Hydrobiologia (2015) 761:85–95
DOI 10.1007/s10750-015-2396-9

BIOLOGY OF THE ROSS SEA

Genome-scale phylogenetic and DNA composition analyses of Antarctic *Pseudoalteromonas* bacteria reveal inconsistencies in current taxonomic affiliation

Emanuele Bosi · Marco Fondi · Isabel Maida · Elena Perrin · Donatella de Pascale · Maria Luisa Tutino · Ermenegilda Parrilli · Angelina Lo Giudice · Alain Filloux · Renato Fani

Received: 7 January 2015 / Revised: 19 June 2015 / Accepted: 28 June 2015 / Published online: 16 July 2015

Abstract Bacteria belonging to the *Pseudoalteromonas* genus have important ecological implications in marine environments, playing a role in the control of microbial community as producers of bioactive molecules endowed with antifouling activity and able to antagonize larvae, fungi and bacteria, including important human pathogens. For these reasons, representatives of this genus are very promising for biotechnological and biomedical applications. In this work, we used different genome-scale approaches to infer the taxonomy of 38 *Pseudoalteromonas* representatives (most of which isolated from Antarctica) and whose complete genome has been sequenced. We show that an accurate re-evaluation of the real taxonomic relationships of *Pseudoalteromonas* representatives is needed since many inconsistencies with the current taxonomic annotation were observed. Moreover, data obtained with different genome-scale methods are consistent, confirming the reliability of the genomic approaches. On the basis of these data, we propose a re-annotation for some *Pseudoalteromonas* species. This proposal should be validated in the future by comparing the phenotypes of these strains.

Keywords *Pseudoalteromonas* · Taxonomy · Antarctic bacteria · Genome analyses · Biotechnology

Guest editors: Diego Fontaneto & Stefano Schiaparelli / Biology of the Ross Sea and Surrounding Areas in Antarctica

Electronic supplementary material The online version of this article (doi:10.1007/s10750-015-2396-9) contains supplementary material, which is available to authorized users.

E. Bosi · M. Fondi · I. Maida · E. Perrin · R. Fani (✉)
Laboratory of Microbial and Molecular Evolution, Department of Biology, University of Florence, Via Madonna del Piano 6, 50019 Sesto Fiorentino, FI, Italy
e-mail: renato.fani@unifi.it; renato.fani@virgilio.it

D. de Pascale
Institute of Protein Biochemistry, CNR, Napoli, Italy

M. L. Tutino · E. Parrilli
Department of Chemical Sciences, University of Naples Federico II, Complesso Universitario M. S. Angelo, Via Cintia, 80126 Napoli, Italy

A. Lo Giudice
Institute for the Coastal Marine Environment, National Research Council (IAMC-CNR), Spianata San Raineri 86, 98124 Messina, Italy

A. Lo Giudice
Department of Biological and Environmental Sciences, University of Messina, Viale F. Stagno d'Alcontrès 31, 98166 Messina, Italy

Introduction

The genus *Pseudoalteromonas* denotes a group of Gram-negative, aerobic marine bacteria belonging to the class Gammaproteobacteria, firstly established in 1995 (Gauthier et al., 1995).

In these two last decades, several *Pseudoalteromonas* strains have been isolated from Polar Regions, inshore waters or surfaces of marine organisms, and were shown to synthesize a wide range of bioactive molecules (Kobayashi, 2003; Feller, 2013; Yu et al., 2013). In this context, it has been recently shown that strains belonging to this genus and isolated from different ecological niches in Antarctica possess the ability to completely inhibit the growth of human opportunistic pathogens belonging to the *Burkholderia cepacia* complex (Bcc) via the synthesis of a plethora of different antimicrobial compounds (Papaleo et al., 2012; 2013); moreover, some of them are also able to synthesize antibiofilm molecules (Papa et al., 2013; Parrilli et al., 2015).

Another intriguing feature of some *Pseudoalteromonas* representatives is the association with marine eukaryotic hosts. Indeed, complex communities with biofouling activities have been shown to be beneficial to the eukaryotic host, effectively playing a role in host defence (Holmström et al., 1992; Egan et al., 2002). The study of the molecular mechanisms of antifouling activity from *Pseudoalteromonas* species is promising for multiple applications, such as biofouling control in aquaculture and novel drug discovery. In particular, the isolation of novel antibiotics is strategically important, considering the global threat for human's health posed by the emergence of multi-drug resistant (MDR) pathogens, mainly due to antibiotic over usage (Paitan & Ron, 2014). Finally, the most studied *Pseudoalteromonas* strain, namely *Pseudoalteromonas haloplanktis* TAC125 (Médigue et al., 2005), has been suggested as an alternative host for the soluble overproduction of heterologous proteins, given its ability to grow fast at low temperatures (Wilmes et al., 2010; Rippa et al., 2012; Corchero et al., 2013; Giuliani et al., 2014). All these features make the representatives of *Pseudoalteromonas* genus interesting biological subjects, with a great and still under-exploited biotechnological potential.

From a taxonomical viewpoint, the *Pseudoalteromonas* genus, together with *Alteromonas*, *Glaciecola*, *Thalassomonas*, *Colwellia*, *Idiomarina*, *Shewanella*, *Moritella*, *Ferrimonas* and *Psychromonas*, forms a group referred to as *Alteromonas*-like bacteria. Historically, the genus was first described by Gauthier et al. in 1995 when, according to a broad-scale phylogenetic analysis based on small subunit ribosomal DNA sequences, it was observed that some *Alteromonas* representatives did not belong to the monophyletic taxon, which included the other *Alteromonas* species (Gauthier et al., 1995). Based on this phylogenetic analysis and phenotypical evidences, a total of 12 *Alteromonas* species, together with *Pseudomonas piscicida*, were assigned to the genus *Pseudoalteromonas*.

