



OPEN

## Comparative genomics of a novel clade shed light on the evolution of the genus *Erysipelothrix* and characterise an emerging species

Ana Laura Grazziotin<sup>1,7</sup>✉, Newton M. Vidal<sup>2,7</sup>✉, Patricia Giovana Hoepers<sup>1,7</sup>, Thais F. M. Reis<sup>1</sup>, Dany Mesa<sup>3</sup>, Luiz Felipe Caron<sup>4</sup>, Max Ingberman<sup>5</sup>, Breno C. B. Beirão<sup>4</sup>, João Paulo Zuffo<sup>6</sup> & Belchiorina Beatriz Fonseca<sup>1</sup>

*Erysipelothrix* sp. isolates obtained from a deadly outbreak in farmed turkeys were sequenced and compared to representatives of the genus. Phylogenetic trees—supported by digital DNA:DNA hybridization and Average Nucleotide Identity—revealed a novel monophyletic clade comprising isolates from pigs, turkeys, and fish, including isolates previously described as *E. sp. Strain 2*. Genes coding for the SpaC protein, typically found in *E. sp. Strain 2*, were detected in all isolates of the clade. Therefore, we confirm *E. sp. Strain 2* represents a unique species, that despite its official name “*Erysipelothrix piscisicarius*” (meaning a killer of fish), may be isolated from a broad host range. Core genome analysis showed that the pathogenic species of this genus, *E. rhusiopathiae* and the clade *E. sp. Strain 2*, are enriched in core functionalities related to nutrient uptake and transport, but not necessarily homologous pathways. For instance, whereas the aerobic DctA transporter may uptake C<sub>4</sub>-dicarboxylates in both species, the anaerobic DcuC transporter is exclusive of the *E. sp. Strain 2*. Remarkably, the pan-genome analysis uncovered that genes related to transport and metabolism, recombination and repair, translation and transcription in the fish isolate, within the novel clade, have undergone a genomic reduction through pseudogenization. This reflects distinct selective pressures shaping the genome of species and strains within the genus *Erysipelothrix* while adapting to their respective niches.

Bacterial comparative genomics analyses have brought to light unprecedented aspects of bacterial physiology, diversity and evolution<sup>1</sup>. Uncovering the genomic repertoire of bacterial organisms has also revealed an extensive intraspecific diversity<sup>2</sup>. Therefore, whole-genome sequencing (WGS) has become a powerful tool not only for detecting genetic features and specific adaptations but also for taxonomy, assisting in species delineation<sup>3</sup>. Phylogenomics and whole-sequence alignment-based metrics, such as digital DNA:DNA hybridization (dDDH) and Average Nucleotide Identity (ANI), have been widely used and supported the identification of novel species and reclassification of known taxa<sup>3–6</sup>. In addition, components of the genomic repertoire (core, pan-genome and unique genes) may provide supporting evidence for bacterial characterization and species definition. For

<sup>1</sup>Programa de Pós-Graduação em Ciências Veterinárias, Faculdade de Medicina Veterinária, Universidade Federal de Uberlândia, Rua Ceará, 1084, Bloco 2D, Sala 54, Umarama, Uberlândia, MG CEP: 38405-240, Brasil. <sup>2</sup>Programa de Pós-Graduação em Biodiversidade Animal, Departamento de Evolução e Ecologia, Universidade Federal de Santa Maria, Avenida Roraima, 1000, Prédio 17, Sala 1140-D, Cidade Universitária, Bairro Camobi, Santa Maria, RS CEP: 97105-900, Brasil. <sup>3</sup>Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, Curitiba, PR, Brasil. <sup>4</sup>Departamento de Patologia Básica, Setor de Ciências Biológicas, Universidade Federal do Paraná, Curitiba, PR, Brasil. <sup>5</sup>Imunova Análises Biológicas, Curitiba, PR, Brasil. <sup>6</sup>Centro de Diagnóstico de Microbiologia Animal, Universidade do Estado de Santa Catarina, Florianópolis, SC, Brasil. <sup>7</sup>These authors contributed equally: Ana Laura Grazziotin, Newton M. Vidal and Patricia Giovana Hoepers. ✉email: analauragrazziotin@gmail.com; nwvidal@gmail.com

instance, the presence of species-specific core genes, lineage-specific expansions or gene losses make up a bacterial genomic identity and reflect adaptive strategies.

A number of complete bacterial genomes of the genus *Erysipelothrix* (family Erysipelotrichaceae, phylum Firmicutes) have been made available in the past years. The first genome, *E. rhusiopathiae* strain Fujisawa, was released in 2011<sup>7</sup> and showed that the organism lacks many biosynthetic pathways, which was also observed in *E. rhusiopathiae* SY1027<sup>8</sup>, indicating a reductive genome evolution. Since then many more genomes of the same and other species have been published, providing an opportunity to assess their genetic variations, functional traits and reconstruct ancestral trajectories. An in depth analysis of *E. rhusiopathiae* genomes from a worldwide population showed that the species comprises three distinct clades with weak association to host or geographic origin<sup>9</sup>. Conversely, a WGS study of *E. rhusiopathiae* from a Japanese swine outbreak showed the strains were closely related with few SNPs (single nucleotide polymorphisms) among them and four main lineages were responsible for the acute disease<sup>10</sup>. Most studies, however, have focused on characterizing *Erysipelothrix* species or strains based mainly on serology, spa proteins, and genotype, based on molecular techniques such as pulsed-field gel electrophoresis<sup>11–13</sup>. The phylogenetic reconstruction, phenotypic characterization and pathogenic potential of the genus *Erysipelothrix* were covered in a study of the family Erysipelotrichaceae<sup>14</sup>, which redefined two genera within the family. However, no comprehensive comparative genomic analysis of the genus *Erysipelothrix* has been carried out to date. Moreover, *E. rhusiopathiae* has been vastly studied whereas studies focusing on other *Erysipelothrix* species are very scarce, limiting our understanding of ecological aspects, diversity, genetic traits and evolutionary scale.

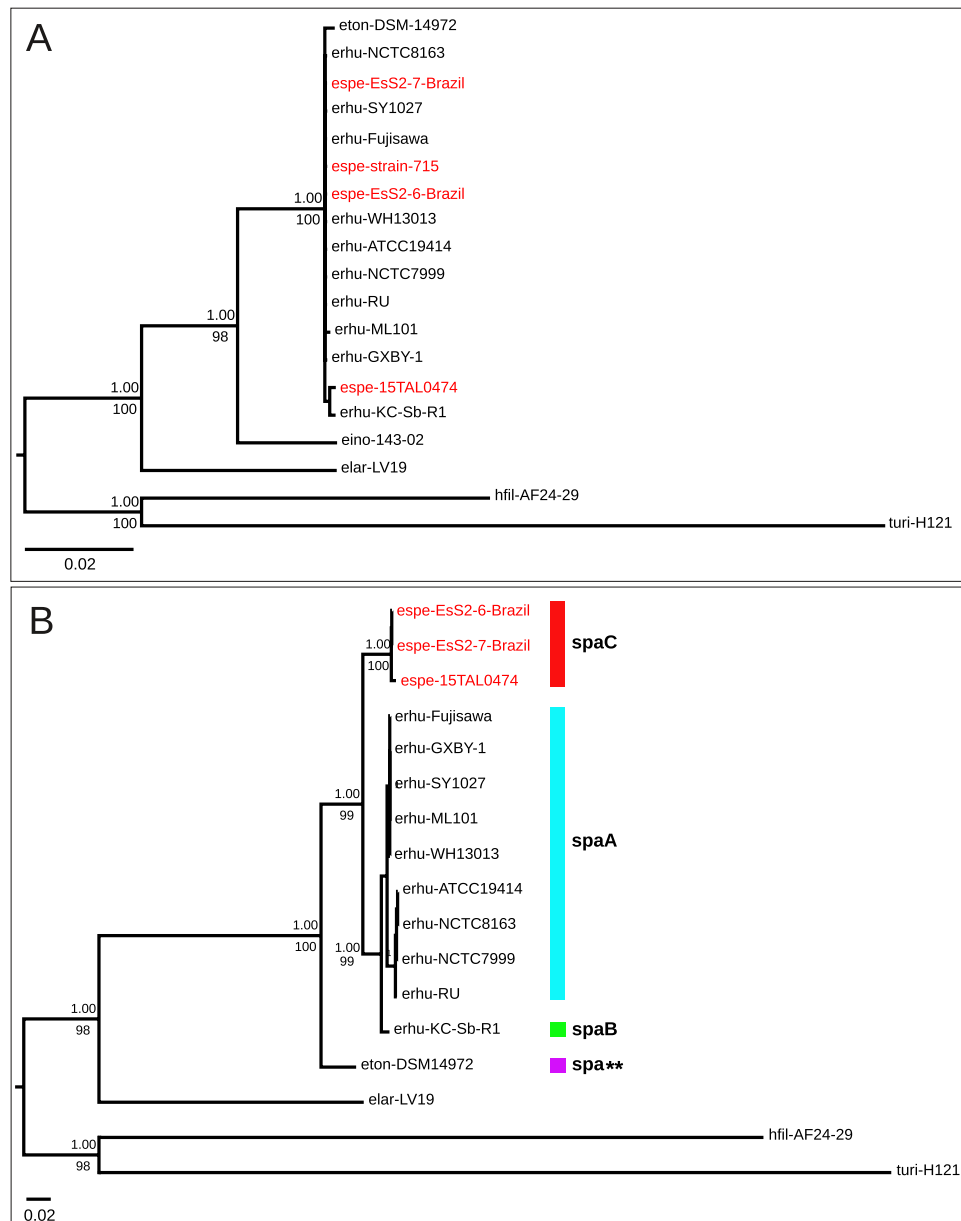
Currently, the *Erysipelothrix* genus comprises five named species, *E. rhusiopathiae*<sup>15</sup>, *E. tonsillarum*<sup>16</sup>, *E. inopinata*<sup>17</sup>, *E. larvae*<sup>18</sup> and *E. piscisarius*<sup>19</sup>. *E. rhusiopathiae* is the best characterized species, responsible for a spectrum of diseases in humans and wild and domestic animals<sup>20</sup>. *E. tonsillarum* has been isolated from healthy swine tonsils<sup>16</sup> and also from dogs with endocarditis<sup>21,22</sup>. *E. inopinata* was isolated from a broth culture<sup>17</sup> and *E. larvae* seems to be a commensal species of a beetle gut<sup>18</sup>. In addition, other potential novel species of the genus have been indicated, such as *E. sp. Strain 1*, *E. sp. Strain 2* and *E. sp. Strain 3*<sup>11,23,24</sup>. The first two, *E. sp. Strain 1* and *E. sp. Strain 2*, were isolated from pigs and previously identified as *E. rhusiopathiae* strain Péc5 56 (serovar 13) and strain 715 (serovar 18), respectively<sup>23</sup> until they were shown to be very dissimilar from either *E. rhusiopathiae* and *E. tonsillarum* type strains as well as from each other based on DDH experiments, suggesting they represented novel species<sup>23</sup>. A third group of distinct isolates, *E. sp. Strain 3*, was also identified<sup>24</sup>. *E. sp. Strain 1* and *Strain 3* have been poorly characterized to date. In contrast, *E. sp. Strain 2* (type strain 715) has been studied and at least three serovars (9, 10 and 18) are associated with this strain, which were found to be pathogenic in mice and pigs<sup>24</sup>; it carries a molecular variant (*spaC*) of the surface protective antigen protein<sup>25</sup> and; it is phylogenetic distinct from *E. rhusiopathiae* and *E. tonsillarum*<sup>9</sup>. Recently, deadly outbreaks in farmed fish and turkeys were associated with *E. sp. Strain 2*<sup>26,27</sup>. Although the ANI analysis between the fish isolate genome (isolate 15TAL0474) and the swine isolate genome (type strain 715) showed they are highly similar (above 99% similarity), slight but consistent differences based on a MLSA tree were observed between the two isolates and thus, authors proposed the fish isolate as a novel species with the name *E. piscisarius*<sup>19</sup>. Given that *E. sp. Strain 2*-related isolates have been shown to cause lesions in pigs and mice<sup>24</sup> and death of farmed fish<sup>19,26</sup> and turkeys<sup>27</sup>, this is likely to be an economically important pathogen in animal production. Nevertheless, limited information is available regarding its biology and, since only recently a representative genome has become available<sup>19</sup>, the understanding of its population diversity and genome evolution is still scarce.

