucher NJ, Turner C, Marttinen P, Cheng L, Pessia A, Aanensen**
569   **DM, Mather AE, Page AJ, Salter SJ, Harris D, Nosten F, Goldblatt D, Corander J, Parkhill J,**
570   **Turner P, Bentley SD.** 2014. Dense genomic sampling identifies highways of pneumococcal
571   recombination. Nat Genet **46:**305-309.
572   10.   **Barnes DM, Whittier S, Gilligan PH, Soares S, Tomasz A, Henderson FW.** 1995. Transmission of
573   multidrug-resistant serotype 23F *Streptococcus pneumoniae* in group day care: evidence
574   suggesting capsular transformation of the resistant strain in vivo. J Infect Dis **171:**890-896.
575   11.   **Hiller NL, Ahmed A, Powell E, Martin DP, Eutsey R, Earl J, Janto B, Boissy RJ, Hogg J, Barbadora**
576   **K, Sampath R, Lonergan S, Post JC, Hu FZ, Ehrlich GD.** 2010. Generation of genic diversity among
577   *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal
578   pediatric infection. PLoS Pathog **6:**e1001108.
579   12.   **Sá-Leão R, Nunes S, Brito-Avô A, Alves CR, Carriço JA, Saldanha J, Almeida JS, Santos-Sanches I,**
580   **de Lencastre H.** 2008. High rates of transmission of and colonization by *Streptococcus*

27

581 *pneumonia*e and *Haemophilus influenzae* within a day care center revealed in a longitudinal
582 study. J Clin Microbiol **46:**225-234.
583 13. **Calix JJ, Nahm MH.** 2010. A new pneumococcal serotype, 11E, has a variably inactivated wcjE
584 gene. J Infect Dis **202:**29-38.
585 14. **Park IH, Geno KA, Yu J, Oliver MB, Kim KH, Nahm MH.** 2015. Genetic, biochemical, and
586 serological characterization of a new pneumococcal serotype, 6H, and generation of a
587 pneumococcal strain producing three different capsular repeat units. Clin Vaccine Immunol
588 **22:**313-318.
589 15. **Auranen K, Mehtala J, Tanskanen A, Kaltoft SM.** 2009. Between-strain competition in
590 acquisition and clearance of pneumococcal carriage--epidemiologic evidence from a longitudinal
591 study of day-care children. Am J Epidemiol **171:**169-176.
592 16. **Lipsitch M, Dykes JK, Johnson SE, Ades EW, King J, Briles DE, Carlone GM.** 2000. Competition
593 among *Streptococcus pneumoniae* for intranasal colonization in a mouse model. Vaccine
594 **18:**2895-2901.
595 17. **Garcia-Rodriguez JA, Fresnadillo Martinez MJ.** 2002. Dynamics of nasopharyngeal colonization
596 by potential respiratory pathogens. J Antimicrob Chemother **50 Suppl S2:**59-73.
597 18. **Guiral S, Mitchell TJ, Martin B, Claverys JP.** 2005. Competence-programmed predation of
598 noncompetent cells in the human pathogen *Streptococcus pneumoniae*: genetic requirements.
599 Proc Natl Acad Sci U S A **102:**8710-8715.
600 19. **Havarstein LS, Martin B, Johnsborg O, Granadel C, Claverys JP.** 2006. New insights into the
601 pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor.
602 Mol Microbiol **59:**1297-1307.
603 20. **Dawid S, Roche AM, Weiser JN.** 2007. The blp bacteriocins of *Streptococcus pneumoniae*
604 mediate intraspecies competition both in vitro and in vivo. Infect Immun **75:**443-451.
605 21. **Reichmann P, Hakenbeck R.** 2000. Allelic variation in a peptide-inducible two-component system
606 of *Streptococcus pneumoniae*. FEMS Microbiol Lett **190:**231-236.
607 22. **Kjos M, Miller E, Slager J, Lake FB, Gericke O, Roberts IS, Rozen DE, Veening JW.** 2016.
608 Expression of *Streptococcus pneumoniae* bacteriocins is induced by antibiotics via regulatory
609 interplay with the competence system. PLoS Pathog **12:**e1005422.
610 23. **Wholey WY, Kochan TJ, Storck DN, Dawid S.** 2016. Coordinated bacteriocin expression and
611 competence in *Streptococcus pneumoniae* contributes to genetic adaptation through neighbor
612 predation. PLoS Pathog **12:**e1005413.
613 24. **Valente C, De Lencastre H, Sá-Leão R.** 2012. Pherotypes of co-colonizing pneumococci among
614 Portuguese children. Microb Drug Resist **18:**550-554.
615 25. **Vestrheim DF, Gaustad P, Aaberge IS, Caugant DA.** 2011. Pherotypes of pneumococcal strains
616 co-existing in healthy children. Infect Genet Evol **11:**1703-1708.
617 26. **de Saizieu A, Gardes C, Flint N, Wagner C, Kamber M, Mitchell TJ, Keck W, Amrein KE, Lange R.**
618 2000. Microarray-based identification of a novel *Streptococcus pneumoniae* regulon controlled
619 by an autoinduced peptide. J Bacteriol **182:**4696-4703.
620 27. **Son MR, Shchepetov M, Adrian PV, Madhi SA, de Gouveia L, von Gottberg A, Klugman KP,**
621 **Weiser JN, Dawid S.** 2011. Conserved mutations in the pneumococcal bacteriocin transporter
622 gene, *blpA*, result in a complex population consisting of producers and cheaters. MBio **2**.
623 28. **Bogaardt C, van Tonder AJ, Brueggemann AB.** 2015. Genomic analyses of pneumococci reveal a
624 wide diversity of bacteriocins - including pneumocyclicin, a novel circular bacteriocin. BMC
625 Genomics **16:**554.

