1 **Article**

2 **Long title:**

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

5 **Short title:**

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

7

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

9 [1]The University of Queensland, School of Biological Sciences and [2]Centre for Marine Science, St Lucia
10 Campus, Brisbane, Queensland 4072, Australia

11

12

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

15

16 ***Correspondence:**

17 **a.barnes@uq.edu.au**

18

**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 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 (*mutY*, *mutM*, *mutT*) genes. We find that *S. iniae* mutation rate phenotype and genotype are strongly associated with phylogenetic diversification and variation in major streptococcal virulence determinants (capsular polysaccharide, hemolysin, cell chain length, resistance to oxidation, and biofilm formation). Furthermore, profound changes in virulence determinants observed in mammalian isolates (atypical host) and vaccine-escape isolates found in bone (atypical tissue) of vaccinated barramundi are linked to multiple MMR and OG variants and unique mutation rates. This implies that adaptation to new host taxa, to new host tissue, and to immunity of a vaccinated host is 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.

## Introduction

Optimal mutation rates in microbial populations are determined by a trade-off between maintenance of genetic integrity and evolvability (Stich, et al. 2010; Ferenci 2015). High fidelity of DNA repair is advantageous in adapted populations since most mutations lead to deviation from the evolved phenotype (Stich, et al. 2010; Sniegowski and Raynes 2013). Conversely, lower fidelity of DNA repair can be advantageous under stressful conditions since increased mutation supply promotes diversity and may thereby accelerate adaptation (Stich, et al. 2010; Wielgoss, et al. 2013). Bacteria can efficiently fluctuate between high and low mutation regimes: While extremely low mutation rates are maintained in favourable conditions (Wielgoss, et al. 2011), elevation of mutation rates known as stress-induced mutagenesis (SIM) occurs in adapting populations (Galhardo, et al. 2007; Sundin and Weigand 2007). SIM is associated with temporary hypermutability during SOS response and increased frequency of heritable hypermutable strains (mutators) (Galhardo, et al. 2007; Sundin and Weigand 2007). Mutators periodically arise within a population via mutations in DNA repair genes (Miller 1996; Boe, et al. 2000) 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 (Mao, et al. 1997; Taddei, et al. 1997). When adaptation is gained, low mutation rate can be restored via anti-mutator compensatory mutations (Wielgoss, et al. 2013).

Mutator strains are abundant among bacterial clinical isolates, especially in chronic cases (Hall and Henderson-Begg 2006; Oliver 2010). Multiple studies have attributed this to antibiotic selection (Negri, et al. 2002; Matsushima, et al. 2010; Wang, et al. 2013), since mutators are a well-established risk factor for the development of antimicrobial resistance (Blazquez 2003; Chopra, et al. 2003). However, the association between frequency of mutators in diagnostic samples and antibiotic treatments is weak (Gutierrez, et al. 2004; Mena, et al. 2008), and isogenic knockout mutator strains have increased ability to colonize an animal host in challenge models devoid of antibiotic exposure (Giraud, et al. 2001; Nilsson, et al. 2004; Healey, et al. 2016). Thus, the prevalence of mutators in clinical isolates most likely results from the selective process during adaptation to the host (Labat, et al. 2005; Mena, et al. 2007; Mena, et al. 2008; Oliver and Mena 2010), where immunity acts as the ultimate selective pressure shaping the diversity of pathogenic strains (Aquino and Nunes 2016). The 'Immunological niche' concept states that, due to the diversity and complexity of immune responses, every individual host represents a unique habitat that requires adaptation (Cobey 2014). 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 (Lukacisinova, et al. 2017). The fitness of a pathogenic

strain in the host hinges on the concept of virulence (the degree of damage to the host), which is a dynamic, complex trait determined by a multitude of virulence factors and host responses (Methot and Alizon 2014). Optimal virulence is a trade-off between acuteness and persistence with higher virulence facilitating extraction of resources and transmission, but too much damage can kill the host and stop transmission altogether (Mackinnon and Read 2004). Additionally, virulence factors are often antigens so reduced virulence may be a means of immune evasion (Deitsch, et al. 1997). Consequently, both loss and gain of virulence may be advantageous, and indicate adaptive shifts towards acute or chronic pathogenesis, respectively (Maurelli 2007; Shrestha, et al. 2014). Exacerbation and attenuation of virulence factors are repeatedly observed in both knockout and naturally occurring mutators (Picard, et al. 2001; Merino, et al. 2002; Smania, et al. 2004; Mena, et al. 2007; Gonzalez, et al. 2012; Canfield, et al. 2013). Thus, it appears that mutator strains/alleles may help pathogenic bacteria to evolve towards the optimal level of virulence.

Most bacterial mutators isolated from human clinical samples are deficient in genes from mismatch (MMR) and oxidised guanine (OG) DNA repair systems (Hall and Henderson-Begg 2006). MMR and OG genes are conserved across all domains of life (Fukui 2010), and single nucleotide polymorphisms in their sequence may produce drastic alterations in mutation rate (Rajanna, et al. 2013; Dai, et al. 2015). Nonetheless, these genes appear to be variable in major bacterial pathogens, confirming a potential role of mutation rate dynamics in the evolution of virulence (Ambur, et al. 2009; Chen, et al. 2010). MMR is coupled to replication and removes mispaired nucleotides as well as short insertion/deletion loops, and prevents recombination between divergent sequences (Li 2008; Tham, et al. 2013).The MMR pathway has been extensively characterised in *Escherichia coli,* where the MutS dimer protein interacts with the replication processivity clamp DnaN (b-clamp), binds to the mismatch, recruits MutL, and the MutS:MutL complex then activates the MutH endonuclease (Fukui 2010). MutH cuts the unmethylated nascent strand at hemimethylated d(GATC) sites. MutL loads the UvrD helicase that unwinds the nicked DNA containing the error, which is followed by excision by one of the single-strand nucleases (Fukui 2010). This MMR organization, however, appears to be limited to *E. coli* and closely related Gammaproteobacteria (Fukui 2010; Lenhart, et al. 2016). In most bacteria and eukaryotes, MutH is absent and MutL has a conserved endonuclease site, dam methylase (acting at GATC sites) is absent and the strand discrimination signal is unknown (Fukui 2010; Lenhart, et al. 2016). Studies in *Bacillus subtilis,* a model species for gram-positive bacteria, suggest that nicks created by ribonuclease *rnhC* during removal of misincorporated ribonucleosides direct MutL endonuclease towards the newly synthethised strand, and that recD2 acts as a MMR helicase instead of uvrD (Lenhart, et al. 2016). The OG system repairs oxidative DNA damage throughout the cell cycle (Lu, et al. 2001; David, et al. 2007). Among all DNA bases, guanine is most sensitive to reactive oxygen species and its oxidised form, 8-oxo-

dG or OG, is highly mutagenic as it preferentially pairs with adenine which, if not corrected, leads to G:C to A:T transversion on a second round of replication (Lu, et al. 2001; David, et al. 2007). OG system organization is conserved in most bacteria and eukaryotes and generally MutT homologs remove 8-oxo-dG from the nucleotide pool, MutM homologs excise 8-oxo-dG incorporated into DNA, and MutY homologs remove adenine from OG:A mispairings (Lu, et al. 2001; David, et al. 2007).

The present study investigates the role of mutators and mutation rate dynamics in the epidemiology and evolution of pathogenic bacteria using the broad host-range pathogen *Streptococcus iniae*. *S. iniae* is a highly adaptable species, evidenced by its global distribution, ability to cross divergent host taxa boundaries and repeated re-infection of previously immunised hosts during disease outbreaks (Agnew and Barnes 2007; Millard, et al. 2012). Moreover, Streptococcal pathogens in general are difficult candidates for immunisation in humans and animals due to high strain diversity evidently driven by immunisation, yet there is little information on the role of mutators in this evolutionary process (Croucher, et al. 2011; Croucher, et al. 2013). We determine variation in mutation rate and identify mutations in MMR (*mutS*, *mutL*, *dnaN*, *recD2*, *rnhC*) and OG (*mutY* , *mutM*, *mutT*) genes among 80 diverse strains of *S. iniae* and discover that they correlate with phylogenetic diversification but are highly conserved within phylogenetic clades. To identify whether this adaptive process might be driven by the host immune response we identify variation in major phenotypic traits that determine virulence in streptococci (capsular polysaccharide, hemolysin, length of cell chains) and bacteria in general (resistance to reactive oxygen species (ROS), and biofilm formation). *S. iniae* polysaccharide capsule and the beta-hemolysin streptolysin S both increase survival under host immune response via reduction of phagocytic killing and other mechanisms (Locke, Colvin, Datta, et al. 2007; Locke, Colvin, Varki, et al. 2007; Kadioglu, et al. 2008; Rajagopal 2009) However, both are antigens that induce antibody production and development of immune memory (Kadioglu, et al. 2008; Rajagopal 2009). In addition, both capsule production and hemolysin activity have many implications that are not fully understood (Kadioglu, et al. 2008; Geno, et al. 2015). For example, non-encapsulated strains exhibit greater ability to adhere to and colonize an epithelium (Hammerschmidt, et al. 2005). Length of cell-chains is also linked to virulence in streptococci: shorter chains are associated with acute infection and increased survival in the blood as they are less likely to activate the complement, whilst longer chains promote adherence and colonization (Rodriguez, et al. 2012). As pathogenic bacteria are continuously exposed to reactive oxygen released by immune cells, resistance to oxidative compounds contributes to virulence in pathogens allowing them to avoid immune clearance (Hegde, et al. 2008; Andisi, et al. 2012). Biofilm formation is also generally associated with virulence in bacteria; detached cells are associated with acute infections whereas biofilm aggregates are a hallmark of chronic conditions (Bjarnsholt 2013). The expression of these traits among *S. iniae* isolates is highly consistent with

phylogenetic affiliation, which confirms that strain diversity is primarily driven by adaptation to the host achieved via adjustment of virulence. Furthermore, we find that occurrence of abnormal virulence determinants is strongly associated with number of mutations in MMR and OG genes and occurrence of infection in unusual immunological landscapes such as mammals and bone tissue of immunised fish. These data support the critical role of mutators in adaptation to the host and evolution of virulence, and suggest that these strains may facilitate the reinfection of immunised animals and transmission between divergent host species.