In this study, we sequenced isolates from the turkey outbreak<sup>27</sup> and compared them to the representative species of the genus *Erysipelothrix*. We hypothesized that the emergent pathogenic *Erysipelothrix* isolates from recent outbreaks in turkey and fish belong to a single genospecies (a species that can be differentiated from other species based on genomic methods), which is apart from the other well characterized *Erysipelothrix* species. Therefore, we investigated the presence of Spa proteins and the phylogenetic relationship amongst all current species of the genus using publicly available genomes. Whole genome-based similarity metrics (dDDH and ANI) were also performed to confirm the taxonomic relationship. After, the genomic repertoires within and among species were assessed, focusing on the novel emergent species, in order to identify shared and specific genetic features related to the species diversity, genome evolution and specific adaptations within the genus.

## Results and discussion

**The 16S rRNA phylogenetic tree is not suitable for delineating *Erysipelothrix* species.** Full length 16S rRNA sequences were retrieved from available genomes (Supplementary Table S1). Sequences from *E. inopinata* and *E. sp. Strain 2* (type strain 715) were retrieved from NCBI Nucleotide since no genome sequences were publicly available. The 16S rRNA gene was used since it has been a long-standing primary choice for bacterial diagnosis and identification. Based on the 16S rRNA gene tree, *Erysipelothrix* species formed three distinct clades (Fig. 1A). *E. larvae* was shown as the most ancestral species of the genus *Erysipelothrix*, followed by *E. inopinata*, each one was placed in a highly supported single branch on the tree. However, the remaining isolates belonging to *E. tonsillarum*, *E. rhusiopathiae* and *E. sp. Strain 2* (isolates 15TAL0474, EsS2-6-Brazil, EsS2-7-Brazil and type strain 715) were clustered all together, supported by pairwise sequence similarities above 99% (Supplementary Table S2), which is higher than the standard threshold value (97%) used as species boundaries<sup>28</sup>. Therefore, 16S rRNA sequences are not recommended to distinguish among *Erysipelothrix* species.

Thus, we used the housekeeping gene *rpoB* (beta subunit of RNA polymerase) to check the phylogenetic relatedness (Fig. 1B). The *rpoB* gene has been suggested as an alternative for the 16S rRNA gene due to its universality, ancient origin and sufficient number of sequence variation to discriminate bacterial species<sup>29</sup> and, therefore, it has been applied for bacterial identification of clinical isolates<sup>30,31</sup>. The *rpoB* gene tree showed a clear distinction of *Erysipelothrix* species (Fig. 1B). Remarkably, the three *E. sp. Strain 2*-related isolates (15TAL0474, EsS2-6-Brazil



**Figure 1.** Phylogenetic reconstruction of *Erysipelothrix* genus using single-genes. Legend: **(A)** Bayesian phylogenetic tree based on 16S rRNA gene under the model GTR+I+G. **(B)** Bayesian phylogenetic tree based on *rpoB* nucleotide sequence under the model GTR+I+G. Posterior probability values of support obtained in Bayesian Analysis (BA) are shown above nodes. Rapid bootstrap values obtained in Maximum Likelihood (ML) analysis are shown below nodes. Species were indicated as follows: *E. sp.* Strain 2 isolates (*espe-EsS2-6-Brazil*, *espe-EsS2-7-Brazil*, *espe-15TAL0474*, *espe-strain-715*); *Erysipelothrix rhusiopathiae* isolates (*erhu-ATCC19414*, *erhu-KC-Sb-R1*, *erhu-NCTC8163*, *erhu-NCTC7999*, *erhu-SY1027*, *erhu-GXBY-1*, *erhu-WH13013*, *erhu-Fujisawa*, *erhu-ML101*, *erhu-RU*); *Erysipelothrix inopinata* strain 143-02 (*eino-143-02*); *E. tonsillarum* DSM 14972 (*eton-DSM14972*); *E. larvae* LV19 (*elar-LV19*); *Holdemania filiformis* AF24-29 (*hfil-AF24-29*); *Turicibacter sp.* H121 (*turi-H121*). Species *hfil-AF24-29* and *turi-H121* were used as outgroups. *E. sp.* Strain 2 isolates are shown in red.

and *EsS2-7-Brazil*) formed a highly supported monophyletic group, indicating that these isolates might represent a new taxon. Accordingly, the three isolates showed 99.61–99.98% identity within the group (Supplementary Table S2), which is above the proposed threshold for a new bacterial species (97.7%)<sup>32,33</sup> and subspecies (98.2%) delineation<sup>29</sup>, indicating that these isolates might belong to the same species. *E. inopinata* and *E. sp.* Strain 2 (type strain 715) were not included in this and further analysis since no *rpoB* gene sequence nor their genome sequences were publicly available during the time this work was performed and manuscript was written.

