

# SCIENTIFIC REPORTS



OPEN

## *Legionella pneumophila* strain associated with the first evidence of person-to-person transmission of Legionnaires' disease: a unique mosaic genetic backbone

Received: 25 February 2016

Accepted: 29 April 2016

Published: 19 May 2016

Vítor Borges<sup>1,\*</sup>, Alexandra Nunes<sup>1,\*</sup>, Daniel A. Sampaio<sup>2</sup>, Luís Vieira<sup>2</sup>, Jorge Machado<sup>3</sup>, Maria J. Simões<sup>4</sup>, Paulo Gonçalves<sup>4</sup> & João P. Gomes<sup>1</sup>

A first strong evidence of person-to-person transmission of Legionnaires' Disease (LD) was recently reported. Here, we characterize the genetic backbone of this case-related *Legionella pneumophila* strain ("PtVFX/2014"), which also caused a large outbreak of LD. PtVFX/2014 is phylogenetically divergent from the most worldwide studied outbreak-associated *L. pneumophila* subspecies *pneumophila* serogroup 1 strains. In fact, this strain is also from serogroup 1, but belongs to the *L. pneumophila* subspecies *fraseri*. Its genomic mosaic backbone reveals eight horizontally transferred regions encompassing genes, for instance, involved in lipopolysaccharide biosynthesis or encoding virulence-associated Dot/Icm type IVB secretion system (T4BSS) substrates. PtVFX/2014 also inherited a rare ~65 kb pathogenicity island carrying virulence factors and detoxifying enzymes believed to contribute to the emergence of best-fitted strains in water reservoirs and in human macrophages, as well as a inter-species transferred (from *L. oakridgensis*) ~37.5 kb genomic island (harboring a *lvh/lvr*T4ASS cluster) that had never been found intact within *L. pneumophila* species. PtVFX/2014 encodes another *lvh/lvr* cluster near to CRISPR-associated genes, which may boost *L. pneumophila* transition from an environmental bacterium to a human pathogen. Overall, this unique genomic make-up may impact PtVFX/2014 ability to adapt to diverse environments, and, ultimately, to be transmitted and cause human disease.

Legionnaires' disease (LD) is a severe pneumonia with a case fatality rate of 8–12%, reaching higher values in elderly, smokers, and people with preexisting medical conditions, such as chronic cardiovascular or respiratory disease and diabetes<sup>1–2</sup>. LD is caused by *Legionella* spp., with *L. pneumophila* serogroup 1 accounting for most of the diagnosed cases<sup>3–5</sup>. *L. pneumophila* is a facultative intracellular Gram-negative aerobic bacterium that is widespread in nature, where water is its major reservoir, growing at the temperature range from 25 °C to 42 °C<sup>6</sup>. Transmission usually occurs by inhalation of aerosols or aspiration of water containing these bacteria<sup>6,7</sup>. Nevertheless, contrarily to what has been historically postulated, we have recently reported that person-to-person transmission of LD may occur<sup>8</sup>. LD has an incubation period typically ranging from 2–10 days, although longer and shorter periods have been described<sup>2</sup>. The vast majority of LD cases is sporadic, but large outbreaks may also occur<sup>9–13</sup>. The first reported outbreak dates back to 1976 and affected the members of an American Legion convention in Philadelphia, from which resulted the current disease designation<sup>14</sup>. The most common sources of outbreaks are contaminated cooling towers and hot water plumbing infrastructures, but decorative fountains, whirlpools or home showers<sup>15</sup> have also been involved<sup>2</sup>. While the standardized sequence-based typing (SBT) has been useful for outbreak source identification and to study the molecular epidemiology of LD<sup>16,17</sup>, whole-genome

<sup>1</sup>Bioinformatics Unit and Research Unit, National Institute of Health, Lisbon, Portugal. <sup>2</sup>Innovation and Technology Unit, National Institute of Health, Lisbon, Portugal. <sup>3</sup>Coordination of the Department of Infectious Diseases, National Institute of Health, Lisbon, Portugal. <sup>4</sup>National Reference Laboratory for Legionella, National Institute of Health, Lisbon, Portugal. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to J.P.G. (email: j.paulo.gomes@insa.min-saude.pt)

sequencing (WGS) is emerging as a more informative technology<sup>18–21</sup> since it enables an in-depth investigation of outbreak-related strains and a higher resolution power for source attribution. It is also imperative to use WGS to deeply genetically characterize the strains associated with either endemic cases or large outbreaks. In fact, the conjugation of these two research areas (molecular epidemiology and comparative genomics) may considerably enhance our knowledge on LD and, thus, improve public health interventions.

One of the largest outbreaks of LD worldwide occurred in 2014 in Vila Franca de Xira (VFX), Portugal, yielding more than 400 cases and 14 deaths<sup>22</sup>. Epidemiological, environmental, microbiological and sequence-based molecular analyses traced industrial wet cooling systems as the potential source of infection and revealed that the outbreak-related strain (here designated PtVFX/2014) was a *L. pneumophila* serogroup 1 displaying the novel sequence type (ST) 1905<sup>13</sup>. Remarkably, it was possible to implicate this *L. pneumophila* ST1905 strain into the strongest evidence to date of person-to-person transmission of LD<sup>8</sup>. Considering the huge genetic diversity within the *L. pneumophila* species<sup>11,23–27</sup>, marked by the presence of key virulence traits (like the well-described Dot/Icm type IVB secretion system – T4BSS) encoded in both core- and accessory-genomes<sup>28,29</sup>, we aimed to deeply characterize the genetic backbone of this novel person-to-person transmission- and outbreak-related strain by integrating it in the frame of the species phylogeny and diversity.

## Results and Discussion

**WGS-based outbreak investigation.** In 2014, Portugal hosted the second largest outbreak worldwide of LD so far, which was caused by a novel ST1905 *L. pneumophila* serogroup 1 strain (PtVFX/2014) likely originated from local industrial cooling towers<sup>13</sup>. WGS was applied to strengthen the investigation of this LD outbreak as it provides a higher level of resolution than the traditional SBT scheme<sup>18–21</sup>, which is based on the sequence of seven loci (involving only ~2500 bp out of the ~3.4 Mbp of the *L. pneumophila* genome)<sup>16,17</sup>. The draft genome sequences of the outbreak-related strains were ~3.47 Mb in length. We found no nucleotide differences after performing pairwise comparisons between the environmental isolate and each clinical sample (using near-complete genome sequences; >99.7% of each draft genome), which suggested an industrial cooling tower as the potential source of the LD outbreak. The scenario of genomic identity among same-outbreak isolates observed here and in several other LD outbreaks is expected<sup>9,10,21</sup>, but contrasts with the recent findings described by Sánchez-Buso and colleagues<sup>12</sup>, who remarkably found non-clonal relationships among isolates epidemiologically linked to the same outbreak. In the LD outbreak occurred in Portugal, the observed clonality likely reflects the existence of a single physical environmental source with a *L. pneumophila* population markedly represented by the causative infection clone.

Of note, confirmation of the SBT-based ST1905 allelic profile highlighted a bias associated with the *in silico* extraction of the allele sequence from WGS data, as previously noticed by other authors<sup>18</sup>. In fact, the PtVFX/2014 strain displays two non-matching *mompS* copies (assigned with distinct allele numbers, 15 *versus* 7), which would hamper a proper ST attribution if the allelic profile was exclusively determined *in silico*.

Taken together, WGS enabled an in-depth investigation of the recent LD outbreak that occurred in Portugal, as it allowed the identification, with a high level of certainty, of an industrial cooling tower as the source of the epidemic.

