Molecular epidemiology of Clostridium neonatale and its relationship with the occurrence of necrotizing enterocolitis in preterm neonates

M. Hosny, E. Baptiste, A. Levasseur, B. La Scola

Association between *Clostridium neonatale* and the occurrence of Necrotizing Enterocolitis in preterm neonates

Michel Hosny[1], Emeline Baptiste[1], Anthony Levasseur[1] and Bernard La Scola[1]*.

[1]Aix-Marseille Université UM63, Institut de Recherche pour le Développement IRD 198, Assistance Publique – Hôpitaux de Marseille (AP-HM), Microbes, Evolution, Phylogeny and Infection (MEΦI), Institut Hospitalo-Universitaire (IHU) - Méditerranée Infection, France (M. Hosny, E. Baptiste, A. Levasseur and B. La Scola).

* **Correspondance:** Bernard La Scola, MD, PhD, Pôle des Maladies Infectieuses, Aix-Marseille Université, IRD, Assistance Publique – Hôpitaux de Marseille (AP-HM), Microbes, Evolution, Phylogeny and Infection (MEΦI), Institut Hospitalo-Universitaire (IHU) - Méditerranée Infection, France, 19-21 Boulevard Jean Moulin, 13385 Marseille Cedex 05, France. Telephone: +33 4 91 32 43 75, Fax: +33 4 91 38 77 72, E-mail: bernard.la-scola@univ-amu.fr

Running title:

Keywords: Necrotizing enterocolitis; Clostridium neonatale; rpoB-based specific quantitative real-time polymerase chain reaction; core-genome phylogeny; clonal lineage; whole genome sequence

Abstract

Clostridia are among taxa most frequently identified from necrotizing enterocolitis preterm neonates' stool samples, especially *Clostridium butyricum*. Recently, *Clostridium neonatale* has also been detected from epidemic cases, but using a culture-based approach, we were unable to confirm this discovery in a local cohort. In order to investigate this link by a molecular approach, a specific *rpoB*-based quantitative real-time polymerase chain reaction was developed to detect *C. neonatale* directly from patients' stool specimens. Design of this *rpoB*-based quantitative real-time polymerase was based on the genomic analysis of seven clinical isolates of *C. neonatale*. It was tested on stool samples from 88 preterm neonates with necrotizing enterocolitis and 71 matched controls. *C. neonatale* was significantly more prevalent in stools from preterm neonates with necrotizing enterocolitis than in controls (respectively 30/88 [34%]; *vs.* 9/71 [13%]; $p = 0.003$). Whole-genome analysis also allowed the identification of three genomic clusters of *C. neonatale*. This clustering was associated with a geographical location regardless of isolation in the NEC or control, suggesting asymptomatic carriage. Although less prevalent than *C. butyricum* in our cohort, *C. neonatale* is significantly associated with the occurrence of necrotizing enterocolitis.

Keywords: Necrotizing enterocolitis, *Clostridium neonatale*, *rpoB*-based specific quantitative real-time polymerase chain reaction, core-genome phylogeny, clonal lineage.

Introduction

Necrotizing enterocolitis (NEC) is a major cause of morbidity and mortality in neonatal intensive care units (NICUs), especially for very low birth weight born (VLBW) [1]. Over the past decade, most studies have focused on intestinal immaturity, feeding strategies and the composition of the gut microbiota, as NEC is recognized as a multifactorial disease [1,2]. Approaches of gut microbiota composition determination was performed using molecular-based tools, such as 16S rRNA pyrosequencing, shotgun metagenomic sequencing and specific quantitative real-time polymerase chain reaction [1,3,4]. These tools have been used to describe dysbiosis-linked alterations in neonates' gut microbiota. In addition, several bacterial species have been specifically associated with NEC including *K. pneumoniae*, *E. cloacae*, Uropathogenic *Escherichia coli*, *Clostridium butyricum* and *Clostridium neonatale* [1,4–7].This latter was first isolated in the context of epidemic NEC cases in a Canadian NICU [8]. Pulsed-field gel electrophoresis identified similar banding profile from blood and stool culture isolates [8]. Later, Roze *et al.* reported correlations between nutritional strategies and gut microbiota composition, where *C. neonatale* was identified from first stool of preterm neonates with NEC [9] and its genome sequence was determined [10,11]. A causal relationship between the presence of *C. neonatale* in stool samples and NEC has not yet been clearly demonstrated, but it warrants further studies. Based on a cohort of neonates and controls enrolled to study the involvement of *C. butyricum* with NEC [3,7], we evaluated the association of *C. neonatale* in the occurrence of NEC. This hypothesis was developed by studying the frequency of *C. neonatale* in the neonates' stools using a specific *rpoB*-based quantitative real-time polymerase chain reaction (qPCR), and then comparing the strains isolated by genomic sequencing.

Materials and Methods

Study design and patients

The study was validated by agreements from the ethics committee of the "*Institut Fédératif de Recherche*, IFR48" and the "*Institut Hospitalo-Universitaire*, IHU-2017-007". Written approval was obtained from the parents of all patients [3]. One hundred and fifty-nine stool samples were collected from preterm neonates treated in five neonatal intensive care units (NICUs), whose parents consented to the study. Enrolled samples were obtained from our previous cohort studies, that included 88 patients with NEC and 71 healthy controls [3,7]. None of the patients were subjected to probiotic therapy. Stool samples were collected, if possible, on the day of symptoms onset and stored at -80°C. Patients and controls were assorted by sex, gestational age (under 37 gestational weeks), birth weight, days of life, feeding strategies, mode of delivery and previous antibiotic therapy (Table 1). Routine microbiological analysis was negative for all samples.

