The Genome Sequence of *Rickettsia felis* Identifies the First Putative Conjugative Plasmid in an Obligate Intracellular Parasite

Hiroyuki Ogata1*, Patricia Renesto2, Stéphane Audic1, Catherine Robert2, Guillaume Blanc1, Pierre-Edouard Fournier1,2, Hugues Parinello2, Jean-Michel Claverie1, Didier Raoult2*

1 Structural and Genomic Information Laboratory, UPR 2589, IBSM, CNRS, Marseille Cedex, France, 2 Unité des Rickettsies, UMR 6020, IFR 48, CNRS, Faculté de Médecine, Marseille Cedex, France

We sequenced the genome of *Rickettsia felis*, a flea-associated obligate intracellular α-proteobacterium causing spotted fever in humans. Besides a circular chromosome of 1,485,148 bp, *R. felis* exhibits the first putative conjugative plasmid identified among obligate intracellular bacteria. This plasmid is found in a short (39,263 bp) and a long (62,829 bp) form. *R. felis* contrasts with previously sequenced *Rickettsia* in terms of many other features, including a number of transposases, several chromosomal toxin–antitoxin genes, many more spoT genes, and a very large number of ankyrin- and tetratricopeptide-motif-containing genes. Host-invasion-related genes for patatin and RickA were found. Several phenotypes predicted from genome analysis were experimentally tested: conjugative pilus and mating were observed, as well as β-lactamase activity, actin-polymerization-driven mobility, and hemolytic properties. Our study demonstrates that complete genome sequencing is the fastest approach to reveal phenotypic characters of recently cultured obligate intracellular bacteria.

Citation: Ogata H, Renesto P, Audic S, Robert C, Blanc G, et al. (2005) The genome sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. PLoS Biol 3(8): e248.
included a remarkably high number of paralogs for transposases, surface cell antigens (sca), global metabolism regulators (spoT), and proteins containing protein–protein interaction motifs such as ankyrin repeats and tetratricopeptide repeats (TPRs). Furthermore, we identified many other ORFs putatively associated with the adaptations of *R. felis* to its host environment or with its pathogenesis.

**Plasmids**

The two *R. felis* plasmids, named pRF and pRF₈, are 62,829 bp and 39,263 bp long, respectively. Their topologies and sizes were confirmed experimentally (Figures S1 and S2). The pRF plasmid contains 68 ORFs, of which 53 (77.9%) exhibited homologs in public databases and 44 (64.7%) were associated with functional attributes. The nucleotide sequences of pRF₈ and pRF are identical, except for an additional 23,566-bp segment that contains 24 ORFs (pRF₁₅–pRF₃₈) in pRF (see Table 3). These plasmids are likely to be *R. felis* specific since all attempts to detect specific plasmid sequences by polymerase chain reaction (PCR) from DNA of available reference rickettsial species were unsuccessful. In contrast, the same assays against 30 fleas naturally infected by *R. felis* resulted in amplification of the plasmid sequences in all cases.

Plasmids are referred to as conjugative or nonconjugative. The former are disseminated by conjugation from cell to cell, while the latter are only vertically transmitted. The pRF plasmid encodes several homologs of proteins involved in the different conjugative steps (see Table 3; Figure S3). First, it exhibits a split gene (pRF₃₈/pRF₃₉) homologous to the *traA₅* of the *Agrobacterium tumefaciens* tumor-inducing plasmid [10]. *TraAT₅* is thought to be a DNA-processing machinery with nickase and helicase activities to generate the transfer strand from the origin of transfer (*oriT*) [10]. Second, the pRF encodes another split gene (pRF₄₃/pRF₄₄) homologous to the *traD₅* in the *Escherichia coli* F plasmid. *TraD₅* is a "coupling protein" that connects the DNA-processing machinery (and transfer strand) to the mating pair formation (Mpf) apparatus, a type IV secretion system (T4SS) [11]. Finally, pRF exhibits an ORF (pRF₄₇) similar to *TraGF*, a protein involved in the F-pilus assembly and aggregate stabilization [12].

Despite the presence of these ORFs linked to the initiation of plasmid transfer, the pRF sequence lacks clear homologs for the proteins involved in the Mpf apparatus found in other bacteria. Nevertheless, the *R. felis* chromosome (as well as other *Rickettsia* genomes) encodes most of the components of T4SS, which are highly similar to the *vir* genes of *A. tumefaciens*. Since the *R. felis* T4SS components (*virB2*...
Table 1. Comparison of R. felis and Other Published Rickettsia Genomes

| Group | Species | Size (bp) | G+C (%) | Coding (%) | ORFs | RNAs | Repeat (%) | RPEs | Gene Families | Transposases | Toxin-Antitoxin | Ankyrin | TPR | Vector |
|-------|---------|-----------|---------|-----------|------|------|------------|------|-------------|--------------|--------------|---------|------|--------|
| RFG   | R. felis | 1,587,240 | 32.5    | 83.6      | 1,512 | 39   | 4.3        | 728  | 88          | 145          | 122          | 0      | 0    | Flea   |
|       | Chromosome | 1,485,148 | 32.5    | 83.8      | 1,400 | 39   | 4.4        | 725  | 88          | 66           | 30           | 0      | 0    |        |
|       | pRF plasmid | 62,829  | 33.6    | 80.3      | 68    | 0    | 4.8        | 1    | 0           | 9            | 0            | 0      | 0    |        |
|       | pRF6 plasmid | 39,263  | 33.2    | 80.6      | 44    | 0    | 1.6        | 1    | 0           | 7            | 0            | 0      | 0    |        |
|       | R. conorii | 1,268,755 | 32.4    | 81.5      | 1,374 | 39   | 0.65       | 559  | 78          | 74           | 0            | 11     | 2    | Tick   |
| Typhus group | R. prowazekii | 1,111,523 | 29.0    | 76.2      | 834   | 39   | 0.30       | 120  | 22          | 68           | 0            | 0      | 3    |        |
|       | R. typhi | 1,111,496 | 28.9    | 76.3      | 838   | 39   | 0.29       | 121  | 25          | 56           | 0            | 0      | 1    | Flea   |

aThe G+C (%) and coding (%) values are average G+C content and coding density in the genome. ORF is Open Reading Frame. RNAs are RNA-coding genes.
bThe repeat (%) value is the percentage of repetitive DNA sequences in the genome.
cThe number of RPEs (Repeat Elements) is the number of repetitive elements in the genome.
dNumber of gene families were obtained by BLASTP (E-value threshold of 10^-5) with a single-linkage clustering method. The number in parentheses corresponds to the number of R. felis gene families computed by omitting the ORFs in the pRF plasmid.
eORFs for toxin-antitoxin systems.
fAnkyrin-repeat-containing ORFs.
gTPR-containing ORFs.

First Putative Conjugative Plasmid Identified

[RF1075], virB3 [RF0087], virB4 [RF0088], virB6 [RF0089], RF0090, RF0091, RF0092, and RF0093, virB8 [RF0463 and RF0465], virB9 [RF0462 and RF0466], virB10 [RF0467], virB11 [RF0468], and virD4 [RF0469]) are conserved in all sequenced Rickettsia genomes that lack plasmids, the primary suspected role of the R. felis T4SS is to translocate virulence factors to hosts. However, the T4SS of R. prowazekii (dot/icm) have been shown to function both as DNA-transfer machineries and as effector translocators [13]. Thus, the R. felis T4SS may also promote the transfer of DNA as in A. tumefaciens. We also noticed that the R. felis chromosome exhibits a DNA primase gene (RF0786) similar to TraC found in the E. coli IncP plasmid. TraC initiates the replication of transferred DNA strands in the recipient cells. Finally, the R. felis chromosome encodes a protein (RF0020) similar to competence protein ComE3, a protein (RF0964) similar to the F-pilin acetylation protein TraX, and a split gene (RF0705/RF0706) homologous to the Pil-pilus assembly protein FimD. In conclusion, the presence of those putative conjugative transfer genes suggests that the R. felis plasmids have been acquired by conjugation and that R. felis may still retain the capacity of transferring plasmids.