From a methodological viewpoint, this taxonomical revision highlighted the central role of the 16S rRNA gene sequence to assess the evolutionary relationships among bacterial species. However, being based on a single gene, this method suffers some limitations, such as the limited ability to resolve closely related species (Stackebrandt and Ebers 2006), the fact that it does not represent the whole gene collection, and the poor correlation with the genome-scale method of DNA–DNA Hybridization (DDH), which is an experimental method measuring the overall similarity between two genome sequences (Schildkraut et al., 1961; McCarthy & Bolton, 1963). A higher resolution can be obtained using the sequences of multiple conserved genes with a method known as Multi-Locus Sequence Analysis (MLSA) (Stackebrandt et al., 2002).

At the early stage of the sequencing era, it was theorized how the whole-genome sequence would be the standard to determine the taxonomy (Wayne et al., 1987). Nowadays, thanks to ever decreasing costs and running time, genome sequencing for prokaryotes has become a routine, to the point that a considerable number of sequenced genomes are available in biological repositories such as GenBank.

This wealth of data can be exploited to infer phylogenetic relationships using genome-scale computational methods to provide quantitative estimation of the genomic similarity analogously equivalent to the experimental method DDH (Goris et al., 2007; Kim et al., 2014). One of the advantages provided by a

A. Filloux
Department of Life Sciences, MRC Centre for Molecular Bacteriology and Infection, Imperial College London, London SW72AZ, UK

computational method is that it can be used to replace DDH, thus allowing performing significant taxonomic studies at a genome scale for a vast number of species.

Although it may be erroneously considered as marginal, a correct and robust taxonomic classification of microbes plays a central role in describing the extent of microbial diversity in relation with different environments and/or eukaryotic hosts. Also, a reliable phylogenetic reconstruction is crucial in guiding the choice of which other novel organisms should be introduced in a Next Generation Sequencing (NGS) pipeline, in order to avoid oversampling a narrow taxonomic space (the "where to add taxa" problem) (Eddy, 2005; McAuliffe et al., 2005; Pardi & Goldman, 2005; Geuten et al., 2007).

In this work, we performed a comprehensive and multi-level study on the taxonomy of a panel consisting of 25 currently available *Pseudoalteromonas* representatives and 13 *de novo* sequenced genomes from Antarctic strains, by integrating different genome-scale methods, namely, genome-scale phylogeny, Average Nucleotide Identity (ANI) and Tetra Nucleotide Frequency (TNF). The benefits and pitfalls of phylogenetic analyses for taxonomic purposes using a genome-scale level set of shared genes in prokaryotes have been discussed by Rossello-Mora (2012). Concerning the ANI, it has been demonstrated how it can be used to discriminate whether two genomes belong either to the same or different species, using a fixed threshold (Goris et al., 2007). Lastly, it has been also shown that oligonucleotide frequencies exhibit species-specific patterns (Karlin & Burge, 1995; Karlin, 1998) and how, in particular, frequencies of tetranucleotides harbour a phylogenetic signal (Pride et al., 2003). This integrated approach revealed that (i) different methods produce consistent results, and (ii) incoherence is observed between genome-scale driven taxonomic annotation and current affiliations of members of the genus *Pseudoalteromonas*.

Materials and methods

Pseudoalteromonas dataset

The available genomic sequences of 25 *Pseudoalteromonas* strains were downloaded from GenBank. Additionally, the genomes of 13 *Pseudoalteromonas* strains isolated from different Antarctic ecological niches (marine sponges, water column, sediments) were sequenced. The genomic DNA was purified according to the protocol described by Papaleo et al. (2013) and sequenced using an Illumina HiSeq 2000 platform (Cock et al., 2010). The resulting reads were trimmed using SolexaQ to obtain a high-quality reads library, which were assembled with the tool AbySS 1.4, choosing the k-mer value leading to the best assembly, defined as the ratio between assembled nucleotides and number of contigs obtained.

Genome coherence measures

To measure the genome coherence between *Pseudoalteromonas* representatives, the Jspecies tool was used (Richter & Rosselló-Móra, 2009). This software implements different methods to provide a quantitative measure of genomic similarity, which can be used to estimate whether two genomes belong to the same species. The metrics implemented in Jspecies are the Average Nucleotide Identity (ANI) and the TetraNucleotide Frequencies (TNF) (Teeling et al., 2004a, b; Goris et al., 2007).

Average nucleotide identity (ANI)

The ANI provides a quantitative measure of the genomic similarity between two organisms. The Jspecies tool implements the ANI computation method described by Goris (Goris et al., 2007), in which a genome is divided into 1024 nucleotide fragments, which are mapped onto the other genome to measure the ANI. To map the sequences, we computed the ANI values using the MUMmer tool (Delcher et al., 2003). According to Goris et al. (2007), a pair of genomes with an ANI value greater than 0.96 is considered to belong to the same species.

Tetra Nucleotide Frequencies (TNF)

The TNF can be used to measure the similarity of two genomes, by computing the significance of the frequency of each oligonucleotide as Z-scores (Schbath, 1997), and by measuring the Pearson's correlation of these values. In particular, it has been demonstrated that there is a good correlation with the similarity measures obtained using the DDH method (Teeling et al., 2004a, b). Specifically, two genomes with TNF value greater than 0.99 should be

considered as belonging to the same species (Teeling et al., 2004a, b). To have a clearer signal and better visualize the species relationships, the TNF values have been transformed into binary values, considering as 1 the TNF values greater than the species threshold, and as 0 the values lower than the threshold.

Conserved genes phylogeny

In this work, a set of 1537 highly conserved genes has been found using the DuctApe suite (Galardini et al., 2014). The amino acid sequences of the proteins encoded by the 1537 conserved genes (i.e. shared by all the 38 *Pseudoalteromonas* genomes) have been aligned using the ClustalW software (Larkin et al., 2007) with default parameters and concatenated in a single sequence spanning 597,947 residues. Neighbour-Joining (NJ) phylogenetic tree was obtained with Mega 6 software (Tamura et al., 2013) with the following parameters: Poisson model, uniform rates among sites and 500 bootstrap replicates.

