Discrimination of species within the *Enterobacter cloacae* complex using MALDI-TOF Mass Spectrometry and Fourier-Transform Infrared Spectroscopy coupled with Machine Learning tools

Ana Candela[1*], Miriam Mateos[2*], Alicia Gómez-Asenjo[1], Manuel J. Arroyo[3], Marta Hernandez-García[2,4], Rosa del Campo[2,4], Emilia Cercenado[1,5,6], Gema Méndez[3], Luis Mancera[3], Juan de Dios Caballero[2,4], Laura Martínez-García[2,7], Desirée Gijón[2,4], María Isabel Morosini[2,4], Patricia Ruiz-Garbajosa[2,4], Rafael Cantón[2,4], Patricia Muñoz[1,5,6], David Rodríguez-Temporal[1†] and Belén Rodríguez-Sánchez[1].

[1]Clinical Microbiology and Infectious Diseases Department, Hospital General Universitario Gregorio Marañón, Madrid, Spain and Institute of Health Research Gregorio Marañón (IiSGM), Madrid, Spain

[2]Servicio de Microbiología. Hospital Universitario Ramón y Cajal and Instituto Ramón y Cajal de Investigación Sanitaria (IRYCIS), Madrid, Spain.

[3]Clover Bioanalytical Software, Av. del Conocimiento, 41, 18016 Granada, Spain

[4]CIBER en Enfermedades Infecciosas, Madrid, Spain

[5]CIBER de Enfermedades Respiratorias (CIBERES CB06/06/0058), Madrid, Spain.

[6]Medicine Department, Faculty of Medicine, Universidad Complutense de Madrid, Madrid, Spain

[7]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain.

Running Title: *E. cloacae* complex analysis with MALDI-TOF MS and FTIR S

Corresponding Author:

David Rodríguez-Temporal, PhD.

Servicio de Microbiología Clínica y Enfermedades Infecciosas. Hospital General Universitario Gregorio Marañón.Dr. Esquerdo 46. 28007 Madrid, Spain

Phone: +34- 91- 426 7163, Fax: +34- 91- 426 9595

E-mail: david.rodriguez@iisgm.com

*These authors contributed equally to the study

ABSTRACT

Enterobacter cloacae complex (ECC) encompasses heterogenic genetic clusters of species that have been associated with nosocomial outbreaks. These species may host different acquired antimicrobial susceptibility patterns and their identification is challenging. DNA-based techniques are laborious and require specific equipment. MALDI-TOF MS has showed low accuracy for the discrimination of ECC species. The aim of this study is to develop machine learning predictive models using MALDI-TOF MS and new diagnostic technologies like Fourier-Transform Infrared Spectroscopy (FTIR-S) for species-level identification of these species.

A total of 163 ECC clinical isolates were included in the study: 47 for the predictive model development and internal validation and 126 for external validation. All spectra obtained by MALDI-TOF MS and FTIR-S were processed using Clover MS Data Analysis software. Two models were created: Model A for differentiate six ECC species and Model B for *E. hormaechei* subspecies.

For MALDI-TOF MS spectra, Model A identified correctly 96.0% of isolates using Random Forest (RF) algorithm, and Model B identified 94.1% using Support Vector Machine (SVM). Regarding FTIR-S, Model A identified 73.0% of isolates by RF and Model B 72.5%. Two new predictive models were created for FTIR-S: Model C for discrimination of *E. hormaechei* from non-*E. hormaechei* (87.3% identification, RF) and Model D for differentiation among non-*E. hormaechei* species (62.7% identification, RF).

MALDI-TOF MS combined with machine learning tools could be a rapid and accurate method for species-level identification within the ECC. FTIR-S differentiated general groups of the ECC although discrimination of non-*E. hormaechei* species was poor.

INTRODUCTION

Enterobacter spp. is a facultative anaerobic Gram-negative genus that can be found as a natural commensal in the gut microbiome of mammals (1). Several species have been associated with nosocomial outbreaks causing urinary tract infection, skin and soft tissues infection, pneumonia and bacteraemia (2, 3). *Enterobacter cloacae* complex (ECC) is of particular clinical interest. This group is composed of 13 heterogenic genetic clusters according to *hsp60* gene sequencing: *E. asburiae* (cluster I), *E. kobei* (cluster II), *E. hormaechei* subsp. *hoffmannii* (cluster III), *E. roggenkampii* (cluster IV), *E. ludwigii* (cluster V), *E. hormaechei* subsp. *oharae* and subsp. *xiangfangensis* (cluster VI), *E. hormaechei* subsp. *hormaechei* (cluster VII), *E. hormaechei* subsp. *steigerwaltii* (cluster VIII), *E. bugandensis* (cluster IX), *E. nimipressuralis* (cluster X), *E. cloacae* subsp. *cloacae* (cluster XI), *E. cloacae* subsp. *dissolvens* (cluster XII) and an heterogeneous group of *E. cloacae* sequences considered as cluster XIII. However, the taxonomy of this genus is still under debate (4, 5). In fact, *E. aerogenes* has been recently reclassified into the *Klebsiella* genus as *K. aerogenes* (6). A more comprehensive study based on whole genome sequencing (WGS) data from ECC isolates yielded a redistribution of the species defined by *hsp60* sequencing (5) into 22 clades (7) and allowed the characterization of new ECC species (8).

