npg

ORIGINAL ARTICLE

Emerging pathogens of gilthead seabream: characterisation and genomic analysis of novel intracellular β-proteobacteria

Helena MB Seth-Smith[1,2], Nancy Dourala[3], Alexander Fehr[2], Weihong Qi[1], Pantelis Katharios[4], Maja Ruetten[2], José M Mateos[5], Lisbeth Nufer[2], Roseline Weilenmann[2], Urs Ziegler[5], Nicholas R Thomson[6], Ralph Schlapbach[1] and Lloyd Vaughan[2]

[1]Functional Genomics Center Zürich, University of Zürich, Zürich, Switzerland; [2]Institute for Veterinary Pathology, Vetsuisse Faculty, University of Zürich, Zürich, Switzerland; [3]Selonda Aquaculture, Athens, Greece; [4]Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Center for Marine Research, Heraklion, Crete, Greece; [5]Center for Microscopy and Image Analysis, University of Zürich, Zürich, Switzerland and [6]Pathogen Genomics, The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK

New and emerging environmental pathogens pose some of the greatest threats to modern aquaculture, a critical source of food protein globally. As with other intensive farming practices, increasing our understanding of the biology of infections is important to improve animal welfare and husbandry. The gill infection epitheliocystis is increasingly problematic in gilthead seabream (*Sparus aurata*), a major Mediterranean aquaculture species. Epitheliocystis is generally associated with chlamydial bacteria, yet we were not able to localise chlamydial targets within the major gilthead seabream lesions. Two previously unidentified species within a novel β-proteobacterial genus were instead identified. These co-infecting intracellular bacteria have been characterised using high-resolution imaging and genomics, presenting the most comprehensive study on epitheliocystis agents to date. Draft genomes of the two uncultured species, *Ca.* Ichthyocystis hellenicum and *Ca.* Ichthyocystis sparus, have been *de novo* sequenced and annotated from preserved material. Analysis of the genomes shows a compact core indicating a metabolic dependency on the host, and an accessory genome with an unprecedented number of tandemly arrayed gene families. This study represents a critical insight into novel, emerging fish pathogens and will be used to underpin future investigations into the bacterial origins, and to develop diagnostic and treatment strategies.
The ISME Journal (2016) **10**, 1791–1803; doi:10.1038/ismej.2015.223; published online 5 February 2016

Introduction

Aquaculture is an expanding industry, growing to meet the consumer demand for high-quality food protein, producing 66 million tonnes globally in 2012 (F.A.O., 2014). Gilthead seabream (*Sparus aurata*) is a high-value fish species, with over 130 000 tonnes farmed globally in 2010 (F.A.O., 2012), and is the dominant species within Mediterranean aquaculture. As intensity of farming and stocking densities increase, aquaculture farms are at ever greater risk from infectious diseases (Segner *et al.*, 2011). Some of the most common diseases to affect gilthead seabream result from bacterial (Vibriosis, Photobacteriosis), viral (Lymphocystis) and parasitic (Monogeneans) infections (F.A.O. (online),

Cultured Aquatic Species Information Programme: Sparus aurata; Colorni and Padros, 2011), and the disease epitheliocystis is now emerging.

Epitheliocystis was first described as cysts in the gill epithelia of the bluegill (Hoffman *et al.*, 1969) and has since been diagnosed within many fish species (Nowak and LaPatra, 2006; Stride *et al.*, 2014). The causative agents of epitheliocystis are commonly from the phylum Chlamydiae (Draghi *et al.*, 2004; Draghi *et al.*, 2007; Karlsen *et al.*, 2008; Schmidt-Posthaus *et al.*, 2012; Fehr *et al.*, 2013; Steigen *et al.*, 2013; Stride *et al.*, 2013a; Stride *et al.*, 2013b), yet recently non-chlamydial bacteria have been implicated (Toenshoff *et al.*, 2012; Mendoza *et al.*, 2013; Contador *et al.*, 2015; Katharios *et al.*, 2015). However, no agents of epitheliocystis have been cultured and to date pathogens have been identified solely through 16S rRNA gene sequencing and *in situ* hybridisation of derived probes to infected tissue sections.

Within gilthead seabream, the disease has previously been investigated off the Red Sea coast

of Israel (Paperna, 1977; Paperna et al., 1981) and Spain (Crespo et al., 1999), including detailed gill pathology, but no molecular work was performed in these studies and no causative agents were identified. Chlamydia-like agents were implicated through analysis of electron micrographs of cysts, yet immunohistochemistry of infected gills with anti-chlamydial LPS antibodies produced no reaction (Crespo et al., 1999).

Epitheliocystis appears to be a seasonal disease in Mediterranean farmed gilthead seabream, occurring during the warmer months and when juvenile fish are first introduced to sea cages. An increasing trend of epitheliocystis has been seen over the past few years (Dr Kantham Papanna, Branch Officer (Greece) of the European Association of Fish Pathologists, personal communication) with the disease also increasing the susceptibility of the host to secondary infections (Dr Panos Varvarigos, VetCare, Athens, Greece, personal communication).

We have studied this disease in several fish farms around Greece, a major aquaculture region (F.A.O. (online), National Aquaculture Sector Overview: Greece). By dismissing previous presuppositions and using rigorous methods, we have identified a novel β-proteobacterial genus as a major cause of epitheliocystis in gilthead seabream. This uncultured emerging pathogen has been characterised molecularly and microscopically from infected gill tissue, with draft genomes providing insights into the diversity and evolution of these new bacteria. The approach we describe could prove invaluable for obtaining detailed morphological and genomic data from other novel uncultured infectious or symbiotic microorganisms.

Materials and methods

Sample collection
Gilthead seabream gill samples were obtained from Selonda SA, a pioneering Greek aquaculture company currently producing over 20 000 tonnes gilthead seabream per year from over 100 million juveniles. Samples from morbid fish were taken at three sites and at three time points: November 2012 (Argolida, Saronikos) samples were taken randomly during a low-level infection, whereas samples from 2013 (Argolida, June, numbered 11–45 reflecting cage number in the first digit, and Arkadia, October, 2 cages) were selected during epitheliocystis outbreaks (Table 1). Infection levels were determined on the basis of the number of cysts per gill arch, number of gill arches infected, number of fish infected per cage and mortality attributed to the disease. Mortality was estimated through the pathology of dead fish retrieved every few days from the bottom of the cages. Gill arches from individual fish were taken in parallel into 10% buffered formalin, RNALater and pure ethanol, with the 2013 samples also taken into M4RT Chlamydia transport medium (Remel Microtest M4RT, Thermo Fisher, Reinach, Switzerland) and sterile sea water (SSW). Samples were transported cooled overnight to Zurich for processing.

Sample processing
Microscopy. Formalin-fixed samples were embedded in paraffin (FFPE) after dehydration in ascending alcohol concentrations ending in xylol, with 3 μm sections mounted on positively charged glass slides and used for routine haematoxylin and eosin (HE) staining, periodic-acid-Schiff staining, and fluorescent in situ hybridisation (FISH, described below).

Culture. On arrival in Zurich, chilled samples in M4RT or SSW were microdissected into gill filaments or single cysts using a Leica M165C dissecting microscope. Cysts from SSW were washed briefly in cold PBS with 50 μg ml^{-1} vancomycin and 1 μg ml^{-1} amphotericin B. Whole or homogenised cysts were placed onto monolayers of fish epithelial cell line EPC (Winton et al., 2010) in minimal essential medium (MEM) containing 10% fetal calf serum (Gibco, Life Technologies, Reinach, Switzerland) or Acanthamoeba castellanii ATCC 30010 in peptone yeast extract glucose (PYG) broth (Greub et al., 2004) in 24-well trays and incubated for up to 1 month at 28 or 32 °C, respectively. Cultures were checked regularly for putative intracellular bacterial growth by phase contrast microscope. Marine broth 221 and marine agar 221 (Himedia, Mumbai, India)

Table 1 Sampling of gilthead seabream and gill pathology

Location	Date	# Fish sampled	Fish size (mean of each cage)	Cage fish mortality	Infection/cyst description	Co-infection
Argolida	Nov 2012	2	ND	0%	Low infection	
Saronikos	Nov 2012	5	ND	0%	Low–medium infection	Ciliate *Trichodina*
Argolida	Jun 2013	20 from 4 sea cages	15–23 g	1.3–15.5%	High infection	Amoebae *Trichodina* Lymphocystis Fungi Flukes
Arkadia	Oct 2013	20 from 2 sea cages	9–11 g	2.9–3.1%	Medium-high infection	Flukes

Abbreviation: ND, not determined.

were used to culture samples with incubation at 22 °C for up to 2 weeks.

DNA extraction. Qiagen DNeasy Blood and Tissue kit (Qiagen, Hombrechtikon, Switzerland) was used on gill fragments or filaments stored in RNALater or ethanol, eluting into 100 µl HPLC water.

