RESEARCH LETTER

Evidence for horizontal gene transfer, gene duplication and genetic variation as driving forces of the diversity of haemolytic phenotypes in *Photobacterium damselae* subsp. *damselae*

Amable J. Rivas[1], Alejandro M. Labella[2], Juan J. Borrego[2], Manuel L. Lemos[1] & Carlos R. Osorio[1]

[1]Departamento de Microbioloxía e Parasitoloxía, Instituto de Acuicultura, Universidade de Santiago de Compostela, Santiago de Compostela, Spain; and [2]Departamento de Microbiología, Facultad de Ciencias, Universidad de Málaga, Málaga, Spain

Correspondence: Carlos R. Osorio, Departamento de Microbioloxía e Parasitoloxía, Instituto de Acuicultura, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain.
Tel.: +34 881 816050;
fax: +34 881 816047;
e-mail: cr.osorio@usc.es

Received 21 March 2014; revised 6 May 2014; accepted 6 May 2014. Final version published online 4 June 2014.

DOI: 10.1111/1574-6968.12464

Editor: Craig Shoemaker

Keywords
haemolysin; *hlyA*; damselysin; *dly*; vibriosis; *Vibrio damsela*.

Abstract

Photobacterium damselae subsp. *damselae*, a marine bacterium that causes infections in marine animals and in humans, produces up to three different haemolysins involved in virulence, which include the pPHDD1 plasmid-encoded damselysin (Dly) and $HlyA_{pl}$, and the chromosome-encoded $HlyA_{ch}$. We screened 45 isolates from different origins, and found a correlation between their haemolytic phenotypes and the differential haemolysin gene content. All highly and medium haemolytic strains harboured pPHDD1, with amino acid substitutions in $HlyA_{pl}$ and $HlyA_{ch}$ being the cause of the medium haemolytic phenotypes in some pPHDD1-harbouring strains. Weakly haemolytic strains contained only $hlyA_{ch}$, whereas nonhaemolytic isolates, in addition to lacking pPHDD1, either lacked $hlyA_{ch}$ or contained a $hlyA_{ch}$ pseudogene. Sequence analysis of the genomic context of $hlyA_{ch}$ uncovered an unexpected genetic diversity, suggesting that $hlyA_{ch}$ is located in an unstable chromosomal region. Phylogenetic analysis suggested that $hlyA_{pl}$ and $hlyA_{ch}$ originated by gene duplication within *P. damselae* subsp. *damselae* following acquisition by horizontal transfer. These observations together with the differential distribution of pPHDD1 plasmid among strains suggest that horizontal gene transfer has played a main role in shaping the haemolysin gene baggage in this pathogen.

Introduction

Photobacterium damselae subsp. *damselae* (hereafter *Pdd*) is pathogenic to marine animals and humans (Rivas *et al.*, 2013b). It is a primary pathogen of fish species of economic importance in aquaculture (Fouz *et al.*, 1992), and as we have described in recent reports, its isolation from new cultured fish species indicates that this bacterium is an emerging pathogen in marine aquaculture (Labella *et al.*, 2011). Most of the infections in humans have their origin in wounds exposed to seawater, or inflicted during use of tools and fish handling (Morris *et al.*, 1982; Goodell *et al.*, 2004; Aigbivbalu & Maraqa, 2009).

Pdd isolates from marine animals and from humans exhibit haemolytic activity (Kreger, 1984; Fouz *et al.*, 1993). Pioneering studies described the production of damselysin (Dly), a phospholipase D with haemolytic activity (Kreger *et al.*, 1987). Recently, we found that Dly is encoded within pPHDD1, a 153-kb conjugative virulence plasmid that also encodes $HlyA_{pl}$, a haemolysin of the pore-forming toxin family (Rivas *et al.*, 2011). A chromosome-encoded HlyA (hereafter $HlyA_{ch}$) that is 92% identical to plasmid-encoded $HlyA_{pl}$ is responsible of haemolysis in naturally occurring pPHDD1-negative strains, and is also present in pPHDD1-harbouring isolates (Rivas *et al.*, 2013a). The haemolytic phenotype in pPHDD1-harbouring strains is a consequence of synergistic and additive effects of Dly and the two HlyAs (Rivas *et al.*, 2013a). However, pPHDD1 plasmid is only found in a fraction of strains, and various diverse haemolytic phenotypes can be found within the subspecies. In addition, the genetic basis of the lack of haemolysis in the rarely reported nonhaemolytic *Pdd* strains is so far unknown. The closely related bacterium *P. damselae*

FEMS MICROBIOLOGY LETTERS

subsp. *piscicida*, the causative agent of fish pasteurellosis, lacks observable haemolytic activity on sheep blood agar plates (Magariños *et al.*, 1996).

In the present study we conducted a PCR screening of the three haemolysin genes in 45 *Pdd* isolates from different origins and with different haemolytic phenotypes. We also studied the genetic context of the $hlyA_{ch}$ gene, and the sequence variability of $hlyA_{pl}$ and $hlyA_{ch}$ alleles in different strains. The aims of this study were to assess the correlation between haemolysin gene content and haemolytic phenotype, and to gain new insight into the genetic variability of haemolysis determinants in *Pdd*. Our results suggest that horizontal gene transfer has played a main role in shaping the haemolysin gene content in this pathogen.

Materials and methods

Bacterial strains, plasmids and culture conditions

The 45 *P. damselae* subsp. *damselae* strains and one subsp. *piscicida* strain used in this study are listed in Table 1. Strains were routinely grown at 25 °C on trypticase soy agar supplemented with 1% NaCl (TSA-1). Sheep blood agar plates (Oxoid, Basingstoke, UK) were used for conjugative matings and haemolysis assays. For haemolysis assays on agar plates, a single colony of each strain grown on a TSA-1 plate was picked with the tip of a rounded wooden pick and seeded on the blood agar plate, and pictures were taken after 15 h. *Escherichia coli* was grown at 37 °C in Luria–Bertani (LB) broth and on LB agar. Antibiotics were used at the following final concentrations: gentamycin, (Gm) at 15 μg mL^{-1} and rifampin (Rf) at 50 μg mL^{-1}.

