BMC Microbiology

RESEARCH ARTICLE   **Open Access**

High-quality genome sequence assembly of R.A73 *Enterococcus faecium* isolated from freshwater fish mucus

Rim El Jeni[1,2†], Kais Ghedira[3†], Monia El Bour[1], Sonia Abdelhak[4], Alia Benkahla[3] and Balkiss Bouhaouala-Zahar[2,5*]

Abstract

Background: Whole-genome sequencing using high throughput technologies has revolutionized and speeded up the scientific investigation of bacterial genetics, biochemistry, and molecular biology. Lactic acid bacteria (LABs) have been extensively used in fermentation and more recently as probiotics in food products that promote health. Genome sequencing and functional genomics investigations of LABs varieties provide rapid and important information about their diversity and their evolution, revealing a significant molecular basis.

This study investigated the whole genome sequences of the *Enterococcus faecium* strain (HG937697), isolated from the mucus of freshwater fish in Tunisian dams. Genomic DNA was extracted using the Quick-GDNA kit and sequenced using the Illumina HiSeq2500 system. Sequences quality assessment was performed using FastQC software. The complete genome annotation was carried out with the Rapid Annotation using Subsystem Technology (RAST) web server then NCBI PGAAP.

Results: The *Enterococcus faecium* R.A73 assembled in 28 contigs consisting of 2,935,283 bps. The genome annotation revealed 2884 genes in total including 2834 coding sequences and 50 RNAs containing 3 rRNAs (one rRNA 16 s, one rRNA 23 s and one rRNA 5 s) and 47 tRNAs. Twenty-two genes implicated in bacteriocin production are identified within the *Enterococcus faecium* R.A73 strain.

Conclusion: Data obtained provide insights to further investigate the effective strategy for testing this *Enterococcus faecium* R.A73 strain in the industrial manufacturing process. Studying their metabolism with bioinformatics tools represents the future challenge and contribution to improving the utilization of the multi-purpose bacteria in food.

Keywords: *Enterococcus faecium*, Bacteriocin, Freshwater fish, Lactic acid bacteria, Whole-genome sequencing

* Correspondence: balkiss.bouhaouala@fmt.utm.tn;
balkiss.bouhaouala@pasteur.utm.tn
†Rim El Jeni and Kais Ghedira contributed equally to this work.
²Laboratory of Venoms and Therapeutic Molecules, Pasteur Institute of Tunis, Tunis, Tunisia
⁵Medical School of Tunis, University of Tunis El Manar, 1007 Tunis, Tunisia
Full list of author information is available at the end of the article

Background

Antibiotic and chemotherapeutic drug use in aquaculture are an important disease control measure in the aquaculture industry [1]. However, antimicrobial use may promote drug-resistant microorganisms emerging and antibiotic residues detection in fish and in the environment [2].

Probiotic LABs are widely used, as an alternative to antibiotics uses, to prevent animal and human bacterial infections [3]. *Enterococcus* is a LABs large genus, ubiquitous, having the capacity to adapt challenging environments. Such species are isolated from different habitats including water (i.e. waste, freshwater, and seawater), soil, plants, and the digestive tract of warm-blooded animals and/or humans [4]. Several studies have demonstrated *Enterococcus faecium* beneficial effects as probiotic in humans, animals, and aquatic culture [5–10].

Strains belonging to the genus *Enterococcus* produce a wide variety of bacteriocins often called enterocins. They have antagonistic properties against a wide range of pathogenic bacteria [11].

This genus of bacteria produces a wide variety of bacteriocins, which are considered to be biological control agents in food, maintaining their organoleptic and nutritional properties. They thus constitute an alternative to the use of chemical additives or physico-chemical treatments used in food industry [12]. In addition, bacteriocins have the advantage of being rapidly digested by proteases in the human digestive tract [13] without producing toxic secondary substances. Bacteriocins can also find applications in the medical sector [14], they can be used as antimicrobial agents in the pharmaceutical industry (Folli et al., 2003). Enterocins (bacteriocins of enterococci) are of bacteriological importance because of their ability to inhibit the growth of members of the genera *Listeria*, *Clostridium*, and *Staphylococcus* responsible of the highest mortality rate (20–30%) compared to other foodborne pathogens [15–17].

Several studies have refined the knowledge on the genomic diversity of probiotic *Enterococcus* strains to elucidate their genomic features responsable for survival in GI tract, antibiotic resistance, virulence factors and the genetic divergence between pathogenic and probiotic *Enterococcus* strains [18–21]. Some knowledge has been acquired on LABs metabolic activities include carbohydrate, protein and lipid metabolisms, and other metabolic activities. LAB needs amino acids and peptides to respond to their nitrogen complex [22]. Amino acids and peptides may be obtained through proteases or proteolysis actions. In such actions, peptides are metabolized to free amino acids and other compounds for further use. Due to the requirements of peptide differences, peptides can either be essential growth promoters or stimulating factors, some strains can grow up independently.

Recently, the preselected *Enterococcus faecium* R. A73 strain isolated from freshwater fish mucus, has proven to have specific probiotic properties [3]. In the current study, the whole-genome sequencing of *Enterococcus faecium* R.A73 strain was performed and investigate the genome contents and gene functions through comparison to related species. All together, results support the findings of the previous study.