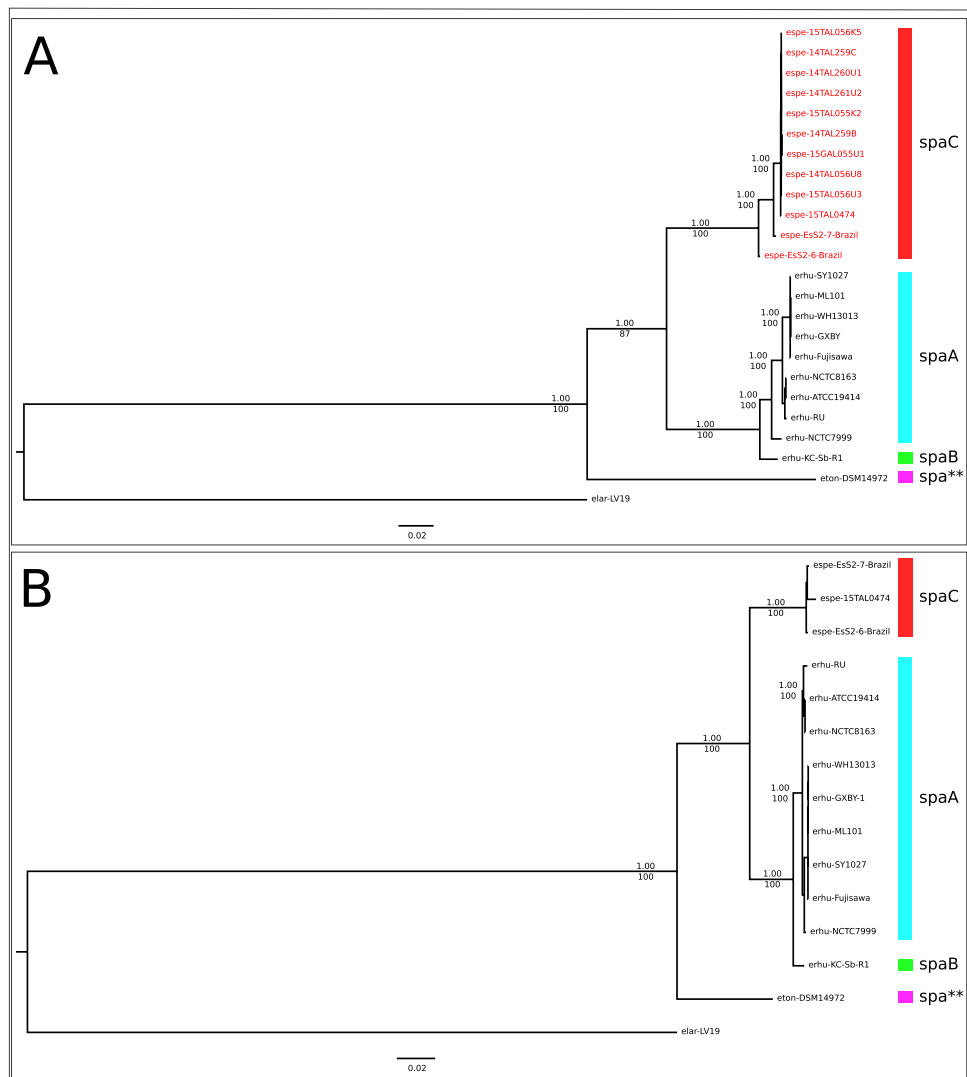
**The SpaC protein sequence is present in all *E. sp. Strain 2*-related isolates and a novel Spa variant is found in *E. tonsillarum*.** We investigated the presence of the surface protective antigen protein (Spa) sequence since the presence of the SpaC variant has been suggested to distinguish *E. sp. Strain 2* from other *Erysipelothrix* spp.<sup>25</sup>. The typical SpaC was found in all *E. sp. Strain 2*-related isolates whereas SpaA and SpaB were found in *E. rhusiopathiae* (Fig. 1B), as expected<sup>25,34</sup>. No Spa sequence was detected in *E. larvae* but surprisingly, a Spa protein sequence was found in *E. tonsillarum* (Supplementary Fig. S1A). The novel Spa protein sequence is distantly related to the other Spa types showing the lowest identities (43.8% Spa A, 41.1% SpaB and 37.9% SpaC) amongst them (Supplementary Fig. S1B). Previous studies of *spa* gene detection based on PCR have not found a *spa* sequence in *E. tonsillarum*<sup>12,25,35,36</sup> and only a single work reported the detection of *spaA* and *spaB* in *E. tonsillarum* by PCR<sup>26</sup>, but the fragments were not sequenced. Experimental or genomic studies assessing the prevalence of Spa protein in other *E. tonsillarum* isolates may clarify the extension of its presence in the species.

**Multilocus sequence analysis (MLSA) and phylogenomics reconstructions show a novel species within the *Erysipelothrix* genus.** Next, we used multilocus sequence approaches to verify the species relatedness within the *Erysipelothrix* genus. In recent years, MLSA and phylogenomics have been widely used to discriminate bacterial species and strains<sup>3,37,38</sup> due to their higher resolution compared to single-locus approaches. The MLSA tree (Fig. 2A) is based on seven slowly evolving gene sequences (*galK*, *gpsA*, *ldhA*, *prsA*, *pta*, *purA* and *recA*) previously proposed for multilocus sequence typing of *E. rhusiopathiae*<sup>13</sup>. In addition to our sequenced genomes and publicly available genomes from various hosts, the MLSA phylogeny included gene sequences from nine other fish isolates (*E. sp. Strain 2*-related isolates), whose genome sequences, although reported, were not made publicly available<sup>19</sup>. The phylogenomic tree (Fig. 2B) is based on the alignment of 506 single-copy orthologous proteins for the *Erysipelothrix* genus. The MLSA and the phylogenomic trees are topologically similar, showing four well-supported clades. *E. larvae* and *E. tonsillarum* form the deepest branches of the trees whereas the two most derived clades split *E. rhusiopathiae* from the newly sequenced *E. sp. Strain 2*-related isolates. The latter group also included all 10 isolates collected from fish during a disease outbreak in the United States<sup>19,26</sup> by MLSA. The consistent monophyletic nature of *E. sp. Strain 2*-related isolates based on three distinct phylogenetic approaches is the main criterion for defining a novel taxon<sup>39</sup>.

**Whole-genome alignment analyses (dDDH and ANI) confirm the phylogenomic relatedness.** To confirm the species relatedness inferred from the phylogenetic trees and ensure an accurate assignment at the species level, the pairwise nucleotide-level comparisons (dDDH and ANI) were calculated for 15 genomes of genus *Erysipelothrix* and closely related genera (Fig. 3A,B). The established same-species delineation thresholds are 70% for dDDH<sup>40,41</sup> and 95% for ANI<sup>42</sup> values. The dDDH and ANI values between all pairs of *E. sp. Strain 2*-related genomes and *E. rhusiopathiae* genomes were below both thresholds (dDDH 31.5–33% and ANI 86.76–87.83%) (Supplementary Table S3), confirming that they represent distinct species at the genome level. Of note, amongst *E. sp. Strain 2*-related genomes all metrics were above the threshold (dDDH 87.1–92.9% and ANI 98.51–99.14%) (Supplementary Table S3), providing further evidence that these isolates comprise a genomospecies, as supported by the monophyletic clade in *rpoB* tree, MLSA and phylogenomics.

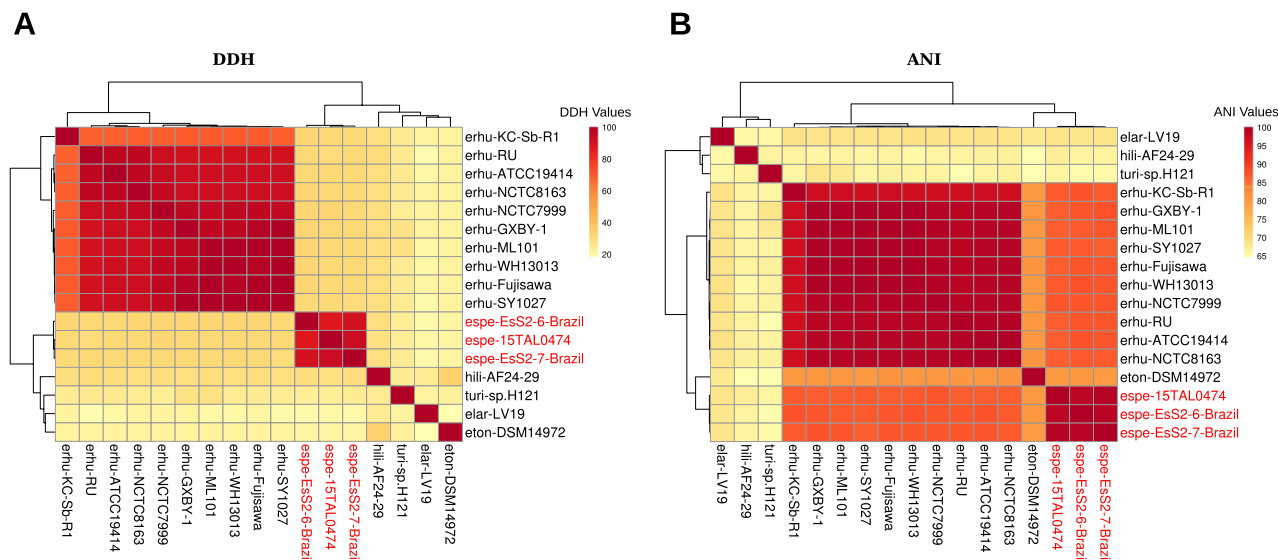
The two combined approaches—phylogenomics and whole-genome nucleotide metrics—demonstrated that isolates related to *E. sp. Strain 2* belong to the same species. The type strain 715 was previously isolated from a swine spleen and distinguished from *E. rhusiopathiae* based on a wet lab DDH approach<sup>23</sup>. At that time, authors suggested that isolate 715 could represent a novel species but to date, no study has comprehensively characterized such isolate. *E. sp. 15TAL0474*, isolated from fish, has been recently sequenced and compared to the pig isolate (type strain 715) by dDDH (90.8%) and ANI (99.01%)<sup>19</sup>, which supported that these strains would belong to the same species. However, due to a slight but consistent variation in MLSA pattern between the pig and the fish isolates, authors considered the fish isolate a novel species, which was named *E. piscisarius*<sup>19</sup>. Intraspecific variation is commonly observed within many species<sup>9,43</sup> and the genotypic diversity within *E. rhusiopathiae* has been already demonstrated<sup>9</sup>. For instance, the variation found between the pig and the fish genomes<sup>19</sup> is no greater than that found within *E. rhusiopathiae*, i.e., between the Clade 1 (more distinct one) and the other clades of *E. rhusiopathiae*<sup>9</sup> (Fig. 3AB; Supplementary Figure S2). The International Code of Nomenclature of Prokaryotes<sup>44</sup> recommends that when choosing a species name (Recommendation 12c), isolates deemed conspecific should retain the species epithet provided on List of Prokaryotic names with Standing in Nomenclature. *E. sp. Strain 2* has been isolated from a broad diversity of hosts, firstly from a pig (type strain 715)<sup>23</sup>, and then from fish (isolate 15TAL0474)<sup>19</sup> and birds (isolates EsS2-6-Brazil and EsS2-7-Brazil)<sup>27</sup>. Nevertheless, though the new species represents a pathogen of multiple distinct hosts (similarly to what is observed for *E. rhusiopathiae*), and the name *E. piscisarius* (meaning a killer of fish) does not represent the bacterium's full host spectrum, as the first taxonomically characterized and validated name for the species<sup>19</sup>, "*Erysipelothrix piscisarius*" should be considered the official species name for *E. sp. Strain 2*. Given that the new species represents a pathogen of multiple distinct hosts (similarly to what is observed for *E. rhusiopathiae*) and that the name *E. piscisarius* (meaning a killer of fish) does not represent the bacterium's full host spectrum, a more generic, unbiased name would be suitable. We suggest "*Erysipelothrix takahashiae*" after Toshio Takahashi who first discovered isolates of this clade and suggested it could represent a novel species<sup>23</sup>.

