Article

Mutators as drivers of adaptation in pathogenic bacteria and a risk factor for host jumps and vaccine escape

Oleksandra Silayeva, Jan Engelstaedter, and Andrew C Barnes*

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

Key words: mutators; DNA repair; pathogen evolution; immune escape, vaccination; Streptococcus; aquaculture; host jumps, zoonoses

***Correspondence:**

a.barnes@uq.edu.au

Abstract

Heritable hypermutable strains deficient in DNA repair genes (mutators) facilitate microbial adaptation as they may rapidly generate beneficial mutations. Bacterial mutators deficient in mismatch (MMR) and oxidised guanine (OG) repair are abundant in clinical samples and show increased adaptive potential in experimental models of infection in mice. However, their role in pathoadaptation is poorly understood. Here we investigate the role of mutators in epidemiology and evolution of the rapidly evolving aquatic pathogen, *Streptococcus iniae*, employing a collection of 80 strains isolated globally over 40 years. We determine phylogenetic relationship among *S. iniae* using 10,267 non-recombinant core genome single nucleotide polymorphisms (SNPs), estimate their mutation rate by fluctuation analysis, and detect variation in major MMR (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*) and OG (*mutY*, *mutM*, *mutT*) genes. We find that *S. iniae* mutation rate phenotype and genotype are strongly associated with phylogenetic diversification and variation in major streptococcal virulence determinants (capsular polysaccharide, hemolysin, cell chain length, resistance to oxidation, and biofilm formation). Furthermore, profound changes in virulence determinants observed in mammalian isolates (atypical host) and vaccine-escape isolates found in bone (atypical tissue) of vaccinated barramundi are linked to multiple MMR and OG variants and unique mutation rates. This implies that adaptation to new host taxa, new host tissue, and to immunity of a vaccinated host is promoted by mutator strains. Our findings support the importance of mutation rate dynamics in evolution of pathogenic bacteria, in particular adaptation to a drastically different immunological setting that occurs during host jump and vaccine escape events.

Introduction

Mutation rates in microbial populations are determined by a trade-off between maintenance of genetic integrity and evolvability (Stich, et al. 2010; Ferenci 2015). High fidelity of DNA repair is advantageous in adapted populations since most mutations lead to deviation from the evolved phenotype (Stich, et al. 2010; Sniegowski and Raynes 2013). Conversely, lower fidelity of DNA repair can be advantageous under stressful conditions since increased mutation supply promotes diversity and may thereby accelerate adaptation (Stich, et al. 2010; Wielgoss, et al. 2013). Bacteria can efficiently fluctuate between high and low mutation regimes: While extremely low mutation rates are maintained in favourable conditions (Wielgoss, et al. 2011), elevation of mutation rates known as stress-induced mutagenesis (SIM) occurs in adapting populations (Galhardo, et al. 2007; Sundin and Weigand 2007). One aspect of SIM is temporary hypermutability associated with changes in expression of multiple genes including error-prone DNA polymerases, recombination-preventing enzymes and enzymes involved in movement of mobile elements (Foster 2007). The most extensively studied example of transient hypermutability is the RecA-mediated general stress response, SOS, associated with up-regulation of error-prone polymerases that can by-pass DNA lesions during replication (Foster 2007). However, it is recognised that increased fitness during SOS response stems from immediate growth advantage rather than from long-term evolution of resistance (Wrande, et al. 2008; Andersson and Hughes 2010; Torres-Barcelo, et al. 2015). Another aspect of SIM, and the focus of the present study, is the evolution of heritable mutator phenotypes, also referred to as constitutive, permanent, or true mutators, that contain mutator alleles and replicate over multiple generations with a high amount of error (Galhardo, et al. 2007; Sundin and Weigand 2007). Heritably hypermutable strains, which periodically arise within a population via mutations in DNA repair genes (Miller 1996; Boe, et al. 2000), are selected under stress indirectly via 'hitchhiking' of mutator alleles with stress-resistance mutations (Mao, et al. 1997; Taddei, et al. 1997), or potentially yet-to-be established pleiotropic effects conferred by inactivation of DNA repair (Torres-Barcelo, et al. 2013; Raynes and Sniegowski 2014). There is much stronger theoretical and experimental evidence demonstrating that heritable mutators facilitate adaptive evolution (Raynes and Sniegowski 2014). However, even heritable mutators are transient phenomena in natural populations as non-mutator variants are selected within already adapted mutator clones (de Visser 2002; Turrientes, et al. 2013). Restoration of mutation rate can occur via copy-number variation leading to full reversion of the mutator allele (Shaver and Sniegowski 2003), disruption of the linkage between mutator alleles and beneficial mutations via recombination (Denamur, et al. 2000), anti-mutator compensatory mutations (Turrientes, et al. 2013; Wielgoss, et al. 2013), or prophage excision when its integration confers increase in mutation rate (Scott, et al. 2008).

Mutator strains are abundant among bacterial clinical isolates, especially in chronic cases (Hall and Henderson-Begg 2006; Oliver 2010). Multiple studies have attributed this to antibiotic selection (Negri,

et al. 2002; Matsushima, et al. 2010; Wang, et al. 2013), since mutators are a well-established risk factor for the development of antimicrobial resistance (Blazquez 2003; Chopra, et al. 2003). However, the association between frequency of mutators in diagnostic samples and antibiotic treatments is weak (Gutierrez, et al. 2004; Mena, et al. 2008), and isogenic knockout mutator strains have increased ability to colonize an animal host in challenge models devoid of antibiotic exposure (Giraud, et al. 2001; Nilsson, et al. 2004; Healey, et al. 2016). Thus, the prevalence of mutators in clinical isolates most likely results from the selective process during adaptation to the host (Labat, et al. 2005; Mena, et al. 2007; Mena, et al. 2008; Oliver and Mena 2010), where immunity acts as the ultimate selective pressure shaping the diversity of pathogenic strains (Aquino and Nunes 2016).

The 'immunological niche' concept states that, due to the diversity and complexity of immune responses, every individual host represents a unique habitat that requires adaptation (Cobey 2014). Considering that the pathogen needs to adapt to every new host, the host population taken as a whole represents an extremely heterogeneous environment where adaptive processes are ongoing and persistence of mutator alleles may be favoured (Lukacisinova, et al. 2017). Further, as a small number of infection units may enter a host, pathogen adaptation is intrinsically associated with population bottlenecks during which the rate at which beneficial variants arise might be crucial (Metzgar and Wills 2000; Abel, et al. 2015). This supports the contention that bacterial evolution is more likely to occur via multiple mutations rather than a single beneficial variant (Fogle, et al. 2008) The fitness of a pathogenic strain in the host hinges on the concept of virulence (the degree of damage to the host), which is a dynamic, complex trait determined by a multitude of virulence factors and host responses (Methot and Alizon 2014). Optimal virulence is a trade-off between acuteness and persistence with higher virulence facilitating extraction of resources and transmission, but too much damage can kill the host and stop transmission altogether (Mackinnon and Read 2004). Additionally, virulence factors are often antigens so reduced virulence may be a means of immune evasion (Deitsch, et al. 1997). Consequently, both loss and gain of virulence may be advantageous, and indicate adaptive shifts towards acute or chronic pathogenesis, respectively (Maurelli 2007; Shrestha, et al. 2014). Exacerbation and attenuation of virulence factors are repeatedly observed in both knockout and naturally occurring mutators (Picard, et al. 2001; Merino, et al. 2002; Smania, et al. 2004; Mena, et al. 2007; Gonzalez, et al. 2012; Canfield, et al. 2013). Thus, it appears that mutator strains/alleles may help pathogenic bacteria to evolve towards the optimal level of virulence.

Most bacterial mutators isolated from human clinical samples are deficient in genes from mismatch (MMR) and oxidised guanine (OG) DNA repair systems (Hall and Henderson-Begg 2006). MMR and OG genes are conserved across all domains of life (Fukui 2010), and single nucleotide polymorphisms in their sequence may produce drastic alterations in mutation rate (Rajanna, et al. 2013; Dai, et al. 2015). Nonetheless, these genes appear to be variable in major bacterial pathogens, confirming a potential role

104 of mutation rate dynamics in the evolution of virulence (Ambur, et al. 2009; Chen, et al. 2010). MMR is
105 coupled to replication and removes mispaired nucleotides as well as short insertion/deletion loops, and
106 prevents recombination between divergent sequences (Li 2008; Tham, et al. 2013).The MMR pathway
107 has been extensively characterised in *Escherichia coli,* where the MutS dimer protein interacts with the
108 replication processivity clamp DnaN (b-clamp), binds to the mismatch, recruits MutL, and the MutS:MutL
109 complex then activates the MutH endonuclease (Fukui 2010). MutH cuts the unmethylated nascent
110 strand at hemimethylated d(GATC) sites. MutL loads the UvrD helicase that unwinds the nicked DNA
111 containing the error, which is followed by excision by one of the single-strand nucleases (Fukui 2010).
112 This MMR organization, however, appears to be limited to *E. coli* and closely related
113 Gammaproteobacteria (Fukui 2010; Lenhart, et al. 2016). In most bacteria and eukaryotes, MutH is
114 absent and MutL has a conserved endonuclease site, dam methylase (acting at GATC sites) is absent and
115 the strand discrimination signal is unknown (Fukui 2010; Lenhart, et al. 2016). Studies in *Bacillus subtilis,*
116 a model species for gram-positive bacteria, suggest that nicks created by ribonuclease *rnhC* during
117 removal of misincorporated ribonucleosides direct MutL endonuclease towards the newly synthethised
118 strand, and that recD2 acts as a MMR helicase instead of uvrD (Lenhart, et al. 2016). The OG system
119 repairs oxidative DNA damage throughout the cell cycle (Lu, et al. 2001; David, et al. 2007). Among all
120 DNA bases, guanine is most sensitive to reactive oxygen species and its oxidised form, 8-oxo-dG or OG,
121 is highly mutagenic as it preferentially pairs with adenine which, if not corrected, leads to G:C to A:T
122 transversion on a second round of replication (Lu, et al. 2001; David, et al. 2007). OG system organization
123 is conserved in most bacteria and eukaryotes and generally MutT homologs remove 8-oxo-dG from the
124 nucleotide pool, MutM homologs excise 8-oxo-dG incorporated into DNA, and MutY homologs remove
125 adenine from OG:A mispairings (Lu, et al. 2001; David, et al. 2007).

126 The present study investigates the role of mutators and mutation rate dynamics in the epidemiology and
127 evolution of pathogenic bacteria using the broad host-range pathogen *Streptococcus iniae*. *S. iniae* is a
128 highly adaptable species, evidenced by its global distribution, ability to cross divergent host taxa
129 boundaries and repeated re-infection of previously immunised hosts during disease outbreaks (Agnew
130 and Barnes 2007; Millard, et al. 2012). Moreover, Streptococcal pathogens in general are difficult
131 candidates for immunisation in humans and animals due to high strain diversity evidently driven by
132 immunisation, yet there is little information on the role of mutators in this evolutionary process
133 (Croucher, et al. 2011; Croucher, et al. 2013). We determine variation in mutation rate and identify
134 mutations in MMR (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*) and OG (*mutY* , *mutM*, *mutT*) genes among 80 diverse
135 strains of *S. iniae.* We discover that both mutation rates and mutations in MMR and OG genes correlate
136 with phylogenetic diversification but are highly conserved within phylogenetic clades. To determine
137 whether this adaptive process might be driven by the host immune response we identify variation in

major phenotypic traits that determine virulence in streptococci (capsular polysaccharide, hemolysin, length of cell chains) and bacteria in general (resistance to reactive oxygen species (ROS), and biofilm formation). *S. iniae* polysaccharide capsule and the beta-hemolysin streptolysin S both increase survival under host immune response via reduction of phagocytic killing and other mechanisms (Locke, Colvin, Datta, et al. 2007; Locke, Colvin, Varki, et al. 2007; Kadioglu, et al. 2008; Rajagopal 2009). However, both are antigens that induce antibody production and development of immune memory (Kadioglu, et al. 2008; Rajagopal 2009). In addition, both capsule production and hemolysin activity have many implications that are not fully understood (Kadioglu, et al. 2008; Geno, et al. 2015). For example, non-encapsulated strains exhibit greater ability to adhere to and colonize an epithelium (Hammerschmidt, et al. 2005). Length of cell-chains is also linked to virulence in streptococci: shorter chains are associated with acute infection and increased survival in the blood as they are less likely to activate the complement, whilst longer chains promote adherence and colonization (Rodriguez, et al. 2012). As pathogenic bacteria are continuously exposed to reactive oxygen released by immune cells, resistance to oxidative compounds contributes to virulence in pathogens allowing them to avoid immune clearance (Hegde, et al. 2008; Andisi, et al. 2012). Biofilm formation is also generally associated with virulence in bacteria; detached cells are associated with acute infections whereas biofilm aggregates are a hallmark of chronic conditions (Bjarnsholt 2013). The expression of these traits among *S. iniae* isolates is highly consistent with phylogenetic affiliation, which supports that strain diversity is primarily driven by adaptation to the host achieved via adjustment of virulence. Furthermore, we find that occurrence of abnormal virulence determinants is strongly associated with variation in MMR and OG genes and occurrence of infection in unusual immunological landscapes such as mammals and bone tissue of immunised fish. These data support the contention that mutators play a critical role in adaptation to the host and evolution of virulence, and suggest that these strains may facilitate the reinfection of immunised animals and transmission between divergent host species.