Bacterial identification
Identification of bacteria used either Chlamydiae-specific (Everett *et al.*, 1999; Draghi *et al.*, 2004) or universal bacterial 16S rRNA gene primers (Weisberg *et al.*, 1991). Positive bands were analysed only when template-free negative controls showed no signal. 16S rRNA gene amplicons were sequenced (Microsynth, Balgach, Switzerland) directly from the PCR product or after cloning into Topo vector pCR2.1 (Thermo Fisher). The resulting reads were assembled (CLC Main Workbench 7.0.2 (CLC bio, Qiagen) and Sequencher 5.2.4 (Gene Codes Corporation, Ann Arbor, MI, USA), compared using blastn against the Genbank database, and analysed using Muscle aligner and PhyML v3 maximum likelihood phylogeny software within Seaview v4 (Gouy *et al.*, 2010). Representative 16S rRNA gene sequences have been deposited with EMBL under accession numbers: LN612726-LN612730 representing 2013Arg42i, 2013Ark19i, 2013Arg22i, 2013Arg32i, 2012Sar4i, respectively (i indicating *Ca.* Ichthyocystis sequences).

Quantitative PCR
A modified Chlamydiae-specific Taqman quantitative PCR (qPCR) protocol (Lienard *et al.*, 2011) (forward: 5′-CCGCCAACACTGGGACT-3′, reverse: 5′-GGAGTTAGCCGGTGCTTCTTTAC-3′, probe: 5′-FAM-CTACGGGAGGCTGCAGTCGAGAATC-BHQ-3′) was used. A qPCR protocol designed for the *Ca.* Ichthyocystis genus (forward: 5′-CAAGGCGA CGATCGGTAGCTG-3′, reverse: 5′-TTACAACCC TAAGGCCTTCTTCACC-3′, probe: 5′-FAM-TTG CTGGATCAGGCTTCCGCCCATTGTCCAAA-BHQ-3′) allowed quantification of high concentrations of bacteria within infected material. All qPCR protocols were validated against serial dilutions of amplicons from target and control species. Reactions were carried out on a StepOne Plus real-time PCR system (Applied Biosystems, Reinach, Switzerland) using Taqman Fast Advanced reagents (Applied Biosystems) according to the manufacturer's instructions with 1–5 µl DNA in a reaction volume of 20 µl.

Fluorescent staining and imaging
FISH was performed on the Ventana Discovery XT automated platform using probes given in Supplementary Table S1. Automated deparaffinisation was followed by pretreatment with 0.2 M HCl in PBS then pepsin (500 µg ml^{-1} in this buffer) for 4 min. Following addition of 50 ng probe per slide

in a hybridisation buffer comprising 6× SSC, 5 × Denhardt's solution and 12% dextran sulphate, samples were denatured at 90 °C for 4 min and hybridised at 48 °C overnight. A wash with 2 × SSC at 48 °C was followed by manual post staining with 4′-6-diamin-2-phenylindole (DAPI) at 10 µg ml^{-1} to visualise bacterial and host DNA, and concanavalin A-AlexaFluor488 or concanavalin A-Alexa-Fluor546 (25 µg ml^{-1}) (Life Technologies) for 10–30 min. Negative controls included sections of fish tissue not infected with *Ca.* Ichthyocystis, as determined by PCR and qPCR. Immunofluorescent (IF) staining with anti-sodium-potassium-ATPase antiporter α5 (Developmental Studies Hybridoma Bank, University of Iowa) and AlexaFluor546-labelled secondary antibody (Life Technologies) used standard antigen retrieval in citrate buffer, or rehydrated sections permeabilised with 0.1% Triton X-100 in PBS with anti-OxPhos complex IV subunit I mouse antibody (Life Technologies) diluted in 2% BSA and AlexaFluor488-labelled secondary antibody (Life Technologies). IF slides were counterstained with DAPI as above. Sections were scanned on a Hamamatsu Nanozoomer 2.0HT scanner for an overview, and high-resolution imaging used a confocal laser scanning microscope (SP5, Leica Microsystems, Heerbruch, Switzerland). Deconvolution was performed using Huygens (Scientific Volume Imaging, The Netherlands) (Ponti *et al.*, 2007), and images were prepared with Imaris 7.6.1 (Bitplane, Oxford Instruments, Zurich, Switzerland) and Photoshop CS4 extended, version 11.0.2 or CS6 extended, version 13.0x32 (Adobe, San Jose, CA, USA).

Electron Microscopy (EM)
Samples for scanning EM (SEM) were fixed in 2.5% glutaraldehyde in 0.2M sodium cacodylate buffer, pH 7.4, then washed with 0.1M sodium cacodylate buffer, pH 7.4, post fixed with 1% osmium tetroxide (OsO$_4$), dehydrated in an ascending alcohol series, mounted on stubs, and sputter coated with gold–palladium. They were viewed using a JEOL JSM-6390LV scanning electron microscope at 15 kV at the Electron Microscopy Laboratory of the University of Crete, Heraklion.

Samples for transmission EM (TEM) were fixed in 2.5% glutaraldehyde and 1% OsO$_4$ buffered with 0.1 M sodium phosphate (pH 7.4). Samples were dehydrated in an ethanol series, embedded in Epon 812 resin, and ultrathin (90 nm) sections were stained with uranyl acetate and lead citrate. Images were acquired using a Philips CM10.

Sample blocks for focused ion beam-SEM (FIB-SEM) were fixed in 2.5% glutaraldehyde in 0.2 M sodium cacodylate buffer, pH 7.4, followed by 1% OsO$_4$, and contrasted with 2% aqueous uranyl acetate. These were dehydrated in an ethanol series, propylene oxide and embedded in Epon 812 resin. Semithin and ultrathin sections were obtained

to identify the regions containing cysts. Selected blocks were attached to 12 mm stubs by conductive carbon cement followed by carbon coating. Three-dimensional data sets were acquired with a FIB-SEM Auriga 40 Crossbeam (Zeiss, Oberkochen, Germany) using the FIBICS Nanopatterning engine (Fibics Inc., Ottawa, Canada). The gallium-ion beam for milling was used at 30 kV, 600 pA current and the images were acquired at an acceleration voltage of 1. 5 kV using an in-lens energy selective backscattered electron detector (ESB) with a grid voltage of 1.3 kV. The resolution was set to 5 nm in the XY axes and 5–10 nm in the Z axis. Image stacks were aligned with TrackEM2 (Cardona et al., 2012). The aligned data set was visualized with Imaris 7.6.1 and segmented with the software Ilastik, 1.1 (Sommer et al., 2011).

Genomics

Cysts from fresh or preserved (RNALater, ethanol, M4RT) gill samples were micromanipulated to remove the majority of host material using a Leica M165C dissecting microscope. DNA was extracted from cysts from individual gills using the Qiagen DNeasy Blood and Tissue kit. Some samples (Supplementary Table S2) were subject to host DNA depletion using the NEBNext Microbiome DNA Enrichment Kit (NEB, MA, USA). Aliquots (1 µl) were subject to Genomiphi V2 multiple displacement whole-genome amplification (MDA) (GE Life Sciences, Glattbruch, Switzerland) and the amount of target DNA in pre- and post- amplification samples determined using qPCR.

Samples were sequenced on the Illumina Miseq platform at 6- to 12-plex with 250 bp paired end reads following Nextera library creation. Raw sequencing reads were adaptor and quality trimmed and filtered using Trimmomatic version 0.32 (Bolger et al., 2014). Resulting fastq files were screened by mapping with SMALT (http://www.sanger.ac.uk/resources/software/smalt/) against the 16S rRNA gene sequences of candidate epitheliocystis pathogens, to assess pathogens present and estimate genome coverage. The data was screened against gilthead seabream mitochondrial 16S rRNA gene sequence (EMBL accession number AF247432) and gilthead seabream 18S rRNA gene sequence (as identified from a preliminary assembly) to confirm sample origin and determine levels of fish DNA contamination.

De novo assembly of quality controlled reads from 2013Ark11 (Ca. I. hellenicum; amplified) and 2013Arg41 (Ca. I. sparus; not amplified) was performed using SPAdes v 3.1.0 (Bankevich et al., 2012). Assemblies were tested using various K-mers in both single-cell mode and multi-cell mode (Nurk et al., 2013) and quality assessment using QUAST (Gurevich et al., 2013) and CGAL (Rahman and Pachter, 2013) (Supplementary Tables S3 and S4). For these data sets, multi-cell mode produced better assemblies than single-cell mode, giving 2743

scaffolds (4.06 Mb) for 2013Ark11 and 1794 scaffolds (3.37 Mb) for 2013Arg41.

Removal of scaffolds representing DNA from the host or other bacterial species was performed initially using SPAdes kmer coverage parameters to define a cutoff above which scaffolds were most likely to belong to the target bacterium (2013Ark11 cutoff × 35; 2013Arg41 cutoff × 30). In addition, scaffolds below 1 kb were discarded as being unlikely to provide valuable information to the assembly. This resulted in 98 scaffolds (2.16 Mb) for 2013Ark11 and 198 scaffolds (2.53 Mb) for 2013Arg41.

