Long title:

Mutators as drivers of adaptation in pathogenic bacteria and a risk factor for host jumps and vaccine escape: insights into evolutionary epidemiology of the global aquatic pathogen *Streptococcus iniae*

Short title:

Mutators as a risk factor for host jumps and immune escape in pathogenic bacteria

Oleksandra Silayeva[1,2], Jan Engelstaedter[1], and Andrew C Barnes[1,2*]

[1]The University of Queensland, School of Biological Sciences and [2]Centre for Marine Science, 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

Hypermutable strains (mutators) occur via mutations in DNA repair genes and facilitate microbial adaptation as they may rapidly generate stress-resistance mutations. Bacterial mutators deficient in mismatch repair (MMR) and oxidised guanine repair (OG) are abundant in clinical settings, however their role in epidemiology and evolution of virulence remains poorly understood. Here we determine phylogenetic relationship among 80 strains of *Streptococcus iniae* isolated globally over 40 years based using non-recombinant core genome single nucleotide polymorphisms (SNPs), estimate their mutation rate by fluctuation analysis, and identify variation in major MMR (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*) and OG (*yfhQ*, *mutM*, *mutT*) genes. We find that *S. iniae* mutation rate phenotype and genotype are strongly associated with phylogenetic diversification and variation in major streptococcal virulence determinants (capsular polysaccharide, hemolysin, cell chain length, resistance to oxidation, and biofilm formation). Furthermore, we find that 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, to new host tissue, and to immunity of a vaccinated host may be facilitated by mutator strains. Our findings highlight the importance of mutators and 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.

Author Summary

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.

Introduction

Optimal mutation rates in microbial populations are determined by a trade-off between maintenance of genetic integrity and evolvability [1, 2]. High fidelity of DNA repair is advantageous in adapted populations since most mutations lead to deviation from the evolved phenotype [2, 3]. Conversely, lower fidelity of DNA repair can be advantageous under stressful conditions since increased mutation supply promotes diversity and may thereby accelerate adaptation [2, 4]. Bacteria can efficiently fluctuate between high and low mutation regimes: While extremely low mutation rates are maintained in favourable conditions [5], elevation of mutation rates known as stress-induced mutagenesis (SIM) occurs in adapting populations [6, 7]. SIM is associated with temporary hypermutability during SOS response and increased frequency of heritable hypermutable strains (mutator strains, or simply mutators) [6, 7]. Mutators periodically arise within a population via mutations in DNA repair genes [8, 9], and become prevalent under stress as, in a clonal population, mutator alleles 'hitchhike' with beneficial mutations that arise at increased rate in the mutator genetic backgrounds [10, 11]. When adaptation is gained low mutation rate can be restored via anti-mutator compensatory mutations [4].

Mutator strains are abundant among bacterial clinical isolates, especially in chronic cases [12, 13]. Multiple studies have attributed this to antibiotic selection [14-16], since mutators are a well-established risk factor for the development of antimicrobial resistance [17, 18]. However, the association between frequency of mutators in diagnostic samples and antibiotic treatments appears to be weak [19, 20], and isogenic knockout mutator strains show an increased ability to colonize an animal host in challenge models devoid of antibiotic exposure [21-23]. Thus, the prevalence of mutators in clinical isolates most likely results from the selective process during adaptation to the host [19, 24-26], where immunity acts as the ultimate selective pressure shaping the diversity of pathogenic strains [27]. The 'Immunological niche' concept states that, due to the diversity and complexity of immune responses, every individual host represents a unique habitat that requires adaptation [28]. Increased mutation supply in a mutator strain could allow rapid adjustment to the particular immunological environment found within each host. Considering that the pathogen needs to adapt to every new host, the host population taken as a whole represents an extremely heterogeneous environment where adaptive processes are ongoing and persistence of mutator alleles is favoured [29]. The fitness of a pathogenic strain in the host hinges on the concept of virulence (the degree of damage to the host), which is a dynamic, complex trait determined by a multitude of virulence factors and host responses [30]. Optimal virulence appears to be a trade-off between acuteness and persistence with higher virulence facilitating extraction of resources and transmission, but too much damage can kill the host and stop transmission altogether [31]. Additionally, virulence factors are often antigens activating the immune response, so lower levels of virulence may be a means of immune evasion [32]. Consequently, both loss and gain of virulence may be

81 advantageous, and indicate adaptive shifts towards acute or chronic pathogenesis, respectively [33, 34].
82 Exacerbation and attenuation of virulence factors are repeatedly observed in both knockout and naturally
83 occurring mutators [25, 35-39]. Thus, it appears that mutator strains/alleles may help pathogenic bacteria
84 to evolve towards the optimal level of virulence.

85 Most bacterial mutators isolated from human clinical samples are deficient in genes from mismatch
86 (MMR) and oxidised guanine (OG) DNA repair systems [12]. MMR and OG genes are conserved across all
87 domains of life [40], and single nucleotide polymorphisms in their sequence may produce drastic
88 alterations in mutation rate [41, 42]. Nonetheless, these genes appear to be variable in major bacterial
89 pathogens, confirming a potential role of mutation rate dynamics in the evolution of virulence [43, 44].
90 MMR is coupled to replication and removes mispaired nucleotides as well as short insertion/deletion
91 loops, and prevents recombination between divergent sequences [45, 46].The MMR pathway has been
92 extensively characterised in *Escherichia coli,* where the MutS dimer protein interacts with the replication
93 processivity clamp DnaN (b-clamp), binds to the mismatch, recruits MutL, and the MutS:MutL complex
94 then activates the MutH endonuclease [40]. MutH cuts the unmethylated nascent strand at
95 hemimethylated d(GATC) sites. MutL loads the UvrD helicase that unwinds the nicked DNA containing
96 the error, which is followed by excision by one of the single-strand nucleases [40]. This MMR organization,
97 however, appears to be limited to *E. coli* and closely related Gammaproteobacteria [40, 47]. In most
98 bacteria and eukaryotes, MutH is absent and MutL has a conserved endonuclease site, dam methylase
99 (acting at GATC sites) is absent and the strand discrimination signal is unknown [40, 47]. Studies in *Bacillus*
100 *subtilis,* a model species for gram-positive bacteria, suggest that nicks created by ribonuclease *rnhC*
101 during removal of misincorporated ribonucleosides direct MutL endonuclease towards the newly
102 synthethised strand, and that recD2 acts as a MMR helicase instead of uvrD [47]. The OG system repairs
103 oxidative DNA damage throughout the cell cycle [48, 49]. Among all DNA bases, guanine is most sensitive
104 to reactive oxygen species and its oxidised form, 8-oxo-dG or OG, is highly mutagenic as it preferentially
105 pairs with adenine which, if not corrected, leads to G:C to A:T transversion on a second round of
106 replication [48, 49]. OG system organization is conserved in most bacteria and eukaryotes and generally
107 MutT homologs remove 8-oxo-dG from the nucleotide pool, MutM homologs excise 8-oxo-dG
108 incorporated into DNA, and MutY homologs remove adenine from OG:A mispairings [48, 49].

109 The present study investigates the role of mutators and mutation rate dynamics in the epidemiology and
110 evolution of pathogenic bacteria using the broad host-range pathogen *Streptococcus iniae*. *S. iniae* is a
111 model highly adaptable species, evidenced by its global distribution, ability to cross divergent host taxa
112 boundaries and repeated re-infection of previously immunised hosts during disease outbreaks in
113 vaccinated fish in aquaculture [50, 51]. Moreover, Streptococcal pathogens in general are difficult
114 candidates for immunisation in humans and animals due to high strain diversity evidently driven by

115 immunisation, yet there is little information on the role of mutators in this evolutionary process [52, 53].
116 We determine variation in mutation rate and identify mutations in MMR (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*)
117 and OG (*yfhQ*, *mutM*, *mutT*) genes among 80 diverse strains of *S. iniae* and discover that they correlate
118 with phylogenetic diversification but are highly conserved within phylogenetic clades. To identify
119 whether adaptive process might be driven by the host immune response we identify variation in major
120 phenotypic traits that determine level of virulence in streptococci (capsular polysaccharide, hemolysin,
121 length of cell chains) and bacteria in general (resistance to reactive oxygen species (ROS), and biofilm
122 formation). Two major streptococcal virulence factors in *S. iniae* infection are the polysaccharide capsule
123 and the beta-hemolysin streptolysin S [54-57]. Both increase survival under native immune response via
124 reduction of phagocytic killing and other mechanisms, however, both are antigens that induce antibody
125 production and development of immune memory [54, 55]. In addition, both capsule production and
126 hemolysin activity have many implications that are not fully understood [55, 58]. For example, non-
127 encapsulated strains exhibit greater ability to adhere to and colonize an epithelium [59]. Length of cell-
128 chains is another phenotype linked to virulence in streptococci: shorter chains are associated with acute
129 infection and increased survival in the blood as they are less likely to activate the complement, whilst
130 longer chains promote adherence and colonization [60]. As pathogenic bacteria are continuously exposed
131 to reactive oxygen released by immune cells (a phenomenon known as 'oxidative burst'), resistance to
132 oxidative compounds contributes to virulence in pathogens allowing them to avoid immune clearance
133 [61, 62]. Another feature generally associated with virulence in bacteria is biofilm forming ability:
134 detached cells are associated with acute infections whereas biofilm aggregates are a hallmark of chronic
135 conditions [63]. The expression of these traits among *S. iniae* isolates is highly consistent with
136 phylogenetic affiliation, which confirms that strain diversity is primarily driven by adaptation to the host
137 achieved via adjustment of virulence. Furthermore, we find that occurrence of abnormal virulence
138 determinants is strongly associated with number of mutations in MMR and OG genes and occurrence of
139 infection in unusual immunological landscapes such as mammals and bone tissue of immunised fish.
140 These data support the critical role of mutators in adaptation to the host and evolution of virulence.
141 Mutators also facilitate the reinfection of immunised animals and transmission between divergent host
142 species.

143

144

145 **Results and Discussion**

146 High mutation rates are costly due to inevitable accumulation of deleterious changes. Once adaptation
147 is gained the mutator strain is predicted to evolve back towards a low mutation rate [64], most likely via
148 compensatory, 'anti-mutator' mutations in DNA repair genes [4, 65]. Considering this volatility of mutator
149 phenotypes, their involvement in adaptation cannot be accurately assessed by *de facto* presence among
150 sporadic isolates, as the latter might have already (partially or fully) adapted and restored a lower
151 mutation rate. However, mutator and anti-mutator mutations that cause adaptive fluctuations in the
152 fidelity of DNA repair could be identified in the genotype of an adapted strain. Also, after loss and gain of
153 functional DNA repair alleles, the mutation rate of an evolved strain is not likely to be exactly the same
154 as before adaptation. Consequently, variation in DNA repair genes and mutation rate determined
155 phenotypically by fluctuation analysis might indicate adaptation via a mutator state when placed in a
156 phylogenetic context.