163 **Results and Discussion**

164 Our work investigates the link between mutation rate genotype and phenotype and phylogenetic
165 diversification among *Streptococcus iniae* strains isolated globally between 1976 and 2016 from different
166 host taxa. Maximum likelihood phylogenetic analysis of 80 *S. iniae* isolates (Table 1) based on non-
167 recombinant core genome SNPs derived from whole genome data resolved 6 major clades (A-F), one
168 lineage with two strains (clade G), and three lineages with a single strain (Figs. 1,2). Clades C-E are
169 endemic; and clades A, B, and F contain strains from diverse geographic regions. Principally, variation in
170 MMR and GO genes occurs only between the lineages, and they are highly conserved within clades (Fig.
171 2, Table 2). In fact, sequences of the major MMR and OG genes are conserved even within phylogenetic
172 groups not readily definable by geographic origin, time, or host species, and these genes appear to be
173 good candidates for Multilocus Sequence Typing (MLST), i.e. identification of phylogenetic position based
174 on selected genes as opposed to a whole genome analysis (Maiden, et al. 2013). Although even
175 synonymous SNPs in any gene locus can affect mRNA and have pronounced phenotypic consequences
176 (Shabalina, et al. 2013), most SNPs in MMR and OG genes were identified in protein functional domains,
177 and most amino acid substitutions were predicted to have a deleterious effect on protein function (Table
178 2). Further, the number of MMR and OG variants correlates significantly with mutation rate (p = 0.0056)
179 and with the number of atypical phenotypic traits contributing to virulence (p = 0.0331), as predicted by
180 a Phylogenetic Generalised Least Squares (PGLS) regression model that accounts for autocorrelation
181 occurring between closely related strains (Symonds and Blomberg 2014). In turn, mutation rate is highly
182 consistent with phylogenetic affiliation and, according to multiple pair-wise strain comparisons by
183 Maximum Likelihood Ratio test (Zheng 2015), with probabilities corrected for multiple comparison using
184 False Discovery Rate (Benjamini and Hochberg 1995), mutation rate difference is insignificant within but
185 significant between major phylogenetic lineages (Suppl. Table 1).

186 Although mutation rate variation between divergent strains is significant, the magnitude of the
187 differences is modest (largely around 5-fold) (Figure 1, 2, Suppl. Table 2), which is consistent with mutator
188 phenotype transience and restoration of low mutation rate in adapted populations. On the other hand,
189 only one order of magnitude or lower differences in mutation rate distinguishes mutator from non-
190 mutator phenotypes in streptococci (around 10^{-7} and 10^{-8} mutation rates respectively) (Hall and
191 Henderson-Begg 2006; Gould, et al. 2007), in contrast to other common pathogens where 2-3 orders of
192 magnitude differences are observed (Hall and Henderson-Begg 2006). Heritably hypermutable
193 phenotypes are costly due to inevitable accumulation of deleterious changes. Once adaptation is gained
194 the mutator strain is predicted to evolve back towards a lower mutation rate (de Visser 2002; Raynes and
195 Sniegowski 2014; Sprouffske, et al. 2018). This may occur via compensatory, 'anti-mutator' mutations in
196 DNA repair genes (Turrientes, et al. 2013; Wielgoss, et al. 2013), or via recombination (Denamur, et al.

197 2000). Our data favour the former mechanism for several reasons. Firstly, we used a maximum likelihood
198 inference method to exclude areas of recombination from our phylogenetic analyses (Croucher, Page, et
199 al. 2015), and the MMR and OG gene regions were not identified as recombinant regions in this analysis.
200 Secondly, we created a pangenome for the 80 strains in our analysis and mapped on putative phage and
201 prophage signatures (Arndt, et al. 2016), and insertion sequence elements (Siguier, et al. 2006). None of
202 the MMR or OG genes, which are all present within the core genome of *S. iniae*, occurred in proximity to
203 phage or ISE (Suppl. Fig. 1). Whilst we cannot rule out switch by recombination, these data suggest that
204 accrual of SNPs via random mutation is a more likely explanation.

205 To investigate whether mutation rate phenotype and inferred phylogenetic association with MMR/OG
206 genotype was also associated with virulence phenotype, we investigated key streptococcal virulence
207 traits amongst the bacterial isolates in the analysis. We observe significant correlation (p = 0.0331, PGLS)
208 of the number of MMR and OG variants with number of atypical phenotypes for traits contributing to
209 virulence in streptococci (capsular polysaccharide, hemolysin, length of cell chains) and bacteria in
210 general (resistance to ROS, and biofilm formation). This suggests that fluctuations in mutation rate may
211 be linked to shifts in virulence, potentially allowing major adaptive events and leading to phylogenetic
212 diversification (Fig. 2, Table 3). Moreover, infections in atypical hosts (mammals) and atypical tissue
213 (bone) within apparently immune vaccinated hosts correlates strongly with number of atypical virulence-
214 associated phenotypes (p = 0 .000) and the number of variants MMR and OG variants (p = 0.002). This
215 suggests that mutator strains may facilitate multiple genetic changes leading to profound shifts in
216 virulence and antigenicity during adaptation to a drastically different immunological setting.

217 Of the clades identified within our collection of *S. iniae* isolates, Clade A is apparently a dominant
218 circulating clade as it has persisted globally for almost two decades (Clade A contains strains isolated
219 from USA, Honduras, and Australia between 1999 to 2016), primarily infecting closely related Perciform
220 fish (barramundi (*Lates calcarifer*), tilapia (*Oreochromis sp*), jade perch (*Scortum barcoo*)). Mutation rates
221 in all isolates among this ancestral lineage (35 strains analysed) fall within a 1.5 - 2 x 10^{-8} range (Figs. 1, 2;
222 Clade A), which is similar to the base mutation rate estimated for other non-mutator gram positive
223 bacteria (Hall and Henderson-Begg 2006; Gould, et al. 2007). All differences in mutation rates within this
224 clade are insignificant except three pairwise comparisons that verge on p <0.05 (Suppl. Table 1). MMR
225 and OG genes in these strains are identical, and only minor deviations from wildtype traits contributing
226 to virulence were observed among the strains (absence of capsule production and longer cell chains in
227 QMA0158, 216) (Figs. 1-3, Table 1). All of the other lineages (with the exceptions of clade C1 and strains
228 QMA0445-46 discussed below) express significantly different mutation rates to clade A, contain unique
229 and often multiple SNPs in *mut* genes, and show peculiar phenotypic variants related to virulence (Figs.
230 1-3, Tables 1,2, Suppl. Table 1). Although phylogenetic relatedness between lineages is fully resolved (Fig.

231 2), they originate almost simultaneously and nested lineages evolve independently sharing little
232 phylogenetic history (Figs. 1,2). Lack of correlation between branch length and sampling date derived by
233 root-to-tip regression analysis with time is good evidence of non-neutral evolution amongst this
234 collection of *S. iniae* isolates (R^2=0.0936, Correlation coefficient 0.306, best-fit root; Suppl. Fig. 2,
235 (Rambaut, et al. 2016)). Deviation from a neutral model is also supported by highly skewed branching in
236 the tree, with some branches comprising single isolates whilst others comprise many epidemiologically
237 unrelated isolates (Figure 1, 2). This is suggestive of frequent strong selection, presumably imposed by
238 heterogeneity of the immune landscape encountered during transfer between host individuals, and
239 resembles genealogical tree topology derived from viral evolution over similar timespans (Bedford, et al.
240 2011; Neher and Hallatschek 2013) although we acknowledge tree topology may also be affected by
241 sampling bias.

242 Clade B contains QMA0084, a strain isolated from a black flying fox, *Pteropus alecto*, in Western Australia
243 (WA) and two strains isolated from ornamental fish (clown loach or tiger botia, *Chromobotia*
244 *macracanthus*) in USA. These isolates share a valine to isoleucine substitution in MutL predicted as
245 neutral (PROVEAN score -0.271), a synonymous SNP in *recD2*, and a 56 bp deletion affecting the predicted
246 promoter of *mutY* (Table 2). The latter mutation encompasses the -35 box, two binding sites of the
247 RpoD17 general transcription factor, and the 1st bp of the -10 box, and was predicted to abolish the
248 original promoter (Suppl. Fig. 3). However, *mutY* might still be transcribed at some level since two other
249 potential promoters were identified in a nearby sequence (Suppl. Fig. 3). It appears that altered *mutY*
250 expression is also affected by the rest of the genetic background as isolates exhibit significant differences
251 (Suppl. Table 1) in mutation rate phenotype, 7 x 10^{-8} in QMA0084 and 4 x 10^{-8} in clown loach strains (Figs.
252 1-2). Phenotypically, clade B has shifted towards decreased haemolytic activity (Fig. 3, B2).

253 Clade C shares a common ancestor with clade B and is comprised of strains obtained from barramundi
254 farmed in WA, Northern Territory (NT) and north QLD fish farms (nested subclades C1 and C2
255 respectively). While a core mutation rate of 1.5-2 x 10^{-8} is observed in subclade C1, a significantly higher
256 mutation rate of 7 x 10^{-8} has been maintained in subclade C2 from north QLD for almost two decades
257 (1995 to 2012) (Figs. 1,2 Suppl. Table 1). We speculate that the latter might be attributable to difference
258 in meteorological conditions: periodic heavy rainfall present in tropical north QLD is likely to bring
259 fluctuations in water salinity and temperature, creating an unstable environment where adaptive
260 processes are ongoing and high mutation rate is advantageous. Minor variation in virulence-related
261 phenotypes (absence of capsule production and longer cell chains in QMA0074, 77)(Fig. 2, Table 1), are
262 supportive that mutation rate evolution in this lineage might be driven by outside-the-host factors. The
263 significant difference in mutation rate between nested clades C1 and C2 is potentially attributable to
264 variation in *dnaN*. In *Bacillus subtilis,* 90% of mismatch repair is dependent on targeting MutS to nascent

265 DNA via the ß sliding clamp (DnaN) zone (Dupes, et al. 2010; Lenhart, et al. 2013). Inactivation of *dnaN1*
266 in *B. anthracis* results in a mutator phenotype with a mutation rate equivalent to a *mutS* mismatch repair-
267 defective strain (Yang and Miller 2008). The tyrosine to isoleucine substitution in one of the critical
268 residues of the loader binding interface of the ß-clamp, predicted as deleterious (PROVEAN score -2,216),
269 is present in all isolates of from clade C (Table 2), which may explain the high mutation rate phenotype
270 in clade C2 strains (Fig. 2, Table 2). In contrast, only subclade C1 harbours T1135C nucleotide substitution
271 in the *dnaN* that changes stop codon TAA into CAA coding for glutamine predicted as neutral (PROVEAN
272 score -0.005), with the next stop codon TAG found immediately downstream. Putatively, T1135C and/or
273 change of the *dnaN* stop codon (Korkmaz, et al. 2014) may act as compensatory to the T326I substitution,
274 reducing the mutation rate to a value not significantly different from the core rate found in clade A (Fig.
275 1, 2) via protein elongation or by affecting translation via stop codon usage bias (Korkmaz, et al. 2014).
276 In addition, a synonymous variant in *recD2* that is only present in subclade C2 and might affect protein
277 expression, may contribute to the difference in mutation rate between C1 and C2 clades.

278 Clade D consists of isolates from trout (*Oncorhynchus mykiss*) from Réunion and Israel. These strains have
279 an estimated mutation rate of 5 x 10^{-8}, contain a glutamate to aspartate substitution in *mutM* predicted
280 to be deleterious to protein function, a synonymous SNP in *RecD2*, and exhibit impeded haemolytic
281 activity (Fig 1,2, Table 2). Also, isolates from Israel form thicker and more dense biofilm structures (Fig.
282 3, D2).