Automated annotation using Prokka (Seeman, 2014) placed ~ 1600–1800 coding sequences (CDSs) per genome. These were manually checked against Prodigal gene prediction (Hyatt et al., 2010) using Artemis (Rutherford et al., 2000) and ACT (Carver et al., 2005), with a six-frame tblastx search used to investigate additional hits. Further scaffolds were discarded when no bacterial CDSs were identified. Annotation was curated manually in Artemis using blastp identities, Pfam, Rfam, SignalP, TMHMM and ncoils and further analysed using KEGG Ontology (Kanehisa and Goto, 2000) performed using BlastKOALA (Kanehisa et al., 2010) with KEGG module evaluation using KEGG Mapper (Kanehisa et al., 2012). Iterative Hidden Markov Model (HMM) searches of gene family members used hmmer-3.1 (Eddy, 2009). For each gene family, HMMs were built based on protein alignments and used to scan assemblies (reporting e-value cutoff 0.0001) iteratively until the set of protein matches stabilized. HMMs were used to scan protein databases NCBInr (http://www.ncbi.nlm.nih.gov/), Uniprot (www.uniprot.org) and the RSCB PDB (http://www.rcsb.org/pdb). Type 3 Secretion System effectors were predicted using Effective (Jehl et al., 2011). Whole genome phylogenies were constructed by mapping sequence data using SMALT against the percentage of conserved proteins (POCP)-defined core genome of the closest reference assembly, determining SNP locations and running PhyML phylogeny from the resulting pseudo-alignment.

Read data for all 10 samples sequenced (Supplementary Table S2) have been submitted to the European Nucleotide Archive (ENA, http://www.ebi.ac.uk/ena/) under study PRJEB7439 and accession number ERR361036 for sample 2013Arg42. The draft genomes have been submitted under this study number.

Results

Sampling and pathology of gilthead seabream

Epitheliocystis is an emerging disease in farmed gilthead seabream in the Mediterranean; data from a commercial aquaculture firm show year-on-year increases in mortality attributed to epitheliocystis (Supplementary Figure S1). Despite being a short-lived

infection lasting ~ 2 weeks, mortality rates of up to 20% occur in affected sea cages.

To determine the severity of the epitheliocystis, and the associated aetiological agents, morbid gilthead seabream were sampled from commercial fish cages at three coastal locations around Greece on three occasions over a year: Argolida (sampled November 2012 randomly and June 2013 during an epitheliocystis outbreak), Arkadia (sampled October 2013 during an outbreak) and Saronikos (sampled November 2012 randomly) (Table 1).

Gill samples were screened for disease (Supplementary Figure S2) and gill arches from individual fish diagnosed with epitheliocystis were stored variously for preservation of morphology, nucleic acid and microbial viability. Histopathology demonstrated the level of epitheliocystis infection, cyst characteristics and presence of co-infecting pathogens (Table 1). Two major cyst types were identified in the fish gills, all located on the tip or on the base of the secondary lamellae (Figures 1a and c). These large cysts, found in high numbers during outbreaks, are likely to affect fish respiration due to epithelial hyperplasia, causing disruption of oxygen diffusion and distress to the host. The first type are up to 100 μm in diameter, have basophilic and finely granular content (Figure 1b) enclosed by a fine 1–2 μm thick eosinophilic membrane (Figure 1d). A layer of epithelial cells often surrounds them, occasionally hyperplastic (Figure 1b). Ruptured cysts can be seen to be surrounded by macrophages, lymphocytes and occasionally neutrophils (Figure 1e). Mild inflammation is often seen around the cysts, and secondary lamellae can be blunted and fused. The second type of cyst is smaller, up to 40 μm in diameter with more basophilic and coarsely granular contents surrounded by a slightly thicker 2–3 μm membrane (Figure 1c). Co-infections with other pathogens were observed in rare cases (Table 1) but these would not contribute to the disease status of the fish.

Pathogen identification
Chlamydiae are the most commonly identified agents of epitheliocystis, and Chlamydiae-specific 16S rRNA gene PCR on gill tissue extracts from Saronikos (November 2012) and Argolida (June 2013) identified many novel sequences belonging to the *Piscichlamydiales* clade. However, exhaustive FISH analysis could not localise these gene sequences to the large cysts identified as the primary pathological lesions (Supplementary Information and Supplementary Figure S3). Amplification of DNA from an isolated cyst (fish 2013Arg42) using universal primers provided two identical cloned sequences representing the 16S rRNA gene of a new genus of β-proteobacterium sharing 89.9% nucleotide identity with the closest previously deposited sequence (Uncultured *Sterolibacterium* sp. clone OTU-X1–28, accession JQ668544), and 89.2% with any cultured bacterium (*Sterolibacterium denitrificans* strain Chol-1S, AJ306683
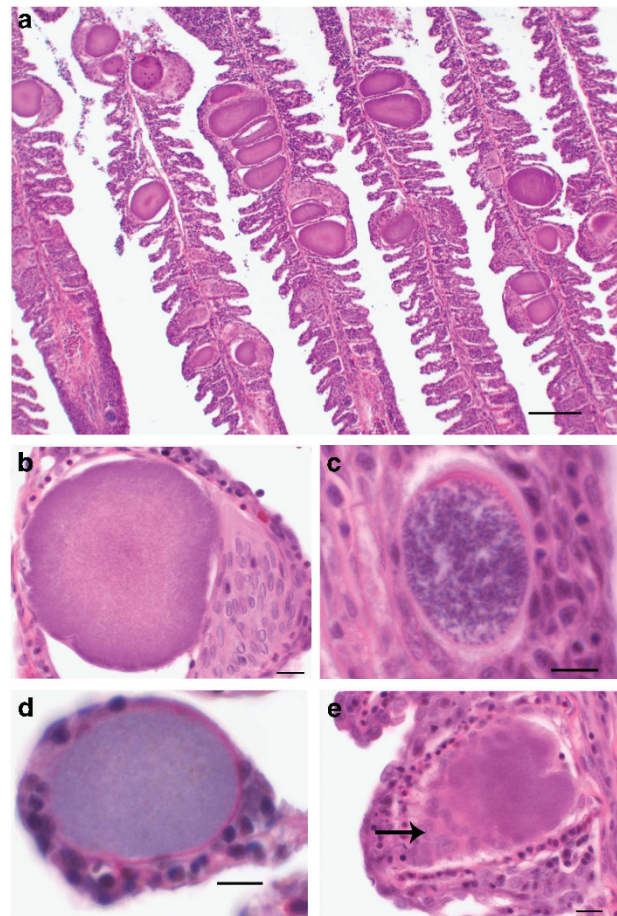

Figure 1 Sections of representative gills stained for histopathology. (**a**) Overview of gill from fish 2013Arg11 with haematoxylin and eosin (HE) staining. The gills are heavily infected with cysts, up to 100 μm in diameter. Scale bar, 100 μm. (**b**) Cyst from fish 2013Arg11 with HE staining, located between two fused secondary lamellae. The cells towards the tip of the lamellae can be seen to be atypically thickened and hyperplastic through epithelial proliferation with squamous metaplasia. The epithelial cells have a large amount of eosinophilic cytoplasm and are tightly connected to each other. The gap around the cyst is thought to be a fixation artefact through shrinkage. Scale bar, 10 μm. (**c**) Cyst from fish 2013Arg11 under HE staining showing coarser granular contents and a thicker eosinophilic membrane. Scale bar, 10 μm. (**d**) Cyst from fish 2013Ark11 under periodic-acid-Schiff staining, showing polysaccharide containing membrane (pink). Within the membranous layer, an elongated nucleus with finely stippled chromatin is visible. Scale bar, 10 μm. (**e**) Cyst from fish 2013Arg11 with HE staining showing a ruptured cyst and associated macrophages (arrow). Scale bar, 10 μm.

(Tarlera, 2003)). (New genera are described by 86.5–94.5% identity between 16S rRNA gene sequences (Yarza *et al.*, 2014)). Phylogenetic analysis (Figure 2) reveals relatedness to two recently described epitheliocystis pathogens: from farmed Atlantic salmon (*Salmo salar*), *Candidatus* Branchiomonas cysticola (Toenshoff *et al.*, 2012) (88% nucleotide identity); and from lake trout (*Salvelinus namaycush*), clone BJC-BK (Contador *et al.*, 2015) (87% nucleotide identity). Further sequences representing this novel genus were amplified and sequenced from samples from fish at each location
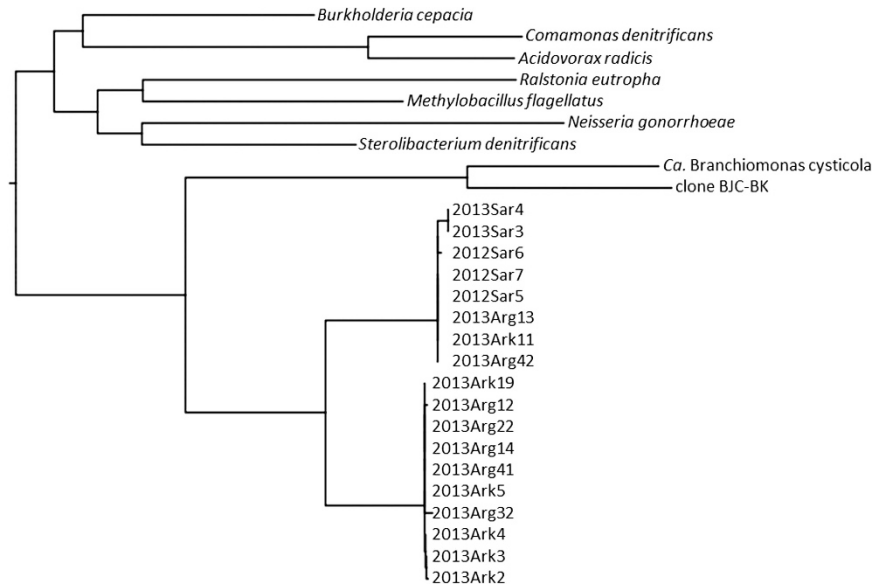
Figure 2 Phylogenetic analysis of novel 16S rRNA gene sequences from gilthead seabream. Amplified 16S rRNA genes provided up to 1500 bp sequence. The alignments and phylogeny were performed using Muscle and PhyML within Seaview with 100 bootstraps. Two distinct clades indicate that there is considerable diversity within the novel genus (*Ca.* Ichthyocystis). Other β-proteobacteria are included as outgroups, with the epitheliocystis agents *Ca.* Branchiomonas cysticola (Toenshoff *et al.*, 2012) and clone BJC-BK (Contador *et al.*, 2015) as the species sharing a common ancestor with the novel bacteria. 2012Arg, 2013Arg, 2013Ark and 2012Sar refer to the year and location of isolation, followed by the fish number. Branches to and between the novel sequences have bootstrap values in excess of 99%.

