Article

Not peer-reviewed version

Converting 16S Amplicon Reads to CFU Equivalents in Processed Foods by Viable Target Metagenomics

Javier Gamboa * and Marco Ferreira

Posted Date: 6 May 2026

doi: 10.20944/preprints202605.0325.v1

Keywords: quantitative metagenomics; viable but non-culturable; amplicon bias correction; 16S copy number; microbial succession; high-pressure processing; *Carnobacterium;* spore-formers; differential lysis; food microbial ecology

 doi:10.20944/preprints202605.0325.v1

Article

Converting 16S Amplicon Reads to CFU Equivalents in Processed Foods by Viable Target Metagenomics

Javier Gamboa [1,*] and Marco Ferreira [2]

[1] Biotecno.org, Vitoria-Gasteiz, Basque Country, Spain

[2] Quality Department, Brasmar Group SGPS, S.A., Portugal

* Correspondence: jgamboa@biotecno.org or jgamboa@biogenetics.es

Abstract

Using a viability-resolved metagenomics protocol coupled with copy-number-corrected absolute quantification, we report three findings from a longitudinal study of eight ready-to-eat (RTE) seafood matrices subjected to combined pasteurization and high-pressure processing (HPP). First, we demonstrate that 16S amplicon relative abundances can be converted into absolute cell counts concordant with ISO culture methods (Lin's CCC = 0.94; Bland-Altman bias = −0.25 $\log_{10}$, non-significant; 93% within ±1 $\log_{10}$). This establishes that the three principal barriers to quantitative amplicon metagenomics — compositional bias, copy-number distortion, and dead-cell DNA contamination — can be jointly overcome within a single analytical workflow. Second, we provide the first community-scale quantification of viable but non-culturable (VBNC) populations in processed foods: 60% of indicator groups with negative culture harbored molecular loads of 10^3–10^7 cells/g. The discrepancy scaled with predicted membrane resilience — maximal for sporulating clostridia (+5.9 $\log_{10}$), intermediate for Gram-negative Enterobacteriaceae (+4.0 $\log_{10}$), minimal for lactic acid bacteria — implicating differential DNase impermeability as a biological mechanism rather than an algorithmic artifact. Third, we uncovered a post-treatment microbial succession qualitatively distinct from that described for HPP alone. Whereas HPP-only products are typically dominated by mesophilic *Leuconostoc* and *Lactobacillus*, the combined thermal-barometric treatment selected spore-formers (*Clostridium* spp., *Sporosarcina*), the thermo-barotolerant lactic acid bacterium *Carnobacterium* as dominant LAB, and post-process recontamination psychrotrophs (*Psychrobacter*, *Acinetobacter*). Combining lethal barriers thus reconfigures surviving community structure rather than merely reducing it. These findings were enabled by DELOREAN (Deep Estimation of Live Organisms by Read Equivalence to Agar Numbers), a framework that integrates an original sample preparation protocol — Viable Target Metagenomics (vtMG) — with a quantitative conversion algorithm. vtMG enriches DNA from membrane-intact cells through differential lysis and DNase treatment as an enzymatic alternative to PMA; the algorithm then anchors copy-number-corrected relative abundances to fluorimetric total DNA quantification via empirical scaling factors. Calibration factors were derived in a prior, independent pilot study and applied without modification to the validation dataset, constituting a legitimate *out-of-distribution* test. The framework is classifier-agnostic, platform-independent, and designed for progressive refinement through collaborative calibration.

Keywords: quantitative metagenomics; viable but non-culturable; amplicon bias correction; 16S copy number; microbial succession; high-pressure processing; *Carnobacterium*; spore-formers; differential lysis; food microbial ecology

1. Introduction

Microbiological safety and quality of foods in the European Union is assessed through colony counts expressed as colony-forming units per gram (CFU/g), as established by Regulation (EC) No 2073/2005 on microbiological criteria applicable to foodstuffs. Under this regulation, acceptability limits (m) and maximum tolerance thresholds (M) are defined for indicator groups including aerobic

 doi:10.20944/preprints202605.0325.v1

mesophilic bacteria, Enterobacteriaceae, coliforms, *Escherichia coli*, sulfite-reducing clostridia, and yeasts and molds, each determined by standardized ISO culture methods on selective and differential media. Despite its age, this culture-based framework remains the legal basis for compliance assessment throughout the European food chain.

In parallel, next-generation sequencing (NGS) of the 16S rRNA gene has transformed microbial ecology by enabling simultaneous identification of hundreds of taxa from a single sample, unconstrained by culturability. Adoption of 16S amplicon metagenomics in food microbiology has accelerated in recent years, yielding insights into spoilage community composition, succession dynamics during shelf life, and detection of taxa refractory to conventional media (De Filippis et al., 2018; Ercolini, 2013). Yet the output of amplicon sequencing — relative abundance profiles expressed as proportions of total reads — is fundamentally incompatible with the regulatory language of absolute colony counts. A relative abundance of 30% for a given taxon conveys no information about whether the absolute load is 10^2 or 10^7 CFU/g; this distinction alone determines whether a product is compliant or must be withdrawn from the market.

This disconnect between the informative output of metagenomics and the quantitative requirements of food safety regulation has impeded integration of molecular methods into routine microbiological monitoring. Several approaches have been proposed to bridge this gap. Spike-in strategies using known quantities of synthetic DNA or cells enable conversion of relative to absolute abundances (Tourlousse et al., 2017; Stämmler et al., 2016) but require additional laboratory steps and are not yet standardized. Quantitative PCR (qPCR) can provide absolute counts for individual targets but lacks the community coverage that makes metagenomics attractive. Viability PCR approaches using propidium monoazide (PMA) selectively suppress amplification of DNA from cells with compromised membranes (Nocker et al., 2007) and can be applied both to specific targets via qPCR and at the community level by combining them with amplicon sequencing (PMA-seq) (Emerson et al., 2017; Nkuipou-Kenfack et al., 2013). However, the efficacy of PMA treatment in complex food matrices is compromised by incomplete dye penetration in samples with high fat or particulate content, dependence on optimal photoactivation conditions, and difficulty of standardization across sample types (Fittipaldi et al., 2012). Moreover, PMA-seq protocols do not include a eukaryotic host DNA depletion step. In meat or seafood products, animal tissue DNA can dominate the extract and compete for sequencing depth, reducing detection sensitivity of the microbial fraction. The vtMG protocol addresses this problem through a prior differential lysis step that selectively disrupts eukaryotic membranes, releasing their DNA to the extracellular medium where it is subsequently degraded by DNase treatment.

A critical technical complication affects any attempt to derive quantitative estimates from 16S amplicon data: the variable copy number of the 16S rRNA gene across bacterial genomes. A taxon carrying seven copies of the 16S gene will generate approximately seven-fold more reads per cell than a single-copy taxon, inflating its apparent relative abundance. Existing tools correct this bias by dividing relative abundances by the estimated copy number of each taxon (CopyRighter, Angly et al., 2014; PICRUSt, Langille et al., 2013; PAPRICA, Bowman & Ducklow, 2015), although the accuracy of these predictions remains debated (Louca et al., 2018; Starke et al., 2021). Such tools produce corrected community profiles — relative abundances closer to the true proportion of cells — but do not yield absolute counts, which food regulation demands.

DELOREAN incorporates its own Random Forest-based copy number prediction model (trained on rrnDB; Section 2.6) and extends the correction to absolute quantification: it combines copy number correction with an independent measure of total viable DNA mass per gram of sample (obtained through the vtMG protocol) to calculate absolute cell counts per taxon. For this conversion, it uses the genome size of each taxon, since a larger genome implies more nanograms of DNA per cell. The result is a direct estimate of CFUeq/g per regulatory indicator group, expressed in the same units as traditional culture and assessable against the *m* and *M* limits of Regulation (EC) No 2073/2005.

High-pressure processing (HPP) is a leading non-thermal preservation technology for extending the shelf life of ready-to-eat (RTE) products, including seafood. Application of pressures typically in

 doi:10.20944/preprints202605.0325.v1

the 400–600 MPa range inactivates vegetative pathogens and spoilage organisms while preserving sensory and nutritional properties (Considine et al., 2008). In industrial practice, HPP is increasingly combined with mild thermal treatments (pasteurization) to achieve broader microbial inactivation, particularly against pressure-resistant organisms. However, the microbial ecology of products subjected to this double treatment (pasteurization followed by HPP) remains poorly characterized. The existing literature on HPP-treated foods describes spoilage communities typically dominated by lactic acid bacteria (LAB), particularly *Leuconostoc*, *Lactobacillus*, and *Weissella* (Li et al., 2020). Whether this pattern holds when HPP is preceded by thermal treatment — which preferentially eliminates thermosensitive LAB and selects spore-forming and thermotolerant organisms — has not been investigated using culture-independent methods.