157 Our work investigates the link between mutation rate genotype and phenotype and phylogenetic
158 diversification among *Streptococcus iniae* strains isolated globally between 1976 and 2016 from different
159 host taxa. Maximum likelihood phylogenetic analysis of 80 *S. iniae* isolates (Table 1) based on non-
160 recombinant core genome SNPs derived from whole genome data resolved 6 major clades (A-F), one
161 lineage with two strains (clade G), and three lineages with a single strain (Figure 1). Principally, variation
162 in MMR and GO genes occurs only between the lineages, and they are highly conserved within clades
163 (Figure 2, Table 2). In fact, sequence of the major MMR and GO genes is conserved even within
164 phylogenetic groups not readily definable by geographic origin, time, or host species, and these genes
165 appear to be good candidates for Multilocus Sequence Typing (MLST, identification of phylogenetic
166 position based on selected genes as opposed to a whole genome analysis). Although even synonymous
167 SNPs in any gene locus could affect mRNA and have pronounced phenotypic consequences [66], most
168 SNPs were identified in protein functional domains, and most amino acid substitutions were predicted to
169 have a deleterious effect on protein function (Table 2). Further, MMR and GO variation correlates
170 significantly with mutation rate (p = 0.0056) and deviations from wildtype traits contributing to virulence
171 (p = 0.0331). In turn, mutation rate is highly consistent with phylogenetic affiliation, and mutation rate
172 difference is classified as insignificant in pair-wise strain comparisons within major phylogenetic lineages
173 but significant between (Figure 1, 2).

174 Isolate QMA0248 (Clade A, used as a reference genome for contig re-ordering in *de novo* genome
175 assembly and read-mapping for SNP detection) a phenotype mutation rate of 1.9×10^{-8}, which is
176 consistent with a base mutation rate estimated for other non-mutator gram positive bacteria [12, 67].
177 Similar mutation rates (within $1.5 - 2 \times 10^{-8}$) are maintained by all isolates (35 strains) along the same

178 ancestral lineage as QMA0248 (Fig 1, 2; clade A), and all the differences are insignificant except three
179 pairwise comparisons that verge on commonly accepted significance level. This lineage appears to have
180 persisted globally in a range of hosts for at least two decades, as it contains strains isolated from 1999 to
181 2016 in USA, Honduras, and Australia from barramundi (*Lates calcarifer*), tilapia (*Oreochromis sp*), jade
182 perch (*Scortum barcoo*), and ornamental fish (*Epalzeorhynchos frenatum, Epalzeorhynchos bicolor*).
183 MMR and GO genes in these strains are identical to QMA0248, and only minor deviations from wildtype
184 traits contributing to virulence were observed among the strains (absence of capsule production and
185 longer cell chains in QMA0158, 216) (Figs. 1-3). The rest of the lineages express significantly different
186 mutation rates (except strains from clade C1 and QMA045-46 discussed below), contain unique and often
187 multiple SNPs in *mut* genes, and show peculiar phenotypic variants related to virulence (Figures 2, 3).
188 Although phylogenetic relatedness between lineages is fully resolved, they originate almost
189 simultaneously and nested lineages evolve independently sharing little phylogenetic history (Figure 1,2).
190 Clade B contains QMA0084, a strain isolated from a black flying fox, *Pteropus alecto*, in Western Australia
191 (WA) and two strains isolated from ornamental fish species (clown loach or tiger botia, *Chromobotia*
192 *macracanthus*) in USA. These isolates share valine to isoleucine substitution in MutL, a synonymous SNP
193 in *recD2*, and 56 bp deletion within the promoter of *yfhQ*. The latter mutation is predicted to drastically
194 changes *yfhQ* transcription pattern, affecting transcription factor binding sites of the original promoter
195 and introducing a second promoter downstream. It appears that transcription is also affected by the rest
196 of the genetic background as isolates exhibit significant difference in mutation rate phenotype, 7 x 10^{-8}
197 in QMA0084 and 4 x 10^{-8} in clown loach strains. Phenotypically, clade B has shifted towards decreased
198 haemolytic activity. Clade B shares a common ancestor with clade C comprised of strains obtained from
199 barramundi farmed in WA, Northern Territory (NT) (nested clade C1), and north QLD fish farms (nested
200 clade C2). While a core mutation rate of 1.5-2 x 10^{-8} is observed in clade C1, a significantly higher mutation
201 rate of 7 x 10^{-8} has been maintained in sister clade C2 from north QLD for almost two decades (1995 to
202 2012) (Fig 1). We speculate that the latter might be attributable to difference in meteorological
203 conditions: periodic heavy rainfall present in tropical north QLD is likely to bring fluctuations in water
204 salinity and temperature, creating an unstable environment where adaptive processes are ongoing and
205 high mutation rate is advantageous. Minor variation in virulence-related phenotypes (absence of capsule
206 production and longer cell chains in QMA0074, 77)(Fig 2), are supportive that mutation rate evolution in
207 this lineage might be driven by outside-the-host factors. The significant difference in mutation rate
208 between nested clades C1 and C2 is potentially attributable to variation in *dnaN*: A predicted deleterious
209 tyrosine to isoleucine substitution (Table 2, PROVEAN score -2.216) is present in all isolates of C2, which
210 may explain the high mutation rate phenotype in clade C2 strains (which also have a synonymous variant
211 in *recD2*)(Fig. 2, Table 2). Indeed, deletion of *dnaN1* in *Bacillus anthracis* results in a mutator phenotype

212 with equivalent mutation rate to a *mutS* mismatch repair-defective strain, supportive of the connection
213 between deleterious mutation in DnaN and mutator phenotype found here [68]. Moreover, the T326I
214 substitution in *S. iniae* clade C2 DnaN occurs in one of the critical residues of the loader binding interface
215 of the ß-clamp, which may impede interaction with MutS. While MutS can perform mismatch repair
216 independently of DnaN in *B. subtilis*, 90% of mismatch repair is dependent on targeting MutS to nascent
217 DNA via the DnaN clamp zone [69, 70]. In contrast the T1135C nucleotide substitution in the *dnaN* stop
218 codon is unique to clade C1 strains and changes stop codon TAA into CAA coding for glutamine, with the
219 next stop codon TAG found immediately downstream. This mutation does not appear to affect mutation
220 and was not predicted to have any affect on the encoded protein function (Table 2) via single amino acid
221 protein elongation [71] producing a mutation rate that is insignificantly different from the core rate found
222 in clade A (Fig 2). Clade D consists of isolates from trout (*Oncorhynchus mykiss*) from Réunion and Israel.
223 These strains have an estimated mutation rate of 5 x 10^{-8}, contain a glutamate to aspartate substitution
224 in *mutM* predicted to be deleterious to protein function, a synonymous SNP in *RecD2*, and exhibit
225 impeded haemolytic activity (Fig 1,2, Table 2).

226 Isolates from humans and related fish strains are found in clade E where multiple phenotypic variants
227 contributing to virulence are observed. This clade contains two nested subclades that share a SNP in
228 *mutS,* but have other unique SNPs and different mutation rates, which is significant in most pairwise
229 comparisons. Subclade E1 contains USA isolates from human (QMA0133-35, 37-38) and hybrid striped
230 bass (QMA0447-48), which have a substitution in the Shine-Dalgarno sequence of *recD2* and a mutation
231 rate of 6.5 x 10^{-8} (Fig 1, Table 2). Fish strains and three human strains (QMA0135, 37-38) show a decreased
232 ability to form cell chains and an elevated ability to withstand oxidative stress. Two human strains
233 QMA0133-34 form somewhat thicker and considerably denser biofilm structures. QMA0133 is non-
234 encapsulated, and QMA0134 appears to differentially express the capsular polysaccharide in a culture.
235 The close phylogenetic relationship presented here implicates a likely transfer from farmed hybrid bass
236 to humans [72], and this host jump may have been facilitated by the mutator phenotype. Subclade E2
237 contains two human strains from Canada (QMA0130-31) and tilapia strain from USA (QMA0466). These
238 strains have unique methionine to isoleucine substitution at the N-terminus of the MutX, and mutation
239 rate of 4.5 x 10^{-8} (Fig 1, Table 2). Both human strains are non-encapsulated (Fig 2). The tilapia strain
240 expresses the capsule but forms short cell chains in common with most strains from the subclade (Fig 2).
241 The close phylogenetic relationship again strongly implicates transfer from tilapia farmed in USA to
242 human patients in Canada via import into the Toronto fish market [73] and again, this host jump may
243 have been facilitated by the mutator state. Clade F is not readily definable by location, time of isolation,
244 or host species and comprises QMA0139 from unknown fish species obtained in Canada, QMA0190
245 isolated from snakehead murrel in Thailand (both of which appear on long branches), and a nested

246 terminal clade comprised of barramundi isolates farmed in Recirculating Aquaculture Systems (RAS) in
247 New South Wales and South Australia. The latter are of a particular interest as they were sampled from
248 barramundi bone lesions during a disease outbreak in vaccinated fish where infection manifested itself
249 as a slowly progressing osteomyelitis instead of typical acute septicaemia and meningitis [50, 51]. Since
250 they are distant to strains used in autogenous vaccines that are found in clade A (QMA0155-57, 160,
251 QMA0250-52), the atypical outbreak might be classified as a case of vaccine-induced serotype
252 replacement (VISR) – spread of co-existing pathogenic strain/s after elimination of the dominant strain/s
253 targeted by the vaccine [74]. In contrast to natural populations where the investigation of VIRS is
254 confounded by ecological and sampling biases, and randomised vaccine trials that lack power to detect
255 the population-wide effect of mass vaccination [74], barramundi RAS represent a relatively controlled
256 and isolated natural environment where vaccination against *S. iniae* fails occasionally but recurrently
257 [51]. RAS are self-contained facilities with minimal water exchange and high standards of water
258 treatment [75]. This ensures that gene flow and, consequently, the probability of new serotype
259 introduction after vaccine implementation is reduced to a minimum. A comparatively uniform immunized
260 host population is exposed to dominant strains against which it was vaccinated along with potentially co-
261 existing lineages in the RAS. In the atypical outbreaks in RAS in Australia, vaccine-induced immunity was
262 partially effective in the infected fish, evident by absence of proliferation of the new strain in typical
263 niduses (blood, brain, and pronephros) and cross-reactivity of vaccine-induced antibodies against bone
264 isolates [51]. This is in contrast with the classical model of serotype replacement where a vaccine is
265 ineffective against the replacing co-existing strain. It has been postulated that VISR can occur even if the
266 vaccine is equally efficient against all circulating strains via other trade-off mechanisms [74]. Apparently,
267 evasion from immune clearance and adaptation to osseous tissue was allowed by shift towards
268 attenuated virulence as suggested by multiple phenotypic changes in bone isolates: absence of capsular
269 production, impeded hemolytic activity, increased cell-chain length, and denser biofilms. All of these
270 changes have previously been associated with chronic infection and increased potential for colonization
271 [55-58, 60, 62, 63], and only absence of polysaccharide capsule is shared with related QMA139 and
272 QMA190. Multiple SNPs in repair genes and the high mutation rate in bone isolates suggest that shift in
273 virulence allowing persistence in the bone tissue might have occurred via mutator phenotype. Serine to
274 arginine substitution in MutL and tyrosine to isoleucine substitution in MutY occur in the ancestor of
275 QMA0139, QMA0190, and bone strains. Both substitutions are predicted to have a deleterious effect on
276 protein function, but their effect on mutation rate is combined with unique SNPs found in each branch:
277 QMA0139 has a SNP in the binding site of the DnaA transcription factor within the *dnaN* promoter, and
278 a mutation rate of 4.5×10^{-8}, QMA0190 has synonymous SNP in *mutL* and mutation rate of 6.5×10^{-8}, and
279 bone strains have glutamine to glycine substitution in MutS and synonymous SNP in *mutL*, and a mutation

280 rate of 3-4 x 10^{-8}. Considering that the mutation rate of QMA0190 is significantly higher compared to
281 rates expressed by QMA0139 and bone isolates, it is likely that variants unique to the latter strains
282 compensate for deleterious variants shared by the isolates [4].

