RESEARCH ARTICLE

REVISED *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; referees: 2 approved]

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

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

v2 **First published:** 01 May 2018, **7**:521 (doi: 10.12688/f1000research.14566.1)
Latest published: 29 Jun 2018, **7**:521 (doi: 10.12688/f1000research.14566.2)

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.

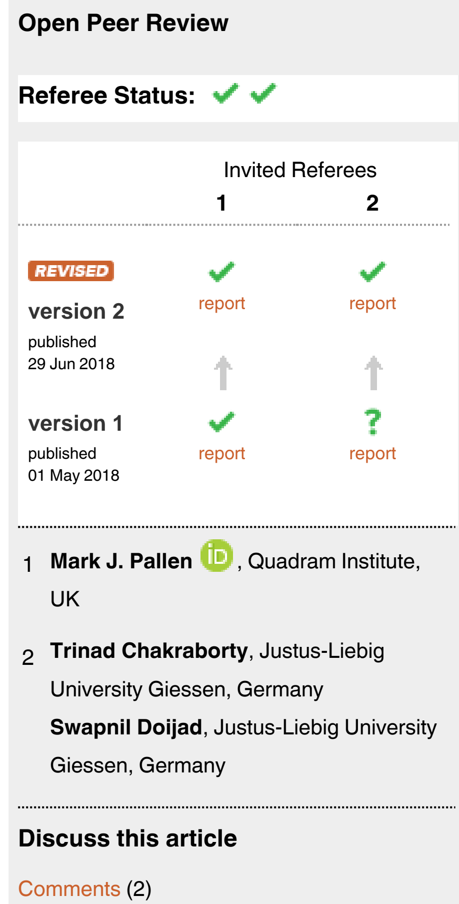
Methods: The genomes of the type strains for *Enterobacter hormaechei* subsp. *oharae*, *E. hormaechei* subsp. *steigerwaltii*, and *E. xiangfangensis*, and two strains from Hoffmann clusters III and IV of the *E. cloacae* complex were sequenced. These genomes, the *E. hormaechei* subsp. *hormaechei* 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 Licence, 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; referees: 2 approved]** *F1000Research* 2018, **7**:521 (doi: 10.12688/f1000research.14566.2)

First published: 01 May 2018, **7**:521 (doi: 10.12688/f1000research.14566.1)

REVISED **Amendments from Version 1**

This version of the paper addresses the referees' concerns. The paper is now in IMRAD format. *E. asburiae* subspecies are now discussed more fully. The "*Candidatus*" designation is discussed. PanOCT Average Nucleotide Identity (ANI) is compared to the Genome to Genome Distance Calculator (GGDC). A more thorough analysis of the outlier genomes is performed. The gene content differences between the *E. hormaechei* subspecies is made clearer. Author Thomas H. Clark's name has been corrected to "Thomas H. Clarke".

See referee reports

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*[1]. defined the type strain to be ATCC 49162[T] 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[2] 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 defined 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*[3]. 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[4].

Gu *et al*[5]. 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*[6]) *aerogenes* strains looking at antimicrobial resistance patterns[7], 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[8,9]. Initially, morphology as viewed through a microscope and later aided by staining such as Gram staining[10] 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[11–19]. 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[17–19]. 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[2,3] and others[20–26] 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

E. hormaechei subspecies not being validly published until recently. Another issue was the lack of sequenced type strains or named species/subspecies for some clades. The definition of what species/subspecies make up the *E. cloacae* complex has been in flux[2,27,28] and even what species are in the genus *Enterobacter*[29–31].

The *E. cloacae* complex was shown to have 18 clades (A-R)[7], 12 of which corresponded to 11 of the 12 clusters defined previously by Hoffmann[2]. Hoffmann cluster X is *E. nimipressuralis* which has been reclassified as *Lelliottia nimipressuralis*[29]. 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*[31].

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[32], 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. lignolyticus* genomes and 1 *E. timonensis* genome deemed to be outside of the *E. cloacae* complex. Two species of *Enterobacter*: *E. siamensis* 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*[33] is a later heterotypic synonym of *E. asburiae*[34] 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 Hoffman cluster VI; clade D containing the Hoffmann cluster III type strain (proposed name *E. hormaechei* subsp. *hoffmannii* subsp. nov.) is Hoffmann cluster III; and clade E containing the *E. hormaechei* subsp. *hormaechei* type strain is Hoffmann cluster VII.

While we believe that ANI and other similar measures recently categorized as overall genome related index (OGRI)[35] 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[36]. 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* subsp. *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 criterium 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[3,5] 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* subsp. *hormaechei* was previously distinguished from *E. hormaechei* subsp. *oharae* and *E. hormaechei* subsp. *steigerwaltii* by growth on dulcitol (a.k.a. galactitol) as the sole carbon source[3]. This phenotype can be explained by the presence of a *gat* operon[7,37] within all 7 of the *hormaechei* subsp. 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)[7,38]. 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. bugandensis*	EB-247	type	R	IX
tercancerogenus	SAMEA104113916	*E. cancerogenus*	*E. cancerogenus*	ATCC 33241	type	U	
1161ECLO	SAMN03197118	*E. cloacae*	*E. cloacae* complex clade K	1161_ECLO	proxy	K	
GN02587	SAMN03732717	*E. cloacae* complex sp. GN02587	*E. cloacae* complex clade L	GN02587	proxy	L	
DS11005	SAMN07448201	*E. cloacae*	*E. cloacae* complex clade N	DS11005	proxy	N	
GN05526	SAMN04578342	*E. cloacae* complex sp. GN05526	*E. cloacae* complex clade O	GN05526	proxy	O	
624ECLO	SAMN03197824	*E. cloacae*	*E. cloacae* complex clade P	624_ECLO	proxy	P	
ND22	SAMN05212257	*E. cloacae*	*E. cloacae* complex clade S	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	SAMN05787340	*E. hormaechei* ssp. *hormaechei*	*E. hormaechei* ssp. *hormaechei*	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. roggenkampii*	DSM 16690	type	M	IV
nterobactersoli	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*					
	No genome	*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 respectively[7,39], account for this difference where all 325 *steigerwaltii* genomes contain these operons but only 1 hoffmannii and no other hormaechei subspp. 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

