Evolutionary epidemiology of *Streptococcus iniae*: linking mutation rate dynamics with adaptation to novel immunological landscapes

Oleksandra Silayeva, Jan Engelstädter, and Andrew C Barnes*

The University of Queensland, School of Biological Sciences, St Lucia Campus, Brisbane, Queensland 4072, Australia

Key words: DNA repair deficiency, heritable mutator phenotype; antigenic variation; immune escape, vaccine-induced serotype replacement; host jumps

***Correspondence:**

a.barnes@uq.edu.au

Abstract

Pathogens continuously adapt to changing host environments where variation in their virulence and antigenicity is critical to their long-term evolutionary success. The emergence of novel variants is accelerated in microbial mutator strains (mutators) deficient in DNA repair genes, most often from mismatch repair and oxidised-guanine repair systems (MMR and OG respectively). Bacterial MMR/OG mutants are abundant in clinical samples and show increased adaptive potential in experimental infection models, yet the role of mutators in the epidemiology and evolution of infectious disease is not well understood. Here we investigated the role of mutation rate dynamics in the evolution of a broad host range pathogen, *Streptococcus iniae,* using a set of 80 strains isolated globally over 40 years. We have resolved phylogenetic relationships using non-recombinant core genome variants, measured *in vivo* mutation rates by fluctuation analysis, identified variation in major MMR/OG genes and their regulatory regions, and phenotyped the major traits determining virulence in streptococci. We found that both mutation rate and MMR/OG genotype are remarkably conserved within phylogenetic clades but significantly differ between major phylogenetic lineages. Further, variation in MMR/OG loci correlates with occurrence of atypical virulence-associated phenotypes, infection in atypical hosts (mammals), and atypical tissue of a vaccinated primary hosts (barramundi bone). These findings suggest that mutators are likely to facilitate adaptations preceding major diversification events, and may promote emergence of variation permitting colonisation of a novel host tissue, novel host taxa (host jumps), and immune-escape in the vaccinated host.

33 **1. *Introduction***

34 Microbial evolution during infection can increase pathogen fitness that, while in the host, is largely determined by
35 interactions with immune responses [1,2]. Variation in virulence and antigenicity determinants allows pathogens to escape
36 immune clearance and spread in a host population [2,3]. Therefore, antigenic variation plays major role in the epidemiology
37 of infectious diseases and often compromises their control by vaccination [4,5]. As vertebrate immune responses are
38 remarkably diverse and complex, every individual host and tissue represents a unique 'immunological niche' that requires
39 adaptation[6]. In bacteria, host-adaptive variation could be obtained via accrual of mutation, recombination, and
40 transposition [7]. Notably, genes involved in host-pathogen interactions are often represent "mutation hot-spots" as they
41 contain sequence physically pre-disposed to errors during replication[8]. Further, overall fidelity of replication in bacteria is
42 influenced by the environmental conditions, and elevation of mutation rate, known as stress-induced mutagenesis (SIM),
43 occurs in adapting populations[9,10]. One aspect of SIM is temporary physiological shifts producing higher mutation rate, such
44 as up-regulation of error-prone polymerases that bypass DNA lesions during the bacterial general stress response, SOS [9,10].
45 Another aspect of SIM is increased frequency of mutators - heritably hypermutable strains containing mutator alleles, also
46 referred to as constitutive or permanent mutators, to distinguish them from temporary (physiologically) hypermutable
47 cultures[9-11]. Mutator strains emerge via disruption of DNA repair genes, most often belonging to the mismatch-repair
48 (MMR) and oxidized guanine (OG) repair systems [12,13]. MMR corrects base mis-parings and short insertion/deletion loops
49 that occur during DNA replication [14,15]. It also prevents incorporation of divergent DNA sequences via non-homologous and
50 homeologous (partially homologous) recombination; thus MMR disfunction produces not only hypermutable but also
51 hyper-recombinable phenotypes[15,16]. The OG system repairs oxidative DNA damage throughout the cell cycle[17,18]. Although
52 relatively very conserved [19], MMR and OG genes are quite variable among major bacterial pathogens implicating mutation
53 rate dynamics in the evolution of virulence and antigenicity[20,21].

54 In adapting populations, mutator alleles are indirectly selected ("hitchhike") along with stress-specific resistance variants,
55 such as antibiotic resistance or host-immunity resistance mutations[22-24]. Considering pathogens adapt to every individual
56 host, the host population represents an extremely heterogeneous environment where adaptive processes are ongoing and
57 mutator alleles are likely to persist[25]. Indeed, bacterial mutators are abundant among clinical isolates, especially in chronic
58 cases[11,26]. Although it may be partly attributable to antibiotic selection[27-29], the association between frequency of mutators
59 in diagnostic samples and antibiotic treatments is weak[30,31]. Further, unexposed to antimicrobials, knockout mutator strains
60 show shifts in virulence [32,33] as well as increased colonization potential in mice[34-36]. Thus, prevalence of mutators in clinical
61 isolates most likely results from selective process for host-adaptive mutations during infection [30,32,37,38].

62 Although mutator alleles can accelerate short-term adaptation they are also indirectly selected against along with
63 deleterious variants[24,39-41] , and emerging anti-mutator genotypes start to dominate in the adapted mutator clone[42,43].
64 When copy-number variant confers a mutator phenotype the reverse mutation can fully restore the non-mutator genotype
65 and phenotype[44]. The latter is also possible when prophage integration confers the mutator phenotype and excision the
66 non-mutator phenotype or vice versa[45]. However, when anti-mutator alleles originate via random local mutation or
67 recombination in a mutator strain, the final adapted mutant is likely to have a mutation rate slightly different to the original
68 mutation rate along with traceable shifts MMR/OG genotype[43,46,47]. Therefore, we hypothesized that mutation rate
69 dynamics produced by emergence of mutator and compensatory anti-mutator alleles could be investigated and linked to

70 adaptive evolution via coupling of comparative phylogenomics with mutation rate and pathogenicity phenomics. Here we
71 used *Streptococcus iniae*, a rapidly evolving aquatic pathogen with global distribution, broad-host range, and record of re-
72 infection of previously immunised farmed fish [48,49], as a model organism to evaluate the potential role of mutators in
73 epidemiology. We determined phylogenetic relationship based on non-recombinant core genome single nucleotide
74 polymorphisms (SNPs), identified variation in the *in vivo* mutation rates estimated by fluctuation analysis [50,51], and variation
75 in major MMR (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*) [52] and OG (*mutY*, *mutM*, *mutT*)[21] coding and regulatory loci among 80
76 diverse strains of *S. iniae* isolated globally over 40 years. This revealed that shifts in mutation rate and MMR/OG genotype
77 correlate with phylogenetic diversifications of *S. iniae*, but, in contrast, are highly conserved within phylogenetic clades. To
78 determine whether diversification is mainly driven by the host-adaptive evolution, we identify variation in major
79 phenotypic traits that determine virulence in streptococci (capsular polysaccharide, hemolysin, length of cell chains) [53-55]
80 and bacteria in general (resistance to reactive oxygen species (ROS), and biofilm formation) [56,57]. We find that occurrence
81 of atypical phenotypes for these traits correlates strongly with number of SNPs in DNA repair loci, and prevails in isolates
82 from unusual immunological landscapes such as mammalian host and bone tissue of vaccinated fish. This suggests that
83 fluctuations in mutation rate may promote evolution of virulence, transmission between divergent host species, and
84 occurrence of vaccine-escape mutants.

85 **2. Materials and methods**

86

87 **Strains and growth conditions**

88 Eighty isolates of *S. iniae* collected in Australia, USA, Canada, Israel, Honduras, and Thailand between 1976 and 2016 from
89 eight fish species (*Lates calcarifer*, *Scortum barcoo, Epalzeorhynchos frenatum, Epalzeorhynchos bicolor, Oreochromis sp.,*
90 *Channa striata, Chromobotia macracanthus, Oncorhynchus mykiss*) and three mammalian species (*Homo sapiens, Inia*
91 *geoffrensis, Pteropus alecto*) were analysed (Table 1). Strains were received from culture collections, veterinarians, or
92 directly from fish farms (Suppl. Table 3) and stored as master seed stocks without further subculture at -80°C in Todd-
93 Hewitt Broth (THB, Oxoid) + 20% glycerol. Frozen stocks were recovered on Columbia agar supplemented with 5%
94 defibrinated sheep blood (Oxoid), and cultured at 28°C on Todd-Hewitt agar (THA) or in THB with agitation 200 rpm unless
95 otherwise specified.

96 **Estimation of mutation rates by fluctuation analysis**

97 A fluctuation analysis assay for spontaneous occurrence of rifampicin resistance was optimized according to Rosche and
98 Foster [58]. A single broth culture was initiated from five separate colonies, recovered on Columbia blood agar from stock,
99 and grown overnight to late-exponential phase in THB. Cultures were adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 1:100,
100 and distributed in 200 µl aliquots into 8 wells of a sterile U-bottom 96-well plate (Greiner). Prior to dilution, 100 µl of each
101 OD_{600}-adjusted culture was spread onto THA containing 0.5 µg/mL rifampicin to confirm absence of resistant mutants.
102 Although comparatively large initial inocula (1-2 x 10^5 CFU per culture, confirmed by Miles and Misra CFU count [59] of OD-
103 adjusted cultures) were used, this minimised variance in final CFU number (N_t) determined in preliminary experiments.
104 Invariability in N_t allowed statistical comparison of mutation rates by the Maximum Likelihood Ratio test [60]. To infer the N_t
105 in final cultures and monitor their growth by optical density, replicate plates containing two cultures per strain were

106 prepared and incubated in a BMG FLUOstar OPTIMA microplate reader. When these representative cultures entered early
107 stationary phase, CFU counts were performed by Miles and Misra method [59] and invariably estimated as 1-2 x 10^8 CFU in
108 each well. Immediately after, entire 200 µL cultures (above) were plated on THA containing 0.5 µg/mL rifampicin for
109 selection of mutants, dried under laminar flow, and incubated until rifampicin resistant colonies appeared. The assay was
110 repeated four times for each strain in blocks of eight cultures using 20 isolates haphazardly chosen from different
111 phylogenetic lineages in each measurement. In some cases, e.g. a single strain representing an independent phylogenetic
112 lineage or with significantly differences in mutation rates detected in closely related strains, assays were repeated 1-2 more
113 times to increase statistical power. The rifampicin-resistant mutant counts were pooled into single data sets representing
114 32 to 48 cultures per strain (Suppl. Table 2).

115 **DNA extraction, preparation and sequencing**

116 Genomic DNA was extracted from cells collected from 10 mL late-exponential phase culture in Todd-Hewitt broth with the
117 DNeasy Blood & Tissue kit (Qiagen) using a modified protocol with an additional lysis step as described previously [61].
118 Sequencing was performed on the Illumina HiSeq2000 platform from Nextera XT pair-end libraries at Australian Genome
119 Research Facility, Melbourne. A reference genome from strain QMA0248 was constructed using both long reads derived
120 from a single Smrt Cell using the PacBio RS II system with P4C2 chemistry and short reads from Illumina HiSeq2000 derived
121 from Nextera XT paired-end libraries as reported elsewhere (NCBI accession no: GCA_002220115.1). All sequence data are
122 deposited at NCBI under Bioproject number PRJNA417543, SRA accession SRP145425. Sample numbers, accession numbers
123 and extended metadata are provided in Suppl. Table 3. Assembly statistics are provided in Suppl. Table 4.