283 Clade G contains two strains from tilapia isolated in USA (QMA0445-46). Notably, despite being
284 phylogenetically distant from clade A, these strains have the same MMR and GO genotype and have
285 retained the core mutation rate and the wildtype virulence phenotype. The first long branch with a single
286 isolate contains a strain from snakehead murrel strain (*Channa striata*) from Thailand. This isolate is non-
287 encapsulated and weakly haemolytic and has a mutation rate of 4.5 x 10^{-8}, potentially attributable to
288 deleterious aspartate to asparagine substitution in rnhC and SNP in the binding site of *fnr* transcription
289 factor within the *dnaN* promoter. A second long branch with a single strain contains the oldest among
290 the analysed strains, QMA0140 isolated in 1976 from dolphin (*Inia geoffrensis*)[76]. This strain exhibits a
291 mutation rate of 1 x 10^{-8}, which is unique among the strains and significantly lower than the core mutation
292 rate, presumably associated with increased translation rate of recD2 helicase produced by a SNP in the
293 ribosome binding site (Fig 2, Table 2). The longest branch on the tree contains a second dolphin strain,
294 QMA0141, isolated two years later in 1978 [77]. This isolate is highly divergent from the rest of strains
295 with around 20 kB of non-recombinant SNPs in pair-wise comparisons with other strains, accounting for
296 around 1% genomic difference. In contrast to QMA140, it mutates at a rate of 1x 10^{-7}, the highest
297 mutation rate phenotype determined among the isolates and significantly different to the rest of the
298 values. Multiple SNPs are observed in all MMR and GO genes: 3 in *dnaN* and *mutX*, 4 in *mutM*, 9 in *mutL*
299 and *rnhC*, 15 in *recD2*, 9 in *mutL* and *rnhC*, and 68 in *mutS*. Both dolphin isolates are non-encapsulated,
300 show increased ability to withstand oxidative stress, and form denser and thicker biofilms (Fig 2, 3).

301 The conservation of *S. iniae* DNA repair genotypes and phenotype within phylogenetic clades and
302 variation between the lineages is consistent with the idea of adaptive fluctuations in mutation rate being
303 a driver of pathogen evolution, and their association with variation in virulence traits supports the
304 contention that adaptive processes and diversification of pathogens are largely driven by host immunity.
305 Moreover, infections in atypical hosts (mammals) and atypical tissues (bone) within apparently immune
306 animals correlates strongly with the degree of deviation from the wildtype virulence phenotype (p=0.000)
307 and the number of variants in DNA repair genes (p=0.002). This indicates that adaptation to a drastically
308 different immunological landscape might be facilitated by elevated mutation supply permitting rapid
309 adjustment of virulence and antigenicity, which has major epidemiological implications. First, mutators
310 might facilitate pathogen transition among divergent host species known as 'host jumps', including higher
311 host taxa (eg. *Osteichthyes* and *Mammalia*) [78] . Second, mutators might present a risk factor to immune
312 escape after vaccination and serotype replacement. Moreover, as evidenced by *S. iniae* colonization of
313 vaccinated barrumundi bone described previously [51], vaccine escape leading to serotype replacement

314 might occur without major disruption of immune recognition when multiple changes attenuating
315 virulence and antigenicity allow persistence in a tissue with lower immune surveillance .

316 To summarize, we estimate mutation rates among the strains by fluctuation analysis, identify the
317 potential genetic bases for the observed differences by comparing MMR and GO genes and their
318 regulatory regions, highlight correlating phenotypic changes associated with virulence, and interpret the
319 data in a context of the phylogeny and the case histories of strain isolation. Our results highlight the role
320 of mutators and mutation rate dynamics in evolution and epidemiology of pathogenic bacteria. In
321 particular, mutators associate with major adaptive events leading to colonization of a new host tissue,
322 infection of novel host taxa, and survival in vaccinated host with specific adaptive immune response.
323 Further experimental studies are required to determine how generalizable these findings are to other
324 pathogens and the evolution of novel strains. It would be of particular interest to conduct similar studies
325 in the Pneumococcus where serotype evolution following vaccination programmes is well documented,
326 but the role of mutators in this process is not yet known [52, 53, 79].

327

328 **Materials and methods**

329 **Strains and growth conditions**

330 Eighty isolates of *S. iniae* collected in Australia, USA, Canada, Israel, Honduras, and Thailand between
331 1976 and 2016 from eight fish species (*Lates calcarifer*, *Scortum barcoo, Epalzeorhynchos frenatum,*
332 *Epalzeorhynchos bicolor, Oreochromis sp., Channa striata, Chromobotia macracanthus, Oncorhynchus*
333 *mykiss*) and three mammalian species (*Homo sapiens, Inia geoffrensis, Pteropus alecto*) were used in this
334 study (Table 1). Strains were received from culture collections, veterinarians, or directly from fish farms
335 (Supplementary Table 1) and stored as master seed stocks without further subculture at -80°C in Todd-
336 Hewitt Broth (THB) supplemented with 20% glycerol. Bacteria were routinely recovered from -80°C stocks
337 on Columbia agar supplemented with 5% defibrinated sheep blood (Oxoid, Australia), and cultured at
338 28°C on Todd-Hewitt agar or in Todd-Hewitt broth (Oxoid) with agitation 200 rpm in a shaking incubator
339 unless otherwise specified.

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

341 To estimate mutation rates of *S. iniae* isolates, a fluctuation analysis assay for spontaneous occurrence
342 of rifampicin resistance was optimized according to Rosche and Foster [80]. A single broth culture was
343 initiated from five separate colonies, recovered on Columbia blood agar from stock cultures stored at
344 -80°C, and grown overnight to late-exponential phase in Todd-Hewitt broth. Cultures were adjusted to
345 OD_{600} = 1 corresponding to ~10^8 viable CFU, diluted 1:100, and distributed in 200 µl aliquots into 8 wells

346 of a sterile U-bottom 96-well plate (Greiner) (initial number of CFU (N_0) ~10^5 per well; lower inocula
347 produced variability in final CFU number, N_t). A replicate plate containing at least two replicate cultures
348 per strain was also prepared in order to monitor the culture growth by optical density with a BMG
349 FLUOstar OPTIMA microplate reader. The 96-well plates were incubated without agitation to early
350 stationary phase. Viable CFU counts were performed for two representative cultures for each strain by
351 Miles and Misra method [81], and estimated as 1-2 x 10^8 CFU per culture. For selection of mutants, whole
352 200 µL cultures were plated on Todd-Hewitt agar containing 0.5 µg/mL rifampicin, dried in a laminar flow
353 hood, and incubated until rifampicin resistant colonies appeared. The assay was repeated three or more
354 times for each strain, and results were pooled into single data sets representing 32 to 120 cultures per
355 strain. Estimation of mutation rate (µ - probability of mutation per cell per generation), was carried out
356 using the Ma-Sandri-Sarkar Maximum Likelihood Estimation (MSS-MLE) method as implemented in
357 FALCOR fluctuation analysis calculator [82] .

358 **DNA extraction, preparation and sequencing**

359 Genomic DNA was extracted with the DNeasy Blood & Tissue kit (Qiagen) using a modified protocol with
360 an additional lysis step. Briefly, 10 mL of late-exponential phase culture was centrifuged and the pellet
361 resuspended in 200 µL lysis buffer comprising 141 µL MilliQ water, 40 µL lysozyme (200 mg/mL), 8 uL of
362 0.5 M EDTA, 5 µL Triton X100, 4 µL RNAase A (100 mg/mL), and 2 µL of 1M Tris-HCl, and incubated at
363 37°C, 700 rpm for 1 h in an Eppendorf thermomixer. Then 200 µL of lysis buffer AL (Qiagen) and 20 µL of
364 proteinase K were added to the sample, mixed by inversion and incubated in the Eppendorf thermomixer
365 at 56°C, 700 rpm for 1 h. Finally, 200 µL of molecular grade absolute ethanol was added, samples
366 vortexed, loaded onto the purification column, left to stand for 2 min, washed with AW1 and AW2 buffers
367 and dried for 3 min. Elution was performed with 80 µL of nuclease-free water pre-heated to 70°C in two
368 steps: first with 50 µL, then with 30 µL. Eluate (70 µL) was transferred into a new tube and stored at
369 -80°C until submission for sequencing. The remaining 10 µL was used for gDNA visualization on agarose
370 gel, quantification by Qubit fluorimetry (Invitrogen), and sequencing 16S rRNA gene using universal 27F
371 and 1492R primers [83], to confirm integrity, identity, and purity of gDNA. When the required
372 concentration for sequencing was not achieved in samples they were dried by vacuum centrifugation
373 (SpeedVac) at room temperature. Sequencing was performed on the Illumina HiSeq2000 platform from
374 Nextera XT pair-end libraries at Australian Genome Research Facility, Melbourne. A reference genome
375 from strain QMA0248 was constructed using both long reads derived from a single Smrt Cell using the
376 PacBio RS II system with P4C2 chemistry and short reads from Illumina HiSeq2000 derived from Nextera
377 XT paired-end libraries as reported elsewhere (NCBI accession no: GCA_002220115.1). All sequence data
378 is deposited at NCBI under Bioproject number PRJNA417543, SRA accession SRP145425. Sample

379 numbers, accession numbers and extended metadata are provided in the supplementary table S1.
380 Assembly statistics are provided in Supplementary table S2

381 **Phylogenetic analysis**

382 Phylogeny was constructed based on core genome SNPs from *de novo* genome assemblies filtered to
383 remove recombination breakpoints. Paired-end reads from Illumina were trimmed with Nesoni clip tool
384 version 0.132 (http://www.vicbioinformatics.com/software.nesoni.shtml), with minimum read length
385 50, and the first 15 bp of each read removed as quality deterioration in this region was observed when
386 assessed with FASTQC version 0.11.5. Assembly was performed using the SPAdes assembler version 3.7.1
387 [84], with minimum read coverage cutoff set to 10. Quality of assemblies was assessed with QUAST 3.2
388 [85]. Contigs were ordered by alignment to QMA248 reference genome (CP022392.1) with Mauve Contig
389 Mover 2.4.0 [86]. Genome annotation was performed using Prokka 1.11 [87]. Rapid alignment of core
390 genomes was carried out using parsnp in the Harvest Tools suite version 1.2 [88], and the resulting
391 alignment provided as an input to Gubbins 1.4.7 [89] for detection and exclusion of variants produced by
392 recombination. Phylogenies were then inferred from post-filtered core genome polymorphic sites by
393 maximum likelihood using RAxML 7.2.8 [90] with the general time reversible nucleotide substitution
394 model GTRGAMMA and bootstrap support from 1000 iterations. Ascertainment bias associated with
395 using only polymorphic sites was accounted for using Felsenstein's correction in RAXML. The resulting
396 phylogenetic tree was visualized using Dendroscope v 3.5.7 [91] with bootstrap node support value cut-
397 off 75. For the phylogram figure, tip labels were hidden for clarity and the edge containing QMA0141 was
398 re-scaled as dotted line representing 100-fold decrease in length (Figure 1). A cladogram based on the
399 inferred phylogeny, showing all tip labels and bootstrap support for each node, was annotated with
400 metadata using Evolview V2 (Figure 2) [92].