Hierarchical clustering

Hierarchical clustering was performed using the Unweighted Pair Group Method with Arithmetic mean (UPGMA).

Orthologous genes dendrogram

Using the DuctApe suite (Galardini et al., 2014), it was possible to identify 2901 groups of orthologs, which were differentially present in the *Pseudoalteromonas* genomes. This information has been used to produce a matrix of presence/absence. The dendrogram of this matrix has been obtained by performing a hierarchical clustering (UPGMA) using Jaccard distance.

Results

Pseudoalteromonas dataset

To gain insights into the genomic similarities between representatives of the *Pseudoalteromonas* genus, a dataset of 38 genomes was used. Thirteen genomes were obtained in this work from *Pseudoalteromonas* strains isolated from different Antarctic ecological niches (sponges, sediments, and seawater). The 38 genomes belong to bacterial strains representative of the 13 currently described *Pseudoalteromonas* species. The genomic features of the strains considered are given in Table 1.

Conserved genes phylogeny

The NJ phylogenetic tree computed using the genes shared by all the 38 *Pseudoalteromonas* strains is given in Fig. 1. The visual inspection of this tree revealed the presence of a main group consisting of 29 sequences (corresponding to 28 strains, since the genome sequence of strain TAC125 was determined twice using two different methodologies). Since this clade comprises all the *P. haloplanktis* strains, this group was referred to as *P. haloplanktis*-like group. However, the *P. haloplanktis* species appeared not to be monophyletic in the clade, since other species (namely *P. undina*, *P. marina* and *P. arctica*) were embedded in this group.

Analysis of ANI

The results of the hierarchical clustering of the ANI matrix, embedding the ANI values computed for each pair of genomes (reported in Additional File 1) revealed the presence of clusters formed by groups of strains sharing highly similar genomes (ANI value $> 96\%$), which might be considered as belonging to the same species (see Fig. 2). Twenty-five strains were split into 8 clusters of variable size (ranging from 2 to 5 strains). The remaining 13 strains were not grouped with any other representatives, and they were referred to as singletons. Interestingly, among the singletons, we found a relatively high number of previously defined *Pseudoalteromonas* species (9 out of the 13 species represented in the dataset).

Notably, the *Pseudoalteromonas* species are found to be consistent with the clusters observed, in that the different *Pseudoalteromonas* species join different clusters; the only exception is represented by *P. haloplanktis*, since the four strains previously affiliated to *P. haloplanktis* species are split into three groups (one of them being a singleton, Fig. 2).

Analysis of TNF

The visual inspection of the hierarchical clustering of the TNF binary matrix (whose data are reported in Additional File 2) shown in Fig. 3 revealed a different

Table 1 Main features of the 38 genomes from the *Pseudoalteromonas* strains analysed in this work

Strains	Contigs/Replicons	Total length	ORFs	G + C (mol%)	Type strain
Pseudoalteromonas sp. AC163	565	4,779,003	4765	39.10	No
P. arctica A 37 1 2 uid168325	68	4,628,018	4094	39.04	Yes
P. atlantica T6c uid58283	1	5,187,005	4281	44.62	No
Pseudoalteromonas sp. BSi20311 uid78647	195	3,979,836	3676	40.33	No
Pseudoalteromonas sp. BSi20429 uid78649	121	4,495,777	4030	39.04	No
Pseudoalteromonas sp. BSi20439 uid78651	243	3,882,800	3612	40.22	No
Pseudoalteromonas sp. BSi20480 uid78653	201	4,149,214	3967	39.60	No
Pseudoalteromonas sp. BSi20495 uid78655	222	4,826,524	4365	38.92	No
Pseudoalteromonas sp. BSi20652 uid78645	298	4,253,936	4085	38.86	No
Pseudoalteromonas sp. Bsw20308 uid179221	146	4,757,001	4172	38.90	No
P. citrea NCIMB 1889 uid168326	114	5337619	4438	41.13	Yes
P. flavipulchra JG1 uid177806	61	5,503,991	4758	43.19	No
P. haloplanktis ANT 505 uid66747	142	4,494,717	4127	39.32	No
P. haloplanktis ATCC 14393 uid198981	56	6,513,609	4329	40.84	Yes
P. haloplanktis TAC125 uid58431	2	3,850,272	3484	40.09	No
P. luteoviolacea B ATCC 29581 uid186644	61	4,046,270	3681	41.95	No
P. marina mano4 uid168327	31	4,177,200	3711	39.65	Yes
Pseudoalteromonas sp. NJ631 uid199000	55	6,943,067	4591	43.36	No
Pseudoalteromonas sp. PAMC 22718 uid179404	56	5,425,171	3821	40.17	No
P. piscicida JCM 20779 uid168328	73	5,281,621	4524	43.24	Yes
P. rubra ATCC 29570 uid168329	64	5,969,931	4893	47.80	Yes
P. ruthenica CP76 uid199935	120	5,225,945	3714	47.59	No
Pseudoalteromonas sp. S838	87	4,990,009	4427	39.19	No
Pseudoalteromonas sp. S88	79	4,911,233	4371	39.21	No
Pseudoalteromonas sp. SM9913 uid61247	2	4,037,671	3712	40.28	No
P. spongiae UST010723 006 uid168330	14	4,724,746	4185	40.81	No
Pseudoalteromonas sp. TAB 23	367	5,139,089	5012	39.11	No
P. haloplanktis TAC125	216	3,888,065	3740	39.98	No
Pseudoalteromonas sp. TAE56	163	4,600,700	4258	39.03	No
Pseudoalteromonas sp. TAE79	298	5,045,088	4940	39.29	No
Pseudoalteromonas sp. TAE80	360	4,971,170	4941	39.29	No
Pseudoalteromonas sp. TB13	254	4,734,094	4489	39.05	No
Pseudoalteromonas sp. TB25	458	4,648,658	4546	39.18	No
Pseudoalteromonas sp. TB41	122	4,632,606	4217	40.34	No
Pseudoalteromonas sp. TB51	369	4,633,324	4625	40.91	No
Pseudoalteromonas sp. TB64	275	4,843,680	4649	37.92	No
P. tunicata D2 uid54181	37	4,994,813	4504	39.75	No
P. undina NCIMB 2128 uid168331	20	4,001,234	3581	39.95	Yes
Average:		4,802,755	4245	40.53	

number of clusters with respect to those identified with the ANI. We detected seven singleton *Pseudoalteromonas* strains and two major clusters, one of which including most of *Pseudoalteromonas* strains. A deeper analysis of the clusters composition revealed a consistency between the clusters found with TNF and those observed in the MLSA-based phylogenetic tree. As an example, the largest cluster found with TNF contains the same strains, forming the *P. haloplanktis*-like group.

