Article

Mutators as drivers of adaptation in *Streptococcus* and a risk factor for host jumps and vaccine escape

Running Title: Mutators as drivers of pathogen evolution

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. Mutators deficient in mismatch (MMR) and oxidised guanine (OG) repair are abundant in clinical samples and show increased adaptive potential in experimental infection models but their role in pathoadaptation is poorly understood. Here we investigate the role of mutators in epidemiology and evolution of the broad host pathogen, *Streptococcus iniae*, employing 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*, *mutX*) genes. *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.

Importance

Host immune response is a powerful selective pressure that drives diversification of pathogenic microorganisms and, ultimately, evolution of new strains. Major adaptive events in pathogen evolution, such as transmission to a new host species or infection of vaccinated hosts, require adaptation to a drastically different immune landscape. Such adaptation may be favoured by hypermutable strains (or mutators) that are defective in normal DNA repair and consequently capable of generating multiple potentially beneficial and compensatory mutations. This permits rapid adjustment of virulence and antigenicity in a new immunological setting. Here we show that mutators, through mutations in DNA repair genes and corresponding shifts in mutation rate, are associated with major diversification events and virulence evolution in the broad host-range pathogen *Streptococcus iniae*. We show that mutators underpin infection of vaccinated hosts, transmission to new host species and the evolution of new strains.

48

49 **Introduction**

50 Mutation rates in microbial populations are a trade-off between maintenance of genetic integrity and
51 evolvability (1, 2). High fidelity of DNA repair is advantageous in adapted populations since most
52 mutations lead to deviation from the evolved phenotype (2, 3). Conversely, lower fidelity of DNA repair
53 can be advantageous under stress since increased mutation supply promotes diversity and may
54 accelerate adaptation (2, 4). Bacteria can efficiently fluctuate between high and low mutation regimes:
55 While extremely low mutation rates are maintained in favourable conditions (5), elevation of mutation
56 rates known as stress-induced mutagenesis (SIM) occurs in adapting populations (6, 7). One aspect of
57 SIM is temporary hypermutability associated with changes in expression of multiple genes including
58 error-prone DNA polymerases, recombination-preventing enzymes and enzymes involved in movement
59 of mobile elements (8). For example transient hypermutability during RecA-mediated general stress
60 response, SOS, is associated with up-regulation of error-prone polymerases that can by-pass DNA
61 lesions during replication (8). However, increased fitness during SOS response stems from immediate
62 growth advantage rather than from long-term evolution of stress resistance (9-11). Another aspect of
63 SIM is the evolution of heritable mutator phenotypes, also referred to as constitutive, permanent, or
64 true mutators, that contain mutator alleles and replicate over multiple generations with a high amount
65 of error (6, 7). Heritably hypermutable strains, which periodically arise within a population via
66 mutations in DNA repair genes (12, 13), are selected under stress indirectly via 'hitchhiking' of mutator
67 alleles with stress-resistance mutations (14, 15), or by undefined pleiotropic effects conferred by
68 inactivation of DNA repair (16, 17). There is much stronger theoretical and experimental evidence
69 demonstrating that heritable mutators facilitate adaptive evolution (17). However, even heritable
70 mutators are transient phenomena in natural populations as non-mutator variants are selected within
71 already adapted mutator clones (18, 19). Restoration of mutation rate can occur via copy-number
72 variation leading to full reversion of the mutator allele (20), disruption of the linkage between mutator
73 alleles and beneficial mutations via recombination (21), anti-mutator compensatory mutations (4, 19),
74 or prophage excision when its integration confers increase in mutation rate (22).

75 Bacterial mutators are abundant among clinical isolates, especially in chronic cases (23, 24). Multiple
76 studies have attributed this to antibiotic selection (25-27), since mutators are a well-established risk
77 factor for antimicrobial resistance (28, 29). However, the association between frequency of mutators in
78 diagnostic samples and antibiotic treatments is weak (30, 31), and isogenic knockout mutator strains
79 have increased ability to colonize mice in challenge models devoid of antibiotic exposure (32-34). Thus,
80 the prevalence of mutators in clinical isolates most likely results from the selective process during

81 adaptation to the host (30, 35-37), where immunity acts as the ultimate selective pressure shaping the
82 diversity of pathogenic strains (38).

83 The 'immunological niche' concept states that, due to the diversity and complexity of immune
84 responses, every individual host represents a unique habitat that requires adaptation (39). As a
85 pathogen needs to adapt to every new host, the host population represents an extremely
86 heterogeneous environment where adaptive processes are ongoing and persistence of mutator alleles
87 may be favoured (40). Moreover, as a small number of infection units may enter a host, pathogen
88 adaptation is intrinsically associated with population bottlenecks during which the rate at which
89 beneficial variants arise might be crucial (41, 42). This supports the contention that bacterial evolution
90 is more likely to occur via multiple mutations rather than a single beneficial variant (43).

91 The fitness of a pathogenic strain in the host hinges on the complex, dynamic trait of virulence,
92 determined by a multitude of virulence factors and host responses (44). Optimal virulence is a trade-off
93 between acuteness and persistence with higher virulence facilitating extraction of resources and
94 transmission, but too much damage can kill the host and stop transmission altogether (45).
95 Additionally, virulence factors are often antigens so reduced virulence may facilitate immune evasion
96 (46). Consequently, both loss and gain of virulence can be advantageous and indicate adaptive shifts
97 towards acute or chronic pathogenesis respectively (47, 48). Exacerbation and attenuation of virulence
98 are repeatedly observed in both knockout and naturally occurring mutators (36, 49-53), thus mutator
99 strains/alleles may help pathogenic bacteria to evolve towards optimal virulence.

100 Most bacterial mutators are deficient in genes from mismatch (MMR) and oxidised guanine (OG) DNA
101 repair systems (23). MMR and OG genes are conserved across all domains of life (54), and single
102 nucleotide polymorphisms in their sequence may drastically alter mutation rate (55, 56). Nonetheless,
103 these genes are variable in major bacterial pathogens, implicating mutation rate dynamics in the
104 evolution of virulence (57, 58). MMR is coupled to replication and removes mispaired nucleotides as
105 well as short insertion/deletion loops, and prevents recombination between divergent sequences (59,
106 60). The MMR pathway in *Escherichia coli* is well-defined, in which MutS dimer protein interacts with
107 the replication processivity clamp DnaN (b-clamp), binds to the mismatch, recruits MutL, and the
108 MutS:MutL complex then activates the MutH endonuclease (54). MutH cuts the unmethylated nascent
109 strand at hemimethylated d(GATC) sites. MutL loads the UvrD helicase that unwinds the nicked DNA
110 containing the error, which is followed by excision by one of the single-strand nucleases (54). This MMR
111 organization, however, appears to be limited to *E. coli* and closely related Gammaproteobacteria (54,
112 61). In most bacteria and eukaryotes, MutH is absent and MutL has a conserved endonuclease site. dam
113 methylase (acting at GATC sites) is absent and the strand discrimination signal is unknown (54, 61). In

114 *Bacillus subtilis,* a model species for gram-positive bacteria, nicks created by ribonuclease *rnhC* during
115 removal of misincorporated ribonucleosides direct MutL endonuclease towards the newly synthethised
116 strand, and recD2 acts as a MMR helicase instead of uvrD (61). The OG system repairs oxidative DNA
117 damage throughout the cell cycle (62, 63). Among all DNA bases, guanine is most sensitive to reactive
118 oxygen species and its oxidised form, 8-oxo-dG or OG, is highly mutagenic as it preferentially pairs with
119 adenine which, if not corrected, leads to G:C to A:T transversion on a second round of replication (62,
120 63). OG system organization is conserved in most bacteria and eukaryotes and generally MutT
121 homologs (eg. MutX in *S. pneumoniae*) remove 8-oxo-dG from the nucleotide pool, MutM homologs
122 excise 8-oxo-dG incorporated into DNA, and MutY homologs remove adenine from OG:A mispairings
123 (62, 63).

124 We investigate the role of mutators and mutation rate dynamics in the epidemiology and evolution of
125 the broad host-range pathogen *Streptococcus iniae*. *S. iniae* is a highly adaptable species, evidenced by
126 its global distribution, ability to cross divergent host taxa boundaries and repeated re-infection of
127 previously immunised hosts during disease outbreaks (64, 65). Moreover, Streptococcal pathogens in
128 general are difficult candidates for immunisation in humans and animals due to high strain diversity
129 evidently driven by immunisation, yet there is little information on the role of mutators in this
130 evolutionary process (66, 67). We determine variation in mutation rate and identify mutations in MMR
131 (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*) and OG (*mutY* , *mutM*, *mutX*) genes among 80 diverse strains of *S.*
132 *iniae.* We discover that both mutation rates and mutations in MMR and OG genes correlate with
133 phylogenetic diversification but are highly conserved within phylogenetic clades. To determine whether
134 variation in mutation rate phenotype and MMR/OG genotype is linked to adaptation to the host
135 immunity, we identify variation in major phenotypic traits that determine virulence in streptococci
136 (capsular polysaccharide, hemolysin, length of cell chains) and bacteria in general (resistance to reactive
137 oxygen species (ROS), and biofilm formation). The expression of these traits among *S. iniae* isolates is
138 highly consistent with phylogenetic affiliation and strongly associated with variation in MMR and OG
139 genes and mutation rate phenotype. This supports the contention that strain diversification driven by
140 adaptation to the host achieved via adjustment of virulence is facilitated by the mutator strains.
141 Furthermore, we find that most changes in virulence determinants and MMR and OG genes occur in
142 strains producing infection in unusual immunological landscapes such as mammals and bone tissue of
143 immunised fish, indicating that mutators may facilitate the reinfection of immunised animals and
144 transmission between divergent host species.

145

146 **Results and Discussion**

147 We investigated the link between mutation rate genotype and phenotype, and phylogenetic
148 diversification among *Streptococcus iniae* strains isolated globally between 1976 and 2016 from
149 different host taxa. Maximum likelihood phylogenetic analysis of 80 *S. iniae* isolates (Table 1) based on
150 non-recombinant core genome SNPs derived from whole genome data resolved six major clades (A-F),
151 one lineage with two strains (clade G), and three lineages with a single strain (Figs. 1,2). Clades C-E
152 show a degree of geographic endemism while clades A, B, and F contain strains from diverse geographic
153 regions. Principally, variation in MMR and GO genes occurs only between the lineages, and they are
154 highly conserved within clades (Fig. 2, Table 2). In fact, sequences of the major MMR and OG genes are
155 conserved even within phylogenetic groups not readily definable by geographic origin, time, or host
156 species, and may be useful for Multilocus Sequence Typing (MLST) (68). Although even synonymous
157 SNPs in any gene locus can affect mRNA and have pronounced phenotypic consequences (69), most
158 SNPs in MMR and OG genes were identified in protein functional domains, and most amino acid
159 substitutions were predicted to have a deleterious effect on protein function (Table 2). Furthermore,
160 the number of MMR and OG variants correlates significantly with mutation rate (p = 0.0056) and with
161 the number of atypical phenotypic traits contributing to virulence (p = 0.0331), as predicted by a
162 Phylogenetic Generalised Least Squares (PGLS) regression model that accounts for autocorrelation
163 occurring between closely related strains (70). In turn, mutation rate variation is highly consistent with
164 phylogenetic affiliation: According to multiple pair-wise strain comparisons by Maximum Likelihood
165 Ratio test (71), with probabilities corrected for multiple comparison using False Discovery Rate (72),
166 mutation rate difference is insignificant within but significant between major phylogenetic lineages
167 (Suppl. Table 1).

