Enterobacter hormaechei subsp. hoffmannii subsp. nov., Enterobacter hormaechei subsp. xiangfangensis comb. nov., Enterobacter roggenkampii sp. nov., and Enterobacter muelleri is a later heterotypic synonym of Enterobacter asburiae based on computational analysis of sequenced Enterobacter genomes. [version 2; peer review: 2 approved]

Granger G. Sutton, Lauren M. Brinkac, Thomas H. Clarke, Derrick E. Fouts

J Craig Venter Institute, Rockville, MD, 20850, USA

Abstract

Background: The predominant species in clinical Enterobacter isolates is E. hormaechei. Many articles, clinicians, and GenBank submissions misname these strains as E. cloacae. The lack of sequenced type strains or named species/subspecies for some clades in the E. cloacae complex complicate the issue.

Methods: The genomes of the type strains for Enterobacter hormaechei subsp. oharae, E. hormaechei subsp. steigerwaltii, and E. xiangfangensis, and two strains from Hoffmann clusters III and IV of the E. cloacae complex were sequenced. These genomes, the E. hormaechei subsp. hormaechei type strain, and other available Enterobacter type strains were analysed in conjunction with all extant Enterobacter genomes in NCBI’s RefSeq using Average Nucleotide Identity (ANI).

Results: There were five recognizable subspecies of E. hormaechei: E. hormaechei subsp. hoffmannii subsp. nov., E. hormaechei subsp. xiangfangensis comb. nov., and the three previously known subspecies. One of the strains sequenced from the E. cloacae complex was not a novel E. hormaechei subspecies but rather a member of a clade of a novel species: E. roggenkampii sp. nov.. E. muelleri was determined to be a later heterotypic synonym of E. asburiae which should take precedence.

Conclusion: The phylogeny of the Enterobacter genus, particularly the cloacae complex, was re-evaluated based on the type strain genome sequences and all other available Enterobacter genomes in RefSeq.
Keywords
Enterobacter, hormaechei, steigerwaltii, oharae, xiangfangensis, hoffmannii, roggenkampii, Prokaryote Code

Corresponding author: Granger G. Sutton (GSutton@jcvi.org)

Author roles: Sutton GG: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Software, Validation, Writing – Original Draft Preparation, Writing – Review & Editing; Brinkac LM: Data Curation, Methodology, Visualization, Writing – Review & Editing; Clarke TH: Formal Analysis, Methodology, Software, Validation, Visualization, Writing – Review & Editing; Fouts DE: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Software, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under award number U19AI110819. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Sutton GG et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the Creative Commons Zero "No rights reserved" data waiver (CC0 1.0 Public domain dedication).

How to cite this article: Sutton GG, Brinkac LM, Clarke TH and Fouts DE. Enterobacter hormaechei subsp. hoffmannii subsp. nov., Enterobacter hormaechei subsp. xiangfangensis comb. nov., Enterobacter roggenkampii sp. nov., and Enterobacter muelleri is a later heterotypic synonym of Enterobacter asburiae based on computational analysis of sequenced Enterobacter genomes. [version 2; peer review: 2 approved] F1000Research 2018, 7:521 https://doi.org/10.12688/f1000research.14566.2

First published: 01 May 2018, 7:521 https://doi.org/10.12688/f1000research.14566.1
Introduction

The name *Enterobacter hormaechei* was created for a taxon at the rank of species that had previously been called Enteric Group 75. O’Hara et al. defined the type strain to be ATCC 49162 from the 23 strains they studied. Twelve of the strains were shown to be closely related via DNA-DNA hybridization (DDH) and less closely related to other *Enterobacter* species. Numerous biochemical assays were performed on the 23 strains to characterize and differentiate the new species.

Hoffmann and Roggenkamp investigated the genetic structure of the *E. cloacae* complex (the set of species included in this complex has varied over time) by a combination of sequencing of the three housekeeping genes *hsp60*, *rpoB*, and *hemB*; and PCR-restriction fragment length polymorphism (PCR-RFLP) analysis of *ampC*. They detected 12 genetic clusters (I-XII) based most exhaustively on the *hsp60* sequencing. Three of the clusters (cluster III, 58 strains; cluster VI, 28 strains; and cluster VIII, 59 strains) accounted for 70% of the 206 strains studied. The authors noted that “Only 3% of our study strains clustered with the type strain of *E. cloacae.*” (cluster XI), “We found that 3% of our study strains clustered around the *E. hormaechei* type strain.” (cluster VII), and “Our clusters VI and VIII were closely related to *E. hormaechei* cluster VII. DDH studies are needed to verify whether these clusters form a common DNA relatedness group allowing emending and broadening of the species description of *E. hormaechei.*”

Hoffmann et al. followed up with a characterization of clusters VI, VII, and VIII asserting based on DDH that these clusters were subspecies of the same species. Since cluster VII contained the type strain for *E. hormaechei* Hoffmann et al. named cluster VII *E. hormaechei* subsp. *hormaechei*, cluster VI *E. hormaechei* subsp. *oharae*, and cluster VIII *E. hormaechei* subsp. *steigerwaltii*. Forty-eight strains were characterized using 129 biochemical tests showing that there were phenotypic differences between the subspecies. Unfortunately the authors did not decide to include the other predominant cluster (III) in their analysis, nor did they validly publish these subspecies names. This was rectified recently in Validation List no. 172.

Gu et al. defined *E. xiangfangensis* using a phylogenetic tree based upon concatenated partial *rpoB*, *atpD*, *gyrB* and *infB* gene sequences from a novel isolate and existing type strains where *E. xiangfangensis* grouped closest to *E. hormaechei*. Biochemical assays were performed and *E. xiangfangensis* strains were differentiable from the *E. hormaechei* type strain.

During analysis of the *E. cloacae* complex and *E.* (now *Klebsiella*) *aerogenes* strains looking at antimicrobial resistance patterns, many of the Hoffmann et al. clusters were rediscovered using whole genome comparisons such as SNP analysis and average nucleotide identity (ANI). The clusters were identifiable by the *hsp60* sequences deposited by the Hoffmann group. The three subspecies of *E. hormaechei* defined by Hoffmann et al. fell within the expected ANI range for bacterial species, being greater than 95% ANI between subspecies and greater than 98% ANI within a subspecies. Unexpectedly Hoffmann cluster III also met the ANI criteria to be an *E. hormaechei* subspecies. Further, genomes named *E. xiangfangensis* in GenBank fell within the *E. hormaechei* subsp. *steigerwaltii* cluster rather than a separate cluster. Moreover, most of the genomes in these clusters were mistakenly identified as *E. cloacae* when they were submitted to GenBank. To resolve the naming inconsistencies of these genomes the type strains for *E. hormaechei* subsp. *steigerwaltii*, *E. hormaechei* subsp. *oharae*, *E. xiangfangensis*, Hoffmann cluster III, and Hoffmann cluster IV were sequenced.

Tools for bacterial species assignment have changed over time. Initially, morphology as viewed through a microscope and later aided by staining such as Gram staining to distinguish cell wall differences was used. Biochemical assays and other methods to determine phenotype followed. Use of the genome started with DNA-DNA hybridization (DDH) where a 70% threshold for species followed later by a 79% threshold for subspecies were proposed. Widespread use of marker genes in particular the 16S rRNA gene made assays easier. A threshold of less than 97% identity for the 16S rRNA gene was used to determine a new species but values above 97% could not guarantee that isolates were the same species. The sequence of other less conserved marker genes such as *hsp60* has also been used to differentiate species. More recently multiple marker genes are sequenced and a combined alignment is used. With the advent of inexpensive genome sequencing, computing ANI, which correlates very closely with DDH, has largely supplanted other methods. Studies have shown that an ANI threshold between 94-96.5% correlates well with existing species definitions and 97-98% for subspecies. DDH has been shown to not only correlate with ANI but also with how many of the genes or what fraction of the genomes are shared in common so some ANI based tools take this measurement into account as well. Most definitions of new species involve sequencing the genome and taking ANI and shared gene content into account in some fashion but many species definitions predate genome sequencing and some type strains have not been sequenced. There is no generally accepted method for reconciling older species definitions with genome comparisons but usually ANI and shared gene content form a basis for the analysis.

As Hoffmann and others discovered the predominant species in clinical *Enterobacter* isolates is *E. hormaechei*. Unfortunately many articles, clinicians, and GenBank submissions misname these strains as *E. cloacae* perhaps as a short hand for the *E. cloacae* complex and possibly due to the
The *E. cloacae* complex was shown to have 18 clades (A-R), 12 of which corresponded to 11 of the 12 clusters defined previously by Hoffmann. Hoffmann cluster X is *E. nimipressuralis* which has been reclassified as *Lelliottia nimipressuralis*\(^{26}\). Table 1 incorporates more recently sequenced genomes and published papers adding four clades (S-V) and incorporating the latest literature. For example, clade R (Hoffmann cluster IX) was recently defined to be *E. bugandensis*\(^{11}\).

### Results

All RefSeq genomes labelled as being in the genus *Enterobacter* were downloaded from NCBI RefSeq resulting in 1,249 genomes. A fast approximate ANI tool, called MASH\(^{13}\), was used to generate a pairwise ANI based distance matrix and average linkage hierarchical clustering was used to generate the tree shown in Figure 1. 1,216 genomes were assigned to 22 clades (A-V Table 1) in the *E. cloacae* complex (Supplemental Table 1) while 30 genomes were deemed to be outliers and not in the *Enterobacter* genus (best MASH matches in Supplemental Table 2) as well as 2 *E. ligiousicus* genomes and 1 *E. timonensis* genome deemed to be outside of the *E. cloacae* complex. Two species of *Enterobacter*: *E. xianensis* and *E. tabaci* do not have sequenced genomes and their type strains’ 16S rRNA sequences while having full length matches at 98% and 99% respectively to some *E. cloacae* complex genomes did not have definitive matches to any particular clade. The type strains for *E. asburiae* and *E. muelleri* fall within the same clade (J – Hoffmann cluster I). All 78 genomes in this clade are above the 95% ANI species cut-off (Table 2) but using a 98% ANI subspecies cut-off produces 8 subclades of sizes 1, 1, 2, 2, 2 (*E. muelleri*), 3 (*E. asburiae*), 24, and 43. Thus *E. muelleri*\(^{15}\) is a later heterotypic synonym of *E. asburiae*\(^{28}\) which should take precedence.