**The core genome of pathogenic species is overrepresented by metabolic genes.** We found 917 protein families in the core genome of *E. rhusiopathiae* and *E. sp. Strain 2* and a total of 2006 families comprising the pan-genome of both species. The core genome, as expected, is enriched ( $p < 0.05$ ) in protein families related to the basic cellular machinery, such as "Translation, ribosomal structure and biogenesis" (Cluster of Ortholo-



**Figure 2.** Phylogenetic reconstruction of *Erysipelothrix* genus using Multilocus Sequence analysis (MLSA) and phylogenomic analysis. Legend: (A) MLSA Bayesian phylogenetic tree based on the partitioned supermatrix of seven genes (galK, gpsA, ldhA, prsA, pta, purA and recA). (B) Phylogenomic Bayesian phylogenetic tree based on the partitioned supermatrix of 506 single-copy core-genome orthologous proteins. Posterior probability values of support obtained in Bayesian Analysis (BA) are shown above nodes. Rapid bootstrap values obtained in Maximum Likelihood (ML) analysis are shown below nodes. Species were indicated as described Supplementary Table S1.

gous Groups—COG category J), “Metabolism and transport of amino acids” (COG category E), “Metabolism and transport of lipids” (COG category I), and “Metabolism and transport of inorganic ions” (COG category P) (Supplementary Fig. S3; Supplementary Table S4). For some isolates functional enrichment was not statistically significant, but still their core genomes clearly showed higher proportion of genes in such categories compared to the accessory genome (Supplementary Fig. S4; Supplementary Table S5), indicating that pathways related to the metabolism of amino acids, lipids and inorganic ions play an important role for the group as a whole. Accordingly, these COG categories have been found to show a considerable number of regulated genes in *E. rhusiopathiae* HX130709 grown in rich medium<sup>45</sup>. After checking the list of regulated genes<sup>45</sup>, we found that most of the regulated genes present in COG E (68.1%), COG P (63.3%), and COG I (80%) in *E. rhusiopathiae* HX130709 belong to the core genes of *E. rhusiopathiae* and *E. sp.* Strain 2. Considering that *E. rhusiopathiae* was grown in a nutrient-rich and stress-free condition<sup>45</sup>, it is expected that most recruited genes are related to cell maintenance. Genes belonging to the core-genome enriched categories maintain the basic cellular machinery, the central metabolism, and mediate transport processes into and out of the cell, which means that shared genes in these categories are needed for cell growth and survival.

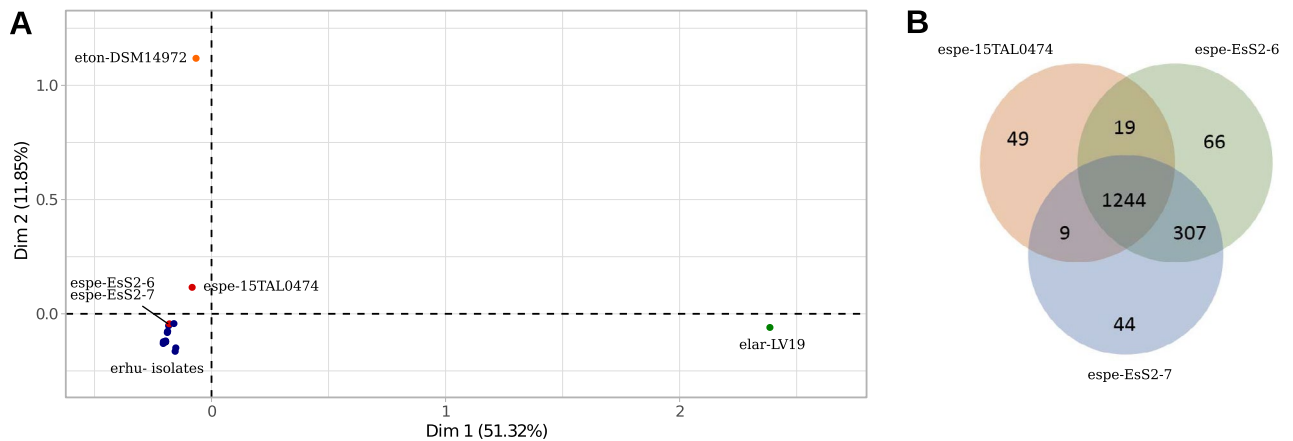


**Figure 3.** Heatmap of whole-genome sequence pairwise comparisons between species of the genus *Erysipelothrix* and two outgroups. Legend: (A) Heatmap of the digital DNA:DNA hybridizations (dDDH). (B) Heatmap of the Average Nucleotide Identity (ANI). The percentages are provided in Supplementary Table S3. Species mentioned in the figures are described Supplementary Table S1.

**Distinct core strategies of nutrient uptake and energy metabolism between *E. rhusiopathiae* and *E. sp. Strain 2*.** When analysing the two species separately, 1,109 and 1,244 protein families comprised the core genome of *E. rhusiopathiae* and *E. sp. Strain 2*, respectively. The core genome represented on average 70.69% of the total coding sequences in *E. rhusiopathiae* and 82.40% for *E. sp. Strain 2* isolates. Differences were found between the two core genomes and we highlight two protein families related to nutrient uptake and energetic metabolism.

$C_4$ -dicarboxylate transporters are secondary carriers for the uptake, exchange or efflux of  $C_4$ -dicarboxylates (fumarate, succinate, aspartate and malate) from the Krebs cycle, which are relevant to the bacterial energetic metabolism when sugars are not available<sup>46</sup>. The DctA family of  $C_4$ -dicarboxylate carriers (COG1301) was found in all studied *Erysipelothrix* species (*E. rhusiopathiae*, *E. tonsillarum*, *E. larvae* and *E. sp. Strain 2*), making up the core genetic repertoire of the genus (Supplementary Table S6). In contrast, the DcuC protein family  $C_4$ -dicarboxylate transporter (COG3069) is a core protein in *E. sp. Strain 2*, which is absent in all *E. rhusiopathiae* isolates (Supplementary Table S6). Similar to *E. sp. Strain 2*, the bacterial pathogen *Campylobacter jejuni* carries both  $C_4$ -dicarboxylate transporter genes (*dctA* and *dcuC*)<sup>47</sup>. DctA was the only  $C_4$ -dicarboxylate carrier required by *C. jejuni* to grow based on dicarboxylate-carbon sources at high oxygen levels<sup>47</sup> whereas under anaerobic conditions, DcuC was upregulated in the pathogen<sup>48</sup>. The *dcuC* gene might be induced in *E. sp. Strain 2*, similarly to other bacteria<sup>46,48</sup>, allowing them to transport aspartate and fumarate under oxygen-limited conditions<sup>49,50</sup>. Although *E. rhusiopathiae* isolates do not share an orthologous *dcuC* gene with *E. sp. Strain 2* and apparently, they would not be able to perform  $C_4$ -dicarboxylate transport under anaerobic condition by this route, we cannot disregard that the function might be played by a non-orthologous gene. Gene knockout mutant and transcriptome experiments of *Erysipelothrix* isolates based on dicarboxylate-carbon sources under aerobic and anaerobic conditions would help to understand the preferential metabolic strategies employed by these organisms and whether *E. rhusiopathiae* strains carry any alternative anaerobic route for dicarboxylate uptake.

