A multilocus sequence analysis for the taxonomic update and identification of the genus *Proteus*

**CURRENT STATUS:** UNDER REVISION

Hang Dai  
National Institute for Communicable Disease Control and Prevention

Binghuai Lu  
China-Japan Friendship Hospital

Zhenpeng Li  
National Institute for Communicable Disease Control and Prevention

Zhenzhou Huang  
National Institute for Communicable Disease Control and Prevention

Hongyan Cai  
National Institute for Communicable Disease Control and Prevention

Keyi Yu  
National Institute for Communicable Disease Control and Prevention

Duochun Wang  
National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention

email: wangduochun@icdc.cn  
**Corresponding Author**

**ORCiD:** https://orcid.org/0000-0003-2372-1532

**DOI:**

10.21203/rs.3.rs-17186/v1

**SUBJECT AREAS**

*Applied & Industrial Microbiology*  
*General Microbiology*

**KEYWORDS**

*Proteus, Multilocus sequence analysis, Taxonomy, Identification*
Abstract

**Background:** The members of the genus *Proteus* are commonly opportunistic pathogens that cause a variety of infections in humans. The molecular evolutionary characteristics and genetic relationships among *Proteus* species are remain unelucidated. In this study, we developed a multilocus sequence analysis (MLSA) approach based on five housekeeping genes (HKGs) to delineate phylogenetic relationships of species within genus of *Proteus*.

**Results:** Of all 223 *Proteus* strains collected in current study, phylogenetic tree of concatenated five HKGs (*dnaJ, mdh, pyrC, recA* and *rpoD*) divided into eleven clusters, which representative of their counterpart species. Meanwhile, phylogenetic trees of the five individual HKGs were also corresponded to that of the concatenated tree, except for *recA*, which clustered four strains at an independent cluster. The evaluation of inter- and intra-species distance of HKGs concatenation, all inter-species indicated more significant different distances than those of intra-species, which revealed these HKGs concatenation can be used as gene marker to distinguish different *Proteus* species. Further web-based DNA-DNA hybridization estimated by genome of type strains confirmed the validity of the MLSA and each of eleven clusters were congruent with eleven different *Proteus* species. In addition, we used the established MLSA method to identify the randomly collected *Proteus*, and found that *P. mirabilis* is the most species of *Proteus*. However, the top second is *P. terrae*, but not *P. vulgaris*. Combined with the genetic, genomic and phenotypic characteristics, three species, *P. terrae*, *P. cibarius* and *Proteus* genospecies 5 should be regarded as the heterotypic synonyms.

**Conclusions:** Our data suggested the MLSA was a powerful method for the discrimination and classification of the *Proteus* at species level. The MLSA scheme provides a rapid, economical and precise identification of *Proteus* strains. The identification of *Proteus* species determined by the MLSA approach plays an important role in clinical diagnosis and treatment of *Proteus* infection.

**Background**
The genus Proteus belonging to the family Enterobacteriaceae, is a motile Gram-negative bacterium that survives in soil, water, and the intestinal tracts of mammals. Most members of the genus Proteus are opportunistic pathogens that cause a variety of infections in humans, including urinary tract
infection [1], wounds as well as respiratory tract, skin, eye, ear, nose, and throat infections [2].

The genus was first described by Hauser and was successively separated into two species, i.e., Proteus mirabilis and Proteus vulgaris, on the basis of their ability to ferment maltose [2]. Strains of P. vulgaris comprised three biogroups, based on three biochemical reactions, namely indole production, salicin fermentation and aesculin hydrolysis. Biogroup 1 was characteristic by being negative for those three reactions, named as P. penneri [3]. By contrast, biogroup 2 was positive for the three reactions and retained the name of P. vulgaris. Biogroup 3 was positive for indole production but negative for salicin fermentation and aesculin hydrolysis [4], and further separated into four groups by DNA-DNA hybridization that were designated Proteus genospecies 3, 4, 5 and 6 [4]. The genospecies 3 can be distinguished from Proteus genospecies 4, 5 and 6 because it is negative for Jordan’s tartrate utilization, and was named by the species of P. hauseri, while genospecies 4, 5 and 6 remained unnamed due to their undistinguishable phenotypic differentiation [4]. In addition, six newly defined species, i.e., P. terrae and P. cibarius, P. alimentorum, P. columbae, P. faecis and P. cibi were proposed recently based on phylogenetic, phenotypic, chemotaxonomic and genotypic analysis [5–9].