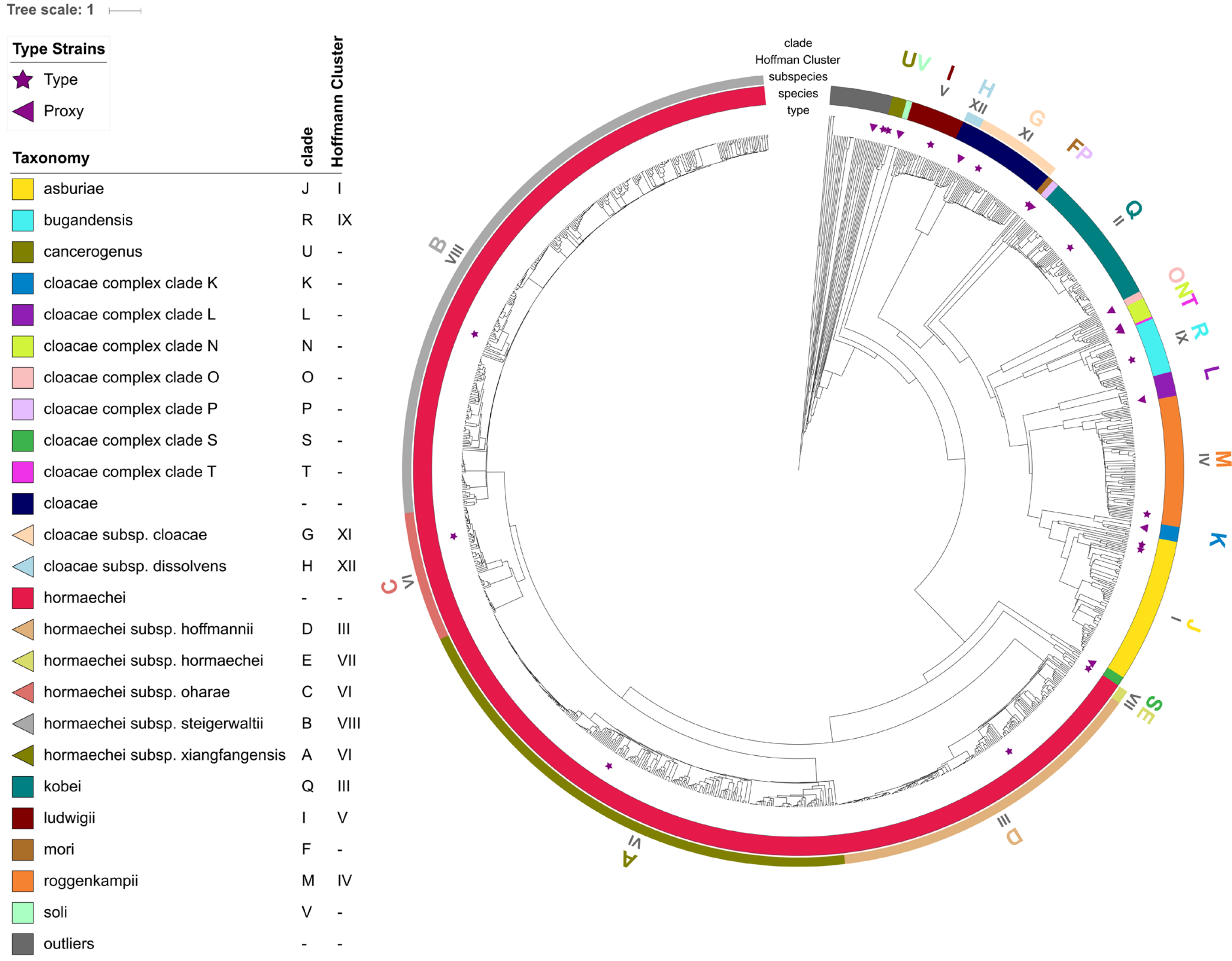


Figure 1. Average nucleotide identity (ANI) based tree for 1,249 NCBI RefSeq *Enterobacter* labelled genomes.