Discrimination of the ECC at the species level is usually performed by DNA-based methods. The most commonly targeted gene is *hsp60*, although multilocus sequence analysis (MLSA) and WGS have also been applied (5, 9, 10). DNA-based diagnostic methods techniques are laborious and require specific equipment. Therefore, new emerging techniques such as MALDI-TOF MS (Matrix Assisted Laser Desorption/Ionization Time-of-flight Mass Spectrometry) have been proposed as an alternative to molecular methods. MALDI-TOF MS has shown to be an excellent methodology for bacterial identification. It can easily identify *E. cloacae* complex

isolates but it showed low discrimination power for the species in this group when using standard analysis (11, 12).

Fourier-Transform Infrared Spectroscopy (FTIR-S) has also emerged recently as a rapid method for bacterial typing (13), although there are still few studies for the discrimination of close-related species. These new technologies, coupled to a machine learning approach has a great potential for fast and cost-effective bacterial typing and discrimination of closely related species.

The aim of this study is to develop and validate prediction models for the differentiation of the species within the ECC using MALDI-TOF MS and FTIR-S spectra combined with machine learning tools. This task is important because of their diverse implications in human pathologies and their involvement in nosocomial outbreaks (4). Besides, *E. hormaechei*, the most commonly encountered ECC species in the clinical settings, has been correlated with enhanced acquisition of antimicrobial resistance mechanisms and the expression of virulence factors (14-16).

MATERIALS AND METHODS

Bacterial isolates

Overall, 163 clinical isolates belonging to the ECC and nine *K. aerogenes* (formerly *E. aerogenes*) were analysed. They belonged to a surveillance study of antimicrobial resistances were collected at the Hospital Universitario Ramón y Cajal (Madrid, Spain) between 2005 and 2018 and identified by partial sequencing of the *hsp60* gene (17). All the analysed isolates host different carbapenemases.

All isolates were incubated overnight at 37ºC and metabolically activated after three subcultures on Columbia Blood Agar (bioMérieux, Marcy l'Etoile, France).

Isolates were analysed with MALDI-TOF MS and FTIR-S at the Hospital General Universitario Gregorio Marañón (Madrid, Spain).

Bacterial identification and spectra acquisition using MALDI-TOF MS

Identification was performed using the MBT Smart MALDI Biotyper (Bruker Daltonics, Bremen, Germany). All strains were spotted in duplicate onto the MALDI target plate and overlaid with 1 µl of 70% formic acid. After drying at room temperature, spots were covered with 1 µl HCCA matrix, prepared according to the manufacturer's indications (Bruker Daltonics) and allowed to dry. Two spectra in the range of 2,000-20,000 Da were obtained on each spot, resulting in 4 spectra per isolate.

Spectra acquisition using Fourier-Transform Infrared Spectroscopy

All strains were also analysed by FTIR-S using an IR Biotyper (Bruker Daltonics) according to the manufacturer's instructions. The procedure consisted on the resuspension of a 1 µl loopful of bacteria in 50 µl of 70% ethanol. After a brief vortexing step, 50 µl of sterile water were added and the mixture was vortexed again. Then, 15 µl of this preparation were transferred onto the 96-well silicon plate and allowed to dry at 40ºC for 20 minutes approximately. Each isolate was analysed in triplicates -technical replicates- and two standards (Bruker Infrared Test Standard 1 and 2, Bruker Daltonics) were included in each run and only spectra with valid target quality controls were further analysed.

Data processing of MALDI-TOF MS protein spectra and development of predictive models

After spectra acquisition, 47 isolates representing all *Enterobacter* species were used as a test set for the evaluation of the predictive model for species differentiation (Table 1). MALDI-TOF MS protein spectra were processed with Clover MS Data Analysis software (Clover Biosoft, Granada, Spain). A pre-processing pipeline was applied to

the MALDI-TOF MS spectra that consisted of: 1) smoothing (Savitzky-Golay Filter: window length=11; polynomial order=3) and baseline substraction (Top-Hat filter method with factor=0.02); 2) creation of an average spectrum per isolate; 3) alignment of average spectra from different isolates (shift: medium; constant tolerance: 2 Da; linear mass tolerance: 600 ppm); 4) normalization by Total Ion Current (TIC).