Genome sequencing, assembly and annotation of *Clostridium neonatale* isolates

Five *C. neonatale* strains, all from Marseille, isolated in a previous work were sequenced and analyzed [7]. Four were NEC-associated isolates (NEC25, NEC26, NEC32, NEC86) and one was from a control neonate (C25). Strains were cultured on 5% Columbia sheep blood agar (Becton Dickinson®, USA) at 37°C for 48 hours under anaerobic conditions before DNA extraction on EZ1 DNA Tissue Kit (QIAGEN, Germany). Genomic DNA was sequenced on MiSeq Technology (Illumina, Inc., San Diego, CA) using the paired-end applications. NEC86 strain was additionally sequenced by MinION technology (Nanopore, Oxford, UK) [12,13]. SPAdes software was used for NEC86 reads assembly obtained from both sequencing methods [14]. Reads of the other *C. neonatale* genomes, were mapped against NEC86 dataset using the CLC genomics workbench 7 (Qiagen Inc., Valencia, CA). Coding DNA sequences (CDS) and annotation were procured using Prokka software [15]. Finally, sequenced genomes were

deposited to the European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI) under accession numbers: NEC25 (UWJD01000001- 01000003), NEC26 (UICU01000001- 01000003), NEC32 (UICW01000001- 01000003), NEC86 (UICR01000001-01000003), C25 (UICQ01000001- 01000003) (Table 2). Further analysis and comparison was conducted by introducing genome sequences from two strains isolated from a Canadian NICU (LCDC99A005: GCA_001458595.1 and LCDC99A006: GCA_002553455.1) [8,11].

Design of specific qPCR for *Clostridium neonatale* detection

As previously performed for the detection of *C. butyricum* [3], we designed a *C. neonatale* specific qPCR targeting the *rpoB* gene. In a first time, we verified that the *rpoB* gene was presented in unique copy within *C. neonatale* genomes (see above). In a second time, these sequences were aligned with those of 4 pathogenic species using MEGA7 software [16]. Other *Clostridium* genomes included in this analysis were *C. butyricum* E4 str. BoNT E BL5262 (GCA_000182605.1), *Clostridium botulinum* A str. (GCA_000017025.1), *Clostridium perfringens* ATCC 13124 (GCA_000013285.1) and *Clostridioides* (*Clostridium*) *difficile* QCD-66c26 (GCA_000003215.1) as an outgroup at the class level (Clostridia). Primers and probe were designed to detect specifically *C. neonatale rpoB* gene using Primer3 software [17]. Finally, a 151 nucleotides *rpoB* sequence was selected: this region is homogenous among *C. neonatale* strains and highly heterogenic compared to other Clostridia. This molecular construction was validated by a Maximum-Likelihood phylogenetic analysis generated by MEGA7 software (Figure 1) [16]. LightCycler® 480 Probes Master (Roche, Germany) was used to achieve qPCR reactions. The optimized 20 µL qPCR mix contained: master mix (10 µL), primers (0.5 µL, 20 nM), probe (0.5 µL, 5 nM), water (3.5 µL) and 5 µL of DNA. The qPCR

cycling protocol started with one cycle of 50°C for 2 minutes, followed by one activation cycle (95°C for 5 min) and 40 amplification cycles (95°C for 1 second; Tm= 60°C for 30 seconds). Tubes were deposed in a thermal cycler CFX96 Touch™ (Bio-rad®, France). Sequences of the qPCR system were designed as follows: forward primer: AATAGTTGATAAAGAACACGGTAGAGC, reverse primer: TAGCGGTTCTTTTGCTTGAG and probe (FAM-TGCAGATGAAGAAGATCAGTTTTT-TAMRA) (Supplementary Figure S1). First, the qPCR system was tested on a collection of 79 bacterial species associated with the gut microbiota, including five locally isolated *C. neonatale* and 24 *Clostridium* species (Supplementary Table S1). Total genomic DNA was extracted directly from stool samples using the NucleoSpin Tissue Kit (Macherey-Nagel, Hoerdt, France) as previously described [7].

Statistical analysis

Statistical analysis was performed using SPSS® statistics 2016 (IBM, NY, USA). Mean and standard deviation were used to describe continuous variables. Percentage and number of events were used for quantitative variables. Student t-test or Mann-Whitney *U* test were used to perform two-group comparisons for quantitative variables. The chi-square (Mantel-Haenszel) test was used to perform two-group comparisons for qualitative variables, or the Fisher exact test was used when the expected count was less than 5. A multivariate analysis using logistic regression was performed to identify independent risk factors for NEC. Variables clinically relevant and associated ($p < 0.10$) with NEC in univariate analysis were used in the model. A p value of < 0.05 was considered statistically significant.

