AIX-MARSEILLE UNIVERSITÉ
FACULTÉ DE MÉDECINE DE MARSEILLE
ÉCOLE DOCTORALE DES SCIENCES DE LA VIE ET DE LA SANTÉ

T H È S E

Présentée et publiquement soutenue devant

LA FACULTÉ DE MÉDECINE DE MARSEILLE

Le 13 décembre 2013

Par M. Erwin SENTAUSA
Né le 16 décembre 1979 à Malang, Indonésie

INTRASPECIES COMPARATIVE GENOMICS OF *RICKETTSIA*

Pour obtenir le grade de DOCTORAT d'AIX-MARSEILLE UNIVERSITÉ

SPÉCIALITÉ : PATHOLOGIE HUMAINE - MALADIES INFECTIEUSES

Membres du Jury de la Thèse :

Dr. Patricia RENESTO	Rapporteur
Pr. Max MAURIN	Rapporteur
Dr. Florence FENOLLAR	Membre du Jury
Pr. Pierre-Edouard FOURNIER	Directeur de thèse

Unité de Recherche sur les Maladies Infectieuses et Tropicales Émergentes

UM63, CNRS 7278, IRD 198, Inserm 1095

Avant Propos

Le format de présentation de cette thèse correspond à une recommandation de la spécialité Maladies Infectieuses et Microbiologie, à l'intérieur du Master de Sciences de la Vie et de la Santé qui dépend de l'Ecole Doctorale des Sciences de la Vie de Marseille. Le candidat est amené à respecter des règles qui lui sont imposées et qui comportent un format de thèse utilisé dans le Nord de l'Europe permettant un meilleur rangement que les thèses traditionnelles. Par ailleurs, la partie introduction et bibliographie est remplacée par une revue envoyée dans un journal afin de permettre une évaluation extérieure de la qualité de la revue et de permettre à l'étudiant de le commencer le plus tôt possible une bibliographie exhaustive sur le domaine de cette thèse. Par ailleurs, la thèse est présentée sur article publié, accepté ou soumis associé d'un bref commentaire donnant le sens général du travail. Cette forme de présentation a paru plus en adéquation avec les exigences de la compétition internationale et permet de se concentrer sur des travaux qui bénéficieront d'une diffusion internationale.

Professeur Didier RAOULT

TABLE OF CONTENT

ABSTRACT

The *Rickettsia* genus is composed of Gram-negative, obligate intracellular bacteria that cause a range of human diseases around the world. New techniques have led to progress in the identification and classification of *Rickettsia*, including the introduction of molecular methods like sequence comparison (16S rRNA, *omp*A, *omp*B, *glt*A, *sca*4 …) and the creation of the subspecies status. Genomics and next-generation sequencing have opened a new way to learn more about the pathogenesis and evolution of *Rickettsia*. The first part of this thesis is a review on the advantages and limitations of genomics in prokaryotic taxonomy, while the second part consists of the genomic analyses of five *Rickettsia* subspecies and a new *Rickettsia* species. Using high-throughput sequencing methods, we obtained the draft genomes of *R. sibirica sibirica*, *R. sibirica mongolitimonae*, *R. conorii indica*, *R. conorii caspia*, *R. conorii israelensis*, and *R. gravesii*. This work can be a basis of further studies to increase the understanding on the disease-causing mechanisms, evolutionary relationships, and taxonomy of rickettsiae.

RÉSUMÉ

Le genre *Rickettsia* est composé de bactéries Gram-négatives, intracellulaires obligatoires qui causent un éventail de maladies humaines à travers le monde. Des nouvelles techniques ont permis de progresser dans l'identification et la classification des *Rickettsia*, y compris l'introduction de méthodes moléculaires comme la comparaison de séquences de gènes (ARNr 16S, *omp*A, *omp*B, *glt*A, *sca*4 …) et la création du statut de sous-espèce. La génomique et les techniques de séquençage de nouvelle génération ont permis d'accéder à une nouvelle façon d'en apprendre davantage sur la pathogenèse et l'évolution de *Rickettsia*. La première partie de cette thèse est une revue sur les avantages et les limites de la génomique en taxonomie des procaryotes, tandis que la seconde partie est constituée des analyses génomiques de cinq sous-espèces de *Rickettsia* et une nouvelle espèce de *Rickettsia*. En utilisant des méthodes de séquençage à haut débit, nous avons obtenu les génomes de *R. sibirica sibirica*, *R. sibirica mongolitimonae*, *R. conorii indica*, *R. conorii caspia*, *R. conorii israelensis* et *R. gravesii*. Ce travail constitue la base d'autres études qui permettront de mieux comprendre les mécanismes physiopathologiques, l'évolution, et la taxonomie des rickettsies.

GENERAL INTRODUCTION

The genus *Rickettsia* consists of Gram-negative, obligate intracellular bacteria that cause a range of human diseases around the world, including tick-borne rickettsioses caused by spotted-fever group rickettsiae and typhus caused by the typhus group rickettsiae [1,2]. Those diseases have had a longstanding reputation as severe infectious diseases leading to death and disability. Moreover, several *Rickettsia* species that were considered nonpathogenic for decades are now associated with human infections, while new species of undetermined pathogenicity continue to be detected.

New techniques have led to progress in the identification and classification of *Rickettsia*. Previously, the term "rickettsia" was used to name many rod-shaped bacteria that could not be cultured and were not otherwise identified [3]. The introduction of molecular methods, such as 16S rRNA gene sequence comparison, has deeply revised the definition of the term and has allowed new taxonomic and phylogenetic inferences. Furthermore, multi-genic approaches have been used to investigate rickettsial species relationships as well as to develop a taxonomic strategy [4]. Using this strategy, several groups have identified new species around the world [5–20], and the creation of the subspecies status has been proposed [21,22].

The advent of genomics and next-generation sequencing has also given us more understanding on the pathogenesis and evolutionary relationships among rickettsial species [23]. As of October 2013, 55 rickettsial genomes are available, either as complete genome sequences, scaffolds, or contigs, in the databases of the National Center for Biotechnology Information at the National Institutes of Health (http://www.ncbi.nlm.nih.gov/genome/). Genomics show that, compared to less virulent species, highly pathogenic *Rickettsia* undergo extreme genome reduction and massive gene loss, contradictory to the concept of bacterial pathogenicity by acquisition of virulence factors. *Rickettsia* have degraded genomes with split genes, high fractions of paralogous genes, repeated sequences, mobile insertion elements involved in host-cell interaction processes, and ability for gene exchange despite their intracellular habitat.

The present thesis deals with the study of several rickettsial genomes. First, we review the usage of genomics in prokaryotic taxonomy. The subsequent part of the thesis pertains to the analyses as well as the comparisons of *R. sibirica* and *R. conorii* subspecies genomes, respectively. *R. sibirica* and *R. conorii* are two pathogenic species suggested to consist of several subspecies that have different virulence [21,22]. At the end, we present the genomic study of *R. gravesii*, a novel *Rickettsia* of unknown pathogenicity isolated in Australia [6]. The rickettsial genomics studies are

presented in six articles which were published or accepted for publication in scientific journals.

My contribution in the articles includes protein-coding sequence identification, manual curation of the annotation for split genes or nonpredicted genes, RNA detection, orthologous gene detection, and manuscript preparation. In addition, for article 6 I carried out the sequence assembly, contig ordering, and protein-coding sequence identification and annotation.

REVIEW: Advantages and Limitations of Genomics in Prokaryotic Taxonomy

Erwin Sentausa and Pierre-Edouard Fournier

Clinical Microbiology and Infection 2013; 19: 790–795

REVIEW 10.1111/1469-0691.12181

Advantages and limitations of genomics in prokaryotic taxonomy

E. Sentausa and P.-E. Fournier
URMITE, UM63, CNRS 7278, IRD 198, Inserm 1095, Aix Marseille Université, Marseille, France

Abstract

Taxonomic classification is an important field of microbiology, as it enables scientists to identify prokaryotes worldwide. Although the current classification system is still based on the one designed by Carolus Linnaeus, the currently available genomic content of several thousands of sequenced prokaryotic genomes represents a unique source of taxonomic information that should not be ignored. In addition, the development of faster, cheaper and improved sequencing methods has made genomics a tool that has a place in the workflow of a routine microbiology laboratory. Thus, genomics has reached a stage where it may be used in prokaryotic taxonomic classification, with criteria such as the genome index of average nucleotide identity being an alternative to DNA–DNA hybridization. However, several hurdles remain, including the lack of genomic sequences of many prokaryotic taxonomic representatives, and consensus procedures to describe new prokaryotic taxa that do not, as yet, accommodate genomic data. We herein review the advantages and disadvantages of using genomics in prokaryotic taxonomy.