The processed spectra were included in two peak matrices used to develop two predictive models: 1) Model A, aiming for the differentiation of the 6 main species -*E. asburiae* (cluster I), *E. kobei* (cluster II), *E. hormaechei* (clusters III, VI and VIII considered together), *E. roggenkampii* (cluster IV), *E. ludwigii* (cluster V) and *E. cloacae* subsp. *dissolvens* (cluster XII)- and 2) Model B, for the differentiation of the three *E. hormaechei* clusters (Figure 1). These two peak matrices were used as input data for three machine learning algorithms: Partial Least Square-Discriminant Analysis (PLS-DA), Support Vector Machine (SVM) and Random Forest (RF). Internal k-fold cross validation using k=10 was performed.

Data processing of FTIR-S spectra and development of predictive models

Following the same approach applied for the analysis of the *Enterobacter* isolates with MALDI-TOF MS, the same group of 47 isolates was used as test set for the creation of a predictive model based on FTIR-S spectra. For this purpose, each isolate was analysed in triplicates - technical replicates - in two consecutive days - biological replicates -. Therefore, the total number of spectra per isolate was six. The spectra were processed using Clover MS Data Analysis with the following pipeline: 1) Standard Normal Variation; 2) Smoothing (Savitzky-Golay Filter: window length=9; polynomial order=2; second derivative); 3) Maximum absorbance normalization.

As for MALDI-TOF MS spectra analysis, discrimination of the six major ECC species was attempted, as well as differentiation of the three *E. hormaechei* clusters.

For this purpose, the algorithms SVM and RF were also applied using internal k-fold cross validation of k=10.

Classification of the *E. cloacae* complex isolates using the developed models for MALDI-TOF MS spectra

MALDI-TOF MS spectra from the external validation set (n=126) were processed using the pipeline described above and blindly classified by the predictive models using Clover MS Data Analysis software.

Classification of the *E. cloacae* complex isolates using the developed model for FTIR-S spectra

FTIR-S spectra from the external validation set (n=126) were processed using the pipeline described above and blindly classified by the predictive models using Clover MS Data Analysis software.

Ethics statement

The Ethics Committee of the Gregorio Marañón Hospital (CEIm) evaluated this project and considered that all the conditions for waiving informed consent were met, since the study was conducted with microbiological samples and not with human products.

RESULTS

MALDI-TOF MS predictive models

Different models were developed by the implementation of the available algorithms (PLS-DA, SVM and RF) to the test set. These models were validated internally by cross-validation (k-fold=10) and externally by the automatic classification of the 126 protein spectra included in the validation set (Table 1). For Model A evaluation, Random Forest (RF) yielded the highest rates of correct classification for the most

common *Enterobacter* species (Model A): 95.7% and 96.0% for internal and external validation, respectively (Table 2). The RF plot for internal validation is showed in Figure 2. Only 5/126 isolates were misidentified by Model A using RF (Table S1): *E. hormaechei* identified as *E. roggenkampii* (n=1), *E. roggenkampii* identified as *E. hormaechei* (n=2), *E. roggenkampii* identified as *E. kobei* (n=1) and *E. kobei* identified as *E. roggenkampii* (n=1). PLS-DA and SVM algorithms also provided high rates of correct classification in the internal validation although PLS-DA showed lower discriminative power in the external validation (70.6%) -Table 2-.

Regarding Model B, designed to differentiate *E. hormaechei* clusters III, VI and VIII, the implementation of PLS-DA yielded 77.8% correct classification in the internal validation (Table 2). Thus, this model was not considered suitable for external validation. However, SVM provided a correct classification of 94.4% and 94.1% of the isolates in the internal and external validation, respectively (Tables S2-S3). Three *E. hormaechei* isolates out of 51 were misidentified: one strain from cluster VIII was identified as cluster VI, and two strains from cluster VI were identified as cluster VIII. All cluster III strains were correctly classified (Table S3). Besides, RF algorithm reached 94.4% of correct classification (Figure 3) and allowed species-level identification of 80.4% of the isolates from the validation set (Table S4).

FTIR-S predictive models

The implementation of FTIR-S allowed the development of classification models based on SVM and RF. Internal cross-validation (k=10) and external validation of Model A using RF reached 98.5% (Table S5) and 73.0% correct classification, respectively. The rate of accuracy yielded by SVM was lower in both internal and external validation (Table 3). For Model B, correct discrimination of the *E. hormaechei* clusters was achieved in 100% and 72.5% of the cases in the internal and external validation, respectively. In that model, also SVM obtained lower identification rates.

Due to the low percentages of correct classification obtained in the external validation of Model A and Model B, two more predictive models were evaluated for FTIR-S spectra: 1) Model C: for the differentiation of *E. hormaechei* vs non-*E. hormaechei* isolates; 2) Model D: for discrimination among non-*E. hormaechei* species. Applying RF algorithm, Model C and D reached 99.1% and 99.0% correct classification of the isolates in the internal validation and 87.3% and 62.7% in the external validation, respectively (Table 4).