DNA techniques

Genomic DNA was extracted with the Easy-DNA kit (Invitrogen, Carlsbad, CA). PCR screening of haemolysin (*dly*, $hlyA_{pl}$, $hlyA_{ch}$) and flanking genes (*kefA* and the oxidoreductase gene) was conducted with primer pairs specifically designed for this study: *dly* (5′-CCTATGGGA CATGAATGG-3′; 5′-GCTCTAGGCTAAATGAATC-3′), $hlyA_{pl}$ (5′-GCTATAAATGAATAAGAAAA-3′; 5′-TTGAA GCTAACTCAAAAA-3′), $hlyA_{ch}$ (5′AATGTTTCTTTCCG TTGGGC-3′; 5′-CCGGAGTTCCACCAGTAAAT-3′), *kefA* (5′-ATAAGCAGCCACAAACTCCC-3′; 5′-ATTTTAGTT GGCGACGCAGG-3′), and oxidoreductase (5′-TTCAAA GCGAGAAATCCACC-3′; 5′-TAATCGGTGGGATGTACG CT-3′). PCR was performed with Kapa *Taq* DNA polymerase (Kapa, Willmington, MA) using a T-gradient termocycler (Biometra, Göttingen, Germany) with the following thermal cycling conditions: 95 °C for 5 min, followed by 30 cycles of 95 °C for 30 s, 57 °C for 30 s, and an elongation step of 1 min at 72 °C. The expected amplicons were 549 bp (*dly*), 767 bp ($hlyA_{pl}$), 353 bp ($hlyA_{ch}$), 340 bp (*kefA*) and 639 bp long (oxidoreductase gene). The genetic context of the $hlyA_{ch}$ gene was amplified by inverse PCR starting from conserved flanking genes. DNA sequences were determined with the CEQ DTCS-Quick Start Kit and a capillary DNA Sequencer CEQ8000 (Beckman Coulter).

Gene complementation

For gene complementation, $hlyA_{pl}$ and $hlyA_{ch}$ ORFs together with their respective promoter sequences were PCR-amplified with Hi-Fidelity Kapa *Taq* (Kapa), cloned into pHRP309 vector and mobilized from *E. coli* S17-1 λ*pir* into the respective *Pdd* strains as previously described (Rivas *et al.*, 2013a). For determination of the specific sequence stretches of the $hlyA_{pl}$ gene of strain RM-71 that complement the MH phenotype of strain ATCC 33539, we constructed partial chimeric $hlyA_{pl}$ genes into an ATCC 33539 background. For this, different overlapping $hlyA_{pl}$ gene fragments were PCR-amplified, cloned into the suicide vector pNidKan (Mouriño *et al.*, 2004) and introduced into strain ATCC 33539 by conjugation from *E. coli* S17-1-λ-*pir*. After insertion of the PCR fragment into the $hlyA_{pl}$ allele of strain ATCC 33539 by a single event of homologous recombination, recombinants were assayed for recovery of the large halo (LH) haemolytic phenotype on sheep blood agar plates.

Molecular phylogenetic analysis

Amino acid sequences of HlyA homologues were aligned using CLUSTALW. Evolutionary analyses were conducted in MEGA5 (Tamura *et al.*, 2011), using the maximum-likelihood method based on the Poisson correction model and 100 bootstrap replicates. Bootstrap values higher than 70 are shown.

Accession numbers

DNA sequences have been deposited in the GenBank database under accession numbers KF984030 to KF984038.

Results and discussion

Haemolysin gene distribution in *Pdd* isolates and its correlation with the haemolytic phenotype

In a previous study we found that *Pdd* strains can be divided into four categories depending on the width of

Table 1. Results of the PCR screening for gene markers, and haemolytic phenotypes for the 45 *Photobacterium damselae* subsp. *damselae* strains and one subsp. *piscicida* strain used in this study

P. damselae strains	Source	*Oxidoreductase*	*kefA*	*hlyA*$_{ch}$	*hlyA*$_{pl}$	*dly*	*Haemolytic halo*
RM-71	Turbot, Spain	+	+	+	+	+	LH
RG-91	Turbot, Spain	+	+	+	+	+	LH
RG-153	Turbot, Spain	+	+	+	+	+	LH
RG-214	Turbot, Spain	+	+	+	+	+	LH
CDC-2227-81	Human, United States	+	+	+	+	+	LH
80077637	Human, Australia	+	+	+	+	+	LH
AZ 245.1	Turbot, Spain	+	+	+	+	+	LH
ATCC 33539	Damselfish, United States	+	+	+	+	+	MH
402O	Redbanded seabream, Spain	+	+	+	+	+	MH
TW294 L2	European seabass, Spain	+	+	+	–	–	SH
TW462/02.1	Gilthead seabream, Spain	+	+	+	–	–	SH
TW250/03	Gilthead seabream, Spain	+	+	+	–	–	SH
AZ 247.1	Turbot, Spain	+	+	+	–	–	SH
ACR 208.1	Turbot, Spain	+	+	+	–	–	SH
USC-Viro-1	Turbot, Spain	+	+	+	–	–	SH
LD-07	Gilthead seabream, Spain	+	+	+	–	–	SH
238	Dolphin, United States	+	+	+	–	–	SH
309	Mussel, Spain	+	+	+	–	–	SH
158	Eel, Belgium	+	+	+	–	–	SH
162	Eel, Belgium	+	+	+	–	–	SH
PG-801	Shrimp, Taiwan	+	+	+	–	–	SH
192	Dolphin, United States	+	+	+	–	–	SH
ATCC 35083	Brown shark, United States	+	+	+	–	–	SH
602 R	Redbanded seabream, Spain	+	+	+	–	–	SH
S503C	White seabream, Spain	+	+	+	–	–	SH
601R	Redbanded seabream, Spain	+	+	+	–	–	SH
LCA24907	Barramundi (*Lates calcarifer*)	+	+	+	–	–	SH
501B	Redbanded seabream, Spain	+	+	+	–	–	SH
501R	Redbanded seabream, Spain	+	+	+	–	–	SH
403D1	Redbanded seabream, Spain	+	+	+	–	–	SH
203H	Redbanded seabream, Spain	+	+	+	–	–	SH
Lb501R	European seabass, Spain	+	+	+	–	–	SH
D408U	Gilthead seabream, Spain	+	+	+	–	–	SH
601R	Redbanded seabream, Spain	+	+	+	–	–	SH
401H	Redbanded seabream, Spain	+	+	+	–	–	SH
9401H	Redbanded seabream, Spain	+	+	+	–	–	SH
NPV26	Plancktom, Adriatic Sea	+	+	+	–	–	SH
9FT1M-3	Shark, United States	+	+	+	–	–	SH
9FT2B-2	Shark, United States	+	+	+	–	–	SH
RS78SPL1	Red snapper, United States	+	+	+	–	–	SH
ST-1	Seatrout, United States	+	+	+	–	–	SH
RS80L1V1	Red snapper, United States	+	+	+	–	–	SH
501H	Redbanded seabream, Spain	+	+	+$^{\Psi}$	–	–	NH
ACRP-72.1	Turbot, Spain	+	+	–	–	–	NH
J3G-801	Shrimp, Taiwan	+	+	–	–	–	NH
DI-21*	Gilthead seabream, Spain	+	+	–	–	–	NH

LH, large haemolytic halo; MH, medium haemolytic halo; SH, small haemolytic halo; NH, no haemolytic halo.

$^{\Psi}$This strain contains an *hlyA*$_{ch}$ pseudogene.