283 Isolates from humans and fish are found in clade E where multiple phenotypic variants contributing to
284 virulence are observed (Figs. 2-3, Table 1). This clade contains two nested subclades that share a SNP in
285 *mutS,* but have other unique SNPs (Fig. 2, Table 2) and exhibit mutation rates that are differently
286 significant in most pairwise comparisons (Suppl. Table 1). Subclade E1 contains USA isolates from humans
287 (QMA0133-35, 37-38) and hybrid striped bass (QMA0447-48), which have a substitution in the Shine-
288 Dalgarno sequence of *recD2* and a mutation rate of 6.5 x 10^{-8} (Fig 1, Table 2). Fish strains and three human
289 strains (QMA0135, 37-38) form shorter cell chains and higher resistance to oxidative stress. Two human
290 strains QMA0133-34 produce thicker and more dense biofilm structures (Fig. 3, D2). QMA0133 is non-
291 encapsulated (Fig 3, A2), and QMA0134 appears to differentially express the capsular polysaccharide in
292 culture (Fig 3, A3). The close phylogenetic relationship presented here implicates a likely transfer from
293 farmed hybrid bass to humans (Facklam, et al. 2005), and this host jump may have been facilitated by the
294 mutator phenotype. Subclade E2 contains two human strains from Canada (QMA0130-31) and a tilapia
295 strain from USA (QMA0466). These strains have unique methionine to isoleucine substitution at the N-
296 terminus of MutX, and a mutation rate of 4.5 x 10^{-8} (Fig 1, Table 2). Both human strains are non-
297 encapsulated (Fig. 2). The tilapia strain expresses the capsule but forms short cell chains in common with
298 most strains from the subclade (Fig 2, Fig 3, C3). The close phylogenetic relationship again strongly

299 implicates transfer from tilapia farmed in USA to human patients in Canada via import into the Toronto
300 fish market (Weinstein, et al. 1997) and again, this host jump may have been facilitated by the mutator
301 state.

302 Clade F is not readily definable by location, time of isolation, or host species and comprises QMA0139
303 from unknown fish species obtained in Canada, QMA0190 isolated from snakehead murrel in Thailand
304 (both of which appear on long branches), and a nested terminal clade comprised of barramundi isolates
305 farmed in Recirculating Aquaculture Systems (RAS) in New South Wales and South Australia (Fig. 2 1,2).
306 The latter are of a particular interest as these mutators were sampled from barramundi bone lesions
307 during a disease outbreak in vaccinated fish where infection manifested itself as a slowly progressing
308 osteomyelitis instead of typical acute septicaemia and meningitis, implicating mutators in adaptation to
309 new tissue in response to host immunity (Agnew and Barnes 2007; Millard, et al. 2012). Since they are
310 distant to strains used in autogenous vaccines in clade A (QMA0155-57, 160, QMA0250-52), the atypical
311 outbreak might be classified as a case of vaccine-induced serotype replacement (VISR) – spread of co-
312 existing pathogenic strain/s after elimination of the dominant strain/s targeted by the vaccine
313 (Weinberger, et al. 2011). In contrast to natural populations where the investigation of VISR is
314 confounded by ecological and sampling biases, and randomised vaccine trials that lack power to detect
315 the population-wide effect of mass vaccination (Weinberger, et al. 2011), barramundi RAS represent a
316 relatively controlled and isolated natural environment where vaccination against *S. iniae* fails occasionally
317 but recurrently (Millard, et al. 2012). RAS are self-contained facilities with minimal water exchange and
318 high standards of water treatment (Losordo, et al. 2009). This ensures that gene flow and, consequently,
319 the probability of new serotype introduction after vaccination is reduced to a minimum. A comparatively
320 uniform immunized host population is exposed to dominant strains against which it was vaccinated along
321 with potentially co-existing lineages in the RAS. In the atypical outbreaks in RAS in Australia, vaccine-
322 induced immunity was partially effective in the infected fish, evident by absence of proliferation of the
323 new strain in typical niduses (blood, brain, and pronephros) and cross-reactivity of vaccine-induced
324 antibodies against bone isolates (Millard, et al. 2012). This is in contrast with the classical model of
325 serotype replacement where a vaccine is ineffective against the replacing co-existing strain. It has been
326 postulated that VISR can occur even if the vaccine is equally efficient against all circulating strains via
327 other trade-off mechanisms (Weinberger, et al. 2011). Apparently, evasion of immune clearance and
328 adaptation to osseous tissue was permitted by shift towards attenuated virulence indicated by multiple
329 phenotypic changes in bone isolates, including absence of capsule, impeded hemolytic activity, increased
330 cell-chain length, and denser biofilms (Fig. 3). All of these changes have previously been associated with
331 chronic infection and increased potential for colonization (Locke, Colvin, Datta, et al. 2007; Locke, Colvin,
332 Varki, et al. 2007; Kadioglu, et al. 2008; Andisi, et al. 2012; Rodriguez, et al. 2012; Bjarnsholt 2013; Geno,

333 et al. 2015), and only absence of polysaccharide capsule is shared with related QMA139 and QMA190, as
334 well as QMA0140-41, QMA0187, and QMA0445-46, which implicates a non-encapsulated ancestor for
335 these strains. The latter acapsular strain potentially also gave rise to clade E, assuming capsular
336 production was regained in encapsulated strains found in this branch. Multiple SNPs in repair genes (Fig.
337 2, Table 2) and the unique mutation rate (Figs. 1-2, Table 1) in bone isolates suggest that shift in virulence
338 allowing persistence in the bone tissue might have occurred via mutator phenotype. Serine to arginine
339 substitution in MutL and tyrosine to isoleucine substitution in MutY occur in the ancestor of QMA0139,
340 QMA0190, and bone strains. Both substitutions are predicted to have a deleterious effect on protein
341 function (PROVEAN scores -2.050 and -2.544 respectively), but their effect on mutation rate is combined
342 with unique SNPs found in each branch: QMA0139 has a SNP in the binding site for the DnaA transcription
343 factor within the *dnaN* promoter, and a mutation rate of 4.5 x 10^{-8}, QMA0190 has synonymous SNP in
344 *mutL* and a mutation rate of 6.5 x 10^{-8}, and bone strains have glutamine to glycine substitution in MutS
345 predicted as neutral (PROVEAN score -0.635), a synonymous SNP in *mutL*, and a mutation rate of 3-4 x
346 10^{-8}. Considering that the mutation rate of QMA0190 is significantly higher compared to rates expressed
347 by QMA0139 and bone isolates (Suppl. Table 1), it is likely that variants unique to the latter strains
348 compensate for deleterious variants shared by the isolates (Wielgoss, et al. 2013).

349 Clade G contains two strains from tilapia isolated in USA (QMA0445-46). Notably, despite being
350 phylogenetically distant from clade A, these strains have the same MMR and OG genotype and have
351 retained the core mutation rate and the wildtype virulence phenotype (Fig. 2, Table 1), with the exception
352 of capsule absence, perhaps attributable to a common non-encapsulated ancestor shared with clade F,
353 QMA0140-41, QMA0187, and clade E.

354 Three isolates of *S. iniae* were identified as independent phylogenetic lineages, with unique mutation
355 rate genotypes and phenotypes and combinations of phenotypes associated with virulence (Figs. 1-2,
356 Tables 1-2). The first long branch with a single isolate contains a strain from snakehead murrel strain
357 (*Channa striata*) from Thailand. This isolate is non-encapsulated and weakly haemolytic and has a
358 mutation rate of 4.5 x 10^{-8}, potentially attributable to a deleterious aspartate to asparagine substitution
359 in rnhC and SNP in the binding site of *fnr* transcription factor within the *dnaN* promoter. A second long
360 branch with a single strain contains the oldest among the analysed strains, QMA0140 isolated in 1976
361 from dolphin (*Inia geoffrensis*) (Pier and Madin 1976). This strain exhibits a mutation rate of 1 x 10^{-8},
362 which is unique among the strains and significantly lower than the core mutation rate, presumably
363 associated with increased translation rate of recD2 helicase produced by a SNP in the ribosome binding
364 site (Fig 2, Table 2). The longest branch on the tree contains a second dolphin strain, QMA0141, isolated
365 two years later in 1978 (Pier, et al. 1978). This isolate is highly divergent from the rest of strains with
366 around 20 kB of non-recombinant SNPs in pair-wise comparisons with other strains, accounting for

367 around 1% genomic difference. In contrast to QMA140, it mutates at a rate of 1x 10^{-7}, the highest
368 mutation rate phenotype determined among the isolates and significantly different to the rest of the
369 values. Multiple SNPs are observed in all MMR and OG genes: 3 in *dnaN* and *mutX*, 4 in *mutM*, 9 in *mutL*
370 and *rnhC*, 15 in *recD2*, 9 in *mutL* and *rnhC*, and 68 in *mutS*. Both dolphin isolates are non-encapsulated,
371 show increased ability to withstand oxidative stress, and form denser and thicker biofilms (Fig 2, 3).

372 The conservation of *S. iniae* DNA repair genotypes and phenotype within phylogenetic clades and
373 variation between the lineages is consistent with the idea of adaptive fluctuations in mutation rate being
374 a driver of pathogen evolution, and their association with variation in virulence traits supports the
375 contention that adaptive processes and diversification of pathogens are largely driven by host immunity.
376 A number of genetic mechanisms have been proposed to potentially underlie the flux of mutator alleles
377 and phenotypes, including random mutations (Wielgoss, et al. 2013), tandem repeat copy-number
378 variation (Shaver and Sniegowski 2003; Chen, et al. 2010), non-homologous recombination (Denamur, et
379 al. 2000), and prophage integration/excision (Scott, et al. 2008). In *S. iniae,* mutation rate dynamics
380 appear to be primarily associated with random mutations and compensatory evolution. Tandem-repeat
381 copy-number variants were not identified in *S. iniae* MMR or OG genes, supporting the contention that
382 occurrence of mutator phenotypes and their full reversion via this mutational type might be limited to *E.*
383 *coli* and closely related species (Shaver and Sniegowski 2003; Chen, et al. 2010). Indeed, the presence of
384 tandem repeats has only been detected in *mutL* in *Proteobacteria* (Shaver and Sniegowski 2003; Chen, et
385 al. 2010), and it is well recognized that *mutL* is structurally and functionally different in gram-positive
386 bacteria and eukaryotes (Fukui 2010; Lenhart, et al. 2016). With the exception of the large deletion
387 affecting the *mutY* promoter in clade B, all identified mutations are single nucleotide substitutions. These
388 SNPs are not likely to result from gene transfer, as they were not identified amongst regions of
389 recombination, nor within proximity to phage or ISE (Suppl. Fig. 1). These MMR and OG nucleotide
390 substitutions most likely result from random mutations which can be sufficient to induce profound
391 changes in mutation rate (Rajanna, et al. 2013; Dai, et al. 2015). Since, in most cases, SNPs occur in
392 multiple genes (Fig. 2) it is likely that compensatory evolution of mutation rate prevails in *S. iniae* where
393 some variants increase and some decrease the background mutation rate (Wielgoss, et al. 2013).

394 The association of MMR and OG variation with changes in virulence traits (Fig. 2) implies that adaptive
395 processes in *S. iniae* population driven by host immune response may be facilitated by the emergence
396 and decline of mutator strains. Furthermore, most profound changes in virulence-associated traits and
397 MMR and OG variants are observed in atypical hosts (mammals) and atypical tissue (bone) within
398 apparently immune vaccinated fish hosts (Fig. 2), which has major epidemiological implications. First,
399 mutators may promote multiple adaptive changes required for pathogen transition among divergent host
400 taxa known as 'host jumps', such as *Osteichthyes* and *Mammalia* discussed in the present study (Baumler

401 and Fang 2013). Second, mutators might present a risk factor to immune escape after vaccination and
402 serotype replacement. Moreover, as evidenced by *S. iniae* colonization of bone tissue of vaccinated
403 barramundi described previously (Millard, et al. 2012), vaccine escape leading to serotype replacement
404 might occur without major disruption of immune recognition when multiple changes attenuating
405 virulence and antigenicity allow persistence in a tissue with lower immune surveillance.

406 We acknowledge that our results do not exclude the possibility that variation in DNA repair genotype and
407 phenotype may potentially represent neutral variation coinciding with shifts in virulence occurring in each
408 phylogenetic lineage. However, complete conservation of MMR and OG genes and mutation rate within
409 *S. iniae* lineages (Fig. 2) and the fact that most SNPs occur in functionally relevant coding and regulatory
410 regions and/or change amino acid sequence (Table 2), supports that the occurrence of mutator strains is
411 likely to precede and facilitate shifts in virulence leading to major diversification events and
412 epidemiologically relevant incidents such as vaccine escape outbreaks and host jumps. Our results
413 indicate that high mutation rate might promote transition to a new immunological setting requiring
414 multiple mutations and phenotypic changes, such as adaptation to a novel host, new host tissue, or a host
415 with vaccine-induced adaptive immune response. Extension to other species such as *Streptococcus*
416 *pneumoniae* may be particularly interesting as serotype evolution following vaccination programmes is
417 well documented, but the role of mutators in this process is not yet known (Croucher, et al. 2011;
418 Croucher, et al. 2013; Croucher, Kagedan, et al. 2015). Future studies should determine how generalisable
419 the role of mutators is in the epidemiology of bacterial pathogens, evolution of virulence and antigenicity,
420 and emergence of new pathogenic strains.