Whether the 8 subclades of *E. asburiae* should be treated as subspecies is beyond the scope of this paper but is revisited in the Discussion section.

Five clades (A-E) are above the 95% ANI cut-off to be considered the same species (Table 2). Almost all within-clade pairwise ANIs are greater than between-clade ANIs (Table 2) and all genomes within a clade had the highest pairwise ANI to the type strain for that clade, supporting that these are distinct subspecies. Based on *hsp60* sequences, clade A containing the *E. xiangfangensis* type strain is Hoffmann cluster VI; clade B containing the *E. hormaechei* subsp. *steigerwaltii* type strain is Hoffmann cluster VIII; clade C containing the *E. hormaechei* subsp. *oharae* type strain is also Hoffmann cluster VI; clade D containing the Hoffmann cluster III type strain (proposed name *E. hormaechei* subspp. *hoffmannii* subspp. *nov.*) is Hoffmann cluster III; and clade E containing the *E. hormaechei* subspp. *hormaechei* type strain is Hoffmann cluster VII.

While we believe that ANI and other similar measures recently categorized as overall genome related index (OGRI)\(^{15}\) should be used for species/subspecies determination, phenotypic differences due to gene content may play a role particularly for delineation of subspecies. To explore the gene content differences of the *E. cloacae* complex and the *E. hormaechei* subspecies in particular, the pan-genome of the 1,216 *E. cloacae* complex genomes was determined using PanOCT\(^{16}\). The pan-genome generates orthologous gene clusters that delineate which genes are in common between the clades and which genes differentiate the clades (Supplemental Table 3 and Supplemental Table 4). There were 2,966 genes in “common to all” of the clades (present in 90% of the genomes of each clade). The number of genes “specific to” a clade (present in 90% of the genomes of that clade and in less than 10% of genomes from any other clade) varied from 0 (L) to 465 (V). The number of genes “missing from” a clade (present in less than 10% of the genomes of that clade and present in at least 90% of the genomes of all other clades) varied from 0 (A,C,H,K,O) to 40 (U). The clades which represent named species and subspecies show no qualitative difference in gene content from clades with no named species (Supplemental Table 4). In particular, clade D which is the proposed *E. hormaechei* subspp. *hoffmannii* has more genes specific to it than 3 of the 4 recognized subspecies. The gene content numbers need to be looked at carefully since they depend on the number of genomes in a clade (T has 187 clade specific genes but this is based on a single genome which means it is really strain specific genes rather than species specific), the distance from other clades (V the most distant clade has 465 specific genes and also has only 3 genomes), and sampling bias such as if most genomes in a clade are from a clonal outbreak. Gene content analysis can also be confounded by misassembly or misannotation of draft genomes which is why we use RefSeq genomes which have passed a quality screen and are consistently annotated. Again we emphasize that ANI as our primary criterion appears to have less of these subjective issues to deal with.

Biochemical and other properties of the *E. hormaechei* subspp. clades have been previously published\(^{5,6}\) except for clade D. These biochemical properties were used to differentiate between the subspecies but not between other species within the *E. cloacae* complex. With the availability of whole genome sequences and pan-genome analysis tools some of the observed phenotypic traits can be assigned to genetic features, such as the presence or absence of protein coding genes for known metabolic pathways. *E. hormaechei* subspp. *hormaechei* was previously distinguished from *E. hormaechei* subspp. *oharae* and *E. hormaechei* subspp. *steigerwaltii* by growth on dulcitol (a.k.a. galactitol) as the sole carbon source\(^{1}\). This phenotype can be explained by the presence of a *gat operon*\(^{37}\) within all 7 of the *hormaechei* subspp. genomes while none of *oharae*, *steigerwaltii*, or *hoffmannii* genomes have the *gat operon*. In the same genomic location, between the D-galactarate dehydratase gene and the 16S rRNA methyltransferase gene, all of the *steigerwaltii*, *oharae*, and *hoffmannii* subspp. genomes have a related, but different operon, encoding for N-acetyl galactosamine metabolism (a.k.a., the *aga operon*)\(^{38}\). For *xiangfangensis* most (222 out of 255) of the genomes have the *aga operon* but 33 have the *gat operon* instead. Similarly, *steigerwaltii* isolates can be distinguished from *hormaechei*, *oharae*, *xiangfangensis*, and *hoffmannii* by their ability to grow on adonitol (a.k.a. ribitol) and D(+)-arabitol;
Table 1. Type and proxy strain genomes for Enterobacter cloacae complex clades. *E. lignolyticus* and *E. timonensis* have not been validly published and are deemed to be outside of the *E. cloacae* complex. *E. siamensis* and *E. tabaci* do not have sequenced genomes but based on their 16S rRNA genes may be in the *E. cloacae* complex. Proxy indicates whether a type or proxy strain was available. The last two columns are for the clade (A-V) and Hoffmann cluster (I-XII).

| Short ID | BioSample ID | Current name | Proposed name | Strain | Proxy |
|----------|--------------|--------------|--------------|--------|-------|
| ATCC35953 | SAMN03742638 | *E. asburiae* | *E. asburiae* | ATCC 35953 | type J I |
| obactermuelleri | SAMEA103972944 | *E. muelleri* | *E. asburiae* | JM-458 | type J I |
| cterbugandensis | SAMEA104115216 | *E. bugandensis* | *E. asburiae* | EB-247 | type R IX |
| ltercancerogenus | SAMEA104113916 | *E. cancerogenus* | *E. cancerogenus* | ATCC 33241 | type U |
| 1161ECLO | SAMN03197118 | *E. cloacae* complex | 1161_ECLO | proxy K |
| GN02587 | SAMN03732717 | *E. cloacae* complex sp. GN02587 | *E. cloacae* complex | GN02587 | proxy L |
| DS11005 | SAMN07448201 | *E. cloacae* complex | DS11005 | proxy N |
| GN05526 | SAMN04578342 | *E. cloacae* complex sp. GN05526 | *E. cloacae* complex | GN05526 | proxy O |
| 624ECLO | SAMN03197824 | *E. cloacae* complex | 624_ECLO | proxy P |
| ND22 | SAMN05212257 | *E. cloacae* complex | ND22 | proxy S |
| C9 | SAMN06237083 | *E. cancerogenus* | *E. cloacae* complex clade T | C9 | proxy T |
| ATCC13047 | SAMN02603901 | *E. cloacae* ssp. cloacae | *E. cloacae* ssp. cloacae | ATCC 13047 | type G XI |
| SDM | SAMN02603521 | *E. cloacae* ssp. dissolvens | *E. cloacae* ssp. dissolvens | SDM | proxy H XII |
| DSM14563 | SAMN05581748 | *E. cloacae* complex Hoffmann cluster III | *E. hormaechei* ssp. hoffmannii | DSM 14563 | type D III |
| ATCC49162 | SAMN05581740 | *E. hormaechei* ssp. hoffmannii | *E. hormaechei* ssp. hoffmannii | ATCC 49162 | type E VII |
| DSM16687 | SAMN05581749 | *E. hormaechei* ssp. oharae | *E. hormaechei* ssp. oharae | DSM 16687 | type C VI |
| DSM16691 | SAMN05581751 | *E. hormaechei* ssp. steigerwaltii | *E. hormaechei* ssp. steigerwaltii | DSM 16691 | type B VIII |
| LMG27195 | SAMN05581746 | *E. xiangfangensis* | *E. hormaechei* ssp. xiangfangensis | LMG27195 | type A VI |
| DSM13645 | SAMN05581747 | *E. kobei* | *E. kobei* | DSM 13645 | type Q II |
| EN119 | SAMN05787341 | *E. ludwigii* | *E. ludwigii* | EN-119 | type I V |
| LMG25706 | SAMN02471025 | *E. mori* | *E. mori* | LMG 25706 | type F |
| DSM16690 | SAMN05581750 | *E. cloacae* complex Hoffmann cluster IV | *E. rogenkampfii* | DSM 16690 | type M IV |
| rterobactersoli | SAMEA104113920 | *E. soli* | *E. soli* | LMG 25861 | type V |
| SCF1 | SAMN00116754 | *E. lignolyticus* | *E. lignolyticus* | SCF1 | type |
| mt20 | SAMEA3859023 | *E. timonensis* | *E. timonensis* | mt20 | type |
| No genome | E. siamensis | E. siamensis | | |
| No genome | E. tabaci | E. tabaci | | |

Both 5 carbon sugar alcohols known as penitols. The rbt and dal operons known from *Klebsiella aerogenes*, which metabolize ribitol and D(+)-arabitol respectively⁷⁻⁹, account for this difference where all 325 steigerwaltii genomes contain these operons but only 1 hoffmannii and no other hormaechei subsp. genomes do. The gat, aga, and rbt/dal operons are not limited to the *E. hormaechei* clades but appear in some other *E. cloacae* complex species as shown in Supplemental Table 6. *E. hormaechei* subsp. *hoffmannii* has 25 clade specific genes 10 of which (clusters 28856-28865 Supplemental Table 3) occur...
as a unit between core clusters (16694-5) and another 6 (15153-15156, 27141-2) occur between core clusters (17653-4). These clusters have no or vague annotation but are intriguing targets to provide functional phenotypic differences.

**Methods**

**MASH**\(^5\) is a very fast tool for determining approximate pairwise ANI values given sequenced genomes. A PERL script was used to invoke the following command to generate a set of MASH (version 2.0) sketches of k-mer size 16 for the 1,249 downloaded *Enterobacter* genomes:

```
mash sketch -k 16 -o Enter.Sketch.file [List of the Genomes]
```

The resulting sketches file was then used to compare all the genomes against each other with an additional PERL script which calls MASH (version 2.0) with the command:

```
Mash dist Enter.Sketch.file [List of the Genomes]
```

which generated data that could be extracted into an all versus all ANI comparison (Supplemental Table 5). We used the GGRaSP\(^6\) R package (version 1.0) which generated an ultrametric tree by using the R hclust function with average linkage from the distance matrix calculated by subtracting 100 from the MASH ANI results. The result was translated into Newick format with the APE\(^4\) R package (Supplemental File 1) rendered with metadata annotated using the Interactive Tree of Life\(^5\) into Figure 1.

