Tropheryma Whipplei Genome at the Beginning of the Post-Genomic Era

N. Crapoulet[1,§], P. Renesto[1,§], J. S. Dumler[2], K. Suhre[3], H. Ogata[3], J-M. Claverie[3] and D. Raoult[1,*]

[1]*Unité des Rickettsies, CNRS UMR 6020, IFR 48, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille, France;* [2]*Division of Medical Microbiology, department of pathology, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA and* [3]*Information Génomique et Structurale, CNRS UPR 2589, 13409 Marseille, France*

Abstract: *Tropheryma whipplei* is a Gram positive human pathogen that is the causative agent of Whipple's disease. Nearly one century elapsed between the first description of the disease in 1907 and the cultivation of this bacterium within eukaryotic cell cultures in 2000. This achievement has made possible genome sequencing of this poorly studied microorganism. This review summarizes post-genomic knowledge resulting from these genomic data. To compare the theoretical genetic capabilities of *T. whipplei* with those of other sequenced bacteria, a virtual microarray representation was generated. This *in silico* analysis supports the concept of independent evolution pathways for microbial pathogens. Concrete post-genomic consequences related to clinical microbiology such as the analysis of antibiotic susceptibility or the design of molecular tools convenient for PCR detection and epidemiology studies are described. Analysis of whole-cell metabolic networks of *T. whipplei* also provide clues for designing axenic media for this pathogen that is particularly recalcitrant to cultivation. This opens the way to investigate transcriptome analysis of *T. whipplei* by microarrays. Future prospects are also discussed.

Received on: 02 April 2005 - Accepted on: 21 April 2005

Key Words: *T. whipplei*, reduced genomes, post-genomics.

INTRODUCTION

Tropheryma whipplei is the bacterial agent of Whipple's disease, a rare multisystemic chronic infection initially described in 1907 by George Whipple [1]. Gastrointestinal compromise causing abdominal pain and diarrhea, which in turn leads to weight loss, malnutrition, and anemia is very common. Therefore, atypical clinical forms are increasingly recognized [2]. These bacteria were characterized as *Actinomycetes* for the first time by means of broad-range *rrs* (16S rRNA gene) amplification and molecular phylogenetic methods [3]. Phylogenetic location of *T. whipplei* is between *Cellulomonas* species and the group of *Actinomycetes* with group B peptidoglycan [4]. Several observations suggest an environmental origin for this bacterium [5]. Thus, epidemiological analysis reported a high frequency of Whipple's disease among farmers and carpenters [6]. In addition, this bacteria has been found in specimens from sewage treatment plants [5].

Over the last few years, sequencing of this bacteria [7, 8] was made possible by its cultivation within fibroblasts [9]. This achievement represents a major step in enhancing the knowledge of this poorly studied intracellular microorganism which replicates within late phagosomal compartment of HeLa cells [10]. This review focuses on recent advances gained from *T. whipplei* genome sequencing. Specific features of this pathogen deduced from its annotation are described. In order to investigate if genomic evolution correlates with bacterial lifestyle, a comparative genomic analysis including 67 other bacterial genomes and based on the clusters of orthologous groups (COG) was performed and presented as a virtual profiling microarray. Preliminary post-genomic analysis of *T. whipplei*, i.e. proteomics and transcriptomics approaches are also discussed as well as progress towards development of tools useful for clinical applications.

ANALYSIS OF *T. WHIPPLEI* GENOME

Genome Size

Prokaryotic genome sizes can vary from ~0.5 to 10 Mbp. To date, the smallest known genome (0.58 Mbp) is that of *Mycoplasma genitalium* [11] and the largest (9.11 Mbp) is that of *Bradyrhizobium japonicum* [12]. Although large variations within prokaryotic genomes are observed, the 20-fold size difference seen is quite small, compared to the more than 1 000 000-fold size range found in eukaryotic microbial genomes [13]. Genome size results from two counteracting processes, namely the acquisition of new genes by gene duplication or by lateral gene transfer and the deletion of non-essential genes. Being the sum of different genetic events, this value is not thought to be a good indicator of evolutionary lineage [14]. With a genome size of less than 1 Mb, *T. whipplei* offers the prime example of genome reduction among *Actinobacteria* and can be classified in the small bacterial genome category. Evolution towards reduced bacterial genome size is often associated with an intracellular lifestyle [15]. As for most intracellular pathogens, bacterial endosymbionts of insects have also undergone severe genome reduction in the context of their obligate associations with hosts [16]. According to current evolutionary thinking, selective lost of genes from bacteria depends on the environmental niche that is occupied [17, 18]. This can be exampli-fied by the fact that essential nutrients are provided by host

*Address correspondence to this author at the Unité des Rickettsies, CNRS UMR 6020, IFR 48, Faculté de Médecine, 27 Bd Jean Moulin, 13385 Marseille, France; E-mail: Didier.Raoult@medecine.univ-mrs.fr

§These authors have equally contributed to this work.

1389-2029/05 $50.00+.00

cells, thus eliminating the pressure to maintain biosynthetic genes.

GC Content

The GC content of sequenced bacterial genomes ranges from 72.1% for *Streptomyces coelicolor* [19] to 26.5% for *Wigglesworthia glossinidia* [20]. In contrast to *Actinobacteria* which are GC-rich [14], *T. whipplei* has a low GC% of 46. A correlation was clearly established between genome size and GC content, reduced genomes tending towards AT-richness [15]. This could result from the fact that GTP and CTP nucleotides are more energetically "expensive" than ATP and UTP. Another hypothesis could be the absence of genes involved in DNA recombination and repair allowing increased rates of random mutation (C to T or G to A) [14]. The fact that the GC content of *T. whipplei* is lower than other members of the *Actinomycetes* clade confirms its evolutionary divergence from ancestral genome leading to its reduction.

Amino Acid Metabolism and Energy-Generating System

In silico reconstruction of the metabolic network of *T. whipplei* showed that genes required for synthesis of key amino acids (histidine, tryptophan, leucine, arginine, proline, lysine, methionine, cysteine, and asparagine) were lacking. In addition, partial deficiencies were predicted for seven other amino acids (glutamate, glutamine, aspartate, threonine, valine, isoleucine, and phenylalanine) [21]. This observation is correlated with the fact that the majority of *T. whipplei* ORFs encoding membrane transporters were identified as amino acids transporters (http://www.membranetransport.org/). Such transporters are likely to mediate uptake of nutrients from host cells. Many core functions such as a major energy-generating system (the TCA cycle) are also absent in *T. whipplei*. It could be that bacteria also use host cells as a source of carbon and energy to grow.

Lack for Type-III and Type-IV Secretion System

Secretion systems are considered as part of pathogenesis machinery of several bacterial pathogens. One of the most studied systems is the type-III secretion system found in various Gram-negative organisms. This apparatus, which is homologous to the bacterial flagellar basal body, allows injection of bacterial proteins across the two bacterial membranes and the eukaryotic cell membrane to destroy host cell defense and signaling systems [22, 23]. Other bacteria use functionally homologous systems known as type-IV secretion systems. It was recently demonstrated that type-IV secretion systems plays an essential role for *Coxiella burnetii* pathogenesis [24, 25], as previously described for *Legionella pneumophila* [26]. In contrast to several other pathogens, *T. whipplei* does not have any type-III nor type-IV secretion machinery.

Regulatory Proteins

As described for other small genomes [15], *T. whipplei* has lost many regulatory elements such as sigma factors. This evolutionary aspect could be the result of protection against extreme environmental conditions among intracellular bacteria in contrast with free-living bacteria.