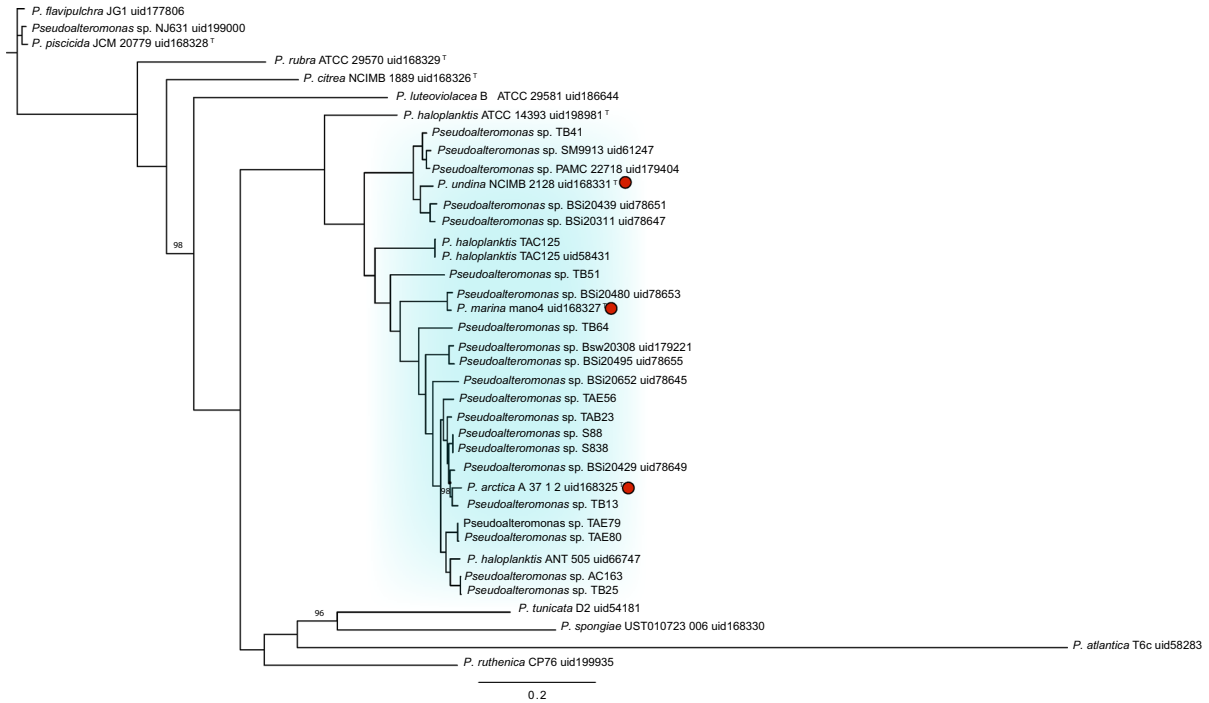

Fig. 1 *Pseudoalteromonas* phylogenetic tree based on a concatenated sequence consisting of 597,947 amino acids from 1537 conserved proteins shared by the 38 genomes. The azure area represents the *P. haloplanktis*-like group, while the red dots stand near to strains not assigned to the *P. haloplanktis* species. Unless specified, bootstrap support is 100. The scale below represents the substitution rate

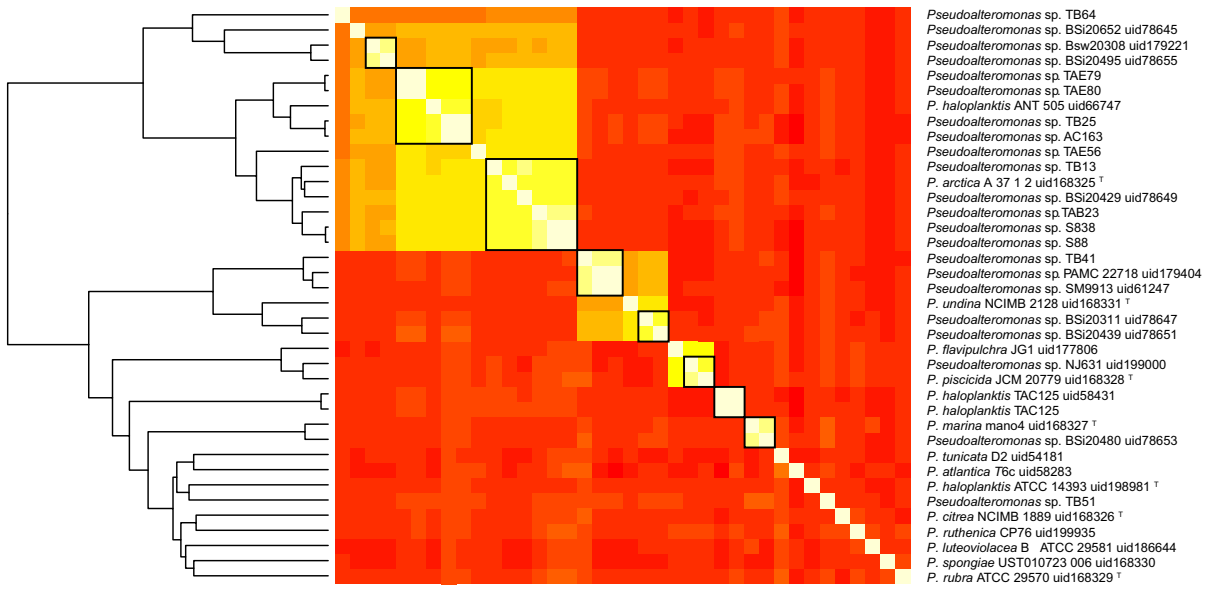

Fig. 2 Heatmap representation of the ANI matrix. The clusters reported representative species according to the ANI method. The cell colours represent the ANI values, i.e. a *dark red colour* stands for a ANI value of 0, whereas a *white colour* stands for ANI value of 1