## Results and Discussion

High mutation rates are costly due to inevitable accumulation of deleterious changes. Once adaptation is gained the mutator strain is predicted to evolve back towards a low mutation rate (de Visser 2002), most likely via compensatory, 'anti-mutator' mutations in DNA repair genes (Turrientes, et al. 2013; Wielgoss, et al. 2013). Considering this volatility of mutator phenotypes, their involvement in adaptation cannot be accurately assessed by *de facto* presence among sporadic isolates, as the latter might have already (partially or fully) adapted and restored a lower mutation rate. However, mutator and anti-mutator mutations that cause adaptive fluctuations in the fidelity of DNA repair could be identified in the genotype of an adapted strain. Also, after loss and gain of functional DNA repair alleles, the mutation rate of an evolved strain is not likely to be exactly the same as before adaptation. Consequently, variation in DNA repair genes and mutation rate determined phenotypically by fluctuation analysis might indicate adaptation via a mutator state when placed in a phylogenetic context.

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

variation between divergent strains is significant, the magnitude of the differences is modest (largely around 5-fold) (Figure 1, 2), which is consistent with mutator phenotype transience and restoration of low mutation rate in adapted populations. Clade A might be a dominant circulating strain as it has persisted globally for almost two decades (Clade A contains strains isolated from USA, Honduras, and Australia between 1999 to 2016) primarily infecting closely related Perciform fish (barramundi (*Lates calcarifer*), tilapia (*Oreochromis sp*), jade perch (*Scortum barcoo*)). Mutation rate in all isolates along the ancestral lineage (total of 35 strains) falls within 1.5 - 2 x $10^{-8}$ range (Fig 1, 2; clade A),which is similar to the base mutation rate estimated for other non-mutator gram positive bacteria (Hall and Henderson-Begg 2006; Gould, et al. 2007), and all the differences are insignificant except three pairwise comparisons that verge on commonly accepted significance level. MMR and GO genes in these strains are identical, and only minor deviations from wildtype traits contributing to virulence were observed among the strains (absence of capsule production and longer cell chains in QMA0158, 216) (Figs. 1-3). The rest of the lineages express significantly different mutation rates (except strains from clade C1 and QMA0445-46 discussed below), contain unique and often multiple SNPs in *mut* genes, and show peculiar phenotypic variants related to virulence (Figures 2, 3). Although phylogenetic relatedness between lineages is fully resolved, they originate almost simultaneously and nested lineages evolve independently sharing little phylogenetic history (Figure 1,2). Lack of correlation between branch length and sampling date derived by root-to-tip regression analysis with time is good evidence of non-neutral evolution amongst this collection of *S. iniae* isolates ($R^2$=0.0936, Correlation coefficient 0.306, best-fit root; Supplementary Figure 1)((Rambaut, et al. 2016). Deviation from a neutral model is also supported by highly skewed branching in the tree, with some branches comprising single isolates whilst others comprise many epidemiologically unrelated isolates (Figure 1, 2). This is suggestive of frequent strong selection, presumably imposed by heterogeneity of the immune landscape encountered during transfer between host individuals, and resembles genealogical tree topology derived from viral evolution over similar timespans (Bedford, et al. 2011; Neher and Hallatschek 2013) although we acknowledge tree topology may also be affected by sampling bias.

Clade B contains QMA0084, a strain isolated from a black flying fox, *Pteropus alecto*, in Western Australia (WA) and two strains isolated from ornamental fish (clown loach or tiger botia, *Chromobotia macracanthus*) in USA. These isolates share valine to isoleucine substitution in MutL, a synonymous SNP in *recD2*, and 56 bp deletion within the promoter of *mutY*. The latter mutation is predicted to drastically changes *mutY* transcription pattern, affecting transcription factor binding sites of the original promoter and introducing a second promoter downstream. It appears that transcription is also affected by the rest of the genetic background as isolates exhibit significant difference in mutation rate phenotype, 7 x $10^{-8}$ in QMA0084 and 4 x $10^{-8}$ in clown loach strains. Phenotypically, clade B has shifted towards

decreased haemolytic activity. Clade B shares a common ancestor with clade C comprised of strains obtained from barramundi farmed in WA, Northern Territory (NT) (nested clade C1), and north QLD fish farms (nested clade C2). While a core mutation rate of 1.5-2 x $10^{-8}$ is observed in clade C1, a significantly higher mutation rate of 7 x $10^{-8}$ has been maintained in sister clade C2 from north QLD for almost two decades (1995 to 2012) (Fig 1). We speculate that the latter might be attributable to difference in meteorological conditions: periodic heavy rainfall present in tropical north QLD is likely to bring fluctuations in water salinity and temperature, creating an unstable environment where adaptive processes are ongoing and high mutation rate is advantageous. Minor variation in virulence-related phenotypes (absence of capsule production and longer cell chains in QMA0074, 77)(Fig 2), are supportive that mutation rate evolution in this lineage might be driven by outside-the-host factors. The significant difference in mutation rate between nested clades C1 and C2 is potentially attributable to variation in *dnaN*: A predicted deleterious tyrosine to isoleucine substitution (Table 2, PROVEAN score -2.216) is present in all isolates of C2, which may explain the high mutation rate phenotype in clade C2 strains (which also have a synonymous variant in *recD2*)(Fig. 2, Table 2). Indeed, deletion of *dnaN1* in *Bacillus anthracis* results in a mutator phenotype with equivalent mutation rate to a *mutS* mismatch repair-defective strain, supportive of the connection between deleterious mutation in DnaN and mutator phenotype found here (Yang and Miller 2008). Moreover, the T326I substitution in *S. iniae* clade C2 DnaN occurs in one of the critical residues of the loader binding interface of the ß-clamp, which may impede interaction with MutS. While MutS can perform mismatch repair independently of DnaN in *B. subtilis*, 90% of mismatch repair is dependent on targeting MutS to nascent DNA via the DnaN clamp zone (Dupes, et al. 2010; Lenhart, et al. 2013). In contrast the T1135C nucleotide substitution in the *dnaN* stop codon is unique to clade C1 strains and changes stop codon TAA into CAA coding for glutamine, with the next stop codon TAG found immediately downstream. This mutation does not appear to affect mutation and was not predicted to have any effect on the encoded protein function (Table 2) via single amino acid protein elongation (Korkmaz, et al. 2014) producing a mutation rate that is insignificantly different from the core rate found in clade A (Fig 2). Clade D consists of isolates from trout (*Oncorhynchus mykiss*) from Réunion and Israel. These strains have an estimated mutation rate of 5 x $10^{-8}$, contain a glutamate to aspartate substitution in *mutM* predicted to be deleterious to protein function, a synonymous SNP in *RecD2*, and exhibit impeded haemolytic activity (Fig 1,2, Table 2).

Isolates from humans and related fish strains are found in clade E where multiple phenotypic variants contributing to virulence are observed. This clade contains two nested subclades that share a SNP in *mutS,* but have other unique SNPs and different mutation rates, which is significant in most pairwise comparisons. Subclade E1 contains USA isolates from human (QMA0133-35, 37-38) and hybrid striped

bass (QMA0447-48), which have a substitution in the Shine-Dalgarno sequence of *recD2* and a mutation rate of 6.5 x $10^{-8}$ (Fig 1, Table 2). Fish strains and three human strains (QMA0135, 37-38) show a decreased ability to form cell chains and an elevated ability to withstand oxidative stress. Two human strains QMA0133-34 form somewhat thicker and considerably denser biofilm structures. QMA0133 is non-encapsulated, and QMA0134 appears to differentially express the capsular polysaccharide in a culture. The close phylogenetic relationship presented here implicates a likely transfer from farmed hybrid bass to humans (Facklam, et al. 2005), and this host jump may have been facilitated by the mutator phenotype. Subclade E2 contains two human strains from Canada (QMA0130-31) and tilapia strain from USA (QMA0466). These strains have unique methionine to isoleucine substitution at the N-terminus of the MutX, and mutation rate of 4.5 x $10^{-8}$ (Fig 1, Table 2). Both human strains are non-encapsulated (Fig 2). The tilapia strain expresses the capsule but forms short cell chains in common with most strains from the subclade (Fig 2). The close phylogenetic relationship again strongly implicates transfer from tilapia farmed in USA to human patients in Canada via import into the Toronto fish market (Weinstein, et al. 1997) and again, this host jump may have been facilitated by the mutator state. Clade F is not readily definable by location, time of isolation, or host species and comprises QMA0139 from unknown fish species obtained in Canada, QMA0190 isolated from snakehead murrel in Thailand (both of which appear on long branches), and a nested terminal clade comprised of barramundi isolates farmed in Recirculating Aquaculture Systems (RAS) in New South Wales and South Australia. The latter are of a particular interest as they were sampled from barramundi bone lesions during a disease outbreak in vaccinated fish where infection manifested itself as a slowly progressing osteomyelitis instead of typical acute septicaemia and meningitis (Agnew and Barnes 2007; Millard, et al. 2012). Since they are distant to strains used in autogenous vaccines in clade A (QMA0155-57, 160, QMA0250-52), the atypical outbreak might be classified as a case of vaccine-induced serotype replacement (VISR) – spread of co-existing pathogenic strain/s after elimination of the dominant strain/s targeted by the vaccine (Weinberger, et al. 2011). In contrast to natural populations where the investigation of VISR is confounded by ecological and sampling biases, and randomised vaccine trials that lack power to detect the population-wide effect of mass vaccination (Weinberger, et al. 2011), barramundi RAS represent a relatively controlled and isolated natural environment where vaccination against *S. iniae* fails occasionally but recurrently (Millard, et al. 2012). RAS are self-contained facilities with minimal water exchange and high standards of water treatment (Losordo, et al. 2009). This ensures that gene flow and, consequently, the probability of new serotype introduction after vaccine implementation is reduced to a minimum. A comparatively uniform immunized host population is exposed to dominant strains against which it was vaccinated along with potentially co-existing lineages in the RAS. In the atypical outbreaks in RAS in Australia, vaccine-induced immunity was partially