124 **Genome assembly, recombination detection and phylogenetic analysis**

125 Phylogeny was constructed based on core genome SNPs from *de novo* genome assemblies filtered to remove
126 recombination breakpoints. Paired-end reads from Illumina were trimmed with Nesoni clip tool version 0.132
127 (http://www.vicbioinformatics.com/software.nesoni.shtml), with minimum read length 50, and the first 15 bp of each read
128 removed as quality deterioration in this region was observed when assessed with FASTQC version 0.11.5. Assembly was
129 performed using the SPAdes assembler version 3.7.1 [62], with minimum read coverage cutoff set to 10. Quality of assemblies
130 was assessed with QUAST 3.2 [63]. Contigs were ordered by alignment to QMA248 reference genome (CP022392.1) with
131 Mauve Contig Mover 2.4.0 [64]. Genome annotation was performed using Prokka 1.11 [65]. Rapid alignment of core genomes
132 was carried out using parsnp in the Harvest Tools suite version 1.2 [66], and the resulting alignment provided as an input to
133 Gubbins 1.4.7 [67] for detection and exclusion of variants produced by recombination. Phylogenies were then inferred from
134 post-filtered core genome polymorphic sites by maximum likelihood using RAxML 8.2.8 [68] with the general time reversible
135 nucleotide substitution model GTRGAMMA and bootstrap support from 1000 iterations. Effect of ascertainment bias
136 associated with using only polymorphic sites on branch length during ML inference was corrected using Felsenstein's
137 correction implemented in RAxML 8 [69]. The resulting phylogenetic tree was visualized using Dendroscope v 3.5.7 [70] with
138 bootstrap node support value cut-off 75. For the phylogram figure, tip labels were hidden for clarity and the edge containing
139 QMA0141 was re-scaled as dotted line representing 100-fold decrease in length (Fig. 1). A cladogram based on the inferred
140 phylogeny, showing all tip labels and bootstrap support for each node, was annotated with metadata using Evolview V2
141 (Fig. 2)[71]. To determine whether there was a strong temporal signal in the phylogenetic data, a root-to-tip regression of

142 branch length against time since isolation was performed in TempEst [72], using genetic distances corrected for ascertainment
143 bias in an unrooted tree estimated by maximum likelihood from the alignment of non-recombinant core-genome SNPs in
144 RAxML 8.2.8.

145 **Variation in MMR/OG genes and their regulatory regions, SNP analysis**

146 Large variants in MMR (*dnaN,* mut*S, muTL, recD2, rnhC*) and OG genes (*mutY* , *mutM*, *mutX*) and surrounding regions were
147 identified within multiple alignments of *de novo* genome assemblies using progressiveMauve [73]. A 56 bp deletion found in
148 the *mutY* promoter of clade B was confirmed by PCR using the CAGAAGGAAGAAACAGAC_F/ACCTCTATTGTAGCAAAG_R
149 primers set. SNP analysis was performed by read-mapping with Geneious version 9.1 [74], using default settings unless
150 otherwise specified. Paired-end reads from each genome were trimmed, merged into a single file with expected distances
151 between the reads set to 250 bp, and mapped to the curated reference genome of strain QMA0248. Mapped reads were
152 used to detect SNPs with minimum coverage set to 10 and frequency to 0.9. A consensus "pseudogenome" was generated
153 for every strain based on the reference sequence incorporating any detected variants [75]. Multiple alignment of these
154 pseudogenomes was carried out with *Geneious* aligning tool to identify SNPs in MMR/OG genes, their ribosome-binding
155 sites, and promoter regions. Promoter regions of repair genes were determined with BPROM [76], protein functional domains
156 by SMART genomic [77], and effect of amino acid substitutions on protein function by PROVEAN Protein with a sensitivity cut-
157 off of 1.3 [78]. To estimate the mutational bias among phylogenetic lineages, SNP types and position per strain were exported
158 as a table and SNPs of each type occurring only in each clade but not in any other clade were counted and plotted using a
159 custom script in R (Suppl. Fig. 1).

160 **Detecting possible association between variation in DNA repair genes and mobilome**

161 A pangenome for the 80 *S. iniae* isolates was constructed with GView server [79] using the QMA0248 reference genbank
162 (.gbk) file as seed by sequentially adding 79 genome assemblies by clade in the order derived from the cladogram (Fig.
163 2). Phage positions in the resulting pangenome were then determined by BLAST using Phaster [80] while IS positions were
164 identified using ISFinder [81] with a BLAST e-value cutoff of 1×10^{-30}. Positions of MMR and OG genes in the pangenome
165 were determined via manual search of the .gbk file. An image was created in BRIG [82] using the pangenome as reference
166 then aligning the 80 genome assemblies by BLAST with an e-value cutoff of 1×10^{-30}. The resulting image was annotated
167 in BRIG from a spreadsheet derived from the positions determined above, and coloured according to clade (Suppl. Fig.
168 2).

169 **Phenotypic variation related to virulence and antigenicity**

170 A multitude of intracellular and secreted enzymes contribute to virulence in the *Streptococcus* genus [83,84]. Considering the
171 large number of strains, we limited analysis to major, easily observable, and distinct phenotypes strongly associated with
172 virulence in streptococci (capsule, hemolysis, cell chains) and bacteria in general (oxidation resistance, biofilms). All assays
173 were performed at least in triplicate per strain.

174 **Buoyant density assay for presence of polysaccharide capsule**

175 Presence/absence of polysaccharide capsule was estimated by Percoll buoyant density assay. Isotonic stock Percoll (ISP)
176 was prepared by mixing nine parts of Percoll with one part of 1.5 M NaCl. Then, 6 parts of ISP was diluted with 4 parts of
177 0.15 M NaCl to make final 50 % Percoll solution, which was distributed by 3 mL into flow cytometry tubes. THB cultures (10
178 mL) grown to late-exponential phase were adjusted at OD_{600} 1 (10^8 CFU/mL), centrifuged at 3220 x *g* for 5 min, resuspended
179 in 0.5 mL of 0.15 M NaCl, and layered onto the Percoll. Tubes were centrifuged at 4°C in a swinging bucket rotor at 4000 x
180 *g* for 3 h with low acceleration and no brake. In this assay, encapsulated cells form a clear compact band in the Percoll
181 gradient (Fig. 3, A1), non-encapsulated cells form a pellet in the bottom of the tube (Figure 3, A2) and, occasionally, strains
182 show differential expression of capsular polysaccharide evidenced by band and a pellet (Fig. 3, A3).

183 **Haemolytic activity assay on sheep blood agar**

184 Rapid high-throughput detection of impaired haemolytic activity was achieved by blood-agar clearance zone assay. Briefly,
185 5 mm wells were made in Columbia agar supplemented with 5% defibrinated sheep blood (Oxoid, Australia). Bacterial
186 cultures grown to late-exponential phase (~10^8 CFU) were diluted 1:1000 and 50 µL of diluted cultures were pipetted into
187 punctures in the agar (initial inoculum ~$5x10^4$ CFU in each puncture). A 3 mm wide clearance zone was generally produced
188 by bacterial lysis of sheep erythrocytes during 24 h incubation (Fig. 3, B1). Where haemolysis was absent or fragmentary,
189 impeded haemolytic activity was recorded (Fig. 3, B2).

190 **Chain formation microscopy**

191 To assess chain formation, *S. iniae* cultures in THB were grown stationarily in 96-well plates at 28 °C for 24 h, mixed and 5
192 µL wet mounts prepared on a glass microscope slide. Slides were observed by bright field microscopy under 40x objective
193 with an Olympus BX40 microscope and captured using an Olympus DP28 digital camera using CellSens software (Olympus
194 Optical Co, Japan). Specimens were observed for at least 3 min and 2-20 cell chains were generally present (Fig. 3, C1).
195 Where over 20 cells in a chain were repeatedly detected (Fig. 3, C2) increased chain formation was recorded, and when
196 more than 10 cells in a chain were not detected (and the culture was mainly composed of detached cells) (Fig. 3, C3)
197 impeded chain formation was recorded. For the figure, 50 µL of cultures were dried onto slides at RT, fixed with methanol,
198 Gram stained, and images captured under a 100x objective.

199 **Oxidation resistance assay**

200 Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) [85] of hydrogen peroxide were
201 measured to assess oxidative stress resistance among *S. iniae* strains. THB cultures grown to late exponential phase were
202 adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 1000-fold, and distributed by 100 µL (~10^4 CFU) into wells of a U-bottom 96-
203 well plate (Greiner). THB (100 µL) with 2-fold excess concentration of hydrogen peroxide was added to the wells and serially
204 diluted twofold, resulting in a range from 0 to 10 mM final peroxide concentrations. Plates were incubated stationarily for
205 24h and examined for presence of observable growth to determine the MIC. Viable cell counts of cultures without visible
206 growth were performed to determine MBC. MIC of hydrogen peroxide was estimated as 3 mM for all strains, and MBC as
207 4 mM for the majority of strains. MBC elevated to 5 mM in dolphin and human isolates from USA was classified deviation
208 from a wildtype phenotype.

209 **Biofilm formation and visualisation assay**

210 *S. iniae* biofilms were grown for 4 days in 8-well Lab-Tek® II Chamber Slide™ Systems. To prepare an initial inoculum of 5 x
211 10^5 CFU, THB cultures grown to late exponential phase were adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 100-fold, and 0.5
212 mL of diluted cultures were placed into Chamber slide wells. After 24 h incubation, THB was removed and replaced with
213 fresh THB every 10-14 h for 3 days. For visualisation, biofilms were washed in PBS, stained for 15 min with 1 µM fluorescent
214 *Bac*Light Red bacterial stain, washed in PBS, fixed for 30 min with 10% formalin, and washed twice PBS. Z-stacks were
215 collected by ZEISS LSM 710 Inverted Laser Scanning Confocal Microscope at 20x objective and visualised using ZEN2012.
216 Quantification of biofilms was performed using COMSTAT [86]. Structures of biomass larger than 3.5 and average thickness
217 over 5 were classified as increased biofilm forming activity (Fig. 3 D).

218 **Identifying wildtype virulence-related phenotypes**

219 We consider prevalent phenotypes as a wildtype, namely: presence of polysaccharide capsule (71.25 % of strains; Figure 3,
220 A1) and haemolytic activity (85% of strains; Figure 3, B1), up to 20 cells in a chain (77.5 % of strains; Figure 3, A1), 3 mM
221 MIC and 4 mM MBC of hydrogen peroxide (91.25 % of strains), and 8 -12 µm thick biofilms covering up to 30% of
222 representative image (86.25 % of strains, Fig. 4A). Other phenotypes are regarded as deviant and discussed.

223 **Statistical analysis**

224 Analysis of co-variance between number of MMR and OG variants, mutation rates, number of atypical phenotype
225 associated with virulence, and atypical places of isolation were tested with a Phylogenetic Generalised Least Squares
226 model that accounts for relatedness among strains under Pagel's λ of 0.6 implemented in R [87]. Estimation of mutation
227 rates was carried out using the Ma-Sandri-Sarkar Maximum Likelihood Estimation (MSS-MLE) method as implemented
228 in FALCOR fluctuation analysis calculator [50]. Differences in mutation rates were compared by Likelihood Ratio Test (LRT)
229 using the rSalvador R package [60,88]. LRT estimates the overlap between confidence intervals calculated for two
230 fluctuation experiment data sets and allows to compare fluctuation assay data with different number of cultures and
231 resolve very small differences in mutation rate [60,88]. The *Compare.LD* function was applied pair-wise to all strains, and
232 obtained *p* values were corrected for multiple comparisons by controlling the False Discovery Rate [89] using p.adjust in
233 R. Adjusted *p* values less than 0.05 were considered to indicate significant difference in mutation rates.

234

235 **3. Results**

236 Maximum likelihood phylogenetic analysis of 80 *S. iniae* isolates (Table 1) based on 10,267 non-recombinant core genome
237 SNPs derived from whole genome data resolved ten major phylogenetic lineages: six lineages with multiple strains (clades
238 A-F), one lineage with two strains (clade G), and three lineages with a single strain (Figs. 1, 2). Some lineages (C-E) show a
239 degree of geographic endemism while clades (A, B, and F) contain strains from diverse geographic regions. Principally,
240 variation in MMR and GO genes occurs only between the lineages, and they are highly conserved within clades (Fig. 2, Table
241 2). Most SNPs in MMR and OG genes were non-synonymous and present in protein functional domains, and most amino
242 acid substitutions were predicted to have a deleterious effect on protein function (Table 2). Furthermore, the number of
243 MMR and OG variants correlates significantly with mutation rate (p = 0.0056), as predicted by a Phylogenetic Generalised
244 Least Squares (PGLS) regression model that accounts for autocorrelation between the closely related strains [87]. In turn,
245 mutation rate variation is highly consistent with phylogenetic affiliation: according to multiple pair-wise strain comparisons
246 by Maximum Likelihood Ratio test [60], with probabilities corrected for multiple comparison using False Discovery Rate [89], in
247 general mutation rate difference is insignificant within but significant between major phylogenetic lineages (Suppl. Table
248 1). Although mutation rate variation between divergent strains is significant, the magnitude of the differences is modest
249 (around 5-fold) (Figure 1, 2, Suppl. Table 2).