A specific complication of HPP is the physiological heterogeneity of surviving cells, which cannot be reduced to a simple "alive/dead" dichotomy. Recent literature distinguishes at least three post-HPP cellular states that coexist in proportions dependent on species and treatment intensity (Yang et al., 2023; Bozoglu et al., 2004). *(i)* Pressure-resistant (PR) cells survive and maintain culturability. *(ii)* Sublethally injured (SLI) cells retain a morphologically recognizable outer membrane under electron microscopy but exhibit altered permeability, denatured cytoplasmic proteins, and nucleoid condensation (Gayán et al., 2017; Moussa et al., 2007; Ritz et al., 2001). *(iii)* Viable but non-culturable (VBNC) cells possess apparently intact membranes but compromised metabolic capacity. The latter two states are invisible to conventional culture but partially retain the ability to resuscitate under favorable conditions during storage, posing a food safety risk that the literature has extensively documented (Yang et al., 2023; Wu, 2008).

Current molecular methods for selective detection of viable DNA — PMA-seq and DNase treatments applied directly to the sample — rely on the dichotomy "intact membrane = viable" / "disrupted membrane = non-viable." For double-treated products, this binary approximation is insufficient: it cannot differentiate a healthy cell from a sublethally injured cell whose membrane remains morphologically recognizable, and may underestimate the VBNC fraction depending on the degree of permeabilization. The distinction matters because these subpopulations carry different regulatory implications. PR cells register in culture; SLI cells do not but can recover and proliferate; VBNC cells constitute a hidden reserve whose reactivation marks the true end of microbiological shelf life.

The vtMG protocol presented here addresses this limitation by introducing a *differential pre-lysis* step prior to DNase treatment. The pre-lysis is calibrated to be sufficiently aggressive to disrupt eukaryotic membranes of the host (animal tissue in the product) and to permeabilize already-compromised membranes of sublethally injured bacterial cells, yet sufficiently mild to avoid lysing bacterial cells with healthy membranes. After this step, DNase degrades both extracellular and host DNA and the DNA of now-permeabilized sublethally injured cells, while DNA from membrane-intact cells remains protected. The result is an estimate of DNA from cells *not sublethally damaged*, a fraction closer to the population with active proliferation capacity in the product — which is the one that determines actual microbiological spoilage. Calibration of the pre-lysis — aggressive enough to permeabilize sublethally damaged membranes without compromising healthy cells — is a critical experimental parameter of the protocol that must be empirically optimized for each matrix type and lethal treatment applied, and constitutes one of the main methodological development lines of vtMG.

In this work, we present DELOREAN (Deep Estimation of Live Organisms by Read Equivalence to Agar Numbers), a framework that converts community-level 16S rRNA amplicon data into CFU equivalent estimates (CFUeq/g) aligned with the regulatory indicator groups of Regulation (EC) No 2073/2005. The framework comprises three integrated components: (i) viable target metagenomics (vtMG), a sample preparation protocol enriching DNA from membrane-intact organisms through differential lysis and DNase treatment; (ii) a conversion algorithm with explicit, correct separation of copy number and genome size corrections; and (iii) empirical calibration factors from paired molecular and culture data. We validate DELOREAN against ISO culture counts in an industrial shelf-life study of eight HPP+pasteurization-treated RTE seafood matrices, sampled at four time

 doi:10.20944/preprints202605.0325.v1

points over 90 days post-treatment, and characterize the microbial succession accompanying shelf-life loss in these double-treated products.

2. Materials and Methods

2.1. Study Design and Sample Collection

The study was conducted within an industrial shelf-life study (SLS) of HPP-treated RTE seafood products manufactured by Brasmar (Portugal). Eight product matrices were included: tuna loin with additive, tuna loin without additive, cod flakes with additive, cod flakes without additive, hake, baby squid (txipirones), octopus, and curled octopus tentacle (rejo). All products were subjected to a double lethal treatment consisting of pasteurization followed by high-pressure processing (HPP) and stored under refrigerated conditions per the manufacturer's standard commercial protocol.

Sampling was performed at four post-HPP time points: T1 (0 days), T2 (45 days), T3 (75 days), and T4 (90 days). At each time point, one 25 g analytical unit per matrix was collected, yielding a total of 32 sample-time combinations. All samples were analyzed by both molecular methods (vtMG + 16S amplicon) and, at T4, by traditional culture-based microbiological methods.

2.2. Traditional Microbiological Analysis

Culture-based enumeration at T4 was performed by Laboratorios Bromatológicos Araba S.A.U. (Mérieux NutriSciences Group, Vitoria-Gasteiz, Spain), an ENAC-accredited laboratory. The following indicators were analyzed:

ENAC-accredited assays followed the corresponding ISO methods: aerobic mesophilic count (ISO 4833-1:2013, plate count at 30 °C), total coliforms (ISO 4832:2006, plate count at 30 °C), β-glucuronidase-positive *Escherichia coli* (ISO 16649-2:2001, plate count at 44 °C), and total Enterobacteriaceae (ISO 21528-2:2017, plate count at 37 °C).

Assays not covered by ENAC accreditation were performed using validated in-house methods: mesophilic lactic acid bacteria (plate count), sulfite-reducing clostridia (in-house method MAL-CS-01), and detection of *Listeria monocytogenes* (in-house method MAL-LI-04, presence/absence in 25 g).

The limit of detection (LOD) for enumeration methods was 1.0×10^1 CFU/g. Results below the LOD were recorded as < 10 CFU/g and assigned a value of 5.0 CFU/g (half LOD) for statistical analysis. The upper counting limit for aerobic mesophilic count was 3.0×10^7 CFU/g; results exceeding this limit were recorded as $> 3.0 \times 10^7$ CFU/g and the value of 3.0×10^7 CFU/g was used in calculations.

2.3. Viable Target Metagenomics (vtMG) — Sample Preparation

The vtMG protocol was designed to selectively extract DNA from bacterial cells that retain membrane integrity, thereby excluding DNA from dead cells with compromised membranes and free extracellular DNA. The procedure comprised the following steps:

(a) *Sample homogenization.* A 25 g analytical portion was homogenized in buffered peptone water in a stomacher to release microbial cells from the food matrix.

(b) *Mild differential lysis.* Host cell membranes (eukaryotic) and food matrix debris were disrupted by a mild lysis step targeting the more fragile eukaryotic membranes, leaving bacterial membranes largely intact. This approach of differential lysis followed by nuclease treatment is well established in clinical metagenomics for host DNA depletion in samples with high proportions of human DNA (Hasan et al., 2016; Charalampous et al., 2019). In the vtMG context, the principle is adapted to food matrices, where eukaryotic DNA originates from the animal tissue of the product rather than human cells. The step reduces host DNA contamination and enriches the microbial fraction.

(c) *DNase treatment (Pretto-DNase v3 protocol).* To 100 µL of the exudate obtained after differential lysis, 100 µL of molecular-grade water was added, followed by the reaction mixture consisting of 11 µL of DNase I in a buffer of 12.5 µL Tris-HCl, 12.5 µL $CaCl_2$, and 12.5 µL $MgCl_2$. Enzymatic digestion was carried out at 37 °C for 60 minutes with gentle agitation, degrading all accessible DNA in the

 doi:10.20944/preprints202605.0325.v1

extracellular medium — including host DNA released in step (b), environmental extracellular DNA, and DNA from bacterial cells with compromised membranes. DNA enclosed within cells with membrane integrity is protected from enzymatic degradation.

(d) *DNase inactivation and bacterial pellet recovery.* The reaction was stopped by adding 90 µL of 0.1 M EDTA to chelate catalytic cofactors (Mg^{2+}, Ca^{2+}) and incubating at 65 °C for 10 minutes with gentle agitation. Samples were centrifuged at 11,000 × g for 5 minutes. The supernatant — containing degraded DNA, inactivated DNase products, and matrix debris — was discarded. The pellet, containing bacterial cells with intact membranes, was resuspended in 200 µL of double-distilled water (ddH_2O) for subsequent bacterial lysis.

(e) *DNA extraction and purification.* Total DNA from the lysate was extracted and purified using the DNeasy PowerFood Microbial Kit (QIAGEN, cat. no. 21000-100), specifically designed for isolation of microbial genomic DNA from complex food matrices. The kit employs mechanical lysis via bead-beating with 0.15 mm garnet PowerBead tubes, combined with chemical lysis (Solution MBL), followed by PCR inhibitor removal using IRT (Inhibitor Removal Technology; Solution IRS, incubation at 2–8 °C for 5 min) and silica membrane column purification. DNA was eluted in 50 µL of Solution EB. This kit has been validated by the manufacturer for DNA recovery from foodborne pathogens including *Clostridium perfringens*, *C. difficile*, *Enterococcus faecalis*, *Escherichia coli*, *Lactobacillus* spp., *Listeria monocytogenes*, and *Salmonella enterica* from meat, dairy, vegetable, and juice matrices.

(f) *Dual DNA quantification.* Concentration and purity of extracted DNA were assessed by both spectrophotometric (NanoDrop; A260/A280) and fluorimetric methods. Spectrophotometric values (ng/µL) were used to calculate DNA concentration per gram of sample (DNA/g), taking into account the elution volume (50 µL standard) and the analytical portion weight (25 g). The DNA/g value serves as the absolute anchor for the DELOREAN conversion algorithm, representing the total mass of DNA derived from membrane-intact organisms per gram of product.

The specific parameters of step (b) — mild differential lysis — including detergent type, concentration, incubation temperature, and duration, were empirically optimized for RTE seafood matrices subjected to combined thermal + HPP treatment. These parameters are proprietary to the vtMG protocol and will be disclosed in a dedicated methodological publication currently in preparation. The optimization criterion was maximization of the ratio of eukaryotic membrane disruption to bacterial membrane disruption, assessed by differential DNA quantification before and after the lysis step.

