MICROBIAL GENOMICS

RESEARCH ARTICLE

Baseggio *et al.*, *Microbial Genomics* 2021;7:000562

DOI 10.1099/mgen.0.000562

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Complete, closed and curated genome sequences of *Photobacterium damselae* subsp. *piscicida* isolates from Australia indicate mobilome-driven localized evolution and novel pathogenicity determinants

Laura Baseggio[1], Oleksandra Silayeva[1], Nicky Buller[2], Matt Landos[3], Jan Englestädter[1] and Andrew C. Barnes[1,*]

Abstract

Despite the recent advances in sequencing technologies, the complete assembly of multi-chromosome genomes of the *Vibrionaceae*, often containing several plasmids, remains challenging. Using a combination of Oxford Nanopore MinION long reads and short Illumina reads, we fully sequenced, closed and curated the genomes of two strains of a primary aquatic pathogen *Photobacterium damselae* subsp. *piscicida* isolated in Australia. These are also the first genome sequences of *P. damselae* subsp. *piscicida* isolated in Oceania and, to our knowledge, in the Southern hemisphere. We also investigated the phylogenetic relationships between Australian and overseas isolates, revealing that Australian *P. damselae* subsp. *piscicida* are more closely related to the Asian and American strains rather than to the European ones. We investigated the mobilome and present new evidence showing that a host specialization process and progressive adaptive evolution to fish are ongoing in *P. damselae* subsp. *piscicida*, and are largely mediated by transposable elements, predominantly in chromosome 2, and by plasmids. Finally, we identified two novel potential virulence determinants in *P. damselae* subsp. *piscicida* – a chorismate mutase gene, which is ubiquitously retained and co-localized with the AIP56 apoptogenic toxin-encoding gene on the pPHDP10 plasmid, and transfer-messenger RNA gene *ssrA* located on the main chromosome, homologous to a critical-to-virulence determinant in *Yersinia pseudotuberculosis*. Our study describes, to our knowledge, the only fully closed and manually curated genomes of *P. damselae* subsp. *piscicida* available to date, offering new insights into this important fish pathogen and its evolution.

DATA SUMMARY

Complete, closed and curated genome and plasmid assemblies for *Photobacterium damselae* subspecies *piscicida* strains QMA0505 and QMA0506 have been deposited at the National Center for Biotechnology Information under BioProject number PRJNA662633, with links to sample metadata and sequence reads from all platforms used. Accession numbers for assembled chromosomes and plasmids for strains QMA0505 and QMA0506 are CP061854–CP061860 and CP061861–CP061867, respectively.

INTRODUCTION

Photobacterium damselae subsp. *piscicida* is a Gram-negative marine bacterium found in warm-temperate waters [1]. Unlike its sister subspecies *P. damselae* subsp. *damselae*, which is a generalist, opportunistic pathogen [2, 3], *P. damselae* subsp. *piscicida* is a highly virulent fish-specialized pathogen with a widespread geographical distribution. In the Northern hemisphere, photobacteriosis caused by *P. damselae* subsp. *piscicida* leads to mortality events in high-value aquaculture species such as gilthead sea bream (*Sparus aurata*), sea bass (*Dicentrarchus labrax*) and sole (*Solea solea* and *Solea senegalensis*) in Europe; striped bass (*Morone saxatilis*), white perch

Received 28 September 2020; Accepted 16 March 2021; Published 22 April 2021

Author affiliations: [1]School of Biological Sciences, University of Queensland, Brisbane, Queensland 4072, Australia; [2]Diagnostic and Laboratory Services (DDLS), Department of Primary Industries and Regional Development (DPIRD), 3 Baron-Hay Court, South Perth, Western Australia 6151, Australia; [3]Future Fisheries Veterinary Services, East Ballina, New South Wales 2478, Australia.

***Correspondence:** Andrew C. Barnes, a.barnes@uq.edu.au

Keywords: Nanopore; *Photobacterium damselae* subspecies *piscicida*; phylogenetics; plasmid assembly.

Abbreviations: HPI, high-pathogenicity island; ISE, insertion sequence element; MGE, mobile genetic element; NCBI, National Center for Biotechnology Information; PTS, phosphotransferase system; ST, sequence type.

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Two supplementary tables and three supplementary figures are available with the online version of this article.

(*Morone americana*) and hybrid striped bass (*Morone saxatilis* × *Morone chrysops*) in the USA; and cobia (*Rachycentron canadum*), golden pompano (*Trachinotus ovatus*) and yellowtail (*Seriola quinqueradiata*) in Asia. Importantly, larvae and juveniles are more susceptible to disease than adults, and acute infection may cause 90–100% mortality in sea bream juveniles [4].

The pathogenicity of *P. damselae* subsp. *piscicida* has not yet been fully elucidated, but the unique exotoxin AIP56 is well characterized and likely critical for virulence as it can trigger apoptosis in fish macrophages, neutrophils and epithelial cells [5]. This A-B metalloprotease is encoded on the plasmid pPHDP10 and neither homologues nor its phylogenetic origin are known at present. Interestingly, this plasmid is highly stable, and it is almost always present in isolates collected over the last 30 years [6]. A second virulence factor, the siderophore piscibactin, is involved in iron uptake. The system is encoded by a gene cluster on the plasmid pPHDP70, and curing of this plasmid reduces *P. damselae* subsp. *piscicida* virulence in fish [7]. Other potential virulence factors are involved in adhesion, such as the lipoprotein PDP_0080 [8], defence against oxidative stress, such as superoxide dismutase [9], and secretion of effector proteins via a type 3 secretion system [6].

Despite the apparent serological homogeneity of *P. damselae* subsp. *piscicida* [10], there is good evidence for phylogenetic diversification. For example, Japanese strains are consistently different from European strains by RAPD analyses [11]. In addition to broad-scale genetic evidence of discrete evolutionary lineages, there are also key differences in the mobilome; in the European strains, two plasmids, 20 MDa (~30 kb) and 7 MDa (~10 kb, possibly pPHDP10), have been found, with a third 50 MDa (~75 kb, possibly pPHDP70 that carries the piscibactin genes cluster) plasmid widely present, while the Japanese isolates harbour a 37 MDa (~56 kb) plasmid [12]. Strains isolated in the USA also carry plasmids of 20 and 7 MDa but, in contrast to the European and Japanese strains, no pPHDP70 plasmid has been identified to date [6]. Furthermore, the pPHDP70 plasmid of the European strains shares four sequence blocks with the Japanese *P. damselae* subsp. *damselae* ~204 kb pAQU1 plasmid [13]. Additional transferable antibiotic-resistance plasmids have also been identified and sequenced in *P. damselae* subsp. *piscicida*, and many complete plasmid assemblies have been submitted to the National Center for Biotechnology Information (NCBI) (Table S1, available with the online version of this article).

In contrast to the plasmid characterization, chromosomal data are very limited and only contig- and scaffold-level genome assemblies are available (Table 1), with only the USA isolate 91-197 from hybrid striped bass being closed (PRJDB5709). In contrast to the majority of eubacteria, in the *Vibrionaceae*, it is common to find two chromosomes of different sizes, with the larger being generally around 3–3.3 Mb and the smaller ranging from 0.8 up to 2.4 Mb [14]. Lack of complete genomic information impairs both epidemiological pathogen surveillance and the development of effective preventative strategies in aquaculture. Vaccines against intracellular pathogens like *P. damselae* subsp. *piscicida* need to elicit not only humoral but also cellular immune responses and, to date, DNA vaccines are considered a relatively safe option for obtaining cellular immune protection in fish, which is not offered by killed whole-cell vaccines traditionally used in aquaculture [15]. This strategy has been successfully employed in experimental vaccines against *P. damselae* subsp. *piscicida* in flounder in Japan [16]. However, identification of candidate DNA vaccine immunogens relies heavily on the availability of accurate genetic information. Such information is limited for *P. damselae* subsp. *piscicida* isolates, reflecting the cost and complexity of sequencing and closing mobilome-rich multi-chromosome genomes in isolates that also carry several plasmids that are critical for virulence. For long-read sequencing appropriate for accurate genome closure, PacBio single molecule real-time sequencing technology is well tested but expensive, and samples require time-consuming preparation before submission to a dedicated facility for sequencing. In contrast, robust and cost-effective Illumina short-read platforms provide high-accuracy base calling, but are unsuited to closing complex genomes. As a consequence, most current *P. damselae* subsp. *piscicida* assemblies comprise very high numbers of short contigs (Table 1).

Recently, photobacteriosis outbreaks have been recorded in yellowtail kingfish (*Seriola lalandi*) farms in South Australia and Western Australia. Here, we sequenced and assembled complete genomes for two Australian isolates of *P. damselae* subsp. *piscicida* using Oxford Nanopore Technology (ONT) sequencing. Nanopore sequencing provides low-cost long-read data suited to closure of complex genomes including plasmids [17]. Moreover, the preparation of the libraries is quick and easy, and the samples can be sequenced in a short time directly in the same laboratory, without the need for

Impact Statement

Photobacterium damselae subsp. *piscicida* causes severe mortality in aquaculture worldwide. Genome closure is complicated by the presence of two chromosomes and several plasmids, a common feature in *Vibrionaceae*. For the first time, to our knowledge, we provide here two high quality, manually curated *P. damselae* subsp. *piscicida* genomes from Australia, which were sequenced and assembled comparing two different long-read sequencing technologies, Oxford Nanopore MinION and PacBio, followed by error correction with Illumina short-read data. These assemblies allowed the in-depth analysis of the pathogen mobilome and the identification of novel chromosomal and plasmid-encoded putative virulence factors. Finally, we provide the first, global phylogenetic structure of *P. damselae* subsp. *piscicida* strains, showing that the European isolates form a cluster that is genetically distant from those from Asia, America and Australia.