Results

E. faecium R.A73 genome annotation

Genome content

The genome of *Enterococcus faecium* R.A73 strain, isolated from Tilapia *Oreochromis niloticus* mucus, has been sequenced using the Illumina HiSeq 2500 system. The present draft genome includes 2,935,283 bases, with a GC content of 38.0%, and was assembled into 28 scaffolds. The Genomic annotations illustrated a total number of 2884 genes, corresponding to 2834 coding sequences (CDSs) and 50 RNAs with single predicted copies of the 16S, 23S, and 5S rRNA genes and 47 predicted tRNAs (Fig. 1). A total of 342 RAST genome sub-systems were identified, with many features of carbohydrates subsystem (Fig. 2), including the genes involved in the metabolism of central carbohydrate, amino sugars, di- and oligosaccharides, the carbon metabolism, organic acids, the fermentation metabolism, sugar alcohols, polysaccharides, and monosaccharides. There are also many amino acids and derivative characteristics of the sub-system, including the lysine, threonine, methionine, and cysteine.
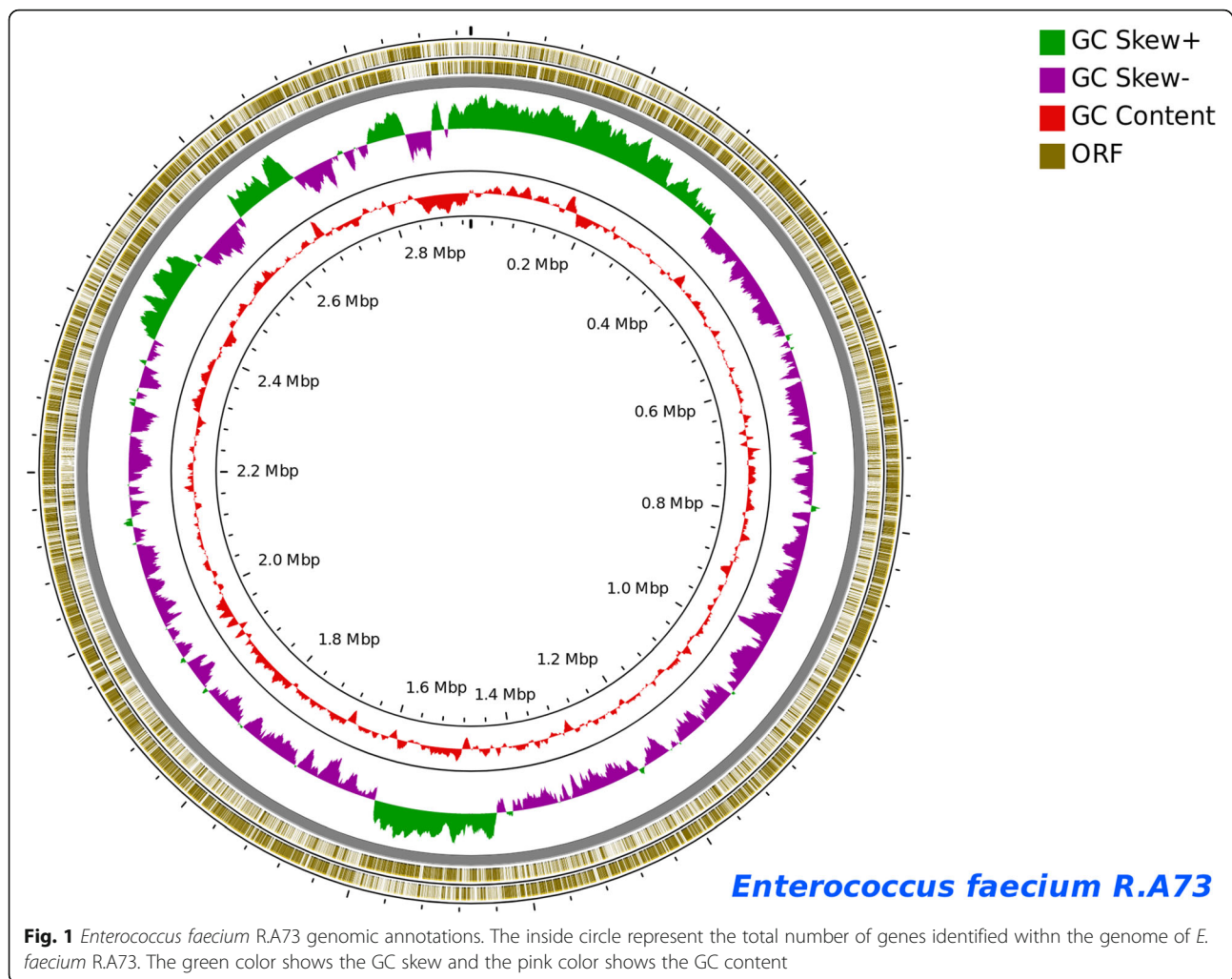
Functional annotation

A total of 2063 protein-coding genes (72.58% of the total protein-coding genes) were assigned a putative function by Clusters of Orthologous Groups (COGs). Genes associated with carbohydrate transport and metabolism (294 Open Reading Frames (ORFs)), translation (206 ORFs), and transcription (205 ORFs) were ranked among the most abundant COG functional categories. The genes distribution into COG functional categories is summarized in (Fig. 2).

Phylogeny and classification

Based on rDNA 16S sequences, the phylogenetic tree showed that the R.A73 strain is more similar to *E. faecium* LMG 11423 and *E. durans* NBRC 100479 than other *Enterococcus* species (Fig. 3).

Moreover, a Genome-to-Genome Distance Calculator (GGDC) was performed for genome-to-genome comparison between R.173 and related strains. DNA-DNA hybridization is considered as the best indicator for distinguishing species. The probabilities of DDH value higher than 70% detected through logistic regression under three formulae indicate that *E. faecium* R.A73 is different from other species of the genus excepting *Enterococcus faecium*. A DDH value > 96% was found

El Jeni *et al. BMC Microbiology*   (2020) 20:322

Page 3 of 12

Fig. 1 *Enterococcus faecium* R.A73 genomic annotations. The inside circle represent the total number of genes identified withfn the genome of *E. faecium* R.A73. The green color shows the GC skew and the pink color shows the GC content

following the comparison against *E. faecium* T110 (Table S1). The later analysis combined to the rDNA 16S based phylogeny method confirmed its identification as *E. faecium* species.

Comparative genomics
Comparative analysis of genome sequences

The comparative genomics help to understand several aspects related to the pathogenicity, the resistance to antibiotics, and probiotic characteristics.
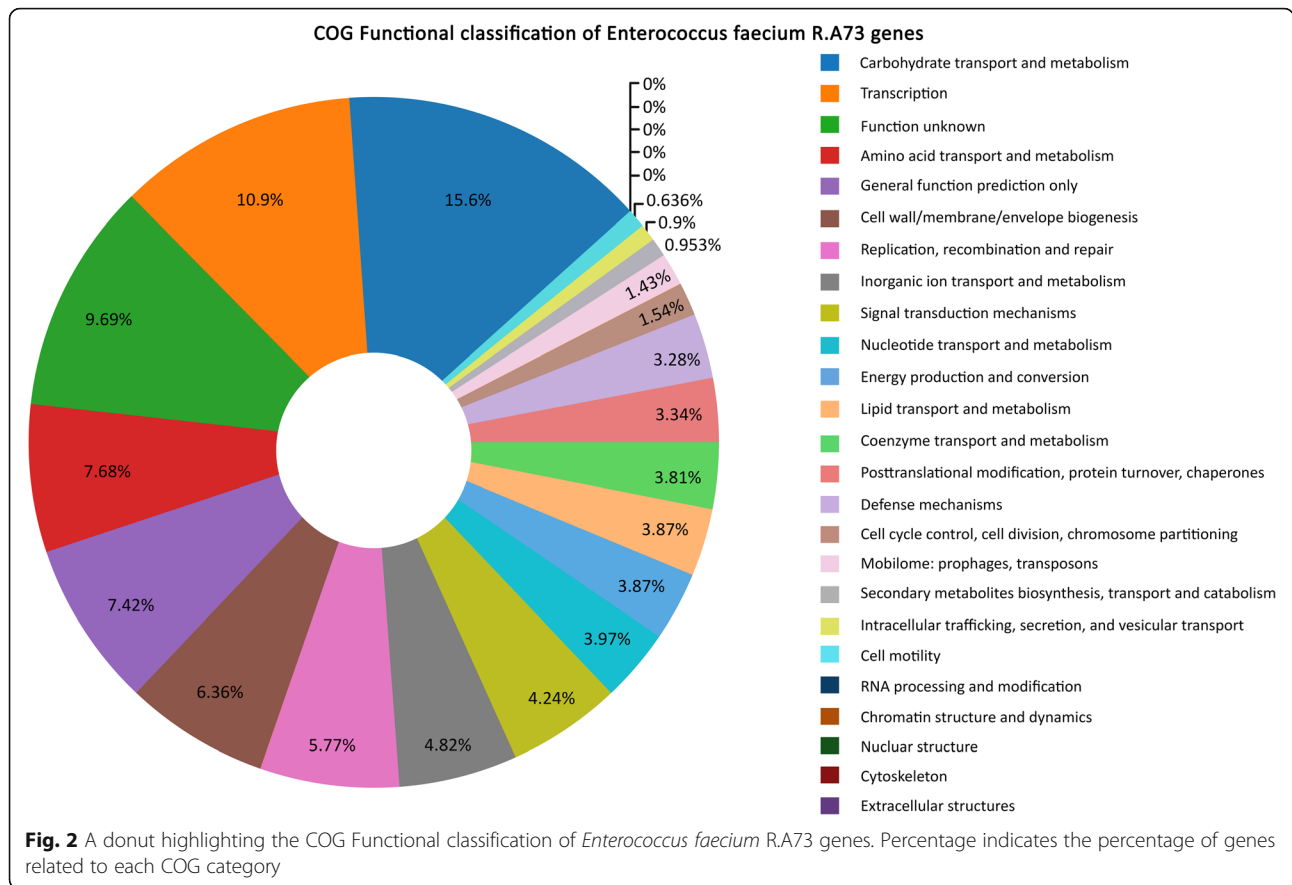
Enterococcus faecium protein sequences predicted by the PGAAP annotation system, have been retrieved and compared with 14 protein sequences of completely sequenced related organisms corresponding to *Enterococcus* 7 L76 uid197170, *Enterococcus casseliflavus* This20 uid55693, *Enterococcus faecalis* 62 159663 uid, *Enterococcus faecalis* D32 171261 uid, *Enterococcus faecalis* og1RF54927 uid, *Enterococcus faecalis Symbioflor* 1 uid183342, *Enterococcus faecalis* V583 uid57669, *Enterococcus faecium* AUS0004 uid87025, *Enterococcus faecium* AUS0085 uid214432, *Enterococcus faecium* do uid55353,

