New approaches in the systematics of rickettsiae

S. N. Shpynov¹, P.-E. Fournier², N. N. Pozdnichenko³, A. S. Gumenuk¹ and A. A. Skiba³

¹) N.F. Gamaleya NRCEM, Moscow, Russia, ²) UMR VITROME, Aix-Marseille Université,IRD, Service de Santé des Armées, Assistance Publique-Hôpitaux de Marseille, Institut Hospitalo-Universitaire Méditerranée-Infection, Marseille, France and ³) OmSTU, Omsk, Russia

Abstract

The development of a formal order analysis (FOA) allowed constructing a classification of 49 genomes of Rickettsiaceae family representatives. Recently FOA has been extended with new tools—’Map of genes,’ ‘Matrix of similarity’ and ‘Locality-sensitive hashing’—for a more in-depth study of the structure of rickettsial genomes. The new classification confirmed and supplemented the previously constructed one by determining the position of Rickettsia africae str. ESF-5, R. heilongjiangensis 054, R. monacensis str. IrR/Munich, R. montanensis str. OSU 85-930, R. raoultii str. Khabarovsk, R. rhipicephali str. 3-7-femal6-CWPP and Rickettsiales bacterium str. Ac37b. The ‘Map of genes’ demonstrated the complete genomes and their components in a graphical form. The ‘Matrix of similarity’ was applied for an in-depth classification to a subtaxonomic category of the strain within the species R. rickettsii (11 strains) and R. prowazekii (ten strains). The ‘Matrix of similarity’ determines the degree of homology of complete genomes by pairwise comparison of their components and identification of those being identical and similar in the arrangement of nucleotides. A new genomosystematics approach is proposed for the study of complete genomes and their components through the development and application of FOA tools. Its applications include the development of principles for the classification of microorganisms, based on the analysis of complete genomes and their annotations. This approach may help in the taxonomic classification and characterization of some Candidatus Rickettsia spp. that are found in large numbers in arthropods worldwide.

© 2018 The Author(s). Published by Elsevier Ltd.

Keywords: Arthropods, ecology, epidemiology, formal order analysis, genome, genomosystematics, rickettsiae, rickettsioses, systematics, virulence

Original Submission: 21 November 2017; Revised Submission: 27 January 2018; Accepted: 7 February 2018

Article published online: 30 March 2018

Introduction

Rickettsia species are strictly intracellular vector-borne bacteria from the order Rickettsiales that cause mild to severe diseases in humans and other animals [1]. Currently 30 Rickettsia species have standing in nomenclature (http://www.bacterio.net/-allnamesmr.html). About 20 of them are pathogenic, including the causative agents of the deadly diseases epidemic typhus and Rocky Mountain spotted fever (RMSF) [2]. About 60 rickettsiae that were isolated or detected in ticks are currently considered as nonpathogenic, not validated, incompletely described and/or uncultivated species. Among uncultivated rickettsiae, 15 are potential new species and may be classified as Candidatus spp. (http://www.bacterio.net/-candidatus.html).

However, the systematics and nomenclature of Rickettsia species are based on a limited number of available phenotypic characteristics as a result of their obligate intracellular location. Members of the genus Rickettsia were initially classified on the basis of their morphologic, antigenic and metabolic characteristics into the following groups: (a) the spotted fever group (SFG), which includes species transmitted by hard ticks such as Rickettsia conorii, the causative agent of Mediterranean spotted fever [3], and R. rickettsii, the agent of RMSF [4]; (b) the typhus group (TG), which includes R. typhi, the flea-transmitted causative agent of murine typhus and R. prowazekii, the louse-transmitted agent of epidemic
typhus [5,6]; and (e) the group containing R. tsutsugamushi, the aetiologic agent of scrub typhus [7]. Then the application of molecular and phylogenetic methods enabled defining three groups within the genus Rickettsia: the TG, the SFG that includes a large collection of mostly tick-borne rickettsiae and an ancestral group (AG), which includes R. bellii and R. canadensis [8]. Rickettsia tsutsugamushi was found to exhibit unique phenotypic and molecular characteristics and was reclassified as Orientia tsutsugamushi [9]. Most recently, whole-genome sequence analysis suggested the existence of another group within the Rickettsia genus, termed the transitional group, consisting of R. felis and R. akari [10], but this group is not widely accepted [1].

Most bioinformatic genome analysis programs such as BLAST [11], MEGA [12] and many others, are all based on mathematical and statistical methods that compare only homologous sequence fragments of genomes or their concatenation. In contrast, the formal order analysis (FOA) method transforms the nucleotide order in the sequence into a numerical sequence (value) of fixed length [13]. FOA takes into account the original arrangement of nucleotides in each genome. Previous FOA results have corroborated the separation of species in three groups within the genus Rickettsia, including TG, SFG and AG, and Orientia as a separate genus. Rickettsia felis and R. akari were not in the same group according to FOA. Therefore, this approach did not confirm the existence of a so-called transitional group [14].

In recent years, the application of new strategies of culturotics and taxonogenomics has made it possible to isolate and identify difficult-to-cultivate bacterial species, including rickettsiae, and to optimize the amount of information for their classification [15,16]. The complete genome of Candidatus Midichloria mitochondrii was sequenced in the female Ixodes ricinus tick [17]. The analysis of intergenic regions encoding noncoding small RNAs has demonstrated their role in the virulence of R. prowazekii [18]. Thus, in the era of high-quality genomes sequences, it is necessary to develop new approaches for obtaining the maximum amount of information from genome structure and its characteristics.

We sought to develop an approach based of FOA for improving the classification of rickettsiae, with an in-depth study of genomic characteristics for differentiating rickettsial strains and an estimation of their phenotypic characteristics, such as virulence.