Based on the tree 30 genomes were deemed to be outliers and probably not in the *Enterobacter* genus as well as 2 *E. lignolyticus* genomes and 1 *E. timonensis* genome deemed to be outside of the *E. cloacae* complex. These 30 genomes were compared to all genome sequenced bacterial type strains from NCBI RefSeq (Supplemental Table 2) using MASH which confirmed that these genomes were likely misnamed as *Enterobacter*. The decision to leave *E. lignolyticus* and *E. timonensis* outside of the *E. cloacae* complex was based on two reasons:
Table 2. Pairwise Average nucleotide identity (ANI) values within and between the Enterobacter cloacae complex clades. Mean and standard deviation are shown above and below the minimum and maximum pairwise identity values below. The last two rows show E. liquifaciens (L) and E. timonensis (T) which have consistently lower ANI values.

| A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 98.77 ±0.46 | 96.96 ±0.13 | 97.01 ±0.15 | 96.17 ±0.32 | 94.53 ±0.43 | 89.80 ±0.29 | 88.63 ±0.31 | 87.65 ±0.31 | 89.49 ±0.28 | 89.39 ±0.28 | 89.16 ±0.28 | 87.86 ±0.27 | 86.94 ±0.29 | 90.03 ±0.29 | 93.77 ±0.19 | 89.85 ±0.15 | 86.89 ±0.38 |
| (97.9-96.2) | (96.3-97.5) | (95.9-96.9) | (93.9-95.2) | (88.9-91.0) | (87.5-90.9) | (86.3-90.8) | (87.1-91.3) | (86.8-89.3) | (90.6-91.2) | (88.8-90.5) | (88.4-89.3) | (90.0-89.5) | (93.2-94.5) | (89.3-90.4) | (88.1-89.8) |
| 96.96 ±0.13 | 96.96 ±0.13 | 97.33 ±0.17 | 95.98 ±0.41 | 94.48 ±0.42 | 98.47 ±0.33 | 87.41 ±0.30 | 89.33 ±0.28 | 89.21 ±0.29 | 89.19 ±0.29 | 89.54 ±0.24 | 88.41 ±0.24 | 88.82 ±0.25 | 87.83 ±0.27 | 89.93 ±0.28 | 89.53 ±0.25 | 89.95 ±0.27 |
| (97.8-97.7) | (96.3-97.6) | (96.9-98.2) | (96.5-96.2) | (95.4-94.7) | (88.5-86.4) | (86.7-88.9) | (87.3-89.1) | (88.4-86.9) | (89.0-91.9) | (88.4-86.9) | (88.4-86.9) | (88.4-86.9) | (87.3-89.1) | (88.3-87.7) | (88.3-88.8) | (88.3-88.8) |
| 97.01 ±0.13 | 97.33 ±0.17 | 97.66 ±0.16 | 96.03 ±0.16 | 94.75 ±0.39 | 93.55 ±0.39 | 88.40 ±0.34 | 87.10 ±0.34 | 89.29 ±0.31 | 89.21 ±0.28 | 89.54 ±0.24 | 88.41 ±0.24 | 88.82 ±0.25 | 87.83 ±0.27 | 89.93 ±0.28 | 89.53 ±0.25 | 89.95 ±0.27 |
| (97.8-97.7) | (96.3-97.6) | (96.9-98.2) | (96.5-96.2) | (95.4-94.7) | (88.5-86.4) | (86.7-88.9) | (87.3-89.1) | (88.4-86.9) | (89.0-91.9) | (88.4-86.9) | (88.4-86.9) | (88.4-86.9) | (87.3-89.1) | (88.3-87.7) | (88.3-88.8) | (88.3-88.8) |
| 97.33 ±0.17 | 96.66 ±0.16 | 96.03 ±0.16 | 94.75 ±0.39 | 93.55 ±0.39 | 88.40 ±0.34 | 87.10 ±0.34 | 89.29 ±0.31 | 89.21 ±0.28 | 89.54 ±0.24 | 88.41 ±0.24 | 88.82 ±0.25 | 87.83 ±0.27 | 89.93 ±0.28 | 89.53 ±0.25 | 89.95 ±0.27 |
| (96.7-97.8) | (96.9-98.2) | (96.5-96.2) | (95.4-94.7) | (88.5-86.4) | (86.7-88.9) | (87.3-89.1) | (88.4-86.9) | (89.0-91.9) | (88.4-86.9) | (88.4-86.9) | (88.4-86.9) | (87.3-89.1) | (88.3-87.7) | (88.3-88.8) | (88.3-88.8) |

F1000Research 2018, 7:521 Last updated: 20 AUG 2021
|   | A  | B  | C  | D  | E  | F  | G  | H  | I  | J  | K  | L  | M  | N  | O  | P  | Q  | R  | S  | T  | U  | V  |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| M | 89.87 ±0.37 | 89.43 ±0.44 | 89.54 ±0.47 | 89.69 ±0.41 | 89.03 ±0.33 | 89.28 ±0.45 | 89.47 ±0.35 | 88.90 ±0.25 | 93.47 ±0.26 | 93.20 ±0.16 | 93.30 ±0.27 | 93.20 ±0.28 | 93.39 ±0.34 | 92.90 ±0.16 | 92.33 ±0.20 | 92.20 ±0.27 | 92.11 ±0.27 | 92.02 ±0.36 | 91.90 ±0.44 | 87.31 ±0.02 | 87.72 ±0.06 |
| N | 89.15 ±0.29 | 89.13 ±0.28 | 89.33 ±0.24 | 88.55 ±0.31 | 90.20 ±0.34 | 90.43 ±0.33 | 88.87 ±0.33 | 91.88 ±0.29 | 92.05 ±0.24 | 91.67 ±0.20 | 92.30 ±0.28 | 98.28 ±0.28 | 93.13 ±0.01 | 90.64 ±0.16 | 90.78 ±0.20 | 91.14 ±0.30 | 86.75 ±0.22 | 93.05 ±0.32 | 88.33 ±0.28 | 87.37 ±0.18 |
| O | 89.11 ±0.30 | 89.09 ±0.36 | 88.96 ±0.38 | 88.48 ±0.42 | 90.00 ±0.29 | 90.78 ±0.20 | 89.17 ±0.29 | 92.99 ±0.24 | 92.90 ±0.22 | 92.24 ±0.13 | 92.39 ±0.20 | 97.90 ±0.14 | 91.95 ±0.19 | 91.77 ±0.20 | 91.42 ±0.22 | 91.74 ±0.20 | 87.80 ±0.28 |
| P | 87.86 ±0.36 | 87.99 ±0.55 | 88.23 ±0.31 | 88.30 ±0.41 | 87.49 ±0.18 | 86.65 ±0.30 | 87.85 ±0.38 | 91.93 ±0.27 | 91.78 ±0.21 | 90.87 ±0.20 | 90.27 ±0.17 | 90.44 ±0.27 | 91.05 ±0.27 | 90.44 ±0.36 | 90.37 ±0.17 | 90.26 ±0.45 | 88.32 ±0.27 |
| Q | 89.64 ±0.35 | 89.43 ±0.36 | 89.53 ±0.39 | 89.28 ±0.25 | 90.60 ±0.28 | 89.35 ±0.27 | 92.03 ±0.26 | 91.63 ±0.28 | 91.93 ±0.34 | 90.30 ±0.28 | 91.17 ±0.28 | 90.94 ±0.26 | 92.01 ±0.25 | 90.99 ±0.29 | 91.21 ±0.23 | 91.10 ±0.30 | 87.80 ±0.27 |
| R | 90.03 ±0.29 | 89.93 ±0.25 | 89.19 ±0.20 | 89.98 ±0.30 | 91.40 ±0.21 | 89.75 ±0.33 | 90.13 ±0.33 | 92.30 ±0.27 | 92.38 ±0.27 | 91.65 ±0.28 | 90.20 ±0.20 | 91.14 ±0.28 | 90.20 ±0.27 | 90.25 ±0.27 | 90.59 ±0.29 | 91.39 ±0.29 | 88.37 ±0.21 |
| S | 93.77 ±0.19 | 93.89 ±0.26 | 93.98 ±0.25 | 93.32 ±0.14 | 93.14 ±0.22 | 88.13 ±0.22 | 88.25 ±0.24 | 87.44 ±0.11 | 89.13 ±0.08 | 89.18 ±0.15 | 89.15 ±0.15 | 89.57 ±0.07 | 89.07 ±0.14 | 89.12 ±0.12 | 89.22 ±0.32 | 88.16 ±0.29 |
| T | 89.85 ±0.15 | 89.95 ±0.27 | 89.66 ±0.18 | 89.95 ±0.21 | 91.19 ±0.17 | 90.61 ±0.20 | 89.80 ±0.16 | 92.46 ±0.17 | 92.37 ±0.17 | 91.26 ±0.19 | 91.90 ±0.20 | 90.90 ±0.17 | 91.42 ±0.19 | 90.00 ±0.16 | 91.89 ±0.14 | 94.18 ±0.29 | 88.76 ±0.06 |
| U | 86.89 ±0.38 | 86.19 ±0.27 | 85.30 ±0.39 | 87.22 ±0.53 | 86.69 ±0.20 | 82.24 ±0.22 | 82.40 ±0.40 | 85.17 ±0.46 | 87.62 ±0.24 | 86.06 ±0.26 | 86.57 ±0.11 | 87.29 ±0.14 | 86.31 ±0.26 | 86.02 ±0.28 | 86.49 ±0.14 | 87.62 ±0.02 | 88.62 ±0.27 |
| V | 87.43 ±0.23 | 87.50 ±0.32 | 87.80 ±0.11 | 87.75 ±0.14 | 87.98 ±0.19 | 88.15 ±0.18 | 87.64 ±0.18 | 88.17 ±0.26 | 87.65 ±0.18 | 87.73 ±0.26 | 87.26 ±0.14 | 87.37 ±0.18 | 88.32 ±0.19 | 87.40 ±0.19 | 88.37 ±0.21 | 86.62 ±0.00 | 86.62 ±0.27 |
| W | 82.06 ±0.30 | 82.22 ±0.33 | 83.03 ±0.16 | 84.03 ±0.25 | 83.69 ±0.34 | 82.55 ±0.20 | 81.56 ±0.31 | 81.73 ±0.23 | 82.36 ±0.13 | 83.02 ±0.05 | 82.41 ±0.15 | 82.68 ±0.19 | 81.54 ±0.34 | 83.16 ±0.01 | 81.12 ±0.32 | 82.76 ±0.01 | 83.13 ±0.00 |
| X | 85.07 ±0.27 | 85.82 ±0.24 | 85.45 ±0.18 | 85.93 ±0.40 | 85.22 ±0.18 | 84.47 ±0.19 | 83.99 ±0.23 | 83.88 ±0.23 | 85.45 ±0.02 | 85.73 ±0.05 | 85.10 ±0.03 | 85.42 ±0.05 | 85.30 ±0.13 | 84.97 ±0.07 | 85.58 ±0.01 | 83.87 ±0.00 | 83.87 ±0.00 |
historically neither has been included in the complex, and there is a quantitative difference in the mean ANI values between genomes of these two species and genomes included in the 22 clades within the complex (last two rows of Table 2). The highest mean ANI for \( E. \) *lignolyticus* and \( E. \) *timonensis* to genomes included in the 22 clades within the complex is 86.2% for \( E. \) *timonensis* to clade S; whereas, the lowest mean ANI within the complex is 86.5% between clades P and U. To further support the decision on what genomes were outliers, we took the 30 outliers, the \( E. \) *lignolyticus* and \( E. \) *timonensis* type strains, the 23 \( E. \) *cloacae* complex type or proxy strains (Table 1), all type strains from other genera closely related to \( Enterobacter \) and generated pairwise ANI values using PanOCT (Supplemental Table 7) to build both UPGMA and Neighbor-Joining trees (Supplemental Figure 2). This analysis supported our decision on what genomes are outliers. One anomaly arose from this analysis: the current type strain genome for \( Lelliottia \) *nimipressuralis* currently in GenBank (ASM187564v1) is the same species as the proposed \( E. \) *roogenkampii* (ASM172980v1) type strain. The type strain 16S sequence (Z96077) for \( Lelliottia \) *nimipressuralis* doesn’t match this purported type strain genome sequence and this genome is an exact duplicate to the previously submitted \( Enterobacter \) sp. FB (ASM80579v1). The duplicate genomes are from the same submitter and the only reasonable conclusion is that this was a submission error for \( Lelliottia \) *nimipressuralis*. This has been reported to NCBI GenBank for resolution (Supplemental File 2). From the all versus all MASH ANI comparison GGRaSP was used to generate average linkage clusters and the medoids of those clusters at both the 95% (species) and 98% (subspecies) levels. If type strains existed at the subspecies level those clusters were used (\( E. \) *hormaechei* and \( E. \) *cloacae*) otherwise species level clusters were used resulting in 22 clades (A-V). If a type strain genome sequence existed for a clade it was selected otherwise the medoid was selected as a proxy. The one exception for this was clade J where two different type strains existed: \( E. \) *asburiae* and \( E. \) *muelleri* where both were retained for the typing. These 23 representative genomes were used to “type” all 1,216 \( Enterobacter \) *cloacae* complex genomes (Supplemental Table 1). For typing the best MASH ANI match was used and resolved to either the species or subspecies level. As expected the typing was in complete agreement with the clades in the MASH ANI tree (Figure 1). The MASH sketches for these 22 clade representatives (after removing the redundant \( E. \) *muelleri*) can be used as a fast categorization tool for novel \( Enterobacter \) *cloacae* complex genomes.