DISCUSSION

In this study, the implementation of classifying algorithms to MALDI-TOF spectra allowed the correct species assignment of 96.0% isolates belonging to the main ECC species and 94.1% of the *E. hormaechei* clusters III, VI and VIII. On the other hand, the discrimination power of FTIR S for the same categories was 73.0% and 72.5% and in 87.3% of the cases this technology discriminated correctly between *E. hormaechei* and non-*E. hormaechei* isolates.

Poor discrimination of *E. cloacae* complex species by MALDI-TOF MS has been previously reported either by using commercial (11, 17) and enriched, in-house databases (18). However, a recent study from a research group with extensive experience in MALDI-TOF MS and the creation of expanded libraries reported 92.0% correct species-level identification by implementing a specific in-house library enriched with well-characterized ECC strains and correct discrimination of 97.0% *E. hormaechei* isolates (19). This approach can be useful for discrimination of close-related species but the construction of a database is cumbersome and requires highly trained personnel.

A more rapid approach for the discrimination of ECC species is the implementation of classification models. A previous study already demonstrated that

the application of hierarchical clustering to protein spectra provided a clear discrimination of *E. hormaechei* isolates (clusters III, VI and VIII) from the rest of the species from the complex in the study developed by Wang et al. (18).

In this study, different supervised algorithms were implemented for the discrimination of six main ECC species and for differentiation of the three *E. hormaechei* clusters (III, VI and VIII). For MALDI-TOF MS spectra, RF algorithm showed highest discrimination capacity both in the cross-validation (95.7%) and in the external validation of the model (96.0%) and SVM demonstrated a high suitability for the differentiation of the *E. hormaechei* clusters (94.4% correct classification in the cross-validation and 94.1% in the external validation). Although a lower specificity in the classification of *E. hormaechei* subspecies was found, the two developed models have shown high discrimination between species of the ECC. Therefore, MALDI-TOF MS combined with machine learning tools could be a rapid and accurate method for species-level identification within the ECC. Only five (3.9%) isolates from the main species and three (5.8%) from the *E. hormaechei* clusters were misclassified, which shows the high specificity of this methodology.

The implementation of FTIR-S and RF modelling allowed the correct classification of 98.5% -cross-validation- and 73.0% -external validation- of the spectra from the main species and 100% and 72.5% of the *E. hormaechei* clusters. The low rate of correct classifications encouraged us to developed different models for the discrimination of *E. hormaechei* isolates from the rest of the ECC species. In this case, 99.1% and 87.3% correct classifications were achieved in the cross-validation and the external validation, respectively. Discrimination of non-*E. hormaechei* species was low using RF (62.7%) and even lower with SVM. Our hypothesis is that FTIR-S accuracy is so high that it may discriminate the analysed isolates based on their sequence type or their genetic lineage since this methodology is designed mainly for bacterial typing.

Unfortunately, the strains of this study were not characterized at that level. This approach has been previously observed in the study by Vogt et al. (20).

One of the limitations of this study was the fact that all isolates host carbapenemases because they source from a previously analysed collection (17). Although this feature does not hamper the results from the present work, further studies will be performed including both resistant and susceptible isolates.

Although further algorithms and modelling is needed in order to implement FTIR-S for the discrimination of *E. cloacae* complex species, MALDI-TOF MS has demonstrated to be a rapid and cost-effective method for this task. The use of spectra analysis tools is becoming user-friendly and easy to apply and its use may provide species-level identification in a fast and inexpensive way.

ACKNOWLEDGMENTS

This work was supported by the projects PI15/01073 and PI18/00997 from the Health Research Fund (FIS. Instituto de Salud Carlos III. Plan Nacional de I+D+I 2013-2016) of the Carlos III Health Institute (ISCIII, Madrid, Spain) partially financed by the European Regional Development Fund (FEDER) ‘A way of making Europe’.

MM was funded by the Community of Madrid (Programa de Garantía Juvenil, PEJD-2017-PRE/BMD-5106) and DRT with a postdoc contract from the Intramural Funding Program of the IiSGM. BRS (CPII19/00002) is a recipient of a Miguel Servet contract supported by the FIS.

AUTHOR CONTRIBUTIONS

Ana Candela: experimental part, formal analysis, data collection, validation, visualization, writing – original draft preparation and review/editing. Miriam Mateos and

Alicia Gómez-Asenjo: experimental part, formal analysis and data collection. Manuel J. Arroyo, Gema Méndez, and Luis Mancera: data analysis, validation, writing – original draft preparation and review/editing. Marta Hernández-García, Rosa del Campo: experimental part, formal analysis and writing, submission of isolates, original draft preparation and review/editing. Juan de Dios Caballero, Laura Martínez-García, Desirée Gijón, María Isabel Morosini, Patricia Ruiz-Garbajosa, Rafael Cantón and Patricia Muñoz: validation, writing and review/editing. Emilia Cercenado: conceptualization, formal analysis, validation, writing and review/editing. David Rodríguez-Termporal: conceptualization, formal analysis, validation, original draft preparation and review/editing; Belén Rodríguez-Sánchez: conceptualization, project administration, formal analysis, supervision, validation, visualization, original draft preparation and review/ editing.

