Mechanisms of Evolution in *Rickettsia conorii* and *R. prowazekii*

Hiroyuki Ogata,[1] Stéphane Audic,[1] Patricia Renesto-Audiffren,[2] Pierre-Edouard Fournier,[2] Valérie Barbe,[3] Delphine Samson,[3] Véronique Roux,[2] Pascale Cossart,[4] Jean Weissenbach,[3] Jean-Michel Claverie,[1]* Didier Raoult[2]*

Rickettsia conorii is an obligate intracellular bacterium that causes Mediterranean spotted fever in humans. We determined the 1,268,755-nucleotide complete genome sequence of *R. conorii*, containing 1374 open reading frames. This genome exhibits 804 of the 834 genes of the previously determined *R. prowazekii* genome plus 552 supplementary open reading frames and a 10-fold increase in the number of repetitive elements. Despite these differences, the two genomes exhibit a nearly perfect colinearity that allowed the clear identification of different stages of gene alterations with gene remnants and 37 genes split in 105 fragments, of which 59 are transcribed. A 38-kilobase sequence inversion was dated shortly after the divergence of the genus.

Rickettsia species live in different ecological niches inside different arthropod hosts (insects or ticks), in which most of them are transmitted vertically from the mother to the progeny (*1*). *R. conorii* naturally infects the dog brown tick *Rhipicephalus sanguineus*. When transmitted to humans through tick bites, the bacterium causes Mediterranean spotted fever (*1*, *2*). *R. conorii* is closely related to the previously sequenced *R. prowazekii* (*3*), the agent of louse-borne typhus. We determined the complete sequence of the *R. conorii* genome (GenBank accession number AE006914) (Table 1) (*4, 5*). Comparative analysis of these two closely related *Rickettsia* sp. (Table 2) provides snapshots of the progression of the gene degradation process, which has been linked to adaptation to intracellular parasitism (*3*, *6–9*).

The overall gene order in the *R. conorii* genome [Figs. 1 and 2 and Web fig. 1 (*10*)] is remarkably similar to that of *R. prowazekii*, except for the translocation/inversion of a few short segments in the region corresponding to the end of replication. The detailed sequence comparison of the two genome sequences revealed numerous cases of apparently orthologous pairs of open reading frames (ORFs) (according to the best reciprocal match criteria) exhibiting large differences in sizes. For instance, the *phoR* gene encoding a 921-residue protein in *R. prowazekii* becomes a set of three consecutive ORFs (RC0702 through RC0704) of 643, 132, and 82 residues in *R. conorii*. Similarly, the *sca1* gene encoding a 1902-residue protein in *R. conorii* corresponds to three consecutive ORFs (RP016 through RP018) in *R. prowazekii* (Fig. 3). In addition to this "gene splitting" phenomenon (*11*), we also identified many additional *R. conorii* genes (229) and fewer *R. prowazekii* genes (6), for which a residual similarity could be found in homologous but noncoding regions in the other genome (*12*), thus representing cases of "decaying orthologs" (Fig. 2). Finally, 323 *R. conorii* ORFs exhibited no orthologous relationship (regular, split, or decaying), whereas 24 *R. prowazekii* ORFs had no equivalent in *R. conorii*. In total, 552 *R. conorii* genes have no orthologous functional counterpart in *R. prowazekii*, whereas 30 *R. prowazekii* genes have no counterpart in *R. conorii* (Table 2). Those genes are most likely to be responsible for the phenotypic differences between the two species.

Out of the 552 ORFs (reduced to 514 when every set of multiple ORFs arising from split genes is counted as one) constituting the gene excess in *R. conorii,* 106 ORFs (reduced to 79 as before) were assigned a putative function. These supplementary genes are overrepresented ($P < 0.05$, χ^2 test) as compared with those of *R. prowazekii* in the categories of DNA replication, transporters, regulatory functions, and drug sensitivity (Table 2). *R. conorii* has three genes (RC0843-4, RC0017, and RC0450) related to DNA transformation functions, including the DNA uptake protein ComF, the competence operon protein ComE3, and the chromosomal transformation protein Smf. The presence of such a DNA transformation gene has not been previously described for other obligate intracellular parasites [*R. prowazekii* (*3*), *Chlamydia* spp. (*13*), and *Mycobacterium leprae* (*9*)]. *R. conorii*'s capability for exogenous DNA uptake is further suggested by the presence of four ORFs of apparently foreign origin: one phage-related protein (RC0490), one insertion element (RC0688), and two lysozymelike proteins from viruses (RC0727 and RC1298). Both *Rickettsia* species are naturally resistant to penicillin and aminoglycoside antibiotics, and *R. conorii* exhibits higher resistance than *R. prowazekii* to antibiotics (*14*). Consistently, its genome contains nine additional genes related to its sensitivity to drugs, including four genes for β-lactamases (RC1243-4) and its regulation (RC0535, RC0788, and RC1358); three drug efflux transporters (RC0301, RC0564-9, and RC1181); an aminoglycoside 3′-phosphotransferase (RC0947); and an acetyltransferase (RC0554). *R. conorii* is known to move around inside host cells by propulsion produced by continuous actin polymerization (*15*). No clear homolog of proteins known to be responsible for the actin-based motility of *Listeria monocytogenes* (ActA) or *Shigella flexneri* (IcsA) (*16*) was found, but an ORF (RC0909) coding for a 520-residue–long protein exhibits an overall organization similar to that of ActA. Both proteins share a highly charged NH_2-terminus (~300 residues) and a central proline-rich region. RC0909 has a weak similarity to the WASP homology domain 2, found in a family of proteins known to regulate the formation of the actin filaments.

As intracellular parasites, *Rickettsia* have small genomes and an evolutionary tendency toward further genomic reduction (*6*). Therefore, genes found as multiple copies may outline their specific adaptations. Using BLAST (*17*), we identified six gene families with more than three paralogs. Comparing their copy numbers with those in other bacterial genomes, we found that five gene families (Tlc, SpoT, ProP, Sca, and AmpG) were

Table 1. Comparison between *R. conorii* and *R. prowazekii*.