Multiple Copies Genes

In contrast to the observation of global genome decay, some *T. whipplei* genes are found in multiple copies. The existence of multiple gene copies in eukaryotes has been known for a long time and is considered an important element in their molecular evolution [27]. Therefore, the existence of such gene families termed as paralogs, as first evidenced in *E. coli,* was not expected [28]. These gene families of diverse size and degree of similarity remain an important and little explored feature of prokaryotes. They are likely to be involved with adaptation of the bacteria to specific niches [29]. In *T. whipplei* Twist genome, 38 families of predicted gene paralogs were found. They include predicted membrane proteins, exodeoxyribonuclease III, inorganic pyrophosphatases and iron ABC transporters as well as a heterogeneous family of surface proteins termed WiSP [7]. The genomic organization of Twist and the other sequenced *T. whipplei* isolate (strain TW08/27) [7] differs by the inversion of a large chromosomal segment which could result from host-bacteria interaction as previously suggested for other Gram-negative species (*Salmonella, Neisseira, Pseudomonas*, and *Bordetella*) [30]. Two genes of the WiSP membrane protein family are located at the extremities of inverted region. This lead to a significant alteration of both TW157 and TW625-encoding WiSP proteins in the TW08/27 isolate compared with those in Twist. This could confer variability of antigen specificity between these strains.

Repeated Sequences and Genomic Rearrangements

It has been observed that some reduced genomes of intracellular bacteria share a low content of repeated sequences [31]. Typical changes associated with host cell restriction and accompanying genome reduction also include chromosomal rearrangements and deletion, pseudogene formation, and increased frequency of mobile elements such as insertion sequences (IS). Proliferation of these IS was shown to be particularly high in recently host-adapted bacteria. In contrast few IS are found in ancient symbionts [15]. To our knowledge, the sole IS characterized within *T. whipplei* genome is a 80 bp actinobacterial insertion type B sequence located within domain III of its 23S rDNA [32]. Similarly, only one pseudogene was identified. Finally, coding content of *T. whipplei* was of 85.6%, a value comparable to that observed in free-living bacteria.

Functional Annotation of *T. whipplei* Genome is Still Incomplete

As a carbon pool, folates are essential for vital functions. While synthesized *de novo* by most microorganims, a few bacteria do not make folates and must acquire them from external sources by using energy-requiring transport systems. This could be the case concerning *T. whipplei* for which incomplete folate biosynthetic pathway has been identified. Indeed, according to KEGG [33] dihydrofolate reductase (DHFR) is lacking. However, this contradicts the well established role of this enzyme as a target of trimethoprim, an antimicrobial agent successfully used in combination with sulfamethoxazole for the treatment of Whipple's disease [34]. In organisms with small genomes, it is relatively straightforward to use direct computational pre-

dictions based on genomic sequences to identify most genes by their long open reading frames (ORFs). However, this can result to incorrect or incomplete information [35, 36]. Several bioinformatic algorithms have been developed to ensure correct functional assignments and can be based for family and superfamily classifications (http://supfam.mrc-lmb.cam.ac.uk/SUPERFAMILY/). This approach recently identified two proteins potentially encoded by *T. whipplei* genome that share homology with DHFR. In the Twist strain, DHFR candidates are TWT687 (5-amino-6-(5-phosphoribosylamino) uracil reductase) and TWT392 (5-amino-6-(5-phosphoribosylamino) uracil reductase), respectively. Similarly, this database allowed identification of three putative thioredoxin-like proteins, TWT256 (thiol-specific antioxidant protein), TWT756 (thiol:disulfide interchange protein), and TWT345 annotated as hypothetical protein. As in other sequenced genomes, a large proportion of *T. whipplei* (30.5%) genes are 'hypothetical', in that no function could be ascribed. Such ORFs could be properly assigned in future to help a better characterization of this microorganism.

Annotation of newly assigned proteins must be experimentally validated by testing enzymatic activity of recombinant proteins cloned into expression vectors. Genetic transformation of *T. whipplei* has never been reported. Progress in this field would be also be helpful to investigate phenotypic characteristics of knock-out mutants.

Comparison of *T. whipplei* with other Genomes

Complete genome sequences are becoming available for a large number of bacterial species, making comparative genomics a particularly promising tool for research. Here, such a comparison was undertaken considering 67 bacterial genomes. Genome content of these bacteria was compared through the COG database, each COG including proteins thought to be orthologous [37]. As shown in Fig. (**1**), a "virtual microarray" showing the subset of data with the functional class and metabolic pathway annotation was constructed.

This analysis revealed several features among genes uniformly conserved or lost in selected genomes. The concept of a minimal gene set necessary and sufficient to ensure bacteria life was long investigated by comparing the first two completed bacterial genomes, *Haemophilus influenzae* and *Mycoplasma genitalium*, the small genome encoding 480 proteins [38]. Further accumulation of sequenced bacterial genomes allowed an estimate of 81 genes as the number conserved in all domains of life [39]. In this analysis, a minimal complement of genes and functional categories were shared among all genomes, comprising core metabolic functions related to transcription, translation, ribosomal structure and aminoacyl-tRNA synthetases. Another striking finding in this analysis is the clustering of bacteria with small genomes that are almost uniquely host-associated parasites or symbionts [40]. This point is illustrated in phylogenic tree presented Fig. (**2**). When considering only reduced bacterial genomes, only 51 genes are ubiquitously conserved (Table **1**). These genes mainly belong to the nucleic acid-based central information pathway (ribosomal proteins, DNA/RNA polymerase subunits, elongation factors). This pattern is close to that deduced when comparing three domains of life, namely Bacteria, Archaea and Eukarya [41, 42].

Upon closer inspection, it becomes apparent that clustering exclusive of large genome organisms occurs not as a result of similarities, but predominantly by virtue of the marked reduction in overall genome size that has led to the loss of a majority of COGs - similarity by "exclusion". The genetic mechanisms leading to genome degradation in obligate intracellular bacteria have been widely debated and it is thought that gene loss is not random [43, 44]. As previously demonstrated, bacterial genome reduction occurs as a function of association with eukaryotic hosts. Presence in the host of a constant environment rich in metabolic intermediates renders useless some genes and these superfluous sequences are eliminated through mutational bias favoring deletion. Thus, as for *T. whipplei* absence of genes involved in amino acid synthesis and energy production is found in all the reduced bacterial genomes. As illustrated in Table **2**, among genomes < 2 Mb, approximately 40% and 16% of the missing genes are involved in metabolic pathways and in cellular processing and signalling, respectively. The remaining 40% correspond to COGs for which functions are unknown.

Comparative genomic analysis of reduced bacterial genomes illustrates clearly that each microorganism has adapted a divergent evolutionary strategy. Observed variable retentions and losses range from aerobic respiratory pathway loss in *Chlamydiales*, *Mollicutes*, *Borrelia burgdorferi* and *Treponema pallidum* [45-48], retention of terpenoid synthesis pathways in *T. whipplei* [7, 8], and retention of nucleoside and nucleotide salvage pathways in *Mycoplasma* and *Ureaplasma* [11, 49]. These examples serve to underscore the variable metabolic capacities that resulted from differential gene loss and retention among conserved COG members.

Other specific features for *T. whipplei* genome such as the absence of both type-III and type-IV secretion systems or the presence of multiple copies of Wisp proteins were detailed above. This can permit phenotypic variations, which can in turn allow the microorganism to counter improved host defences, to compete with other bacterial strains that co-infect the same host, or to facilitate transmission. Diverse evolutionary strategies are previously described for insect endosymbionts, specific genetic characteristics regulating the mode of symbiont transmission, the number of symbionts transmitted, and the sex of host offspring [50]. Thus, it is conceivable that specific selection might vary between several bacteria, according to their respective lifestyles.

