

RESEARCH ARTICLE

Open Access

The genome sequence of *Brucella pinnipedialis* B2/94 sheds light on the evolutionary history of the genus *Brucella*

Stéphane Audic^{1,2,3*}, Magali Lescot¹, Jean-Michel Claverie¹, Axel Cloeckaert⁴ and Michel S Zygmunt^{4*}

Abstract

Background: Since the discovery of the Malta fever agent, *Brucella melitensis*, in the 19th century, six terrestrial mammal-associated *Brucella* species were recognized over the next century. More recently the number of novel *Brucella* species has increased and among them, isolation of species *B. pinnipedialis* and *B. ceti* from marine mammals raised many questions about their origin as well as on the evolutionary history of the whole genus.

Results: We report here on the first complete genome sequence of a *Brucella* strain isolated from marine mammals, *Brucella pinnipedialis* strain B2/94. A whole gene-based phylogenetic analysis shows that five main groups of host-associated *Brucella* species rapidly diverged from a likely free-living ancestor close to the recently isolated *B. microti*. However, this tree lacks the resolution required to resolve the order of divergence of those groups. Comparative analyses focusing on a) genome segments unshared between *B. microti* and *B. pinnipedialis*, b) gene deletion/fusion events and c) positions and numbers of *Brucella* specific IS711 elements in the available *Brucella* genomes provided enough information to propose a branching order for those five groups.

Conclusions: In this study, it appears that the closest relatives of marine mammal *Brucella* sp. are *B. ovis* and *Brucella* sp. NVSL 07-0026 isolated from a baboon, followed by *B. melitensis* and *B. abortus* strains, and finally the group consisting of *B. suis* strains, including *B. canis* and the group consisting of the single *B. neotomae* species. We were not able, however, to resolve the order of divergence of the two latter groups.

Keywords: Brucella, Bacterial Genome Evolution, Comparative Genomics

Background

Brucellae are Gram-negative, facultative, intracellular bacteria that can infect many species of animals and man. Six species were initially recognized within the genus *Brucella*: *B. abortus*, *B. melitensis*, *B. suis*, *B. ovis*, *B. canis*, and *B. neotomae* [1-3]. This classification is mainly based on differences in pathogenicity, host preference, and phenotypic characteristics. Four additional species have been included in the genus *Brucella* since 2007. These comprise the species *B. ceti* and *B. pinnipedialis*

isolated from marine mammals, with cetaceans (dolphin, porpoise, and whale species) and pinnipeds (various seal species) as preferred host respectively [4,5]. *B. microti* described in 2008 was first isolated from the common vole and then from the red fox, and from soil [6-8]. The latest described species is *B. inopinata*, isolated from an infected human breast implant, and currently the most divergent *Brucella* species at the phenotypic and molecular level [9,10]. The animal or environmental reservoir of *B. inopinata* is not known. New *Brucella* species will likely be described in the future such as for isolates from baboons [11], from wild rodents in Australia [12] and for strain BO2 isolated from a patient with chronic destructive pneumonia [13]. Strain BO2 and strains from wild Australian rodents have been proposed as a novel lineage of the *B. inopinata* species [12,13].

* Correspondence: stephane.audic@sb-roscoff.fr; Michel.Zygmunt@tours.inra.fr

¹Laboratoire Information Génomique et Structurale, CNRS-UPR2589, Aix-Marseille University, Institut de Microbiologie de la Méditerranée (IM2, IFR-88), Parc Scientifique de Luminy-163 Avenue de Luminy-Case 934-FR-13288, Marseille cedex 09, France

⁴INRA, UR1282, Infectiologie Animale et Santé Publique, IASP, Nouzilly, F-37380, France

Full list of author information is available at the end of the article

Molecular and phenotypic typing of marine mammal *Brucella* strains led to their classification into two species, *B. ceti* and *B. pinnipedialis*, according to their preferred host, cetaceans and pinnipeds respectively [5]. However, several subgroups were identified within each species by molecular typing methods such as multilocus sequence analysis (MLSA), multilocus VNTR (Variable Number of Tandem Repeats) analysis (MLVA), or *omp2a* and *omp2b* porin genes [14-19]. Among them one subgroup within *B. ceti*, exclusively composed of strains isolated from various dolphin species, was proposed to constitute a separate species with the name *B. delphini* [3,14,18]. The isolates from cetaceans from the Pacific may also constitute a separate species [19]. Three human cases with naturally acquired infection by *Brucella* strains presumably from marine origin were reported, one case of spinal osteomyelitis from a patient in New Zealand [20] and two neurobrucellosis cases from Peruvian patients [21]. Interestingly, these human isolates exhibited the same genotype as strains from cetaceans from the Pacific [22].

Among their distinctive characteristics at the molecular and genomic level, marine mammal *Brucella* strains were shown to carry in their genomes a higher number of the insertion sequence element IS711 (or IS6501) [23,24] than terrestrial mammal *Brucella* species and biovars with the exception of *B. ovis* [15,17,25]. Consequently, infrequent restriction site-PCR (IRS-PCR) methods and more recently ligation-mediated PCR (LM-PCR) were applied, taking into account this higher number of IS711 elements, to study the genomic diversity of marine mammal strains [26-28]. These studies confirmed the classification into two marine mammal *Brucella* species, each divided in subgroups. In addition, six specific IS711-containing DNA fragments were detected allowing the molecular identification of *B. ceti* and its subgroup composed exclusively of dolphin isolates [17,26-28]. Besides these specific IS711-containing fragments another DNA fragment was detected that was exclusively found in *B. pinnipedialis* strains, with the exception of hooded seal isolates, consisting of a putative genomic island [26,27]. The size of this island was estimated at 62 kbp according to the physical maps made from the genomes of marine mammal strains by macrorestriction analyses [14].