Species	*R. conorii*	*R. prowazekii*
Genome size (bp)	1,268,755	1,111,523
G+C content (%)	32.4	29.0
ORFs	1374	834
tRNAs	33	33
rRNAs	3	3
Other RNAs	2	2
Pseudogenes	2	12
Coding content	81%	76%

[1]Information Génétique & Structurale, CNRS–AVENTIS UMR 1889, 31 chemin Joseph Aiguier, 13402 Marseille Cedex 20, France. [2]Unité des Rickettsies, Université de la Méditerranée, Faculté de Médecine, CNRS UMR 6020, 27 boulevard Jean Moulin, 13385 Marseille Cedex 05, France. [3]Genoscope and CNRS UMR 8030, 2 rue Gaston Crémieux, 91000 Evry, France. [4]Institut Pasteur, Unité des Interactions Bactéries-Cellules, 28 rue du Docteur Roux, 75724 Paris, Cedex 15, France.

*To whom correspondence should be addressed. E-mail: Jean-Michel.Claverie@igs.cnrs-mrs.fr; Didier.Raoult@medecine.univ-mrs.fr

significantly overrepresented ($P < 0.05$, Fisher's exact test) in *Rickettsia* species [Web table 1 (*10*)]. Adenosine triphosphate (ATP)/adenosine diphosphate (ADP) translocases are known to be unique to *Rickettsia* spp. and *Chlamydia* spp. among bacteria and may be of plant origin (*18*). This gene allows the importation of ATP from the infected host cell. Five copies were found in *R. conorii* and *R. prowazekii*. Also, four SpoT copies were found in *R. conorii* and *R. prowazekii* (*19*). The SpoT protein hydrolyzes the nucleotide (p)ppGpp, also known as alarmone. This compound plays a major role in processes related to starvation in various bacteria (*20*). Alarmone also initiates the expression of virulence genes in *Legionella pneumophila* (*21*), the production of antibiotics in *Streptomyces coelicolor* (*22*), and the change in the cell density of *Myxococcus xanthus* (*23*). The four copies of SpoT may be related to the adaptation of *Rickettsia* to long starvation periods in pausing ticks or louse feces. Both *Rickettsia* species have significantly large numbers of ProP proline/betain transporter paralogs: 11 and 7 ORFs for *R. conorii* and *R. prowazekii*, respectively. In many organisms, including bacteria and plants, proline transporters play critical roles in the response to osmotic changes in the environment (*24*). *Leishmania donovani*, an intracellular protozoan parasitizing both arthropod and mammalian cells, uses temperature-regulated proline transporters to adapt to different host temperature conditions (*25*). The numerous ProP paralogs in *Rickettsia* might be linked to their adaptation to osmotic stress or to the temperature-dependent regulation of their virulence known as "reactivation" (*26*, *27*). *R. conorii* harbors five genes (four in *R. prowazekii*) encoding for outer membrane proteins of the Sca family, including rOmpA, which accounts for antigenic differences between *Rickettsia* species. *R. conorii* exhibits four copies of AmpG (three in *R. prowazekii*) that are likely to contribute to its natural resistance to β-lactam antibiotics. Finally, genes for the ATP-binding protein of multidrug resistance ABC transporters are present as four copies in *R. conorii* (three in *R. prowazekii*), but these numbers are not significantly higher than in other bacteria.

The genome of *R. conorii* exhibits a much higher density of interspersed repetitive DNA than that of *R. prowazekii* (Fig. 2). In the *R. conorii* genome, we identified 10 families (656 elements) of repeated DNA (*28*, *29*). Those repeats vary in size between 19 and 172 base pairs (bp) and constitute 3.2% of the entire genome. Overall, the repeat fraction of the genome is G+C rich (40%) and is in part responsible for the higher G+C content of *R. conorii* as compared with *R. prowazekii*. The distribution of the repeated elements is essentially random throughout the genome (Fig. 1). The quasiperfect colinearity maintained between the two *Rickettsia* genomes contrasts with the view that the multiplication of interspersed repeats promotes genomic rearrangements (*30*, *31*).

The analysis of the *R. conorii* putative coding regions revealed numerous cases of consecutive ORFs matching consecutive segments of a single longer ORF in other species, including *R. prowazekii.* Such gene fragmentations (e.g., internal stop codons) are usually associated with "pseudogenes." However, a truncated form of the outer-membrane protein (rOmpA) is normally expressed in *R. felis* (*32*). In addition, most of the split ORFs retained the statistical properties (coding potential and codon bias) of normal coding regions and a good similarity with intact protein orthologs. This prompted us to anno-