and time point (Table 1). Phylogenetic analysis shows that they fall into two distinct species within the novel genus, sharing only 95.2% sequence identity (species distinction defined at 94.5–98.7% 16S rRNA gene nucleotide identity (Yarza *et al.*, 2014)).

Using qPCR to determine bacterial load, low levels of chlamydial sequence were quantified in all samples (mean 2701 copies per µl, range 30–23 400), but levels of the new pathogen genus were higher in every case (mean 1 324 651 copies per µl, range 2415–10 470 000), particularly in the heavily infected samples from 2013Arg and 2013Ark (up to 3000 × higher, Supplementary Table S5). Bacterial load values cannot be compared between samples, as DNA extraction was not performed on standardised material. An approximate mean figure of 400 000 16S rRNA gene copies per cyst was determined from analysing extracts of multiple microdissected cysts.

Culture was attempted unsuccessfully onto fish epithelial cell monolayers, in *Acanthamoeba,* in marine broth and on marine agar.

Localisation of bacteria in infected tissue
FISH was performed on sections from 18 fish, producing very strong signals in the large cysts using probes against the novel bacterial genus (Figure 3) compared with no signal against *Ca.* Piscichlamydia clade probes (Supplementary Figure S3) and control probes targeting more diverse bacteria (E-474, data not shown). These novel pathogens are therefore the dominant cause of epitheliocystis in these gilthead seabream around the Greek coastline, over at least

1 year. As uncultured, novel pathogens, the two species were named *Candidatus* Ichthyocystis hellenicum and *Candidatus* Ichthyocystis sparus (see Taxonomy).

Further FISH using probes to discriminate between the two species screened gill sections from 16 fish. Both species can be identified in the same gill sections indicating that they are co-circulating pathogens (Figures 4a and b) in the heavily infected Argolida 2013 and Arkadia 2013 samples, whereas only *Ca.* I. hellenicum could be identified in 2012 samples (Supplementary Table S2). Serial sections stained with HE and FISH show that the smoother cysts contain *Ca.* I. hellenicum (Figure 1b) and the more granular cysts contain *Ca.* I. sparus (Figure 1c). Individual labelled bacterial particles can be seen within each cyst (Figures 4c and d).

Epitheliocysts are often associated with mitochondrion rich chloride cells (Paperna and Alves Dematos, 1984; Crespo *et al.*, 1990), heavily involved in osmoregulation and ion transport. Fluorescent labels to target mitochondria (anti-oxphos complex IV antibody) and ion channels (anti-sodium–potassium antiporter antibodies) found no association between cells containing mitochondria, or chloride cells, and those infected with *Ca.* Ichthyocystis (data not shown).

Ca. Ichthyocystis species distribution
We investigated the geographical and temporal distribution of the two *Ca.* Ichthyocystis species in Mediterranean gilthead seabream using FISH and 16S rRNA gene sequencing (Supplementary
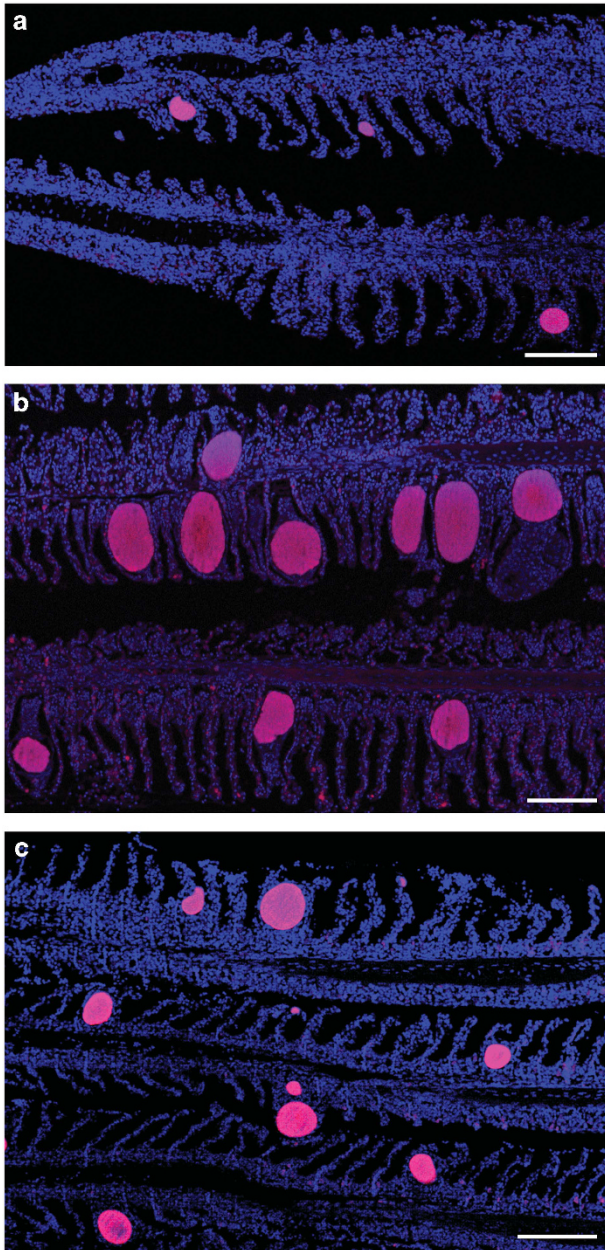
Figure 3 Images showing the hybridisation of probes targeting novel bacterial sequences to the large cysts. Gill sections from (**a**) 2012Sar5; (**b**) 2013Arg12; (**c**) 2013Ark11. All sections are labelled with Ichthyo290-Cy3 (red) and counterstained with DAPI for DNA (blue). Scale bar, 100 μm.

Figure 4 Fluorescent images distinguishing between *Ca.* I. hellenicum and *Ca.* I. sparus species. (**a**, **b**) Serial sections of fish 2013Arg11 probed using (**a**) Ichthyo290-Cy3 probe against the whole novel genus or (**b**) Ichthyo230-Cy3 probe, possessing five mismatches to the *Ca.* I. sparus 16S rRNA gene sequence, thus targeting just *Ca.* I. hellenicum. *Ca.* I sparus cysts can be identified as being labelled in **a** and unlabelled in **b** (arrows). (**c**, **d**) Deconvolved images of typical cysts hybridised to probe Ichthyo290-Cy3. (**c**) *Ca.* I. hellenicum containing cyst from fish 2013Arg42; (**d**) *Ca.* I. sparus containing cyst from fish 2013Arg23. All FISH probes are labelled with Cy3 (red) and all sections are counterstained with DAPI for DNA (blue). Scale bars, 50 μm (**a**, **b**), 10 μm (**c**, **d**).