421 **Materials and methods**

422 **Strains and growth conditions**

423 Eighty isolates of *S. iniae* collected in Australia, USA, Canada, Israel, Honduras, and Thailand between
424 1976 and 2016 from eight fish species (*Lates calcarifer*, *Scortum barcoo, Epalzeorhynchos frenatum,*
425 *Epalzeorhynchos bicolor, Oreochromis sp., Channa striata, Chromobotia macracanthus, Oncorhynchus*
426 *mykiss*) and three mammalian species (*Homo sapiens, Inia geoffrensis, Pteropus alecto*) were used in this
427 study (Table 1). Strains were received from culture collections, veterinarians, or directly from fish farms
428 (Suppl. Table 1) and stored as master seed stocks without further subculture at -80°C in Todd-Hewitt
429 Broth (THB) supplemented with 20% glycerol. Bacteria were routinely recovered from -80°C stocks on
430 Columbia agar supplemented with 5% defibrinated sheep blood (Oxoid, Australia), and cultured at 28°C
431 on Todd-Hewitt agar or in Todd-Hewitt broth (Oxoid) with agitation 200 rpm in a shaking incubator unless
432 otherwise specified.

433 **Estimation of mutation rate phenotype by fluctuation analysis**

434 To estimate mutation rates of *S. iniae* isolates, a fluctuation analysis assay for spontaneous occurrence
435 of rifampicin resistance was optimized according to Rosche and Foster (Rosche and Foster 2000). A single
436 broth culture was initiated from five separate colonies, recovered on Columbia blood agar from stock
437 cultures stored at -80°C, and grown overnight to late-exponential phase in Todd-Hewitt broth (THB).
438 Cultures were adjusted to OD_{600} = 1 corresponding to 10^8 viable CFU/mL, diluted 1:100, and distributed
439 in 200 µl aliquots into 8 wells of a sterile U-bottom 96-well plate (Greiner). Prior to dilution of 100 µl of
440 overnight OD_{600}-adjusted cultures were spread onto THA supplemented with 0.5 µg/mL rifampicin to
441 confirm absence of rifampicin-resistant mutants. Although comparatively large initial inoculum (1-2 x 10^5
442 CFU per culture, confirmed by Miles and Misra CFU count (Hedges 2002) of OD-adjusted cultures) was
443 used, this reduced to the minimum variation in final CFU number (N_t) as shown by the preliminary
444 experiments. This approach was undertaken since final population size is critical for estimation of the
445 mutation rate and measuring N_t for such a large number of strains and replicated cultures was technically
446 prohibiting. Also, invariability in N_t allowed statistical comparison of mutation rates by the Maximum
447 Likelihood Ratio test (Zheng 2015). To infer the N_t in final cultures and monitor their growth by optical
448 density, replicate plates containing two cultures per strain were prepared and incubated in a BMG
449 FLUOstar OPTIMA microplate reader. When these representative cultures were reaching an early
450 stationary phase, CFU counts were performed by Miles and Misra method (Hedges 2002) and invariably
451 estimated as 1-2 x 10^8 CFU in each well. Immediately after, whole 200 µL cultures were plated on Todd-
452 Hewitt agar containing 0.5 µg/mL rifampicin for selection of mutants, dried in a laminar flow hood, and
453 incubated until rifampicin resistant colonies appeared. The assay was repeated four times for each strain
454 in blocks of eight cultures using 20 isolates haphazardly chosen from different phylogenetic lineages in
455 each measurement. Processing of all 80 strains at the same time was not only time-consuming, but also
456 compromised the uniformity of initial inoculum size due to growth of OD_{600}-adjusted cultures after the
457 addition of the fresh THB. In some cases, e. g. a single strain representing an independent phylogenetic
458 lineage or with significantly differences in mutation rates detected in closely related strains, assays were
459 repeated 1-2 more times to increase a statistical power. The rifampicin-resistant mutant counts were
460 pooled into single data sets representing 32 to 48 cultures per strain (Suppl. Table 2).

461 **DNA extraction, preparation and sequencing**

462 Genomic DNA was extracted from cells collected from 10 mL late-exponential phase culture in Todd-
463 Hewitt broth with the DNeasy Blood & Tissue kit (Qiagen) using a modified protocol with an additional
464 lysis step as described previously (Kawasaki, et al. 2018). gDNA was analysed on agarose gel, quantified
465 by Qubit fluorimetry (Invitrogen), and 16S rRNA gene sequenced using universal 27F and 1492R primers
466 (Amann, et al. 1995), to confirm integrity, identity, and purity. When the required concentration for
467 sequencing was not achieved in samples they were dried by vacuum centrifugation (SpeedVac) at room

468 temperature. Sequencing was performed on the Illumina HiSeq2000 platform from Nextera XT pair-end
469 libraries at Australian Genome Research Facility, Melbourne. A reference genome from strain QMA0248
470 was constructed using both long reads derived from a single Smrt Cell using the PacBio RS II system with
471 P4C2 chemistry and short reads from Illumina HiSeq2000 derived from Nextera XT paired-end libraries
472 as reported elsewhere (NCBI accession no: GCA_002220115.1). All sequence data is deposited at NCBI
473 under Bioproject number PRJNA417543, SRA accession SRP145425. Sample numbers, accession numbers
474 and extended metadata are provided in Suppl. Table 3. Assembly statistics are provided in Suppl. Table
475 4.

476 **Phylogenetic analysis**

477 Phylogeny was constructed based on core genome SNPs from *de novo* genome assemblies filtered to
478 remove recombination breakpoints. Paired-end reads from Illumina were trimmed with Nesoni clip tool
479 version 0.132 (http://www.vicbioinformatics.com/software.nesoni.shtml), with minimum read length
480 50, and the first 15 bp of each read removed as quality deterioration in this region was observed when
481 assessed with FASTQC version 0.11.5. Assembly was performed using the SPAdes assembler version 3.7.1
482 (Bankevich, et al. 2012), with minimum read coverage cutoff set to 10. Quality of assemblies was assessed
483 with QUAST 3.2 (Gurevich, et al. 2013). Contigs were ordered by alignment to QMA248 reference genome
484 (CP022392.1) with Mauve Contig Mover 2.4.0 (Rissman, et al. 2009). Genome annotation was performed
485 using Prokka 1.11 (Seemann 2014). Rapid alignment of core genomes was carried out using parsnp in the
486 Harvest Tools suite version 1.2 (Treangen, et al. 2014), and the resulting alignment provided as an input
487 to Gubbins 1.4.7 (Croucher, Page, et al. 2015) for detection and exclusion of variants produced by
488 recombination. Phylogenies were then inferred from post-filtered core genome polymorphic sites by
489 maximum likelihood using RAxML 8.2.8 (Stamatakis 2014) with the general time reversible nucleotide
490 substitution model GTRGAMMA and bootstrap support from 1000 iterations. Effect of ascertainment bias
491 associated with using only polymorphic sites on branch length during ML inference was corrected using
492 Felsenstein's correction implemented in RAxML 8 (Leache, et al. 2015). The resulting phylogenetic tree
493 was visualized using Dendroscope v 3.5.7 (Huson, et al. 2007) with bootstrap node support value cut-off
494 75. For the phylogram figure, tip labels were hidden for clarity and the edge containing QMA0141 was
495 re-scaled as dotted line representing 100-fold decrease in length (Fig. 1). A cladogram based on the
496 inferred phylogeny, showing all tip labels and bootstrap support for each node, was annotated with
497 metadata using Evolview V2 (Fig. 2) (He, et al. 2016). To determine whether there was a strong temporal
498 signal in the phylogenetic data, a root-to-tip regression of branch length against time since isolation was
499 performed in TempEst (Rambaut, et al. 2016), using genetic distances, corrected for ascertainment bias
500 in an unrooted tree estimated by maximum likelihood from the alignment of non-recombinant core-
501 genome SNPs in RAxML 8.2.8.

502 **Variation in DNA repair genes and their regulatory regions**

503 SNP analysis of MMR and GO genes was performed by read-mapping with Geneious version 9.1 (Kearse,
504 et al. 2012), using default settings unless otherwise specified. Paired-end reads from each genome were
505 trimmed, merged into a single file with expected distances between the reads set to 250 bp, and mapped
506 to the curated reference genome of strain QMA0248. Mapped reads were used to detect SNPs with
507 minimum coverage set to 10 and frequency to 0.9. A consensus "pseudogenome" was generated for
508 every strain based on the reference sequence incorporating any detected variants (Ben Zakour, et al.
509 2016). Multiple alignment of these pseudogenomes was carried out with *Geneious* aligning tool.
510 Sequences of MMR genes (*dnaN,* mut*S, muTL, recD2, rnhC*) and OG genes (*mutY* , *mutM*, *mutX*) annotated
511 in the QMA0248 reference genome were identified in each strain genome sequence. Promoter regions
512 of repair genes were determined with BPROM (Umarov and Solovyev 2017), protein functional domains
513 by SMART genomic (Letunic and Bork 2018), and effect of amino acid substitutions on protein function
514 by PROVEAN Protein with a sensitivity cut-off of 1.3 (Choi and Chan 2015). To detect large variants,
515 alignment of these regions extracted from *de novo* genome assemblies was carried out, and a 56 bp
516 deletion detected in *mutY* promoter of clade B was confirmed by PCR with
517 CAGAAGGAAGAAACAGAC_F/ACCTCTATTGTAGCAAAG_R primers.

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

519 A pangenome for the 80 *S. iniae* isolates was constructed with GView server (Petkau, et al. 2010) using
520 the QMA0248 reference genbank (.gbk) file as seed by sequentially adding 79 genome assemblies by
521 clade in the order derived from the cladogram (Fig. 2). Phage positions in the resulting pangenome
522 were then determined by BLAST using Phaster (Arndt, et al. 2016) while IS positions were identified
523 using ISFinder (Siguier, et al. 2006) with a BLAST e-value cutoff of 1×10^{-30}. Positions of MMR and OG
524 genes in the pangenome were determined via manual search of the .gbk file. An image was created in
525 BRIG (Alikhan, et al. 2011) using the pangenome as reference then aligning the 80 genome assemblies
526 by BLAST with an e-value cutoff of 1×10^{-30}. The resulting image was annotated in BRIG from a
527 spreadsheet derived from the positions determined above, and coloured according to clade (Suppl.
528 Fig. 1).

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

530 A multitude of intracellular and secreted enzymes contribute to virulence in the *Streptococcus* genus
531 (Kadioglu, et al. 2008; Rajagopal 2009). Considering the large number of strains, we limited analysis to
532 major, easily observable, and distinct phenotypes strongly associated with virulence in streptococci

533 (capsule, hemolysis, cell chains) and bacteria in general (oxidation resistance, biofilms). All assays were
534 performed at least in triplicate per strain.

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

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

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

547 Rapid high-throughput detection of impaired haemolytic activity was achieved by blood-agar clearance
548 zone assay. Briefly, 5 mm wells were made in Columbia agar supplemented with 5% defibrinated sheep
549 blood (Oxoid, Australia). Bacterial cultures grown to late-exponential phase (~10^8 CFU) were diluted
550 1:1000 and 50 µL of diluted cultures were pipetted into punctures in the agar (initial inoculum ~$5x10^4$
551 CFU in each puncture). A 3 mm wide clearance zone was generally produced by bacterial lysis of sheep
552 erythrocytes during 24 h incubation (Fig. 3, B1). Where haemolysis was absent or fragmentary, impeded
553 haemolytic activity was recorded (Fig. 3, B2).

554 **Chain formation microscopy**

555 To assess chain formation, *S. iniae* cultures in THB were grown stationarily in 96-well plates at 28 °C for
556 24 h, mixed and 5 µL wet mounts prepared on a glass microscope slide. Slides were observed by bright
557 field microscopy under 40x objective with an Olympus BX40 microscope and captured using an Olympus
558 DP28 digital camera using CellSens software (Olympus Optical Co, Japan). Specimens were observed for
559 at least 3 min and 2-20 cell chains were generally observed (Fig. 3, C1). Where over 20 cells in a chain
560 were repeatedly detected (Fig. 3, C2) increased chain formation was recorded, and when more than 10
561 cells in a chain were not detected by similar observation and the culture was mainly composed of
562 detached cells (Fig. 3, C3) impeded chain formation was recorded. For the figure, 50 µL of cultures were
563 dried onto slides at RT, fixed with methanol, Gram stained, and images captured under 100x objective.

564 **Oxidation resistance assay**

565 Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) (Andrews 2001)
566 of hydrogen peroxide were measured to assess oxidative stress resistance among *S. iniae* strains. THB
567 cultures grown to late exponential phase were adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 1000-fold, and
568 distributed by 100 µL (~10^4 CFU) into wells of a U-bottom 96-well plate (Greiner). THB (100 µL) with 2-
569 fold excess concentration of hydrogen peroxide was added to the wells and serially diluted twofold,
570 resulting in a range from 0 to 10 mM final peroxide concentrations. Plates were incubated stationarily
571 for 24h and examined for presence of observable growth to determine the MIC. Viable cell counts of
572 cultures without visible growth were performed to determine MBC. MIC of hydrogen peroxide was
573 estimated as 3 mM for all strains, and MBC as 4 mM for the majority of strains. MBC elevated to 5 mM
574 in dolphin and human isolates from USA was classified deviation from a wildtype phenotype.