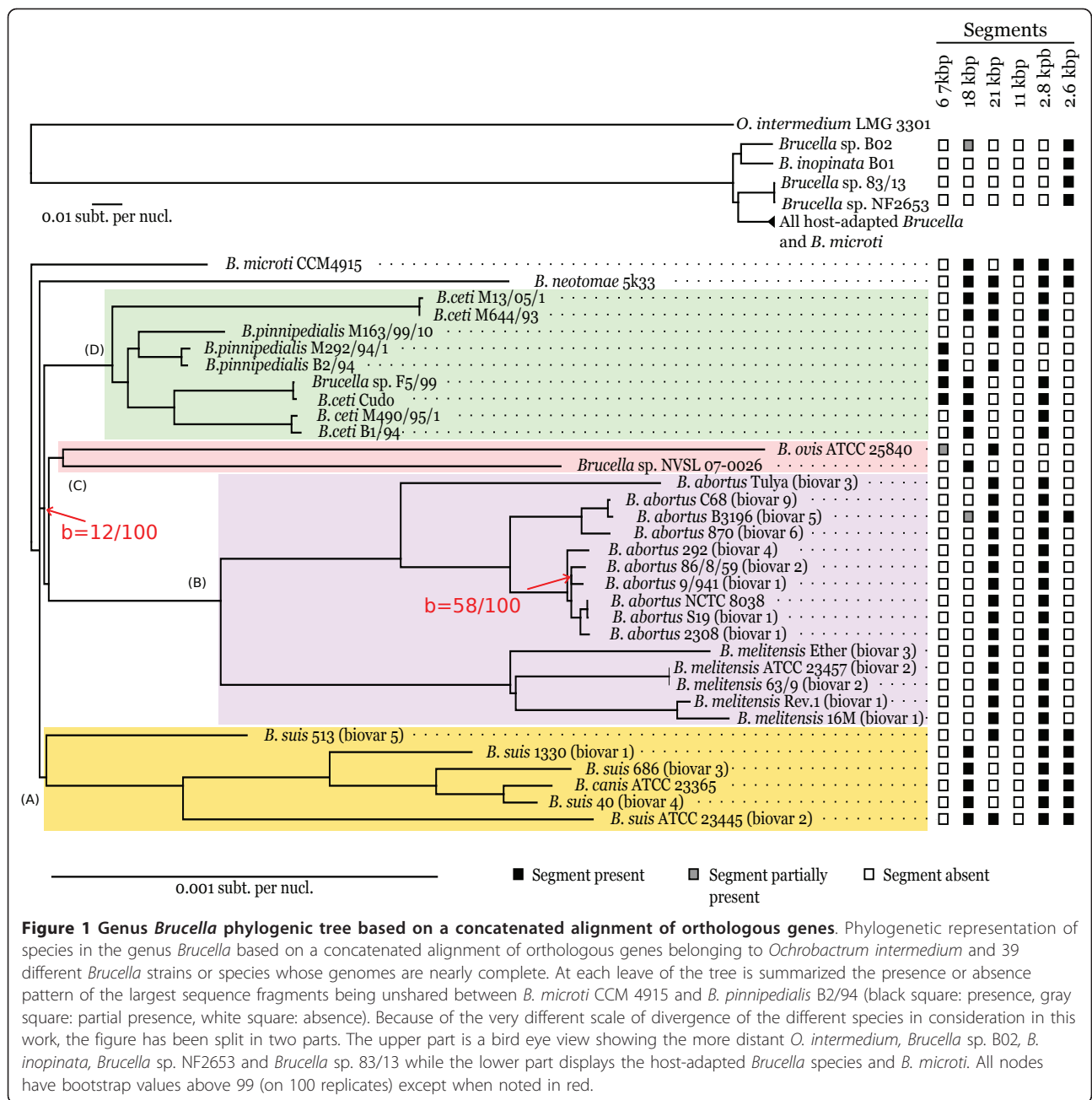
The taxonomy of *Brucella* is still controversial, with an ongoing debate on whether they should be considered as distinct species or distinct strains of *B. melitensis*, considering the close proximity of their genomes [29]. We determined and analyzed the complete genome of *B. pinnipedialis* B2/94 to bring new insights into the origin of *Brucella* isolated from marine mammals as well as their time of divergence from *Brucella* isolated from terrestrial animals.

Results and Discussion

The genome sequence of *B. pinnipedialis* B2/94 was determined (30× coverage) by shotgun sequencing with the GS-FLX technology and the remaining gaps filled using the standard Sanger technology. Like that of other *Brucella* strains, the genome is composed of two circular chromosomes, of 2,138,342 bp (base-pairs) and 1,260,926 bp in lengths, respectively. Bioinformatic annotation predicted the presence of 3,342 protein coding genes, 55 tRNAs and 9 ribosomal RNAs. The comparison with the known genomes of other *Brucella* species revealed the presence of 90 pseudogenes. The 23S rDNA sequence of *B. pinnipedialis* B2/94 was found similar to that of other *Brucella* species, in contrast with the anomalous and unexpected 23S ribosomal RNA sequence previously described for *B. microti* [30].

Genome structure and whole gene set phylogeny

The phylogenetic placement of *B. pinnipedialis* B2/94 was assessed with the help of a whole gene tree including all orthologous genes from *Brucella* strains available in the complete genome division of GenBank, completed by sequence data from selected *Brucella* strains of particular phylogenetic relevance, either deposited in the whole genome shotgun (WGS) division of GenBank, or available from other sources (see Methods). This includes sequences from *B. neotomae*, 10 sequences from marine mammal *Brucella* strains, *Brucella* sp. B02, *B. inopinata* BO1, *Brucella* sp. 83/13 or *Brucella* sp. NF2653. The resulting tree (Figure 1) readily shows that the *Brucella* species with the highest sequence similarity with *B. pinnipedialis* B2/94 and whose complete genome is available is *B. microti* CCM 4915. This is a consequence of the relatively slow evolution rate of those two species rather than reflecting a particular phylogenetic affinity. The central position occupied by *B. microti*, just at the root of host-associated *Brucella* species, is also confirmed. This tree distinguishes five major groups, the *B. suis* strains (shaded rectangle A, Figure 1), the different *B. melitensis* and *B. abortus* strains, also well separated from each other (rectangle B), another group with *B. ovis* and *Brucella* sp. NVSL 07-0026 (rectangle C), and marine mammal *Brucella* strains (rectangle D). *B. neotomae* remains isolated in this tree and constitutes the fifth group. However, the resolution of this tree is insufficient to infer the relative order of divergence of those different groups. Moreover, while bootstrap values in this tree are in general very strong (99 or 100%), the branching (D,(C, B)) proposed in the maximum likelihood tree is generally not supported, with a bootstrap value of 12%, prompting for a finer analysis of the events shaping the *Brucella* evolutionary tree.



The genome sequences of *B. microti* CCM 4915 and *B. pinnipedialis* B2/94 were found to be remarkably conserved even at the nucleotide level, allowing to generate a complete alignment of the chromosomes of the two species from which a list of all indels (insertions and deletions) (available as Additional file 1, Table S1 and Additional file 2, Table S2) was easily obtained. In spite of the overall similarity of these two genomes, this alignment revealed major changes in genome structure, as large segments unshared between the two species. The alignment from the largest chromosome exhibited 238

gapped positions with the largest insert in *B. microti* being 2,653 bp long, and the largest in *B. pinnipedialis* being 21,713 bp long. The alignment from the small chromosome exhibits 151 gapped positions, the largest insert being 18,341 bp long in *B. microti* and 67,389 bp long in *B. pinnipedialis*. The total number of indels (389) found in comparing these two species is thus smaller than the 405 indels found in comparing *B. microti* CCM 4915 and *B. suis* 1330 [30]. The complete alignment contains a total of 3,290,621 aligned positions (2,104,923 and 1,185,698 per chromosome), with 0.10% of nucleotide

changes in aligned regions (2,195 and 1,200 nucleotide changes, respectively). This fraction of nucleotide substitution is also smaller than what was observed between *B. microti* CCM 4915 and *B. suis* 1330, where a 0.16% divergence was reported [30].

Detailed analysis of the largest unshared sequence regions and the branching order of host associated *Brucella* species

The evolutionary history of the largest segments unshared between *B. microti* CCM 4915 and *B. pinnipedialis* B2/94 was analyzed by examining the structure of the orthologous loci in the other *Brucella* species for which sequence data was available (see Methods). On the large chromosome, the 2 largest indels are 21 kbp, and 2.6 kbp in length. On the small chromosome, the largest indels are 67 kbp, 18 kbp, 11 kbp and 2.8 kbp in length. All other indels are at most the size of an IS711 insertion sequence (843 bp). The presence/absence of the above genomic inserts was assessed in other *Brucella* strains (see Methods). For each segment, we recorded the number of nucleotides with homologues in the other *Brucella* genomes (Additional file 3, Table S3). The presence/absence of those segments is reported at the leaves of the tree (Figure 1) as filled squares. The evolutionary history of those unshared genome segments, treated as discrete characters, was reconstructed by parsimony analysis using the Mesquite software (see Methods), and represented as a cladogram in Additional file 4, Figure S1. The proposed evolutionary scenario corresponding to those unshared segments is discussed below, the largest ones first.