Table S2). In the 2012Arg and 2012Sar samples, only *Ca.* I. hellenicum can be detected. These samples also show a low load of *Ca.* Ichthyocystis by qPCR, and only 1–10 cysts can be observed per section. FISH on 2013Arg and 2013Ark gill arches show the presence of both species, with cysts containing *Ca.* I. hellenicum more numerous in all samples tested. However, PCR of these gill arches produces sequences from *Ca.* I. sparus more often than *Ca.* I. hellenicum, with little correlation with FISH results (Su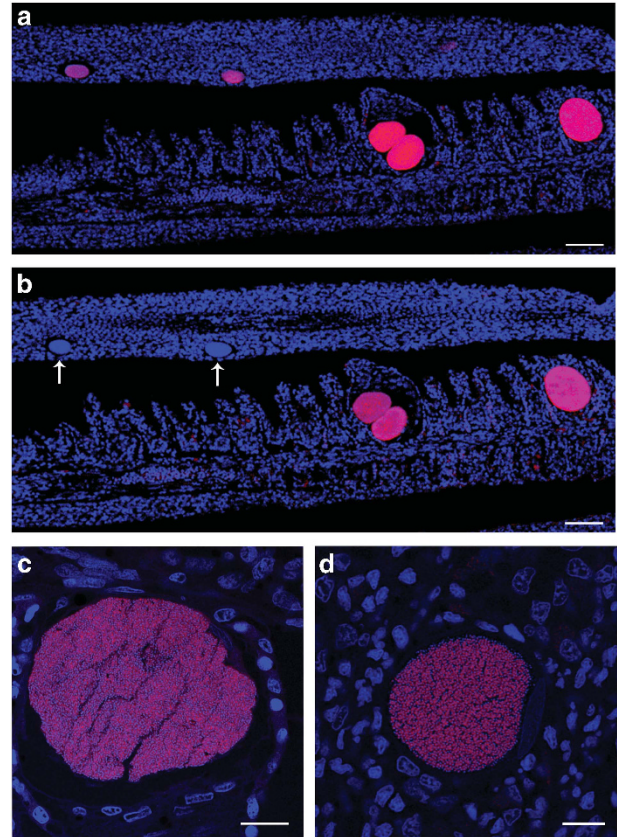pplementary Table S2). As different gill arches from the same fish were used in the analyses, this indicates that adjacent gill arches contain different ratios of the two novel species, but that both species are present in the same locations, often co-infecting the same host.

Description of Ca. Ichthyocystis epitheliocysts
Further imaging of the cysts gave deeper insights into these uncultured bacteria and their interaction with the host tissue. SEM of 2012Sar samples show cysts bulging through the epithelium (Figures 5a and b), with hyperplastic gills and fused lamellae (Figure 5a) coated in an excess of mucus (Figure 5b).

TEM was performed on sections from 14 fish (Supplementary Table S2) and three-dimensional imaging, using FIB-SEM, allowed specific analysis of a sample in a chosen plane that aids the visualisation of specific structures. EM images
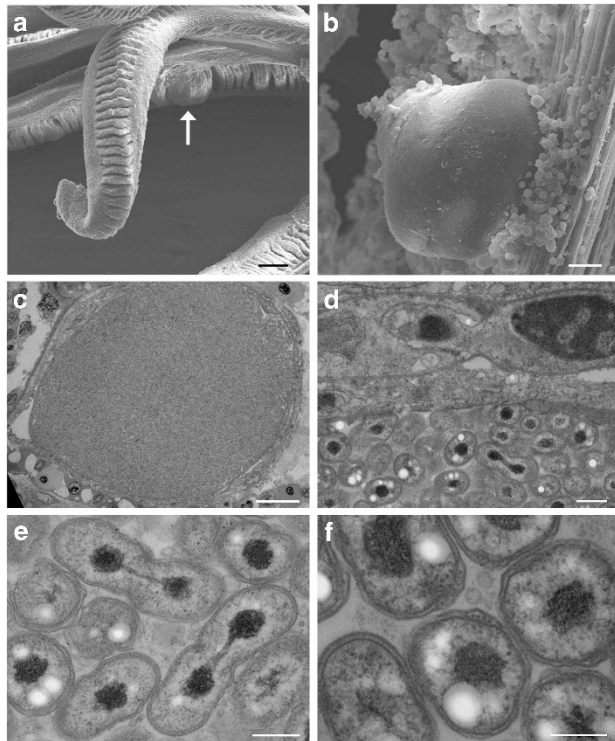
Figure 5 SEM and TEM images of infected gills. (**a**) SEM of gill filaments from 2012Sar showing hyperplasia. A cyst is indicated (arrow). Scale bar, 100 μm. (**b**) SEM of a cyst from a 2012Sar emerging and surrounded by mucus. (**c**) Overview of cyst from fish 2013Arg14, anchored to the primary lamella and embedded between two secondary lamellae, with nucleated erythrocytes visible in the capillary running up the centre of the secondary lamellae. (**d**) Edge of a cyst from fish 2013Arg13, showing the interdigitated epithelial cells, with no continuous membrane visible separating the bacterial compartment from the host cell cytoplasm. (**e**) Dividing bacteria within a cyst from fish 2013Arg13, with the partitioning DNA still connected. Numerous vesicles are present between the bacteria. (**f**) High-resolution image from fish 2013Arg12, showing the bacterial double membrane, between which rows of periplasmic small electron dense particles can often be seen. Large electron lucent vesicles within the bacterial cytoplasm are common. Also in this image, multiple vesicles between bacteria are visible. Scale bars, 100 μm (**a**), 10 μm (**b** and **c**), 0.5 μm (**d**), 0.3 μm (**e**) and 0.2 μm (**f**).

Figure 6 Images selected from FISH and FIB-SEM data of a cyst from fish 2013Arg42. (**a**) *Ca.* I. hellenicum from fish 2013Arg23 hybridised with Ichthyo-290 (red) and subsequent concanavalin A-Alexa488 staining (green) for glycoprotein membrane proteins on plasma membranes, counterstained with DAPI for DNA (blue). Membranes can be seen completely enveloping the cyst. The wave-like patterns within the cyst result from sectioning artefacts. (**b**) Edge of a cyst showing interdigitating epithelial cells (arrow) and desmosomes connecting cell processes (asterisk). (**c**) A dividing bacterium (centre). Some bacteria contain a dense body (arrows) not connected to the major body of suspected nuclear material. (**d**) The clearly defined double membrane is visible, connected to small vesicles (arrow). (**e**) A long bacterium is indicated (arrow). (**f**) Small clear vacuoles are present in many bacteria (arrows). Scale bars, 20 μm (**a**); 500 nm (**b,c,e,f**); 200 nm (**d**). Videos of the three-dimensional data of Figures 6b and c are presented in Supplementary Videos S1 and S2, respectively.

cannot distinguish between cysts containing *Ca.* I. hellenicum or *Ca.* I sparus.

There is no evidence of a *Chlamydia*-like developmental cycle, with replicating bacteria towards the outside of the cyst and differentiated infectious forms in the middle, as similar bacterial forms are seen throughout the cyst (Figures 5c and d). The membranes surrounding the cysts seen under histological (Figure 1d) and concanavalin A (Figure 6a) staining for host cell membranes can be seen in EM images to consist of interdigitating epithelial cell processes, with cell membranes tightly connected by desmosomes (Figures 5d and 6b; Supplementary Video S1). The bacteria-containing compartment is clearly differentiated from the host cell cytoplasm, although a continuous membrane barrier is not visible. We were not able to
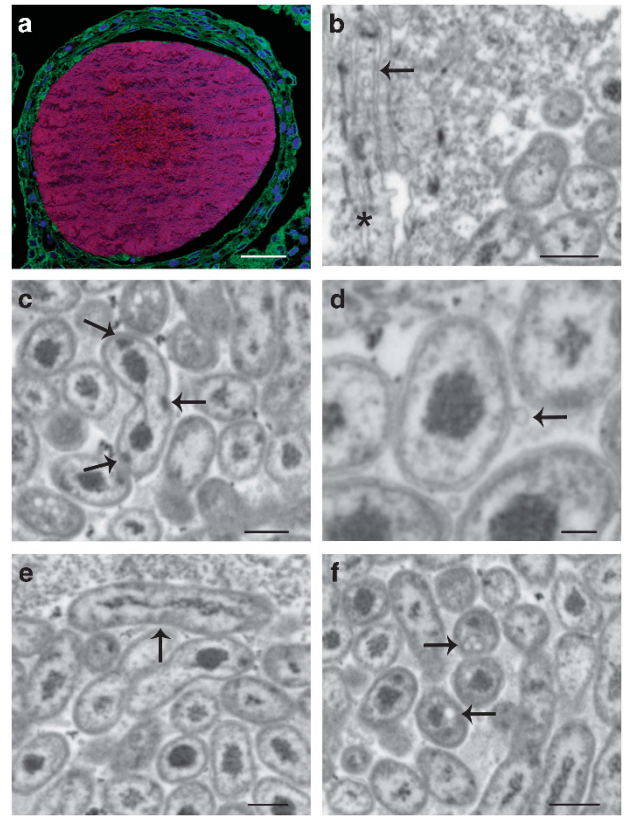
determine whether the cyst is intracellular within a single epithelial cell which then becomes enveloped by additional cells, or whether multiple epithelial cells combine to enclose the developing cyst.

The bacteria are slightly elongated with a diameter of 0.5 μm and a length of 0.7 μm (Figures 5e, f and 6c). A double membrane surrounds each bacterium (Figures 5f and 6d), which may bud off into the small vesicles found in the cyst matrix (Figures 5e, f and 6d). Four forms of the bacteria are visible: with a centrally located dense nucleoid (Figures 5e, f, 6d and e); in the process of dividing, showing the nucleoids connected by a thin bridge (Figures 5e and 6c, Supplementary Video S2); long

bacteria (Figure 6e); and with a darker cytoplasm and small clear vacuoles (Figures 5e, f and 6f). The use of FIB-SEM allowed visualisation of many dividing bacteria possibly indicating rapid growth of the cyst (Supplementary Video S2). Some bacteria also carry an additional dense body adjacent to the membrane, separate from the putative nucleoids (Figure 6c).