Draft genome-based and core-genome phylogeny

For draft genome-based phylogeny, only mapped genomes were included. Scaffolds were concatenated, and aligned using the MAFFT software [18]. For core-genome analysis, a dataset

comprising seven genomes were generated. Orthologous proteins were obtained from ProteinOrtho software using the following parameters: identity 50%, coverage 60% and E-value 1.10^{-10}. CDSs of core-genes were inferred from the pan-genome, then concatenated and aligned using a Python script. The phylogenetic trees were generated using the maximum-likelihood method within PhyML [19] and edited by TreeGraph 2 software [20]. BLAST was used to identify hemolysin (A, B, C), beta-hemolysin sequences, *dlt* operon (*dlt A*, *B*, *C*, *D*) and the Clusters of Orthologous Groups (COG) identification (E-value $1e^{-03}$, coverage 0,7 and identity percent 30% for this later).

Results

Evaluation of patients features and *Clostridium neonatale* specific *rpoB*-based qPCR system

There was no statistically significant difference between NEC and control groups regarding gestational age, days of life, sex ratio, birth weight and feeding strategies. Only vaginal delivery was significantly more prevalent in controls than in NEC (37/71 [52%]; *vs* 32/88 [36%]; $p < 0.05$). This statistical analysis improved the rationality of clinical features between NEC and healthy controls. Specificity of primers and probe were validated due to positive amplification for all *C. neonatale* strains (5/5) and the absence of amplification for other bacterial strains tested isolated from gut microbiota (0/74). This result excluded cross-reaction between *C. neonatale* qPCR system and bacterial species inhabited within the gut microbiota. *C. neonatale* was more frequently detected in stools from preterm neonates with NEC than in controls (respectively 30/88 [34%]; *vs.* 9/71 [13%]; $p = 0.003$). The presence of *C. neonatale* in stool samples was significantly associated with the occurrence of NEC. Statistical results are summarized in Table 1. Means of *C. neonatale* qPCR cycle thresholds (Ct) were statistically non-significant between NEC and controls (26.9 and 25.4 respectively, p value = 0.5292, unpaired student t test),

suggesting that the density of *C. neonatale* is similar in both cohorts. The prevalence of *C. neonatale* was also evaluated with that of *C. butyricum* reported in Hosny *et al.* [7]. In brief, the prevalence of *C. butyricum*, evaluated by both culture and qPCR, were assessed from each cohort and compared with *C. neonatale* frequency of detection (present work). In the majority of NEC cases, *C. neonatale* was frequently detected in association with *C. butyricum* when compared to controls (respectively 20/30, [66.7%] *vs* 1/9, [11.1%], p=0.003) (Supplementary Table S2).

Genomic analysis of *Clostridium neonatale*

Final draft genomes of *C. neonatale* isolates sequenced herein consist of 3 scaffolds. The average length of included genomes was 4,431,047 bp, NEC86 had the longest genome length (4,739,641 bp) and NEC26 the shortest one (4,282,644 bp). GC content varied between 28.4% (LCDC99A006) and 28.9% (NEC26). Predicted CDSs ranged between 3,795 genes (LCDC99A005) and 4,188 genes (NEC86) with genes average of 3,945. General features of these genomes are summarized in Table 2. The draft genome-based phylogeny of *C. neonatale* isolates enabled the identification of three clusters, revealing clonality among isolates from the same NICU by following: (1) strain isolated from NICU-1 (LCDC99A005), (2) strain isolated from NICU-2 (NEC86) and (3) strains isolated from NICU-3 (NEC25, NEC26, NEC32 and C25) (Figure 2A). Furthermore, core-genome analysis distinguished the same clustering, where the strict pangenome consisted of 5,150 genes, including 3,157 core-genes (Figure 2B). Predicted COG categories were homogenous in all *C. neonatale* genomes. The COGs encoding carbohydrate metabolism and transport, and therefore general functional prediction (G and R), were over-represented, on the contrary, no treatment or modification of RNA, and therefore the

155 structure and dynamics of chromatin were identified (A and B) (Supplementary Figure S2 and
156 Table S3).

157 **Identification of virulent and unique genes**

158 Comparative analyses identified hemolysin-encoding genes in all strains as follows:
159 hemolysin A (coverage: 100%, identity: 84% to 92%), two protein sequences encoded for
160 hemolysin B (coverage: 100%, identity: 44% to 99%) for the first protein and (coverage: 99%,
161 identity: 76% to 77%) for the second one, hemolysin C (coverage: 96%, identity: 81%) and beta-
162 hemolysin (coverage: 100%, identity: 92%). We also identified genes encoding for the secretion
163 of *C. difficile* toxin A and toxin B. Moreover, sequences of *dlt* operon (*dlt A*, *B*, *C*, *D*) were not
164 detected in all draft genomes of *C. neonatale* (Supplementary Table S4). Unique genes also
165 existed and mostly represented by hypothetical proteins, as follows: NEC25 (1/4, 25%), NEC26
166 (77/113, 68%), NEC32 (28/39, 72%), C25 (2/3, 67%), NEC86 (139/212, 65.5%) and
167 LCDC99A005 (31/59, 52.5%) LCDC99A006 (Supplementary Table S5).