Keywords: Average nucleotide identity, DNA–DNA hybridization, genome-based phylogeny, genomics, prokaryotic taxonomy
Article published online: 13 March 2013
Clin Microbiol Infect 2013; **19:** 790–795

Corresponding author: P.-E. Fournier, URMITE, UM63, CNRS 7278, IRD 198, INSERM U1095, Aix Marseille Université, 27 Blvd Jean Moulin, 13385 Marseille Cedex 5, France
E-mail: Pierre-Edouard.Fournier@univmed.fr

Introduction

Taxonomy, the study of organism classification, is a part of systematics, the study of the diversity and relationships among organisms. Prokaryotic taxonomy is traditionally regarded as consisting of three separate, but interrelated, areas: classification, nomenclature, and characterization. Classification is the arrangement of organisms into taxonomic groups on the basis of similarities; nomenclature is the assignment of names to the taxonomic groups identified in the classification; and characterization is the determination of whether an isolate is a member of a taxon defined in the classification and named in the nomenclature [1]. The influence of prokaryotic taxonomy is tremendous: attaching a name to a microbial strain conveys assumptions and implications associated with that organism, such as routine identification from clinical samples, pathogenicity potential, safety of handling, and cost [2]. However, there is no universal agreement on the rules and criteria used for microorganism classification.

Taxonomic classification has long been based solely on phenotypic characteristics, genetic data having being used only since the 1960s. However, the sequencing of the first bacterial genome in 1995 [3] substantially changed microbiology, by giving access to the whole genetic repertoire of a strain. It is now possible to generate whole prokaryotic genome sequences in a very short period of time, offering the possibility of using the whole genomic sequence of a prokaryote for its taxonomic description. In this review, we explore the benefits and shortcomings of using genomic data in prokaryotic taxonomy.

Historical Overview and Current Practice in Prokaryotic Taxonomy

Although Carolus Linnaeus set the bases of modern taxonomy in the 18th century by studying plants, it was not before the late 19th century that Ferdinand Cohn classified bacteria into genera and species. Cohn and his contemporaries used

morphology, growth requirements, chemical reactions and pathogenic potential as the basis for bacterial classification [4]. Later, biochemical and physiological properties were also used by the Society of American Bacteriologists (which later became the American Society for Microbiology) in a report on bacterial characterization and classification that became the basis for the first edition of *Bergey's Manual of Determinative Bacteriology* in 1923. In 1947, a Code of Bacteriological Nomenclature was approved at the 4th International Congress for Microbiology [5]. In the 1960s, the technique of DNA–DNA hybridization (DDH) was introduced to measure genetic relatedness [6], but it was only widely accepted for classification purposes more than 20 years later [7]. In the 1980s, the development of PCR and sequencing of the 16S rRNA gene led to major changes in prokaryotic taxonomy [8], and this tool, although already commonly used for the description of new species in the 1990s, was recommended in 2002 as a key parameter in taxonomic classification [9,10].

Although prokaryotic nomenclature is regulated in the International Code of Nomenclature of Prokaryotes or the 'Bacteriological Code' [11], which is the latest edition of the Code of Bacteriological Nomenclature and is overseen by the International Committee on Systematics of Prokaryotes (ICSP), there has been no officially recognized system for the characterization and classification of prokaryotes until now. However, the most widely used system of characterization relies on a polyphasic approach, which is also used in the most widely accepted classification presented in *Bergey's Manual of Systematic Bacteriology* [4,12].

The term 'polyphasic taxonomy' was introduced in 1970 to refer to a taxonomy that brings together and incorporates many levels of information, from ecological to molecular, and includes several distinct types of information to yield a multidimensional classification. Currently, polyphasic taxonomy refers to a taxonomy that aims to utilize all available data [13]. These data include both phenotypic information, such as chemotaxonomic features (cell wall compounds, quinones, polar lipids, etc.), morphology, staining behaviour, and culture characteristics (medium, temperature, incubation time, etc.), and genetic properties, such as G+C content, DDH value, and 16S rRNA gene sequence identity with other closely related species with validated names [14].

Currently, the most commonly used tool for evaluating the phylogenetic position of a prokaryote is 16S rRNA gene sequence comparison. Likewise, the latest whole taxonomic schema for prokaryotic diversity presented in *Bergey's Manual* uses 16S rRNA phylogeny as its main basis [15]. However, there is growing interest in the use of other genes (protein-encoding genes) to resolve issues that are not solved by 16S rRNA gene sequencing. For example, some housekeeping genes (e.g. *dnaJ*, *dnaK*, *gyrB*, *recA*, and *rpoB*) have been used instead in multilocus sequence typing/multilocus sequence analysis (MLSA) [16]. One limitation of 16S rRNA is that it is rather conserved, and hence is not universally reliable for determination of taxonomic relationships at the species level. Furthermore, both nucleotide variations within multiple rRNA operons in a single genome and the possibility of 16S rRNA genes being derived from horizontal gene transfer may distort relationships between taxa in phylogenetic trees [17]. Nevertheless, 16S rRNA is currently the first-line tool for evaluating the taxonomic status of a prokaryotic strain at the same genus or species levels. It is currently assumed that two strains are members of the same species if their 16S rRNA gene sequence identity is >99%, and it may provide the first indication that a novel species has been isolated if an identity of <98.7% is found [18]. Similarly, a 16S rRNA identity of <95% with the phylogenetically closest species with a validated name may suggest that the isolate is a representative of a new genus.

Another widely used taxonomic criterion is DDH. A DDH value of $\geq$70% has been recommended as a threshold for the definition of members of a species, and DDH is deemed necessary when strains share >98.7% 16S rRNA gene sequence identity [12,14]. However, the DDH cut-off used is not applicable to all prokaryotic genera. For example, when applied to *Rickettsia* species, a DDH of 70% would not discriminate *Rickettsia rickettsii*, *Rickettsia conorii*, *Rickettsia sibirica*, and *Rickettsia montanensis* [19]. In addition, DDH protocols are considered to be tedious and complicated, with inherently large degrees of error, and only a few laboratories are equipped for this method, which remains expensive and is clearly not adapted to routine microbiology [2,20]. Furthermore, DDH studies can provide only a rough measurement of average genetic relationship, only closely related species or subspecies can be distinguished, and incremental databases cannot be developed for this method [4].

The Prokaryotic Genomic Era

The sequencing of the *Haemophilus influenzae* genome in 1995 by conventional Sanger sequencing was a landmark in modern biology, as it marked the beginning of the genomic era [3]. However, in the next decade, bacterial genome sequencing remained time-consuming and expensive, and was reserved to a few sequencing centres worldwide. Thanks to the next-generation sequencing (NGS) technologies introduced from 2005, the number of sequenced prokaryotic genomes has rapidly increased, as new platforms are much faster and cheaper [21]. As of 18 September 2012, the Genome online Database listed 3381 prokaryotic genomes available as either

full genome sequences, scaffolds, or contigs, and 11 789 other prokaryotic genome projects are ongoing (http://www.genomesonline.org/cgi-bin/GOLD/index.cgi).

The current commercially available NGS platforms can be divided into two categories: the high-end instruments and the bench-top instruments [21]. The high-end instruments can produce long reads and deliver dozens to thousands of prokaryotic genomes per run, but are too expensive for the average research laboratory; the bench-top instruments are modestly priced, and have lower throughput, but are also fast and considered to be better for most applications in microbiology [22]. The 454 GS FLX+, Illumina's HiSeq 2000/2500, Life Technologies' 5500xl SOLiD and Pacific Biosciences' PacBio *RS* are the latest high-end instruments, one of which has an output of up to 600 Gb per run, whereas 454 GS Junior, Life Technologies' Ion PGM and Ion Proton and Illumina's MiSeq are bench-top instruments that are able to sequence a complete prokaryotic genome in a few days.