Materials and methods

Genome sequences of Rickettsia and Orientia species
Forty-nine complete genome sequences of Rickettsia (n = 47) and Orientia (n = 2) spp. (Table 1) were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genome).

Formal order analysis (FOA) tools
FOA was used to analyse rickettsial genomes as previously described [13,14,19]. FOA uses a high-precision and unambiguous numerical representation of the original arrangement of nucleotides in the sequence. In order to do this, numerical characteristics (average remoteness, depth, etc.) based on intersymbol intervals (internucleotide distance [20]) were developed.

Recently FOA has been improved with the new tools ‘Map of genes’ (MG) [21], ‘Matrix of similarity’ (MS) [22] and ‘Locality-sensitive hashing’ (LSH) (Pozdnichenko NN et al., paper presented at 11th International Conference of Computer-Aided Technologies in Applied Mathematics) for a more in-depth study of rickettsial genome structure.

Pairs of numeric values of order characteristics from studied genomes and their components (<g, G>) are mapped into pillars of dots on the MG. Components representing individual genomes are placed vertically, and some horizontal lines are formed with similar components in different genomes. The MG tool kit enables interactive obtaining of a detailed description of any component of the genome. Automated identification of similar components is also possible.

The MS represents the similarity values for each pair of analysed genomes. Genome similarity is determined by comparing the order characteristic values from their components. The MS tool kit enables interactive obtaining of a list of only similar components of any pair of genomes, and when necessary, a sliding window characteristics for those components can be obtained. The latter enables visualization of element-by-element similarity of genomic components.

The comparison of coding and noncoding sequences in genomes from different strains of Rickettsia spp. is possible using the MG tool, which can also be useful in the analysis of genomes from different microorganisms in order to find homologous genes and orthologs. The MG tool identifies interstrain genomic differences within Rickettsia species. The MG tool also provides a complete representation of genomes and their components in a graphical form. Genomes are placed (classified) according to the index g on the x-axis, and their components (coding and noncoding sequences) are sited according to the depth index (G) on the y-axis. Sequences that are completely similar by a characteristic are 100% homologous. Strains of the same species show a high degree of homology of their components. It is necessary to introduce a criterion (characteristic) for identification of components with a degree of homology <100% in order to compare genomes of different species. Using this approach, it is possible to conduct a selective analysis of all genomic components, individually or by grouping them by feature (coding DNA sequence, rRNA, tRNA, noncoding RNA, pseudogenes, repeat regions, etc.).
The UPGMA algorithm (unweighted pair-family and criteria for the formation of taxa within the genus group average) was used for the analysis of average distances.

All genomes were imported from GenBank NCBI (USA): http://www.ncbi.nlm.nih.gov/genome/.

Each component can be identified among all compared organisms by its name in the annotation, which allows checking its presence in each genome. The LSH of nucleotide sequences is provided by differing values of numerical characteristics. All genomes were analysed using FOA [13] software, which is available online (http://foarlab.org).

Cluster analysis

Cluster analysis was carried out using PAST software (http://folk.uio.no/ohammer/past/) for verification of the obtained classification scheme for representatives of the Rickettsiaceae family and criteria for the formation of taxa within the genus Rickettsia (Fig. 1). The UPGMA algorithm (unweighted pair-group average) was used for the analysis of average distances. Clusters were formed on the basis of the average distance between members of all groups.

| No. | Access Number in GeneBank | Genotype size (bp) | G+C% | g |
|-----|---------------------------|--------------------|------|---|
| 1   | NC_017050                 | 111 454            | 29   | 4.1823242027653 |
| 2   | NC_017056                 | 111 445            | 29   | 4.182413821358 |
| 3   | NC_000963                 | 111 523            | 29   | 4.1824702490549 |
| 4   | NC_017560                 | 111 612            | 29   | 4.18264321159974 |
| 5   | NC_017048                 | 111 969            | 29   | 4.18314008733500 |
| 6   | NC_017057                 | 112 101            | 29   | 4.18325015358528 |
| 7   | NC_017047                 | 109 804            | 29   | 4.18363689393900 |
| 8   | NC_020993                 | 109 301            | 29   | 4.18494808769293 |
| 9   | NC_014865                 | 111 769            | 29   | 4.187674360422 |
| 10  | NC_006162                 | 111 520            | 29   | 4.184807421873 |
| 11  | NC_017062                 | 112 957            | 28.9 | 4.189792583639  |
| 12  | NC_017066                 | 112 372            | 28.9 | 4.189961796300  |
| 13  | NC_006164                 | 111 496            | 28.9 | 4.1990993275420 |
| 14  | NC_009217                 | 851 238            | 30.8 | 4.236960175481  |
| 15  | NC_009883                 | 528 980            | 31.6 | 4.246105909798  |
| 16  | NC_007940                 | 522 076            | 31.6 | 4.247946077370   |
| 17  | NC_016929                 | 150 228            | 31.0 | 4.250585941161  |
| 18  | NC_009879                 | 159 772            | 31.0 | 4.2576175261109  |
| 19  | NC_0179417                | 353 450            | 32.4 | 4.26391576761608 |
| 20  | NC_007109                 | 485 148            | 32.6 | 4.291847695814   |
| 21  | NC_015866                 | 278 468            | 32.3 | 4.307561843097   |
| 22  | NC_017042                 | 290 368            | 32.4 | 4.308807051608   |
| 23  | NC_016050                 | 283 087            | 32.4 | 4.3117915307534 |
| 24  | NC_017058                 | 296 670            | 32.3 | 4.31236851109199 |
| 25  | NC_017054                 | 279 796            | 32.6 | 4.3172615975566 |
| 26  | NC_012633                 | 278 530            | 32.4 | 4.3180838121582 |
| 27  | NC_017065                 | 275 720            | 32.5 | 4.3190033264176 |
| 28  | NC_016539                 | 275 089            | 32.5 | 4.3200765486268 |
| 29  | NC_017044                 | 300 386            | 32.4 | 4.3209421751330 |
| 30  | NC_010263                 | 268 755            | 32.4 | 4.3213975511604 |
| 31  | NC_016909                 | 268 188            | 32.4 | 4.3246659062342 |
| 32  | NC_016913                 | 265 681            | 32.5 | 4.3272794620026 |
| 33  | NC_016914                 | 269 837            | 32.5 | 4.327946822827 |
| 34  | NC_016908                 | 270 083            | 32.5 | 4.3286649454322 |
| 35  | NC_016915                 | 270 751            | 32.5 | 4.3286698637378 |
| 36  | NC_009882                 | 257 710            | 32.5 | 4.3291630859912 |
| 37  | NC_008000                 | 257 005            | 32.5 | 4.328800163848 |
| 38  | NC_008000                 | 268 201            | 32.4 | 4.328608339081 |
| 39  | NC_008000                 | 268 220            | 32.4 | 4.328637678714 |
| 40  | NC_008000                 | 268 242            | 32.4 | 4.3286429678664 |
| 41  | NC_009900                 | 360 898            | 32.5 | 4.3314587356848 |
| 42  | NC_016930                 | 287 740            | 32.5 | 4.3325054686125 |
| 43  | NC_017020                 | 407 796            | 32.45 | 4.3459070799744 |
| 44  | NC_009127                 | 288 492            | 32.6 | 4.3514665884247 |
| 45  | NC_009881                 | 281 061            | 32.3 | 4.3517339590928 |
| 46  | NC_009880                 | 212 051            | 30.5 | 4.4594960730303 |
| 47  | NC_009889                 | 200 987            | 30.5 | 4.4642319139296 |