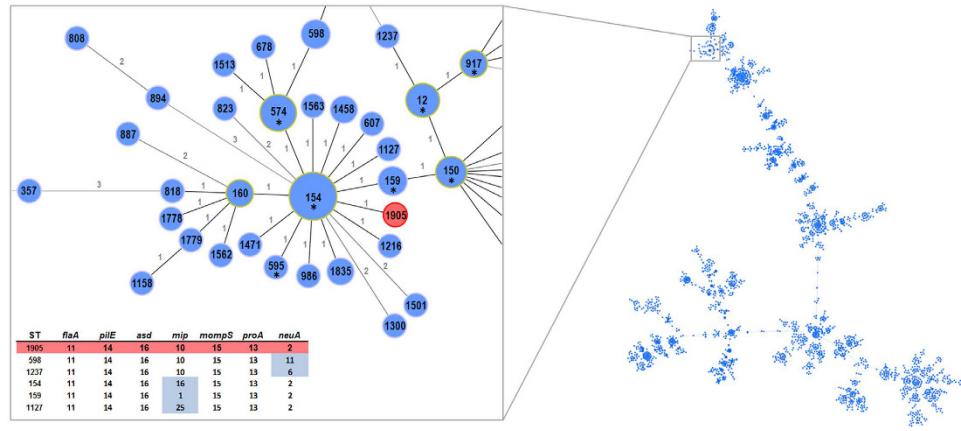
### SBT-based integration of PtVFX/2014 within the worldwide *L. pneumophila* genetic diversity.

In order to integrate the novel ST1905 in the worldwide epidemiological picture of the *L. pneumophila*, the allelic profile of PtVFX/2014 was compared with all STs available at the European Working Group for *Legionella* Infections (EWGLI) SBT database<sup>16,17</sup> (Fig. 1). ST1905 was found to differ just by one allele from ST154, ST159 and ST1127 (in *mip* locus), as well as from ST598 and ST1237 (in *neuA* locus). Nevertheless, these allelic profiles differ by seven to eight SNPs to ST1905, with exception of ST159 which displays only one nucleotide difference, being thus the closest ST to ST1905 (data not shown). No linkage can be made between the geographical source and the cluster enrolling these most close related STs, as isolates from these STs were already detected in worldwide spread countries (e.g., USA, Canada, UK, Italy, France, China, Japan, among others). For instance, ST159 has been described both in Europe and Asia.

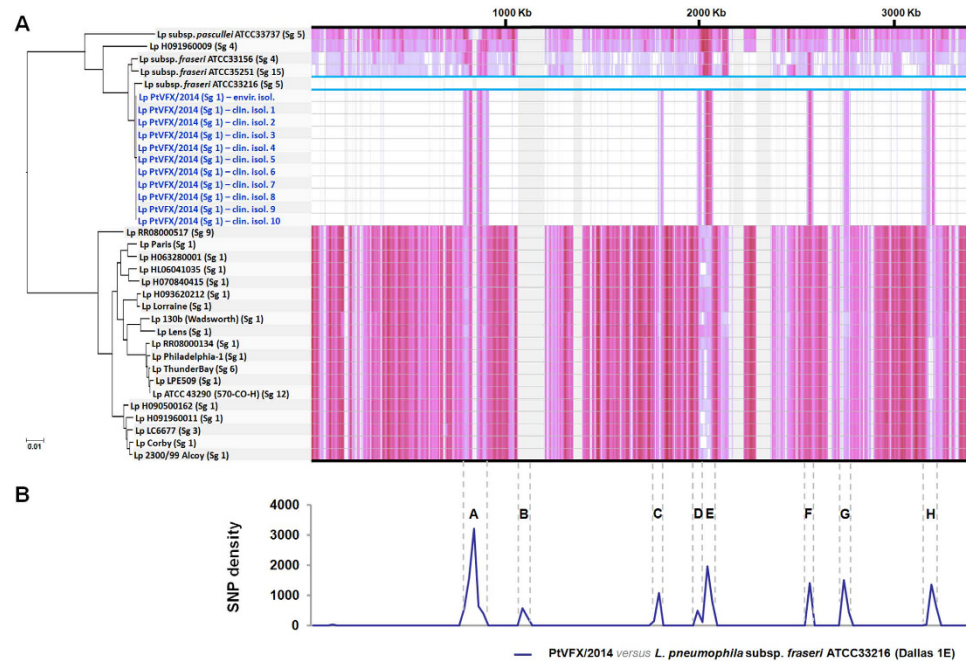
### Integration of the genetic backbone of the PtVFX/2014 strain in the frame of the species phylogeny and diversity.

To get insight on the genetic backbone of the PtVFX/2014 strain and to infer its position within the phylogeny of *L. pneumophila* species, we constructed a core-genome alignment (encompassing both coding and non-coding regions) that comprises ~64% of the genome sequence of each of the 11 outbreak-associated clones and of the 24 selected publicly available strains. This comparative analysis enrolls isolates from different *L. pneumophila* subspecies (*pneumophila*, *fraseri* and *pasculli*) that were specifically selected in order to handle as much genetic diversity as possible. This contrasts with most of the genome-based comparative studies performed so far, which have been mainly focused on *L. pneumophila* subsp. *pneumophila* (mainly through the analysis of core genes)<sup>24–26,30,31</sup>.

Intriguingly, the serogroup 1 PtVFX/2014 clones were found to be phylogenetically close related with strains from *L. pneumophila* subsp. *fraseri* (Fig. 2A), and were clearly distinct from the most worldwide studied *L. pneumophila* serogroup 1 strains (Philadelphia-1, Paris, Lens, Corby, and 2300/99 Alcoy). For instance, we detected 224376 variant sites (within a ~87% core-genome) when comparing the PtVFX/2014 strain with the well-studied serogroup 1 *L. pneumophila* Philadelphia-1 strain, whereas the pairwise genome comparison against the most close related *L. pneumophila* subsp. *fraseri* ATCC33216 (or Dallas 1E) strain revealed just up to 17000 nucleotide differences within a ~98% core-genome. SNP density analysis throughout the genome revealed a non-random distribution, as those SNPs are clustered into eight distinct chromosome regions (Fig. 2A,B), which strongly suggests their inheritance by horizontal gene transfer (HGT). Overall, these HGT-inherited regions involve around 250 genes and represent ~8% of the genome length (Table 1). One of the clearest examples of a crossover region



**Figure 1.** SBT-based representation of the worldwide genetic diversity among *L. pneumophila* isolates. Relationships between allelic profiles of 10085 isolates (from the EWGLI SBT database) are shown in a goeBURST full minimum spanning tree (MST) constructed using PHYLOViZ software. The MST connects the ST profiles in a way that the summed distance of all links of the tree is the minimum. STs are indicated by numbers within circles, where connecting lines are labeled with the number of allelic differences between STs. The cluster including the PtVFX/2014 ST1905 strain (in red) is zoomed, where STs found in more than one country are labeled with an asterisk. The matrix describes the ST profiles differing by a single allele from the ST1905.



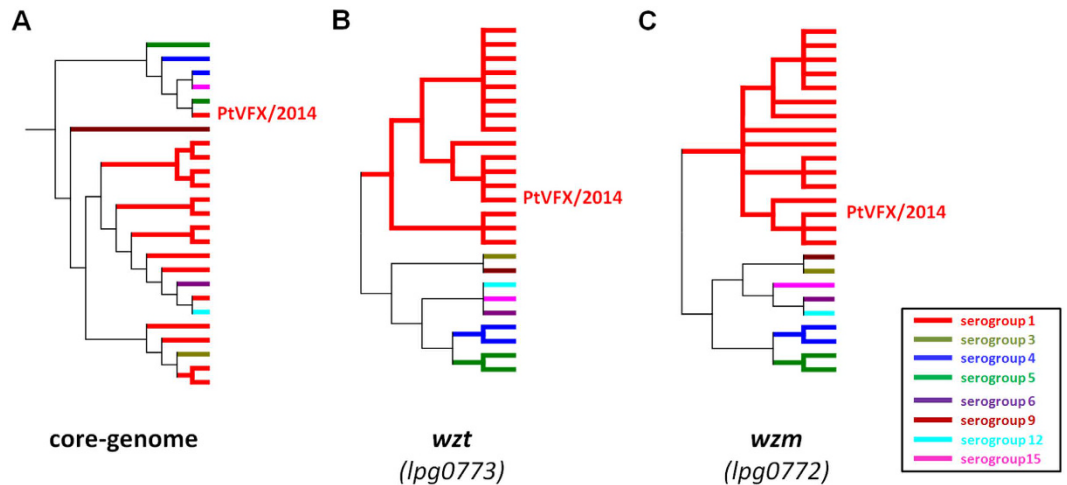
**Figure 2.** Phylogenomic analysis of PtVFX/2014 strain within the *L. pneumophila* species. (A) A phylogenetic reconstruction enrolling genome sequences from 24 *L. pneumophila* strains and from 11 outbreak-associated PtVFX/2014 clones (ten clinical isolates plus one environmental isolate; in dark blue) is shown paired with the corresponding SNP density plot across the length of the genome. Serogroup (Sg) classification is also shown (see Supplementary Table S1 for details). The tree is drawn to scale, with branch lengths reflecting number of base substitutions *per site*. SNP density in the core-genome (~64% of the genome) is depicted proportionally to the color intensity and reflects SNPs against the genome of the *L. pneumophila* subsp. *fraseri* ATCC33216 strain (surrounded with a light blue box). Areas in grey refer to regions in the accessory genome. Genome orientation is according to that of the *L. pneumophila* Philadelphia-1 strain (GenBank accession number AE017354). (B) Detailed pairwise SNP density plot between PtVFX/2014 and the most closely related *L. pneumophila* subsp. *fraseri* ATCC33216 strain showing putative HGT-inherited regions (labeled from A to H; see details in Table 1). Calculations of SNP density were performed across the genome alignment over a sliding window (window size = 25000 bp; step size = 25000 bp). Vertical dashed lines delimiting each region make the correspondence with the SNP density plot of panel (A).