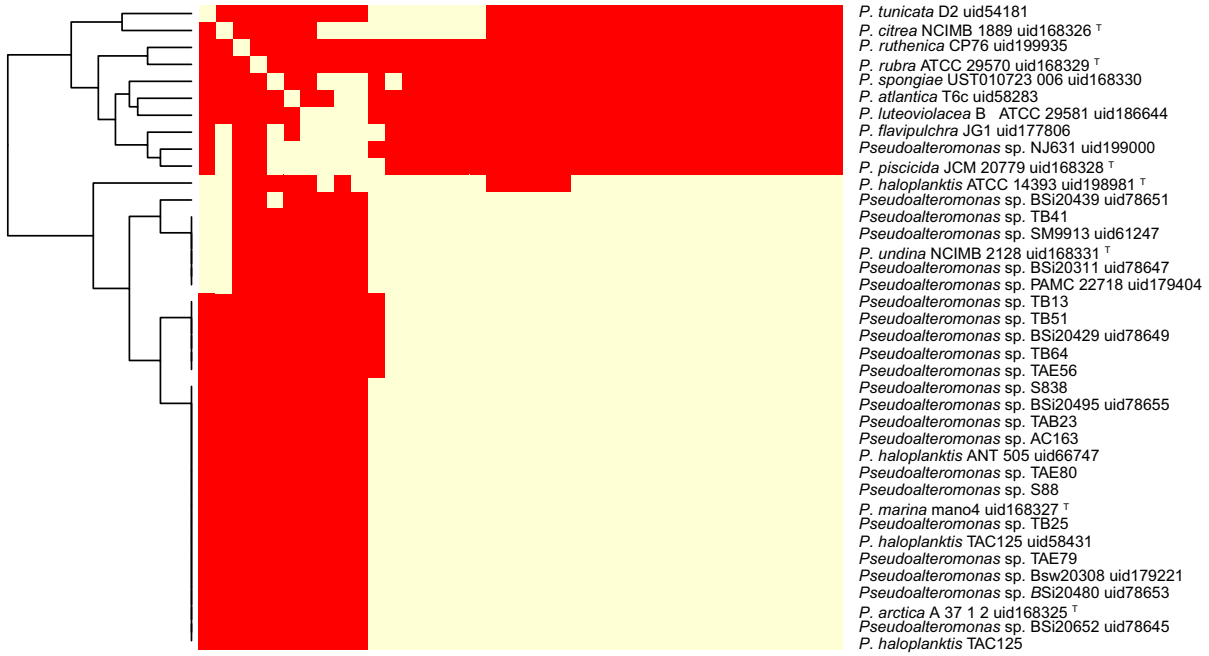

Fig. 3 Heatmap representation of the TNF matrix. *White* cells represent pair of strains with TNF value greater than the species threshold. Clustered strains represent species found according to the TNF method

To have a clearer picture of the similarity of the dendrograms produced with different methods, each pair is reported in Additional File 3, 4 and 5, while the number and composition of cluster obtained with the two methods are given in Table 2.

Orthologous genes analysis

In order to fully exploit the information embedded in the genomic sequences, the gene content information has been used to make phylogenetic inferences. The dendrogram of the matrix of gene presence/absence clustering, embedding a total of 2901 genes differentially present in the *Pseudoalteromonas* strains reported in Additional File 6, has a topology that is very similar to that of the phylogenetic tree obtained with the concatamer of the conserved genes.

Discussion

In this work, we used the genomic sequences of 38 *Pseudoalteromonas* representatives to depict their taxonomic relationships, using two different (genome-scale) approaches. To the best of our knowledge, this dataset constitutes the most comprehensive and recent source of information for this genus. The first approach used was a phylogenetic analyses based on a comprehensive set of common proteins (genome-scale phylogeny). The topology of the phylogenetic tree showed no monophyly for the type species of the *Pseudoalteromonas* genus, i.e. *P. haloplanktis*, even though it was possible to detect a well-defined clade comprising all the *P. haloplanktis* representatives, which was named *P. haloplanktis*-like group. This group contains 28 strains, i.e. the majority of the *Pseudoalteromonas* strains considered in this work, representative of four *Pseudoalteromonas* species, namely *P. haloplanktis*, *P. undina*, *P. marina* and *P. arctica*. This result might be explained as follows:

(1) The topology of the phylogenetic tree is misleading. In fact, it has been argued how by concatenating genes, the information about individual *loci* may be hidden, leading to loss of resolution power and, potentially, to misleading results (Rosselló-Móra, 2012 and references therein).
(2) The taxonomic assignments are not consistent with the tree topology, which might be due to

Table 2 Cluster composition (number and composition of clusters) of dendrograms obtained with ANI and TNF