250 The accrual of random mutations in MMR/OG loci appears to be the primary mechanism of mutation rate molecular
251 evolution within analysed set of *S. iniae* genomes. Firstly, we used a maximum likelihood inference method to exclude areas
252 of recombination from our phylogenetic analyses [67], and none of the MMR and OG gene loci were excluded as recombinant.
253 Within the core genome, where MMR and OG gene loci are located, recombination rates were extremely low with mean
254 $\mu_{r/m}$ 0.032 and only 5 from the 80 strains containing recombination blocks (Suppl. Table 5). Secondly, we created a
255 pangenome for the 80 strains in our analysis and mapped on putative phage and prophage signatures [80], and insertion
256 sequence elements [81]. None of the MMR or OG genes occurred in proximity to phage or insertion sequence elements (Suppl.
257 Fig. 2). Also, tandem-repeat copy-number variants identified in *mutL* in *E. coli* [44,90] were not detected in MMR or OG genes
258 in *S. iniae* (Table 2). All mutations in MMR/OG loci in our dataset were SNPs, except for the large deletion affecting the *mutY*
259 promoter in clade B (Table 2). As most *S. iniae* lineages contain variants pertaining to both MMR and OG systems (Table 2),
260 the detection of mutational bias produced by either dysfunctional MMR or OG demonstrated in experimental populations
261 of *E.coli* [91] is limited for our natural isolates. Nonetheless, consistent with findings in *E. coli* where OG mutants showed
262 increased frequency of AT to CG transversions [91], the latter bias was detected in the only *S. iniae* lineage (clade D) where
263 mutations are constrained to a single OG variant (Suppl. Fig. 1). This variant is an amino acid substitution in the DNA
264 glycolyase domain of mutM that is predicted to deleteriously affect the protein function (PROVEAN score of -2.990, Table
265 2).

266 To investigate whether variation in mutation rate phenotype and MMR/OG genotype is associated with adaptation to the
267 host, we identified variation in key virulence traits in streptococci (capsular polysaccharide, hemolysin, length of cell chains)
268 and bacteria in general (resistance to ROS, and biofilm formation). We found that number of atypical phenotypes for these
269 traits correlates significantly (p = 0.0331, PGLS) with the number of MMR and OG variants (Fig. 2, Table 3). Moreover,
270 infections in atypical hosts (mammals) and tissue within apparently immune primary hosts (bone of vaccinated barramundi)

271 correlates strongly with both number of atypical virulence-associated phenotypes ($p < 0.0001$) and the number of MMR/OG
272 variants ($p = 0.002$).

273 Of the clades identified among *S. iniae*, clade A is apparently a dominant circulating clade primarily infecting Perciform fish
274 that has persisted globally for almost two decades (clade A contains strains isolated from USA, Honduras, and Australia
275 between 1999 to 2016). Mutation rates in all isolates (*n*=35) among this ancestral lineage fall within a 1.5 - 2 x 10^{-8} range
276 (Figs. 1, 2; Clade A), which is similar to the base mutation rate estimated for other non-mutator gram positive bacteria [11,92].
277 All differences in mutation rates within this clade are insignificant except three pairwise comparisons that verge on $p < 0.05$
278 (Suppl. Table 1). MMR and OG genes in these strains are identical, and only minor deviations from wildtype traits
279 contributing to virulence were observed (absence of capsule production and longer cell chains in QMA0158, 216) (Figs. 1-
280 3, Table 1). All other lineages (with the exceptions of clade C1 and strains QMA0445-46 discussed below) have significantly
281 different mutation rates to clade A, contain unique and often multiple SNPs in MMR/OG loci, and show peculiar phenotypic
282 variants related to virulence (Figs. 1-3, Tables 1, 2, Suppl. Table 1). Although phylogenetic relatedness between lineages is
283 fully resolved (Fig. 2), they originate almost simultaneously, and nested lineages evolve independently sharing little
284 phylogenetic history (Figs. 1, 2). Lack of correlation between branch length and sampling date derived by root-to-tip
285 regression analysis with time is evidence of non-neutral evolution amongst this collection of *S. iniae* isolates (R^2=0.0936,
286 Correlation coefficient 0.306, best-fit root; Suppl. Fig. 3, [72]). Deviation from a neutral model is supported by highly skewed
287 branching, with some branches comprising single isolates whilst others comprise many epidemiologically unrelated isolates
288 (Figure 1, 2). Although potentially affected by the sampling bias, this is suggestive of frequent strong selection, presumably
289 imposed by heterogeneity of the immune landscape encountered during transfer between host individuals, and resembles
290 genealogical tree topologies arising from viral evolution over similar timespans [93,94].

291 Clade B contains isolates from ornamental fish, QMA0084 from flying fox, *Epalzeorhynchos kalopterus*, in Western Australia
292 (WA) and QMA0462-63 from clown loach *Chromobotia macracanthus*, in USA. These isolates share a valine to isoleucine
293 substitution in MutL predicted as neutral (PROVEAN score -0.271), a synonymous SNP in *recD2*, and a 56 bp deletion
294 affecting the predicted promoter of *mutY* (Table 2). The latter mutation encompasses the -35 box, two binding sites of the
295 RpoD17 general transcription factor, and the 1st bp of the -10 box, and was predicted to abolish the original promoter
296 (Suppl. Fig. 4). However, *mutY* might still be transcribed at some level since two other potential promoters were identified
297 in a nearby sequence (Suppl. Fig. 3). It appears that altered *mutY* expression is also affected by the rest of the genetic
298 background as isolates exhibit significant differences (Suppl. Table 1) in mutation rate phenotype, 7 x 10^{-8} in QMA0084 and
299 4 x 10^{-8} in clown loach strains (Figs. 1-2). Phenotypically, clade B has shifted towards decreased haemolytic activity (Fig. 3,
300 B2).

301 Clade C shares a common ancestor with clade B and is comprised of strains obtained from barramundi farmed in WA,
302 Northern Territory (NT) and north QLD fish farms (nested subclades C1 and C2 respectively). While a core mutation rate of
303 1.5-2 x 10^{-8} is observed in subclade C1, a significantly higher mutation rate of 7 x 10^{-8} has been maintained in subclade C2
304 from north QLD for almost two decades (1995 to 2012) (Figs. 1,2 Suppl. Table 1). Minor variation in virulence-related
305 phenotypes (absence of capsule production and longer cell chains in QMA0074, 77) (Fig. 2, Table 1) suggests that evolution
306 in this lineage might be driven by outside-the-host factors. The significant difference in mutation rate between nested
307 clades C1 and C2 may result from variation in *dnaN*. In *Bacillus subtilis,* 90% of mismatch repair is dependent on targeting

308 MutS to nascent DNA via the ß sliding clamp (DnaN) zone [95,96]. Inactivation of *dnaN1* in *B. anthracis* results in a mutator
309 phenotype with a mutation rate equivalent to a *mutS* mismatch repair-defective strain [97]. The tyrosine to isoleucine
310 substitution in one of the critical residues of the loader binding interface of the ß-clamp, predicted as deleterious (PROVEAN
311 score -2,216), is present in all isolates of from clade C (Table 2), which may explain the high mutation rate in clade C2 strains
312 (Fig. 2, Table 2). Potentially, rapid fluctuation in water salinity, oxygen and temperature accompanying periodic heavy
313 rainfall in tropical north QLD represent an unstable environment that favours persistence of mutator alleles [25]. In contrast,
314 only subclade C1 harbours T1135C nucleotide substitution in the *dnaN* that changes stop codon TAA into CAA coding for
315 glutamine predicted as neutral (PROVEAN score -0.005), with the next stop codon TAG found immediately downstream.
316 Putatively, T1135C and/or change of the *dnaN* stop codon [98] may act as compensatory to the T326I substitution, reducing
317 the mutation rate to a value not significantly different from the core rate found in clade A (Fig. 1, 2) via protein elongation
318 or by affecting translation via stop codon usage bias [98]. In addition, a synonymous variant in *recD2* is found in subclade C2
319 only (Table 2).

320 Clade D consists of strains from trout (*Oncorhynchus mykiss*) isolated in Réunion and Israel. These strains have an estimated
321 mutation rate of 5 x 10^{-8}, contain a glutamate to aspartate substitution in *mutM* predicted to be deleterious to protein
322 function, a synonymous SNP in *RecD2*, and exhibit impeded haemolytic activity (Fig 1,2, Table 2). Also, isolates from Israel
323 form thicker and more dense biofilm structures (Fig. 3, D2).

324 Isolates from humans and fish are found in clade E where multiple phenotypic profiles for the virulence-associated traits
325 are observed (Figs. 2-3, Table 1). This clade contains two nested subclades that share a SNP in *mutS,* but have other unique
326 MMR/OG SNPs (Fig. 2, Table 2) and exhibit mutation rates that are significantly different in most pairwise comparisons
327 (Suppl. Table 1). Subclade E1 contains USA isolates from humans (QMA0133-35, 37-38) and hybrid striped bass (QMA0447-
328 48), which have a substitution in the Shine-Dalgarno sequence of *recD2* and a mutation rate of 6.5 x 10^{-8} (Fig 1, Table 2).
329 Fish strains and three human strains (QMA0135, 37-38) form shorter cell chains and are less sensitive to H_2O_2. Two human
330 strains QMA0133-34 produce thicker and more dense biofilm structures (Fig. 3, D2). QMA0133 is non-encapsulated (Fig 3,
331 A2), and QMA0134 appears to differentially express the capsular polysaccharide in culture (Fig 3, A3). Subclade E2 contains
332 two human strains from Canada (QMA0130-31) and a tilapia strain from USA (QMA0466). These strains have unique
333 methionine to isoleucine substitution at the N-terminus of MutX, and a mutation rate of 4.5 x 10^{-8} (Fig 1, Table 2). Both
334 human strains are non-encapsulated (Fig. 2). In contrast, the tilapia strain expresses the capsule but forms short cell chains
335 in common with most strains from the subclade E1 (Fig 2, Fig 3, C3).

336 Clade F is not readily definable by location, time of isolation, or host species and comprises QMA0139 from unidentified
337 fish in Canada, QMA0190 isolated from snakehead murrel in Thailand (both of which appear on long branches), and nested
338 terminal clade containing strains QMA0233-36, 249, 254-55 from barramundi farmed in Recirculating Aquaculture Systems
339 (RAS) in New South Wales and South Australia (Fig. 2). The latter were isolated from barramundi bone lesions during a
340 disease outbreak in vaccinated fish where infection manifested itself as a slowly progressing osteomyelitis [49] in contrast to
341 typical acute septicaemia and meningitis [48]. Multiple phenotypic changes associated with chronic infection and increased
342 potential for colonization were present in bone isolates, including absence of capsule, impeded hemolytic activity,
343 increased cell-chain length, and denser biofilms [53-56,84,99,100] (Fig. 3). Bone strains have a unique MMR/OG genotype (Fig. 2,
344 Table 2) and a mutation rate that is different from QMA0139 and QMA0190 (Figs. 1, 2, Table 1). A serine to arginine

345 substitution in MutL and tyrosine to isoleucine substitution in MutY seems to have occurred in the ancestor of QMA0139,
346 QMA0190, and bone strains. Both substitutions are predicted to have a deleterious effect on protein function (PROVEAN
347 scores -2.050 and -2.544 respectively), but their effect on mutation rate is combined with unique SNPs found in each
348 branch: QMA0139 has a SNP in the binding site for the DnaA transcription factor within the *dnaN* promoter, and a mutation
349 rate of 4.5 x 10^{-8}, QMA0190 has synonymous SNP in *mutL* and a mutation rate of 6.5 x 10^{-8}, and bone strains have glutamine
350 to glycine substitution in MutS predicted as neutral (PROVEAN score -0.635), a synonymous SNP in *mutL*, and a mutation
351 rate of 3-4 x 10^{-8}. Considering that the mutation rate of QMA0190 is significantly higher compared to rates expressed by
352 QMA0139 and bone isolates (Suppl. Table 1), it is likely that variants unique to each of the latter strains compensate for
353 deleterious variants shared by the isolates [46].

354 Clade G contains two strains from tilapia isolated in USA (QMA0445-46). Notably, despite being phylogenetically distant
355 from clade A, these strains have the same MMR and OG genotype and have retained the core mutation rate and the typical
356 phenotypic profile for virulence-associated traits (with the exception of capsule absence most likely inherited from common
357 ancestor shared with strains from clade F, QMA0140-41, 187 that are also non-encapsulated) (Fig. 2, Table 1).