Genome Plasticity

We identified 333 repeated DNA sequences (50 to 2,645 bp long) in the R. felis genome, accounting for 4.3% of the sequence, a proportion markedly higher than in other sequenced Rickettsia genomes (see Table 1; Figure 2B). The major source of those repeats is the proliferation of transposase genes, for which we identified 82 copies (or inactivated derivatives). Among other obligate intracellular bacteria, only W. pipientis uMel [14] and Parachlamydia sp. UWE25 [15] exhibit such a high number of large mobile genetic elements. The occurrence of highly similar transposase sequences appears to play a major role in the plasticity of the R. felis genome (see Figure 2A). Transposase ORFs were identified at most extremities of the R. felis genomic segments colinear with the R. conorii genome, suggesting that the R. felis chromosome has been rearranged many times through recombination mediated by these mobile sequences. With the use of the GRAPPA software inferring the most parsimonious genome-rearrangement scenario, we estimated at least 11 inversion events between R. felis and R. conorii. In contrast, only four inversions are required to associate more distantly related R. conorii and R. prowazekii genomes. In addition to transposases, we identified eight phage-related ORFs (see Table 2). The R. felis genome thus appears to have been invaded more frequently by such foreign DNAs than other Rickettsia species. Besides long repeats, Rickettsia genomes are known to contain a number of small palindromic repeats (Rickettsia palindromic elements [RPEs]) capable of invading both coding and noncoding regions [16]. We identified 728 RPEs in the R. felis genome. Of these RPEs, 85 were found within ORFs and three were found in RNA-coding genes.

The R. felis chromosome and plasmids share several homologs, suggesting gene exchanges between these replicons. Of 68 ORFs in pRF, 11 have a close homolog (>50% amino acid sequence identity) in the chromosome; these are seven transposases, patatin-like phospholipase (pRF11), thymidylate kinase (pRF13), and two small heat-shock proteins (pRF51 and pRF52). Among these, patatin-like proteins exhibit the most intriguing phylogeny (Figure S4). The genomes of five Rickettsia species (R. prowazekii, R. typhi, R. conorii, R. sibirica, and R. felis) exhibit chromosomal patatin-like phospholipase gene (pat1). Gene organization around pat1 is similar between these Rickettsia. Interestingly, a phylogenetic analysis for these Pat1 and the plasmid-encoded Pat2 indicates a close relationship between Pat1 (RF0360) and Pat2 of R. felis, together being an outgroup of Pat1 sequences of other Rickettsia, suggesting a gene replacement of the chromosomally encoded pat1 by the plasmid-encoded pat2 in the lineage leading to R. felis.

Most R. felis genes with orthologs in other Rickettsia have probably been inherited vertically from a common ancestor. On the other hand, genes without orthologs in other Rickettsia may have been acquired by lateral gene transfer. To test this hypothesis, we analyzed the taxonomic distribution of
Self-comparison of felis and in horizontal green lines indicate the positions of transposase ORFs in significantly different (distributions of the best matches for these two ORF sets were 58 ORFs) and cyanobacteria (18%; 33 ORFs). The taxonomic Rickettsia hits for the chromosomal ORFs lacking orthologs in other as partially (64%) with sequences from the same taxonomic group.

DOI: 10.1371/journal.pbio.0030248.g002

Comparison between related organisms by lateral transfer. However, methods that many Rickettsia ORFs with orthologs in other database (excluding rickettsial sequences) (Figure S5). BLASTP best hits of Genome Comparisons of Figure 2. Red dots represent homologous genomic segments greater than 150 bp identified by BLASTN (E-value < 10^-5).

(A) Comparison between R. felis and R. conorii chromosomes. Vertical and horizontal green lines indicate the positions of transposase ORFs in R. felis and in R. conorii, respectively.

(B) Self-comparison of R. felis chromosome.

BLASTP best hits of R. felis ORFs against the nonredundant database (excluding rickettsial sequences) (Figure S5). R. felis ORFs with orthologs in other Rickettsia matched preferentially (64%) with sequences from the same taxonomic group as R. felis (i.e., γ-proteobacteria). In contrast, the BLAST best hits for the chromosomal ORFs lacking orthologs in other Rickettsia were found preferentially in γ-proteobacteria (31%; 58 ORFs) and cyanobacteria (18%; 33 ORFs). The taxonomic distributions of the best matches for these two ORF sets were significantly different (p < 0.001; χ² test). This result suggests that many R. felis-specific genes may originate from distantly related organisms by lateral transfer. However, methods based on nucleotide composition bias failed to identify unambiguous candidates for lateral gene acquisition in R. felis.

Surface Antigens
The 16S family is one of the largest paralogous gene families in Rickettsia [8]. Five 16S members have been identified in the previously published Rickettsia genomes. Several Sca proteins are known to account for major antigenic differences between Rickettsia species [17] and may play important roles in adhesion to host cells [18]. Sca proteins are characterized by highly variable N-terminal sequences and a conserved C-terminal autotransporter β-domain, which translocates the N-terminal part outside the outer membrane. The R. felis genome exhibits the highest number of sca genes among currently available Rickettsia genomes. We identified nine intact sca paralogs (sca1, sca2, sca3, sca4, sca5/ompB, sca8, sca9, sca12, and sca13) as well as four fragmented or split paralogs (sca0/ompA, sca7, sca10, and sca11). Reverse transcriptase–polymerase chain reaction (RT-PCR) experiments demonstrated that, under mild log growth phase, all R. felis sca paralogs were transcribed, including split ones. Phylogenetic analyses suggest that ancient duplication events gave rise to these paralogs before the divergence of Rickettsia species. We noticed that sca genes exhibit highly different patterns of presence/absence across different Rickettsia species (Table S1). Only ompB and sca4 are conserved in all available Rickettsia genomes [19], remaining members being degraded or absent in one or more species. Together with the accelerated amino acid changes, differential gene degradation of sca paralogs probably contributes to the intra-species variation of those cell-surface proteins and might be linked with their adaptation to different host environments.

R. felis is genetically and serologically classified into the SFG of Rickettsia [20]. However, cross-reactivities caused by both proteins and lipopolysaccharides have been found with R. typhi using mouse sera [2] and human sera (Figure S6). R. conorii rarely cross-reacts with R. typhi. We therefore suspected that genes found in both R. felis and R. typhi, but missing in R. conorii, might be responsible for the cross-reactivities of R. felis and R. typhi. A list of such genes includes a sca family gene (sca3), encoding a protein with a predicted molecular weight of 319 kDa, and rfaJ for the lipopolysaccharide 1,2-glucosyltransferase (Table 4).

Adaptation to Environment
Transcriptional regulation may be of critical importance in R. felis, as the numbers of spoT, the gene regulating “alarmone,” and chromosomal toxin–antitoxin modules are higher in the R. felis genome than in any other sequenced bacterial genome.