2.4. 16S rRNA Gene Amplicon Sequencing and Bioinformatic Processing

DNA extracted via vtMG was amplified using primers 515F and 1492R, generating amplicons of approximately 977 bp spanning hypervariable regions V4–V9 of the 16S rRNA gene. Libraries were prepared with the 16S Microbial Sequencing Kit (Oxford Nanopore Technologies, ONT) and sequenced on the MinION platform by Laboratorios Bromatológicos Araba S.A.U. (Mérieux NutriSciences Group, Vitoria-Gasteiz, Spain).

Bioinformatic processing followed the HelixCore pipeline integrating peer-reviewed open-source tools. Basecalling used Dorado (ONT; https://github.com/nanoporetech/dorado) with the SUP (super-accuracy) model. Chopper (De Coster & Rademakers, 2023) performed quality control and filtering ($Q \geq 10$, minimum length 200 bp). Taxonomic classification employed EMU v3 (Curry et al., 2022), an expectation-maximization classifier designed for long 16S reads that provides species-level assignments with relative abundance estimates against a SILVA-derived reference database, as evidenced by the resulting nomenclature (e.g., *Clostridium sensu stricto* groups 1, 5, 13).

Taxa represented by less than 1% of total reads in a given sample were retained in the dataset but flagged as low-confidence assignments following the approach of Bokulich et al. (2013). Samples with insufficient DNA concentration to generate sequencing libraries (DNA < 0.01 ng/µL) were annotated as "no abundance data" in the corresponding panels — an operational stopping point of the vtMG protocol that avoids generating unreliable data and reduces unnecessary costs.

 doi:10.20944/preprints202605.0325.v1

Although EMU was the primary classifier in this study, the DELOREAN framework is designed to be classifier-agnostic. It operates downstream of any taxonomic classification tool, requiring only a table of relative abundances per taxon as input. Compatibility with Kraken2/Bracken, DADA2, Minimap2, and other classifiers is maintained by design.

2.5. The DELOREAN Algorithm

2.5.1. Rationale and Overview

The DELOREAN algorithm converts community-level 16S amplicon relative abundance profiles into absolute CFUeq estimates for each regulatory indicator group defined in Regulation (EC) No 2073/2005. The conversion proceeds in three steps: (1) correction of relative abundances for variation in 16S rRNA gene copy number; (2) conversion of total viable DNA mass into total cell count and distribution among taxa; and (3) application of an empirical calibration factor per indicator group and matrix. A critical design principle is the explicit separation of copy number correction (Step 1) from genome size correction (Step 2), each applied at the mathematically appropriate stage of the calculation.

2.5.2. Step 1 — Copy-Number-Corrected Relative Abundance

In 16S amplicon sequencing, each sequencing read originates from amplification of a single copy of the 16S rRNA gene. A taxon *i* carrying *CN_i* copies of this gene per genome will therefore generate *CN_i* times more reads per cell than a taxon with a single copy. To obtain relative abundances that reflect the proportion of cells rather than the proportion of gene copies, the observed relative abundance *RA_i* must be divided by the copy number:

$$RA\_corrected\_i = (RA\_i \,/\, CN\_i) \,/\, \Sigma\_k(RA\_k \,/\, CN\_k)$$

This correction **does not** involve genome size. The rationale is direct: in an amplicon experiment, the sequencing library is constructed from PCR products of a single locus (the 16S gene), not from randomly fragmented whole genomes. The probability that a given cell contributes a read to the library depends exclusively on the number of 16S copies it carries, not on the total size of its genome. Genome size would influence read representation in whole-genome shotgun metagenomics, where DNA is randomly fragmented along the genome, but not in amplicon sequencing. Genome size enters the conversion in Step 2 (Section 2.5.3), where it serves a different function: determining the DNA mass per cell for conversion of total DNA mass to cell number.

Copy number values were obtained from the rrnDB database (Stoddard et al., 2015) at the genus level when available. For taxa not represented in rrnDB, copy numbers were estimated using a Random Forest predictive model (Section 2.6).

2.5.3. Step 2 — DNA Mass to Cell Count Conversion

The total viable bacterial DNA mass per gram of product (*DNA/g*), obtained from the vtMG protocol (Section 2.3), provides the absolute quantitative anchor for the conversion. This total DNA mass is distributed among taxa in proportion to their copy-number-corrected relative abundances (from Step 1), and converted to cell counts using the relationship between genome size and DNA mass per cell.

The DNA mass per cell for taxon *i* is given by:

$$DNA\_mass\_per\_cell\_i = GS\_i \times (1.096 \times 10^{-6}\ ng/Mb)$$

where *GS_i* is the genome size in megabases and the constant 1.096×10^{-6} ng/Mb represents the mass of one megabase of double-stranded DNA. Here genome size enters the calculation because it determines how much of the total extracted DNA mass corresponds to each cell. A taxon with a 5 Mb genome contributes five times more DNA per cell than a taxon with a 1 Mb genome.

The total number of cells of taxon *i* per gram of product is then:

 doi:10.20944/preprints202605.0325.v1

$$Cells\_i/g = [DNA/g \times RA\_corrected\_i] / [GS\_i \times 1.096 \times 10^{-6}]$$

which can be expressed in the aggregated form:

$$CFUeq\_i/g = [DNA\_total \times (RA\_i / CN\_i)] / [1.096 \times 10^{-6} \times \Sigma\_k(RA\_k \times GS\_k / CN\_k)] \times K(j,m)$$

In this formulation, the denominator Σ_k(RA_k × GS_k / CN_k) represents the weighted average genomic contribution of the community, integrating both the copy-number-corrected community structure and the genome size of each taxon to convert total DNA mass into total cell equivalents.

Genome size values were obtained from the NCBI Genome database, using median genus-level values from complete genome assemblies when available.

2.5.4. Step 3 — Empirical Calibration Factor K

Conversion of estimated cell counts to CFUeq requires an empirical calibration factor *K(j)* specific to each indicator group *j*. This factor absorbs systematic biases from multiple sources: differential DNA extraction efficiency across taxa, PCR amplification bias, the proportion of membrane-intact but non-culturable cells in a given indicator group, and differences in growth efficiency on the selective medium used for the corresponding culture method.

The *K(j)* factors used in this study were derived in a prior, independent pilot experiment conducted on cod flakes (n = 6 paired vtMG + ISO culture samples), calibrated as the median ratio of observed culture counts (CFU/g) to estimated cell counts (cells/g from Step 2) for each indicator group. Using the median rather than the mean provides robustness against outliers. These factors were applied without modification to the BRAESP-SLS validation dataset (eight diverse matrices, T4 at 90 days post-treatment), thus constituting empirical priors independent of the evaluation dataset. This design allows the T4 validation to be considered a legitimate *out-of-distribution* test, in which the K values were never fitted to the validation dataset.

The factors employed were: aerobic mesophiles K = 0.95; Enterobacteriaceae K = 1.20; total coliforms K = 1.15; *Escherichia coli* K = 1.10; sulfite-reducing clostridia K = 1.30; lactic acid bacteria K = 0.90. The proximity of these factors to unity indicates that the DELOREAN algorithm (copy number correction + DNA mass → cell conversion) already produces estimates in the correct range, and that K factors absorb minor residual biases specific to each indicator group.

We acknowledge as a relevant methodological limitation that these priors were derived from a single source matrix (cod flakes). Their application to the remaining matrices in the validation dataset (tuna loin, hake, baby squid, octopus, rejo) assumes that systematic biases per indicator group are transferable across RTE seafood matrices with similar double treatment — an assumption that this work tests and that the results support (Section 3). Extension to other product categories (meat, dairy, vegetables, beverages) will require specific calibration with dedicated paired data.

2.5.5. Taxonomic Assignment to Regulatory Indicator Groups

16S-based taxonomic assignments were mapped to the regulatory indicator groups defined in Regulation (EC) No 2073/2005 using the following criteria:

Aerobic mesophilic bacteria: All bacterial taxa detected in the sample, as this indicator is designed to capture the total culturable aerobic flora.

Lactic acid bacteria (LAB): Taxa belonging to the families Lactobacillaceae (including *Lactobacillus*, *Lactiplantibacillus*, *Pediococcus*), Enterococcaceae (*Enterococcus*), Carnobacteriaceae (*Carnobacterium*), Leuconostocaceae (*Leuconostoc*, *Weissella*), and Streptococcaceae (*Streptococcus*, *Lactococcus*), following the broad definition used in food microbiology for enumeration on MRS agar.

Coliforms: Lactose-fermenting Enterobacteriaceae at the genus level, including *Escherichia*, *Klebsiella*, *Enterobacter*, *Citrobacter*, and *Serratia*.

Escherichia coli: Reads assigned to *Escherichia coli* at the species level. The 16S rRNA gene cannot reliably discriminate *E. coli* from *Shigella* spp. regardless of the region sequenced, as these taxa may

 doi:10.20944/preprints202605.0325.v1

share identical 16S sequences (Chakravorty et al., 2007). The V4–V9 amplicon employed in this study lacks V1–V3, the subregion with the best relative resolution for this complex (Johnson et al., 2019). Reads assigned to the *Escherichia-Shigella* group were treated conservatively and flagged for confirmation.