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

402 SNP analysis of MMR and GO genes was performed by read-mapping with Geneious version 9.1 [93],
403 using default settings unless otherwise specified. Paired-end reads from each genome were trimmed,
404 merged into a single file with expected distances between the reads set to 250 bp, and mapped to
405 reference genome of strain QMA0248. Mapped reads were used to detect SNPs with minimum coverage
406 set to 10 and frequency to 0.9. A consensus pseudogenome was generated for every strain based on the
407 reference sequence including any detected variants. Multiple alignment of pseudogenomes was carried
408 out with *Geneious* aligning tool. Sequences of MMR genes (*dnaN,* mut*S, muTL, recD2, rnhC*) and OG genes
409 (*yfhQ*, *mutM*, *mutX*) annotated in the QMA0248 reference genome were identified in each strain genome
410 sequence. Promoter regions of repair genes were determined with BPROM [94], protein functional
411 domains by SMART genomic [95], and effect of amino acid substitutions on protein function by PROVEAN

412 Protein with a sensitivity cut-off of 1.3 [96]. To detect large variants, alignment of these regions extracted
413 from *de novo* genome assemblies was carried out, and a 56 bp deletion detected in *yfhQ* promoter of
414 clade B was confirmed by PCR using CAGAAGGAAGAAACAGAC_F/ACCTCTATTGTAGCAAAG_R primers
415 followed by Sanger sequencing.

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

417 A multitude of intracellular and secreted enzymes contribute to virulence in the *Streptococcus* genus [54,
418 55]. Considering the large number of strains, we limited analysis to major, easily observable, and distinct
419 phenotypes strongly associated with virulence in streptococci (capsule, hemolysis, cell chains) and
420 bacteria in general (oxidation resistance, biofilms). All assays were performed at least in triplicate per
421 strain.

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

423 Presence/absence of polysaccharide capsule was estimated by Percoll buoyant density assay. Isotonic
424 stock Percoll (ISP) was prepared by mixing nine parts of Percoll with one part of 1.5 M NaCl. Then, 6 parts
425 of ISP was diluted with 4 parts of 0.15 M NaCl to make final 50 % Percoll solution, which was distributed
426 by 3 mL into flow cytometry tubes. 10 mL of THB cultures grown to late-exponential phase was adjusted
427 at OD_{600} 1 (10^8 CFU/mL), centrifuged at 3220 x *g* for 5min, resuspended in 0.5 mL of 0.15 M NaCl, and
428 layered onto the Percoll solution. Tubes were centrifuged at 4°C in a swinging bucket rotor at 4000 x *g*
429 for 3 h with low acceleration and no brake. In this assay, encapsulated cells form a clear compact band in
430 the Percoll gradient (Figure 3, A1), non-encapsulated cells form a pellet in the bottom of the tube (Figure
431 3, A2) and, occasionally, strains show differential expression of capsular polysaccharide evidenced by
432 band and a pellet (Figure 3, A3).

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

434 Rapid high-throughput detection of impaired haemolytic activity was achieved by blood-agar clearance
435 zone assay. Briefly, 5 mm wells were made in Columbia agar supplemented with 5% defibrinated sheep
436 blood (Oxoid, Australia). Bacterial cultures grown to late-exponential phase (~10^8 CFU) were diluted
437 1:1000 and 50 µL of diluted cultures were pipetted into punctures in the agar (initial inoculum ~$5x10^4$
438 CFU in each puncture). A 3 mm wide clearance zone was generally produced by bacterial lysis of sheep
439 erythrocytes during 24 h incubation (Figure 3, B1). Where haemolysis was absent or fragmentary,
440 impeded haemolytic activity was recorded (Figure 3, B2).

441 **Chain formation microscopy**

442 To assess chain formation, *S. iniae* cultures in THB were grown stationarily in 96-well plates at 28 °C for
443 24 h, mixed and 5 µL wet mounts prepared on a glass microscope slide. Slides were observed by bright
444 field microscopy under 40x objective with an Olympus BX40 microscope and captured using an Olympus
445 DP28 digital camera using CellSens software (Olympus Optical Co, Japan). Specimens were observed for
446 at least 3 min and 2-20 cell chains were generally observed (Figure 3, C1). Where over 20 cells in a chain
447 were repeatedly detected (Figure 3, C2) increased chain formation was recorded, and when more than
448 10 cells in a chain were not detected by similar observation and the culture was mainly composed of
449 detached cells (Figure 3, C3) impeded chain formation was recorded. For the figure, 50 µL of cultures
450 were dried onto slides at RT, fixed with methanol, Gram stained, and images captured under 100x
451 objective.

452 **Oxidation resistance assay**

453 Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) [97] of
454 hydrogen peroxide were measured to assess oxidative stress resistance among *S. iniae* strains. THB
455 cultures grown to late exponential phase were adjusted to OD_{600} = 1 (10^8 CFU/mL), diluted 1000-fold, and
456 distributed by 100 µL (~10^4 CFU) into wells of a U-bottom 96-well plate (Greiner). THB (100 µL) with 2-
457 fold excess concentration of hydrogen peroxide was added to the wells and serially diluted twofold,
458 resulting in a range from 0 to 10 mM final peroxide concentrations. Plates were incubated stationarily
459 for 24h and examined for presence of observable growth to determine the MIC. Viable cell counts of
460 cultures without visible growth were performed to determine MBC. MIC of hydrogen peroxide was
461 estimated as 3 mM for all strains, and MBC as 4 mM for the majority of strains. MBC elevated to 5 mM
462 in dolphin and human isolates from USA was classified deviation from a wildtype phenotype.

463 **Biofilm formation and visualisation assay**

464 *S. iniae* biofilms were grown for 4 days in 8-well Lab-Tek® II Chamber Slide™ Systems. To prepare an initial
465 inoculum of 5 x 10^5 CFU, THB cultures grown to late exponential phase were adjusted to OD_{600} = 1 (10^8
466 CFU/mL), diluted 100-fold, and 0.5 mL of diluted cultures were placed into Chamber slide wells. After 24
467 h incubation, THB was removed and replaced with fresh THB every 10-14 h for 3 days. For visualisation,
468 biofilms were washed in PBS, stained for 15 min with 1 µM fluorescent *Bac*Light Red bacterial stain,
469 washed in PBS, fixed for 30 min with 10% formalin, and washed twice PBS. Z-stacks were collected by
470 ZEISS LSM 710 Inverted Laser Scanning Confocal Microscope at 20x objective and visualised using
471 ZEN2012. 8 -12 µm thick biofilms covering up to 30% of representative 200 x 200 µm surface produced
472 by most strains were classified as typical (Figure 3, D1), denser (Figure 3, D2) or/and thicker (Figure 3, D3)
473 structures were recorded as deviant.

474 **Identifying wildtype virulence-related phenotypes**

475 We consider prevalent phenotypes as a wildtype, namely: presence of polysaccharide capsule (71.25 %
476 of strains; Figure 3, A1) and haemolytic activity (85% of strains; Figure 3, B1), up to 20 cells in a chain
477 (77.5 % of strains; Figure 3, A1), 3 mM MIC and 4 mM MBC of hydrogen peroxide (91.25 % of strains),
478 and 8 -12 µm thick biofilms covering up to 30% of representative image (86.25 % of strains, Fig. 4A). Other
479 phenotypes are regarded as deviant and discussed.

480 **Statistical analysis**

481 Estimation of mutation rates was performed by Ma-Sandri-Sarkar-Maximum Likelihood Estimator (MSS-
482 MLE) method. Fluctuation assay results were adjusted to equal number of mutants in the final culture
483 and allowed comparison of mutation rates by Likelihood Ratio test [98] using rSalvador R package [99].
484 *Compare.LD* function was applied pair-wise to all strains. Pairwise comparisons resulting in *p* values less
485 than 0.05 were considered significantly different. Associations between variation in DNA repair genes,
486 mutation rates, deviations from the wildtype virulence phenotype, and typical places of isolation were
487 tested with a Phylogenetic Generalised Least Squares model that accounts for relatedness among strains
488 under Pagel's λ of 0.6 implemented in R [100].

489

490 **Acknowledgements**

491 For provision of strains into our collection over 15 years we thank: Judy Forbes-Faulkner, Oonoonba
492 Veterinary Laboratory, QLD, Australia; Lynn Shewmaker, CDC, Atlanta, USA; Christian Michel, INRA,
493 France; Suresh Benedict, Berrimah Veterinary Laboratories, NT, Australia; Nicky Buller, WA Fisheries, WA,
494 Australia; Mark White, Tréidlia Biovet Pty (formerly Allied Biotechnology), NSW, Australia; Matt Landos,
495 Future Fisheries Veterinary Services, NSW, Australia; Spencer Russel, Novartis Animal Health, Canada;
496 Ruben Arturo Lopez Crespo, Regal Springs Tilapia, Honduras. ACB and OS extend their thanks to Mr Nazar
497 Rudenko for assistance with writing software pipelines and for use of his unix server.