GGRaSP was similarly used to select the 250 most diverse genomes including the outliers from the 1,249 downloaded genomes while eliminating very closely related genomes. PanOCT \(^{9,42}\) run at the nucleotide level was used to generate the orthologous clusters for a pan-genome. The primary use of this was to validate the approximate MASH ANI values. PanOCT determines pairwise ANI values by looking at every orthologous cluster shared by a pair of genomes. The percent identity of each match is weighted by the length of the match, summed over all relevant clusters, and divided by the sum of match lengths which is consistent with previous calculations of ANI. Supplemental Figure 1 shows that the MASH ANI estimate is very strongly correlated (98.9) with the PanOCT ANI measurement. For PanOCT ANI values greater than 94% the estimate is very tight (mean error 0.34±0.22) versus less than 94% (1.15±0.70). The clades and tree at the clade level remained the same using PanOCT ANI values.

The reason we use MASH to estimate ANI is that few other tools such as Genome-to-Genome Distance Calculator (GGDC) \(^{48}\) are efficient enough to compute 1249x1249 pairwise comparisons. To our knowledge GGDC is only available as a web based application with a limit of submitting 75 comparisons at one time. MASH is only an approximation of ANI based on sampling but as we showed for species level comparisons (> 94% ANI) provides a quite accurate estimate. For final determination of novel species boundaries MASH should be supported by an exact ANI calculation as we did using PanOCT which determines ANI based on orthologous matches similar to OrthoANI \(^{44}\). Comparison of MASH and PanOCT ANI to GGDC which has been carefully validated with respect to actual laboratory DDH results increases confidence in our methods. We chose four reasonable size datasets to compare GGDC to PanOCT ANI by generating all versus all comparisons omitting self comparisons: 21 of the most diverse of the 1,216 \( Enterobacter \) *cloacae* complex genomes as determined by MASH and GGRaSP, 10 \( E. \) *hormaechei* genomes chosen similarly, 10 \( E. \) *roogenkampii* genomes chosen similarly, and 10 \( E. \) *asburiae/E. \) *muelleri* genomes chosen similarly. In order to easily compare GGDC to PanOCT ANI we converted PanOCT ANI into a distance measure \( d_{\text{pani}} = 1 - (\text{PanOCT ANI/100}) \). GGDC returns three distance measures: Formula 1: length of all HSPs divided by total genome length, Formula 2: sum of all identities found in HSPs divided by overall HSP length, and Formula 3: sum of all identities found in HSPs divided by total genome length. Total genome length is the sum of the two genomes being compared. Formula 1 is a measure of what percentage of the two genomes are shared in common. Formula 2 is basically one variation of how to calculate ANI. Formula 3 is a combination of formulas 1 and 2. The GGDC recommends Formula 2 for draft genomes since it is affected least by genome completeness. The GGDC then uses some statistical modeling to approximate a predicted laboratory DDH value. Supplemental Figure 3 and Supplemental Table 8 shows that for the combined four datasets \( d_{\text{pani}} \) is practically indistinguishable from GGDC Formula 2.
In the Results section we noted that the type strains for E. asburiae and E. muelleri fall within the same clade which could be separated into subspecies by ANI but we declined to do so. For E. hormaechei we did propose new subspecies but this was because subspecies for E. hormaechei had already been defined. We believe that there must be a cogent reason for delineating beyond the species level. We agree with Chun et al.\(^6\) who state: “At this stage, we do not have sufficient data to provide a general guideline for defining subspecies using genome data. However, a good practice should involve the following criteria: (i) OGRIs between subspecies and other species should be lower than the species-level cutoff value, (ii) OGRIs between subspecies should be higher than the species-level cutoff, (iii) strains belonging to different subspecies should be genomically coherent and form distinguishable clades by OGRIs and phylogenomic treeing, (iv) subspecies should be differentiated by a sufficient number of phenotypes, and (v) there should be a sound rationale why subspecies should be created and separately recognized, such as showing different host specificity in the case of pathogens.”. An
overall genome related index (OGRI) is a computational measure of genome similarity or distance of which ANI is one such. Our ANI analysis possibly fullfill criteria i-iii although given how few strains are in most of the putative subspecies this does not seem robust and criteria iv-v are clearly not met. We only raised the subspecies issue for Enterobacter hormaechei and E. muelleri because often in the past when two competing names exist for a species if the type strains can be separated into clear clades they become subspecies. Since the type strains fall into neither of the major clades for this species and certainly do not cleanly divide the species we did not feel this was appropriate.

Computational analysis supports the reassignment of Enterobacter xiangfangensis to E. hormaechei subsp. xiangfangensis. We propose to name clade D/Hoffman cluster III as E. hormaechei subsp. hoffmannii in honor of Harald Hoffmann’s work elucidating the phylogenetic structure of the E. cloaca subspecies. In particular the subspecies of E. hormaechei. We propose to name clade M/Hoffman cluster IV Enterobacter roggenkampii after Andreas Roggenkamp for his work on elucidating the phylogenetic structure of the E. cloaca complex. The analysis also shows that E. muelleri is a later heterotypic synonym of E. asburiae which should take precedence.

Description of Enterobacter hormaechei subsp. xiangfangensis subsp. nov., comb. nov.

E. hormaechei subsp. xiangfangensis (xiang.fang.en’sis. N.L. gen. m. adj. xiangfangensis pertaining to Xiangfang, a district located in Harbin, Heilongjiang Province, where the bacterium was first isolated).

Basonym: Enterobacter xiangfangensis.

The species description is unchanged from its description as Enterobacter xiangfangensis.

The type strain is strain 10–17 (=LMG 27195=NCIMB 14836=CUCU 62994), isolated from traditional sourdough in Heilongjiang Province, China.

The GenBank accessions for the complete genome sequence of Enterobacter hormaechei subsp. xiangfangensis are PRJNA259658, SAMN05581746, ASM172978v1, and CP017183.1.

Description of Enterobacter hormaechei subsp. hoffmannii subsp. nov.

E. hormaechei subsp. hoffmannii (hoff.mann’i. N.L. gen. m. Hoffmann, in honor of Harald Hoffmann, a German microbiologist who helped elucidate the phylogenetic structure of the E. cloaca complex in particular the subspecies of E. hormaechei).