168 Although mutation rate variation between divergent strains is significant, the magnitude of the
169 differences is modest (around 5-fold) (Figure 1, 2, Suppl. Table 2). This may be inherent to streptococci,
170 since only one order of magnitude or lower differences in mutation rate distinguish mutator from non-
171 mutator phenotypes in other streptococcal species (around 10^{-7} and 10^{-8} mutation rates respectively)
172 (23, 73). This is in contrast to other common pathogens where 2-3 orders of magnitude differences are
173 observed (23). On the other hand, small differences in mutation rate are consistent with transience of
174 mutators in natural populations and restoration of low mutation rate after adaptation is gained.
175 Evolutionary theory predicts that heritably hypermutable phenotypes are costly due to inevitable
176 accumulation of deleterious changes thus an adapted mutator clone would evolve back towards a
177 lower mutation rate (17, 18, 74). This may occur via compensatory, 'anti-mutator' mutations in DNA
178 repair genes (4, 19), or via recombination (21). Our data favour the former mechanism for several
179 reasons. Firstly, we used a maximum likelihood inference method to exclude areas of recombination

180 from our phylogenetic analyses (75), and the MMR and OG gene regions were not identified as
181 recombinant regions in this analysis. Secondly, we created a pangenome for the 80 strains in our
182 analysis and mapped on putative phage and prophage signatures (76), and insertion sequence elements
183 (77). None of the MMR or OG genes, which are all present within the core genome of *S. iniae*, occurred
184 in proximity to phage or ISE (Suppl. Fig. 1). Nevertheless, in laboratory *E. coli* populations,
185 recombination seems to be the dominant process by which alterations in MMR and OG genotype occur,
186 as determined by incongruence between phylogenetic trees constructed from the variant MMR/OG
187 gene sequences and those inferred from genome-wide data (21). We were unable to apply similar
188 analysis to our data set as the low number of substitutions in *S. iniae* MMR and OG genes (one or two
189 non-synonymous mutations per gene compared with 1-53% non-synonymous base substitutions in the
190 *E. coli* population (21)) provided insufficient support for robust phylogeny when used in isolation. The
191 present strains originate from complex natural ecosystems where recombination (and mutation) may
192 be more constrained by multifactorial selection pressures in contrast to a monoculture experiment.
193 Moreover, recombination rates in richly capsulated streptococci are lower than in those with reduced
194 capsule (78). Indeed, recombination rates within the core genome across the *S. iniae* strains in our
195 analysis were extremely low with mean $\mu_{r/m}$ 0.032 and only 5 from the 83 strains containing
196 recombination blocks within the core genome where all of the MMR and OG genes analysed are
197 located (Suppl. Table 5). Whilst we cannot rule out switch in MMR and OG type by recombination,
198 accrual of SNPs via random mutation is a more parsimonious explanation here.

199 To investigate whether variation in mutation rate phenotype and MMR/OG genotype is associated with
200 adaptation to the host, we identified variation in key virulence traits in streptococci (capsular
201 polysaccharide, hemolysin, length of cell chains) and bacteria in general (resistance to ROS, and biofilm
202 formation). Employing Phylogenetic Generalised Least Squares regression model accounting for
203 autocorrelation due to phylogeny, we observe significant correlation (p = 0.0331, PGLS) of the number
204 of MMR and OG variants with number of atypical phenotypes for traits contributing to virulence. This
205 suggests that fluctuations in mutation rate are linked to shifts in virulence, potentially facilitating major
206 adaptive events that lead to strain diversification (Fig. 2, Table 3). Moreover, infections in atypical hosts
207 (mammals) and tissue (bone) within apparently immune vaccinated hosts correlates strongly with
208 number of atypical virulence-associated phenotypes ($p < 0.0001$) and the number of variants MMR and
209 OG variants ($p = 0.002$). This implies that mutator strains facilitate multiple genetic changes leading to
210 profound shifts in virulence and antigenicity during adaptation to a drastically different immunological
211 setting.

212 Of the clades identified among *S. iniae*, clade A is apparently a dominant circulating clade primarily
213 infecting Perciform fish that has persisted globally for almost two decades (clade A contains strains

214 isolated from USA, Honduras, and Australia between 1999 to 2016). Mutation rates in all isolates (*n*=35)
215 among this ancestral lineage fall within a 1.5 - 2 x 10^{-8} range (Figs. 1, 2; Clade A), which is similar to the
216 base mutation rate estimated for other non-mutator gram positive bacteria (23, 73). All differences in
217 mutation rates within this clade are insignificant except three pairwise comparisons that verge on p
218 <0.05 (Suppl. Table 1). MMR and OG genes in these strains are identical, and only minor deviations
219 from wildtype traits contributing to virulence were observed (absence of capsule production and longer
220 cell chains in QMA0158, 216) (Figs. 1-3, Table 1). All other lineages (with the exceptions of clade C1 and
221 strains QMA0445-46 discussed below) have significantly different mutation rates to clade A, contain
222 unique and often multiple SNPs in *mut* genes, and show peculiar phenotypic variants related to
223 virulence (Figs. 1-3, Tables 1,2, Suppl. Table 1). Although phylogenetic relatedness between lineages is
224 fully resolved (Fig. 2), they originate almost simultaneously and nested lineages evolve independently
225 sharing little phylogenetic history (Figs. 1,2). Lack of correlation between branch length and sampling
226 date derived by root-to-tip regression analysis with time is evidence of non-neutral evolution amongst
227 this collection of *S. iniae* isolates (R^2=0.0936, Correlation coefficient 0.306, best-fit root; Suppl. Fig. 2,
228 (79)). Deviation from a neutral model is supported by highly skewed branching, with some branches
229 comprising single isolates whilst others comprise many epidemiologically unrelated isolates (Figure 1,
230 2). This is suggestive of frequent strong selection, presumably imposed by heterogeneity of the immune
231 landscape encountered during transfer between host individuals, and resembles genealogical tree
232 topology derived from viral evolution over similar timespans (80, 81) although we acknowledge tree
233 topology may also be affected by sampling bias.

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

245 Clade C shares a common ancestor with clade B and is comprised of strains obtained from barramundi
246 farmed in WA, Northern Territory (NT) and north QLD fish farms (nested subclades C1 and C2
247 respectively). While a core mutation rate of 1.5-2 x 10^{-8} is observed in subclade C1, a significantly higher

248 mutation rate of 7 x 10^{-8} has been maintained in subclade C2 from north QLD for almost two decades
249 (1995 to 2012) (Figs. 1,2 Suppl. Table 1). We speculate that the latter is attributable to rapid fluctuation
250 in water salinity, oxygen and temperature accompanying periodic heavy rainfall in tropical north QLD,
251 creating an unstable environment where adaptive processes are ongoing and high mutation rate is
252 advantageous. Minor variation in virulence-related phenotypes (absence of capsule production and
253 longer cell chains in QMA0074, 77) (Fig. 2, Table 1), are supportive that mutation rate evolution in this
254 lineage might be driven by outside-the-host factors. The significant difference in mutation rate between
255 nested clades C1 and C2 may result from variation in *dnaN*. In *Bacillus subtilis,* 90% of mismatch repair
256 is dependent on targeting MutS to nascent DNA via the ß sliding clamp (DnaN) zone (82, 83).
257 Inactivation of *dnaN1* in *B. anthracis* results in a mutator phenotype with a mutation rate equivalent to
258 a *mutS* mismatch repair-defective strain (84). The tyrosine to isoleucine substitution in one of the
259 critical residues of the loader binding interface of the ß-clamp, predicted as deleterious (PROVEAN
260 score -2,216), is present in all isolates of from clade C (Table 2), which may explain the high mutation
261 rate in clade C2 strains (Fig. 2, Table 2). In contrast, only subclade C1 harbours T1135C nucleotide
262 substitution in the *dnaN* that changes stop codon TAA into CAA coding for glutamine predicted as
263 neutral (PROVEAN score -0.005), with the next stop codon TAG found immediately downstream.
264 Putatively, T1135C and/or change of the *dnaN* stop codon (85) may act as compensatory to the T326I
265 substitution, reducing the mutation rate to a value not significantly different from the core rate found
266 in clade A (Fig. 1, 2) via protein elongation or by affecting translation via stop codon usage bias (85). In
267 addition, a synonymous variant in *recD2* that is only present in subclade C2 may affect protein
268 expression thereby contributing to the difference in mutation rate between C1 and C2 subclades.

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

274 Isolates from humans and fish are found in clade E where multiple virulence trait phenotypes are
275 observed (Figs. 2-3, Table 1). This clade contains two nested subclades that share a SNP in *mutS,* but
276 have other unique SNPs (Fig. 2, Table 2) and exhibit mutation rates that are significantly different in
277 most pairwise comparisons (Suppl. Table 1). Subclade E1 contains USA isolates from humans
278 (QMA0133-35, 37-38) and hybrid striped bass (QMA0447-48), which have a substitution in the Shine-
279 Dalgarno sequence of *recD2* and a mutation rate of 6.5 x 10^{-8} (Fig 1, Table 2). Fish strains and three
280 human strains (QMA0135, 37-38) form shorter cell chains and are less sensitive to H_2O_2. Two human
281 strains QMA0133-34 produce thicker and more dense biofilm structures (Fig. 3, D2). QMA0133 is non-

282 encapsulated (Fig 3, A2), and QMA0134 appears to differentially express the capsular polysaccharide in
283 culture (Fig 3, A3). The close phylogenetic relationship presented here implicates a likely transfer from
284 farmed hybrid bass to humans (86), a host jump that may have been facilitated by the mutator
285 phenotype. Subclade E2 contains two human strains from Canada (QMA0130-31) and a tilapia strain
286 from USA (QMA0466). These strains have unique methionine to isoleucine substitution at the N-
287 terminus of MutX, and a mutation rate of 4.5 x 10^{-8} (Fig 1, Table 2). Both human strains are non-
288 encapsulated (Fig. 2). The tilapia strain expresses the capsule but forms short cell chains in common
289 with most strains from the subclade (Fig 2, Fig 3, C3). The close phylogenetic relationship again strongly
290 implicates transfer from tilapia farmed in USA to human patients in Canada via import into the Toronto
291 fish market (87) and again, the mutator state may underlie this host jump.