498 **References**

499

500 1. Ferenci T. Trade-off Mechanisms Shaping the Diversity of Bacteria. Trends in microbiology.
501 2015. doi: 10.1016/j.tim.2015.11.009. PubMed PMID: 26705697.
502 2. Stich M, Manrubia SC, Lazaro E. Variable Mutation Rates as an Adaptive Strategy in
503 Replicator Populations. PloS one. 2010;5(6). doi: ARTN e11186

504 10.1371/journal.pone.0011186. PubMed PMID: WOS:000278886300028.
505 3. Sniegowski P, Raynes Y. Mutation Rates: How Low Can You Go? Current Biology.
506 2013;23(4):R147-R9. doi: Doi 10.1016/J.Cub.2013.01.018. PubMed PMID: WOS:000315178400007.
507 4. Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, et al. Mutation rate
508 dynamics in a bacterial population reflect tension between adaptation and genetic load. Proc Natl
509 Acad Sci U S A. 2013;110(1):222-7. doi: 10.1073/pnas.1219574110. PubMed PMID: 23248287;
510 PubMed Central PMCID: PMCPMC3538217.
511 5. Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Medigue C, et al.
512 Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution Experiment With
513 Escherichia coli. G3. 2011;1(3):183-6. doi: 10.1534/g3.111.000406. PubMed PMID: 22207905;
514 PubMed Central PMCID: PMC3246271.
515 6. Galhardo RS, Hastings PJ, Rosenberg SM. Mutation as a stress response and the regulation of
516 evolvability. Critical reviews in biochemistry and molecular biology. 2007;42(5):399-435. doi:
517 10.1080/10409230701648502. PubMed PMID: 17917874; PubMed Central PMCID: PMC3319127.
518 7. Sundin GW, Weigand MR. The microbiology of mutability. Fems Microbiol Lett.
519 2007;277(1):11-20. doi: Doi 10.1111/J.1574-6968.2007.00901.X. PubMed PMID:
520 WOS:000250619200002.
521 8. Boe L, Danielsen M, Knudsen S, Petersen JB, Maymann J, Jensen PR. The frequency of
522 mutators in populations of Escherichia coli. Mutation research. 2000;448(1):47-55. PubMed PMID:
523 10751622.
524 9. Miller JH. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and
525 repair. Annual review of microbiology. 1996;50:625-43. doi: 10.1146/annurev.micro.50.1.625.
526 PubMed PMID: 8905093.
527 10. Mao EF, Lane L, Lee J, Miller JH. Proliferation of mutators in A cell population. Journal of
528 bacteriology. 1997;179(2):417-22. PubMed PMID: 8990293; PubMed Central PMCID: PMC178711.
529 11. Taddei F, Radman M, MaynardSmith J, Toupance B, Gouyon PH, Godelle B. Role of mutator
530 alleles in adaptive evolution. Nature. 1997;387(6634):700-2. doi: Doi 10.1038/42696. PubMed PMID:
531 WOS:A1997XD86900052.
532 12. Hall LM, Henderson-Begg SK. Hypermutable bacteria isolated from humans--a critical
533 analysis. Microbiology. 2006;152(Pt 9):2505-14. doi: 10.1099/mic.0.29079-0. PubMed PMID:
534 16946246.
535 13. Oliver A. Mutators in cystic fibrosis chronic lung infection: Prevalence, mechanisms, and
536 consequences for antimicrobial therapy. International journal of medical microbiology : IJMM.
537 2010;300(8):563-72. doi: 10.1016/j.ijmm.2010.08.009. PubMed PMID: 20837399.
538 14. Matsushima A, Takakura S, Fujihara N, Saito T, Ito I, Iinuma Y, et al. High prevalence of
539 mutators among Enterobacter cloacae nosocomial isolates and their association with antimicrobial
540 resistance and repetitive detection. Clinical microbiology and infection : the official publication of
541 the European Society of Clinical Microbiology and Infectious Diseases. 2010;16(9):1488-93. doi:
542 10.1111/j.1469-0691.2009.03145.x. PubMed PMID: 20041896.
543 15. Wang S, Wang Y, Shen J, Wu Y, Wu C. Polymorphic mutation frequencies in clinical isolates
544 of Staphylococcus aureus: the role of weak mutators in the development of fluoroquinolone
545 resistance. Fems Microbiol Lett. 2013;341(1):13-7. doi: 10.1111/1574-6968.12085. PubMed PMID:
546 23330696.

547 16. Negri MC, Morosini MI, Baquero MR, del Campo R, Blazquez J, Baquero F. Very low
548 cefotaxime concentrations select for hypermutable Streptococcus pneumoniae populations.
549 Antimicrobial agents and chemotherapy. 2002;46(2):528-30. doi: Doi 10.1128/Aac.46.2.528-
550 530.2002. PubMed PMID: WOS:000173427900042.
551 17. Chopra I, O'Neill AJ, Miller K. The role of mutators in the emergence of antibiotic-resistant
552 bacteria. Drug resistance updates : reviews and commentaries in antimicrobial and anticancer
553 chemotherapy. 2003;6(3):137-45. PubMed PMID: 12860461.
554 18. Blazquez J. Hypermutation as a factor contributing to the acquisition of antimicrobial
555 resistance. Clin Infect Dis. 2003;37(9):1201-9. PubMed PMID: WOS:000185916200010.
556 19. Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL, et al. Genetic Adaptation of
557 Pseudomonas aeruginosa to the Airways of Cystic Fibrosis Patients Is Catalyzed by Hypermutation.
558 Journal of bacteriology. 2008;190(24):7910-7. doi: Doi 10.1128/Jb.01147-08. PubMed PMID:
559 WOS:000261217900007.
560 20. Gutierrez O, Juan C, Perez JL, Oliver A. Lack of association between hypermutation and
561 antibiotic resistance development in Pseudomonas aeruginosa isolates from intensive care unit
562 patients. Antimicrobial agents and chemotherapy. 2004;48(9):3573-5. doi: Doi
563 10.1128/Aac.48.9.3573-3575.2004. PubMed PMID: WOS:000223625800055.
564 21. Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, et al. Costs and benefits of high
565 mutation rates: Adaptive evolution of bacteria in the mouse gut. Science. 2001;291(5513):2606-8.
566 doi: Doi 10.1126/Science.1056421. PubMed PMID: WOS:000167861000043.
567 22. Nilsson AI, Kugelberg E, Berg OG, Andersson DI. Experimental adaptation of Salmonella
568 typhimurium to mice. Genetics. 2004;168(3):1119-30. doi: 10.1534/genetics.104.030304. PubMed
569 PMID: 15579674; PubMed Central PMCID: PMC1448776.
570 23. Healey KR, Zhao Y, Perez WB, Lockhart SR, Sobel JD, Farmakiotis D, et al. Prevalent mutator
571 genotype identified in fungal pathogen Candida glabrata promotes multi-drug resistance. Nature
572 communications. 2016;7:11128. doi: 10.1038/ncomms11128. PubMed PMID: 27020939.
573 24. Labat F, Pradillon O, Garry L, Peuchmaur M, Fantin B, Denamur E. Mutator phenotype
574 confers advantage in Escherichia coli chronic urinary tract infection pathogenesis. FEMS Immunol
575 Med Microbiol. 2005;44(3):317-21. doi: 10.1016/j.femsim.2005.01.003. PubMed PMID: 15907455.
576 25. Mena A, Macia MD, Borrell N, Moya B, de Francisco T, Perez JL, et al. Inactivation of the
577 mismatch repair system in Pseudomonas aeruginosa attenuates virulence but favors persistence of
578 oropharyngeal colonization in cystic fibrosis mice. Journal of bacteriology. 2007;189(9):3665-8. doi:
579 10.1128/JB.00120-07. PubMed PMID: 17307847; PubMed Central PMCID: PMC1855878.
580 26. Oliver A, Mena A. Bacterial hypermutation in cystic fibrosis, not only for antibiotic
581 resistance. Clin Microbiol Infec. 2010;16(7):798-808. doi: 10.1111/j.1469-0691.2010.03250.x.
582 PubMed PMID: WOS:000278645700002.
583 27. Aquino T, Nunes A. Host immunity and pathogen diversity: A computational study.
584 Virulence. 2016;7(2):121-8. doi: 10.1080/21505594.2016.1149284. PubMed PMID: 26836775;
585 PubMed Central PMCID: PMC4994829.
586 28. Cobey S. Pathogen evolution and the immunological niche. Annals of the New York Academy
587 of Sciences. 2014;1320:1-15. doi: 10.1111/nyas.12493. PubMed PMID: 25040161; PubMed Central
588 PMCID: PMC4141700.
589 29. Lukacisinova M, Novak S, Paixao T. Stress-induced mutagenesis: Stress diversity facilitates
590 the persistence of mutator genes. PLoS computational biology. 2017;13(7):e1005609. doi:
591 10.1371/journal.pcbi.1005609. PubMed PMID: 28719607; PubMed Central PMCID: PMC5538753.
592 30. Methot PO, Alizon S. What is a pathogen? Toward a process view of host-parasite
593 interactions. Virulence. 2014;5(8):775-85. doi: 10.4161/21505594.2014.960726. PubMed PMID:
594 25483864.
595 31. Mackinnon MJ, Read AF. Virulence in malaria: an evolutionary viewpoint. Philosophical
596 transactions of the Royal Society of London Series B, Biological sciences. 2004;359(1446):965-86.
597 doi: 10.1098/rstb.2003.1414. PubMed PMID: 15306410; PubMed Central PMCID: PMC1693375.

598 32. Deitsch KW, Moxon ER, Wellems TE. Shared themes of antigenic variation and virulence in
599 bacterial, protozoal, and fungal infections. Microbiology and molecular biology reviews : MMBR.
600 1997;61(3):281-93. PubMed PMID: 9293182; PubMed Central PMCID: PMC232611.
601 33. Shrestha S, Bjornstad ON, King AA. Evolution of acuteness in pathogen metapopulations:
602 conflicts between "classical" and invasion-persistence trade-offs. Theoretical ecology. 2014;7(3):299-
603 311. doi: 10.1007/s12080-014-0219-7. PubMed PMID: 25214895; PubMed Central PMCID:
604 PMC4157682.
605 34. Maurelli AT. Black holes, antivirulence genes, and gene inactivation in the evolution of
606 bacterial pathogens. Fems Microbiol Lett. 2007;267(1):1-8. doi: 10.1111/j.1574-6968.2006.00526.x.
607 35. Merino D, Reglier-Poupet H, Berche P, Charbit A, European Listeria Genome C. A
608 hypermutator phenotype attenuates the virulence of Listeria monocytogenes in a mouse model.
609 Molecular microbiology. 2002;44(3):877-87. PubMed PMID: 11994166.
610 36. Gonzalez K, Faustoferri RC, Quivey RG, Jr. Role of DNA base excision repair in the mutability
611 and virulence of Streptococcus mutans. Molecular microbiology. 2012;85(2):361-77. doi:
612 10.1111/j.1365-2958.2012.08116.x. PubMed PMID: 22651851; PubMed Central PMCID:
613 PMC3391326.
614 37. Picard B, Duriez P, Gouriou S, Matic I, Denamur E, Taddei H. Mutator natural Escherichia coli
615 isolates have an unusual virulence phenotype. Infection and immunity. 2001;69(1):9-14. doi: Doi
616 10.1128/Iai.69.1.9-14.2001. PubMed PMID: WOS:000165943500002.
617 38. Smania AM, Segura I, Pezza RJ, Becerra C, Albesa I, Argarana CE. Emergence of phenotypic
618 variants upon mismatch repair disruption in Pseudomonas aeruginosa. Microbiol-Sgm.
619 2004;150:1327-38. doi: Doi 10.1099/Mic.0.26751-0. PubMed PMID: WOS:000221538000030.
620 39. Canfield GS, Schwingel JM, Foley MH, Vore KL, Boonanantanasarn K, Gill AL, et al. Evolution
621 in fast forward: a potential role for mutators in accelerating Staphylococcus aureus pathoadaptation.
622 Journal of bacteriology. 2013;195(3):615-28. doi: 10.1128/JB.00733-12. PubMed PMID: 23204459;
623 PubMed Central PMCID: PMC3554003.
624 40. Fukui K. DNA mismatch repair in eukaryotes and bacteria. Journal of nucleic acids.
625 2010;2010. doi: 10.4061/2010/260512. PubMed PMID: 20725617; PubMed Central PMCID:
626 PMC2915661.
627 41. Dai L, Muraoka WT, Wu ZW, Sahin O, Zhang QJ. A single nucleotide change in mutY increases
628 the emergence of antibiotic-resistant Campylobacter jejuni mutants. J Antimicrob Chemoth.
629 2015;70(10):2739-48. doi: 10.1093/jac/dkv190. PubMed PMID: WOS:000363249800009.
630 42. Rajanna C, Ouellette G, Rashid M, Zemla A, Karavis M, Zhou C, et al. A strain of Yersinia
631 pestis with a mutator phenotype from the Republic of Georgia. Fems Microbiol Lett.
632 2013;343(2):113-20. doi: 10.1111/1574-6968.12137. PubMed PMID: 23521061.
633 43. Ambur OH, Davidsen T, Frye SA, Balasingham SV, Lagesen K, Rognes T, et al. Genome
634 dynamics in major bacterial pathogens. FEMS microbiology reviews. 2009;33(3):453-70. doi:
635 10.1111/j.1574-6976.2009.00173.x. PubMed PMID: WOS:000264956900002.
636 44. Chen F, Liu WQ, Eisenstark A, Johnston RN, Liu GR, Liu SL. Multiple genetic switches
637 spontaneously modulating bacterial mutability. BMC evolutionary biology. 2010;10:277. doi:
638 10.1186/1471-2148-10-277. PubMed PMID: 20836863; PubMed Central PMCID: PMC2955026.
639 45. Li GM. Mechanisms and functions of DNA mismatch repair. Cell Res. 2008;18(1):85-98. doi:
640 10.1038/cr.2007.115. PubMed PMID: WOS:000252920000009.
641 46. Tham KC, Hermans N, Winterwerp HHK, Cox MM, Wyman C, Kanaar R, et al. Mismatch
642 Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA
643 Structures. Mol Cell. 2013;51(3):326-37. doi: Doi 10.1016/J.Molcel.2013.07.008. PubMed PMID:
644 WOS:000323086400008.
645 47. Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA. Mismatch repair in Gram-positive
646 bacteria. Research in microbiology. 2016;167(1):4-12. doi: 10.1016/j.resmic.2015.08.006. PubMed
647 PMID: 26343983.