SpoT and RelA are two hallmark enzymes regulating global cellular metabolism of E. coli in response to starvation [21]. These enzymes control the concentration of alarmone, (p)ppGpp (guanosine tetra- and pentaphosphates), which in turn acts as an effector of transcription. Remarkably, R. felis exhibits 14 spoT (spoT1–13 and 15) paralogs (Figure S7). Using RT-PCR, we examined the transcription status of 14 R. felis spoT genes. All the spoT ORFs were transcribed. We classified these ORFs into two groups, based on their alignment against the sequence of the Streptococcus dysgalactiae RelE, that possesses both (p)ppGpp hydrolase and synthetase activities [22]. The first group (SpoT1–10, 14, and 15) was aligned with

Figure 2. Genome Comparisons of R. felis and R. conorii
Red dots represent homologous genomic segments greater than 150 bp identified by BLASTN (E-value < 10^-5).
## Table 2. *R. felis*-Specific Genes Encoded in the Chromosome

| Category                        | ORFs | Gene Name | Annotation                                                                 |
|---------------------------------|------|-----------|-----------------------------------------------------------------------------|
| **Large gene families**         | 65 ORFs | tnp       | Transposase or inactivated derivative                                         |
|                                 | 16 ORFs | ank       | Ankyrin-repeat-containing gene                                               |
|                                 | 10 ORFs (five genes) | sco | Surface cell antigen homolog (Sca7–9, 11, 13)                              |
|                                 | 7 ORFs | tpr       | TPR-containing protein                                                       |
|                                 | 7 ORFs | spoT      | Guanosine polyphosphate pyrophosphohydrolases/synthetase homolog (SpoT5–10, 13) |
|                                 | 7 ORFs |           | MFS-type transporter                                                        |
| **Toxin–antitoxin system**      | 6 ORFs | reIE      | Cytoplasmic translational repressor of toxin–antitoxin system ReIE           |
|                                 | 5 ORFs | sbtD      | Antitoxin of toxin–antitoxin system SbtD                                    |
|                                 | 3 ORFs | phd       | Antitoxin of toxin–antitoxin system Phd                                     |
|                                 |       |           | Probable toxin of toxin–antitoxin stability                                 |
|                                 | RF0787 | parE      | Toxin of toxin–antitoxin system ParE                                        |
|                                 | RF0095 | vspC2     | Toxin of toxin–antitoxin system, containing PIN domain                       |
|                                 | RF0094 | vspB2     | Antitoxin of toxin–antitoxin system VapB                                    |
| **Drug resistance**             | RF0127 |           | Tellurite-resistance-protein-related protein                                |
|                                 | RF0774 |           | Streptomycin 6-kinase                                                       |
|                                 | RF0981 | mdtB      | ABC-type multidrug transport system, ATPase, and permease components        |
|                                 | RF1137 |           | Penicillin acylase (EC 3.5.1.11)                                             |
|                                 | RF1275 |           | Class D β-lactamase                                                         |
|                                 | RF1367 |           | Class C β-lactamase                                                         |
| **Phage-associated genes**      | RF0471 |           | Phage portal protein                                                        |
|                                 | RF0749/RF0750 |           | Phage prophage protease (HK97 family) and phage major capsid protein (HK97 family) |
|                                 | RF0570 |           | Phage-uncharacterized protein                                               |
|                                 | RF0793 |           | Phage-associated protein                                                    |
|                                 | RF0933 |           | Phage-related lysozyme                                                       |
|                                 | RF1287 |           | Phage-related transcriptional regulator                                      |
|                                 | RF1397 |           | Prophage antirepressor                                                      |
| **Plasmid/conjugation-related genes** | RF0020 |           | ComEC/Rec2-related protein                                                  |
|                                 | RF0021 |           | Similar to ComEC/Rec2 family protein                                         |
|                                 | RF0786 | traC      | Possible DNA primase (for the initiation of the replication of transferred DNA stands in the recipient cells) |
| **Macromolecule metabolism**    | 4 ORFs | dam       | Site-specific DNA adenine methylase (EC 2.1.1.72)                           |
|                                 | RF0123 | dcm       | Site-specific DNA methylase                                                 |
|                                 | RF0165 |           | Similar to superfamily I DNA and RNA helicases                              |
|                                 | RF0259 |           | Protein phosphatase                                                         |
|                                 | RF0335 | reIB      | DNA-damage-inducible protein J                                              |
|                                 | RF0359 |           | Site-specific recombinases (cassette chromosome recombinase B)              |
|                                 | RF0555 | rimM      | 16S rRNA-processing protein RimM                                            |
|                                 | RF0795 |           | rRNA methylase (partial)                                                    |
|                                 | RF0796 |           | rRNA methylase                                                             |
|                                 | RF0915 |           | Methylated-DNA–protein-cysteine methyltransferase (EC:2.1.1.63)            |
|                                 | RF1004 | hspC2     | Small heat-shock protein                                                    |
|                                 | RF1310 | radC      | DNA-repair protein (RadC)                                                   |
| **Small molecule metabolism**   | RF0036 |           | Pyroloquinoline quinone (coenzyme PQO) biosynthesis protein C              |
|                                 | RF0039 | folKP     | Folate synthesis bifunctional protein (EC 2.7.6.3) (PPPK) and (EC 2.5.1.15) (DHP5) |
|                                 | RF0078 | prs       | Ribose-phosphate pyrophosphokinase (EC:2.7.6.1)                            |
|                                 | RF0166 |           | Sugar kinases, ribokinase family                                            |
|                                 | RF0241 | manC      | Mannose-1-phosphate guanylyltransferase                                     |
|                                 | RF0374 | scoA      | Succinyl-CoA:3-ketoacid-coenzyme A transferase                              |
|                                 | RF0527 | bioB      | Biotin synthase (EC 2.8.1.6)                                                |
|                                 | RF0531 | dprA      | DNA-processing protein DprA, putative                                       |
|                                 | RF0597 |           | Alkylated DNA-repair protein                                                |
|                                 | RF0811 |           | Predicted aminomethyltransferase related to GcvT                            |
|                                 | RF0949 |           | Similar to predicted glutamine amidotransferases                           |
|                                 | RF0996 |           | D-alanyl-D-alanine dipeptidase                                               |
|                                 | RF0997 |           | Putative pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96) (PHS) (Pterin carbinolamine dehydratase) |
| **Transporters**                | 2 ORFs |           | ABC transporter, ATP-binding protein                                         |
|                                 | RF0322 |           | Transporter                                                                 |
|                                 | RF0643 |           | RND efflux system, outer-membrane protein                                   |
|                                 | RF0862 |           | Similar to amino acid permeases                                             |
|                                 | RF0970 |           | Na⁺/proline symporter, signal transduction histidine kinase                 |
| **Regulatory functions**        | RF1381 | nhaA      | Na⁺/H⁺ antiporter NhaA                                                      |
|                                 | RF0312 |           | Tryptophan-repressor-binding protein                                         |
|                                 | RF0537 |           | Transcriptional regulator, AbRß family                                       |
the sequence of the hydrolase domain, and the second group (SpoT11, 12, and 13) with the sequence of the synthetase domain. Being consistent with the previous observation [23], our phylogenetic analyses suggest that each paralogous gene group originated in early duplication events before the divergence of *Rickettsia* species. Notably, every sequenced *Rickettsia* genome encodes at least one ORF exhibiting hydrolase catalytic residues and one ORF exhibiting synthetase catalytic residues, suggesting that both hydrolase and synthetase functions are required for *Rickettsia*. We also found that seven *spoT* (spoT1–4 and 7–9) genes were located in the *R. felis* chromosome next to a gene encoding a transporter of the major facilitator superfamily (MFS) including proline/betaine transporters. MFS is also a large paralogous gene family composed of at least 23 ORF members in *R. felis*.

Toxin–antitoxin systems are composed of tightly linked toxin and antitoxin gene pairs and ensure stable plasmid inheritance when they are encoded in plasmids. In these systems, the toxic effect of a long-lived toxin is continuously inhibited by a short-lived antitoxin only when whole systems are maintained. The toxin–antitoxin modules have also been found on the chromosomes of many free-living prokaryotes, but have rarely been found in obligate intracellular bacteria [24,25]. In the *R. felis* chromosome, we identified 16 toxin genes (RF0016, RF0095, RF0271, RF0456, RF0490, RF0602, RF0701, RF0732, RF0787, RF0792, RF0898, RF0911, RF0956, RF1272, RF1286, and RF1368) and 14 antitoxin genes (RF0015, RF0094, RF0272, RF0457, RF0489, RF0601, RF0702, RF0731, RF0779, RF0788, RF0899, RF0910, RF0957, and RF1369), comprising at least 13 modules in operon structures. It is suggested that toxin–antitoxin systems, when encoded on the bacterial chromosome, might be involved in selective killing (a primitive form of bacterial apoptosis) or reversible stasis of bacterial subpopulations during periods of starvation or other stress [26,27]. It is also tempting to speculate that the toxin–antitoxin system could be targeted to the eukaryotic host cells. In this case, this system may help to maintain the presence of bacteria in the host. Notably, in the chromosomally encoded *mazEF* system of *E. coli*, the toxin action is regulated by (p)ppGpp. The large number of toxin–antitoxin modules in *R. felis*, as well as a number of *spoT* paralogs, might thus be linked to the synchronization of its multiplication within eukaryotic hosts.