Region	Estimated length (Kbp)	Genome location		Presence of Dot/Icm effectors coding genes <sup>c</sup>		Relevant genes included <sup>a, d</sup>
		<i>loci</i> Philadelphia-1 strain <sup>a</sup>	<i>loci</i> PtVFX/2014 <sup>b</sup>	Philadel-phia-1 <sup>a</sup>	PtVFX/2014 <sup>b</sup>	
A	113	<i>lpg0831-lpg0723</i>	PtVFX2014_03240-03730	<i>lpg0733</i>	PtVFX2014_03780	LPS biosynthesis loci ( <i>lpg0748-lpg0779</i> ) <sup>c</sup> (VFDB <sup>39</sup> ); <i>lpg0791/mip</i> <sup>40</sup> ; The iron assimilation- and virulence-associated <i>lpg0746/iraAB</i> locus (VFDB <sup>37</sup> ); <i>lpg0742/ldsA</i> <sup>38</sup>
B	21	<i>lpg1025(yegE)-lpg0986</i>	PtVFX2014_08445-08505	—	—	Genes potentially associated with resistance to antibiotics and toxic compounds
C	25	<i>lpg1640-lpg1619(ftsI4)</i>	PtVFX2014_04280-04395	<i>lpg1639</i> ; <i>lpg1625</i> ; <i>lpg1621</i>	PtVFX2014_04285; PtVFX2014_04365; PtVFX2014_04385	MecI family transcriptional regulator associated with β-lactams resistance ( <i>lpg1620</i> )
D	8	<i>lpg1794-lpg1800(recX)</i>	PtVFX2014_11245-11275	<i>lpg1798</i>	PtVFX2014_11260	Transcriptional regulator LysR ( <i>lpg1796</i> ); Regulatory protein RecX ( <i>lpg1800</i> ).
E	33	<i>lpg1817-lpg1845(vacI)</i>	PtVFX2014_11360-11500	<i>lpg1822</i> ; <i>lpg1836</i>	PtVFX2014_11385; PtVFX2014_11455	<i>lpg1845/vacI</i> : encodes a lipoprotein associated with peptidoglycan stabilization (potential virulence factor in cell to cell migration <sup>36</sup> ); <i>lpg1818/msbA</i> : encodes the Lipid A export ATP-binding/permease protein MsbA <sup>41</sup>
F	20	<i>lpg2282(asnS)-lpg2262(ackA2)</i>	PtVFX2014_01325-01425	<i>lpg2271</i>	PtVFX2014_01380	—
G	34	<i>lpg2462-lpg2426(mdcA)</i>	PtVFX2014_00340-00520	<i>lpg2461</i> ; <i>lpg2456</i> ; <i>lpg2452</i> ; <i>lpg2444</i> ; <i>lpg2443</i> ; <i>lpg2434</i> ; <i>lpg2433</i>	PtVFX2014_00345; PtVFX2014_00370; PtVFX2014_00395; PtVFX2014_00425; PtVFX2014_00430; PtVFX2014_00480; PtVFX2014_00485	<i>lpg2438</i> : encodes a florfenicol efflux pump
H	33	<i>lpg2815-lpg2842(phoH)</i>	PtVFX2014_14890-15010	<i>lpg2815</i> ; <i>lpg2826</i> ; <i>lpg2828</i> ; <i>lpg2832</i>	PtVFX2014_14890; PtVFX2014_14945; PtVFX2014_14955; PtVFX2014_14960	<i>lpg2839/smpB</i> (gene likely essential for viability of <i>L. pneumophila</i> <sup>42</sup> ); <i>lpg2825</i> : gene associated with cold shock; <i>lpg2841</i> : putative antibiotic resistance-associated gene.

**Table 1. Description of HGT-inherited regions in PtVFX/2014.** <sup>a</sup>*Loci* designations are based on the Philadelphia-1 strain genome annotation (GenBank accession number AE017354). <sup>b</sup>*Loci* designations are based on the PtVFX/2014 strain genome annotation (GenBank accession number LORH00000000) (this study). <sup>c</sup>Presence/absence analysis based on 303 known Dot/Icm substrates coding genes that were recently gathered in the literature for six sequenced *L. pneumophila* strains (Philadelphia-1, 2300/99 Alcoy, Corby, Paris, Lens and 130b). <sup>d</sup>Based on specific literature and/or on the Virulence Factors Database (VFDB) (<http://www.mgc.ac.cn/VFs/>). <sup>e</sup>Correspond to the *loci* PtVFX2014\_03470-03605. The genes *lpg0778* and *lpg0675* are absent in the PtVFX/2014 strain.

(affecting the gene *lpg1817*), where PtVFX/2014 genome abruptly exhibits a high concentration of homoplasmic mutations, is shown in Supplementary Figure S1. This is not surprising as it is well-known that recombination events and HGT are major driving forces in shaping the highly dynamic *L. pneumophila* genomes<sup>12,23–25</sup>, with each strain possessing its own unique genomic mosaic as an outcome of its environmental context and its specific adaptive needs to infect different hosts or to persist in the environment<sup>32,33</sup>. For instance, recombination was found to be responsible for introducing genetic variation even in a *L. pneumophila* population markedly associated with recurrent outbreaks, with implications in public health interventions for outbreak investigation and source attribution<sup>12</sup>.