NGS technology has already transformed microbiology and the way in which people study prokaryotes. Genome sequencing has made possible the development of specific culture media for several prokaryotes, and enabled us to more easily identify bacterial pathogens, test their antibiotic resistance and virulence, and track their emergence and spread [22,23]. Sequencing is now replacing microarrays as the method of choice for studying gene expression (with RNA sequencing), mutant libraries (with Tn-seq and transposon-directed insertion site sequencing), and protein–DNA interactions (with chromatin immunoprecipitation followed by sequencing) [21]. Finally, it is no longer an absolute requirement to obtain large quantities of highly purified DNA for sequencing of a prokaryotic genome, as full genome sequencing from a complex microbial community and sequencing from a single cell are also possible, although the former method provides only an average sequence of a group of a closely related but not necessarily clonal population [24,25].

Can Genome Sequences be Used in Prokaryotic Taxonomy?

Over the past 10 years, scientists have attempted to use genomes to assess the phylogenetic relationships between organisms, with a variety of techniques being used, including examination of the order of the genes, analysis of core genes (presence or absence or sequence alignment), indels or single-nucleotide polymorphisms in core genes, and the construction of super-trees (phylogenetic trees assembled from a combination of smaller phylogenetic trees) [17,26]. As argued by Klenk and Göker [27], genome-scale data for phylogenetic reconstruction are advantageous, as genome sequences provide more characters to be analysed, and this, in general, improves the phylogenetic signal/noise ratio. Moreover, genomic information such as gene content, gene order and rare genomic rearrangements is complementary to the data provided by the nucleotide sequence. It was also argued that, although horizontal gene transfer might be very widespread in prokaryotes, it has not been proven to hinder phylogenetic reconstruction from genomic data. The vast majority of genes and genetic markers that are distinctive of higher prokaryotic taxa are vertically inherited, and a solid foundation for microbial systematics can be developed on the basis of these [28]. Indeed, Zhi *et al.* argued that trees based on the comparison of orthologous genes have reasonably good congruence with those built by comparison of 16S rRNA sequences, and, to some extent, with trees based on the presence and absence of genes [17]. For some recent examples, Thompson *et al.* demonstrated that the phylogenetic tree of vibrios obtained with the 16S rRNA gene is similar to that obtained with MLSA [29], and Bennet *et al.* found similar results for *Neisseria* when using multilocus sequence typing of 53 ribosomal protein subunits [30]. However, there is also an opposing view that a phylogenetic tree based on a single gene does not necessarily reflect the history of prokaryotes, as pointed out by Doolittle and Bapteste [31].

Whereas genome-base phylogeny has been the subject of a substantial number of publications, data on genome-based taxonomy remain scarce. In 2011, Whitman [32] recommended the routine description of prokaryotic species on the basis of their genomic sequences. In this way, type strains would be uniquely and unambiguously identified, and redundancy of nomenclature would be impossible. The genomic sequences would not only establish the genetic identity, but would also provide a diagnosis of the species with a precision unimaginable at the time when the Code was written. However, Kämpfer and Glaeser argued that genes and genomes do not function on their own, and can only display their potential within the cell as the basic unit of evolution and hence taxonomy [13]. Therefore, the 'minimalist' and/or genomic approach to descriptions of novel taxa must not abandon fundamental principles of taxonomy, including the incorporation of phenotypic data and requirements for strain deposition in culture collections.

Current genetic taxonomic criteria include several numerical cut-offs, notably DDH. Therefore, several authors studied the correlation between the percentage of nucleotide sequence similarity at the core genome level and DNA–DNA reassociation results. In particular, the average nucleotide identity (ANI) and MLSA have been suggested to be valid alternatives to DDH [33,34]. ANI, defined as the mean percentage of nucleotide sequence identity of orthologous genes shared by two genomes, seems to reproduce DDH

results with more accuracy. Two prokaryotic strains may be considered as belonging to the same species if they share a $\geq$96% ANI value, this cut-off being equivalent to the 70% DDH value. In addition, ANI studies can be performed *in silico* with public databases, and Richter and Rosselló-Móra even proposed that reliable ANI values may be obtained from the comparison of sequences covering ~20% of each genome [35]. In addition to ANI, other parameters, such as the maximal unique matches index, defined as a genomic distance index based on both DNA conservation of the core genome and the proportion of DNA shared by two genomes [36], and 'tetranucleotide regression', defined as the differences between observed and expected values of the frequencies of all 256 possible tetranucleotide (A, T, G, C) combinations [35], have been proposed to help evaluate the species status of a strain based on genome data. Furthermore, the genome-to-genome distance calculator can be used to calculate the genomic distance on the basis of the total length of all high-scoring segment pairs identified by a BLAST search of the genome [37,38]. The results of ANI, the maximal unique matches index and the genome-to-genome distance calculator have been suggested to have a high correlation with DNA–DNA relatedness. However, the value of ANI is, at present, unbeatable, because it most probably reflects what experimentally occurs when two DNAs are hybridized in DDH experiments [39]. In 2010, Tindall *et al.* [14] suggested, in a 'taxonomic note' on the characterization of prokaryote strains published in the *International Journal of Systematic and Evolutionary Microbiology*, the official publication of the ICSP, that ANI may substitute for DDH analyses in the near future. With the rapid development and decreasing cost of high-throughput prokaryotic genome sequencing technology (with the imminent possibility of having a $1 bacterial genome sequence [21]), this proposition seems reasonable. ANI has been used recently, for instance, to describe new species of *Burkholderia*, *Geobacter*, and *Vibrio*, as well as to help characterize a new subspecies of *Francisella*, a new genus of *Sphaerochaeta*, and a new class of *Dehalococcoidetes* [40–45].

However, several current drawbacks limit the use of genomics for systematics. First, Klenk and Göker pointed out that completely sequenced genomes for many of the major lineages of prokaryotes are lacking [27]. The currently available genome sequences have been obtained mostly from three phyla (*Proteobacteria*, *Firmicutes*, and *Actinobacteria*). Thus, many phyla are poorly represented in genomics (http://www.genomesonline.org/cgi-bin/GOLD/index.cgi). Furthermore, the same authors noted that, even if the genome sequences of the species of interest are available, in many cases they are not type strains, and, therefore must be used with caution, as prokaryote taxonomy is based on type strains only [14]. However, efforts such as the phylogeny-driven Genomic Encyclopedia of *Bacteria* and *Archaea* programme, which aims to sequence all type strains [46], should help to fill the gaps, even though Zhi *et al.* argued that the increasing number of available genomes currently remains highly biased towards organisms of biotechnological and medical importance [17]. Another problem is that existing genomic sequences vary greatly in their finished quality, often being available only as unfinished draft assemblies that, according to Ricker *et al.* and Klassen *et al.*, may be less informative than finished whole genome sequences [47,48]. For that reason, minimal sequencing quality should be defined for genomes to be included in taxonomic analyses. For example, the guidelines developed by the Next-generation Sequencing: Standardization of Clinical Testing work group might be utilized for this purpose [49]. Moreover, Ozen *et al.* argued that the results obtained with whole genome-based tools such as ANI do not consistently agree with current taxonomy, and different methods should be used for the different levels of taxonomy, as they stated that there is not one universal method with which to naturally classify prokaryotes [50]. However, Sutcliffe *et al.* emphasized that, indeed, the current principles and practice of prokaryotic systematics have not yet fully accommodated genomic data, and that significant revision of the procedures used to describe novel prokaryotic taxa is needed, including the likely introduction of new publication formats [51]. Furthermore, Figueras *et al.* pointed out that consensus genome comparison criteria that are acceptable in prokaryotic taxonomic classification remain to be defined [52].