POST-GENOMIC ANALYSIS OF *T. WHIPPLEI*

Proteome Analysis of *T. whipplei*

As a consequence of AT enrichment, proteins encoded by small genome bacteria are enriched in amino acids containing more A or T in corresponding codons such as lysine. Such AT bias modifies the isoelectric point (pI) of polypeptides. Thus, the average pIs for polypeptides of *Buchnera* relative *E. coli* are 9.6 and 7.2, respectively [51]. The *Buchnera* genome also has a very low GC% (26.3%). This is not the case concerning *Coxiella burnetti* (GC% = 42.7), while approximately 45% of *C. burnetii* proteins were found to have a pI value of ≥9 [52]. It has been hypothesized that this could correlate with its ability to replicate in eukaryotic cell acidic phagolysosomes [52, 53]. Similarly, a high aver-

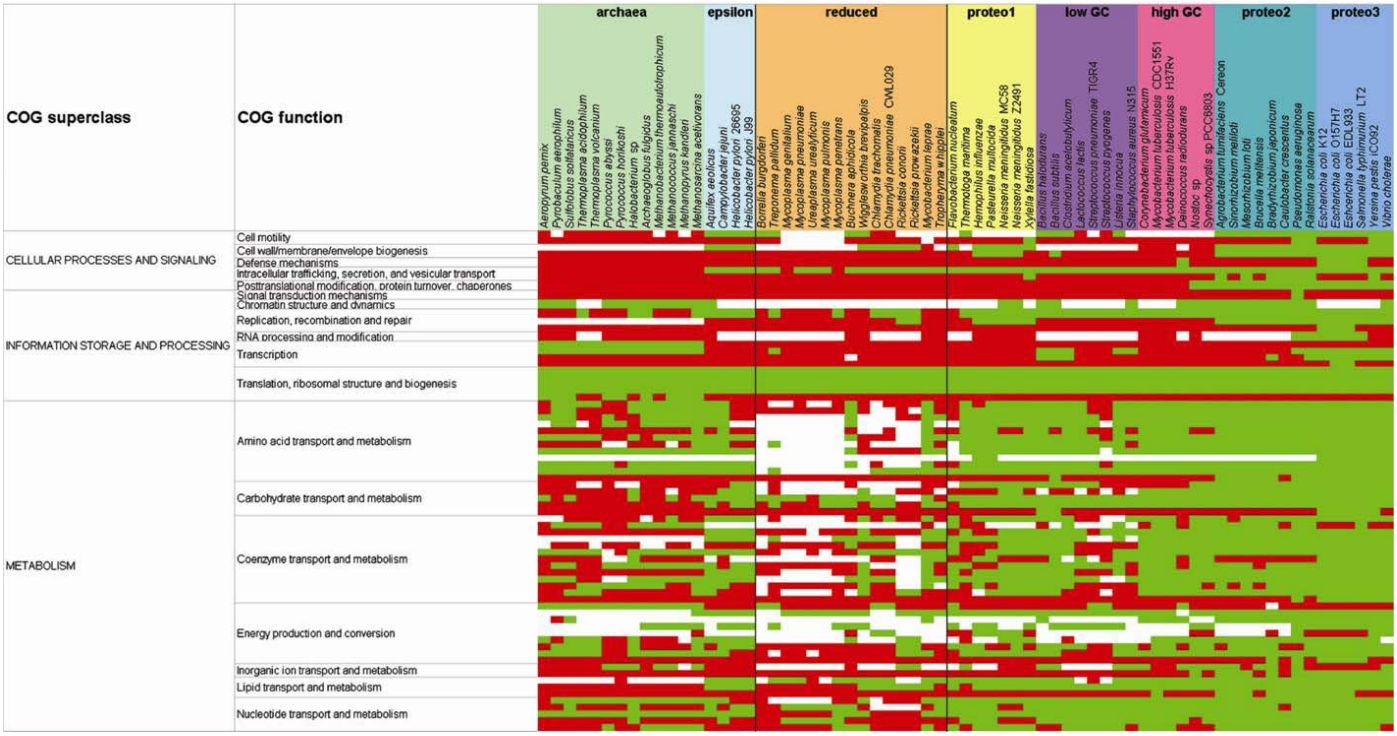

Fig. (1). Virtual or profiling microarray of COG superclass and functional groups in the complete genome sequences of 67 bacteria, separated based upon bacterial grouping (heading). Each functional group is colored according to proportional content of the functional group, where green depicts ≥ 50% of COG functional group member present, red depicts < 50% of COG functional group member present, and white depicts no COG functional group member detected in the genome. Note that only constant feature among all genomes is the presence of ≥ 50% of COGs in the translation, ribosomal structure and biogenesis functional group.

age pI value was noticed for *Helicobacter pylori* which replicates in the low-pH environment of the gastric mucosa [52]. Previous experiments performed in our laboratory demonstrated that intracellular survival of *T. whipplei* required acidic pH [10]. Interestingly, and as illustrated in Fig. (3), the pI value distribution of proteins encoded by *T. whipplei* is similar to that of *Coxiella burnetti*.

Preliminary analysis of *T. whipplei* proteome was investigated using two-dimensional polyacrylamide gel electrophoresis combined with MALDI-TOF mass spectrometry (unpublished data). This established a reference map of *T. whipplei* proteome for identifying bacterial antigens specifically recognized by sera of infected patients. Comparison of protein profiles of different isolates or of bacteria maintained under different experimental conditions could be analyzed by such an approach.

Evidence for Intervening Sequences Within 23S rRNA

A concrete and immediate application resulting from the establishment of axenic medium was the optimization of *T. whipplei* RNA extraction. When considering obligate intracellular bacteria, contamination of bacterial RNA with host cell nucleic acids is a significant problem. The electropherogram pattern of RNA extracted from Vero cells infected with *R. conorii*, a strict intracellular bacteria, was determined by using the Bioanalyzer 2100. We thus observed 4 peaks corresponding to 16S and 23S bacterial rRNA and to 18S and 30S eukaryotic rRNA, respectively, with a ratio prokaryotic *vs* eukaryotic material of only 10% [54]. In our hands, preliminary experiments carried out in order to extract *T. whipplei* RNA from infected fibroblasts were difficult to analyze. Indeed, using this pathogen an atypical RNA profile was observed. This suggested that *T. whipplei* RNA was degraded, but contamination with eukaryotic material was an impediment to further analysis. This was resolved using cell-free cultivation of this bacterium which permitted optimization of the RNA extraction procedure. To our knowledge, extraction of *T. whipplei* RNA has never been described. This achievement is of importance when considering that it opens the way to microarray application discussed below. Moreover, purification of *T. whipplei* RNA makes possible the determination of 23S/16S rRNA ratios and northern blot assays. By this latter approach we demonstrated that the 80 bp insertion sequence located within 23S rRNA [32] was excised, leading to 23S rRNA degradation. This cleavage is likely to involve a ribonuclease III identified within the genome of this microorganism. Phylogenetic analysis between *T. whipplei* and other described bacterial intervening sequences (IVSs) suggests that *T. whipplei* 23S rRNA IVS was acquired by lateral gene transfer between enteric bacteria (submitted for publication).