358 Three isolates of *S. iniae* were identified as independent phylogenetic lineages, with unique mutation rate genotypes and
359 phenotypes, and virulence traits (Figs. 1-2, Tables 1-2). The first long branch with a single isolate contains a strain from
360 snakehead murrel strain (*Channa striata*) from Thailand. This isolate is non-encapsulated and weakly haemolytic and has a
361 mutation rate of 4.5 x 10^{-8}, potentially attributable to a deleterious aspartate to asparagine substitution in rnhC and SNP in
362 the binding site of for the *fnr* transcription factor within the *dnaN* promoter. A second long branch with a single strain
363 contains the oldest among the analysed strains, QMA0140 isolated in 1976 from dolphin (*Inia geoffrensis*) [101]. This strain
364 exhibits a mutation rate of 1 x 10^{-8}, which is unique among the strains and significantly lower than the core mutation rate,
365 perhaps linked to increased translation of recD2 helicase resulting from a SNP in the ribosome binding site (Fig 2, Table 2).
366 The longest branch on the tree contains a second dolphin strain, QMA0141, isolated two years later in 1978 [102]. This isolate
367 is highly divergent from the rest of strains with around 20 kB of non-recombinant SNPs in pair-wise comparisons with other
368 strains, accounting for around 1% genomic difference. In contrast to QMA140, it mutates at a rate of 1x 10^{-7}, the highest
369 mutation rate phenotype determined among the isolates and significantly different to the rest of the values. Multiple SNPs
370 are observed in all MMR and OG genes: 3 in *dnaN* and *mutX*, 4 in *mutM*, 9 in *mutL* and *rnhC*, 15 in *recD2*, 9 in *mutL* and
371 *rnhC*, and 68 in *mutS*. Both dolphin isolates are non-encapsulated, show increased ability to withstand oxidative stress, and
372 form denser and thicker biofilms (Fig 2, 3).

373

374

375 **4. Discussion**

376 Mutator alleles from MMR and OG systems are abundant among infectious isolates and are likely to promote pathogen
377 evolution since they can greatly increase the supply of pathoadaptive mutations in changing host environments [11,30,32,37,38].
378 The dynamics of mutation rate modifiers – increase in mutator alleles in times of adaptation and their decline in adapted
379 communities – is not fully understood and continues to be the focus of multiple experimental studies [24]. Investigation of

380 such mutation rate dynamics in natural populations is a challenging task considering mutator alleles often revert to wildtype
381 level via back-mutation or recombination/translocation that fully restores the wiltype mutation rate genotype and
382 phenotype [44,45]. However, when compensatory evolution of mutation rates rather than reversion occurs, i.e. when anti-
383 mutator variants with secondary-site mutations in MMR/OG genes emerge that lower mutation rates – this is likely to
384 produce traceable shifts in DNA repair genotype and phenotype of an adapted strain [46]. Here we have investigated the
385 mutation rate dynamics in natural populations of the rapidly evolving aquatic pathogen *S. iniae* [48], using 80 strains isolates
386 globally over 40 years. We performed whole-genome sequencing, constructed phylogenetic tree based on non-
387 recombinant core genome SNPs [103], identified variation in major MMR and OG loci, estimated mutation rates by fluctuation
388 analysis [58], and compared the latter by the Likelihood Ratio Test [51]. As hypothesized, we have found that MMR/OG genotype
389 and mutation rate variation occurs between major phylogenetic lineages and is minimal within phylogenetic clades. While
390 each lineage has unique MMR/OG profile, MMR and OG loci are conserved even within phylogenetic groups not readily
391 definable by geographic origin, time, or host species. Therefore, these loci can be used for rapid diagnostics via Multilocus
392 Sequence Typing (MLST) [104]. In turn, mutation rates are conserved within phylogenetic clades, while significant differences
393 occur among major phylogenetic lineages. Although significant, differences in mutation rate are largely small (around 5-
394 fold), which is consistent with transience of mutators in natural populations and restoration of low mutation rate after
395 adaptation is gained [46]. On the other hand, one order of magnitude or lower differences in mutation rate distinguish
396 mutator from non-mutator phenotypes in other streptococcal species [11,92].

397 Arguably, the accrual of local mutations in MMR/OG loci is a primary mechanism of molecular evolution driving mutation
398 rate dynamics in *S. iniae*. Firstly, all variants in MMR and OG genes were single nucleotide substitutions with the exception
399 of the large deletion affecting the *mutY* promoter in clade B. Secondly, we did not find evidence of other mechanisms that
400 underlie the flux of mutator alleles and phenotypes including recombination [47], and prophage integration/excision [45], and
401 tandem repeat copy-number variation [44,90]. Indeed, the easily reversible copy-number variation prone to back mutation and
402 full restoration of the wiltype non-mutator genotype and phenotype, to date has been only found in *mutL* in *Proteobacteria*
403 [44,90], which is structurally and functionally different from its homologs in gram-positive bacteria and eukaryotes [19,52]. Notably,
404 a single SNP in MMR or OG loci can be sufficient to induce profound changes in mutation rate [105,106], and most of the SNPs
405 identified in our isolates were found within coding regions, non-synonymous, and predicted to negatively affect the protein
406 function. The dysfunctional MMR and OG can leave molecular signatures within the genome in the form of mutational bias:
407 as observed in a long term evolutionary experiment in *E.coli* defective MMR and OG repair increased frequency of AT to GC
408 transitions and AT to CG transversions respectively [91]. Such analysis is limited for our dataset since most of our natural *S.*
409 *iniae* isolates contain both MMR and OG variants. Nonetheless, we quantified the mutational spectra by clade, and,
410 consistent with findings in *E. coli* experimental mutators, the highest frequency of AT to CG transversions was observed in
411 the only lineage containing a single OG variant (clade D with the amino acid substitution in MutM in predicted as
412 deleterious). Thus, it appears that mutation rate dynamics in *S. iniae is* primarily driven by local mutations and compensatory
413 mutator/anti-mutator evolution where some variants increase and some decrease the mutation rate [46].

414 The conservation of MMR/OG genotypes within *S. iniae* clades and their variation among the major phylogenetic lineages,
415 linkage of unique MMR/OG profiles to unique mutation rate values, and the likelihood of most identified SNPs to affect
416 protein function or translation rate are consistent with an idea that mutator phenotypes may be facilitating adaptations

417 preceding major diversification events in *S. iniae* evolution. The tree topology with highly skewed branching deviates from
418 a neutral model and resembles a genealogical tree topology derived from viral evolution over similar timespans, suggesting
419 evolution under strong selection [93,94]. Since selection by the host immune responses is the major force shaping strain
420 diversity in pathogens [6], we phenotyped our strains for major traits determining virulence in streptococci - capsular
421 polysaccharide, hemolysin, length of cell chains, resistance to reactive oxygen species, and biofilm formation[53-57]. As
422 estimated by a Phylogenetic Least Squares (PGLS) model accounting for autocorrelation among phylogenetically close
423 strains [87], the occurrence of non-typical phenotypes for these traits correlates with number of MMR/OG variants. The
424 variation in virulence phenotypes is minimal and relatively low mutation rate (core rate of 1.5-2 x 10^{-8}) is observed in
425 apparently dominant globally distributed clade A and also clade G with MMR/OG profile and mutation rate identical to clade
426 A. Other lineages contain MMR/OG variants, express 3-5 times higher mutation rates (except mutator strain from dolphin
427 with 10 times higher mutation rate), and contain isolates with non-typical virulence-associated variants. This is consistent
428 with an idea of selection for lower mutation rates in adapted populations [46]. One the other hand, in contrast to other
429 common pathogens where 2-3 orders of magnitude differences in mutation rate are observed between mutators and non-
430 mutators [11], only one order of magnitude or lower differences appear to distinguish mutator from non-mutators in
431 streptococci (around 10^{-7} and 10^{-8} mutation rates respectively) [11,92] .

432 Furthermore, the most variants in MMR/OG SNPs and changes in virulence phenome are observed in strains from atypical
433 immune settings - atypical hosts (mammals) and atypical tissue within apparently immune vaccinated fish hosts
434 (barramundi bone). Firstly, this implies that mutators may promote pathogen transmission to novel hosts ('host jumps'),
435 including transmission between the higher host taxa such as from fish to mammals discussed in the present study [107]. The
436 adaptation to an ultimately different immune responses of a novel host species is likely to require multiple genetic and
437 phenotypic variants, whose occurrence on the same genetic background is most readily achievable by the mutator strains.
438 Therefore, it is conceivable that transfers from wild fish to dolphins (strains QMA0140-41) and farmed hybrid bass to
439 humans [108] in lineage E have been facilitated by the mutator phenotype. Second, mutators might present a risk factor to
440 immune escape after vaccination and vaccine-induced serotype replacement (VIRS) – the spread of co-existing pathogen
441 serotypes after elimination of the dominant serotypes vaccinated against [109]. An atypical pathology outbreak caused by
442 strains from clade F has occurred on RAS in New South Wales and South Australia where fish were immunised using whole-
443 cell killed bacterins prepared from strains found in dominant clade A [49]. Notably, vaccine-induced immunity was partially
444 effective in the infected fish, evident by cross-reactivity of vaccine-induced antibodies against clade F strains and absence
445 of their proliferation in typical niduses (blood, brain, and pronephros) [49]. Nonetheless, colonization of the barramundi bone
446 tissue by clade F strains caused slowly progressing osteomyelitis instead of typical acute septicaemia and meningitis [48,49].
447 These bone isolates express multiple non-typical phenotypes - absence of capsule, impeded hemolytic activity, increased
448 cell-chain length, and denser biofilms. All these shifts are likely to be adaptive as they were previously associated with
449 chronic infection and increased potential for colonization [53-56,84,99,100]. Multiplicity of virulence-associate phenotypic shifts
450 in bone strains indicates that evasion from the immune clearance and adaptation to osseous tissue are likely to have
451 occurred via the mutator phenotype. Indeed, bone isolates have unique mutation rates and multiple MMR and OG SNPs
452 where some variants appear to have mutator and some anti-mutator effect. Thus, mutator alleles may potentially promote
453 evolution of novel serotypes and VISR in populations of farmed fish. Notably, as evidenced by bone isolates, VIRS might
454 occur without major disruption of immune recognition when multiple changes permit colonisation of and persistence in

455 the tissue with lower immune surveillance. Extension to other species such as *Streptococcus pneumoniae* may be
456 particularly interesting as serotype evolution following vaccination programmes is well documented, but the role of
457 mutators in this process is unknown [110-112]. Since mutators may promote emergence of novel serotypes, development of
458 long-term efficient generic polyvalent vaccine protective against all co-existing strains may not be readily achievable with
459 rapidly evolving microorganisms. Thus, a flexible approach to veterinary vaccine licencing and re-formulation may be
460 recommended to locally eliminate the emergent virulent serotypes of antigenically variable pathogens before they spread
461 in the population.

462 In summary, our results broadly support the notion that mutators promote evolution of virulence and antigenicity, leading
463 to major diversification events in *S. iniae*. More specifically, they may facilitate adaptation to a new immunological setting,
464 such as one required for exploitation of a new host taxon, new host tissue/s, or a host with vaccine-induced adaptive
465 immune response. Therefore, dynamics of mutator alleles in natural populations of pathogens could be associated with
466 epidemiologically relevant incidents such as vaccine escape outbreaks and host jumps that should be explored further.

467

468 **Acknowledgements**

469 This work was supported by The Australian Research Council Discovery Project Grant DP120102755 awarded to ACB. JE
470 was supported by an ARC Future Fellowship (FT140100907). We thank the University of Queensland School of Biological
471 Sciences for funding a tuition fee waiver for OS. For provision of strains into our collection over 15 years we thank: Judy
472 Forbes-Faulkner, Oonoonba Veterinary Laboratory, QLD, Australia; Lynn Shewmaker, CDC, Atlanta, USA; Christian Michel,
473 INRA, France; Suresh Benedict, Berrimah Veterinary Laboratories, NT, Australia; Nicky Buller, DPIRD WA, WA, Australia;
474 Mark White, Tréidlia Biovet Pty (formerly Allied Biotechnology), NSW, Australia; Matt Landos, Future Fisheries Veterinary
475 Services, NSW, Australia; Spencer Russell, Novartis Animal Health, Canada; Ruben Arturo Lopez Crespo, Regal Springs
476 Tilapia, Honduras. We thank Dr Simone Blomberg for advice on statistical analyses and Mr Nazar Rudenko for assistance
477 with writing software pipelines and for use of his Unix server.