In our laboratory, we recently included genome sequence analysis in a polyphasic strategy to describe new bacterial species, together with phenotypic data including their matrix-assisted laser desorption ionization time-of-flight mass spectrum, and main phenotypic characteristics (habitat, Gram stain reaction, culture and metabolic characteristics, and, when applicable, pathogenicity) [53]. In our scheme, the degree of nucleotide sequence similarity of orthologous genes between the genome of a putative new bacterial species and the genomes of its most closely related and validly published species should be similar to that observed among these validly published species. Our method differed from the ANI calculation, as we first determined the orthologous protein set between two genomes by BLASTP, using a coverage of $\geq$50% and a degree of amino acid identity of $\geq$30%, and then calculated the mean percentage of nucleotide sequence identity between these orthologous genes (Fig. 1). In contrast, orthologous genes used for ANI determination are identified by a BLASTN search. As an example, the genome from *Peptoniphilus senegalensis* sp. nov., isolated from a Senegalese patient's stool, shared 976, 977 and 1195 orthologous genes

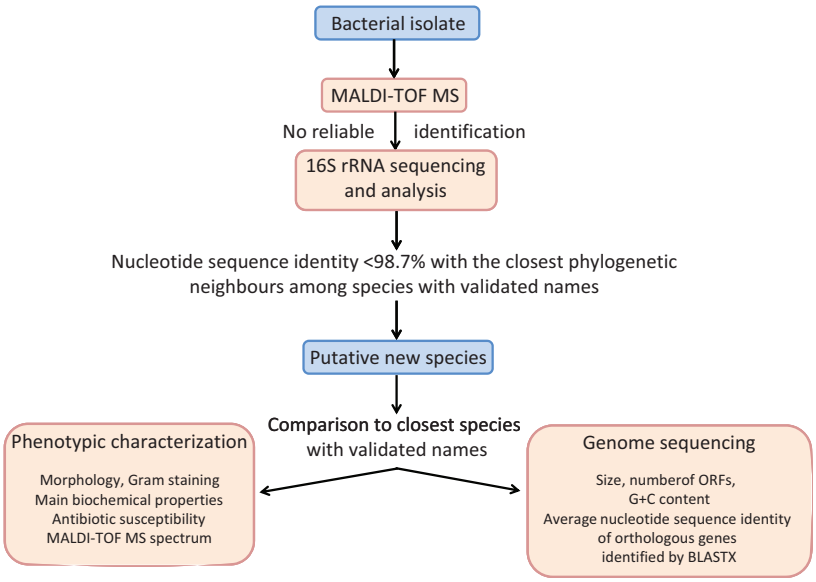

FIG. 1. Current strategy used in our laboratory to describe novel prokaryotes. MALDI-TOF MS, matrix-assisted laser desorption ionization time-of-flight mass spectrometry; ORF, open reading frame.

(86.9%, 87.08% and 86.48% mean orthologous gene nucleotide similarity) with *Peptoniphilus lacrimalis*, *Peptoniphilus indolicus*, and *Peptoniphilus harei*, respectively [54]. These values were similar to those observed among validly published *Peptoniphilus* genomes, as *P. indolicus* shared 942 and 1078 orthologous genes (87.06% and 86.78% mean similarity) with *P. lacrimalis* and *P. harei*, respectively, and *P. harei* and *P. lacrimalis* shared 1 095 orthologous genes and 87.35% mean similarity. Therefore, both genomic and phenotypic data were consistent with the new species status of *P. senegalensis* sp. nov.

Conclusions

The current availability of >3000 prokaryotic genome sequences, including those from most of the major human pathogens, offers the opportunity to make use of the total genetic content of prokaryotes for their taxonomic classification. However, as ANI or other genomic comparison markers may replace DDH as a standard to circumscribe prokaryotic species in the very near future, several challenges remain, in particular the need to define a genomic-based method that is agreed upon by microbiologists, and cut-offs that either apply to most prokaryotes or vary according to taxonomic groups. In addition, although the integration of genomic data into prokaryotic taxonomic classification seems unavoidable in the near future, genome sequences should always be included in a polyphasic strategy in combination with phenotypic data. Thus, procedures to describe new prokaryotic taxa need a reassessment to accommodate genomic data while genome sequences of more prokaryotic taxonomic representatives or under-represented taxa are looked for.

Transparency Declaration

None of the authors of the present manuscript have a commercial or other association that might pose a conflict of interest (e.g. pharmaceutical stock ownership or consultancy). This work was supported by a grant from the Méditerranée Infection Foundation.

References

1. Brenner DJ, Staley JT, Krieg NR. Classification of procaryotic organisms and the concept of bacterial speciation. In: Brenner DJ, Krieg NR, Staley JT, eds. *Bergey's manual of systematic bacteriology. 2nd edn. Volume 2: The proteobacteria. Part A: Introductory essays*. New York: Springer, 2005; 27–32.
2. Moore ERB, Mihaylova SA, Vandamme P, Krichevsky MI, Dijkshoorn L. Microbial systematics and taxonomy: relevance for a microbial commons. *Res Microbiol* 2010; 161: 430–438.
3. Fleischmann RD, Adams MD, White O *et al.* Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 1995; 269: 496–512.
4. Schleifer KH. Classification of Bacteria and Archaea: past, present and future. *Syst Appl Microbiol* 2009; 32: 533–542.
5. Stackebrandt E. Forces shaping bacterial systematics. *Microbe* 2007; 2: 283–288.
6. McCarthy BJ, Bolton ET. An approach to the measurement of genetic relatedness among organisms. *Proc Natl Acad Sci USA* 1963; 50: 156–164.
7. Wayne LG, Brenner DJ, Colwell RR *et al.* Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* 1987; 37: 463–464.
8. Woese CR. Bacterial evolution. *Microbiol Rev* 1987; 51: 221–271.
9. Stackebrandt E, Frederiksen W, Garrity GM *et al.* Report of the ad hoc committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* 2002; 52: 1043–1047.
10. Rosselló-Móra R. Updating prokaryotic taxonomy. *J Bacteriol* 2005; 187: 6255–6257.