292 Clade F is not readily definable by location, time of isolation, or host species and comprises QMA0139
293 from unidentified fish in Canada, QMA0190 isolated from snakehead murrel in Thailand (both of which
294 appear on long branches), and a nested terminal clade comprised of barramundi isolates farmed in
295 Recirculating Aquaculture Systems (RAS) in New South Wales and South Australia (Fig. 2). The latter are
296 mutators and were sampled from barramundi bone lesions during a disease outbreak in vaccinated fish
297 where infection manifested itself as a slowly progressing osteomyelitis instead of typical acute
298 septicaemia and meningitis, implicating mutators in adaptation to new tissue in response to host
299 immunity (64, 65). Since they are distant to strains used in autogenous vaccines in clade A (QMA0155-
300 57, 160, QMA0250-52), the atypical outbreak might be classified as a case of vaccine-induced serotype
301 replacement (VISR) – spread of co-existing pathogenic strain/s after elimination of the dominant
302 strain/s in the vaccine (88). In contrast to natural populations where the investigation of VISR is
303 confounded by ecological and sampling biases, and randomised vaccine trials that lack power to detect
304 the population-wide effect of mass vaccination (88), barramundi RAS represent a relatively controlled
305 environment where vaccination against *S. iniae* fails occasionally but recurrently (65). RAS are self-
306 contained with minimal water exchange and high standards of water treatment (89). This ensures that
307 gene flow and, consequently, the probability of new serotype introduction after vaccination is reduced
308 to a minimum. A comparatively uniform immunized host population is exposed to dominant strains
309 against which it was vaccinated along with potentially co-existing lineages in the RAS. In the atypical
310 outbreaks in RAS in Australia, vaccine-induced immunity was partially effective in the infected fish,
311 evident by absence of proliferation of the new strain in typical niduses (blood, brain, and pronephros)
312 and cross-reactivity of vaccine-induced antibodies against bone isolates (65). This is in contrast with the
313 classical model of serotype replacement where a vaccine is ineffective against the replacing co-existing
314 strain. However, VISR can occur even if the vaccine is equally efficient against all circulating strains via
315 other trade-off mechanisms (88). Apparently, evasion of immune clearance and adaptation to osseous

316 tissue was permitted by shift towards attenuated virulence indicated by multiple phenotypic changes in
317 bone isolates, including absence of capsule, impeded hemolytic activity, increased cell-chain length, and
318 denser biofilms (Fig. 3). All of these changes have previously been associated with chronic infection and
319 increased potential for colonization (90-96). Multiple SNPs in repair genes (Fig. 2, Table 2) and the
320 unique mutation rate (Figs. 1-2, Table 1) in bone isolates suggest that shift in virulence allowing
321 persistence in the bone tissue might have occurred via mutator phenotype. Serine to arginine
322 substitution in MutL and tyrosine to isoleucine substitution in MutY occur in the ancestor of QMA0139,
323 QMA0190, and bone strains. Both substitutions are predicted to have a deleterious effect on protein
324 function (PROVEAN scores -2.050 and -2.544 respectively), but their effect on mutation rate is
325 combined with unique SNPs found in each branch: QMA0139 has a SNP in the binding site for the DnaA
326 transcription factor within the *dnaN* promoter, and a mutation rate of 4.5 x 10^{-8}, QMA0190 has
327 synonymous SNP in *mutL* and a mutation rate of 6.5 x 10^{-8}, and bone strains have glutamine to glycine
328 substitution in MutS predicted as neutral (PROVEAN score -0.635), a synonymous SNP in *mutL*, and a
329 mutation rate of 3-4 x 10^{-8}. Considering that the mutation rate of QMA0190 is significantly higher
330 compared to rates expressed by QMA0139 and bone isolates (Suppl. Table 1), it is likely that variants
331 unique to the latter strains compensate for deleterious variants shared by the isolates (4).

332 Clade G contains two strains from tilapia isolated in USA (QMA0445-46). Notably, despite being
333 phylogenetically distant from clade A, these strains have the same MMR and OG genotype and have
334 retained the core mutation rate and the wildtype virulence phenotype (Fig. 2, Table 1), with the
335 exception of capsule absence.

336 Three isolates of *S. iniae* were identified as independent phylogenetic lineages, with unique mutation
337 rate genotypes and phenotypes, and virulence traits (Figs. 1-2, Tables 1-2). The first long branch with a
338 single isolate contains a strain from snakehead murrel strain (*Channa striata*) from Thailand. This isolate
339 is non-encapsulated and weakly haemolytic and has a mutation rate of 4.5 x 10^{-8}, potentially
340 attributable to a deleterious aspartate to asparagine substitution in rnhC and SNP in the binding site of
341 for the *fnr* transcription factor within the *dnaN* promoter. A second long branch with a single strain
342 contains the oldest among the analysed strains, QMA0140 isolated in 1976 from dolphin (*Inia*
343 *geoffrensis*) (97). This strain exhibits a mutation rate of 1 x 10^{-8}, which is unique among the strains and
344 significantly lower than the core mutation rate, perhaps linked to increased translation of recD2
345 helicase resulting from a SNP in the ribosome binding site (Fig 2, Table 2). The longest branch on the
346 tree contains a second dolphin strain, QMA0141, isolated two years later in 1978 (98). This isolate is
347 highly divergent from the rest of strains with around 20 kB of non-recombinant SNPs in pair-wise
348 comparisons with other strains, accounting for around 1% genomic difference. In contrast to QMA140,
349 it mutates at a rate of 1x 10^{-7}, the highest mutation rate phenotype determined among the isolates and

350 significantly different to the rest of the values. Multiple SNPs are observed in all MMR and OG genes: 3
351 in *dnaN* and *mutX*, 4 in *mutM*, 9 in *mutL* and *rnhC*, 15 in *recD2*, 9 in *mutL* and *rnhC*, and 68 in *mutS*. Both
352 dolphin isolates are non-encapsulated, show increased ability to withstand oxidative stress, and form
353 denser and thicker biofilms (Fig 2, 3).

354 The conservation of *S. iniae* mutation rate and MMR/OG genotype within phylogenetic clades and
355 variation between the lineages support the importance of mutation rate dynamics in pathogen
356 evolution. Mechanisms that underlie the flux of mutator alleles and phenotypes include random
357 mutations (4), tandem repeat copy-number variation (20, 58), non-homologous recombination (21),
358 and prophage integration/excision (22). In *S. iniae,* mutation rate dynamics appear to be primarily
359 driven by random mutations and compensatory evolution. Tandem-repeat copy-number variants were
360 not identified in *S. iniae* MMR or OG genes, supporting the contention that occurrence of mutator
361 phenotypes and their full reversion via this mutational type is limited to *E. coli* and closely related
362 species (20, 58). Indeed, the presence of tandem repeats has only been detected in *mutL* in
363 *Proteobacteria* (20, 58), and it is well recognized that *mutL* is structurally and functionally different in
364 gram-positive bacteria and eukaryotes (54, 61). With the exception of the large deletion affecting the
365 *mutY* promoter in clade B, all identified mutations are single nucleotide substitutions. These MMR and
366 OG SNPs most likely result from random mutations which can be sufficient to induce profound changes
367 in mutation rate (55, 56). Since, in most cases, SNPs occur in multiple genes (Fig. 2) it is likely that
368 compensatory evolution of mutation rate prevails in *S. iniae* where some variants increase and some
369 decrease the background mutation rate (4). The association of MMR/OG variation with changes in
370 virulence traits (Fig. 2) supports the contention that emergence and decline of mutator strains may
371 facilitate adaptive processes in *S. iniae* populations driven by host immune response. Furthermore, the
372 most profound changes in virulence-associated traits and MMR and OG variants are observed in
373 atypical hosts (mammals) and atypical tissue (bone) within apparently immune vaccinated fish hosts
374 (Fig. 2), which has major epidemiological implications. First, mutators may promote multiple adaptive
375 changes required for pathogen transition among divergent host taxa known as 'host jumps', such as
376 *Osteichthyes* and *Mammalia* discussed in the present study (99). Second, mutators might present a risk
377 factor to immune escape after vaccination and serotype replacement. Moreover, as evidenced by *S.*
378 *iniae* colonization of bone tissue of vaccinated barramundi described previously (65), vaccine escape
379 leading to serotype replacement might occur without major disruption of immune recognition when
380 multiple changes attenuating virulence and antigenicity allow persistence in a tissue with lower
381 immune surveillance.

382 We cannot eliminate the possibility that variation in mutation rate and MMR/OG genotype in *S. iniae* is
383 neutral and simply coincides with phylogenetic diversification and variation in traits determining

384 virulence. Presence of dysfunctional MMR and OG genes within the evolutionary history of bacterial
385 strains may leave signatures within the genome in the form of mutational bias. For example, in all but
386 the most AT-rich bacteria there is a predisposition for GC-AT transitions over time (100), but in a long
387 term evolutionary experiment in *E.coli,* defective OG repair resulted in strong bias towards AT to CG
388 transversions while AT to GC transitions were dominant in MMR defective mutators (101). We
389 quantified the mutational spectra by clade across the natural *S. iniae* isolates in this study (Fig. 4). Two
390 clades of non-mutator strains (A and C1) had highly similar SNP type distribution (Fig. 4). Of the mutator
391 clades, only clade D had mutations constrained to a single MMR/OG, a SNP in *mutM* in the OG repair
392 system (Fig. 2), resulting in an amino acid substitution in the DNA glycolyase domain with PROVEAN
393 score of -2.990 (deleterious) (Table 2). Here we found the highest frequency of AT to CG transversions,
394 consistent with findings in *E. coli* experimental mutators defective in OG repair, albeit at much lower
395 frequency in our natural *S. iniae* mutator isolates (101). Two mutator clades (B and E2) had higher
396 frequency of AT to GC transitions than the non-mutator clades, and indeed other mutator clades (Fig 4).
397 Clade B contains mutations in both *mutL* (V-I substitution, PROVEAN -0.27, Table 2) and *recD2*
398 (synonymous) in the MMR system. Clade B strains also contain a conserved mutation in the promoter of
399 *mutY* in the OG repair system (Table 2). We acknowledge the weakness of these signals compared to
400 those reported in *E.coli* (101). However, we include only natural mutator isolates in which changes occur
401 in multiple genes in both MMR and OG systems. Moreover, *S. iniae* genomes are already AT-rich (mean
402 36.55% GC compared to *E. coli* ~50%) which may constrain transition and transversion mutations that
403 further bias the nucleotide content. Consequently, complete conservation of MMR and OG genes linked
404 with consistent mutation rates within *S. iniae* lineages (Fig. 2) and the fact that most SNPs occur in
405 functionally relevant coding and regulatory regions and/or change amino acid sequence (Table 2),
406 supports that the occurrence of mutator strains is likely to precede and facilitate shifts in virulence. This
407 may in turn lead to the major diversification events and epidemiologically relevant incidents such as
408 vaccine escape outbreaks and host jumps in these bacterial lineages.

409 Our results support the contention that high mutation rates promote transition to a new immunological
410 setting requiring multiple mutations and phenotypic changes, such as adaptation to a novel host, new
411 host tissue, or a host with vaccine-induced adaptive immune response. Extension to other species such
412 as *Streptococcus pneumoniae* may be particularly interesting as serotype evolution following vaccination
413 programmes is well documented, but the role of mutators in this process is unknown (66, 67, 102).
414 Future studies should determine how generalisable the role of mutators is in the epidemiology of
415 bacterial pathogens, evolution of virulence and antigenicity, and emergence of new pathogenic strains.

416 **Materials and methods**

417 **Strains and growth conditions**

418 Eighty isolates of *S. iniae* collected in Australia, USA, Canada, Israel, Honduras, and Thailand between
419 1976 and 2016 from eight fish species (*Lates calcarifer*, *Scortum barcoo, Epalzeorhynchos frenatum,*
420 *Epalzeorhynchos bicolor, Oreochromis sp., Channa striata, Chromobotia macracanthus, Oncorhynchus*
421 *mykiss*) and three mammalian species (*Homo sapiens, Inia geoffrensis, Pteropus alecto*) were analysed
422 (Table 1). Strains were received from culture collections, veterinarians, or directly from fish farms
423 (Suppl. Table 3) and stored as master seed stocks without further subculture at -80°C in Todd-Hewitt
424 Broth (THB, Oxoid) + 20% glycerol. Frozen stocks were recovered on Columbia agar supplemented with
425 5% defibrinated sheep blood (Oxoid), and cultured at 28°C on Todd-Hewitt agar (THA) or in THB with
426 agitation 200 rpm unless otherwise specified.