284 effective in the infected fish, evident by absence of proliferation of the new strain in typical niduses
285 (blood, brain, and pronephros) and cross-reactivity of vaccine-induced antibodies against bone isolates
286 (Millard, et al. 2012). This is in contrast with the classical model of serotype replacement where a
287 vaccine is ineffective against the replacing co-existing strain. It has been postulated that VISR can occur
288 even if the vaccine is equally efficient against all circulating strains via other trade-off mechanisms
289 (Weinberger, et al. 2011). Apparently, evasion from immune clearance and adaptation to osseous
290 tissue was allowed by shift towards attenuated virulence as suggested by multiple phenotypic changes
291 in bone isolates: absence of capsular production, impeded hemolytic activity, increased cell-chain
292 length, and denser biofilms. All of these changes have previously been associated with chronic infection
293 and increased potential for colonization (Locke, Colvin, Datta, et al. 2007; Locke, Colvin, Varki, et al.
294 2007; Kadioglu, et al. 2008; Andisi, et al. 2012; Rodriguez, et al. 2012; Bjarnsholt 2013; Geno, et al.
295 2015), and only absence of polysaccharide capsule is shared with related QMA139 and QMA190, as well
296 as QMA0140-41, QMA0187, and QMA0445-46, which implicates a non-encapsulated ancestor for these
297 strains, and possibly clade E (assuming capsular production was regained in encapsulated strains found
298 in this branch). Multiple SNPs in repair genes and the high mutation rate in bone isolates suggest that
299 shift in virulence allowing persistence in the bone tissue might have occurred via mutator phenotype.
300 Serine to arginine substitution in MutL and tyrosine to isoleucine substitution in MutY occur in the
301 ancestor of QMA0139, QMA0190, and bone strains. Both substitutions are predicted to have a
302 deleterious effect on protein function, but their effect on mutation rate is combined with unique SNPs
303 found in each branch: QMA0139 has a SNP in the binding site of the DnaA transcription factor within
304 the *dnaN* promoter, and a mutation rate of 4.5 x $10^{-8}$, QMA0190 has synonymous SNP in *mutL* and
305 mutation rate of 6.5 x $10^{-8}$, and bone strains have glutamine to glycine substitution in MutS and
306 synonymous SNP in *mutL*, and a mutation rate of 3-4 x $10^{-8}$. Considering that the mutation rate of
307 QMA0190 is significantly higher compared to rates expressed by QMA0139 and bone isolates, it is likely
308 that variants unique to the latter strains compensate for deleterious variants shared by the isolates
309 (Wielgoss, et al. 2013). However, fluctuation analysis in streptococci is intrinsically confounded when
310 colonies in plate counts are formed by a cell chain rather than a single cell. Therefore, mutation rate of
311 long-chained bone isolates might be proportionately underestimated in these experiments.

312 Clade G contains two strains from tilapia isolated in USA (QMA0445-46). Notably, despite being
313 phylogenetically distant from clade A, these strains have the same MMR and GO genotype and have
314 retained the core mutation rate and the wildtype virulence phenotype. The first long branch with a
315 single isolate contains a strain from snakehead murrel strain (*Channa striata*) from Thailand. This isolate
316 is non-encapsulated and weakly haemolytic and has a mutation rate of 4.5 x $10^{-8}$, potentially
317 attributable to deleterious aspartate to asparagine substitution in rnhC and SNP in the binding site of

318 *fnr* transcription factor within the *dnaN* promoter. A second long branch with a single strain contains
319 the oldest among the analysed strains, QMA0140 isolated in 1976 from dolphin (*Inia geoffrensis*) (Pier
320 and Madin 1976). This strain exhibits a mutation rate of 1 x $10^{-8}$, which is unique among the strains and
321 significantly lower than the core mutation rate, presumably associated with increased translation rate
322 of recD2 helicase produced by a SNP in the ribosome binding site (Fig 2, Table 2). The longest branch on
323 the tree contains a second dolphin strain, QMA0141, isolated two years later in 1978 (Pier, et al. 1978).
324 This isolate is highly divergent from the rest of strains with around 20 kB of non-recombinant SNPs in
325 pair-wise comparisons with other strains, accounting for around 1% genomic difference. In contrast to
326 QMA140, it mutates at a rate of 1x $10^{-7}$, the highest mutation rate phenotype determined among the
327 isolates and significantly different to the rest of the values. Multiple SNPs are observed in all MMR and
328 GO genes: 3 in *dnaN* and *mutX*, 4 in *mutM*, 9 in *mutL* and *rnhC*, 15 in *recD2*, 9 in *mutL* and *rnhC*, and 68
329 in *mutS*. Both dolphin isolates are non-encapsulated, show increased ability to withstand oxidative
330 stress, and form denser and thicker biofilms (Fig 2, 3).

331 The conservation of *S. iniae* DNA repair genotypes and phenotype within phylogenetic clades and
332 variation between the lineages is consistent with the idea of adaptive fluctuations in mutation rate
333 being a driver of pathogen evolution, and their association with variation in virulence traits supports
334 the contention that adaptive processes and diversification of pathogens are largely driven by host
335 immunity. Moreover, infections in atypical hosts (mammals) and atypical tissues (bone) within
336 apparently immune animals correlates strongly with the degree of deviation from the wildtype
337 virulence phenotype (p=0.000) and the number of variants in DNA repair genes (p=0.002). This indicates
338 that adaptation to a drastically different immunological landscape might be facilitated by elevated
339 mutation supply permitting rapid adjustment of virulence and antigenicity, which has major
340 epidemiological implications. First, mutators might facilitate pathogen transition among divergent host
341 species known as 'host jumps', including to higher host taxa (eg. *Osteichthyes* and *Mammalia*)(Baumler
342 and Fang 2013) . Second, mutators might present a risk factor to immune escape after vaccination and
343 serotype replacement. Moreover, as evidenced by *S. iniae* colonization of vaccinated barramundi bone
344 described previously (Millard, et al. 2012), vaccine escape leading to serotype replacement might occur
345 without major disruption of immune recognition when multiple changes attenuating virulence and
346 antigenicity allow persistence in a tissue with lower immune surveillance.

347 To summarize, we estimate mutation rates among the strains by fluctuation analysis, identify the
348 potential genetic bases for the observed differences by comparing MMR and GO genes and their
349 regulatory regions, highlight correlating phenotypic changes associated with virulence, and interpret
350 the data in a context of the phylogeny and the case histories of strain isolation. Our results highlight
351 that mutators and mutation rate dynamics play a critical role in epidemiology of bacterial pathogens,

352 evolution of virulence and antigenicity, and emergence of new pathogenic strains and serotypes. In
353 particular, mutators might associate with major adaptive events leading to colonization of a new
354 host tissue, infection of novel host taxa, and survival in vaccinated hosts with specific adaptive
355 immune response Further experimental research involving isogenic MMR/GO gene knockout
356 mutants is required to support our conclusions and determine how generalizable these findings are
357 to other pathogens. It would be of particular interest to conduct similar studies in the
358 Pneumococcus where serotype evolution following vaccination programmes is well documented,
359 but the role of mutators in this process is not yet known (Croucher, et al. 2011; Croucher, et al.
360 2013; Croucher, Kagedan, et al. 2015).

361

362 **Materials and methods**

363 **Strains and growth conditions**

364 Eighty isolates of *S. iniae* collected in Australia, USA, Canada, Israel, Honduras, and Thailand between
365 1976 and 2016 from eight fish species (*Lates calcarifer*, *Scortum barcoo, Epalzeorhynchos frenatum,*
366 *Epalzeorhynchos bicolor, Oreochromis sp., Channa striata, Chromobotia macracanthus, Oncorhynchus*
367 *mykiss*) and three mammalian species (*Homo sapiens, Inia geoffrensis, Pteropus alecto*) were used in
368 this study (Table 1). Strains were received from culture collections, veterinarians, or directly from fish
369 farms (Supplementary Table 1) and stored as master seed stocks without further subculture at -80°C in
370 Todd-Hewitt Broth (THB) supplemented with 20% glycerol. Bacteria were routinely recovered from -
371 80°C stocks on Columbia agar supplemented with 5% defibrinated sheep blood (Oxoid, Australia), and
372 cultured at 28°C on Todd-Hewitt agar or in Todd-Hewitt broth (Oxoid) with agitation 200 rpm in a
373 shaking incubator unless otherwise specified.