11. Lapage SP, Sneath PHA, Lessel EF, Skerman VBD, Seeliger HPR, Clark WA, eds. *International code of nomenclature of bacteria (1990 revision)*. Washington, DC: American Society for Microbiology, 1992.
12. Gillis M, Vandamme P, Vos P, Swings J, Kersters K. Polyphasic taxonomy. In: Brenner DJ, Krieg NR, Staley JT, eds. *Bergey's manual of systematic bacteriology. 2nd edn. Volume 2: The proteobacteria. Part A: Introductory essays*. New York: Springer, 2005; 43–48.
13. Kämpfer P, Glaeser SP. Prokaryotic taxonomy in the sequencing era —the polyphasic approach revisited. *Environ Microbiol* 2012; 14: 291–317.
14. Tindall BJ, Rosselló-Móra R, Busse HJ, Ludwig W, Kämpfer P. Notes on the characterization of prokaryote strains for taxonomic purposes. *Int J Syst Evol Microbiol* 2010; 60: 249–266.
15. Ludwig W, Klenk HP. Overview: a phylogenetic backbone and taxonomic framework for procaryotic systematics. In: Brenner DJ, Krieg NR, Staley JT, eds. *Bergey's manual of systematic bacteriology. 2nd edn. Volume 2: The proteobacteria. Part A: Introductory essays*. New York: Springer, 2005; 49–66.
16. Kämpfer P. Systematics of prokaryotes: the state of the art. *Antonie Van Leeuwenhoek* 2012; 101: 3–11.
17. Zhi XY, Zhao W, Li WJ, Zhao GP. Prokaryotic systematics in the genomics era. *Antonie Van Leeuwenhoek* 2012; 101: 21–34.
18. Stackebrandt E, Ebers J. Taxonomic parameters revisited: tarnished gold standards. *Microbiol Today* 2006; 33: 152–155.
19. Drancourt M, Raoult D. Taxonomic position of the Rickettsiae: current knowledge. *FEMS Microbiol Rev* 1994; 13: 13–24.
20. Rosselló-Móra R, Urdiain M, López-López A. DNA–DNA hybridization. *Methods Microbiol* 2011; 38: 325–347.
21. Loman NJ, Constantinidou C, Chan JZM *et al.* High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat Rev Microbiol* 2012; 10: 599–606.
22. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet* 2012; 13: 601–612.
23. Fournier PE, Raoult D. Prospects for the future using genomics and proteomics in clinical microbiology. *Annu Rev Microbiol* 2011; 65: 169–188.
24. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV. Untangling genomes from metagenomes: revealing an uncultured class of marine euryarchaeota. *Science* 2012; 335: 587–590.
25. Lasken RS. Genomic sequencing of uncultured microorganisms from single cells. *Nat Rev Microbiol* 2012; 10: 631–640.
26. Jones AL. The future of taxonomy. *Adv Appl Microbiol* 2012; 80: 23–35.
27. Klenk HP, Göker M. En route to a genome-based classification of Archaea and Bacteria? *Syst Appl Microbiol* 2010; 33: 175–182.
28. Gao B, Gupta R. Microbial systematics in the post-genomics era. *Antonie Van Leeuwenhoek* 2012; 101: 45–54.
29. Thompson C, Vicente A, Souza R *et al.* Genomic taxonomy of vibrios. *BMC Evol Biol* 2009; 9: 258.
30. Bennett JS, Jolley KA, Earle SG *et al.* A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology* 2012; 158: 1570–1580.
31. Doolittle WF, Bapteste E. Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci USA* 2007; 104: 2043–2049.
32. Whitman W. Intent of the nomenclatural Code and recommendations about naming new species based on genomic sequences. *Bull BISMiS* 2011; 2: 135–139.
33. Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA* 2005; 102: 2567–2572.
34. Goris J, Konstantinidis KT, Klappenbach JA, Coenye T, Vandamme P, Tiedje JM. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. *Int J Syst Evol Microbiol* 2007; 57: 81–91.
35. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* 2009; 106: 19126–19131.
36. Deloger M, El Karoui M, Petit MA. A genomic distance based on MUM indicates discontinuity between most bacterial species and genera. *J Bacteriol* 2009; 191: 91–99.
37. Auch AF, Klenk HP, Göker M. Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. *Stand Genomic Sci* 2010; 2: 142–148.
38. Auch AF, Von Jan M, Klenk HP, Göker M. Digital DNA–DNA hybridization for microbial species delineation by means of genome-to-genome sequence comparison. *Stand Genomic Sci* 2010; 2: 117–134.
39. Rosselló-Móra R. Towards a taxonomy of Bacteria and Archaea based on interactive and cumulative data repositories. *Environ Microbiol* 2012; 14: 318–334.
40. Vanlaere E, Baldwin A, Gevers D *et al.* Taxon K, a complex within the *Burkholderia cepacia* complex, comprises at least two novel species, *Burkholderia contaminans* sp. nov. and *Burkholderia lata* sp. nov. *Int J Syst Evol Microbiol* 2009; 59: 102–111.
41. Prakash O, Gihring TM, Dalton DD *et al. Geobacter daltonii* sp. nov., an Fe(III)- and uranium(VI)-reducing bacterium isolated from a shallow subsurface exposed to mixed heavy metal and hydrocarbon contamination. *Int J Syst Evol Microbiol* 2010; 60: 546–553.
42. Hoffmann M, Monday SR, Allard MW *et al. Vibrio caribbeanicus* sp. nov., isolated from the marine sponge *Scleritoderma cyanea*. *Int J Syst Evol Microbiol* 2012; 62: 1736–1743.
43. Mikalsen J, Olsen AB, Tengs T, Colquhoun DJ. *Francisella philomiragia* subsp. *noatunensis* subsp. nov., isolated from farmed Atlantic cod (*Gadus morhua* L.). *Int J Syst Evol Microbiol* 2007; 57: 1960–1965.
44. Ritalahti KM, Justicia-Leon SD, Cusick KD *et al. Sphaerochaeta globosa* gen. nov., sp. nov. and *Sphaerochaeta pleomorpha* sp. nov., free-living, spherical spirochaetes. *Int J Syst Evol Microbiol* 2012; 62: 210–216.
45. Löffler FE, Yan J, Ritalahti KM *et al. Dehalococcoides mccartyi* gen. nov., sp. nov., obligate organohalide-respiring anaerobic bacteria, relevant to halogen cycling and bioremediation, belong to a novel bacterial class, *Dehalococcoidetes* classis nov., within the phylum *Chloroflexi*. *Int J Syst Evol Microbiol* 2012. doi:10.1099/ijs.0.034926-0 [Epub ahead of print].
46. Wu D, Hugenholtz P, Mavromatis K *et al.* A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature* 2009; 462: 1056–1060.
47. Ricker N, Qian H, Fulthorpe RR. The limitations of draft assemblies for understanding prokaryotic adaptation and evolution. *Genomics* 2012; 100: 167–175.
48. Klassen J, Currie C. Gene fragmentation in bacterial draft genomes: extent, consequences and mitigation. *BMC Genomics* 2012; 13: 14.
49. Gargis AS, Kalman L, Berry MW *et al.* Assuring the quality of next-generation sequencing in clinical laboratory practice. *Nat Biotechnol* 2012; 30: 1033–1036.
50. Ozen AI, Vesth T, Ussery DW. From genome sequence to taxonomy —a skeptic's view. In: Rosenberg E, DeLong EF, Stackebrandt E, Lory S, Thompson F, eds. *The prokaryotes*, 4th edn. Berlin: Springer, 2012.
51. Sutcliffe I, Trujillo M, Goodfellow M. A call to arms for systematists: revitalising the purpose and practises underpinning the description of novel microbial taxa. *Antonie Van Leeuwenhoek* 2012; 101: 13–20.
52. Figueras M, Beaz-Hidalgo R, Collado L, Martínez-Murcia A. Recommendations for a new bacterial species description based on analyses of the unrelated genera *Aeromonas* and *Arcobacter*. *Bull BISMiS* 2011; 2: 1–16.
53. Lagier JC, Armougom F, Million M *et al.* Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect* 2012; 18: 1185–1193.
54. Mishra AK, Lagier JC, Robert C, Raoult D, Fournier PE. Non contiguous-finished genome sequence and description of *Peptoniphilus timonensis* sp. nov. *Stand Genomic Sci* 2012; 7: 1–11.

ARTICLE 1: Sequence and Annotation of *Rickettsia sibirica sibirica* Genome

Erwin Sentausa, Khalid El Karkouri, Catherine Robert, Didier Raoult, and Pierre-Edouard Fournier

Journal of Bacteriology 2012; 194: 2377

Commentary to Article 1

Rickettsia sibirica has been known as the etiological agent of Siberian tick typhus (also known as North Asian tick typhus), a spotted fever group rickettsiosis that was first described in Russia in the 1930s. Based on the similarities in 16S rRNA and citrate synthase (*gltA*) and the *Rickettsia*-specific *ompA* and *ompB* genes and gene D, Fournier *et al*. showed that the then newly identified "*R. mongolotimonae*" belongs to the *R. sibirica* species [4]. Based on this and dissimilarities in intergenic sequences as well as serotypic and epidemio-clinical traits, they then proposed that these two rickettsiae be reclassified as subspecies, namely *R. sibirica* subsp. *sibirica* and *R. sibirica* subsp. *mongolitimonae* [22].

This article reports the draft genome sequence and annotation of *R. sibirica sibirica* strain BJ-90. This strain was isolated in 1990 from *Dermacentor sinicus* ticks collected in Beijing, China and was recently demonstrated to cause severe human illness with multiorgan dysfunction [24]. The genome was also compared to the genome of *R. sibirica sibirica* strain 246, the species' type strain, which was previously sequenced.

ARTICLE 2: Genome Sequence of "*Rickettsia sibirica* subsp. *mongolitimonae*"

Erwin Sentausa, Khalid El Karkouri, Catherine Robert, Didier Raoult, and Pierre-Edouard Fournier

Journal of Bacteriology 2012; 194: 2389–2390

Commentary to Article 2

In 1991, a new spotted-fever group rickettsial strain named HA-91 was isolated from *Hyalomma asiaticum* ticks collected in Inner Mongolia, China and designated as a new species with the name "*R. mongolotimonae*" [25,26]. In 1996, it was found to cause a human disease in France and, afterward, in other Mediterranean countries [27]. The organism was then identified as a member of the *R. sibirica* species complex [4], but further genotypic analyses grouped it as a subspecies of *R. sibirica* [22]. This article reports the draft genome sequence and annotation of *R. sibirica mongolitimonae* strain HA-91.

ARTICLE 3: Genome Sequence of *Rickettsia conorii* subsp. *indica*, the Agent of Indian Tick Typhus

Erwin Sentausa, Khalid El Karkouri, Catherine Robert, Didier Raoult, and Pierre-Edouard Fournier

Journal of Bacteriology 2012; 194: 3288–3289

Commentary to Article 3

Indian tick typhus is a tick-borne rickettsiosis prevalent in India and Pakistan and it has been clinically recognized at the beginning of the 20th century. Although the etiologic agent has never been isolated in patients, a spotted-fever group rickettsia that was isolated in 1950 from a *Rhipicephalus sanguineus* tick collected in India was considered to be the cause of the disease [28]. This so called Indian tick typhus rickettsia was classified as *R. conorii*, but then reclassified as *R. conorii* subsp. *indica* based on dissimilarities of 16S rDNA, *glt*A, *omp*A, *omp*B, and *sca*4 genes as well as serotypic and epidemio-clinical characteristics [21]. It was recently detected in a human patient in Sicily [29], while the presumable first death by this subspecies has also been reported [30].