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

428 A fluctuation analysis assay for spontaneous occurrence of rifampicin resistance was optimized
429 according to Rosche and Foster (103). A single broth culture was initiated from five separate colonies,
430 recovered on Columbia blood agar from stock, and grown overnight to late-exponential phase in THB.
431 Cultures were adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 1:100, and distributed in 200 µl aliquots into
432 8 wells of a sterile U-bottom 96-well plate (Greiner). Prior to dilution, 100 µl of each OD_{600}-adjusted
433 culture was spread onto THA containing 0.5 µg/mL rifampicin to confirm absence of resistant mutants.
434 Although comparatively large initial inocula (1-2 x 10^5 CFU per culture, confirmed by Miles and Misra
435 CFU count (104) of OD-adjusted cultures) were used, this minimised variance in final CFU number (N_t)
436 determined in preliminary experiments. Invariability in N_t allowed statistical comparison of mutation
437 rates by the Maximum Likelihood Ratio test (71). To infer the N_t in final cultures and monitor their
438 growth by optical density, replicate plates containing two cultures per strain were prepared and
439 incubated in a BMG FLUOstar OPTIMA microplate reader. When these representative cultures entered
440 early stationary phase, CFU counts were performed by Miles and Misra method (104) and invariably
441 estimated as 1-2 x 10^8 CFU in each well. Immediately after, entire 200 µL cultures (above) were plated
442 on THA containing 0.5 µg/mL rifampicin for selection of mutants, dried under laminar flow, and
443 incubated until rifampicin resistant colonies appeared. The assay was repeated four times for each
444 strain in blocks of eight cultures using 20 isolates haphazardly chosen from different phylogenetic
445 lineages in each measurement. In some cases, e.g. a single strain representing an independent
446 phylogenetic lineage or with significantly differences in mutation rates detected in closely related
447 strains, assays were repeated 1-2 more times to increase statistical power. The rifampicin-resistant
448 mutant counts were pooled into single data sets representing 32 to 48 cultures per strain (Suppl. Table
449 2).

450 **DNA extraction, preparation and sequencing**

451 Genomic DNA was extracted from cells collected from 10 mL late-exponential phase culture in Todd-
452 Hewitt broth with the DNeasy Blood & Tissue kit (Qiagen) using a modified protocol with an additional
453 lysis step as described previously (105). Sequencing was performed on the Illumina HiSeq2000 platform
454 from Nextera XT pair-end libraries at Australian Genome Research Facility, Melbourne. A reference
455 genome from strain QMA0248 was constructed using both long reads derived from a single Smrt Cell
456 using the PacBio RS II system with P4C2 chemistry and short reads from Illumina HiSeq2000 derived
457 from Nextera XT paired-end libraries as reported elsewhere (NCBI accession no: GCA_002220115.1). All
458 sequence data are deposited at NCBI under Bioproject number PRJNA417543, SRA accession
459 SRP145425. Sample numbers, accession numbers and extended metadata are provided in Suppl. Table
460 3. Assembly statistics are provided in Suppl. Table 4.

461 **Alignment, recombination detection and phylogenetic analysis**

462 Phylogeny was constructed based on core genome SNPs from *de novo* genome assemblies filtered to
463 remove recombination breakpoints. Paired-end reads from Illumina were trimmed with Nesoni clip tool
464 version 0.132 (http://www.vicbioinformatics.com/software.nesoni.shtml), with minimum read length
465 50, and the first 15 bp of each read removed as quality deterioration in this region was observed when
466 assessed with FASTQC version 0.11.5. Assembly was performed using the SPAdes assembler version
467 3.7.1 (106), with minimum read coverage cutoff set to 10. Quality of assemblies was assessed with
468 QUAST 3.2 (107). Contigs were ordered by alignment to QMA248 reference genome (CP022392.1) with
469 Mauve Contig Mover 2.4.0 (108). Genome annotation was performed using Prokka 1.11 (109). Rapid
470 alignment of core genomes was carried out using parsnp in the Harvest Tools suite version 1.2 (110),
471 and the resulting alignment provided as an input to Gubbins 1.4.7 (75) for detection and exclusion of
472 variants produced by recombination. Phylogenies were then inferred from post-filtered core genome
473 polymorphic sites by maximum likelihood using RAxML 8.2.8 (111) with the general time reversible
474 nucleotide substitution model GTRGAMMA and bootstrap support from 1000 iterations. Effect of
475 ascertainment bias associated with using only polymorphic sites on branch length during ML inference
476 was corrected using Felsenstein's correction implemented in RAxML 8 (112). The resulting phylogenetic
477 tree was visualized using Dendroscope v 3.5.7 (113) with bootstrap node support value cut-off 75. For
478 the phylogram figure, tip labels were hidden for clarity and the edge containing QMA0141 was re-
479 scaled as dotted line representing 100-fold decrease in length (Fig. 1). A cladogram based on the
480 inferred phylogeny, showing all tip labels and bootstrap support for each node, was annotated with
481 metadata using Evolview V2 (Fig. 2)(114). To determine whether there was a strong temporal signal in
482 the phylogenetic data, a root-to-tip regression of branch length against time since isolation was
483 performed in TempEst (79), using genetic distances corrected for ascertainment bias in an unrooted

484 tree estimated by maximum likelihood from the alignment of non-recombinant core-genome SNPs in
485 RAxML 8.2.8.

486 **SNP analysis, variation in DNA repair genes and their regulatory regions**

487 SNP analysis was performed by read-mapping with Geneious version 9.1 (115), using default settings
488 unless otherwise specified. Paired-end reads from each genome were trimmed, merged into a single
489 file with expected distances between the reads set to 250 bp, and mapped to the curated reference
490 genome of strain QMA0248. Mapped reads were used to detect SNPs with minimum coverage set to 10
491 and frequency to 0.9. A consensus "pseudogenome" was generated for every strain based on the
492 reference sequence incorporating any detected variants (116). Multiple alignment of these
493 pseudogenomes was carried out with *Geneious* aligning tool. SNP types and position per strain were
494 exported as a table and SNPs of each type occurring only in each clade but not in any other clade were
495 counted and plotted using a custom script in R. Sequences of MMR genes (*dnaN,* mut*S, muTL, recD2,*
496 *rnhC*) and OG genes (*mutY* , *mutM*, *mutX*) annotated in the QMA0248 reference genome were
497 identified in each strain genome sequence. Promoter regions of repair genes were determined with
498 BPROM (117), protein functional domains by SMART genomic (118), and effect of amino acid
499 substitutions on protein function by PROVEAN Protein with a sensitivity cut-off of 1.3 (119). To detect
500 large variants, alignment of these regions extracted from *de novo* genome assemblies was carried out,
501 and a 56 bp deletion detected in *mutY* promoter of clade B was confirmed by PCR with
502 CAGAAGGAAGAAACAGAC_F/ACCTCTATTGTAGCAAAG_R primers.

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

504 A pangenome for the 80 *S. iniae* isolates was constructed with GView server (120) using the
505 QMA0248 reference genbank (.gbk) file as seed by sequentially adding 79 genome assemblies by
506 clade in the order derived from the cladogram (Fig. 2). Phage positions in the resulting pangenome
507 were then determined by BLAST using Phaster (76) while IS positions were identified using ISFinder
508 (77) with a BLAST e-value cutoff of 1×10^{-30}. Positions of MMR and OG genes in the pangenome were
509 determined via manual search of the .gbk file. An image was created in BRIG (121) using the
510 pangenome as reference then aligning the 80 genome assemblies by BLAST with an e-value cutoff of
511 1×10^{-30}. The resulting image was annotated in BRIG from a spreadsheet derived from the positions
512 determined above, and coloured according to clade (Suppl. Fig. 1).

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

514 A multitude of intracellular and secreted enzymes contribute to virulence in the *Streptococcus* genus
515 (90, 122). Considering the large number of strains, we limited analysis to major, easily observable, and

516 distinct phenotypes strongly associated with virulence in streptococci (capsule, hemolysis, cell chains)
517 and bacteria in general (oxidation resistance, biofilms). All assays were performed at least in triplicate
518 per strain.

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

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

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

531 Rapid high-throughput detection of impaired haemolytic activity was achieved by blood-agar clearance
532 zone assay. Briefly, 5 mm wells were made in Columbia agar supplemented with 5% defibrinated sheep
533 blood (Oxoid, Australia). Bacterial cultures grown to late-exponential phase (~10^8 CFU) were diluted
534 1:1000 and 50 µL of diluted cultures were pipetted into punctures in the agar (initial inoculum ~$5x10^4$
535 CFU in each puncture). A 3 mm wide clearance zone was generally produced by bacterial lysis of sheep
536 erythrocytes during 24 h incubation (Fig. 3, B1). Where haemolysis was absent or fragmentary, impeded
537 haemolytic activity was recorded (Fig. 3, B2).

538 **Chain formation microscopy**

539 To assess chain formation, *S. iniae* cultures in THB were grown stationarily in 96-well plates at 28 °C for
540 24 h, mixed and 5 µL wet mounts prepared on a glass microscope slide. Slides were observed by bright
541 field microscopy under 40x objective with an Olympus BX40 microscope and captured using an
542 Olympus DP28 digital camera using CellSens software (Olympus Optical Co, Japan). Specimens were
543 observed for at least 3 min and 2-20 cell chains were generally observed (Fig. 3, C1). Where over 20
544 cells in a chain were repeatedly detected (Fig. 3, C2) increased chain formation was recorded, and when
545 more than 10 cells in a chain were not detected by similar observation and the culture was mainly
546 composed of detached cells (Fig. 3, C3) impeded chain formation was recorded. For the figure, 50 µL of

547 cultures were dried onto slides at RT, fixed with methanol, Gram stained, and images captured under
548 100x objective.

549 **Oxidation resistance assay**

550 Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) (123) of
551 hydrogen peroxide were measured to assess oxidative stress resistance among *S. iniae* strains. THB
552 cultures grown to late exponential phase were adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 1000-fold,
553 and distributed by 100 µL (~10^4 CFU) into wells of a U-bottom 96-well plate (Greiner). THB (100 µL) with
554 2-fold excess concentration of hydrogen peroxide was added to the wells and serially diluted twofold,
555 resulting in a range from 0 to 10 mM final peroxide concentrations. Plates were incubated stationarily
556 for 24h and examined for presence of observable growth to determine the MIC. Viable cell counts of
557 cultures without visible growth were performed to determine MBC. MIC of hydrogen peroxide was
558 estimated as 3 mM for all strains, and MBC as 4 mM for the majority of strains. MBC elevated to 5 mM
559 in dolphin and human isolates from USA was classified deviation from a wildtype phenotype.

560 **Biofilm formation and visualisation assay**

561 *S. iniae* biofilms were grown for 4 days in 8-well Lab-Tek® II Chamber Slide™ Systems. To prepare an
562 initial inoculum of 5 x 10^5 CFU, THB cultures grown to late exponential phase were adjusted to OD_{600} = 1
563 (10^8 CFU/mL), diluted 100-fold, and 0.5 mL of diluted cultures were placed into Chamber slide wells.
564 After 24 h incubation, THB was removed and replaced with fresh THB every 10-14 h for 3 days. For
565 visualisation, biofilms were washed in PBS, stained for 15 min with 1 µM fluorescent *Bac*Light Red
566 bacterial stain, washed in PBS, fixed for 30 min with 10% formalin, and washed twice PBS. Z-stacks were
567 collected by ZEISS LSM 710 Inverted Laser Scanning Confocal Microscope at 20x objective and visualised
568 using ZEN2012. Quantification of biofilms was performed using COMSTAT (124). Structures of biomass
569 larger than 3.5 and average thickness over 5 were classified as increased biofilm forming activity (Fig. 3
570 D).