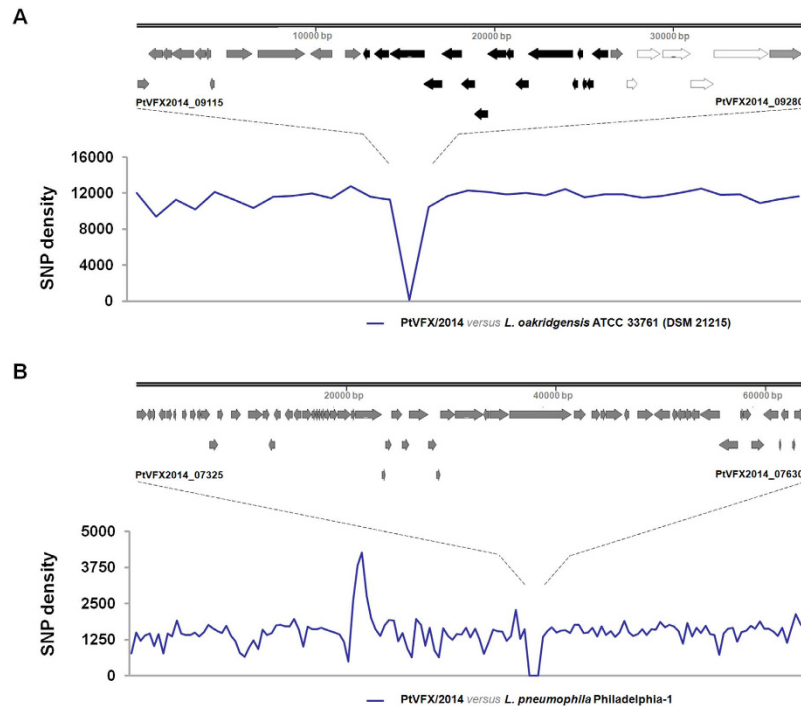
Among the *loci* imported from the detected HGT events in the PtVFX/2014 strain (Table 1), we highlight genes coding for proteins: *i*) from the lipopolysaccharide (LPS) biosynthesis cluster; *ii*) potentially associated with resistance to antibiotics and toxic compounds, or with iron acquisition; *iii*) that likely impact the bacterium ability to infect macrophages; and, *iv*) that are potentially essential for viability and cell to cell migration of *L. pneumophila*. Phylogenetic analysis of two genes (*lpg0773/wzm* and *lpg0772/wzt*) from the LPS cluster that have been used for genetic serogroup classification<sup>34,35</sup> strongly sustains that the HGT inheritance of the LPS cluster is responsible for the serogroup 1 profile of PtVFX/2014 (Fig. 3). This is especially relevant since *L. pneumophila* serogroup 1 isolates are predominantly reported in LD cases worldwide<sup>3,5</sup>, although PtVFX/2014 is evolutionarily divergent from the *L. pneumophila* subsp. *pneumophila* serogroup 1 strains that have caused most of large outbreaks of LD described so far<sup>9,10,24,26,36</sup>. On the other hand, it corroborates that the LPS cluster determining serogroup 1 can be present in highly diverse genomic backbones, and that itself likely constitutes a major determinant of human disease<sup>35</sup>. It is also worth noting that this HGT-derived region also involves the *inaAB* locus, which was shown to be required for assimilation, intracellular infection and virulence<sup>37</sup>, as well as a gene (*lpg0742/ldsA*) coding for a protein that may affect virulence by interfering with the assembly and/or the activation of the Dot/Icm system<sup>38</sup>.



**Figure 3. Genetic basis underlying the serogroup classification of PtVFX/2014.** Schematic non-scaled condensed trees showing the genetic relatedness among 25 *L. pneumophila* strains (Supplementary Table S1) based on the core-genome alignment from Fig. 2A (panel A), as well as on the individual alignments of two genes that have been applied for genetic serogroup classification: *lpg0773/wzt* (panel B) and *lpg0772/wzm* (panel C). For simplification purposes, only one clone of the serogroup 1 PtVFX/2014 strain is presented in all trees. The core-genome tree topology matches the one of Fig. 2A, while the *lpg0773/wzt* and *lpg0772/wzm* trees topologies were inferred in MEGA5 using the Neighbor-Joining method, with Kimura 2-parameter model for estimating evolutionary distances (Bootstrapping = 1000 replicates). Branches are colored according to the serogroup classification.

Also important was the exchange of multiple genes encoding substrates of the Dot/Icm T4BSS (Table 1). It is known that critical virulence traits of bacterial pathogens rely on a wide variety of secretion systems that translocate effector proteins into the host cell in order to promote infection and survival<sup>43,44</sup>, where the Dot/Icm T4BSS is the major essential virulence factor in *L. pneumophila*<sup>28,44,45</sup>. In this regard, those recombination events involving Dot/Icm T4BSS effectors may impact the PtVFX/2014 ability to infect, to be transmitted, and to cause disease. For instance, we highlight two exchanged Dot/Icm effectors (Lpg0733/RavH and Lpg2461) that were previously implicated in the modulation of eukaryotic vesicular trafficking<sup>46</sup>, and the protective B-cell antigen Lpg2271<sup>47,48</sup> (Table 1). Of note, multiple homoplasic mutations shared with the most worldwide studied outbreak-associated *L. pneumophila* serogroup 1 strains were detected in the genes encoding the Dot/Icm substrates Lpg2271 and Lpg1639. The mutational scenario observed for *lpg1639* (Supplementary Fig. S2) was even more marked, since PtVFX/2014 is clearly divergent from the close-related strains from *L. pneumophila* subsp. *fraseri* or *L. pneumophila* subsp. *pascullei* in this gene. Given their relevance, a detailed description of the repertoire of known Dot/Icm effectors carried by the PtVFX/2014 strain is presented below. Taken together, these results point that the outbreak-related PtVFX/2014 strain acquired by HGT multiple genetic features that may be associated with specific virulence or transmission traits.

**Other virulence-associated traits of the *L. pneumophila* PtVFX/2014 strain.** It is well-known that the *L. pneumophila* genome is highly plastic, with an accessory genome marked by several pathogenicity islands, plasmid-like elements, phage-related genes and other virulence-associated loci potentially related to its ability to adapt and survive in diverse environments and to cause human disease<sup>12,20,24,25,49</sup>. From the inspection of the PtVFX/2014 accessory genome for either known genomic islands or strain-specific genetic traits, we highlight the presence of an ~37.5 kb region that was found to be intact and highly similar (BLASTn, cover 100%, identity 99%, E-value 0.0) in only one strain (ATCC 33761 = DSM 21215) of *L. oakridgensis*, a species that rarely causes LD<sup>50</sup>. The SNP density analysis throughout the core-genome alignment of *L. pneumophila* PTVFX/2014 versus *L. oakridgensis* ATCC 33761 (DSM 21215) (Fig. 4A) confirms the similarity, and clearly points that this region was acquired by inter-species HGT. It carries genes potentially involved in conjugation, such as a *lvh/lvr* cluster coding for a typical T4SS (T4ASS) and some *tra* conjugal transfer genes<sup>26,36,51</sup>, as well as a gene cluster homolog to a type I-like restriction-modification system. While the *lvh/lvr* T4ASS is thought to be involved in *L. pneumophila* entry, delay of phagosome acidification and intracellular multiplication, which are virulence-related mechanisms that may boost its transition from an environmental bacterium to a human pathogen<sup>52</sup>, little is known about the role of the restriction-modification systems in *L. pneumophila*, although they have been suggested to be involved in its adaptation to different environmental niches<sup>32,53</sup>. Noteworthy, this ~37.5 kb genomic island was not found to be intact in any *L. pneumophila* strain with the genome available at the GenBank (including the closest related *L. pneumophila* subsp. *fraseri* ATCC33216/Dallas 1E strain). In fact, BLASTn analyses of this island only revealed partial homology (cover < 67%, identity < 98%, E-value 0.0) mainly targeting the region involving the T4ASS *lvh/lvr* cluster in other *L. pneumophila* strains. However, this might be explained by the presence of T4ASS *lvh/lvr* cluster integrated in either other genomic islands or genome locations. This is not surprising as *L. pneumophila* genomes may carry more than one *lvh/lvr* T4ASS cluster<sup>27,30,31,36,49</sup>. Indeed, for instance, PtVFX2014



**Figure 4. Relevant putative pathogenicity islands inherited by PtVFX/2014.** (A) The graph shows the SNP density across a core-genome alignment (encompassing about 39–49% of each whole-genome sequence) between PtVFX/2014 and the *L. oakridgensis* ATCC 33761 (or DSM 21215) strain over a sliding window (window size = 38000 bp; step size = 38000 bp). Both gene content and organization of the ~37.5 kb genomic island is highlighted above the graph, where the *lvh/lvr* cluster is shown in black arrows, while the gene cluster homolog to a type I-like restriction-modification system is displayed in white arrows. (B) The graph shows the SNP density across a core-genome alignment (encompassing about 86–88% of each whole-genome sequence) between PtVFX/2014 and the *L. pneumophila* Philadelphia-1 strain over a sliding window (window size = 20000 bp; step size = 20000 bp). Gene organization of the ~65 kb genomic island is highlighted above the graph. For both graphs, core-genome alignments were extracted (using MAUVE software) by keeping and concatenating regions where genomes aligned over at least 500 bp, x-axis do not reflect any genome orientation. Genes delimiting each genomic island are labeled in each panel.

strain possesses an additional *lvh/lvr* T4ASS cluster upstream to this ~37.5 kb genomic island. Interestingly, in close proximity of this additional *lvh/lvr* cluster, we detected three genes (*PtVFX2014\_08925/cas1*, *PtVFX2014\_08930/cas2* and *PtVFX2014\_08930/cas4*) associated with the clustered regularly interspaced short palindromic repeats (CRISPR) system (CRISPR-associated genes – *cas*). The CRISPR-Cas system is a RNA-mediated immunity system used by some bacteria to resist to phages and other invading DNA<sup>54</sup>. In *L. pneumophila*, although little is known about the impact of the CRISPR system on regulation of virulence-related traits<sup>55</sup>, it has been demonstrated that the CRISPR-associated gene *cas2* of *L. pneumophila* strain 130b is required for intracellular infection of amoebae, indicating that it might play a role in the environment persistence and transmission of LD<sup>56</sup>. Of note, *cas* genes in the PtVFX/2014 strain resemble those annotated in the *L. pneumophila* strains 130b and Paris<sup>26,27</sup>, but not in the 2300/99 Alcoy and Lens strains<sup>55</sup>.