Enterobacteriaceae: All taxa assigned to the family Enterobacteriaceae, including both lactose-fermenting and non-fermenting genera (*Salmonella*, *Proteus*, *Providencia*, *Morganella*, *Yersinia*, *Serratia*, *Plesiomonas*, etc.).

Sulfite-reducing clostridia: Taxa assigned to *Clostridium sensu stricto* (groups 1, 5, 11, 12, 13, 18) and classical *Clostridium* species (*C. perfringens*, *C. bowmanii*, *C. pasteurianum*, *C. subterminale*, *C. intestinale*), corresponding to organisms capable of growing on TSN/SPS agar under anaerobic conditions with blackening of the medium.

Yeasts and molds: Not assessable by 16S rRNA gene sequencing. The 16S marker is exclusively bacterial and does not amplify fungal DNA. Detection of yeasts and molds requires the ITS (Internal Transcribed Spacer) marker (Schoch et al., 2012). This limitation is explicitly noted in all DELOREAN panels.

Listeria monocytogenes: Presence/absence assessment based on detection of reads assigned to the genus *Listeria* at any abundance level, including below the 1% threshold. Detections are flagged for confirmatory testing by ISO 11290-1 or species-specific PCR.

2.5.6. Pathogen Surveillance

Beyond the regulatory indicator groups, all detected taxa were cross-referenced against a curated list of foodborne pathogens derived from Regulation (EC) No 2073/2005 Annex I and relevant EFSA Scientific Opinions. Any taxon matching a known pathogen was flagged in the DELOREAN panel regardless of its relative abundance, as even a minority pathogen population is of regulatory relevance. Flagged pathogens were annotated with their estimated CFUeq/g, the indicator group in which they were detected, and a recommendation for confirmatory testing by specific PCR or ISO reference method.

2.6. Random Forest Model for 16S Copy Number Prediction

For taxa not represented in the rrnDB database, 16S rRNA gene copy numbers were predicted using a supervised Random Forest regression model. The training dataset was constructed from rrnDB entries with known copy numbers, using taxonomic lineage information (phylum, class, order, family, genus) as encoded categorical variables.

Model performance was evaluated using leave-one-genus-out cross-validation, in which all entries from a given genus were excluded from training and the model was used to predict their copy numbers. This strategy ensures that the evaluation reflects the realistic use case of predicting copy numbers for taxa not present in the reference database. Performance was compared against two baselines: (a) lookup of mean copy number at the family level, and (b) global mean copy number of all bacteria.

Prediction of 16S copy numbers from taxonomic information is a recognized source of uncertainty in quantitative amplicon analyses (Louca et al., 2018). DELOREAN addresses this limitation at two levels. First, the Random Forest model was evaluated via genus-level cross-validation (10 partitions, complete genera assigned to a single partition, never seen during training) on the rrnDB v5.10 dataset (Stoddard et al., 2015; April 2025 update), which contains 46,920 clean records aggregated to 7,314 species distributed across 1,887 genera, 493 families, and 41 phyla. The Random Forest achieved MAE = 1.25 copies, RMSE = 1.76, $R^2 = 0.62$, with 69.0% of predictions within ±1 copy and 88.8% within ±2 copies. A family-level lookup baseline (mean copy number of all genera in the same family, excluding the target genus) offered virtually identical performance (MAE = 1.25; $R^2 = 0.62$; 67.2% within ±1), while a global mean baseline was clearly inferior (MAE = 2.26; $R^2 = -0.01$). Feature importance of the final model revealed that mid-rank taxonomic signals dominate prediction (order = 0.33; family = 0.30; class = 0.23; phylum = 0.13), with negligible contribution from genome

 doi:10.20944/preprints202605.0325.v1

size (< 0.01) due to its strong collinearity with taxonomic ranks. The equivalence between Random Forest and family lookup confirms previous observations (Starke et al., 2021; Louca et al., 2018) that 16S copy number prediction is a partially solved problem, in which simple taxonomic aggregation strategies capture nearly all the predictable signal available. Second, and more determinatively, the empirical calibration factors K (Step 3, Section 2.5.4) are calculated from paired culture data and absorb residual copy number prediction errors along with other sources of systematic pipeline bias. This means that DELOREAN's final accuracy does not depend exclusively on the predictive model's precision but on the integral calibration of the complete system against reference data.

2.7. Biodiversity and Ecology Analyses

Alpha diversity was characterized using the Shannon index (H'), Pielou's evenness (J' = H'/ln S, where S is observed taxon richness), and the Chao1 richness estimator. These metrics were calculated on copy-number-corrected relative abundance profiles.

Beta diversity between samples within each time point was assessed using the Bray-Curtis dissimilarity index, expressed as percentage similarity (100% = identical composition, 0% = no shared taxa). Samples with a mean Bray-Curtis similarity below 50% relative to other samples at the same time point were flagged as outlier profiles.

2.8. Validation Metrics

Concordance between DELOREAN estimates (CFUeq/g) and traditional culture counts (CFU/g) was evaluated on $\log_{10}$-transformed values using the following metrics: Pearson correlation coefficient (r), mean absolute error (MAE, in $\log_{10}$ units), mean bias (mean difference $\log_{10}$ CFUeq − $\log_{10}$ CFU), root mean square error (RMSE, in $\log_{10}$ units), and the percentage of paired values falling within ±1.0 and ±0.5 $\log_{10}$ units of each other. For the aerobic mesophilic indicator, the ratio of molecular to culture estimate (CFUeq/CFU) was additionally calculated for each sample.

Concordance between methods was further classified into four categories: (a) concordant ($|\Delta\log_{10}| \leq 0.5$); (b) molecular overestimation ($\Delta\log_{10} > 0.5$, both methods above LOD); (c) culture overestimation ($\Delta\log_{10} < -0.5$, both methods above LOD); and (d) culture below LOD with molecular detection (culture = LOD but CFUeq > 100), interpreted as possible evidence of sublethally damaged or VBNC populations.

2.9. Data Presentation: DELOREAN Regulatory Panels

Results for each sample were compiled into standardized Microbiological Criteria Panels, formatted according to the indicator groups and limits (*m*, *M*) of Regulation (EC) No 2073/2005. Each panel comprises: sample metadata (matrix, treatment, DNA concentration, analysis date), a logarithmic bar scale for each indicator group with color-coded zones (green ≤ *m*, amber *m*–*M*, red > *M*), taxonomic breakdown of dominant organisms within each indicator group, alpha diversity metrics (Shannon H', Pielou J'), inter-sample Bray-Curtis similarity, integrated pathogen alerts, and interpretive notes on concordance or methodological discrepancies. When paired culture data were available (T4), culture values were overlaid on molecular estimates for direct visual comparison.

3. Results

3.1. DNA Yield and Viable Biomass Across Shelf Life

Viable DNA concentration extracted via vtMG followed a consistent biphasic pattern across all 8 matrices (Table 1). At T1 (0 days post-HPP) and T2 (45 days), concentrations were very low (0.0001–2.00 ng/g), reflecting the efficacy of the double lethal treatment; three T1 samples (Cod_CON, Baby_squid, Octopus) fell below the operational sequencing threshold (< 0.01 ng/µL) and were annotated as "no abundance data."

Between T2 and T3 (45–75 days), a biomass increase of approximately 10^2–10^3-fold occurred across all matrices, reaching concentrations of 46–170 ng/g at T3. At T4 (90 days), concentrations

 doi:10.20944/preprints202605.0325.v1

stabilized or continued to increase (8–274 ng/g), with the notable exception of Cod_SIN (16 ng/g), which maintained the lowest load of the T4 group.

This pattern identifies a critical window between days 45 and 75 post-HPP in which the double treatment loses its barrier efficacy across all matrices simultaneously.

Table 1. DNA concentration (ng/µL, ng/g) by sample and time point (32 panels).

Matrix	T1 (0 d) ng/g	T2 (45 d) ng/g	T3 (75 d) ng/g	T4 (90 d) ng/g
Tuna_CON	0.04	0.40	170.0	274.0
Tuna_SIN	0.04	0.36	134.0	198.0
Cod_CON	< LOQ*	0.08	128.0	248.0
Cod_SIN	0.02	2.00	46.0	16.0
Hake	0.06	0.52	96.0	178.0
Baby_squid	< LOQ*	0.28	166.0	212.0
Octopus	< LOQ*	0.16	112.0	256.0
Rejo	0.06	0.44	88.0	8.0

<LOQ: Below operational quantification limit (DNA < 0.01 ng/µL); annotated as "no abundance data.".

3.2. DELOREAN–Culture Concordance Validation: Quantifiable Pairs

Concordance between DELOREAN (CFUeq/g) and culture (CFU/g) was assessed on pairs where both methods returned quantifiable results above their respective detection limits, excluding pairs in which culture reported < LOD but the molecular method detected microbial load (differential detection; Section 3.4). This yielded 14 quantifiable pairs at T4 across four indicator groups: aerobic mesophiles (n = 8), LAB (n = 4), Enterobacteriaceae (n = 1), and coliforms (n = 1) (Table 2).