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[32] 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[40] 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[41] R package (Supplemental File 1) rendered with metadata annotated using the Interactive Tree of Life[42] 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 the minimum and maximum pairwise values below. The last two rows show *E. lignolyticus* (Li) and *E. timonensis* (Ti) 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	V
A	98.77 ±0.46 (97.9-100)	96.96 ±0.13 (96.2-97.5)	97.01 ±0.13 (96.3-97.6)	96.17 ±0.15 (95.3-96.9)	94.53 ±0.18 (93.9-95.2)	89.80 ±0.32 (88.9-91.0)	88.63 ±0.43 (87.5-90.9)	88.18 ±0.29 (87.1-89.4)	87.65 ±0.31 (86.3-88.8)	89.49 ±0.40 (87.8-91.2)	89.39 ±0.28 (88.4-90.3)	89.16 ±0.28 (88.0-90.3)	89.87 ±0.37 (88.4-91.6)	89.15 ±0.29 (88.1-90.4)	89.11 ±0.40 (88.3-90.7)	87.86 ±0.36 (86.8-88.7)	89.64 ±0.35 (88.5-91.2)	90.03 ±0.29 (89.0-91.2)	93.77 ±0.19 (93.2-94.5)	89.85 ±0.15 (89.3-90.4)	86.89 ±0.38 (85.3-88.1)	87.43 ±0.23 (86.9-88.2)
B	96.96 ±0.13 (96.2-97.5)	98.61 ±0.29 (97.8-100)	97.33 ±0.13 (96.7-97.8)	95.98 ±0.17 (95.3-96.9)	94.51 ±0.21 (94.0-95.2)	89.48 ±0.41 (88.5-90.8)	88.48 ±0.42 (87.0-90.6)	88.14 ±0.37 (86.8-89.4)	88.28 ±0.33 (87.0-89.5)	89.28 ±0.44 (87.5-91.1)	88.98 ±0.27 (88.1-89.9)	89.13 ±0.30 (88.1-90.1)	89.43 ±0.44 (87.9-91.3)	89.13 ±0.28 (88.3-90.2)	89.09 ±0.39 (88.2-90.5)	87.92 ±0.55 (86.6-89.1)	89.43 ±0.36 (88.3-91.1)	90.01 ±0.29 (89.1-91.3)	93.89 ±0.25 (93.0-94.6)	89.57 ±0.24 (89.0-90.2)	87.27 ±0.42 (85.9-88.3)	87.65 ±0.25 (86.9-88.4)
C	97.01 ±0.13 (96.3-97.6)	97.33 ±0.13 (96.7-97.8)	98.66 ±0.84 (97.6-100)	96.03 ±0.16 (95.5-96.6)	94.75 ±0.16 (94.4-95.2)	89.35 ±0.39 (88.5-90.5)	88.84 ±0.43 (87.9-90.6)	88.25 ±0.32 (87.3-88.9)	88.10 ±0.30 (87.3-89.1)	89.29 ±0.44 (87.9-90.8)	89.21 ±0.31 (88.1-89.8)	89.19 ±0.26 (88.4-89.9)	89.54 ±0.47 (88.0-91.4)	89.33 ±0.24 (88.5-90.0)	89.08 ±0.36 (88.1-90.3)	88.23 ±0.49 (87.3-89.0)	89.53 ±0.39 (88.5-91.2)	89.93 ±0.25 (89.0-91.1)	93.98 ±0.25 (93.4-94.6)	89.95 ±0.27 (89.4-90.3)	87.50 ±0.48 (86.3-88.7)	87.80 ±0.32 (87.1-88.3)
D	96.17 ±0.15 (95.3-96.9)	95.98 ±0.17 (95.3-96.9)	96.03 ±0.16 (95.5-96.6)	98.89 ±0.51 (97.7-100)	94.18 ±0.16 (93.7-94.7)	89.54 ±0.35 (88.8-90.8)	88.79 ±0.42 (88.0-90.6)	88.47 ±0.31 (87.8-89.4)	87.71 ±0.27 (86.7-89.1)	89.53 ±0.39 (88.2-91.2)	88.94 ±0.42 (87.8-89.8)	89.11 ±0.20 (88.5-90.0)	89.69 ±0.41 (88.4-91.5)	89.14 ±0.31 (88.3-90.3)	88.96 ±0.38 (88.3-90.6)	88.30 ±0.44 (87.2-89.1)	89.08 ±0.39 (88.1-91.0)	90.19 ±0.29 (89.1-91.7)	93.96 ±0.14 (93.5-94.5)	89.86 ±0.18 (89.4-90.4)	87.14 ±0.39 (85.9-88.4)	87.75 ±0.22 (87.3-88.4)
E	94.53 ±0.18 (93.9-95.2)	94.51 ±0.21 (94.0-95.2)	94.75 ±0.16 (94.4-95.2)	94.18 ±0.16 (93.7-94.7)	99.08 ±0.54 (98.2-100)	88.89 ±0.34 (88.4-89.7)	88.51 ±0.34 (87.8-89.9)	88.01 ±0.32 (87.4-88.9)	87.30 ±0.40 (86.5-88.4)	88.85 ±0.39 (87.5-91.0)	88.40 ±0.49 (87.5-89.6)	88.45 ±0.24 (87.8-88.9)	89.03 ±0.44 (87.8-90.5)	88.55 ±0.44 (87.6-89.6)	88.48 ±0.45 (87.8-89.5)	87.41 ±0.46 (86.7-88.4)	89.28 ±0.25 (88.5-90.0)	89.98 ±0.30 (89.4-90.8)	93.32 ±0.24 (92.8-93.9)	89.95 ±0.21 (89.7-90.3)	87.22 ±0.53 (85.9-88.3)	87.19 ±0.14 (86.9-87.5)
F	89.80 ±0.32 (88.9-91.0)	89.48 ±0.41 (88.5-90.8)	89.35 ±0.39 (88.5-90.5)	89.54 ±0.35 (88.8-90.8)	88.89 ±0.34 (88.4-89.7)	97.79 ±0.44 (97.4-98.3)	89.10 ±0.44 (88.1-90.7)	89.60 ±0.27 (89.2-90.1)	88.82 ±0.22 (88.4-89.3)	91.22 ±0.31 (90.3-92.1)	91.08 ±0.29 (90.6-91.6)	90.20 ±0.24 (89.8-90.6)	90.85 ±0.33 (90.1-91.5)	90.20 ±0.35 (89.6-90.9)	91.33 ±0.42 (90.4-91.8)	89.49 ±0.18 (89.2-89.7)	90.60 ±0.28 (90.0-91.3)	91.40 ±0.21 (90.8-92.0)	89.23 ±0.22 (88.8-89.6)	91.19 ±0.21 (91.0-91.4)	88.69 ±0.34 (88.0-89.2)	87.98 ±0.35 (87.5-88.3)