Microarray Perspectives

The avaibility of *T. whipplei* genome sequences coupled with optimization of RNA extraction now allows transcriptome analysis. Global gene expression profiling with microarrays is a powerful tool for assessment of bacterial gene expression under different physiological conditions [55, 56].

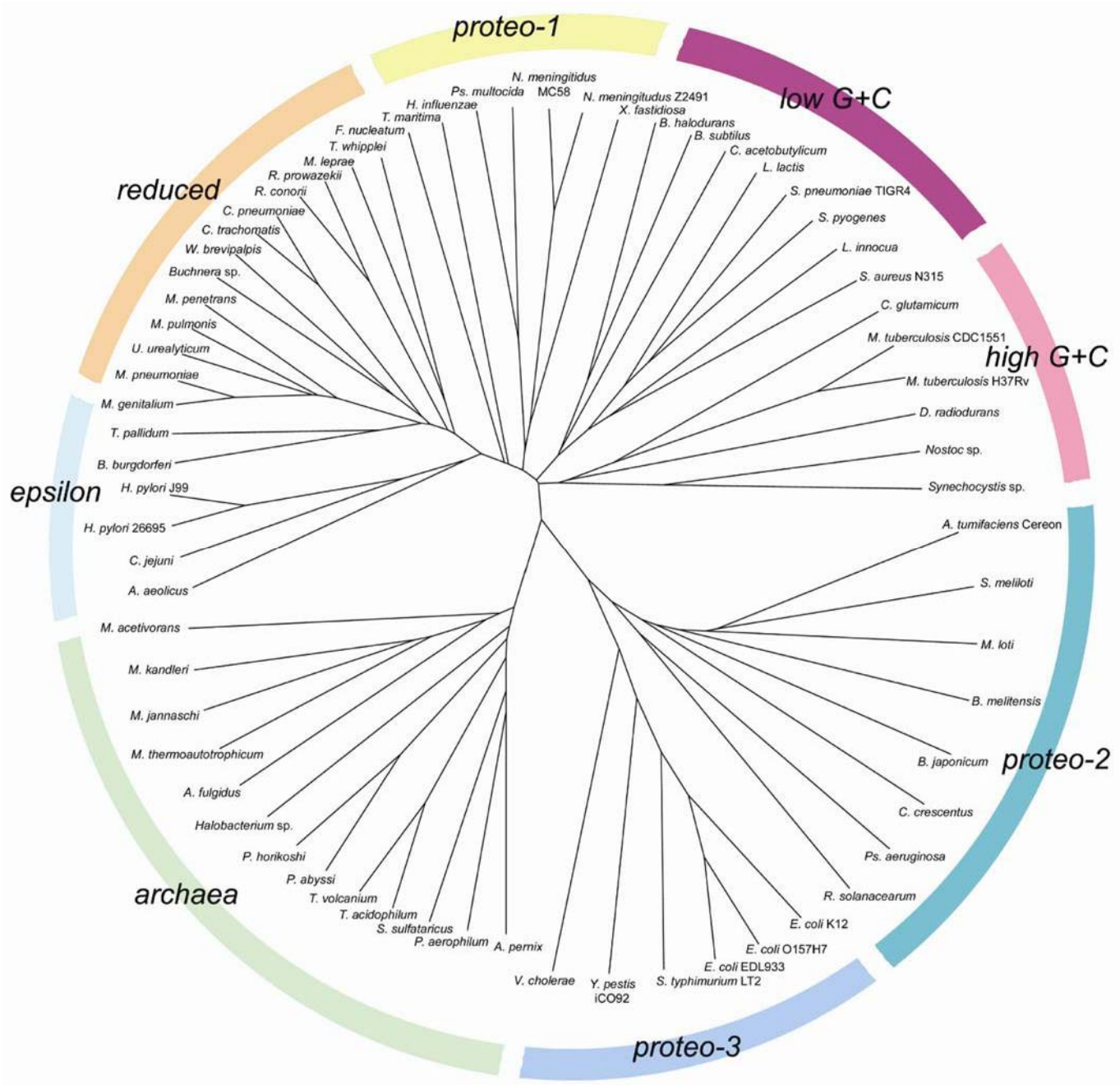

Fig. (2). UPGMA tree develop from alignment of COG content for 67 completed bacterial genomes. Note the clustering of small genome bacteria to the exclusion of those with larger genomes. The major genome groupings are indicated by the labels. The names and abbreviations for the bacterial species genomes are shown in Fig. (**1**).

This approach would undoubtedly contribute to a more global view of the factors involved in *T. whipplei* adaptation and pathogenicity.

Analysis of the Transcriptome of T. whipplei Grown in Various Experimental Conditions

Most microarray studies use hybridization of RNA extracted from bacteria cultured in axenic medium. Selected *in vitro* growth conditions are believed to simulate host microenvironments with respect to one or more parameters like growth phase, pH and osmolarity. During its life, this bacterium is suspected to have an environmental origin [5], most probably encountering different environments and therefore must adapt to various temperatures. Consequently, transcriptome analysis of *T. whipplei* maintained under different temperatures should shed light on the pathogenicity. This was successfully achieved for several pathogens such as *Shigella flexneri* (environmental reservoir), *Bordetella pertussis* (unknown reservoir) and *Yersinia* pestis (arthropod vector) [57]. For these microorganisms, virulence genes were up-regulated at 37°C and repressed at lower temperatures, this parameter acting as a "on-off switch" signal.

Table 1. Genes Conserved in Reduced Bacterial Genomes

COG	COG annotation	COG function	COG class	COG class
COG0013	Alanyl-tRNA synthetase	Translation, ribosomal structure and biogenesis	INFORMATION STORAGE AND PROCESSING	48
COG0018	Arginyl-tRNA synthetase			
COG0030	Dimethyladenosine transferase (rRNA methylation)			
COG0008	Glutamyl- and glutaminyl-tRNA synthetases			
COG0124	Histidyl-tRNA synthetase			
COG0060	Isoleucyl-tRNA synthetase			
COG0495	Leucyl-tRNA synthetase			
COG0016	Phenylalanyl-tRNA synthetase alpha subunit			
COG0012	Predicted GTPase, probable translation factor			
COG0442	Prolyl-tRNA synthetase			
COG0081	Ribosomal protein L1			
COG0244	Ribosomal protein L10			
COG0080	Ribosomal protein L11			
COG0102	Ribosomal protein L13			
COG0093	Ribosomal protein L14			
COG0200	Ribosomal protein L15			
COG0197	Ribosomal protein L16/L10E			
COG0256	Ribosomal protein L18			
COG0090	Ribosomal protein L2			
COG0089	Ribosomal protein L23			
COG0198	Ribosomal protein L24			
COG0255	Ribosomal protein L29			
COG0087	Ribosomal protein L3			
COG0088	Ribosomal protein L4			
COG0094	Ribosomal protein L5			
COG0097	Ribosomal protein L6P/L9E			
COG0100	Ribosomal protein S11			
COG0048	Ribosomal protein S12			
COG0099	Ribosomal protein S13			
COG0199	Ribosomal protein S14			
COG0184	Ribosomal protein S15P/S13E			
COG0185	Ribosomal protein S19			
COG0052	Ribosomal protein S2			
COG0092	Ribosomal protein S3			
COG0522	Ribosomal protein S4 and related proteins			
COG0098	Ribosomal protein S5			
COG0049	Ribosomal protein S7			
COG0441	Threonyl-tRNA synthetase			
COG0231	Translation elongation factor P (EF-P)/translation initiation factor 5A (eIF-5A)			
COG0480	Translation elongation factors (GTPases)			
COG0361	Translation initiation factor 1 (IF-1)			
COG0180	Tryptophanyl-tRNA synthetase			
COG0162	Tyrosyl-tRNA synthetase			
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	Transcription		
COG0195	Transcription elongation factor			
COG0250	Transcription antiterminator			
COG0358	DNA primase (bacterial type)	Replication, recombination and repair		
COG0592	DNA polymerase sliding clamp subunit (PCNA homolog)			
COG0201	Preprotein translocase subunit SecY	Intracellular trafficking, secretion, and vesicular transport	CELLULAR PROCESSES AND SIGNALING	3
COG0541	Signal recognition particle GTPase			
COG0552	Signal recognition particle GTPase			