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

375 To estimate mutation rates of *S. iniae* isolates, a fluctuation analysis assay for spontaneous occurrence
376 of rifampicin resistance was optimized according to Rosche and Foster (Rosche and Foster 2000). A
377 single broth culture was initiated from five separate colonies, recovered on Columbia blood agar from
378 stock cultures stored at -80°C, and grown overnight to late-exponential phase in Todd-Hewitt broth.
379 Cultures were adjusted to $OD_{600}$ = 1 corresponding to ~$10^8$ viable CFU, diluted 1:100, and distributed in
380 200 µl aliquots into 8 wells of a sterile U-bottom 96-well plate (Greiner) (initial number of CFU ($N_0$) ~$10^5$
381 per well; lower inocula produced variability in final CFU number, $N_t$). A replicate plate containing at
382 least two replicate cultures per strain was also prepared in order to monitor the culture growth by
383 optical density with a BMG FLUOstar OPTIMA microplate reader. The 96-well plates were incubated

384 without agitation to early stationary phase. Viable CFU counts were performed for two representative
385 cultures for each strain by Miles and Misra method (Hedges 2002), and estimated as 1-2 x $10^8$ CFU per
386 culture. For selection of mutants, whole 200 µL cultures were plated on Todd-Hewitt agar containing
387 0.5 µg/mL rifampicin, dried in a laminar flow hood, and incubated until rifampicin resistant colonies
388 appeared. The assay was repeated three or more times for each strain, and results were pooled into
389 single data sets representing 32 to 120 cultures per strain. Estimation of mutation rate (µ - probability
390 of mutation per cell per generation), was carried out using the Ma-Sandri-Sarkar Maximum Likelihood
391 Estimation (MSS-MLE) method as implemented in FALCOR fluctuation analysis calculator (Hall, et al.
392 2009) .

393 **DNA extraction, preparation and sequencing**

394 Genomic DNA was extracted from cells collected from 10 mL late-exponential phase culture in Todd-
395 Hewitt broth with the DNeasy Blood & Tissue kit (Qiagen) using a modified protocol with an additional
396 lysis step as described previously (Kawasaki, et al. 2018). gDNA was analysed on agarose gel, quantified
397 by Qubit fluorimetry (Invitrogen), and 16S rRNA gene sequenced using universal 27F and 1492R primers
398 (Amann, et al. 1995), to confirm integrity, identity, and purity. When the required concentration for
399 sequencing was not achieved in samples they were dried by vacuum centrifugation (SpeedVac) at room
400 temperature. Sequencing was performed on the Illumina HiSeq2000 platform from Nextera XT pair-end
401 libraries at Australian Genome Research Facility, Melbourne. A reference genome from strain QMA0248
402 was constructed using both long reads derived from a single Smrt Cell using the PacBio RS II system
403 with P4C2 chemistry and short reads from Illumina HiSeq2000 derived from Nextera XT paired-end
404 libraries as reported elsewhere (NCBI accession no: GCA_002220115.1). All sequence data is deposited
405 at NCBI under Bioproject number PRJNA417543, SRA accession SRP145425. Sample numbers, accession
406 numbers and extended metadata are provided in Supplementary Table S1. Assembly statistics are
407 provided in Supplementary Table S2

408 **Phylogenetic analysis**

409 Phylogeny was constructed based on core genome SNPs from *de novo* genome assemblies filtered to
410 remove recombination breakpoints. Paired-end reads from Illumina were trimmed with Nesoni clip tool
411 version 0.132 (http://www.vicbioinformatics.com/software.nesoni.shtml), with minimum read length
412 50, and the first 15 bp of each read removed as quality deterioration in this region was observed when
413 assessed with FASTQC version 0.11.5. Assembly was performed using the SPAdes assembler version
414 3.7.1 (Bankevich, et al. 2012), with minimum read coverage cutoff set to 10. Quality of assemblies was
415 assessed with QUAST 3.2 (Gurevich, et al. 2013). Contigs were ordered by alignment to QMA248
416 reference genome (CP022392.1) with Mauve Contig Mover 2.4.0 (Rissman, et al. 2009). Genome

annotation was performed using Prokka 1.11 (Seemann 2014). Rapid alignment of core genomes was carried out using parsnp in the Harvest Tools suite version 1.2 (Treangen, et al. 2014), and the resulting alignment provided as an input to Gubbins 1.4.7 (Croucher, Page, et al. 2015) for detection and exclusion of variants produced by recombination. Phylogenies were then inferred from post-filtered core genome polymorphic sites by maximum likelihood using RAxML 7.2.8 (Stamatakis 2014) with the general time reversible nucleotide substitution model GTRGAMMA and bootstrap support from 1000 iterations. Ascertainment bias associated with using only polymorphic sites was accounted for using Felsenstein's correction in RAXML. The resulting phylogenetic tree was visualized using Dendroscope v 3.5.7 (Huson, et al. 2007) with bootstrap node support value cut-off 75. For the phylogram figure, tip labels were hidden for clarity and the edge containing QMA0141 was re-scaled as dotted line representing 100-fold decrease in length (Figure 1). A cladogram based on the inferred phylogeny, showing all tip labels and bootstrap support for each node, was annotated with metadata using Evolview V2 (Figure 2) (He, et al. 2016). To determine whether there was a strong temporal signal in the phylogenetic data, a root-to-tip regression of branch length against time since isolation was performed in TempEst (Rambaut, et al. 2016), using genetic distances in an unrooted tree estimated by maximum likelihood from the alignment of non-recombinant core-genome SNPs in RAxML.

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

SNP analysis of MMR and GO genes was performed by read-mapping with Geneious version 9.1 (Kearse, et al. 2012), using default settings unless otherwise specified. Paired-end reads from each genome were trimmed, merged into a single file with expected distances between the reads set to 250 bp, and mapped to reference genome of strain QMA0248. Mapped reads were used to detect SNPs with minimum coverage set to 10 and frequency to 0.9. A consensus pseudogenome was generated for every strain based on the reference sequence including any detected variants. Multiple alignment of pseudogenomes was carried out with *Geneious* aligning tool. Sequences of MMR genes (*dnaN,* mut*S, muTL, recD2, rnhC*) and OG genes (*mutY mutY* , *mutM*, *mutX*) annotated in the QMA0248 reference genome were identified in each strain genome sequence. Promoter regions of repair genes were determined with BPROM (V. Solovyev 2011), protein functional domains by SMART genomic (Letunic and Bork 2018), and effect of amino acid substitutions on protein function by PROVEAN Protein with a sensitivity cut-off of 1.3 (Choi and Chan 2015). To detect large variants, alignment of these regions extracted from *de novo* genome assemblies was carried out, and a 56 bp deletion detected in *mutY* promoter of clade B was confirmed by PCR using CAGAAGGAAGAAACAGAC_F/ACCTCTATTGTAGCAAAG_R primers.

**Phenotypic variation related to virulence and antigenicity**

450 A multitude of intracellular and secreted enzymes contribute to virulence in the *Streptococcus* genus
451 (Kadioglu, et al. 2008; Rajagopal 2009). Considering the large number of strains, we limited analysis to
452 major, easily observable, and distinct phenotypes strongly associated with virulence in streptococci
453 (capsule, hemolysis, cell chains) and bacteria in general (oxidation resistance, biofilms). All assays were
454 performed at least in triplicate per strain.

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

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

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

467 Rapid high-throughput detection of impaired haemolytic activity was achieved by blood-agar clearance
468 zone assay. Briefly, 5 mm wells were made in Columbia agar supplemented with 5% defibrinated sheep
469 blood (Oxoid, Australia). Bacterial cultures grown to late-exponential phase (~$10^8$ CFU) were diluted
470 1:1000 and 50 µL of diluted cultures were pipetted into punctures in the agar (initial inoculum ~$5x10^4$
471 CFU in each puncture). A 3 mm wide clearance zone was generally produced by bacterial lysis of sheep
472 erythrocytes during 24 h incubation (Figure 3, B1). Where haemolysis was absent or fragmentary,
473 impeded haemolytic activity was recorded (Figure 3, B2).

474 **Chain formation microscopy**

475 To assess chain formation, *S. iniae* cultures in THB were grown stationarily in 96-well plates at 28 °C for
476 24 h, mixed and 5 µL wet mounts prepared on a glass microscope slide. Slides were observed by bright
477 field microscopy under 40x objective with an Olympus BX40 microscope and captured using an
478 Olympus DP28 digital camera using CellSens software (Olympus Optical Co, Japan). Specimens were
479 observed for at least 3 min and 2-20 cell chains were generally observed (Figure 3, C1). Where over 20
480 cells in a chain were repeatedly detected (Figure 3, C2) increased chain formation was recorded, and
481 when more than 10 cells in a chain were not detected by similar observation and the culture was mainly

482 composed of detached cells (Figure 3, C3) impeded chain formation was recorded. For the figure, 50 µL
483 of cultures were dried onto slides at RT, fixed with methanol, Gram stained, and images captured under
484 100x objective.