571 **Identifying wildtype virulence-related phenotypes**

572 We consider prevalent phenotypes as a wildtype, namely: presence of polysaccharide capsule (71.25 %
573 of strains; Figure 3, A1) and haemolytic activity (85% of strains; Figure 3, B1), up to 20 cells in a chain
574 (77.5 % of strains; Figure 3, A1), 3 mM MIC and 4 mM MBC of hydrogen peroxide (91.25 % of strains),
575 and 8 -12 µm thick biofilms covering up to 30% of representative image (86.25 % of strains, Fig. 4A).
576 Other phenotypes are regarded as deviant and discussed.

577 **Statistical analysis**

578 Analysis of co-variance between number of MMR and OG variants, mutation rates, number of
579 atypical phenotype associated with virulence, and atypical places of isolation were tested with a
580 Phylogenetic Generalised Least Squares model that accounts for relatedness among strains under
581 Pagel's λ of 0.6 implemented in R (70). Estimation of mutation rates was carried out using the Ma-
582 Sandri-Sarkar Maximum Likelihood Estimation (MSS-MLE) method as implemented in FALCOR
583 fluctuation analysis calculator (125). Differences in mutation rates were compared by Likelihood
584 Ratio Test (LRT) using the rSalvador R package (71, 126). LRT estimates the overlap between
585 confidence intervals calculated for two fluctuation experiment data sets and allows to compare
586 fluctuation assay data with different number of cultures and resolve very small differences in
587 mutation rate (71, 126). The *Compare.LD* function was applied pair-wise to all strains, and obtained
588 *p* values were corrected for multiple comparisons by controlling the False Discovery Rate (72) using
589 p.adjust in R. Adjusted *p* values less than 0.05 were considered to indicate significant difference in
590 mutation rates.

591

592 **Acknowledgements**

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

603 **References**

604 1. Ferenci T. 2015. Trade-off Mechanisms Shaping the Diversity of Bacteria. Trends Microbiol
605 doi:10.1016/j.tim.2015.11.009.
606 2. Stich M, Manrubia SC, Lazaro E. 2010. Variable Mutation Rates as an Adaptive Strategy in
607 Replicator Populations. Plos One 5.
608 3. Sniegowski P, Raynes Y. 2013. Mutation Rates: How Low Can You Go? Current Biology
609 23:R147-R149.
610 4. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B,
611 Medigue C, Lenski RE, Schneider D. 2013. Mutation rate dynamics in a bacterial population
612 reflect tension between adaptation and genetic load. Proc Natl Acad Sci U S A 110:222-7.
613 5. Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE,
614 Schneider D. 2011. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term
615 Evolution Experiment With Escherichia coli. G3 (Bethesda) 1:183-186.
616 6. Galhardo RS, Hastings PJ, Rosenberg SM. 2007. Mutation as a stress response and the
617 regulation of evolvability. Crit Rev Biochem Mol Biol 42:399-435.
618 7. Sundin GW, Weigand MR. 2007. The microbiology of mutability. Fems Microbiology Letters
619 277:11-20.
620 8. Foster PL. 2007. Stress-induced mutagenesis in bacteria. Crit Rev Biochem Mol Biol 42:373-
621 97.
622 9. Wrande M, Roth JR, Hughes D. 2008. Accumulation of mutants in "aging" bacterial colonies
623 is due to growth under selection, not stress-induced mutagenesis. Proc Natl Acad Sci U S A
624 105:11863-8.
625 10. Torres-Barcelo C, Kojadinovic M, Moxon R, MacLean RC. 2015. The SOS response increases
626 bacterial fitness, but not evolvability, under a sublethal dose of antibiotic. Proc Biol Sci
627 282:20150885.
628 11. Andersson DI, Hughes D. 2010. Antibiotic resistance and its cost: is it possible to reverse
629 resistance? Nat Rev Microbiol 8:260-71.
630 12. Boe L, Danielsen M, Knudsen S, Petersen JB, Maymann J, Jensen PR. 2000. The frequency of
631 mutators in populations of Escherichia coli. Mutat Res 448:47-55.
632 13. Miller JH. 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis
633 and repair. Annu Rev Microbiol 50:625-43.
634 14. Mao EF, Lane L, Lee J, Miller JH. 1997. Proliferation of mutators in A cell population. J
635 Bacteriol 179:417-22.
636 15. Taddei F, Radman M, MaynardSmith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of
637 mutator alleles in adaptive evolution. Nature 387:700-702.
638 16. Torres-Barcelo C, Cabot G, Oliver A, Buckling A, Maclean RC. 2013. A trade-off between
639 oxidative stress resistance and DNA repair plays a role in the evolution of elevated mutation
640 rates in bacteria. Proc Biol Sci 280:20130007.
641 17. Raynes Y, Sniegowski PD. 2014. Experimental evolution and the dynamics of genomic
642 mutation rate modifiers. Heredity (Edinb) 113:375-80.
643 18. de Visser JA. 2002. The fate of microbial mutators. Microbiology 148:1247-52.
644 19. Turrientes MC, Baquero F, Levin BR, Martinez JL, Ripoll A, Gonzalez-Alba JM, Tobes R,
645 Manrique M, Baquero MR, Rodriguez-Dominguez MJ, Canton R, Galan JC. 2013. Normal
646 mutation rate variants arise in a Mutator (Mut S) Escherichia coli population. PLoS One
647 8:e72963.
648 20. Shaver AC, Sniegowski PD. 2003. Spontaneously arising mutL mutators in evolving
649 Escherichia coli populations are the result of changes in repeat length. J Bacteriol 185:6076-
650 82.

651 21. Denamur E, Lecointre G, Darlu P, Tenaillon O, Acquaviva C, Sayada C, Sunjevaric I, Rothstein
652 R, Elion J, Taddei F, Radman M, Matic I. 2000. Evolutionary implications of the frequent
653 horizontal transfer of mismatch repair genes. Cell 103:711-21.
654 22. Scott J, Thompson-Mayberry P, Lahmamsi S, King CJ, McShan WM. 2008. Phage-associated
655 mutator phenotype in group A streptococcus. J Bacteriol 190:6290-301.
656 23. Hall LM, Henderson-Begg SK. 2006. Hypermutable bacteria isolated from humans--a critical
657 analysis. Microbiology 152:2505-14.
658 24. Oliver A. 2010. Mutators in cystic fibrosis chronic lung infection: Prevalence, mechanisms,
659 and consequences for antimicrobial therapy. Int J Med Microbiol 300:563-72.
660 25. Matsushima A, Takakura S, Fujihara N, Saito T, Ito I, Iinuma Y, Ichiyama S. 2010. High
661 prevalence of mutators among Enterobacter cloacae nosocomial isolates and their
662 association with antimicrobial resistance and repetitive detection. Clin Microbiol Infect
663 16:1488-93.
664 26. Wang S, Wang Y, Shen J, Wu Y, Wu C. 2013. Polymorphic mutation frequencies in clinical
665 isolates of Staphylococcus aureus: the role of weak mutators in the development of
666 fluoroquinolone resistance. FEMS Microbiol Lett 341:13-7.
667 27. Negri MC, Morosini MI, Baquero MR, del Campo R, Blazquez J, Baquero F. 2002. Very low
668 cefotaxime concentrations select for hypermutable Streptococcus pneumoniae populations.
669 Antimicrobial Agents and Chemotherapy 46:528-530.
670 28. Chopra I, O'Neill AJ, Miller K. 2003. The role of mutators in the emergence of antibiotic-
671 resistant bacteria. Drug Resist Updat 6:137-45.
672 29. Blazquez J. 2003. Hypermutation as a factor contributing to the acquisition of antimicrobial
673 resistance. Clinical Infectious Diseases 37:1201-1209.
674 30. Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL, Oliver A. 2008. Genetic
675 Adaptation of Pseudomonas aeruginosa to the Airways of Cystic Fibrosis Patients Is
676 Catalyzed by Hypermutation. Journal of Bacteriology 190:7910-7917.
677 31. Gutierrez O, Juan C, Perez JL, Oliver A. 2004. Lack of association between hypermutation and
678 antibiotic resistance development in Pseudomonas aeruginosa isolates from intensive care
679 unit patients. Antimicrobial Agents and Chemotherapy 48:3573-3575.
680 32. Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, Taddei F. 2001. Costs and
681 benefits of high mutation rates: Adaptive evolution of bacteria in the mouse gut. Science
682 291:2606-2608.
683 33. Nilsson AI, Kugelberg E, Berg OG, Andersson DI. 2004. Experimental adaptation of
684 Salmonella typhimurium to mice. Genetics 168:1119-30.
685 34. Healey KR, Zhao Y, Perez WB, Lockhart SR, Sobel JD, Farmakiotis D, Kontoyiannis DP,
686 Sanglard D, Taj-Aldeen SJ, Alexander BD, Jimenez-Ortigosa C, Shor E, Perlin DS. 2016.
687 Prevalent mutator genotype identified in fungal pathogen Candida glabrata promotes multi-
688 drug resistance. Nat Commun 7:11128.
689 35. Labat F, Pradillon O, Garry L, Peuchmaur M, Fantin B, Denamur E. 2005. Mutator phenotype
690 confers advantage in Escherichia coli chronic urinary tract infection pathogenesis. FEMS
691 Immunol Med Microbiol 44:317-21.
692 36. Mena A, Macia MD, Borrell N, Moya B, de Francisco T, Perez JL, Oliver A. 2007. Inactivation
693 of the mismatch repair system in Pseudomonas aeruginosa attenuates virulence but favors
694 persistence of oropharyngeal colonization in cystic fibrosis mice. J Bacteriol 189:3665-8.
695 37. Oliver A, Mena A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic
696 resistance. Clinical Microbiology and Infection 16:798-808.
697 38. Aquino T, Nunes A. 2016. Host immunity and pathogen diversity: A computational study.
698 Virulence 7:121-8.
699 39. Cobey S. 2014. Pathogen evolution and the immunological niche. Ann N Y Acad Sci 1320:1-
700 15.