The phosphoenolpyruvate (PEP)-dependent sugar phosphotransferase system (PTS) is the major carbohydrate (glucose, glucitol, mannose and ascorbate) transport system in bacteria. The PTS superfamilies comprise two cytoplasmic phosphotransferases (HPr and enzyme I—EI) and a sugar-specific permease complex (enzyme II—EII). Genes coding for HPr and EI were found in the core genome of *E. rhusiopathiae*, as well as in the other species (*E. larvae*, *E. tonsillarum* and *E. sp. Strain 2*) since their products are used to phosphorylate enzymes of all PTS superfamilies. Genes of the anaerobic L-ascorbate degradation pathway (from L-ascorbate to D-xylulose-5P (Ko00053)) belong to the operon *ulaABCDE*<sup>51</sup> and are regulated by operon *ulaGR*<sup>52</sup>. The anaerobic L-ascorbate degradation pathway is complete in all *E. rhusiopathiae* isolates, but two (Supplementary Fig. S5). Gene *ulaD* was missing in *E. rhusiopathiae* strain RU whereas *ulaD* and *ulaF* were missing in strain SY1027. These missing genes would be part of the core genome, however, they were considered pseudogenes due to multiple frameshift mutations. Many bacteria have been reported to ferment L-ascorbate under anaerobic conditions<sup>51,53</sup> and this route may provide energy supply for survival when other sources are limited in natural environments for *E. rhusiopathiae*. In contrast, genes of the anaerobic pathway for L-ascorbate degradation were not found in *E. sp. Strain 2*. Similarly, typical L-ascorbate-related genes have not been found in *Ralstonia eutropha* genome, although the species is capable of using L-ascorbate as a sole source of carbon, which is performed via a novel catabolic pathway<sup>54</sup>. Genes of this novel pathway were not identified in *E. sp. Strain 2* after sequence searches. Further experimental investigations may help elucidate whether the species might use another distinct strategy for L-ascorbate metabolism or might not uptake this nutrient at all.



**Figure 4.** Pan-genome analysis of the genus *Erysipelothrix* and of *E. sp.* Strain 2 isolates. Legend: (A) Multiple correspondence analysis (MCA) of the genus *Erysipelothrix* pan-genome. Species mentioned in the figures are described in Supplementary Table S1. Blue dots belong to *E. rhusiopathiae* isolates. Red dots belong to *E. sp.* Strain 2 isolates. (B) Venn diagram of *E. sp.* Strain 2 pan-genome. Number of orthologous groups (OGs) that belong to the core genome (shared among the three isolates—union of all circles), number of accessory OGs (shared between two isolates—intersection of two circles), and number of singletons (exclusive/unique genes—remaining area of each circle) are described.

#### The pan-genome of *Erysipelothrix* genus shows a reduced accessory genome in the fish isolate 15TAL0474.

We examined the relationship among *Erysipelothrix* species based on a multiple correspondence analysis (MCA) of the pan-genome (Fig. 4A). *E. larvae* and *E. tonsillarum* were distantly related from the other most derived species, *E. rhusiopathiae* and *E. sp.* Strain 2, as expected (Fig. 4A). The most ancestral species are not only distantly related from the others based on the core protein sequence and whole nucleotide divergences (Fig. 2A,B, Fig. 3A,B), but also on gene content diversity (Fig. 4A). Surprisingly, *E. sp.* Strain 2 isolate 15TAL0474 was shown apart from the other two Strain 2 isolates (EsS2-6-Brazil and EsS2-7-Brazil), which fell within the *E. rhusiopathiae* group (Fig. 4A). Isolate 15TAL0474 shows the smallest proteome (1,352 protein coding genes) among all studied genomes (Supplementary Table S1). Thus, the core genome represents almost the totality (93.4%) of its proteome whereas for the other two related isolates (EsS2-6-Brazil and EsS2-7-Brazil), it comprises about 75% of their proteomes. This is likely a result of a reduced accessory genome (28 OGs) in isolate 15TAL0474 compared to the other two genomes (316 and 326 OGs) (Fig. 4B) and apparently, the missing set might explain the distance seen among these isolates in the MCA. Particularly, 15TAL0474 has 286 pseudogenes whereas EsS2-6-Brazil and EsS2-7-Brazil have only 21 and 16, respectively. In addition, among the 307 OGs shared between EsS2-6-Brazil and EsS2-7-Brazil, 293 OGs are also shared with *E. rhusiopathiae* group and most of them (~80% or 232/293 OGs) are consistently present in the accessory genome of *E. rhusiopathiae* (9 out of 10 strains), indicating that the accessory set was probably present in the core genome of the ancestral organism but has been under distinct pressures among strains. In addition to the missing accessory genes in 15TAL0474, the number of shared accessory genes between EsS2-strains and *E. rhusiopathiae* might explain their proximity in MCA.

We hypothesized that the extensive accessory reduction in 15TAL0474 could be related to an ongoing pseudogenization process. To check our hypothesis, we performed a reciprocal best hit (RBH) analysis of 15TAL0474 pseudogenes against the proteomes of all other genomes. A total of 200 (70%) pseudogenes had a RBH within the *E. sp.* Strain 2 group (with EsS2-6-Brazil and/or EsS2-7-Brazil) (Supplementary Table S7). Among them, 184 pseudogenes had hits with both EsS2-6-Brazil and EsS2-7-Brazil, and therefore, the core genome of *E. sp.* Strain 2 would be considerably raised from 1244 to 1428 OGs if the set of 15TAL0474 was functional. Genes related to transport and metabolism (carbohydrate [COG category G] and amino acid [E]) and information storage and processing (replication, recombination and repair [L]; translation, ribosomal structure and biogenesis [J]; transcription [K]) were the most represented (44.02% or 81/184) among decayed genes in 15TAL0474. The remaining pseudogenes had RBH with (1) only one EsS2-strain (16 pseudogenes), comprising the accessory genome of *E. sp.* Strain 2; (2) with itself (70 pseudogenes), comprising the exclusive set of 15TAL0474; or (3) with a gene outside the *E. sp.* Strain 2 group (16 pseudogenes). Therefore, the pan-genome analysis reveals the impact of gene reduction in 15TAL0474 as well suggesting the diverse genetic evolution for the pan- and core genomes among *Erysipelothrix* strains.

Genome downsizing has been shown in many bacterial species, which have undergone a transition from a free-living to a parasitic lifestyle. For instance, *Mycobacterium lepraemurium*<sup>55</sup>, *M. uberis*<sup>56</sup>, *Staphylococcus saccharolyticus*, *Shigella* spp.<sup>57</sup>, and *Rickettsia* spp.<sup>58</sup> show reduced genomes that have been shrinking through gene decay and tend to minimize their gene content to the strictly required set as seen in *Mycoplasma genitalium*<sup>59</sup>. While the bacteria is adapting to the host niche, many genes become no longer major contributors for fitness in such environment and may be subject to gene decay. Since the host may provide required nutrients or machinery, genes of the core metabolism and DNA repair<sup>57,58</sup> are commonly lost by the pathogen, which might explain their fastidious growth outside the host and mutation rate leading to pseudogenization. It is likely that

*E. sp.* 15TAL0474 is under an ongoing reductive genome process to essentiality during its adaptation to a novel aquatic host whereas the orthologous genes remain needed in other isolates within the species, which colonize a distinct host. *E. rhusiopathiae* has also been described to have a wide-host spectrum<sup>9</sup> as *E. sp.* Strain 2 and evidence of host-adapted strains are still scarce. Only recently, genetic determinants of *E. rhusiopathiae* strains were shown to be associated with pigs and wild boars, indicating host-associated strains<sup>60</sup>. We acknowledge that the small number of *E. sp.* Strain 2 isolates, including two epidemiologically related isolates, may not reflect the full genetic background of the species population and its diversity. Therefore, sequencing of further *E. sp.* Strain 2 isolates from distinct hosts might eventually help clarify the relationship between host and variants within this emerging species.

Here we reported a comprehensive comparative genomic analysis of the genus *Erysipelothrix*. Previous studies focused on *E. rhusiopathiae* whereas other species in the genus have been neglected. Thus, based on phylogenomics, and supported by dDDH and ANI values, we confirmed that the genus comprises a novel species, formerly known as *E. sp.* Strain 2, and recently named "*Erysipelothrix pisciscarius*". We also showed that core functionalities shared by *E. rhusiopathiae* and *E. sp.* Strain 2 may be performed by homologous or analogous pathways, as illustrated by the C<sub>4</sub>-dicarboxylate transport. This reveals the complex biology of these organisms, which may employ distinct or alternative strategies to reach a similar purpose. Our work also uncovered distinct lineage-specific adaptations that have occurred within *E. sp.* Strain 2, resulting in a massive gene decay in the fish isolate. Considering the wide range of ecotypes in which *Erysipelothrix* species have been isolated, it is possible that a variety of survival strategies co-evolved with the respective bacterial hosts. However, further studies are still needed to find out which selective forces might be acting over members of this novel clade isolated from distinct environments and also shaping their genomes. Finally, the findings reported here provide new insights into *Erysipelothrix* genome evolution and diversification that contribute to understanding the unique characteristics within the genus and may aid with new control strategies or prospective vaccine targets.

## Methods

**Whole genome sequencing.** Two isolates of *Erysipelothrix sp.* Strain 2 from a farm turkey outbreak were randomly selected for whole genome sequencing and comparative genomics. Selected samples had been previously isolated from the lung and liver of deceased farm turkeys during the outbreak and confirmed as *Erysipelothrix sp.* Strain 2 by PCR, as described elsewhere<sup>27</sup>. Genomic DNA was extracted using Wizard Genomic DNA Purification kit (Promega, Wisconsin, USA) and quantified using Qubit HS dsDNA kit (Life Technologies, California, USA). DNA sequencing libraries were prepared using Illumina Nextera XT kit (Illumina, California, USA). Libraries were quantified and their quality was verified with Bioanalyzer (Agilent, California, USA). Whole genome sequencing was performed in a Illumina MiSeq platform (Illumina), using paired-end sequencing and 250 bp read length, which was conducted at the WEWSeq Biotecnologia (Curitiba, Brazil). Raw read quality was checked using FastQC<sup>61</sup>. Genomes were de novo assembled using SPAdes v. 3.12<sup>62</sup> and annotated using NCBI Prokaryotic Genome Annotation Pipeline<sup>63</sup>.