Table 1. *P. damselae* subsp. *piscicida* assemblies and related information available to date from the NCBI

Strain	Date (M/YY), level	GenBank acc. no.	Assembly method	Coverage (×)	Technology	Host	Isolation (year/location)
PP3	4/19, contig	GCA_004683985.1	Megahit v. 1.0	100	Illumina MiSeq	*Seriola quinqueradiata*	NA, Japan
DI21*	7/15, scaffold	GCA_000300355.4	Newbler v. 2.6	60	454	*Sparus aurata*	1990, Spain
91-197	5/17, complete genome	GCA_002356235.1	HGAP v. 3	343.5	PacBio RSII	*Morone chrysops* × *Morone saxatilis*	1991, USA
MT1415	5/19, contig	GCA_005048985.1	Megahit v. 1.0	200	Illumina MiSeq	*Dicentrarchus labrax*	1993, Italy
OT-51443	4/17, contig	GCA_002157645.1	HGAP v. 3	343.5	PacBio RS II	*Seriola quinqueradiata*	1995, Japan
ATCC 29688	3/18, contig	GCA_003026775.1	SPAdes v. 3.11.0	88	Illumina MiSeq	*Seriola quinqueradiata*	1995, Japan
ATCC 29689	3/18, contig	GCA_003026245.1	SPAdes v. 3.11.0	70	Illumina MiSeq	*Seriola quinqueradiata*	1995, Japan
SNW-8.1	4/19, contig	GCA_004684005.1	SPAdes v. 3.6	50	Illumina MiSeq	*Salmo salar*	2014, Spain
L091106-03H	1/17, scaffold	GCA_001715165.2	CABOT vs JAN-2015; SOAPdenovo vs JAN-2015; SSPACE vs JAN-2015	50	454	*Solea senegalensis*	2017, Spain

NA: Isolation date not available

*Strain not included in the phylogenetic analyses. Only pPHDP70 assembly (NZ_KP100338.1) was included.

external facilities. As ONT is currently inferior to Illumina in terms of base-calling accuracy, closed long-read draft assemblies were polished with Illumina MiSeq short reads. For both strains, polished Nanopore assemblies were compared to polished PacBio assemblies and, when possible, mismatches were manually resolved using mapped Illumina reads. We closed both chromosomes and five plasmids belonging to the Australia outbreak isolates and manually curated the annotations providing high-quality reference sequences for *P. damselae* subsp. *piscicida*. Using these data, we then explored the evolutionary relationships between Australian and overseas isolates, with a focus on the role of plasmids and other mobile genetic elements (MGEs).

METHODS

Bacterial strains

P. damselae subsp. *piscicida* strains QMA0505 (AS-16-0540 #1) and QMA0506 (AS-16-0555 #7) were isolated at the Department of Primary Industries and Regional Development (DPIRD) in Western Australia from yellowtail kingfish (*Seriola lalandi*) farms in South Australia and Western Australia, respectively, during 2016 disease outbreaks and kept frozen at −80 °C. The identity of the strains was confirmed by PCR with *P. damselae* subsp. *piscicida* specific primers [18], and by phenotypic assays with API20E tests (BioMérieux), following the manufacturer's instructions, conventional biochemical tests incubated for 48 h at 24 °C according to Buller (2014) [19] and MALDI-TOF MS, BioTyper and the MBT 7854 MSP library (Bruker). Micro-organisms were routinely grown on tryptone soy agar 2 (Oxoid) for 24 h at 27 °C or blood agar (BA) for 48 h at 24 °C.

DNA extraction and library preparation

DNA was extracted from fresh cultures using the *N*-cetyl-*N*,*N*,*N*-trimethylammonium bromide (CTAB) method [20] and quantified by Qubit fluorimetry (Invitrogen). An aliquot (1 ng) of DNA from each strain was then used to prepare Nextera XT libraries for MiSeq v2 Illumina sequencing (Ramaciotti Centre for Genomics, University of New South Wales, Sydney, Australia). For Nanopore sequencing, a 12-sample barcoded library was prepared using rapid sequencing kit SQK-RBK004. The library was sequenced on an Oxford Nanopore MinION device equipped with a FLO-MIN106D (R9) flow cell.

For PacBio long-read sequencing, DNA was extracted from 10 ml 48 h tryptone soy broth with 2% NaCl (TSB2) cultures of *P. damselae* subsp. *piscicida* using the Genomic-Tip G/20 kit and associated buffers (Qiagen). Aliquots (100 µl) containing 62 µg DNA for QMA0505 and 20.7 µg for QMA0506 were sent for PacBio 15–20 kb SMRT Bell library preparation and Blue Pippin size selection prior to sequencing on four SMRT cells using the P6-C4 chemistry on the RSII platform (Ramaciotti Centre, University of New South Wales, Sydney, Australia). In each case, DNA was quantified by Qubit fluorimetry (Invitrogen), and the quality and purity checked by gel electrophoresis, followed by amplification and sequencing of

the almost complete 16S rRNA gene using universal 27F and 1492R primers [21]. Derived 16S sequences were assembled in Sequencher v5.2.2 (Gene Codes), and then analysed by BLASTN to confirm species identity before genome sequencing.

Nanopore–Illumina hybrid genome assembly and annotation

The quality of short Illumina reads was assessed using FastQC v.0.11.8 [22] and then trimmed using FastP v.0.20.0 [23], using the flags --trim_front=15 and --trim_tail=5. Long Nanopore reads were demultiplexed and trimmed with Porechop v0.2.3_seqan2.1.1 [24]. Draft genomes were assembled with long reads using Flye v.2.6 [25], using arguments --nano-raw, --genome-size=4.8 m and --plasmids. For assembly of chromosomes 1 and 2 from QMA0505 the default parameters were used, while for plasmids the --min-overlap was set at 8000. For QMA0506, --min-overlap was set at 7000, while only for plasmid 5 (pPHDP10) --min-overlap was set to 6200. The quality of draft assemblies was assessed by inspecting the assembly graph in Bandage v.0.8.1 [26]. These draft backbones were then polished to a coverage depth of ~60–70× with Illumina short reads using Unicycler_polish in Unicycler v.0.4.8 [27]. The bacterial origin of replication (*dnaA*) was set at position 0 with the fixstart command in Circlator [28]. The genome QMA0505 was annotated using Prokka [29] and further curated manually by BLASTX of all hypothetical proteins against the NCBI database and accepting results with identity and coverage greater or equal to 93%. The curated annotated genome for QMA0505 was considered a reference and used to construct a database in Prokka for the annotations of QMA0506. Finally, operon predictions were performed with FgenesB tool in SoftBerry, using *Vibrio parahaemolyticus* RIMD 2210633 as a reference (http://www.softberry.com/berry.phtml).

Validation and curation of Nanopore chromosome assemblies

The validation of chromosomal assemblies using Nanopore reads was performed on strain QMA0505. Chromosomes 1 and 2 were assembled from PacBio sequencing reads from two SMRT cells using SMRT Portal version 2.3.0 (Pacific Biosciences) and the hierarchical genome assembly process (HGAP3), with default parameters for a coverage of 30-fold and based on estimated genome size 4.8 Mb, then polished with Quiver [30]. Following draft assembly, contigs were analysed for terminal overlaps with Circlator. Overlapping regions were trimmed and the resulting circular chromosomes reoriented and cut such that the origin of replication was at the start. Nanopore assemblies using Flye and PacBio HGAP3 assemblies were then polished using the Illumina short reads with Unicycler_polish in Unicycler v.0.4.8 [27]. Polished assemblies were aligned in Geneious Prime v.2020.1.2 and SNPs detected using default settings. Short Illumina reads were mapped onto the polished Nanopore assemblies with medium sensitivity. For each SNP in the PacBio/Nanopore alignment that was not a tRNA nor rRNA coding region, the consensus in Illumina reads was investigated manually. When at least 90% of Illumina data indicated the same base, the SNP was corrected accordingly. tRNA and rRNA coding sequences were not included in this process, because the high similarity between multiple copies of these genes made the read-mapping step inconsistent and coverage uneven. Once manually polished, the chromosomal assemblies of QMA0505 were annotated again in Prokka using the previously manually curated annotations as a database. All the locus tags refer to QMA0505, but the loci are also present in QMA0506, unless otherwise stated.

Curation of the Australian plasmids

Illumina reads of both QMA0505 and QMA0506 were mapped on the polished Nanopore plasmid assemblies. For each SNP found between the pairs of plasmids aligned with MAFFT in Geneious Prime v.2020.1.2, the consensus in the mapped Illumina reads was examined. When at least 90% of Illumina data indicated the same base, the SNP was corrected accordingly.