G	88.63 ±0.43 (87.5-90.9)	88.48 ±0.42 (87.0-90.6)	88.84 ±0.43 (87.9-90.6)	88.79 ±0.42 (88.0-90.6)	88.51 ±0.34 (87.8-89.9)	89.10 ±0.44 (88.1-90.7)	98.42 ±0.31 (97.7-100)	95.70 ±0.18 (95.2-96.2)	88.82 ±0.31 (87.6-89.5)	89.93 ±0.38 (88.8-91.3)	89.18 ±0.32 (88.5-90.2)	89.47 ±0.25 (89.0-90.4)	89.28 ±0.39 (88.0-90.9)	90.14 ±0.31 (89.3-91.1)	90.00 ±0.28 (89.3-90.9)	88.65 ±0.29 (87.7-89.3)	89.86 ±0.32 (88.7-91.5)	89.75 ±0.34 (88.9-91.2)	88.13 ±0.31 (87.5-88.7)	89.99 ±0.17 (89.6-90.4)	87.86 ±0.20 (87.3-88.3)	87.48 ±0.19 (87.0-88.0)
H	88.18 ±0.29 (87.1-89.4)	88.14 ±0.37 (86.8-89.4)	88.25 ±0.32 (87.3-88.9)	88.47 ±0.31 (87.8-89.4)	88.01 ±0.32 (87.4-88.9)	89.60 ±0.27 (89.2-90.1)	95.70 ±0.18 (95.2-96.2)	98.82 ±0.24 (98.6-100)	89.09 ±0.30 (88.2-89.8)	89.97 ±0.38 (88.7-90.9)	89.85 ±0.37 (89.0-90.5)	89.69 ±0.25 (89.2-90.1)	89.47 ±0.45 (88.2-90.6)	90.43 ±0.33 (89.5-91.0)	90.78 ±0.19 (90.4-91.1)	88.75 ±0.30 (88.2-89.4)	89.35 ±0.34 (88.3-90.1)	90.13 ±0.33 (89.3-91.1)	88.25 ±0.22 (87.7-88.7)	90.61 ±0.16 (90.3-90.8)	88.24 ±0.31 (87.6-89.0)	88.15 ±0.19 (87.8-88.4)
I	87.65 ±0.31 (86.3-88.8)	88.28 ±0.33 (87.0-89.5)	88.10 ±0.30 (87.3-89.1)	87.71 ±0.27 (86.7-89.1)	87.30 ±0.40 (86.5-88.4)	88.82 ±0.22 (88.4-89.3)	88.82 ±0.31 (87.6-89.5)	89.09 ±0.30 (88.2-89.8)	98.63 ±0.24 (98.0-100)	88.98 ±0.33 (87.7-89.9)	89.01 ±0.24 (88.3-89.6)	88.65 ±0.33 (87.8-89.7)	88.90 ±0.35 (87.8-89.9)	88.87 ±0.33 (88.0-89.6)	89.17 ±0.27 (88.5-89.8)	88.46 ±0.38 (87.7-89.1)	89.32 ±0.27 (88.5-90.1)	88.95 ±0.33 (87.9-89.6)	87.44 ±0.40 (86.5-88.6)	88.90 ±0.20 (88.3-89.3)	87.17 ±0.23 (86.5-87.7)	87.64 ±0.18 (87.3-88.0)
J	89.49 ±0.40 (87.8-91.2)	89.28 ±0.44 (87.5-91.1)	89.29 ±0.44 (87.9-90.8)	89.53 ±0.39 (88.2-91.2)	88.85 ±0.39 (87.5-91.0)	91.22 ±0.31 (90.3-92.1)	89.93 ±0.38 (88.8-91.3)	89.97 ±0.38 (88.7-90.9)	88.98 ±0.33 (87.7-89.9)	97.20 ± 1.15 (94.5-100)	94.38 ±0.28 (93.6-95.2)	93.50 ±0.21 (92.8-94.0)	93.47 ±0.25 (92.5-94.3)	91.88 ±0.29 (91.1-92.6)	92.99 ±0.29 (92.2-94.0)	91.93 ±0.27 (91.3-92.7)	92.03 ±0.26 (91.2-93.0)	92.30 ±0.33 (91.1-93.3)	89.13 ±0.41 (87.5-90.0)	92.46 ±0.25 (91.8-92.9)	87.79 ±0.46 (86.0-88.8)	88.17 ±0.26 (87.5-88.7)
K	89.39 ±0.28 (88.4-90.3)	88.98 ±0.27 (88.1-89.9)	89.21 ±0.31 (88.1-89.8)	88.94 ±0.42 (87.8-89.8)	88.40 ±0.49 (87.5-89.6)	91.08 ±0.29 (90.6-91.6)	89.18 ±0.32 (88.5-90.2)	89.85 ±0.37 (89.0-90.5)	89.01 ±0.24 (88.3-89.6)	94.38 ±0.28 (93.6-95.2)	98.26 ±0.77 (97.5-100)	93.29 ±0.26 (92.8-93.7)	94.14 ±0.36 (93.3-95.5)	92.05 ±0.44 (90.9-92.6)	92.90 ±0.24 (92.5-93.3)	91.78 ±0.21 (91.4-92.1)	91.63 ±0.28 (90.5-92.2)	92.38 ±0.27 (91.6-92.9)	89.18 ±0.30 (88.7-89.8)	92.37 ±0.22 (92.0-92.7)	87.62 ±0.34 (86.8-88.2)	87.65 ±0.16 (87.3-87.8)
L	89.16 ±0.28 (88.0-90.3)	89.13 ±0.30 (88.1-90.1)	89.19 ±0.26 (88.4-89.9)	89.11 ±0.20 (88.5-90.0)	88.45 ±0.24 (87.8-88.9)	90.20 ±0.24 (89.8-90.6)	89.47 ±0.25 (89.0-90.4)	89.69 ±0.25 (89.2-90.1)	88.65 ±0.33 (87.8-89.7)	93.50 ±0.21 (92.8-94.0)	93.29 ±0.26 (92.8-93.7)	97.80 ± 1.94 (95.6-100)	93.20 ±0.21 (92.5-93.8)	91.67 ±0.27 (91.0-92.2)	92.24 ±0.13 (92.0-92.5)	90.87 ±0.21 (90.5-91.3)	91.93 ±0.23 (91.2-92.7)	91.65 ±0.25 (91.0-92.4)	89.15 ±0.26 (88.6-89.7)	91.26 ±0.25 (90.8-91.6)	88.06 ±0.32 (87.2-88.7)	87.73 ±0.28 (87.0-88.1)