It is probable that five *R. felis*–specific ORFs are related to its capacity of antibiotic resistance. We identified a streptomycin resistance protein homolog (RF0774), a class C β-lactamase, AmpC (RF1367), a class D β-lactamase (RF1275), a penicillin acylase homolog with conserved catalytic residues (RF1137), and an ABC-type multidrug transport-system protein, MdlB (RF0981). AmpC β-lactamase is known to be induced by AmpG of the MFS, which was also identified in the *R. felis* genome (RF0265, RF0608, RF0834, and RF1247). In vivo β-lactamase activity of *R. felis* was measured using high-performance liquid chromatography (see below).
Adaptation to Eukaryotic Hosts

*R. felis* may have developed a specific mechanism to cross-talk with its eukaryotic hosts. It exhibits 22 ankyrin-repeat-containing proteins and 11 TPR-containing proteins. These two protein motifs are frequently found in eukaryotic proteins, but their distributions are rather limited in viruses and bacteria, in both of which they appear to be linked with pathogenicity.

The ankyrin repeat is a protein–protein interaction motif, involved in transcription initiation, cell cycle regulation, cytoskeletal integrity, and cell-to-cell signaling [28]. *Anaplasma phagocytophilum*, a closely related intracellular *α*-proteobacterium, exhibits a protein containing ankyrin repeats (AnkA), which was detected in the cytoplasm and the nucleus of infected eukaryotic cells (human leukemia-60) [29]. According to the Superfamily database [30], only 15 bacterial species possess more than three ankyrin-repeat-containing proteins, and two species exhibiting the highest number of ankyrin repeats are obligate intracellular bacteria, *W. pipientis* (21 proteins) and *Coxiella burnetii* (20 proteins), although Wu et al. [14] reported slightly different numbers of ankyrin-repeat-containing proteins for these species. A recent genome analysis of a facultative intracellular bacterium, *L. pneumophila*, revealed 20 proteins with ankyrin repeats [31]. Ankyrin repeats were also found in more than 30 ORFs of the giant virus *Acanthamoeba polyphaga* Mimivirus [32].

TPR, composed of a motif of 34 amino acids organized in tandem, is also recruited by different proteins and facilitates protein–protein interactions [33]. Its role in the adaptation of parasites to their hosts has been suggested. The *R. felis* genome exhibits 11 TPR-containing ORFs (seven in the chromosome and four in the pRF plasmid). Only *Leptospira interrogans* (the agent of leptospirosis), *Treponema* species (including the agent of syphilis), and *L. pneumophila* [31] exhibit a high number of both TPR and ankyrin repeats. These organisms are eukaryotic parasites. The cryptococcal crooked neck 1 gene of *Cryptococcus neoformans* (a yeast), containing 16 copies of TPR, appears associated with its virulence [34].

Host Invasion/Pathogenesis

Plasmids often carry out functions that benefit bacteria in their survival or expression of virulence. pRF exhibits two ORFs that are possibly associated with the pathogenesis of *R. felis*: a hyaluronidase and a patatin-like protein. The hyaluronidase homolog (pRF56) exhibits a significant homology to hyaluronidase NagI (1,297 aa) of *Clostridium perfringens*. Hyaluronidases, which depolymerize hyaluronic acid—an unbranched polysaccharide ubiquitously present in the extracellular matrix of animal tissues—are known as “spreading factors” [35]. Another ORF (*pat2*) exhibits a significant homology to patatin-like phospholipases. Its paralog (*pat1*) was also identified in the chromosome, as already mentioned. Patatin is the major storage glycoprotein found in potato tubers, but also exhibits phospholipase A2 activity for protection from infection. Proteins containing patatin-like domains are more frequently found in pathogenic than in

---

### Table 3. *R. felis* ORFs Encoded in pRF Plasmid

| Category                          | ORFs | Gene Name | Annotation                              | Presence in pRF |
|----------------------------------|------|-----------|-----------------------------------------|-----------------|
| **Gene families**                | 7    | *trp*     | Transposase or inactivated derivative   | +               |
|                                  | 2    | *trp*     | Transposase or inactivated derivative   |                 |
|                                  | 2    | *ank*     | Ankyrin-repeat-containing gene          |                 |
|                                  | 3    | *tpr*     | TPR-containing gene                     |                 |
| **Conjugative transfer proteins**| prRF38/prRF39 | *traA*  | Conjugative transfer protein similar to *A. tumefaciens* Ti plasmid *TraA* |                 |
|                                  | prRF37 | *traD*  | Conjugative transfer protein similar to *A. tumefaciens* Ti plasmid *TraD* |                 |
|                                  | prRF44/prRF43 | *traD*  | Conjugative transfer protein similar to *E. coli* F plasmid *TraD* | +               |
|                                  | prRF47 | *traG*  | Conjugative transfer protein similar to *E. coli* F plasmid *TraG* | +               |
| **Partitioning and DNA metabolism** | prRF03 | *parA*  | Chromosome-partitioning ParA family protein |                 |
|                                  | prRF07 | *parB*  | Type I restriction-modification system, R (restriction) subunit |                 |
|                                  | prRF06 | *tmk*   | Thymidylate kinase                      |                 |
|                                  | prRF19 |         | Similar to chromosomal replication initiator protein *DnaA* |                 |
|                                  | prRF23 | *parA*  | Similar to *Pseudomonas syringae* plasmid *psrR1* stability protein ParA |                 |
|                                  | prRF34 | *dnaA*  | DNA polymerase III, epsilon-subunit-like protein |                 |
|                                  | prRF35 | *parB*  | ParB-like nuclease domain               |                 |
|                                  | prRF53 |         | DNA polymerase III, epsilon-subunit-like protein |                 |
|                                  | prRF66 |         | Similar to site-specific recombinases    |                 |
|                                  | prRF32 | *tnpR*  | TnpR resolvase (plasmid-encoded site-specific recombinase) |                 |
| **Host invasion**                | prRF11 | *pat2*  | Patatin-like phospholipase              | +               |
|                                  | prRF56 |         | Hyaluronidase                           |                 |
| **Other functions**              | prRF08 |         | Similar to CHE-like receiver domain    |                 |
|                                  | prRF22 |         | Similar to *P. syringae* plasmid *psrR1* ORF12 |                 |
|                                  | prRF25 | *sca12* | Cell-surface antigen homolog Sca12 (52 kDa) |                 |
|                                  | prRF26 | *lon*   | ATP-dependent protease La (TPR-containing) |                 |
|                                  | prRF49 |         | Similar to integrase                    |                 |
|                                  | prRF51 | *hsp2*  | Small heat-shock protein                 |                 |
|                                  | prRF52 | *hsp1*  | Small heat-shock protein                 |                 |
|                                  | 19    | ORFs    | Hypothetical proteins                   |                 |
|                                  | 9     | ORFs    | Hypothetical proteins                   |                 |