Another relevant feature of PtVFX/2014 is the presence of a ~65 kb pathogenicity island that was previously reported to be absent in the most worldwide studied *L. pneumophila* strains<sup>24,57</sup>, with exception of the Philadelphia-1 strain, in which it was firstly found<sup>58</sup>. Like above, this PtVFX/2014 island was likely imported via HGT, being highly similar (Fig. 4B) to the one described in the Philadelphia-1 strain<sup>58</sup>. This island was recently found to mediate oxidative stress resistance *in vitro* and in macrophages<sup>57</sup>, carrying a plasmid F-like element comprising homologs of the T4SS (*tra/trb* genes), a *prpA-lvrABC* gene cluster (*PtVFX2014\_07415-PtVFX2014\_07430*), several membrane transporters, and multiple other potential virulence factors<sup>58</sup>. For instance, it harbors a homolog of an efflux pump YdhE/NorM belonging to the multidrug and toxin extrusion (MATE) family (*PtVFX2014\_07605*), which has been associated with resistance to several antimicrobial agents in other bacteria<sup>59</sup>, and a gene encoding a macrophage infectivity potentiator-related protein (*lpg2112/PtVFX2014\_07340*)<sup>32,60</sup>. The *prpA-lvrABC* gene cluster includes a *csrA* paralog (*lpg2094/PtVFX2014\_07430*) encoding a RNA-binding protein that belongs to a superfamily of global regulators that are widely distributed among bacterial species. In *L. pneumophila*, CsrA-like proteins have been implicated in several mechanisms, including the bacterial switch to a resilient state yielding persistence in water supplies or the regulation of the transition between the replicative and transmissive phases of its pathogenic life cycle<sup>46,61–63</sup>. In conclusion, similarly to the outbreak-related Philadelphia-1 strain, the presence of this island endows PtVFX/2014 with machinery for excision and transfer as well as with a number of putative virulence factors and detoxifying enzymes that are believed

to contribute to the emergence and persistence of best-fitted strains in natural or human-built water reservoirs and in human macrophages<sup>57,58</sup>.

**Repertoire of Dot/Icm T4BSS substrates in PtVFX/2014 strain.** *L. pneumophila* encodes over 300 Dot/Icm T4BSS effectors that are translocated during the bacterial life-cycle into the host cell to subvert cellular functions, and thus, assuring intracellular survival and replication<sup>47,49,64,65</sup>. For instance, Dot/Icm effectors have been implicated in the bacterial uptake into the host cells, evasion of lysosome fusion, recruitment of host vesicles, lipid remodeling of the *Legionella*-containing vacuole, modulation of important cellular pathways (such as the host ubiquitin pathways) or host cell exit<sup>29,66–68</sup>. The majority of Dot/Icm effectors are shared by most *L. pneumophila* strains sequenced so far<sup>47,49,64,65</sup>. Nevertheless, there are still a considerable number of effectors that have been suggested to be present in just few strains or even to be strain-specific, making the investigation of the repertoire of Dot/Icm T4BSS effectors encoded in novel sequenced genomes important to understand potential strain-specific virulence or transmission traits.

In this regard, we screened the PtVFX/2014 genome for the presence/absence of 303 known Dot/Icm substrates coding genes that were recently gathered in the literature for six sequenced *L. pneumophila* strains (Philadelphia-1, 2300/99 Alcoy, Corby, Paris, Lens and 130b)<sup>49,64</sup>. This kind of presence/absence analysis has been mostly based on pairwise comparisons with data from the *L. pneumophila* Philadelphia-1 strain, so potential Dot/Icm effectors identified for the other sequenced strains were also not taken into account in the current study. Overall, 232 out of the 303 Dot/Icm substrates (~77%) are shared by the seven strains (Supplementary Table S2), which corroborates previous comparisons<sup>47,49,64,65</sup>. This observation gains even more relevance, since PVFX/2014 belongs to *L. pneumophila* subsp. *fraseri*, being phylogenetically distant from the strains used so far for Dot/Icm make-up comparisons. Thus, it is tempting to speculate that this set of ~230 shared substrates may constitute the “core” Dot/Icm arsenal within the *L. pneumophila* species.

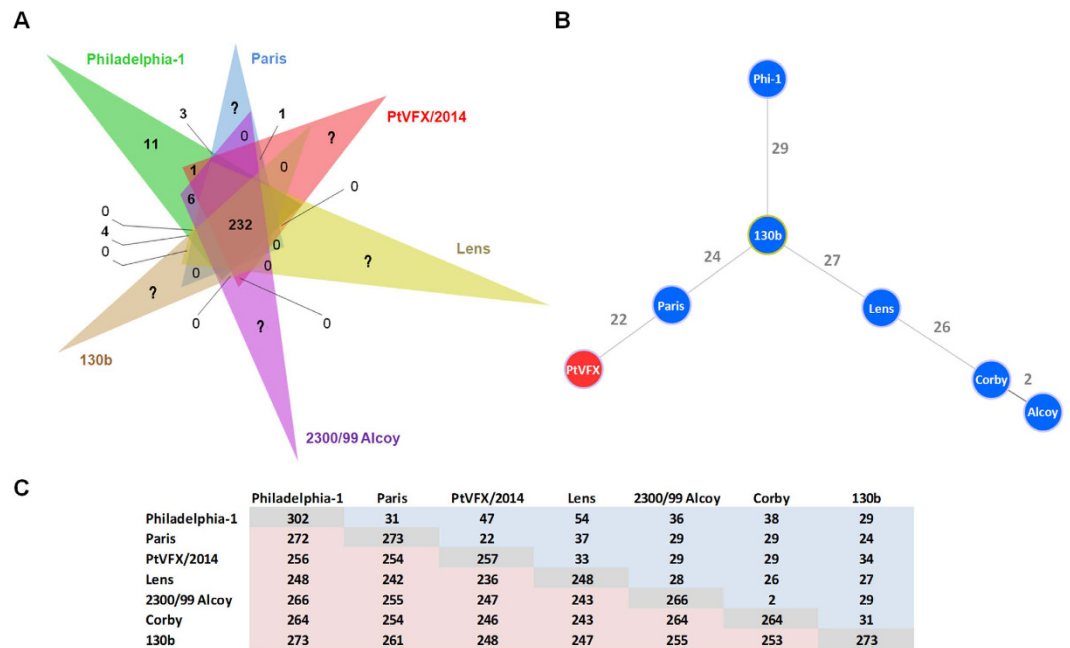
Regarding PtVFX/2014, we identified 257 known Dot/Icm substrates (Supplementary Table S2), whose repertoire is most similar to the one carried by the *L. pneumophila* Paris strain (Fig. 5), differing by only 8% (22/276) of the substrates. In particular, PtVFX/2014 exclusively shares the Dot/Icm effector coding gene *lpp2486/PtVFX2014\_00550* with the Paris strain, which encodes an eukaryotic-like protein with an N-terminal F-box domain and a C-terminal coiled-coil motif that may play an important role in infection of *Acanthamoeba castellanii* and human macrophages<sup>69</sup>. Besides this substrate, we found just another Dot/Icm exclusively shared by PtVFX/2014 and a single strain, namely the gene *PtVFX2014\_12325* homolog of the *lpg0402/ankY/legA9* of the Philadelphia-1 strain, which encodes a protein with an ankyrin motif. Despite its functional role is still unknown, ankyrin-containing proteins are thought to be important for *L. pneumophila* infection of eukaryotic host cells<sup>70</sup>. Indeed, Anky/LegA9 was suggested to be involved in autophagy uptake of *Legionella*-containing vacuole and avoidance of lysosomal fusion<sup>71</sup>.