Thus, the genus Proteus comprises at present ten validly published species and three unnamed genospecies (4, 5 and 6).

Except for those six newly defined species, the classification of other Proteus species and genospecies were based on the difference of biochemical reactions and DNA-DNA hybridization, which were designed nineteen years ago or even more time [4, 10]. Particularly, the molecular evolutionary characteristics and genetic relationships among those Proteus phenospecies and genospecies are remain unelucidated, due to absence of a molecular typing method in Proteus genus. Multilocus sequence analysis (MLSA) based on several housekeeping genes (HKGs), has previously been successfully employed to delineate boundaries between closely-related bacterial species, subspecies and component strains [11–13]. Partial sequences of protein-encoding genes have been proven useful for species identification and as phylogenetic markers in the family Enterobacteriaceae [14, 15].

In the present study, we developed a five-gene MLSA approach to delineate genetic similarities and differences among Proteus species. We used this MLSA method to type the genotypic species of 223
Proteus strains that were identified by phenotypes. Our data indicated this MLSA was a powerful method for the discrimination, classification and phylogenetic analysis of the Proteus at species level, meanwhile, we revealed taxonomic relationship between phenotypic and genotypic species, specially, modifying two phenotypic taxonomy using this MLSA method.

Methods

Bacterial isolates and biochemical identification

A total of 223 Proteus strains were analyzed in this study. Among them, 210 strains were collected and isolated from seven provinces in China, year from 2002 to 2015. Specifically, 181 strains were isolated from clinical (faeces of diarrhoea patients, patients of nosocomial infection, like blood, urine and wound, et al) and 29 strains were isolated from food. The 210 strains were identified as Proteus species by biochemical tests (API 20E). Meanwhile, type strains of ten Proteus species and three genospecies were also obtained and used for MLSA and biochemical tests, i.e, P. mirabilis KCTC 2566T, P. vulgaris KCTC 2579T, P. penneri ATCC 33519T, P. hauseri JCM 1668T, P. cibarius JCM 30699T, P. terrae LMG 28659T, P. columbae 08MAS2615T, P. alimentorum 08MAS0041T, P. faecis TJ1636T, P. cibi FJ2001126-3T, genospecies 4 (ATCC 51469T), 5 (ATCC 51470T) and 6 (ATCC 51471T). Separate biochemical tests were assessed using agents (Guangdong HuanKai Microbial Technology Co., Ltd) in accordance with the manufacturer’s instructions. Uncertain experiments were performed in triplicate.

DNA Extraction, PCR Amplification And Sequencing

The genomic DNA from Proteus strains were extracted using a genomic DNA purification kit (Tiangen Biotech, Beijing, China) in accordance with the manufacturer’s instructions. Extracted DNA was dissolved in TE buffer and stored at -20 °C until use as PCR templates. Five candidate HKGs were used for MLSA analysis, i.e, dnaJ, mdh, pyrC, recA and rpoD. The primer sets were designed and listed in Table 1. For PCR amplification, each reaction was performed in a final volume of 50 µl containing 25 µl of 2 x Taq PCR MasterMix (Tiangen Biotech, Beijing, China), 1.5 µl 10 µM of each forward and reverse primer, 2 µl DNA template, and 20 µl ddH₂O. The reaction mixture was subjected to denaturation at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 30 sec, annealing at 52 to 55 °C for 30 sec and extension at 72 °C for 1 min/kb. An extension step of 10 min at 72 °C
was carried out following the last cycle to ensure full-length synthesis of the fragment. All PCR products of the five HKGs were commercially direct sequenced in both directions (TsingKe Biological technology, Beijing, China).