Enterococcus faecium NRRL B 2354 uid188477, *Enterococcus hirae ATCC* 9790 uid70619, *Enterococcus mundtii* that 25 uid229420 and *Enterococcus faecium* T110.

The comparative proteome among enterococcus genomes (Table 1) showed a high similarity between *E. faecium* HG937697 and *E. faecium* T110 genomes with 2,318 common orthologs genes (80.37%). This similarity was confirmed using The BRIG tool (Fig. 4). Specific protein-coding genes (208) were identified in *E. faecium* R.A73 strain.

Comparative analysis of virulence genes

The presence of genes related to virulence in *Enterococcus faecium* R.A73 strain was investigated. Among several *Enterococcus* virulence genes available in the virulence factor database VFDB (http://www.mgc.ac.cn/VFs/), almost 30 genes, including the virulence factor esp. gene (enterococcal surface protein), were absent in *Enterococcus faecium* R.A73 while ebpA (DTX73_01685), ebpB (DTX73_01690), ebpC (DTX73_01695), srtC (DTX73_01700), ecbA (DTX73_00685), efaA (DTX73_03830) were noted.

Fig. 2 A donut highlighting the COG Functional classification of *Enterococcus faecium* R.A73 genes. Percentage indicates the percentage of genes related to each COG category

Bacteriocin and antibacterial peptide production genes

Several genes involved in bacteriocin production as well antibacterial peptides were identified in *Enterococcus faecium* R.A73 strain (Table S2). These genes include colicin V CvpA family protein (DTX73_02515), production bacteriocin pole, antibacterial peptides agents synthesis, bacteriocin-associated protein (DTX73_02925), bacteriocin immunity protein (DTX73_04250, DTX73_06025, DTX73_06505), bacteriocin (DTX73_04255, DTX73_06475, DTX73_06480, DTX73_07350, DTX73_09680, DTX73_09720), EntF family bacteriocin induction factor (DTX73_06500), TmhB bacteriocin enhancer peptide (DTX73_09690), ThmA bacteriocin (DTX73_09695), ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain (DTX73_09710), class IIb bacteriocin, lactobin A/cerein 7B family (DTX73_09720). Other genes possess different roles implicated in amidophosphoribosyl transferase (EC 2.4.2.14) (DTX73_12820), acetyl-coenzyme A chain carboxyl beta transferase (EC 6.4.1.2) (DTX73_04140), a synthase dihydrofolate (EC 6.3.2.12) (DTX73_05510), an rRNA pseudouridine synthase a (EC 4.2.1.70) (DTX73_10445) and the bifunctional folylpolyglutamate synthase/dihydrofolate synthase (EC 6.3.2.17) (DTX73_05510).
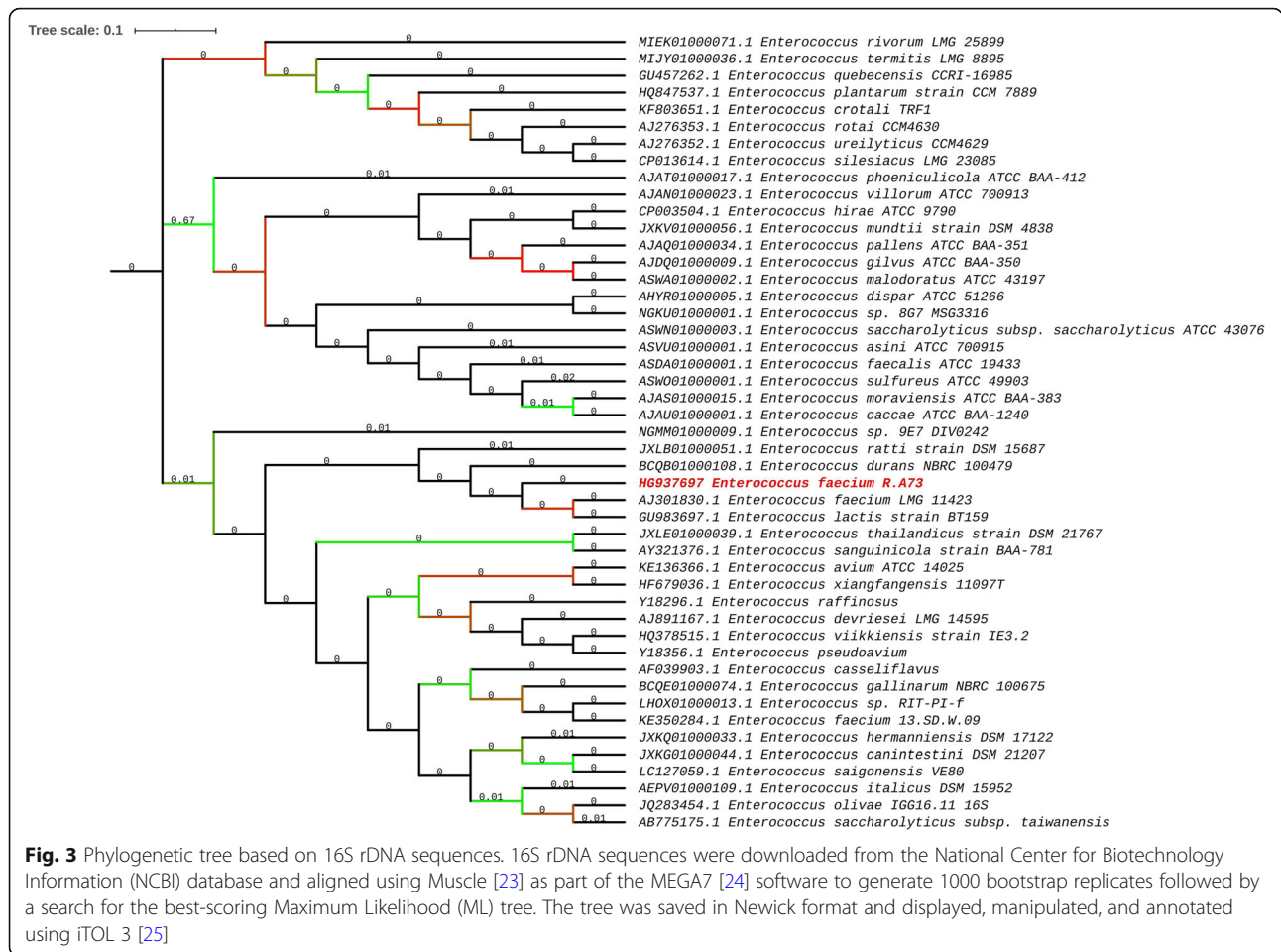
Furthermore, the genome revealed the presence of a gene encoding for one enterocin (DTX73_06510).