701 40. Lukacisinova M, Novak S, Paixao T. 2017. Stress-induced mutagenesis: Stress diversity
702 facilitates the persistence of mutator genes. PLoS Comput Biol 13:e1005609.
703 41. Abel S, Abel zur Wiesch P, Davis BM, Waldor MK. 2015. Analysis of Bottlenecks in
704 Experimental Models of Infection. PLoS Pathog 11:e1004823.
705 42. Metzgar D, Wills C. 2000. Evidence for the adaptive evolution of mutation rates. Cell
706 101:581-4.
707 43. Fogle CA, Nagle JL, Desai MM. 2008. Clonal interference, multiple mutations and adaptation
708 in large asexual populations. Genetics 180:2163-73.
709 44. Methot PO, Alizon S. 2014. What is a pathogen? Toward a process view of host-parasite
710 interactions. Virulence 5:775-85.
711 45. Mackinnon MJ, Read AF. 2004. Virulence in malaria: an evolutionary viewpoint. Philos Trans
712 R Soc Lond B Biol Sci 359:965-86.
713 46. Deitsch KW, Moxon ER, Wellems TE. 1997. Shared themes of antigenic variation and
714 virulence in bacterial, protozoal, and fungal infections. Microbiol Mol Biol Rev 61:281-93.
715 47. Shrestha S, Bjornstad ON, King AA. 2014. Evolution of acuteness in pathogen
716 metapopulations: conflicts between "classical" and invasion-persistence trade-offs. Theor
717 Ecol 7:299-311.
718 48. Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of
719 bacterial pathogens. FEMS Microbiology Letters 267:1-8.
720 49. Merino D, Reglier-Poupet H, Berche P, Charbit A, European Listeria Genome C. 2002. A
721 hypermutator phenotype attenuates the virulence of Listeria monocytogenes in a mouse
722 model. Mol Microbiol 44:877-87.
723 50. Gonzalez K, Faustoferri RC, Quivey RG, Jr. 2012. Role of DNA base excision repair in the
724 mutability and virulence of Streptococcus mutans. Mol Microbiol 85:361-77.
725 51. Picard B, Duriez P, Gouriou S, Matic I, Denamur E, Taddei H. 2001. Mutator natural
726 Escherichia coli isolates have an unusual virulence phenotype. Infection and Immunity 69:9-
727 14.
728 52. Smania AM, Segura I, Pezza RJ, Becerra C, Albesa I, Argarana CE. 2004. Emergence of
729 phenotypic variants upon mismatch repair disruption in Pseudomonas aeruginosa.
730 Microbiology-Sgm 150:1327-1338.
731 53. Canfield GS, Schwingel JM, Foley MH, Vore KL, Boonanantanasarn K, Gill AL, Sutton MD, Gill
732 SR. 2013. Evolution in fast forward: a potential role for mutators in accelerating
733 Staphylococcus aureus pathoadaptation. J Bacteriol 195:615-28.
734 54. Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids 2010.
735 55. Dai L, Muraoka WT, Wu ZW, Sahin O, Zhang QJ. 2015. A single nucleotide change in mutY
736 increases the emergence of antibiotic-resistant Campylobacter jejuni mutants. Journal of
737 Antimicrobial Chemotherapy 70:2739-2748.
738 56. Rajanna C, Ouellette G, Rashid M, Zemla A, Karavis M, Zhou C, Revazishvili T, Redmond B,
739 McNew L, Bakanidze L, Imnadze P, Rivers B, Skowronski EW, O'Connell KP, Sulakvelidze A,
740 Gibbons HS. 2013. A strain of Yersinia pestis with a mutator phenotype from the Republic of
741 Georgia. FEMS Microbiol Lett 343:113-20.
742 57. Ambur OH, Davidsen T, Frye SA, Balasingham SV, Lagesen K, Rognes T, Tonjum T. 2009.
743 Genome dynamics in major bacterial pathogens. Fems Microbiology Reviews 33:453-470.
744 58. Chen F, Liu WQ, Eisenstark A, Johnston RN, Liu GR, Liu SL. 2010. Multiple genetic switches
745 spontaneously modulating bacterial mutability. BMC Evol Biol 10:277.
746 59. Li GM. 2008. Mechanisms and functions of DNA mismatch repair. Cell Research 18:85-98.
747 60. Tham KC, Hermans N, Winterwerp HHK, Cox MM, Wyman C, Kanaar R, Lebbink JHG. 2013.
748 Mismatch Repair Inhibits Homeologous Recombination via Coordinated Directional
749 Unwinding of Trapped DNA Structures. Molecular Cell 51:326-337.
750 61. Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA. 2016. Mismatch repair in Gram-
751 positive bacteria. Res Microbiol 167:4-12.

752 62. David SS, O'Shea VL, Kundu S. 2007. Base-excision repair of oxidative DNA damage. Nature
753 447:941-950.
754 63. Lu AL, Li X, Gu Y, Wright PM, Chang DY. 2001. Repair of oxidative DNA damage: mechanisms
755 and functions. Cell Biochem Biophys 35:141-70.
756 64. Agnew W, Barnes AC. 2007. Streptococcus iniae: an aquatic pathogen of global veterinary
757 significance and a challenging candidate for reliable vaccination. Vet Microbiol 122:1-15.
758 65. Millard CM, Baiano JC, Chan C, Yuen B, Aviles F, Landos M, Chong RS, Benedict S, Barnes AC.
759 2012. Evolution of the capsular operon of Streptococcus iniae in response to vaccination.
760 Appl Environ Microbiol 78:8219-26.
761 66. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage
762 WP, Lipsitch M. 2013. Population genomics of post-vaccine changes in pneumococcal
763 epidemiology. Nat Genet 45:656-63.
764 67. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von
765 Gottberg A, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai
766 DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. 2011.
767 Rapid Pneumococcal Evolution in Response to Clinical Interventions. Science 331:430-434.
768 68. Maiden MCJ, van Rensburg MJJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. 2013.
769 MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Micro 11:728-
770 736.
771 69. Shabalina SA, Spiridonov NA, Kashina A. 2013. Sounds of silence: synonymous nucleotides as
772 a key to biological regulation and complexity. Nucleic Acids Res 41:2073-94.
773 70. Symonds MRE, Blomberg SP. 2014. A Primer on Phylogenetic Generalised Least Squares, p
774 105-130. *In* Garamszegi LZ (ed), Modern Phylogenetic Comparative Methods and Their
775 Application in Evolutionary Biology: Concepts and Practice doi:10.1007/978-3-662-43550-
776 2_5. Springer Berlin Heidelberg, Berlin, Heidelberg.
777 71. Zheng Q. 2015. Methods for comparing mutation rates using fluctuation assay data. Mutat
778 Res 777:20-2.
779 72. Benjamini Y, Hochberg Y. 1995. Controlling the False Discovery Rate - a Practical and
780 Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B-
781 Methodological 57:289-300.
782 73. Gould CV, Sniegowski PD, Shchepetov M, Metlay JP, Weiser JN. 2007. Identifying mutator
783 phenotypes among fluoroquinolone-resistant strains of Streptococcus pneumoniae using
784 fluctuation analysis. Antimicrob Agents Chemother 51:3225-9.
785 74. Sprouffske K, Aguilar-Rodriguez J, Sniegowski P, Wagner A. 2018. High mutation rates limit
786 evolutionary adaptation in Escherichia coli. PLoS Genet 14:e1007324.
787 75. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR.
788 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome
789 sequences using Gubbins. Nucleic Acids Res 43:e15.
790 76. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang YJ, Wishart DS. 2016. PHASTER: a better,
791 faster version of the PHAST phage search tool. Nucleic Acids Research 44:W16-W21.
792 77. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference
793 centre for bacterial insertion sequences. Nucleic Acids Research 34:D32-D36.
794 78. Chaguza C, Andam CP, Harris SR, Cornick JE, Yang M, Bricio-Moreno L, Kamng'ona AW,
795 Parkhill J, French N, Heyderman RS, Kadioglu A, Everett DB, Bentley SD, Hanage WP. 2016.
796 Recombination in Streptococcus pneumoniae Lineages Increase with Carriage Duration and
797 Size of the Polysaccharide Capsule. mBio 7:e01053-16.
798 79. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of
799 heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution 2.
800 80. Bedford T, Cobey S, Pascual M. 2011. Strength and tempo of selection revealed in viral gene
801 genealogies. BMC Evol Biol 11:220.

802 81. Neher RA, Hallatschek O. 2013. Genealogies of rapidly adapting populations. Proceedings of
803 the National Academy of Sciences of the United States of America 110:437-442.
804 82. Dupes NM, Walsh BW, Klocko AD, Lenhart JS, Peterson HL, Gessert DA, Pavlick CE, Simmons
805 LA. 2010. Mutations in the Bacillus subtilis beta clamp that separate its roles in DNA
806 replication from mismatch repair. J Bacteriol 192:3452-63.
807 83. Lenhart JS, Sharma A, Hingorani MM, Simmons LA. 2013. DnaN clamp zones provide a
808 platform for spatiotemporal coupling of mismatch detection to DNA replication. Mol
809 Microbiol 87:553-68.
810 84. Yang H, Miller JH. 2008. Deletion of dnaN1 generates a mutator phenotype in Bacillus
811 anthracis. DNA Repair (Amst) 7:507-14.
812 85. Korkmaz G, Holm M, Wiens T, Sanyal S. 2014. Comprehensive analysis of stop codon usage in
813 bacteria and its correlation with release factor abundance. J Biol Chem 289:30334-42.
814 86. Facklam R, Elliott J, Shewmaker L, Reingold A. 2005. Identification and characterization of
815 sporadic isolates of Streptococcus iniae isolated from humans. Journal of Clinical
816 Microbiology 43:933-937.
817 87. Weinstein MR, Litt M, Kertesz DA, Wyper P, Rose D, Coulter M, McGeer A, Facklam R, Ostach
818 C, Willey BM, Borczyk A, Low DE. 1997. Invasive infections due to a fish pathogen,
819 Streptococcus iniae. New England Journal of Medicine 337:589-594.
820 88. Weinberger DM, Malley R, Lipsitch M. 2011. Serotype replacement in disease after
821 pneumococcal vaccination. Lancet 378:1962-73.
822 89. Losordo T, DeLong D, Guerdat T. 2009. Advances in technology and practice for land-based
823 aquaculture systems: tank-based recirculating systems for finfish production. New
824 Technologies in Aquaculture doi:Doi 10.1533/9781845696474.6.945:945-983.
825 90. Kadioglu A, Weiser JN, Paton JC, Andrew PW. 2008. The role of Streptococcus pneumoniae
826 virulence factors in host respiratory colonization and disease. Nat Rev Microbiol 6:288-301.
827 91. Locke JB, Colvin KM, Datta AK, Patel SK, Naidu NN, Neely MN, Nizet V, Buchanan JT. 2007.
828 Streptococcus iniae capsule impairs phagocytic clearance and contributes to virulence in
829 fish. J Bacteriol 189:1279-87.
830 92. Locke JB, Colvin KM, Varki N, Vicknair MR, Nizet V, Buchanan JT. 2007. Streptococcus iniae
831 beta-hemolysin streptolysin S is a virulence factor in fish infection. Diseases of Aquatic
832 Organisms 76:17-26.
833 93. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH.
834 2015. Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev
835 28:871-99.
836 94. Rodriguez JL, Dalia AB, Weiser JN. 2012. Increased chain length promotes pneumococcal
837 adherence and colonization. Infect Immun 80:3454-9.
838 95. Andisi VF, Hinojosa CA, de Jong A, Kuipers OP, Orihuela CJ, Bijlsma JJ. 2012. Pneumococcal
839 gene complex involved in resistance to extracellular oxidative stress. Infect Immun 80:1037-
840 49.
841 96. Bjarnsholt T. 2013. The role of bacterial biofilms in chronic infections. APMIS Suppl
842 doi:10.1111/apm.12099:1-51.
843 97. Pier GB, Madin SH. 1976. Streptococcus iniae sp nov, a beta-hemolytic Streptococcus
844 isolated from an amazon freshwater dolphin, Inia geoffrensis. . International Journal of
845 Systematic Bacteriology 26:545-553.
846 98. Pier GB, Madin SH, Alnakeeb S. 1978. Isolation and characterisation of a 2nd isolate of
847 Streptococcus iniae. International Journal of Systematic Bacteriology 28:311-314.
848 99. Baumler A, Fang FC. 2013. Host specificity of bacterial pathogens. Cold Spring Harb Perspect
849 Med 3:a010041.
850 100. Hildebrand F, Meyer A, Eyre-Walker A. 2010. Evidence of selection upon genomic GC-
851 content in bacteria. PLoS Genet 6:e1001107.