478 **References**

479 1 Diard, M. & Hardt, W. D. Evolution of bacterial virulence. *FEMS microbiology reviews* **41**, 679-697,
480 doi:10.1093/femsre/fux023 (2017).
481 2 Frank, S. A. & Barbour, A. G. Within-host dynamics of antigenic variation. *Infection Genetics and*
482 *Evolution* **6**, 141-146, doi:10.1016/j.meegid.2004.10.005 (2006).
483 3 Deitsch, K. W., Moxon, E. R. & Wellems, T. E. Shared themes of antigenic variation and virulence in
484 bacterial, protozoal, and fungal infections. *Microbiology and molecular biology reviews : MMBR* **61**,
485 281-293 (1997).
486 4 Palmer, G. H. & Brayton, K. A. Antigenic variation and transmission fitness as drivers of bacterial
487 strain structure. *Cell Microbiol* **15**, 1969-1975, doi:10.1111/cmi.12182 (2013).
488 5 Barnett, T. C. *et al.* Host-pathogen interaction during bacterial vaccination. *Current opinion in*
489 *immunology* **36**, 1-7, doi:10.1016/j.coi.2015.04.002 (2015).
490 6 Cobey, S. Pathogen evolution and the immunological niche. *Annals of the New York Academy of*
491 *Sciences* **1320**, 1-15, doi:10.1111/nyas.12493 (2014).
492 7 Arber, W. Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS*
493 *microbiology reviews* **24**, 1-7, doi:10.1111/j.1574-6976.2000.tb00529.x (2000).
494 8 Moxon, R., Bayliss, C. & Hood, D. Bacterial contingency loci: the role of simple sequence DNA
495 repeats in bacterial adaptation. *Annual review of genetics* **40**, 307-333,
496 doi:10.1146/annurev.genet.40.110405.090442 (2006).
497 9 Sundin, G. W. & Weigand, M. R. The microbiology of mutability. *FEMS microbiology letters* **277**, 11-
498 20, doi:Doi 10.1111/J.1574-6968.2007.00901.X (2007).
499 10 Foster, P. L. Stress-induced mutagenesis in bacteria. *Critical reviews in biochemistry and molecular*
500 *biology* **42**, 373-397, doi:10.1080/10409230701648494 (2007).
501 11 Hall, L. M. & Henderson-Begg, S. K. Hypermutable bacteria isolated from humans--a critical analysis.
502 *Microbiology* **152**, 2505-2514, doi:10.1099/mic.0.29079-0 (2006).
503 12 Boe, L. *et al.* The frequency of mutators in populations of Escherichia coli. *Mutation research* **448**,
504 47-55 (2000).
505 13 Miller, J. H. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair.
506 *Annual review of microbiology* **50**, 625-643, doi:10.1146/annurev.micro.50.1.625 (1996).
507 14 Jiricny, J. The multifaceted mismatch-repair system. *Nature reviews. Molecular cell biology* **7**, 335-
508 346, doi:10.1038/nrm1907 (2006).
509 15 Harfe, B. D. & Jinks-Robertson, S. DNA mismatch repair and genetic instability. *Annual review of*
510 *genetics* **34**, 359-399, doi:10.1146/annurev.genet.34.1.359 (2000).
511 16 Surtees, J. A., Argueso, J. L. & Alani, E. Mismatch repair proteins: key regulators of genetic
512 recombination. *Cytogenet Genome Res* **107**, 146-159, doi:Doi 10.1159/000080593 (2004).
513 17 David, S. S., O'Shea, V. L. & Kundu, S. Base-excision repair of oxidative DNA damage. *Nature* **447**,
514 941-950, doi:10.1038/nature05978 (2007).
515 18 Lu, A. L., Li, X., Gu, Y., Wright, P. M. & Chang, D. Y. Repair of oxidative DNA damage: mechanisms
516 and functions. *Cell biochemistry and biophysics* **35**, 141-170, doi:10.1385/CBB:35:2:141 (2001).
517 19 Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *Journal of nucleic acids* **2010**,
518 doi:10.4061/2010/260512 (2010).
519 20 Ambur, O. H. *et al.* Genome dynamics in major bacterial pathogens. *FEMS microbiology reviews* **33**,
520 453-470, doi:10.1111/j.1574-6976.2009.00173.x (2009).
521 21 van der Veen, S. & Tang, C. M. The BER necessities: the repair of DNA damage in human-adapted
522 bacterial pathogens. *Nature reviews. Microbiology* **13**, 83-94, doi:10.1038/nrmicro3391 (2015).
523 22 Mao, E. F., Lane, L., Lee, J. & Miller, J. H. Proliferation of mutators in A cell population. *Journal of*
524 *bacteriology* **179**, 417-422 (1997).
525 23 Taddei, F. *et al.* Role of mutator alleles in adaptive evolution. *Nature* **387**, 700-702, doi:Doi
526 10.1038/42696 (1997).
527 24 Raynes, Y. & Sniegowski, P. D. Experimental evolution and the dynamics of genomic mutation rate
528 modifiers. *Heredity*, doi:10.1038/hdy.2014.49 (2014).

529 25 Lukacisinova, M., Novak, S. & Paixao, T. Stress-induced mutagenesis: Stress diversity facilitates the
530 persistence of mutator genes. *PLoS computational biology* **13**, e1005609,
531 doi:10.1371/journal.pcbi.1005609 (2017).
532 26 Oliver, A. Mutators in cystic fibrosis chronic lung infection: Prevalence, mechanisms, and
533 consequences for antimicrobial therapy. *International journal of medical microbiology : IJMM* **300**,
534 563-572, doi:10.1016/j.ijmm.2010.08.009 (2010).
535 27 Matsushima, A. *et al.* High prevalence of mutators among Enterobacter cloacae nosocomial isolates
536 and their association with antimicrobial resistance and repetitive detection. *Clinical microbiology*
537 *and infection : the official publication of the European Society of Clinical Microbiology and Infectious*
538 *Diseases* **16**, 1488-1493, doi:10.1111/j.1469-0691.2009.03145.x (2010).
539 28 Wang, S., Wang, Y., Shen, J., Wu, Y. & Wu, C. Polymorphic mutation frequencies in clinical isolates
540 of Staphylococcus aureus: the role of weak mutators in the development of fluoroquinolone
541 resistance. *FEMS microbiology letters* **341**, 13-17, doi:10.1111/1574-6968.12085 (2013).
542 29 Negri, M. C. *et al.* Very low cefotaxime concentrations select for hypermutable Streptococcus
543 pneumoniae populations. *Antimicrobial agents and chemotherapy* **46**, 528-530, doi:Doi
544 10.1128/Aac.46.2.528-530.2002 (2002).
545 30 Mena, A. *et al.* Genetic Adaptation of Pseudomonas aeruginosa to the Airways of Cystic Fibrosis
546 Patients Is Catalyzed by Hypermutation. *Journal of bacteriology* **190**, 7910-7917, doi:Doi
547 10.1128/Jb.01147-08 (2008).
548 31 Gutierrez, O., Juan, C., Perez, J. L. & Oliver, A. Lack of association between hypermutation and
549 antibiotic resistance development in Pseudomonas aeruginosa isolates from intensive care unit
550 patients. *Antimicrobial agents and chemotherapy* **48**, 3573-3575, doi:Doi 10.1128/Aac.48.9.3573-
551 3575.2004 (2004).
552 32 Mena, A. *et al.* Inactivation of the mismatch repair system in Pseudomonas aeruginosa attenuates
553 virulence but favors persistence of oropharyngeal colonization in cystic fibrosis mice. *Journal of*
554 *bacteriology* **189**, 3665-3668, doi:10.1128/JB.00120-07 (2007).
555 33 Canfield, G. S. *et al.* Evolution in fast forward: a potential role for mutators in accelerating
556 Staphylococcus aureus pathoadaptation. *Journal of bacteriology* **195**, 615-628,
557 doi:10.1128/JB.00733-12 (2013).
558 34 Giraud, A. *et al.* Costs and benefits of high mutation rates: Adaptive evolution of bacteria in the
559 mouse gut. *Science* **291**, 2606-2608, doi:Doi 10.1126/Science.1056421 (2001).
560 35 Nilsson, A. I., Kugelberg, E., Berg, O. G. & Andersson, D. I. Experimental adaptation of Salmonella
561 typhimurium to mice. *Genetics* **168**, 1119-1130, doi:10.1534/genetics.104.030304 (2004).
562 36 Healey, K. R. *et al.* Prevalent mutator genotype identified in fungal pathogen Candida glabrata
563 promotes multi-drug resistance. *Nature communications* **7**, 11128, doi:10.1038/ncomms11128
564 (2016).
565 37 Labat, F. *et al.* Mutator phenotype confers advantage in Escherichia coli chronic urinary tract
566 infection pathogenesis. *FEMS Immunol Med Microbiol* **44**, 317-321,
567 doi:10.1016/j.femsim.2005.01.003 (2005).
568 38 Oliver, A. & Mena, A. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance.
569 *Clin Microbiol Infec* **16**, 798-808, doi:10.1111/j.1469-0691.2010.03250.x (2010).
570 39 Wielgoss, S. *et al.* Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution
571 Experiment With Escherichia coli. *G3* **1**, 183-186, doi:10.1534/g3.111.000406 (2011).
572 40 Sprouffske, K., Aguilar-Rodriguez, J., Sniegowski, P. & Wagner, A. High mutation rates limit
573 evolutionary adaptation in Escherichia coli. *Plos Genet* **14**, e1007324,
574 doi:10.1371/journal.pgen.1007324 (2018).
575 41 Sniegowski, P. D., Gerrish, P. J., Johnson, T. & Shaver, A. The evolution of mutation rates: separating
576 causes from consequences. *BioEssays : news and reviews in molecular, cellular and developmental*
577 *biology* **22**, 1057-1066, doi:10.1002/1521-1878(200012)22:12<1057::AID-BIES3>3.0.CO;2-W (2000).
578 42 de Visser, J. A. The fate of microbial mutators. *Microbiology* **148**, 1247-1252 (2002).

579 43 Turrientes, M. C. *et al.* Normal mutation rate variants arise in a Mutator (Mut S) Escherichia coli
580 population. *PloS one* **8**, e72963, doi:10.1371/journal.pone.0072963 (2013).
581 44 Shaver, A. C. & Sniegowski, P. D. Spontaneously arising mutL mutators in evolving Escherichia coli
582 populations are the result of changes in repeat length. *Journal of bacteriology* **185**, 6076-6082
583 (2003).
584 45 Scott, J., Thompson-Mayberry, P., Lahmamsi, S., King, C. J. & McShan, W. M. Phage-associated
585 mutator phenotype in group A streptococcus. *Journal of bacteriology* **190**, 6290-6301,
586 doi:10.1128/JB.01569-07 (2008).
587 46 Wielgoss, S. *et al.* Mutation rate dynamics in a bacterial population reflect tension between
588 adaptation and genetic load. *Proceedings of the National Academy of Sciences of the United States*
589 *of America* **110**, 222-227, doi:10.1073/pnas.1219574110 (2013).
590 47 Denamur, E. *et al.* Evolutionary implications of the frequent horizontal transfer of mismatch repair
591 genes. *Cell* **103**, 711-721 (2000).
592 48 Agnew, W. & Barnes, A. C. Streptococcus iniae: an aquatic pathogen of global veterinary
593 significance and a challenging candidate for reliable vaccination. *Veterinary microbiology* **122**, 1-15,
594 doi:10.1016/j.vetmic.2007.03.002 (2007).
595 49 Millard, C. M. *et al.* Evolution of the capsular operon of Streptococcus iniae in response to
596 vaccination. *Applied and environmental microbiology* **78**, 8219-8226, doi:10.1128/AEM.02216-12
597 (2012).
598 50 Hall, B. M., Ma, C. X., Liang, P. & Singh, K. K. Fluctuation analysis CalculatOR: a web tool for the
599 determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics* **25**, 1564-
600 1565, doi:10.1093/bioinformatics/btp253 (2009).
601 51 Zheng, Q. Methods for comparing mutation rates using fluctuation assay data. *Mutat Res-Fund Mol*
602 *M* **777**, 20-22, doi:10.1016/j.mrfmmm.2015.04.002 (2015).
603 52 Lenhart, J. S., Pillon, M. C., Guarne, A., Biteen, J. S. & Simmons, L. A. Mismatch repair in Gram-
604 positive bacteria. *Research in microbiology* **167**, 4-12, doi:10.1016/j.resmic.2015.08.006 (2016).
605 53 Locke, J. B. *et al.* Streptococcus iniae capsule impairs phagocytic clearance and contributes to
606 virulence in fish. *Journal of bacteriology* **189**, 1279-1287, doi:10.1128/JB.01175-06 (2007).
607 54 Locke, J. B. *et al.* Streptococcus iniae beta-hemolysin streptolysin S is a virulence factor in fish
608 infection. *Dis Aquat Organ* **76**, 17-26, doi:Doi 10.3354/Dao076017 (2007).
609 55 Rodriguez, J. L., Dalia, A. B. & Weiser, J. N. Increased chain length promotes pneumococcal
610 adherence and colonization. *Infect Immun* **80**, 3454-3459, doi:10.1128/IAI.00587-12 (2012).
611 56 Bjarnsholt, T. The role of bacterial biofilms in chronic infections. *APMIS. Supplementum*, 1-51,
612 doi:10.1111/apm.12099 (2013).
613 57 Kohchi, C., Inagawa, H., Nishizawa, T. & Soma, G. ROS and innate immunity. *Anticancer research* **29**,
614 817-821 (2009).
615 58 Rosche, W. A. & Foster, P. L. Determining mutation rates in bacterial populations. *Methods* **20**, 4-
616 17, doi:10.1006/meth.1999.0901 (2000).
617 59 Hedges, A. J. Estimating the precision of serial dilutions and viable bacterial counts. *International*
618 *journal of food microbiology* **76**, 207-214 (2002).
619 60 Zheng, Q. Methods for comparing mutation rates using fluctuation assay data. *Mutation research*
620 **777**, 20-22, doi:10.1016/j.mrfmmm.2015.04.002 (2015).
621 61 Kawasaki, M. *et al.* Microevolution of aquatic Streptococcus agalactiae ST-261 from Australia
622 indicates dissemination via imported tilapia and ongoing adaptation to marine hosts or
623 environment. *Appl Environ Microbiol*, doi:10.1128/AEM.00859-18 (2018).
624 62 Bankevich, A. *et al.* SPAdes: a new genome assembly algorithm and its applications to single-cell
625 sequencing. *Journal of computational biology : a journal of computational molecular cell biology* **19**,
626 455-477, doi:10.1089/cmb.2012.0021 (2012).
627 63 Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome
628 assemblies. *Bioinformatics* **29**, 1072-1075, doi:10.1093/bioinformatics/btt086 (2013).