### Table 1

| Primer | Nucleotide sequence (5' to 3') | Amplicon size (bp) | Tm  |
|--------|--------------------------------|-------------------|-----|
| dnaJ-F | CRATGAAATATCACCCAGAYCG         | 790               | 55 °C |
| dnaJ-R | ACACGRCCATCMAGWTT              | 769               | 55 °C |
| mdh-F  | GCAAAGAAACGGGCAATT            | 745               | 52 °C |
| mdh-R  | CRGGTGGTATTGGGCAGG            | 707               | 52 °C |
| pyrC-F | TGGATTGCAATGGATG              | 775               | 52 °C |
| pyrC-R | GATTCTTTGCGATGTTGT            | 775               | 52 °C |
| recA-F | CTRTACCAWGCACCGGCT            | 745               | 52 °C |
| recA-R | AGGKTCTATCATGGG                | 745               | 52 °C |
| rpoD-F | CGGGAAGGGTGAATTTGAT           | 745               | 52 °C |
| rpoD-R | CGTAGACATACGACG               | 745               | 52 °C |

### MLSA Analysis

Phylogenetic trees were constructed by MLSA of the concatenated sequence of five HKGs fragments (dnaJ-mdh-pyrC-recA-rpoD, 3,157 bp) and the five individual HKGs. The total length of the alignments used were: 629 bp (dnaJ), 635 bp (mdh), 647 bp (pyrC), 701 bp (recA) and 545 bp (rpoD). Comparison analyses of the sequences were conducted with BioEdit software (Ibis Biosciences, Carlsbad, CA, USA). Clustal-W was used to perform multiple alignments of the nucleotide sequences. The phylogenetic analysis was performed using MEGA 7.0 for the maximum-likelihood (ML) method. In the ML method, the General Time Reversible model was selected and the rate matrix, the base frequencies, the invariable site proportion and the gamma distribution were determined via likelihood. Phylogenetic tree branch support estimation, 1000 replications were calculated to obtain the bootstrap values.

### Intra- And Inter-species Phylogenetic Distance Of HKGs

Intra-species phylogenetic distance was defined as the phylogenetic distance within the strains from the same species, and inter-species phylogenetic distance was defined as the phylogenetic distance of strains from a species with strains from other species. The phylogenetic distance between strains was calculated using MEGA 7.0 with Kimura 2 parameter model. The minimum, median, and maximum of intra- and inter-species for each species were calculated. Variance of compacted or
dispersive distance of species analyzed using Fisher’s exact test.

Genomic Relatedness Among Isolates Of Different Species

The genomic relatedness among isolates of different species was further evaluated by web-based DNA-DNA hybridizations (DDH), like in silico DDH (is DDH) and average nucleotide identity (ANI) to detect their similarity values [16, 17]. Is DDH values were determined using the genome-to-genome distance calculator (GGDC) web server (http://ggdc.dsmz.de/) and ANI values were measured by EZ BioCloud platform (http://www.ezbiocloud.net/tools/ani), with similarity values of 70% and 95% as the standard threshold for species boundaries, i.e., two isolates represented different species when their is DDH and ANI values were below 70% and 95% threshold, respectively [16, 17].

Results

MLSA of the five concatenated HKGs

Of all 223 Proteus strains collected in this study, phylogenetic tree of concatenated 5-gene divided them into eleven clusters (Fig. 1), representing of thirteen species. Among them, ten clusters contained one type strain of each. However, cluster 5 was comprised of three type strains, i.e, Proteus genospecies ATCC51470^T, P. cibarius JCM 30699^T and P. terra LMG 28659^T.

Expectedly, among the 223 Proteus strains, P. mirabilis (cluster 1) is the largest cluster (n = 157, 70.4%) distinctly separated from others, and there are three subclusters within this cluster. Cluster 5 is the second largest among Proteus strains (n = 33, 14.8%), followed by P. vulgaris (cluster 2) (n = 14, 6.3%) and P. penneri (cluster 7) (n = 6, 2.7%)

Identification of Proteus species by phylogenetic analysis of five individual genes

Phylogenetic trees based on five individual HKGs were also constructed (Fig. 2). Phylogenetic trees of the five HKGs (dnaJ, mdh, pyrC, recA and rpoD) can be divided into eleven clusters, representative thirteen species, and corresponding to that of the concatenated tree. Meanwhile, phylogenetic trees of four individual HKGs (dnaJ, mdh, pyrC and rpoD) were the same as that of the concatenated tree, both in numbers of species (cluster) and strain numbers within each species (cluster). There is one inconsistence between trees of recA and concatenated 5-gene: recA identified four strains as uncluster, whereas the four strains were identified by concatenated 5-gene and other four HKGs as species 6 (Fig. 2).
Inter- And Intra-species Distance Of HKGs

The inter- and intra-species distances of HKGs were summarized in boxplot of the concatenated 5-gene (Fig. 3). All inter-species indicated the obviously different distances than that of intra-species.