Discussion

Dysbiosis remains a main risk factor inducing the establishment of NEC, by promoting the translocation of pathogenic bacteria. Specific microorganisms were involved; the predominance of γ-proteobacteria, generating excessive inflammatory response and irregular colonization by strictly anaerobic bacteria, including Clostridia [1]. Multidisciplinary approaches suspected the link between this class and the occurrence of NEC, especially *C. butyricum*, *C. neonatale* and *C. perfringens* [1,21]. To date, *C. butyricum* is the unique *Clostridium* species among the cases of Clostridia-associated NEC clearly described in correlation with NEC [3,7,22]. However, several studies have already reported the involvement of *C. neonatale* in NEC disease [1,8]. Using a culture-based strategy, we observed that the prevalence of *C. neonatale* in stool samples from patients with NEC was not significantly different from that of controls (respectively 4/88 [4.5%]; *vs.* 1/71 [1.14%]; $p = 0.26$) [7]. During the review process of this previous work, a reviewer proposed that this discrepancy could be due to the use of heat-shock-based protocol. If it is supposed to kill all bacteria species except those spore-forming like *C. neonatale,* we could not exclude that the method skips the isolation of *in vivo* non-spore-form of Clostridia species. Therefore, we decided to investigate our cohort using *C. neonatale* specific *rpoB*-based qPCR. With a strong phylogeny and taxonomy index, *rpoB* is a conserved gene whose efficacy for the identification of fastidious microorganisms directly from samples has been previously demonstrated [23,24] and which was used for the heterogeneous detection of BoNT producing Clostridia [25] and *C. butyricum* directly from stool samples [3]. A significant frequency of *C. neonatale* in stools from NEC was detected compared to controls (respectively 34% *vs* 13%, $p = 0.003$). In contrast with the study conducted by Rozé *et al.,* in which the authors reported links between feeding strategies and NEC and the abundance of *C. neonatale* in NEC patients [9], the

191 feeding strategy was herein not associated with the development of NEC. The only correlation
192 identified was the vaginal delivery mode, a feature already observed in NEC cases [2]. Beside *C.*
193 *neonatale* and *C. butyricum*, several NEC outbreaks have been described to be associated with
194 other bacterial species [1,2]. A case report described toxin-producing *C. perfringens*, where NEC
195 severity was independent from the α-toxin concentration [26]. Also, nosocomial colonization by
196 *Klebsiella pneumoniae* type 26 and antibiotic typing was used to specify strain characteristics
197 [27]. Furthermore, genetically similar clones of *Enterobacter sakazakii* were distinguished from
198 powdered milk formula and neonates with NEC [28]. Few other bacterial species were suspected
199 in association with NEC like UPEC and *C. paraputrificum* [4,29].
200 The use of WGS allowed us to further compare and characterize isolates even if the number
201 of strains and available genomes was limited. This study disclosed genetic similarity between
202 strains isolated from the same NICU. Through phylogenetic analysis, we identified three
203 distinct clusters, highlighted by their geographical area of isolation, suggesting the same clone
204 of *C. neonatale* spreading in the same NICU-3 as we previously observed with *C. butyricum*
205 [7]. This is in agreement with the discovery of this species as it was firstly isolated from a
206 Canadian NICU (strains LCDC) and clonality between isolates was proved by PFGE from
207 stool and blood cultures [8]. Geographic and temporal clustering of NEC were described in
208 several studies with various etiologies. Temporal clustering was reported by Faustini *et al.* in
209 diverse NICUs [30]. The similarity between control (C25) and NEC-associated *C. neonatale* in
210 cluster [3] presupposes the existence of asymptomatic carriage. This same kind of mechanism
211 was described in case of *C. difficile*-associated pseudomembranous colitis [31] and we
212 suggested a similar mechanism for *C. butyricum* in neonates [7]. It should be noted that there
213 appears to be a strong association between *C. neonatale* and neonates/children as it is a species

214 that has never been isolated from adults in our microbiota studies [32]. On the contrary, *C.*
215 *neonatale* has been identified by sequencing of 16S rRNA in the microbiota of young children
216 with a strong association of developing asthma [33].

217 If the association of Clostridia with NEC was increasingly reported, the pathogeny of this
218 disease remains elusive in spite production of toxins is usually suggested, as in the case of *C.*
219 *difficile* infections (CDI) [33]. Herein, genes encoding for the secretion of bacterial toxins
220 were predicted, especially hemolysin and *C. difficile* toxin A/B (TcdA/B). These later, is a
221 leading cause of CDI, where the pathogenic mechanism was the consequence of TcdA
222 (enterotoxin) and TcdB (cytotoxin) productions inducing colonic tissue damage [25,26]. The
223 idea of toxin-mediated disease rather than invasive-mediated disease is supported by the work
224 of Heida *et al.* [34]. Furthermore, hemolysins sequences with highly conserved domain found
225 in a toxin of *Brachyspira hyodysenteriae* that disclosed cytolytic effect on several cell lines
226 [13]. Cassir *et al.* identified cytotoxic activity of *C. butyricum* supernatant on Jurkat cells [4].
227 In conclusion, this study highlights the association between *C. neonatale* and NEC as the
228 possible existence NEC-associated geographic clones. Further genomic analysis is required on
229 a larger number of sequenced genomes.

Acknowledgments

We are grateful to the medical teams at Assistance publique - Hôpitaux de Marseille (AP-HM), CHU Nîmes, CHU Montpellier and CHU Nice for providing the stool samples. We thank the genomic and biobank platforms at the Institut Hospitalo-Universitaire (IHU) - Méditerranée Infection for technical assistance.