FIGURE LEGENDS

Figure 1. *E. cloacae* complex identification algorithm according to models developed.

Figure 2. MALDI-TOF MS spectra in Random Forest distance plot for Model A.

Figure 3. MALDI-TOF MS spectra in Random Forest distance plot for Model B.

REFERENCES

1. Mezzatesta ML, Gona F, Stefani S. 2012. Enterobacter cloacae complex: clinical impact and emerging antibiotic resistance. Future Microbiol 7:887-902.
2. Akbari M, Bakhshi B, Najar Peerayeh S. 2016. Particular Distribution of Enterobacter cloacae Strains Isolated from Urinary Tract Infection within Clonal Complexes. Iran Biomed J 20:49-55.
3. Kremer A, Hoffmann H. 2012. Prevalences of the Enterobacter cloacae complex and its phylogenetic derivatives in the nosocomial environment. Eur J Clin Microbiol Infect Dis 31:2951-5.
4. Davin-Regli A, Lavigne JP, Pages JM. 2019. Enterobacter spp.: Update on Taxonomy, Clinical Aspects, and Emerging Antimicrobial Resistance. Clin Microbiol Rev 32.
5. Hoffmann H, Roggenkamp A. 2003. Population genetics of the nomenspecies Enterobacter cloacae. Appl Environ Microbiol 69:5306-18.
6. Tindall BJ, Sutton G, Garrity GM. 2017. 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 67:502-504.
7. Sutton GG, Brinkac LM, Clarke TH, Fouts DE. 2018. 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. F1000Res 7:521.
8. Wu W, Feng Y, Zong Z. 2019. Characterization of a strain representing a new Enterobacter species, Enterobacter chengduensis sp. nov. Antonie Van Leeuwenhoek 112:491-500.
9. Singh NK, Bezdan D, Checinska Sielaff A, Wheeler K, Mason CE, Venkateswaran K. 2018. Multi-drug resistant Enterobacter bugandensis species isolated from the International Space Station and comparative genomic analyses with human pathogenic strains. BMC Microbiol 18:175.
10. Hoffmann H, Stindl S, Ludwig W, Stumpf A, Mehlen A, Heesemann J, Monget D, Schleifer KH, Roggenkamp A. 2005. Reassignment of enterobacter dissolvens to Enterobacter cloacae as E. cloacae subspecies dissolvens comb. nov. and emended description of Enterobacter asburiae and Enterobacter kobei. Syst Appl Microbiol 28:196-205.
11. Pavlovic M, Konrad R, Iwobi AN, Sing A, Busch U, Huber I. 2012. A dual approach employing MALDI-TOF MS and real-time PCR for fast species identification within the Enterobacter cloacae complex. FEMS Microbiol Lett 328:46-53.
12. De Florio L, Riva E, Giona A, Dedej E, Fogolari M, Cella E, Spoto S, Lai A, Zehender G, Ciccozzi M, Angeletti S. 2018. MALDI-TOF MS Identification and Clustering Applied to Enterobacter Species in Nosocomial Setting. Front Microbiol 9:1885.
13. Quintelas C, Ferreira EC, Lopes JA, Sousa C. 2018. An Overview of the Evolution of Infrared Spectroscopy Applied to Bacterial Typing. Biotechnol J 13.
14. Garinet S, Fihman V, Jacquier H, Corvec S, Le Monnier A, Guillard T, Cattoir V, Zahar JR, Woerther PL, Carbonnelle E, Wargnier A, Kerneis S, Morand PC. 2018. Elective distribution of resistance to beta-lactams among Enterobacter cloacae genetic clusters. J Infect 77:178-182.