Phylogenetic analyses

For phylogenetic reconstruction, available *P. damselae* subsp. *piscicida* genomes were retrieved from the NCBI (Table 1). For *P. damselae* isolates with no indication of the subspecies, a BLAST search for the *ureC* gene was performed and its absence was accepted as evidence that the subspecies was *P. damselae* subsp. *piscicida* [18]. For draft assemblies at the contig level, sequences were mapped to the reference chromosomes QMA0505 and consensus generated in Geneious Prime v.2020.1.1. Reordered genomes were aligned with Parsnp in the HarvestTools suite v.1.1.2 [31], and regions of recombination in the resulting core-genome alignments were then inferred and removed using Gubbins v.2.4.1 [32]. Phylogenies based on 30484 non-recombinant core-genome SNPs were reconstructed by maximum-likelihood method using RAxML v.8.2.11 in Geneious Prime v.2020.1.1, with the general time reversible nucleotide substitution model (GTR) and bootstrap of 1000 iterations. As the dataset contained only variable sites, an ascertainment bias correction was included with the flags --m ASC_GTRGAMMA and --asc-corr=felsenstein with the number of invariant sites specified. As outgroup, the complete curated genome sequence of *P. damselae* subsp. *damselae* strain QMA0509 (AS-15-3942 #7) from the laboratory collection was included to guide the rooting of the phylogeny. The tree was annotated with metadata using Evolview v.2 [33]. Finally, for strain OT-51443, the contigs deposited in the NCBI database were reordered with Mauve Contig Mover, using QMA0505 as a reference.

RESULTS

Case diagnosis and observations

The MALDI-TOF scores for QMA0505 (AS-16-0540 #1) and QMA0506 (AS-16-0555 #7) were significant to the species level at 2.0 and 2.1, respectively. Both isolates were positive for arginine dihydrolase, fermentation of 1% glucose and 1% mannose, and growth on agar containing 3% NaCl. The result

Table 2. Australian *P. damselae* subsp. *piscicida* strains (a) QMA0505 and (b) QMA0506 genome assembly data

Comparison between assemblies, all polished with the same Illumina short reads and with different technologies and assemblers. Increasing quantities of DNA sequenced by PacBio (1, 12 and 15 nM) and Nanopore long reads were assembled into a draft using Flye, while the 1 nM concentration of PacBio long-reads was also assembled with HGAP3. Nanopore assemblies were further manually curated. Assembled chromosome and plasmid sizes are in base-pairs

Long-read assemby method	Chr. 1	Chr. 2	Plas. 1	Plas. 2 pPHDP70	Plas. 3	Plas. 4	Plas. 5 pPHDP10
(a) QMA0505							
PacBio 1 nM (Flye)	3230005	1136179	136327	81408	37525	23188	–
PacBio 12 nM (Flye)	3230071	1136170	136327	81412	36448	23189	–
PacBio 15 nM (Flye)	3230074	1136181	136328	81412	36447	23189	–
PacBio 1 nM (HGAP3)	3230051	1135725	136328	81412	75215	–	–
Nanopore (not curated, Flye)	3230449	1136362	136317	81403	36456	23188	7991
Nanopore (curated, Flye)	3230241	1136236	136326	81411	36456	23190	7991
(b) QMA0506							
PacBio 1 nM (Flye)	3171776	1133618	136327	81411	38781	23189	–
PacBio 12 nM (Flye)	3171786	1133152	136328	81412	38782	23189	–
PacBio 15 nM (Flye)	3171764	1133550	136328	81412	39858	23190	–
PacBio 1 nM (HGAP3)	3198996	1135525	161014	109463	–	–	–
Nanopore (Flye, not curated)	3230247	1134456	136333	81431	36447	23192	7987
Nanopore (Flye, curated)	3230122	1134735	136321	81418	36447	23189	7991

for methyl red was weak. There was no growth on thiosulphate-citrate-bile salts-sucrose (TCBS) agar after 48 h incubation at 24 °C. Both isolates were negative for the following: β-haemolysis on BA containing 5% horse red blood cells, swarming on BA and marine salt agar (MSA) containing 5% horse red blood cells, ornithine decarboxylase, lysine decarboxylase, nitrate reduction, indole, urease, Voges–Proskaüer reaction (acetoin production), hydrolysis of aesculin, gelatin and DNA, ONPG (*o*-nitrophenyl-β-galactoside), growth on MacConkey agar, and negative fermentation of L-arabinose, D-glucose, inositol, lactose, maltose, mannitol, mannose, salicin, sorbitol, sucrose, trehalose and xylose. The isolates were sensitive to vibriostatic agent 0/129 discs of 10 and 150 µg (Oxoid). The API20E bionumber was 2004004. Both isolates were negative by PCR for the 267 bp *ureC* amplicon, consistent with *P. damselae* subsp. *piscicida* identification.

Mortalities in farmed yellowtail kingfish during outbreaks in Australia have been relatively low compared to those reported in Japan and Europe, with per cage mortality rates not exceeding 10% and generally around ~1–2% across an entire farm lease. It is also worth noting that outbreaks, and associated isolations of *P. damselae* subsp. *piscicida*, have not occurred every season in Australia, even when leases have remained stocked throughout with no fallow period.

Genome assemblies

Oxford Nanopore reads permitted assembly and closure of draft genomes of QMA0505 and QMA0506 using Flye, with a coverage of at least 90× and 26×, respectively. Additionally, short Illumina reads used to polish the drafts provided an additional 107× and 361× coverage, respectively. For each strain, both chromosomes and plasmids were fully assembled and closed (Table 2a, b). Once polished, the sequences were annotated and manually curated, resulting in complete, high-quality genomes. For QMA0505, 589 and 240 operons were predicted in chromosome 1 and chromosome 2, respectively, with the largest on chromosome 1 (~12 kbp) encoding a type II secretion system between IC627_14670 and IC627_14735, and the largest on chromosome 2 (~15 kbp) encoding a phosphotransferase system (PTS) predicted to be a mannose/fructose/sorbose/*N*-acetylgalactosamine transport system from IC627_17835 to IC627_17895.

Validation and curation of Nanopore chromosome assemblies

In QMA0505, the alignment of polished PacBio and polished Nanopore assemblies revealed 410 SNPs outside the rRNA and tRNA coding sequences, corresponding to 0.013% in the chromosome 1 assembly. Of these, 45 SNPs could not be resolved due to insufficient Illumina read coverage and 57 were ambiguous. In chromosome 1, polished-PacBio assembly bases were consistent with Illumina data in 239 of the cases (214 deletions, 24 insertions and 1 mismatch compared to the Nanopore assembly), while for Nanopore assemblies it occurred in 69 SNPs (14 deletions, 43 insertions, 12 mismatches). In chromosome 2 of QMA0505, 359 SNPs were found (0.0326%), with 10.2% of unidentified bases and 17.5% of bases lacking Illumina coverage. PacBio base

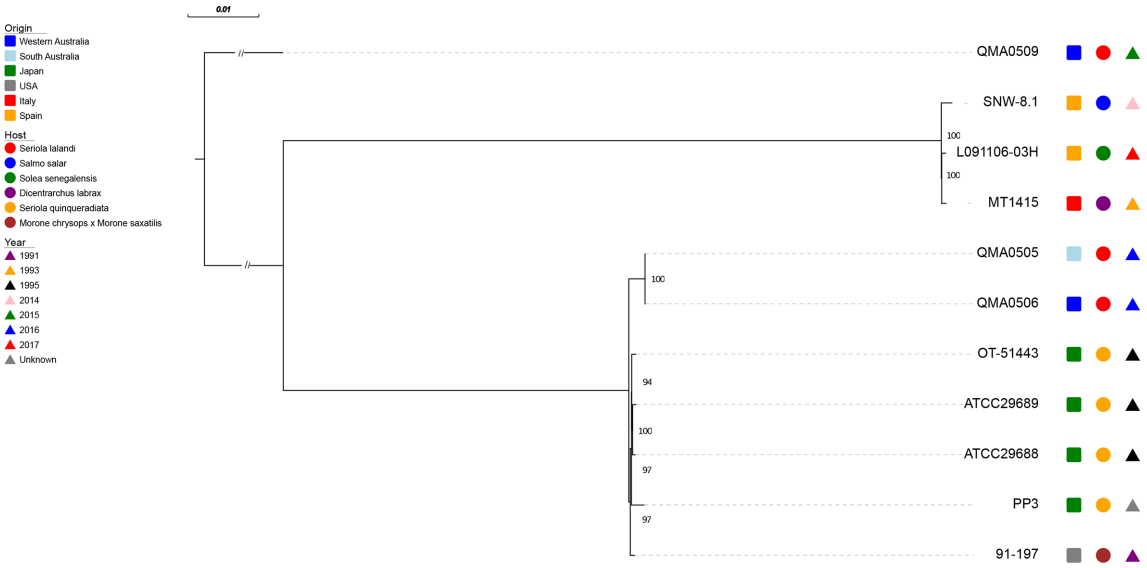

Fig. 1. Phylogram of *P. damselae* subsp. *piscicida* chromosomes, with *P. damselae* subsp. *damselae* QMA0509 as an outgroup. Bootstrap value support is indicated next to the nodes. Scale bar indicates nucleotide substitutions per site.

calling was consistent with Illumina data in 40.1% of the SNPs (with 141 deletions and 3 mismatches), while the Nanopore consensus was consistent with Illumina reads in 32.2% of the cases. For chromosome 1 of QMA0506, the number of SNPs outside rRNA and tRNA coding sequences was 882, but only 184 (20.9%) were resolved with Illumina reads. Of these, Illumina reads were consistent in 123 deletions, 3 mismatches and 7 insertions with PacBio assemblies, and in 4 insertions, 3 deletions and 44 mismatches with Nanopore. The remaining 698 SNPs comprised 501 ambiguous bases and 197 were not covered by Illumina reads. For chromosome 2 of QMA0506, 372 SNPs were detected between the PacBio and Nanopore assembly alignments, with 277 uncertain bases and 5 without Illumina coverage. Polished PacBio assemblies were consistent with Illumina base calling in 14.5% of the cases (48 deletions, 4 insertions and 2 mismatches), while for Nanopore assemblies it occurred in 9.7% of the SNPs (2 deletions, 4 insertions, 30 mismatches). The regions that were not resolved by manual curation with Illumina were corrected using PCR and Sanger sequencing.