Among the inter-species boxplot, two species, P. mirabilis, and P. hauseri indicated compacted distance ranges (both standard deviation, SD = 0.004), whereas the remaining nine species shared dispersive distance ranges (SD ranges from 0.024 to 0.065); On the other hand, among intra-species boxplot, P. hauseri possessed compacted distance range (SD = 0.000) compared to that of five species (SD range from 0.012 to 0.058). Meanwhile, boxplots of the five individual genes (Figure S1) indicated same trends of intra and inter-species distance as that of the concatenated 5-gene, although there are small part overlapping in species 5 and 6 of pyrC. The detailed genetic distance and median values of individual genes and the concatenated 5-genes were summarized in Table S1.

Web-based DNA-DNA Hybridizations Among Species

To confirm the correctness of strains among the eleven species, we used web-based DDH, like is DDH and ANI to detect their similarity values. Among the eleven species defined in this study, the is DDH and ANI values of the type /representative strains were 23.5–51.4% and 80.8–94.4% (Table 2), less than the proposed cutoff level for species delineation, i.e., 70% and 95%, respectively. Notably, among the three subclusters within cluster 5 (Fig. 1), either among the three published type strains (Proteus genospecies ATCC51470T, P. cibarius JCM30699T and P. terrae LMG28659T) or representative strain (CA142267) among the three subclusters, their is DDH and ANI values were more than the proposed cutoff level for species delineation. The results indicated strains within the cluster 5 actually belonged to same species.
Table 2

is DDH relatedness and ANI values among the eleven species.

| Strains | a | b | c | d | e | f | g | h | i | j | k | l | m | n |
|---------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Species | P. mirabilis | P. penneri | P. vulgaris | P. hauseri | Proteus genosp. 4 | Proteus genosp. 5 | P. cibariu | P. terrae | CA14267 | Proteus genosp. 6 | P. alimentor | P. faecis | P. cib |
| DDH/ANI | DDH | ANI |
| Strains | 10 | 0.00 |
| b | 1.010 | 0.00 |
| c | 24.81 | 37.89 | 37.89 | 6.81 | 0.00 |
| d | 23.80 | 25.85 | 25.38 | 9.85 | 0.00 |
| e | 24.81 | 41.90 | 44.91 | 6.83 | 0.00 |
| f | 24.81 | 44.91 | 37.89 | 25.83 | 40.90 | 10.10 |
| g | 25.82 | 44.91 | 37.89 | 25.83 | 39.90 | 72.96 | 10.10 |
| h | 24.81 | 44.91 | 37.89 | 25.83 | 39.90 | 93.99 | 10.10 |
| i | 24.82 | 44.91 | 37.89 | 25.83 | 39.90 | 97.94 | 10.10 |
| j | 25.82 | 44.91 | 36.89 | 26.83 | 37.89 | 97.92 | 47.92 | 10.10 |
| k | 25.82 | 44.91 | 36.89 | 26.83 | 37.89 | 97.92 | 47.92 | 10.10 |
| l | 24.82 | 44.91 | 36.89 | 25.83 | 37.89 | 97.92 | 47.92 | 10.10 |
| m | 24.81 | 35.83 | 35.83 | 25.83 | 36.89 | 95.81 | 45.91 | 10.10 |
| n | 24.81 | 35.83 | 35.83 | 25.83 | 36.89 | 95.81 | 45.91 | 10.10 |

# Strain: a, P. mirabilis ATCC 29906 (GenBank accession no. ACLE00000000.1); b, P. penneri ATCC 33519 (PHF00000000); c, P. vulgaris KCTC 2579 (PHHN00000000); d, P. hauseri JCM 1668 (PGWU00000000); e, Proteus genosp. 4 ATCC 51469 (PENV00000000); f, Proteus genosp. 5 ATCC 51470 (PENU00000000); g, P. cibariu JCM 30699 (PGWT00000000); h, P. terrae LMG 28659 (PENS00000000); i, CA142267; j, Proteus genosp. 6 ATCC 51971 (PENT00000000); k, P. columb 08MAS2615 (NGVR00000000); l, P. alimentorum 08MAS041 (NBVR00000000); m, P. faecis TJ1636 (PENZ00000000); n, P. cib J2001126-3 (PENW00000000).