852 101. Couce A, Caudwell LV, Feinauer C, Hindre T, Feugeas JP, Weigt M, Lenski RE, Schneider D,
853 Tenaillon O. 2017. Mutator genomes decay, despite sustained fitness gains, in a long-term
854 experiment with bacteria. Proc Natl Acad Sci U S A 114:E9026-E9035.
855 102. Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, Lipsitch M,
856 Hanage WP. 2015. Selective and genetic constraints on pneumococcal serotype switching.
857 PLoS Genet 11:e1005095.
858 103. Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations. Methods
859 20:4-17.
860 104. Hedges AJ. 2002. Estimating the precision of serial dilutions and viable bacterial counts. Int J
861 Food Microbiol 76:207-14.
862 105. Kawasaki M, Delamare-Deboutteville J, Bowater RO, Walker MJ, Beatson S, Ben Zakour NL,
863 Barnes AC. 2018. Microevolution of aquatic Streptococcus agalactiae ST-261 from Australia
864 indicates dissemination via imported tilapia and ongoing adaptation to marine hosts or
865 environment. Appl Environ Microbiol doi:10.1128/AEM.00859-18.
866 106. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
867 Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner
868 PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell
869 sequencing. J Comput Biol 19:455-77.
870 107. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome
871 assemblies. Bioinformatics 29:1072-5.
872 108. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of
873 draft genomes using the Mauve aligner. Bioinformatics 25:2071-3.
874 109. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-9.
875 110. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-
876 genome alignment and visualization of thousands of intraspecific microbial genomes.
877 Genome Biology 15:524.
878 111. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
879 large phylogenies. Bioinformatics 30:1312-3.
880 112. Leache AD, Banbury BL, Felsenstein J, de Oca AN, Stamatakis A. 2015. Short Tree, Long Tree,
881 Right Tree, Wrong Tree: New Acquisition Bias Corrections for Inferring SNP Phylogenies.
882 Systematic Biology 64:1032-47.
883 113. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: An
884 interactive viewer for large phylogenetic trees. BMC Bioinformatics 8:460.
885 114. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. 2016. Evolview v2: an online visualization
886 and management tool for customized and annotated phylogenetic trees. Nucleic Acids Res
887 44:W236-41.
888 115. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A,
889 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic:
890 An integrated and extendable desktop software platform for the organization and analysis of
891 sequence data. Bioinformatics 28:1647-1649.
892 116. Ben Zakour NL, Alsheikh-Hussain AS, Ashcroft MM, Khanh Nhu NT, Roberts LW, Stanton-
893 Cook M, Schembri MA, Beatson SA. 2016. Sequential Acquisition of Virulence and
894 Fluoroquinolone Resistance Has Shaped the Evolution of Escherichia coli ST131. MBio
895 7:e00347-16.
896 117. Umarov RK, Solovyev VV. 2017. Recognition of prokaryotic and eukaryotic promoters using
897 convolutional deep learning neural networks. Plos One 12.
898 118. Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic
899 Acids Res 46:D493-D496.
900 119. Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino
901 acid substitutions and indels. Bioinformatics 31:2745-7.

902 120. Petkau A, Stuart-Edwards M, Stothard P, Van Domselaar G. 2010. Interactive microbial
903 genome visualization with GView. Bioinformatics 26:3125-6.
904 121. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG):
905 simple prokaryote genome comparisons. BMC Genomics 12:402.
906 122. Rajagopal L. 2009. Understanding the regulation of Group B Streptococcal virulence factors.
907 Future Microbiol 4:201-21.
908 123. Andrews JM. 2001. Determination of minimum inhibitory concentrations. J Antimicrob
909 Chemother 48 Suppl 1:5-16.
910 124. Verotta D, Haagensen J, Spormann AM, Yang K. 2017. Mathematical Modeling of Biofilm
911 Structures Using COMSTAT Data. Comput Math Methods Med 2017:7246286.
912 125. Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation analysis CalculatOR: a web tool for the
913 determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics
914 25:1564-5.
915 126. Zheng Q. 2017. rSalvador: An R Package for the Fluctuation Experiment. G3 (Bethesda)
916 7:3849-3856.

917

918

919 **Supporting Information Legends**

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

926 Supplementary Figure 2. Root-to-tip regression analysis (TempEst, (79)of branch lengths from best-fit-
927 rooted tree against time (year of isolation). The tree was derived from alignment of non-recombinant
928 core-genome SNPs corrected for ascertainment bias in RAxML.

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

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

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

938 Supplementary Table 3. Extended metadata for 80 *Streptococcus iniae* isolates sequenced in the
939 present study

940 Supplementary Table 4. Assembly statistics and quality data for 80 *de novo* genome assemblies
941 prepared in this study.

942 Supplementary Table 5. Recombination analysis (Gubbins) statistics for 83 *S. iniae* strains, ordered by
943 Clade.

944 **Figures and tables**

945 Figure 1: Unrooted summary phylogram of *Streptococcus iniae* strains based on alignment of core
946 genome SNPs filtered to remove recombination. Genetic distances were inferred by maximum
947 likelihood in RAXML and node support is indicated as percentage of 1000 bootstrap replicates.
948 Geographic origin, time range (year) of isolation, and mutation rate were added manually *post hoc*. All
949 nodes with bootstrap support below 75% were collapsed. The phylogram is drawn with scale

950 proportionate to genetic distance except the dashed branch supporting QMA0141, which is depicted at
951 10% scale. The inset shows the phylogram structure with QMA0141 branch drawn to the same scale.

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

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

963 Figure 4: Proportions of single nucleotide polymorphism type across the core genomes of *Streptococcus*
964 *iniae* clades. Numbers to the right of the plot indicate the total numbers of SNPs that are unique to
965 each clade.

966

967 Table 1. *Streptococcus iniae* strains used in this study. Includes origin details (host species, time, site
968 of isolation), phylogenetic affiliation, virulence-associated phenotypes, and mutation rate. Atypical
969 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

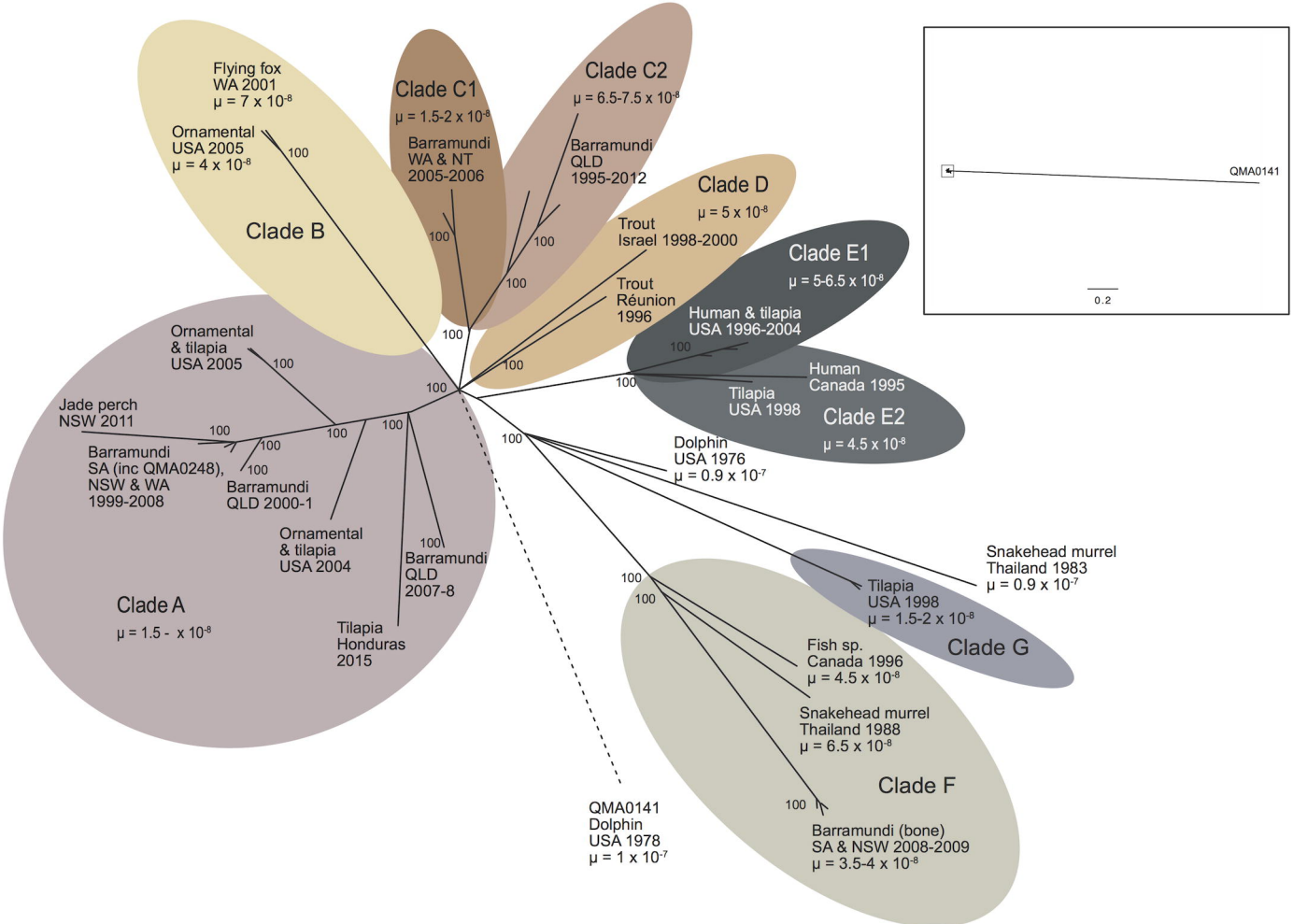
972 Table 2: Variants in MMR and OG genes found among *S. iniae* isolates. CDS – protein coding
973 sequence, BS – predicted binding sites of transcription factors, PROVEAN score – predicted effect of
974 amino acid substitutions on the protein function; scores lower that -2 are considered to indicate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