**TABLE 1.** Genome features of sequenced and average remoteness (g) of Rickettsia spp. and Orientia tsutsugamushi.

Results

The classification obtained in this work confirmed and improved the previous one, using only one single characteristic of the order that is the average remoteness (g) [14]. The Rickettsiales bacterium str. Ac37b and _R. monocensis_ were localized on the borders of the AG, beside _R. bellii_ str. IrR/München OSU 85-389 and _R. canadensis_ str. McKiel. _Rickettsia helongiangiensis_ str. 054, _R. hipchephalii_ str. 3-7-female6-CWP, _R. montanensis_ str. OSU 85-930, _R. africae_ str. ESF-5 and _R. raoultii_ str. Kharakov were placed in the SFG (Fig. 1).

By using cluster analysis, sets of genomes from members of the family Rickettsiaceae were grouped into disjointed subset clusters consisting of genomes that are close by index g and represented as a dendrogram (Fig. 1). The family Rickettsiaceae.
FIG. 1. Systematics of *Rickettsia* spp. and *Orientia tsutsugamushi* using characteristics of average remoteness (g) of their genomes, as well as ecological, epidemiologic and nosologic (etiologic) features (genomosystematics of rickettsiae).
was divided into the genera Rickettsia and Orientia, with a distance index ranging from 0.016 to 0.018. Two major groups are formed inside the genus Rickettsia in the range of 0.008 to 0.01. The major TG (MTG) includes the previous TG and AG. The TG and AG are formed by a value index of 0.006. The AG comprised two subgroups: R. bellii (R. bacteriophila. Ac37b and R. bellii str. OSU 85-389) and R. canadensis (R. canadensis str. McKiel and R. monacensis str. IrR/Munich). The major SFG (MSFG) group included the R. akari group, the R. felis group and the classical SFG. The R. akari group (mite-borne, rickettsialpox) exhibits a distance value index of 0.004. The R. felis and classical SFG were discriminated by a value index of 0.003. The classical SFG group is also subdivided into three subgroups: R. rickettsii, R. conori and R. australis. The R. rickettsii subgroup (R. massiliae str. MTUS, R. philipii str. 364D and R. amblyommatis str. GAT-30V) exhibits a distance value index of 0.0018, followed by the subgroups R. conori (R. raoauli str. Khabarovsk, R. montanensis str. OSU 85-930, R. africase str. ESF-5, R. slovaca str. 13-B and R. parkeri str. Portsmouth) and R. australis (R. helioniangensis str. 054, R. rhipicephali str. 7-3-female6-CWPP and R. japonica str. YH), with distance value indexes of 0.0015.

The MG tool detected the homology of rRNA genes (SS rRNA, 16S rRNA and 23S rRNA) of rickettsiae. The analysis of rRNA demonstrated that all three R. typhi strains were 100% homologous for these genes, while nine of ten R. prowazekii strains had complete SS rRNA and 16S rRNA homology, with the exception of R. prowazekii str. Madrid E. This strain differed in 23S rRNA from all other strains, whereas the remaining nine strains were divided into two groups: Breinl, Chernikova and BuV67-CWPP formed one group, and all remaining strains formed another group. All R. rickettsii strains were 100% homologous with R. philipii, R. conori and R. parkeri for the SS rRNA gene. All R. rickettsii strains were also 100% homologous for 16S rRNA, except R. rickettsii str. Hlpt2, which was identical to R. philipii.

The analysis of the 23S rRNA gene enabled discriminating R. rickettsii strains into three groups: (a) Sheila Smith, R. Brazil and Colombia, (b) Hlpt2 and (c) all remaining strains. Two R. canadensis strains exhibited complete homology with each other only for the SS rRNA gene and were also 100% homologous with R. bellii, R. rhipicephali, R. montanensis, R. monacensis, R. felis and R. japonica. Two R. amblyommatis strains exhibited 100% absolute homology for SS rRNA.