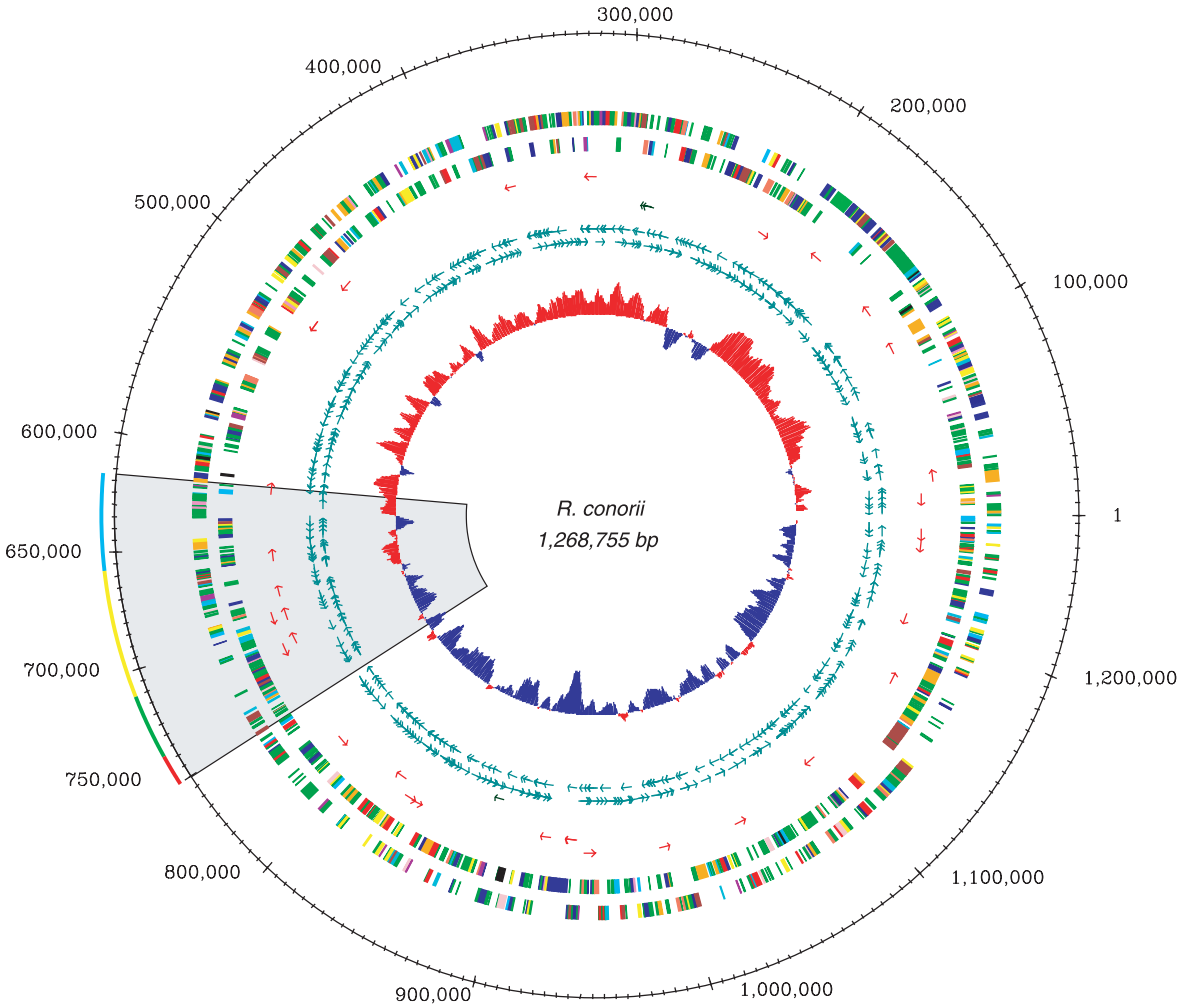

Fig. 1. Circular representation of the *R. conorii* genome (strain Malish 7). The outermost circle indicates the nucleotide positions. The second and third circles locate the ORFs on the plus and minus strands, respectively. Function categories are color-coded [see Web fig. 1 (*10*)]. The fourth and fifth circles locate tRNAs. The locations of three rRNAs are indicated by black arrows. The sixth and seventh circles indicate the locations of repeats. The eighth circle shows the G-C skew (G-C/G+C) with a window size of 10 kb. The region locally breaking the genome colinearity with *R. prowazekii* is indicated by a shaded sector. The four major genomic segments involved in this rearrangement are colored in blue, yellow, green, and red [see Fig. 3 and Web fig. 1 (*10*) for details].

tate these altered genes by the more neutral designation of "split genes," pending further experimental evidence. Thirty-seven split genes (*11*) (resulting in 105 total ORFs) were identified in *R. conorii* (Table 2). Among them, 14 have intact orthologs in *R. prowazekii*, 4 exhibit intact paralogs in *R. conorii*, and 19 have intact homologs in other prokaryotes. In *R. prowazekii*, we identified 11 split genes (resulting into 23 ORFs) that all have intact orthologs in *R. conorii*. By reverse transcriptase–polymerase chain reaction (RT-PCR) (*33*), we examined the detailed transcription pattern of all 37 *R. conorii* split genes [Web table 2 (*10*)]. We observed at least one transcript for 30 of 37 genes and RT-PCR products (all of the expected size) for 59 of 105 ORFs. All ORFs were transcribed for 11 of 31 genes, and the sole 5′ ORF was transcribed for 8 genes. These cases are consistent with the continued usage of the promoter of the original gene. However, the RT-PCR results on the other split genes suggested more complex transcription patterns. In seven cases, only the 3′ ORFs were found to be transcribed; in four cases, the 5′ and 3′ segments were found to be transcribed but not the middle segments; and in one case, only the middle segment was detected. Transcripts were much more likely to be detected for larger ORFs (≥70 residues) than for smaller ORFs (<70 residues) (55 of 78 versus 4 of 27; $P < 10^{-6}$, Fisher's exact test). These results suggest that these split genes might have retained some of their original