Antibiotics resistance

Two genes involved in resistance to antibiotics and toxic compounds were identified. These genes correspond to an homolog of aac (6′)-Ii involved in Aminoglycoside resistance (% identity: 98.36; Query/HSP length: 549/549; Accession number: L12710) and a homolog to msr(C) involved in MLS - Macrolide, Lincosamide and Streptogramin B (% identity: 97.70; Query/HSP length: 1479/1479; Accession number: AF313494). Besides, PGAAP and RAST annotation systems were also able to detect 52 other genes potentially involved in virulence, disease, and defense mechanisms. These genes found in the HG937697 genome are presented in (Table 2).

Discussion

A genomics study was performed in a preselected *Enterococcus faecium* R.A73 strain, isolated from freshwater fish mucus, displaying potential probiotic characteristics and significant efficiency as food additives. The complete genome annotation revealed that the bacteria R.A73

El Jeni *et al. BMC Microbiology* (2020) 20:322

Page 5 of 12

Fig. 3 Phylogenetic tree based on 16S rDNA sequences. 16S rDNA sequences were downloaded from the National Center for Biotechnology Information (NCBI) database and aligned using Muscle [23] as part of the MEGA7 [24] software to generate 1000 bootstrap replicates followed by a search for the best-scoring Maximum Likelihood (ML) tree. The tree was saved in Newick format and displayed, manipulated, and annotated using iTOL 3 [25]

genome did not have any plasmid which may be due to growing temperature, copies number, or even isolation methods [26].

Several carbohydrate subsystem features were identified in *Enterococcus faecium* R.A73 strain genome. It has been proven that carbohydrates degradation and their related compounds are mainly responsible for the primary metabolic activity of LAB, generating energy and carbon source molecules [27, 28]. The genome annotation for the strain under study suggests an abundance of metabolic activities such as proteins, lipids, and other compounds decomposition, which are important for LAB growth. Interestingly, many amino acids and derivatives characteristic of the subsystem, including lysine, threonine, methionine, and cysteine, are found in the genoma of *Enterococcus faecium* R.A73 strain. LAB amino acid requirements are strain-dependent with a large range of species differences [29, 30]. *Enterococcus faecium* have the ability to use a wide range of mono-, di-, oligo-saccharides and therefore they have an enriched carbohydrate metabolism [4, 31] as well as using a variety of carbohydrates has been shown to be among

properties associated to probiotic strains [32]. Furthermore, 51 genes out of 208 were assigned to COG functional categories associated with carbohydrate transport and metabolism (6 genes), amino acid transport and metabolism (6 genes), and cell wall/membrane/envelope biogenesis (5 genes).

The presence of the prophage in the genome of *E. faecium* R.A73 strain was predictable. Bacteriophages contribute to the evolution of bacteria through their integration into the genome, *E. faecium* bacteria are known to harbour bacteriophages [33].

Protein-coding for ABC transporters have been detected, they are known to have an antibacterial activity that may contribute to probiotic potential in such strains [34].

Enterococcus faecium R.A73 strain genome identified 22 genes involved in bacteriocin production as well as antimicrobial peptides. The gene involved in colicin V (Col V) has been identified. Col V is an antibiotic-like peptide that kills susceptible cells by disrupting their potential membrane once it reaches the periplasmic inner membrane. It is secreted by some members of enterobacteria to kill closely related bacterial cells, thus reducing competition for essential nutrients [35, 36].

Table 1 Genome size and gene count of 14 pathogens and probiotics *Enterococcus* species used in genome comparative study

Species	Genome size (Mb)	Gene count
Enterococcus 7 L76 uid197170	3.09	2295
Enterococcus casseliflavus This20 uid55693		
Enterococcus faecalis 62,159,663 uid	3.13	3158
Enterococcus faecalis D32 171,261 uid	3.06	3174
Enterococcus faecalis og1RF54927 uid	2.73	2676
Enterococcus faecalis Symbioflor 1 uid183342	2.81	2761
Enterococcus faecalis V583 uid57669	3.35	3412
Enterococcus faecium AUS0004 uid87025	3.01	3118
Enterococcus faecium AUS0085 uid214432	3.23	3318
Enterococcus faecium do uid55353	3.05	3209
Enterococcus faecium NRRL B 2354 uid188477	2.84	2704
Enterococcus hirae ATCC 9790 uid70619	2.85	2752
Enterococcus mundtii that 25 uid229420	3.35	3229
Enterococcus faecium T110	2.73	2606

This protein was shared by several Enteroccus strains including *Enterococcus faecium* DO (WP_002295088.1).

The comparative proteomes analysis showed 208 unique genes detected in *E. faecium* R.A73 strain which including five bacteriocins (bacteriocin (DTX73_07350, DTX73_09680), ThmB bacteriocin enhancer peptide (DTX73_09690), ThmA bacteriocin (DTX73_09695), ABC-type bacteriocin/lantibiotic exporters, contain a N-terminal double-glycine peptidase domain (DTX73_09710), class IIb bacteriocin, lactobin A/cerein 7B family (DTX73_09720)). Lantibiotics that constitute a group of bacteriocins were shown to have several pharmaceutical applications including Blood pressure treatment, inflammations and allergies treatment, Skin, mastitis, herpes infections treatment, dental caries treatment, and peptic ulcer treatment. In R.A73 an ABC-type bacteriocin/lantibiotic exporters-like wasfound that contains an N-terminal double-glycine peptidase domain (DTX73_09710). Moreover, ThmA/ThmB (DTX73_09695/DTX73_09690) which are known as termophilin 13 that are produced by *S. thermophiles* SPi13 possesses natural antimicrobial activities [36–38].

Comparative proteome analysis showed that R.A73 strain was closely related to the probiotic strain T110 (Fig. 4). This latter is a commercially probiotic widely prescribed for humans, animals, and aquaculture [8]. It is a content of many commercial available probiotics and no cause of illness or death has been reported [8].

In ordre to understand if *Enterococcus faecium* R.A73 harboured resistance genes, the screening of antibiotic-resistance was done. Some virulent genes were found, in *Enterococcus faecium* R. A73 strain, highly homologs to ebpA (DTX73_01685), ebpB (DTX73_01690), ebpC (DTX73_01695), srtC (DTX73_017000), ecbA (DTX73_ 00685), efaA (DTX73_03830), aac (6′)-Ii and msr(C). The virulence gene scm, efaA and srtC are not well characterized as virulence determinants in *E. faecium* [8]. Likewise, aac (6′)-Ii and msr(C) genes are species specific and could be useful for detection and identification of *E. faecium* species [39, 40].

The R.A73 strain may be categorized as antimicrobial resistance (AMR) because in previous study [3] it was found to be resistant to several antibiotics (oxacillin, streptomycin, cefazolin and clindamycin). However, Enterococcus may acquire resistance to some antibiotics via the presence of intrinsic genes related to their innate resistance as well as through horizontal genes transfer [41, 42]. The latter mechanism can lead as well as the ability to aquire certain adaptive genetic traits, such as (AMR) determinants [43]. In Japan, Enterococcus strains used as probiotics have shown resistance to tetracyclines and betalactams [44].