The MS tool was used for in-depth intraspecific analysis of the most pathogenic Rickettsia species, R. prowazekii (ten strains) and R. rickettsii (11 strains). We studied the correlation between the MS score (genotypic characteristic) obtained by analysing their genomes with FOA and virulence. The MS tool determines the degree of homology of complete genomes by a pairwise comparison of their components and identification of identical and similar components in the arrangement of nucleotides. The results are presented in Tables 2 and 3, respectively. Strains were ranked within these species according to the decreasing MS score with R. prowazekii str. Breinl and R. rickettsii str. Sheila Smith, which are type strains that have a well-characterized high degree of virulence.

According to their decreasing MS score with str. Breinl, R. prowazekii strains were classified as follows: Chernikova (85.7%), Naples (80.37%), Rp22 (80.11%), BuV67-CWPP (74.42%), Katsinyian (73.9%), NMRC Madrid E (64.55%), Madrid E (60.07%), RpGvF24 (52.31%) and GvV257 (50.88%; Table 2).

Similarly, according to their decreasing MS score with str. Sheila Smith, R. rickettsii strains were classified as follows: R (98.75%), Brazil (80.79%), Colombia (79.23%), Arizona (75.68%), Hino (75.05%), Morgan (75.01%), Iowa (74.31%), Iowa isolate Large Clone (63.32%), Iowa isolate Small Clone (63.32%) and Hlpt2 (26.22%; Table 3).

Numerical characteristics of order can be used for compact representation and LSH of complete nucleotidic genome sequences. Thus, the genome from R. prowazekii str. Madrid E is 1 111 520 bp long and requires 300 pages in FASTA format, but it can be represented in FOA characteristic by a 14-decimal number (e.g., 1.41848807421873).

**TABLE 2. Study of homology degree of components for genomes of strains of Rickettsia prowazekii using ‘Matrix of similarity’**

| No. | Rickettsia prowazekii strain | Katsinyian | BuV67-CWPP | Madrid E | Rp22 | Naples | GvV257 | RpGvF24 | Chernikova | NMRC madrid E | Breinl |
|-----|----------------------------|------------|------------|----------|------|--------|--------|---------|------------|----------------|-------|
| 1   | Katsinyian                 | 100.00%    | 91.44%     | 79.40%   | 86.36% | 85.67% | 60.11% | 61.65%  | 84.71%     | 77.01%         | 73.90% |
| 2   | BuV67-CWPP                 | 91.44%     | 100.00%    | 74.90%   | 87.86% | 85.87% | 60.07% | 62.38%  | 85.68%     | 73.04%         | 74.42% |
| 3   | Madrid E                  | 79.40%     | 74.90%     | 100.00%  | 63.72% | 70.52% | 47.79% | 48.91%  | 68.41%     | 66.63%         | 60.08% |
| 4   | Rp22                      | 86.36%     | 86.78%     | 69.75%   | 100.00%| 96.09% | 58.65% | 60.19%  | 91.24%     | 69.21%         | 80.11% |
| 5   | Naples-1                  | 85.67%     | 85.87%     | 70.52%   | 96.09% | 100.00%| 58.12% | 59.67%  | 91.24%     | 69.30%         | 80.37% |
| 6   | GvV257                    | 60.11%     | 60.06%     | 47.79%   | 58.65% | 58.12% | 100.00%| 85.56%  | 58.32%     | 49.59%         | 50.88% |
| 7   | RpGvF24                   | 61.65%     | 62.38%     | 48.91%   | 60.19% | 59.67% | 85.56% | 100.00% | 60.30%     | 50.68%         | 52.31% |
| 8   | Chernikova               | 84.71%     | 85.68%     | 68.41%   | 92.44% | 91.24% | 58.32% | 60.30%  | 100.00%    | 67.58%         | 85.70% |
| 9   | NMRC Madrid E             | 77.00%     | 73.04%     | 66.63%   | 69.30% | 49.59% | 50.68% | 67.58%  | 76.58%     | 100.00%        | 64.55% |
| 10  | Breinl                    | 73.90%     | 74.42%     | 60.08%   | 80.11% | 80.37% | 50.88% | 52.30%  | 85.70%     | 64.55%         | 100.00% |
TABLE 3. Study of homology degree of components for genomes of strains of Rickettsia rickettsii using ‘Matrix of similarity’