Table 2. Paired data for Population A (n = 14): quantifiable pairs with both methods above LOD.

Indicator	Matrix	CFUeq/g	CFU/g	$\log_{10}$ CFUeq	$\log_{10}$ CFU	$\Delta\log_{10}$
Aer. mesophiles	Tuna_CON	2.1×10^7	2.1×10^7	7.32	7.32	0.00
Aer. mesophiles	Tuna_SIN	1.8×10^7	1.8×10^7	7.26	7.26	0.00
Aer. mesophiles	Cod_CON	3.0×10^7	$> 3.0 \times 10^7$	7.48	7.48	0.00
Aer. mesophiles	Cod_SIN	9.4×10^2	9.7×10^2	2.97	2.99	−0.01
Aer. mesophiles	Hake	3.0×10^7	$> 3.0 \times 10^7$	7.48	7.48	0.00
Aer. mesophiles	Baby_squid	3.0×10^7	$> 3.0 \times 10^7$	7.48	7.48	0.00
Aer. mesophiles	Octopus	3.0×10^7	$> 3.0 \times 10^7$	7.48	7.48	0.00
Aer. mesophiles	Rejo	3.0×10^7	$> 3.0 \times 10^7$	7.48	7.48	0.00
LAB	Hake	2.8×10^7	2.8×10^7	7.45	7.45	0.00
LAB	Baby_squid	1.2×10^7	1.2×10^7	7.08	7.08	0.00
LAB	Octopus	3.0×10^7	$> 3.0 \times 10^7$	7.48	7.48	0.00
LAB	Rejo	2.1×10^7	2.1×10^7	7.32	7.32	0.00
Enterobact.	Cod_SIN	9.0×10^2	9.0×10^2	2.95	2.95	0.00
Coliforms	Cod_SIN	5.4	6.1×10^2	0.73	2.79	−2.05

Lin's concordance correlation yielded CCC = 0.941 (95% CI: 0.819–0.982), indicating substantial agreement (McBride, 2005). Pearson r was 0.970; accuracy correction factor Cb = 0.970 (Table 3).

 doi:10.20944/preprints202605.0325.v1

Table 3. Concordance metrics — Population A (n = 14).

Metric	Value	95% CI
Lin's CCC	0.941	0.819–0.982
Pearson r	0.970	—
Accuracy Cb	0.970	—
Passing-Bablok slope	1.296	0.877–2.305
Passing-Bablok intercept	−2.391	−9.841 to +0.703
CUSUM linearity test	7.00 > 5.09 (rejected)	—
Bland-Altman bias	−0.254 $\log_{10}$	−0.557 to +0.049
B-A 95% limits of agreement	−1.389 to +0.881 $\log_{10}$	—
B-A proportional bias slope	+0.21	p = 0.012
R^2	0.900	—
RMSE	0.613 $\log_{10}$	—
MAE	0.302 $\log_{10}$	—
Skill score vs. null model	+0.900	—
% within ±0.5 $\log_{10}$	78.6%	—
% within ±1.0 $\log_{10}$	92.9%	—

The Passing-Bablok regression estimated a slope of 1.296 (95% CI: 0.877–2.305) and an intercept of −2.391 (95% CI: −9.841 to +0.703). The confidence intervals of both parameters include identity (slope = 1, intercept = 0), so the hypothesis of method equivalence is not rejected. The CUSUM test rejected linearity (7.00 > 5.09 critical), likely influenced by an outlier identified in coliforms (see below).

The Bland-Altman analysis showed a mean bias of −0.254 $\log_{10}$ CFU/g (95% CI: −0.557 to +0.049), which includes zero and is therefore not statistically significant. The 95% limits of agreement were −1.389 to +0.881 $\log_{10}$. A weak proportional bias was detected (slope +0.21, p = 0.012), indicating that the slight molecular underestimation tends to increase at higher loads — an expected effect when culture operates near its counting ceiling (> 3.0×10^7 CFU/g).

One outlier was identified by IQR analysis: the coliform pair in Cod_SIN-T4 (CFUeq = 5.4 vs. culture = 610 CFU/g; Δlog = −2.05). This inverse discrepancy — where culture exceeds the molecular estimate — is consistent with non-specific growth on VRBL selective medium of organisms that are not true coliforms, a documented phenomenon in samples with advanced spoilage flora (Wu, 2008).

3.3. Aerobic Mesophile Concordance: Resolution Across 5 Orders of Magnitude

For aerobic mesophiles — integrating all detected bacterial taxa — concordance at T4 (n = 8) was exceptional. Lin's CCC reached 1.000 (confidence interval not computable owing to zero residual variance); Passing-Bablok regression was similarly not computable, as all points fell on the identity line.

Bland-Altman analysis yielded a bias of −0.002 $\log_{10}$ (SD = 0.005 $\log_{10}$; 95% limits of agreement: −0.013 to +0.009 $\log_{10}$) — a range 200-fold narrower than the regulatory ±1 $\log_{10}$ tolerance. MAE was 0.002 $\log_{10}$, RMSE 0.005 $\log_{10}$, and the molecular/culture ratio ranged from 0.969 (Cod_SIN) to 1.000 (7 of 8 matrices).

These results span a dynamic range of 5 orders of magnitude: from 9.4×10^2 CFUeq/g (Cod_SIN, the only sample with low load at T4) to 3.0×10^7 CFUeq/g (multiple matrices at the counting ceiling). The perfect concordance in Cod_SIN — the most informative sample being far from the counting ceiling — confirms that the concordance is not an artifact of the ceiling effect.

A caveat applies: 7 of 8 culture counts for aerobic mesophiles were reported as > 3.0×10^7 CFU/g (upper counting limit of ISO 4833-1:2013). Concordance in these samples partially reflects that both

 doi:10.20944/preprints202605.0325.v1

culture and DELOREAN converge at the same maximum value, and resolution above this threshold cannot be assessed. Cod_SIN (9.7×10^2 CFU/g vs. 9.4×10^2 CFUeq/g; ratio = 0.969) is the only sample providing a quantitative comparison free of this ceiling effect.

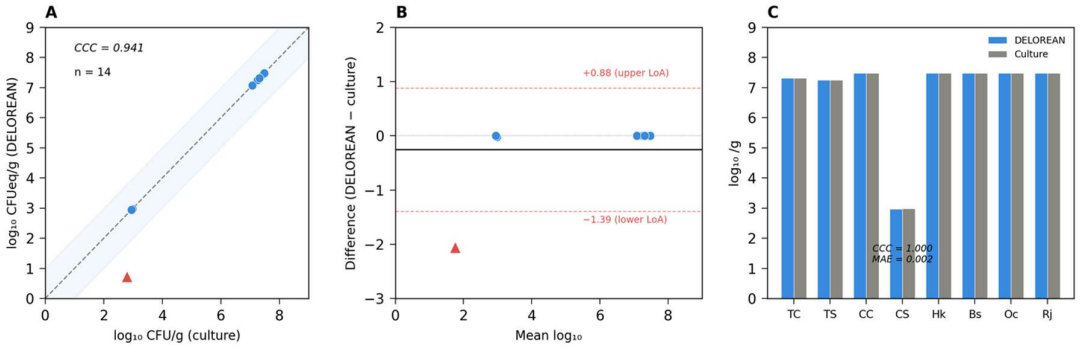

Figure 1. Concordance. (A) Scatter $\log_{10}$ CFUeq vs CFU (n=14); gray=±1 $\log_{10}$. (B) Bland-Altman; bias=−0.25. (C) Aerobic mesophiles paired bars (T4). Red=outlier.

3.4. Molecular Detection of Non-Culturable Populations

Of 40 pairs evaluated at T4, 24 (60%) showed culture < LOD (< 10 CFU/g) alongside DELOREAN estimates of 10^1–10^7 CFUeq/g. These discrepant pairs spanned three indicator groups: coliforms (7/8 matrices), Enterobacteriaceae (7/8), and sulfite-reducing clostridia (8/8), plus 3 LAB pairs (Table 4).

Table 4. Discrepant pairs — Population B: culture < LOD with molecular detection.

Indicator	n matrices	Mean molecular bias ($\log_{10}$)	Range ($\log_{10}$)	Proposed mechanism
Sulfite-reducing clostridia	8/8	+5.9	+4.2 to +6.85	Spore DNase impermeability
Coliforms	7/8	+4.6	+3.1 to +5.8	Sublethally damaged Gram-negatives
Enterobacteriaceae	7/8	+4.0	+2.3 to +5.4	Loss of metabolic selectivity
LAB	3/8	+2.8	+2.3 to +3.5	VBNC state

Mean molecular bias was +4.7 $\log_{10}$ (range: +2.3 to +6.85). The discrepancy was greatest for sulfite-reducing clostridia (+5.9 $\log_{10}$), followed by coliforms (+4.6 $\log_{10}$) and Enterobacteriaceae (+4.0 $\log_{10}$) — a hierarchy consistent with the biological properties of each group: *Clostridium* spores resist DNase degradation due to their multilamellar structure, so the vtMG protocol detects DNA from inactivated spores retaining intact membranes but metabolically inert.

The pattern of negative proportional bias (Bland-Altman slope = −0.72, p = 0.0014 in the global analysis) confirms that the discrepancy increases as culture decreases — exactly as expected if the molecular signal originates from cells with intact membranes but unable to grow on selective media.