575 **Biofilm formation and visualisation assay**

576 *S. iniae* biofilms were grown for 4 days in 8-well Lab-Tek® II Chamber Slide™ Systems. To prepare an initial
577 inoculum of 5 x 10^5 CFU, THB cultures grown to late exponential phase were adjusted to OD_{600} = 1 (10^8
578 CFU/mL), diluted 100-fold, and 0.5 mL of diluted cultures were placed into Chamber slide wells. After 24
579 h incubation, THB was removed and replaced with fresh THB every 10-14 h for 3 days. For visualisation,
580 biofilms were washed in PBS, stained for 15 min with 1 µM fluorescent *Bac*Light Red bacterial stain,
581 washed in PBS, fixed for 30 min with 10% formalin, and washed twice PBS. Z-stacks were collected by
582 ZEISS LSM 710 Inverted Laser Scanning Confocal Microscope at 20x objective and visualised using
583 ZEN2012. Quantification of biofilms was performed using COMSTAT (Verotta, et al. 2017). Structures of
584 biomass larger than 3.5 and average thickness over 5 were classified as increased biofilm forming activity
585 (Fig. 3 D).

586 .

587 **Identifying wildtype virulence-related phenotypes**

588 We consider prevalent phenotypes as a wildtype, namely: presence of polysaccharide capsule (71.25 %
589 of strains; Figure 3, A1) and haemolytic activity (85% of strains; Figure 3, B1), up to 20 cells in a chain
590 (77.5 % of strains; Figure 3, A1), 3 mM MIC and 4 mM MBC of hydrogen peroxide (91.25 % of strains),
591 and 8 -12 µm thick biofilms covering up to 30% of representative image (86.25 % of strains, Fig. 4A). Other
592 phenotypes are regarded as deviant and discussed.

593 **Statistical analysis**

594 Analysis of co-variance between number of MMR and OG variants, mutation rates, number of atypical
595 phenotype associated with virulence, and atypical places of isolation were tested with a Phylogenetic
596 Generalised Least Squares model that accounts for relatedness among strains under Pagel's λ of 0.6

implemented in R (Symonds and Blomberg 2014). Estimation of mutation rates was carried out using the Ma-Sandri-Sarkar Maximum Likelihood Estimation (MSS-MLE) method as implemented in FALCOR fluctuation analysis calculator (Hall, et al. 2009). Differences in mutation rates were compared by Likelihood Ratio Test (LRT) using the rSalvador R package (Zheng 2015, 2017). LRT estimates the overlap between confidence intervals calculated for two fluctuation experiment data sets and allows to compare fluctuation assay data with different number of cultures and resolve very small differences in mutation rate (Zheng 2015, 2017). The *Compare.LD* function was applied pair-wise to all strains, and obtained *p* values were corrected for multiple comparisons by controlling the False Discovery Rate (Benjamini and Hochberg 1995) using p.adjust in R. Adjusted *p* values less than 0.05 were considered to indicate significant difference in mutation rates.

Acknowledgements

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

619 **References**

620 Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. 2015. Analysis of Bottlenecks in Experimental
621 Models of Infection. PLoS Pathog 11:e1004823.
622 Agnew W, Barnes AC. 2007. Streptococcus iniae: an aquatic pathogen of global veterinary
623 significance and a challenging candidate for reliable vaccination. Vet Microbiol 122:1-15.
624 Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple
625 prokaryote genome comparisons. BMC Genomics 12:402.
626 Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of
627 individual microbial cells without cultivation. Microbiology Reviews 59:143-169.
628 Ambur OH, Davidsen T, Frye SA, Balasingham SV, Lagesen K, Rognes T, Tonjum T. 2009. Genome
629 dynamics in major bacterial pathogens. FEMS Microbiol Rev 33:453-470.
630 Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse resistance?
631 Nat Rev Microbiol 8:260-271.
632 Andisi VF, Hinojosa CA, de Jong A, Kuipers OP, Orihuela CJ, Bijlsma JJ. 2012. Pneumococcal gene
633 complex involved in resistance to extracellular oxidative stress. Infection and Immunity 80:1037-
634 1049.
635 Andrews JM. 2001. Determination of minimum inhibitory concentrations. J Antimicrob Chemother
636 48 Suppl 1:5-16.
637 Aquino T, Nunes A. 2016. Host immunity and pathogen diversity: A computational study. Virulence
638 7:121-128.
639 Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang YJ, Wishart DS. 2016. PHASTER: a better, faster
640 version of the PHAST phage search tool. Nucleic Acids Research 44:W16-W21.
641 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S,
642 Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-
643 cell sequencing. J Comput Biol 19:455-477.
644 Baumler A, Fang FC. 2013. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med
645 3:a010041.
646 Bedford T, Cobey S, Pascual M. 2011. Strength and tempo of selection revealed in viral gene
647 genealogies. BMC Evol Biol 11:220.
648 Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-Cook M,
649 Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and Fluoroquinolone Resistance
650 Has Shaped the Evolution of Escherichia coli ST131. MBio 7:e00347-00316.
651 Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and Powerful
652 Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-Methodological
653 57:289-300.
654 Bjarnsholt T. 2013. The role of bacterial biofilms in chronic infections. APMIS Suppl:1-51.
655 Blazquez J. 2003. Hypermutation as a factor contributing to the acquisition of antimicrobial
656 resistance. Clinical Infectious Diseases 37:1201-1209.
657 Boe L, Danielsen M, Knudsen S, Petersen JB, Maymann J, Jensen PR. 2000. The frequency of
658 mutators in populations of Escherichia coli. Mutat Res 448:47-55.
659 Canfield GS, Schwingel JM, Foley MH, Vore KL, Boonanantanasarn K, Gill AL, Sutton MD, Gill SR.
660 2013. Evolution in fast forward: a potential role for mutators in accelerating Staphylococcus aureus
661 pathoadaptation. J Bacteriol 195:615-628.
662 Chen F, Liu WQ, Eisenstark A, Johnston RN, Liu GR, Liu SL. 2010. Multiple genetic switches
663 spontaneously modulating bacterial mutability. Bmc Evolutionary Biology 10:277.
664 Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid
665 substitutions and indels. Bioinformatics 31:2745-2747.
666 Chopra I, O'Neill AJ, Miller K. 2003. The role of mutators in the emergence of antibiotic-resistant
667 bacteria. Drug Resist Updat 6:137-145.
668 Cobey S. 2014. Pathogen evolution and the immunological niche. Ann N Y Acad Sci 1320:1-15.

669 Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP,
670 Lipsitch M. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat
671 Genet 45:656-663.
672 Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A,
673 Song JH, Ko KS, et al. 2011. Rapid pneumococcal evolution in response to clinical interventions.
674 Science 331:430-434.
675 Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, Lipsitch M, Hanage WP.
676 2015. Selective and genetic constraints on pneumococcal serotype switching. PLoS Genet
677 11:e1005095.
678 Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid
679 phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using
680 Gubbins. Nucleic Acids Res 43:e15.
681 Dai L, Muraoka WT, Wu ZW, Sahin O, Zhang QJ. 2015. A single nucleotide change in mutY increases
682 the emergence of antibiotic-resistant Campylobacter jejuni mutants. Journal of Antimicrobial
683 Chemotherapy 70:2739-2748.
684 David SS, O'Shea VL, Kundu S. 2007. Base-excision repair of oxidative DNA damage. Nature 447:941-
685 950.
686 de Visser JA. 2002. The fate of microbial mutators. Microbiology 148:1247-1252.
687 Deitsch KW, Moxon ER, Wellems TE. 1997. Shared themes of antigenic variation and virulence in
688 bacterial, protozoal, and fungal infections. Microbiol Mol Biol Rev 61:281-293.
689 Denamur E, Lecointre G, Darlu P, Tenaillon O, Acquaviva C, Sayada C, Sunjevaric I, Rothstein R, Elion
690 J, Taddei F, et al. 2000. Evolutionary implications of the frequent horizontal transfer of mismatch
691 repair genes. Cell 103:711-721.
692 Dupes NM, Walsh BW, Klocko AD, Lenhart JS, Peterson HL, Gessert DA, Pavlick CE, Simmons LA.
693 2010. Mutations in the Bacillus subtilis beta clamp that separate its roles in DNA replication from
694 mismatch repair. J Bacteriol 192:3452-3463.
695 Facklam R, Elliott J, Shewmaker L, Reingold A. 2005. Identification and characterization of sporadic
696 isolates of Streptococcus iniae isolated from humans. Journal of Clinical Microbiology 43:933-937.
697 Ferenci T. 2015. Trade-off Mechanisms Shaping the Diversity of Bacteria. Trends Microbiol.
698 Fogle CA, Nagle JL, Desai MM. 2008. Clonal interference, multiple mutations and adaptation in large
699 asexual populations. Genetics 180:2163-2173.
700 Foster PL. 2007. Stress-induced mutagenesis in bacteria. Crit Rev Biochem Mol Biol 42:373-397.
701 Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids 2010.
702 Galhardo RS, Hastings PJ, Rosenberg SM. 2007. Mutation as a stress response and the regulation of
703 evolvability. Crit Rev Biochem Mol Biol 42:399-435.
704 Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH. 2015.
705 Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev 28:871-899.
706 Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, Taddei F. 2001. Costs and benefits of high
707 mutation rates: Adaptive evolution of bacteria in the mouse gut. Science 291:2606-2608.
708 Gonzalez K, Faustoferri RC, Quivey RG, Jr. 2012. Role of DNA base excision repair in the mutability
709 and virulence of Streptococcus mutans. Molecular Microbiology 85:361-377.
710 Gould CV, Sniegowski PD, Shchepetov M, Metlay JP, Weiser JN. 2007. Identifying mutator
711 phenotypes among fluoroquinolone-resistant strains of Streptococcus pneumoniae using fluctuation
712 analysis. Antimicrob Agents Chemother 51:3225-3229.
713 Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome
714 assemblies. Bioinformatics 29:1072-1075.
715 Gutierrez O, Juan C, Perez JL, Oliver A. 2004. Lack of association between hypermutation and
716 antibiotic resistance development in Pseudomonas aeruginosa isolates from intensive care unit
717 patients. Antimicrob Agents Chemother 48:3573-3575.

718 Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation analysis CalculatOR: a web tool for the
719 determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics 25:1564-
720 1565.
721 Hall LM, Henderson-Begg SK. 2006. Hypermutable bacteria isolated from humans--a critical analysis.
722 Microbiology 152:2505-2514.
723 Hammerschmidt S, Wolff S, Hocke A, Rosseau S, Muller E, Rohde M. 2005. Illustration of
724 pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infection
725 and Immunity 73:4653-4667.
726 He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. 2016. Evolview v2: an online visualization and
727 management tool for customized and annotated phylogenetic trees. Nucleic Acids Res 44:W236-241.
728 Healey KR, Zhao Y, Perez WB, Lockhart SR, Sobel JD, Farmakiotis D, Kontoyiannis DP, Sanglard D, Taj-
729 Aldeen SJ, Alexander BD, et al. 2016. Prevalent mutator genotype identified in fungal pathogen
730 Candida glabrata promotes multi-drug resistance. Nat Commun 7:11128.
731 Hedges AJ. 2002. Estimating the precision of serial dilutions and viable bacterial counts. Int J Food
732 Microbiol 76:207-214.
733 Hegde A, Bhat GK, Mallya S. 2008. Effect of exposure to hydrogen peroxide on the virulence of
734 Escherichia coli. Indian J Med Microbiol 26:25-28.
735 Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: An interactive
736 viewer for large phylogenetic trees. BMC Bioinformatics 8:460.
737 Kadioglu A, Weiser JN, Paton JC, Andrew PW. 2008. The role of Streptococcus pneumoniae virulence
738 factors in host respiratory colonization and disease. Nat Rev Microbiol 6:288-301.
739 Kawasaki M, Delamare-Deboutteville J, Bowater RO, Walker MJ, Beatson S, Ben Zakour NL, Barnes
740 AC. 2018. Microevolution of aquatic Streptococcus agalactiae ST-261 from Australia indicates
741 dissemination via imported tilapia and ongoing adaptation to marine hosts or environment. Appl
742 Environ Microbiol.
743 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S,
744 Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for
745 the organization and analysis of sequence data. Bioinformatics 28:1647-1649.
746 Korkmaz G, Holm M, Wiens T, Sanyal S. 2014. Comprehensive analysis of stop codon usage in
747 bacteria and its correlation with release factor abundance. Journal of Biological Chemistry
748 289:30334-30342.
749 Labat F, Pradillon O, Garry L, Peuchmaur M, Fantin B, Denamur E. 2005. Mutator phenotype confers
750 advantage in Escherichia coli chronic urinary tract infection pathogenesis. FEMS Immunol Med
751 Microbiol 44:317-321.
752 Leache AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. 2015. Short Tree, Long Tree, Right
753 Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies. Systematic
754 Biology 64:1032-1047.
755 Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA. 2016. Mismatch repair in Gram-positive
756 bacteria. Res Microbiol 167:4-12.
757 Lenhart JS, Sharma A, Hingorani MM, Simmons LA. 2013. DnaN clamp zones provide a platform for
758 spatiotemporal coupling of mismatch detection to DNA replication. Mol Microbiol 87:553-568.
759 Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids
760 Res 46:D493-D496.
761 Li GM. 2008. Mechanisms and functions of DNA mismatch repair. Cell Research 18:85-98.
762 Locke JB, Colvin KM, Datta AK, Patel SK, Naidu NN, Neely MN, Nizet V, Buchanan JT. 2007.
763 Streptococcus iniae capsule impairs phagocytic clearance and contributes to virulence in fish. J
764 Bacteriol 189:1279-1287.
765 Locke JB, Colvin KM, Varki N, Vicknair MR, Nizet V, Buchanan JT. 2007. Streptococcus iniae beta-
766 hemolysin streptolysin S is a virulence factor in fish infection. Diseases of Aquatic Organisms 76:17-
767 26.