*DI-21 is a subsp. *piscicida* strain.

the haemolytic halo when present (LH, large halo; MH, medium halo; SH, small halo) or on its absence (NH, nonhaemolytic) (Rivas *et al.*, 2011). Here we screened 45 strains to analyse the presence of the *dly*, *hlyA*$_{pl}$ and *hlyA*$_{ch}$ genes and the results were compared with their haemolytic phenotypes. As shown in Table 1 and Fig. 1(a), we found that each of the four distinct haemolytic phenotypes can be correlated with a specific gene

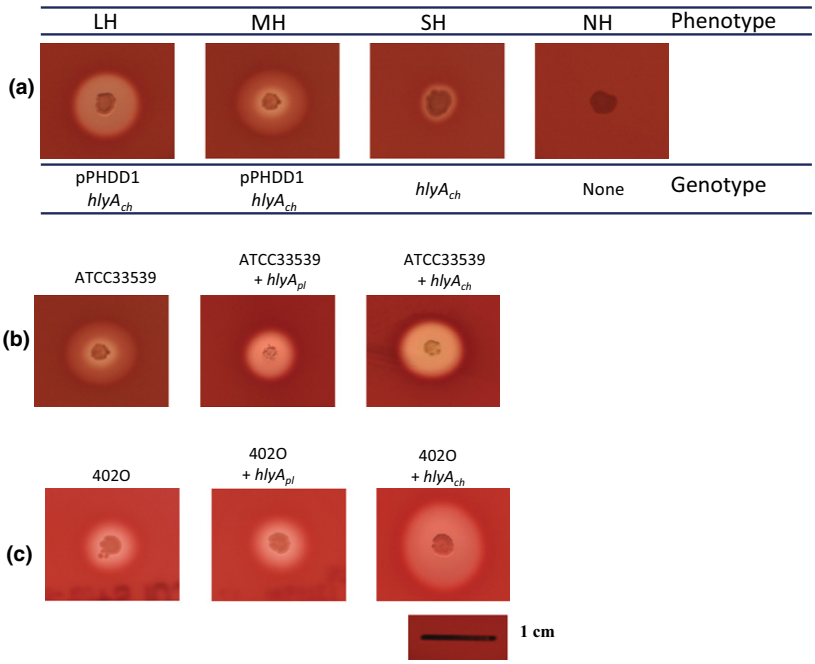

Fig. 1. (a) Representation of the four distinct haemolytic phenotypes observed in *Photobacterium damselae* subsp. *damselae* strains on sheep blood agar and correlation with a specific genotype. LH, large halo; MH, medium halo; SH, small halo; NH, nonhaemolytic. Below each picture the genotype is represented in terms of presence or absence of plasmid pPHDD1 and of the $hlyA_{ch}$ gene. (b,c) Results of complementation of strains ATCC 33539 (b) and 402O (c) (showing an MH phenotype) with plasmids containing either $hlyA_{pl}$ or $hlyA_{ch}$ genes from strain RM-71 (LH phenotype). Scale bar = 1 cm.

combination. All the LH strains tested positive for the two pPHDD1-encoded haemolysin genes and for $hlyA_{ch}$, a gene combination that produces the maximum haemolytic activity reported in this subspecies (Rivas *et al.*, 2013a). The $hlyA_{ch}$ gene proved to be ubiquitous in all the haemolytic strains regardless of their LH, MH or SH phenotype, which suggests that $HlyA_{ch}$ is responsible for haemolysis in strains showing an SH phenotype (Table 1). In support of this, we previously demonstrated that deletion of $hlyA_{ch}$ in the naturally plasmidless strain LD-07 caused elimination of detectable haemolysis against sheep erythrocytes (Rivas *et al.*, 2013a). Finally, among the three plasmidless strains that showed no haemolysis (NH), two tested negative for the three markers (ACRP -72.1 and J3G801), and one (501H) contained a $hlyA_{ch}$ pseudogene.

To find an explanation for the MH phenotypes in pPHDD1-harbouring strains, we complemented strains ATCC 33539 and 402O with $hlyA_{pl}$ or $hlyA_{ch}$ genes from an LH strain (RM-71). We observed that $hlyA_{pl}$ caused a recovery of the LH phenotype in strain ATCC 33539 but not in 402O, whereas complementation with $hlyA_{ch}$ produced a recovery of the LH phenotype in the two strains (Fig. 1b and c). To identify candidate substitutions affecting the activity of this protein, we PCR-amplified different fragments of the RM-71 $hlyA_{pl}$ gene, cloned them into the suicide vector pNidKan and selected for recombination of the constructions into the strain ATCC 33539 $hlyA_{pl}$ allele to generate chimeric proteins. Interestingly, recombination of an RM-71 $hlyA_{pl}$ gene PCR fragment encoding the last 50 aa residues into strain ATCC 33539 rendered an LH phenotype (data not shown). Sequence analysis showed only one amino acid change between strains ATCC 33539 and RM-71 in that region of the protein, and corresponded to an alanine substitution of serine at position 575 (24 positions from the C terminus of the protein). This position is located in the C-terminal 14-kDa β-trefoil lectin domain of $HlyA_{pl}$ (Rivas *et al.*, 2011), and is thought to be related to carbohydrate binding (Zhang *et al.*, 1999).

These results suggest that MH phenotypes are caused by amino acid substitutions in $HlyA_{pl}$, $HlyA_{ch}$ or both, which probably decrease the activity of the haemolysins. Together, the results described here clearly show the existence of a correlation between haemolysin gene content and the type of haemolytic halo produced in *Pdd*.

The genomic context of $hlyA_{ch}$ reveals unexpected gene diversity and contains features related to mobile DNA

Identification of nonhaemolytic *Pdd* strains that test negative for the $hlyA_{ch}$ gene and the fact that the closely related bacterium *P. damselae* subsp. *piscicida* is nonhaemolytic on sheep blood agar suggest that the genome region encoding $HlyA_{ch}$ might be a hot spot for insertion–deletion (indel) events. Comparing the genome sequences of subsp. *damselae* ATCC 33539 (GenBank accession no. ADBS00000000) and subsp. *piscicida* DI21 (AKYG00000000.1) we found that $hlyA_{ch}$ is located in

chromosome I of strain ATCC 33539 within a variable region. This region shows a differential gene content between strains ATCC 33539 and DI21 and is flanked by conserved genes: *kefA* (encoding a putative ion channel) (VDA_002421 in the strain ATCC 33539 genome annotation) plus two serine tRNA genes, and a oxidoreductase gene (VDA_002412) (Fig. 2).

To confirm the hypothesis that $hlyA_{ch}$ is located within a variable genome region, we sequenced the DNA flanked by *kefA* and oxidoreductase genes in eight additional strains, including representatives of the different haemolytic phenotypes. We found that each strain had a unique gene repertoire, with the number of variable genes ranging from two to 12. Nucleotide sequence similarity of the *kefA* and the oxidoreductase gene among strains was 98%, indicating that they are conserved genes in this species, and PCR screening demonstrated that the two are present in all the *P. damselae* strains (Table 1).