Previous study has investigated the probiotic properties of Enterococcus strains isolated from artisanal dairy products [45]. The most important virulence factors investigated include cylA, cylB and cylM, esp., agg, gelE, cpd, ccf, and cad genes. These later are responsible for the cytosilin transportation and activation, application in modification of post-translational proteins, immune evasion, adherence to eukaryotic cells, the production of toxin which hydrolyzes gelatin, and finally sex pheromones which are responsible for facilitating conjugation [6, 46]. No genes belonging to the aforementioned list was found in R.A73. The same study showed that probiotics investigated strains demonstrated hydrophobicity activity, auto-aggregation, and adhesion ability to the human intestinal cell line contributing to the gut colonization.
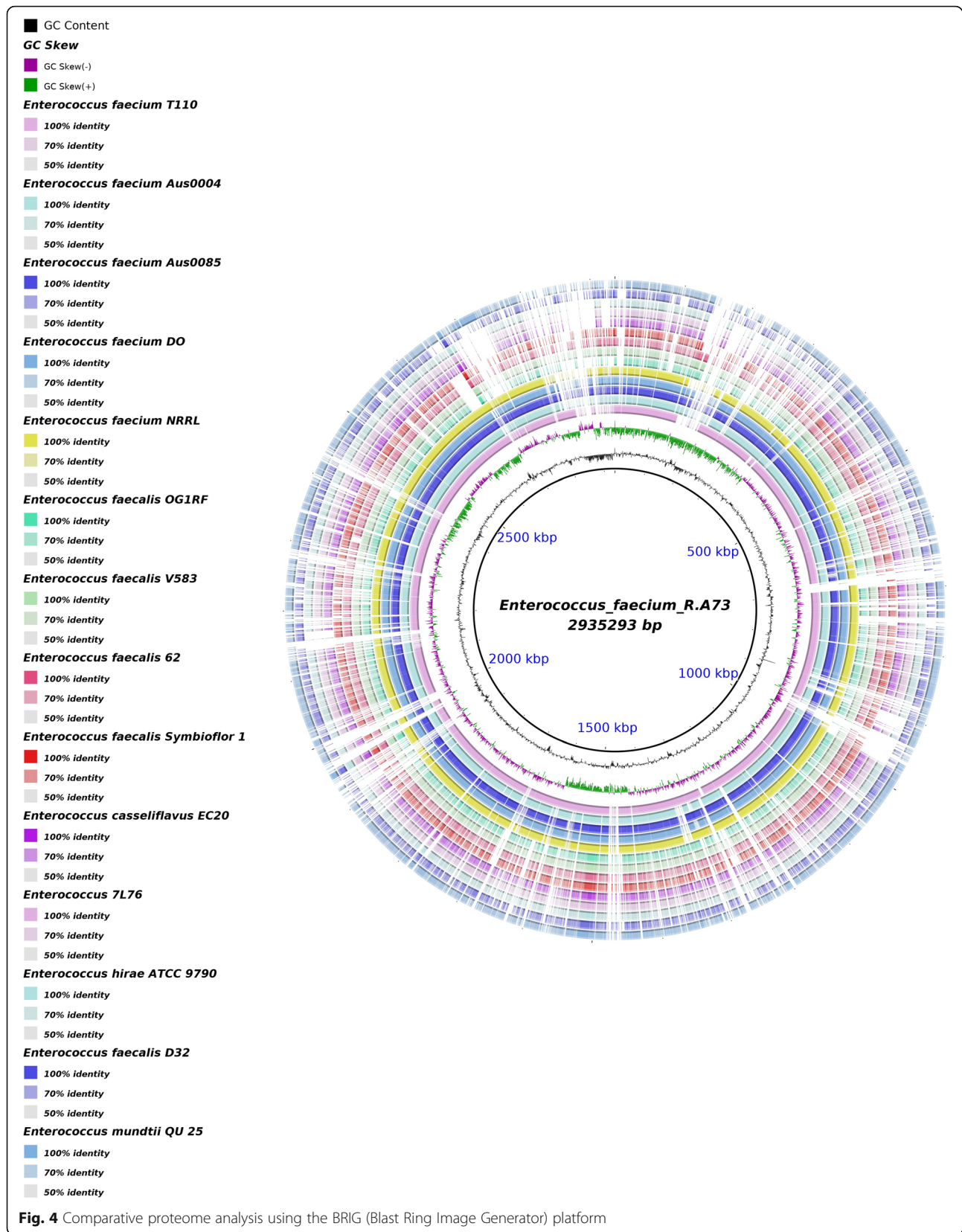
El Jeni *et al. BMC Microbiology*   (2020) 20:322

Page 7 of 12

GC Content

GC Skew

GC Skew(-)

GC Skew(+)

Enterococcus faecium T110

100% identity

70% identity

50% identity

Enterococcus faecium Aus0004

100% identity

70% identity

50% identity

Enterococcus faecium Aus0085

100% identity

70% identity

50% identity

Enterococcus faecium DO

100% identity

70% identity

50% identity

Enterococcus faecium NRRL

100% identity

70% identity

50% identity

Enterococcus faecalis OG1RF

100% identity

70% identity

50% identity

Enterococcus faecalis V583

100% identity

70% identity

50% identity

Enterococcus faecalis 62

100% identity

70% identity

50% identity

Enterococcus faecalis Symbioflor 1

100% identity

70% identity

50% identity

Enterococcus casseliflavus EC20

100% identity

70% identity

50% identity

Enterococcus 7L76

100% identity

70% identity

50% identity

Enterococcus hirae ATCC 9790

100% identity

70% identity

50% identity

Enterococcus faecalis D32

100% identity

70% identity

50% identity

Enterococcus mundtii QU 25

100% identity

70% identity

50% identity

Fig. 4 Comparative proteome analysis using the BRIG (Blast Ring Image Generator) platform

Table 2 Number of genes in *Enterococcus faecium* R.A73 genome associated with the general COG functional categories

Code	Number of genes	% Of total features	Description
J	206	7.14	Translation
A	0	0.0	RNA processing and modification
K	205	7.10	Transcription
L	109	3.77	Replication, recombination, and repair
B	0	0.0	Chromatin structure and dynamics
D	29	1.00	Cell cycle control, mitosis and meiosis
Y	0	0.0	Nuclear structure
V	62	2.14	Defense mechanisms
T	80	2.77	Signal transduction mechanisms
M	120	4.16	Cell wall/membrane biogenesis
N	12	0.41	Cell motility
Z	0	0.0	Cytoskeleton
W	0	0.0	Extracellular structures
U	17	0.58	Intracellular trafficking and secretion
O	63	2.18	Posttranslational modification, protein turnover, chaperones
C	73	2.53	Energy production and conversion
G	294	10.19	Carbohydrate transport and metabolism
E	145	5.02	Amino acid transport and metabolism
F	75	2.60	Nucleotide transport and metabolism
H	72	2.49	Coenzyme transport and metabolism
I	73	2.53	Lipid transport and metabolism
P	91	3.15	Inorganic ion transport and metabolism
Q	18	0.62	Secondary metabolites biosynthesis transport and catabolism
R	140	4.8	General function prediction only
S	183	6.34	Function unknown
–	821	28.48	Not in COGs

Indeed, some of the main selection criteria for potential probiotics is their ability to adhere to the gastrointestinal tract in order to exert their probiotic effects for an extended time [47]. However, adhesion is as well considered as a potential virulence factor for pathogenic bacteria [48]. Therefore, ebpA, ebpB and ebpC are classified as virulence determinants but they are in fact adherence factors. *Ebp* genes may play a role during colonization of the mammalian host, adherence to abiotic surfaces, or bacterial surface components [49].