| No. | Rickettsia rickettsii strain | Arizona | Iowa | Iowa isolate, large clone | Iowa isolate, small clone | Brazil | Morgan | Hino | Colombia | Hpl#2 | R | Sheilasmith |
|-----|-----------------------------|---------|------|--------------------------|--------------------------|--------|--------|------|----------|--------|---------|-------------|
| 1   | Arizona                     | 100.00% | 92.50% | 78.18%                   | 78.11%                   | 74.91% | 93.92% | 94.00% | 75.85%   | 26.43% | 75.81% | 75.68%     |
| 2   | Iowa                        | 92.49%  | 100.00% | 82.66%                   | 82.60%                   | 74.20% | 94.73% | 96.63% | 74.19%   | 25.80% | 74.44% | 74.31%     |
| 3   | Iowa isolate, large clone   | 78.16%  | 82.66% | 100.00%                  | 99.87%                   | 63.44% | 80.10% | 81.24% | 63.03%   | 21.52% | 63.64% | 63.32%     |
| 4   | Iowa isolate, small clone   | 78.11%  | 82.59% | 99.87%                   | 100.00%                  | 63.44% | 80.03% | 81.24% | 63.03%   | 21.52% | 63.64% | 63.32%     |
| 5   | Brazil                      | 74.91%  | 74.20% | 63.44%                   | 63.44%                   | 100.00% | 73.95% | 74.20% | 77.94%   | 26.26% | 80.93% | 80.80%     |
| 6   | Morgan                      | 93.92%  | 94.73% | 80.10%                   | 80.03%                   | 73.95% | 100.00% | 96.14% | 74.60%   | 26.26% | 75.28% | 75.01%     |
| 7   | Hino                        | 94.00%  | 96.63% | 81.24%                   | 81.24%                   | 74.20% | 96.14% | 100.00% | 75.07%   | 26.22% | 75.17% | 75.05%     |
| 8   | Colombia                    | 75.85%  | 74.19% | 63.03%                   | 63.03%                   | 77.94% | 74.60% | 75.07% | 100.00%  | 26.01% | 79.29% | 79.24%     |
| 9   | Hpl#2                       | 26.43%  | 23.80% | 21.52%                   | 21.52%                   | 26.26% | 26.26% | 26.22% | 26.01%   | 100.00% | 26.31% | 26.22%     |
| 10  | R                           | 75.81%  | 74.44% | 63.64%                   | 63.64%                   | 80.93% | 75.28% | 75.17% | 79.20%   | 26.31% | 100.00% | 98.75%     |
| 11  | Sheila Smith                | 75.68%  | 74.31% | 63.32%                   | 63.32%                   | 80.79% | 75.01% | 75.05% | 79.24%   | 26.22% | 98.75% | 100.00%    |

Discussion

The classification obtained using the FOA method showed a significant divergence of the genera Rickettsia and Orientia within the family Rickettsiaceae. Initially two groups were formed in the Rickettsia genus: MTG and MSFG. The MTG was divided into the TG and AG. This classification was supported by the detection of antigenic cross-reactions between R. canadensis and TG [23,24]. Furthermore, R. conorii was also suspected to be responsible for cases of acute cerebral vasculitis published by Zdrodovskii and Golinevich in 1960 [29], as follows.

I. Louse- or flea-borne typhus fever group (aka typhus fever group)

1. Epidemic, or louse-borne, European or historic typhus fever—agent: R. prowazekii (Rocha Lima, 1916) or R. prowazekii var. prowazekii (Pinkerton, 1936).
2. Endemic or murine typhus fever (Marcy, 1926)—agent: R. mooseri (Monteiro, 1931) or R. prowazekii var. mooseri (Pinkerton, 1936). Synonym: R. typhi (Wolbach and Todd, 1920).

II. Tick-borne SFG

1. New World subgroup—(1) Rocky Mountain spotted fever (Maxey, 1899); (2) Brazilian or Sao Paolo typhus fever (Monteiro, 1935), agent: D. rickettsii (R. rickettsii) (Wolbach, 1919).
2. Old World subgroup—(1) Marseilles or Mediterranean Fever (Fièvre boutonneuse, pimple fever) (Conor and Bruch, 1910), agent: D. conorii (R. conorii) (Brumpt, 1932); (2) South and East African tick typhus, agent: D. rickettsii var. p. p. (Alexander and Mason, 1939); (3) North Asian tick rickettsiosis or tick typhus (Velik, Savul’kin, Shmatikov, Krontovskaia et al., 1935–1938), agent: D. sibiricus (R. sibirica) (Zdrodovskii and Golinevich, 1949); (4) North Australian tick typhus (Andrew, Bonnin, and Williams, 1946), agent: D. nov. spec. (Plotz et al., 1946).
3. Subgroup of gamasid rickettsioses—Varioiiform or vesicular rickettsiosis (Huebner, Greenberg et al., 1946–1947; Drobinskii, Zhdanov, Kulagin, 1948–1950), agent: R. acari (Huebner et al., 1946).

III. Mite-borne fever group (aka, in Japanese terminology, tsutsugamushi group)

Tsutsugamushiki fever or Japanese river fever (Baelz and Kawakami 1879), agent: R. orientalis (Nagayo et al., 1930). Synonyms: R. tsutsugamushi (Ogata, 1931).

IV. Pneumotropic group of rickettsioses (Q fever group)
Rickettsiae that are associated with Arachnida order Mesostigmata are the hosts of AG and SFG rickettsiae, and gamasid mites from the Rickettsia akari phalus, Amblyomma, Haemaphysalis of the order Ixodida are vectors of scrub typhus exhibiting different ecoepidemiologic and clinical features (host, transmission, seasonal manifestation, etc.). In contrast with other rickettsioses and scrub typhus, which are zoonoses, epidemic typhus is an anthroposis. Rickettsioses are divided into two groups: insect-borne and acariborne rickettsioses (Fig. 1). Insect-borne rickettsioses include louse-borne rickettsiosis (epidemic typhus caused by R. prowazekii) and flea-borne rickettsioses (murine typhus and flea-borne spotted fever caused by R. typhi and R. felis, respectively). Acari-borne rickettsioses consist of tick-borne spotted fevers (among others, RMSF, Mediterranean spotted fever, Siberian tick typhus, Queensland tick typhus caused by R. rickettsii, R. conorii, R. sibirica and R. australis), mite-borne rickettsiosis (rickettsialpox caused by R. akari) and chigger-borne disease (scrub typhus caused by O. tsutsugamushi).

Rickettsiae are associated with arthropods, which can transmit them to vertebrates via saliva or faeces. Rickettsia prowazekii is transmitted by the human body louse (Pediculus hominis corporis), and its main reservoir is humans [31]. Transmission of this bacterium does not occur directly by a bite but by contamination of scratch sites with the faeces or the crushed bodies of infected lice [5,32]. A similar infection mode occurs for R. typhi and R. felis infections, transmitted by fleas [33,34]. Transmission of R. typhi to humans occurs by contamination of the skin or respiratory tract by aerosols of dust containing infective material or via contamination of the conjunctivae of the host with infected flea faeces [32]. Thus, the human infection caused by R. prowazekii, R. typhi and R. felis is carried out as transmission via contamination.