768 Losordo T, DeLong D, Guerdat T. 2009. Advances in technology and practice for land-based
769 aquaculture systems: tank-based recirculating systems for finfish production. New Technologies in
770 Aquaculture:945-983.
771 Lu AL, Li X, Gu Y, Wright PM, Chang DY. 2001. Repair of oxidative DNA damage: mechanisms and
772 functions. Cell Biochem Biophys 35:141-170.
773 Lukacisinova M, Novak S, Paixao T. 2017. Stress-induced mutagenesis: Stress diversity facilitates the
774 persistence of mutator genes. PLoS Comput Biol 13:e1005609.
775 Mackinnon MJ, Read AF. 2004. Virulence in malaria: an evolutionary viewpoint. Philos Trans R Soc
776 Lond B Biol Sci 359:965-986.
777 Maiden MCJ, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013. MLST
778 revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Micro 11:728-736.
779 Mao EF, Lane L, Lee J, Miller JH. 1997. Proliferation of mutators in A cell population. J Bacteriol
780 179:417-422.
781 Matsushima A, Takakura S, Fujihara N, Saito T, Ito I, Iinuma Y, Ichiyama S. 2010. High prevalence of
782 mutators among Enterobacter cloacae nosocomial isolates and their association with antimicrobial
783 resistance and repetitive detection. Clin Microbiol Infect 16:1488-1493.
784 Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of
785 bacterial pathogens. FEMS Microbiol Lett 267:1-8.
786 Mena A, Macia MD, Borrell N, Moya B, de Francisco T, Perez JL, Oliver A. 2007. Inactivation of the
787 mismatch repair system in Pseudomonas aeruginosa attenuates virulence but favors persistence of
788 oropharyngeal colonization in cystic fibrosis mice. J Bacteriol 189:3665-3668.
789 Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL, Oliver A. 2008. Genetic Adaptation
790 of Pseudomonas aeruginosa to the Airways of Cystic Fibrosis Patients Is Catalyzed by Hypermutation.
791 J Bacteriol 190:7910-7917.
792 Merino D, Reglier-Poupet H, Berche P, Charbit A, European Listeria Genome C. 2002. A hypermutator
793 phenotype attenuates the virulence of Listeria monocytogenes in a mouse model. Molecular
794 Microbiology 44:877-887.
795 Methot PO, Alizon S. 2014. What is a pathogen? Toward a process view of host-parasite interactions.
796 Virulence 5:775-785.
797 Metzgar D, Wills C. 2000. Evidence for the adaptive evolution of mutation rates. Cell 101:581-584.
798 Millard CM, Baiano JC, Chan C, Yuen B, Aviles F, Landos M, Chong RS, Benedict S, Barnes AC. 2012.
799 Evolution of the capsular operon of Streptococcus iniae in response to vaccination. Appl Environ
800 Microbiol 78:8219-8226.
801 Miller JH. 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair.
802 Annu Rev Microbiol 50:625-643.
803 Negri MC, Morosini MI, Baquero MR, del Campo R, Blazquez J, Baquero F. 2002. Very low cefotaxime
804 concentrations select for hypermutable Streptococcus pneumoniae populations. Antimicrob Agents
805 Chemother 46:528-530.
806 Neher RA, Hallatschek O. 2013. Genealogies of rapidly adapting populations. Proc Natl Acad Sci U S A
807 110:437-442.
808 Nilsson AI, Kugelberg E, Berg OG, Andersson DI. 2004. Experimental adaptation of Salmonella
809 typhimurium to mice. Genetics 168:1119-1130.
810 Oliver A. 2010. Mutators in cystic fibrosis chronic lung infection: Prevalence, mechanisms, and
811 consequences for antimicrobial therapy. Int J Med Microbiol 300:563-572.
812 Oliver A, Mena A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance.
813 Clinical Microbiology and Infection 16:798-808.
814 Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G. 2010. Interactive microbial genome
815 visualization with GView. Bioinformatics 26:3125-3126.
816 Picard B, Duriez P, Gouriou S, Matic I, Denamur E, Taddei H. 2001. Mutator natural Escherichia coli
817 isolates have an unusual virulence phenotype. Infection and Immunity 69:9-14.

818 Pier GB, Madin SH. 1976. Streptococcus iniae sp nov, a beta-hemolytic Streptococcus isolated from
819 an amazon freshwater dolphin, Inia geoffrensis. . International Journal of Systematic Bacteriology
820 26:545-553.
821 Pier GB, Madin SH, Alnakeeb S. 1978. Isolation and characterisation of a 2nd isolate of Streptococcus
822 iniae. International Journal of Systematic Bacteriology 28:311-314.
823 Rajagopal L. 2009. Understanding the regulation of Group B Streptococcal virulence factors. Future
824 Microbiol 4:201-221.
825 Rajanna C, Ouellette G, Rashid M, Zemla A, Karavis M, Zhou C, Revazishvili T, Redmond B, McNew L,
826 Bakanidze L, et al. 2013. A strain of Yersinia pestis with a mutator phenotype from the Republic of
827 Georgia. FEMS Microbiol Lett 343:113-120.
828 Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of
829 heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution 2.
830 Raynes Y, Sniegowski PD. 2014. Experimental evolution and the dynamics of genomic mutation rate
831 modifiers. Heredity (Edinb) 113:375-380.
832 Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft
833 genomes using the Mauve Aligner. Bioinformatics 25:2071-2073.
834 Rodriguez JL, Dalia AB, Weiser JN. 2012. Increased chain length promotes pneumococcal adherence
835 and colonization. Infection and Immunity 80:3454-3459.
836 Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations. Methods 20:4-17.
837 Scott J, Thompson-Mayberry P, Lahmamsi S, King CJ, McShan WM. 2008. Phage-associated mutator
838 phenotype in group A streptococcus. J Bacteriol 190:6290-6301.
839 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.
840 Shabalina SA, Spiridonov NA, Kashina A. 2013. Sounds of silence: synonymous nucleotides as a key to
841 biological regulation and complexity. Nucleic Acids Res 41:2073-2094.
842 Shaver AC, Sniegowski PD. 2003. Spontaneously arising mutL mutators in evolving Escherichia coli
843 populations are the result of changes in repeat length. J Bacteriol 185:6076-6082.
844 Shrestha S, Bjornstad ON, King AA. 2014. Evolution of acuteness in pathogen metapopulations:
845 conflicts between "classical" and invasion-persistence trade-offs. Theor Ecol 7:299-311.
846 Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for
847 bacterial insertion sequences. Nucleic Acids Research 34:D32-D36.
848 Smania AM, Segura I, Pezza RJ, Becerra C, Albesa I, Argarana CE. 2004. Emergence of phenotypic
849 variants upon mismatch repair disruption in Pseudomonas aeruginosa. Microbiology-Sgm 150:1327-
850 1338.
851 Sniegowski P, Raynes Y. 2013. Mutation Rates: How Low Can You Go? Current Biology 23:R147-R149.
852 Sprouffske K, Aguilar-Rodriguez J, Sniegowski P, Wagner A. 2018. High mutation rates limit
853 evolutionary adaptation in Escherichia coli. Plos Genetics 14:e1007324.
854 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
855 phylogenies. Bioinformatics 30:1312-1313.
856 Stich M, Manrubia SC, Lazaro E. 2010. Variable Mutation Rates as an Adaptive Strategy in Replicator
857 Populations. PLoS One 5.
858 Sundin GW, Weigand MR. 2007. The microbiology of mutability. FEMS Microbiol Lett 277:11-20.
859 Symonds MRE, Blomberg SP. 2014. A Primer on Phylogenetic Generalised Least Squares. In:
860 Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in
861 Evolutionary Biology: Concepts and Practice. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 105-
862 130.
863 Taddei F, Radman M, MaynardSmith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator
864 alleles in adaptive evolution. Nature 387:700-702.
865 Tham KC, Hermans N, Winterwerp HHK, Cox MM, Wyman C, Kanaar R, Lebbink JHG. 2013. Mismatch
866 Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA
867 Structures. Mol Cell 51:326-337.

868 Torres-Barcelo C, Cabot G, Oliver A, Buckling A, Maclean RC. 2013. A trade-off between oxidative
869 stress resistance and DNA repair plays a role in the evolution of elevated mutation rates in bacteria.
870 Proc Biol Sci 280:20130007.
871 Torres-Barcelo C, Kojadinovic M, Moxon R, MacLean RC. 2015. The SOS response increases bacterial
872 fitness, but not evolvability, under a sublethal dose of antibiotic. Proc Biol Sci 282:20150885.
873 Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome
874 alignment and visualization of thousands of intraspecific microbial genomes. Genome Biology
875 15:524.
876 Turrientes MC, Baquero F, Levin BR, Martinez JL, Ripoll A, Gonzalez-Alba JM, Tobes R, Manrique M,
877 Baquero MR, Rodriguez-Dominguez MJ, et al. 2013. Normal mutation rate variants arise in a Mutator
878 (Mut S) Escherichia coli population. PLoS One 8:e72963.
879 Umarov RK, Solovyev VV. 2017. Recognition of prokaryotic and eukaryotic promoters using
880 convolutional deep learning neural networks. PLoS ONE 12.
881 V. Solovyev AS. 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In.
882 Metagenomics and its Applications in Agriculture, Biomedicine and Environmental: Nova Science
883 Publishers. p. 61-78.
884 Verotta D, Haagensen J, Spormann AM, Yang K. 2017. Mathematical Modeling of Biofilm Structures
885 Using COMSTAT Data. Comput Math Methods Med 2017:7246286.
886 Wang S, Wang Y, Shen J, Wu Y, Wu C. 2013. Polymorphic mutation frequencies in clinical isolates of
887 Staphylococcus aureus: the role of weak mutators in the development of fluoroquinolone resistance.
888 FEMS Microbiol Lett 341:13-17.
889 Weinberger DM, Malley R, Lipsitch M. 2011. Serotype replacement in disease after pneumococcal
890 vaccination. Lancet 378:1962-1973.
891 Weinstein MR, Litt M, Kertesz DA, Wyper P, Rose D, Coulter M, McGeer A, Facklam R, Ostach C,
892 Willey BM, et al. 1997. Invasive infections due to a fish pathogen, Streptococcus iniae. New England
893 Journal of Medicine 337:589-594.
894 Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE,
895 Schneider D. 2011. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution
896 Experiment With Escherichia coli. G3 (Bethesda) 1:183-186.
897 Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B,
898 Medigue C, Lenski RE, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect
899 tension between adaptation and genetic load. Proc Natl Acad Sci U S A 110:222-227.
900 Wrande M, Roth JR, Hughes D. 2008. Accumulation of mutants in "aging" bacterial colonies is due to
901 growth under selection, not stress-induced mutagenesis. Proc Natl Acad Sci U S A 105:11863-11868.
902 Yang H, Miller JH. 2008. Deletion of dnaN1 generates a mutator phenotype in Bacillus anthracis. DNA
903 Repair (Amst) 7:507-514.
904 Zheng Q. 2015. Methods for comparing mutation rates using fluctuation assay data. Mutat Res
905 777:20-22.
906 Zheng Q. 2017. rSalvador: An R Package for the Fluctuation Experiment. G3 (Bethesda) 7:3849-3856.

907

908

909 **Supporting Information Legends**

910 Supplementary Figure 1. *Streptococcus iniae* pangenome derived from 80 strains, coloured by clade
911 (clade A-G innermost - outer) showing positions of MMR and OG genes relative to putative
912 phage/prophage and insertion sequence elements. Phage positions were predicted using Phaster and are
913 indicated as arcs in the outermost ring (green, complete phage; blue, incomplete phage; red, uncertain).
914 ISE were located with ISfinder and are indicated directionally in the outermost ring as black arrows.