	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 (88.4-91.6)	89.43 ±0.44 (87.9-91.3)	89.54 ±0.47 (88.0-91.4)	89.69 ±0.41 (88.4-91.5)	89.03 ±0.44 (87.8-90.5)	90.85 ±0.33 (90.1-91.5)	89.28 ±0.39 (88.0-90.9)	89.47 ±0.45 (88.2-90.6)	88.90 ±0.35 (87.8-89.9)	93.47 ±0.25 (92.5-94.3)	94.14 ±0.36 (93.3-95.5)	93.20 ±0.21 (92.5-93.8)	97.72 ±0.86 (95.1-100)	92.30 ±0.30 (91.5-93.1)	92.39 ±0.28 (91.6-93.0)	90.92 ±0.27 (90.1-91.4)	91.47 ±0.34 (90.2-92.9)	92.21 ±0.29 (91.1-93.0)	89.57 ±0.36 (88.5-90.5)	91.90 ±0.22 (91.3-92.3)	87.31 ±0.44 (86.2-88.3)	87.72 ±0.26 (87.3-88.3)
N	89.15 ±0.29 (88.1-90.4)	89.13 ±0.28 (88.3-90.2)	89.33 ±0.24 (88.5-90.0)	89.14 ±0.31 (88.3-90.3)	88.55 ±0.44 (87.6-89.6)	90.20 ±0.35 (89.6-90.9)	90.14 ±0.31 (89.3-91.1)	90.43 ±0.33 (89.5-91.0)	88.87 ±0.33 (88.0-89.6)	91.88 ±0.29 (91.1-92.6)	92.05 ±0.44 (90.9-92.6)	91.67 ±0.27 (91.0-92.2)	92.30 ±0.30 (91.5-93.1)	98.28 ±0.41 (97.6-99.9)	93.13 ±0.20 (92.8-93.5)	90.64 ±0.16 (90.2-90.9)	90.78 ±0.30 (90.0-91.5)	91.14 ±0.28 (90.3-91.8)	88.72 ±0.41 (87.7-89.5)	90.90 ±0.39 (90.3-91.5)	86.75 ±0.38 (85.7-87.4)	87.37 ±0.14 (87.2-87.7)
O	89.11 ±0.40 (88.3-90.7)	89.09 ±0.39 (88.2-90.5)	89.08 ±0.36 (88.1-90.3)	88.96 ±0.38 (88.3-90.6)	88.48 ±0.45 (87.8-89.5)	91.33 ±0.42 (90.4-91.8)	90.00 ±0.28 (89.3-90.9)	90.78 ±0.19 (90.4-91.1)	89.17 ±0.27 (88.5-89.8)	92.99 ±0.29 (92.2-94.0)	92.90 ±0.24 (92.5-93.3)	92.24 ±0.13 (92.0-92.5)	92.39 ±0.28 (91.6-93.0)	93.13 ±0.20 (92.8-93.5)	97.90 ±0.88 (97.0-98.8)	91.05 ±0.27 (90.5-91.4)	91.77 ±0.28 (91.0-92.5)	91.74 ±0.20 (91.3-92.2)	89.07 ±0.36 (88.5-89.8)	91.42 ±0.21 (91.2-91.6)	88.02 ±0.36 (87.3-88.7)	87.80 ±0.14 (87.6-88.0)
P	87.86 ±0.36 (86.8-88.7)	87.92 ±0.55 (86.6-89.1)	88.23 ±0.49 (87.3-89.0)	88.30 ±0.44 (87.2-89.1)	87.41 ±0.46 (86.7-88.4)	89.49 ±0.18 (89.2-89.7)	88.65 ±0.29 (87.7-89.3)	88.75 ±0.30 (88.2-89.4)	88.46 ±0.38 (87.7-89.1)	91.93 ±0.27 (91.3-92.7)	91.78 ±0.21 (91.4-92.1)	90.87 ±0.21 (90.5-91.3)	90.92 ±0.27 (90.1-91.4)	90.64 ±0.16 (90.2-90.9)	91.05 ±0.27 (90.5-91.4)	98.77 ±0.91 (98.2-100)	90.44 ±0.36 (89.7-91.2)	90.34 ±0.37 (89.6-91.1)	88.14 ±0.40 (87.5-88.9)	90.00 ±0.26 (89.8-90.2)	86.49 ±0.45 (85.7-87.2)	88.32 ±0.19 (88.1-88.5)
Q	89.64 ±0.35 (88.5-91.2)	89.43 ±0.36 (88.3-91.1)	89.53 ±0.39 (88.5-91.2)	89.08 ±0.39 (88.1-91.0)	89.28 ±0.25 (88.5-90.0)	90.60 ±0.28 (90.0-91.3)	89.86 ±0.32 (88.7-91.5)	89.35 ±0.34 (88.3-90.1)	89.32 ±0.27 (88.5-90.1)	92.03 ±0.26 (91.2-93.0)	91.63 ±0.28 (90.5-92.2)	91.93 ±0.23 (91.2-92.7)	91.47 ±0.34 (90.2-92.9)	90.78 ±0.30 (90.0-91.5)	91.77 ±0.28 (91.0-92.5)	90.44 ±0.36 (89.7-91.2)	98.67 ±0.39 (97.9-100)	92.01 ±0.25 (91.2-92.9)	89.12 ±0.30 (88.4-89.9)	91.89 ±0.20 (91.3-92.3)	88.42 ±0.34 (87.2-89.0)	87.40 ±0.21 (87.0-88.0)
R	90.03 ±0.29 (89.0-91.2)	90.01 ±0.29 (89.1-91.3)	89.93 ±0.25 (89.0-91.1)	90.19 ±0.29 (89.1-91.7)	89.98 ±0.30 (89.4-90.8)	91.40 ±0.21 (90.8-92.0)	89.75 ±0.34 (88.9-91.2)	90.13 ±0.33 (89.3-91.1)	88.95 ±0.33 (87.9-89.6)	92.30 ±0.33 (91.1-93.3)	92.38 ±0.27 (91.6-92.9)	91.65 ±0.25 (91.0-92.4)	92.21 ±0.29 (91.1-93.0)	91.14 ±0.28 (90.3-91.8)	91.74 ±0.20 (91.3-92.2)	90.34 ±0.37 (89.6-91.1)	92.01 ±0.25 (91.2-92.9)	98.24 ±0.79 (95.6-100)	90.92 ±0.24 (90.2-91.4)	94.18 ±0.12 (94.0-94.6)	88.49 ±0.34 (87.5-89.3)	88.37 ±0.21 (87.9-88.7)
S	93.77 ±0.19 (93.2-94.5)	93.89 ±0.25 (93.0-94.6)	93.98 ±0.25 (93.4-94.6)	93.96 ±0.14 (93.5-94.5)	93.32 ±0.24 (92.8-93.9)	89.23 ±0.22 (88.8-89.6)	88.13 ±0.31 (87.5-88.7)	88.25 ±0.22 (87.7-88.7)	87.44 ±0.40 (86.5-88.6)	89.13 ±0.41 (87.5-90.0)	89.18 ±0.30 (88.7-89.8)	89.15 ±0.26 (88.6-89.7)	89.57 ±0.36 (88.5-90.5)	88.72 ±0.41 (87.7-89.5)	89.07 ±0.36 (88.5-89.8)	88.14 ±0.40 (87.5-88.9)	89.12 ±0.30 (88.4-89.9)	90.92 ±0.24 (90.2-91.4)	98.52 ±0.89 (97.8-100)	89.75 ±0.29 (89.5-90.2)	88.18 ±0.32 (87.4-88.7)	87.66 ±0.33 (87.2-88.1)
T	89.85 ±0.15 (89.3-90.4)	89.57 ±0.24 (89.0-90.2)	89.95 ±0.27 (89.4-90.3)	89.86 ±0.18 (89.4-90.4)	89.95 ±0.21 (89.7-90.3)	91.19 ±0.21 (91.0-91.4)	89.99 ±0.17 (89.6-90.4)	90.61 ±0.16 (90.3-90.8)	88.90 ±0.20 (88.3-89.3)	92.46 ±0.25 (91.8-92.9)	92.37 ±0.22 (92.0-92.7)	91.26 ±0.25 (90.8-91.6)	91.90 ±0.22 (91.3-92.3)	90.90 ±0.39 (90.3-91.5)	91.42 ±0.21 (91.2-91.6)	90.00 ±0.26 (89.8-90.2)	91.89 ±0.20 (91.3-92.3)	94.18 ±0.12 (94.0-94.6)	89.75 ±0.29 (89.5-90.2)	100 ±0.00 (100-100)	88.62 ±0.40 (87.7-88.9)	88.76 ±0.04 (88.7-88.8)
U	86.89 ±0.38 (85.3-88.1)	87.27 ±0.42 (85.9-88.3)	87.50 ±0.48 (86.3-88.7)	87.14 ±0.39 (85.9-88.4)	87.22 ±0.53 (85.9-88.3)	88.69 ±0.34 (88.0-89.2)	87.86 ±0.20 (87.3-88.3)	88.24 ±0.31 (87.6-89.0)	87.17 ±0.23 (86.5-87.7)	87.79 ±0.46 (86.0-88.8)	87.62 ±0.34 (86.8-88.2)	88.06 ±0.32 (87.2-88.7)	87.31 ±0.44 (86.2-88.3)	86.75 ±0.38 (85.7-87.4)	88.02 ±0.36 (87.3-88.7)	86.49 ±0.45 (85.7-87.2)	88.42 ±0.34 (87.2-89.0)	88.49 ±0.34 (87.5-89.3)	88.18 ±0.32 (87.4-88.7)	88.62 ±0.40 (87.7-88.9)	98.80 ±0.65 (98.3-100)	86.62 ±0.27 (86.2-86.9)
V	87.43 ±0.23 (86.9-88.2)	87.65 ±0.25 (86.9-88.4)	87.80 ±0.32 (87.1-88.3)	87.75 ±0.22 (87.3-88.4)	87.19 ±0.14 (86.9-87.5)	87.98 ±0.35 (87.5-88.3)	87.48 ±0.19 (87.0-88.0)	88.15 ±0.19 (87.8-88.4)	87.64 ±0.18 (87.3-88.0)	88.17 ±0.26 (87.5-88.7)	87.65 ±0.16 (87.3-87.8)	87.73 ±0.28 (87.0-88.1)	87.72 ±0.26 (87.3-88.3)	87.37 ±0.14 (87.2-87.7)	87.80 ±0.14 (87.6-88.0)	88.32 ±0.19 (88.1-88.5)	87.40 ±0.21 (87.0-88.0)	88.37 ±0.21 (87.9-88.7)	87.66 ±0.33 (87.2-88.1)	88.76 ±0.04 (88.7-88.8)	86.62 ±0.27 (86.2-86.9)	99.99 ±0.00 (100-100)
Li	82.06 ±0.30 (81.1-83.0)	82.87 ±0.33 (81.8-83.8)	83.03 ±0.16 (82.6-83.5)	82.22 ±0.25 (81.4-83.0)	84.03 ±0.17 (83.8-84.3)	82.55 ±0.30 (82.2-82.8)	82.25 ±0.61 (80.9-83.3)	81.56 ±0.33 (81.1-82.0)	81.73 ±0.39 (80.9-82.6)	82.36 ±0.34 (81.6-83.5)	83.02 ±0.31 (82.6-83.3)	82.41 ±0.15 (82.2-82.6)	82.68 ±0.49 (81.4-83.5)	81.49 ±0.45 (80.9-82.4)	83.16 ±0.34 (83.0-83.7)	81.12 ±0.28 (80.9-81.4)	81.61 ±0.21 (81.1-82.0)	82.76 ±0.31 (81.8-83.3)	83.72 ±0.31 (83.3-84.1)	82.99 ±0.00 (83.0-83.0)	83.13 ±0.31 (82.4-83.5)	82.23 ±0.00 (82.2-82.2)
Ti	85.07 ±0.27 (83.8-85.9)	85.82 ±0.25 (85.1-86.4)	85.45 ±0.24 (85.1-86.3)	85.93 ±0.18 (85.2-86.6)	84.70 ±0.40 (84.0-85.3)	85.22 ±0.18 (85.1-85.4)	84.47 ±0.32 (83.8-85.1)	83.99 ±0.40 (83.3-84.7)	83.88 ±0.23 (83.3-84.3)	85.45 ±0.44 (84.3-86.3)	84.59 ±0.40 (84.1-85.2)	85.73 ±0.52 (84.9-86.2)	85.10 ±0.39 (84.0-86.0)	84.93 ±0.33 (84.3-85.3)	85.42 ±0.22 (85.2-85.7)	83.50 ±0.19 (83.3-83.7)	84.71 ±0.37 (84.0-85.5)	84.97 ±0.22 (84.7-85.5)	86.16 ±0.16 (86.0-86.4)	84.94 ±0.00 (84.9-84.9)	85.58 ±0.17 (85.3-85.8)	83.87 ±0.09 (83.8-84.0)