Ticks are the main vectors and reservoirs of SFG rickettsiae. Rickettsiae infecting the ticks’ salivary glands can be transmitted to vertebrate hosts during feeding [32]. R. akari is responsible for rickettsialpox, which is an urban disease involving mites of the genus Allogenedermatys (Liponyssoides), the house mouse Mus musculus and, accidentally, humans [32,35]. Humans get rickettsialpox after being bitten by an infected mite. Orientia tsutsugamushi is transmitted by bites of feeding larval trombiculid mites (chiggers), which are the reservoir of the agent and the only life stage that feeds on a vertebrate host [36–38]. Therefore, human infection by SFG rickettsiae R. akari and O. tsutsugamushi is carried out as transmission via inoculation. Rickettsia felis has been proposed to be one of the most ancient Rickettsia species [39]. It was identified worldwide in more than 20 different haematophagous species of fleas, mosquitoes, ticks and mites [40]. Although C. felis fleas were initially considered to be the only vector of R. felis, evidence supports the role of other vectors, notably Anopheles, in the transmission of the bacterium [41,42]. The acquisition and persistence of R. felis in Anopheles have been demonstrated, and live bacteria were detected in mosquito faeces and their salivary glands, gut and ovaries [41]. The transmission of R. felis to vertebrates by A. gambiae and Liposcelis bostrychophilus is experimentally proven, but transmission to humans is only hypothetical. Vertebrate infection by fleas may occur by blood feeding or contamination of excoriations by faeces [40]. Probably R. felis is transmitted to humans both through transmission via contamination and transmission via inoculation modes.

© 2018 The Author(s). Published by Elsevier Ltd. NMNI. 23, 93–102

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
Using the MG tool enabled building a rickettsial classification for each of the genes available online. The presence or absence of a gene is a classification feature for closely related species of rickettsia.

We attempted to rank R. rickettsii and R. prowazekii strains for virulence by analysing the number of components of the genome having complete (100%) MS homology (Tables 2 and 3). Strains from R. rickettsii differ significantly in virulence [43–46]. Virulence varies from the most virulent Sheila Smith strain to the avirulent Iowa. Strains Sheila Smith, Brazil and Morgan were defined as highly cytopathic isolates based on an in vitro model, whereas the Colombia and Hlp#2 strains caused lower reactions [43]. The phylogenetic tree constructed using the analysis of multilocus sequences showed that strains Sheila Smith and R are closely related and differ significantly from strains Iowa and Morgan, which are close to each other [46]. Strain Hlp#2 exhibited differences in the ompA gene compared to 12 other R. rickettsii strains isolated from ticks and patient's blood in the Americas [44]. The virulence loss of strain Iowa may be associated with the disruption of ompA and the defect in processing of ompB, which are demonstrated to cause protection of guinea pigs in subsequent infection by R. rickettsii str. Sheila Smith [47]. The Sheila Smith, R and Brazil strains exhibit a 10 kbp deletion, unlike the others. The significance of this deletion is unclear, since this region is present both in str. Iowa, Morgan, Hino, Hauke, Arizona, Colombia and Hlp#2 [46]. It is believed that the presence or absence of this region does not have a direct effect on virulence, since it is present in the virulent Morgan strain. The indel genotyping system enabled the identification of 25 genotypes of R. rickettsii in 4 groups [45]. Strain Hlp#2, which is often considered as nonpathogenic, showed the greatest diversity compared to other strains. Our results support the phylogenetic trees obtained by Clark and colleagues [46], and Genome Tree is publicly available online (https://www.ncbi.nlm.nih.gov/genome/tree/674/). Furthermore, the division of R. rickettsii strains into three groups confirms the MG analysis results of the 23S rRNA gene.

Our results support the current understanding of the virulence of rickettsial strains based on phenotypic and genotypic characteristics. Rickettsia prowazekii strains exhibit various degrees of virulence [48]. The Breinl strain and a more recent isolate, Rp22, are considered to be highly virulent. The Madrid I strain was isolated in 1941 during the Madrid outbreak of epidemic typhus. After passages in embryonated eggs, strain Madrid I has lost its virulence and has been used under the name of Madrid E as a vaccine in humans since 1944 [49]. A comparative genomic microarray study revealed a highly conserved genome content between str. Breinl and Madrid E (only 3% nucleotide variations) [50]. The draft genomes from the flying squirrel strain GvF12 was found to differ from those of str. GvVF27 and GvVF24 at 226 and 11 positions, respectively, whereas the GvF12 and Madrid E genomes were found to vary at 869 positions. In comparison, the Breinl and Madrid E genomes were found to differ at 292 positions. These preliminary data indicate that flying squirrel isolates may be more similar to each other than to human isolates [51]. Furthermore, rRNA screening with the MG tool demonstrated differences between R. prowazekii str. Madrid E and R. rickettsii str. Hlp#2 and other strains.

It has been shown that the ranking of the genomes from R. rickettsii and R. prowazekii strains is correlated to the Gapped Identity (%) on the National Center for Biotechnology Information (NCBI) website (https://www.ncbi.nlm.nih.gov/genome/neighbors/674?&genome_assembly_id=300283 and https://www.ncbi.nlm.nih.gov/genome/neighbors/737?genome_assembly_id=168378, respectively). This correlation requires additional research. A comparative analysis of the degree of virulence in the MG analysis showed that the percentage of coincident genome components correlated with an index of Gapped Identity (%) to a high degree. Thus, the elimination (decrease in the degree) of the virulence of strains is associated with the accumulation of the gapped effect from the highly virulent strains R. rickettsii str. Sheila Smith and R. prowazekii str. Breinl.