626 29. **Mato R, Sanches IS, Simas C, Nunes S, Carriço JA, Sousa NG, Frazão N, Saldanha J, Brito-Avô A,**
627   **Almeida JS, Lencastre HD.** 2005. Natural history of drug-resistant clones of *Streptococcus*
628   *pneumoniae* colonizing healthy children in Portugal. Microb Drug Resist **11:**309-322.
629 30. **Sá-Leão R, Nunes S, Brito-Avô A, Frazão N, Simões AS, Crisóstomo MI, Paulo AC, Saldanha J,**
630   **Santos-Sanches I, de Lencastre H.** 2009. Changes in pneumococcal serotypes and antibiotypes
631   carried by vaccinated and unvaccinated day-care centre attendees in Portugal, a country with
632   widespread use of the seven-valent pneumococcal conjugate vaccine. Clin Microbiol Infect
633   **15:**1002-1007.
634 31. **Simões AS, Pereira L, Nunes S, Brito-Avô A, de Lencastre H, Sá-Leão R.** 2011. Clonal evolution
635   leading to maintenance of antibiotic resistance rates among colonizing pneumococci in the PCV7
636   era in Portugal. J Clin Microbiol **49:**2810-2817.
637 32. **Nunes S, Valente C, Simões AS, Paulo AC, Brito-Avô A, de Lencastre H, Sá-Leão R.** 2012. *In the*
638   *era of pneumococcal conjugate vaccines, antibiotic consumption remains a main driving force of*
639   *antimicrobial resistance*. Abstract presented at the - 8th International Symposium on
640   Pneumococci & Pneumococcal Diseases (ISPPD8), Iguaçu Falls, Brazil. March 11 - 15. Abstract
641   no.164. Available at the internet:
642   http://www2.kenes.com/ISPPD/Scientific/Documents/FinalAbstractbook.pdf.
643 33. **Pai R, Gertz RE, Beall B.** 2006. Sequential multiplex PCR approach for determining capsular
644   serotypes of *Streptococcus pneumoniae* isolates. J Clin Microbiol **44:**124-131.
645 34. **Brito DA, Ramirez M, de Lencastre H.** 2003. Serotyping *Streptococcus pneumoniae* by multiplex
646   PCR. J Clin Microbiol **41:**2378-2384.
647 35. **Sorensen UB.** 1993. Typing of pneumococci by using 12 pooled antisera. J Clin Microbiol
648   **31:**2097-2100.
649 36. **Valente C, Hinds J, Pinto F, Brugger SD, Gould K, Mühlemann K, de Lencastre H, Sá-Leão R.**
650   2012. Decrease in pneumococcal co-colonization following vaccination with the seven-valent
651   pneumococcal conjugate vaccine. PLoS One **7:**e30235.
652 37. **Brugger SD, Hathaway LJ, Muhlemann K.** 2009. Detection of *Streptococcus pneumoniae* strain
653   cocolonization in the nasopharynx. J Clin Microbiol **47:**1750-1756.
654 38. **Huebner RE, Dagan R, Porath N, Wasas AD, Klugman KP.** 2000. Lack of utility of serotyping
655   multiple colonies for detection of simultaneous nasopharyngeal carriage of different
656   pneumococcal serotypes. Pediatr Infect Dis J **19:**1017-1020.
657 39. **Simões AS, Valente C, de Lencastre H, Sá-Leão R.** 2011. Rapid identification of noncapsulated
658   *Streptococcus pneumoniae* in nasopharyngeal samples allowing detection of co-colonization and
659   reevaluation of prevalence. Diagn Microbiol Infect Dis **71:**208-216.
660 40. **Enright MC, Spratt BG.** 1998. A multilocus sequence typing scheme for *Streptococcus*
661   *pneumoniae*: identification of clones associated with serious invasive disease. Microbiology **144 (**
662   **Pt 11):**3049-3060.
663 41. **Carrolo M, Pinto FR, Melo-Cristino J, Ramirez M.** 2009. Pherotypes are driving genetic
664   differentiation within *Streptococcus pneumoniae*. BMC Microbiol **9:**191.
665 42. **Kochan TJ, Dawid S.** 2013. The HtrA protease of *Streptococcus pneumoniae* controls density-
666   dependent stimulation of the bacteriocin blp locus via disruption of pheromone secretion. J
667   Bacteriol **195:**1561-1572.
668 43. **Pinchas MD, LaCross NC, Dawid S.** 2015. An electrostatic interaction between BlpC and BlpH
669   dictates pheromone specificity in the control of bacteriocin production and immunity in
670   *Streptococcus pneumoniae*. J Bacteriol **197:**1236-1248.