The putative functions encoded by the variable genes are highly diverse (Table 2), and include hypothetical proteins, transcriptional regulators (AraC- and LysR-type regulators), ATP-binding proteins, DNA metabolism enzymes (timidylate kinases), degradative enzymes (hydrolases and lipases) and functions related to mobile elements (restric-

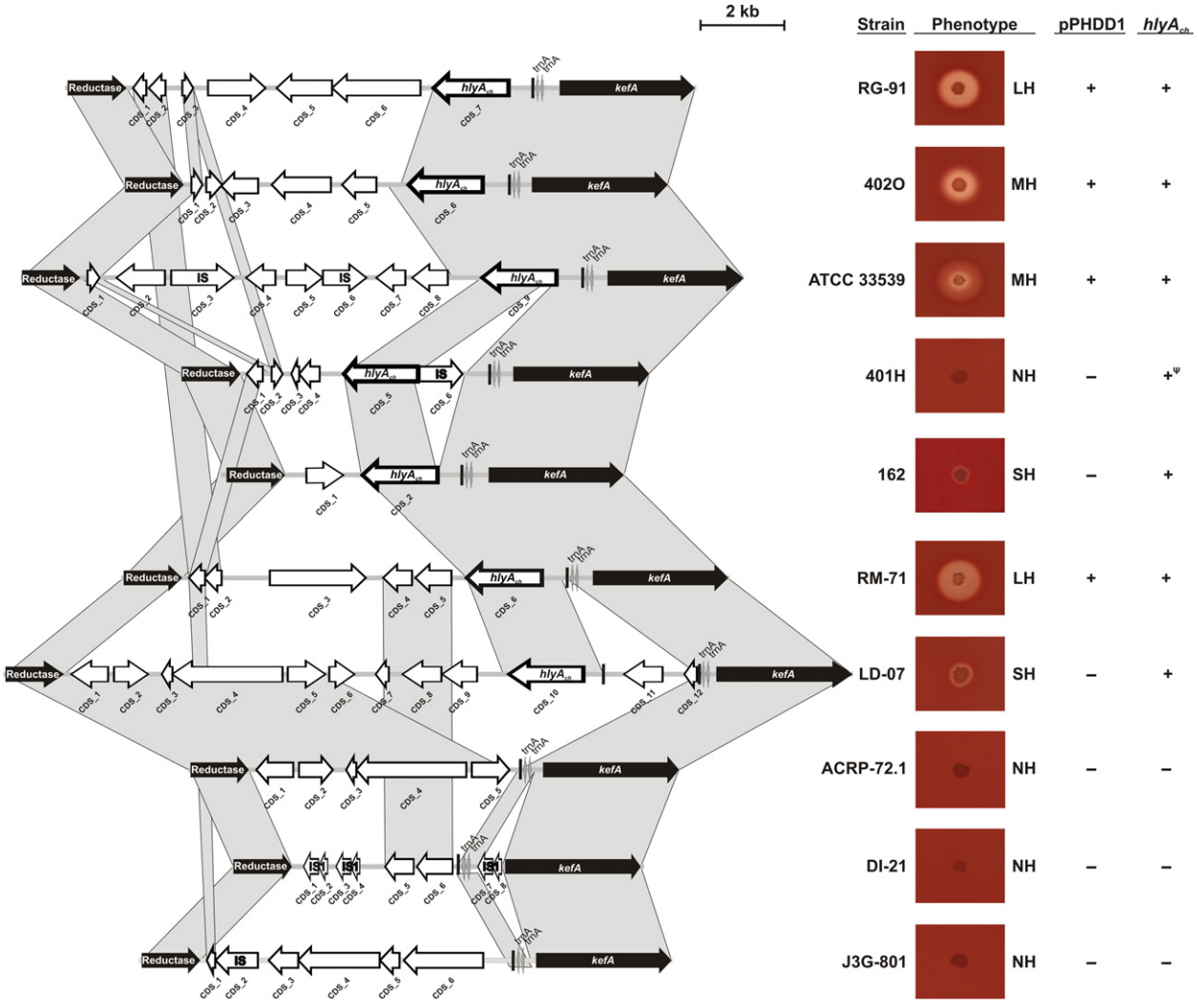

Fig. 2. Schematic representation of the variable chromosomal region (variable genes represented as white arrows) flanked by the conserved genes *kefA* and oxidoreductase (represented as black arrows) in nine *Photobacterium damselae* subsp. *damselae* strains and one subsp. *piscicida* (DI-21) strain. Homologous regions are connected by diagonal grey bands between pairs of strains. Two serine tRNA genes (represented by grey arrows upstream from *kefA*) are present in all strains. This variable region is probably a genome assembly hot spot for indel events, including those that affected $hlyA_{ch}$ gene gain/loss. The haemolytic phenotypes and the genotypes are depicted in the right-hand panels. Note that of the four nonhaemolytic strains, three lack $hlyA_{ch}$ and one (401H) contains an *IS* element that disrupts $hlyA_{ch}$ (ψ). Strains RG-91, 4020, ATCC 33539 and RM-71 harbour plasmid pPHDD1.

Table 2. Predicted proteins encoded by the variable genes located within the region flanked by conserved *kefA* and oxidoreductase genes in nine *Photobacterium damselae* subsp. *damselae* strains and one (DI-21) subsp. *piscicida* strain