Use of LSH is important for reduction of the amount of information stored and is required for identification of genetic texts. The LSH is an algorithm that results in the nucleotide sequence being redecoded into a numerical sequence.

Bacterial taxonomy relies on a polyphasic approach based on the combination of phenotypic and genotypic characteristics (DNA-DNA hybridization, 16S rRNA gene sequence similarity and phylogeny, DNA G+C content). Recently, the polyphasic approach has been adapted to culturomics [52] through the development of taxonomics [16]. Taxonomics is a polyphasic strategy combining phenotypic characteristics obtainable and comparable by most laboratories with matrix-assisted desorption ionization–time of flight mass spectrometry analysis, genome sequence characteristics and comparison using average genomic identity of orthologous gene sequences (AGIOS), average nucleotide identity (ANI) and/or other genome comparison software for the taxonomic description of new bacterial taxa [15,16,53].

The complete genome of Candidatus Midichloria mitochondrii IrVCA was sequenced (GenBank accession no. NC_015722.1) from the ovarian tissue of a female Ixodes ricinus tick collected in nature (Varese, Italy) [17]. Bioinformatic analysis enabled studying the genotypic characteristics and modelling the phenotypic characteristics of this uncultivated microorganism from the order Rickettsiales. We believe that sequencing the genome of an uncultivated microorganism directly from the organs of a naturally infected female tick or using a experimentally infected tick model will develop. Then, after genome sequence annotation, the
genotypic characteristics may be studied and the phenotypic characteristics modelled using bioinformatics analysis. To classify these uncultivated bacteria taxonomically, taxonogenomics may be used. However, to determine the position at the family, genus and species ranks, FOA, which takes into account the arrangement of nucleotides in the genome, may be included in the taxonogenomic analysis. The application of this method may provide an occasion for the International Committee of Systematic Bacteriology to implement a new procedure for recognizing the status of new species among noncultivated bacteria [54].

For the first time, an attempt to compare the genomes from prokaryotes (bacteria and archaea) was adapted from the works of E. Kunin and coauthors in 1997 [55]. The term ‘phylogenomics,’ proposed by Eisen and Fraser in 2003 [56], covers systems-based study of genes and genomes, as well as analysis of the evolution of gene families within genomes. Genosystematics has been successfully applied to rickettsiae using 16S rRNA analysis and four protein-coding genes [57].

**Conclusion**

In this study, a new genosystematics approach is proposed for the study of complete genomes and their components through the development and application of FOA tools. Its applications include the development of principles for the classification of microorganisms based on the analysis of complete genomes and their annotations. The classification of rickettsial genomes obtained on the basis of FOA has a strong correlation with the taxonomy of arthropods, which are the hosts of rickettsia and which is confirmed by their ecological associations. The objectivity of the classification of rickettsial genomes is confirmed by the classification of rickettsioses, built on the basis of mechanisms of infection by various groups of rickettsia, with these mechanisms being of great importance in the epidemiology and aetiology of rickettsioses.

Thus, genosystematics (systematics of genomes) underlies the classification of rickettsiae and rickettsioses based on ecological, epidemiologic and aetiologic principles.

**Conflict of interest**

None declared.

**References**

[1] Merhej V, Raoult D. Rickettsial evolution in the light of comparative genomics. Biol Rev Camb Philos Soc 2011;86:379–405.

[2] Parola P, Paddock CD, Socolovschi C, Labruna MB, Mediannikov O, Kernif T, et al. Update on tick-borne rickettsioses around the world: a geographic approach. Clin Microbiol Rev 2011;24:657–75.

[3] Raoult D, Weiller PJ, Chagnon A, Chaudet H, Gallais H, Canasov P. Mediterranean spotted fever, clinical, laboratory and epidemiological features of 199 cases. Am J Trop Med Hyg 1986;35:845–50.

[4] Dumler JS, Walker DH. Rocky Mountain spotted fever—changing ecology and persisting virulence. N Engl J Med 2005;353:551–3.

[5] Bechay Y, Capo C, Mege JL, Raoult D. Epidemic typhus. Lancet Infect Dis 2008;8:417–26.

[6] Zinsser H. Rats, lice, and history. London: Broadway House; 1935.

[7] Merhej V, Angelakis E, Socolovschi C, Raoult D. Genotyping, evolution and epidemiological findings of Rickettsia species. Infect Genet Evol 2014;25:122–37.

[8] Stothard DR, Clark JB, Fuerst PA. Ancestral divergence of Rickettsia bellii from the spotted fever and typhus groups of Rickettsia and antiquity of the genus Rickettsia. Int J Syst Bacteriol 1994;44:798–804.

[9] Tamura A, Ohashi N, Urakami H, Miyamura S. Classification of Rickettsia tsutsugamushi in a new genus, Orientia gen. nov., as Orientia tsutsugamushi comb. nov. Int J Syst Bacteriol 1995;45:589–91.

[10] Gillespie JJ, Beier MS, Rahman NM, Ammerman NC, Shallom JM, Purkayastha A, et al. Plasmodium and rickettsial evolution: insight from Rickettsia fels. PLoS One 2007;2: e266.

[11] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.

[12] Kumar S, Tamura K, Nei M. MEGA: Moleculer evolutionary genetics analysis. Pennsylvania State University: University Park; 1993, version 1.0.1.

[13] Gumenyuk AS, Pozdnichenko N, Rodionov I, Shpynov SN. About the means of formal analysis system of nucleotide chains [in Russian]. Mat Bio Bioinform 2013;8:373–97.