Ti-plasmid, tumor-inducing plasmid.
DOI: 10.1371/journal.pbio.0030248.t003
Table 4. *R. felis* ORFs Present in *R. typhi* but Absent or Degraded in *R. conorii* and *R. sibirica*

| ORF | Gene Name | Annotation |
|-----|-----------|------------|
| RF0096 | pta | Phosphate acetyltransferase (Pta) (EC 2.3.1.8) |
| RF0097 | ackA | Acetate kinase (EC 2.7.2.1) |
| RF0162 | phIC | Poly-beta-hydroxybutyrate polymerase |
| RF0163 | paoJ | Acetyl-CoA acetyltransferase (EC 2.3.1.9) |
| RF0183 | Hypothetical protein |
| RF0183 | Hypothetical protein |
| RF022 | Hypothetical protein |
| RF0228 | poly-beta-hydroxyalkanoate depolymerase |
| RF0257 | Hypothetical protein |
| RF0358 | bcr1 | Bicyclomycin resistance protein (MFS drug exporter) |
| RF0410 | spoT12 | Guanosine polyphosphate pyrophosphohydrolases/synthetases homolog |
| RF0526 | bioY | BioY family protein |
| RF0585 | Hypothetical protein |
| RF0693 | sac3 | Cell-surface antigen Sca3 (319 kDa) |
| RF0636 | rfaJ | Lipopolysaccharide 1,2-glucosyltransferase (RfaJ) |
| RF0800 |fadB | 3-hydroxyacyl-CoA dehydrogenase (FadB) |
| RF0994 | Hypothetical protein |
| RF1057 | atm1 | Multidrug resistance protein Atm1 |
| RF1271 | Stress-induced DNA-binding protein (Dps family) |
| RF1298 | phIC | Poly(3-hydroxyalkanoate) synthetase |
| RF1349 | proP6 | Proline/betaine transporter (MFS transporter) |

DOI: 10.1371/journal.pbio.0030248.g003

Figure 3. Visualization of *R. felis* Pili by Transmission Electron Microscopy

Bacteria collected from the supernatant of *R. felis*-infected XTC cells were negatively stained. (A) Sexual pilus observed between two bacteria. (B) *R. felis* also possesses small appendages likely to be fimbriae pili.

DOI: 10.1371/journal.pbio.0030248.g003

nonpathogenic bacteria. McLeod et al. [7] suggested that patatin-like proteins might be responsible for the phospholipase A2 activity identified some years ago in rickettsiae [36]. Potent host-invasion capacity is also provided by *R. felis*-specific ORFs found on the chromosome, for instance, a chitinase homolog (RF0413) and a chitin-binding protein homolog (RF0710). Chitin is a homopolymer of N-acetylglucosamine and a major component of the exoskeleton of arthropods and of the peritrophic envelope of insects, a lining layer of the midgut. These genes may facilitate the access of bacteria to the insects’ gut epithelial cells. *R. felis* may also use chitin as a nutrient source, as does *Vibrio cholerae* [37]. We identified a homolog (RF0268) for ecotin, an antifreeze protein and a chaperonin implicated in a neurodegenerative disease. Finally, *R. felis* exhibits an ortholog (RF0371) for *R. conorii* RickA, which induces its actin-based motility [40].

Phenotypic Post-Genomics Analysis

The obligate intracellular nature of *R. felis* hindered progress in the detailed characterization of its phenotypic diversity. Here, we envisaged post-genomics as a way of associating in vivo phenotypes of these bacteria to genomic features. The presence of pili-associated genes prompted us to investigate, by electron microscopy, the presence of such appendages on the cell surface. This approach led to the first characterization of pili on the surface of a *Rickettsia*; we observed two forms of pili at the surfaces of *R. felis* (Figure 3). One form of pili establishes direct contact between bacteria, providing a very typical figure of Mpf apparatus; these pili are probably specialized in conjugation. The other form of pili forms small hair-like projections emerging out from the cell surface; these pili are probably involved in the attachment of the bacteria to other cells. Without pili, many disease-causing bacteria lose their invasion capability. The latter type of pili might be considered as virulence factors, as described for *Francisella tularensis* [41,42]. As previously mentioned, we also found a RickA homolog in the *R. felis* genome [40]. Based on this finding, we performed immunofluorescence assays. The orientations of actin filaments beside bacteria are distinct from the stress fibers of the host. This further suggests that *R. felis* is probably capable of using the actin cytoskeleton to disseminate through eukaryotic cells, a method exploited by other SFG.

Table 4. *R. felis* ORFs Present in *R. typhi* but Absent or Degraded in *R. conorii* and *R. sibirica*
rickettsiae [40] (Figure S8). Another R. felis phenotypic character suggested from genomic analyses (three ORFs for patatin-like proteins) was its hemolytic capacity. We confirmed experimentally that R. felis lyases erythrocytes, this effect being inhibited by dithiothreitol. Another genome-guided discovery was β-lactam inhibition, which reached 57% and 53% of the concentration and the minimal inhibitory concentration, respectively, following 2 h incubation of R. felis with amoxicillin. Despite being preliminary results, these findings illustrate the fact that whole-genome sequencing offers opportunities to rapidly gain a better understanding of the phenotypic characters of a fastidious microorganism.

Discussion

R. felis is the first obligate intracellular bacterium exhibiting a possible conjugative plasmid. Of the nine previously published studies of members of the Order Rickettsiales (six ing a possible conjugative plasmid. Of the nine previously [44]. The possible conjugative plasmid identified in R. felis was the detection of repetitive mobile DNA elements. Many that ankyrin repeats and several TPRs were also found. It appears that ankyrin repeats and several TPRs were also found. Many such recently identified/cultured fastidious organisms, complete genome sequencing is a very potent and timesaving strategy to identify unrecognized phenotypic properties.

Materials and Methods

**Bacterial purification and DNA extraction.** R. felis (strain California 2) was cultivated on XTC cells growing on RPMI with 5% fetal bovine serum, supplemented with 5 mM L-glutamine. The purification of the bacteria was performed by different steps. First, the bacteria were treated in the presence of 1% trypsin in K36 buffer for 1 h at 37 °C, then centrifuged and digested by DNaseI for 1 h at 37 °C to reduce the euarkyotic DNA contamination. The sample was loaded on a renografin gradient and the bands of the purified bacteria were washed in K36, treated again by DNaseI. After inactivation with EDTA (50 mM), the bacteria were resuspended in TE, dispatched in 150-µl tubes and stored at −80 °C. Depending on this initial concentration, one or two tubes were diluted in 1 ml of TNE (10 mM Tris (pH 7.5), 150 mM NaCl, 2 mM EDTA) and incubated for 5 h at 37 °C in the presence of lysozyme (2 mg/ml). Lysis was performed for 2 h at 37 °C by adding 1% SDS and RNaseA (25 µg/ml). Overnight treatment with 1 mg/ml of proteinase K followed at 37 °C. After three phenol–chloroform extractions and alcoholic precipitation, the DNA was resuspended in 30 µl of TE and its concentration was estimated by agarose gel electrophoresis.

**Pulsed-field agarose gel electrophoresis.** The concentrated bacterial suspension was included in 1% (vol/vol) Incert agarose gel blocks (BMA, Rockland, Maryland, United States). The agarose blocks were digested by Proteinase K (1 mg/ml) (Eurobio Laboratories, Paris, France) in 1% laurylsarcosine and 0.5 M EDTA (pH 8) (Sigma-Aldrich, St. Louis, Missouri, United States) for 24 h at 50 °C. Fresh Proteinase K was then added and the incubation was continued for 24 h. The blocks were then washed twice in TE (pH 7.6) for 30 min at room temperature. Proteinase K inactivation was performed through incubation in a 4% phenylmethylsulfonyl fluoride (MBI Fermentas, Burlington, Canada) solution for 1 h at 50 °C. This inactivation step was carried out twice. The blocks were then washed two to three times in TE and stored in 0.5 M EDTA (pH 8) at 4 °C. Before restriction enzyme digestion, the agarose blocks were equilibrated twice with TE for 15 min. Digestion was carried out for 4 h, then fresh enzyme was added and the incubation was continued overnight. The digested agarose blocks and molecular-weight markers (Low Range PFG Marker, Lambda Ladder PFG Marker [New England Biolabs, Beverly, Massachusetts, United States]) were equilibrated in 0.5× TBE (50 mM Tris, 50 mMboric acid, 1 mM EDTA).