Funding

This work was supported by the French Government under the "*Investissements d'avenir*" program managed by the *Agence Nationale de la Recherche (ANR)*, [reference: *Méditerranée-Infection* 10-IAHU-03], by Région Provence-Alpes-Côte d'Azur and European funding FEDER PRIMI. M. Hosny was supported by *Fondation de Coopération Scientifique, Méditerranée-Infection* [*Infectiopôle-Sud* 2015].

Conflict of interest

None of the authors have any conflicts of interest to report.

References

[1] Hosny M, Cassir N, La Scola B. Updating on gut microbiota and its relationship with the occurrence of necrotizing enterocolitis. Hum Microbiome J 2017;4:14–9. doi:10.1016/j.humic.2017.04.001.

[2] Cassir N, Simeoni U, Scola B La. Gut microbiota and the pathogenesis of necrotizing enterocolitis in preterm neonates. Future Microbiol 2016;11:273–92. doi:10.2217/fmb.15.136.

[3] Cassir N, Benamar S, Khalil JB, Croce O, Saint-Faust M, Jacquot A, et al. *Clostridium butyricum* strains and dysbiosis linked to necrotizing enterocolitis in preterm neonates. Clin Infect Dis 2015;61:1107–15. doi:10.1093/cid/civ468.

[4] Ward D V, Scholz M, Zolfo M, Taft DH, Schibler KR, Tett A, et al. Metagenomic sequencing with strain-level resolution implicates uropathogenic *E. coli* in necrotizing enterocolitis and mortality in preterm infants. Cell Rep 2016;14:2912–24.

[5] Boccia D, Stolfi I, Lana S, Moro ML. Nosocomial necrotising enterocolitis outbreaks:epidemiology and control measures. Eur J Pediatr 2001;160:385–91.

[6] Smith B, Bodé S, Petersen BL, Jensen TK, Pipper C, Kloppenborg J, et al. Community analysis of bacteria colonizing intestinal tissue of neonates with necrotizing enterocolitis. BMC Microbiol 2011;11. doi:10.1186/1471-2180-11-73.

[7] Hosny M, Bou Khalil JY, Caputo A, Abdallah RA, Levasseur A, Colson P, et al. Multidisciplinary evaluation of *Clostridium butyricum* clonality isolated from preterm neonates with necrotizing enterocolitis in South France between 2009 and 2017. Sci Rep 2019;9:2077.

263 doi:10.1038/s41598-019-38773-7.

264 [8] Alfa MJ, Robson D, Davi M, Bernard K, Caeseele P Van, Harding GKM. An outbreak of
265 necrotizing enterocolitis associated with a novel *Clostridium* species in a neonatal intensive care
266 unit. Clin Infect Dis 2002;35:101–5.

267 [9] Roze J-C, Ancel P-Y, Lepage P, Martin-Marchand L, Nabhani Z Al, Delannoy J, et al.
268 Nutritional strategies and gut microbiota composition as risk factors for necrotizing enterocolitis
269 in very-preterm infants. Am J Clin Nutr 2017;36:821–830. doi:10.3945/ajcn.

270 [10] Benamar S, Cassir N, La Scola B. Genome sequence of a *Clostridium neonatale* strain
271 isolated in a Canadian neonatal intensive care unit. Genome Announc 2016;4.
272 doi:10.1128/genomeA.01431-15.

273 [11] Bernard K, Burdz T, Wiebe D, Alfa M, Bernier A-M. *Clostridium neonatale* sp. nov.
274 linked to necrotizing enterocolitis in neonates and a clarification of species assignable to the
275 genus *Clostridium* (Prazmowski 1880) emend. Lawson and Rainey 2016. Int J Syst Evol
276 Microbiol 2018;68:2416–23. doi:10.1099/ijsem.0.002827.

277 [12] Lu H, Giordano F, Ning Z. Oxford Nanopore MinION sequencing and genome assembly.
278 Genomics, Proteomics Bioinforma 2016;14:265–79. doi:10.1016/j.gpb.2016.05.004.

279 [13] Hosny M, Benamar S, Durand G, Armstrong N, Michelle C, Cadoret F, et al. Description
280 of *Clostridium phoceensis* sp. nov., a new species within the genus *Clostridium*. New Microbes
281 New Infect 2016;14:85–92. doi:10.1016/j.nmni.2016.09.008.

282 [14] Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes:

283 A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol
284 2012;19:455–77. doi:10.1089/cmb.2012.0021.

285 [15] Seemann T. Prokka: Rapid prokaryotic genome annotation. Bioinformatics
286 2014;30:2068–9. doi:10.1093/bioinformatics/btu153.

287 [16] Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis
288 Version 7.0 for bigger datasets. Mol Biol Evol 2016;33:1870–4. doi:10.1093/molbev/msw054.

289 [17] Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-
290 new capabilities and interfaces. Nucleic Acids Res 2012;40:1–12. doi:10.1093/nar/gks596.

291 [18] Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
292 Improvements in performance and usability. Mol Biol Evol 2013;30:772–80.
293 doi:10.1093/molbev/mst010.

294 [19] Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large
295 phylogenies by maximum likelihood. Syst Biol 2003;52:696–704.
296 doi:10.1080/10635150390235520.