We cannot discard that PtVFX/2014 carries additional Dot/Icm substrates as, for instance, deep analysis of the recently sequenced 130b strain revealed the presence of 10 novel effectors (not addressed in Supplementary Table S2)<sup>27</sup>, where four of which (*lpw\_00581*, *lpw\_16311*, *lpw\_20091* and *lpw\_25791*) are also encoded in the PtVFX/2014 genome. Moreover, we detected some PtVFX/2014-exclusive genes that are adjacent to known Dot/Icm coding genes. One of those examples refers to *PtVFX2014\_10440* (coding for a hypothetical protein), which is flanked by the two *dot/icm* genes *lpg0171/legU1* and *lpg0172*. Whether these exclusive genes constitute novel *L. pneumophila* Dot/Icm substrates cannot be stated, although they may constitute potential targets for future functional studies, since *dot/icm* genes often appear in clusters spread throughout the genome. Similarly to what has been shown for other *L. pneumophila* strains, the novel outbreak-related PtVFX/2014 strain revealed a unique composition of Dot/Icm effectors that may underlie a distinct virulence profile.

Overall, in this study, we analyzed the genetic backbone of the PtVFX/2014 strain, which was associated with the first strong evidence of person-to-person transmission and with the second worldwide largest outbreak of LD so far. Comparative genomics revealed that PtVFX/2014 possesses a unique genomic make-up, marked by relevant HGT events enrolling multiple virulence factors, which endow it with specific traits that may impact its ability to adapt and persist in diverse environments, and, ultimately, to be transmitted and to cause human disease. It is our goal to apply phenotypic approaches to dissect, in near future, the transmission and virulence skills of PtVFX/2014 strain.

## Methods

**WGS for outbreak investigation.** To investigate the genetic relatedness of ST1905 isolates and to confirm the source of LD outbreak occurred in Portugal, 10 clinical isolates and one environmental isolate obtained from an industrial cooling tower were selected for WGS. Briefly, high-quality genomic DNA samples from pure bacterial cultures of both the clinical and the environmental isolates were used to prepare Nextera XT Illumina libraries that were subjected to paired-end sequencing (2 × 150 bp) on a MiSeq system (Illumina Inc.), according to the manufacturer’s instructions, using a depth of coverage >100-fold for all samples. Obtained reads were subjected to quality assessment and improvement (using FastQC and FASTX tools), and further assembled using the Velvet version 1.2.10 (<https://www.ebi.ac.uk/~zerbino/velvet/>). The *de novo* assembly process was optimized using the VelvetOptimiser script version 2.2.5. To study the genetic relatedness between each clinical isolate and the environmental isolate, a pairwise strategy was used in order to maximize the extent of core-genome to be compared. Briefly, pairwise alignments of the draft genome sequences were performed using the progressive algorithm of Mauve software (version 2.3.1) (<http://darlinglab.org/mauve/mauve.html>), and core-alignments were extracted by keeping and concatenating regions where sequences aligned over at least 500 bp. Both DnaSP v5 (<http://www.ub.edu/dnasp/>) and MEGA5 (<http://www.megasoftware.net/>) software were then used to check potential nucleotide variants, with Tablet 1.14.04.10 (<https://ics.hutton.ac.uk/tablet/>) being further applied to exclude false positive polymorphisms.



**Figure 5. Repertoire of Dot/Icm T4BSS substrates in PtVFX/2014 strain.** (A) Venn diagram [constructed using *jvenn*] showing the distribution of 303 known Dot/Icm substrates among six *L. pneumophila* strains. Given that any counting of strain-specific substrates would be an underestimation (as deep functional studies have not been systematically performed for all strains), those substrates were labeled with a question mark, except for the Philadelphia-1 strain as it was used as a query in the pairwise comparisons. Corby strain is not represented in the Venn diagram as it possesses a Dot/Icm repertoire almost identical to that of Alcoy. The only exceptions occur in the genes *lpg0080* and *lpg1965*, which are absent in Corby (*lpg0080* is also absent in Lens, whereas *lpg1965* is present in 130b and Philadelphia-1 besides 2300/99 Alcoy). Details regarding the Dot/Icm make-up of each strain are given in Supplementary Table S2. (B) goeBURST full minimum spanning tree showing that PtVFX/2014 (in red) carries a unique the Dot/Icm repertoire. Relationships based on the Dot/Icm make-up among strains were estimated using the goeBURST algorithm implemented in the PHYLOViZ software. Connecting lines are labeled with the number of Dot/Icm substrates differences between each strain. (C) Pairwise matrix displaying the number of Dot/Icm substrates that are shared (light red) or divergent (light blue) among strains. Diagonal values (in grey) represent the total number of Dot/Icm substrates inferred for each strain.

**goeBURST analysis based on allelic profiles of *L. pneumophila* STs.** The European Working Group for *Legionella* Infections (EWGLI) SBT database (<http://www.ewgli.org/>)<sup>16,17</sup> included data of 10085 isolates (with ST attributed) from 60 countries representing 2108 distinct STs, at the time this analysis was conducted (24.11.2015). To display the relationships between the allelic profile of ST1905 and the other STs described worldwide, we constructed a goeBURST full minimum spanning tree (MST)<sup>72</sup> implemented in the PHYLOViZ software<sup>73</sup>.

**Whole-genome comparative analyses within the *L. pneumophila* species.** To integrate the genetic backbone of the *L. pneumophila* PtVFX/2014 strain in the frame of the species phylogeny, 10 close chromosome sequences (including those from the most studied outbreak-associated *L. pneumophila* serogroup 1 strains), four draft genomes from *L. pneumophila* subspecies other than the *L. pneumophila* subsp. *pneumophila* available at GenBank, and sequence data (retrieved from European Nucleotide Archive) from 10 isolates selected for increasing genetic diversity (see Supplementary Table S1) were also analyzed. This selection criterion allowed us to work with the 15 highly diverse phylogenetic clusters described by Underwood and colleagues<sup>20</sup>. A core-genome alignment (enrolling a total of 35 genome sequences) followed by phylogenetic inferences were performed with Parsnp implemented on Harvest suite<sup>74</sup>, using the default parameters, with exception of parameter *-C*, which was adjusted to 2000 in order to maximize the reference coverage. Whole-genome pairwise alignments performed using the progressive algorithm of MAUVE 2.3.1 software were also analyzed through the DnaSP v5 software in order to identify genomic regions with high single nucleotide polymorphism (SNP) density. SimPlot/BootScan software (<http://sray.med.som.jhmi.edu/SCRoftware/simplot/>) was used to precisely identify breakpoint regions of putative recombination events. Multiple additional MAUVE alignments, RAST (<http://rast.nmpdr.org/>) and BLAST analyses were also performed to characterize the genome of the outbreak-associated ST1905 strain in regard to the presence of specific genomic islands, plasmids, prophages, virulence factors, and other relevant features.



**Genetic analysis of *wzt* and *wzm* loci.** In order to investigate the genetic basis determining the serogroup type of the strain PtVFX/2014 (serogroup 1), we analyzed the nucleotide sequences of two genes (*wzt* and *wzm*) belonging to the lipopolysaccharide (LPS) cluster of *L. pneumophila*<sup>39</sup>, since these loci were found to be good genetic markers for discriminating *L. pneumophila* serogroups<sup>34,35</sup>. In this regard, phylogenetic reconstructions over individual gene alignments were performed through MEGA5 by using the neighbor-joining method with bootstrapping (1000 replicates) and the Kimura-2-parameter for distance estimates.