This article reports the draft genome sequence and annotation of *R. conorii indica*. The genome was also compared to the genome of *R. conorii conorii*.

ARTICLE 4: Genome Sequence of *Rickettsia conorii* subsp. *caspia*, the Agent of Astrakhan Fever

Erwin Sentausa, Khalid El Karkouri, Catherine Robert, Didier Raoult, and Pierre-Edouard Fournier

Journal of Bacteriology 2012; 194: 4763–4764

Commentary to Article 4

Rickettsia conorii caspia was first identified as a *R. conorii*-like bacterium that causes Astrakhan fever, a rickettsiosis that has been reported since 1970s in Astrakhan, a region of Russia located by the Caspian Sea [31,32]. Subsequently found in Kosovo [33], Chad [34], and, recently, southern France [35], it was discovered to be antigenically and genomically related to Israel tick typhus rickettsia [36]. In 2005, Zhu et al. [21] proposed that these rickettsiae, together with *R. conorii* sensu stricto and *R. conorii* subsp. *indica*, be grouped as subspecies of *R. conorii*.

This article reports the draft genome sequence and annotation of *R. conorii caspia*. The genome was also compared to the genome of *R. conorii conorii* and *R. conorii indica*.

ARTICLE 5: Genome Sequence of *Rickettsia conorii* subsp. *israelensis*, the Agent of Israeli Spotted Fever

Erwin Sentausa, Khalid El Karkouri, Catherine Robert, Didier Raoult, and Pierre-Edouard Fournier

Journal of Bacteriology 2012; 194: 5130–5131

Commentary to Article 5

The first cases of rickettsial spotted fever in Israel were reported in the late 1940s, while the agent was isolated from a patient in 1971. The bacterium was characterized as closely related to but slightly different from *R. conorii*, and it was later found in patients in Italy [37], Portugal [38,39], Tunisia [40], and, supposedly, Libya [41]. Based on 16S rDNA, *glt*A, *omp*A, *omp*B, and *sca*4 genes as well as serotypic and epidemio-clinical characteristics, this rickettsia has been classified as a subspecies of *R. conorii*, namely *R. conorii israelensis* [21].

This article reports the draft genome sequence of *R. conorii israelensis*. The genome was also compared to the genome of *R. conorii conorii*, *R. conorii indica*, and *R. conorii caspia*.

ARTICLE 6: Genome Sequence of *Rickettsia gravesii*, Isolated from Western Australian Ticks

Erwin Sentausa, Mohammad Yazid Abdad, Catherine Robert, John Stenos, Didier Raoult, and Pierre-Edouard Fournier

Genome Announcements 2013; 1: e00975-13

Commentary to Article 6

Rickettsia gravesii is a new rickettsia isolated from *Amblyomma triguttatum triguttatum* ticks removed from humans on Barrow Island, Western Australia, reported in 2006. Based on its 16S rRNA sequence, it is closely related to the human pathogen *R. massiliae* [6].

This article reports the draft genome sequence and annotation of *R. gravesii*. The genome was also compared to the genome of *R. massiliae*.

genomeA
Journals.ASM.org

Genome Sequence of *Rickettsia gravesii*, Isolated from Western Australian Ticks

Erwin Sentausa,[a] Mohammad Yazid Abdad,[b] Catherine Robert,[a] John Stenos,[b] Didier Raoult,[a] Pierre-Edouard Fournier[a]

Aix-Marseille Université, URMITE, UM63, CNRS 7278, IRD 198, Inserm U1095, Marseille, France[a]; Australian Rickettsial Reference Laboratory, Geelong, Victoria, Australia[b]

***Rickettsia gravesii* is a new *Rickettsia* species closely related to the human pathogen *Rickettsia massiliae*. Here, we describe the genome sequence of *R. gravesii* strain BWI-1, isolated from *Amblyomma triguttatum triguttatum* ticks collected from humans on Barrow Island, Western Australia.**

Received 21 October 2013 **Accepted** 31 October 2013 **Published** 27 November 2013

Citation Sentausa E, Abdad MY, Robert C, Stenos J, Raoult D, Fournier P-E. 2013. Genome sequence of *Rickettsia gravesii*, isolated from Western Australian ticks. Genome Announc. 1(6):e00975-13. doi:10.1128/genomeA.00975-13.

Address correspondence to Pierre-Edouard Fournier, pierre-edouard.fournier@univ-amu.fr.

Rickettsiae are obligate intracellular alphaproteobacteria and are the etiological agents of several arthropod-borne diseases in humans. *Rickettsia gravesii* is a novel species isolated from *Amblyomma triguttatum triguttatum* ticks removed from humans on Barrow Island, Western Australia, after there was anecdotal evidence of a disease possibly of rickettsial origin in the region (1). It was also found in other tick species, such as *Amblyomma limbatum* (2); its distribution so far is recognized to coincide with that of *A. triguttatum triguttatum* (3), and it was found to be highly prevalent in members of the latter tick species collected from feral pigs in the southern part of Western Australia (4). Although its pathogenic potential is currently unknown, *R. gravesii* is closely related to the spotted-fever group species *Rickettsia massiliae* (1), which is pathogenic to humans and prevalent in Europe and Africa (5, 6). Here, we describe the genome sequence of *R. gravesii* strain BWI-1^T.

R. gravesii (deposited in the Collection de Souches de l'Unite des Rickettsies [CSUR] under reference R172) was grown in XTC and L929 cells, and its genomic DNA was extracted using a phenol-chloroform protocol. Sequencing was performed using the MiSeq platform (Illumina, San Diego, CA) with a 2 × 250-bp paired-end run after library preparation with the Nextera XT sample preparation kit (Illumina). *De novo* genome assembly was done using the CLC Genomics Workbench 4.9 (CLC bio, Aarhus, Denmark). The resulting contigs were reordered in Mauve 2.3.1 (7) using the genome sequence from *R. massiliae* strain MTU5 (GenBank accession no. CP000683) (8) as a reference. Open reading frame (ORF) prediction and gene annotation were carried out using RAST 4.0 (9). rRNAs, tRNAs, and other RNAs were identified using BLASTn (10), tRNAscan-SE 1.21 (11), and RNAmmer 1.2 (12), respectively. The orthologous genes between *R. gravesii* and *R. massiliae* MTU5 were identified using OrthoMCL (13), with a BLASTp *E* value cutoff of 1×10^{-5} and the default Markov cluster algorithm (MCL) inflation parameter of 1.5.

The draft genome sequence of *R. gravesii* BWI-1^T is made up of 28 chromosomal contigs exhibiting an average length and coverage of 47,415 bp and 185×, respectively, arranged in a single scaffold, for a chromosome size of 1,327,625 bp (G+C content, 32.2%). We also detected a 19,874-bp plasmid (pRgr) with a G+C content of 31.8% and 91% sequence identity (36% coverage; *E* value, 0.0) to *Rickettsia monacensis* strain IrR/Munich plasmid pRM (accession no. EF564599). The chromosome contains 1,675 protein-encoding genes and, like other *Rickettsia* species, 3 noncontiguous rRNAs (5S, 16S, and 23S rRNA), 33 tRNAs, and 3 other RNAs. In addition, the pRgr plasmid contains 24 protein-encoding genes, including a split *sca12* gene and a proline-betaine transporter gene, but no RNAs.

Compared to that of *R. massiliae* MTU5, the *R. gravesii* chromosome exhibits a high level of synteny with the exception of four inversions of 37,978 bp, 4,791 bp, 2,782 bp, and 1,339 bp. Moreover, several genes are lacking in the *R. gravesii* genome, including *paaJ* (acetyl-coenzyme A [CoA] acetyltransferase), *def*3 (polypeptide deformylase), and genes for several transposases and inactivated derivatives.

Nucleotide sequence accession number. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. AWXL00000000.

ACKNOWLEDGMENT

This research was funded by the Mediterranee-Infection Foundation.