Genomes of Ca. Ichthyocystis

Genomes were obtained directly from heavily infected material stored in RNALater or ethanol. Ten DNA extracts from micromanipulated individual or pooled cysts were subjected to whole-genome sequencing with or without host DNA depletion and MDA (Supplementary Table S2). Data from all samples mapped to the *Ca.* Ichthyocystis 16S rRNA gene, with no coverage of chlamydial sequences, confirming that these large cysts contain *Ca.* Ichthyocystis genus bacteria.

Two samples were chosen for *de novo* assembly, representing *Ca.* I. hellenicum (2013Ark11, estimated 130× coverage) and *Ca.* I. sparus (2013Arg41, estimated 240×). The draft genomes are relatively small for β-proteobacteria at ∼ 2.3 Mb (Supplementary Table S6; mean β-proteobacterial genome size 5.1 Mb, range 0.2–9.7 Mb (Genome Atlas Database: Betaproteobacteria (online)).

By comparing the assemblies, scaffolds were rearranged to determine a putative shared core genome for the genus (Figure 7) comprising ∼ 1.5 Mb (65–69% of the full assemblies respectively), with ∼ 70% average nucleotide identity (ANI), represented by 17–20 scaffolds in the two samples. As such, *Ca.* I. hellenicum and *Ca.* I. sparus can be defined as separate species, using the definition that members of the same species should share over 69% genome coverage at 95% ANI (Goris *et al.*, 2007). Furthermore, a POCP method with 50% cutoff (Qin *et al.*, 2014) places these species within the same genus, sharing 56.6% of their CDSs. The sequence data from the remaining samples was mapped against the core genes: the comparative phylogeny mirrors the 16S rRNA gene phylogeny and gives an indication of the high diversity of this genus within Mediterranean epitheliocysts (Supplementary Figure S4).

We used KEGG analysis to look at the functions encoded within the draft genomes of 2013Ark11 and 2013Arg41, both of which gave highly similar results. Genetic information processing pathways (transcription, translation, replication and repair) are complete, as are central metabolic pathways (carbohydrate, lipid, nucleotide, cofactor and vitamin metabolism), and aminoacyl-tRNA synthases, giving confidence that we have almost entire genomes. However, very few amino-acid biosynthesis
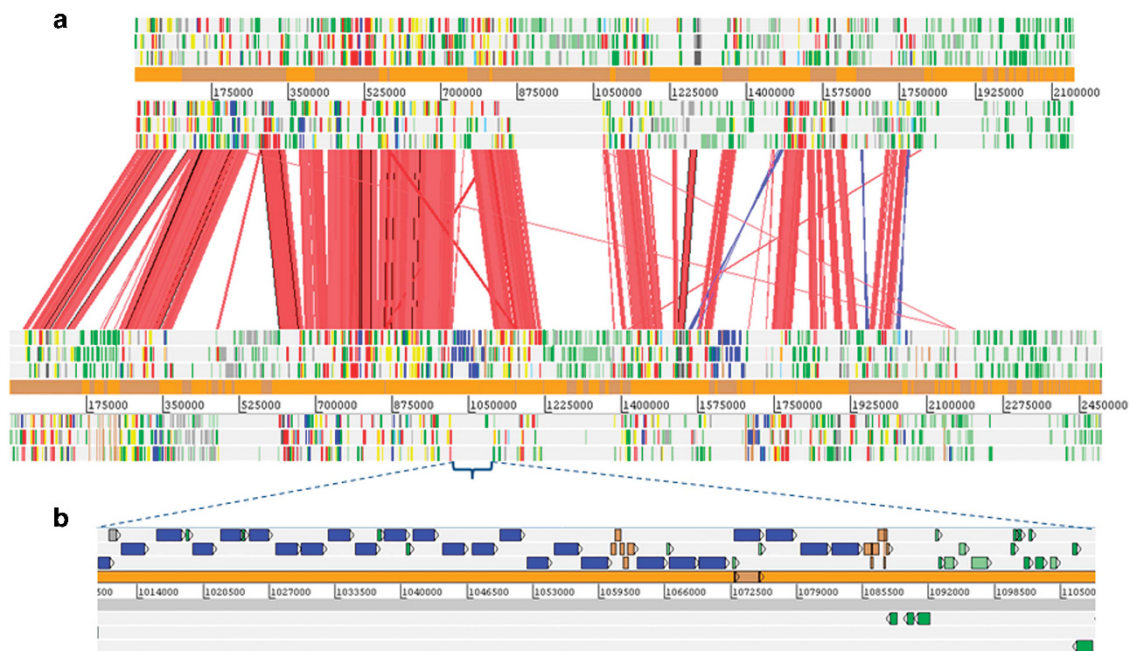

Figure 7 Comparison of *Ca.* I. hellenicum and *Ca.* I. sparus draft genomes. (**a**) Scaffolds from draft genomes of 2013Ark11 (top) and 2013Arg41 (bottom) were ordered against each other, compared using blastn and visualised using ACT. Orange and brown strips along the genome indicate the scaffolds, numbered in bp along the genome. Deep-red bars between the genomes indicate high levels of nucleotide identity (>70%) in the core regions of the genome of this genus, which get paler as nucleotide identity falls. Blue bars indicate inversions. Sections with no matches indicate regions of difference between the genomes. Note that within the ordered contigs, the genomes contain few rearrangements. Smaller scaffolds on the right are unique to each species. Manually annotated CDSs are indicated for each genome in the six forward and reverse frames, with colour coding for CDS functions: dark blue for pathogenicity/adaptation, dark grey for energy metabolism, red for information transfer, dark green for surface associated, yellow for central/intermediary metabolism, pale green for hypothetical proteins, pale blue for regulators, orange for conserved hypothetical proteins, brown for pseudogenes, pink for phage and IS elements and grey for miscellaneous. (**b**) Region unique to *Ca.* I. sparus showing an array of CDSs representing the ShET2 enterotoxin domain containing gene family (dark blue).

pathways are complete, implying that these are obligate intracellular bacteria which scavenge amino acids from their hosts using amino-acid transport systems which have been identified. Both genomes also harbour some of the structural subunits of Type Two (T2SS) and Three (T3SS) Secretion Systems.

A prominent genomic feature in these species is the occurrence of expanded families of genes, first observed as large regions of difference between the two genomes (Figure 7). We have identified 433 CDSs in 2013Ark11 (28% of all annotated CDSs) and 544 in 2013Arg41 (31%), belonging to 30 novel gene families, many of which occur within tandem arrays. The genomes of each species contain nine unique families, with 12 shared families. The largest family contains up to 200 members, indicating that the duplication and diversification of these families is a significant evolutionary strategy in this genus. Very few of the families have any putative function assigned to them: one with ~96 family members in the *Ca.* I. sparus genome, but absent from *Ca.* I. hellenicum, contains the Pfam PF07906 domain related to the ShET2 enterotoxin N-terminal domain (Figure 7). Many of the proteins encoded by these families are predicted to be secreted and may be effectors of the identified secretion systems.

No insertion sequences or phages were found within these genomes, and very few pseudogenes have been identified. Putative Major Facilitator Superfamily transporters have been found in both species, although there is no other evidence of the acquisition of antibiotic resistance within this genus. A detailed genome description is being prepared for separate publication.

Discussion

We have identified and characterised novel emerging pathogens of the economically important aquaculture species, gilthead seabream. Although epitheliocystis has been periodically described in this fish species, the increase in incidence and mortality associated with this disease is a current cause for concern (Dr Kantham Papanna, personal communication). Although chlamydial organisms are responsible for the vast majority of previous cases of epitheliocystis, we describe previously unknown β-proteobacteria as the major causative agents in fish farms around Greece. This study represents the most thorough investigation of this disease yet, using advanced microscopy and genomics to full advantage to describe this uncultured bacterial genus. Two distinct species, *Ca.* Ichthyocystis hellenicum and *Ca.* I. sparus, co-circulate during heavy infections, and we provide an insight into the diversity within and between these species.

We identified the pathogens as a novel genus of β-proteobacteria, related to organisms identified within epitheliocysts of Atlantic salmon and lake trout (Toenshoff *et al.*, 2012; Mitchell *et al.*, 2013;

Contador *et al.*, 2015). To confirm the identity of these infectious organisms, we used genomic data obtained from infected gill material that proves the novelty of these bacteria. Two genomes have been assembled and a further seven have been analysed by mapping. The nine genomes provide information on the diversity of these bacteria, and indicate that there may be many additional strains involved in epitheliocystis in the Mediterranean to explore.