Conclusion

Marine microbiology fields are still evolving and significant progress can be expected on marine pollution issues including bacterial oil degradation, which is under investigation at present. The current results respond to potential probiotic properties. *Enterococcus faecium* R.A73 strain can be safely used as bio-ingredients in conservation and fish processing consumed by humans and animals. However, further studies are needed for comprehensive identification of AMR genes in the probiotic strains.

Methods

Bacterial strain

In total, 177 LABs have been isolated from different organs (intestine, skin, gills and mucus) in freshwater fish (*Mugil cephalis* and *Oreochromis niloticus*). Within this collection, the novel R.A73, isolated from Tilapia *Oreochromis niloticus* mucus, was identified as *Enterococcus faecium*, exhibited high inhibitory activities against food-borne pathogens and spoilage microbial species and has significant in vitro probiotic profiles [3].

Growth conditions and DNA preparation/isolation

Enterococcus faecium R.A73 was inoculated in De Man-Rogosa-Sharpe (MRS) broth for 48 h at 20 °C. Pure genomic DNA was then extracted using the Quick-GDNA

kit (Zymo Research) and subsequently sent it to the platform service "BaseClear" in Netherlands, for whole genome sequencing.

Genome sequencing
Enterococcus faecium R.A73 strain genome has been sequenced using the Illumina HiSeq 2500 system. FAST Q paired-end sequence data files have been generated using the Illumina CASAVA pipeline version 1.8.3. Initial quality assessment was based on the data that passed Illumina chastity filtering. Readings with adapters and/or the PhiX control signal were then deleted. The second assessment of quality based on the remaining reads was performed using the FASTQC quality control tool version 0.10.0. FASTQ sequence quality has been enhanced by removing the low-quality bases, with the" Trim Sequences" options from CLC Genomics Version 7.0.4.

De novo assembly
The quality-filtered sequence reads were assembled in some contig sequences. The analysis was carried out by using the option "*De novo* Assembly" in the genomics workshop CLC version 7.0.4. The optimal k-mer size was automatically determined using KmerGenie [50]. Contigs were then linked to each other's and put into scaffolds or supercontigs. The orientation, order, or distance between the contigs was estimated by using the insert size between the paired-end.

The scaffolding has been performed using the SSPACE Premium scaffolder version 2.3 [51]. Gapped regions within the scaffolds were partially closed in an automated manner using GapFiller version 1.10 [52]. The method takes advantage of the insert size between the paired-end reads.

Genome annotation
The RAST web server was used [53] to perform genome annotation. Briefly, protein-coding genes were predicted using the Classic RAST annotation scheme [53]. RNAmmer tool [54] was used to predict ribosomal RNAs, while tRNAs can-SE [55] was used to detect transfer RNAs. The NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) was used to perform a final annotation.

Functional annotation
Clusters of Orthologous Group were assigned based on comparative proteomes analysis against the COG database [56] using protein sequences that have previously been predicted by PGAAP. Briefly, using the best reciprocal hits approach with an e-value <= 1E-05, protein sequences were retrieved and compared against the protein sequences available in the COG database.

Phylogenetic analysis and genome-to-genome distance calculation
Identification of closely related strains to *E. faecium* R.A73 was performed based on Basic Local Alignment Search Tool (BLAST) searches and pairwise global sequence alignments through the well-curated EzTaxon database; which covers not only type strains of prokaryotic species with validly published names but also phylotypes that may represent species in nature. The 16S rDNA gene sequences with pairwise similarity higher than 96% to *E. faecium* R.A73 (locus_tag = "DTX73_13310") were chosen for phylogenetic tree construction. 16S rDNA sequences were downloaded from the National Center for Biotechnology Information (NCBI) database. They were aligned using Muscle [23] as part of the MEGA7 [24] software to generate 1000 bootstrap replicates followed by a search for the best-scoring Maximum Likelihood (ML) tree. This latter was displayed, manipulated, and annotated using iTOL 3 [25]. Digital DDH similarities between *the E. faecium* R.A73 genome and those of other *Enterococcus* species were calculated using the GGDC web server version 2.0 under the recommended setting [57].

Comparative genomics
Genome comparison of *E. faecium* HG937697 strain with related species was performed using BRIG (Blast Ring Image Generator), an open-source multi-platform software application, which displays multi-genome comparisons and similarity between the reference genome at the center of one image compared to other related genomes listed in (Table 1), in the form of a concentric colored ring set according to BLAST identity [58].

Furthermore, protein sequences of *E. faecium* R.A73 strain that were predicted by RAST and PGAAP annotation system were extracted and compared to protein sequences of the proteomes of related *Enterococcus* cited in (Table 1). The comparison was computed using Inparanoid (http://InParanoid.sbc.su.se) [59] then MultiParanoid (http://multiparanoid.cgb.ki.se/) [60] Perl programs to identify the cluster of orthologous genes between pairs of species than between all the species, respectively.

Bacteriocin genes identification
Gene annotaion performed with PGAAP and RAST server annotation [61] allowed to identify genes encoding for bacteriocins and related products in the *E. faecium* R.A73 strain. The comparison of protein sequences between related probiotic enterococcus strains led to the identification of bacteriocins orthologuous proteins in R.A73 strain. Furthemore, R.A73 protein sequences were compared to all bacteriocins protein sequences available in Bactibase database (http://bactibase.hammamilab.org/bacteriocinslist.php?view=GeneralView) [62].

El Jeni *et al. BMC Microbiology* (2020) 20:322

Page 10 of 12

Antibiotics resistance and virulence genes

The ResFinder-2.1 server [63] available at cge.cbs.dtu.dk/ services/ResFinder/ in combination with PGAAP and RAST server annotation [61] was used to investigate genes involved in resistance to antibiotics and toxic compounds in the *E. faecium* R.A73 strain.

Genbank submission

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession QOVC00000000. The version described in this paper is version QOVC01000000.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10. 1186/s12866-020-01980-8.

> **Additional file 1: Table S1.** The DDH probabilities to distinguish between R.A73 strain and reference strains that belonged to the similar Enterococcus genus.
>
> **Additional file 2: Table S2.** The Blastp output results between R.A73 protein sequences and BACTIBASE (a database dedicated to bacteriocins) protein sequences.

Abbreviations

RAST: Rapid Annotation using Subsystem Technology; NCBI: National Center for Biotechnology Information; ML: Maximum Likelihood; BRIG: Blast Ring Image Generator; LAB: Lactic Acid Bacteria; CDSs: Coding Sequences; COGs: Clusters of Orthologous Groups; ORFs: Open Reading Frames; GGDC: Genome-to-Genome Distance Calculator; Col V: Colicin V; AMR: Antimicrobial Resistance; MRS: De Man-Rogosa-Sharpe; PGAAP: Prokaryotic Genomes Automatic Annotation Pipeline; BLAST: Basic Local Alignment Search Tool

Authors' contributions

REJ carried out the microbiological, molecular experiments and generated the data; REJ, KG carried out bioinformatic data analysis and interpretation and drafted the manuscript; MEB contributed to the microbiological data analysis/interpretation in accordance with genomic data; SA, ABK facilitated the NGS designed experiments; BBZ designed the experiments and contributed to the data analysis/interpretation and final manuscript preparation. All authors read and approved the final manuscript.