915 Supplementary Figure 2. Root-to-tip regression analysis (TempEst, (Rambaut, et al. 2016)of branch
916 lengths from best-fit-rooted tree against time (year of isolation). The tree was derived from alignment of
917 non-recombinant core-genome SNPs corrected for ascertainment bias in RAxML.

918 Supplementary Figure 3: Alignment of the sequences found upstream from *mutY* gene in QMA0084
919 (representative of clade B) and QMA0248 (representative of all other analysed *S. iniae* strains) and
920 predicted changes in transcription pattern produced by 56 bp deletion upstream of *mutY* gene identified
921 in clade B strains.

922 Supplementary Table 1. 80-by-80 matrix of *p* values resulting from pair-wise comparisons of
923 fluctuation analysis data by Maximum Ratio Test corrected for multiple comparisons by controlling for
924 the False Discovery Rate (Microsoft Excel spreadsheet available as download).

925 Supplementary Table 2. Results of fluctuation analysis (Microsoft Excel spreadsheet available as
926 download).

927 Supplementary Table 3. Extended metadata for 80 *Streptococcus iniae* isolates sequenced in the present
928 study (Microsoft Excel spreadsheet)

929 Supplementary Table 4. Assembly statistics and quality data for 80 *de novo* genome assemblies prepared
930 in this study (Microsoft Excel spreadsheet).

931 **Figures and tables**

932 Figure 1: Unrooted summary phylogram of *Streptococcus iniae* strains based on alignment of core
933 genome SNPs filtered to remove recombination. Genetic distances were inferred by maximum likelihood
934 in RAXML and node support is indicated as percentage of 1000 bootstrap replicates. Geographic origin,
935 time range (year) of isolation, and mutation rate were added manually *post hoc*. All nodes with bootstrap
936 support below 75% were collapsed. The phylogram is drawn with scale proportionate to genetic distance
937 except the dashed branch supporting QMA0141, which is depicted at 10% scale. The inset shows the
938 phylogram structure with QMA0141 branch drawn to the same scale.

939 Figure 2: Cladogram of *Streptococcus iniae* strains derived from the same dataset as Figure 1. Clade, host
940 species, variants in MMR and OG DNA repair genes, mutation rate phenotype, and phenotypic variants
941 associated with virulence were annotated with Evolview. Mutation rate (μ) phenotypes, determined by
942 fluctuation analysis, are depicted as box blots representing upper and lower 95% confidence limits, with
943 the interface representing mean values. Node support is indicated as percentage of 1000 bootstrap
944 replicates.

945 Figure 3: Phenotypic variants associated with virulence among *Streptococcus iniae* strains. A) Buoyant
946 density assay for capsule (CPS) presence, A1: CPS+, A2 CPS- A3: CPS+/-. B) Haemolysis of sheep red blood
947 cells by agar diffusion B1: +, B2:-. C) Streptococcal chain length C1: normal, C2: Long, C3: short. D) Biofilm
948 formation in chamber slides determined by confocal microscopy, D1: normal, D2: denser, D3: thicker.

949

950 Table 1. *Streptococcus iniae* strains used in this study. Includes origin details (host species, time, site
951 of isolation), phylogenetic affiliation, virulence-associated phenotypes, and mutation rate. Atypical
952 places of isolation (hosts, 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	***Pteropus alecto***	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

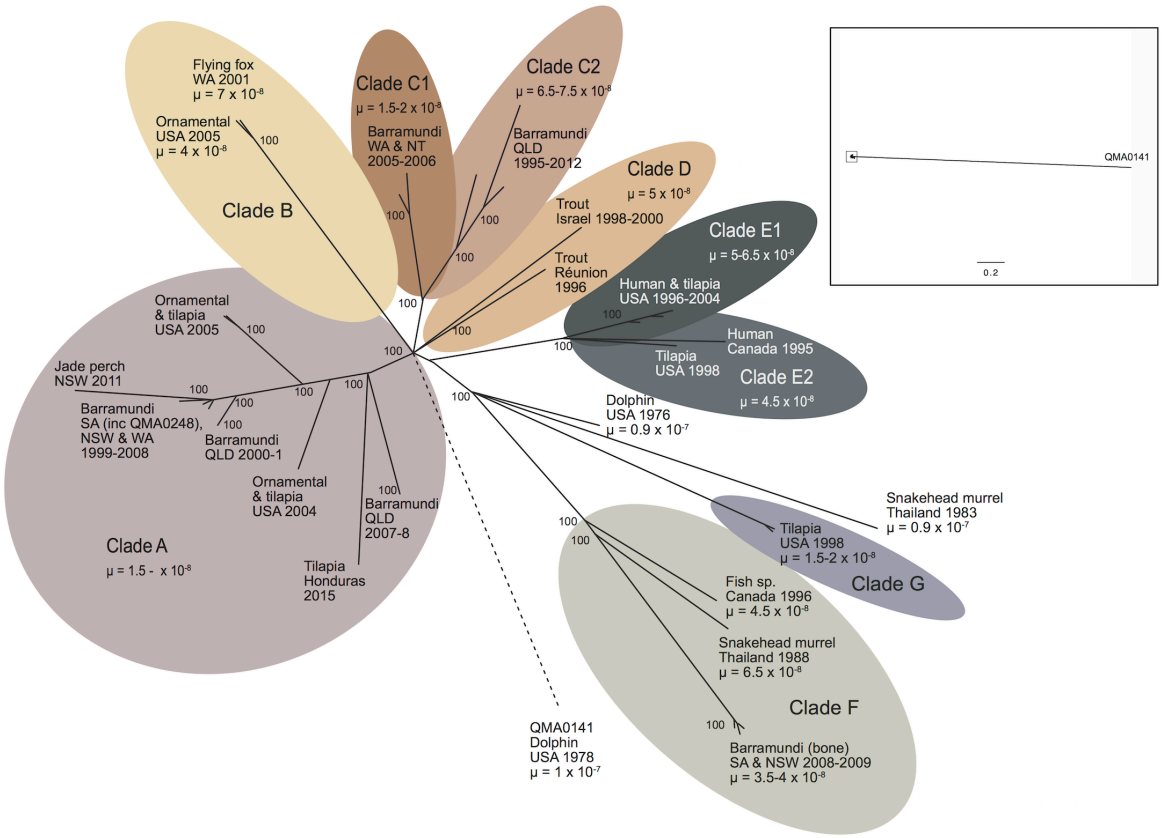
955 Table 2: Variants in MMR and OG genes found among *S. iniae* isolates. CDS – protein coding sequence,
956 BS – predicted binding sites of transcription factors, PROVEAN score – predicted effect of amino acid
957 substitutions on the protein function; scores lower that -2 are considered to indicate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

959

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	TAA->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)	atcataagatttcctcctat aaatctgtgtcatatgtcat ccttctttatagtttc 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

960

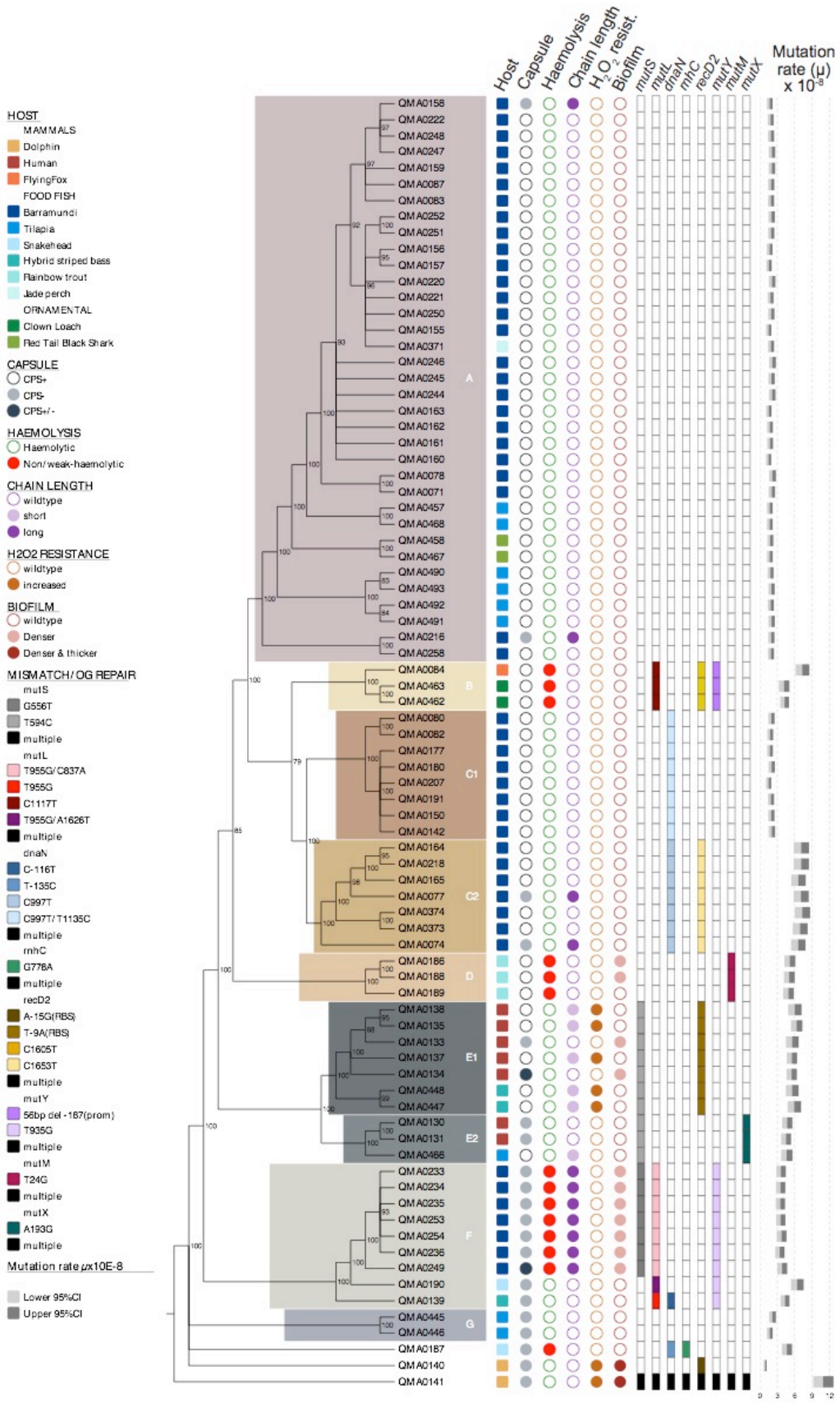
961

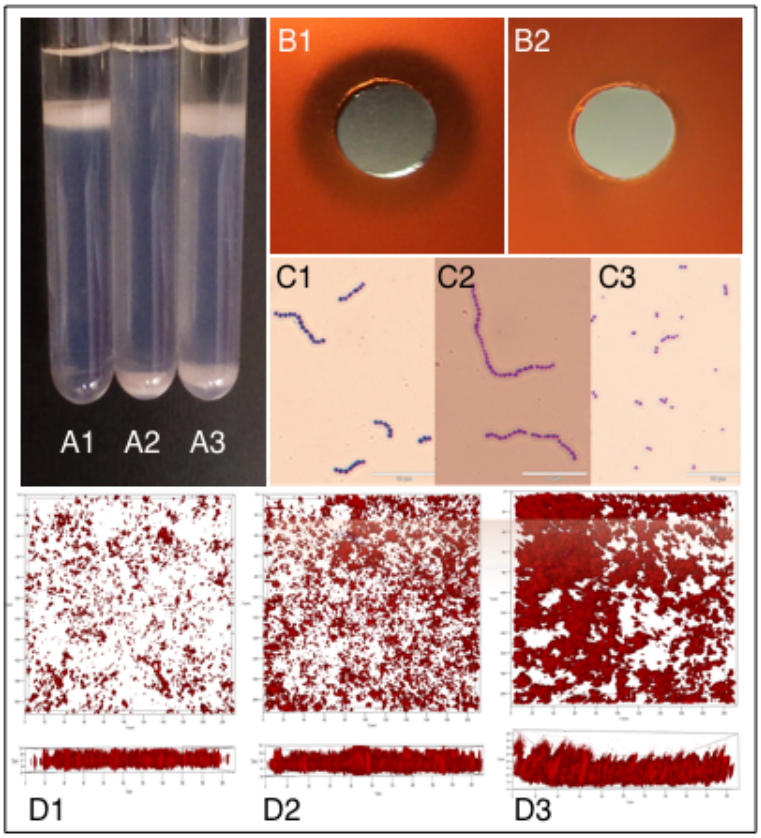

Flying fox
WA 2001
μ = 7 x 10-8
Ornamental
USA 2005
μ = 4 x 10-8
Clade B
Clade C1
μ = 1.5-2 x 10-8
Barramundi
WA & NT
2005-2006
Clade C2
μ = 6.5-7.5 x 10-8
Barramundi
QLD
1995-2012
Clade D
μ = 5 x 10-8
Trout
Israel 1998-2000
Trout
Réunion
1996
Clade E1
μ = 5-6.5 x 10-8
Human & tilapia
USA 1996-2004
Human
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