Hoffmann and Roggenkamp determined clusters within the E. cloaca complex using marker genes, primarily hsp60. Hoffman et al. followed up on three closely grouping clusters to define the three current subspecies of E. hormaechei based on DDH and phenotypic tests. Chavda et al. determined groups for the E. cloaca complex using SNPs from whole genome alignments. ANI analysis showed that the Chavda groups were highly similar at levels associated with species or subspecies groupings. This paper performs a more detailed analysis of gene content and ANI across a larger set of genomes supporting the Chavda groups A-E as E. hormaechei subspecies. E. hormaechei subsp. hoffmannii subsp. nov. has similar gene content and ANI characteristics as the previously defined four subspecies.

Hoffmann deposited the type strain, EN-114, for Enterobacter hormaechei subsp. hoffmannii in Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, accession DSM-14563, and recently the strain was also deposited in BCCM/LMG Bacteria Collection, accession LMG-30171. The GenBank accessions for the complete genome sequence are PRJNA259658, SAMN05581748, ASM172974v1, CP017186.1, and CP017187.1.

According to 2, the strain was isolated from the respiratory tract of a clinical patient. The DSMZ database indicates that the sample was isolated prior to 2002 in Bavaria, Germany.

Description of Enterobacter roggenkampii sp. nov.

E. roggenkampii (rog.gen.kamp’i. N.L. gen. m. Roggenkamp, in honor of Andreas Roggenkamp, a German microbiologist who helped elucidate the phylogenetic structure of the E. cloaca complex).

Hoffmann and Roggenkamp determined clusters within the E. cloaca complex using marker genes, primarily hsp60. Chavda et al. determined groups for the E. cloaca complex using SNPs from whole genome alignments. ANI analysis showed that the Chavda groups were highly similar at levels associated with species or subspecies groupings. Enterobacter roggenkampii sp. nov. is the type strain for Hoffmann cluster IV and Chavda group M. This paper performs a more detailed analysis of gene content and ANI across a larger set of genomes supporting the Chavda groups A-R and adding S-V. E. roggenkampii sp. nov. has similar gene content and ANI characteristics as previously defined species in the E. cloaca complex.

Hoffmann deposited the type strain, EN-117, for Enterobacter roggenkampii in Leibniz-Institut DSMZ-Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH, accession DSM-16690, and recently the strain was also deposited in BCCM/LMG Bacteria Collection, accession LMG-30172. The GenBank accessions for the complete genome sequence are PRJNA259658, SAMN05581750, ASM172980v1, CP017184.1, and CP017185.1.

According to 2, the strain was isolated from the stool of a clinical patient. The DSMZ database indicates that the sample was isolated in 2000 in Germany.

The GenBank accessions for the complete genome sequence of Enterobacter hormaechei subsp. oharae are PRJNA259658, SAMN05581749, ASM172970v1, and CP017180.1.
Data availability
All data underlying the results are available as part of the article and no additional source data are required.

Competing interests
No competing interests were disclosed.

Grant information
This work has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under award number U19AI110819.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
We would like to thank: Jason Inman from JCVI for help with pan-genome runs; Karen Beeri, Karrie Goglin, and Kelly Colt from the JCVI sequencing core for growth and sequencing of the type strains; and Elke Lang and Claudine Vereecke for help getting the type strains into the BCCM/LMG Bacteria Collection.

Supplementary material
Supplemental Table 1. ANI clades compared to MASH best match assignment for 1,216 Enterobacter cloacae complex genomes.
Click here to access the data.

Supplemental Table 2. MASH typing of 30 outlier genomes falling outside of the Enterobacter cloacae complex but labelled as Enterobacter in RefSeq.
Click here to access the data.

Supplemental Table 3. PanOCT generated orthologous clusters for 1,216 Enterobacter cloacae complex genomes. Rows are clusters, columns are genomes, cells contain RefSeq gene identifiers.
Click here to access the data.

Supplemental Table 4. Gene counts for genes common to all genomes, specific to a clade, or missing from a clade.
Click here to access the data.

Supplemental Table 5. Pairwise MASH Average Nucleotide Identity (ANI) values for 1,249 genomes labelled Enterobacter in RefSeq.
Click here to access the data.

Supplemental Table 6. Distribution of gat, aga, and rbi/dal operons across E cloacae complex clades.
Click here to access the data.

Supplemental Table 7. PanOCT Average Nucleotide Identity (ANI) pairwise values for 30 outlier genomes and related type species genomes.
Click here to access the data.

Supplemental Table 8. Genome distance measure values for representative genomes using GGCD, PanOCT ANI, and MASH.
Click here to access the data.

Supplemental Figure 1. Graph of MASH estimated versus PanOCT calculated Average Nucleotide Identity (ANI) for 250 representative genomes.
Click here to access the data.

Supplemental Figure 2. Phylogenetic trees of 30 outlier genomes and related type species genomes.
Click here to access the data.
Supplemental Figure 3. Graphs of PanOCT ANI, MASH, and GGCD genome distance measures.

Click here to access the data.

Supplemental File 1. Newick formatted tree generated from Supplemental Table 5 and used to generate Figure 1.

Click here to access the data.

Supplemental File 2. Details about mistaken Lelliottia ninipressuralis type strain genome.

Click here to access the data.

References

1. O'Hara CM, Steigerwalt AG, Hill BC, et al.: Enterobacter hormaechei, a new species of the family Enterobacteriaceae formerly known as enteric group 75. J Clin Microbiol. 1989; 27(9): 2046–9. Published Abstract | Free Full Text

2. Hoffmann H, Roggenkamp A: Population genetics of the nomenclature Enterobacter cloaceae. Appl Environ Microbiol. 2003; 69(9): 5306–18. Published Abstract | Publisher Full Text | Free Full Text

3. Hoffmann H, Strid S, Ludwig W, et al.: Enterobacter hormaechei subsp. oharae subspp. nov., E. hormaechei subspp. hormaechei comb. nov., and E. hormaechei subspp. steigerwaldii subspp. nov., three new subspecies of clinical importance. J Clin Microbiol. 2000; 38(7): 2397–303. Published Abstract | Publisher Full Text | Free Full Text

4. Oren A, Garrity GM: List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2016; 66(11): 4299–305. Published Abstract | Publisher Full Text | Free Full Text

5. Hu CT, Li CY, Yang LJ, et al.: Enterobacter hormaechei, a new species of the family Enterobacteriaceae. Int J Syst Evol Microbiol. 2014; 64(Pt 8): 2491–7. Published Abstract | Publisher Full Text | Free Full Text

6. Tindal BJ, Sutton G, Garrity GM: Enterobacter aerogenes Hormaeche and Edwards 1960 (Approved Lists 1980) and Klebsiella mobilis Bascomb et al. 1971 (Approved Lists 1980) share the same nomenclatural type (ATCC 13048) on the Approved Lists and are homotypic synonyms, with consequences for the name Klebsiella mobilis Bascomb et al. 1971 (Approved Lists 1980). Int J Syst Evol Microbiol. 2017; 67(2): 502–504. Published Abstract | Publisher Full Text

7. Chavda KD, Chen L, Fouts DE, et al.: Comprehensive Genome Analysis of Carbapenemase-Producing Enterobacter spp.: New Insights into Phylogeny, Population Structure, and Resistance Mechanisms. mBio. 2016; 7(6): e02093–16. Published Abstract | Publisher Full Text | Free Full Text

8. Stanley JT: The bacterial species dilemma and the genomic-phylogenetic species concept. Philos Trans R Soc Lond B Biol Sci. 2000; 361(1475): 1899–909. Published Abstract | Publisher Full Text

9. Georgiades K, Raoult D: Defining pathogenic bacterial species in the genomic era. Front Microbiol. 2011; 1: 151. Published Abstract | Publisher Full Text | Free Full Text

10. Bartholomew JW, Mittwer T: The Gram stain. Bacterial Rev. 1952; 16(1): 1–29. Published Abstract | Free Full Text

11. Kim M, Oh HS, Park SC, et al.: Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. Int J Syst Evol Microbiol. 2014; 64(Pt 2): 346–51. Published Abstract | Publisher Full Text | Free Full Text

12. Konstantinidis KT, Ramette A, Tiedje JM: The bacterial species definition in the genomic era. Philos Trans R Soc Lond B Biol Sci. 2006; 361(1475): 1929–40. Published Abstract | Publisher Full Text | Free Full Text

13. Chan JZ, Halachev MR, Loman NJ, et al.: Defining bacterial species in the genomic era: insights from the genus Acinetobacter. BMC Microbiol. 2012; 12: 302. Published Abstract | Publisher Full Text | Free Full Text

14. Gons J, Konstantinidis KT, Klappenbach JA, et al.: DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007; 57(Pt 1): 81–91. Published Abstract | Publisher Full Text

15. Richter M, Rosselló-Móra R: Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009; 106(45): 19126–31. Published Abstract | Publisher Full Text | Free Full Text

16. Zhang W, Du P, Zheng H, et al.: Whole-genome sequence comparison as a method for improving bacterial species definition. J Gen Appl Microbiol. 2014; 60(2): 75–8. Published Abstract | Publisher Full Text

17. Varghese NJ, Mukherjee S, Ivanova N, et al.: Microbial species delineation using whole genome sequences. Nucleic Acids Res. 2015; 43(14): 6761–71. Published Abstract | Publisher Full Text | Free Full Text

18. Abreu-Klopfh JR, Auff AF, Klerk HP, et al.: Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013; 14: 60. Published Abstract | Publisher Full Text | Free Full Text

19. Colston SM, Fulmer MK, Beka L, et al.: Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using Aeromonas as a test case. mBio. 2014; 5(6): e02136. Published Abstract | Publisher Full Text | Free Full Text

20. Davin-Regli A, Bost C, Charrel R, et al.: A nosocomial outbreak due to Enterobacter cloacae strains with the E. hormaechei genotype in patients treated with fluoroquinolones. J Clin Microbiol. 1997; 35(4): 1008–10. Published Abstract | Publisher Full Text | Free Full Text