355 15. Barnes AI, Paraje MG, del CBP, Albesa I. 2001. Molecular properties and metabolic
356 effect on blood cells produced by a new toxin of Enterobacter cloacae. Cell Biol Toxicol
357 17:409-18.
358 16. Paauw A, Caspers MPM, Leverstein-van Hall MA, Schuren FHJ, Montijn RC, Verhoef J,
359 Fluit AC. 2009. Identification of resistance and virulence factors in an epidemic
360 Enterobacter hormaechei outbreak strain. Microbiology (Reading) 155:1478-1488.
361 17. Mateos M, Hernandez-Garcia M, Del Campo R, Martinez-Garcia L, Gijon D, Morosini
362 MI, Ruiz-Garbajosa P, Canton R. 2020. Emergence and Persistence over Time of
363 Carbapenemase-Producing Enterobacter Isolates in a Spanish University Hospital in
364 Madrid, Spain (2005-2018). Microb Drug Resist.
365 18. Wang YQ, Xiao D, Li J, Zhang HF, Fu BQ, Wang XL, Ai XM, Xiong YW, Zhang JZ, Ye CY.
366 2018. Rapid Identification and Subtyping of Enterobacter cloacae Clinical Isolates Using
367 Peptide Mass Fingerprinting. Biomed Environ Sci 31:48-56.
368 19. Godmer A, Benzerara Y, Normand AC, Veziris N, Gallah S, Eckert C, Morand P, Piarroux
369 R, Aubry A. 2021. Revisiting Species Identification within the Enterobacter cloacae
370 Complex by Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass
371 Spectrometry. Microbiol Spectr 9:e0066121.
372 20. Vogt S, Loffler K, Dinkelacker AG, Bader B, Autenrieth IB, Peter S, Liese J. 2019. Fourier-
373 Transform Infrared (FTIR) Spectroscopy for Typing of Clinical Enterobacter cloacae
374 Complex Isolates. Front Microbiol 10:2582.

375

376

377

378

379

380

381

382

383

384

385

386

387

Table 1. Number of isolates of each species used for model creation and validation for MALDI-TOF MS and FTIR S.

Species included in the study	Number of strains used for model development	Number of strains used for external validation
K. aerogenes	6	3
E. asburiae	6	1
E. dissolvens	1	0
E. hormaechei III	6	13
E. hormaechei VI	6	22
E. hormaechei VIII	6	16
E. kobei	6	9
E. ludwigii	3	0
E. roggenkampii	7	62
Total	**47**	**126**

Table 2. Internal and external validation for Model A and Model B in all algorithms used for MALDI-TOF MS spectra.

	Model A		Model B	
Algorithm	**Internal validation**	**External validation**	**Internal validation**	**External validation**
PLS-DA	91.5%	70.6%	77.8%	NR
SVM	91.5%	92.1%	94.4%	94.1%
RF	95.7%	96.0%	94.4%	80.4%

PLS-DA, Partial Least Square Discriminant Analysis; SVM, Support Vector Machine; RF, Random Forest; NR, Not Realized

Table 3. Internal and external validation for Model A and B for all algorithms used for FTIR-S spectra.

	Model A		Model B	
Algorithm	**Internal validation**	**External validation**	**Internal validation**	**External validation**
SVM	95.1%	69.8%	99.2%	72.5%
RF	98.5%	73.0%	100%	72.5%

Table 4. Internal and external validation for new FTIR-S models.

	Model C		Model D	
Algorithm	**Internal validation**	**External validation**	**Internal validation**	**External validation**
SVM	96.7%	88.1%	95.1%	56.0%
RF	99.1%	87.3%	99.0%	62.7%