	ANI		TNF
Singleton 1	*P. citrea* NCIMB 1889 uid168326	Singleton 1	*P. tunicata* D2 uid54181
Singleton 2	*Pseudoalteromonas* sp. BSi20652 uid78645	Singleton 2	*P. citrea* NCIMB 1889 uid168326
Singleton 3	*Pseudoalteromonas* sp. TB64	Singleton 3	*P. ruthenica* CP76 uid199935
Singleton 4	*P. spongiae* UST010723 006 uid168330	Singleton 4	*P. rubra* ATCC 29570 uid168329
Singleton 5	*P. atlantica* T6c uid58283	Singleton 5	*P. spongiae* UST010723 006 uid168330
Singleton 6	*Pseudoalteromonas* sp. TAE56	Singleton 6	*P. atlantica* T6c uid58283
Singleton 7	*P. luteoviolacea* B ATCC 29581 uid186644	Singleton 7	*P. luteoviolacea* B ATCC 29581 uid186644
Singleton 8	*P. undina* NCIMB 2128 uid168331	Cluster 1	*P. flavipulchra* JG1 uid177806
Singleton 9	*P. haloplanktis* ATCC 14393 uid198981	Cluster 1	*P. piscicida* JCM 20779 uid168328
Singleton 10	*Pseudoalteromonas* sp. TB51	Cluster 1	Pseudoalteromonas sp. NJ631 uid199000
Singleton 11	*P. tunicata* D2 uid54181	Cluster 2	*Pseudoalteromonas* sp. BSi20652 uid78645
Singleton 12	*P. ruthenica* CP76 uid199935	Cluster 2	*Pseudoalteromonas* sp. TB64
Singleton 13	*P. rubra* ATCC 29570 uid168329	Cluster 2	*Pseudoalteromonas* sp. TAE56
Cluster 1	*Pseudoalteromonas* sp. BSi20439 uid78651	Cluster 2	*P. undina* NCIMB 2128 uid168331
Cluster 1	*Pseudoalteromonas sp.* BSi20311 uid78647	Cluster 2	*P. haloplanktis* ATCC 14393 uid198981
Cluster 2	*P. haloplanktis* TAC125	Cluster 2	*Pseudoalteromonas* sp. TB51
Cluster 2	*P. haloplanktis* TAC125 uid58431	Cluster 2	*Pseudoalteromonas* sp. BSi20439 uid78651
Cluster 3	*P. marina* mano4 uid168327	Cluster 2	*Pseudoalteromonas* sp. BSi20311 uid78647
Cluster 3	*Pseudoalteromonas* sp. BSi20480 uid78653	Cluster 2	*P. haloplanktis* TAC125
Cluster 4	*Pseudoalteromonas* sp. BSi20495 uid78655	Cluster 2	*P. haloplanktis* TAC125 uid58431
Cluster 4	*Pseudoalteromonas* sp. Bsw20308 uid179221	Cluster 2	*P. marina* mano4 uid168327
Cluster 5	*P. flavipulchra* JG1 uid177806	Cluster 2	*Pseudoalteromonas* sp. BSi20480 uid78653
Cluster 5	*P. piscicida* JCM 20779 uid168328	Cluster 2	*Pseudoalteromonas* sp. BSi20495 uid78655
Cluster 5	*Pseudoalteromonas* sp. NJ631 uid199000	Cluster 2	*Pseudoalteromonas* sp. Bsw20308 uid179221
Cluster 6	*Pseudoalteromonas* sp. SM9913 uid61247	Cluster 2	*Pseudoalteromonas* sp. SM9913 uid61247
Cluster 6	*Pseudoalteromonas* sp. PAMC 22718 uid179404	Cluster 2	*Pseudoalteromonas* sp. PAMC 22718 uid179404
Cluster 6	*Pseudoalteromonas* sp. TB41	Cluster 2	*Pseudoalteromonas* sp. TB41
Cluster 7	*Pseudoalteromonas* sp. TAE79	Cluster 2	*Pseudoalteromonas* sp. TAE79
Cluster 7	*Pseudoalteromonas* sp. TB25	Cluster 2	*Pseudoalteromonas* sp. TB25
Cluster 7	*Pseudoalteromonas* sp. TAE80	Cluster 2	*Pseudoalteromonas* sp. TAE80
Cluster 7	*P. haloplanktis* ANT 505 uid66747	Cluster 2	*P. haloplanktis* ANT 505 uid66747
Cluster 7	*Pseudoalteromonas* sp. AC163	Cluster 2	*Pseudoalteromonas* sp. AC163
Cluster 8	*Pseudoalteromonas* sp. S88	Cluster 2	*Pseudoalteromonas* sp. S88
Cluster 8	*P. arctica* A 37 1 2 uid168325	Cluster 2	*P. arctica* A 37 1 2 uid168325
Cluster 8	*Pseudoalteromonas* sp. BSi20429 uid78649	Cluster 2	*Pseudoalteromonas* sp. BSi20429 uid78649
Cluster 8	*Pseudoalteromonas* sp. TB13	Cluster 2	*Pseudoalteromonas* sp. TB13
Cluster 8	*Pseudoalteromonas* sp. TAB 23	Cluster 2	*Pseudoalteromonas* sp. TAB 23
Cluster 8	Pseudoalteromonas sp. S838	Cluster 2	Pseudoalteromonas sp. S838

assignments based on a not full taxonomic characterization of *Pseudoalteromonas* isolates.

The other approach used was the genome coherence approach, which relies on the idea that members of the same species share the same genomic features such as nucleotide identity and/or similar tetranucleotide composition. Therefore, to explore the likelihood of the two different scenarios, we analysed the genome coherence of the *Pseudoalteromonas* strains using the ANI and TNF. It has been demonstrated (Pride et al.,

2003; Goris et al., 2007) how both these methods can be effectively used to discriminate whether two genomes belong to the same or different species. Hence, we used these approaches to investigate whether the taxon and/or the branching order obtained are consistent with each other and to test the reliability of the results obtained with the MLSA phylogeny. Data obtained revealed that i) the results obtained with these methods are consistent with the genome-scale phylogeny and that ii) the two methods have a different resolution; in particular, the ANI identifies more (putative) species that are found to be clustered together when using the TNF, where the topologies of the dendrograms obtained with these methods show substantial agreement. The fact that the three different methods used in this work provided such similar conclusions is essentially a strong confirmation of the results obtained that can be summarized as a catalogue of inconsistencies with the current taxonomic annotation. In order to fully exploit the information embedded in the genomic sequences, the gene content information has been used to make phylogenetic inferences, which again were found to be coherent with the previous methods. Overall, this suggests that the information embedded in the pattern of orthologs presence can be used to capture the actual phylogenetic relationships.