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 genera with best MASH matches to the 30 outliers (Supplemental Table 2), and 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. roggenkampii* (**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[36,43] 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)[18] 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. roggenkampii* 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 $d_{PANI} = 1 - (\text{PanOCT ANI}/100)$. 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_{PANI} is practically indistinguishable from GGDC Formula 2.

For the PanOCT run with 1,216 genomes to determine gene content similarities, PanOCT was run as part of the JCVI pan-genome pipeline in hierarchical fashion with the following batches of genomes run by PanOCT at level 1: (combined 3 *E. mori*, 3 *E. soli*, 8 *E. cancerogenus*, 8 *E. cloacae* complex clade K, 13 *E. cloacae* complex clade L, 11 *E. cloacae* complex clade N, 4 *E. cloacae* complex clade O, 4 *E. cloacae* complex clade P, 5 *E. cloacae* complex clade S, 1 *E. cloacae* complex clade T); (combined 45 *E. cloacae* subsp. *cloacae*, 9 *E. cloacae* subsp. *dissolvens*); (randomly split into 4 groups 169 *E. hormaechei* subsp. *hoffmannii*); (7 *E. hormaechei* subsp. *hormaechei*); (68 *E. hormaechei* subsp. *oharae*); (randomly split into 8 groups 325 *E. hormaechei* subsp. *steigerwaltii*); (randomly split into 6 groups

255 *E. hormaechei* subsp. *xiangfangensis*); (78 *E. asburiae*); (30 *E. bugandensis*); (71 *E. kobei*); (29 *E. ludwigii*); and (70 *E. roggenkampii*). The level 1 clusters were then combined using PanOCT at level 2 and the final output generated using the PanOCT (version 3.27) command line:

panoct.pl -R matchtable.txt -f genomes.list -g combined.att_file -P combined.fasta -b final_panoct_run -c 0,95

The diverse 250 genome PanOCT run and the level 1 PanOCT batch runs used the PanOCT (version 3.27) command line:

panoct.pl -b results -t combined.blast -f genomes.list -g combined.att -P combined.fasta -S yes -L 1 -M Y -H Y -V Y -N Y -F 1.33 -G y -c 0,50,95,100 -T

The hierarchical PanOCT run of 1,216 genomes produced a matrix of orthologous gene clusters (Supplemental Table 3) where the rows are clusters and the columns are genomes with the cells containing the RefSeq IDs for the gene from the corresponding genome. This matrix was used to determine genes common to all, specific to, and missing from clades A-V. Individual PanOCT runs were also done for clade J, D, and M. Clade J to insure that PanOCT ANI values confirmed MASH ANI values that *E. asburiae* and *E. muelleri* are the same species which they did and these ANI values were used to determine the 8 subclades at 98% ANI using hierarchical clustering (hclust in R) average linkage. Clade D to confirm the MASH ANI values for *E. hormaechei* subsp. *hoffmannii* which they did. Clade M was done likewise to confirm *E. roggenkampii* which they did.