Results were percentages based on Formula 2, calculate distances and DDH estimates with GGDC 2; ANI values were estimated using the web-based service ANI calculator (http://www.ezbiocloud.net/tools/ani). The grey shadow marked strains with is DDH > 70% and ANI > 95%, respectively, indicating they belong to the same species.

Reclassification of Proteus genospes 5 and P. cibariu to P. terrae

Since either MLSA of the five concatenated HKGs or phylogenetic analysis of five individual genes indicated that three type strains, i.e, Proteus genospes ATCC51470, P. cibariu JCM 30699 and P. terrae LMG 28659 were fell into one cluster (cluster 5 in Fig. 1), further web-based DNA-DNA hybridizations, like is DDH and ANI confirmed that, among the three subclusters within cluster 5, either among the three type strains or representative strain (CA142267) among the three subclusters, their is DDH and ANI values were higher than the proposed cutoff level for species delineation (70%
for is DDH and 95% for ANI, Table 2). The genomic analysis provided evidences that strains within the cluster 5 actually belonged to same species.

Further phenotypic characteristics were detected among type strains of Proteus genospecies 5, P. cibarius and P. terra, and slight distinctive properties were observed (Table 3). Only minor differences were obtained between type strains of the three species, this including growth in optimum temperature, growth range in NaCl and pH, the utilization of Dnase, lipase and citric acid, and DNA G + C content. Combined with the genetic, genomic and phenotypic characteristics, three species, P. terra reported by Behrendt et al. 2015, P. cibarius reported by Hyun et al. 2016 and Proteus genospecies 5 reported by O’Hara et al. 2000 should be regarded as the heterotypic synonyms of Proteus terra reported by Behrendt et al. 2015.

Table 3

| Characteristic                        | P. terra | P. cibarius | Proteus genospecies 5 |
|---------------------------------------|----------|-------------|-----------------------|
| Growth in optimum temperature (°C)   | 37       | 35          | 37                    |
| Growth range in NaCl (%w/v)          | 0–15     | 0–12        | 0–15                  |
| Growth range in pH                   | 4–9      | 4–9         | 4–9                   |
| Dnase (25 °C) (3 days)               | +        | +           | +                     |
| Lipase (olive oil) (7 days)          | -        | +           | -                     |
| CIT                                   | -        | -           | +                     |
| DNA G + C content (mol %)            | 37.9     | 37.8        | 37.8                  |

#Species strain: P. terra LMG 28659T; P. cibarius JCM 30699T; Proteus genospecies 5 ATCC 51470T.

Discussion

MLSA has been used for classification at species level in numerous Enterobacteriaceae [14, 15, 18–23]. Normally, four to seven HKGs were selected for MLSA to determine phylogenetic relationships. It has been advised to use sequence data from more than one gene, to reduce the possibility of ambiguities caused by genetic recombination or specific selection. MLSA is increasingly applied in order to obtain a higher resolution power between species within a genus [24]. In this study, when amplified by PCR of 223 tested Proteus strains collected, the five HKGs (dnaJ, mdh, pyrC, recA, and rpoD) have a good corresponding relationship of consistency among different species. Thus, we established the MLSA method with the five genes for taxonomic analysis of Proteus genus. Our MLSA-based approach can be used to effectively discriminate Proteus sp., and enable the delineation
of species boundaries with high confidence.

Our MLSA method divided all 223 Proteus strains into eleven clusters, representative of eleven species. Among the eleven species, P. mirabilis was the majority species collected in this study, which agree with numerous reports of Proteus genus classified by phenotypic methods [2]. However, even all P. mirabilis isolates were phenotypic with the same distinguishing biochemical features, i.e, positive for ornithine decarboxylase but negative for sucrose and maltose only. Species P. mirabilis can be further divided into three dominant subcluster. In contrary, species P. vulgaris was the most conservative cluster than any others, which exhibited one of the minimum intra-species distances of HKGs among eleven species (Fig. 2). Interestingly, P. hauseri was phylogenetically more closely to P. mirabilis than any other species (Fig. 1), although P. hauseri was once belonged to biogroup 3 of P. vulgaris [4].