**Nucleotide sequence accession number.** A draft genome sequence of one representative clone of the outbreak-related ST1905 strain (PtVFX/2014) was submitted to GenBank, annotated using the NCBI Prokaryotic Genomes Annotation Pipeline 2.3, and is available under the accession number LORH00000000. Raw sequence reads of each PtVFX/2014 strain were deposited in Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>) under the accession numbers: SRR3176625, SRR3176631, SRR3176714, SRR3176716, SRR3176831, SRR3176838, SRR3176884, SRR3176894, SRR3176900, SRR3176901 and SRR3176904.

## References

- Dominguez, A. *et al.* Factors influencing the case-fatality rate of Legionnaires' disease. *Int J Tuberc Lung Dis* **13**, 407–412 (2009).
- Phin, N. *et al.* Epidemiology and clinical management of Legionnaires' disease. *Lancet Infect Dis* **14**, 1011–1021 (2014).
- Yu, V. L. *et al.* Distribution of *Legionella* species and serogroups isolated by culture in patients with sporadic community-acquired legionellosis: an international collaborative survey. *J Infect Dis* **186**, 127–128 (2002).
- Centers for Disease Control and Prevention. Legionellosis—United States, 2000–2009. *MMWR Morb Mortal Wkly Rep* **60**, 1083–1086 (2011).
- European Centre for Disease Prevention and Control. Legionnaires' disease in Europe, 2012. (ECDC, Stockholm, 2014).
- Fields, B. S., Benson, R. F. & Besser, R. E. *Legionella* and Legionnaires' disease: 25 years of investigation. *Clin Microbiol Rev* **15**, 506–526 (2002).
- Muder, R. R., Yu, V. L. & Woo, A. H. Mode of transmission of *Legionella pneumophila*. A critical review. *Arch Intern Med* **146**, 1607–1612 (1986).
- Correia, A. M. *et al.* Probable Person-to-Person Transmission of Legionnaires' Disease. *N Engl J Med* **374**, 497–498 (2016).
- Garcia-Fulgueiras, A. *et al.* Legionnaires' disease outbreak in Murcia, Spain. *Emerg Infect Dis* **9**, 915–921 (2003).
- Levesque, S. *et al.* Genomic characterization of a large outbreak of *Legionella pneumophila* serogroup 1 strains in Quebec City, 2012. *PLoS One* **9**, e103852 (2014).
- McAdam, P. R. *et al.* Gene flow in environmental *Legionella pneumophila* leads to genetic and pathogenic heterogeneity within a Legionnaires' disease outbreak. *Genome Biol* **15**, 504 (2014).
- Sanchez-Buso, L., Comas, I., Jorques, G. & Gonzalez-Candelas, F. Recombination drives genome evolution in outbreak-related *Legionella pneumophila* isolates. *Nat Genet* **46**, 1205–1211 (2014).
- Shivaji, T. *et al.* A large community outbreak of Legionnaires disease in Vila Franca de Xira, Portugal, October to November 2014. *Euro Surveill* **19**, 20991 (2014).
- Fraser, D. W. *et al.* Legionnaires' disease: description of an epidemic of pneumonia. *N Engl J Med* **297**, 1189–1197 (1977).
- van Heijnsbergen, E. *et al.* Confirmed and Potential Sources of *Legionella* Reviewed. *Environ Sci Technol* **49**, 4797–4815 (2015).
- Gaia, V. *et al.* Consensus sequence-based scheme for epidemiological typing of clinical and environmental isolates of *Legionella pneumophila*. *J Clin Microbiol* **43**, 2047–2052 (2005).
- Ratzow, S., Gaia, V., Helbig, J. H., Fry, N. K. & Luck, P. C. Addition of *neuA*, the gene encoding N-acylneuraminate cytidylyl transferase, increases the discriminatory ability of the consensus sequence-based scheme for typing *Legionella pneumophila* serogroup 1 strains. *J Clin Microbiol* **45**, 1965–1968 (2007).
- Moran-Gilad, J. *et al.* Design and application of a core genome multilocus sequence typing scheme for investigation of Legionnaires' disease incidents. *Euro Surveill* **20**, 21186 (2015).
- Reuter, S. *et al.* A pilot study of rapid whole-genome sequencing for the investigation of a *Legionella* outbreak. *BMJ Open* **3**, e002175 (2013).
- Underwood, A. P., Jones, G., Mentasti, M., Fry, N. K. & Harrison, T. G. Comparison of the *Legionella pneumophila* population structure as determined by sequence-based typing and whole genome sequencing. *BMC Microbiol* **13**, 302 (2013).
- Graham, R. M., Doyle, C. J. & Jennison, A. V. Real-time investigation of a *Legionella pneumophila* outbreak using whole genome sequencing. *Epidemiol Infect* **142**, 2347–2351 (2014).
- Pinto, C. S. A large outbreak of Legionnaires' Disease in Vila Franca de Xira, Portugal. Paper presented at Third Annual Scientific Conference of the the European Working Group for Legionella Infections (ESCMID) Study Group for *Legionella* infections (ESGLI) (ESGLI 2015), ESCMID. London. 2015, 16–17 September.
- Gomez-Valero, L. *et al.* Extensive recombination events and horizontal gene transfer shaped the *Legionella pneumophila* genomes. *BMC Genomics* **12**, 536 (2011).
- D'Auria, G., Jimenez-Hernandez, N., Peris-Bondia, F., Moya, A. & Latorre, A. *Legionella pneumophila* pangenome reveals strain-specific virulence factors. *BMC Genomics* **11**, 181 (2010).
- Coscolla, M., Comas, I. & Gonzalez-Candelas, F. Quantifying nonvertical inheritance in the evolution of *Legionella pneumophila*. *Mol Biol Evol* **28**, 985–1001 (2011).
- Cazalet, C. *et al.* Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat Genet* **36**, 1165–1173 (2004).
- Schroeder, G. N. *et al.* *Legionella pneumophila* strain 130b possesses a unique combination of type IV secretion systems and novel Dot/Icm secretion system effector proteins. *J Bacteriol* **192**, 6001–6016 (2010).
- Ensminger, A. W. & Isberg, R. R. *Legionella pneumophila* Dot/Icm translocated substrates: a sum of parts. *Curr Opin Microbiol* **12**, 67–73 (2009).
- Luo, Z. Q. & Isberg, R. R. Multiple substrates of the *Legionella pneumophila* Dot/Icm system identified by interbacterial protein transfer. *Proc Natl Acad Sci USA* **101**, 841–846 (2004).
- Glockner, G. *et al.* Identification and characterization of a new conjugation/type IVA secretion system (*trb/tra*) of *Legionella pneumophila* Corby localized on two mobile genomic islands. *Int J Med Microbiol* **298**, 411–428 (2008).
- Steinert, M., Heuner, K., Buchrieser, C., Albert-Weissenberger, C. & Glockner, G. *Legionella* pathogenicity: genome structure, regulatory networks and the host cell response. *Int J Med Microbiol* **297**, 577–587 (2007).
- Jules, M. & Buchrieser, C. *Legionella pneumophila* adaptation to intracellular life and the host response: clues from genomics and transcriptomics. *FEBS Lett* **581**, 2829–2838 (2007).
- Hilbi, H., Hoffmann, C. & Harrison, C. F. *Legionella* spp. outdoors: colonization, communication and persistence. *Environ Microbiol Rep* **3**, 286–296 (2011).
- Cao, B., Yao, F., Liu, X., Feng, L. & Wang, L. Development of a DNA microarray method for detection and identification of all 15 distinct O-antigen forms of *Legionella pneumophila*. *Appl Environ Microbiol* **79**, 6647–6654 (2013).