485 **Oxidation resistance assay**

486 Minimum inhibitory concentration (MIC) and minimum bactericidal concentration (MBC) (Andrews
487 2001) of hydrogen peroxide were measured to assess oxidative stress resistance among *S. iniae* strains.
488 THB cultures grown to late exponential phase were adjusted to $OD_{600}$ = 1 ($10^8$ CFU/mL), diluted 1000-
489 fold, and distributed by 100 µL (~$10^4$ CFU) into wells of a U-bottom 96-well plate (Greiner). THB (100 µL)
490 with 2-fold excess concentration of hydrogen peroxide was added to the wells and serially diluted
491 twofold, resulting in a range from 0 to 10 mM final peroxide concentrations. Plates were incubated
492 stationarily for 24h and examined for presence of observable growth to determine the MIC. Viable cell
493 counts of cultures without visible growth were performed to determine MBC. MIC of hydrogen
494 peroxide was estimated as 3 mM for all strains, and MBC as 4 mM for the majority of strains. MBC
495 elevated to 5 mM in dolphin and human isolates from USA was classified deviation from a wildtype
496 phenotype.

497 **Biofilm formation and visualisation assay**

498 *S. iniae* biofilms were grown for 4 days in 8-well Lab-Tek® II Chamber Slide™ Systems. To prepare an
499 initial inoculum of 5 x $10^5$ CFU, THB cultures grown to late exponential phase were adjusted to $OD_{600}$ = 1
500 ($10^8$ CFU/mL), diluted 100-fold, and 0.5 mL of diluted cultures were placed into Chamber slide wells.
501 After 24 h incubation, THB was removed and replaced with fresh THB every 10-14 h for 3 days. For
502 visualisation, biofilms were washed in PBS, stained for 15 min with 1 µM fluorescent *Bac*Light Red
503 bacterial stain, washed in PBS, fixed for 30 min with 10% formalin, and washed twice PBS. Z-stacks were
504 collected by ZEISS LSM 710 Inverted Laser Scanning Confocal Microscope at 20x objective and visualised
505 using ZEN2012. 8 -12 µm thick biofilms covering up to 30% of representative 200 x 200 µm surface
506 produced by most strains were classified as typical (Figure 3, D1), denser (Figure 3, D2) or/and thicker
507 (Figure 3, D3) structures were recorded as deviant.

508 **Identifying wildtype virulence-related phenotypes**

509 We consider prevalent phenotypes as a wildtype, namely: presence of polysaccharide capsule (71.25 %
510 of strains; Figure 3, A1) and haemolytic activity (85% of strains; Figure 3, B1), up to 20 cells in a chain
511 (77.5 % of strains; Figure 3, A1), 3 mM MIC and 4 mM MBC of hydrogen peroxide (91.25 % of strains),
512 and 8 -12 µm thick biofilms covering up to 30% of representative image (86.25 % of strains, Fig. 4A).
513 Other phenotypes are regarded as deviant and discussed.

## Statistical analysis

Estimation of mutation rates was performed by Ma-Sandri-Sarkar-Maximum Likelihood Estimator (MSS-MLE) method. Fluctuation assay results were adjusted to equal number of mutants in the final culture and allowed comparison of mutation rates by Likelihood Ratio test (Zheng 2015) using rSalvador R package (Zheng 2017). *Compare.LD* function was applied pair-wise to all strains. Pairwise comparisons resulting in $p$ values less than 0.05 were considered significantly different. Associations between variation in DNA repair genes, mutation rates, deviations from the wildtype virulence phenotype, and typical places of isolation were tested with a Phylogenetic Generalised Least Squares model that accounts for relatedness among strains under Pagel's $\lambda$ of 0.6 implemented in R (Symonds and Blomberg 2014).

## Acknowledgements

This work was supported by The Australian Research Council Discovery Project Grant DP120102755 awarded to ACB. OS received a fee waiver from the University of Queensland School of Biological Sciences to whom we extend out thanks. For provision of strains into our collection over 15 years we thank: Judy Forbes-Faulkner, Oonoonba Veterinary Laboratory, QLD, Australia; Lynn Shewmaker, CDC, Atlanta, USA; Christian Michel, INRA, France; Suresh Benedict, Berrimah Veterinary Laboratories, NT, Australia; Nicky Buller, WA Fisheries, WA, Australia; Mark White, Tréidlia Biovet Pty (formerly Allied Biotechnology), NSW, Australia; Matt Landos, Future Fisheries Veterinary Services, NSW, Australia; Spencer Russel, Novartis Animal Health, Canada; Ruben Arturo Lopez Crespo, Regal Springs Tilapia, Honduras. ACB and OS thank Mr Nazar Rudenko for assistance with writing software pipelines and for use of his unix server. Sequence reads and de novo draft sequence assemblies derived from this work are deposited with NCBI and are accessible via BioProject PRJNA417543. Biosample numbers for raw reads and accession numbers of genome assemblies are provided in Supplementary Tables 1 and 2 respectively.

538 **References**

539 Agnew W, Barnes AC. 2007. Streptococcus iniae: an aquatic pathogen of global veterinary
540 significance and a challenging candidate for reliable vaccination. Vet Microbiol 122:1-15.
541 Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of
542 individual microbial cells without cultivation. Microbiology Reviews 59:143-169.
543 Ambur OH, Davidsen T, Frye SA, Balasingham SV, Lagesen K, Rognes T, Tonjum T. 2009. Genome
544 dynamics in major bacterial pathogens. FEMS Microbiol Rev 33:453-470.
545 Andisi VF, Hinojosa CA, de Jong A, Kuipers OP, Orihuela CJ, Bijlsma JJ. 2012. Pneumococcal gene
546 complex involved in resistance to extracellular oxidative stress. Infection and Immunity 80:1037-
547 1049.
548 Andrews JM. 2001. Determination of minimum inhibitory concentrations. J Antimicrob Chemother
549 48 Suppl 1:5-16.
550 Aquino T, Nunes A. 2016. Host immunity and pathogen diversity: A computational study. Virulence
551 7:121-128.
552 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S,
553 Prjibelski AD, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-
554 cell sequencing. J Comput Biol 19:455-477.
555 Baumler A, Fang FC. 2013. Host specificity of bacterial pathogens. Cold Spring Harb Perspect Med
556 3:a010041.
557 Bedford T, Cobey S, Pascual M. 2011. Strength and tempo of selection revealed in viral gene
558 genealogies. BMC Evol Biol 11:220.
559 Bjarnsholt T. 2013. The role of bacterial biofilms in chronic infections. APMIS Suppl:1-51.
560 Blazquez J. 2003. Hypermutation as a factor contributing to the acquisition of antimicrobial
561 resistance. Clinical Infectious Diseases 37:1201-1209.
562 Boe L, Danielsen M, Knudsen S, Petersen JB, Maymann J, Jensen PR. 2000. The frequency of
563 mutators in populations of Escherichia coli. Mutat Res 448:47-55.
564 Canfield GS, Schwingel JM, Foley MH, Vore KL, Boonanantanasarn K, Gill AL, Sutton MD, Gill SR.
565 2013. Evolution in fast forward: a potential role for mutators in accelerating Staphylococcus aureus
566 pathoadaptation. J Bacteriol 195:615-628.
567 Chen F, Liu WQ, Eisenstark A, Johnston RN, Liu GR, Liu SL. 2010. Multiple genetic switches
568 spontaneously modulating bacterial mutability. Bmc Evolutionary Biology 10:277.
569 Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid
570 substitutions and indels. Bioinformatics 31:2745-2747.
571 Chopra I, O'Neill AJ, Miller K. 2003. The role of mutators in the emergence of antibiotic-resistant
572 bacteria. Drug Resist Updat 6:137-145.
573 Cobey S. 2014. Pathogen evolution and the immunological niche. Ann N Y Acad Sci 1320:1-15.
574 Croucher NJ, Finkelstein JA, Pelton SI, Mitchell PK, Lee GM, Parkhill J, Bentley SD, Hanage WP,
575 Lipsitch M. 2013. Population genomics of post-vaccine changes in pneumococcal epidemiology. Nat
576 Genet 45:656-663.
577 Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, von Gottberg A,
578 Song JH, Ko KS, et al. 2011. Rapid Pneumococcal Evolution in Response to Clinical Interventions.
579 Science 331:430-434.
580 Croucher NJ, Kagedan L, Thompson CM, Parkhill J, Bentley SD, Finkelstein JA, Lipsitch M, Hanage WP.
581 2015. Selective and genetic constraints on pneumococcal serotype switching. PLoS Genet
582 11:e1005095.
583 Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid
584 phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using
585 Gubbins. Nucleic Acids Res 43:e15.