**Comparative genomics.** Comparative genome analyses were performed for a total of 15 *Erysipelothrix* genomes plus two outgroups belonging to the Erysipelotrichaceae family: *Holdemania filiformis* AF24-29 and *Turicibacter sp.* H121. In addition to our two *E. sp.* Strain 2 isolates (EsS2-6-Brazil and EsS2-7-Brazil), publicly available RefSeq genomes were retrieved from FTP-NCBI on December 14, 2018. At least one representative of *E. rhusiopathiae* clades (Clade 1, Clade 2 and Intermediate), according to Forde et al.<sup>9</sup>, were represented among selected genomes (Supplementary Material). Species and accession numbers for public available genomes used in this work are (Supplementary Table S1): *Erysipelothrix sp.* 15TAL0474 (NZ\_CP034234.1), *E. rhusiopathiae* strains Fujisawa (NC\_015601.1), NCTC8163 (NZ\_LR134439.1), GXBY-1 (NZ\_CP014861.1), ML101 (NZ\_CP029804.1), WH13013 (NZ\_CP017116.1), KC-Sb-R1 (NZ\_CP033601.1), SY1027 (NC\_021354.1), ATCC 19414 (NZ\_ACLK00000000.2), NCTC7999 (NZ\_UFYF00000000.1), and RU (NZ\_RJTK00000000.1), *E. tonsillarum* DSM 14972 (NZ\_AREO00000000.1), *E. larvae* LV19 (NZ\_CP013213.1), *Holdemania filiformis* AF24-29 (NZ\_QRUP01000001.1) and *Turicibacter sp.* H121 (NZ\_CP013476.1). Genome accessions for *Erysipelothrix sp.* EsS2-6-Brazil and EsS2-7-Brazil, sequenced in this study, are: SBAR00000000.1 and SCFT00000000.1.

**Orthologous inference.** FastOrtho software<sup>64</sup> (<https://github.com/olsonanl/FastOrtho>) was used to define the orthologous groups. FastOrtho is a reimplement of the OrthoMCL program<sup>65</sup> that does not require the use of databases or Perl. Briefly, it uses BLASTP (v. 2.7.1+)<sup>66</sup> to perform all-against-all homology search and also the MCL Markov Clustering algorithm<sup>67</sup> to construct orthologous groups. BLASTP parameters were set as: -num\_threads 7 -outfmt 7 -evalue 1e-05 -max\_target\_seqs 1000 and the remaining parameters were used as default. The MCL algorithm was used with default parameters.

**Functional annotation.** Clusters of Orthologous Groups (COGs) were assigned to protein sequences using the Batch CD-Search online tool<sup>68,69</sup> against the COG v1.0-4873 PSSMs database. COG annotations and functional categories (A-Z letter code) were attributed based on the most updated COG version<sup>70</sup>. Functional category enrichment analyses were calculated using the Fisher's exact test ( $P < 0.05$ ). Pfam Domain annotations were obtained running hmmscan (v. 3.2.1) locally against the Pfam database release 32.0 (17,929 protein families)<sup>71</sup> considering E-value  $\leq 0.01$ . KEGG annotations were obtained from BlastKOALA<sup>72</sup> and KofamKOALA<sup>73</sup>.

**Single-gene phylogenetic analysis.** Single-gene phylogenetic trees were constructed using 16S rRNA gene and rpoB nucleotide sequences from 15 *Erysipelothrix* species with genomes available and from two outgroup species, *Holdemania filiformis* AF24-29 and *Turicibacter sp.* H121. For the 16S rRNA gene tree, sequences



from *Erysipelothrix* sp. strain 715 and *E. inopinata* (whose genome sequences are not available to date) were included in the analysis. Sequences of these species were retrieved using an online BLASTN search<sup>74</sup> with default parameters, using *E. rhusiopathiae* strain Fujisawa sequence as query. Sequences for each dataset were aligned with MUSCLE (v. 3.8.31)<sup>75</sup> using default parameters, and poorly aligned columns were removed using trimAl (v. 1.4.rev22)<sup>76</sup> with option -automated1. Best-fit nucleotide substitution models were selected using Model-Test-NG<sup>77</sup> according to the corrected Akaike Information Criterion (AICc) implemented on Cipres Science Gateway<sup>78</sup>. Phylogenetic analyses were performed using Maximum Likelihood (ML) and Bayesian Analysis (BA) on Cipres Science Gateway<sup>78</sup>. ML search for the best-scoring ML tree was performed on RAxML (v. 8.2.12)<sup>79</sup> under rapid bootstrap and stop bootstrap automatically (autoMRE) with majority rule criteria. BA analysis was performed on MrBayes (v. 3.2.7a)<sup>80</sup>, running two Markov Chain Monte Carlo (MCMC) runs of four chains each for 2,000,000 generations, sampling trees every 1000 generations with a burn-in of 25%. Phylogenetic trees were visualized and edited in FigTree (v. 1.4.2)<sup>81</sup>.

**Multilocus sequence analysis (MLSA).** MLSA phylogenetic tree was constructed based on the concatenated alignments of seven housekeeping genes (galK, gpsA, ldhA, prsA, pta, purA and recA) that have been previously proposed for multilocus sequence typing of *E. rhusiopathiae*<sup>13</sup>. Orthologous sequences for each individual genome were retrieved as previously described for 16S rRNA and rpoB. In addition, sequences from nine *Erysipelothrix* sp. Strain 2 isolated from fish (isolates 14TAL261U2, 14TAL260U1, 14TAL056U8, 14TAL259B, 15TAL055K2, 15TAL056U3, 15GAL055U1, 15TAL056K5, 14TAL259C) described elsewhere<sup>26</sup> were included in this dataset. Sequences were aligned with MUSCLE (v. 3.8.31)<sup>75</sup> and trimmed with trimAl (v. 1.4.rev22)<sup>76</sup> as described above. Sequences were concatenated using FASconCAT-G (v. 1.04)<sup>82</sup> and the best-fit partitioning schemes and nucleotide models of evolution were selected using PartitionFinder (v. 2.1.1)<sup>83</sup> implemented on Cipres Science Gateway<sup>78</sup>. PartitionFinder settings used were: datatype=DNA, phylogeny program=raxml, branchlengths=linked, models=all, model\_selection=aicc, search=all. Phylogenetic analyses were carried out using both ML and BA, under the respective partition schemes and models of evolution defined by PartitionFinder, with remaining parameters as described previously. Phylogenetic trees were visualized and edited in FigTree (v. 1.4.2)<sup>81</sup>.

**Phylogenomic analysis.** Protein sequences of 618 single-copy core-genome orthologous groups from the 15 *Erysipelothrix* complete genomes were retrieved from the FastOrtho output file. We identified 112 genes potentially involved in horizontal gene transfer (HGT) events and removed their respective orthologous group (OG) to avoid their impact in the phylogenomic analysis (see details in the Supplementary Material). We ended up with a 506 OGs single-copy core genome dataset that was used to perform the phylogenomic analysis. For each individual orthologous group, sequences were aligned with MUSCLE (v. 3.8.31)<sup>75</sup> and trimmed with trimAl (v. 1.4.rev22)<sup>76</sup> as described above. The best-fit partitioning schemes and amino acid models of evolution were selected using PartitionFinder (v. 2.1.1)<sup>83</sup> implemented on Cipres Science Gateway<sup>78</sup>, with the following settings: datatype=protein, phylogeny program=raxml, branchlengths=linked, models=all, model\_selection=aicc, rcluster-max=100, rcluster-percent=10.0, search=rcluster<sup>84</sup>. Phylogenetic analyses were carried out using both ML and BA, under the respective partition schemes and models of evolution defined by PartitionFinder, with remaining parameters as described previously. Phylogenetic trees were visualized and edited in FigTree (v. 1.4.2)<sup>81</sup>.

**Analysis of pseudogenes in *Erysipelothrix* sp. 15TAL0474.** In order to understand the evolution of pseudogenes in *Erysipelothrix* sp. 15TAL0474, putative amino acid sequences of the 286 pseudogenes (as annotated in the RefSeq version of the genome) were used as queries to run BLASTP (v. 2.7.1+)<sup>65</sup> searches against the 15 *Erysipelothrix* complete genomes. For every query, the best hit in each distinct genome was retrieved to run a reciprocal BLASTP (v. 2.7.1+)<sup>65</sup> against the genome of *Erysipelothrix* sp. 15TAL0474. When the best hit for the reciprocal BLASTP was the same initial pseudogene in *Erysipelothrix* sp. 15TAL0474, the two sequences were considered reciprocal best hits (RBH) and therefore, orthologous genes.

**Average nucleotide identity and digital DNA–DNA hybridization.** The average nucleotide identity (ANI) and digital DNA–DNA hybridization (dDDH) values were calculated for all 17 genomes used in this study. ANI values were calculated for all pairwise comparisons using OrthoANIu algorithm<sup>85</sup> available at the EzGenome web service<sup>86</sup>. Digital DDH values were calculated using the Genome-to-Genome Distance Calculator v. 2.1 available at the GGDC website service<sup>41</sup>. Matrices of ANI and dDDH values were visualized in heatmaps using Clustvis<sup>87</sup>, with a Manhattan distance calculation and a complete linkage for rows and columns.