35. Cazalet, C. *et al.* Multigenome analysis identifies a worldwide distributed epidemic *Legionella pneumophila* clone that emerged within a highly diverse species. *Genome Res* **18**, 431–441 (2008).
36. Chien, M. *et al.* The genomic sequence of the accidental pathogen *Legionella pneumophila*. *Science* **305**, 1966–1968 (2004).
37. Viswanathan, V. K., Edelstein, P. H., Pope, C. D. & Cianciotto, N. P. The *Legionella pneumophila* *iraAB* locus is required for iron assimilation, intracellular infection, and virulence. *Infect Immun* **68**, 1069–1079 (2000).
38. Vincent, C. D. *et al.* Identification of non-dot/icm suppressors of the *Legionella pneumophila* DeltadotL lethality phenotype. *J Bacteriol* **188**, 8231–8243 (2006).
39. Petzold, M. *et al.* A structural comparison of lipopolysaccharide biosynthesis loci of *Legionella pneumophila* serogroup 1 strains. *BMC Microbiol* **13**, 198 (2013).
40. Wintermeyer, E. *et al.* Influence of site specifically altered Mip proteins on intracellular survival of *Legionella pneumophila* in eukaryotic cells. *Infect Immun* **63**, 4576–4583 (1995).
41. Shevchuk, O., Jager, J. & Steinert, M. Virulence properties of the *Legionella pneumophila* cell envelope. *Front Microbiol* **2**, 74 (2011).
42. Charpentier, X., Faucher, S. P., Kalachikov, S. & Shuman, H. A. Loss of RNase R induces competence development in *Legionella pneumophila*. *J Bacteriol* **190**, 8126–8136 (2008).
43. Cornelis, G. R. The type III secretion injectisome. *Nat Rev Microbiol* **4**, 811–825 (2006).
44. Isberg, R. R., O'Connor, T. J. & Heidtman, M. The *Legionella pneumophila* replication vacuole: making a cosy niche inside host cells. *Nat Rev Microbiol* **7**, 13–24 (2009).
45. Ninio, S. & Roy, C. R. Effector proteins translocated by *Legionella pneumophila*: strength in numbers. *Trends Microbiol* **15**, 372–380 (2007).
46. Nevo, O., Zusman, T., Rasis, M., Lifshitz, Z. & Segal, G. Identification of *Legionella pneumophila* effectors regulated by the LetAS-RsmYZ-CsrA regulatory cascade, many of which modulate vesicular trafficking. *J Bacteriol* **196**, 681–692 (2014).
47. Zhu, W. *et al.* Comprehensive identification of protein substrates of the Dot/Icm type IV transporter of *Legionella pneumophila*. *PLoS One* **6**, e17638 (2011).
48. Weber, S. S. *et al.* Identification of protective B cell antigens of *Legionella pneumophila*. *J Immunol* **189**, 841–849 (2012).
49. Gomez-Valero, L. *et al.* Comparative analyses of *Legionella* species identifies genetic features of strains causing Legionnaires' disease. *Genome Biol* **15**, 505 (2014).
50. Brzuszkiewicz, E. *et al.* *Legionella oakridgensis* ATCC 33761 genome sequence and phenotypic characterization reveals its replication capacity in amoebae. *Int J Med Microbiol* **303**, 514–528 (2013).
51. Segal, G., Russo, J. J. & Shuman, H. A. Relationships between a new type IV secretion system and the icm/dot virulence system of *Legionella pneumophila*. *Mol Microbiol* **34**, 799–809 (1999).
52. Bandyopadhyay, P., Liu, S., Gabbai, C. B., Venitelli, Z. & Steinman, H. M. Environmental mimics and the Lvh type IVA secretion system contribute to virulence-related phenotypes of *Legionella pneumophila*. *Infect Immun* **75**, 723–735 (2007).
53. Marra, A. & Shuman, H. A. Isolation of a *Legionella pneumophila* restriction mutant with increased ability to act as a recipient in heterospecific matings. *J Bacteriol* **171**, 2238–2240 (1989).
54. Horvath, P. & Barrangou, R. CRISPR/Cas, the immune system of bacteria and archaea. *Science* **327**, 167–170 (2010).
55. Faucher, S. P. & Shuman, H. A. Small Regulatory RNA and *Legionella pneumophila*. *Front Microbiol* **2**, 98 (2011).
56. Gunderson, F. F. & Cianciotto, N. P. The CRISPR-associated gene *cas2* of *Legionella pneumophila* is required for intracellular infection of amoebae. *MBio* **4**, e00074–00013 (2013).
57. Flynn, K. J. & Swanson, M. S. Integrative conjugative element ICE-betaox confers oxidative stress resistance to *Legionella pneumophila* *in vitro* and in macrophages. *MBio* **5**, e01091–01014 (2014).
58. Brassinga, A. K. *et al.* A 65-kilobase pathogenicity island is unique to Philadelphia-1 strains of *Legionella pneumophila*. *J Bacteriol* **185**, 4630–4637 (2003).
59. Long, F., Rouquette-Loughlin, C., Shafer, W. M. & Yu, E. W. Functional cloning and characterization of the multidrug efflux pumps NorM from *Neisseria gonorrhoeae* and YdhE from *Escherichia coli*. *Antimicrob Agents Chemother* **52**, 3052–3060 (2008).
60. Miyamoto, H. *et al.* Protein profiles of *Legionella pneumophila* Philadelphia-1 grown in macrophages and characterization of a gene encoding a novel 24 kDa *Legionella* protein. *Microb Pathog* **15**, 469–484 (1993).
61. Abbott, Z. D., Yakhnin, H., Babitzke, P. & Swanson, M. S. *csrR*, a paralog and direct target of CsrA, promotes *Legionella pneumophila* resilience in water. *MBio* **6**, e00595 (2015).
62. Forsbach-Birk, V., McNealy, T., Shi, C., Lynch, D. & Marre, R. Reduced expression of the global regulator protein CsrA in *Legionella pneumophila* affects virulence-associated regulators and growth in *Acanthamoeba castellanii*. *Int J Med Microbiol* **294**, 15–25 (2004).
63. Molofsky, A. B. & Swanson, M. S. *Legionella pneumophila* CsrA is a pivotal repressor of transmission traits and activator of replication. *Mol Microbiol* **50**, 445–461 (2003).
64. Gomez-Valero, L., Rusniok, C., Cazalet, C. & Buchrieser, C. Comparative and functional genomics of *Legionella* identified eukaryotic like proteins as key players in host-pathogen interactions. *Front Microbiol* **2**, 208 (2011).
65. Lifshitz, Z. *et al.* Computational modeling and experimental validation of the *Legionella* and *Coxiella* virulence-related type-IVB secretion signal. *Proc Natl Acad Sci USA* **110**, E707–715 (2013).
66. de Felipe, K. S. *et al.* *Legionella* eukaryotic-like type IV substrates interfere with organelle trafficking. *PLoS Pathog* **4**, e1000117 (2008).
67. Ge, J. & Shao, F. Manipulation of host vesicular trafficking and innate immune defence by *Legionella* Dot/Icm effectors. *Cell Microbiol* **13**, 1870–1880 (2011).
68. Isaac, D. T. & Isberg, R. Master manipulators: an update on *Legionella pneumophila* Icm/Dot translocated substrates and their host targets. *Future Microbiol* **9**, 343–359 (2014).
69. Lomma, M. *et al.* The *Legionella pneumophila* F-box protein Lpp2082 (AnkB) modulates ubiquitination of the host protein parvin B and promotes intracellular replication. *Cell Microbiol* **12**, 1272–1291 (2010).
70. Pan, X., Luhrmann, A., Satoh, A., Laskowski-Arce, M. A. & Roy, C. R. Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320**, 1651–1654 (2008).
71. Khweek, A. A. *et al.* A bacterial protein promotes the recognition of the *Legionella pneumophila* vacuole by autophagy. *Eur J Immunol* **43**, 1333–1344 (2013).
72. Francisco, A. P., Bugalho, M., Ramirez, M. & Carrico, J. A. Global optimal eBURST analysis of multilocus typing data using a graphic matrix approach. *BMC Bioinformatics* **10**, 152 (2009).
73. Francisco, A. P. *et al.* PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. *BMC Bioinformatics* **13**, 87 (2012).
74. Treangen, T. J., Ondov, B. D., Koren, S. & Phillippy, A. M. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* **15**, 524 (2014).

## Author Contributions

V.B., A.N. and J.P.G. conceived the study design and wrote the manuscript, V.B., A.N., D.A.S., L.V., J.M., M.J.S. and P.G. performed the experiments. All authors discussed the results and reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <http://www.nature.com/srep>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Borges, V. *et al.* *Legionella pneumophila* strain associated with the first evidence of person-to-person transmission of Legionnaires' disease: a unique mosaic genetic backbone. *Sci. Rep.* **6**, 26261; doi: 10.1038/srep26261 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>