Feature ID	Length (bp)	Function of the predicted protein	Homologue matches			
			Accesion no.	Species	Identities	E value
RM71_CDS1	393	Hypothetical protein	YP_002019309.1	*Pelodictyon phaeoclathratiforme*	40%	2.00E-30
RM71_CDS2	411	Hypothetical protein	YP_002019310.1	*P. phaeoclathratiforme*	57%	1.00E-47
RM71_CDS3	2310	ATPase	WP_007467808.1	*Photobacterium* sp.	95%	0.0
RM71_CDS4	696	Aspartate racemase	WP_010431798.1	*Vibrio cyclitrophicus*	78%	2.00E-129
RM71_CDS5	873	AraC family transcriptional regulator	YP_005049609.1	*Vibrio furnissii*	61%	9.00E-121
RM71_CDS6	1809	HlyA^{ch}	EEZ41388.1	*Photobacterium damselae* subsp. *damselae*	99%	0.0
ATCC33539_CDS1	291	Hypothetical protein	YP_004188639.1	*Vibrio vulnificus*	78%	1.00E-40
ATCC33539_CDS2	1179	Hypothetical protein	EEZ41381.1	*P. damselae* subsp. *damselae*	100%	0.0
ATCC33539_CDS3	1530	Transposase	EEZ41382.1	*P. damselae* subsp. *damselae*	100%	0.0
ATCC33539_CDS4	717	Hypothetical protein	EEZ41383.1	*P. damselae* subsp. *damselae*	100%	1.00E-178
ATCC33539_CDS5	876	Phage integrase family protein	EEZ41384.1	*P. damselae* subsp. *damselae*	100%	0.0
ATCC33539_CDS6	1059	Transposase	EEZ41385.1	*P. damselae* subsp. *damselae*	100%	0.0
ATCC33539_CDS7	714	Hypothetical protein	EEZ41386.1	*P. damselae* subsp. *damselae*	100%	4.00E-168
ATCC33539_CDS8	873	Hypothetical protein	EEZ41387.1	*P. damselae* subsp. *damselae*	100%	0.0
ATCC33539_CDS9	1809	HlyA^{ch}	EEZ41388.1	*P. damselae* subsp. *damselae*	100%	0.0
RG91_CDS1 (Ψ)	327	Hypothetical protein	YP_002019309.1	*P. phaeoclathratiforme*	38	7.00E-20
RG91_CDS2	411	Hypothetical protein	YP_002019310.1	*P. phaeoclathratiforme*	56%	5.00E-45
RG91_CDS3	258	Hypothetical protein	YP_004188639.1	*Vibrio vulnificus*	84%	1.00E-41
RG91_CDS4	1380	Deoxyguanosine tri phosphate triphosphohydrolase	YP_004313207.1	*Marinomonas mediterranea*	77%	0.0
RG91_CDS5	1335	Putative DNA processing chain A	WP_007464740.1	*Photobacterium* sp.	78%	0.0
RG91_CDS6	2127	ATP-dependent DNA helicase RecQ	WP_007464742.1	*Photobacterium* sp.	88%	0.0
RG91_CDS7	1809	HlyA^{ch}	EEZ41388.1	*P. damselae* subsp. *damselae*	99%	0.0
LD07_CDS1	900	LysR family transcriptional regulator	WP_004744032.1	*Vibrio tubiashii*	60%	1.00E-127
LD07_CDS2	829	Metal dependent phosphohydrolase	YP_001472395.1	*Shewanella sediminis*	70%	2.00E-143
LD07_CDS3 (Ψ)	267	GCN5-like N-acetyltransferase	YP_004111988.1	*Desulfurispirillum indicum*	78%	9.00E-07
LD07_CDS4	2679	Ankyrin repeat-containing protein	NP_762834.1	*Vibrio vulnificus*	82%	0.0
LD07_CDS5	915	Molybdenum cofactor biosynthesis protein	NP_762835.2	*V. vulnificus*	92%	0.0
LD07_CDS6	616	Thymidylate kinase	WP_006230234.1	*Photobacterium profundum*	74%	6.00E-106
LD07_CDS7 (Ψ)	321	Hypothetical protein	YP_003145259.1	*Kangiella koreensis*	63%	9.00E-44
LD07_CDS8	942	ATP-binding protein	WP_010302393.1	*Pectobacterium carotovorum*	32%	1.00E-42
LD07_CDS9	847	Glutamyl-tRNA synthetase	WP_004537850.1	*Burkholderia pseudomallei*	37%	5.00E-42
LD07_CDS10	1809	HlyA^{ch}	EEZ41388.1	*P. damselae* subsp. *damselae*	99%	0.0
LD07_CDS11 (Ψ)	942	HNH endonuclease	YP_751179.1	*Shewanella frigidimarina*	41%	2.00E-38
LD07_CDS12 (Ψ)	162	Integrase	YP 001448934.1	*Vibrio harveyi*	55%	2.00E-08
ACRP-72.1_CDS1	900	LysR family transcriptional regulator	WP_004744032.1	*V. tubiashii*	59%	2.00E-126
ACRP-72.1_CDS2	828	Metal dependent phosphohydrolase	YP_001472395.1	*S. sediminis*	70%	2.00E-143
ACRP-72.1_CDS3 (Ψ)	246	GCN5-like N-acetyltransferase	YP_004111988.1	*D. indicum*	78%	1.00E-06
ACRP-72.1_CDS4	2685	Ankyrin repeat-containing protein	NP_762834.1	*V. vulnificus*	81%	0.0

Table 2. Continued

Feature ID	Length (bp)	Function of the predicted protein	Homologue matches			
			Accesion no.	Species	Identities	E value
Acrp 72.1_CDS5	916	Molybdenum cofactor biosynthesis	NP_762835.2	*V. vulnificus*	91%	0.0
J3G801_CDS1 (Ψ)	210	Hypothetical protein	YP_002019309.1	*P. phaeoclathratiforme*	35%	0.32
J3G801_CDS2	1038	Transposase	WP_005489462.1	*Vibrio parahaemolyticus*	96%	0.0
J3G801_CDS3 (Ψ)	726	Hypothetical protein	WP_001923066.1	*V. cholerae*	30%	7.00E-10
J3G801_CDS4	1995	Lipase	WP_001976335.1	*V. cholerae*	29%	2.00E-37
J3G801_CDS5	482	Unknown protein	No homologues			
J3G801 CDS6	1941	Type IV secretion protein Rhs	WP 001894897.1	*V. cholerae*	36%	8.00E-107
501H_CDS1	393	Hypothetical protein	YP_002019309.1	*P. phaeoclathratiforme*	41%	1.00E-30
501H_CDS2	270	Hypothetical protein	YP_004188639.1	*V. vulnificus*	80%	3.00E-39
501H_CDS3 (Ψ)	191	ATP-binding protein	WP_006879706.1	*Vibrio brasiliensis*	74%	1.00E-15
501H_CDS4	495	Tetrapyrrole methylase	WP_009441834.1	*Selenomonas* sp.	30%	0.23
501H_CDS5	1791	$HlyA^{ch}$	EEZ41388.1	*P. damselae* subsp. *damselae*	99%	0.0
501H CDS6	1041	Transposase	YP_001436144.1	*V. harveyi*	92%	0.0
402O_CDSl	270	Hypothetical protein	YP_004188639.1	*V. vulnificus*	80%	3.00E-40
402O_CDS2 (Ψ)	390	Aldose 1-epimerase	YP_634029.1	*Myxococcus xanthus*	38%	0.57
402O_CDS3	903	Divergent AAA domain protein	WP_001459743.1	*Escherichia coli*	44%	6.00E-78
402O_CDS4	1449	Hypothetical protein	YP_003157719.1	*Desulfomicrobium baculatum*	57%	0.0
402O_CDS5	837	Hypothetical protein	WP_010699753.1	*Pantoea agglomerans*	69%	3.00E-146
402O_CDS6	1809	$HlyA^{ch}$	EEZ41388.1	*P. damselae* subsp. *damselae*	98%	0.0
162_CDS1	900	Hypothetical protein	YP_005049289.1	*V. furnissii*	60%	4.00E-112
162_CDS2	1809	$HlyA^{ch}$	EEZ41388.1	*P. damselae* subsp. *damselae*	98%	0.0
DI21_CDS1	426	Transposase IS1	WP_006712141.1	*Vibrio ichthyoenteri*	84%	9.00E-77
DI21_CDS2	276	Putative transposase	CAH04677.1	*P. damselae* subsp. *damselae*	100%	2.00E-52
DI21_CDS3	426	Transposase IS1	WP_006712141.1	*Vibrio ichthyoenteri*	84%	9.00E-77
DI21_CDS4	276	Putative transposase	CAH04677.1	*P. damselae* subsp. *damselae*	100%	2.00E-52
DI21_CDS5	696	Aspartate racemase	WP_010431798.1	*Vibrio cyclitrophicus*	79%	5.00E-131
DI21_CDS6	873	AraC family transcriptional regulator	YP_005049609.1	*Vibrio furnissii*	60%	7.00E-119
DI21_CDS7	426	Transposase IS1	WP_006712141.1	*Vibrio ichthyoenteri*	84%	9.00E-77
DI21_CDS8	276	Putative transposase	CAH04677.1	*P. damselae* subsp. *damselae*	100%	2.00E-52

Ψ, Pseudogene.

tion endonucleases, phage integrases, Rhs elements and insertion sequences). Three genes in strain RG-91, *dgt*, *dprA* and *recQ*, encode functions related to the processing of single-strand DNA following transformation. A transposase gene truncating $hlyA_{ch}$ in strain 501H is likely to be the cause of the lack of haemolytic phenotype in this strain (Fig. 2).