671   44.   **Kim JO, Weiser JN.** 1998. Association of intrastrain phase variation in quantity of capsular
672   polysaccharide and teichoic acid with the virulence of *Streptococcus pneumoniae*. J Infect Dis
673   **177:**368-377.
674   45.   **Miller EL, Abrudan MI, Roberts IS, Rozen DE.** 2016. Diverse ecological strategies are encoded by
675   *Streptococcus pneumoniae* bacteriocin-like peptides. Genome Biol Evol **8:**1072-1090.
676   46.   **Riley MA, Wertz JE.** 2002. Bacteriocins: evolution, ecology, and application. Annu Rev Microbiol
677   **56:**117-137.
678   47.   **Donkor ES, Bishop CJ, Gould K, Hinds J, Antonio M, Wren B, Hanage WP.** 2011. High levels of
679   recombination among *Streptococcus pneumoniae* isolates from the Gambia. MBio **2:**e00040-
680   00011.
681   48.   **Hathaway LJ, Brugger SD, Morand B, Bangert M, Rotzetter JU, Hauser C, Graber WA, Gore S,**
682   **Kadioglu A, Muhlemann K.** 2012. Capsule type of *Streptococcus pneumoniae* determines growth
683   phenotype. PLoS Pathog **8:**e1002574.
684   49.   **Weinberger DM, Trzcinski K, Lu YJ, Bogaert D, Brandes A, Galagan J, Anderson PW, Malley R,**
685   **Lipsitch M.** 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence.
686   PLoS Pathog **5:**e1000476.
687   50.   **Li Y, Weinberger DM, Thompson CM, Trzcinski K, Lipsitch M.** 2013. Surface charge of
688   *Streptococcus pneumoniae* predicts serotype distribution. Infect Immun **81:**4519-4524.
689   51.   **Trzcinski K, Li Y, Weinberger DM, Thompson CM, Cordy D, Bessolo A, Malley R, Lipsitch M.**
690   2015. Effect of serotype on pneumococcal competition in a mouse colonization model. MBio **6**.
691   52.   **Lux T, Nuhn M, Hakenbeck R, Reichmann P.** 2007. Diversity of bacteriocins and activity
692   spectrum in *Streptococcus pneumoniae*. J Bacteriol **189:**7741-7751.
693   53.   **Dawid S, Sebert ME, Weiser JN.** 2009. Bacteriocin activity of *Streptococcus pneumoniae* is
694   controlled by the serine protease HtrA via posttranscriptional regulation. J Bacteriol **191:**1509-
695   1518.
696   54.   **Kuramitsu HK, He XS, Lux R, Anderson MH, Shi WY.** 2007. Interspecies interactions within oral
697   microbial communities. Microbiology and Molecular Biology Reviews **71:**653.
698   55.   **Kommineni S, Bretl DJ, Lam V, Chakraborty R, Hayward M, Simpson P, Cao Y, Bousounis P,**
699   **Kristich CJ, Salzman NH.** 2015. Bacteriocin production augments niche competition by
700   enterococci in the mammalian gastrointestinal tract. Nature **526:**719-722.
701