[14] Shpynov S, Pozdnichenko N, Gumenuk A. Approach for classification of microorganisms based on the analysis of comparative genomics. Biol Rev Camb Philos Soc 2011;86:379–405.

[15] Shpynov S, Pozdnichenko N, Gumenyuk AS, Pozdnichenko NN, Shpynov SN. Computer program in the Register of Computer Programs 2017616679, © 2018 The Author(s). Published by Elsevier Ltd, NMNI, 23, 93–102. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
[23] Ignatovich VF. Antigenic relations of Rickettsia prowazekii and Rickettsia Canad. established in the study of sera of patients with Brill’s disease [in Russian]. J Hyg Epidemiol Microbiol Immunol 1977;21:55–60.

[24] Linne mann Jr CC, Pretzman CL, Peterson ED. Acute febrile cerebrovascularitis. A non–spotted fever group rickettsial disease. Arch Intern Med 1989;149:1682–4.

[25] Wenzel RP, Hayden FG, Gröschel DH, Salata RA, Young WS, Greenlee JE, et al. Acute febrile cerebrovascularitis: a syndrome of unknown, perhaps rickettsial, cause. Ann Intern Med 1986;104:606–15.

[26] Jado I, Oteo JA, Aldímis Z, Gil H, Escudero R, Ibarra V, et al. Rickettsia monacensis and human disease. Spain Emerg Infect Dis 2007;9:1405–7. V. 13.

[27] Simeser JA, Rahman MS, Dreher-Lesnick SM, Azad AF. A novel and naturally occurring transposon, ISRpe1 in the Rickettsia pecosii genome disrupting the nickA gene involved in actin-based motility. Mol Microbiol 2005;71–9. V. 58.I.

[28] Felsheim RF, Kurtsi TJ, Munderloh UG. Genome sequence of the endosymbiont Rickettsia pecosii and comparison with virulent Rickettsia rickettsii: identification of virulence factors. PLoS One 2009;4(12): e8361. V. 21..

[29] Zdrodovskii PF, Golinevich HM. The rickettsial diseases. London: Pergamon Press; 1960.

[30] Krantz GW, Walter DE. A manual of acarology. 3rd ed. Lubbock: Texas Tech University Press; 2009.

[31] Weiss E, Moulder JW, Order I. Rickettsiales, gieszczkiewicz 1939. In: Krieg NR, Holt JG, editors. Bergey’s manual of systematic bacteriology. Baltimore, MD: Williams & Wilkins; 1984. p. 687–703. 1.

[32] Eremeeva ME, Dasch GA. Genetic analysis of isolates of Rickettsia rickettsii that differ in virulence. Ann N Y Acad Sci 2003;990:717–22.

[33] Eremeeva ME, Dasch GA. Closing the gaps between genotype and phenotype in Rickettsia rickettsii. Ann N Y Acad Sci 2009;1166:12–26.

[34] Jackson WR, Arbuthnot EC, McGrath EP. Comparative genome sequencing of Rickettsia streptococci strains that differ in virulence. Infect Immun 2015;83:1568–76.

[35] Ellison DW, Clark TR, Sturdevant DE, Virtaneva K, Porcella SF, Hackett T. Genomic comparison of virulent Rickettsia rickettsii Sahela Smith and avirulent Rickettsia rickettsii Iowa. Infect Immun 2008;76: 542–50.

[36] Turco J, Winkler HH. Differentiation between virulent and avirulent strains of Rickettsia prowazekii by macrophage-like cell lines. Infect Immun 1982;35:783–91.

[37] Beach Y, El Karkouri K, Mediannikov O, Leroy Q, Pelletier N, Robert C, et al. Genomic, proteomic, and transcriptomic analysis of virulent and avirulent Rickettsia prowazekii reveals its adaptive mutation capabilities. Genome Res 2010;20:655–63.

[38] Ge H, Chuang YY, Zhao S, Tong M, Tsai MH, Temenak JJ, et al. Comparative genomics of Rickettsia prowazekii Madrid E and Bre nil strains. J Bacteriol 2004;186:566–65.

[39] Bishop-Lilly KA, Ge H, Butani A, Osborne B, Verrazile K, Mokashi V, et al. Genome sequencing of four strains of Rickettsia prowazekii, the causative agent of epidemic typhus, including one flying squirrel isolate. Genome Announc 2013;1. e00399–13.

[40] Lagier JC, Armougom F, Million M, Hugon P, Pagnier I, Robert C, et al. Microbial culturoomics: paradigm shift in the human gut microbiome study. Clin Microbiol Infect 2012;18:185–93.

[41] Ramasamy D, Mishra AK, Lagier JC, Padmanabhan R, Rossi M, Sentausa E, et al. A polyphasic strategy incorporating genomic data for the taxonomic description of novel bacterial species. Int J Syst Evol Microbiol 2014;64(Pr 2):384–91.

[42] Samoilienko IE, Kumpa LV, Shpynov SN, Obert AS, Butakov OV, et al. Methods of isolation and cultivation of new rickettsiae from the Nenosaisa of the North Asian tick typhus in siberia. Ann N Y Acad Sci 2006;1078:613–6.

[43] Eisen JA, Fraser CM. Phylogenomics: intersection of evolution and genomics. Science 2003;300(5626):1706–7.

[44] Koonin EV, Makarova KS, Galperin MY, Walker DR. Comparison of archaeal and bacterial genomes: computer analysis of protein sequence profiles predicts novel functions and suggests a chimeric origin for the Archaea. Mol Microbiol 1997;25:619–37.

[45] Fournier PE, Dumler JS, Greub G, Zhang J, Wu Y, Raoult D. Gene sequence–based criteria for identification of new rickettsia isolates and description of Rickettsia heilongiangensis sp. nov. J Clin Microbiol 2003;41:5456–65.