21. Paauw A, Caspers MP, Leverstein-van Hall MA, et al.: Identification of resistance and virulence factors in an epidemic Enterobacter hormaechei outbreak strain. Microbiology. 2009; 155(Pt 5): 1478–88. Published Abstract | Publisher Full Text | Free Full Text

22. Campos LC, Lobianco LF, Seki LM, et al.: Outbreak of Enterobacter hormaechei septicaemia in newborns caused by contaminated parenteral nutrition in Brazil. J Hosp Infect. 2007; 66(1): 55–7. Published Abstract | Publisher Full Text

23. Wengen PN, Tokars JI, Brennan P, et al.: An outbreak of Enterobacter hormaechei infection and colonization in an intensive care nursery. Clin Infect Dis. 1997; 24(6): 1243–4. Published Abstract | Publisher Full Text

24. Ohad S, Block C, Kravitz V, et al.: Rapid identification of Enterobacter hormaechei and Enterobacter cloacae genetic cluster III. J Appl Microbiol. 2014; 116(3): 1315–21. Published Abstract | Publisher Full Text

25. Guerin F, Isnard C, Sinel C, et al.: Cluster-dependent colistin hetero-resistance in Enterobacter cloacae complex. J Antimicrob Chemother. 2016; 71(11): 3058–61. Published Abstract | Publisher Full Text

26. Morand PC, Billotet A, Rotman M, et al.: Specific distribution within the Enterobacter cloacae complex of strains isolated from infected orthopedic implants. J Clin Microbiol. 2009; 47(8): 2489–95. Published Abstract | Publisher Full Text | Free Full Text

27. Mezzatesta ML, Gona F, Stefanis F: Enterobacter cloacae complex: clinical impact and emerging antibiotic resistance. Future Microbiol. 2012; 7(7): 887–902. Published Abstract | Publisher Full Text

28. Paauw A, Caspers MP, Schuren FH, et al.: Genomic diversity within the Enterobacter cloacae complex. PLoS One. 2008; 3(8): e3018. Published Abstract | Publisher Full Text | Free Full Text

29. Brady C, Cleaveonek I, Venter S, et al.: Taxonomic evaluation of the genus Enterobacter based on multilocus sequence analysis (MLSA): proposal to reclassify E. ninipressuralis and E. amnigenus into Lelliottia gen. nov. as Lelliottia ninipressuralis comb. nov. and Lelliottia amnigena comb. nov., respectively, E. gergoviae and E. pyrus into Pluralibacter gen. nov. as Pluralibacter gergoviae comb. nov. and Pluralibacter pyrus comb. nov., respectively, E. cowanii, E. radicincitans, E. oryzae, E. radicicinicans, E. oryzaceae, E. arachidis into Kosakonia gen. nov. as Kosakonia cowanii comb. nov., Kosakonia radicicinicans comb. nov., Kosakonia oryzaceae comb. nov. and Kosakonia arachidis comb. nov., respectively, and E. turicensis, E. helveticus and E. pulvisere into Cronobacter as Cronobacter zurichensis nov. gen., 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): 393–9.

30. Stephan R, Grimb CJ, Goginad GR, et al.: Re-examination of the taxonomic status of Enterobacter helveticus, Enterobacter pulveris and Enterobacter turicensis as members of the genus Cronobacter and their reclassification in the genera Francobacteriaceae, nov. and Siccibacteriaceae, nov. as Francobacter helveticus comb. nov., Francobacter pulveris comb. nov. and Siccibacter turicensis comb. nov., respectively. Int J Syst Evol Microbiol. 2014; 64(10): 3402–10.

31. Ondov BD, Treangen TJ, Melsted P, et al.: Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016; 17(1): 132.

32. Kämpfer P, McInroy JA, Glaeser SP: Enterobacter muelleri sp. nov., isolated from the rhizosphere of Zea mays. Int J Syst Evol Microbiol. 2015; 65(11): 4093–9.

33. Brenner DJ, McWhorter AC, Kai A, et al.: Enterobacter asburiae sp. nov., a new species found in clinical specimens, and reassignment of Erwinia dissolvens and Erwinia nimipressuralis to the genus Enterobacter as Enterobacter dissolvens comb. nov. and Enterobacter nimipressuralis comb. nov. J Clin Microbiol. 1986; 23(6): 1114–20.

34. Chun J, Oren A, Ventosa A, et al.: Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2018; 68(1): 461–466.

35. Fouts DE, Brinkac L, Beck E, et al.: PanOCT: automated clustering of orthologs using conserved gene neighborhood for pan-genomic analysis of bacterial strains and closely related species. Nucleic Acids Res. 2012; 40(22): e172.

36. Nobelmann B, Lengeler JW: Molecular analysis of the gal genes from Escherichia coli and of their roles in galactitol transport and metabolism. J Bacteriol. 1996; 178(23): 6790–5.
From our point of view, digital DDH (dDDH) calculated by GGDC tool correlates better with empirical DDH than the various implementations of ANI (Meier-Kolthoff 2014\textsuperscript{1}, Meier-Kolthoff et al., 2013\textsuperscript{2}) and thus dDDH can safely be preferred over ANI. GGDC estimates at 70% dDDH are widely accepted, have repeatedly proven to be accurate, and is recommended for the discrimination of the species by bacterial taxonomical experts (and not the preliminary ANI derived from GGDC algorithm) (Chun et al., 2018\textsuperscript{3}).

In bacterial taxonomy, only type strains are valid, while strains belong to phylogenomically diverse clades are less important during defining a species. As there are limitations in the use of the GGDC tool to compare a large number of strains, it would have been more logical to estimate GGDC-dDDH values (only) for the type strains for which the ANI cutoff is ambiguous.

The GGDC tool shows \textit{E. xiangfangensis} and \textit{E. hormaechei} are in fact two different species (DDH estimate (GLM-based): 59.80% [56.9 - 62.5%]). Indeed, there is only a 12% probability that these two type strain can be classified as subspecies of a either species. Extension of this estimate to clade level indicates clade A-D and E are two different species and represents \textit{E. xiangfangensis} and \textit{E. hormaechei}, respectively. Just because several strains of clade A-D were historically identified as \textit{E. hormaechei} subspecies, there is no reason to ignore the recently validated \textit{E. xiangfangensis} species.

Currently, phenotypic identification remains the gold standard for identification of microorganisms in standard diagnostic laboratories and provides the bulk of the data for taxonomic classification. WGS is presently largely done in research laboratories or as pilot endeavors in specialized diagnostic laboratories. As genotype-phenotype correlations are at present incomplete, current classification schemes would give phenotypic data priority.
References
1. Meier-Kolthoff JP, Klenk HP, Göker M: Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol*. 2014; 64 (Pt 2): 352-6 PubMed Abstract | Publisher Full Text
2. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M: Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. 2013; 14: 60 PubMed Abstract | Publisher Full Text
3. Chun J, Oren A, Ventosa A, Christensen H, et al.: Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol*. 2018; 68 (1): 461-466 PubMed Abstract | Publisher Full Text

Competing Interests: No competing interests were disclosed.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 17 Jul 2018

Granger Sutton, J Craig Venter Institute, Rockville, USA

We agree that dDDH (Refs 1,2) has been shown to correlate slightly better with laboratory DDH than ANI does. We do not agree that this should provide the basis for preferring dDDH over ANI as the basis for determining species or subspecies level relatedness. The proposed criteria from Chun et al (Ref 3) also do not give a preference for dDDH over ANI as an OGRI. The question we need to ask is what defines a species? Genome relatedness is certainly a primary component of that and because laboratory DDH was the first method for calculating genome relatedness it became the gold standard but with current genome sequencing there is no reason for it to remain the gold standard. dDDH combines shared genomic content with the ANI of the shared genomic content. PanOCT computes multiple pairwise relatedness measures: two of which are the ANI of orthologous genes and the Jaccard similarity of the gene content. We have shown (Ref 4) that the gene content similarity measure can be significantly affected by horizontally transferred genes such as plasmids which raises the question of whether that should be part of a species relatedness measure. Chun et al (Ref 3) argue that at levels above species and certainly above genera that OGRI measures are not useful and rather that a set of core genes with low horizontal transfer potential be used for phylogenetic tree construction. This is much more consistent with ANI which tends to measure the core gene similarity rather than dDDH which includes variable gene content. We believe that evolutionary relatedness including species definitions is best measured with ANI while gene content provides a somewhat orthogonal measurement to capture horizontal transfer events. We recognize that horizontal transfers are also evolutionary events and strongly correlated with ANI hence the "somewhat orthogonal". We welcome the discussion of what should define a species and understand that the views of Drs. Chakraborty and Doijad are as valid as our own.

Using DDH Hoffman et al (Ref 5) showed that *Enterobacter hormaechei* subsp. *oharae*, *E. hormaechei* subsp. *hormaechei*, and *E. hormaechei* subsp. *steigerwaltii* are the same species: "The close DNA-DNA relatedness within clusters VI and VII was reflected by ΔTm values below 0.5. The relatively higher heterogeneity of cluster VIII was indicated by higher within-
group ΔTm values of up to 2.7. By evaluating the DNA relatedness among the clusters, we found that clusters VI and VIII are closely related (mean ΔTm value = 2.2), while a relatively longer distance for E. hormaechei cluster VII from the members of clusters VI and VIII was indicated by the mean ΔTm value of 4.0. However, all three genetic clusters could still be assigned to the same species (14). They could be genetically distinguished from the other species of the E. cloacae complex, which had ΔTm values of 5.6 to 10.3 (Table 2)."

Unfortunately they did not report DNA-DNA relatedness values but only ΔTm values. They did cite previous work which gave DNA-DNA relatedness: "Davin-Regli et al. (4) reported an outbreak with an “E. cloacae" strain with the E. hormaechei genotype” but an aberrant biotype. The strain exhibited all of the characteristics of E. hormaechei and was 80% related to the type strain in DNA-DNA reassociation experiments but was positive for growth on D-sorbitol and α-D-melibiose. Obviously, this outbreak was caused by a strain of genetic cluster VI. Hence, these studies are in agreement with our observation that genetic clusters VI and VIII belong to the species E. hormaechei (4, 6).". We agree that by ANI and dDDH that E. hormaechei subsp. hormaechei is borderline at best to be grouped as the same species as the other E. hormaechei subspecies but Drs. Chakraborty and Doijad cannot have it both ways. Hoffman et al showed phenotypic data supporting there grouping of the subspecies and delineation from other subspecies as well as genotypic support using marker genes which has since been used in clinical papers to differentiate the subspecies from each other and other species. Certainly one could propose making these separate species but the bar for undoing historical precedent is much higher than arguing that the ANI or dDDH values are borderline.