648 48. David SS, O'Shea VL, Kundu S. Base-excision repair of oxidative DNA damage. Nature.
649 2007;447(7147):941-50. doi: 10.1038/nature05978. PubMed PMID: WOS:000247373100036.
650 49. Lu AL, Li X, Gu Y, Wright PM, Chang DY. Repair of oxidative DNA damage: mechanisms and
651 functions. Cell biochemistry and biophysics. 2001;35(2):141-70. doi: 10.1385/CBB:35:2:141. PubMed
652 PMID: 11892789.
653 50. Agnew W, Barnes AC. Streptococcus iniae: an aquatic pathogen of global veterinary
654 significance and a challenging candidate for reliable vaccination. Veterinary microbiology.
655 2007;122(1-2):1-15. doi: 10.1016/j.vetmic.2007.03.002. PubMed PMID: 17418985.
656 51. Millard CM, Baiano JC, Chan C, Yuen B, Aviles F, Landos M, et al. Evolution of the capsular
657 operon of Streptococcus iniae in response to vaccination. Appl Environ Microbiol. 2012;78(23):8219-
658 26. Epub 2012/09/25. doi: AEM.02216-12 [pii]

659 10.1128/AEM.02216-12. PubMed PMID: 23001668; PubMed Central PMCID: PMC3497394.
660 52. Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, et al. Population
661 genomics of post-vaccine changes in pneumococcal epidemiology. Nat Genet. 2013;45(6):656-63.
662 Epub 2013/05/07. doi: 10.1038/ng.2625. PubMed PMID: 23644493; PubMed Central PMCID:
663 PMCPMC3725542.
664 53. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, et al. Rapid
665 Pneumococcal Evolution in Response to Clinical Interventions. Science. 2011;331(6016):430-4. doi:
666 10.1126/science.1198545. PubMed PMID: WOS:000286635900044.
667 54. Rajagopal L. Understanding the regulation of Group B Streptococcal virulence factors. Future
668 microbiology. 2009;4(2):201-21. doi: 10.2217/17460913.4.2.201. PubMed PMID: 19257847; PubMed
669 Central PMCID: PMC2691590.
670 55. Kadioglu A, Weiser JN, Paton JC, Andrew PW. The role of Streptococcus pneumoniae
671 virulence factors in host respiratory colonization and disease. Nat Rev Microbiol. 2008;6(4):288-301.
672 doi: 10.1038/nrmicro1871. PubMed PMID: 18340341.
673 56. Locke JB, Colvin KM, Datta AK, Patel SK, Naidu NN, Neely MN, et al. Streptococcus iniae
674 capsule impairs phagocytic clearance and contributes to virulence in fish. J Bacteriol.
675 2007;189(4):1279-87. doi: 10.1128/JB.01175-06. PubMed PMID: 17098893; PubMed Central PMCID:
676 PMC1797360.
677 57. Locke JB, Colvin KM, Varki N, Vicknair MR, Nizet V, Buchanan JT. Streptococcus iniae beta-
678 hemolysin streptolysin S is a virulence factor in fish infection. Diseases of aquatic organisms.
679 2007;76(1):17-26. doi: Doi 10.3354/Dao076017. PubMed PMID: WOS:000247715900003.
680 58. Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, et al. Pneumococcal
681 Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev. 2015;28(3):871-99. Epub
682 2015/06/19. doi: 10.1128/cmr.00024-15. PubMed PMID: 26085553; PubMed Central PMCID:
683 PMCPmc4475641.
684 59. Hammerschmidt S, Wolff S, Hocke A, Rosseau S, Muller E, Rohde M. Illustration of
685 pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infection
686 and immunity. 2005;73(8):4653-67. doi: 10.1128/IAI.73.8.4653-4667.2005. PubMed PMID:
687 16040978; PubMed Central PMCID: PMC1201225.
688 60. Rodriguez JL, Dalia AB, Weiser JN. Increased chain length promotes pneumococcal
689 adherence and colonization. Infection and immunity. 2012;80(10):3454-9. doi: 10.1128/IAI.00587-
690 12. PubMed PMID: 22825449; PubMed Central PMCID: PMC3457561.
691 61. Hegde A, Bhat GK, Mallya S. Effect of exposure to hydrogen peroxide on the virulence of
692 Escherichia coli. Indian journal of medical microbiology. 2008;26(1):25-8. PubMed PMID: 18227593.
693 62. Andisi VF, Hinojosa CA, de Jong A, Kuipers OP, Orihuela CJ, Bijlsma JJ. Pneumococcal gene
694 complex involved in resistance to extracellular oxidative stress. Infection and immunity.
695 2012;80(3):1037-49. doi: 10.1128/IAI.05563-11. PubMed PMID: 22215735; PubMed Central PMCID:
696 PMC3294666.
697 63. Bjarnsholt T. The role of bacterial biofilms in chronic infections. APMIS Supplementum.
698 2013;(136):1-51. doi: 10.1111/apm.12099. PubMed PMID: 23635385.

64. de Visser JA. The fate of microbial mutators. Microbiology. 2002;148(Pt 5):1247-52. PubMed PMID: 11988499.
65. Turrientes MC, Baquero F, Levin BR, Martinez JL, Ripoll A, Gonzalez-Alba JM, et al. Normal mutation rate variants arise in a Mutator (Mut S) Escherichia coli population. PloS one. 2013;8(9):e72963. doi: 10.1371/journal.pone.0072963. PubMed PMID: 24069167; PubMed Central PMCID: PMC3771984.
66. Shabalina SA, Spiridonov NA, Kashina A. Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity. Nucleic Acids Res. 2013;41(4):2073-94. doi: 10.1093/nar/gks1205. PubMed PMID: 23293005; PubMed Central PMCID: PMC3575835.
67. Gould CV, Sniegowski PD, Shchepetov M, Metlay JP, Weiser JN. Identifying mutator phenotypes among fluoroquinolone-resistant strains of Streptococcus pneumoniae using fluctuation analysis. Antimicrob Agents Chemother. 2007;51(9):3225-9. Epub 2007/08/01. doi: AAC.00336-07 [pii]

10.1128/AAC.00336-07 [doi]. PubMed PMID: 17664329.
68. Yang H, Miller JH. Deletion of dnaN1 generates a mutator phenotype in Bacillus anthracis. DNA Repair (Amst). 2008;7(3):507-14. Epub 2008/02/05. doi: 10.1016/j.dnarep.2007.10.003. PubMed PMID: 18242150; PubMed Central PMCID: PMCPMC2367095.
69. Dupes NM, Walsh BW, Klocko AD, Lenhart JS, Peterson HL, Gessert DA, et al. Mutations in the Bacillus subtilis beta clamp that separate its roles in DNA replication from mismatch repair. J Bacteriol. 2010;192(13):3452-63. Epub 2010/05/11. doi: 10.1128/JB.01435-09. PubMed PMID: 20453097; PubMed Central PMCID: PMCPMC2897676.
70. Lenhart JS, Sharma A, Hingorani MM, Simmons LA. DnaN clamp zones provide a platform for spatiotemporal coupling of mismatch detection to DNA replication. Mol Microbiol. 2013;87(3):553-68. Epub 2012/12/12. doi: 10.1111/mmi.12115. PubMed PMID: 23228104; PubMed Central PMCID: PMCPMC5938748.
71. Korkmaz G, Holm M, Wiens T, Sanyal S. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. The Journal of biological chemistry. 2014;289(44):30334-42. doi: 10.1074/jbc.M114.606632. PubMed PMID: 25217634; PubMed Central PMCID: PMC4215218.
72. Facklam R, Elliott J, Shewmaker L, Reingold A. Identification and characterization of sporadic isolates of Streptococcus iniae isolated from humans. Journal of Clinical Microbiology. 2005;43(2):933-7. PubMed PMID: ISI:000227045600071.
73. Weinstein MR, Litt M, Kertesz DA, Wyper P, Rose D, Coulter M, et al. Invasive infections due to a fish pathogen, Streptococcus iniae. N Engl J Med. 1997;337(9):589-94. PubMed PMID: ISI:A1997XT39400002.
74. Weinberger DM, Malley R, Lipsitch M. Serotype replacement in disease after pneumococcal vaccination. Lancet. 2011;378(9807):1962-73. doi: 10.1016/S0140-6736(10)62225-8. PubMed PMID: 21492929; PubMed Central PMCID: PMC3256741.
75. Losordo T, DeLong D, Guerdat T. Advances in technology and practice for land-based aquaculture systems: tank-based recirculating systems for finfish production. Basics. 2009:945-83. doi: Doi 10.1533/9781845696474.6.945. PubMed PMID: WOS:000276209900032.
76. Pier GB, Madin SH. Streptococcus iniae sp nov, a beta-hemolytic Streptococcus isolated from an amazon freshwater dolphin, Inia geoffrensis. . International Journal of Systematic Bacteriology. 1976;26(4):545-53. PubMed PMID: ISI:A1976CT37500020.
77. Pier GB, Madin SH, Alnakeeb S. Isolation and characterisation of a 2nd isolate of Streptococcus iniae. International Journal of Systematic Bacteriology. 1978;28(2):311-4. PubMed PMID: ISI:A1978EW90500021.
78. Baumler A, Fang FC. Host specificity of bacterial pathogens. Cold Spring Harbor perspectives in medicine. 2013;3(12):a010041. Epub 2013/12/04. doi: 10.1101/cshperspect.a010041. PubMed PMID: 24296346; PubMed Central PMCID: PMCPmc3839602.