The locus of BMI_II545-6 in *B. microti* CCM 4915 is occupied in the *B. pinnipedialis* B2/94 small chromosome by a 67,389 bp insert (position: 520,136 to 587,524) which is present in a number of *Brucella* strains from marine mammals and partially in *B. ovis* ATCC 25840. The gene arrangements of this genomic region in *B. pinnipedialis*, *B. microti* and *B. ovis* are depicted in Figure 2. This region is flanked on both sides by an IS711 element (respectively BPI_II536-7 and BPI_II602-3), at a location where a sole copy of the IS711 element is found in *B. microti* CCM 4915 (BMI_II545-6), as in all *Brucella* strains lacking this insert. This region is partially found in *B. ovis* ATCC 25840, except in the interval between BPI_II545 and BPI_II576 deleted from *B. ovis*. The deletion in *B. ovis* ATCC 25840 occurred inside the gene BOV_A0488, whose 85 first amino acids are similar to BPI_II577 and whose last 100 amino acids are similar to BPI_II544. This interval exhibits a few protein coding genes or their remnants, with similarities mainly found in distant alphaproteobacteria i.e. excluding *Brucella* and *Ochrobactrum*. In particular, two pseudogenes, BPI_II545 and BPI_II548, separated by one IS711 element, appear to be the remnants of two consecutive *Paracoccus*

denitrificans genes coding for 5-oxoprolinase (Pden_4408-9) and are preceded by an AraC-family transcriptional regulator (BPI_II544) with similarity into the same organism (Pden_4410). The same three genes are also found grouped together but in different orders in the genomes of several betaproteobacteria e.g. *Bordetella avium* (BAV3293-5) or *Verminephrobacter eiseniae* (Veis_1518-20). Two consecutive genes, *etfA* (BPI_II555) and *etfB* (BPI_II556) are also present on the large chromosome with 100% nucleotide identity. We also observed 4 components of a spermidine/putrescine ABC-transporter (BPI_II564-7) found colinearly in the genomes of e.g. *Roseobacter denitrificans* (RD1_3862-5) or *Silicibacter pomeroyi* (SPOA381-4). We found similarities to two consecutive genes of *Parvibaculum lavamentivorans* (Plav_0386-7), similarities to Plav_0387 being noticed on both sides of the BPI_II574-5 IS*Bm1* transposase gene. Other genes like the aldehyde dehydrogenase BPI_II571 have best matches in *Brucella* and *Ochrobactrum*, but with 73% similarity only at the amino acid level. Several transposase genes are also present. This large 67 kbp fragment probably found its origin in the common ancestor of the marine mammal *Brucella* species and *B. ovis*, as attested by its partial presence in *B. ovis*. The structure of this segment suggests a grouping of *B. ovis* with marine mammal *Brucella* sp. (Additional file 4, Figure S1-a). It may have been lost thereafter several times, in *Brucella* sp. strain NVSL 07-006 isolated from a baboon and along the *B. ceti* and *B. pinnipedialis* lineages, putatively by recombination of the two IS711 copies.

A 21,713 bp fragment on the large chromosome (position: 259,190 to 280,902 in *B. pinnipedialis* B2/94) is not found in *B. microti* CCM 4915. This fragment is present in all *B. melitensis* and *B. abortus* strains and in *B. neotomae*, but its occurrence is quite variable among *B. ceti* and *B. pinnipedialis* strains and even among *B. suis* strains, where it is observed in the earliest diverging *B. suis* 513 and *B. suis* ATCC 23445, but absent otherwise. This fragment starts with a phage integrase gene (BPI_I248) and ends with a tRNA (BPI_I278) which was the likely insertion site. An IS711 element is inserted within a gene (BPI_II256) that remained intact in *B. melitensis* 16 M (BMEI1694) and *B. ovis* ATCC 25840 (BOV_0245). This fragment encodes a flagellar protein FlgJ (BPI_I260). The other putative genes in this region have no convincing similarities to annotated proteins. This 21 kbp fragment (Additional file 4, Figure S1-b) probably entered the *Brucella* genomes after *B. microti* divergence, and disappeared separately on several branches. It confirms the grouping of *B. ceti* Cudo, *B. ceti* B1/94, *B. ceti* M490/95/1 and *Brucella* sp. F5/99. Its presence in the genome of *B. neotomae* 5K33 and absence in *B. suis* strains except for *B. suis* ATCC 23445 and *B. suis* 513 suggest its insertion prior to the