Drs. Chakraborty and Doijad state: "Currently, phenotypic identification remains the gold standard for identification of microorganisms in standard diagnostic laboratories and provides the bulk of the data for taxonomic classification. WGS is presently largely done in research laboratories or as pilot endeavors in specialized diagnostic laboratories. As genotype-phenotype correlations are at present incomplete, current classification schemes would give phenotypic data priority.". Not being clinicians we are not sure if this is true but based on our reading of the literature if it is true it is likely to not be true in the near future. We are not against phenotypic characterization if it is economical and reliable. We look forward to a robust discussion of the pros and cons of phenotypic versus genotypic diagnostic methods. Regardless, assignment of species and species delineation has long been genotype based since DDH is a genotypic measure as well as marker genes and OGRI. We are not against phenotypic characterization of type strains although one could argue that this only really makes sense if a clade of strains of the same species is characterized to evaluate variability. We reached out to DSMZ to inquire about phenotypic characterization services which they are willing to provide at some level on a case by case basis but they could not tell us what minimal characterization is necessary for a type strain. Perhaps Drs. Chakraborty and Doijad could intervene on our behalf with DSMZ and have the appropriate characterization performed and placed in the DSMZ supported “The Bacterial Diversity Metadatabase” (BacDive). This could be the first step towards some form of phenotypic characterization standard for type strains.

1. Meier-Kolthoff JP, Klenk HP, Göker M: Taxonomic use of DNA G+C content and DNA-DNA hybridization in the genomic age. *Int J Syst Evol Microbiol.* 2014; **64** (Pt 2): 352-6 PubMed Abstract | Publisher Full Text
2. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M: Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*. 2013; 14: 60 PubMed Abstract | Publisher Full Text

3. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR, da Costa MS, Rooney AP, Yi H, Xu XW, De Meyer S, Trujillo ME: Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol*. 2018; 68 (1): 461-466 PubMed Abstract | Publisher Full Text

4. Chavda KD, Chen L, Fouts DE, et al.: Comprehensive Genome Analysis of Carbapenemase-Producing *Enterobacter* spp.: New Insights into Phylogeny, Population Structure, and Resistance Mechanisms. *mBio*. 2016; 7(6): pii: e02093-16. PubMed Abstract | Publisher Full Text | Free Full Text

5. Hoffmann H, Stindl S, Ludwig W, et al.: *Enterobacter hormaechei* subsp. *oharae* subsp. nov., *E. hormaechei* subsp. *hormaechei* comb. nov., and *E. hormaechei* subsp. *steigerwaltii* subsp. nov., three new subspecies of clinical importance. *J Clin Microbiol*. 2005; 43(7): 3297–303. PubMed Abstract | Publisher Full Text | Free Full Text

**Competing Interests:** No competing interests were disclosed.

---

Reviewer Report 12 July 2018

https://doi.org/10.5256/f1000research.16861.r35632

© 2018 Pallen M. This is an open access peer review report distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mark J. Pallen

Quadram Institute, Norwich, UK

I am happy with the changes.

**Competing Interests:** No competing interests were disclosed.

**I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

---

Version 1

Reviewer Report 07 June 2018

https://doi.org/10.5256/f1000research.15853.r34159
Considerable genome data is now available for isolates of many members of the family Enterobacteriaceae. As we move away from well-defined species such as E. coli and Salmonella, taxonomic assignments become blurred and there is now a great need to develop standardized tools for proper classification. A particular case is that of the species Enterobacter, where only 12 of the 35 historically classified species in this genus are valid.

The present manuscript reevaluates taxonomic allocation of members of the Enterobacter cloacae complex using whole genome sequences (WGS). It is important to remember that the dataset comprises primarily of draft genome sequences of varying quality and with only a very small number representing truly closed genomes.

Isolates of the E. hormaechei complex are often associated with clinical disease. Based on the data from this study there are now two novel subspecies of E. hormaechei designated as E. hormaechei subsp. hoffmannii and E. hormaechei subsp. xiangfangensis respectively. In addition, a new species E. roggenkampii is proposed. Overall the study predicts the existence of 7 additional species within the genus Enterobacter.

The bulk of the analysis is based on a single tool viz. MASH-based ANI and is supplemented by the panOCT tool developed by the authors. The authors should consider the use of additional software tools to determine the overall genome-related index (OGRI).

Specific comments:
- Clade A-E represent the five subspecies of E. hormaechei. The average nucleotide identity (ANI) for the clades A-D and E are at the borderline ANI-species definition.
- In view of the fact that data is based mainly on draft genomes, the utility of supportive assignments based on the total numbers of unique genes must be considered carefully.
- For such closely related clades, multi-tool-based analysis of taxonomy are helpful to reassure the claims. To support the species/subspecies distinction, particularly for those closely related clades, the use of widely used taxonomic tools such as the digital DNA-DNA hybridization tool, GGDC should be employed to strengthen the claims.
- ANI values can vary when using different calculation tools as for e.g. with JSpecies and ANI calculator. The use of MASH algorithm leads to minor variation in ANI values and makes the borderline species definitions presented here difficult to interpret.
- To confirm separation of E. timonensis and E. lignolyticus from the genus Enterobacter,
comparison with members of the closest genera (for e.g., *Klebsiella, Citrobacter* etc.) should be added.

Finally, biochemical and fermentation characteristics are key indicators for phenotypic characterization of isolates in diagnostic laboratories.

The final paragraph on biochemical properties is inadequate and could lead to confusion of phenotypes and undo the very purpose of the proposed classification scheme. Thus the *gat* operon is not exclusive to *E. hormaechei* subspecies *hormaechei* as stated, but is also present for e.g. in type strain *E. bugandensis* EB-247T.

**Is the work clearly and accurately presented and does it cite the current literature?**
Yes

**Is the study design appropriate and is the work technically sound?**
Partly

**Are sufficient details of methods and analysis provided to allow replication by others?**
Yes

**If applicable, is the statistical analysis and its interpretation appropriate?**
Not applicable

**Are all the source data underlying the results available to ensure full reproducibility?**
Yes

**Are the conclusions drawn adequately supported by the results?**
Partly

**Competing Interests:** No competing interests were disclosed.

We confirm that we have read this submission and believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Author Response 26 Jun 2018

**Granger Sutton**, J Craig Venter Institute, Rockville, USA

We thank Dr. Pallen for the thoughtful review and respond to issues below.

“I don't see the need for the separate Introduction and Background sections. According to the guidelines for authors, papers in this journal should follow the usual IMRAD format, so I think that the two sections should simply become sub-sections of the Introduction, perhaps with brief explanatory headers.”

We removed the Background and Conclusion section headings to conform to the IMRAD format.

“I am not sure why the authors abdicate responsibility for determining whether "8 subclades of *E. asburiae* should be treated as subspecies". Why not roll their approach out
to cover these lineages too?”

We now address this in the Discussion section.

“The authors discuss the concept of "placeholder" species and subspecies in the Discussion, but fail to mention the "Candidatus" designation, which is recognised by the current bacterial taxonomy apparatchiks:

http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-45-1-186
https://en.wikipedia.org/wiki/Candidatus

They should include some discussion of this designation that includes a recognition of its major shortcoming in requiring phenotypic data in addition to genome sequence.”

We thank Dr. Pallen for pointing this out to us and have included this in the Discussion section.

We thank Dr. Chakraborty for the thoughtful review and respond to issues below.

“The bulk of the analysis is based on a single tool viz. MASH-based ANI and is supplemented by the panOCT tool developed by the authors. The authors should consider the use of additional software tools to determine the overall genome-related index (OGRI).” and “For such closely related clades, multi-tool-based analysis of taxonomy are helpful to reassure the claims. To support the species/subspecies distinction, particularly for those closely related clades, the use of widely used taxonomic tools such as the digital DNA-DNA hybridization tool, GGDC should be employed to strengthen the claims.”

We have included the comparison of GGDC to MASH and PanOCT ANI in the Methods section.

“Clade A-E represent the five subspecies of E. hormaechei. The average nucleotide identity (ANI) for the clades A-D and E are at the borderline ANI-species definition.”

This is certainly true but is also true of the already existing E. hormaechei subspecies: clade B E. hormaechei ssp. steigerwaltii, clade C E. hormaechei ssp. oharae, and clade E E. hormaechei ssp. hormaechei. While in the absence of previous taxonomic assignments one might choose to be reluctant to combine clades B, C, and E into a single species based on ANI because they have already been grouped as a species the borderline ANI values are not strong enough to argue for changing this. Given this adding clades A and D to E. hormaechei is strongly confirmed by the ANI values between clades A, B, C, and D.

“In view of the fact that data is based mainly on draft genomes, the utility of supportive assignments based on the total numbers of unique genes must be considered carefully.”

We have noted this concern in the results section. Gene content is not a primary consideration in our proposed new species designation but rather a possible reason to delineate at the subspecies level. In our experience most recent draft genome sequences are of high quality and the RefSeq genomes we used are screened by NCBI to meet certain quality requirements. Draft genome breaks tend to be at and due to repetitive elements such as transposons which would not affect the representation of most genes. We also try to take this into account by using a 90% rather than a 100% threshold.

“ANI values can vary when using different calculation tools as for e.g. with JSpecies and ANI calculator. The use of MASH algorithm leads to minor variation in ANI values and makes the borderline species definitions presented here difficult to interpret.”

ANI values for the newly proposed type strains were backed up by PanOCT ANI and now by GGDC and are not borderline except as consistent with previous taxonomy.