Each agarose block was laid in a 1% PFG agarose (Sigma-Aldrich) solution in 0.5× TBE. Pulsed-field gel electrophoresis was carried out on a CHEF-DR II device (Bio-Rad, Hercules, California, United States) under different electrophoresis conditions. The 1% agarose gel was run at 200 V using ramped pulse times from 1 to 5 s for 10 h to observe the pattern of small DNA fragments (2–48 kb). The migration was taking place under the following two consecutive conditions: (i) a ramping time from 3 to 10 s at 200 V for 12 h, with the pattern representative for 48- to 242-kb fragments, then (ii) a ramping time from 20 to 40 s at 180 V for 15 h, with the pattern representative for 145- to 610-kb fragments.

**Shotgun of R. felis genome and sequencing strategy.** Three shotgun genomic libraries were constructed by mechanical shearing of the genomic DNA using a Hydroshear device (GeneMachine, http://genome.nhgri.nih.gov/genemachine/). DNA fragments were blunt-ended using T4 DNA polymerase (New England Biolabs) and ligated to the BstXI adapter. Fragments of 3, 4.5, and 7 kb were separated on a preparative agarose gel (FMC BioProducts, Rockland, Maryland, United States), extracted with Qiaquick kit (Qiagen, Valencia, California, United States), and ligated into pCDNA2.1 (Invitrogen, Carlsbad, California, United States) for the two smaller inserts and into pCNS (a low copy number vector; C. R., unpublished data) for the largest one. DNA cloning was performed using electroporant E. coli DH10B Electromax cells (Invitrogen). Plasmid DNAs were purified and sequenced. 96 clones were analyzed by gel electrophoresis to validate the libraries. DNA sequencing of insert ends was carried out using Big Dye 3.1 terminator chemistry on an automated capillary ABI3700 sequencer (Applied Biosystems, Foster City, California, United States).
Sequences were analyzed and assembled into contigs using Phred, Phrap, and Consed software [51] taking all sequences into account. Sequences were considered valid when at least 75% of the nucleotides had a Phred score of more than 20. The finishing of the genome sequencing included only additional directed reactions that were performed on an ABI3100 sequencer. Two circular plasmid molecules of 59 and 38 kb were sequenced, respectively, in 5‘-3‘ direction and 3‘-5‘ direction for each sequence. On the chromosome, three small regions of 41, 155, and 64 bp failed by dropping of sequence. A number of parameters (DMSO, glycerol, hybridization, and elongation temperature) were tested one by one or were combined to sequence over these gaps. We finally succeeded with the association of another type of chemistry, D- rhodamine with 2 M betaine. We designed and used 420 primers (i) to close the sequencing gaps by walking either on shotgun subclones or on the chromosome and (ii) to improve sequence regions of low quality.

The integrity of the assembly was validated by comparing the restriction patterns obtained by pulsed-field electrophoresis with those deduced from the electronic consensus sequence. The selection of restriction enzymes was based on rare sites. We analyzed single digests of R. felis DNA. The main restriction enzymes used for these studies were Apal, AfeI, FspI, and SphI. This comparative study confirmed the predicted length of the R. felis DNA fragments.

The structures for pRF and pRFβ plasmids were controlled by specific primer amplifications (see Figure S1). Three PCRs were performed and the amplification results were in agreement with the expected length of these PCR products. These PCR products were cloned into pGEM-T vector (Promega, Madison, Wisconsin, United States) and sequenced on both strands using the T7 primer and specific primers (see Table S1).

Repeated DNA sequences

Repeated DNA sequences were identified with the use of RepeatFinder [58], by ignoring the repeats identified in the pRF plasmids using tRNAscan-SE [53]. Database searches were performed using BLAST programs [54] against Swiss-Prot/TrEMBL [55], the NCBI CDD database [60], and Low Range FPF Marker (New England Biolabs) as described above, with a pulse time from 1 s to 5 s for 12 h at 180 V. The gel was treated and transferred onto Hybond N+ (Amersham Biosciences, Little Chalfont, United Kingdom) with a vacuum blot. The DNA on the filter was fixed by heating for 2 h at 80 °C, and the membranes were cut into two pieces. Two probes were derived from two PCR products. The first, pRF-pRFβ (726 bp), was derived within the pRF-specific insert, and the second, pRFα-pRFγ (251 bp), was designed to encompass the deletion site of the pRFβ. These two probes were labeled with dCTP32 (Amersham Biosciences, Little Chalfont, United Kingdom) with a vacuum blot. The DNA on the filter was fixed by heating for 2 h at 80 °C, and the membranes were cut into two pieces. Two probes were derived from two PCR products. The first, pRF-pRFβ (726 bp), was designed within the pRF-specific insert, and the second, pRFα-pRFγ (251 bp), was designed to encompass the deletion site of the pRFβ. These two probes were labeled with dCTP32 and hybridized at 65 °C for 17 h on each membrane. Membranes were washed three times with 1× SSC, 0.1% SDS at 65 °C. The exposure time ranged from 6 h to overnight at −80 °C in ECL film. The hybridizations were clearly established on R. felis digested by PvuI and led to one signal with the pRFβ-pRFβ probe and two signals for the two plasmid structures with the pRFα-pRFγ probe at a predicted molecular weight compatible with our prediction (see Figure S2).

We tested 30 samples of fleas naturally infected by R. felis obtained from different geographic areas (Algeria [11 fleas], France [15 fleas], and New Zealand [four fleas]) with three pairs of primers: (i) primers designed in the class I ORF gene (pRFα, pRFβ, pRFγ), (ii) primers in the pRF plasmid (pRFα-pRFβ, pRFβ-pRFγ), and (iii) primers in the pRF plasmid (pRFα-pRFγ). We confirmed positive PCR products of (i) 196 bp, (ii) 208 bp, and (iii) 251 bp for all the 30 cases.

**Annotation.** We predicted protein-coding genes (ORFs) using SeqFit [52] as previously described [8]. tRNA genes were identified using tRNAscan-SE [53]. Database searches were performed using BLAST programs [54] against Swiss-Prot/TrEMBL [55], the NCBI CDD database [56], and SMART [57]. The number of transposases, ankyrin/TPR-containing genes, autotransporter domains, and integrases were computed using TIGR software (http://www.tigr.org) with the TIGRFAMs database. The total number of inversions to associate a pair of Rickettsia genomes using GRAPPA release 2.0 [65].

**Ultrastructural characterization of pili by electronic microscopy.** R. felis cells were carefully collected from the supernatant of XTC cells infected for 5 d and grown at 28 °C. Following centrifugation (400 g, 10 min), bacteria were fixed for 1 h at 4 °C in glutaraldehyde (2.5% in phosphate-buffered saline [PBS]). Cells were then washed in PBS and placed on a carbon–formvar-coated 400-mesh copper grid (Electron Microscopy Sciences, Hatfield, Pennsylvania, United States) for 15 min. Cells were then negatively stained with 2% phosphotungstic acid for 10 s, before analysis by electron microscopy (Philips Morgagni 268D, Philips Electronics, Eindhoven, the Netherlands).

**Estimation of β-lactamase activity.** To evaluate the level of β-lactamase activity, 105 R. felis cells grown on XTC cells and then sonicated were mixed with amoxicillin to a final concentration of 20 µg/ml and incubated for 2 h at 28 °C. The concentration of amoxicillin was measured in the R. felis + amoxicillin suspension as well as in a suspension of XTC cells without bacteria + amoxicillin, as determined by high-performance liquid chromatography. In addition, the minimum inhibitory concentrations of these four suspensions were estimated by growth inhibition of a Micrococcus luteus strain.