Some pairwise comparisons between strains suggest that genes have been inserted or deleted during evolution. Strains LD-07 and ACRP-72.1 share a unique cluster of five genes, but LD-07 has an additional six-gene cluster that includes $hlyA_{ch}$ (Fig. 2). Overall, this variable region might be considered as a genome assembly hot spot where indel events have occurred, including those that affected $hlyA_{ch}$ acquisition or loss. Similar situations of putative genome assembly hot spots have been reported recently in other species, as for *Oenococcus oeni* genes encoding phosphotransferase functions (Jamal *et al.*, 2013). Coinfection of marine animals (Zhao *et al.*, 2009; Labella *et al.*, 2010) and humans (Clarridge & Zighelboimdaum, 1985) with different phenotypes of *Pdd* simultaneously in a host has been demonstrated. It is thus tempting to hypothesize that some genetic variants in the $hlyA_{ch}$ region may have originated by recombination between different *Pdd* strains coexisting in the same host, in a similar way as described for *Helicobacter pylori* (Didelot *et al.*, 2013).

The presence of genes related to DNA mobility as integrases, insertion sequences, Rhs elements and tRNA genes (Table 2) suggests that mobile elements may have driven the hypervariability in this genomic region. It is well known that tRNA genes play a general role in DNA

acquisition (Hou, 1999) and direct repeats and insertion elements flanking pathogenicity islands often overlap with or are close to tRNA genes (Blum *et al.*, 1994). A comparative analysis between *Salmonella enterica* and *E. coli* isolates showed that the region associated with $tRNA^{leuX}$ is extremely variable, often encoding completely unrelated DNA sequences and genes among strains (Bishop *et al.*, 2005). This is the case of the $hlyA_{ch}$-containing DNA region in *P. damselae*, where there is no apparent relationship among genes. Close homologues of the variable genes belong to different species, and many of them share little or no relationship to *Photobacterium* or *Vibrio* genes (Table 2).

One characteristic of mobile elements is the presence of repetitive DNA flanking their sequences (Blum *et al.*, 1994). By analysing the variable DNA regions close to $hlyA_{ch}$ we found immediately downstream of the two serine tRNAs a 7-pb DNA sequence (GCAATAC) that occurs either in single copy or in multicopy in all strains tested (Fig. 3). Interestingly, while the $hlyA_{ch}$-negative subsp. *piscicida* strain DI-21 contained a single copy of this heptamer, the subsp. *damselae* strains RM-71 and LD-07 showed two copies of this heptamer flanking the $hlyA_{ch}$ gene as a discrete DNA module, and in the case of LD-07 there was an extra copy of this heptamer (Fig. 3). *Pdd* strain ATCC 33539 also showed this 7-pb repeat flanking $hlyA_{ch}$ and *cds8* genes.

Integrated genomic islands are usually located downstream of a tRNA locus and repetitive sequences comprising 20–30 nt of the 3′ end of the tRNA are usually found (Dufour *et al.*, 2011). Interestingly, we observed that strain LD-07 shows two direct repeats that correspond to partial serine tRNA sequences of 26 bp (Fig. 3), between CDS2 and CDS3 and between CDS10 and CDS11. Interestingly CDS11 is a pseudogene of an integrase, and it is well known that integrase genes are commonly associated with tRNA genes in genomic islands. Strain ACRP-72.1 also contains this 26-bp partial tRNA sequence between CDS2 and CDS3 (Fig. 3), a trait that is conserved between LD-07 and ACRP-72.1 together with a collection of five genes of the variable fraction. Due to the presence of genes related to DNA mobility (integrases, transposases, tRNA loci) and to recombination processes and point mutations, genomic islands constantly undergo structural