The biological interpretation of this discrepant population distinguishes three non-exclusive mechanisms:

(i) *Sublethal HPP damage.* Bacterial cells subjected to high pressure may suffer sufficient damage to their metabolic or envelope structures to prevent colony formation on selective media (VRBG, VRBL), while retaining sufficient membrane integrity to protect their DNA from DNase degradation (Huang et al., 2014; Bozoglu et al., 2004).

(ii) *Inactivated spores. Clostridium* spp. spores possess multiple layers (cortex, coat, exosporium) that render them impermeable to DNase even when metabolically inactive. The vtMG protocol would detect DNA from non-viable spores as if they were viable.

 doi:10.20944/preprints202605.0325.v1

(iii) *Loss of metabolic selectivity.* Enterobacteriaceae sublethally damaged by HPP may retain membrane integrity but lose the ability to ferment glucose on VRBG or lactose on VRBL, a requirement for their detection by culture.

These molecularly detectable but non-culturable populations represent a finding of health relevance: they constitute a reservoir of organisms with cellular integrity that could, under favorable conditions, recover growth capacity.

3.5. Microbial Succession in Double-Treated RTE Seafood

Taxonomic profiles obtained by DELOREAN across the four time points revealed a coherent and reproducible microbial succession across all 8 matrices, with four differentiated phases.

3.5.1. Phase I — Post-Treatment Sterility and Recontamination (T1, 0 Days)

At T1, most matrices carried extremely low (< 10^3 CFUeq/g) or undetectable loads. Quantifiable samples were dominated by post-process recontamination psychrotrophs: *Psychrobacter* (97.7–97.9%) in Tuna_CON/SIN, Hake, and Rejo; *Pseudomonas* (97.5%) in Cod_SIN. Diversity was minimal (Shannon H' = 0.13; Pielou J' = 0.09–0.12; 3–4 effective taxa), indicating single-colonizer dominance.

Clostridium was detected at low levels (10^1 CFUeq/g) in all quantifiable samples, consistent with persistence of spores surviving the double treatment.

All samples were COMPLIANT for all regulatory indicators.

3.5.2. Phase II — Stable Colonization at Low Load (T2, 45 Days)

At T2, biomass increased moderately (DNA 0.01–2.00 ng/g) while maintaining the same dominant taxa. The community expanded slightly with the appearance of *Acinetobacter* (1.4%) and *Enterococcus* as the first detectable LAB (100% of the LAB fraction in samples that presented it: Tuna_CON/SIN, Hake, Rejo). Clostridia remained stable at 10^1 CFUeq/g.

Diversity remained low (H' = 0.26–0.31; J' = 0.17–0.19). All samples remained COMPLIANT.

3.5.3. Phase III — Ecological Collapse and Community Explosion (T3, 75 Days)

Between T2 and T3, biomass increased 10^2–10^3-fold (DNA 46–170 ng/g), and total load reached 10^7 CFUeq/g across all matrices. Three ecological guilds emerged simultaneously:

Spore-formers: Clostridium spp. (sensu stricto groups 1, 5, 13), *Sporosarcina* (61.5% in Cod_CON), *Psychrobacillus*. These organisms, resistant to thermal treatment in spore form, germinated and proliferated once refrigeration conditions and absence of competitors permitted.

Thermotolerant LAB: Carnobacterium (dominant in Hake and Octopus, 25–81% of the LAB fraction), *Enterococcus* (dominant in Rejo and Baby_squid). The predominance of *Carnobacterium* — the most baro- and thermotolerant LAB — over the typical HPP-alone genera (*Leuconostoc*, *Lactobacillus*) confirms that the double treatment selects a qualitatively distinct LAB community.

Recontamination psychrotrophs: Psychrobacter and *Acinetobacter* persisted as significant components (5–87%) of the community.

Each matrix developed a differentiated taxonomic profile: Tuna_CON was co-dominated by *Clostridium* and *Psychrobacter*; Cod_CON by *Sporosarcina*; Hake by *Psychrobacter* + *Carnobacterium* + *Clostridium*; Cod_SIN by *Pseudomonas* + *Clostridium*; and Rejo by *Lachnoclostridium* (90.8%), a strict anaerobe of the family Lachnospiraceae from the cephalopod intestinal tract.

Diversity increased markedly (H' = 0.56–1.87; 35–75 taxa), with the highest richness in cephalopod samples (Baby_squid: 75 taxa; Octopus: 55 taxa).

All matrices were NON-COMPLIANT for aerobic mesophiles. Enterobacteriaceae (10^4–10^5 CFUeq/g), coliforms (*Klebsiella*, *Enterobacter*, *Serratia*, *Citrobacter*), and pathogen alerts were detected: *C. perfringens* (3.2×10^5 CFUeq/g in Tuna_CON), *Salmonella* (1.6×10^4 in Octopus), *Escherichia-Shigella* (multiple matrices).

 doi:10.20944/preprints202605.0325.v1

3.5.4. Phase IV — Late Spoilage Community (T4, 90 Days)

At T4, diversity reached its maximum (H' up to 2.83; 165 taxa in Hake) with complex communities where spore-formers dominated most matrices: *Clostridium sensu stricto 5* (52.5% in Tuna_CON), *Clostridium sensu stricto 13* (40.7–59.7% in Tuna_SIN and Cod_CON), with *Luteococcus* as an unexpected and consistent presence (6–63%).

LAB diversified: *Carnobacterium*, *Enterococcus*, *Lactobacillus*, *Weissella*, and *Streptococcus* coexisted in most matrices, with LAB loads of 10^5–10^7 CFUeq/g.

Enterobacteriaceae reached 10^5–10^6 CFUeq/g — dominated by *Serratia*, *Klebsiella*, *Enterobacter*, and *Yersinia* — although culture reported them as < LOD in 7 of 8 matrices (Section 3.4).

3.6. Effect of Preservative Additive

Matrices with additive (Tuna_CON, Cod_CON) and without additive (Tuna_SIN, Cod_SIN) showed differentiated ecological trajectories. In Cod, the effect was particularly marked: Cod_SIN maintained the lowest load of the entire study at T4 (940 CFUeq/g vs. 3.0×10^7 in Cod_CON), with a community dominated by *Clostridium* and low diversity, while Cod_CON developed a complex community dominated by *Clostridium sensu stricto 13* and *Escherichia-Shigella* with the highest Enterobacteriaceae load of the study (3.2×10^6 CFUeq/g).

In Tuna, differences were less dramatic but consistent: CON showed higher dominance of *Clostridium sensu stricto 5* (52.5%) and lower *Clostridium sensu stricto 13* relative to SIN (40.7%).

The additive did not prevent final spoilage but modulated the ecological trajectory, dominant taxonomic composition, and timing of microbial explosion.

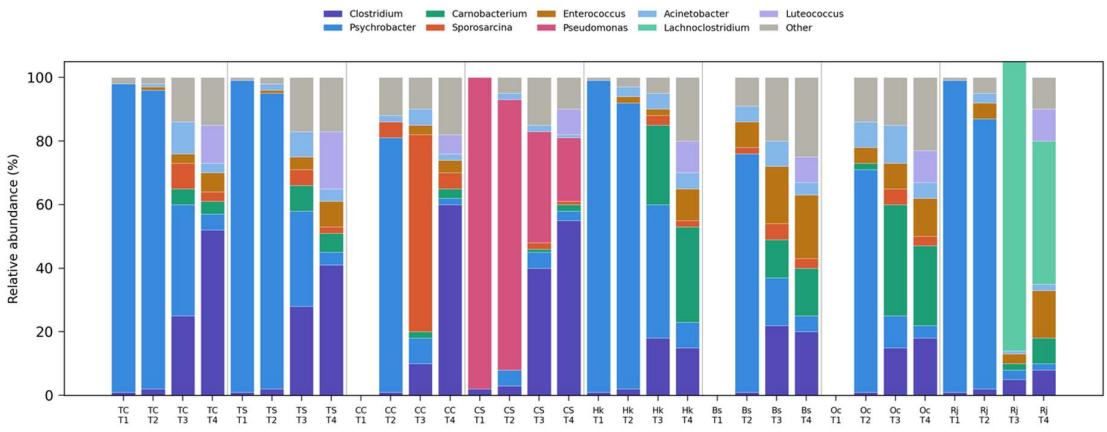

Figure 2. Community composition (stacked bars), 8 matrices × 4 time points. Copy-number-corrected relative abundance (%).

3.7. Pathogen Surveillance

Cross-referencing of all detected taxa against the foodborne pathogen list (Reg. EC 2073/2005 Annex I + EFSA Opinions) identified pathogen signals in 13 of 32 panels (Table 5). No pathogens were detected at T1 or T2. Detections were concentrated at T3 and T4:

 doi:10.20944/preprints202605.0325.v1

Table 5. Pathogen alert inventory across all panels.