**RNA extraction and RT-PCR.** Approximately 6.5 × 106 bacteria were used to infect one 25-cm2 flask of confluent XTC cells maintained at 28 °C. The infected cells were harvested 48 h later, centrifuged (12,000 g, 10 min), and pellets were immediately frozen in liquid nitrogen before being stored at −80 °C. Total RNA was isolated by using the RNasy Mini Kit (Qiagen) according to the manufacturer’s instructions. At the end of the extraction procedure, all samples were treated with RNase-Free DNase Set (Qiagen) for 30 min. The concentration and quality of isolated RNA were determined with the Agilent 2100 bioanalyzer (Agilent Technologies, Edenwoood, New Jersey, United States). Aliquots of the DNase-treated total RNA samples were stored at −80 °C until use. RT-PCR was performed from 2 µl of RNA (25 µl final reaction volume) with the Superscript One-Step RT-PCR with Platinum Taq (Invitrogen). Possible DNA contamination was assessed with the Expand high-fidelity polymerase (Roche, Basel, Switzerland). Cycling conditions were 30 min at 50 °C, 5 min at 95 °C, and 40 cycles at 30 s at 95 °C, 30 s at 50 °C, and 1 min at 72 °C, followed by a final extension cycle of 7 min at 72 °C. The RT-PCRs were conducted on the PTC-100 thermocycler (Bio-Rad). Amplification products were run on 2% (wt/vol) agarose gels, and the DNA was stained with ethidium bromide. The size of the PCR product was determined by comparison with DNA molecular-weight marker VI (Boehringer, Ingelheim, Germany).

**Detection of F-actin and immunofluorescence staining.** Vero cells grown to confluence on glass coverslips were infected with R. felis for 24–48 h at 28 °C in a humidified CO2 incubator (5% CO2). Infected cells were then fixed for 1 h at 4 °C with formaldehyde (3% wt/vol in PBS supplemented with 1 mM MgCl2 and 1 mM CaCl2) washed three times in PBS, and then made permeable with 0.2% Triton X-100 in PBS for 1 min. After three washings in PBS, the coverslips were incubated for 1 h with a monoclonal anti-R. felis antibody. Bacteria were visualized by staining with anti-mouse-Alexa (555) (Molecular Probes, Eugene, Oregon, United States) and FITC-phalloidin (1:250) (Sigma). The coverslips were mounted using Fluorrep (BioMérieux, Marcy-Étoile, France) and were examined with a confocal laser scanning microscope using a 100× oil immersion objective lens.

**Hemolysis experiments.** Human blood (10 ml) was centrifuged (1,500 g, 10 min), and after removal of the plasma washing erythrocytes were resuspended in 20 ml of PBS. This suspension (100 µl) was mixed with 800 µl of PBS and 100 µl of rickettsial suspension (106, 105, and 104 bacteria, respectively). In some experiments, rickettsiae were incubated for 1 h at 35 °C in the presence of 2 mM DTT. Complete hemolysis was assumed when the optical density at 545 nm was equal to 1.0. Following 3 h of incubation at 35 °C, the samples were fixed using paraformaldehyde (0.5% final concentration) and centrifuged. Hemoglobin release was estimated by measurement of the optical density at 545 nm. This experiment was performed in duplicate.

**Primers.** The sequences of the primers for PCR and RT-PCR are provided in Table S3.
Supporting Information

Figure S1. Confirmation of Plasmid Topologies for prF and prF6 by PCR
(A) The locations of the three primer sets (pRFA–pRFb, pRFc–pRFd, and pRFa–pRFf) used to validate the presence of the two distinct plasmid forms are indicated.
(B) The result of the PCR assay with these primers. Two pairs of primers (pRFa–pRFg and pRFh–pRFi) were used to obtain the probes for the Southern blot (see Figure S2), as well as another pair of primers (pRF37F1/R1) used in plasmid detection in fleas infected by R. felis, as indicated in (A).

Found at DOI: 10.1371/journal.pbio.0030248.sg001 (1.5 MB TIF).

Figure S2. Characterization of R. felis Plasmids by Southern Blot
The two membranes (A and B) loaded with R. felis genomic DNA (#1/4) and R. felis DNA digested by PvuI (#2/3) were hybridized either by the probe pRFa–pRFg or by pRFh–pRFi.

Found at DOI: 10.1371/journal.pbio.0030248.sg002 (39 KB PDF).

Figure S3. A Model for the Conjugative Plasmid Transfer of R. felis
This model is based on gram-negative bacterial conjugation systems involving T4SS. Homologs responsible for different steps of conjugation were identified in the R. felis genome. DNA-processing machinery, plasmid-encoded pRF38/39 (TraATi); the coupling protein, plasmid-encoded pRF43/44 (TraDF) or chromosomally encoded (pRF37F1/R1) used in plasmid detection in fleas infected by R. felis, are also indicated in (A).

Found at DOI: 10.1371/journal.pbio.0030248.sg003 (57 KB PDF).

Figure S4. Phylogenetic Trees for the Pataxin-Like Proteins, Thymidylate Kinases, and Small Heat-Shock Proteins
Phylogenetic trees were constructed using the neighbor-joining method with Jones-Taylor-Thornton model.

Found at DOI: 10.1371/journal.pbio.0030248.sg004 (16 KB PDF).

Figure S5. Taxonomic Distribution of BLAST Best Hits of R. felis ORFs
R. felis ORFs were searched against the nonredundant database (excluding rickettsial sequences). The distribution difference between ORFs with rickettsial orthologs and ORFs lacking rickettsial orthologs remained significant even after the removal of transposase ORFs.

Found at DOI: 10.1371/journal.pbio.0030248.sg005 (16 KB PDF).

Figure S6. Cross-Reactivity of R. felis
Western immunoblot showing the preferential cross-reactivity of antibodies with R. felis and R. typhi in a patient with murine typhus (lanes a and b), with R. felis and R. conorii in a patient with Mediterranean spotted fever (lanes d–f), Lanes a and d, R. conorii antigen; lanes b and e, R. typhi antigen; lanes c and f, R. felis antigen; MM, molecular mass.

Found at DOI: 10.1371/journal.pbio.0030248.sg006 (30 KB PDF).

Figure S7. Domain Structures and the Presence/Absence Patterns of sptT Genes in Different Rickettsia
With reference to the S. dysgalactiae Rel_008 (H: 53H, 77H, 78D, and 144D) and five (241R, 243K, 251K, 264D, and 323E) catalytic residues were examined for the (p)ppGpp hydrolase and synthetase domains, respectively. ORF sizes were those for R. felis genes, except SpoT14, for which the R. prowazekii ORF size is indicated. a, absent; Ch, conserved hydrolase catalytic residues; Cs, conserved synthetase catalytic residues; s, split or fragmented genes.

Found at DOI: 10.1371/journal.pbio.0030248.sg007 (17 KB PDF).

Figure S8. Confocal Laser Analysis of R. felis-Infected Vero Cells
Bacteria were stained by indirect immunofluorescence using a monoclonal anti-R. felis antibody followed by an anti-mouse-Alexa 594 antibody (red). F-actin was stained with FITC-phalloidin (green). Arrows indicate R. felis with actin tails.

Found at DOI: 10.1371/journal.pbio.0030248.sg008 (29 KB PDF).

Table S1. Distribution of ssa Genes among Rickettsia Genomes
Found at DOI: 10.1371/journal.pbio.0030248.st001 (30 KB DOC).

Table S2. Comparison of Different Features of Bacteria Infecting Fleas with Their Close Relatives
Found at DOI: 10.1371/journal.pbio.0030248.st002 (32 KB DOC).

Accession Numbers
The genome sequence of R. felis is accessible via GenBank (http://www.ncbi.nlm.nih.gov/genbank) under the accession numbers: CP000053, CP000054, and CP000055. The EMBL Nucleotide Sequence Database (http://www.ebi.ac.uk/emb/ accession number for Cl. perfringens; Q9XM99). The Pfam (http://www.sanger.ac.uk/ Software/Pfam/) accession number for ComE3 is PF03772. The Protein Data Bank (http://www.rcsb.org/pdb/) accession number for S. dysgalactiae Rel_008 is 1VJ7.