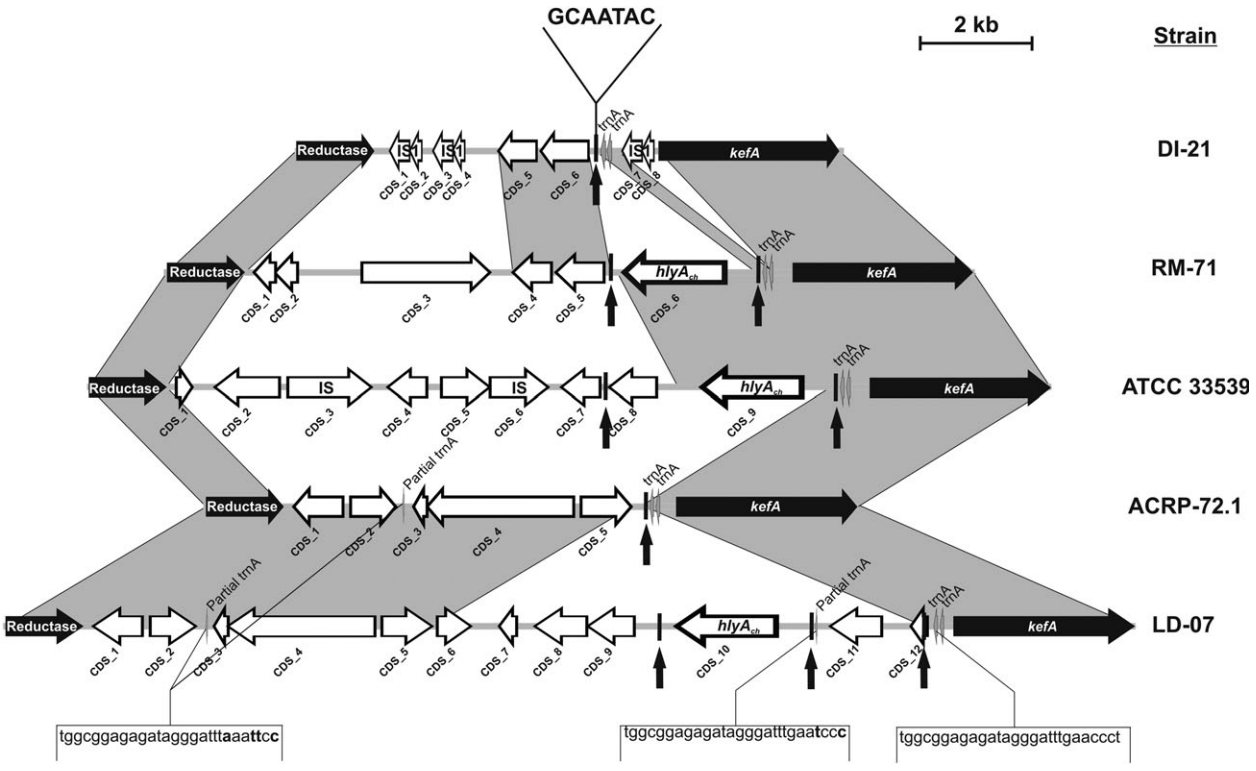

Fig. 3. Detailed scheme of the chromosome region containing the variable loci flanked by the conserved genes *kefA* and oxidoreductase in five of the strains studied, highlighting DNA features putatively related to insertion of foreign DNA. The copies of the heptamer GCAATAC are denoted by a vertical black bar pinpointed by a vertical black arrow. Note that this heptamer occurs once in DI-21 and ACRP-72.1, twice in RM-71 and ATCC 33539, and three times in LD-07. The 'partial tRNA' sequence repeated in strains ACRP-72.1 and LD-07 and the corresponding homologous sequence in the complete tRNA gene are also highlighted.

changes (Schneider *et al.*, 2011). In *Bradyrhizobium japonicum*, duplications of fragments of tRNA genes, flanked by DNA insertions ranging from 4 to 97 kb, have been found along the genome, strongly suggesting that insertion of DNA elements into tRNA genes took place (Kaneko *et al.*, 2002). Thus, the repeated 26-bp sequence of a partial tRNA gene found in *Pdd* might represent a scar of former genomic islands that have undergone events of DNA insertions and losses that led to the current gene composition.

Phylogenetic analysis suggests that *hlyA$_{pl}$* and *hlyA$_{ch}$* have originated by gene duplication within *Pdd* following acquisition by horizontal transfer

Homologues of the *Pdd* HlyA proteins are encoded in several species of *Vibrionaceae* and *Aeromonadaceae*, most of them pathogenic for marine animals and some opportunistic human pathogens. In the present study we constructed a phylogenetic tree with the sequences of ten *Pdd* HlyA proteins (seven $HlyA_{ch}$ and three $HlyA_{pl}$) and with HlyA homologues from other species, retrieved from protein databases (Fig. 4). The large genetic distance that separates the HlyA homologues from different species suggests a long-term association of *hlyA* genes with *Vibrionaceae* and *Aeromonadaceae* species.

Phylogenetic analysis yielded two well-defined branches in *Pdd*, one comprising $HlyA_{ch}$ sequences and one that includes $HlyA_{pl}$ sequences (Fig. 4). The high similarity at both the nucleotide and the amino acid levels between the two HlyA toxins of *Pdd* compared with their homologues in other species suggests that the two gene versions, *hlyA$_{ch}$* and *hlyA$_{pl}$*, have originated from a gene duplication event and started to diverge within *Pdd* upon acquisition of the original gene copy by horizontal transfer. It is interesting to note that so far no plasmid-borne *hlyA* homologues have been found outwith *Pdd*. In addition, while *hlyA$_{ch}$* is located in chromosome I of *Pdd*, other *Vibrionaceae* species harbour this gene in chromosome II, such as *Vibrio anguillarum* (Naka *et al.*, 2011), *Vibrio cholerae* (Heidelberg *et al.*, 2000) and *Vibrio fluvialis* (Khatri *et al.*, 2013).

The location of the *hlyA$_{pl}$* gene immediately upstream of the damselysin toxin-encoding gene *dly* in plasmid pPHDD1 seems to obey to a selection pressure for maintaining (and hence for coregulating) together two genes whose products have been demonstrated to act in a synergistic manner to produce haemolysis (Rivas *et al.*, 2013a). The plasmid location of *hlyA$_{pl}$* and the strong variability of the genetic context surrounding *hlyA$_{ch}$* might support the hypothesis of acquisition of the *hlyA* gene by horizontal transfer and a further gene duplication event.

Conclusions

The three haemolysins are encoded within unstable genome regions, namely plasmids and potential genome assembly hot spots that have contributed to strain differentiation and subspecies evolution in *Pdd*. Phylogenetic analysis suggests that *hlyA$_{pl}$* and *hlyA$_{ch}$* originated by gene duplication within *Pdd*. These observations, together with the differential distribution of plasmid pPHDD1 among

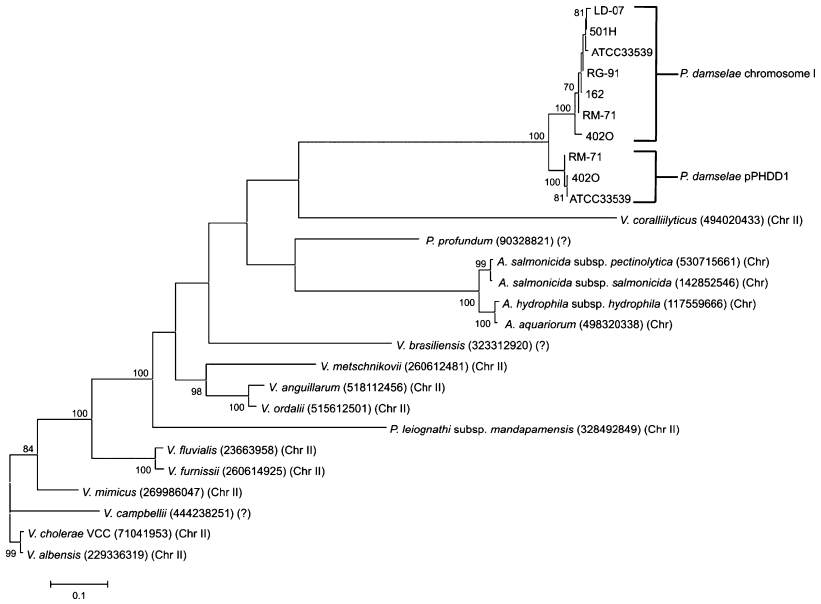

Fig. 4. Molecular phylogenetic analysis of *Photobacterium damselae* subsp. *damselae* pPHDD1 plasmid- and chromosome I-encoded HlyA proteins and of homologues in *Vibrionaceae* and *Aeromonadaceae* species, with the maximum-likelihood method (log likelihood –3591.8451) with 100 bootstrap replicates. Bootstrap values higher than 70 are shown. *Vibrio cholerae* cytolysin VCC was used as an external group. GenBank sequence identification numbers (GI) are shown in parentheses. Information on the chromosomal location of the *hlyA* gene is provided (Chr II, chromosome II; Chr, bacteria with a single chromosome; ?, location unknown). The tree is drawn to scale, with branch lengths measured according to the number of substitutions per site.

strains, suggest that horizontal gene transfer has played a main role in shaping the haemolysin gene baggage in this pathogen.