30

702 **FIGURE LEGENDS**

703 **Figure 1. Strategy used for identification of co-colonized samples and isolation of co-colonizing**

704 **pneumococcal strains.**

705

706 **Figure 2. Genetic diversity and characteristics of the blp locus of co-colonizing pneumococci.**

707 Interpretation of results was performed using the goeBURST algorithm to determine possible

708 evolutionary relationships between isolates: strains sharing 5 out of 7 alleles were considered

709 genetically related. Clonal complexes (CC) were assigned considering only the collection of co-

710 colonizing pneumococci. Each circle represents a sequence type (ST). The size of the circle is

711 proportional to the number of isolates of that ST. The serotype(s) of the isolates belonging to a

712 given ST is indicated inside each circle. Related ST are grouped in clonal complexes (CC, Panel A)

713 and unrelated ST are represented as singletons (S, Panel B). For each ST information on CSP type

714 and genetic characterization of *blp* locus is represented by the bars; the first section represents

715 CSP type (CSP1 in black, CSP2 in light gray, other type in dark gray); the second section

716 represents *blpC* type (dark green for *blpC*$_{P164}$, light green for *blpC*$_{R6}$, dark blue for *blpC*$_{6A}$, light

717 blue for *blpC*$_{T4,}$ and pink for *blpC*$_{P155}$); the third section represents presence (red cross) or

718 absence (white) of disruption in *blpA*; the fourth section represents the BIR profile, numbered 1

719 to 39. BIR RFLP profiles matched to genome sequenced strains (GenBank designations): 1

720 (GA13723), 2 (AP200), 3 (GA13856), 4 (GA54354), 5 (20703335), 7 (GA47439), 9 (670-6B), 10

721 (OXO141), 11 (GA17971), 12 (CDC1087), 14 (46518), 18 (SP14-BS69), 19 (GA47502), 20 (70585),

722   21 (8a-SA64), and 22 (5a-14-3). ND, BIR profiles not determined due to failure to amplify the BIR

723   region.

724

725   **Figure 3. Distribution of *blpC* types in the co-colonization strains.** Asterisks indicate statistically

726   significant differences. Black, active secretion; dark gray, no secretion, *cheaters*; white, no

727   secretion, putative non *cheaters*; light gray, not tested for active secretion, putative non

728   *cheaters*. Error bars represent standard error.