586 Dai L, Muraoka WT, Wu ZW, Sahin O, Zhang QJ. 2015. A single nucleotide change in mutY increases
587 the emergence of antibiotic-resistant Campylobacter jejuni mutants. Journal of Antimicrobial
588 Chemotherapy 70:2739-2748.
589 David SS, O'Shea VL, Kundu S. 2007. Base-excision repair of oxidative DNA damage. Nature 447:941-
590 950.
591 de Visser JA. 2002. The fate of microbial mutators. Microbiology 148:1247-1252.
592 Deitsch KW, Moxon ER, Wellems TE. 1997. Shared themes of antigenic variation and virulence in
593 bacterial, protozoal, and fungal infections. Microbiol Mol Biol Rev 61:281-293.
594 Dupes NM, Walsh BW, Klocko AD, Lenhart JS, Peterson HL, Gessert DA, Pavlick CE, Simmons LA.
595 2010. Mutations in the Bacillus subtilis beta clamp that separate its roles in DNA replication from
596 mismatch repair. J Bacteriol 192:3452-3463.
597 Facklam R, Elliott J, Shewmaker L, Reingold A. 2005. Identification and characterization of sporadic
598 isolates of Streptococcus iniae isolated from humans. Journal of Clinical Microbiology 43:933-937.
599 Ferenci T. 2015. Trade-off Mechanisms Shaping the Diversity of Bacteria. Trends Microbiol.
600 Fukui K. 2010. DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids 2010.
601 Galhardo RS, Hastings PJ, Rosenberg SM. 2007. Mutation as a stress response and the regulation of
602 evolvability. Crit Rev Biochem Mol Biol 42:399-435.
603 Geno KA, Gilbert GL, Song JY, Skovsted IC, Klugman KP, Jones C, Konradsen HB, Nahm MH. 2015.
604 Pneumococcal Capsules and Their Types: Past, Present, and Future. Clin Microbiol Rev 28:871-899.
605 Giraud A, Matic I, Tenaillon O, Clara A, Radman M, Fons M, Taddei F. 2001. Costs and benefits of high
606 mutation rates: Adaptive evolution of bacteria in the mouse gut. Science 291:2606-2608.
607 Gonzalez K, Faustoferri RC, Quivey RG, Jr. 2012. Role of DNA base excision repair in the mutability
608 and virulence of Streptococcus mutans. Molecular Microbiology 85:361-377.
609 Gould CV, Sniegowski PD, Shchepetov M, Metlay JP, Weiser JN. 2007. Identifying mutator
610 phenotypes among fluoroquinolone-resistant strains of Streptococcus pneumoniae using fluctuation
611 analysis. Antimicrob Agents Chemother 51:3225-3229.
612 Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome
613 assemblies. Bioinformatics 29:1072-1075.
614 Gutierrez O, Juan C, Perez JL, Oliver A. 2004. Lack of association between hypermutation and
615 antibiotic resistance development in Pseudomonas aeruginosa isolates from intensive care unit
616 patients. Antimicrob Agents Chemother 48:3573-3575.
617 Hall BM, Ma CX, Liang P, Singh KK. 2009. Fluctuation analysis CalculatOR: a web tool for the
618 determination of mutation rate using Luria-Delbruck fluctuation analysis. Bioinformatics 25:1564-
619 1565.
620 Hall LM, Henderson-Begg SK. 2006. Hypermutable bacteria isolated from humans--a critical analysis.
621 Microbiology 152:2505-2514.
622 Hammerschmidt S, Wolff S, Hocke A, Rosseau S, Muller E, Rohde M. 2005. Illustration of
623 pneumococcal polysaccharide capsule during adherence and invasion of epithelial cells. Infection
624 and Immunity 73:4653-4667.
625 He Z, Zhang H, Gao S, Lercher MJ, Chen WH, Hu S. 2016. Evolview v2: an online visualization and
626 management tool for customized and annotated phylogenetic trees. Nucleic Acids Res 44:W236-241.
627 Healey KR, Zhao Y, Perez WB, Lockhart SR, Sobel JD, Farmakiotis D, Kontoyiannis DP, Sanglard D, Taj-
628 Aldeen SJ, Alexander BD, et al. 2016. Prevalent mutator genotype identified in fungal pathogen
629 Candida glabrata promotes multi-drug resistance. Nat Commun 7:11128.
630 Hedges AJ. 2002. Estimating the precision of serial dilutions and viable bacterial counts. Int J Food
631 Microbiol 76:207-214.
632 Hegde A, Bhat GK, Mallya S. 2008. Effect of exposure to hydrogen peroxide on the virulence of
633 Escherichia coli. Indian J Med Microbiol 26:25-28.
634 Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: An interactive
635 viewer for large phylogenetic trees. BMC Bioinformatics 8:460.

636 Kadioglu A, Weiser JN, Paton JC, Andrew PW. 2008. The role of Streptococcus pneumoniae virulence
637 factors in host respiratory colonization and disease. Nat Rev Microbiol 6:288-301.
638 Kawasaki M, Delamare-Debouteville J, Bowater RO, Walker MJ, Beatson S, Ben Zakour NL, Barnes
639 AC. 2018. Microevolution of aquatic Streptococcus agalactiae ST-261 from Australia indicates
640 dissemination via imported tilapia and ongoing adaptation to marine hosts or environment. Appl
641 Environ Microbiol.
642 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S,
643 Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for
644 the organization and analysis of sequence data. Bioinformatics 28:1647-1649.
645 Korkmaz G, Holm M, Wiens T, Sanyal S. 2014. Comprehensive analysis of stop codon usage in
646 bacteria and its correlation with release factor abundance. Journal of Biological Chemistry
647 289:30334-30342.
648 Labat F, Pradillon O, Garry L, Peuchmaur M, Fantin B, Denamur E. 2005. Mutator phenotype confers
649 advantage in Escherichia coli chronic urinary tract infection pathogenesis. FEMS Immunol Med
650 Microbiol 44:317-321.
651 Lenhart JS, Pillon MC, Guarne A, Biteen JS, Simmons LA. 2016. Mismatch repair in Gram-positive
652 bacteria. Res Microbiol 167:4-12.
653 Lenhart JS, Sharma A, Hingorani MM, Simmons LA. 2013. DnaN clamp zones provide a platform for
654 spatiotemporal coupling of mismatch detection to DNA replication. Mol Microbiol 87:553-568.
655 Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids
656 Res 46:D493-D496.
657 Li GM. 2008. Mechanisms and functions of DNA mismatch repair. Cell Research 18:85-98.
658 Locke JB, Colvin KM, Datta AK, Patel SK, Naidu NN, Neely MN, Nizet V, Buchanan JT. 2007.
659 Streptococcus iniae capsule impairs phagocytic clearance and contributes to virulence in fish. J
660 Bacteriol 189:1279-1287.
661 Locke JB, Colvin KM, Varki N, Vicknair MR, Nizet V, Buchanan JT. 2007. Streptococcus iniae beta-
662 hemolysin streptolysin S is a virulence factor in fish infection. Diseases of Aquatic Organisms 76:17-
663 26.
664 Losordo T, DeLong D, Guerdat T. 2009. Advances in technology and practice for land-based
665 aquaculture systems: tank-based recirculating systems for finfish production. New Technologies in
666 Aquaculture:945-983.
667 Lu AL, Li X, Gu Y, Wright PM, Chang DY. 2001. Repair of oxidative DNA damage: mechanisms and
668 functions. Cell Biochem Biophys 35:141-170.
669 Lukacisinova M, Novak S, Paixao T. 2017. Stress-induced mutagenesis: Stress diversity facilitates the
670 persistence of mutator genes. PLoS Comput Biol 13:e1005609.
671 Mackinnon MJ, Read AF. 2004. Virulence in malaria: an evolutionary viewpoint. Philos Trans R Soc
672 Lond B Biol Sci 359:965-986.
673 Mao EF, Lane L, Lee J, Miller JH. 1997. Proliferation of mutators in A cell population. J Bacteriol
674 179:417-422.
675 Matsushima A, Takakura S, Fujihara N, Saito T, Ito I, Iinuma Y, Ichiyama S. 2010. High prevalence of
676 mutators among Enterobacter cloacae nosocomial isolates and their association with antimicrobial
677 resistance and repetitive detection. Clin Microbiol Infect 16:1488-1493.
678 Maurelli AT. 2007. Black holes, antivirulence genes, and gene inactivation in the evolution of
679 bacterial pathogens. FEMS Microbiol Lett 267:1-8.
680 Mena A, Macia MD, Borrell N, Moya B, de Francisco T, Perez JL, Oliver A. 2007. Inactivation of the
681 mismatch repair system in Pseudomonas aeruginosa attenuates virulence but favors persistence of
682 oropharyngeal colonization in cystic fibrosis mice. J Bacteriol 189:3665-3668.
683 Mena A, Smith EE, Burns JL, Speert DP, Moskowitz SM, Perez JL, Oliver A. 2008. Genetic Adaptation
684 of Pseudomonas aeruginosa to the Airways of Cystic Fibrosis Patients Is Catalyzed by Hypermutation.
685 J Bacteriol 190:7910-7917.