Table 2. Genes Lost in Reduced Bacterial Genomes

COG	COG annotation	COG function	COG class	COG class
COG3143	Flagellar hook-length control protein	Cell motility	CELLULAR PROCESSES AND SIGNALING	4
COG3852 COG2766	Uncharacterized protein involved in tellurite resistance Uncharacterized Fe-S center protein	Signal transduction mechanisms		
COG2747	Lysophospholipase L1 and related esterases	Transcription		
COG1410 COG2066	Uncharacterized protein related to proFAR isomerase (HisA) Long-chain fatty acid transport protein	Amino acid transport and metabolism	METABOLISM	10
COG3265	Uncharacterized protein conserved in bacteria	Carbohydrate transport and metabolism		
COG4206	ABC-type sulfate transport system, permease component	Coenzyme transport and metabolism		
COG2224 COG1251	Malate synthase NADH dehydrogenase, FAD-containing subunit	Energy production and conversion		
COG0376	Catalase (peroxidase I)	Inorganic ion transport and metabolism		
COG3154	Uncharacterized protein involved in an early stage of isoprenoid biosynthesis	Lipid transport and metabolism		
COG4181 COG3127	Uncharacterized protein conserved in bacteria Uncharacterized iron-regulated protein	Secondary metabolites biosynthesis, transport and catabolism		
COG2719 COG3045 COG3384 COG2718 COG3126 COG4681 COG3159 COG3132	Uncharacterized vancomycin resistance protein Uncharacterized protein related to deoxyribodipyrimidine photolyase FOG: Transposase and inactivated derivatives Uncharacterized conserved protein Predicted ABC-type transport system involved in lysophospholipase L1 biosynthesis, permease component Predicted membrane protein Regulator of sigma D Outer membrane lipoprotein	Function unknown	POORLY CHARACTERIZED	10
COG3217 COG0384	ABC-type uncharacterized transport system, auxiliary component Predicted epimerase, PhzC/PhzF homolog	General function prediction only		

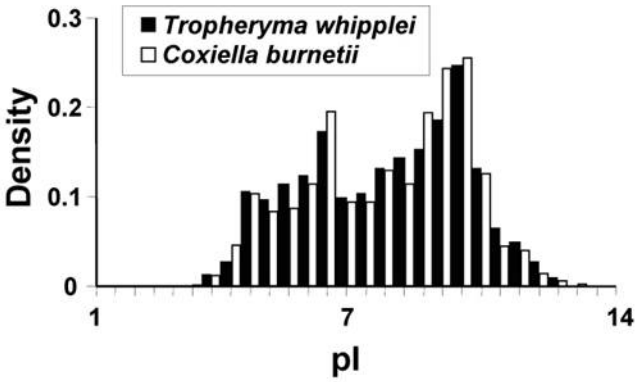

Fig. (3). Histogram for pI of *Coxiella burnetii* and *Tropheryma whipplei*. pI of all the ORFs annotated in both genomes were calculated by using 'iep' tool from the European Molecular Biology Open Software Suite (EMBOSS).

Comparative gene regulation analysis of *T. whipplei* grown either under axenic conditions or in association with fibroblasts should also provide insights into the relationships of this intracellular pathogen with host cells. When grown in presence fibroblasts, two forms of *T. whipplei* are observed, an intracellular form located within vacuoles of infected cells and an extracellular form in aggregates [58]. Comparison of their respective expression profiles could allow identification of genes responsible for bacterial sporulation or starvation.

Another powerful application is profiling of antibiotic-induced changes in gene expression. This approach has been used to characterize defence mechanisms against antibiotic stress for several pathogens. Such applications can also be useful for identifying new therapeutic targets [59]. Finally, microarray analysis could also permit the prediction of operons from co-expression patterns obtained from distinct experimental conditions and the reconstruction of transcriptional regulatory networks [60, 61].

Microarrays and Comparative Genomics

Studies on microarray-based comparative genomic hybridization (array CGH) can led to a reconstruction of the genetic profile of a bacterium by highlighting both deletions and differences in the number of multicopy genes between various strains [62, 63]. While the different *T. whipplei*

strains Twist, Endo 5, and Slow 2, have distinct replication times when grown in association with MRC5, their growth is similar in axenic medium [64]. This difference could result from deletion of some genes. The array CGH application could provide evidence of differences in genetic content among these strains.

CLINICAL MICROBIOLOGY CONSEQUENCES

Design of an Axenic Medium Allowing Cell-Free Cultivation of *T. whipplei*

As mentioned above, *T. whipplei* was for years recalcitrant to cultivation. *In vitro* propagation of this microorganism was achieved by inoculating human fibroblasts [58], consistent with. the strict intracellular nature of this pathogen [9]. However, from the *in silico* functional analysis of *T. whipplei* genome, we successfully designed a comprehensive medium allowing its growth in absence of eukaryotic cells [21]. This medium, obtained by addition of exogenous amino acids and L-glutamine required to compensate predicted deficiencies in *T. whipplei* metabolic pathways, was inoculated with a cell-free supernatant of fibroblasts infected with *T. whipplei* strain Twist [21]. Kinetic growth was monitored by Gimenez staining, flow cytometry counting and quantitative polymerase chain reaction analysis. From resulting data, the replication time of this bacteria was estimated to be of 28 hours, a value close to that observed in host-dependent conditions [65]. Possible genetic modifications of the bacteria selected with axenic culture, as compared with fibroblast cultures, were excluded by comparison of their respective restriction profiles analyzed through pulse-field gel pattern following digestion with two distinct enzymes. We also showed that the rope-like structures hypothesized to be associated with a starvation state of the bacterium are not present in axenic cultures. To date two other established strains and four clinical samples of *T. whipplei* have been propagated axenically in our laboratory.

This paradigm illustrates the application of genomic information toward rational design of a physiological medium based on understanding cellular metabolism rather than by empirical "trial and error". In addition, as discussed below, axenic cultivation of such micro-organisms should permit rapid advancements in knowledge of the molecular mechanisms responsible for the pathogenicity of *T. whipplei* and which were previously hampered by their host niche location.

Antibiotic Susceptibility and Trimetoprim

Antibiotic treatment for Whipple's disease was initially empirically determined on the basis of a few clinical observations. Isolation and culture of the infectious agent allowed determination of its susceptibility to a large panel of antibiotics, bolstered by the use of a quantitative PCR assay. Thus, it was demonstrated that doxycycline, macrolides, ketolides, aminoglygosides, penicillin, rifampin, teicoplanin, chloramphenicol, and trimethoprim-sulfamethoxazole (co-trimoxazole) were active, with MICs ranging from 0.25 to 2 μg/ml [66]. Due to the absence of the DHFR trimethoprim target within the *T. whipplei* genome [34], the activity of tri-methoprim-sulfamethoxazole confirmed using bacteria grown in axenic conditions was hypothesized to result from the effect of the sulfonamide alone [64]. When tested separately, trimethoprim failed to inhibit *T. whipplei* growth, while sulfamethoxazole was inhibitory [64]. Accordingly, the dihydropteroate synthetase gene, encoding the target for sulfamethoxazole, is present in *T. whipplei* genome.