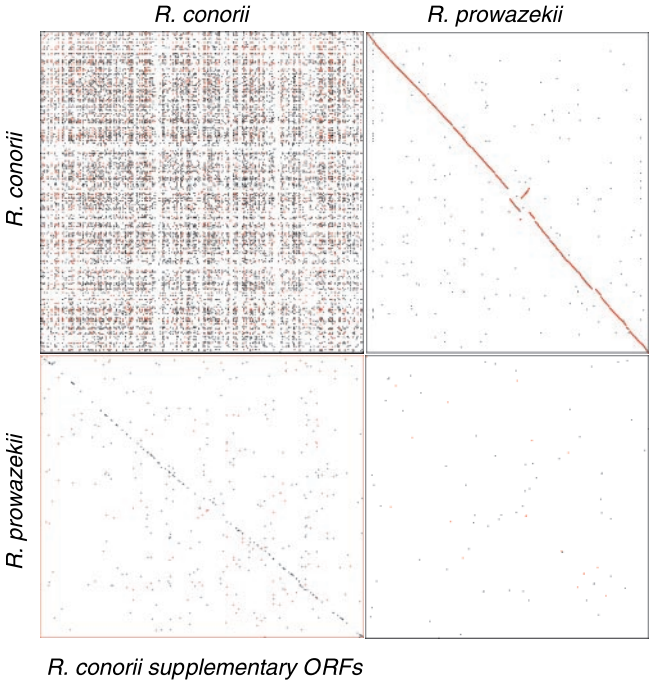

Fig. 2. Repeat density and colinearity of the *R. conorii* and *R. prowazekii* genomes. The two self-genome comparisons and the cross-genome comparison are presented in the upper left, upper right, and lower right panels. Each dot represents a high-scoring segment pair (HSP) identified by BLASTN (with a fixed database size parameter of 1 Mbp). Self matches are not shown. Red and black dots correspond to HSPs of *E* value $< 10^{-4}$ and *E* value $< 10^{-10}$, respectively. In addition, the lower left panel presents the similarities of *R. conorii* 552 supplementary ORFs with the intergenic regions of *R. prowazekii*. Each dot represents an HSP detected by BLASTN (*E* value < 0.1). The black/red dots correspond to matches on the same/reverse strand, respectively. For the sake of readability, dot sizes are standard in all the panels and do not correspond to the actual size of the HSP.

Table 2. Numbers of ORFs (each split case was counted once). Numbers in parentheses indicate the corresponding number of intact ORFs. Rp, *R. prowazekii*; Rc, *R. conorii*.

Function	Orthologous ORFs			*R. conorii* ORFs in excess						*R. prowazekii* ORFs in excess		
	Reciprocal best match	Split in Rp	Split in Rc	Remnant in Rp	Remnant in Rp, split in Rc	Similarity in Rp (paralog)	Similarity in Rp (paralog), split in Rc	No remnant in Rp	No remnant in Rp, split in Rc	Remnant in Rc	Similarity in Rc (paralog)	No remnant in Rc
Replication	45			4	5 (2)	6		1				
Transcription	22											
Translation	92	2 (1)		1		2						
Protein modification	14						2 (1)					
Degradation of protein	16	2 (1)	4 (1)			3		2				
Transporters	52		3 (1)	7	6 (3)	1	5 (2)					
Regulatory functions	10		3 (1)	2		3	2 (1)		2 (1)			
Energy metabolism	69		3 (1)				2 (1)					1
Fatty acid and phospholipid	21		11 (3)	3			2 (1)			1		
Nucleotide metabolism	16				2 (1)							
Amino acid metabolism	6			2								
Cell envelope	61	5 (2)	4 (2)	1	2 (1)	2	2 (1)	1				
Metabolism of cofactors	27	2 (1)		2	4 (1)							
Cellular process	36			2								
Drug sensitivity	14	2 (1)	7 (2)	4	8 (2)	2		1				
Other category	14											
Unclassified	36			3	2 (1)	4		1			1	
Unknown	216	10 (5)	9 (3)	169		134	7 (2)	128	8 (2)	5	5	17
R. conorii total 1374 (1306)	767	11 (11)	44 (14)	200	29 (11)	157	22 (9)	134	10 (3)	–	–	–
R. prowazekii total 834 (822)	767	23 (11)	14 (14)	–	–	–	–	–	–	6	6	18

functions [as already discussed for SpoT genes in *R. prowazekii* (*19*)]. The complete assessment of the physiological significances of these genes will require a detailed characterization of their translation products in *Rickettsia*.

It has been previously suggested that most of the intergenic sequences in *R. prowazekii* consist of decayed genes that are no longer active but are not yet totally eliminated from the genome (*3*, *8*, *19*). Through a systematic survey, we identified noncoding remnant sequences for 229 ORFs (out of the 552 *R. conorii* supplementary genes) at their homologous locations in the *R. prowazekii* genome (Fig. 2). For example, the *R. conorii* gene (RC1273) for the outer membrane protein rOmpA is 6063 bp long and is located between the cell division protein FtsK (RC1274) and a hypothetical ORF (RC1272) (Fig. 3). *R. prowazekii* exhibits the orthologs for FtsK and the hypothetical ORF but not for rOmpA. Part of the intergenic sequence between the *R. prowazekii* orthologs exhibits a significant similarity to the rOmpA gene of *R. conorii*, although the remnant sequence identified in *R. prowazekii* (369 bp) contains several in-frame stop codons. Our comparative genome analysis thus strongly supports a model of massive gene decay in *R. prowazekii*.

We found one clear case of a gene decaying after its horizontal transfer from (or, less likely, to) *Chlamydia*. *R. conorii* exhibits a split form (RC0035-38) of the gene for a bifunctional folate synthesis protein described in Chlamydiae as composed of two distinct domains: 7,8-dihydro-6-hydroxymethylpterin-pyrophosphokinase domain (HPPK) and dihydropteroate synthase domain (DHPS). Homologs of this enzyme with the same domain organization are found only in Chlamydiae, plants, and fungi. *R. prowazekii* exhibits remnant sequences corresponding to the *R. conorii* ORFs. The proximity of the *R. conorii* ORFs and Chlamydia genes supported by a phylogenetic tree analysis and the exclusive presence of this gene in *R. conorii* and *Chlamydia* species among known bacteria suggest a gene exchange between *Rickettsia* and *Chlamydia*, as proposed for ATP/ADP translocases (*18*). The alteration of this folate synthesis gene in both *Rickettsia* species—detected as a split gene in *R. conorii* and as a remnant sequence in *R. prowazekii*—suggests that significant changes of evolutionary constraints occurred after exchanging the gene with *Chlamydia*.