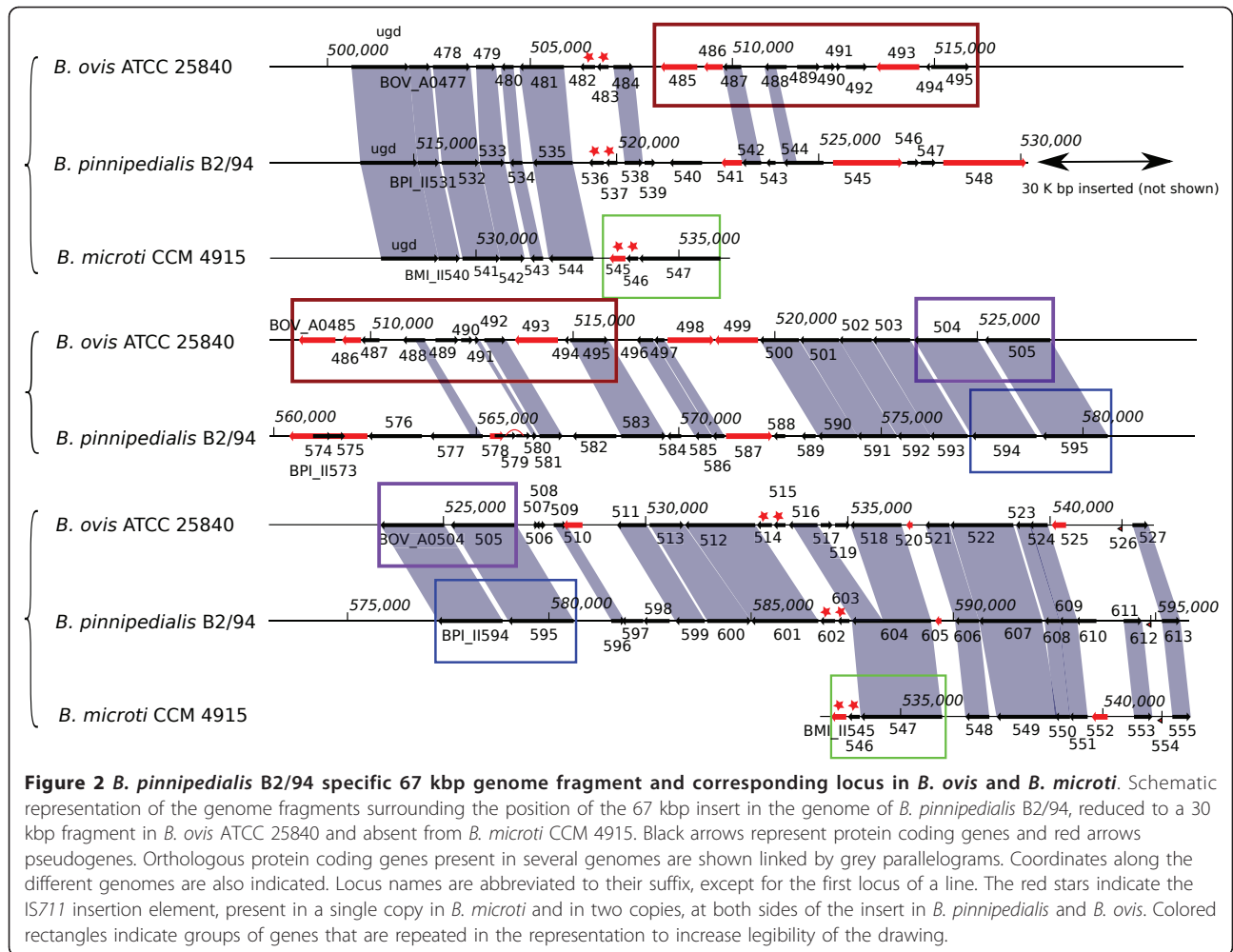


Figure 2 *B. pinnipedialis* B2/94 specific 67 kbp genome fragment and corresponding locus in *B. ovnis* and *B. microti*. Schematic representation of the genome fragments surrounding the position of the 67 kbp insert in the genome of *B. pinnipedialis* B2/94, reduced to a 30 kbp fragment in *B. ovnis* ATCC 25840 and absent from *B. microti* CCM 4915. Black arrows represent protein coding genes and red arrows pseudogenes. Orthologous protein coding genes present in several genomes are shown linked by grey parallelograms. Coordinates along the different genomes are also indicated. Locus names are abbreviated to their suffix, except for the first locus of a line. The red stars indicate the IS711 insertion element, present in a single copy in *B. microti* and in two copies, at both sides of the insert in *B. pinnipedialis* and *B. ovnis*. Colored rectangles indicate groups of genes that are repeated in the representation to increase legibility of the drawing.

divergence of the *B. suis* and *B. neotomae* lineage, and a subsequent loss in the *B. suis* lineage.

A 18,341 bp region on the small chromosome (between BMI_II357 and BMI_II381) is absent from the *B. pinnipedialis* B2/94 genome (position: 344,671-363,011 in *B. microti* CCM 4915). This region is present in *B. microti*, *B. suis*, *B. canis*, *B. ceti*, *Brucella* sp. F5/99, *B. neotomae* and absent in *B. abortus*, *B. melitensis*, *B. ovnis* and *B. pinnipedialis*. Additionally, a partial match was found in the genome of *Brucella* sp. B02. In this region, the presence of genes encoding TraI-J proteins involved in bacterial conjugation can be noted. Like the above 67 kbp fragment, this 18 kbp fragment (Additional file 4, Figure S1-c) supports a divergence of the *B. melitensis*/*B. abortus* clade prior to the separation of *B. ovnis*, *Brucella* sp. strain NVSL 07-006 and marine mammal *Brucella* strains. It might have appeared before *B. microti* divergence, and disappeared several times, in particular from the branch leading to the *B. abortus*/*B. melitensis* clade and from the branch leading to the *B. pinnipedialis* clade.

The 11 kbp region on the small chromosome is a phage related region discussed in [30] and unique to *B. microti* (position: 1,038,883-1,050,624 in *B. microti* CCM 4915). It will not be discussed further (Additional file 4, Figure S1-d).

Still on the small chromosome, a 2,881 bp region (position: 1,082,391-1,085,271 in *B. microti* CCM 4915) encodes genes BMI_II1086-8. This fragment (Additional file 4, Figure S1-e) appeared before *B. microti* divergence, and is absent from the branch leading to *B. ovnis* and *Brucella* sp. strain NVSL 07-0026, but also from the branch leading to *B. pinnipedialis* M292/94/1 and *B. pinnipedialis* B2/94, separating those two species from *B. pinnipedialis* M163/99/10. Closer examination reveals that this 2.8 kbp fragment is absent in *B. ovnis* because it belongs to a much larger 44.5 kbp region deleted from *B. ovnis* [31] and also partially deleted from *Brucella* sp. NVSL 07-0026, where an approximately 30 kbp long region is missing (pos: 1,117,180 to 1,143,394 in *B. pinnipedialis* B2/94 small chromosome). This finding supports both grouping of the

B. pinnipedialis M292/94/1 and B2/94 strains, and that of *B. ovis* with *Brucella* sp. strain NVSL 07-0026.

Between BMI_I949 and BMI_I953, there is a 2,653 bp region deleted from the *B. pinnipedialis* B2/94 large chromosome (position: 928,716-931,368 in *B. microti* CCM 4915) but also from *B. ovis* ATCC 25840, *Brucella* sp. NVSL 07-0026 and all *B. melitensis* and *B. abortus* strains. Interestingly, this deletion occurred inside a gene (encoding an ABC transporter), thus showing that it is a deletion and not an acquisition event. This region (Additional file 4, Figure S1-f) is particularly informative because it clearly separates *Brucella* strains into two groups. This fragment was present in ancestral *Brucella*, and then lost after the divergence of the *B. suis* clade, of *B. neotomae*, of *B. microti* and prior to the divergence of *B. abortus* and *B. melitensis*, *B. ovis* and *Brucella* sp. strain NVSL 07-2026, and finally marine mammal *Brucella* species. An interesting exception is found in *B. abortus* B3196, where this ABC transporter gene is intact. This feature clearly suggests a divergence of the *B. suis* and *B. neotomae* group before that of the other host-associated *Brucella*.

IS711 insertion sequences

IS711 insertion sequences [32] account for a large number of gaps within the complete genome alignment of *B. pinnipedialis* B2/94 and *B. microti* CCM 4915. In total, 19 IS711 elements were identified on the large chromosome and 12 on the small chromosome of *B. pinnipedialis*. We looked for the occurrence of those insertion sequences in the available complete genomes of *Brucella* (Methods) and reported their position in Figure 3 and Additional file 5, Table S4. Although it is difficult to analyze transposable repetitive elements in unfinished genomes (as they are often the ones precluding genome assembly) [33], we included two other partial genomes in our analysis, but where an abundant number of IS711 elements were observed, *B. ceti* Cudo and *Brucella* sp. strain NVSL 07-0026.

Figure 3 shows that 6 IS711 insertion sequences are found in all selected *Brucella* strains (IS711 groups numbered 1, 4, 12, 13, 26, and 31). We also observed that 4 IS711 elements were uniquely shared by *B. ovis* ATCC 25840 and *Brucella* sp. strain NVSL 07-006 (groups 41, 43, 48 and 54), supporting the grouping of those two species, already noticed when discussing the whole gene tree, and confirmed by the large deletion around the 2.8 kbp fragment discussed previously. Those two species were the likely subjects of an intense IS711 transposition activity, with 14 elements being uniquely found in *B. ovis* and 9 being uniquely found in *Brucella* sp. NVSL 07-006. Among the 31 IS711 positions found in *B. pinnipedialis* B2/94, 21 are shared with *B. ceti* Cudo among which 12 are found only in *B. pinnipedialis* B2/94 and *B. ceti* Cudo,

and absent from all other strains (groups 5-9, 17, 18, 19, 21, 23 and 24).

Difference in gene content

Comparing orthologous loci (present in both *B. microti* CCM 4915 and *B. pinnipedialis* B2/94 genomes), we found that the number of pseudogenes in the genome of *B. pinnipedialis* B2/94 was larger than that in the genome of *B. microti* CCM 4915 (30 for *B. microti* on chromosome I and 48 for *B. pinnipedialis*, 16 for *B. microti* on chromosome II and 42 for *B. pinnipedialis*). This was also noticed when comparing *B. microti* CCM 4915 and *B. suis* 1330, and was attributed to a slower evolution rate in *B. microti*. The genes that are potentially different in both species are reported in Additional file 6, Table S5.