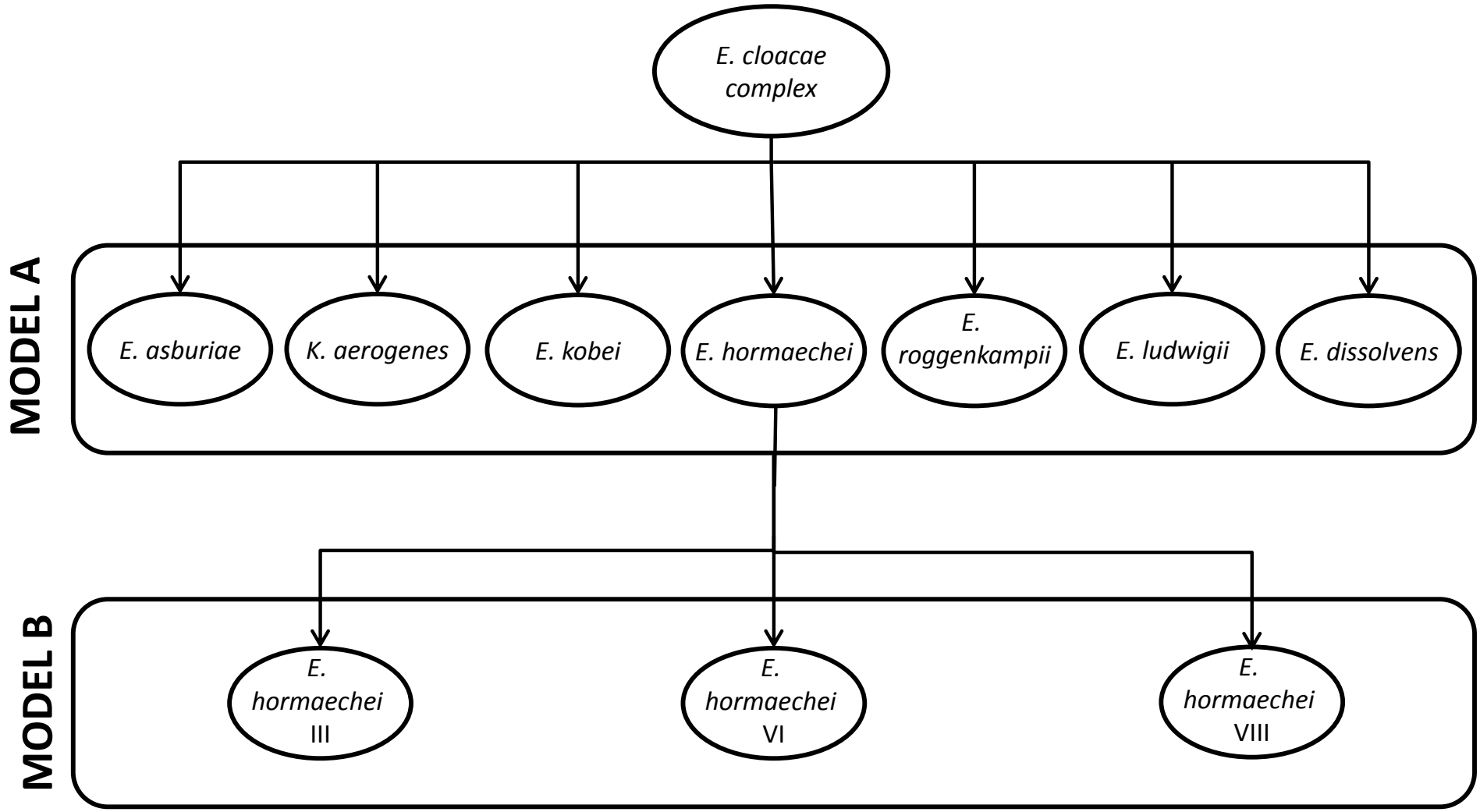

Figure 1. *E. cloacae* complex identification algorithm according to models developed.