64 Rissman, A. I. *et al.* Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* **25**, 2071-2073, doi:10.1093/bioinformatics/btp356 (2009).
65 Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068-2069, doi:10.1093/bioinformatics/btu153 (2014).
66 Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biology* **15**, 524, doi:10.1186/PREACCEPT-2573980311437212 (2014).
67 Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic acids research* **43**, e15, doi:10.1093/nar/gku1196 (2015).
68 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).
69 Leache, A. D., Banbury, B. L., Felsenstein, J., de Oca, A. N. & Stamatakis, A. Short Tree, Long Tree, Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. *Systematic Biology* **64**, 1032-1047, doi:10.1093/sysbio/syv053 (2015).
70 Huson, D. H. *et al.* Dendroscope: An interactive viewer for large phylogenetic trees. *BMC bioinformatics* **8**, 460, doi:10.1186/1471-2105-8-460 (2007).
71 He, Z. *et al.* Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic acids research* **44**, W236-241, doi:10.1093/nar/gkw370 (2016).
72 Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2** (2016).
73 Darling, A. E., Mau, B. & Perna, N. T. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PloS one* **5**, e11147, doi:10.1371/journal.pone.0011147 (2010).
74 Kearse, M. *et al.* Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, doi:10.1093/bioinformatics/bts199 (2012).
75 Ben Zakour, N. L. *et al.* Sequential Acquisition of Virulence and Fluoroquinolone Resistance Has Shaped the Evolution of Escherichia coli ST131. *MBio* **7**, e00347-00316, doi:10.1128/mBio.00347-16 (2016).
76 Umarov, R. K. & Solovyev, V. V. Recognition of prokaryotic and eukaryotic promoters using convolutional deep learning neural networks. *Plos One* **12**, doi:ARTN e0171410

10.1371/journal.pone.0171410 (2017).
77 Letunic, I. & Bork, P. 20 years of the SMART protein domain annotation resource. *Nucleic acids research* **46**, D493-D496, doi:10.1093/nar/gkx922 (2018).
78 Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745-2747, doi:10.1093/bioinformatics/btv195 (2015).
79 Petkau, A., Stuart-Edwards, M., Stothard, P. & Van Domselaar, G. Interactive microbial genome visualization with GView. *Bioinformatics* **26**, 3125-3126, doi:10.1093/bioinformatics/btq588 (2010).
80 Arndt, D. *et al.* PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research* **44**, W16-W21, doi:10.1093/nar/gkw387 (2016).
81 Siguier, P., Perochon, J., Lestrade, L., Mahillon, J. & Chandler, M. ISfinder: the reference centre for bacterial insertion sequences. *Nucleic Acids Research* **34**, D32-D36, doi:10.1093/nar/gkj014 (2006).
82 Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, 402, doi:10.1186/1471-2164-12-402 (2011).
83 Rajagopal, L. Understanding the regulation of Group B Streptococcal virulence factors. *Future microbiology* **4**, 201-221, doi:10.2217/17460913.4.2.201 (2009).
84 Kadioglu, A., Weiser, J. N., Paton, J. C. & Andrew, P. W. The role of Streptococcus pneumoniae virulence factors in host respiratory colonization and disease. *Nat Rev Microbiol* **6**, 288-301, doi:10.1038/nrmicro1871 (2008).

85 Andrews, J. M. Determination of minimum inhibitory concentrations. *The Journal of antimicrobial chemotherapy* **48 Suppl 1**, 5-16 (2001).

86 Verotta, D., Haagensen, J., Spormann, A. M. & Yang, K. Mathematical Modeling of Biofilm Structures Using COMSTAT Data. *Comput Math Methods Med* **2017**, 7246286, doi:10.1155/2017/7246286 (2017).

87 Symonds, M. R. E. & Blomberg, S. P. in *Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice* (ed László Zsolt Garamszegi) 105-130 (Springer Berlin Heidelberg, 2014).

88 Zheng, Q. rSalvador: An R Package for the Fluctuation Experiment. *G3* **7**, 3849-3856, doi:10.1534/g3.117.300120 (2017).

89 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57**, 289-300 (1995).

90 Chen, F. *et al.* Multiple genetic switches spontaneously modulating bacterial mutability. *Bmc Evol Biol* **10**, 277, doi:10.1186/1471-2148-10-277 (2010).

91 Couce, A. *et al.* Mutator genomes decay, despite sustained fitness gains, in a long-term experiment with bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **114**, E9026-E9035, doi:10.1073/pnas.1705887114 (2017).

92 Gould, C. V., Sniegowski, P. D., Shchepetov, M., Metlay, J. P. & Weiser, J. N. Identifying mutator phenotypes among fluoroquinolone-resistant strains of Streptococcus pneumoniae using fluctuation analysis. *Antimicrobial agents and chemotherapy* **51**, 3225-3229, doi:10.1128/AAC.00336-07 (2007).

93 Bedford, T., Cobey, S. & Pascual, M. Strength and tempo of selection revealed in viral gene genealogies. *Bmc Evol Biol* **11**, 220, doi:10.1186/1471-2148-11-220 (2011).

94 Neher, R. A. & Hallatschek, O. Genealogies of rapidly adapting populations. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 437-442, doi:Doi 10.1073/Pnas.1213113110 (2013).

95 Dupes, N. M. *et al.* Mutations in the Bacillus subtilis beta clamp that separate its roles in DNA replication from mismatch repair. *J Bacteriol* **192**, 3452-3463, doi:10.1128/JB.01435-09 (2010).

96 Lenhart, J. S., Sharma, A., Hingorani, M. M. & Simmons, L. A. DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. *Mol Microbiol* **87**, 553-568, doi:10.1111/mmi.12115 (2013).

97 Yang, H. & Miller, J. H. Deletion of dnaN1 generates a mutator phenotype in Bacillus anthracis. *DNA Repair (Amst)* **7**, 507-514, doi:10.1016/j.dnarep.2007.10.003 (2008).

98 Korkmaz, G., Holm, M., Wiens, T. & Sanyal, S. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. *J Biol Chem* **289**, 30334-30342, doi:10.1074/jbc.M114.606632 (2014).

99 Geno, K. A. *et al.* Pneumococcal Capsules and Their Types: Past, Present, and Future. *Clinical microbiology reviews* **28**, 871-899, doi:10.1128/CMR.00024-15 (2015).

100 Andisi, V. F. *et al.* Pneumococcal gene complex involved in resistance to extracellular oxidative stress. *Infect Immun* **80**, 1037-1049, doi:10.1128/IAI.05563-11 (2012).

101 Pier, G. B. & Madin, S. H. Streptococcus iniae sp nov, a beta-hemolytic Streptococcus isolated from an amazon freshwater dolphin, Inia geoffrensis. . *International Journal of Systematic Bacteriology* **26**, 545-553 (1976).

102 Pier, G. B., Madin, S. H. & Alnakeeb, S. Isolation and characterisation of a 2nd isolate of Streptococcus iniae. *International Journal of Systematic Bacteriology* **28**, 311-314 (1978).

103 Lees, J. A. *et al.* Evaluation of phylogenetic reconstruction methods using bacterial whole genomes: a simulation based study. *Wellcome Open Res* **3**, 33, doi:10.12688/wellcomeopenres.14265.2 (2018).

104 Maiden, M. C. J. *et al.* MLST revisited: the gene-by-gene approach to bacterial genomics. *Nat Rev Micro* **11**, 728-736, doi:10.1038/nrmicro3093

729 http://www.nature.com/nrmicro/journal/v11/n10/abs/nrmicro3093.html#supplementary-information
730 (2013).
731 105 Dai, L., Muraoka, W. T., Wu, Z. W., Sahin, O. & Zhang, Q. J. A single nucleotide change in mutY
732 increases the emergence of antibiotic-resistant Campylobacter jejuni mutants. *J Antimicrob*
733 *Chemoth* **70**, 2739-2748, doi:10.1093/jac/dkv190 (2015).
734 106 Rajanna, C. *et al.* A strain of Yersinia pestis with a mutator phenotype from the Republic of Georgia.
735 *FEMS microbiology letters* **343**, 113-120, doi:10.1111/1574-6968.12137 (2013).
736 107 Baumler, A. & Fang, F. C. Host specificity of bacterial pathogens. *Cold Spring Harbor perspectives in*
737 *medicine* **3**, a010041, doi:10.1101/cshperspect.a010041 (2013).
738 108 Facklam, R., Elliott, J., Shewmaker, L. & Reingold, A. Identification and characterization of sporadic
739 isolates of Streptococcus iniae isolated from humans. *Journal of Clinical Microbiology* **43**, 933-937
740 (2005).
741 109 Weinberger, D. M., Malley, R. & Lipsitch, M. Serotype replacement in disease after pneumococcal
742 vaccination. *Lancet* **378**, 1962-1973, doi:10.1016/S0140-6736(10)62225-8 (2011).
743 110 Croucher, N. J. *et al.* Population genomics of post-vaccine changes in pneumococcal epidemiology.
744 *Nat Genet* **45**, 656-663, doi:10.1038/ng.2625 (2013).
745 111 Croucher, N. J. *et al.* Rapid pneumococcal evolution in response to clinical interventions. *Science*
746 **331**, 430-434, doi:10.1126/science.1198545 (2011).
747 112 Croucher, N. J. *et al.* Selective and genetic constraints on pneumococcal serotype switching. *PLoS*
748 *Genet* **11**, e1005095, doi:10.1371/journal.pgen.1005095 (2015).

749

750

751 Table 1. *Streptococcus iniae* strains used in this study. Includes origin details (host species, time, site of isolation),
752 phylogenetic affiliation, virulence-associated phenotypes, and mutation rate. Atypical places of isolation (hosts,
753 tissues) and deviant phenotypes are in bold.