Initial analysis of the gene content within this genus shows that the two species share only 65–69% of their genome, representing the core genes. The small genome size, paucity of pseudogenes and absence of amino-acid biosynthesis pathways indicates that these are compact, highly adapted genomes, consistent with an intracellular lifestyle (Merhej *et al.*, 2009). The regions unique to each species show a fascinating feature of massive duplication and diversification of several gene families with unknown functions. Although horizontal gene transfer is involved in gene-family expansion in most genomes studied to date (Lerat *et al.*, 2005; Treangen and Rocha, 2011), the lack of mobile elements and the tandem locations of these paralogues indicate that those described here more likely arose from gene duplication and subsequent mutation. Such a system of genome diversification has been previously noted within pathogens, such as the variable surface lipoproteins Vsp family in *Mycobacterium bovis* (Nussbaum, 2002) and Vlp family in *Mycoplasma hyorhinis* (Citti *et al.*, 2000), and appears to be a mechanism commonly used by obligate intracellular bacteria, with examples including the polymorphic membrane proteins within *Chlamydia* species (Stephens *et al.*, 1998; Thomson *et al.*, 2005), effector proteins in *Protochlamydia* and *Neochlamydia* (Domman *et al.*, 2014), and several gene families in *Ehrlichia ruminantium* (Collins *et al.*, 2005). However, the scale of this within the genomes of *Ca.* Ichthyocystis, with the gene families comprising over a quarter of the annotated CDSs, appears to be unprecedented. Further studies are essential to investigate the functions of these paralogues; if some are associated with the predicted T2SS and T3SS, bacterial effectors must have a very important role in the pathogenicity of these bacteria.

This work presents the first genomes of epitheliocystis-causing bacteria to be published, and the draft genomes will be improved, thoroughly analysed and published in their own right. The relatively small genome is comparable in size to that of many members of the obligate intracellular Chlamydiae and it will be intriguing to see whether further parallels can be found when the first chlamydial epitheliocystis genome becomes available. Future analysis will provide insights into the metabolic capacity, host adaptation, virulence and genome dynamics of this genus. Culture would be essential to generate a complete genome of this bacterium, which we shall continue to pursue.

We have shown the power of the use of FISH in correctly identifying causative agents of disease: targeting 16S rRNA gene sequences from previous candidate bacteria (Chlamydiae) in the absence of further work would have resulted in false identification of these pathogens. Using fluorescent imaging, EM and the latest techniques in three-dimensional-EM we have been able to characterise these novel bacteria in much greater detail than is normally possible for uncultured species.

EM shows that the bacteria are intracellular, contained either within a single epithelial cell or a construct built up of multiple cells. The encapsulation by overlapping membranes from surrounding cells, themselves connected by multiple desmosomal-like structures, indicates that the cysts enjoy a unique environment, in close association with blood capillaries, and well screened from the outside environment. The recruitment and reorganisation of these cells is presumably controlled by the bacteria, which would permit the controlled and efficient supply of cellular nutrients, necessary for the rapid cyst growth evidenced by the apparent constant division of bacteria throughout the cysts. How this is achieved must await further investigations which, for the first time in research into epitheliocystis, can be driven by genomic data. Around intact as well as ruptured cysts we also see macrophages and neutrophils indicating that they are not completely shielded from the host immune system. This immune reaction may lead to resolution of the infection, and protection of the host against subsequent infection.

The reason for the increasing incidence of epitheliocystis is still an open question and a cause for considerable concern. Further sampling is needed to elucidate the distribution and roles of these pathogens, in addition to the co-circulating Chlamydiae bacteria, among Mediterranean fish farms. Several of the features in the *Ca.* Ichthyocystis EM images are shared with those in previous cases of gilthead seabream infections since 1975 (Paperna, 1977; Paperna *et al.*, 1978; Paperna *et al.*, 1981; Crespo *et al.*, 1999), indicating that these pathogens may have been circulating, unidentified, for decades. The first characterisation of this disease in this host was within the Gulf of Eilat (Paperna, 1977), indicating a possible effect of the Suez channel in Lessepsian migration of non-native species from the Red Sea to the Mediterranean. The role of increasingly intensive aquaculture practices should also not be ignored, and the part played by animal husbandry (Segner *et al.*, 2011). Epitheliocystis may also increase host susceptibility to co-infections, meaning that a vaccine against these bacteria could be valuable to the aquaculture industry: inoculations against vibriosis and pasteurellosis are already routinely performed. It is essential to understand the diversity of these pathogens in order for a successful vaccine to be developed. We have made a critical first step towards this goal.

In this work we have identified a novel genus of bacterial pathogens that cause epitheliocystis in farmed fish. Characterisation through imaging gives evidence of rapidly dividing cocci forming intracellular cysts, and genomics on these uncultured bacteria shows a small core genome and large arrays of novel gene families. Ongoing studies over wider geographical areas will help to clarify their diversity, pathogenicity and impact.

Taxonomy
'*Candidatus* Ichthyocystis hellenicum' and '*Candidatus* Ichthyocystis sparus' spp. nov., recovered from gilthead seabream (*Sparus aurata*). Ichthyocystis (Ich'thy'o.cy'stis. Gr. masc. n. ichthys [ikhthýs], fish; N.L. fem. cystis from Gr. fem. n. kustis, bladder; N. L. fem. N. Ichthyocystis, cyst of fish), hellenicum (Hel'le'nic.um. N.L. neut. hellenicum, from Greece), sparus (Spa'rus. N.L. n. –a zoological genus name of the fish host). The 16S rRNA gene sequences of the new species place them within a novel bacterial genus. β-Proteobacteria causing cysts within fish gills, not cultured. Cysts are membrane enclosed, staining basophilic under HE, and react with an intense and specific labelling when a FISH probe designed to be specific to *Ca.* Ichthyocystis is applied.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS *et al.* (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455–477.

Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.

Cardona A, Saalfeld S, Schindelin J, Arganda-Carreras I, Preibisch S, Longair M *et al.* (2012). TrakEM2 Software for Neural Circuit Reconstruction. *PLoS One* **7**: e38011.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* **21**: 3422–3423.

Citti C, Watson-McKown R, Droesse M, Wise KS. (2000). Gene families encoding phase- and size-variable surface lipoproteins of *Mycoplasma hyorhinis*. *J Bacteriol* **182**: 1356–1363.

Collins NE, Liebenberg J, de Villiers EP, Brayton KA, Louw E, Pretorius A *et al.* (2005). The genome of the heartwater agent Ehrlichia ruminantium contains multiple tandem repeats of actively variable copy number. *Proc Natl Acad Sci USA* **102**: 838–843.

Colorni A, Padros F. (2011) Diseases and health management. In: Pavlidis MA, Mylonas CC (eds) *Sparidae: Biology and Aquaculture of Gilthead Sea Bream and other Species, Biology and Aquaculture of Gilthead Sea Bream and other Species.* Wiley-Blackwell: Oxford, UK.

Contador E, Methner P, Ryerse I, Huber P, Lillie BN, Frasca S Jr *et al.* (2015). Epitheliocystis in lake trout *Salvelinus namaycush* (Walbaum) is associated with a beta-proteobacteria. *J Fish Dis*; e-pub ahead of print 4 May 2015; doi:10.1111/jfd.12369.

Crespo S, Grau A, Padros F. (1990). Epitheliocystis disease in the cultured amberjack, *Seriola dumerili* Risso (Carangidae). *Aquaculture* **90**: 197–207.

Crespo S, Zarzo C, Padros F, de Mateo MM. (1999). Epitheliocystis agents in sea bream *Sparus aurata*: morphological evidence for two distinct chlamydia-like developmental cycles. *Dis Aquat Org* **37**: 61–72.

Domman D, Collingro A, Lagkouvardos I, Gehre L, Weinmaier T, Rattei T *et al.* (2014). Massive Expansion of Ubiquitination-Related Gene Families within the Chlamydiae. *Mol Biol Evol* **31**: 2890–2904.

Draghi A, Popov VL, Kahl MM, Stanton JB, Brown CC, Tsongalis GJ *et al.* (2004). Characterization of "*Candidatus* piscichlamydia salmonis" (order *Chlamydiales*), a chlamydia-like bacterium associated with epitheliocystis in farmed Atlantic salmon (*Salmo salar*). *J Clin Microbiol* **42**: 5286–5297.

Draghi A, Bebak J, Popov VL, Noble AC, Geary SJ, West AB *et al.* (2007). Characterization of a *Neochlamydia*-like bacterium associated with epitheliocystis in cultured Arctic charr *Salvelinus alpinus*. *Dis Aquat Org* **76**: 27–38.

Eddy SR. (2009). A new generation of homology search tools based on probabilistic inference. *Genome Inform* **23**: 205–211.

Everett KDE, Hornung LJ, Andersen AA. (1999). Rapid detection of the *Chlamydiaceae* and other families in the order *Chlamydiales*: three PCR tests. *J Clin Microbiol* **37**: 575–580.

F.A.O. [online], Cultured Aquatic Species Information Programme: Sparus aurata. Available from http://www.fao.org/fishery/culturedspecies/Sparus_aurata/en (accessed 3 November 2014).

F.A.O. (online). National Aquaculture Sector Overview: Greece. Available from http://www.fao.org/fishery/countrysector/naso_greece/en (accessed 3 November 2014).

F.A.O. (2012). *The State of World Fisheries and Aquaculture.* Food and Agriculture Organisation of the United Nations: Rome.

F.A.O. (2014). *The State of World Fisheries and Aquaculture.* Food and Agriculture Organisation of the United Nations: Rome.