Pathogen	Time	Matrices	CFUeq/g range	Culture result
C. perfringens	T3	Tuna_CON	3.2×10^5	Not tested
C. perfringens	T4	5/8 matrices	$2.1 \times 10^2 – 1.9 \times 10^6$	< LOD (all)
Salmonella spp.	T3	Octopus	1.6×10^4	Not tested
Salmonella spp.	T4	6/8 matrices	$3 \times 10^2 – 10^5$	Negative (all)
Cronobacter	T4	Tuna_CON, Cod_CON	$5.7 \times 10^3 – 4.7 \times 10^4$	Not tested
S. aureus	T4	Rejo	9.0×10^3	Not tested
Escherichia-Shigella	T3–T4	Multiple	$9.2 \times 10^3 – 2.1 \times 10^6$	< LOD (*E. coli*)
L. monocytogenes	—	None	Not detected	Not detected

The V4–V9 amplicon (515F/1492R) includes the V6–V9 regions, which provide the highest discriminatory power for *Clostridium* and *Staphylococcus* (Johnson et al., 2019) — the dominant taxa in this study — but lacks V1–V3, the subregion with the best resolution for the *Escherichia/Shigella* complex (Johnson et al., 2019). However, even the complete 16S gene (V1–V9) cannot reliably discriminate *E. coli* from *Shigella*, as they share > 99% sequence identity and may present identical 16S sequences (Chakravorty et al., 2007). All detections of this complex are reported at the group level and require confirmation by alternative markers.

All pathogen detections are flagged in DELOREAN panels with a recommendation for confirmation by specific PCR or ISO reference method, following the principle that molecular screening does not replace the confirmatory assay but rather guides it.

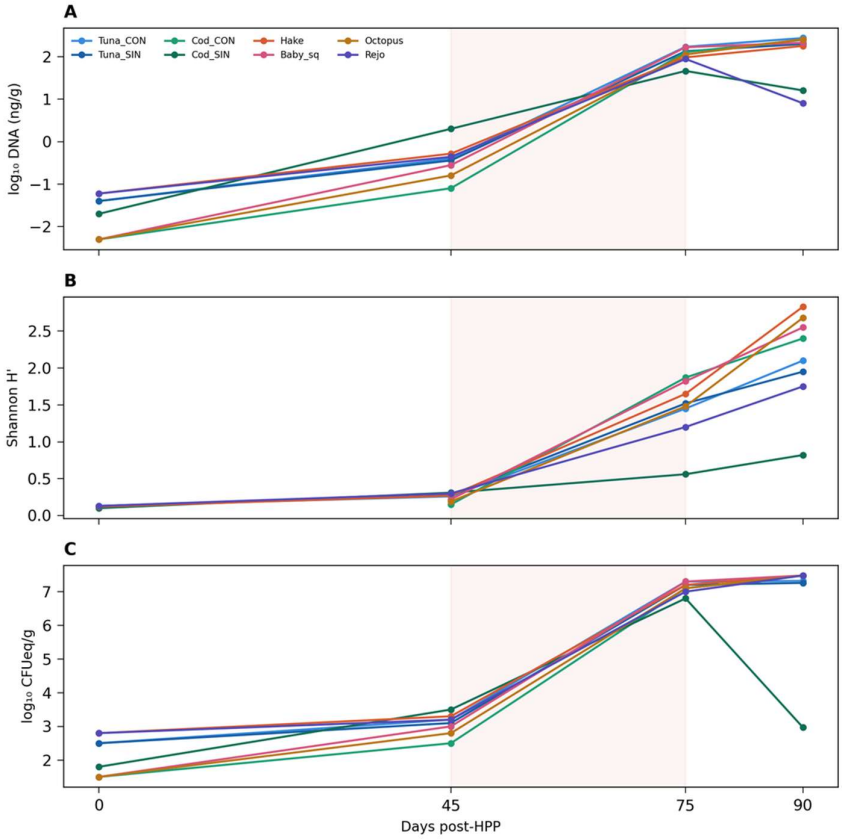

Figure 3. Temporal trajectory. (A) DNA ($\log_{10}$ ng/g); (B) Shannon H'; (C) Aerobic mesophiles ($\log_{10}$ CFUeq/g). Red shading = critical window (days 45–75).

 doi:10.20944/preprints202605.0325.v1

3.8. Global Validation Metrics (Transparency)

For transparency, concordance metrics were calculated on the full set of 40 T4 pairs (Table S1, supplementary). Global Lin's CCC was 0.167 (95% CI: −0.152 to 0.455), Bland-Altman bias +3.07 $\log_{10}$ (95% CI: 2.16–3.99), and RMSE 4.24 $\log_{10}$. R^2 against the identity model was −4.58; the skill score against the null model was equally negative.

These poor global metrics arise from the inclusion of 24 pairs (60%) in which culture reported < LOD while the molecular method detected 10^1–10^7 CFUeq/g — pairs representing differential detection of VBNC/sublethally damaged populations (Section 3.4), not algorithm failure. Stratification into quantifiable pairs (Section 3.2; CCC = 0.941) and differential detection pairs (Section 3.4) reveals that the unstratified metrics conflate two biologically distinct populations, and reporting them without context would misrepresent method performance.

4. Discussion

4.1. DELOREAN as a Quantitative Bridge: Scope and Limitations of Concordance

The validation results demonstrate that DELOREAN reproduces traditional culture counts with high precision when both methods operate within their quantifiable ranges. Lin's concordance (CCC = 0.941) on the 14 quantifiable pairs, with a non-significant Bland-Altman bias (−0.25 $\log_{10}$, CI includes zero) and 93% of pairs within ±1 $\log_{10}$, places DELOREAN in the performance range of validated alternative methods according to ISO 16140-2:2016, which establishes ±0.5 $\log_{10}$ as the acceptability criterion for quantitative methods. Seventy-nine percent of pairs met this stricter criterion.

Concordance proved strongest for the aerobic mesophilic indicator (CCC = 1.000; MAE = 0.002 $\log_{10}$). This has a structural explanation: by integrating all bacterial taxa, the mesophilic indicator dilutes taxonomic misclassification, copy number prediction error, and regulatory group assignment error that may individually affect narrower indicators. In practice, aerobic mesophiles are the primary microbial load indicator in European regulation (Reg. EC 2073/2005), and their perfect concordance with culture has direct operational value.

Interpreting this concordance requires clarifying the status of the K calibration factors. These factors were derived in an independent pilot study on cod flakes (n = 6 paired vtMG + ISO culture samples, Section 2.5.4) and applied without modification to the BRAESP-SLS validation dataset comprising eight different matrices. Because the K values were not fitted on T4 data, the concordance results reported here constitute a legitimate *out-of-distribution* test: empirical priors from a previous experiment applied to a validation dataset never seen during calibration. This architecture is methodologically distinct from in-sample fitting and eliminates the circularity objection. The results further demonstrate *transferability* of K factors across RTE seafood matrices with double treatment — a non-trivial finding that supports broader applicability of the framework.

Two caveats qualify these results. First, the K values derive from a single source category (cod flakes). Although validation across seven additional matrices is positive, generalization to radically different product categories (meat, dairy, vegetables, beverages) will require dedicated paired calibrations. Second, 7 of 8 culture values for aerobic mesophiles reached the counting ceiling (> 3.0 × 10^7 CFU/g), so concordance in the high range partially reflects convergence of both methods at the same upper limit. Only Cod_SIN-T4 (molecular/culture ratio = 0.969 at 10^2) provides a comparison unaffected by this ceiling effect, and its concordance confirms that the framework performs equally at low loads.

4.2. The "Dark Flora": VBNC Populations as a Central Finding

The central finding of this study is not the concordance but the discrepancy: 60% of T4 pairs showed molecular detection of 10^3–10^7 CFUeq/g in samples where culture reported < LOD. This discrepancy is not random — it is systematic, reproducible across matrices, and concentrated in indicators that rely on selective media (VRBG, VRBL, TSN/SPS).

 doi:10.20944/preprints202605.0325.v1

The parsimonious interpretation is that the double treatment (pasteurization + HPP) generates a microbial population retaining sufficient membrane integrity to protect its DNA from DNase degradation, yet bearing sufficient metabolic or structural damage to prevent colony formation on selective media. Such sublethally damaged or VBNC populations are well documented in HPP products (Huang et al., 2014; Considine et al., 2008) but have rarely been quantified at the community scale.

The discrepancy gradient across indicator groups reinforces this interpretation. Sulfite-reducing clostridia show the highest bias (+5.9 $\log_{10}$), consistent with spore impermeability to DNase. Coliforms and Enterobacteriaceae show intermediate biases (+4.0 to +4.6 $\log_{10}$), consistent with sublethal HPP damage to Gram-negative membranes. LAB show the smallest discrepancy, consistent with their higher sensitivity to HPP and greater probability of complete cell death with membrane lysis (Considine et al., 2008; Huang et al., 2014).

This "dark flora" — molecularly detectable but invisible to culture — poses a regulatory question that the current framework cannot address: should microbiological criteria encompass only organisms capable of forming colonies, or also those retaining cellular integrity with potential for recovery? DELOREAN does not resolve this question, but it provides, for the first time, the quantitative tool to frame it.