749 79. Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, et al. Selective
750 and genetic constraints on pneumococcal serotype switching. PLoS Genet. 2015;11(3):e1005095.
751 Epub 2015/04/01. doi: 10.1371/journal.pgen.1005095. PubMed PMID: 25826208; PubMed Central
752 PMCID: PMCPmc4380333.
753 80. Rosche WA, Foster PL. Determining mutation rates in bacterial populations. Methods.
754 2000;20(1):4-17. doi: 10.1006/meth.1999.0901. PubMed PMID: 10610800; PubMed Central PMCID:
755 PMC2932672.
756 81. Hedges AJ. Estimating the precision of serial dilutions and viable bacterial counts.
757 International journal of food microbiology. 2002;76(3):207-14. PubMed PMID: 12051477.
758 82. Hall BM, Ma CX, Liang P, Singh KK. Fluctuation analysis CalculatOR: a web tool for the
759 determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics.
760 2009;25(12):1564-5. doi: 10.1093/bioinformatics/btp253. PubMed PMID: 19369502; PubMed
761 Central PMCID: PMC2687991.
762 83. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and in situ detection of
763 individual microbial cells without cultivation. Microbiology Reviews. 1995;59(1):143-69.
764 84. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new
765 genome assembly algorithm and its applications to single-cell sequencing. Journal of computational
766 biology : a journal of computational molecular cell biology. 2012;19(5):455-77. doi:
767 10.1089/cmb.2012.0021. PubMed PMID: 22506599; PubMed Central PMCID: PMC3342519.
768 85. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome
769 assemblies. Bioinformatics. 2013;29(8):1072-5. doi: 10.1093/bioinformatics/btt086. PubMed PMID:
770 23422339; PubMed Central PMCID: PMC3624806.
771 86. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft
772 genomes using the Mauve aligner. Bioinformatics. 2009;25(16):2071-3. doi:
773 10.1093/bioinformatics/btp356. PubMed PMID: 19515959; PubMed Central PMCID:
774 PMCPMC2723005.
775 87. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics.
776 2014;30(14):2068-9. doi: 10.1093/bioinformatics/btu153. PubMed PMID: 24642063.
777 88. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome
778 alignment and visualization of thousands of intraspecific microbial genomes. Genome Biology.
779 2014;15(11):524. doi: 10.1186/PREACCEPT-2573980311437212. PubMed PMID: 25410596.
780 89. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, et al. Rapid phylogenetic
781 analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic
782 Acids Res. 2015;43(3):e15. doi: 10.1093/nar/gku1196. PubMed PMID: 25414349; PubMed Central
783 PMCID: PMCPMC4330336.
784 90. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
785 phylogenies. Bioinformatics. 2014;30(9):1312-3. doi: 10.1093/bioinformatics/btu033. PubMed PMID:
786 24451623; PubMed Central PMCID: PMCPMC3998144.
787 91. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. Dendroscope: An interactive
788 viewer for large phylogenetic trees. BMC bioinformatics. 2007;8:460. doi: 10.1186/1471-2105-8-460.
789 PubMed PMID: 18034891; PubMed Central PMCID: PMC2216043.
790 92. He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. Evolview v2: an online visualization and
791 management tool for customized and annotated phylogenetic trees. Nucleic Acids Res.
792 2016;44(W1):W236-41. doi: 10.1093/nar/gkw370. PubMed PMID: 27131786; PubMed Central
793 PMCID: PMCPMC4987921.
794 93. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An
795 integrated and extendable desktop software platform for the organization and analysis of sequence
796 data. Bioinformatics. 2012;28(12):1647-9. doi: 10.1093/bioinformatics/bts199. PubMed PMID:
797 WOS:000305419800052.

94. V. Solovyev AS. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Nova Science Publishers; 2011. p. 61-78.
95. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46(D1):D493-D6. doi: 10.1093/nar/gkx922. PubMed PMID: 29040681; PubMed Central PMCID: PMC5753352.
96. Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinformatics. 2015;31(16):2745-7. doi: 10.1093/bioinformatics/btv195. PubMed PMID: 25851949; PubMed Central PMCID: PMCPMC4528627.
97. Andrews JM. Determination of minimum inhibitory concentrations. The Journal of antimicrobial chemotherapy. 2001;48 Suppl 1:5-16. PubMed PMID: 11420333.
98. Zheng Q. Methods for comparing mutation rates using fluctuation assay data. Mutation research. 2015;777:20-2. doi: 10.1016/j.mrfmmm.2015.04.002. PubMed PMID: 25912079.
99. Zheng Q. rSalvador: An R Package for the Fluctuation Experiment. G3. 2017;7(12):3849-56. doi: 10.1534/g3.117.300120. PubMed PMID: 29084818; PubMed Central PMCID: PMC5714482.
100. Symonds MRE, Blomberg SP. A Primer on Phylogenetic Generalised Least Squares. In: Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in Evolutionary Biology: Concepts and Practice. Berlin, Heidelberg: Springer Berlin Heidelberg; 2014. p. 105-30.

819 **Supporting Information Legends**

820 Supplementary Table 1. Extended metadata for 80 *Streptococcus iniae* isolates sequenced in the present
821 study (Microsoft Excel spreadsheet

822 Supplementary Table 2. Assembly statistics and quality data for 80 *de novo* genome assemblies prepared
823 in this study (Microsoft Excel spreadsheet).

824 **Figures and tables**

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

832 Figure 2: Cladogram of *Streptococcus iniae* strains derived from the same dataset as Figure 1. Clade, host
833 species, variants in MMR and GO DNA repair genes, mutation rate phenotype, and phenotypic variants
834 associated with virulence were annotated with Evolview. Node support is indicated as percentage of 1000
835 bootstrap replicates.

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

840

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

Strain	Clade	Host	Year	Location	Capsule	Hemolysis	Chains	Oxidation resistance	Biofilms	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	**denser**	0.5539
QMA0134	E1	***Homo sapiens***	2001	USA	**differential**	wildtype	wildtype	wildtype	**denser**	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	G	***Inia geoffrensis***	1976	USA	**absent**	wildtype	wildtype	**increased**	**denser, thicker**	0.0901
QMA0141		***Inia geoffrensis***	1978	USA	**absent**	wildtype	wildtype	**increased**	**denser, thicker**	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	wildtype	0.5135
QMA0187	G	*Channa striata*	1983	Thailand	**absent**	**weak**	wildtype	wildtype	wildtype	0.4635
QMA0188	D	*Oncorhynchus mykiss*	1998	Israel	wildtype	**weak**	wildtype	wildtype	wildtype	0.5042

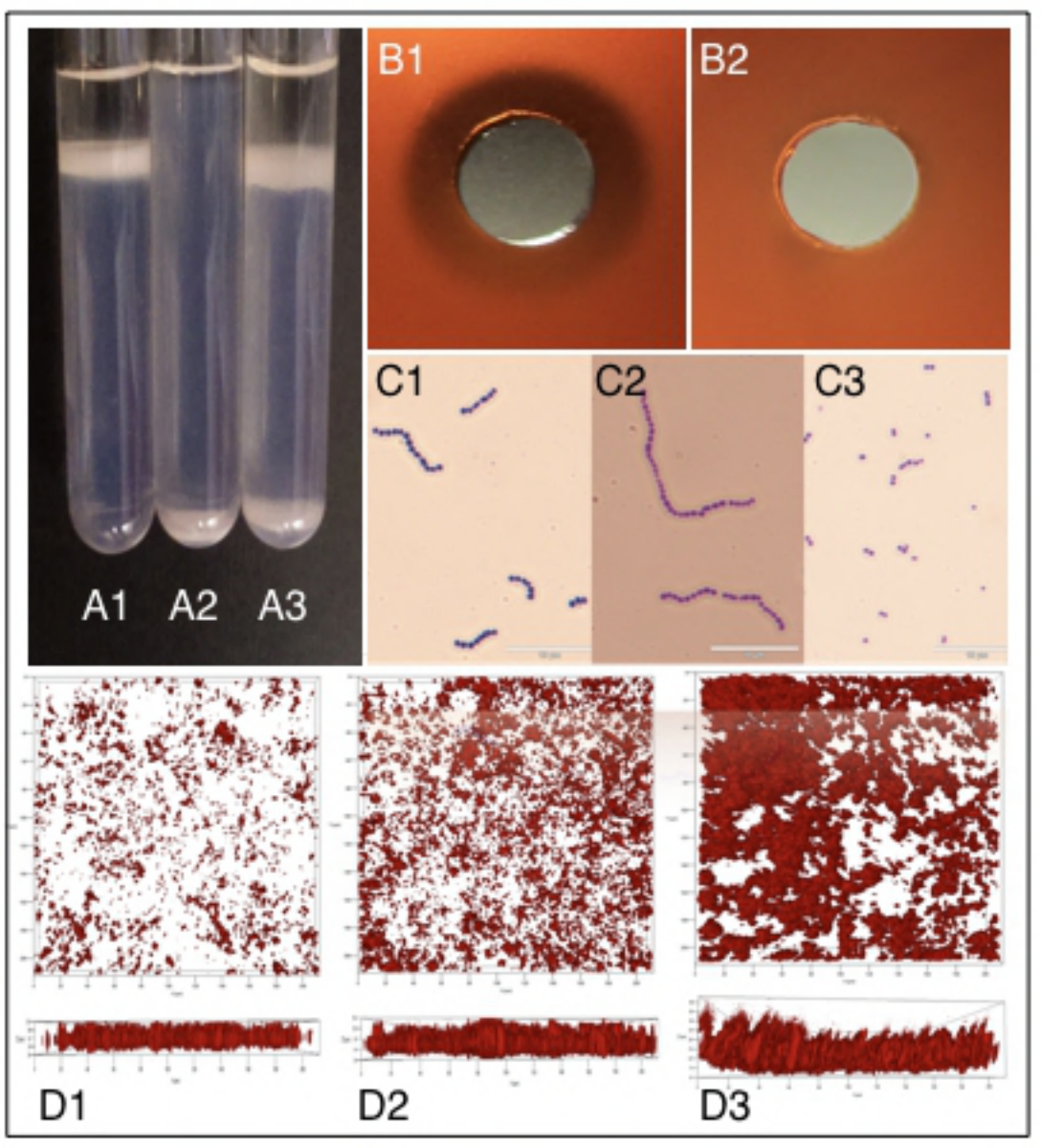
QMA0189	D	*Oncorhynchus mykiss*	1996	Reunion	wildtype	**weak**	wildtype	wildtype	wildtype	0.4907
QMA0190	F	*Channa striata*	1988	Thailand	**absent**	wildtype	wildtype	wildtype	wildtype	0.6413
QMA0191	C1	*Lates calcarifer*	2005	NT	wildtype	wildtype	wildtype	wildtype	wildtype	0.1883
QMA0207	C1	*Lates calcarifer*	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	**denser**	0.362
QMA0234	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**denser**	0.354
QMA0235	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**denser**	0.3496
QMA0236	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	**weak**	**long**	wildtype	**denser**	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	**denser**	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	**denser**	0.3555
QMA0254	F	*Lates calcarifer,* ***bone***	2009	NSW RAS	**absent**	wildtype	**long**	wildtype	**denser**	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