Cluster 5 included three subclusters, our web-based DDH indicated that strains within the cluster (including three type strains, Proteus genospecies ATCC 51470T, P. cibarius JCM 30699T and P. terra LMG 28659T) actually belong to the same species. P. cibarius and P. terra were defined as new species of the genus Proteus maybe because both studies were excluded type strain of Proteus genospecies 5 (such as ATCC 51470 T) [4], meanwhile, papers of the two species were accepted for publication in a very near time (year of 2015 and 2016) at different journals [5, 6], so that they did not cited each other.

Proteus is the most common opportunistic pathogen, of which P. mirabilis and P. vulgaris have long been considered as the two most common species [2, 25, 26]. Clinically, different treatment schemes may adopt according to different species of Proteus [27, 28]. In this study, we used the established MLSA method to identify the randomly collected Proteus, and found that P. mirabilis is the most species of Proteus. However, the top second is P. terra, but not P. vulgaris, this result is quite different from that of clinical phenotype identification [2]. The reason is that in clinical, strains of Proteus genospecies 4, 5 and 6 have long been identified as P. vulgaris by phenotypic biochemical reactions [4]; meanwhile, as the result of this study, Proteus genospecies 5 accounts for a large
proportion (Fig. 1); moreover, P. penneri and P. hauseri are initially classified as different biogroup of P. vulgaris [4]. In view of the accurate identification of species level is of great significance for the clinical treatment of Proteus infection, therefore, MLSA based identification should be suggested in the classification of Proteus genus.

We therefore established a MLSA method for genus of Proteus spp. and figured out the relationship between phenospecies and MLSA based clusters (genospecies). Firstly, in compare with the phenotypic biochemical classification (species) method, all tested strains can be divided into eleven clusters (genospecies) by the MLSA method. The web-based DDH further confirmed the validity of the MLSA analysis and the eleven clusters (genospecies) are congruent with eleven species. Second, our study revealed the phenospecies of strains composed of different genotypes at different phylogenetic scales, for instance, the most common phenospecies of P. mirabilis can be further divided into three genotypes; similarly, cluster 5 included three different evolutionary genotypes. Thirdly, our MLSA method proposed the emendation of description of the genus Proteus: two proposed Proteus species recently, P. cibarius [5] and P. terrae [6], should actually be the same species as Proteus genospecies 5.

Declarations

Acknowledgments

This work was supported by grants from the National Natural Science Foundation of China (31570134) and the National Sci-Tech Key Project (2018ZX10734404, 2018ZX10102001) from the Ministry of Health, China.

Author’ contributions

DW and HD conceived and designed experiments. BL and HD collected strains. HD, ZH, HC and KY performed laboratory experiments. ZL analyzed and interpreted data. ZL and ZH contributed reagents and materials. DW and HD wrote the manuscript. DW revised the manuscript. All authors have read and approved the final version of the manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (31570134)
and the National Sci-Tech Key Project (2018ZX10102001, 2018ZX10734404) from National Health Commission, China.

### Availability of data and materials

The nucleotide sequences of five HKGs are deposited in GenBank nucleotide sequence database, under accession numbers of *dnaJ*: MG492023-MG492065, MG492068-MG492222, MG492228, MG492230-MG492232, MG492234-MG492250; *mdh*: MG492251-MG492295, MG492298-MG492452, MG492458, MG492460, MG492462-MG492478; *pyrC*: MG492479-MG492515, MG492518-MG492672, MG492678, MG492680, MG492682-MG492706; *recA*: MG492707-MG492733, MG492735, MG492737-MG492738, MG492743-MG492897, MG492900-MG492934; *rpoD*: MG492935-MG492959, MG492961, MG492963, MG492969-MG493125, MG493128-MG493162.

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declared that they have no competing interests.

### References

1. Li X, Zhao H, Lockatell CV, Drachenberg CB, Johnson DE, Mobley HL: *Visualization of Proteus mirabilis within the matrix of urease-induced bladder stones during experimental urinary tract infection*. Infect Immun. 2002; 70(1):389-394.

2. O'Hara CM, Brenner FW, Miller JM: *Classification, identification, and clinical significance of Proteus, Providencia, and Morganella*. Clin Microbiol Rev. 2000; 13(4):534-546.