Bioinformatic analysis of the *T. whipplei* genome also allowed identification of not only *gyrA* but also *parC* encoding the alpha subunit of the natural fluoroquinolone targets DNA gyrase and topoisomerase IV, respectively. From this observation it was hypothesized that *T. whipplei* is naturally relatively resistant to fluoroquinolones [65].

Enhanced Sensitivity of PCR Detection

Seven repeated sequences of 677 bp length with a percentage of nucleotide sequence similarity ranging from 99.4 to 100% were found in both *T. whipplei* sequenced genomes [7, 8]. Higher PCR detection sensitivity of such multiple copies genes was demonstrated for *Coxiella burnetii* [67]. Thus, *T. whipplei* repeated sequences were chosen as targets in PCR assays performed with samples from patients with Whipple's disease and from a control group as template. Data were compared to those obtained using regular diagnostic PCR targeting the 16S-23S rRNA gene intergenic spacer and the *rpoB* gene [68]. This comparison demonstrated that DNA detection sensitivity was significantly enhanced using primers targeting repeated sequences in the *T. whipplei* genome [69]. Thus, availability of *T. whipplei* genome sequence allowed rational choice of primers that would significantly improve the diagnosis of such infectious diseases.

Multi-Spacer Typing-Based Epidemiology

Availability of genome sequences would also permit rational, rather than empirical determination of bacterial genotype at the strain level. The development of such a typing method for rickettsial strains has recently been established in our laboratory [70]. The high degree of co-linearity between the *R. conorii* and *R. prowazekii* genomes enabled comparison of the interspecies variability of coding genes, degraded genes, conserved intergenic spacers, and variable spacers. From *in silico* identification of the most variable sequences between these closely related bacterial genomes, intergenic spacers were selected as target sequences for strain genotyping. The combined use of variable spacer sequences, named multispacer typing (MST), was found to be significantly more discriminatory than multigene sequencing. Preliminary data suggested that MST, which shares several advantages including discriminatory power, reproducibility and simplicity of interpretation, could be useful for *T. whipplei.*

CONCLUSION

It is now clearly established that genome sequencing facilitates research on infectious agents. This is particularly true when considering sequencing of fastidious microorganisms such as intracellular bacteria. Thus, *W. pipientis* was first cultured in 1997 and sequenced in 2004 [71], in turn allowing identification of several genes involved in host-symbiont interactions. Another example is that of *Parachlamydia* first identified in amoeba in 1997 and se-

quenced in 2004 [72] which permitted the unexpected identification of genes for conjugation [73]. *T. whipplei* which was for years recalcitrant to cultivation, was isolated and cultured for the first time in 2000 [9] and the genome sequences of 2 different strains were published 3 years later [7, 8]. Rather than the genome sequence being an end in itself, it is the start of much work to put the information into context with regard to the biology of the microorganism. In this review, we summarized progress in understanding *T. whipplei* that results from its genome sequence. Sequence data can be exploited to reveal how species evolved, strain diversity being based both on the loss of non-essential genes and acquisition of others. We described characteristics of the reduced genome of *T. whipplei* examined alone or in the context of a large scale comparisons including 67 bacteria. This analysis provides evidence that a common pathway or method in genome size reduction does not exist.

We then focused our interest on post-genomic approaches, including proteomics and transcriptomics that would allow to rapid identification of candidate targets for new diagnostic assays and therapeutics. Concrete clinical microbiology improvements have already been achieved, among which the establishment of cell-free culture [21], the determination of antibiotic susceptibility [64-66] and the design of specific molecular tools for enhanced PCR detection [69] have been reported. As for other microorganisms, the availability of *T. whipplei* genome sequence represents a powerful resource to help understanding of the complex biology of this pathogen as well as its relationships with host cells. To our knowledge, genetic manipulation of *T. whipplei* has never been reported. Progress in this field would undoubtedly be required in the future.

ACKNOWLEDGEMENTS

This work was supported by a grant form the European Community (Project Whipple's disease, QLG1-CT-2002-01049).

REFERENCES

[1] Whipple, G.H. A hitherto undescribed disease characterized anatomically by deposits of fat and fatty acids in the intestinal and mesenteric lymphatic tissues. *Bull. Johns Hopkins Hosp.* **1907**, 382-393.

[2] Marth, T. and Raoult, D. Whipple's disease. *Lancet* **2003**, *361*: 239-246.

[3] Relman, D.A.; Schmidt, T.M.; MacDermott, R.P. and Falkow, S. Identification of the uncultured bacillus of Whipple's disease. *N. Engl. J. Med.* **1992**, *327*: 293-301.

[4] Maiwald, M.; Ditton, H.J.; von Herblay, A.; Rainey, F.A. and Stackebrandt, E. Reassessment of the phylogenetic position of the bacterium associated with Whipple's disease and determination of the 16S-23S ribosomal intergenic spacer sequence. *Int. J. Syst. Bacteriol.* **1996**, *46*: 1078-1082.

[5] Maiwald, M.; Schuhmacher, F.; Ditton, H.J. and von Herblay, A. Environmental occurrence of the Whipple's disease bacterium (*Tropheryma whippelii*). *Appl. Environ. Microbiol.* **1998**, *64*: 760-762.

[6] Dobbins, W.O., III Whipple's Disease, Hardcover edn. Charles C Thomas Pub Ltd. **1987**.

[7] Bentley, S.D.; Maiwald, M.; Murphy, L.D.; Pallen, M.J.; Yeats, C.A.; Dover, L.G.; Norbertczak, H.T.; Besra, G.S.; Quail, M.A.; Harris, D.E.; von Herblay, A.; Goble, A.; Rutter, S.; Squares, R.; Squares, S.; Barrell, B.G.; Parkhill, J. and Relman, D.A. Sequencing and analysis of the genome of the Whipple's disease bacterium *Tropheryma whipplei*. *Lancet* **2003**, *361*: 637-644.

[8] Raoult, D.; Ogata, H.; Audic, S.; Robert, C.; Suhre, K.; Drancourt, M. and Claverie, J.M. *Tropheryma whipplei* Twist: a human pathogenic Actinobacteria with a reduced genome. *Genome Res.* **2003**, *13*: 1800-1809.

[9] Raoult, D.; Birg, M.L.; La Scola, B.; Fournier, P.E.; Enea, M.; Lepidi, H.; Roux, V.; Piette, J.C.; Vandenesch, F.; Vital-Durand, D. and Marrie T.J. Cultivation of the bacillus of Whipple's disease. *N. Engl. J. Med.* **2000**, *342*: 620-625.

[10] Ghigo, E.; Capo, C.; Aurouze, M.; Tung, C.H.; Gorvel, J.P.; Raoult, D. and Mege, J.L. Survival of *Tropheryma whipplei*, the agent of Whipple's disease, requires phagosome acidification. *Infect. Immun.* **2002**, *70*: 1501-1506.

[11] Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, R.D.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison, C.A. and Venter, J.C. The minimal gene complement of *Mycoplasma genitalium*. *Science* **1995**, *270*: 397-403.