Curation of the Australian plasmids

MAFFT alignment of QMA0505 and QMA0506 pPHDP10 plasmids allowed the detection of 12 putative SNPs. However, read-mapping of Illumina reads onto the polished Nanopore assemblies and manually calling consensus resulted in identical final sequences for the plasmid from both isolates. For the other plasmids, 46, 966, 18 and 22 putative SNPs were identified between plasmids 1, 2, 3 and 4, respectively. Overall, ~95% of the SNPs were resolved, with the majority being located in homopolymeric regions.

Phylogenetic reconstruction

The phylogenetic analyses of non-recombinant sites of the core genome common to all isolates revealed that Australian *P. damselae* subsp. *piscicida* strains are distinct from, but more closely related to, the Japanese and American strains, while the European isolates cluster in a separate clade (Fig. 1). The four closely related strains within the Australia/Asia/USA clade, QMA0505, QMA0506, 91-197 and OT-51443, contain 145 and 72 non-recombinant SNPs in chromosomes 1 and 2, respectively, with the highest proportion of recombination occurring in chromosome 2.

Insertion sequence elements (ISEs), large rearrangements and pseudogenes in *P. damselae* subsp. *piscicida* chromosomes

Both chromosomes belonging to the Australian isolates are larger than the almost complete genome of the Japanese OT-51443 (3 132 149 and 1 010 592 bp) and the fully closed American 91-197 (3 172 118 and 1 054 589 bp). The comparison between the genomes reveals that insertion sequences greatly influence the architecture both of the chromosomes and plasmids, resulting in large-scale rearrangements and functional gene content (Fig. 2a, b).

In 91-197 and OT-51443, various types of ISE account for 1.463 and 1.485% of the coding sequences in chromosome 1, and 14.372 and 14.416% of the coding sequences in chromosome 2, respectively (Table 3). The Australian strains show a similar difference between the two chromosomes, but there is more substantial proliferation of ISEs. In chromosome 1 and 2, respectively, of QMA0505, 9.219 and 22.772% of coding sequences are contained in putative ISEs, representing 6.229

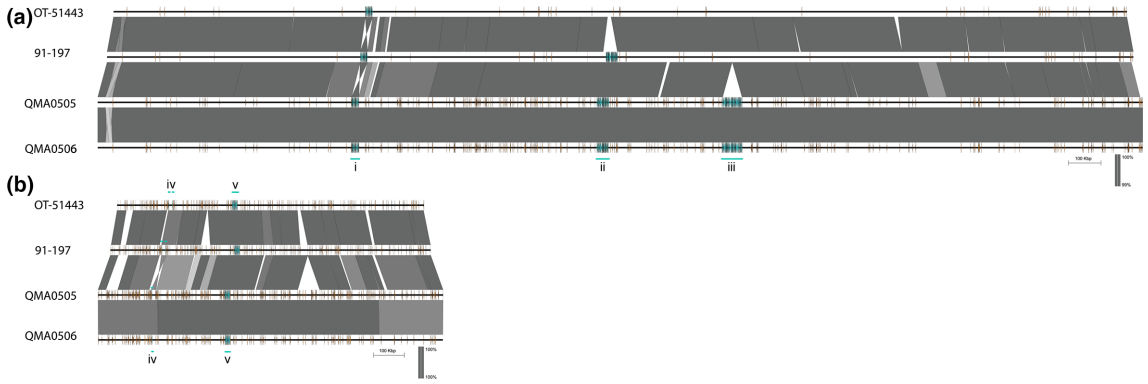

Fig. 2. Comparative structures of chromosomes 1 (a) and 2 (b) belonging to strains OT-51443 (Japan), 91-197 (USA), QMA0505 (Australia) and QMA0506 (Australia). Insertion sequences are represented as orange arrows whose orientation reflects the position of the gene on either the forward or reverse strand. Cyan arrows and bars represent regions of interest flanked by insertion sequence (refer to the text).

and 17.623% of the chromosome size. In QMA0506, ISEs are 9.245 and 23.019% of the coding sequences, and 6.228 and 17.549% of the chromosomes size. Interestingly, the IS*1*-like element IS*Pda1* family transposases that are the most abundant ISE class in both Australian strains and almost equally represented on both chromosomes were not identified in 91-197 and OT-51443 (Table 3).

As a result of the proliferation of ISEs, the four genomes have undergone a process of general pseudogenization. In chromosome 1 of 91-197, 9.371% of genes have lost their function, while in chromosome 2, the frequency increased to 13.503%. However, the Australian isolates show again a more extensive presence of pseudogenes; in chromosome 1 of QMA0505, 342 out of 3135 (10.909%) genes were inactivated, and in chromosome 2, 246 of 1212 (20.297%) were inactivated.

With respect to the large-scale genetic rearrangements mediated by ISEs, inversion of 22 kbp on chromosome 1 (Fig. 2a, region i) and 15 kbp on chromosome 2 (Fig. 2b, region iv) flanked by ISEs were identified in all four isolates. A ~34 kbp region present in chromosome 1 of both Australian and USA strains is absent in the Japanese strain (IC627_07795 to IC627_07640) (Fig. 2a, region ii). Here, genes encode a set of essential proteins, including a type VI secretion system protein VgrG (IC627_07640), tellurite-resistance protein TerB (IC627_07675), zinc transport protein ZntB (IC627_07680), miniconductance mechanosensitive channel (IC627_07745) and YdcF family protein (IC627_07795). A second, ~60 kbp region of chromosome 1, is unique to the two Australian strains (Fig. 2a, region iii). Located between two tRNA-(ms[2]io[6]A)-hydroxylases (IC627_09455 and IC627_09735), 56 putative coding sequences share highest BLAST identity (94.07%) with *Providencia stuartii* strain AR_0026 (CP026704.1) but only 8% coverage (Table S2). Finally, the largest predicted operon in chromosome 2, the PTS, is also flanked by insertion sequences (IC627_17835 to IC627_17895) (Fig. 2b, region v)

P. damselae subsp. *piscicida* plasmids and their insertion sequences

In both Australian strains, the plasmid pPHDP10 comprises 7991 bp with a G+C content of 39.8 mol% and contains seven coding sequences: a DNA invertase *hin*, one hypothetical protein, a chorismate mutase, *aip56* and three insertion sequences. The Italian MT1415 strain reference plasmid (DQ069059.1) is 9631 bp and the Spanish L091106-03H is 9736 bp, exceeding the Australian plasmid by more than 1.6 kb and 1.7 kb, respectively. A second distinguishing feature of the Australian and the European pPHDP10 is the inversion that has occurred in the region of *aip56* and the chorismate mutase. The insertion sequence content in the pPHDP10 plasmid is variable amongst strains. In the Australian strains, the oppositely oriented insertion sequences are downstream of the aip56/chorismate mutase region, while the Spanish strain contains only one insertion element further downstream. In contrast, the Italian strain MT1415 contains multiple, mostly fragmented, insertion elements on both sides of the region. Finally, MT1415 also has a DUF4158 domain-containing protein instead of a complete insertion sequence, and this insertion sequence is absent in L091106-03H (Fig. 3).

In addition to pPHDP10, four other plasmids with highly similar features, but differing in ISEs, were found in both Australian strains (Table 2b). The plasmids bear some similarities with large plasmids found in some overseas isolates.