686 Merino D, Reglier-Poupet H, Berche P, Charbit A, European Listeria Genome C. 2002. A hypermutator
687 phenotype attenuates the virulence of Listeria monocytogenes in a mouse model. Molecular
688 Microbiology 44:877-887.
689 Methot PO, Alizon S. 2014. What is a pathogen? Toward a process view of host-parasite interactions.
690 Virulence 5:775-785.
691 Millard CM, Baiano JC, Chan C, Yuen B, Aviles F, Landos M, Chong RS, Benedict S, Barnes AC. 2012.
692 Evolution of the capsular operon of Streptococcus iniae in response to vaccination. Appl Environ
693 Microbiol 78:8219-8226.
694 Miller JH. 1996. Spontaneous mutators in bacteria: insights into pathways of mutagenesis and repair.
695 Annu Rev Microbiol 50:625-643.
696 Negri MC, Morosini MI, Baquero MR, del Campo R, Blazquez J, Baquero F. 2002. Very low cefotaxime
697 concentrations select for hypermutable Streptococcus pneumoniae populations. Antimicrob Agents
698 Chemother 46:528-530.
699 Neher RA, Hallatschek O. 2013. Genealogies of rapidly adapting populations. Proc Natl Acad Sci U S A
700 110:437-442.
701 Nilsson AI, Kugelberg E, Berg OG, Andersson DI. 2004. Experimental adaptation of Salmonella
702 typhimurium to mice. Genetics 168:1119-1130.
703 Oliver A. 2010. Mutators in cystic fibrosis chronic lung infection: Prevalence, mechanisms, and
704 consequences for antimicrobial therapy. Int J Med Microbiol 300:563-572.
705 Oliver A, Mena A. 2010. Bacterial hypermutation in cystic fibrosis, not only for antibiotic resistance.
706 Clinical Microbiology and Infection 16:798-808.
707 Picard B, Duriez P, Gouriou S, Matic I, Denamur E, Taddei H. 2001. Mutator natural Escherichia coli
708 isolates have an unusual virulence phenotype. Infection and Immunity 69:9-14.
709 Pier GB, Madin SH. 1976. Streptococcus iniae sp nov, a beta-hemolytic Streptococcus isolated from
710 an amazon freshwater dolphin, Inia geoffrensis. . International Journal of Systematic Bacteriology
711 26:545-553.
712 Pier GB, Madin SH, Alnakeeb S. 1978. Isolation and characterisation of a 2nd isolate of Streptococcus
713 iniae. International Journal of Systematic Bacteriology 28:311-314.
714 Rajagopal L. 2009. Understanding the regulation of Group B Streptococcal virulence factors. Future
715 Microbiol 4:201-221.
716 Rajanna C, Ouellette G, Rashid M, Zemla A, Karavis M, Zhou C, Revazishvili T, Redmond B, McNew L,
717 Bakanidze L, et al. 2013. A strain of Yersinia pestis with a mutator phenotype from the Republic of
718 Georgia. FEMS Microbiol Lett 343:113-120.
719 Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of
720 heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evolution 2.
721 Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. 2009. Reordering contigs of draft
722 genomes using the Mauve aligner. Bioinformatics 25:2071-2073.
723 Rodriguez JL, Dalia AB, Weiser JN. 2012. Increased chain length promotes pneumococcal adherence
724 and colonization. Infection and Immunity 80:3454-3459.
725 Rosche WA, Foster PL. 2000. Determining mutation rates in bacterial populations. Methods 20:4-17.
726 Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068-2069.
727 Shabalina SA, Spiridonov NA, Kashina A. 2013. Sounds of silence: synonymous nucleotides as a key to
728 biological regulation and complexity. Nucleic Acids Res 41:2073-2094.
729 Shrestha S, Bjornstad ON, King AA. 2014. Evolution of acuteness in pathogen metapopulations:
730 conflicts between "classical" and invasion-persistence trade-offs. Theor Ecol 7:299-311.
731 Smania AM, Segura I, Pezza RJ, Becerra C, Albesa I, Argarana CE. 2004. Emergence of phenotypic
732 variants upon mismatch repair disruption in Pseudomonas aeruginosa. Microbiology-Sgm 150:1327-
733 1338.
734 Sniegowski P, Raynes Y. 2013. Mutation Rates: How Low Can You Go? Current Biology 23:R147-R149.
735 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
736 phylogenies. Bioinformatics 30:1312-1313.

737 Stich M, Manrubia SC, Lazaro E. 2010. Variable Mutation Rates as an Adaptive Strategy in Replicator
738 Populations. PLoS One 5.
739 Sundin GW, Weigand MR. 2007. The microbiology of mutability. FEMS Microbiol Lett 277:11-20.
740 Symonds MRE, Blomberg SP. 2014. A Primer on Phylogenetic Generalised Least Squares. In:
741 Garamszegi LZ, editor. Modern Phylogenetic Comparative Methods and Their Application in
742 Evolutionary Biology: Concepts and Practice. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 105-
743 130.
744 Taddei F, Radman M, MaynardSmith J, Toupance B, Gouyon PH, Godelle B. 1997. Role of mutator
745 alleles in adaptive evolution. Nature 387:700-702.
746 Tham KC, Hermans N, Winterwerp HHK, Cox MM, Wyman C, Kanaar R, Lebbink JHG. 2013. Mismatch
747 Repair Inhibits Homeologous Recombination via Coordinated Directional Unwinding of Trapped DNA
748 Structures. Mol Cell 51:326-337.
749 Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome
750 alignment and visualization of thousands of intraspecific microbial genomes. Genome Biology
751 15:524.
752 Turrientes MC, Baquero F, Levin BR, Martinez JL, Ripoll A, Gonzalez-Alba JM, Tobes R, Manrique M,
753 Baquero MR, Rodriguez-Dominguez MJ, et al. 2013. Normal mutation rate variants arise in a Mutator
754 (Mut S) Escherichia coli population. PLoS One 8:e72963.
755 V. Solovyev AS. 2011. Automatic Annotation of Microbial Genomes and Metagenomic Sequences. In.
756 Metagenomics and its Applications in Agriculture, Biomedicine and Environmental: Nova Science
757 Publishers. p. 61-78.
758 Wang S, Wang Y, Shen J, Wu Y, Wu C. 2013. Polymorphic mutation frequencies in clinical isolates of
759 Staphylococcus aureus: the role of weak mutators in the development of fluoroquinolone resistance.
760 FEMS Microbiol Lett 341:13-17.
761 Weinberger DM, Malley R, Lipsitch M. 2011. Serotype replacement in disease after pneumococcal
762 vaccination. Lancet 378:1962-1973.
763 Weinstein MR, Litt M, Kertesz DA, Wyper P, Rose D, Coulter M, McGeer A, Facklam R, Ostach C,
764 Willey BM, et al. 1997. Invasive infections due to a fish pathogen, Streptococcus iniae. New England
765 Journal of Medicine 337:589-594.
766 Wielgoss S, Barrick JE, Tenaillon O, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE,
767 Schneider D. 2011. Mutation Rate Inferred From Synonymous Substitutions in a Long-Term Evolution
768 Experiment With Escherichia coli. G3 (Bethesda) 1:183-186.
769 Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B,
770 Medigue C, Lenski RE, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect
771 tension between adaptation and genetic load. Proc Natl Acad Sci U S A 110:222-227.
772 Yang H, Miller JH. 2008. Deletion of dnaN1 generates a mutator phenotype in Bacillus anthracis. DNA
773 Repair (Amst) 7:507-514.
774 Zheng Q. 2015. Methods for comparing mutation rates using fluctuation assay data. Mutat Res
775 777:20-22.
776 Zheng Q. 2017. rSalvador: An R Package for the Fluctuation Experiment. G3 (Bethesda) 7:3849-3856.

777

778

779 **Supporting Information Legends**

780 Supplementary Figure 1. Root-to-tip regression analysis (TempEst, (Rambaut, et al. 2016)of branch
781 lengths from best-fit-rooted tree against time (year of isolation). The tree was derived from alignment
782 of non-recombinant core-genome SNPs corrected for ascertainment bias in RAxML.

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

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

787 **Figures and tables**

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

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

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

804

805 Table 1. *Streptococcus iniae* strains used in this study. Includes origin details (host species, time, site
806 of isolation), phylogenetic affiliation, virulence-associated phenotypes, and mutation rate. Atypical
807 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 |