297 [20] Stöver BC, Müller KF. TreeGraph 2: Combining and visualizing evidence from different
298 phylogenetic analyses. BMC Bioinformatics 2010;11:7. doi:10.1186/1471-2105-11-7.

299 [21] Zhao-Fleming H, Dissanaike S, Rumbaugh K. Are anaerobes a major, underappreciated
300 cause of necrotizing infections? Anaerobe 2017;45:65–70. doi:10.1016/j.anaerobe.2017.04.012.

301 [22] Benamar S, Cassir N, Merhej V, Jardot P, Robert C, Raoult D, et al. Multi-spacer typing
302 as an effective method to distinguish the clonal lineage of *Clostridium butyricum* strains isolated

303 from stool samples during a series of necrotizing enterocolitis cases. J Hosp Infect 2017;95:300–
304 5. doi:10.1016/j.jhin.2016.10.026.

305 [23] Adékambi T, Drancourt M, Raoult D. The rpoB gene as a tool for clinical
306 microbiologists. Trends Microbiol 2009;17:37–45. doi:10.1016/j.tim.2008.09.008.

307 [24] Fenollar F, Raoult D. Molecular genetic methods for the diagnosis of fastidious
308 microorganisms. APMIS 2004;112:785–807. doi:10.1111/j.1600-0463.2004.apm11211-1206.x.

309 [25] Fach P, Micheau P, Mazuet C, Perelle S, Popoff M. Development of real-time PCR tests
310 for detecting botulinum neurotoxins A, B, E, F producing *Clostridium botulinum*, *Clostridium*
311 *baratii* and *Clostridium butyricum*. J Appl Microbiol 2009;107:465–73. doi:10.1111/j.1365-
312 2672.2009.04215.x.

313 [26] Dittmar E, Beyer P, Fischer D, Schäfer V, Schoepe H, Bauer K, et al. Necrotizing
314 enterocolitis of the neonate with *Clostridium perfringens*: Diagnosis, clinical course, and role of
315 alpha toxin. Eur J Pediatr 2008;167:891–5. doi:10.1007/s00431-007-0614-9.

316 [27] Hill HR, Hunt CE, Matsen JM. Nosocomial colonization with *Klebsiella*, type 26, in a
317 neonatal intensive-care unit associated with an outbreak of sepsis, meningitis, and necrotizing
318 enterocolitis. J Pediatr 1974;85:415–9. doi:10.1016/S0022-3476(74)80133-2.

319 [28] Acker J Van, Smet F De, Muyldermans G, Naessens A, Lauwers S, Muyldermans T a N,
320 et al. Outbreak of necrotizing enterocolitis associated with *Enterobacter sakazakii* in powdered
321 milk formula. J Clin Microbiol 2001;39:7–12. doi:10.1128/JCM.39.1.293.

322 [29] Ferraris L, Butel MJ, Campeotto F, Vodovar M, Rozé JC, Aires J. Clostridia in premature

323 neonates' gut: Incidence, antibiotic susceptibility, and perinatal determinants influencing
324 colonization. PLoS One 2012;7. doi:10.1371/journal.pone.0030594.

325 [30] Faustini A, Forastiere F, Giorgi Rossi P, Perucci CA. An epidemic of gastroenteritis and
326 mild necrotizing enterocolitis in two neonatal units of a University Hospital in Rome, Italy.
327 Epidemiol Infect 2004;132:455–65.

328 [31] Eyre DW, Cule ML, Wilson DJ, Griffiths D, Vaughan A, O'Connor L, et al. Diverse
329 sources of *C. difficile* infection identified on whole-genome sequencing. N Engl J Med
330 2013;369:1195–205. doi:10.1056/NEJMoa1216064.

331 [32] Lagier JC, Dubourg G, Million M, Cadoret F, Bilen M, Fenollar F, et al. Culturing the
332 human microbiota and culturomics. Nat Rev Microbiol 2018;16:540–50. doi:10.1038/s41579-
333 018-0041-0.

334 [33] Stiemsma LT, Arrieta M-C, Dimitriu PA, Cheng J, Thorson L, Lefebvre DL, et al. Shifts
335 in *Lachnospira* and *Clostridium* sp. in the 3-month stool microbiome are associated with
336 preschool age asthma. Clin Sci 2016;130:2199–207. doi:10.1042/CS20160349.

337

338 **Table 1: Factors associated with the occurrence of necrotizing enterocolitis when compared with control preterm neonates**

Factors	NEC (n=88) No. (%)	Controls (n=71) No. (%)	Univariate analysis *p* value	Multivariate analysis OR [95% CI]	Multivariate analysis *p* value
Gestational age +/- SD (days)	28.36 (+/-2.82)	28.24 (+/- 2.93)	0.7864	NA	NA
Days of life +/- SD (days)	25.99 (+/- 13.16)	23.99 (+/- 11.61)	0.3166	NA	NA
Male sex	49 (55.7)	43 (60.5)	0.594	NA	NA
Birth weight mean +/- SD (g)	1179 (+/- 380)	1174 (+/- 432)	0.950	NA	NA
Very low birth weight (≤1500g)	69 (78.4)	58 (81.7)	0.617	NA	NA
Pasteurized breast milk	59 (67)	49 (69)	0.372	NA	NA
Formula fed	9 (10)	13 (18)	0.139	NA	NA
Vaginal delivery	32 (36)	37 (52)	0.023	2.3 [1.2-4.5]	0.018
Antibiotics before collection	64 (73)	40 (56)	0.092	1.9 [0.9-3.8]	0.07
C. neonatale q-PCR positive	30 (34)	9 (13)	0.003	3.7 [1.6-8.8]	0.003