“To confirm separation of E. timonensis and E. lignolyticus from the genus Enterobacter, comparison with members of the closest genera (for e.g., Klebsiella, Citrobacter etc.) should
We have added this analysis to the Methods section. “Finally, biochemical and fermentation characteristics are key indicators for phenotypic characterization of isolates in diagnostic laboratories.” As the paper mentions we are not opposed to the biochemical characterization of type strains but need a standard that can be implemented by culture collections so that computationalists can acquire this data. The DSMZ for instance supports doing some of this characterization but does not claim it to be standard. In addition, DSMZ supports storing this characterization data in “The Bacterial Diversity Metadatabase” (BacDive) such as for the *E. bugandensis* type strain ([https://bacdive.dsmz.de/strain/132404](https://bacdive.dsmz.de/strain/132404)). What is interesting is that most biochemical characterization is not used to define a species in current practice. Researchers no longer collect phenotypic features and cluster based on a feature vector. Rather, genotypic characteristics are captured such as 16S or hsp60 or rpoB or WGS which are used to define a cluster of strains and then phenotypic characterization of those strains is performed and used as part of the species definition no matter how divergent those features may be. Computational taxonomy provides a structure by which strains can be clustered, named, referenced, discussed and compared to related clades. Biologists should follow up on clinically or otherwise interesting clades. We are not sure whether Dr. Chakraborty is arguing for historical consistency in what characterization is minimally required for a type strain or is arguing that there is little or no value in computational taxonomy without phenotypic characterization because it is required for clinical diagnosis. We would disagree with both since with the advent of whole genome sequences (or even DDH) phenotype is not needed to define species and clinical diagnosis can be done with molecular markers.

“The final paragraph on biochemical properties is inadequate and could lead to confusion of phenotypes and undo the very purpose of the proposed classification scheme. Thus the gat operon is not exclusive to *E. hormaechei* subspecies *hormaechei* as stated, but is also present for e.g. in type strain *E. bugandensis* EB-247T.”

We apologize for being unclear. We were summarizing what is already in the literature for distinguishing *E. hormaechei* subspecies from each other. We have been more precise and clarified this issue in the Results section.

**Competing Interests:** No competing interests were disclosed.
Mark J. Pallen

Quadram Institute, Norwich, UK

This is in general a well written and well argued paper that represents a valuable addition to attempts to bring bacterial taxonomy into the genomic age. I can find no fault with the methodologies used nor with the general interpretation of results. I agree with the authors that all future bacterial taxonomy and nomenclature should be based on genomic data and they have fallen in line with an emerging consensus of how to make that work using ANI.

It is clear that bacterial taxonomy is broken and needs fixing and the only suitable response to the tyranny of The International Committee on Systematic Bacteriology is subversion by publishing papers like this that ignore its ridiculous and outdated requirements.

To quote Darwin: "Our classifications will come to be, as far as they can be so made, genealogies".

I have just a handful of minor criticisms/suggestions for improvement:

1. I don't see the need for the separate Introduction and Background sections. According to the guidelines for authors, papers in this journal should follow the usual IMRAD format, so I think that the two sections should simply become sub-sections of the Introduction, perhaps with brief explanatory headers.

2. I am not sure why the authors abdicate responsibility for determining whether "8 subclades of E. asburiae should be treated as subspecies". Why not roll their approach out to cover these lineages too?

3. The authors discuss the concept of "placeholder" species and subspecies in the Discussion, but fail to mention the "Candidatus" designation, which is recognised by the current bacterial taxonomy apparatchiks:

   http://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/00207713-45-1-186

   https://en.wikipedia.org/wiki/Candidatus

   They should include some discussion of this designation that includes a recognition of its major shortcoming in requiring phenotypic data in addition to genome sequence.

Is the work clearly and accurately presented and does it cite the current literature?  
Yes

Is the study design appropriate and is the work technically sound?  
Yes

Are sufficient details of methods and analysis provided to allow replication by others?  
Yes

If applicable, is the statistical analysis and its interpretation appropriate?  
Yes
Are all the source data underlying the results available to ensure full reproducibility?
Yes

Are the conclusions drawn adequately supported by the results?
Yes

**Competing Interests:** No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 17 May 2018

**Granger Sutton, J Craig Venter Institute, Rockville, USA**

The three reviewer suggestions all have merit and we will try to address them once more reviews are received. 1) We can certainly conform to IMRaD by using subheadings. 2) The issue of species versus subspecies should be addressed in the discussion. Our feeling was that when it is already problematic to validly publish names for species it is even more burdensome to do so for subspecies. What is the appropriate criteria to go to the trouble to differentiate subspecies: clinical significance, number of exemplars of each subspecies, and/or amount of core gene content difference between subspecies (this can only be determined once there are enough exemplars of each subspecies)? 3) We were unaware of the Candidatus designation and appreciate this being pointed out. While it does not appear to be a good fit for the case where genome sequences exist and species/subspecies are determined computationally since it was designed for environmental or unculturable samples with limited sequence data but at least some phenotypic or morphological data, it does suggest that some similar designation be used for "placeholder" names. We do not want to assign potentially permanent names with a notation indicating they are provisional but would like the name itself to indicate it is provisional and to be replaced when someone does the hard work of depositing a type strain and any required minimal phenotypic information. Again we should address this in the discussion.

**Competing Interests:** No competing interests were disclosed.

---

**Comments on this article**

**Version 2**

Author Response 31 Jan 2019

**Granger Sutton, J Craig Venter Institute, Rockville, USA**

There have been three recent papers from the same group on four new Enterobacter species - three of which are in the NCBI taxonomy:
Enterobacter huaxiensis and Enterobacter chuandaensis https://protect-us.mimecast.com/s/gvM8CL9no8s8XxuqHnU5?domain=ijs.microbiologyresearch.org

Enterobacter chengduensis https://www.ncbi.nlm.nih.gov/pubmed/30302649

Enterobacter sichuanensis https://ijs.microbiologyresearch.org/content/journal/ijsem/10.1099/ijsem.0.003089#tab5

The most recent of these papers references this paper but makes no real attempt to use what is in it. As pointed out in the paper we gave temporary names to species clades using NCBI's preferred genomospecies format and these temporary names should be updated. Below are the four type strains from the papers plus two strains from the NCBI taxonomy.

From BioProject PRJNA415108:

| GenBank ID          | Sample ID       | SAMN ID       | WCHEN ID       |
|---------------------|-----------------|---------------|----------------|
| GCA_003944645.1     | RWHU00000000    | SAMN10525001  | WCHEHu045002   |
| Enterobacter mori   | 90.82           | Enterobacter huaxiensis (NCBI taxonomy) |
| GCA_003594935.1     | QZCT00000000    | SAMN09845186  | WCHEN090008    |
| Enterobacter mori   | 90.69           | Enterobacter sp. WCHEN090008 (huaxiensis type strain from paper) |
| GCA_003594915.1     | QZCS00000000    | SAMN09845205  | WCHEN090028    |
| Enterobacter genomosp. T | 98.26        | Enterobacter sp. WCHEN090028 (chuandaensis type strain from paper) |
| GCA_003944655.1     | RWHT00000000    | SAMN10525011  | WCHECh090071  |
| Enterobacter genomosp. L | 95.69        | Enterobacter chengduensis (NCBI taxonomy) |
| GCA_002939185.1     | POVL00000000    | SAMN08357870  | WCHECI1597    |
| Enterobacter genomosp. N | 98.56        | Enterobacter sp. WCHECI1597 (sichuanensis type strain from paper) |

From BioProject PRJNA355403:

| GenBank ID          | Sample ID       | SAMN ID       | WCHEN ID       |
|---------------------|-----------------|---------------|----------------|
| GCA_001984825.1     | MTSO00000000    | SAMN06249239  | WCHECI-C4     |
| Enterobacter genomosp. L | 95.68        | Enterobacter sp. WCHECI-C4 (chengduensis type strain from paper) |

Columns 5 and 6 are MASH ANI best hits to the type and proxy type strains from our paper. So we would like genomosp. T to become chuandaensis, genomosp. L to become chengduensis, and genomosp. N to become sichuanensis. Enterobacter huaxiensis appears to be a novel species with no corresponding genomosp. clade.

The authors also seem concerned that E. roggenkampii is L. nimipressuralis even though we argued for why this is not the case in this paper.

It would have been nice if the authors and/or IJSEM had looked at our paper and kept NCBI in the loop about the new type strains and reassigning the genomosp. clades.

**Competing Interests:** No competing interests were disclosed.
Author Response 27 Jun 2018

Granger Sutton, J Craig Venter Institute, Rockville, USA

We thank Florian Plaza Onate for pointing this out. To confirm this observation we started with the PanOCT run of the 250 most diverse genomes including the outlier genomes. We selected all clusters which were present in more than 151 genomes which would include all core clusters and many others. We extracted the medoid fasta sequences for these 3833 clusters. We then used our LOCUST tool to search for and extract homologous sequences from the three Enterobacter mori strains (LMG25796, 80072117, ECC1766). For LMG25796, 208 genes were missing and 328 were short. For 80072117, 95 genes were missing and 331 were short. For ECC1766, 72 genes were missing and 332 were short. For default LOCUST parameters, short genes are ones missing more than 5bp from either end of a Blast match so some short genes can be due to divergence from the medoid sequence rather than genome incompleteness. For missing genes, a small fragment may be present but was not significant enough to be found by Blast using LOCUST's blast parameters. Regardless of these caveats, it is clear that LMG25796 is the most incomplete of the three strains and for analyses needing more complete genomes should be handled with caution. However LMG25796 is the type strain and has full length genes for 3297 of the 3833 genes we selected which is more than enough for Average Nucleotide Identity calculations.

Competing Interests: No competing interests were disclosed.

Reader Comment 06 Jun 2018

Florian Plaza Oñate, Enterome, France

Enterobacter mori strain LMG 25706 is probably not a good representative of the clade. 50% (20/40) of the universal phylogenetic marker genes defined by Sunagawa et al. are missing in this genome.
In the representatives of the other clades, almost all the markers are detected (>=39/40)

Competing Interests: No competing interests were disclosed.
The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com