Acknowledgments
We wish to thank Jean Weissenhach and his team at Genoscope for the shotgun sequencing, Bernadette Giannelli and Thi Tien N. Gyun for technical assistance in sequencing the R. felis genome, Bernard Campagna for electronic microscopy analysis, Sylvianne Robineau for RT-PCR experiments, Claude Nappaz for production of monoclonal antibodies against R. felis, and Deborah Byrne for carefully reading the manuscript. We gladly acknowledge the financial support of Marie-Ville-Nice Genopole.

Competing interests.
The authors have declared that no competing interests exist.

Author contributions.
HO performed genome annotation and coordinated the informatics analyses. PR performed electron microscopic analysis of pili, immunofluorescence assay of actin-based motility, and hemolytic activity assay. SA and CR assembled genome sequences, performed sequence finishing, and carried out experiments to characterize the plasmids. CR constructed genomic libraries. GB performed most phyllogenetic analyses, PEF performed β-lactamase activity assay and contributed to the bioinformatics analyses. HP contributed to the sequencing. JMC provided (computing) support and supplied ideas. DR provided laboratory (experimental) support, supplied ideas, and coordinated the experimental aspects of the work. All authors contributed in drafting the manuscript.

References
1. Adams JR, Schmidtman ET, Azad AF (1999) Infection of colonized cat fleas, Ctenocephalides felis (Bouche), with a Rickettsia-like microorganism. Am J Trop Med Hyg 43: 400–409.
2. Azad AF, Sacci JB Jr, Nelson WM, Dasch GA, Schmidtman ET, et al. (1999) Genetic characterization and transovarial transmission of a typhus-like Rickettsia found in cat fleas. Proc Natl Acad Sci U S A 89: 43–46.
3. Raoult D, La Scola B, Enea M, Fournier PE, Ross V, et al. (2001) A flea-associated Rickettsia pathogenic for humans. Emerg Infect Dis 7: 75–81.
4. Schriefer ME, Sacci JB Jr, Dumler JS, Bullen MG, Azad AF (1994) Identification of a novel rickettsial infection in a patient diagnosed with murine typhus. J Clin Microbiol 32: 949–954.
5. Rolain JM, Franc M, Davoust B, Raoult D (2003) Molecular detection of Bartonella quintana, B. henselae, B. clavuligera, Rickettsia felis, and Wolbachia pipientis in cat fleas. Emerg Infect Dis 9: 538–542.
6. Anderson SG, Zornompidour A, Andersson JO, Sicheritz-Ponten T, Alsmark UC, et al. (1998) The genome sequence of Rickettsia prowazekii and the origin of mitochondria. Nature 396: 133–140.
7. McLeod MP, Qin X, Kazubaty SE, Gioia J, Highlander SK, et al. (2004) Complete genome sequence of Rickettsia typhi and comparison with sequences of other rickettsiae. J Bacteriol 186: 5842–5855.
8. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, et al. (2001) Complete genome sequence of R. prowazekii and comparison with other rickettsiae. J Bacteriol 183: 5842–5855.
9. Malek JA, Wierzbowski JM, Tao W, Bosak SA, Saranga DJ, et al. (2004) Protein interaction mapping on a functional shotgun sequence of Rickettsia sibirica. Nucleic Acids Res 32: 1059–1064.
10. Alt-Morbe J, Stryker JL, Fuqua C, Li PL, Farrand SK, et al. (1996) The conjugal transfer system of Agrobacterium tumefaciens octopine-type Ti plasmids is closely related to the transfer system of an IncP plasmid and distantly related to Ti plasmid vir genes. J Bacteriol 178: 4248–4257.

11. Beranek A, Zettl M, Lorenzoni K, Schauer A, Manhart M, et al. (2004) Thirty-eight C-terminal amino acids of the coupling protein TraD of the F-like conjugative resistance plasmid R1 are required and sufficient to confer binding to the substrate selector protein TraM. J Bacteriol 186: 7690–7706.

12. Frost LS, Ippen-Ihler K, Skurray RA (1994) Analysis of the sequence and functional domains of the chromosomal conjugal DNA transfer system. Microbiol Rev 58: 162–210.

13. Cascales E, Christie PJ (2003) The versatile type IV secretion systems. Nat Rev Microbiol 1: 137–149.

14. Wu M, Sun LV, Mamataeva J, Reimer M, Manhart M, et al. (2004) Thirty-eight C-terminal amino acids of the coupling protein TraD of the F-like conjugative resistance plasmid R1 are required and sufficient to confer binding to the substrate selector protein TraM. J Bacteriol 186: 7690–7706.

15. Horn M, Collingro A, Schmitz-Esser S, Beier CL, Purkhold U, et al. (2004) Illuminating the evolutionary history of chlamydiae. Science 304: 728–730.

16. Ogata H, Audic S, Barve V, Artiguenave F, Fournier PE, et al. (2000) Selfish DNA: The complex organization of mobile genetic elements in the genome of the unicellular eukaryote Tetrahymena thermophila. Science 290: 91–96.

17. Raoult D, Roux V (1997) Rickettsiotes as paradigms of new or emerging infectious diseases. Clin Microbiol Rev 10: 694–719.

18. Uchiyama T (2003) Adherence to and invasion of Vero cells by recombinant Rickettsia conorii expressing the outer membrane protein ompB. Int J Syst Evol Microbiol 50: 1449–1455.

19. Roux V, Raoult D (2000) Phylogenetic analysis of members of the genus Rickettsia using the gene encoding the outer-membrane protein ompB. J Bacteriol 182: 1449–1455.

20. Fang R, Raoult D (2003) Antigenic classification of Rickettsia felis by using monoclonal and polyclonal antibodies. Clin Diagn Lab Immunol 10: 221–228.

21. Romero C, Renesto P, Crapolet N, Matsumoto K, Parola P, et al. (2005) Thirty-eight C-terminal amino acids of the coupling protein TraD of the F-like conjugative resistance plasmid R1 are required and sufficient to confer binding to the substrate selector protein TraM. J Bacteriol 186: 7690–7706.

22. Pandey DP, Gerdes K (2005) Toxin-antitoxin loci are highly abundant in genomes of prokaryotes and eukaryotes. Trends Genet 21: 16029–16034.

23. Silverman DJ, Santucci LA, Meyers N, Sekeyova Z (1992) Penetration of host cells by Rickettsia rickettsii appears to be mediated by a phospholipase of rickettsial origin. Infect Immun 60: 2775–2783.

24. Silverman DJ, Santucci LA, Meyers N, Sekeyova Z (1992) Penetration of host cells by Rickettsia rickettsii appears to be mediated by a phospholipase of rickettsial origin. Infect Immun 60: 2775–2783.

25. Chain PS, Barnell CM, Gerdes K (2005) Prophage-whipple twist: A human pathogenic actinobacteria with a reduced genome. Genome Res 15: 1380–1381.

26. Bentley SD, Maitland M, Murphy LD, Paulen MJ, Yeats CA, et al. (2005) Sequencing and analysis of the genome of the Whipple's disease bacterium Tropheryma whippelii. Lancet 361: 637–644.

27. Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8: 175–185.

28. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, et al. (2004) UniProt: The Universal Protein knowledgebase. Nucleic Acids Res 32: 276–277.

29. Ponting CP, Schultz J, Milpetz F, Bork P (1999) SMART: Identification and functional annotation of proteins for known and predicted mammalian proteins. Genome Res 9: 539–546.

30. Notredame C, Higgins DG, Heringa J (2000) T-Coffee: A novel method for multiple sequence alignment. J Mol Biol 302: 205–217.

31. Lawrence JG, Ochman H (1998) Molecular archaeology of the Escherichia coli genome. Proc Natl Acad Sci U S A 95: 1344–1350.