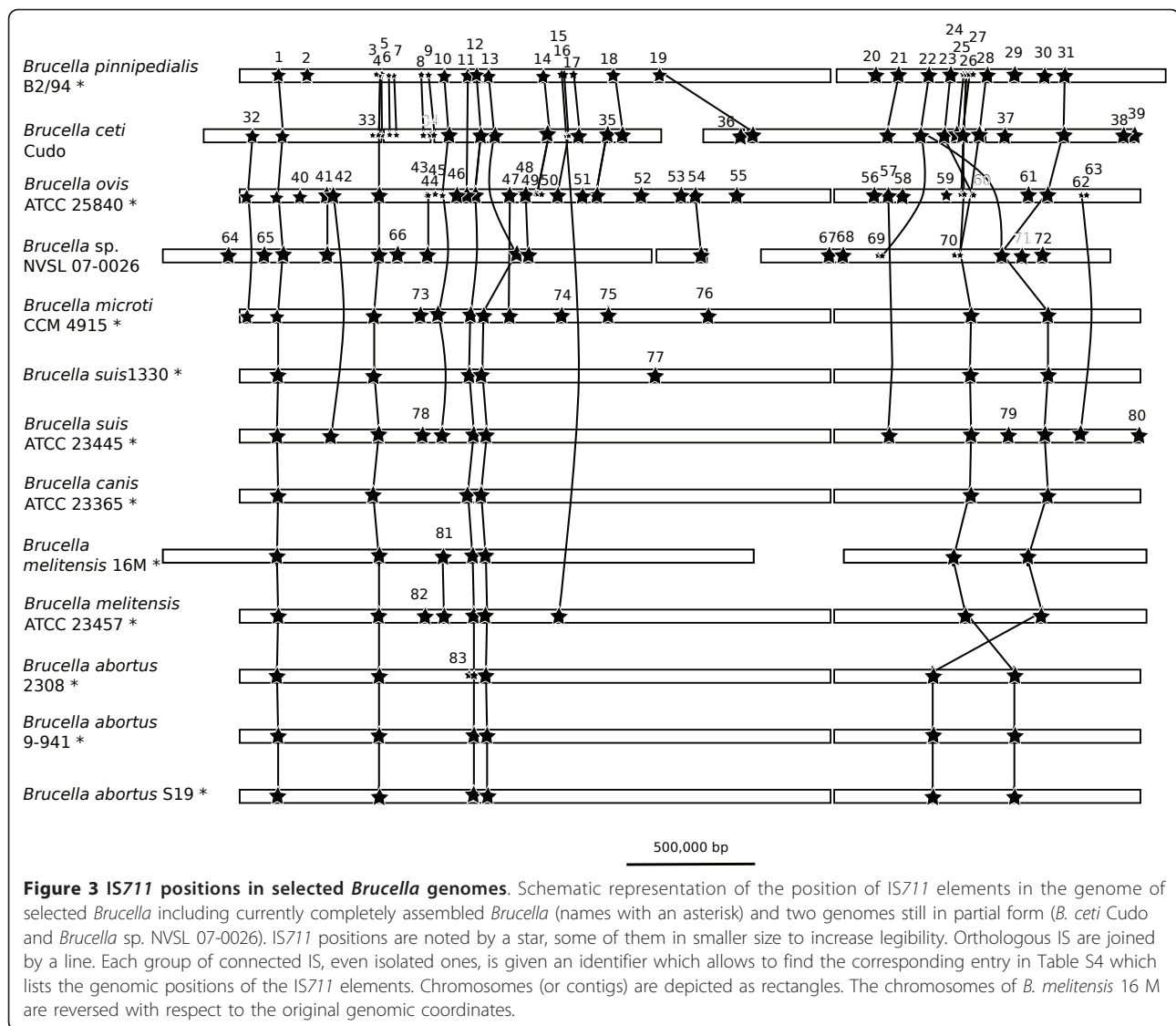
Among the genes altered in *B. pinnipedialis* B2/94, many are components of ABC transporters. Those genes are highlighted in Additional file 6, Table S5. We found 17 genes related to ABC transporters that are impaired, on a total of 90 impaired genes. In the *B. pinnipedialis* genome we identified approximately 249 ABC transporter-related intact genes on a total of 3,342 protein coding genes. This output is highly improbable (p-value = 1.e-4, chi-square test) and strongly suggest that ABC transporters were specifically degraded in *B. pinnipedialis* and more generally in host-associated *Brucella* species. Beside ABC type transport systems, many other genes involved in transport are found impaired in *B. pinnipedialis* B2/94: a CorA family transporter BPI_I592 (ortholog BMI_I558), an EmrB/Qaca family drug resistance transporter BPI_I1098 (ortholog BMI_I1064), the dipeptide transport system permease protein DppC (BPI_I1637, ortholog BMI_I1597), an outer membrane autotransporter BPI_I2072 (ortholog BMI_I2035), a glucose/galactose transporter BPI_II188 (ortholog BMI_III187), a cadmium-translocating P-type ATPase BPI_II1260 (ortholog BMI_II1204), and finally a putative transport protein BPI_II453 (ortholog BMI_II468).

One of those ABC transporter genes, BPI_I1818, exhibits an interesting feature. A frameshift difference between the *B. pinnipedialis* B2/94 and *B. microti* CCM 4915 sequences merges the membrane and ATP-binding components (BMI_I1778-9) of a thiamin ABC transporter into a single reading frame. A blast (tblastn) search of the BPI_I1818 gene sequence against the nucleotide sequences of the other *Brucella* strains shows that this gene fusion occurred in all marine mammal strains studied, except for *B. ceti* M13/05/1 and M644/93 which represent the distinct dolphin subgroup of strains within *B. ceti* mentioned above.

Conclusions

Time of appearance of marine mammal *Brucella* species

It has been suggested [14] that the divergence of species in the genus *Brucella* could have been concomitant with



the divergence of their mammalian hosts, 60 millions years (my) ago. However this is inconsistent with the fact that the hosts of *B. ceti* and *B. pinnipedialis* did not diverge at the same time. The ancestors of pinnipeds where carnivores and Higdon *et al.* [34] used molecular data to estimate the split between ursids and pinnipeds to 35.7 ± 2.63 (= mean \pm SE) my, and fossil records report early pinnipeds 35 my ago [35]. Cetaceans went back to the sea much earlier, the oldest known cetaceans date back to the Eocene, 55 my ago [36]. If we consider divergence in the 16S rRNA gene sequence, and referring to *B. microti* which has a central position, *B. pinnipedialis*, *B. melitensis*, *B. abortus*, *B. suis* (perhaps with the exception of *B. suis* 513, which has 2 (nt) differences, C- > T at position 11 and G- > T at position 1468), *B. ovis*, *B. canis* have all identical sequences, and *B. neotomae*

has 1 bp difference (C- > T at position 541). Using the estimate of 1-2% of change in 16S rRNA sequence per 50 my, 1 bp difference (which really should be considered as a maximal) corresponds to 0.07% change, and a divergence time of 1.75-3.5 my. This time estimate is probably a crude overestimation and recent work [37], using single nucleotide polymorphisms from 13 genomes, showed that most *Brucella* species probably diverged 86,000 to 296,000 years ago. This analysis reveals that the divergence time of *Brucella* sp. found today in marine mammals is totally incompatible with the divergence time of their hosts. A fortuitous contamination of cetaceans and pinnipeds, probably via the food chain, may explain better this transmission of *Brucella* to the marine mammals. This also opens the remote possibility of marine *Brucella* infecting terrestrial mammals.