[12] Kaneko, T.; Nakamura, Y.; Sato, S.; Minamisawa, K.; Uchiumi, T.; Sasamoto, S.; Watanabe, A.; Idesawa, K.; Iriguchi, M.; Kawashima, K.; Kohara, M.; Matsumoto, M.; Shimpo, S.; Tsuruoka, H.; Wada, T.; Yamada, M. and Tabata, S. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **2002**, *9*: 189-197.

[13] McGrath, C.L. and Katz, L.A. Genome diversity in microbial eukaryotes. *Trends Ecol. Evol.* **2004**, *19*: 32-38.

[14] Bentley, S.D. and Parkhill, J. Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* **2004**, *38*: 771-792.

[15] Moran, N.A. Microbial minimalism: genome reduction in bacterial pathogens. *Cell* **2002**, *108*: 583-586.

[16] Wernegreen, J.J.; Lazarus, A.B. and Degnan, P.H. Small genome of *Candidatus* Blochmannia, the bacterial endosymbiont of *Camponotus*, implies irreversible specialization to an intracellular lifestyle. *Microbiology* **2002**, *148*: 2551-2556.

[17] Sakharkar, K.R.; Dhar, P.K. and Chow, V.T. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *Int. J. Syst. Evol. Microbiol.* **2004**, *54*: 1937-1941.

[18] Peterson, S.N. and Fraser, C.M. The complexity of simplicity. *Genome Biol.* **2001**, *2*: comment2002.1–comment2002.8.

[19] Bentley, S.D.; Chater, K.F.; Cerdeno-Tarraga, A.M.; Challis, G.L.; Thomson, N.R.; James, K.D.; Harris, D.E.; Quail, M.A.; Kieser, H.; Harper, D.; Bateman, A.; Brown, S.; Chandra, G.; Chen, C.W.; Collins, M.; Cronin, A.; Fraser, A.; Goble, A.; Hidalgo, J.; Hornsby, T.; Howarth, S.; Huang, C.H.; Kieser, T.; Larke, L.; Murphy, L.; Oliver, K.; O'Neil, S.; Rabbinowitsch, E.; Rajandream, M.A.; Rutherford, K.; Rutter, S.; Seeger, K.; Saunders, D.; Sharp, S.; Squares, R.; Squares, S.; Taylor, K.; Warren, T.; Wietzorrek, A.; Woodward, J.; Barrell, B.G.; Parkhill, J. and Hopwood, D.A. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* **2002**, *417*: 141-147.

[20] Akman, L.; Yamashita, A.; Watanabe, H.; Oshima, K.; Shiba,T.; Hattori, M. and Aksoy, S. Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat. Genet.* **2002**, *32*: 402-407.

[21] Renesto, P.; Crapoulet, N.; Ogata, H.; La Scola, B.; Vestris, G.; Claverie, J.M. and Raoult, D. Genome-based design of a cell-free culture medium for *Tropheryma whipplei*. *Lancet* **2003**, *362*: 447-449.

[22] Hueck, C.J. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **1998**, *62*: 379-433.

[23] Cornelis, G.R. and Van Gijsegem, F. Assembly and function of type III secretory systems. *Annu. Rev. Microbiol.* **2000**, *54*: 735-774.

[24] Zusman, T.; Yerushalmi, G. and Segal, G. Functional similarities between the icm/dot pathogenesis systems of *Coxiella burnetii* and *Legionella pneumophila*. *Infect. Immun.* **2003**, *71*: 3714-3723.

[25] Zamboni, D.S.; McGrath, S.; Rabinovitch, M. and Roy, C.R. *Coxiella burnetii* express type IV secretion system proteins that function similarly to components of the *Legionella pneumophila* Dot/Icm system. *Mol. Microbiol.* **2003**, *49*: 965-976.

[26] Vogel, J.; Andrews, H.L.; Wong, S.K. and Isberg, R.R. Conjugative transfer by the virulence system of *Legionella pneumophila. Science* **1998**, *279*: 873-876.

[27] Gogarten, J.P. and Olendzenski, L. Orthologs, paralogs and genome comparisons. *Curr. Opin. Genet. Dev.* **1999**, *9*: 630-636.

[28] Blattner, F.R.; Plunkett, G., III; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B. and Shao, Y. The complete genome sequence of *Escherichia coli* K-12. *Science* **1997**, *277*: 1453-1474.

[29] Zhang, J. Evolution by gene duplication: an update. *Trends Ecol. Evol.* **2003**, *18*: 292-298.

[30] Hughes, D. Evaluating genome dynamics: the constraints on rearrangements within bacterial genomes. *Genome Biol.* **2000**, *1*: reviews0006.1–reviews0006.8.

[31] Frank, A.C.; Amiri, H. and Andersson, S.G. Genome deterioration: loss of repeated sequences and accumulation of junk DNA. *Genetica* **2002**, *115*: 1-12.

[32] Hinrikson, H.P.; Dutly, F. and Altwegg, M. Analysis of the actinobacterial insertion in domain III of the 23S rRNA gene of uncultured variants of the bacterium associated with Whipple's disease using broad-range and '*Tropheryma whippelii*'-specific PCR. *Int. J. Syst. Evol. Microbiol.* **2000**, *50* Pt *3*: 1007-1011.

[33] Kanehisa, M.; Goto, S.; Kawashima, S. and Nakaya, A. The KEGG databases at GenomeNet. *Nucleic Acids Res.* **2002**, *30*: 42-46.

[34] Cannon, W.R. Whipple's disease, genomics, and drug therapy. *Lancet* **2003**, *361*: 1916.

[35] Bork, P. and Koonin, E.V. Predicting functions from protein sequences-where are the bottlenecks? *Nat. Genet.* **1998**, *18*: 313-318.

[36] Brenner, S.E. Errors in genome annotation. *Trends Genet.* **1999**, *15*: 132-133.

[37] Tatusov, R.L.; Galperin, M.Y.; Natale, D.A. and Koonin, E.V. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **2000**, *28*: 33-36.

[38] Mushegian, A.R. and Koonin, E.V. A minimal gene set for cellular life derived by comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*: 10268-10273.

[39] Koonin, E.V. How many genes can make a cell: The minimal-gene-set concept. *Ann. Rev. Genomics Hum. Genet.* **2000**, *1*: 99-116.

[40] Doolittle, R.F. The parasite genome: The grand assault. *Nature* **2002**, *419*: 493-494.

[41] Harris, J.K.; Kelley, S.T.; Spiegelman, G.B. and Pace, N.R. The Genetic Core of the Universal Ancestor. *Genome Res.* **2003**, *13*: 407-412.

[42] Koonin, E.V. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat. Rev. Microbiol.* **2003**, *1*: 127-136.

[43] Andersson, S.G.E. and Kurland, C.G. Reductive evolution of resident genomes. *Trends Microbiol.* **1998**, *6*: 263-268.

[44] Lawrence, J.G.; Hendrix, R.W. and Casjens, S. Where are the pseudogenes in bacterial genomes? *Trends Microbiol.* **2001**, *9*: 535-540.

[45] Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwinn, M.; Hickey, E.K.; Clayton, R.; Ketchum, K.A.; Sodergren, E.; Hardham, J.M.; McLeod, M.P.; Salzberg, S.; Peterson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDonald, L.; Artiach, P.; Bowman, C.; Cotton, M.D.; Fujii, C.; Garland, S.; Hatch, B.; Horst, K.; Roberts, K.; Sandusky, M.; Weidman, J.; Smith, H.O. and Venter, J.C. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **1998**, *281*: 375-388.

[46] Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Gwinn, M.; Dougherty B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; van V.R.; Palmer, N.; Adams, M.D.; Gocayne, J. and Venter, J.C. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi. Nature* **1997**, *390*: 580-586.