Both QMA0505 and QMA0506 have the pPHDP70 plasmid, but the Australian strains carry a form that is larger than those found in the Spanish DI21 and the L091106-03H strains. The plasmid is 81411 bp in QMA0505 and 81418 bp in QMA0506, while in DI21 it is 68686 bp and in L091106-03H it is 68595 bp (Fig. 4). There are also several major rearrangements. Firstly, the pPHDP70 plasmids from Spanish isolates lack a ~5 kbp region in which a fimbrial/pilus periplasmic chaperone and a hypothetical protein are present in the Australian strains (IC628_22550 to IC628_22530 in QMA0506 and IC627_22570 to IC627_22595 in QMA0505). This 5 kbp

Table 3. Features table of chromosomes 1 and 2 of the Australian strains (QMA0505 and QMA0506), 91-197 and OT-51443

Feature	QMA0505		QMA0506		91-197		OT-51443	
	Chr. 1	**Chr. 2**	**Chr. 1**	**Chr. 2**	**Chr. 1**	**Chr. 2**	**Chr. 1**	**Chr. 2**
Size (base-pairs)	3230241	1136236	3230122	1134735	3172118	1054589	3132149	1010592
G+C content (mol%)	41.4	39.3	41.4	39.3	41.6	39.3	41.6	39.3
ISEs								
IS*1* family transposase	53	30	55	31	2	152	2	144
IS*1*-like element IS*Pda1* family transposase	179	166	177	165	0	0	0	0
IS*3* family transposase	13	22	13	21	0	0	0	0
IS*30* family transposase	1	0	1	0	0	0	0	0
IS*6* family transposase	1	0	1	0	0	0	0	0
IS*91* family transposase	27	39	25	37	0	0	0	0
IS*91*-like element IS*Pda2* family transposase	7	14	9	17	0	0	0	0
ISNCY family transposase	1	0	1	0	0	0	0	0
Transposase	5	1	5	1	8	10	8	10
Transposase zinc-binding domain-containing protein	2	4	2	4	0	0	0	0
Putative transposase	0	0	0	0	36	53	35	52
Transposase DDE domain protein	0	0	0	0	0	0	1	0
Transposon Tn*7* transposition protein TnsA	0	0	0	0	1	0	1	0
ISE total length (base-pairs)	201226	200241	201183	199136	35800	78022	35638	78130
% IS/CDS	9.219	22.772	9.245	23.019	1.463	14.372	1.485	14.416
% IS/size	6.229	17.623	6.228	17.549	1.129	7.398	1.138	7.731
Pseudogenes								
No. of pseudogenes	342	246	326	212	301	202	NA	NA
%Pseudogenes/ CDS	10.909	20.297	10.429	17.681	9.371	13.503	NA	NA

CDS, Coding sequence; IS, insertion sequence; NA, Not annotated

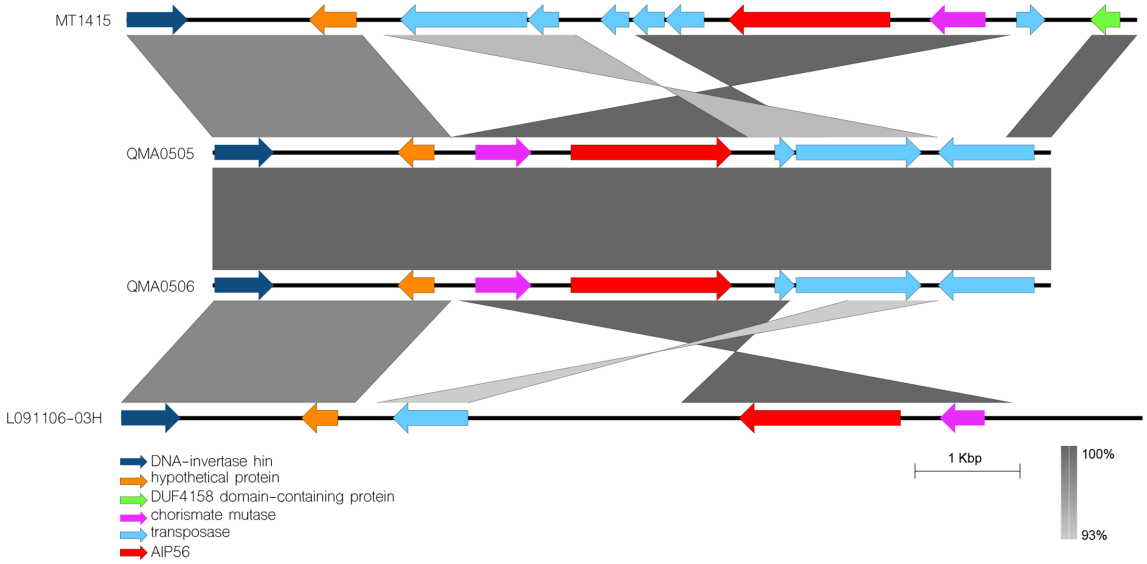

Fig. 3. Comparative analysis of pPHDP10 plasmids from *P. damselae* subsp. *piscicida* strains L091106-03H, MT1415, QMA0505 and QMA0506.

region has a 47% coverage and 99.59% identity by BLAST with IS*91*-like element IS*Pda2* family transposase from *Vibrio toranzoniae* (WP_161680561.1). The plasmids also have an inversion between IC627_22690 and IC627_22665, where a ParA family protein (IC627_22695), a ParB/RepB/SpoOJ family partition protein (IC627_22690), a helix-turn-helix domain containing protein (IC627_22685) and the transposases flanking them (IC627_22680 and IC627_22705) are present. A ~3 kbp region containing only a RhsD family protein is present in L091106-03H, while in the other strains this region also includes insertion sequences upstream and downstream of the gene (IC627_22710, IC627_22720, IC628_22600 and IC628_22610). In the Australian strains, a further rearrangement at position 61 783 (IC627_22725) to position 69434 (IC627_22760) includes an additional insertion sequence, hypothetical protein and a conjugal transfer protein TraF. Finally, a DNA replication terminus site-binding protein Tus and a DNA-binding protein coding sequence that

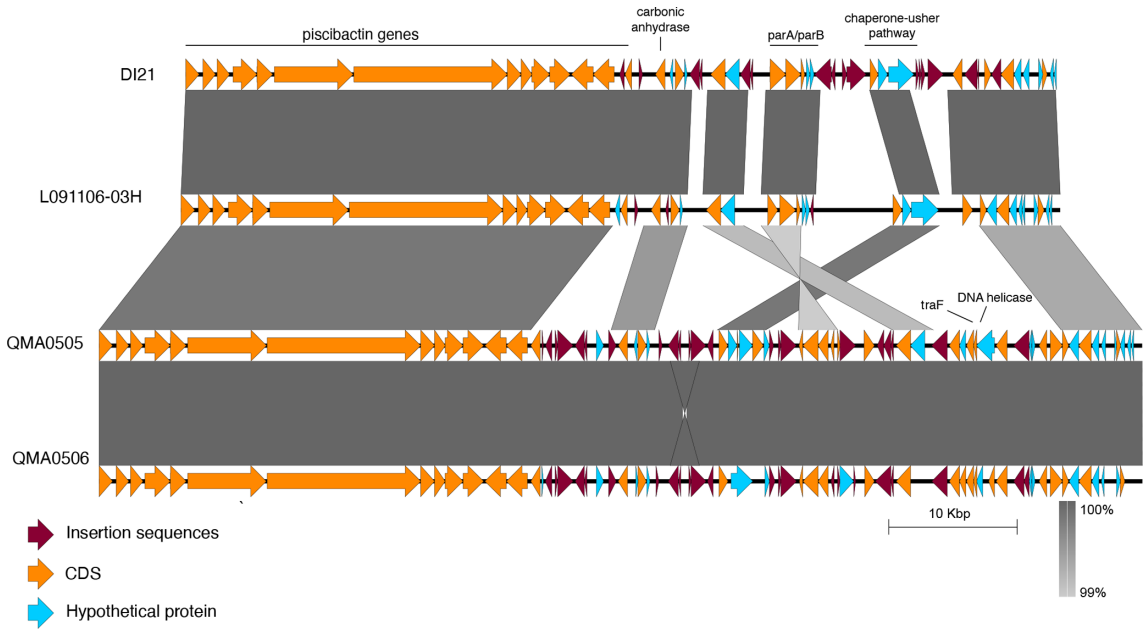

Fig. 4. Comparative analysis of pPHDP70 plasmids from *P. damselae* subsp. *piscicida* strains DI21, L091106-03H, QMA0505 and QMA0506.

are both similar to those found in *Vibrio parahaemolyticus* are present in the Australian strains (99% coverage and 99.31% identity, and 99% coverage and 98.73% identity, respectively).

The largest plasmids in QMA0505 and QMA0506 are 136326 and 136321 bp, respectively. Among the available *P. damselae* subsp. *piscicida* genomes, only the Japanese strain OT-51443 carries a plasmid of a similar size, 145279 bp. All three plasmids encode type III and IV secretion systems, a recombinase family protein, and chaperone proteins Sycd/LcrH family and VscY (Fig. S1).

A plasmid of ~36 kb is highly similar in both QMA0505 and QMA0506, with 17 SNPs out of 18 SNPs in coding regions being as insertion sequences (Fig. S2). The most similar plasmid reported to date is a ~40 kb plasmid found in the Japanese strain OT-51443. In all three plasmids, a type II toxin–antitoxin system HigB family toxin, a type III secretion protein and type IV secretion system proteins are encoded. Moreover, a Hok/Gef family protein and the partition proteins belonging to the ParAB family are present.

There is also an approximately 23 kb plasmid that is similar to one found in the Japanese strain OT-51443. The latter is ~25 kb and the difference in size is accounted for by the presence two additional ISEs (Fig. S3).