Discussion

The Introduction section reviews how the tools for defining a species have evolved. In a recent review of the genus *Mycobacterium*, the authors proposed that any newly defined bacterial species must have a genome sequence and an ANI comparison carried out against existing sequenced type strains to justify a novel species assignment[45]. ANI analysis should not be relied on in isolation for defining a species since historical or clinical phenotypic distinctions may be important for example in distinguishing between *E. coli* and *Shigela* which by ANI are the same species. However, genome sequencing appears to be outstripping the taxonomic definition of species within some genera. For the 22 clades of the E. cloacae complex identified here 9 do not have named type strains (7 if the two proposed here are adopted). For important pathogens where clinical practice may rely on proper classification the ability to name these clades/species and provide resources for identifying them could be pivotal. Unfortunately, the current established journal for validly publishing bacterial species' names, IJSEM, insists on phenotypic characterization and deposition of the type strain before naming is valid. This prevents computational based methods from moving quickly. Paradoxically almost all species identifying diagnostic tests are genotype not phenotype based so genotype is good enough for diagnosis but not species definition. Further, delineating what is acceptable to define as a new species is also genotype not phenotype based whether via DDH, marker genes, or more recently ANI. Worse there are no published standards for what defines the minimal set of phenotypic biochemical assays that must be performed. As the *Mycobacterium* review authors state: "The easy and affordable availability of reliable whole-genome sequences raises doubts about the real added value of investigating phenotypic traits when a new species is described. Actually, different taxonomists use their own panels of tests, often not standardized, to produce results of no use for colleagues and absolutely incomprehensible to the community of mycobacteriologists who have dismissed such approach since the '90s. For the genus Mycobacterium the major phenotypic traits that cannot be disregarded should include growth rate and pigmentation of colonies, while the classical investigation of biochemical activities is clearly obsolete.". If there were accepted standards for minimal phenotypic characterization then culture collection repositories could choose to provide the characterization as fee for service or even for free for type strains as an incentive for deposition. With the rapid growth in synthetic genomics capabilities one could argue that the deposition of a high quality complete genome might suffice rather than a culture.

We propose allowing "placeholder" species or subspecies names such as "*E. cloacae* complex clade S" in order to enable the most robust use of computational and genomic resources for clinical diagnosis. IJSEM currently recognizes provisional species names under the *Candidatus* designation[46]. *Candidatus* was designed for unculturable organisms where a type strain could not be maintained but phenotypic data is still required to be submitted. This is not a good fit for the case where genome sequences exist and species/subspecies are determined computationally because it was designed for environmental or unculturable samples with limited sequence data but at least some phenotypic or morphological data. We suggest that some similar designation be used for our proposed "placeholder" names. We do not want to computationally assign permanent names with a provisional status, but would rather have the name itself 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.

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.*[35] 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 *E. asburiae* 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 *E. 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. cloacae* complex[2] in particular the subspecies of *E. hormaechei*[3]. We propose to name clade M/Hoffmann cluster IV *Enterobacter roggenkampii* after Andreas Roggenkamp for his work on elucidating the phylogenetic structure of the *E. cloacae* complex[2]. The analysis also shows that *E. muelleri*[33] is a later heterotypic synonym of *E. asburiae*[34] which should take precedence.

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

E. hormaechei subsp. *xiangfangensis* (xi.ang.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*[5].

The species description is unchanged from its description as *Enterobacter xiangfangensis*[5].

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

The GenBank accessions for the complete genome sequence of *E. 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.i. N.L. gen. m. *Hoffmann*, in honor of Harald Hoffmann, a German microbiologist who helped elucidate the phylogenetic structure of the *E. cloacae* complex in particular the subspecies of *E. hormaechei*).

Hoffmann and Roggenkamp[2] determined clusters within the *E. cloacae* complex using marker genes, primarily *hsp60*. Hoffman *et al*[3]. 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*[7]. determined groups for the *E. cloacae* 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.i. N.L. gen. m. *Roggenkamp*, in honor of Andreas Roggenkamp, a German microbiologist who helped elucidate the phylogenetic structure of the *E. cloacae* complex).

Hoffmann and Roggenkamp[2] determined clusters within the *E. cloacae* complex using marker genes, primarily *hsp60*. Chavda *et al*[7]. determined groups for the *E. cloacae* 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. cloacae* 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 *E. hormaechei* subsp. *steigerwaltii* are PRJNA259658, SAMN05581751, ASM172972v1, and CP017179.1.

The GenBank accessions for the complete genome sequence of *E. 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, specifc 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 *rbt/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 nimipressuralis* 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.
PubMed Abstract | Free Full Text
2. Hoffmann H, Roggenkamp A: **Population genetics of the nomenspecies *Enterobacter cloacae*.** *Appl Environ Microbiol.* 2003; **69**(9): 5306–18.
PubMed Abstract | Publisher Full Text | Free Full Text
3. 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
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.
PubMed Abstract | Publisher Full Text
5. Gu CT, Li CY, Yang LJ, *et al.*: ***Enterobacter xiangfangensis* sp. nov., isolated from Chinese traditional sourdough, and reclassification of *Enterobacter sacchari* Zhu *et al.* 2013 as *Kosakonia sacchari* comb. nov.** *Int J Syst Evol Microbiol.* 2014; **64**(Pt 8): 2650–6.
PubMed Abstract | Publisher Full Text
6. Tindall 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.
PubMed 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): pii: e02093-16.
PubMed Abstract | Publisher Full Text | Free Full Text
8. Staley JT: **The bacterial species dilemma and the genomic-phylogenetic species concept.** *Philos Trans R Soc Lond B Biol Sci.* 2006; **361**(1475): 1899–909.
PubMed Abstract | Publisher Full Text | Free Full Text
9. Georgiades K, Raoult D: **Defining pathogenic bacterial species in the genomic era.** *Front Microbiol.* 2011; **1**: 151.
PubMed Abstract | Publisher Full Text | Free Full Text
10. Bartholomew JW, Mittwer T: **The Gram stain.** *Bacteriol Rev.* 1952; **16**(1): 1–29.
PubMed 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.
PubMed Abstract | Publisher 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.
PubMed 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.
PubMed Abstract | Publisher Full Text | Free Full Text
14. Goris 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.
PubMed 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.
PubMed 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.
PubMed 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.
PubMed Abstract | Publisher Full Text | Free Full Text
18. Meier-Kolthoff JP, Auch AF, Klenk HP, *et al.*: **Genome sequence-based species delimitation with confidence intervals and improved distance functions.** *BMC Bioinformatics.* 2013; **14**: 60.
PubMed Abstract | Publisher Full Text | Free Full Text
19. Colston SM, Fullmer MS, Beka L, *et al.*: **Bioinformatic genome comparisons for taxonomic and phylogenetic assignments using *Aeromonas* as a test case.** *mBio.* 2014; **5**(6): e02136.
PubMed Abstract | Publisher Full Text | Free Full Text
20. Davin-Regli A, Bosi 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.
PubMed Abstract | 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.
PubMed Abstract | Publisher 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): 95–7.
PubMed Abstract | Publisher Full Text
23. Wenger 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.
PubMed 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**(5): 1315–21.
PubMed Abstract | Publisher Full Text
25. Guérin F, Isnard C, Sinel C, *et al.*: **Cluster-dependent colistin hetero-resistance in *Enterobacter cloacae* complex.** *J Antimicrob Chemother.* 2016; **71**(11): 3058–61.
PubMed Abstract | Publisher Full Text
26. Morand PC, Billoet A, Rottman 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.
PubMed Abstract | Publisher Full Text | Free Full Text
27. Mezzatesta ML, Gona F, Stefani S: ***Enterobacter cloacae* complex: clinical impact and emerging antibiotic resistance.** *Future Microbiol.* 2012; **7**(7): 887–902.
PubMed 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.
PubMed Abstract | Publisher Full Text | Free Full Text
29. Brady C, Cleenwerck I, Venter S, *et al.*: **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): 309–19.
PubMed Abstract | Publisher Full Text