Strain	Clade	Host	Year	Location	Capsule	Hemolysis	Chains	Oxidation resistance	Biofilm formation	Mutation rate (10^{-7})
QMA0071	A	*Lates calcarifer*	2000	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.2041
QMA0074	C2	*Lates calcarifer*	1998	QLD	**absent**	wildtype	**long**	wildtype	wildtype	0.6579
QMA0077	C2	*Lates calcarifer*	1995	QLD	**absent**	wildtype	**long**	wildtype	wildtype	0.7067
QMA0078	A	*Lates calcarifer*	2001	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.2153
QMA0080	C1	*Lates calcarifer*	2004	WA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1966
QMA0082	C1	*Lates calcarifer*	2004	WA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1761
QMA0083	A	*Lates calcarifer*	2004	WA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1894
QMA0084	B	*Epalzeorhynchos kalopterus*	2001	WA	wildtype	**weak**	wildtype	wildtype	wildtype	0.7293
QMA0087	A	*Lates calcarifer*	2004	WA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1925
QMA0130	E2	***Homo sapiens***	1995	Canada	**absent**	wildtype	wildtype	wildtype	wildtype	0.468
QMA0131	E2	***Homo sapiens***	1995	Canada	**absent**	wildtype	wildtype	wildtype	wildtype	0.4488
QMA0133	E1	***Homo sapiens***	2001	USA	**absent**	wildtype	wildtype	wildtype	**increased**	0.5539
QMA0134	E1	***Homo sapiens***	2001	USA	**differential**	wildtype	wildtype	wildtype	**increased**	0.5457
QMA0135	E1	***Homo sapiens***	2002	USA	wildtype	wildtype	**short**	**increased**	wildtype	0.631
QMA0137	E1	***Homo sapiens***	2004	USA	wildtype	wildtype	**short**	**increased**	wildtype	0.5495
QMA0138	E1	***Homo sapiens***	2004	USA	wildtype	wildtype	**short**	**increased**	wildtype	0.5989
QMA0139	F	*Fish sp.*	1996	Canada	**absent**	wildtype	wildtype	wildtype	wildtype	0.4311
QMA0140	-	***Inia geoffrensis***	1976	USA	**absent**	wildtype	wildtype	**increased**	**increased**	0.0901
QMA0141	-	***Inia geoffrensis***	1978	USA	**absent**	wildtype	wildtype	**increased**	**increased**	1.0884
QMA0142	C1	*Lates calcarifer*	2005	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.2044
QMA0150	C1	*Lates calcarifer*	2005	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.1952
QMA0155	A	*Lates calcarifer*	2005	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1493
QMA0156	A	*Lates calcarifer*	2005	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1666
QMA0157	A	*Lates calcarifer*	2005	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1545
QMA0158	A	*Lates calcarifer*	2006	SA RAS	**absent**	wildtype	**long**	wildtype	wildtype	0.1678
QMA0159	A	*Lates calcarifer*	2006	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.2054
QMA0160	A	*Lates calcarifer*	1999	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1481
QMA0161	A	*Lates calcarifer*	2000	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1742
QMA0162	A	*Lates calcarifer*	2000	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1718
QMA0163	A	*Lates calcarifer*	2000	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1523
QMA0164	C2	*Lates calcarifer*	2006	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.7102
QMA0165	C2	*Lates calcarifer*	2006	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.6606
QMA0177	C1	*Lates calcarifer*	2006	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.1718
QMA0180	C1	*Lates calcarifer*	2006	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.2011
QMA0186	D	*Oncorhynchus mykiss*	2000	Israel	wildtype	**weak**	wildtype	wildtype	**increased**	0.5135
QMA0187	-	*Channa striata*	1983	Thailand	**absent**	**weak**	wildtype	wildtype	wildtype	0.4635
QMA0188	D	*Oncorhynchus mykiss*	1998	Israel	wildtype	**weak**	wildtype	wildtype	**increased**	0.5042

QMA0189	D	*Oncorhynchus mykiss*	1996	Reunion	wildtype	**weak**	wildtype	wildtype	wildtype	0.4907
QMA0190	F	*Channa striata*	1988	Thailand	**absent**	wildtype	wildtype	wildtype	wildtype	0.6413
QMA0191	C1	*Lates calcarifer*	2005	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.1883
QMA0207	C1	*Lates calcarifer*	2006	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.1486
QMA0216	A	*Lates calcarifer*	2007	QLD	**absent**	wildtype	**long**	wildtype	wildtype	0.1949
QMA0218	C2	*Lates calcarifer*	2007	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.7093
QMA0220	A	*Lates calcarifer*	2006	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.2086
QMA0221	A	*Lates calcarifer*	2007	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1814
QMA0222	A	*Lates calcarifer*	2006	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1829
QMA0233	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**increased**	0.362
QMA0234	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**increased**	0.354
QMA0235	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**increased**	0.3496
QMA0236	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**increased**	0.3413
QMA0244	A	*Lates calcarifer*	2008	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.2101
QMA0245	A	*Lates calcarifer*	2008	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1988
QMA0246	A	*Lates calcarifer*	2009	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.2149
QMA0247	A	*Lates calcarifer*	2009	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.2059
QMA0248	A	*Lates calcarifer*	2009	SA RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1893
QMA0249	F	*Lates calcarifer,* ***bone***	2009	SA RAS	**differential**	**weak**	**long**	wildtype	**increased**	0.3817
QMA0250	A	*Lates calcarifer*	2007	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.195
QMA0251	A	*Lates calcarifer*	2008	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.1927
QMA0252	A	*Lates calcarifer*	2008	NSW RAS	wildtype	wildtype	wildtype	wildtype	wildtype	0.2006
QMA0253	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	wildtype	**long**	wildtype	**increased**	0.3555
QMA0254	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	wildtype	**long**	wildtype	**increased**	0.3712
QMA0258	A	*Lates calcarifer*	2008	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.1876
QMA0371	A	*Scortum barcoo*	2011	NSW	wildtype	wildtype	wildtype	wildtype	wildtype	0.1889
QMA0373	C2	*Lates calcarifer*	2012	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.6908
QMA0374	C2	*Lates calcarifer*	2012	QLD	wildtype	wildtype	wildtype	wildtype	wildtype	0.729
QMA0445	G	*Oreochromis sp.*	1998	USA	**absent**	wildtype	wildtype	wildtype	wildtype	0.2149
QMA0446	G	*Oreochromis sp.*	1998	USA	**absent**	wildtype	wildtype	wildtype	wildtype	0.1687
QMA0447	E1	*Hybrid striped bass*	1996	USA	wildtype	wildtype	**short**	**increased**	wildtype	0.5902
QMA0448	E1	*Hybrid striped bass*	1998	USA	wildtype	wildtype	**short**	**increased**	wildtype	0.5529
QMA0457	A	*Oreochromis sp.*	2005	USA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1682
QMA0458	A	*Epalzeorhynchos bicolor*	2004	USA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1778
QMA0462	B	*Chromobotia macracanthus*	2005	USA	wildtype	**weak**	wildtype	wildtype	wildtype	0.4242
QMA0463	B	*Chromobotia macracanthus*	2005	USA	wildtype	**weak**	wildtype	wildtype	wildtype	0.414
QMA0466	E2	*Oreochromis sp.*	-	USA	wildtype	wildtype	**short**	wildtype	wildtype	0.4602

QMA0467	A	*Epalzeothynchos frenatum*	2004	USA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1695
QMA0468	A	*Oreochromis sp.*	2005	USA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1689
QMA0490	A	*Oreochromis sp.*	2015	Honduras	wildtype	wildtype	wildtype	wildtype	wildtype	0.1812
QMA0491	A	*Oreochromis sp.*	2015	Honduras	wildtype	wildtype	wildtype	wildtype	wildtype	0.2
QMA0492	A	*Oreochromis sp.*	2015	Honduras	wildtype	wildtype	wildtype	wildtype	wildtype	0.1849
QMA0493	A	*Oreochromis sp.*	2016	Honduras	wildtype	wildtype	wildtype	wildtype	wildtype	0.1987

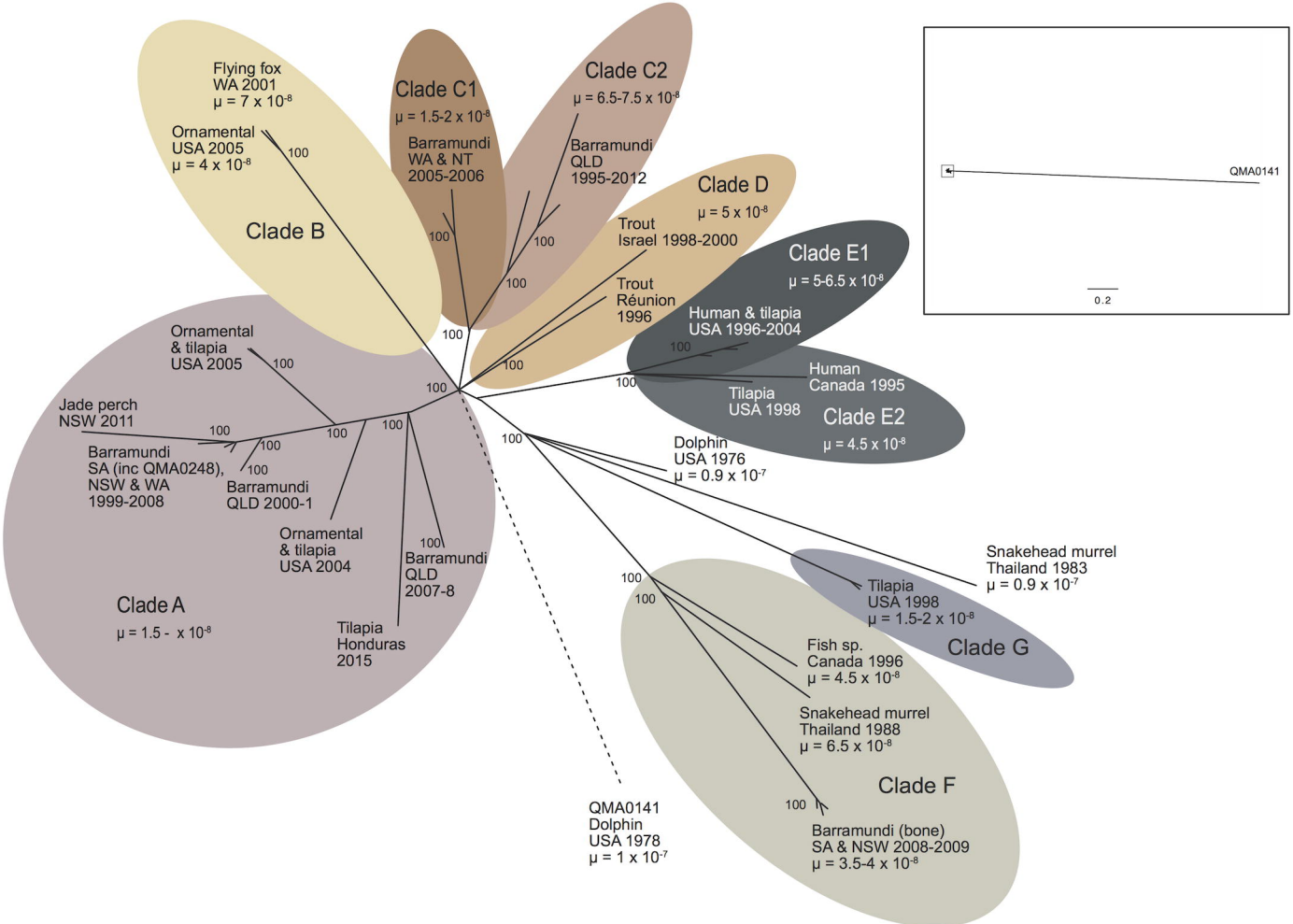
756 Table 2: Variants in MMR and OG genes found among *S. iniae* isolates. CDS – protein coding sequence, BS – predicted
757 binding sites of transcription factors, PROVEAN score – predicted effect of amino acid substitutions on the protein
758 function; scores lower that -2 are considered to indicate a deleterious effect.