729

730   **Figure 4. Characterization of the *blp* locus of the strains isolated from single carriage events.**
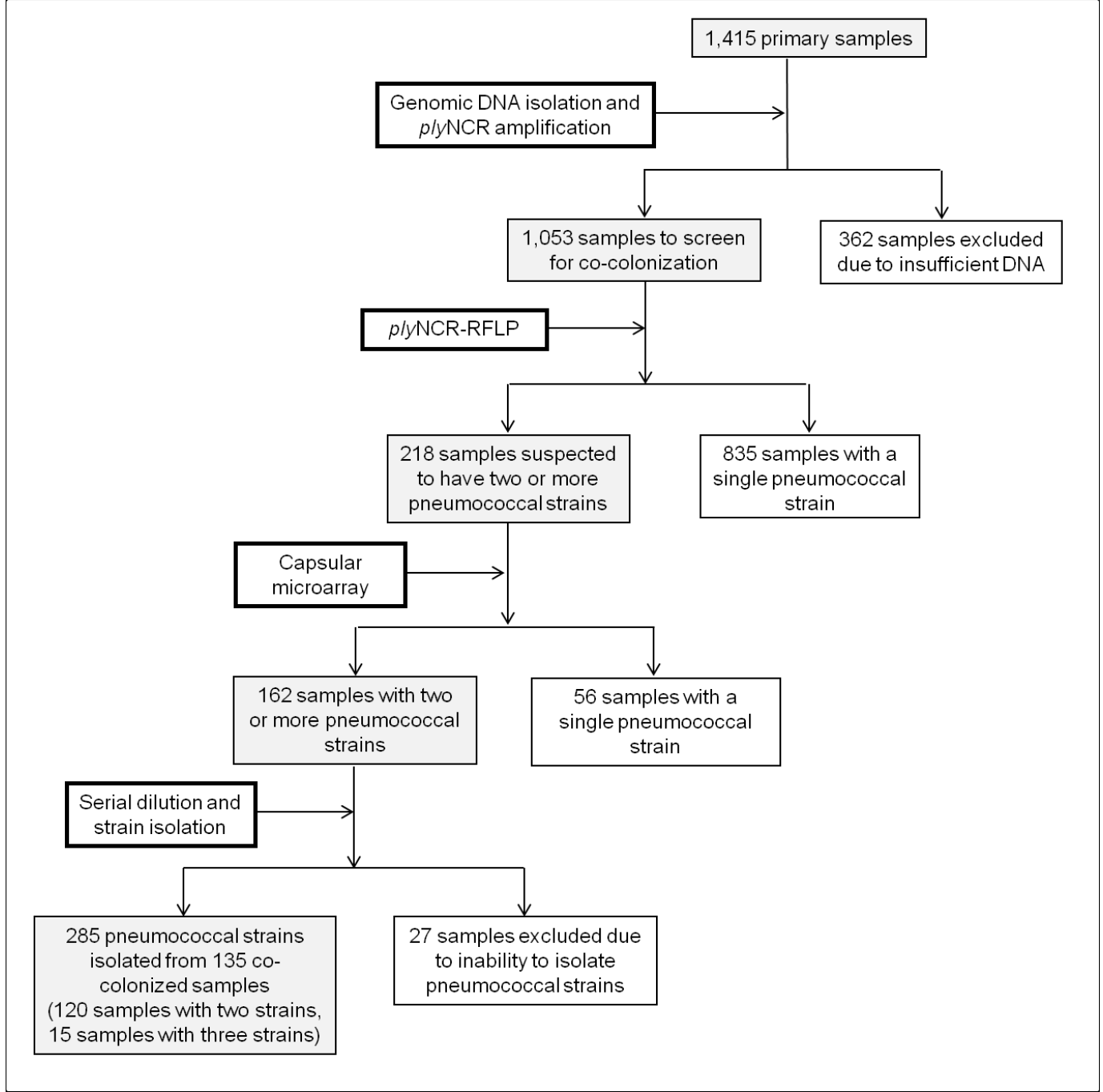
731   (A) Distribution of *blpC* types in the co-colonization strains. Asterisk indicates statistically

732   significant differences. Black, active secretion; dark gray, no secretion, *cheaters*; white, no

733   secretion, putative non *cheaters*; light gray, not tested for active secretion, putative non

734   *cheaters*. Error bars represent standard error. (B) Comparison of the single and co-colonized

735   collections regarding proportions of strains with inhibitory phenotype, cheater genotype and

736   locus inducibility. Dark gray, co-colonization strains; white, single carriage strains. Error bars

737   represent standard error.