[47] Shira, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ishii, K.; Hattori, M.; Kuhara, S. and Nakazawa, T. Comparison of whole genome sequences of *Chlamydia pneumoniae* J138 from Japan and CWL029 from USA. *Nucleic Acids Res.* **2000**, *28*: 2311-2314.

[48] Read, T.D.; Brunham, R.C.; Shen,C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey, E.K.; Peterson, J.; Utterback, T.; Berry, K.; Bass, S.; Linher, K.; Weidman, J.; Khouri, H.; Craven, B.; Bowman, C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; Eisen, J. and Fraser, C.M. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **2000**, *28*: 1397-1406.

[49] Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y. and Cassell, G.H. The complete sequence of the mucosal pathogen *Ureaplasma urealyticum. Nature* **2000**, *407*: 757-762.

[50] Wernegreen, J.J. Endosymbiosis: Lessons in Conflict Resolution. *PLoS Biol.* **2004**, *2*: 307-311.

[51] Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y. and Ishikawa, H. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. *Nature* **2000**, *407*: 81-86.

[52] Seshadri, R.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Nelson, K.E.; Nelson, W.C.; Ward, N.L.; Tettelin, H.; Davidsen, T.M.; Beanan, M.J.; Deboy, R.T.; Daugherty, S.C.; Brinkac, L.M.; Madupu, R.; Dodson, R.J.; Khouri, H.M.; Lee, K.H.; Carty, H.A.; Scanlan, D.; Heinzen, R.A.; Thompson, H.A.; Samuel, J.E.; Fraser, C.M. and Heidelberg, J.F. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii. Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*: 5455-5460.

[53] Hackstadt, T. and Williams, J.C. Biochemical stratagem for obligate parasitism of eukaryotic cells by *Coxiella burnetii. Proc. Natl. Acad. Sci. U. S. A* **1981**, *78*: 3240-3244.

[54] Rovery, C.; Renesto, P.; Crapoulet, N.; Matsumoto, K.; Parola, P.; Ogata, H. and Raoult, D. Transcriptional response of *Rickettsia conorii* exposed to temperature variation and stress starvation. *Res. Microbiol.* **2005**, *156*: 211-218.

[55] Tao, H.; Bausc, C.; Richmond, C.; Blattner, F.R. and Conway, T. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. *J. Bacteriol.* **1999**, *181*: 6425-6440.

[56] Wei, Y.; Lee, J.M.; Richmond, C.; Blattner, F.R.; Rafalski, J.A. and LaRossa, R.A. High-density microarray-mediated gene expression profiling of *Escherichia coli. J. Bacteriol.* **2001**, *183*: 545-556.

[57] Konkel, M.E. and Tilly, K. Temperature-regulated expression of bacterial virulence genes. *Microbes Infect.* **2000**, *2*: 157-166.

[58] La Scola, B.; Fenollar, F.; Fournier, P.E.; Altwegg, M.; Mallet, M.N. and Raoult, D. Description of *Tropheryma whipplei* gen. nov., sp. nov., the Whipple's disease bacillus. *Int. J. Syst. Evol. Microbiol.* **2001**, *51*: 1471-1479.

[59] Freiberg, C.; Brotz-Oesterhelt, H. and Labischinski, H. The impact of transcriptome and proteome analyses on antibiotic drug discovery. *Curr. Opin. Microbiol.* **2004**, *7*: 451-459.

[60] Herrgard, M.J.; Covert, M.W. and Palsson, B.O. Reconstruction of microbial transcriptional regulatory networks. *Curr. Opin. Biotechnol.* **2004**, *15*: 70-77.

[61] Sabatti, C.; Rohlin, L.; Oh, M.K. and Liao, J.C. Co-expression pattern from DNA microarray experiments as a tool for operon prediction. *Nucleic Acids Res.* **2002**, *30*: 2886-2893.

[62] Salama, N.; Guillemin, K.; McDaniel, T.K.; Sherlock, G.; Tompkins, L. and Falkow, S. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*: 14668-14673.

[63] Ochman, H. and Moran, N.A. Genes lost and genes found: evolution of bacterial pathogenesis and symbiosis. *Science* **2001**, *292*: 1096-1099.

[64] Boulos, A.; Rolain, J.M.; Mallet, M.N. and Raoult, D. Molecular evaluation of antibiotic susceptibility of *Tropheryma whipplei* in axenic medium. *J. Antimicrob. Chemother.* **2005**, *55*: 178-181.

[65] Masselot, F.; Boulos, A.; Maurin, M.; Rolain, J.M. and Raoult, D. Molecular evaluation of antibiotic susceptibility: *Tropheryma whipplei* paradigm. *Antimicrob. Agents Chemother.* **2003**, *47*: 1658-1664.

[66] Boulos, A.; Rolain, J.M. and Raoult, D. Antibiotic susceptibility of *Tropheryma whipplei* in MRC5 cells. *Antimicrob. Agents Chemother.* **2004**, *48*: 747-752.

[67] Willems, H.; Thiele, D.; Frolich-Ritter, R. and Krauss, H. Detection of *Coxiella burnetii* in cow's milk using the polymerase chain reaction (PCR). *Zentralbl. Veterinarmed. B* **1994**, *41*: 580-587.

[68] Fenollar, F.; Fournier, P.E.; Raoult, D.; Gerolami, R.; Lepidi, H. and Poyart, C. Quantitative detection of *Tropheryma whipplei* DNA by real-time PCR. *J. Clin. Microbiol.* **2002**, *40*: 1119-1120.

[69] Fenollar, F.; Fournier, P.E.; Robert, C. and Raoult, D. Use of genome selected repeated sequences increases the sensitivity of PCR detection of *Tropheryma whipplei*. *J. Clin. Microbiol.* **2004**, *42*: 401-403.

[70] Fournier, P.E.; Zhu, Y.; Ogata, H. and Raoult, D. Use of highly variable intergenic spacer sequences for multispacer typing of *Rickettsia conorii* strains. *J. Clin. Microbiol.* **2004**, *42*: 5757-5766.

[71] Wu, M.; Sun, L.V.; Vamathevan, J.; Riegler, M.; Deboy, R.; Brownlie, J.C.; McGraw, E.A.; Martin, W.; Esser, C.; Ahmadinejad, N.; Wiegand, C.; Madupu, R.; Beanan, M.J.; Brinkac, L.M.; Daugherty, S.C.; Durkin, A.S.; Kolonay, J.F.; Nelson, W.C.; Mohamoud, Y.; Lee, P.; Berry, K.; Young, M.B.; Utterback, T.; Weidman, J.; Nierman, W.C.; Paulsen, I.T.; Nelson, K.E.; Tettelin, H.; O'Neill, S.L. and Eisen, J.A. Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2004**, *2*: E69.

[72] Horn, M.; Collingro, A.; Schmitz-Esser, S.; Beier, C.L.; Purkhold, U.; Fartmann, B.; Brandt, P.; Nyakatura, G.J.; Droege, M.; Frishman, D.; Rattei, T.; Mewes, H.W. and Wagner, M. Illuminating the evolutionary history of chlamydiae. *Science* **2004**, *304*: 728-730.

[73] Greub, G.; Collyn, F.; Guy, L. and Roten, C.A. A genomic island present along the bacterial chromosome of the *Parachlamydiaceae* UWE25: an obligate amoebal endosymbiont, encodes a potentially functional F-like conjugative DNA transfer system. *BMC. Microbiol.* **2004**, *4*: 48.