3. Hickman FW, Steigerwalt AG, III JJF, Brenner. DJ: *Identification of Proteus penneri sp. nov., formerly known as Proteus vulgaris indole negative or as Proteus vulgaris biogroup 1*. J Clin Microbiol. 1982; 15:1097-1102.
4. O'Hara CM, Brenner FW, Steigerwalt AG, Hill BC, Holmes B, Grimont PA, Hawkey PM, Penner JL, Miller JM, Brenner DJ: Classification of *Proteus vulgaris* biogroup 3 with recognition of *Proteus hauseri* sp. nov., nom. rev. and unnamed *Proteus* genomospecies 4, 5 and 6. Int J Syst Evol Microbiol. 2000; 50 Pt 5:1869-1875.

5. Hyun DW, Jung MJ, Kim MS, Shin NR, Kim PS, Whon TW, Bae JW: *Proteus cibarius* sp. nov., a swarming bacterium from Jeotgal, a traditional Korean fermented seafood, and emended description of the genus *Proteus*. Int J Syst Evol Microbiol. 2016; 66(6):2158-2164.

6. Behrendt U, Augustin J, Sproer C, Gelbrecht J, Schumann P, Ulrich A: Taxonomic characterisation of *Proteus terrae* sp. nov., a N2O-producing, nitrate-ammonifying soil bacterium. Antonie van Leeuwenhoek. 2015; 108(6):1457-1468.

7. Dai H, Wang Y, Fang Y, Xiao T, Huang Z, Kan B, Wang D: *Proteus columbae* sp. nov., isolated from a pigeon in Ma'anshan, China. Int J Syst Evol Microbiol. 2018; 68(2):552-557.

8. Dai H, Wang Y, Fang Y, Huang Z, Kan B, Wang D: *Proteus alimentorum* sp. nov., isolated from pork and lobster in Ma'anshan city, China. Int J Syst Evol Microbiol. 2018; 68(4):1390-1395.

9. Dai H, Chen A, Wang Y, Lu B, Chen J, Huang Y, Li Z, Fang Y, Xiao T, Cai H et al: *Proteus faecis* sp. nov., and *Proteus cibi* sp. nov., two new species isolated from food and clinical samples in China. Int J Syst Evol Microbiol. 2019; 69(3):852-858.

10. Brenner DJ, Hickman-Brenner FW, Holmes B, Hawkey PM, Penner JL, Grimont PA, O'Hara CM: Replacement of NCTC 4175, the current type strain of *Proteus vulgaris*, with ATCC 29905. Request for an opinion. Int J Syst Bacteriol. 1995; 45(4):870-871.
11. Perez-Cataluna A, Lucena T, Tarazona E, Arahal DR, Macian MC, Pujalte MJ: **An MLSA approach for the taxonomic update of the Splendidus clade, a lineage containing several fish and shellfish pathogenic Vibrio spp.** Syst Appl Microbiol. 2016; 39(6):361-369.

12. Tampakaki AP, Fotiadis CT, Ntatsi G, Savvas D: **Phylogenetic multilocus sequence analysis of indigenous slow-growing rhizobia nodulating cowpea (Vigna unguiculata L.) in Greece.** Syst Appl Microbiol. 2017; 40(3):179-189.

13. Yanokura E, Oki K, Makino H, Modesto M, Pot B, Mattarelli P, Biavati B, Watanabe K: **Subspeciation of Bifidobacterium longum by multilocus approaches and amplified fragment length polymorphism: Description of B. longum subsp. suillum subsp. nov., isolated from the faeces of piglets.** Syst Appl Microbiol. 2015; 38(5):305-314.

14. Martinez-Murcia AJ, Monera A, Saavedra MJ, Oncina R, Lopez-Alvarez M, Lara E, Figueras MJ: **Multilocus phylogenetic analysis of the genus Aeromonas.** Syst Appl Microbiol. 2011; 34(3):189-199.

15. Hall M, Chattaway MA, Reuter S, Savin C, Strauch E, Carniel E, Connor T, Van Damme I, Rajakaruna L, Rajendram D et al: **Use of whole-genus genome sequence data to develop a multilocus sequence typing tool that accurately identifies Yersinia isolates to the species and subspecies levels.** J Clin Microbiol. 2015; 53(1):35-42.

16. Meier-Kolthoff JP, Auch AF, Klenk HP, Goker M: **Genome sequence-based species delimitation with confidence intervals and improved distance functions.** BMC bioinformatics. 2013; 14:60.