On the order of appearance of host-associated *Brucella* species

The analysis of distinctive genomic regions between *B. microti* CCM 4915 and *B. pinnipedialis* B2/94 as well as the study of additional markers reveal the order of appearance of the different *Brucella* species. It is clearly apparent that most of the events following the divergence of *B. microti* from the classical *Brucella* species occurred in a very small amount of time, as if something caused a sudden radiation in this lineage and a subsequent adaptation of the organisms to their hosts. Here we summarize some of the major evolutionary events that highlight the evolutionary history of the genus *Brucella*.

Following the divergence of *B. microti*, the next evolutionary event that we can trace is the 2.6 kbp fragment clear disappearance, which tells us that the two next *Brucella* groups to diverge were *B. suis* and *B. neotomae*. We did not find any good marker in favor of a prior divergence of one versus the other.

Marine mammal *Brucella* species as well as *B. ovis* and *Brucella* sp. NVSL 07-0026 share the presence of a high number of IS711 elements, and it has been demonstrated that IS711 transposition is still an active process in *B. ovis* and *B. pinnipedialis* [38]. Those insertion elements are much less numerous in *B. melitensis* and *B. abortus* strains and we thus assume that IS711 transposition events occurred quite abundantly after the divergence of *B. melitensis*/*B. abortus*. In this group, *B. abortus* strains share a unique feature which is a genomic 600 kbp inversion in the small chromosome, as clearly depicted on Figure 1 of reference [39].

The whole gene tree, where *B. ovis* and *Brucella* sp. NVSL 07-0026 cluster together, the 4 IS711 elements positions that they have in common and not shared with *B. pinnipedialis* B2/94 as well as the large deletion that those two species share, in the region surrounding the 2.8 kbp deletion in *B. pinnipedialis*, all those facts support the grouping of *B. ovis* with *Brucella* sp. NVSL 07-0026. The grouping of marine mammal *Brucella* strains that we observe in the tree with early divergence of *B. ceti* M13/05/1 and M644/93 is also reflected by the gene fusion event mentioned earlier. Grouping of *B. pinnipedialis* M292/94 and B2/94 is supported by the loss of the 18 kbp fragment in these two strains.

The explosive radiation in the genus *Brucella*

There is a clear transition in the genus *Brucella* evolutionary tree. The first *Brucella* discovered were the host-associated *Brucella* species, but more recently, *B. microti* was isolated as the first representative of a fast growing list of free-living *Brucella*. This biochemically highly active bacteria was found to share more phenotypic traits with *Ochrobactrum* than with the host-associated *Brucella* species [6]. We proposed earlier [30] that the transition

between a free-living and an host-associated life style could have resulted from the modification in the 23S ribosomal RNA gene sequence with putative effects on the growth rate of the bacteria. A slow growth rate has often been advocated for intra-cellular bacteria, as their survival is often dependent on the survival of their hosts [40]. *Brucella* with this change in 23S structure and its impact on growth rate became suddenly more adapted to an host-associated life style than to a free-living style, and progressively adapted to distinct sets of hosts, giving rise to the main lineages of host-associated *Brucella* species that are encountered today.

Methods

Sequencing and origin of sequence data used for comparative work

Genome was assembled from 430,042 paired GS-FLX reads of average length 229, giving approximately a 30× coverage of the genome, and directed sequencing of the remaining gaps was performed using 193 additional Sanger sequencing reactions. The genome sequence is deposited in the complete genome division of GenBank under project ID 41867 and accession numbers CP002078 and CP002079. Genomic sequence data of all *Brucella* strains mentioned in this work can be conveniently downloaded from a unique location at the Pathosystems Resource Integration Center web site [41], or otherwise from the complete genome and whole genome shotgun divisions of GenBank [42]. Origin of the sequence data is listed in Supplementary Table 3. Most sequences originate from the *Brucella* group project at the Broad Institute, conducted by Davis O'Callaghan, Adrian Whatmore and Renee Tsolis [43] or from the Pathosystems Resource Integration Center of the Virginia Bioinformatics Institute [41].

Whole gene tree

Whole gene tree was build using gene sequences from all available *Brucella* as well as *Ochrobactrum intermedium*. The procedure used to build the tree is similar to that reported in [30]. Briefly, 1125 orthologous genes from 39 *Brucella* plus genes from *O. intermedium* were used. Orthologous genes were selected using the following procedure. A file containing nucleotide gene sequences for the selected organisms was compared to itself using blastn [44] (parameters: -b 100 -v 100 -F F -e 1.e-20). The resulting output file was subjected to clustering using the Markov chain clustering algorithm [45]. In the resulting cluster list, we selected the clusters with only one unique member per species. Genes from each cluster were then aligned using MUSCLE [46] (default parameters). The following alignments were concatenated, resulting in an alignment of 40 sequences, with 1,078,083 positions, cleaned with Gblock [47] (default parameters), which

reduced it to 945,578 positions. From this multiple alignment, on which 767,738 sites without polymorphisms and 2189 distinct patterns were found, a tree was inferred by maximum likelihood using PhyML [48], with 100 bootstrap replicates.

Unshared genome segments, tracing segments history

The nucleotide sequence of the genome fragment coming either from *B. pinnipedialis* B2/94 or *B. microti* CCM 4915 was used as a query for blastn search (e-value = 1.e-100, no filter) against the nucleotide sequences of the remaining *Brucella*. For each target genome, the number of distinct nucleotide positions that had a hit was recorded in Additional file 3, Table S3. Presence or absence of a given fragment is represented in Figure 1 alongside the whole gene tree. The history of those genome segments was subsequently traced using the Mesquite software package [49] on cladograms reflecting the topology obtained from the whole gene tree (Figure 1). Character history was computed using parsimony analysis, with presence or absence of a genome segment treated as a discrete category character.

Across genomes identification of IS711 element positions

A database containing the complete genomes of the *Brucella* species under study was compiled. We searched this database for the occurrence of IS711 using the nucleotide sequence of one element from *B. pinnipedialis* as query, using the blastn program (parameters: -b 1000 -v 1000, limiting to hits with a score of 500). We extracted the corresponding segments, adding 500 nucleotides of context on both side. We then masked this file so that the IS711 sequence itself was replaced by x in the sequence. A subsequently blastn search of this file against itself allowed us to recover the orthologous IS711 positions along the different *Brucella* genomes (listed in Additional file 5, Table S4), used for the representation in Figure 3.

Additional material

Additional file 1: Indels in *B. pinnipedialis* B2/94 and *B. microti* large chromosome alignment. List of insertions and deletions (indels) in the complete genome alignment of *B. pinnipedialis* B2/94 and *B. microti* CCM 4915, large chromosome.

Additional file 2: Indels in *B. pinnipedialis* B2/94 and *B. microti* small chromosome alignment. List of insertions and deletions (indels) in the complete genome alignment of *B. pinnipedialis* B2/94 and *B. microti* CCM 4915, small chromosome.

Additional file 3: Occurrences of the largest genome fragments unshared between *B. pinnipedialis* B2/94 and *B. microti* CCM 4915 in selected *Brucella* genomes. Occurrences of the largest genome fragments unshared between *B. pinnipedialis* B2/94 and *B. microti* CCM 4915 in selected *Brucella* genomes. For each organism, the number of positions in its genome that are found similar to the original fragment is reported. The column "Origin of sequence data" indicates "GenBank" for sequences deposited in the complete genome division of GenBank,

"Patric" if the sequence originates from the Pathosystems Resource Integration Center of the Virginia Bioinformatics Institute, and "Broad" if the sequence originates from the *Brucella* genome sequencing project at the Broad Institute.