Although the golden standard for species definition in microbes relies on a polyphasic approach (Vandamme et al., 1996), i.e. requires a combination of molecular and phenotypic tests, these analyses may not always be possible, due to complications in cultivability or due to the experimental efforts required by these taxonomic methods.

By contrast, in the landscape of the post-genomic era, approaches based on the genomic sequences retain several advantages. These methods are faster and cheaper than traditional taxonomic methods and, most importantly, can be easily replicated and applied to uncultivable organisms for which the genomic sequence can be retrieved with technologies like single-cell genomics and/or metagenomics.

In conclusion, the major findings of this work are that (i) a group of three *Pseudoalteromonas* representatives assigned to different species (*P. flavipulchra* JG1, *Pseudoalteromonas* sp. NJ631 and *P. piscicida* JCM 20779) has been consistently found to belong to the same species according to the three methods used. For these reasons, we propose that these strains might be assigned to the same species; (ii) the presence of a group of similar strains probably belonging to the species *P. haloplanktis* (*P. haloplanktis*-like group). Interestingly, most of these strains share a common isolation site (Antarctica) and similar environment/lifestyle (marine environment/association with marine sponges). On the basis of these evidences, we propose to include these strains in the species *P. haloplanktis*; (iii) a group of three strains belonging to different species (*P. undina* NCIMB 2128, *P. marina* mano4 and *P. arctica* A37) is found in the *P. haloplanktis*-like group, possibly meaning that they might be included in the species *P. haloplanktis*.

Acknowledgments This work is supported by 2013 MIUR/PNRA grants (Programma Nazionale di Ricerche in Antartide) 2013/B4.02 and 2013/AZ1.04. Elena Perrin is financed by a "Fondazione Adriano Buzzati-Traverso" fellowship.We are very grateful to two anonymous reviewers for their helpful suggestions and comments in improving the manuscript.

References

Cock, P. J., C. J. Fields, N. Goto, M. L. Heuer & P. M. Rice, 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research 38: 1767–1771.

Corchero, J. L., B. Gasser, D. Resina, W. Smith, E. Parrilli, F. Vázquez, I. Abasolo, M. Giuliani, J. Jäntti, P. Ferrer, M. Saloheimo, D. Mattanovich, S. Schwartz Jr, M. L. Tutino & A. Villaverde, 2013. Unconventional microbial systems for the cost-efficient production of high-quality protein therapeutics. Biotechnology Advances 31: 140–153.

Delcher, A. L, S. L. Salzberg & A. M. Phillippy, 2003. Using MUMmer to identify similar regions in large sequence sets. Current Protocols in Bioinformatics, 10–3.

Eddy, S. R., 2005. A model of the statistical power of comparative genome sequence analysis. PLoS Biology 3: e10.

Egan, S., S. James & S. Kjelleberg, 2002. Identification and characterization of a putative transcriptional regulator controlling the expression of fouling inhibitors in *Pseudoalteromonas tunicata*. Applied and Environmental Microbiology 68: 372–378.

Feller, G., 2013. Psychrophilic enzymes: from folding to function and biotechnology. Scientifica 2013: 1–28.

Galardini, M., A. Mengoni, E. G. Biondi, R. Semeraro, A. Florio, M. Bazzicalupo, A. Benedetti & S. Mocali, 2014. DuctApe: a suite for the analysis and correlation of genomic and OmniLog™ Phenotype Microarray data. Genomics 103: 1–10.

Gauthier, G., M. Gauthier & R. Christen, 1995. Phylogenetic analysis of the genera *Alteromonas*, *Shewanella*, and *Moritella* using genes coding for small-subunit rRNA sequences and division of the genus *Alteromonas* into two genera, *Alteromonas* (emended) and *Pseudoalteromonas*

gen. nov., and proposal of twelve new species combinations. International Journal of Systematic Bacteriology 45: 755–761.

Geuten, K., T. Massingham, P. Darius, E. Smets & N. Goldman, 2007. Experimental design criteria in phylogenetics: where to add taxa. Systematic Biology 56: 609–622.

Giuliani, M., E. Parrilli, F. Sannino, G. A. Apuzzo, G. Marino & M. L. Tutino, 2014. Recombinant production of a single-chain antibody fragment in *Pseudoalteromonas haloplanktis* TAC125. Applied Microbiology and Biotechnology 98: 4887–4895.

Goris, J., K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme & J. M. Tiedje, 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. International Journal of Systematic and Evolutionary Microbiology 57: 81–91.

Holmström, C., D. Rittschof & S. Kjelleberg, 1992. Inhibition of settlement by larvae of *Balanus amphitrite* and *Ciona intestinalis* by a surface-colonizing marine bacterium. Applied and Environmental Microbiology 58: 2111–2115.

Karlin, S., 1998. Global dinucleotide signatures and analysis of genomic heterogeneity. Current Opinion in Microbiology 1: 598–610.

Karlin, S. & C. Burge, 1995. Dinucleotide relative abundance extremes: a genomic signature. Trends in Genetics 11: 283–290.

Kim, M., H. S. Oh, S. C. Park & J. Chun, 2014. Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. International Journal of Systematic and Evolutionary Microbiology 64: 346–351.

Kobayashi, T., 2003. *Pseudoalteromonas sagamiensis* sp. nov., a marine bacterium that produces protease inhibitors. International Journal of Systematic and Evolutionary Microbiology 53: 1807–1811.

Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson & D. G. Higgins, 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23: 2947–2948.