738

739
740
741

32

1,415 primary samples

Genomic DNA isolation and *ply*NCR amplification

1,053 samples to screen for co-colonization

362 samples excluded due to insufficient DNA

*ply*NCR-RFLP

218 samples suspected to have two or more pneumococcal strains

835 samples with a single pneumococcal strain

Capsular microarray

162 samples with two or more pneumococcal strains

56 samples with a single pneumococcal strain

Serial dilution and strain isolation

285 pneumococcal strains isolated from 135 co-colonized samples (120 samples with two strains, 15 samples with three strains)

27 samples excluded due to inability to isolate pneumococcal strains

% of pneumococci

45
40
35
30
25
20
15
10
5
0

BlpC$_{P164}$ BlpC$_{R6}$ BlpC$_{6A}$ BlpC$_{T4}$ BlpC$_{P155}$

Values on bars:
- BlpC$_{P164}$: 9.5 (white), 5.3 (dark grey), 7.7 (black)
- BlpC$_{R6}$: 1.1 (white), 15.1 (dark grey), 0.4 (black)
- BlpC$_{6A}$: 8.8 (white), 2.8 (dark grey), 10.5 (black)
- BlpC$_{T4}$: 3.5 (white), 31.2 (dark grey), 2.1 (black), *
- BlpC$_{P155}$: 2.1 (light grey), *

☐ No secretion, putative non *cheaters*

▨ No secretion, *cheaters* (disrupted *blpA*)