339 SD = standard deviation; NA = not applicable.

340 **Table 2: Characteristics of *Clostridium neonatale* draft genomes**

Isolate	Location, date of isolation	Sequencing technique	Scaffolds number	Genome length (base pair)	GC content (%)	ORFs number	Accession number
LCDC 99A005	NICU-1, 2000	Mate-pair	3	4,304,459	28.6	3,795	GCA_001458595.1
LCDC 99A006	NICU-1, 2000	NA	58	4,658,596	28.4	4,139	GCA_002553455.1
NEC25	NICU-3, December 2011	Paired-end	3	4,352,344	28.7	3,883	UWJD01000001- 01000003
NEC26	NICU-3, July 2012	Paired-end	3	4,282,644	28.9	3,840	UICU01000001- 01000003
NEC32	NICU-3, December 2012	Paired-end	3	4,321,416	28.7	3,859	UICW01000001- 01000003
NEC86	NICU-2, December 2010	Paired-end, Nanopore	3	4,739,641	28.6	4,188	UICR01000001- 01000003

C25	NICU-3, February 2012	Paired-end	3	4,358,232	28.6	3,913	UICQ01000001- 01000003

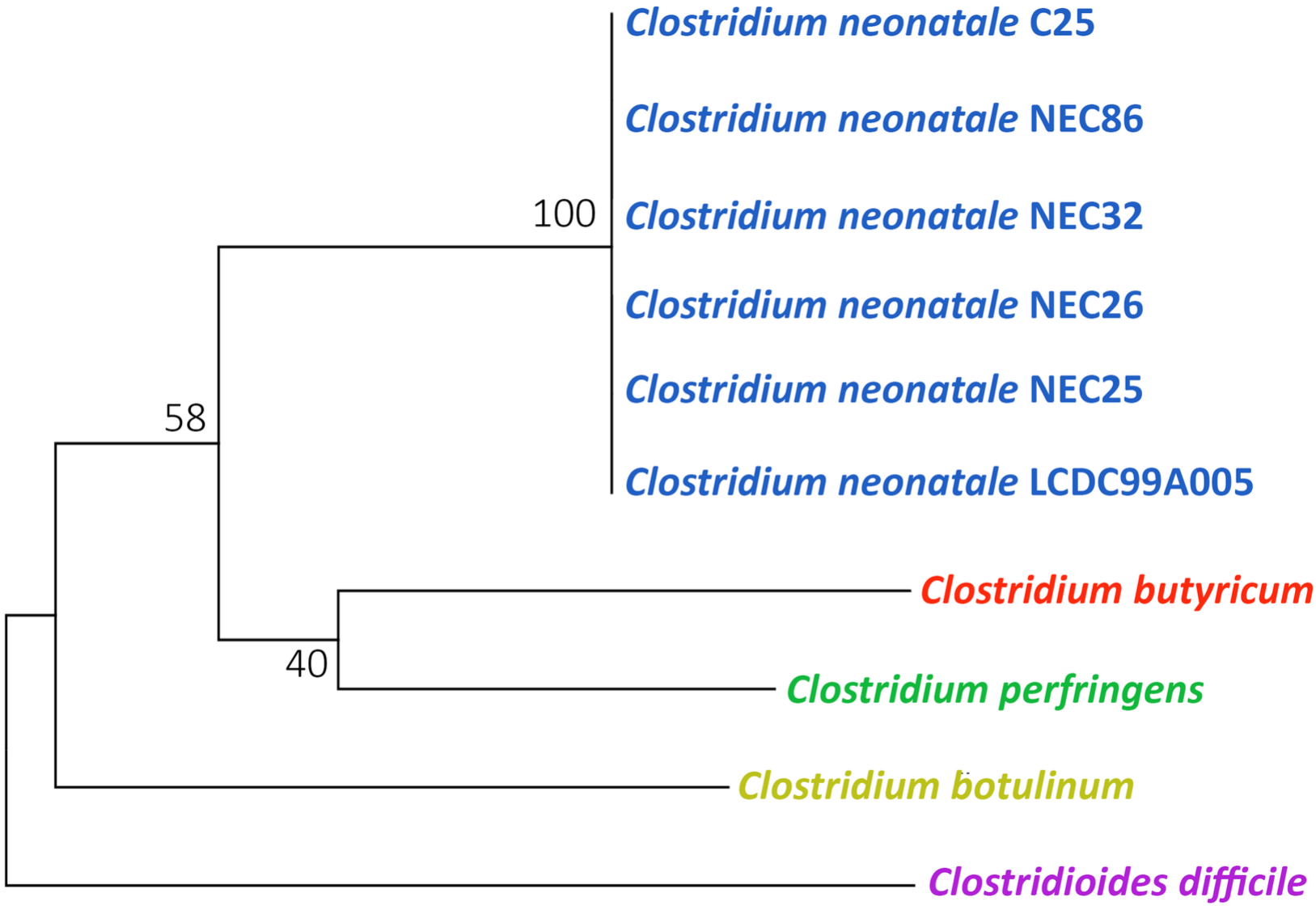
ORF: Open-Reading Frame number, NEC: Necrotizing enterocolitis; C25: control; NICU-1: Canada; NICU-2: Marseille, France; NICU-3: Marseille, France, NA: Not available