17. Lee I, Kim YO, Park SC, Chun J: **OrthoANI: An improved algorithm and software for calculating average nucleotide identity.** Int J Syst Evol Microbiol. 2015;
18. Ioannidou V, Ioannidis A, Magiorkinis E, Bagos P, Nicolaou C, Legakis N, Chatzipanagiotou S: **Multilocus sequence typing (and phylogenetic analysis) of Campylobacter jejuni and Campylobacter coli strains isolated from clinical cases in Greece.** BMC research notes. 2013; 6:359.

19. Alvarez-Perez S, de Vega C, Herrera CM: **Multilocus sequence analysis of nectar pseudomonads reveals high genetic diversity and contrasting recombination patterns.** PloS one. 2013; 8(10):e75797.

20. Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S, Hoste B, Munn CB, Swings J: **Phylogeny and molecular identification of vibrios on the basis of multilocus sequence analysis.** Appl Environ Microbiol. 2005; 71(9):5107-5115.

21. Brady C, Cleenwerck I, Venter S, Coutinho T, De Vos P: **Taxonomic evaluation of the genus Enterobacter based on multilocus sequence analysis (MLSA): proposal to reclassify E. nimipressuralis and E. amnigenus into Lelliottia gen. nov. as Lelliottia nimipressuralis comb. nov. and Lelliottia amnigena comb. nov., respectively, E. gergoviae and E. pyrinus into Pluralibacter gen. nov. as Pluralibacter gergoviae comb. nov. and Pluralibacter pyrinus comb. nov., respectively, E. cowanii, E. radicincitans, E. oryzae and E. arachidis into Kosakonia gen. nov. as Kosakonia cowanii comb. nov., Kosakonia radicincitans comb. nov., Kosakonia oryzae comb. nov. and Kosakonia arachidis comb. nov., respectively, and E. turicensis, E. helveticus and E. pulveris into Cronobacter as Cronobacter zurichensis nom. nov., Cronobacter helveticus comb. nov. and Cronobacter pulveris comb. nov., respectively, and emended description of the genera Enterobacter and Cronobacter.** Syst Appl Microbiol. 2013; 36(5):309-319.
22. Dikow RB: Genome-level homology and phylogeny of *Shewanella* 
   (*Gammaproteobacteria: Iteromonadales: Shewanellaceae*). BMC genomics. 
   2011; 12:237.

23. Colston SM, Fullmer MS, Beka L, Lamy B, Gogarten JP, Graf J: Bioinformatic genome 
   comparisons for taxonomic and phylogenetic assignments using *Aeromonas* 
   as a test case. mBio. 2014; 5(6):e02136.

24. Glaeser SP, Kampfer P: Multilocus sequence analysis (MLSA) in prokaryotic 
   taxonomy. Syst Appl Microbiol. 2015, 38(4):237-245.

25. Rozalski A, Sidorczyk Z, Kotelko K: Potential virulence factors of *Proteus bacilli*. 
   Microbiol Mol Biol Rev. 1997; 61(1):65-89.

26. Muller HE: Occurrence and pathogenic role of *Morganella-Proteus-Providencia* 
   group bacteria in human feces. J Clin Microbiol. 1986; 23(2):404-405.

27. Cohen MM, Jr.: *Proteus* syndrome review: molecular, clinical, and pathologic 
   features. Clin Genet. 2014; 85(2):111-119.

28. Hamilton AL, Kamm MA, Ng SC, Morrison M: *Proteus* spp. as Putative 
   Gastrointestinal Pathogens. Clin Microbiol Rev. 2018; 31(3). pii: e00085-17.

Figures
Figure 1

Phylogenetic reconstruction of Proteus strains based on concatenated dnaJ, mdh, pyrC, recA and rpoD gene sequences. The tree is based on 3,157 nt of common sequences. Analysis was done using the for the maximum-likelihood method. The scale bar indicates substitutions per site. The type strains were indicated in each cluster, if included. Strain number of each species is shown in parentheses.
Figure 2
Phylogenetic reconstructions of Proteus strains based on five individual genes and their identification with five concatenated gene of species. Strain number of each species is shown in parentheses. The scale bar indicates substitutions per site.
Figure 3

Intra- and interspecies distances of eleven species infer by concatenated 5-gene MLSA. In each boxplot, from bottom to top: minimum, median and maximum. Inter: inter-species standard deviation (SD); Intra: intra-species SD; Species 4, 7-10 including only one value and thus no intra-species SD, marked NA.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Supplementary materials.pdf