■ Active secretion

☐ Not tested for activity, putative non *cheaters*

A

B

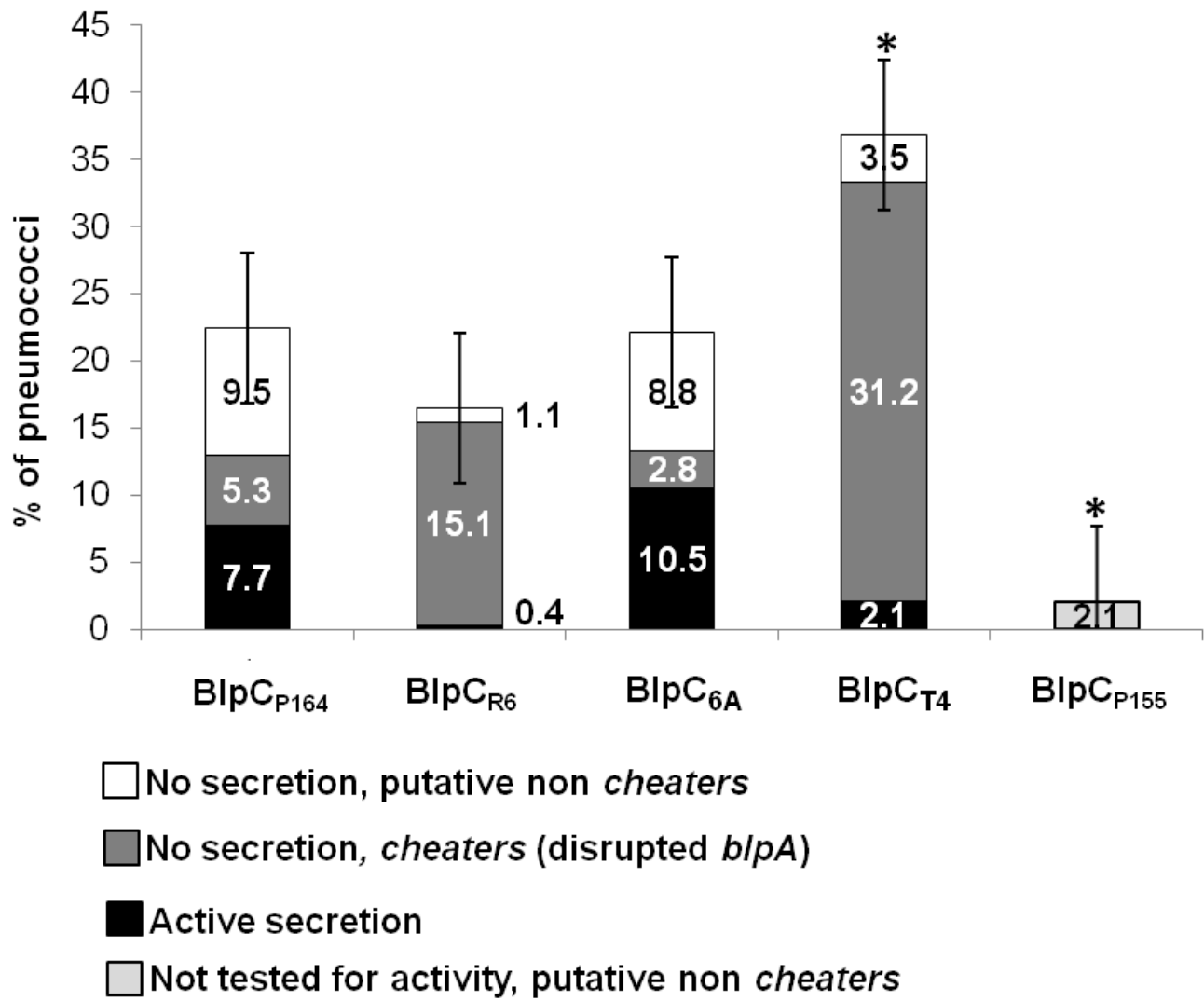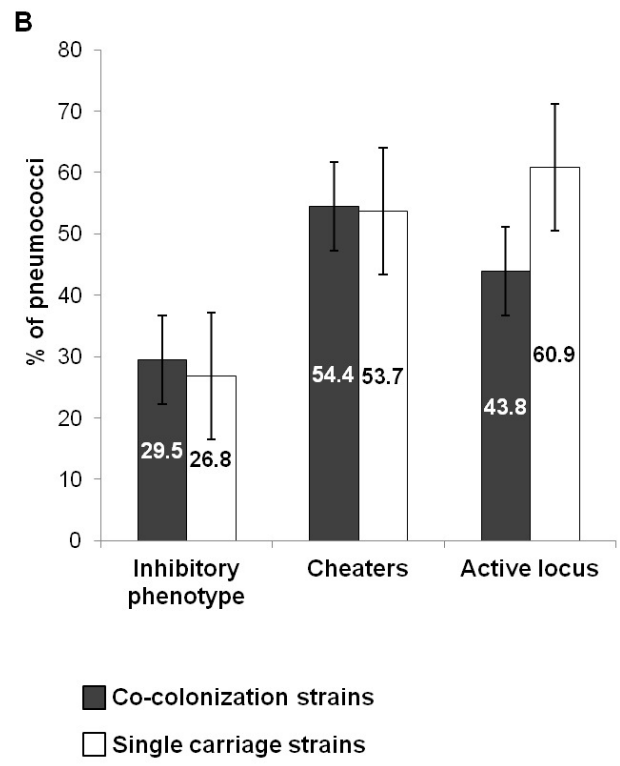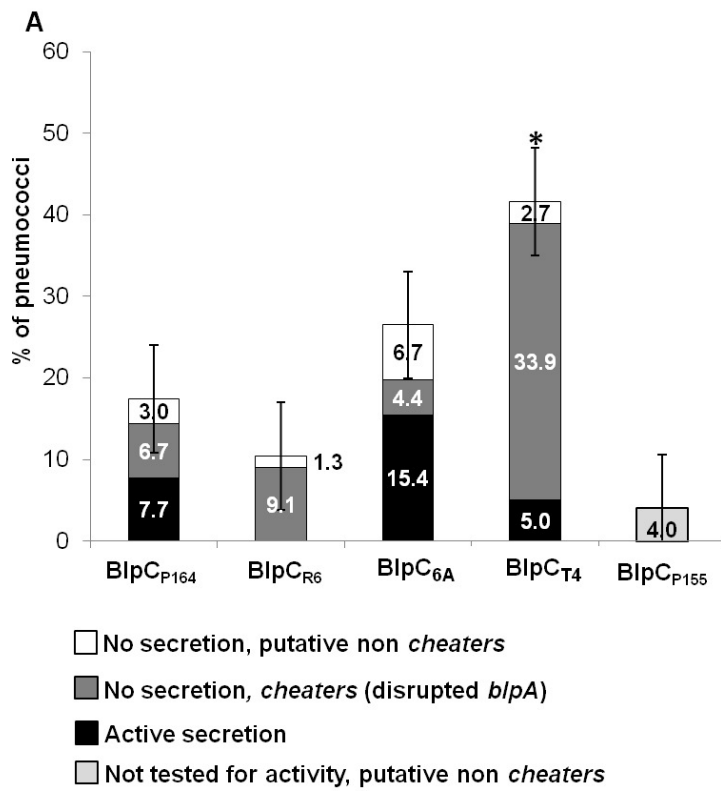**Table 1. Reporter and control strains for characterization of the *blp* locus**