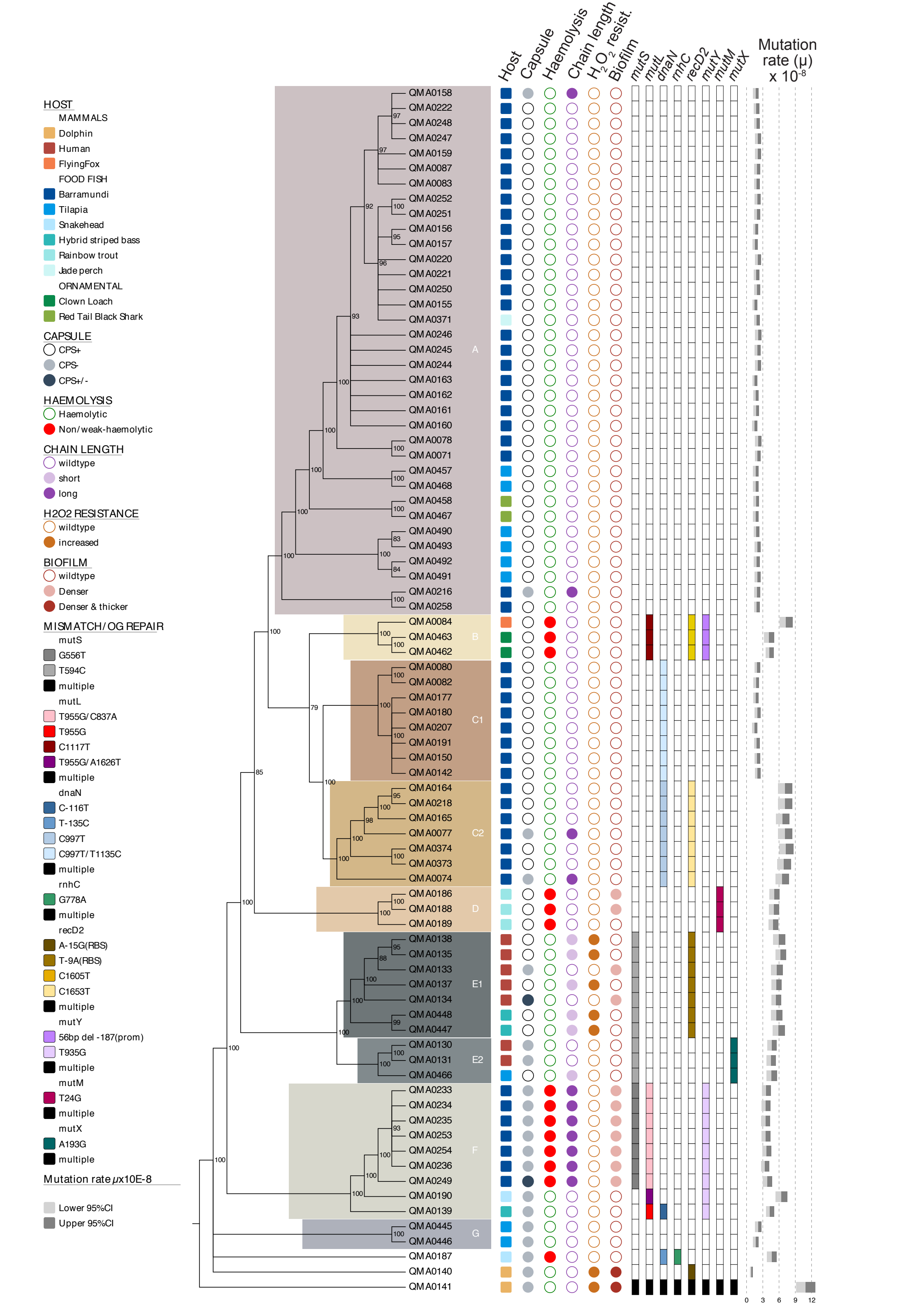
976

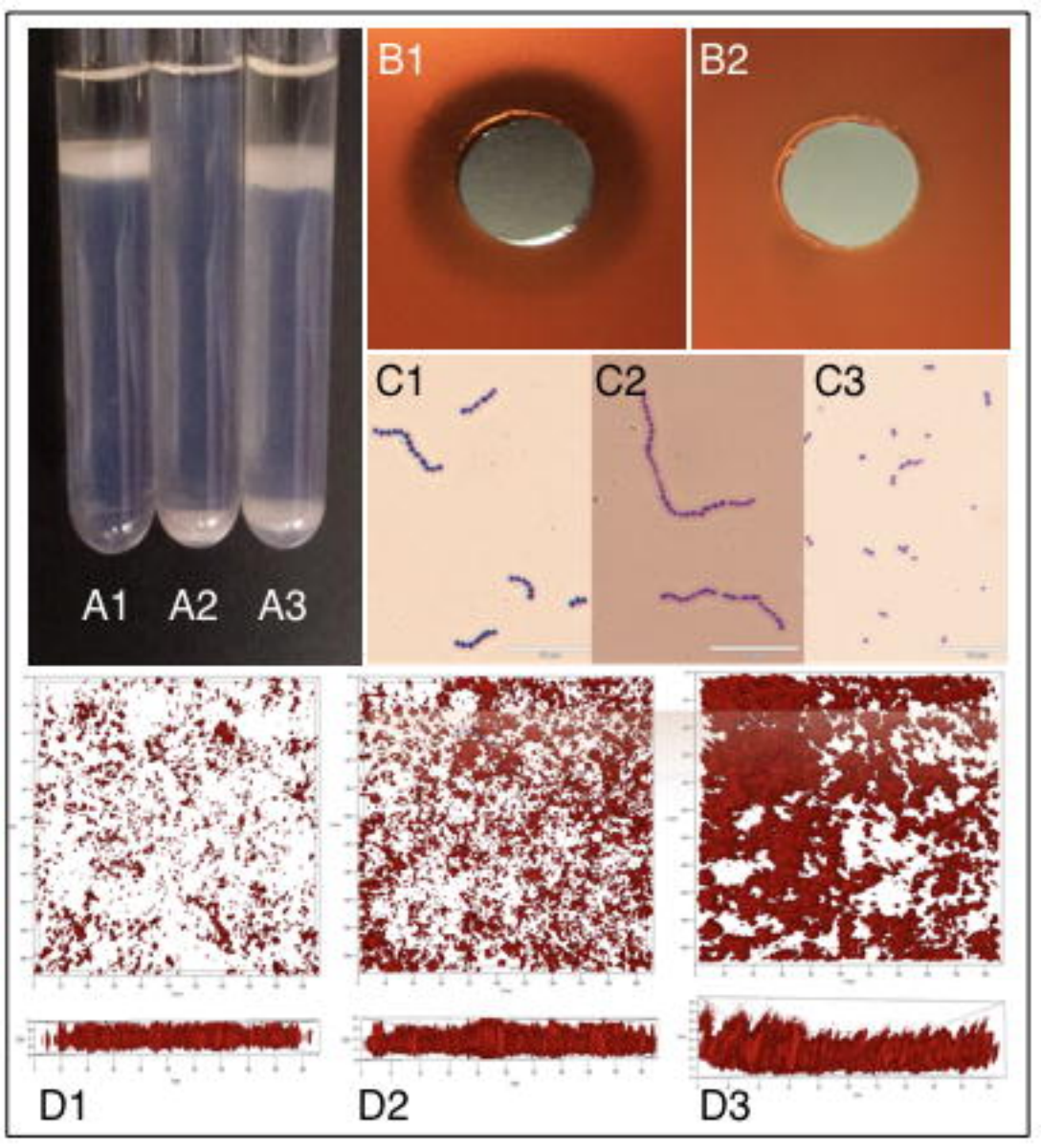
Gene	Clade	Strains	Location	Mutation type	Nucleotide change	Amino acid change	Protein functional domain	PROVEAN score
mutS	F	from bone	CDS	SNP	G556T	Q -> K	Pfam MutS_II	-0.635
	E	all	CDS	SNP	T594C	-	Pfam MutS_II	n/a
mutL	F	from bone	CDS	SNP	C837A	-	DNA mis_repair	n/a
	F	all	CDS	SNP	T955G	S -> R	DNA mis_repair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

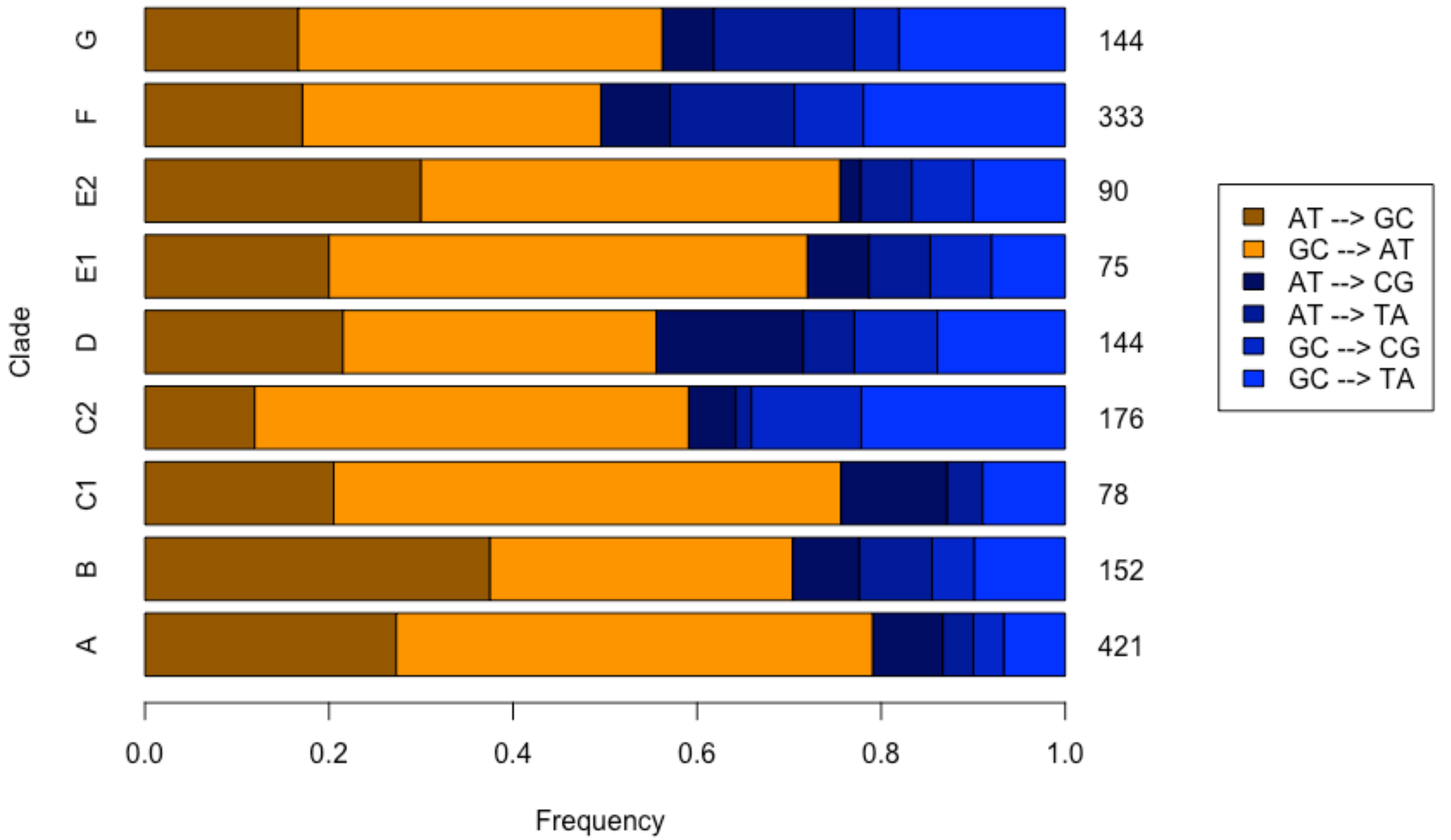
977

978


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


Host
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
G778A
multiple
recD2
A-15G(RBS)
T-9A(RBS)
C1605T
C1653T
multiple
mutY
56bp del -187(prom)
T935G
multiple
mutM
T24G
multiple
mutX
A193G
multiple
Mutation rate μx10E-8
Lower 95%CI
Upper 95%CI
A
B
C1
C2
D
E1
E2
F
G
0
3
6
9
12


A1
A2
A3
B1
B2
C1
C2
C3
D1
D2
D3


Clade
G
F
E2
E1
D
C2
C1
B
A
144
333
90
75
144
176
78
152
421
0.0
0.2
0.4
0.6
0.8
1.0
Frequency
AT --> GC
GC --> AT
AT --> CG
AT --> TA
GC --> CG
GC --> TA