Virulence repertoire of Australian *P. damselae* subsp. *piscicida* strains

Both Australian strains carry the pPHDP10 plasmid containing the *aip56* gene (NC_013775.1), encoding the best-characterized virulence factor in *P. damselae* subsp. *piscicida*, as well as plasmid pPHDP70 bearing the piscibactin gene cluster. PDP_0080 and p55, potential virulence factors identified to date in overseas strains, were also located in both Australian isolates. While the PDP_0080 gene (HQ599846.1) was annotated on chromosome 1 as a type II secretion system pilot lipoprotein GspS-beta (IC627_08880 and IC628_08855), the p55 genes in QMA0505 (IC627_21025) and in QMA0506 (IC628_18250) were annotated on chromosome 2 as SH3 domain-containing proteins. Furthermore, chromosomes 1 of both isolates contain a transfer-messenger RNA gene *ssrA* (IC627_12340 and IC628_12300) preceded by the SsrA-binding protein SmpB gene.

DISCUSSION

High-quality and cost-effective genome sequences of Australian *P. damselae* subsp. *piscicida*

The combination of long and short reads, obtained from Oxford Nanopore MinION and Illumina technologies, respectively, were at least as accurate as PacBio assemblies polished with the same Illumina data (Table 2a, b), and even more efficient in resolving plasmid sequences. Although PacBio reads gave more accurate sequences at a single nucleotide level, we were unable to complete the assembly of the plasmids present in the Australian strains from these reads, although this may reflect the efficiency of the Blue Pippin size selection prior to the sequencing run (Table 2a). Comparing costs and resources needed for PacBio sequencing, Nanopore has significant time and cost advantages in addition to the ability to complete all steps in a small laboratory, with the only disadvantage being slightly less accurate resolution of homopolymers.

Information from *P. damselae* subsp. *piscicida* genomes supports a vector-borne *P. damselae* subsp. *piscicida* life-cycle hypothesis

P. damselae subsp. *piscicida* infections in seabass in Egypt coincided with *Caligus elongatus* (sea lice) infestations and the latter agent was hypothesized to serve as a disease transmission vector for *P. damselae* subsp. *piscicida* [34]. Caligid sea lice are also reported on yellowtail in Australia [35]. Indeed, genomes of *P. damselae* subsp. *piscicida* support the idea of an arthropod disease vector. The absence of the urease gene, a feature distinguishing *P. damselae* subsp. *piscicida* from sister subspecies *P. damselae* subsp. *damselae*, is essential for infection of flea vectors by *Yersinia pestis* [36, 37]. The *cry8A* gene encoding pesticidal protein from *Bacillus thuringiensis*, a pathogen characterized by continuous mobilome-mediated expansion of invertebrate host range [38], is found in *P. damselae* subsp. *piscicida* chromosome 1 (IC627_12360). Moreover, AIP56, the virulence factor unique to *P. damselae* subsp. *piscicida*, has been intensively studied, in terms of its mechanism of action [39, 40], cellular pathways [41] and other auxiliary molecules involved in host intoxication [42]. However, little is known about its origin. The C-terminal portion of the toxin is similar to a hypothetical protein found in the bacteriophage APSE-2 [5]. This virus infects *Hamiltonella defensa*, a bacterial endosymbiont of some aphids that protects them from parasitoids [43], but APSE is also reported in *Arsenophonus*, another endosymbiont of arthropods but with a wider host range compared with '*H. defensa*' [44]. It is worth noting that yellowtail kingfish farmed in Australia have occasional infestations with caligid sea lice and that biting isopods are also observed from time to time.

Support for vector-borne transmission is indirectly provided in the phylogenetic reconstruction of non-recombinant sites that clearly groups Australian strains together with the Japanese and American (Atlantic) strains (Fig. 1). No major surface-water exchanges between the Northern and Southern hemispheres can explain extensive gene flow between Japan, USA and Australia [45], and yellowtail (*Seriola quinqueradiata* in the Northern hemisphere and *Seriola lalandi* in Australia), a major host of *P. damselae* subsp. *piscicida*, are different species belonging to different populations [46]. They do not undergo exchanges nor migrations, probably due to the hostile environment they encounter close to the equator. Besides, *P. damselae* subsp. *piscicida* has a very limited ability to survive outside its host [47]. Also, transmission by vector can sustain both a high-level of horizontal gene transfer and the extremely high virulence that are found in *P. damselae* subsp. *piscicida*. Taken together, these support the need for further experimental evaluation to test the hypothesis of a vector-borne lifestyle for *P. damselae* subsp. *piscicida*.

MGEs drive the independent adaptation process in *P. damselae* subsp. *piscicida*

MGEs play a fundamental role in the *P. damselae* subsp. *piscicida* genome. Australian *P. damselae* subsp. *piscicida* strains carry five plasmids, requiring a substantial allocation of resources for their replication and maintenance. Among them, pPHDP10 and pPHDP70 are key components of the *P. damselae* subsp. *piscicida* genome and define its pathogenicity by encoding AIP56 and piscibactin, respectively. The latter is located on a pathogenicity island that closely resembles the *Yersinia* high-pathogenicity island (HPI) [13]. Pathogenicity islands and plasmids are well known for their ability to severely alter the phenotype and, as a result, the virulence of several bacteria. Some of them are essential to the emergence of novel or highly pathogenic strains and species that stably maintain 'foreign' genetic elements in their genomes, such as the type III secretion system in *Salmonella* pathogenicity island 1 (SPI-1) [48]. Bacteriophages also play a substantial role in the evolution and adaptation of pathogens. The presence of a unique ~60 kbp region between two tRNA-(ms[2]io[6]A)-hydroxylases in QMA0505 may indicate that a phage integration event has taken place. Importantly, their presence has been shown to alter the host phenotype to the point of transforming it to a highly lethal pathogen. For example, in *Vibrio cholerae*, the two key virulence factors have been acquired by horizontal gene transfer and are encoded on the integrated filamentous phage CTXφ [49, 50]. Interestingly, similarly to AIP56, cholera toxin (CT) is an A-B toxin that is indispensable for *Vibrio cholerae* virulence. Also, the presence of the Shiga toxin prophage in *Shigella dysenteriae* determines the onset of severe complications, such as haemolytic uraemic syndrome in humans. Some phages modify *Shigella flexneri* Y by altering the O-antigen [51], changing the serotype and compromising acquired type-specific immunity. The genomic plasticity of *P. damselae* subsp. *piscicida* is highlighted by the considerable number of transposases that frequently flank clusters of related genes. For example, on chromosome 1 of QMA0505, the PTS genes and the flagellar machinery are both enclosed within two insertions sequences. The abundance of insertion sequences, however, also balances the acquisition of novel genes by causing a generalized loss-of-function in other coding sequences, a process that has been elucidated previously in *P. damselae* subsp. *piscicida* [52, 53]. The Australian strains clearly follow the trend of genome reduction, but there are differences in term of locations or types of insertion sequences compared with isolates from the USA, Japan and Europe, indicative of independent and localized evolution. The same process of expansion of ISEs and subsequent inactivation of multiple genes is found in many specialized pathogens, including the human pathogens, *Yersinia pestis*, *Salmonella enterica* subsp. *enterica* serovar Typhi, *Mycobacterium tuberculosis* and *Shigella flexneri* [54]. A single event appears responsible for the shift from humans to chickens when a subtype of the human *Staphylococcus aureus* sequence type (ST) 5 lineage in Poland 'jumped' to the new host and then spread worldwide via the global poultry value chain [55]. The reductive genome evolution and gene loss that are probably ongoing in *P. damselae* subsp. *piscicida* represent a major feature of the host-specialization process in other fish pathogens. The Gram-positive *Streptococcus agalactiae* includes types such as ST7 and ST283 that infect both human and fish hosts, but other types which have substantially reduced genomes, for example ST260 and ST261, can infect only fish [56, 57]. Similar differences are found between the Gram-negative *Edwardsiella tarda* and *Edwardsiella ictaluri*, with the first infecting a variety of animals including humans and the second being restricted to fish hosts, again with a smaller genome [58].

Potential additional plasmid-borne virulence factors

The Australian isolates carry the main identified and potential virulence factors of *P. damselae* subsp. *piscicida* described to date: AIP56, PDP_0080, p55, SOD and Irp1. Here, we propose two additional virulence factors that may play a fundamental role in *P. damselae* subsp. *piscicida* pathogenicity. The pPHDP10 plasmids in Australian strains are considerably smaller than those found in strains from the USA, Europe and Japan. However, the plasmid retains the chorismate mutase gene. This enzyme is part of the shikimate pathway, catalysing an equilibrium reaction that interconverts chorismate and prephenate in the production of phenylalanine and tyrosine [59–61]. In pathogenic yersinias, *Yersinia pestis*, *Yersinia pseudotuberculosis* and *Yersinia enterocolitica* biogroup 1B, the gene *irp9* can convert chorismate into salicylate, from which the siderophore yersiniabactin is then synthesized [62]. The pPHDP70 plasmid has a HPI similar to the HPI of *Yersinia*, where a cluster of *irp* genes encodes the siderophore piscibactin [13]. Furthermore, pPHDP70, along with pPHDP10, is widely present in *P. damselae* subsp. *piscicida* [6]. It is possible, therefore, that the chorismate mutase encoded on pPHDP10 plays a key role in the pathogenicity of *P. damselae* subsp. *piscicida* by providing the substrate that *irp9* on pPHDP70 can convert to salicylate for the production of piscibactin, providing an explanation for the sustained co-presence of these two plasmids. In order to accomplish this, however, the tyrosine and/or phenylalanine concentrations need to favour the reaction equilibrium in the reverse direction. *P. damselae* subsp. *piscicida* mainly targets sentinel cells, such as macrophages, that are important melanization sites involved in the immune response against pathogens and iron storage [63]. These cells try and isolate the non-self-entities by forming melanomacrophage centres that appear as black nodules in the infected area. However, the granulomas usually found in fish affected by photobacteriosis are whitish [64, 65], indicating that the melanization process is somehow impaired. Melanomacrophage centres are rich in tyrosine and other melanin precursors that may feed the chorismate mutase reverse reaction for piscibactin biosynthesis. Whether or not *P. damselae* subsp. *piscicida* can produce melanization-inhibiting proteins that both

prevent the pigmentation of granulomas and increase siderophore production needs further investigation.