To investigate gene degradation in a wider set of *Rickettsia* species for which no sequence was available, we performed PCR assays on the genomic DNA of eight different *Rickettsia* species (*R. typhi*, *R. canadensis*, *R. helvetica*, *R. felis*, *R. australis*, *R. akari*, *R. rickettsii*, and *R. massiliae*) using primers derived from seven *R. conorii* and seven *R. prowazekii* supplementary genes [Web table 3 (*10*)]. With the primers corresponding to the seven *R. conorii* genes, two or more genomic segments were amplified in seven species [*R. rickettsii* (7 of 7), *R. australis* (6 of 7), *R. felis* (6 of 7), *R. helvetica* (5 of 7) *R. massiliae* (5 of 7), *R. akari* (3 of 7), and *R.*

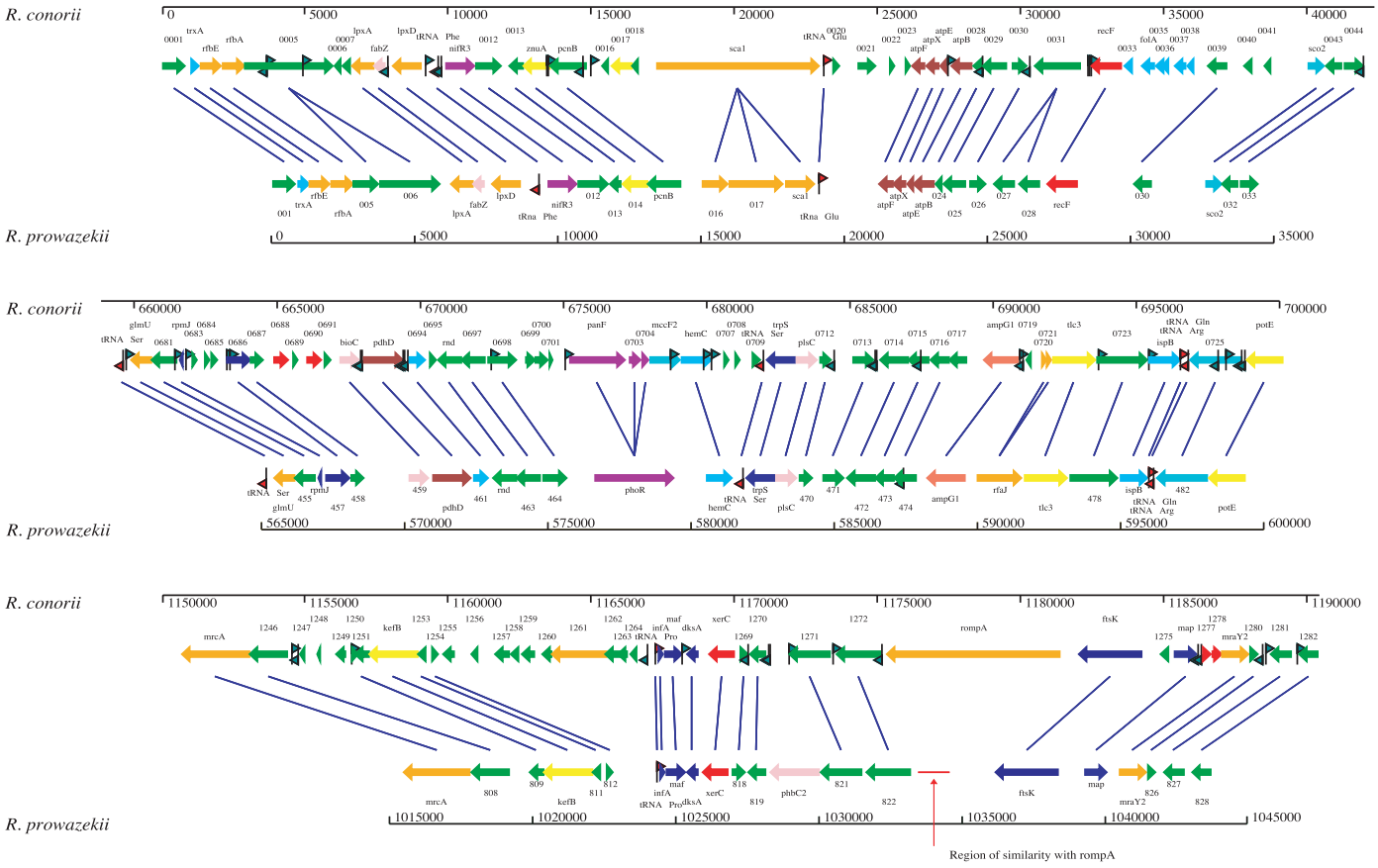

Fig. 3. Illustration of the colinearity. Three distinct segments from the *R. conorii* genome aligned with the homologous segments from the *R. prowazekii* genome are shown. These segments were chosen to show three types of gene alteration: split genes in *R. prowazekii* (top), a split gene in *R. conorii* (middle), and a gene remnant in *R. prowazekii* (bottom). A complete comparative map is available at (*10*).

canadensis (2 of 7)]. With the primers derived from the *R. prowazekii* genes, one or more genomic segments were amplified in four species [*R. typhi* (3 of 7), *R. australis* (1 of 7), *R. felis* (1 of 7), and *R. akari* (1 of 7)]. Thus, the supplementary genes observed in *R. conorii* and *R. prowazekii* do not originate from recent, species-specific, horizontal acquisitions, although the detailed pattern of PCR amplification does not exactly fit the standard classification of the *Rickettsia* genus [Web fig. 2 (*10*)]. Three out of the 14 supplementary genes were found in a split form (insertions/deletions generating stop codons) in one or more of the tested species. Thus, gene degradation appears to be a common feature of *Rickettsia*, targeting overlapping subsets of potentially dispensable genes while adapting to the selective pressures of different niches.

The few inversions/translocations locally breaking the otherwise perfect colinearity of the *R. conorii* and *R. prowazekii* genomes occur in the termination region of DNA replication. We identified several rearranged DNA segments, including a 38-kb segment containing 45 ORFs in *R. conorii* (Fig. 1). To date this inversion/translocation event within the phylogeny of *Rickettsia*, we used PCR on a set of primers designed from highly conserved adjacent sequences in the above eight species. A *R. prowazekii*–like arrangement was observed for *R. typhi*, whereas that of *R. conorii* was observed for *R. felis*, *R. rickettsii*, and *R. massiliae*. The result is consistent with the biphyletic division of *Rickettsia* and suggests that the genome rearrangement event would have happened relatively shortly after the initial divergence of the genus *Rickettsia*.