Acknowledgements

This work was supported by grant EM2012/043 from the Xunta de Galicia, Spain, and by grant AGL2012-39274-C02-01 from the Ministry of Economy and Competitiveness (MINECO) of Spain (cofunded by the FEDER Programme from the European Union). We thank Oliver Nolte and Jay Grimes for providing *Pdd* strains. The authors declare no conflict of interest.

References

Aigbivbalu L & Maraqa N (2009) *Photobacterium damsela* wound infection in a 14-year-old surfer. *South Med J* **102**: 425–426.

Bishop AL, Baker S, Jenks S *et al.* (2005) Analysis of the hypervariable region of the *Salmonella enterica* genome associated with tRNA(leuX). *J Bacteriol* **187**: 2469–2482.

Blum G, Ott M, Lischewski A, Ritter A, Imrich H, Tschape H & Hacker J (1994) Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect Immun* **62**: 606–614.

Clarridge JE & Zighelboimdaum S (1985) Isolation and characterization of 2 hemolytic phenotypes of *Vibrio damsela* associated with a fatal wound infection. *J Clin Microbiol* **21**: 302–306.

Didelot X, Nell S, Yang I, Woltemate S, van der Merwe S & Suerbaum S (2013) Genomic evolution and transmission of *Helicobacter pylori* in two South African families. *P Natl Acad Sci USA* **110**: 13880–13885.

Dufour D, Germon P, Brusseaux E, Le Roux Y & Dary A (2011) First evidence of the presence of genomic islands in *Escherichia coli* P4, a mammary pathogen frequently used to induce experimental mastitis. *J Dairy Sci* **94**: 2779–2793.

Fouz B, Larsen JL, Nielsen B, Barja JL & Toranzo AE (1992) Characterization of *Vibrio damsela* strains isolated from turbot *Scophthalmus maximus* in Spain. *Dis Aquat Org* **12**: 155–166.

Fouz B, Barja JL, Amaro C, Rivas C & Toranzo AE (1993) Toxicity of the extracellular products of *Vibrio damsela* isolated from diseased fish. *Curr Microbiol* **27**: 341–347.

Goodell KH, Jordan MR, Graham R, Cassidy C & Nasraway SA (2004) Rapidly advancing necrotizing fasciitis caused by *Photobacterium* (*Vibrio*) *damsela*: a hyperaggressive variant. *Crit Care Med* **32**: 278–281.

Heidelberg JF, Eisen JA, Nelson WC *et al.* (2000) DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature* **406**: 477–483.

Hou YM (1999) Transfer RNAs and pathogenicity islands. *Trends Biochem Sci* **24**: 295–298.

Jamal Z, Miot-Sertier C, Thibau F, Dutilh L, Lonvaud-Funel A, Ballestra P, Le Marrec C & Dols-Lafargue M (2013) Distribution and functions of phosphotransferase system genes in the genome of the lactic acid bacterium *Oenococcus oeni*. *Appl Environ Microbiol* **79**: 3371–3379.

Kaneko T, Nakamura Y, Sato S *et al.* (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* **9**: 189–197.

Khatri I, Mahajan S, Dureja C, Subramanian S & Raychaudhuri S (2013) Evidence of a new metabolic capacity in an emerging diarrheal pathogen: lessons from the draft genomes of *Vibrio fluvialis* strains PG41 and I21563. *Gut Pathog* **5**: 1757–4749.

Kreger AS (1984) Cytolytic activity and virulence of *Vibrio damsela*. *Infect Immun* **44**: 326–331.

Kreger AS, Bernheimer AW, Etkin LA & Daniel LW (1987) Phospholipase D activity of *Vibrio damsela* cytolysin and its interaction with sheep erythrocytes. *Infect Immun* **55**: 3209–3212.

Labella A, Manchado M, Alonso MC, Castro D, Romalde JL & Borrego JJ (2010) Molecular intraspecific characterization of *Photobacterium damselae* ssp *damselae* strains affecting cultured marine fish. *J Appl Microbiol* **108**: 2122–2132.

Labella A, Berbel C, Manchado M, Castro D & Borrego JJ (2011) *Photobacterium damselae* subsp. *damselae*, an emerging pathogen affecting new cultured marine fish species in Southern Spain. *Recent Advances in Fish Farms* (Faruk A, ed.), pp. 135–142. InTech, Rijeka, Croatia.

Magariños B, Toranzo AE & Romalde JL (1996) Phenotypic and pathobiological characteristics of *Pasteurella piscicida*. *Annu Rev Fish Dis* **6**: 41–64.

Morris JG, Wilson R, Hollis DG, Weaver RE, Miller HG, Tacket CO, Hickman FW & Blake PA (1982) Illness caused by *Vibrio damsela* and *Vibrio hollisae*. *Lancet* **1**: 1294–1297.

Mouriño S, Osorio CR & Lemos ML (2004) Characterization of heme uptake cluster genes in the fish pathogen *Vibrio anguillarum*. *J Bacteriol* **186**: 6159–6167.

Naka H, Dias GM, Thompson CC, Dubay C, Thompson FL & Crosa JH (2011) Complete genome sequence of the marine fish pathogen *Vibrio anguillarum* harboring the pJM1 virulence plasmid and genomic comparison with other virulent strains of *V. anguillarum* and *V. ordalii*. *Infect Immun* **79**: 2889–2900.

Rivas AJ, Balado M, Lemos ML & Osorio CR (2011) The *Photobacterium damselae* subsp. *damselae* hemolysins damselysin and HlyA are encoded within a new virulence plasmid. *Infect Immun* **79**: 4617–4627.

Rivas AJ, Balado M, Lemos ML & Osorio CR (2013a) Synergistic and additive effects of chromosomal and plasmid-encoded hemolysins contribute to hemolysis and virulence in *Photobacterium damselae* subsp. *damselae*. *Infect Immun* **81**: 3287–3299.

Rivas AJ, Lemos ML & Osorio CR (2013b) *Photobacterium damselae* subsp. *damselae*, a bacterium pathogenic for

marine animals and humans. *Front Microbiol* **4**: 283. doi: 10.3389/fmicb.2013.00283.

Schneider G, Dobrindt U, Middendorf B, Hochhut B, Szijarto V, Emody L & Hacker J (2011) Mobilisation and remobilisation of a large archetypal pathogenicity island of uropathogenic *Escherichia coli in vitro* support the role of conjugation for horizontal transfer of genomic islands. *BMC Microbiol* **11**: 1471–2180.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M & Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.

Zhang D, Takahashi J, Seno T, Tani Y & Honda T (1999) Analysis of receptor for *Vibrio cholerae* El tor hemolysin with a monoclonal antibody that recognizes glycophorin B of human erythrocyte membrane. *Infect Immun* **67**: 5332–5337.

Zhao DH, Sun JJ, Liu L, Zhao HH, Wang HF, Liang LQ, Liu LB & Li GF (2009) Characterization of two phenotypes of *Photobacterium damselae* subsp. *damselae* isolated from diseased juvenile *Trachinotus ovatus* reared in cage mariculture. *J World Aquac Soc* **40**: 281–289.