McAuliffe, J. D., M. I. Jordan & L. Pachter, 2005. Subtree power analysis and species selection for comparative genomics. Proceedings of the National Academy of Sciences 102: 7900–7905.

McCarthy, B. J. & E. T. Bolton, 1963. An approach to the measurement of genetic relatedness among organisms. Proceedings of the National Academy of Sciences of the United States of America 50: 156–164.

Médigue, C., E. Krin, G. Pascal, V. Barbe, A. Bernsel, P. N. Bertin, F. Cheung, S. Cruveiller, S. D'Amico, A. Duilio, G. Fang, G. Feller, C. Ho, S. Mangenot, G. Marino, J. Nilsson, E. Parrilli, E. P. Rocha, Z. Rouy, A. Sekowska, M. L. Tutino, D. Vallenet, G. von Heijne & A. Danchin, 2005. Coping with cold: the genome of the versatile marine Antarctica bacterium *Pseudoalteromonas haloplanktis* TAC125. Genome Research 15: 1325–1335.

Paitan, Y. & E. Z. Ron, 2014. Gram-Negative Pathogens: Overview of Novel and Emerging Resistant Pathogens and Drugs. In Antimicrobials, Springer, Berlin Heidelberg.

Papa, R., E. Parrilli, F. Sannino, G. Barbato, M. L. Tutino, M. Artini & L. Selan, 2013. Anti-biofilm activity of the Antarctic marine bacterium *Pseudoalteromonas haloplanktis* TAC125. Research in Microbiology 164: 450–456.

Papaleo, M. C., M. Fondi, I. Maida, E. Perrin, A. Lo Giudice, L. Michaud, S. Mangano, G. Bartolucci, R. Romoli & R. Fani, 2012. Sponge-associated microbial Antarctic communities exhibiting antimicrobial activity against *Burkholderia cepacia* complex bacteria. Biotechnology Advances 30: 272–293.

Papaleo, M. C., R. Romoli, G. Bartolucci, I. Maida, E. Perrin, M. Fondi, V. Orlandini, A. Mengoni, G. Emiliani, M. L. Tutino, E. Parrilli, D. de Pascale, L. Michaud, A. Lo Giudice & R. Fani, 2013. Bioactive volatile organic compounds from Antarctic (sponges) bacteria. New Biotechnology 30: 824–838.

Pardi, F. & N. Goldman, 2005. Species choice for comparative genomics: being greedy works. PLoS Genetics 1: e71.

Parrilli, E., R. Papa, S. Carillo, M. Tilotta, A. Casillo, F. Sannino, A. Cellini, M. Artini, L. Selan, M. M. Corsaro & M. L. Tutino, 2015. Anti-biofilm activity of *Pseudoalteromonas haloplanktis* TAC125 against *Staphylococcus epidermidis* biofilm: evidence of a signal molecule involvement? International Journal of Immunopathology and Pharmacology 28(1): 104–113.

Pride, D. T., R. J. Meinersmann, T. M. Wassenaar & M. J. Blaser, 2003. Evolutionary implications of microbial genome tetranucleotide frequency biases. Genome Research 13: 145–158.

Richter, M. & R. Rosselló-Móra, 2009. Shifting the genomic gold standard for the prokaryotic species definition. Proceedings of the National Academy of Sciences 106: 19126–19131.

Rippa, V., R. Papa, M. Giuliani, C. Pezzella, E. Parrilli, M. L. Tutino, G. Marino & A. Duilio, 2012. Regulated Recombinant Protein Production in the Antarctic bacterium *Pseudoalteromonas haloplanktis* TAC125. In Recombinant Gene Expression. Humana Press, New York.

Rosselló-Móra, R., 2012. Towards a taxonomy of bacteria and archaea based on interactive and cumulative data repositories. Environmental Microbiology 14: 318–334.

Schbath, S., 1997. An efficient statistic to detect over-and under-represented words in DNA sequences. Journal of Computational Biology 4: 189–192.

Schildkraut, C. L., J. Marmur & P. Doty, 1961. The formation of hybrid DNA molecules and their use in studies of DNA homologies. Journal of Molecular Biology 3: 595–616.

Stackebrandt, E. & J. Ebers, 2006. Taxonomic parameters revisited: tarnished gold standards. Microbiology Today 33(4): 152–155.

Stackebrandt, E., W. Frederiksen, G. M. Garrity, P. A. Grimont, P. Kämpfer, M. C. Maiden, X. Nesme, R. Rossello-Mora, J. Swings, H. G. Truper, L. Vauterin, A. C. Ward & W. B. Whitman, 2002. Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. International Journal of Systematic and Evolutionary Microbiology 52: 1043–1047.

Tamura, K., G. Stecher, D. Peterson, A. Filipski & S. Kumar, 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30: 2725–2729.

Teeling, H., J. Waldmann, T. Lombardot, M. Bauer & F. O. Glöckner, 2004a. TETRA: a web-service and a stand-alone program for the analysis and comparison of

tetranucleotide usage patterns in DNA sequences. BMC Bioinformatics 5: 163.

Teeling, H., A. Meyerdierks, M. Bauer, R. Amann & F. O. Glöckner, 2004b. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environmental Microbiology 6: 938–947.

Vandamme, P., B. Pot, M. Gillis, P. De Vos, K. Kersters & J. Swings, 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiological Reviews 60: 407–438.

Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimoxt, O. Kaniiler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr & H. G. Truper, 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. International Journal of Systematic Bacteriology 37: 463–464.

Wilmes, B., A. Hartung, M. Lalk, M. Liebeke, T. Schweder & P. Neubauer, 2010. Fed-batch process for the psychrotolerant marine bacterium *Pseudoalteromonas haloplanktis*. Microbial Cell Factories 9: 72.

Yu, M., K. Tang, J. Liu, X. Shi, T. Gulder, A. M. Tobias & X. H. Zhang, 2013. Genome analysis of *Pseudoalteromonas flavipulchra* JG1 reveals various survival advantages in marine environment. BMC Genomics 14: 707.