Genome reduction is thought to be a main force behind the evolution of parasitic and/or intracellular bacteria (*6–9*). The sequence of the *R. conorii* genome is consistent with this view, and *R. prowazekii* essentially appears as a subset of *R. conorii*. However, the genomes of *R. conorii* and *R. prowazekii* exhibit large differences in size, as well as in gene and G+C content, thus suggesting an adaptation to their specific niches rather than a simple model of random gene loss. Our analysis pointed out 137 *R. conorii* genes without any sequence similarity within the *R. prowazekii* genome. This provides an upper limit on the number of potential genes laterally acquired since their divergence, 40 to 80 million years ago (*34*). A single gene has its best match in eukaryotes (RC0781 to the NH_2-terminal part of yeast biotin-protein ligase), suggesting that *Rickettsia* have no particular tendency to evolve by acquiring genes from their hosts. Given their genetic isolation, it is tempting to postulate that *Rickettsia* had to rely on internal mechanisms such as duplication to acquire or modify some of the gene functions required for a better adaptation to their niche. We saw evidence of the very gradual nature of the genome reduction process by identifying all possible intermediates from intact ORFs: transcribed split ORFs, further split ORFs no longer transcribed, fully decayed but still recognizable ORFs, and complete gene disappearance. Similar mechanisms probably occur in the evolution of all bacterial species but have remained undetected because of more active recombination and a faster evolutionary rate (*35*).

References and Notes

1. D. Raoult, V. Roux, *Clin. Microbiol. Rev.* **10**, 694 (1997).
2. P. Yagupsky, B. Wolach, *Clin. Infect. Dis.* **17**, 850 (1993).
3. S. G. E. Andersson *et al.*, *Nature* **396**, 133 (1998).
4. *R. conorii* (strain Malish 7) was grown and purified as previously reported (*36*). A preliminary library A (3-kb inserts cloned in pcdna2.1 with Bst XI adaptators) was constructed and one genomic equivalent was sequenced. After the pilot study, we generated two other libraries of different insert sizes, F (5 kb) and G (10 kb) cloned in the same vector as A. Plasmid clones were sequenced at both ends of the insert with flanking vector sequences as primers. Dye primer reactions were analyzed on a LI-COR 4200 L (library A) and dye-terminator reactions were analyzed on a capillary ABI3700 (libraries F and G). The whole genome sequence assembly was performed by means of the PHRED and PHRAP software packages (*37*, *38*). A total of 9603 (6.4X), 5855 (2X), and 2352 (0.8X) end sequences (X's indicate genome equivalents), respectively, from libraries A, F, and G were incorporated into contigs. Cloning gaps and several sensitive regions were resolved or confirmed by sequencing duplicated PCR products. The final sequence includes 99.98% positions with PHRED scores over 40. None of the remaining 278 lower quality positions were found to coincide with a stop codon in our final annotation of ORFs nor were they related to split genes. Nucleotides in the *R. conorii* genome were numbered according to the *R. prowazekii* genome, where position one corresponds to the predicted origin of replication. The coherence of the assembly was verified by comparison to the Dag I and Sma I restriction site pattern previously obtained by pulse-field gel electrophoresis (*36*).
5. ORFs were delineated with a bioinformatics protocol similar to the procedure used to annotate the *R. prowazekii* genome sequence (*3*). Following Andersson *et al.*, we classified candidate ORFs in three categories: (i) ORFs at least 150 nucleotides (nt) long associated with high coding potential, (ii) ORFs at least 150 nt long exhibiting significant sequence similarity [E value < 0.005, with BLASTP (*39*) against the National Center for Biotechnology Information nonredundant protein database], and (iii) the remaining ORFs that are at least 300 nt long with neither coding potential nor similarity. Coding potential was evaluated with the SelfID program (*40*). Two ORFs for ribosomal proteins shorter than 50 amino acid residues were identified later. After this first pass, pairs of ORFs overlapping by more than 50% of the size of the shorter ORF were further handled as follows: The ORF exhibiting the highest scoring database matches was kept, irrespective of length. In case of no matches occuring, the longest ORF was selected. The final annotation contains 1374 ORFs. Genes for transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), and other RNAs were identified by BLASTN similarity searches against multiple databases. Two interrupted ORFs in *R. conorii* were annotated as pseudogenes on the basis of their colinearity and sequence similarity with two pseudogenes already annotated as such in *R. prowazekii*. Tentative ORF functions were assigned on the basis of sequence similarity against protein sequence databases [KEGG (*41*), SWISSPROT (*42*), and the nonredundant protein database] and domain/motif databases [Pfam (*43*) and PROSITE (*44*)]. Orthologous versus paralogous relationships were identified with multiple sequence alignments and neighbor-joining trees constructed by ClustalW (*45*). The functional classification was based on the scheme provided by the *R. prowazekii* genome project (*3*) as well as that provided by KEGG.
6. S. G. E. Andersson, C. G. Kurland, *Trends Microbiol.* **6**, 263 (1998).
7. J. O. Andersson, S. G. E. Andersson, *Curr. Opin. Genet. Dev.* **9**, 664 (1999).
8. _____, *Mol. Biol. Evol.* **16**, 1178 (1999).
9. S. T. Cole *et al.*, *Nature* **409**, 1007 (2001).
10. Supplementary Web material is available on *Science* Online at www.sciencemag.org/cgi/content/full/293/5537/2093/DC1.
11. We introduce the concept of split genes to define the cases where a given protein sequence usually encoded as a single ORF appears divided into a set of two or more colinear ORFs (>150 nt) on the same strand of the chromosome of another species. Most (48 of 68 for *R. conorii* and 11 of 12 for *R. prowazekii*) of the newly introduced stop codons are followed by the putative initiation codon of the next ORF within a distance of 150 nt. Split genes were identified with a program originally developed to extract evolutionarily conserved gene strings (*46*).
12. Noncoding remnant sequences in genome *A* corresponding to bona fide genes in the other genome *B* were identified as follows (*A* and *B* denote *R. conorii* or *R. prowazekii*). For every given ORF index gene (g_i) only found in genome *B* (supplementary genes), the colinearity of adjacent genes was used to delineate their homologous noncoding regions g_i' in genome *A*. g_i' was then diagnosed as containing a remnant sequence of gene g_i when the segment most similar to g_i was found within g_i' after a FASTA (*47*) search against the entire genome *A*. Additional cases were identified by comparing g_i and g_i' with LALIGN (*48*) (with a similarity threshold of E value $<10^{-10}$) and a subsequent analysis by dot-plot.
13. S. Kalman *et al.*, *Nature Genet.* **21**, 385 (1999).
14. V. Roux, *Rickettsiae and Rickettsial Diseases at the Turn of the Third Millennium*, D. Raoult, P. Brouqui, Eds. (Elsevier, Paris, France, 1999), p. 52.
15. N. Teysseire, C. Chiche-Portiche, D. Raoult, *Res. Microbiol.* **143**, 821 (1992).
16. P. Cossart, *Curr. Opin. Cell Biol.* **7**, 94 (1995).
17. Gene families were initially identified with the single linkage clustering method with BLASTP (with a threshold of E value $< 10^{-30}$). Out of the resulting clusters, six were composed of four or more ORFs: ATP/ADP translocases (five ORFs), proline/betaine transporters ProP (seven ORFs), (p)ppGpp 3-pyrophosphohydrolase SpoT (four ORFs), AmpG proteins (four ORFs), outer membrane proteins (four ORFs), and multidrug-resistant ABC transporter ATP-binding proteins (including alkaline protease secretion proteins) (five ORFs). According to the additional PSI-BLAST searches and careful examinations of the results, we identified four more ORFs for ProP, one more ORF for SpoT, and one more ORF for the Sca proteins.
18. Y. I. Wolf, L. Aravind, E. V. Koonin, *Trends Genet.* **15**, 173 (1999).
19. J. O. Andersson, S. G. E. Andersson, *Mol. Biol. Evol.* **18**, 829 (2001).
20. G. Schreiber, E. Z. Ron, G. Glaser, *Curr. Microbiol.* **30**, 27 (1995).
21. B. K. Hammer, M. S. Swanson, *Mol. Microbiol.* **33**, 721 (1999).
22. R. Chakraburtty, M. Bibb, *J. Bacteriol.* **179**, 5854 (1997).
23. B. Z. Harris, D. Kaiser, M. Singer, *Genes Dev.* **12**, 1022 (1998).
24. T. Leisinger, in Escherichia coli *and* Salmonella, F. C. Neidhardt *et al.*, Eds. (American Society for Microbiology, Washington, DC, 1996), pp. 434–440.
25. S. Mazareb, Z. Y. Fu, D. Zilberstein, *Exp. Parasitol.* **91**, 341 (1999).