A second novel potential virulence gene that we identified in *P. damselae* subsp. *piscicida* is *ssrA*, located on chromosome 1. It encodes a small stable transfer-messenger RNA that, with *smpB*, is involved in many processes, including stalled ribosome rescue, tagging abnormal proteins for degradation and regulation of phage growth [66, 67]. The *ssrA–smpB* system is ubiquitously present among bacteria and retained even in the smallest genomes, including *Mycoplasma genitalium* and *Mycoplasma pneumoniae* [67]. Notably, although *ssrA* deletion has detectable effects on *Escherichia coli*, it has much more pronounced consequences in species capable of intracellular replication. The *ssrA–smpB* system is essential for growth in *Neisseria gonorrhoeae*, *Mycoplasma genitalium* and *Mycoplasma pneumoniae* [67]. *Yersinia pseudotuberculosis* [68, 69] and *Salmonella typhimurium* [70] mutant strains deficient in SsrA–SmpB are not able to proliferate in macrophages and become avirulent in mice.

Conclusions

P. damselae subsp. *piscicida* is a highly host-adapted fish pathogen infecting a diversity of species in warm and temperate marine water in the Northern hemisphere, and the high-quality genomes of Southern hemisphere isolates are provided here, to our knowledge, for the first time. The Australian isolates are distinct from Northern hemisphere strains according to non-recombinant core-genome SNP phylogeny, but are more closely related to Japanese and American isolates than their European counterparts. Virulence of *P. damselae* subsp. *piscicida* centres around a cohort of plasmids with rather high conservation, particularly pPHDP10 and pPHDP70. We were able to resolve the complete cohort of five plasmids in Australian isolates by Nanopore sequencing, coupled with correction using Illumina reads. Both pPHDP10 and pPHDP70 plasmids are present in Australian strains, they carry the critical AIP56 and Irp virulence factors, but differ from overseas strains in type and orientation of insertion sequences. Evolution of chromosomes and other plasmids are also largely driven by transposable elements. We suggest that, as in other primary pathogens, the *ssrA–smpB* system may be critical to *P. damselae* subsp. *piscicida* virulence and intracellular survival. Also, we hypothesize that chorismate mutase, universally encoded on pPHDP10 immediately upstream of *aip56*, may supply of precursors for siderophore piscibactin biosynthesis, the genes for which are located on pPHDP70.

Funding information
This research was conducted as part of the project '2018-101 – a trivalent vaccine for sustainable yellowtail Kingfish growout' funded by the Fisheries Research and Development Corporation on behalf of the Australian Government.

Acknowledgements
We thank the Fisheries Research and Development Corporation for financial support of this project. We are particularly grateful to Cleanseas Seafood Ltd and Erica Starling of Indian Ocean Fresh for their financial support prior to the full application, which was also supported by Huon Aquaculture. We thank Mark White of Treidlia Biovet for support and Jeremy Carson for advice during the project.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Osorio CR, Collins MD, Toranzo AE, Barja JL, Romalde JL. 16S rRNA gene sequence analysis of *Photobacterium damselae* and nested PCR method for rapid detection of the causative agent of fish pasteurellosis. *Appl Environ Microbiol* 1999;65:2942–2946.
2. Terceti MS, Vences A, Matanza XM, Barca AV, Noia M *et al*. The RstAB system impacts virulence, motility, cell morphology, penicillin tolerance and prodprodprodSystem Impacts Virulence, Motility, Cell Morphology, Penicillin Tolerance and uction of type II secretion system-dependent factors in the fish and human pathpathpathSecretion System-Dependent Factors in the Fish and Human ogen *Photobacterium damselae* subsp. *damselae*. *Front Microbiol* 2019;10:897.
3. Rivas AJ, Balado M, Lemos ML, Osorio CR. Synergistic and additive effects of chromosomal and plasmid-encoded hemolysins contribute to hemolysis and virulence in *Photobacterium damselae* subsp. *damselae*. *Infect Immun* 2013;81:3287–3299.
4. Andreoni F, Magnani M. Photobacteriosis: prevention and diagnosis. *J Immunol Res* 2014;2014:793817.
5. Silva MT, Dos Santos NMS, Do Vale A. AIP56: a novel bacterial apoptogenic toxin. *Toxins* 2010;2:905–918.
6. Abushattal S, Vences A, Osorio CR. A virulence gene typing scheme for *Photobacterium damselae* subsp. *piscicida*, the causative agent of fish photobacteriosis, reveals a high prevalence of plasmid-encoded virulence factors and of type III secretion system genes. *Aquaculture* 2020;521:735057.
7. Osorio CR, Juiz-Río S, Lemos ML. A siderophore biosynthesis gene cluster from the fish pathogen *Photobacterium damselae* subsp. *piscicida* is structurally and functionally related to the *Yersinia* high-pathogenicity island. *Microbiology* 2006;152:3327–3341.
8. Andreoni F, Boiani R, Serafini G, Amagliani G, Dominici S *et al*. Isolation of a novel gene from *Photobacterium damselae* subsp. *piscicida* and analysis of the recombinant antigen as promising vaccine candidate. *Vaccine* 2013;31:820–826.
9. Barnes AC, Balebona MC, Horne MT, Ellis AE. Superoxide dismutase and catalase in *Photobacterium damselae* subsp. *piscicida* and their roles in resistance to reactive oxygen species. *Microbiology* 1999;145:483–494.
10. Magariños B, Romalde JL, Bandín I, Fouz B, Toranzo AE. Phenotypic, antigenic, and molecular characterization of *Pasteurella piscicida* strains isolated from fish. *Appl Environ Microbiol* 1992;58:3316–3322.
11. Magariños B, Toranzo AE, Barja JL, Romalde JL. Existence of two geographically-linked clonal lineages in the bacterial fish pathogen *Photobacterium damselae* subsp. *piscicida* evidenced by random amplified polymorphic DNA analysis. *Epidemiol Infect* 2000;125:213–219.
12. Romalde JL. *Photobacterium damselae* subsp. *piscicida*: an integrated view of a bacterial fish pathogen. *Int Microbiol* 2002;5:3–9.
13. Osorio CR, Rivas AJ, Balado M, Fuentes-Monteverde JC, Rodríguez J *et al*. A transmissible plasmid-borne pathogenicity island confers piscibactin biosynthesis in the fish pathogen *Photobacterium damselae* subsp. *piscicida*. *Appl Environ Microbiol* 2015;81:5867–5879.
14. Okada K, Iida T, Kita-Tsukamoto K, Honda T. Vibrios commonly possess two chromosomes. *J Bacteriol* 2005;187:752–757.
15. Lorenzen N, LaPatra SE. DNA vaccines for aquacultured fish. *Rev Sci Tech* 2005;24:201–213.
16. Kato G, Yamashita K, Kondo H, Hirono I. Protective efficacy and immune responses induced by a DNA vaccine encoding codon-optimized PPA1 against *Photobacterium damselae* subsp. piscicida in Japanese flounder. *Vaccine* 2015;33:1040–1045.