**Ethical approval.** This study was certified by the Animal Ethics Committee of Universidade Federal de Uberlândia, which was approved under the number A004/19. All procedures were performed in accordance with institutional guidelines and regulations of animal research.

### Data availability

The accession numbers for genomes used in this study are provided in Supplementary Table S1. Genome for de novo sequenced isolates *Erysipelothrix* sp. EsS2-6-Brazil and EsS2-7-Brazil will be made available upon publication of the manuscript.

## References

- Castelle, C. J. & Banfield, J. F. Major new microbial groups expand diversity and alter our understanding of the tree of life. *Cell* **172**, 1181–1197 (2018).
- Pallen, M. J. & Wren, B. W. Bacterial pathogenomics. *Nature* **449**, 835–842 (2007).
- Diene, S. M. *et al.* The rhizome of the multidrug-resistant *Enterobacter aerogenes* genome reveals how new “Killer Bugs” are created because of a sympatric lifestyle. *Mol. Biol. Evol.* **30**, 369–383 (2013).
- Millan-Aguinaga, N. *et al.* Phylogenomic insight into *Salinospira* (Bacteria, Actinobacteria) species designations. *Sci. Rep.* **7**, 3564. <https://doi.org/10.1038/s41598-017-02845-3> (2017).
- Orata, F. D., Meier-Kolthoff, J. P., Sauvageau, D. & Stein, L. Y. Phylogenomic analysis of the gammaproteobacterial methanotrophs (order Methylococcales) calls for the reclassification of members at the genus and species levels. *Front. Microbiol.* **9**, 3162. <https://doi.org/10.3389/fmicb.2018.03162> (2018).
- Diallo, K. *et al.* Genomic characterization of novel *Neisseria* species. *Sci. Rep.* **9**, 13742. <https://doi.org/10.1038/s41598-019-50203-2> (2019).
- Ogawa, Y. *et al.* The genome of *Erysipelothrix rhusiopathiae*, the causative agent of swine erysipelas, reveals new insights into the evolution of Firmicutes and the organism's intracellular adaptations. *J. Bacteriol.* **193**, 2951–2959 (2011).
- Kwok, A. H., Li, Y., Jiang, J., Jiang, P. & Leung, F. C. Complete genome assembly and characterization of an outbreak strain of the causative agent of swine erysipelas—*Erysipelothrix rhusiopathiae* SY1027. *BMC Microbiol.* **14**, 176. <https://doi.org/10.1186/1471-2180-14-176> (2014).
- Forde, T. *et al.* Genomic analysis of the multi-host pathogen *Erysipelothrix rhusiopathiae* reveals extensive recombination as well as the existence of three generalist clades with wide geographic distribution. *BMC Genom.* **17**, 461. <https://doi.org/10.1186/s12864-016-2643-0> (2016).
- Ogawa, Y. *et al.* Clonal lineages of *Erysipelothrix rhusiopathiae* responsible for acute swine erysipelas in Japan identified by using genome-wide single-nucleotide polymorphism analysis. *Appl. Environ. Microbiol.* **83**, e00130–e217. <https://doi.org/10.1128/AEM.00130-17> (2017).
- Bender, J. S., Shen, H. G., Irwin, C. K., Schwartz, K. J. & Opriessnig, T. Characterization of *Erysipelothrix* species isolates from clinically affected pigs, environmental samples, and vaccine strains from six recent swine erysipelas outbreaks in the United States. *Clin. Vaccine Immunol.* **17**, 1605–1611 (2010).
- Bender, J. S., Irwin, C. K., Shen, H. G., Schwartz, K. J. & Opriessnig, T. *Erysipelothrix* spp genotypes, serotypes, and surface protective antigen types associated with abattoir condemnations. *J. Vet. Diagn. Invest.* **23**, 139–142 (2011).
- Janßen, T. *et al.* A combinational approach of multilocus sequence typing and other molecular typing methods in unravelling the epidemiology of *Erysipelothrix rhusiopathiae* strains from poultry and mammals. *Vet. Res.* **46**, 84 (2015).
- Verborg, S., Göker, M., Scheuner, C., Schumann, P. & Stackebrandt, E. The families Erysipelotrichaceae emend., Coprobacillaceae fam. Nov., and Turcibacteraceae fam. Nov. In *The Prokaryotes—Firmicutes and Tenericutes* (eds Rosenberg, E. *et al.*) (Springer, Berlin, 2014).
- Rosenbach, F. J. Experimentelle, morphologische und klinische studien über krankheitsserregende mikroorganismen des schweinerotlaufs, des Erysipeloids und der mausesepicämie. *Z. Hyg. Infekt.* **63**, 343–371 (1909).
- Takahashi, T. *et al.* *Erysipelothrix tonsillarum* sp. Nov. isolated from tonsils of apparently healthy pigs. *Int. J. Syst. Evol. Microbiol.* **37**, 166–168 (1987).
- Verborg, S. *et al.* *Erysipelothrix inopinata* sp. Nov., isolated in the course of sterile filtration of vegetable peptone broth, and description of Erysipelotrichaceae fam. Nov.. *Int. J. Syst. Evol. Microbiol.* **54**, 221–225 (2004).
- Bang, B. H., Rhee, M. S., Chang, D. H., Park, D. S. & Kim, B. C. *Erysipelothrix* larvae sp. Nov., isolated from the larval gut of the rhinoceros beetle, *Trypoxylus dichotomus* (Coleoptera: Scarabaeidae). *Antonie Van Leeuwenhoek* **107**, 443–451 (2004).
- Pomaranski, E. K. *et al.* Description of *Erysipelothrix piscisicarius* sp. nov., an emergent fish pathogen, and assessment of virulence using a tiger barb (*Puntigrus tetrazona*) infection model. *Int. J. Syst. Evol. Microbiol.* **70**, 857–867 (2020).
- Wang, Q., Chang, B. J. & Riley, T. *Erysipelothrix rhusiopathiae*. *Vet. Microbiol.* **140**, 405–417 (2010).
- Takahashi, T. *et al.* *Erysipelothrix tonsillarum* isolated from dogs with endocarditis in Belgium. *Res. Vet. Sci.* **54**, 264–265 (1993).
- Takahashi, T., Fujisawa, T., Yamamoto, K., Kijima, M. & Takahashi, T. Taxonomic evidence that serovar 7 of *Erysipelothrix* strains isolated from dogs with endocarditis are *Erysipelothrix tonsillarum*. *J. Vet. Med. B Infect. Dis. Vet. Public Health* **47**, 311–313 (2000).
- Takahashi, T. *et al.* DNA relatedness among *Erysipelothrix rhusiopathiae* strains representing all twenty-three serovars and *Erysipelothrix tonsillarum*. *Int. J. Syst. Bact.* **42**, 469–473 (1992).
- Takahashi, T. *et al.* A taxonomic study on *erysipelo*thrix by DNA-DNA hybridization experiments with numerous strains isolated from extensive origins. *Microbiol. Immunol.* **52**, 469–478 (2008).
- To, H. & Nagai, S. Genetic and antigenic diversity of the surface protective antigen proteins of *Erysipelothrix rhusiopathiae*. *Clin. Vaccine Immunol.* **14**, 813–820 (2007).
- Pomaranski, E. K. *et al.* Characterization of spaC-type *Erysipelothrix* sp. isolates causing systemic disease in ornamental fish. *J. Fish. Dis.* **41**, 49–60 (2018).
- Hoepers, P. G. *et al.* First outbreak reported caused by *Erysipelothrix* species strain 2 in turkeys from poultry-producing farms in Brazil. *Ann. Microbiol.* **69**, 1211–1215 (2019).
- Stackebrandt, E. & Goebel, B. Taxonomic note: A place for DNA–DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *Int. J. Syst. Evol. Microbiol.* **44**, 846–849 (1994).
- Adékambi, T., Shinnick, T. M., Raoult, D. & Drancourt, M. Complete rpoB gene sequencing as a suitable supplement to DNA–DNA hybridization for bacterial species and genus delineation. *Int. J. Syst. Evol. Microbiol.* **58**, 1807–1814 (2008).
- Rowland, G. C., Aboshkiwa, M. & Coleman, G. Comparative sequence analysis and predicted phylogeny of the DNA-dependent RNA polymerase beta subunits of *Staphylococcus aureus* and other eubacteria. *Biochem. Soc. Trans.* **21**, 40S (1993).
- Volokhov, D. V. *et al.* Genetic analysis of housekeeping genes of members of the genus *Acholeplasma*: Phylogeny and complementary molecular markers to the 16S rRNA gene. *Mol. Phylogenet. Evol.* **44**, 699–710 (2007).
- Adékambi, T., Colson, P. & Drancourt, M. rpoB-based identification of nonpigmented and late-pigmenting rapidly growing mycobacteria. *J. Clin. Microbiol.* **41**, 5699–5708 (2003).
- Adékambi, T., Berger, P., Raoult, D. & Drancourt, M. rpoB gene sequence-based characterization of emerging non-tuberculous mycobacteria with descriptions of *Mycobacterium bolletii* sp. nov., *Mycobacterium phocaicum* sp. nov. and *Mycobacterium aubagnense* sp. Nov.. *Int. J. Syst. Evol. Microbiol.* **56**, 133–143 (2006).
- Forde, T. L. *et al.* Genomic and immunogenic protein diversity of *Erysipelothrix rhusiopathiae* isolated from pigs in great Britain: Implications for vaccine protection. *Front. Microbiol.* **11**, 418 (2020).
- Shen, H. G., Bender, J. S. & Opriessnig, T. Identification of surface protective antigen (spa) types in *Erysipelothrix* reference strains and diagnostic samples by spa multiplex real-time and conventional PCR assays. *J. Appl. Microbiol.* **109**, 1227–1233 (2010).
- Harada, K., Furui, Y. & Takahashi, T. Spa type of *Erysipelothrix* strains and its association with virulence of *Erysipelothrix* strains in mice and swine. *Afr. J. Microbiol. Res.* **6**, 7123–7127 (2012).