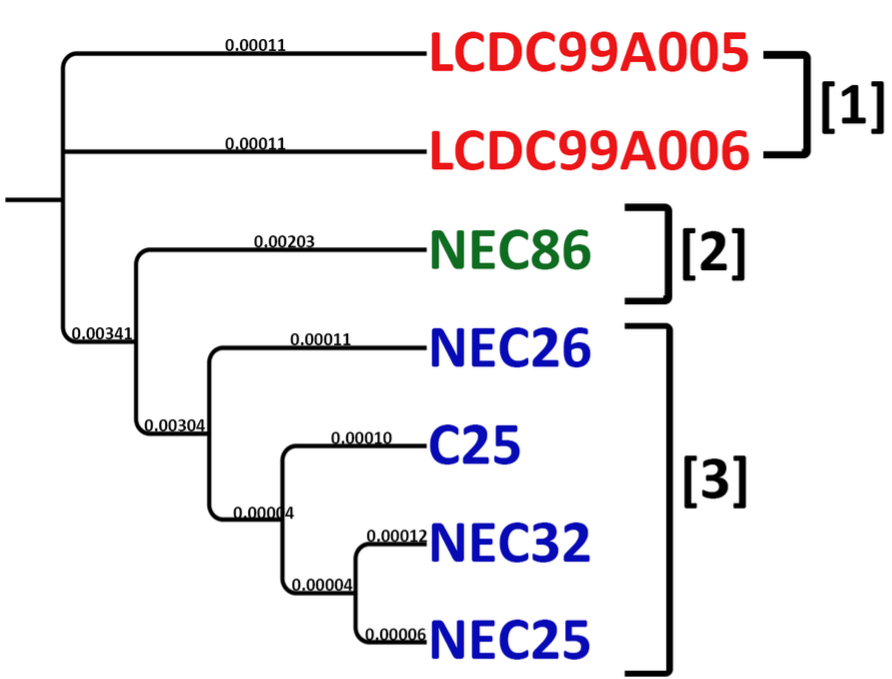
Figure 1: *In silico* analysis of *Clostridium neonatale rpoB* specific region. c This analysis was performed on the selected region of 151 nucleotides. Phylogenetic analysis showed that the *rpoB* sequence is homogenous among *C. neonatale* and classified it among *Clostridium* species. *C. difficile* was out grouped.

NEC: Necrotizing enterocolitis, C25: control.

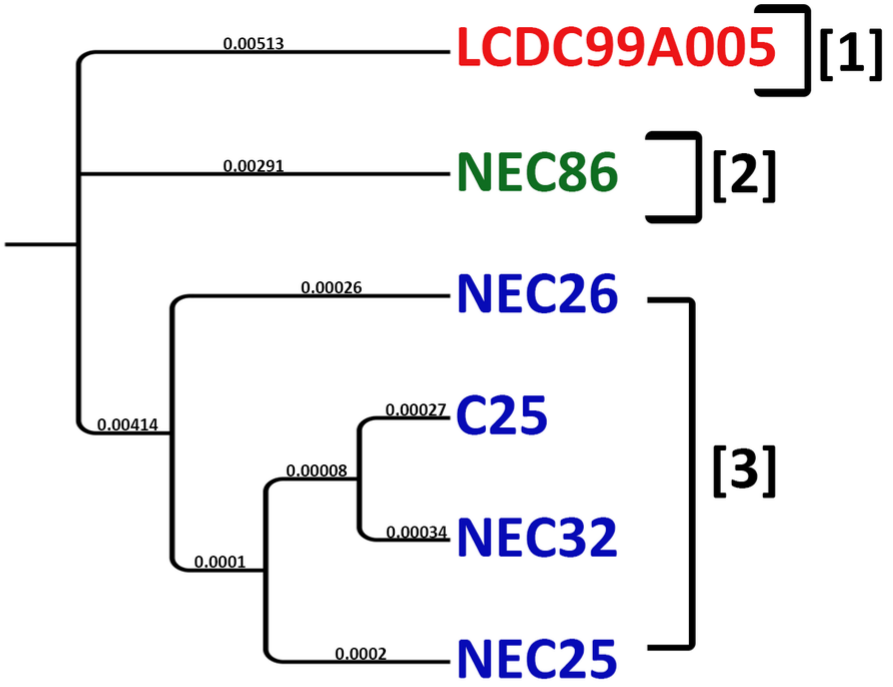
Figure 2: Relationship of *Clostridium neonatale* strains based on genomic analysis

(A): phylogenetic tree based on core-genome, (B): phylogenetic tree based on whole-genome alignment, NEC: Necrotizing enterocolitis, C25: control**.**


Clostridium neonatale C25
Clostridium neonatale NEC86
100
Clostridium neonatale NEC32
Clostridium neonatale NEC26
Clostridium neonatale NEC25
58
Clostridium neonatale LCDC99A005
Clostridium butyricum
40
Clostridium perfringens
Clostridium botulinum
Clostridioides difficile



(A)



(B)

NICU-1 (Canada)

NICU-2 (Marseille)

NICU-3 (Marseille)