26. P. F. Policastro, U. G. Munderloh, E. R. Fischer, T. E. D. Hackstadt, *J. Med. Microbiol.* **46**, 839 (1997).
27. S. F. Hayes, W. Burgdorfer, *Infect. Immun.* **37**, 779 (1982).
28. The repeated elements in the *R. conorii* genome were initially identified on the basis of the direct self-comparison of the genomic DNA by the BLASTN program (E value $< 10^{-4}$). The BLASTN result was then analyzed by the repeat identification program Mocca (*49*). An exhaustive analysis (types, locations, insertion in ORFs, etc.) of the repeats identified in the *R. conorii* genomic sequence is available under the *Rickettsia* database section at http://igs-server.cnrs-mrs.fr/RicBase/.
29. Ten repeat families include the previously identified *Rickettsia* palindromic element described in H. Ogata *et al.*, *Science* **290**, 347 (2000), and H. Ogata, S. Audic, J.-M. Claverie, *Science* **291**, 252 (2001).
30. D. R. Leach, *Bioessays* **16**, 893 (1994).
31. P. M. Sharp, D. R. Leach, *Mol. Microbiol.* **22**, 1055 (1996)
32. D. Raoult *et al.*, *Emerg. Infect. Dis.* **7**, 73 (2001).
33. The presence of transcripts associated with all *R. conorii* split genes was assessed by RT-PCR with primers designed from their predicted coding regions. In addition, we tested five genes (RC0873, RC0426, RC0888, RC0374, and RC0909) of *R. conorii* with no functional counterparts in *R. prowazekii*. Primer sequences and expected amplicon sizes are listed in Web table 2 (*10*). RNA was prepared as follows: A suspension of fresh *R. conorii* was adjusted to 10^8/ml and the bacteria were separated from cells with a sucrose gradient. RNA was extracted with the RN easy Midi kit (Qiagen, Hilden, Germany) as recommended by the manufacturer. RT-PCR was performed on the resulting RNA with the OneStep RT-PCR kit (Qiagen) according to the manufacturer's protocol. Reverse transcription and amplifications were performed in PTC-200 thermocyclers (MJ Research, Watertown, MA) with 40 PCR cycles. RT-PCR products were run in 1% agarose gels, stained with ethidium bromide, and revealed on an ultraviolet box. All amplicons exhibited the expected size. Once RT-PCR was performed, all primer pairs were validated by positive PCR reactions on genomic DNA. The absence of genomic DNA contamination in the RNA preparation was controlled by attempting PCR (with the Gibco-Brl Elongase polymerase) with the primers for ORF RC0909 (WASP-like gene) and ORF RC0843 (comF gene). The PCR assay was negative in both cases.
34. The divergence time was estimated on the basis of 16*S* rRNA sequences, assuming a relatively constant rate of evolution of 1 to 2% per 50 million years (*50*).
35. E. J. Feil *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 182 (2001).
36. V. Roux, D. Raoult, *J. Bacteriol.* **175**, 4895 (1993).
37. B. Ewing, L. Hillier, M. C. Wendl, P. Green, *Genome Res.* **8**, 175 (1998).
38. B. Ewing, P. Green, *Genome Res.* **8**, 186 (1998).
39. S. F. Altschul *et al.*, *Nucleic Acids Res.* **25**, 3389 (1997).
40. S. Audic, J.-M. Claverie, *Proc. Natl. Acad. Sci. U.S.A* **95**, 10026 (1998).
41. M. Kanehisa, S. Goto, *Nucleic Acids Res.* **28**, 27 (2000).
42. A. Bairoch, R. Apweiler, *Nucleic Acids Res.* **28**, 45 (2000).
43. A. Bateman *et al.*, *Nucleic Acids Res.* **28**, 263 (2000).
44. K. Hofmann, P. Bucher, L. Falquet, A. Bairoch, *Nucleic Acids Res.* **27**, 215 (1999).
45. J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **22**, 4673 (1994).
46. H. Ogata, W. Fujibuchi, S. Goto, M. Kanehisa, *Nucleic Acids Res.* **28**, 4021 (2000).
47. W. R. Pearson, D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2444 (1988).
48. X. Huang, W. Miller, *Adv. Appl. Math.* **12**, 373 (1991).
49. C. Notredame, *Bioinformatics* **17**, 373 (2001).
50. H. Ochman, S. Elwyn, N. A. Moran, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 12638 (1999).