808

809

810

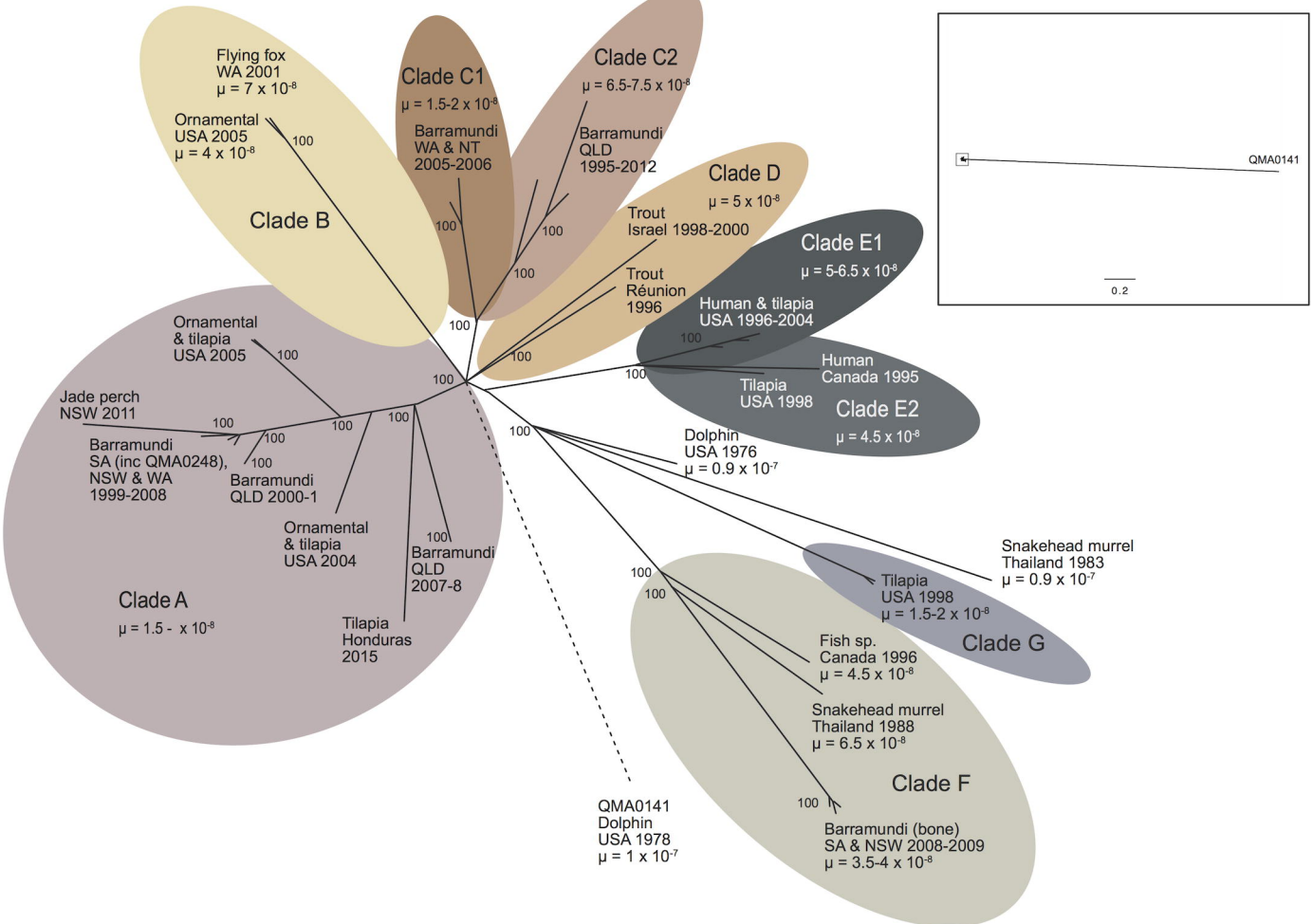
811

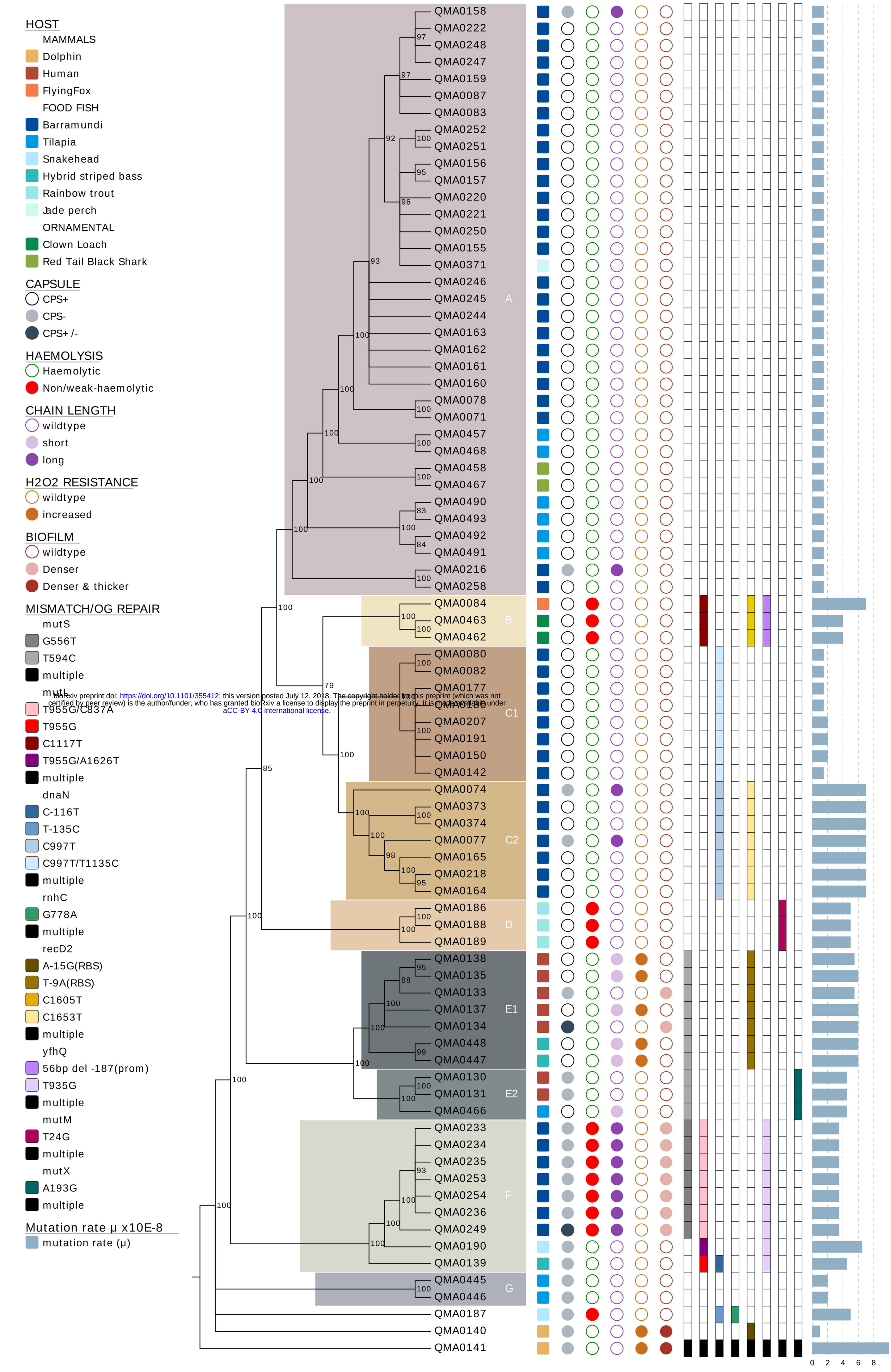
812

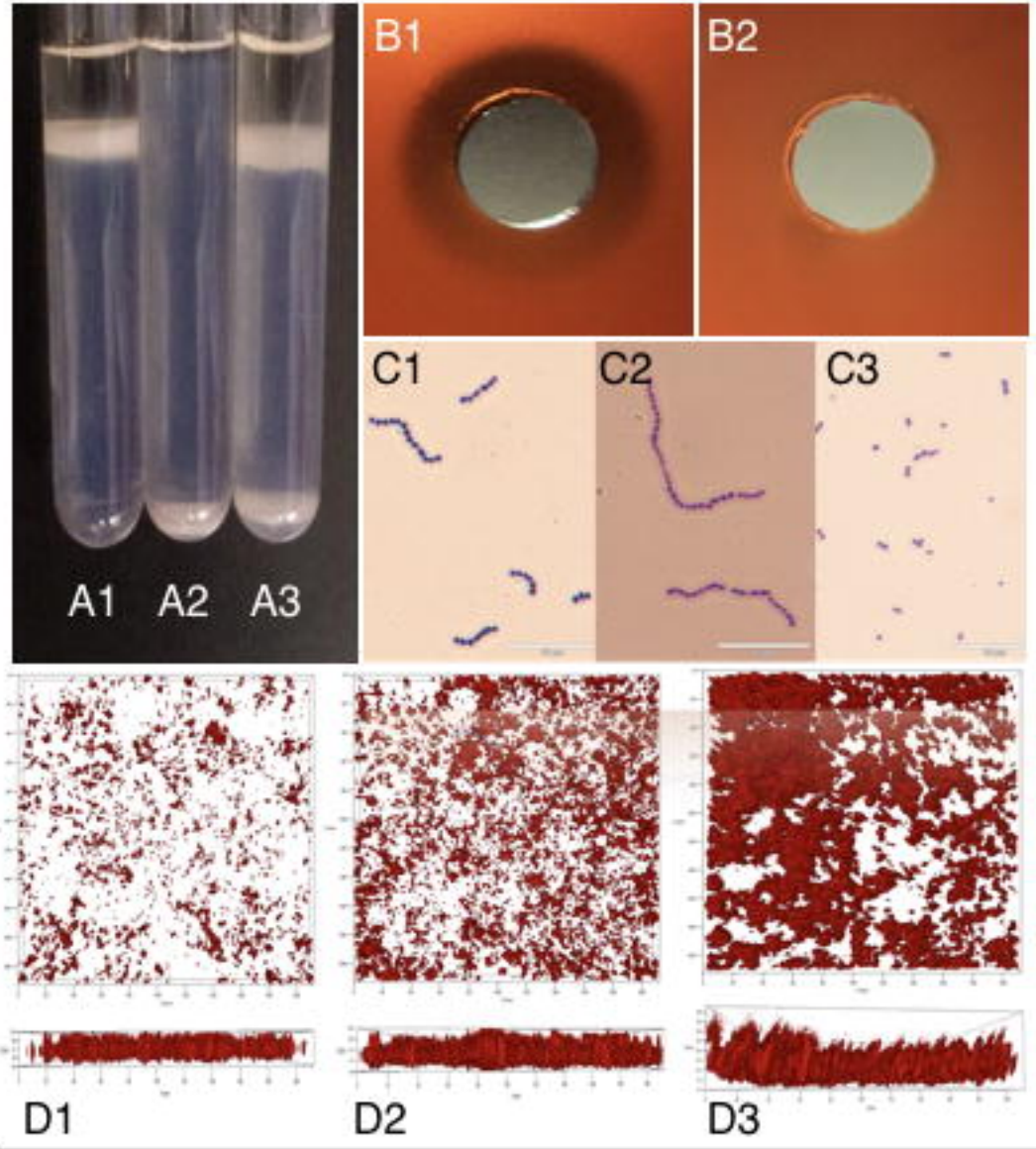
813 Table 2: Variants in MMR and GO genes found among *S. iniae* isolates. D letter next to PROVEAN
814 output value indicates predicted deleterious effect of amino acid change on protein function.

| 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 | PolBc | -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 |
| *mutY* *mutY* | 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 |

815


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




A1
A2
A3
B1
B2
C1
C2
C3
D1
D2
D3