Fehr A, Walther E, Schmidt-Posthaus H, Nufer L, Wilson A, Svercel M *et al.* (2013). Candidatus *Syngnamydia venezia*, a novel member of the phylum *Chlamydiae* from the broad nosed pipefish, *Syngnathus typhle*. *PLoS One* **8**: e70853.

Genome Atlas Database: Betaproteobacteria (online). Available from http://www.cbs.dtu.dk/services/Genome Atlas-3.0/?action=pt&k=Bacteria&ph=20&pt=g (accessed 23 September 2014).

Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. (2007). DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* **57**: 81–91.

Gouy M, Guindon S, Gascuel O. (2010). SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**: 221–224.

Greub G, La Scola B, Raoult D. (2004). Amoebae-resisting bacteria isolated from human nasal swabs by amoebal coculture. *Emerg Infect Dis* **10**: 470–477.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075.

Hoffman GL, Dunbar CE, Wolf K, Zwillenberg LO (1969). Epitheliocystis, a new infectious disease of the bluegill (*Lepomis macrochirus*). *Antonie Van Leeuwenhoek* **35**: 146–158.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* **11**: 119.

Jehl MA, Arnold R, Rattei T. (2011). Effective-a database of predicted secreted bacterial proteins. *Nucleic Acids Res* **39**: D591–D595.

Kanehisa M, Goto S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28**: 27–30.

Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* **38**: D355–D360.

Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. (2012). KEGG for integration and interpretation of large-scale molecular datasets. *Nucleic Acids Res* **40**: D109–D114.

Karlsen M, Nylund A, Watanabe K, Helvik JV, Nylund S, Plarre H. (2008). Characterization of '*Candidatus* Clavochlamydia salmonicola': an intracellular bacterium infecting salmonid fish. *Environ Microbiol* **10**: 208–218.

Katharios P, Seth-Smith HMB, Fehr A, Mateos JM, Qi W, Richter D *et al.* (2015). Environmental marine pathogen isolation using mesocosm culture of sharpsnout seabream: striking genomic and morphological features of novel Endozoicomonas sp. *Sci Rep*; e-pub ahead of print 7 December 2015; doi:10.1038/ismej.2015.223.

Lerat E, Daubin V, Ochman H, Moran NA. (2005). Evolutionary origins of genomic repertoires in bacteria. *PLoS Biol* **3**: e130.

Lienard J, Croxatto A, Aeby S, Jaton K, Posfay-Barbe K, Gervaix A *et al.* (2011). Development of a new *Chlamydiales*-specific real-time PCR and its application to respiratory clinical samples. *J Clin Microbiol* **49**: 2637–2642.

Mendoza M, Guiza L, Martinez X, Caraballo X, Rojas J, Aranguren LF et al. (2013). A novel agent (*Endozoicomonas elysicola*) responsible for epitheliocystis in cobia *Rachycentrum canadum* larvae. *Dis Aquat Organ* **106**: 31–37.

Merhej V, Royer-Carenzi M, Pontarotti P, Raoult D. (2009). Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol Direct* **4**: 13.

Mitchell SO, Steinum T, Toenshoff ER, Kvellestad A, Falk K, Horn M et al. (2013). '*Candidatus* Branchiomonas cysticola' is a common agent of epitheliocysts in seawater-farmed Atlantic salmon *Salmo salar* in Norway and Ireland. *Dis Aquat Org* **103**: 35–43.

Nowak BF, LaPatra SE. (2006). Epitheliocystis in fish. *J Fish Dis* **29**: 573–588.

Nurk S, Bankevich A, Antipov D, Gurevich AA, Korobeynikov A, Lapidus A et al. (2013). Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. *J Comput Biol* **20**: 714–737.

Nussbaum S. (2002). Extended repertoire of genes encoding variable surface lipoproteins in Mycoplasma bovis strains. *Infect Immun* **70**: 2220–2225.

Paperna I. (1977). Epitheliocystis infection in wild and cultured sea bream (*Sparus aurata*, Sparidae) and grey mullets (*Liza Ramada*, Mugiligae). *Aquaculture* **10**: 169–176.

Paperna I, Sabnai I, Castel M. (1978). Ultrastructural study of epitheliocystis organisms from gill epithelium of the fish *Sparus aurata* (L.) and *Liza ramada* (Risso) and their relation to the host cell. *J Fish Dis* **1**: 181–189.

Paperna I, Sabnai I, Zachary A. (1981). Ultrastructural stuides in piscine epitheliocystis: evidence for a pleomorphic developmental cycle. *J Fish Dis* **4**: 459–472.

Paperna I, Alves Dematos AP (1984). The developmental cycle of epitheliocystis in carp, Cyprinus carpio L. *J Fish Dis* **7**: 137–147.

Ponti A, Schwarb P, Gulati A, Bäcker V. (2007). A web interface for high-volume batch deconvolution. *Imaging Microsc* **9**: 57–58.

Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, Zhou J et al. (2014). A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* **196**: 2210–2215.

Rahman A, Pachter L. (2013). CGAL: computing genome assembly likelihoods. *Genome Biol* **14**: R8.

Rutherford KM, Parkhill J, Crook J, Hornsnell T, Rice P, Rajandream MA et al. (2000). Artemis: sequence visualization and annotation. *Bioinformatics* **16**: 944–945.

Schmidt-Posthaus H, Polkinghorne A, Nufer L, Schifferli A, Zimmermann DR, Segner H et al. (2012). A natural freshwater origin for two chlamydial species, *Candidatus* Piscichlamydia salmonis and *Candidatus* Clavochlamydia salmonicola, causing mixed infections in wild brown trout (*Salmo trutta*). *Environ Microbiol* **14**: 2048–2057.

Seeman T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Segner H, Sundh H, Buchmann K, Douxfils J, Sundell KS, Mathieu C et al. (2011). Health of farmed fish: its relation to fish welfare and its utility as welfare indicator. *Fish Physiol Biochem* **38**: 85–105.

Sommer C, Strähle C, Köthe U, Hamprecht FA. (2011). ilastik: Interactive Learning and Segmentation Toolkit. *Eighth IEEE International Symposium on Biomedical Imaging (ISBI), Proceedings,* 230–233.

Steigen A, Nylund A, Karlsbakk E, Akoll P, Fiksdal IU, Nyland S et al. (2013). '*Cand.* Actinochlamydia clariae' gen. nov., sp. nov., a Unique Intracellular Bacterium Causing Epitheliocystis in Catfish (*Clarias gariepinus*) in Uganda. *PLoS One* **8**: e66840.

Stephens RS, Kalman S, Lammel C, Fan J, Marathe R, Aravind L et al. (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**: 754–759.

Stride MC, Polkinghorne A, Miller TL, Groff JM, LaPatra SE, Nowak BF. (2013a). Molecular characterization of "*Candidatus* Parilichlamydia carangidicola", a novel *Chlamydia*-Like epitheliocystis agent in yellowtail kingfish, *Seriola lalandi* (Valenciennes), and the proposal of a new family, "*Candidatus* Parilichlamydiaceae" fam. nov. (Order *Chlamydiales*). *Appl Environ Microbiol* **79**: 1590–1597.

Stride MC, Polkinghorne A, Miller TL, Nowak BF. (2013b). Molecular characterization of "*Candidatus* Similichlamydia latridicola" gen. nov., sp. nov. (*Chlamydiales*: "*Candidatus* Parilichlamydiaceae"), a novel *Chlamydia*-like epitheliocystis agent in the striped trumpeter, *Latris lineata* (Forster). *Appl Environ Microbiol* **79**: 4914–4920.

Stride MC, Polkinghorne A, Nowak BF. (2014). Chlamydial infections of fish: diverse pathogens and emerging causes of disease in aquaculture species. *Vet Microbiol* **170**: 19–27.

Tarlera S. (2003). *Sterolibacterium denitrificans* gen. nov., sp. nov., a novel cholesterol-oxidizing, denitrifying member of the *beta-Proteobacteria*. *Int J Syst Evol Microbiol* **53**: 1085–1091.

Thomson NR, Yeats C, Bell K, Holden MTG, Bentley SD, Livingstone M et al. (2005). The *Chlamydophila abortus* genome sequence reveals an array of variable proteins that contribute to interspecies variation. *Genome Res* **15**: 629–640.

Toenshoff ER, Kvellestad A, Mitchell SO, Steinum T, Falk K, Colquhoun DJ et al. (2012). A novel betaproteobacterial agent of gill epitheliocystis in seawater farmed Atlantic salmon (*Salmo salar*). *PLoS One* **7**: e32696.

Treangen TJ, Rocha EP. (2011). Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet* **7**: e1001284.

Weisberg WG, Barns SM, Pelletier DA, Lane DJ. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**: 697–703.

Winton J, Batts W, deKinkelin P, LeBerre M, Bremont M, Fijan N. (2010). Current lineages of the epithelioma papulosum cyprini (EPC) cell line are contaminated with fathead minnow, *Pimephales promelas*, cells. *J Fish Dis* **33**: 701–704.

Yarza P, Yilmaz P, Pruesse E, Glockner FO, Ludwig W, Schleifer KH et al. (2014). Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**: 635–645.