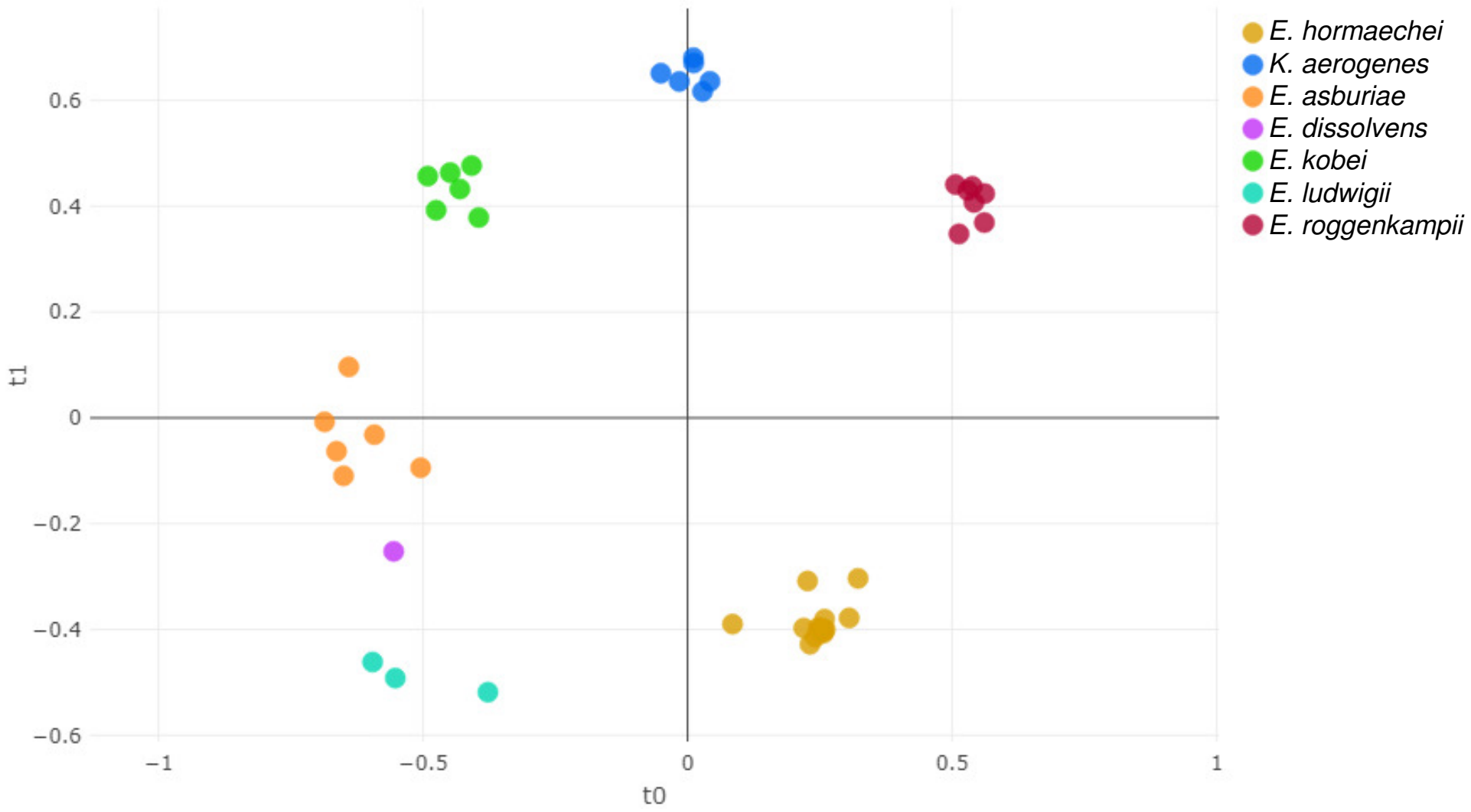

Figure 2. MALDI-TOF MS spectra in Random Forest distance plot for Model A.

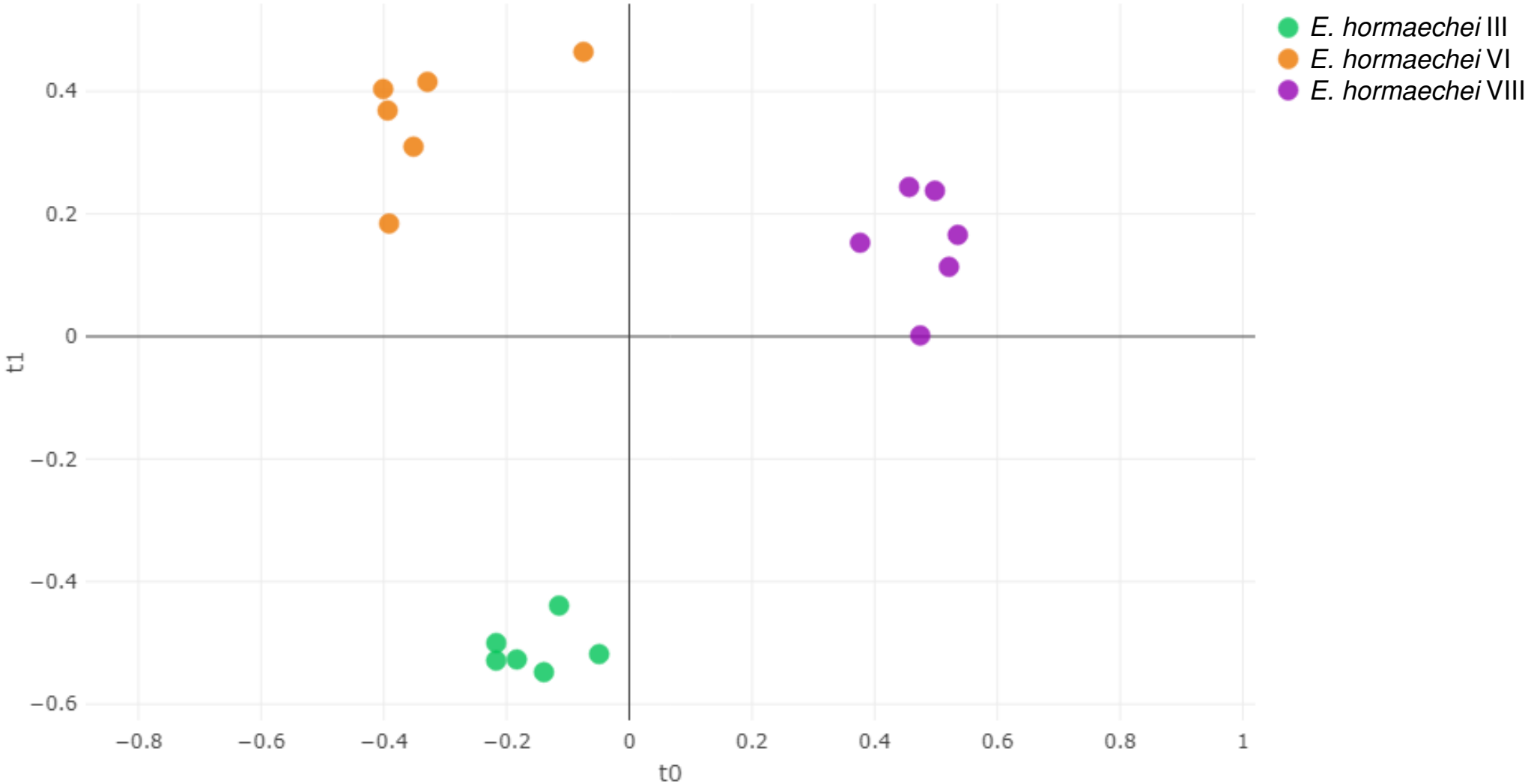

Figure 3. MALDI-TOF MS spectra in Random Forest distance plot for Model B.