37. Facey, P. D. *et al.* Draft genomes, phylogenetic reconstruction, and comparative genomics of two novel cohabiting bacterial symbionts isolated from *Frankliniella occidentalis*. *Genome Biol. Evol.* **7**, 2188–2202 (2015).
38. Liu, Y. *et al.* Genomics insights into the taxonomic status of the *Bacillus cereus* group. *Sci. Rep.* **5**, 14082. <https://doi.org/10.1038/srep14082> (2015).
39. Rosselo-Mora, R. & Amann, R. The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25**, 39–67 (2001).
40. Auch, A. F., von Jan, M., Klenk, H. P. & Göker, M. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand. Genom. Sci.* **2**, 117–134 (2010).
41. Meier-Kolthoff, J. P., Auch, A. F., Klenk, H. & Göker, M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinform.* **21**, 14–60 (2013).
42. Goris, J. *et al.* DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int. J. Syst. Evol. Microbiol.* **57**, 81–91 (2007).
43. Lan, R. & Reeves, P. R. Intraspecies variation in bacterial genomes: The need for a species genome concept. *Trends Microbiol.* **8**, 396–401. [https://doi.org/10.1016/s0966-842x\(00\)01791-1](https://doi.org/10.1016/s0966-842x(00)01791-1) (2000).
44. Parker, C. T., Tindall, B. J. & Garrity, G. M. International code of nomenclature of prokaryotes. *Int. J. Syst. Evol. Microbiol.* **69**, S1–S111 (2019).
45. Li, Y. *et al.* Proteomic and transcriptomic analyses of swine pathogen *Erysipelothrix rhusiopathiae* reveal virulence repertoire. *PLoS One* **11**, e0159462 (2016).
46. Janausch, I. G., Zientz, E., Tran, Q. H., Kröger, A. & Unden, G. C4-dicarboxylate carriers and sensors in bacteria. *Biochim. Biophys. Acta* **1553**, 39–56 (2002).
47. Wösten, M. M., van de Lest, C. H., van Dijk, L. & van Putten, J. P. Function and regulation of the C4-dicarboxylate transporters in *Campylobacter jejuni*. *Front. Microbiol.* **8**, 174. <https://doi.org/10.3389/fmicb.2017.00174> (2017).
48. Overton, T. W. *et al.* Microarray analysis of gene regulation by oxygen, nitrate, nitrite, FNR, NarL and NarP during anaerobic growth of *Escherichia coli*: New insights into microbial physiology. *Biochem. Soc. Trans.* **34**, 104–107 (2006).
49. Woodall, C. A. *et al.* *Campylobacter jejuni* gene expression in the chick cecum: Evidence for adaptation to a low-oxygen environment. *Infect. Immun.* **73**, 5278–5285 (2005).
50. Guccione, E. *et al.* Amino acid-dependent growth of *Campylobacter jejuni*: Key roles for aspartase (AspA) under microaerobic and oxygen-limited conditions and identification of AspB (Cj0762), essential for growth on glutamate. *Mol. Microbiol.* **69**, 77–93 (2008).
51. Yew, W. S. & Gerlt, J. A. Utilization of L-ascorbate by *Escherichia coli* K-12: Assignments of functions to products of the yif-sga and yia-sbg operons. *J. Bacteriol.* **184**, 302–306 (2002).
52. Campos, E., Baldoma, L., Aguilar, J. & Badia, J. Regulation of expression of the divergent ulaG and ulaABCDEFG operons involved in LaAscorbate dissimilation in *Escherichia coli*. *J. Bacteriol.* **186**, 1720–1728 (2004).
53. Campos, E. *et al.* The yiaK LX1X2PQRS and ulaABCDEF G gene systems are required for the aerobic utilization of l-ascorbate in *Klebsiella pneumoniae* strain 13882 with l-ascorbate-6-phosphate as the inducer. *J. Bacteriol.* **190**, 6615–6624. <https://doi.org/10.1128/JB.00815-08> (2008).
54. Stack, T. M. M. *et al.* Characterization of an l-ascorbate catabolic pathway with unprecedented enzymatic transformations. *J. Am. Chem. Soc.* **142**, 1657–1661 (2020).
55. Benjak, A. *et al.* Insights from the genome sequence of *Mycobacterium lepraemurium*: Massive gene decay and reductive evolution. *MBio* **8**, e01283–e1317. <https://doi.org/10.1128/mBio.01283-17> (2017).
56. Benjak, A. *et al.* Highly reduced genome of the new species *Mycobacterium uberis*, the causative agent of nodular thelitis and tubercloid scrotoitis in livestock and a close relative of the leprosy bacilli. *MSphere* **3**, e00405–e418. <https://doi.org/10.1128/mSphere.00405-18> (2018).
57. Feng, Y., Chen, Z. & Liu, S. L. Gene decay in *Shigella* as an incipient stage of host-adaptation. *PLoS One* **6**, e27754. <https://doi.org/10.1371/journal.pone.0027754> (2011).
58. Blanc, G. *et al.* Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet.* **3**, e14. <https://doi.org/10.1371/journal.pgen.0030014> (2007).
59. Fraser, C. M. *et al.* The minimal gene complement of *Mycoplasma genitalium*. *Science* **270**, 397–403 (1995).
60. Soderlund, R. *et al.* Comparative genome analysis of *Erysipelothrix rhusiopathiae* isolated from domestic pigs and wild boars suggests host adaptation and selective pressure from the use of antibiotics. *Microbial. Genom.* <https://doi.org/10.1099/mgen.0.000412> (2020).
61. Andrews, S. FastQC: A quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (2010).
62. Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
63. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).
64. FastOrtho. <https://github.com/olsonanl/FastOrtho>.
65. Li, L., Stoeckert, C. J. Jr. & Roos, D. S. OrthoMCL: Identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).
66. Camacho, C. *et al.* BLAST+: Architecture and applications. *BMC Bioinform.* **10**, 421. <https://doi.org/10.1186/1471-2105-10-421> (2009).
67. Stijn van Dongen. Graph clustering by flow simulation. PhD thesis, University of Utrecht. <http://www.library.uu.nl/digearchief/dip/diss/1895620/inhoud.htm> (2000).
68. Marchler-Bauer, A. *et al.* CDD: A Conserved domain database for the functional annotation of proteins. *Nucleic Acids Res.* **39**, D225–D229 (2010).
69. Batch CD-Search Tool. <https://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>.
70. Galperin, M. Y., Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Res.* **43**, D261–D269 (2015).
71. El-Gebali, S. *et al.* The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432 (2019).
72. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726–731 (2016).
73. Aramaki, T. *et al.* KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btz859> (2019).
74. Johnson, M. *et al.* NCBI BLAST: A better web interface. *Nucleic Acids Res.* **36**, W5–W9 (2008).
75. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
76. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
77. Darrriba, D. *et al.* ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Mol. Biol. Evol.* **37**, 291–294 (2020).
78. Miller, M. A., Pfeiffer, W. & Schwartz, T. Creating the CIPRES science gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)*, 14 Nov. 2010, New Orleans, LA pp 1–8 (2010).

79. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
80. Ronquist, F. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
81. Rambaut, A. FigTree v. 1.4.2. Institute of Evolutionary Biology, University of Edinburgh. <https://github.com/rambaut/figtree/releases> (2014).
82. Kück P & Longo GC (2016) FASconCAT-G v. 1.04. The Zoological Research Museum Alexander Koenig. <https://github.com/PatrickKueck/FASconCAT-G>.
83. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
84. Lanfear, R., Calcott, B., Kainer, D., Mayer, C. & Stamatakis, A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol. Biol.* **14**, 82 (2014).
85. Lee, I., Ouk Kim, Y., Park, S. C. & Chun, J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int. J. Syst. Evol. Microbiol.* **66**, 1100–1103 (2016).
86. Yoon, S. H., Ha, S. M., Lim, J. M., Kwon, S. J. & Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **110**, 1281–1286 (2017).
87. Metsalu, T. & Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using principal component analysis and heatmap. *Nucleic Acids Res.* **43**, W566–W570 (2015).

## Acknowledgements

We would like to thank the Brazilian agencies that supported students with fellowships: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). We are grateful to all local farmers who supported this research providing samples for this study.

## Author contributions

A.L.G. and N.M.V. designed the study, performed bioinformatics analyses, analysed the results and wrote the manuscript. P.G.H., T.M.R., B.B.F., J.P.Z. collected samples, performed bacterial isolation and molecular identification and contributed with reagents. D.M., B.C.B., M.I. and L.F.C. performed the whole genome sequencing and assembling and contributed with reagents. All authors read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-82959-x>.

**Correspondence** and requests for materials should be addressed to A.L.G. or N.M.V.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021