962

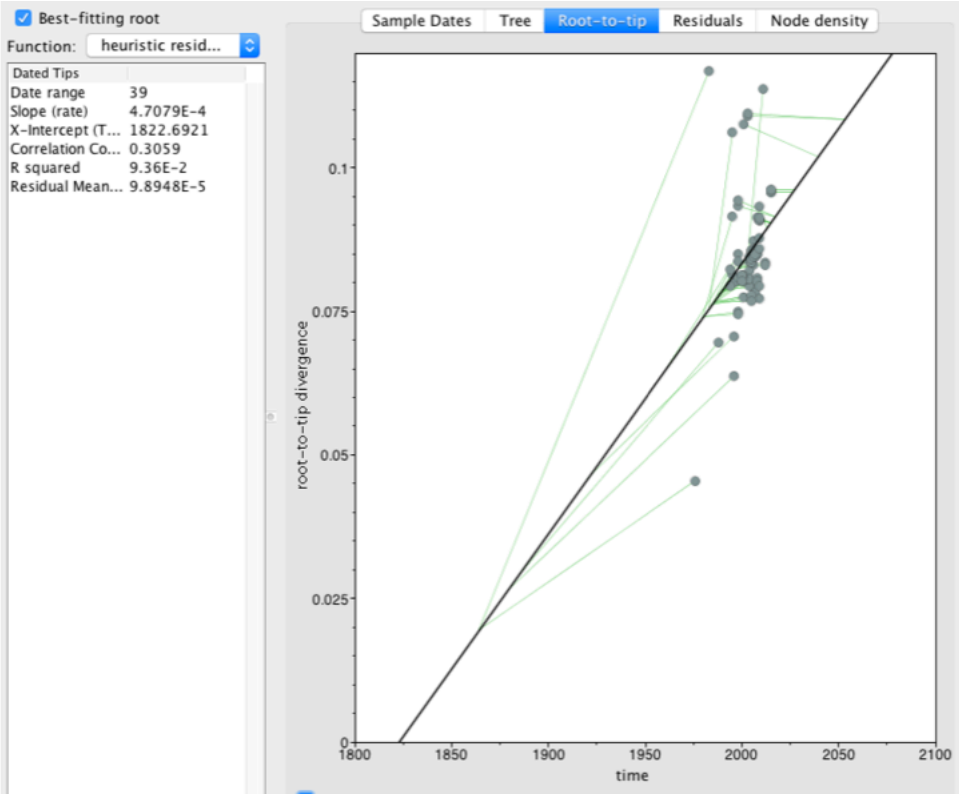
963

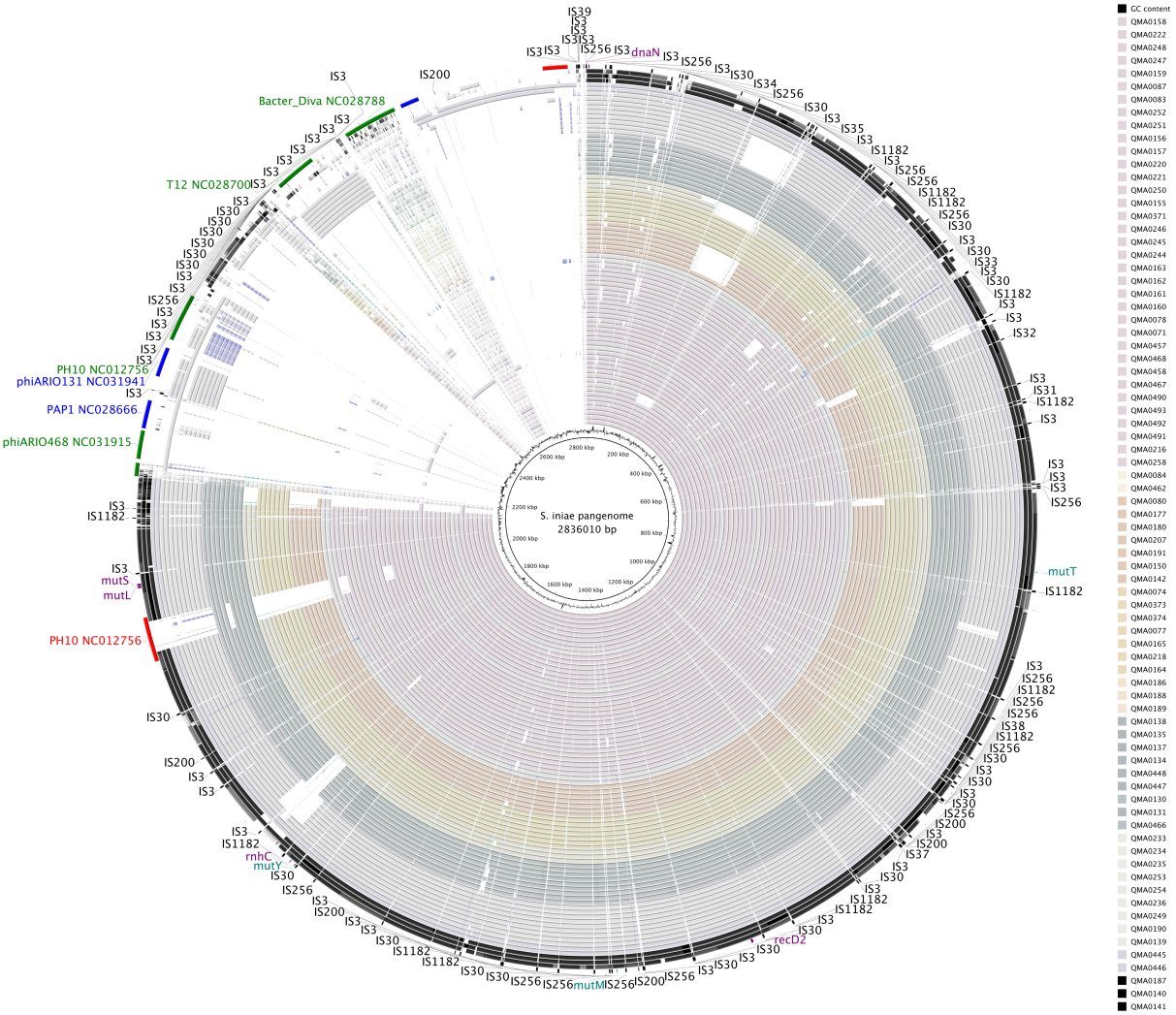
964


Host
Capsule
Haemolysis
Chain length
H_2O_2 resist.
Biofilm
mutS
mutL
dnaN
rnhC
recD2
mutY
mutM
mutX
Mutation rate (μ) x 10^{-8}
HOST
MAMMALS
Dolphin
Human
FlyingFox
FOOD FISH
Barramundi
Tilapia
Snakehead
Hybrid striped bass
Rainbow trout
Jade perch
ORNAMENTAL
Clown Loach
Red Tail Black Shark
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
G776A
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 μx10E-8
Lower 95%CI
Upper 95%CI
A
B
C1
C2
D
E1
E2
F
G
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
QMA0084
QMA0463
QMA0462
QMA0080
QMA0082
QMA0177
QMA0180
QMA0207
QMA0191
QMA0150
QMA0142
QMA0164
QMA0218
QMA0165
QMA0077
QMA0374
QMA0373
QMA0074
QMA0186
QMA0188
QMA0189
QMA0138
QMA0135
QMA0133
QMA0137
QMA0134
QMA0448
QMA0447
QMA0130
QMA0131
QMA0466
QMA0233
QMA0234
QMA0235
QMA0253
QMA0254
QMA0236
QMA0249
QMA0190
QMA0139
QMA0445
QMA0446
QMA0187
QMA0140
QMA0141
0
3
6
9
12


A1
A2
A3
B1
B2
C1
C2
C3
D1
D2
D3

Supplementary Figures




S. iniae pangenome
2836010 bp
200 kbp
400 kbp
600 kbp
800 kbp
1000 kbp
1200 kbp
1400 kbp
1600 kbp
1800 kbp
2000 kbp
2200 kbp
2400 kbp
2600 kbp
2800 kbp
Bacter_Diva NC028788
T12 NC028700
PH10 NC012756
phiARIO131 NC031941
PAP1 NC028666
phiARIO468 NC031915
PH10 NC012756
dnaN
mutT
mutS
mutL
rnhC
mutY
mutM
recD2
GC content
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
QMA0084
QMA0462
QMA0080
QMA0177
QMA0180
QMA0207
QMA0191
QMA0150
QMA0142
QMA0074
QMA0373
QMA0374
QMA0077
QMA0165
QMA0218
QMA0164
QMA0186
QMA0188
QMA0189
QMA0138
QMA0135
QMA0137
QMA0134
QMA0448
QMA0447
QMA0130
QMA0131
QMA0466
QMA0233
QMA0234
QMA0235
QMA0253
QMA0254
QMA0236
QMA0249
QMA0190
QMA0139
QMA0445
QMA0446
QMA0187
QMA0140
QMA0141

```
QMA248_500      cagattatgaataactaatctctggaaaaattagtctccattcttgtaacgtcctttatt   60
QMA0084_444     cagattatgaataactaatctctggaaaaattagtctccattcttgtaacgtcctttatt   60
                ************************************************************

QMA248_500      gaatttaagcatatactatacctttcaatattggagattttccaaaacttatactgtaaa   120
QMA0084_444     gaatttaagcatatactatacctttcaatattggagattttccaaaacttatactgtaaa   120
                ************************************************************

QMA248_500      gagcgctcctttggccaaataacgcatacttcaataagttcaacaatgtgatcactttgt   180
QMA0084_444     gagcgctcctttggccaaataacgcatacttcaataagttcaacaatgtgatcactttgt   180
                ************************************************************

QMA248_500      tagttgtaaaagtcagtttgcctaccgttgttggagctacttagtctccgttcttgtacc   240
QMA0084_444     tagttgtaaaagtcagtttgcctaccgttgttggagctacttagtctccgttcttgtacc   240
                ************************************************************

QMA248_500      gtgcgacatgtagtttcatcataagatttcctcctataaatctgtgtcatatgtcatcct   300
QMA0084_444     gtgcgacatgtagtttc-------------------------------------------   257
                *****************

QMA248_500      tctttatagtttcattataatgtaagcgtttgcttttgtcaagaaaaaatatttccctat   360
QMA0084_444     -------------attataatgtaagcgtttgcttttgtcaagaaaaaatatttccctat   304
                              ***********************************************

QMA248_500      gacattttccacttgcctaataaaaagataagaaaaaaaagctgttatcattgtttttg   420
QMA0084_444     gacattttccacttgcctaataaaaagataagaaaaaaaagctgttatcattgtttttg   364
                ************************************************************

QMA248_500      aaaactctgatagaagctattattttgactttctagcctagccttttcgcatattgaaaa   480
QMA0084_444     aaaactctgatagaagctattattttgactttctagcctagccttttcgcatattgaaaa   424
                ************************************************************

QMA248_500      aactttgctacaatagaggt   500
QMA0084_444     aactttgctacaatagaggt   444
                ********************
```

BROM prediction:

> QMA248_500

Length of sequence- 500
Threshold for promoters - 0.20

Number of predicted promoters - 1

Promoter Pos: 328 LDF- 7.23
10 box at pos. 313 cattataat Score 75
35 box at pos. 291 atgtca Score 23

Oligonucleotides from known TF binding sites:

For promoter at 328:
rpoD17: TTCCTCCT at position 268 Score - 8
rpoD17: GTGTCATA at position 284 Score - 7
rpoD17: TAATGTAA at position 318 Score - 12
galR: TGTAAGCG at position 321 Score - 8
rpoD18: TTGCTTTT at position 330 Score - 6
rpoD16: TGTCAAGA at position 337 Score - 19

> QMA0084_444

Length of sequence- 444
Threshold for promoters - 0.20

Number of predicted promoters - 2

Promoter Pos: 362 LDF- 6.79
10 box at pos. 347 tgttatcat Score 75
35 box at pos. 324 ataaaa Score 7

Promoter Pos: 62 LDF- 3.55
10 box at pos. 44 ttgtaacgt Score 38
35 box at pos. 23 tggaaa Score 23

Oligonucleotides from known TF binding sites:

For promoter at 362:
arcA: AATAAAAA at position 323 Score - 12
purR: ATAAAAAG at position 324 Score - 11
cpxR: TAAAAAGA at position 325 Score - 9
rpoD18: ATCATTGT at position 351 Score - 5
rpoD19: ATTGTTTT at position 354 Score - 7
fnr: TTTTTTGA at position 358 Score - 9

For promoter at 62:
lrp: ATAACTAA at position 11 Score - 14
tus: TAACTAAT at position 12 Score - 17
tus: TTGTAACG at position 44 Score - 7
rpoD17: ATACTATA at position 73 Score - 11

975