Availability of data and materials

The data has been deposited into EMBL GenBank and the strain *E. faecium* R.A73 identified with the accession number EMBL: HG937697.

Ethics approval and consent to participate

A local ethics committee ruled that no formal ethics approval was required in this particular case.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Laboratory of Microbiology and Pathology of Aquatic Organisms, Institut National des Sciences et Technologies de la Mer (INSTM), Tunis, Tunisia. [2]Laboratory of Venoms and Therapeutic Molecules, Pasteur Institute of Tunis, Tunis, Tunisia. [3]Bioinformatics and Biostatistics Laboratory (LR16IPT09), Pasteur Institute of Tunis, Tunis, Tunisia. [4]Biomedical Genomics and Oncogenetics Laboratory LR16IPT05, Pasteur Institute of Tunis, Tunis, Tunisia. [5]Medical School of Tunis, University of Tunis El Manar, 1007 Tunis, Tunisia.

References

1. Amábile-cuevas ACF, Cárdenas-garcía M, Ludgar M, Am CF, Maura C, Ludgar M. Much faster than ways to control resistance. Am Sci. 1995;83:320–9.
2. van der Waaij D, Nord CE. Development and persistence of multi-resistance to antibiotics in bacteria; an analysis and a new approach to this urgent problem. Int J Antimicrob Agents. 2000;16:191–7. https://doi.org/10.1016/S0924-8579(00)00227-2.
3. El-Jeni R, El Bour M, Calo-Mata P, Böhme K, Fernández-No IC, Barros-Velázquez J, et al. In vitro probiotic profiling of novel Enterococcus faecium and Leuconostoc mesenteroides from Tunisian freshwater fishes. Can J Microbiol. 2015;62:60–71.
4. Manero A, Blanch AR. Identification of Enterococcus spp. with a biochemical key. Appl Environ Microbiol. 1999;65:4425–30.
5. Bednorz C, Oelgeschläger K, Kinnemann B, Hartmann S, Neumann K, Pieper R, et al. The broader context of antibiotic resistance: zinc feed supplementation of piglets increases the proportion of multi-resistant Escherichia coli in vivo. Int J Med Microbiol. 2013;303:396–403. https://doi.org/10.1016/j.ijmm.2013.06.004.
6. Liu W, Zhang L, Shi J, Yi H, Zhang Y, Zhang S, et al. Assessment of the safety and applications of bacteriocinogenic Enterococcus faecium Y31 as an adjunct culture in north-eastern Chinese traditional fermentation paocai. Food Control. 2015;50:637–44. https://doi.org/10.1016/j.foodcont.2014.10.004.
7. Barbosa J, Borges S, Teixeira P. Selection of potential probiotic Enterococcus faecium isolated from Portuguese fermented food. Int J Food Microbiol. 2014;191:144–8. https://doi.org/10.1016/j.ijfoodmicro.2014.09.009.
8. Natarajan P, Parani M. First complete genome sequence of a probiotic Enterococcus faecium strain T-110 and its comparative genome analysis with pathogenic and non-pathogenic Enterococcus faecium genomes. J Genet Genomics. 2015;42:43–6. https://doi.org/10.1016/j.jgg.2014.07.002.
9. Lodemann U, Strahlendorf J, Schierack P, Klingspor S, Aschenbach JR, Martens H. Effects of the probiotic Enterococcus faecium and pathogenic Escherichia coli strains in a pig and human epithelial intestinal cell model. Scientifica (Cairo). 2015;2015:1–10.
10. Klingspor S, Martens H, Çaushi D, Twardziok S, Aschenbach JR, Lodemann U. Characterization of the effects of Enterococcus faecium on intestinal epithelial transport properties in piglets. J Anim Sci. 2013;91:1707–18.
11. Campos CA, Rodríguez Ó, Calo-Mata P, Prado M, Barros-Velázquez J. Preliminary characterization of bacteriocins from Lactococcus lactis, Enterococcus faecium and Enterococcus mundtii strains isolated from turbot (Psetta maxima). Food Res Int. 2006;39:356–64. https://doi.org/10.1016/j.foodres.2005.08.008.
12. Javed I, Ahmed S, Manam S, Riaz M, Ahmad B, Ishtiaq Ali M, et al. Production, characterization, and antimicrobial activity of a bacteriocin from newly isolated enterococcus faecium IJ-31. J Food Prot. 2010;73:44–52.
13. Chen H, Hoover D. Bacteriocins and their food applications. Compr Rev Food Sci Food Saf. 2003;2:82–100.
14. Turcotte C, Lacroix C, Kheadr E, Grignon L, Fliss I. A rapid turbidometric microplate bioassay for accurate quantification of lactic acid bacteria bacteriocins. Int J Food Microbiol. 2004;90:283–93. https://doi.org/10.1016/S0168-1605(03)00315-5.

15. Aymerich T, Garriga M, Ylla J, Vallier J, Monfort JM, Hugas M. Application of enterocins as biopreservatives against Listeria innocua in meat products. J Food Prot. 2000;63:721–6.

16. Giraffa G, Carminati D, Tarelli GT. Inhibition of listeria innocua in milk by bacteriocin-producing enterococcus faecium 7C5. J Food Prot. 1995; 58:621–3.

17. Moreno MRF, Leisner JJ, Tee LK, Ley C, Radu S, Rusul G, et al. Microbial analysis of Malaysian tempeh, and characterization of two bacteriocins produced by isolates of Enterococcus faecium. J Appl Microbiol. 2002;92:147–57.

18. Aziz F, Khan MN, Ahmed S, Andrews SC. Draft genome sequence of Enterococcus faecium SP15, a potential probiotic strain isolated from spring water. BMC Res Notes. 2019;12:10–2. https://doi.org/10.1186/s13104-019-4136-0.

19. Domann E, Hain T, Ghai R, Billion A, Kuenne C, Zimmermann K, et al. Comparative genomic analysis for the presence of potential enterococcal virulence factors in the probiotic Enterococcus faecalis strain Symbioflor 1. Int J Med Microbiol. 2007;297:533–9.

20. Li B, Zhan M, Evivie SE, Jin D, Zhao L, Chowdhury S, et al. Evaluating the safety of potential probiotic Enterococcus durans KLDS6.0930 using whole genome sequencing and oral toxicity study. Front Microbiol. 2018;9:1–15.

21. Bonacina J, Suárez N, Hormigo R, Fadda S, Lechner M, Saavedra L. A genomic view of food-related and probiotic Enterococcus strains. DNA Res. 2017;24:11–24.

22. Savijoki K, Ingmer H, Varmanen P. Proteolytic systems of lactic acid bacteria. Appl Microbiol Biotechnol. 2006;71:394–406.

23. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004;32:1792–7.

24. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33:1870–4.

25. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 2016;44:W242–5.

26. Casey J, Daly C, Fitzgerald GF. Chromosomal integration of plasmid DNA by homologous recombination in Enterococcus faecalis and Lactococcus lactis subsp. lactis hosts harboring Tn919. Appl Environ Microbiol. 1991;57:2677–82.

27. Sánchez B, Chaignepain S, Schmitter JM, Urdaci MC. A method for the identification of proteins secreted by lactic acid bacteria grown in complex media. FEMS Microbiol Lett. 2009;295:226–9.

28. Hoefnagel MHN, Starrenburg MJC, Martens DE, Hugenholtz J, Kleerebezem M, Van Swam II, et al. Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology. 2002;148:1003–13.