51. We thank P. Sansonetti for strongly supporting the initiation of this project and F. Gosse for reading the manuscript. Supported in part by the Hoechst, Marion, Roussel research fund (grant no. FRHMR1/9777) and by Aventis Pharma, France. A more comprehensive database on the *R. conorii* genome and proteome is available at http://igs-server.cnrs-mrs.fr/RicBase/.

9 April 2001; accepted 23 July 2001

Role of Nonimmune IgG Bound to PfEMP1 in Placental Malaria

Kirsten Flick,[1] Carin Scholander,[1] Qijun Chen,[1] Victor Fernandez,[1] Bruno Pouvelle,[2] Jurg Gysin,[2] Mats Wahlgren[1]*

Infections with *Plasmodium falciparum* during pregnancy lead to the accumulation of parasitized red blood cells (infected erythrocytes, IEs) in the placenta. IEs of *P. falciparum* isolates that infect the human placenta were found to bind immunoglobulin G (IgG). A strain of *P. falciparum* cloned for IgG binding adhered massively to placental syncytiotrophoblasts in a pattern similar to that of natural infections. Adherence was inhibited by IgG-binding proteins, but not by glycosaminoglycans or enzymatic digestion of chondroitin sulfate A or hyaluronic acid. Normal, nonimmune IgG that is bound to a duffy binding–like domain β of the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) might at the IE surface act as a bridge to neonatal Fc receptors of the placenta.

Malaria infection with *P. falciparum* during pregnancy is an important cause of maternal morbidity and mortality. It may induce premature delivery, spontaneous abortion, or lead to a low birth weight (*1, 2*). Infections often cause more severe symptoms in primiparous than in multiparous women. The incidence of placental malaria similarly diminishes with increasing parity (*3, 4*), probably due to the acquisition of immunity to the infecting parasites (*5, 6*).

IEs are not passed from the mother to the fetus, but accumulate in the placenta which can experience high parasite densities (>50% IEs) while the peripheral circulation is almost free of IEs. Placental malaria may thus be caused by IEs that are selected for and expanded on receptors only present in the placenta (*7–10*), as opposed to those in other vascular beds.

Certain strains of *P. falciparum* bind nonimmune immunoglobulins onto the surface of the host erythrocyte, a fact that made us investigate their role in sequestration, in particular the possibility that IgG could bridge the IEs to Fc-receptors present in the placenta. We thus examined the frequency of IgG- (and IgM-) binding IEs accumulated during pregnancy in the placenta. Small pieces of snap-frozen placental tissues were obtained from six malaria-infected Cameroonese women after approved consent. The parasitemia of the placentas ranged from three to 23% (Fig. 1A), and all of them were classified as having active or active-chronic infections (*11*). IgG-binding IEs (Fig. 1B) were found in all of the placentas (10 to 75% IgG positive, mean 44%), whereas IgM-binding IEs (Fig. 1C) were more rare (2 to 34%, mean 18%) (*12*). IEs attached to the syncytiotrophoblasts bound only IgG (20 to 80%, mean 50%), except those of placenta CP42DJ, where the number of IgG-binding IEs was equal to that of the IgM-binding IEs. By studying the Ig-binding phenotype of IEs eluted from the placentas (*13, 14*), we confirmed that a majority of parasites causing active placental infection bound IgG (Table

Table 1. The phenotype of *P. falciparum* infecting the human placenta. IEs were eluted from the infected placentas and scored for their immunoglobulin binding.

Placenta	IgG-binding IE/IE tested*	Percent
CP24	18/23	78
CP42	14/14	100
CP42DJ†	–	–
CP193	205/276	74
CP939	32/247	13
CP940†	–	–

*Fractions of the IEs were also studied for their capacity for binding to Sca1D cells including inhibition with soluble CSA (100 μg/ml) and treatment with CSAnase ABC (0.5 U). About 50% of the IEs were specific for CSA (*11*). The populations of parasites studied for CSA-binding were not identical to those studied for immunoglobulin-binding since not all eluted IEs were scored in the CSA-assays. †The number of eluted parasites obtained from the placentas CP42DJ and CP940 were not sufficient for the assay.

[1]Microbiology and Tumor Biology Center (MTC), Karolinska Institutet and Swedish Institute for Infectious Disease Control, Box 280, S-171 77 Stockholm, Sweden. [2]Unité de Parasitologie Expérimentale, Faculté de Médecine, Université de la Méditerranée (Aix-Marseille II), 13385 Marseille Cedex 5, France.

*To whom correspondence should be addressed. E-mail: Mats.wahlgren@smi.ki.se