17. Tyler AD, Mataseje L, Urfano CJ, Schmidt L, Antonation KS *et al*. Evaluation of Oxford Nanopore's MinION sequencing device for microbial whole genome sequencing applications. *Sci Rep* 2018;8:10931.
18. Osorio CR, Toranzo AE, Romalde JL, Barja JL. Multiplex PCR assay for ureC and 16S rRNA genes clearly discriminates between both subspecies of *Photobacterium damselae*. *Dis Aquat Organ* 2000;40:177–183.
19. Buller NB. *Bacteria and Fungi from Fish and Other Aquatic Animals*. Wallingford: CABI; 2014.
20. Wilson K. Preparation of genomic DNA from bacteria. *Curr Protoc Mol Biol* 2001;56:2.4.1–2.4.5.
21. Amann RI, Ludwig W, Schleifer KH. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* 1995;59:143–169.
22. Andrews S. FastQC: a quality control tool for high throughput sequence data. Cambridge: Babraham Bioinformatics; 2010.
23. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
24. Wick R. Porechop: adapter trimmer for Oxford Nanopore reads; 2018. https://github.com/rrwick/Porechop
25. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 2019;37:540–546.
26. Wick RR, Schultz MB, Zobel J, Holt KE. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 2015;31:3350–3352.
27. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
28. Hunt M, Silva ND, Otto TD, Parkhill J, Keane JA *et al*. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol* 2015;16:294.
29. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
30. Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J *et al*. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 2013;10:563–569.
31. Treangen TJ, Ondov BD, Koren S, Phillippy AM. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* 2014;15:524.
32. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.
33. He Z, Zhang H, Gao S, Lercher MJ, Chen W-H *et al*. Evolview v2: an online visualization and management tool for customized and annotated phylogenetic trees. *Nucleic Acids Res* 2016;44:W236–W241.
34. Elgendy MY, Abdelsalam M, Moustafa M, Kenawy AM, Seida A. *Caligus elongatus* and *Photobacterium damselae* subsp. *piscicida* concomitant infections affecting broodstock European seabass, *Dicentrarchus labrax*, with special reference to histopathological responses. *J Aquac Res Dev* 2015;6:346.
35. Hutson KS, Ernst I, Whittington ID. Risk assessment for metazoan parasites of yellowtail kingfish *Seriola lalandi* (Perciformes: Carangidae) in South Australian sea-cage aquaculture. *Aquaculture* 2007;271:85–99.
36. Chouikha I, Hinnebusch BJ. Silencing urease: a key evolutionary step that facilitated the adaptation of *Yersinia pestis* to the flea-borne transmission route. *Proc Natl Acad Sci USA* 2014;111:18709–18714.
37. Achtman M, Zurth K, Morelli G, Torrea G, Guiyoule A *et al*. *Yersinia pestis*, the cause of plague, is a recently emerged clone of *Yersinia pseudotuberculosis*. *Proc Natl Acad Sci USA* 1999;96:14043–14048.
38. Zheng J, Gao Q, Liu L, Liu H, Wang Y *et al*. Comparative genomics of *Bacillus thuringiensis* reveals a path to specialized exploitation of multiple invertebrate hosts. *mBio* 2017;8:e00822-17.
39. do Vale A, Silva MT, dos Santos NMS, Nascimento DS, Reis-Rodrigues P *et al*. AIP56, a novel plasmid-encoded virulence factor of *Photobacterium damselae* subsp. *piscicida* with apoptogenic activity against sea bass macrophages and neutrophils. *Mol Microbiol* 2005;58:1025–1038.
40. Costa-Ramos C, do Vale A, Ludovico P, Dos Santos NMS, Silva MT. The bacterial exotoxin AIP56 induces fish macrophage and neutrophil apoptosis using mechanisms of the extrinsic and intrinsic pathways. *Fish Shellfish Immunol* 2011;30:173–181.
41. Pereira LMG, Pinto RD, Silva DS, Moreira AR, Beitzinger C *et al*. Intracellular trafficking of AIP56, an NF-κB-cleaving toxin from *Photobacterium damselae* subsp. *piscicida*. *Infect Immun* 2014;82:5270–5285.
42. Pereira C, Rodrigues IS, Pereira LMG, Lisboa J, Pinto RD *et al*. Role of AIP56 disulphide bond and its reduction by cytosolic redox systems for efficient intoxication. *Cell Microbiol* 2020;22:e13109.
43. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA. *Hamiltonella defensa*, genome evolution of protective bacterial endosymbiont from pathogenic ancestors. *Proc Natl Acad Sci USA* 2009;106:9063–9068.
44. Duron O. *Arsenophonus* insect symbionts are commonly infected with APSE, a bacteriophage involved in protective symbiosis. *FEMS Microbiol Ecol* 2014;90:184–194.
45. Wijeratne S, Pattiaratchi C, Proctor R. Estimates of surface and subsurface boundary current transport around Australia. *J Geophys Res Ocean* 2018;123:3444–3466.
46. Martinez-Takeshita N, Purcell CM, Chabot CL, Craig MT, Paterson CN *et al*. A tale of three tails: cryptic speciation in a globally distributed marine fish of the genus *Seriola*. *Copeia* 2015;103:357–368.
47. Janssen WA, Surgalla MJ. Morphology, physiology, and serology of a *Pasteurella* species pathogenic for white perch. (*Roccus americanus*). *J Bacteriol* 1968;96:1606–1610.
48. Lerminiaux NA, MacKenzie KD, Cameron ADS. *Salmonella* pathogenicity island 1 (SPI-1): the evolution and stabilization of a core genomic type three secretion system. *Microorganisms* 2020;8:576.
49. Waldor MK, Mekalanos JJ. Lysogenic conversion by a filamentous phage encoding cholera toxin. *Science* 1996;272:1910–1914.
50. Davis BM, Waldor MK, Lambowitz AM. Mobile genetic elements and bacterial pathogens. In: Craig NL, Craigie R, Gellert M, Lambowitz AM (editors). *Mobile DNA II*. Washington, DC: American Society for Microbiology; 2002. pp. 1040–1059.
51. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol* 2016;14:235–250.
52. Balado M, Benzekri H, Labella AM, Claros MG, Manchado M *et al*. Genomic analysis of the marine fish pathogen *Photobacterium damselae* subsp. *piscicida*: insertion sequences proliferation is associated with chromosomal reorganisations and rampant gene decay. *Infect Genet Evol* 2017;54:221–229.
53. Osorio CR. *Photobacterium damselae*: how horizontal gene transfer shaped two different pathogenic lifestyles in a marine bacterium. In: Villa TG, Viñas M (editors). *Horizontal Gene Transfer*. Cham: Springer; 2019. pp. 175–199.
54. Georgiades K, Raoult D. Genomes of the most dangerous epidemic bacteria have a virulence repertoire characterized by fewer genes but more toxin-antitoxin modules. *PLoS One* 2011;6:e17962.
55. Lowder BV, Guinane CM, Ben Zakour NL, Weinert LA, Conway-Morris A *et al*. Recent human-to-poultry host jump, adaptation, and pandemic spread of *Staphylococcus aureus*. *Proc Natl Acad Sci USA* 2009;106:19545–19550.
56. Kawasaki M, Delamare-Deboutteville J, Bowater RO, Walker MJ, Beatson S. Microevolution of *Streptococcus agalactiae* ST-261 from Australia indicates dissemination via imported tilapia and ongoing adaptation to marine hosts or environment. *Appl Environ Microbiol* 2018;84:e00859-18.

57. Delannoy CMJ, Crumlish M, Fontaine MC, Pollock J, Foster G *et al*. Human *Streptococcus agalactiae* strains in aquatic mammals and fish. *BMC Microbiol* 2013;13:41.
58. Yang M, Lv Y, Xiao J, Wu H, Zheng H *et al*. *Edwardsiella* comparative phylogenomics reveal the new intra/inter-species taxonomic relationships, virulence evolution and niche adaptation mechanisms. *PLoS One* 2012;7:e36987.
59. Sasso S, Ramakrishnan C, Gamper M, Hilvert D, Kast P. Characterization of the secreted chorismate mutase from the pathogen *Mycobacterium tuberculosis*. *FEBS J* 2005;272:375–389.
60. Vanholme B, Kast P, Haegeman A, Jacob J, Grunewald W *et al*. Structural and functional investigation of a secreted chorismate mutase from the plant-parasitic nematode *Heterodera schachtii* in the context of related enzymes from diverse origins. *Mol Plant Pathol* 2009;10:189–200.
61. Sträter N, Schnappauf G, Braus G, Lipscomb WN. Mechanisms of catalysis and allosteric regulation of yeast chorismate mutase from crystal structures. *Structure* 1997;5:1437–1452.
62. Pelludat C, Brem D, Heesemann J. Irp9, encoded by the high-pathogenicity island of *Yersinia enterocolitica*, is able to convert chorismate into salicylate, the precursor of the siderophore yersiniabactin. *J Bacteriol* 2003;185:5648–5653.
63. Agius C, Roberts RJ. Melano-macrophage centres and their role in fish pathology. *J Fish Dis* 2003;26:499–509.
64. Toranzo AE, Barreiro Sene´n, Casal JF, Figueras A, Magarin˜os B *et al*. Pasteurellosis in cultured gilthead seabream (*Sparus aurata*): first report in Spain. *Aquaculture* 1991;99:1–15.
65. Austin B, Austin DA. *Bacterial Fish Pathogens: Disease of Farmed and Wild Fish*. New York: Springer; 2012.
66. Lee AK, Detweiler CS, Falkow S. OmpR regulates the two-component system SsrA-SsrB in Salmonella pathogenicity island 2. *J Bacteriol* 2000;182:771–781.
67. Karzai AW, Roche ED, Sauer RT. The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat Struct Biol* 2000;7:449–455.
68. Okan NA, Mena P, Benach JL, Bliska JB, Karzai AW. The smpB-ssrA mutant of *Yersinia pestis* functions as a live attenuated vaccine to protect mice against pulmonary plague infection. *Infect Immun* 2010;78:1284–1293.
69. Okan NA, Bliska JB, Karzai AW. A role for the SmpB-SsrA system in *Yersinia pseudotuberculosis* pathogenesis. *PLoS Pathog* 2006;2:e6.
70. Julio SM, Heithoff DM, Mahan MJ. ssrA (tmRNA) plays a role in *Salmonella enterica* serovar *Typhimurium* pathogenesis. *J Bacteriol* 2000;182:1558–1563.