Additional file 4: Evolutionary history of the unshared sequence fragments along the *Brucella* evolutionary tree. Evolutionary history of the unshared sequence fragments along the *Brucella* evolutionary tree. The tree is represented as a cladogram with the same topology as that of the whole gene tree (Figure 1). Plots are generated using Mesquite.

Additional file 5: Position of IS711 insertion sequences in the genome of completely sequenced *Brucella*, *B. ceti* Cudo and *Brucella* sp. NVSL 07-0026. Position of IS711 insertion sequences in the genome of completely sequenced *Brucella*, *B. ceti* Cudo and *Brucella* sp. NVSL 07-0026. Each IS711 element position as well as the accession number of the genome segment to which it belongs are listed in the Table. Clusters of IS711 elements are identified by a number allowing to locate them in Figure 3.

Additional file 6: List of genes showing a large change between *B. pinnipedialis* B2/94 and *B. microti* CCM 4915. List of genes showing a large change between *B. pinnipedialis* B2/94 and *B. microti* CCM 4915. ABC transporter genes are indicated.

Author details

¹Laboratoire Information Génomique et Structurale, CNRS-UPR2589, Aix-Marseille University, Institut de Microbiologie de la Méditerranée (IM2, IFR-88), Parc Scientifique de Luminy-163 Avenue de Luminy-Case 934-FR-13288, Marseille cedex 09, France. ²CNRS, UMR 7144, Equipe Evolution du Plancton et Paléo-Océans, Station Biologique de Roscoff, 29682 Roscoff, France. ³UPMC Univ Paris 06, UMR 7144, Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, 29682 Roscoff, France. ⁴INRA, UR1282, Infectiologie Animale et Santé Publique, IASP, Nouzilly, F-37380, France.

Authors' contributions

SA participated in the design of the study, performed genome assembly, annotation and comparisons with other *Brucella* and wrote the manuscript. ML participated in genome annotation and analyzed the data. JMC participated in the design of the study, analyzed data and wrote the manuscript. MZ and AC designed the study, coordinated the research, analyzed data and wrote the manuscript. All authors read and approved the final manuscript.

Received: 1 April 2011 Accepted: 11 July 2011 Published: 11 July 2011

References

- Godfroid J, Cloeckaert A, Liautard JP, Kohler S, Fretin D, Walravens K, Garin-Bastuji B, Letesson JJ: **From the discovery of the Malta fevers agent to the discovery of a marine mammal reservoir, brucellosis has continuously been a re-emerging zoonosis.** *Vet Res* 2005, **36**:313-326.
- Moreno E, Cloeckaert A, Moriyón I: ***Brucella* evolution and taxonomy.** *Vet Microbiol* 2002, **90**:209-227.
- Whatmore AM: **Current understanding of the genetic diversity of *Brucella*, an expanding genus of zoonotic pathogens.** *Infect Genet Evol* 2009, **9**:1168-1184.
- Foster G, MacMillan AP, Godfroid J, Howie F, Ross HM, Cloeckaert A, Reid RJ, Brew S, Patterson IAP: **A review of *Brucella* sp. infection of sea mammals with particular emphasis on isolates from Scotland.** *Vet Microbiol* 2002, **90**:563-580.
- Foster G, Osterman BS, Godfroid J, Jacques I, Cloeckaert A: ***Brucella ceti* sp. nov. and *Brucella pinnipedialis* sp. nov. for *Brucella* strains with cetaceans and seals as their preferred hosts.** *Int J Syst Evol Microbiol* 2007, **57**:2688-2693.
- Scholz HC, Hubalek Z, Sedláček I, Vergnaud G, Tomaso H, Al Dahouk S, Melzer F, Kämpfer P, Neubauer H, Cloeckaert A, Maquart M, Zygmunt MS, Whatmore AM, Falsen E, Bahn P, Göllner C, Pfeffer M, Huber B, Busse HJ, Nöckler K: ***Brucella microti* sp. nov., isolated from the common vole *Microtus arvalis*.** *Int J Syst Evol Microbiol* 2008, **58**:375-382.
- Scholz HC, Hofer E, Vergnaud G, Le Fleche P, Whatmore AM, Al Dahouk S, Pfeffer M, Krüger M, Cloeckaert A, Tomaso H: **Isolation of *Brucella microti***