Strain designation	Strain characteristics	Reference
P537	Serotype 6A strain with deletion of *blpRHCBA-BIR*. Susceptible to all bacteriocins secretion. Reporter of inhibitory activity	Son *et al*., 2011 (27)
PSD121	R6 background; insertion of a type P164 *blp* locus, with *blpC* deletion and *lac-Z* reporter under control of the BIR promoter. Type P164 BlpC/BlpH reporter	Kochan *et al.*, 2013 (42)
PSD101	R6 background; insertion of a type R6 *blp* locus, with *blpC* deletion and *lac-Z* reporter under control of the BIR promoter. Type R6 BlpC/BlpH reporter	Pinchas *et al.*, 2015 (43)
PMP105	R6 background; insertion of a type 6A *blp* locus, with *blpC* deletion and *lac-Z* reporter under control of the BIR promoter. Type 6A BlpC/BlpH reporter	Pinchas *et al.*, 2015 (43)
PMP105	R6 background; insertion of a type T4 *blp* locus, with *blpC* deletion and *lac-Z* reporter under control of the BIR promoter.	Pinchas *et al.*, 2015 (43)

	Type T4 BlpC/BlpH reporter	
P1	Serotype 6A, BIR with MNO, BlpC$_{6A}$. Control for MNO BIR content	Kim *et al.*, 1998 (44)
P4	Serotype 6B, BIR with QMNO, BlpC$_{P164}$. Control for QMNO BIR content	Kim *et al.*, 1998 (44)
P132	Serotype 29, BIR with K, type 3 BlpC$_{6A}$. Control for K BIR content	Son *et al.*, 2011 (27)
P133	Serotype 6A, BIR with IJK MNO, BlpC$_{R6}$. Control for IJK MNO BIR content	Son *et al.*, 2011 (27)
P140	Serotype 35B, BIR with IJK, BlpC$_{R6}$. Control for IJK BIR content	Son *et al.*, 2011 (27)

Table 2. Evidence for putative *in vivo* horizontal gene transfer among co-colonizing strains in nasopharyngeal samples characterized in this study.

NP sample	Serotype	ST	aroE	gdh	gki	recP	spi	xpt	ddl
Sample 8043									
Strain 1	19A	**9160**	**1**	60	9	8	6	3	29
Strain 2	6C	395	**1**	5	7	12	17	1	14
Strain 3	31	1766	**1**	5	29	1	46	14	18
Closest ST to strain 1[a]	19A	1151	7	60	9	8	6	3	29
Sample 8058									
Strain 1	3	**9162**	**2**	15	2	10	6	1	22
Strain 2	23A	**9163**	**2**	8	9	9	6	4	6
Closest ST to strain 1[a]	3	180	7	15	2	10	6	1	22
Closest ST to strain 2[a]	23A/23F	190	8	8	9	9	6	4	6
Sample 8098									
Strain 1	6B	**9164**	**7**	8	4	18	15	4	31
Strain 2	3	180	**7**	15	2	27	2	11	71
Closest ST to strain 1[a]	6A/6B	5516	2	8	4	18	15	4	31

[a] Closest ST found in the MLST database (http://spneumoniae.mlst.net/). NP, nasopharyngeal; ST, sequence type. Novel STs and alleles shared with co-colonizing strain that generated novel STs represented in bold. In three samples, at least one of the strains generated a novel MLST profile not previously described in the MLST database. In all cases, the closest match was a single-locus variant. The novel allele was shared with its co-colonizing strain, suggesting within host recombination.