REFERENCES

1. **Owen H, Unsworth N, Stenos J, Robertson I, Clark P, Fenwick S.** 2006. Detection and identification of a novel spotted fever group rickettsia in Western Australia. Ann. N. Y. Acad. Sci. **1078:**197–199.
2. **Graves S, Stenos J.** 2009. Rickettsioses in Australia. Ann. N. Y. Acad. Sci. **1166:**151–155.
3. **Owen H, Clark P, Stenos J, Robertson I, Fenwick S.** 2006. Potentially pathogenic spotted fever group rickettsiae present in Western Australia. Aust. J. Rural Health **14:**284–285.
4. **Li AY, Adams PJ, Abdad MY, Fenwick SG.** 2010. High prevalence of *Rickettsia gravesii* sp. nov. in *Amblyomma triguttatum* collected from feral pigs. Vet. Microbiol. **146:**59–62.
5. **Brouqui P, Parola P, Fournier PE, Raoult D.** 2007. Spotted fever rickettsioses in southern and eastern Europe. FEMS Immunol. Med. Microbiol. **49:**2–12.
6. **Oteo JA, Portillo A.** 2012. Tick-borne rickettsioses in Europe. Ticks Tick Borne Dis. **3:**271–278.
7. **Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT.** 2009.

Reordering contigs of draft genomes using the Mauve aligner. Bioinformatics **25:**2071–2073.
8. **Blanc G, Ogata H, Robert C, Audic S, Claverie JM, Raoult D.** 2007. Lateral gene transfer between obligate intracellular bacteria: evidence from the *Rickettsia massiliae* genome. Genome Res. **17:**1657–1664.
9. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. doi:10.1186/1471-2164-9-75.
10. **Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
11. **Lowe TM, Eddy SR.** 1997. t-RNAscan-SE: a program for improved detection of transfer RNA gene in genomic sequence. Nucleic Acids Res. **25:**955–964.
12. **Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35:**3100–3108.
13. **Li L, Stoeckert CJ, Jr, Roos DS.** 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. **13:**2178–2189.

CONCLUSIONS AND PERSPECTIVES

This thesis is organized into two major parts. The first part consists of a review on the usage of genomics in prokaryotic taxonomy, while the second part deals with genomic studies of five subspecies and a new species of *Rickettsia*.

From the first part, we can conclude that there are several advantages as well as limitations of using genomics in prokaryotic taxonomy. The advantages include (i) the keep-increasing accessibility of prokaryotic genomes and (ii) the availability of genomic comparison markers such as average nucleotide identity (ANI) value that can replace the cumbersome DNA–DNA hybridization as a standard to define prokaryotic species. Nevertheless, there is still a need of genome sequences from more prokaryotic taxonomic representatives or under-represented taxa and a genomic-based method that is agreed upon by microbiologists, while consensus procedures to describe new prokaryotic taxa do not, as yet, accommodate genomic data.

In the second part, by using next-generation sequencing technologies, we managed to sequence the genomes of two *R. sibirica* subspecies, three *R. conorii* subspecies, and a new *Rickettsia* species of *R. gravesii*. We also carried out comparative genomics analyses among the *R. sibirica* and *R. conorii* subspecies.

The draft genome of *R. sibirica sibirica* BJ-90 consists of eight contigs, containing 1,254,013 base pairs with a G+C content of 32.5%. It has 1,539 protein-coding sequences, 33 tRNAs, three rRNAs, and three other RNAs. On the other hand, the draft genome of *R. sibirica mongolitimonae* HA-91 consists of 21 contigs, containing 1,252,337 base pairs (G+C content of 32.4%), 1,538 protein-coding sequences, 33 tRNAs, three rRNAs, and three other RNAs. Further analyses showed that superfamily I DNA and RNA helicase is absent from *R. sibirica sibirica* genome, while cell surface antigen Sca9 is absent from that of *R. sibirica mongolitimonae*. However, it remains to be determined whether this difference explains the differences in clinical expression observed between the subspecies.

The draft genome of *R. conorii indica* ITTR consists of 10 contigs, containing 1,249,482 base pairs (G+C content of 32.5%), 1,527 protein-coding sequences, 33 tRNAs, three rRNAs, and three other RNAs. Similarly, the draft genome of *R. conorii caspia* A-167 consists of 25 contigs, containing 1,260,331 base pairs (G+C content of 33%), 1,636 protein-coding sequences, and 39 RNAs. Furthermore, the draft genome of *R. conorii israelensis* ISTT CDC1 consists of 33 contigs, containing 1,252,815 base pairs (G+C content of 32%), 1,806 protein-coding sequences, and 39 RNAs. Differentially present genes among the subspecies should be scrutinized further to understand the differences in clinical expression observed among them.

Finally, the draft genome of *R. gravesii* BWI-1 consists of 28 chromosomal contigs, containing 1,347,499 base pairs (G+C content of 32.2%), and a 19,874 base-pair plasmid pRgr (G+C content of 31.8%). The chromosome has 1,675 protein-coding sequences, 33 tRNAs, three rRNAs, and three other RNAs, while the plasmid contains 24 protein-coding sequences.

Our work has added to the knowledge of *Rickettsia* genomes. We expect that our *in-silico* data be utilized as a basis of further studies to increase the understanding on the pathogenesis, evolutionary relationships, and taxonomy of rickettsiae. Moreover, we anticipate that genome sequence data like our own be used as one of the standards to describe new prokaryotic species.

REFERENCES

1. Parola P, Paddock CD, Socolovschi C, Labruna MB, Mediannikov O, Kernif T, Abdad MY, Stenos J, Bitam I, Fournier P-E, Raoult D (2013) Update on Tick-Borne Rickettsioses around the World: a Geographic Approach. Clin Microbiol Rev 26: 657-702.
2. Sahni SK, Narra HP, Sahni A, Walker DH (2013) Recent molecular insights into rickettsial pathogenesis and immunity. Future Microbiol 8: 1265-1288.
3. Yu X-J, Walker DH (2005) Family I. Rickettsiaceae. In: Brenner DJ, Krieg NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. New York, NY: Springer. pp. 96-116.
4. Fournier P-E, Dumler JS, Greub G, Zhang J, Wu Y, Raoult D (2003) Gene sequence-based criteria for identification of new *Rickettsia* isolates and description of *Rickettsia heilongjiangensis* sp. nov. J Clin Microbiol 41: 5456-5465.
5. Jiang J, Blair PJ, Felices V, Moron C, Cespedes M, Anaya E, Schoeller GB, Sumner JW, Olson JG, Richards AL (2005) Phylogenetic analysis of

a novel molecular isolate of spotted fever group Rickettsiae from northern Peru: *Candidatus* Rickettsia andeanae. Ann N Y Acad Sci 1063: 337-342.

6. Owen H, Unsworth N, Stenos J, Robertson I, Clark P, Fenwick S (2006) Detection and identification of a novel spotted fever group rickettsia in Western Australia. Ann N Y Acad Sci 1078: 197-199.

7. Rolain J-M, Mathai E, Lepidi H, Somashekar HR, Mathew LG, Prakash JAJ, Raoult D (2006) "*Candidatus Rickettsia kellyi*," India. Emerg Infect Dis 12: 483-485.

8. Gottlieb Y, Ghanim M, Chiel E, Gerling D, Portnoy V, Steinberg S, Tzuri G, Horowitz AR, Belausov E, Mozes-Daube N, Kontsedalov S, Gershon M, Gal S, Katzir N, Zchori-Fein E (2006) Identification and localization of a *Rickettsia* sp. in *Bemisia tabaci* (Homoptera: Aleyrodidae). Appl Environ Microbiol 72: 3646-3652.

9. Fournier P-E, Takada N, Fujita H, Raoult D (2006) *Rickettsia tamurae* sp. nov., isolated from *Amblyomma testudinarium* ticks. Int J Syst Evol Microbiol 56: 1673-1675.

10. Fujita H, Fournier P-E, Takada N, Saito T, Raoult D (2006) *Rickettsia asiatica* sp. nov., isolated in Japan. Int J Syst Evol Microbiol 56: 2365-2368.

11. Matsumoto K, Parola P, Rolain J-M, Jeffery K, Raoult D (2007) Detection of "*Rickettsia* sp. strain Uilenbergi" and "*Rickettsia* sp. strain Davousti" in *Amblyomma tholloni* ticks from elephants in Africa. BMC Microbiol 7: 74.

12. Pacheco RC, Moraes-Filho J, Nava S, Brandão PE, Richtzenhain L, Labruna M (2007) Detection of a novel spotted fever group rickettsia in *Amblyomma parvum* ticks (Acari: Ixodidae) from Argentina. Exp Appl Acarol 43: 63-71.

13. Mura A, Masala G, Tola S, Satta G, Fois F, Piras P, Rolain J-M, Raoult D, Parola P (2008) First direct detection of rickettsial pathogens and a new rickettsia, '*Candidatus* Rickettsia barbariae', in ticks from Sardinia, Italy. Clin Microbiol Infect 14: 1028-1033.