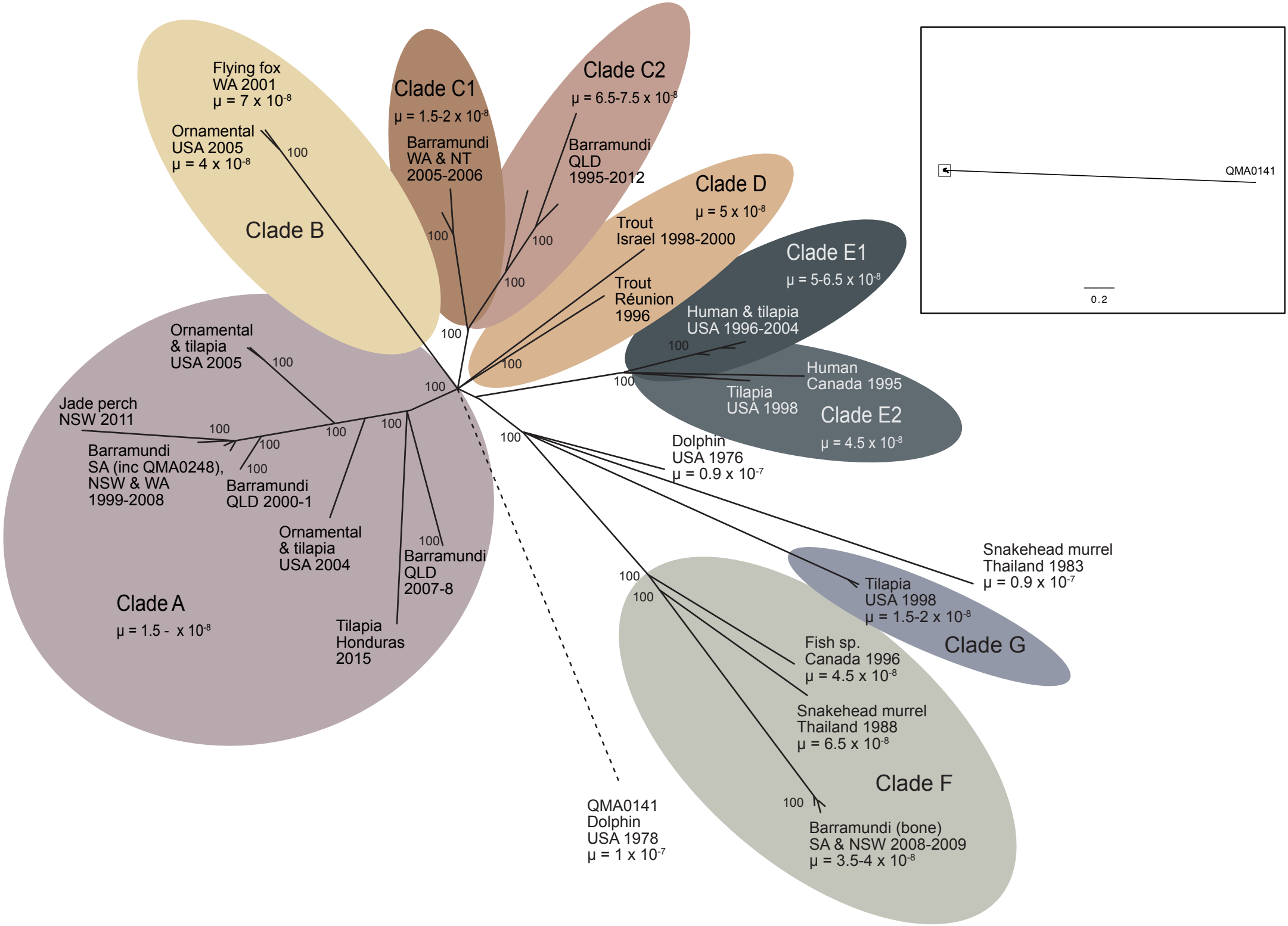
QMA0468	A	*Oreochromis sp.*	2005	USA	wildtype	wildtype	wildtype	wildtype	wildtype	0.1689
QMA0490	A	*Oreochromis sp.*	2015	Honduras	wildtype	wildtype	wildtype	wildtype	wildtype	0.1812
QMA0491	A	*Oreochromis sp.*	2015	Honduras	wildtype	wildtype	wildtype	wildtype	wildtype	0.2
QMA0492	A	*Oreochromis sp.*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

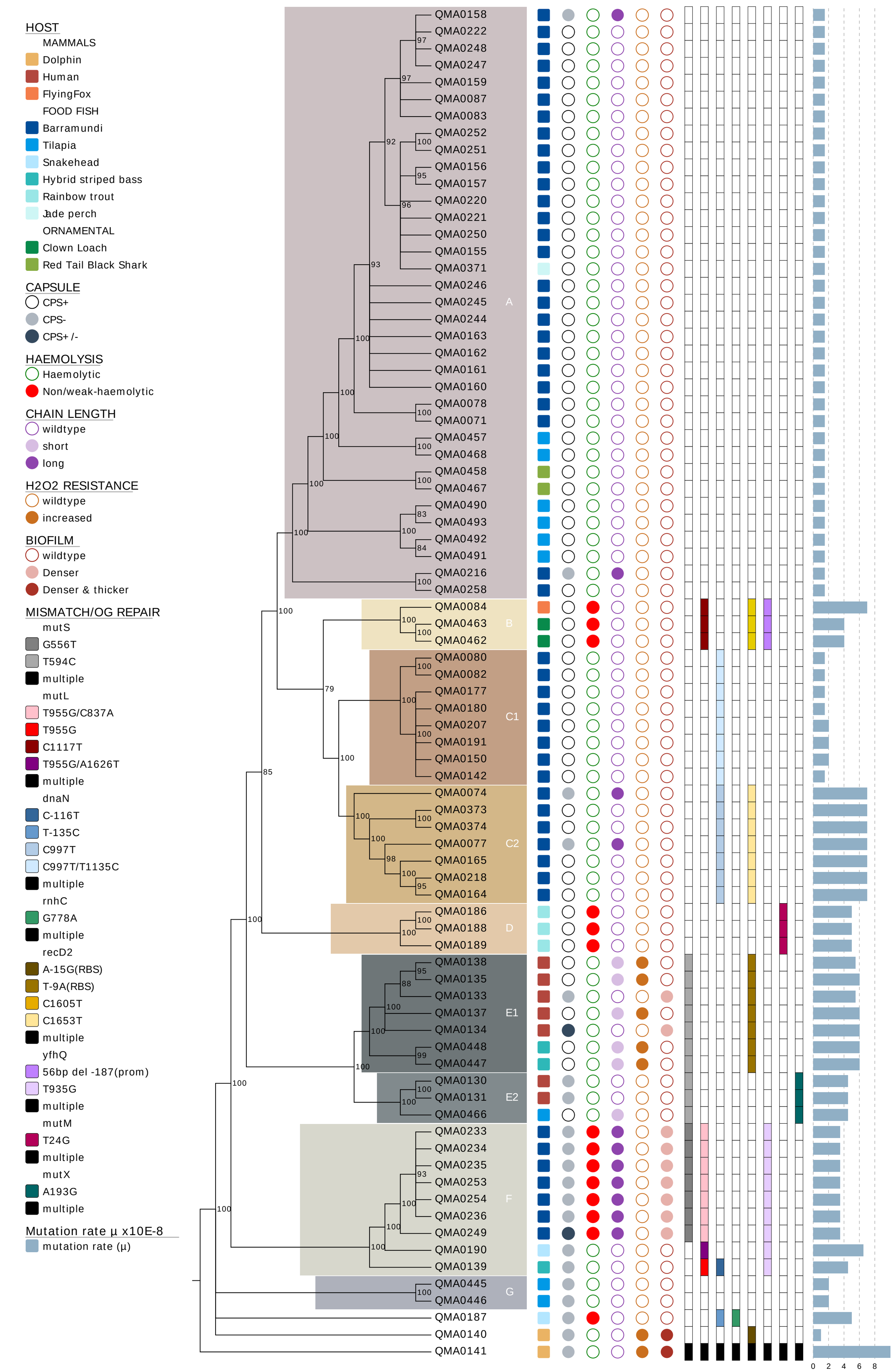
849 Table 2: Variants in MMR and GO genes found among *S. iniae* isolates. D letter next to PROVEAN
850 output value indicates predicted deleterious effect of amino acid change on protein function.

Gene	Clade	Isolates	Mutation	Codon/Amino acid	Locus	PROVEAN score
mutS	G	bone strains	G556T	Q186K	Pfam MutS_II	-0.635
	E1,2	all strains	T594C	198	Pfam MutS_II	n/a
mutL	G	bone strains	C837A	279	DNA mis_repair	n/a
	F	all strains	T955G	S319R	DNA mis_repair	-2.050, D
	B	all strains	C1117T	V393I	-	-0.271
	F	QMA0190	A1626T	542	MutL_C	n/a
dnaN	F	QMA0139	C->T	n/a	promoter, *dnaA* BS 116 bp upstream	n/a
	G	QMA0187	T->C	n/a	promoter, *fnr* BS 135 bp upstream	n/a
	C2	all strains	C977T	T326I	Pol3Bc	-2,216, D
	C1	all strains	T1135C	TAA->Q	Stop codon	n/a
rnhC	G	QMA0187	G778A	D260N	Pfam: RNase_HII	-4.593, D
recD2	G	QMA0140	A -> G	n/a	RBS 15 b. p. upstream	n/a
	E1	all strains	T -> A	n/a	RBS 9 b. p. upstream	n/a
	B	all strains	C1605T	535	-	n/a
	C2	all strains	C1653T	551	-	n/a
yfhQ	B	all strains	56 b.p. deletion	n/a	promoter 187 b. p. upstream	n/a
	F	all strains	T935G	I312T	Pfam: NUDIX_4	-2.544, D
mutM	D	all strains	T24G	E8D	Fapy_DNAglyco	-2.990, D
mutX	E2	all strains	A193G	M65I	Pfam: NUDIX	-0.539

851


A1
A2
A3
B1
B2
C1
C2
C3
D1
D2
D3


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
QMA0141
Dolphin
USA 1978
μ = 1 x 10-7
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
100
QMA0141
0.2


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
yfhQ
56bp del -187(prom)
T935G
multiple
mutM
T24G
multiple
mutX
A193G
multiple
Mutation rate µ x10E-8
mutation rate (µ)
A
B
C1
C2
D
E1
E2
F
G
0 2 4 6 8