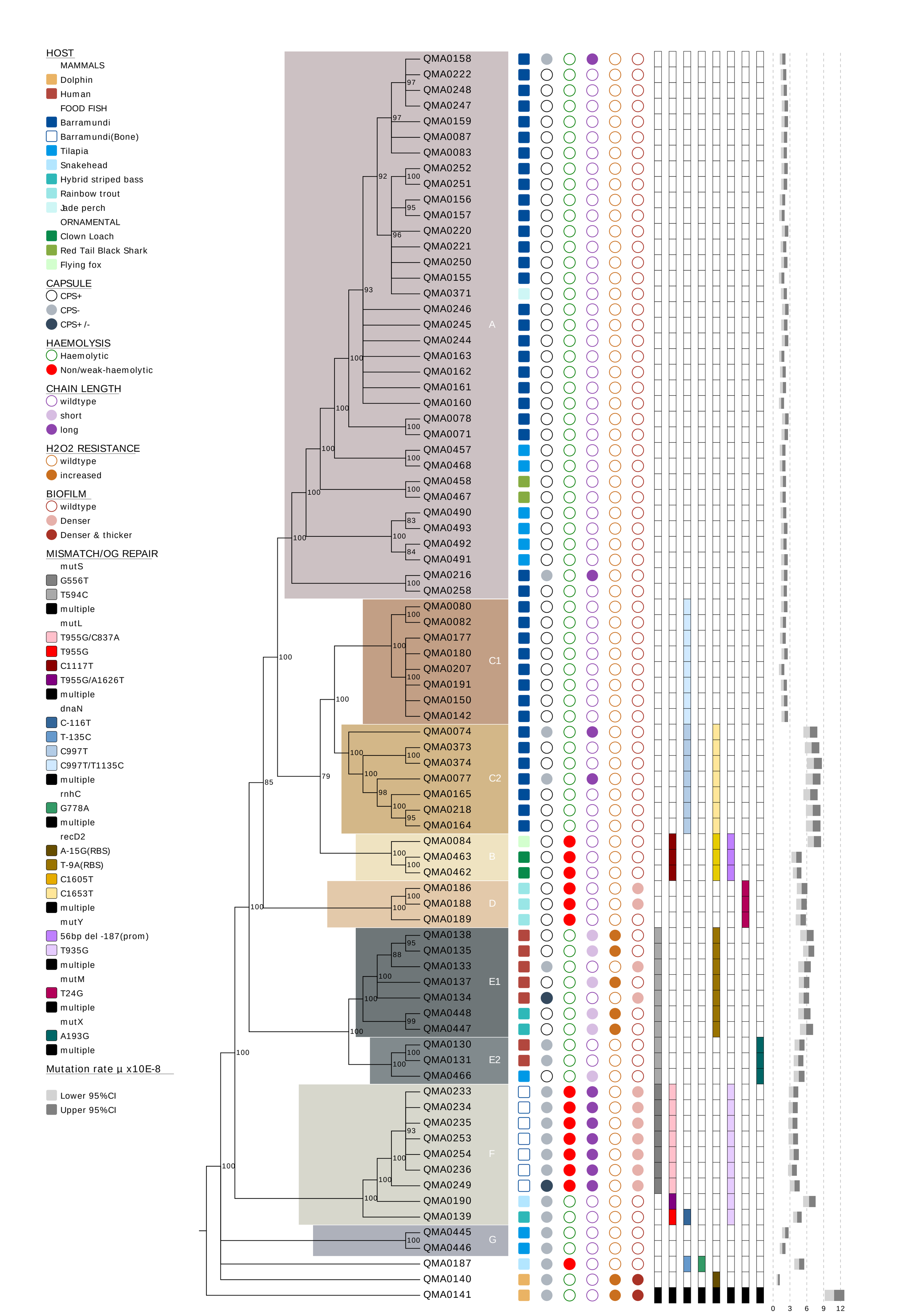
759

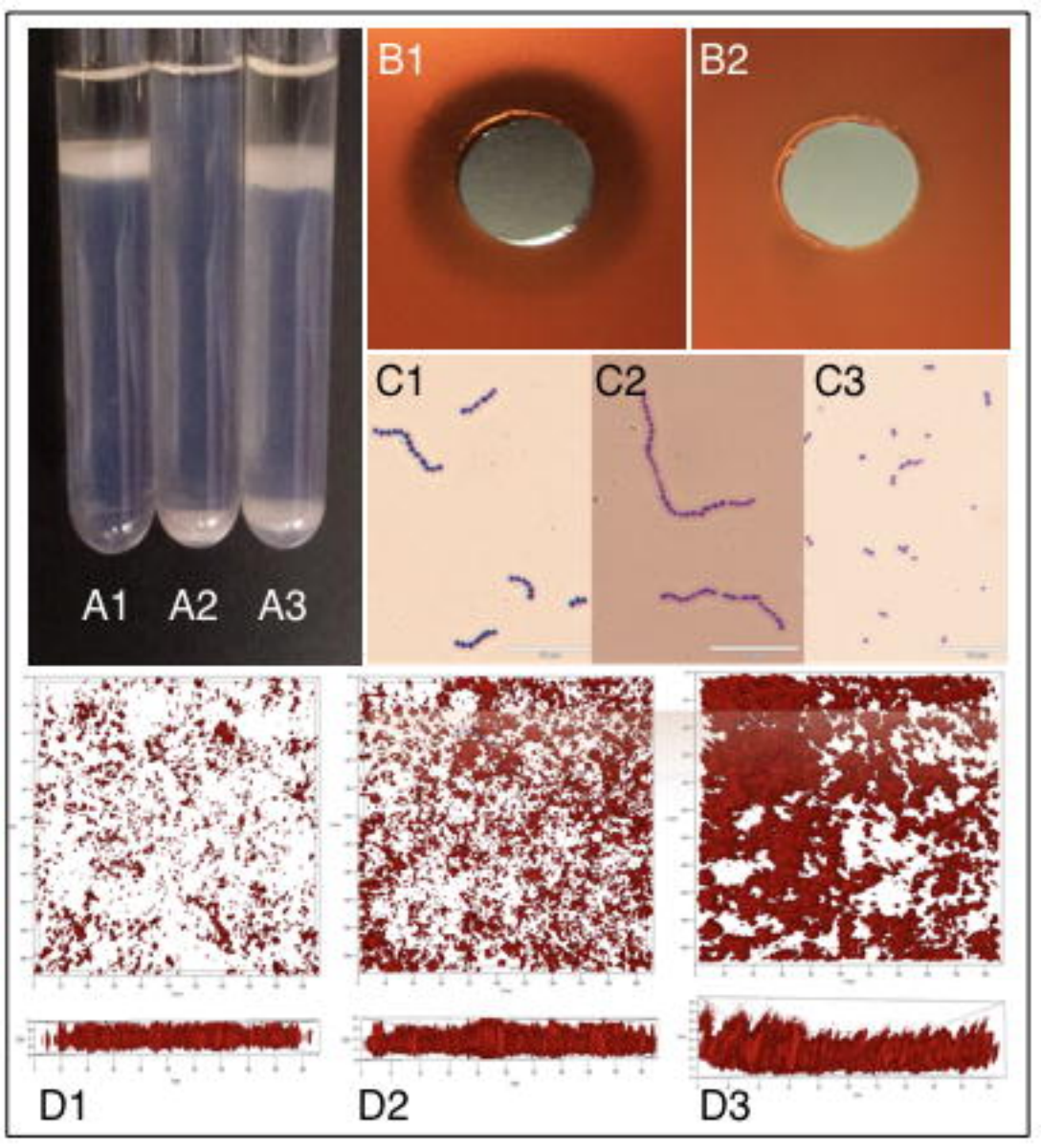
Gene	Clade	Strains	Location	Mutation type	Nucleotide change	Amino acid change	Protein functional domain	PROVEAN score
mutS	F	from bone	CDS	SNP	G556T	Q -> K	Pfam MutS_II	-0.635
	E	all	CDS	SNP	T594C	-	Pfam MutS_II	n/a
mutL	F	from bone	CDS	SNP	C837A	-	DNA mis_repair	n/a
	F	all	CDS	SNP	T955G	S -> R	DNA mis_repair	-2.050
	B	all	CDS	SNP	C1117T	V -> I	-	-0.271
	F	QMA0190	CDS	SNP	A1626T	-	MutL_C	n/a
dnaN	F	QMA0139	promoter, BS of DnaA	SNP	C->T 116 bp upstream	n/a	n/a	n/a
	G	QMA0187	promoter, BS of FNR	SNP	T->C 135 bp upstream	n/a	n/a	n/a
	C2	all strains	CDS	SNP	C977T	T -> I	Pol3Bc	-2,216
	C1	all strains	stop codon	SNP	T1135C	STOP> Q	n/a	-0.005
rnhC	-	QMA0187	CDS	SNP	G778A	D ->N	Pfam: RNase_HII	-4.593
recD2	-	QMA0140	RBS	SNP	A -> G 15 bp upstream	n/a	n/a	n/a
	E1	all strains	RBS	SNP	T -> A 9 bp upstream	n/a	n/a	n/a
	B	all strains	CDS	SNP	C1605T	-	-	n/a
	C2	all strains	CDS	SNP	C1653T	-	-	n/a
mutY	B	all strains	promoter, -35 box, 1st bp of -10 box, 2 BSs of RpoD17	large deletion (56 bp)	atcataagatttcctcctataaatctgtgtcatatgtcatccttcttatagtttc 247-187 bp upstream	n/a	n/a	n/a
	F	all strains	CDS	SNP	T935G	I -> T	Pfam: NUDIX_4	-2.544
mutM	D	all strains	CDS	SNP	T24G	E -> D	Fapy_DNAglyco	-2.990
mutX	E2	all strains	CDS	SNP	A193G	M -> I	Pfam: NUDIX	-0.539

760

761


Flying fox
WA 2001
μ = 7 x 10-8
Ornamental
USA 2005
μ = 4 x 10-8
Clade B
Clade C1
μ = 1.5-2 x 10-8
Barramundi
WA & NT
2005-2006
Clade C2
μ = 6.5-7.5 x 10-8
Barramundi
QLD
1995-2012
Clade D
μ = 5 x 10-8
Trout
Israel 1998-2000
Trout
Réunion
1996
Clade E1
μ = 5-6.5 x 10-8
Human & tilapia
USA 1996-2004
Human
Canada 1995
Tilapia
USA 1998
Clade E2
μ = 4.5 x 10-8
Dolphin
USA 1976
μ = 0.9 x 10-7
Ornamental
& tilapia
USA 2005
Jade perch
NSW 2011
Barramundi
SA (inc QMA0248),
NSW & WA
1999-2008
Barramundi
QLD 2000-1
Ornamental
& tilapia
USA 2004
Barramundi
QLD
2007-8
Tilapia
Honduras
2015
Clade A
μ = 1.5 - x 10-8
Snakehead murrel
Thailand 1983
μ = 0.9 x 10-7
Tilapia
USA 1998
μ = 1.5-2 x 10-8
Clade G
Fish sp.
Canada 1996
μ = 4.5 x 10-8
Snakehead murrel
Thailand 1988
μ = 6.5 x 10-8
Clade F
Barramundi (bone)
SA & NSW 2008-2009
μ = 3.5-4 x 10-8
QMA0141
Dolphin
USA 1978
μ = 1 x 10-7
QMA0141
0.2
100


HOST
MAMMALS
Dolphin
Human
FOOD FISH
Barramundi
Barramundi(Bone)
Tilapia
Snakehead
Hybrid striped bass
Rainbow trout
Jade perch
ORNAMENTAL
Clown Loach
Red Tail Black Shark
Flying fox
CAPSULE
CPS+
CPS-
CPS+ /-
HAEMOLYSIS
Haemolytic
Non/weak-haemolytic
CHAIN LENGTH
wildtype
short
long
H2O2 RESISTANCE
wildtype
increased
BIOFILM
wildtype
Denser
Denser & thicker
MISMATCH/OG REPAIR
mutS
G556T
T594C
multiple
mutL
T955G/C837A
T955G
C1117T
T955G/A1626T
multiple
dnaN
C-116T
T-135C
C997T
C997T/T1135C
multiple
rnhC
G778A
multiple
recD2
A-15G(RBS)
T-9A(RBS)
C1605T
C1653T
multiple
mutY
56bp del -187(prom)
T935G
multiple
mutM
T24G
multiple
mutX
A193G
multiple
Mutation rate μ x10E-8
Lower 95%CI
Upper 95%CI
QMA0158
QMA0222
QMA0248
QMA0247
QMA0159
QMA0087
QMA0083
QMA0252
QMA0251
QMA0156
QMA0157
QMA0220
QMA0221
QMA0250
QMA0155
QMA0371
QMA0246
QMA0245
QMA0244
QMA0163
QMA0162
QMA0161
QMA0160
QMA0078
QMA0071
QMA0457
QMA0468
QMA0458
QMA0467
QMA0490
QMA0493
QMA0492
QMA0491
QMA0216
QMA0258
A
QMA0080
QMA0082
QMA0177
QMA0180
QMA0207
QMA0191
QMA0150
QMA0142
C1
QMA0074
QMA0373
QMA0374
QMA0077
QMA0165
QMA0218
QMA0164
C2
QMA0084
QMA0463
QMA0462
B
QMA0186
QMA0188
QMA0189
D
QMA0138
QMA0135
QMA0133
QMA0137
QMA0134
QMA0448
QMA0447
E1
QMA0130
QMA0131
QMA0466
E2
QMA0233
QMA0234
QMA0235
QMA0253
QMA0254
QMA0236
QMA0249
QMA0190
QMA0139
F
QMA0445
QMA0446
G
QMA0187
QMA0140
QMA0141
0
3
6
9
12


A1
A2
A3
B1
B2
C1
C2
C3
D1
D2
D3