14. Mediannikov O, Matsumoto K, Samoylenko I, Drancourt M, Roux V, Rydkina E, Davoust B, Tarasevich I, Brouqui P, Fournier P-E (2008)

Rickettsia raoultii sp. nov., a spotted fever group rickettsia associated with *Dermacentor* ticks in Europe and Russia. Int J Syst Evol Microbiol 58: 1635-1639.

15. Izzard L, Graves S, Cox E, Fenwick S, Unsworth N, Stenos J (2009) Novel rickettsia in ticks, Tasmania, Australia. Emerg Infect Dis 15: 1654-1656.

16. Duh D, Punda-Polic V, Avsic-Zupanc T, Bouyer D, Walker DH, Popov VL, Jelovsek M, Gracner M, Trilar T, Bradaric N, Kurtti TJ, Strus J (2010) *Rickettsia hoogstraalii* sp. nov., isolated from hard- and soft-bodied ticks. Int J Syst Evol Microbiol 60: 977-984.

17. Palomar AM, Portillo A, Santibáñez P, Santibáñez S, García-Álvarez L, Oteo JA (2012) Genetic characterization of *Candidatus* Rickettsia vini, a new rickettsia amplified in ticks from La Rioja, Spain. Ticks Tick Borne Dis 3: 319-321.

18. Socolovschi C, Pages F, Ndiath MO, Ratmanov P, Raoult D (2012) *Rickettsia* species in African *Anopheles* mosquitoes. PLoS ONE 7: e48254.

19. Sandalakis V, Chochlakis D, Ioannou I, Psaroulaki A (2013) Identification of a novel uncultured *Rickettsia* species strain (*Rickettsia* species strain Tselenti) in Cyprus. Am J Trop Med Hyg 88: 698-700.

20. Jiang J, Maina AN, Knobel DL, Cleaveland S, Laudisoit A, Wamburu K, Ogola E, Parola P, Breiman RF, Njenga MK, Richards AL (2013) Molecular detection of *Rickettsia felis* and *Candidatus* Rickettsia asemboensis in fleas from human habitats, Asembo, Kenya. Vector Borne Zoonotic Dis 13: 550-558.

21. Zhu Y, Fournier P-E, Eremeeva M, Raoult D (2005) Proposal to create subspecies of *Rickettsia conorii* based on multi-locus sequence typing and an emended description of *Rickettsia conorii*. BMC Microbiol 5: 11.

22. Fournier P-E, Zhu Y, Yu X, Raoult D (2006) Proposal to create subspecies of *Rickettsia sibirica* and an emended description of *Rickettsia sibirica*. Ann N Y Acad Sci 1078: 597-606.

23. Merhej V, Raoult D (2011) Rickettsial evolution in the light of comparative genomics. Biol Rev Camb Philos Soc 86: 379-405.

24. Jia N, Jiang J-F, Huo Q-B, Jiang B-G, Cao W-C (2013) *Rickettsia sibirica* subspecies *sibirica* BJ-90 as a cause of human disease. N Engl J Med 369: 1176-1178.

25. Yu X, Jin Y, Fan M, Xu G, Liu Q, Raoult D (1993) Genotypic and antigenic identification of two new strains of spotted fever group rickettsiae isolated from China. J Clin Microbiol 31: 83-88.

26. Fournier P-E, Roux V, Raoult D (1998) Phylogenetic analysis of spotted fever group rickettsiae by study of the outer surface protein rOmpA. Int J Syst Bacteriol 48: 839-849.

27. Foissac M, Socolovschi C, Raoult D (2013) « Lymphangitis-associated rickettsiosis » due à *Rickettsia sibirica mongolitimonae*. Ann Dermatol Venereol 140: 521-527.

28. Parola P, Paddock CD, Raoult D (2005) Tick-borne rickettsioses around the world: emerging diseases challenging old concepts. Clin Microbiol Rev 18: 719-756.

29. Torina A, Fernández de Mera IG, Alongi A, Mangold AJ, Blanda V, Scarlata F, Di Marco V, de la Fuente J (2012) *Rickettsia conorii* Indian tick typhus strain and *R. slovaca* in humans, Sicily. Emerg Infect Dis 18: 1008-1010.

30. Joshi HS, Thomas M, Warrier A, Kumar S (2012) Gangrene in cases of spotted fever: a report of three cases. BMJ Case Rep 2012. DOI: 10.1136/bcr-2012-007295.

31. Tarasevich IV, Makarova VA, Fetisova NF, Stepanov AV, Miskarova ED, Raoult D (1991) Studies of a "new" rickettsiosis "Astrakhan" spotted fever. Eur J Epidemiol 7: 294-298.

32. Tarasevich IV, Makarova VA, Fetisova NF, Stepanov AV, Miskarova ED, Balayeva N, Raoult D (1991) Astrakhan fever, a spotted-fever rickettsiosis. Lancet 337: 172-173.

33. Fournier P-E, Durand J-P, Rolain J-M, Camicas J-L, Tolou H, Raoult D (2003) Detection of Astrakhan fever rickettsia from ticks in Kosovo. Ann N Y Acad Sci 990: 158-161.

34. Fournier P-E, Xeridat B, Raoult D (2003) Isolation of a rickettsia related to Astrakhan fever rickettsia from a patient in Chad. Ann N Y Acad Sci 990: 152-157.

35. Renvoisé A, Delaunay P, Blanchouin E, Cannavo I, Cua E, Socolovschi C, Parola P, Raoult D (2012) Urban family cluster of spotted fever rickettsiosis linked to *Rhipicephalus sanguineus* infected with *Rickettsia conorii* subsp. *caspia* and *Rickettsia massiliae*. Ticks Tick Borne Dis 3: 389-392.

36. Drancourt M, Beati L, Tarasevich I, Raoult D (1992) Astrakhan fever rickettsia is identical to Israel tick typhus rickettsia, a genotype of the *Rickettsia conorii* complex. J Infect Dis 165: 1167-1168.

37. Giammanco GM, Vitale G, Mansueto S, Capra G, Caleca MP, Ammatuna P (2005) Presence of *Rickettsia conorii* subsp. *israelensis*, the causative agent of Israeli spotted fever, in Sicily, Italy, ascertained in a retrospective study. J Clin Microbiol 43: 6027-6031.

38. de Sousa R, Ismail N, Dória-Nóbrega S, Costa P, Abreu T, França A, Amaro M, Proença P, Brito P, Poças J, Ramos T, Cristina G, Pombo G,

Vitorino L, Torgal J, Bacellar F, Walker D (2005) The presence of eschars, but not greater severity, in Portuguese patients infected with Israeli spotted fever. Ann N Y Acad Sci 1063: 197-202.

39. Bacellar F, Beati L, França A, Poças J, Regnery R, Filipe A (1999) Israeli spotted fever rickettsia (*Rickettsia conorii* complex) associated with human disease in Portugal. Emerg Infect Dis 5: 835-836.

40. Znazen A, Hammami B, Lahiani D, Ben JM, Hammami A (2011) Israeli spotted fever, Tunisia. Emerg Infect Dis 17: 1328-1330.

41. Boillat N, Genton B, D'Acremont V, Raoult D, Greub G (2008) Fatal case of Israeli spotted fever after Mediterranean cruise. Emerg Infect Dis 14: 1944-1946.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Didier Raoult for accepting me in his unit.

My sincerest gratefulness goes to Professor Pierre-Edouard Fournier for his patient guidance and advice throughout my time as his student. I have been extremely lucky to have a supervisor who is so kind, caring, and understanding.

I would like to show my distinct appreciation to Doctor Khalid El Karkouri for his help and direction for my work. I learned so much from him, without whom this thesis might not be here.

I gratefully acknowledge the members of my thesis committee; Doctor Patricia Renesto and Professor Max Maurin for their acceptance to be my thesis reviewers, and Doctor Florence Fenollar to be a member of the committee.

For their support during my stay in the unit, I am very grateful to all personnel of URMITE, including the technicians, the engineers, and the administrative staffs. Special thanks go to Francine Simula, Francine Verin, and Valérie Filosa for their never ending assistance.

I thank fellow students (PhD and master) and postdocs for their support and friendship. I apologize that I cannot mention them one by one here.

Finally, I would like to thank my family for being there for me; my parents in Indonesia, my brother, my wife Inge, and our future baby (who is surprisingly coming at the end of my study).