- from Mandibular Lymph Nodes of Red Foxes, *Vulpes vulpes*, in Lower Austria. *Vector Borne Zoonotic Dis* 2009, **9**:153-156.
8. Scholz HC, Hubalek Z, Nesvadbova J, Tomaso H, Vergnaud G, Le Flèche P, Whatmore AM, Al Dahouk S, Krüger M, Lodri C, Pfeffer M: **Isolation of *Brucella microti* from soil.** *Emerg Infect Dis* 2008, **14**: 1316-1317.
 9. De BK, Stauffer L, Koylass MS, Sharp SE, Gee JE, Helsel LO, Steigerwalt AG, Vega R, Clark TA, Daneshvar MI, Wilkins PP, Whatmore AM: **Novel *Brucella* strain (BO1) associated with a prosthetic breast implant infection.** *J Clin Microbiol* 2008, **46**:43-49.
 10. Scholz HC, Nöckler K, Göllner C, Bahn P, Vergnaud G, Tomaso H, Al-Dahouk S, Kämpfer P, Cloeckaert A, Maquart M, Zygmunt MS, Whatmore AM, Pfeffer M, Huber B, Busse HJ, De BK: ***Brucella inopinata* sp. nov., isolated from a breast implant infection.** *Int J Syst Evol Microbiol* 2010, **60**:801-808.
 11. Schlabititz-Loutsevitch NE, Whatmore AM, Quance CR, Koylass MS, Cummins LB, Dick EJ Jr, Snider CL, Cappelli D, Ebersole JL, Nathanielsz PW, Hubbard GB: **A novel *Brucella* isolate in association with two cases of stillbirth in non-human primates-first report.** *J Med Primatol* 2009, **38**:70-73.
 12. Tiller RV, Gee JE, Frace MA, Taylor TK, Setubal JC, Hoffmaster AR, De BK: **Characterization of novel *Brucella* strains originating from wild native rodent species in North Queensland, Australia.** *Appl Environ Microbiol* 2010, **76**:5837-5845.
 13. Tiller RV, Gee JE, Lonsway DR, Gribble S, Bell SC, Jennison AV, Bates J, Coulter C, Hoffmaster AR, De BK: **Identification of an unusual *Brucella* strain (BO2) from a lung biopsy in a 52 year-old patient with chronic destructive pneumonia.** *BMC Microbiology* 2010, **10**:23.
 14. Bourg G, OCallaghan D, Boschiroli ML: **The genomic structure of *Brucella* strains isolated from marine mammals gives clues to evolutionary history within the genus.** *Vet Microbiol* 2007, **125**:375-380.
 15. Clavareau C, Wellemans V, Walravens K, Tryland M, Verger JM, Grayon M, Cloeckaert A, Letesson JJ, Godfroid J: **Phenotypic and molecular characterization of a *Brucella* strain isolated from a minke whale (*Balaenoptera acutorostrata*).** *Microbiology* 1998, **144**:3267-3273.
 16. Cloeckaert A, Verger JM, Grayon M, Paquet JY, Garin-Bastuji B, Foster G, Godfroid J: **Classification of *Brucella* spp. isolated from marine mammals by DNA polymorphism at the omp2 locus.** *Microbes Infect* 2001, **3**:729-738.
 17. Dawson CE, Stubberfield EJ, Perrett LL, King AC, Whatmore AM, Bashiruddin JB, Stack JA, Macmillan AP: **Phenotypic and molecular characterisation of *Brucella* isolates from marine mammals.** *BMC Microbiol* 2008, **8**:224.
 18. Groussaud P, Shankster SJ, Koylass MS, Whatmore AM: **Molecular typing divides marine mammal strains of *Brucella* into at least three groups with distinct host preferences.** *J Med Microbiol* 2007, **56**:1512-1518.
 19. Maquart M, Le Flèche P, Foster G, Tryland M, Ramisse F, Djønne B, Al Dahouk S, Jacques I, Neubauer H, Walravens K, Godfroid J, Cloeckaert A, Vergnaud G: **MLVA-16 typing of 295 marine mammal *Brucella* isolates from different animal and geographic origins identifies 7 major groups within *Brucella ceti* and *Brucella pinnipedialis*.** *BMC Microbiol* 2009, **9**:145.
 20. McDonald WL, Jamaludin R, Mackereth G, Hansen M, Humphrey S, Short P, Taylor T, Swingle J, Dawson CE, Whatmore AM, Stubberfield E, Perrett LL, Simmons G: **Characterization of a *Brucella* sp. strain as a marine-mammal type despite isolation from a patient with spinal osteomyelitis in New Zealand.** *J Clin Microbiol* 2006, **44**:4363-4370.
 21. Sohn AH, Probert WS, Glaser CA, Gupta N, Bollen AW, Wong JD, Grace EM, McDonald WC: **Human neurobrucellosis with intracerebral granuloma caused by a marine mammal *Brucella* spp.** *Emerg Infect Dis* 2003, **9**:485-488.
 22. Whatmore AM, Dawson CE, Groussaud P, Koylass MS, King AC, Shankster SJ, Sohn AH, Probert WS, McDonald WL: **Marine mammal *Brucella* genotype associated with zoonotic infection.** *Emerg Infect Dis* 2008, **14**:517-518.
 23. Halling SM, Tatum FM, Bricker BJ: **Sequence and characterization of an insertion sequence, IS711, from *Brucella ovis*.** *Gene* 1993, **133**:123-127.
 24. Ouahrani S, Michaux S, Sri Widada J, Bourg G, Tournebize R, Ramuz M, Liautaud JP: **Identification and sequence analysis of IS6501, an insertion sequence in *Brucella* spp.: relationship between genomic structure and the number of IS6501 copies.** *J Gen Microbiol* 1993, **139**:3265-3273.
 25. Bricker BJ, Ewalt DR, MacMillan AP, Foster G, Brew S: **Molecular characterization of *Brucella* strains isolated from marine mammals.** *J Clin Microbiol* 2000, **38**:1258-1262.
 26. Cloeckaert A, Grayon M, Grépinet O, Boumedine KS: **Classification of *Brucella* strains isolated from marine mammals by infrequent restriction site-PCR and development of specific PCR identification tests.** *Microbes Infect* 2003, **5**:593-602.
 27. Maquart M, Fardini Y, Zygmunt MS, Cloeckaert A: **Identification of novel DNA fragments and partial sequence of a genomic island specific of *Brucella pinnipedialis*.** *Vet Microbiol* 2008, **132**:181-189.
 28. Zygmunt MS, Maquart M, Bernardet N, Doublet B, Cloeckaert A: **Novel IS711-specific chromosomal locations useful for identification and classification of marine mammal *Brucella* strains.** *J Clin Microbiol* 2010, **48**:3765-3769.
 29. Bohlin J, Snipen L, Cloeckaert A, Lagesen K, Ussery D, Kristoffersen AB, Godfroid J: **Genomic comparisons of *Brucella* spp. and closely related bacteria using base compositional and proteome based methods.** *BMC Evol Biol* 2010, **10**:249.
 30. Audic S, Lescot M, Claverie JM, Scholz HC: ***Brucella microti*: the genome sequence of an emerging pathogen.** *BMC Genomics* 2009, **10**:352.
 31. Tsois RM, Seshadri R, Santos RL, Sangari FJ, Lobo JM, de Jong MF, Ren Q, Myers G, Brinkac LM, Nelson WC, Deboy RT, Angiuoli S, Khouri H, Dimitrov G, Robinson JR, Mulligan S, Walker RL, Elzer PE, Hassan KA, Paulsen IT: **Genome degradation in *Brucella ovis* corresponds with narrowing of its host range and tissue tropism.** *PLoS ONE* 2009, **4**:e5519.
 32. Sigquier P, Perochon J, Lestrade L, Mahillon J, Chandler M: **ISfinder: the reference centre for bacterial insertion sequences.** *Nucleic Acids Res* 2006, **34**:D32-36.
 33. Lerat E: **Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs.** *Heredity* 2010, **104**:520-533.
 34. Higdon J, Bininda-Emonds O, Beck R, Ferguson S: **Phylogeny and divergence of the pinnipeds (Carnivora: Mammalia) assessed using a multigene dataset.** *BMC Evol Biol* 2007, **7**:216.
 35. Rybczynski N, Dawson MR, Tedford RH: **A semi-aquatic Arctic mammalian carnivore from the Miocene epoch and origin of Pinnipedia.** *Nature* 2009, **458**:1021-1024.
 36. Thewissen JGM, Cooper LN, Clementz MT, Bajpai S, Tiwari BN: **Whales originated from aquatic artiodactyls in the Eocene epoch of India.** *Nature* 2007, **450**:1190-1194.
 37. Foster JT, Beckstrom-Sternberg SM, Pearson T, Beckstrom-Sternberg JS, Chain PS Roberto FF, Hnath J, Brettin T, Keim P: **Whole Genome-Based Phylogeny and Divergence of the Genus *Brucella*.** *J Bacteriol* 2009, **191**:2864-2870.
 38. Ocampo-Sosa A, Garcia-Lobo J: **Demonstration of IS711 transposition in *Brucella ovis* and *Brucella pinnipedialis*.** *BMC Microbiol* 2008, **8**:17.
 39. Wattam AR, Williams KP, Snyder EE, Almeida NF Jr, Shukla M, Dickerman AW, Crasta OR, Kenyon R, Lu J, Shallom JM, Yoo H, Ficht TA, Tsois RM, Munk C, Tapia R, Han CS, Dettler JC, Bruce D, Brettin TS, Sobral BW, Boyle SM, Setubal JC: **Analysis of ten *Brucella* genomes reveals evidence for horizontal gene transfer despite a preferred intracellular lifestyle.** *J Bacteriol* 2009, **191**:3569-3579.
 40. Winkler HH: ***Rickettsia prowazekii*, ribosomes and slow growth.** *Trends Microbiol* 1995, **3**:196-198.
 41. **Pathosystems resource integration center.** [http://brcdownloads.vbi.vt.edu/patric2/PATRIC/].
 42. **National Center for Biotechnology Information.** [http://www.ncbi.nlm.nih.gov/].
 43. ***Brucella* group project at the Broad Institute.** [http://www.broadinstitute.org/annotation/genome/brucella_group/Info.html].
 44. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25**:3389-33402.
 45. Enright AJ, Van Dongen S, Ouzounis CA: **An efficient algorithm for large-scale detection of protein families.** *Nucleic Acids Res* 2002, **30**:1575-1584.
 46. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792-1797.
 47. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17**:540-552.
 48. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Syst Biol* 2003, **52**:696-704.
 49. **Mesquite: a modular system for evolutionary analysis.** [http://mesquiteproject.org/].
- doi:10.1186/1471-2148-11-200
- Cite this article as:** Audic et al.: The genome sequence of *Brucella pinnipedialis* B2/94 sheds light on the evolutionary history of the genus *Brucella*. *BMC Evolutionary Biology* 2011 **11**:200.