29. Letort C, Juillard V. Development of a minimal chemically-defined medium for the exponential growth of Streptococcus thermophilus. J Appl Microbiol. 2001;91:1023–9.

30. Tomita S, Saito K, Nakamura T, Sekiyama Y, Kikuchi J. Rapid discrimination of strain-dependent fermentation characteristics among Lactobacillus strains by NMR-based metabolomics of fermented vegetable juice. PLoS One. 2017; 12(7):e0182229.

31. Devriese LA, Van De Kerckhove A, Kilpper-Balz R, Schleifer KH. Characterization and identification of Enterococcus species isolated from the intestines of animals. Int J Syst Bacteriol. 1987;37:257–9.

32. Ghattargi VC, Nimonkar YS, Burse SA, Davray D, Kumbhare SV, Shetty SA, et al. Genomic and physiological analyses of an indigenous strain, Enterococcus faecium 17OM39. Funct Integr Genomics. 2018;18:385–99.

33. Matos RC, Lapaque N, Rigottier-Gois L, Debarbieux L, Meylheuc T, Gonzalez-Zorn B, et al. Enterococcus faecalis Prophage dynamics and contributions to pathogenic traits. PLoS Genet. 2013;9:e1003539.

34. Lebeer S, Vanderleyden J, De Keersmaecker SCJ. Genes and molecules of lactobacilli supporting probiotic action. Microbiol Mol Biol Rev. 2008; 72:728–64.

35. Gérard F, Pradel N, Wu LF. Bactericidal activity of colicin V is mediated by an inner membrane protein, SdaC, of Escherichia coli. J Bacteriol. 2005;187: 1945–50.

36. Marciset O, Jeronimus-Stratingh MC, Mollet B, Poolman B. Thermophilin 13, a nontypical antilisterial poration complex bacteriocin, that functions without a receptor. J Biol Chem. 1997;272:14277–84.

37. Gilbreth SE, Somkuti GA. Thermophilin 110: a bacteriocin of Streptococcus thermophilus ST110. Curr Microbiol. 2005;51:175–82.

38. Rossi F, Marzotto M, Cremonese S, Rizzotti L, Torriani S. Diversity of Streptococcus thermophilus in bacteriocin production; inhibitory spectrum and occurrence of thermophilin genes. Food Microbiol. 2013;35:27–33. https://doi.org/10.1016/j.fm.2013.02.006.

39. Costa Y, Galimand M, Leclercq R, Duval J, Courvalin P. Characterization of the chromosomal aac (6′)-ii gene specific for Enterococcus faecium. Antimicrob Agents Chemother. 1993;37:1896–903.

40. Portillo A, Ruiz-Larrea F, Zarazaga M, Alonso A, Martinez JL, Torres C. Macrolide resistance genes in Enterococcus spp. Antimicrob Agents Chemother. 2000;44:967–71.

41. Tong J, Lu X, Zhang J, Sui Q, Wang R, Chen M, et al. Occurrence of antibiotic resistance genes and mobile genetic elements in enterococci and genomic DNA during anaerobic digestion of pharmaceutical waste sludge with different pretreatments. Bioresour Technol. 2017;235:316–24. https://doi.org/10.1016/j.biortech.2017.03.104.

42. Hegstad K, Mikalsen T, Coque TM, Werner G, Sundsfjord A. Mobile genetic elements and their contribution to the emergence of antimicrobial resistant Enterococcus faecalis and Enterococcus faecium. Clin Microbiol Infect. 2010; 16:541–54. https://doi.org/10.1111/j.1469-0691.2010.03226.x.

43. Mikalsen T, Pedersen T, Willems R, Coque TM, Werner G, Sadowy E, et al. Investigating the mobilome in clinically important lineages of enterococcus faecium and enterococcus faecalis. BMC Genomics. 2015;16:1–16.

44. Yamaguchi T, Miura Y, Matsumoto T. Antimicrobial susceptibility of Enterococcus strains used in clinical practice as probiotics. J Infect Chemother. 2013;19:1109–15. https://doi.org/10.1007/s10156-013-0633-6.

45. Nami Y, Bakhshayesh RV, Jalaly HM, Lotfi H, Eslami S, Hejazi MA. Probiotic properties of enterococcus isolated from artisanal dairy products. Front Microbiol. 2019;10:1–13.

46. Ben BZ, Abriouel H, Ben ON, Lucas R, Martínez-Canamero M, Gálvez A, et al. Antimicrobial activity, safety aspects, and some technological properties of bacteriocinogenic Enterococcus faecium from artisanal Tunisian fermented meat. Food Control. 2010;21:462–70. https://doi.org/10.1016/j.foodcont.2009.07.007.

47. Ouwehand AC, Kirjavainen PV, Shortt C, Salminen S. Probiotics: mechanisms and established effects. Int Dairy J. 1999;9:43–52. https://doi.org/10.1016/S0958-6946(99)00043-6.

48. Apostolou E, Kirjavainen PV, Saxelin M, Rautelin H, Valtonen V, Salminen SJ, et al. Good adhesion properties of probiotics: a potential risk for bacteremia? FEMS Immunol Med Microbiol. 2001;31:35–9.

49. Börner RA, Kandasamy V, Axelsen AM, Nielsen AT, Bosma EF. Genome editing of lactic acid bacteria: opportunities for food, feed, pharma and biotech. FEMS Microbiol Lett. 2019;36:1–12.

50. Chikhi R, Medvedev P. Informed and automated k-mer size selection for genome assembly. Bioinformatics. 2014;30:31–7.

51. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics. 2011;27:578–9.

52. Boetzer M, Pirovano W. Toward almost closed genomes with GapFiller. Genome Biol. 2012;13:R56.

53. Aziz RK, Bartels D, Best A, DeJongh M, Disz T, Edwards RA, et al. The RAST server: rapid annotations using subsystems technology. BMC Genomics. 2008;9:1–15.

54. Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8.

55. Lowe TM, Eddy SR. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1996;25: 955–64.

56. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, et al. The COG database: an updated vesion includes eukaryotes. BMC Bioinformatics. 2003;4:1–14.

57. Auch AF, von Jan M, Klenk HP, Göker M. Digital DNA-DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. Stand Genomic Sci. 2010;2:117–34.

58. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST ring image generator (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011;12:402.

59. O'Brien KP, Remm M, ELL S. Inparanoid: A comprehensive database of eukaryotic orthologs. Nucleic Acids Res. 2005;33(DATABASE ISS):476–80.

60. Alexeyenko A, Tamas I, Liu G, Sonnhammer ELL. Automatic clustering of orthologs and inparalogs shared by multiple proteomes. Bioinformatics. 2006;22:e9–15. https://doi.org/10.1093/bioinformatics/btl213.

61. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). Nucleic Acids Res. 2014;42:206–14.

El Jeni *et al. BMC Microbiology*   (2020) 20:322

Page 12 of 12

62. Hammami R, Zouhir A, Le Lay C, Ben Hamida J, Fliss I. BACTIBASE second release: a database and tool platform for bacteriocin characterization. BMC Microbiol. 2010;10:22.

63. Kleinheinz KA, Joensen KG, Larsen MV. Applying the ResFinder and VirulenceFinder web-services for easy identification of acquired antibiotic resistance and *E. coli* virulence genes in bacteriophage and prophage nucleotide sequences. Bacteriophage. 2014;4:e27943.

Publisher's Note