30. Stephan R, Grim CJ, Gopinath 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 *Franconibacter* gen. nov. and *Siccibacter* gen. nov. as *Franconibacter helveticus* comb. nov., *Franconibacter pulveris* comb. nov. and *Siccibacter turicensis* comb. nov., respectively.** *Int J Syst Evol Microbiol.* 2014; **64**(Pt 10): 3402–10.
PubMed Abstract | Publisher Full Text | Free Full Text

31. Doijad S, Imirzalioglu C, Yao Y, *et al.*: ***Enterobacter bugandensis* sp. nov., isolated from neonatal blood.** *Int J Syst Evol Microbiol.* 2016; **66**(2): 968–74.
PubMed Abstract | Publisher Full Text

32. Ondov BD, Treangen TJ, Melsted P, *et al.*: **Mash: fast genome and metagenome distance estimation using MinHash.** *Genome Biol.* 2016; **17**(1): 132.
PubMed Abstract | Publisher Full Text | Free Full Text

33. 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.
PubMed Abstract | Publisher Full Text

34. 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.
PubMed Abstract | Free Full Text

35. 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.
PubMed Abstract | Publisher Full Text

36. 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.
PubMed Abstract | Publisher Full Text | Free Full Text

37. Nobelmann B, Lengeler JW: **Molecular analysis of the *gat* genes from *Escherichia coli* and of their roles in galactitol transport and metabolism.** *J Bacteriol.* 1996; **178**(23): 6790–5.
PubMed Abstract | Publisher Full Text | Free Full Text

38. Reizer J, Ramseier TM, Reizer A, *et al.*: **Novel phosphotransferase genes revealed by bacterial genome sequencing: a gene cluster encoding a putative N-acetylgalactosamine metabolic pathway in *Escherichia coli*.** *Microbiology.* 1996; **142**(Pt 2): 231–50.
PubMed Abstract | Publisher Full Text

39. Wu J, Anderton-Loviny T, Smith CA, *et al.*: **Structure of wild-type and mutant repressors and of the control region of the *rbt* operon of *Klebsiella aerogenes*.** *EMBO J.* 1985; **4**(5): 1339–44.
PubMed Abstract | Free Full Text

40. Clarke TH, Brinkac LM, Sutton G, *et al.*: **GGRaSP: A R-package for selecting representative genomes using Gaussian mixture models.** *Bioinformatics.* 2018; bty300.
PubMed Abstract | Publisher Full Text

41. Paradis E, Claude J, Strimmer K: **APE: Analyses of Phylogenetics and Evolution in R language.** *Bioinformatics.* 2004; **20**(2): 289–90.
PubMed Abstract | Publisher Full Text

42. Letunic I, Bork P: **Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees.** *Nucleic Acids Res.* 2016; **44**(W1): W242–5.
PubMed Abstract | Publisher Full Text | Free Full Text

43. Chan AP, Sutton G, DePew J, *et al.*: **A novel method of consensus pan-chromosome assembly and large-scale comparative analysis reveal the highly flexible pan-genome of *Acinetobacter baumannii*.** *Genome Biol.* 2015; **16**: 143.
PubMed Abstract | Publisher Full Text | Free Full Text

44. Lee I, Ouk Kim Y, Park SC, *et al.*: **OrthoANI: An improved algorithm and software for calculating average nucleotide identity.** *Int J Syst Evol Microbiol.* 2016; **66**(2): 1100–1103.
PubMed Abstract | Publisher Full Text

45. Tortoli E, Fedrizzi T, Meehan CJ, *et al.*: **The new phylogeny of the genus *Mycobacterium*: The old and the news.** *Infect Genet Evol.* 2017; **56**: 19–25.
PubMed Abstract | Publisher Full Text

46. Murray RG, Stackebrandt E: **Taxonomic note: implementation of the provisional status Candidatus for incompletely described procaryotes.** *Int J Syst Bacteriol.* 1995; **45**(1): 186–7.
PubMed Abstract | Publisher Full Text

Open Peer Review

Current Referee Status:

Version 2

Referee Report 13 July 2018

doi:10.5256/f1000research.16861.r35633

Trinad Chakraborty , **Swapnil Doijad**
Institute of Medical Microbiology, Justus-Liebig University Giessen, Giessen, Germany

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[1], Meier-Kolthoff et al., 2013[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[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 *E. xiangfangensi*s and *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 *E. xiangfangensi*s and *E. hormaechei,* respectively. Just because several strains of clade A-D were historically identified as *E. hormaechei* subspecies, there is no reason to ignore the recently validated *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, 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

Competing Interests: No competing interests were disclosed.

We have read this submission. We 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, 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 phenotyic 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.

Referee Report 12 July 2018

doi:10.5256/f1000research.16861.r35632

Mark J. Pallen
Quadram Institute, Norwich, UK

I am happy with the changes.

Competing Interests: No competing interests were disclosed.

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

Version 1

Referee Report 07 June 2018

doi:10.5256/f1000research.15853.r34159

?

Trinad Chakraborty , **Swapnil Doijad**
Institute of Medical Microbiology, Justus-Liebig University Giessen, Giessen, Germany

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-247^T.

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 have read this submission. We 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, 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 be added.”
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). 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.

Referee Report 17 May 2018

doi:10.5256/f1000research.15853.r33699

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 have read this submission. I 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, 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.

Discuss this Article

Author Response 27 Jun 2018

Granger Sutton, J. Craig Venter Institute, 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