4.3. Ecology of the Double Treatment: An Unprecedented Community

Microbial succession in RTE seafood subjected to pasteurization + HPP differs qualitatively from that described for HPP alone. In HPP-only products, spoilage microbiota is typically dominated by mesophilic LAB — *Leuconostoc*, *Lactobacillus*, *Weissella* — representing the flora surviving barometric treatment (Li et al., 2020). By contrast, double-treated products harbor a community dominated by three guilds that prior pasteurization selects:

First, spore-formers (*Clostridium* spp., *Sporosarcina*, *Psychrobacillus*), whose spores survive both heat and pressure and germinate once competition is eliminated. Second, thermotolerant LAB (*Carnobacterium*, *Enterococcus*), which survive both pasteurization and HPP due to their intrinsic resistance to both stresses. Third, post-process recontamination psychrotrophs (*Psychrobacter*, *Acinetobacter*, *Moraxella*), which do not survive the treatment but colonize the product in the post-packaging phase.

The near-complete absence of *Leuconostoc* and the dominance of *Carnobacterium* as the main LAB are the taxonomic markers distinguishing the double treatment ecology. *Carnobacterium* is recognized as the most baro- and thermotolerant LAB in refrigerated seafood products (Leisner et al., 2007; Afzal et al., 2010), and its dominance in the present study confirms that the double treatment selects a fundamentally different LAB subset from that characterizing HPP-only products.

Rejo maintained a singular profile throughout the study, with dominance of *Lachnoclostridium* (90.8% at T3) — a strict anaerobe of the Lachnospiraceae family, characteristic of the intestinal flora of marine vertebrates and invertebrates. This profile reflects persistence of the cephalopod digestive tract microbiota, which survives the double treatment presumably in spore form, and emerges as the dominant community in the absence of competition.

4.4. The Critical Shelf-Life Window

The T2-to-T3 transition (days 45–75 post-HPP) marks the collapse of microbiological control, with a 10^2–10^3-fold biomass increase occurring simultaneously across all 8 matrices. DELOREAN quantifies this transition with temporal resolution that conventional endpoint culture sampling cannot provide.

The practical implications are direct: the 90-day post-HPP shelf-life target exceeds the microbiological collapse point by at least 15 days for all evaluated matrices. An intermediate sampling at day 60 would have been informative to more precisely delimit this critical window, and is recommended for future shelf-life studies of double-treated products.

 doi:10.20944/preprints202605.0325.v1

4.5. Pathogen Surveillance: Molecular Screening as a Complement

DELOREAN detected *C. perfringens*, *Salmonella* spp., *Cronobacter*, *S. aureus*, and *Escherichia-Shigella* in samples where traditional culture was negative. The relevance of these detections is conditioned by two factors: the taxonomic resolution of the 16S marker and the biological state of the detected organisms.

Regarding taxonomic resolution: the V4–V9 amplicon (515F/1492R) was selected because the V6–V9 regions offer the best resolution for *Clostridium* and *Staphylococcus* (Johnson et al., 2019) — the dominant taxa in the studied communities — and because Matsuo et al. (2021) demonstrated that the V3-V4 region is insufficient for species-level identification of these genera. The cost of this choice is the absence of V1–V3, the subregion with the best resolution for the *Escherichia/Shigella* complex (Johnson et al., 2019). However, even the complete 16S gene (V1–V9) does not reliably discriminate *E. coli* from *Shigella*, as they share > 99% identity and can present identical sequences (Chakravorty et al., 2007). Detections of this complex are reported at the group level and require confirmation by alternative markers.

In light of these considerations, for general DELOREAN applications we recommend the use of primers amplifying the complete 16S gene (V1–V9, ~1500 bp) rather than subregions. The 27F/1492R combination is the standard for Nanopore and generates amplicons compatible with EMU and other long-read classifiers. Johnson et al. (2019) demonstrated that V1–V9 consistently produces the best species-level classification results across all evaluated taxa. To maximize taxonomic coverage, the degenerate version of the forward primer — S-D-Bact-0008-c-S-20 (5'-AGRGTTYGATYMTGGCTCAG-3'; Klindworth et al., 2013) combined with S-D-Bact-1492-a-A-22 (5'-CGGYTACCTTGTTACGACTT-3') — has been shown to capture significantly greater diversity than the conventional 27F from the ONT kit, which has three mismatches with *Bifidobacterium* and underestimates certain Lactobacillales (Waechter et al., 2023; Tandon et al., 2023). In the present study, 515F/1492R (V4–V9) primers were used because they were available in the sequencing kit used by the laboratory, and because the V6–V9 regions provide the best resolution for *Clostridium* — the dominant genus in post-double-treatment communities. Future DELOREAN studies will prioritize V1–V9 as the default configuration.

Regarding biological state: as discussed in Section 4.2, the vtMG protocol detects DNA from membrane-intact organisms, which includes VBNC cells and potentially inactivated spores. That *C. perfringens* or *Salmonella* are detected molecularly does not necessarily imply they are viable or pathogenically active, but it does indicate they maintain cellular integrity — a state from which resuscitation is biologically possible under favorable conditions.

The DELOREAN framework adopts an explicit position on this: pathogen detections are flagged as alerts requiring confirmation by reference methods (specific PCR, corresponding ISO), not as definitive regulatory results. This screening-followed-by-confirmation architecture is the most appropriate use of metagenomics in the current regulatory context.

4.6. Limitations

The present study has several limitations that should be considered when interpreting results:

(a) *Restricted origin of K factors.* The K calibration factors used in this study were derived from a prior, independent pilot set conducted exclusively on cod flakes (n = 6 pairs). Although this eliminates the circularity objection regarding the BRAESP-SLS validation dataset — the K values are empirical priors not fitted to T4 — it constrains the domain of demonstrated validity. The successful transfer of these factors across seven additional RTE seafood matrices (tuna loin, hake, baby squid, octopus, rejo), without matrix-specific recalibration, suggests that within the category of double-treated (thermal + HPP) seafood products the dominant sources of systematic bias are shared across matrices and adequately captured by a single set of indicator-level K values. Whether this transferability extends to seafood products with fundamentally different composition (high-fat species, smoked products) or to entirely different food categories (meat, dairy, vegetables, beverages) remains untested. Each such extension will require its own paired calibration dataset, ideally comprising

 doi:10.20944/preprints202605.0325.v1

multiple matrices and time points to estimate both the central K value and its inter-matrix variance. We envision DELOREAN's calibration architecture as inherently collaborative: as independent laboratories generate paired vtMG + culture data on new product categories, their K estimates can be pooled into a growing reference table analogous to the rrnDB for copy numbers — progressively expanding the framework's applicability without requiring any modification to the core algorithm.

(b) *Single replicate per sample-time combination.* The experimental design included a single analytical unit per matrix and time point (n = 1), without biological replicates. This reflects the industrial origin of the study — shelf-life trials in commercial production environments rarely afford the luxury of replicated destructive sampling at each time point — but it limits the ability to estimate within-sample variability for either the vtMG or the culture method. Consequently, the concordance metrics reported here characterize inter-method agreement across matrices and indicators, not repeatability within a single matrix. Two observations partially mitigate this concern. First, the consistency of the ecological succession pattern across all eight matrices — which were processed, packaged, and stored independently — provides a form of biological replication at the community level: the same phased trajectory (*Psychrobacter* → *Clostridium* + *Carnobacterium* → complex spoilage) emerged independently in matrices ranging from tuna loin to cephalopod tentacle, making it unlikely that the observed patterns are artifacts of stochastic sampling. Second, the aerobic mesophile concordance (CCC = 1.000 across 8 matrices spanning 5 orders of magnitude) would be improbable if within-sample variance were large relative to between-method variance. Nonetheless, formal repeatability and reproducibility studies with replicated designs are warranted before DELOREAN can be proposed for routine regulatory application.

(c) *Ceiling effect in aerobic mesophiles.* The perfect concordance in 7/8 samples is partially conditioned by the culture counting ceiling ($> 3.0 \times 10^7$). Cod_SIN is the only sample providing a comparison free of this artifact.

(d) *Yeasts and molds.* Not assessable by the 16S marker. Fungal flora detection would require incorporation of the ITS marker (Schoch et al., 2012).

(e) *Amplicon taxonomic resolution.* The V4–V9 amplicon (515F/1492R) is optimized for resolution of *Clostridium* and *Staphylococcus* (Johnson et al., 2019; Matsuo et al., 2021), the dominant taxa in this study, but lacks V1–V3, the subregion with the best resolution for *Escherichia*/*Shigella*. The *E. coli*–*Shigella* discrimination is an intrinsic limitation of the complete 16S gene (Chakravorty et al., 2007), exacerbated in our case by the absence of V1–V3.

(f) *Potential bias from differential lysis.* The literature reports that differential lysis protocols with saponin may differentially affect Gram-negative bacteria (Menghi et al., 2024). If this bias operates in vtMG, it could underestimate the Gram-negative proportion. However, the K calibration factors would absorb this systematic bias.

(g) *Copy number prediction.* The Random Forest model for taxa not represented in rrnDB introduces additional uncertainty. As argued in Section 2.6, the K factors absorb residual prediction errors.

5. Conclusions

1. DELOREAN provides a validated framework for converting 16S amplicon metagenomics data into CFUeq/g estimates compatible with the microbiological criteria of Regulation (EC) No 2073/2005.

2. When both methods operate within their quantifiable ranges, concordance with traditional culture is substantial (Lin's CCC = 0.941; non-significant Bland-Altman bias; 93% of pairs within ±1 $\log_{10}$). For aerobic mesophiles — the primary regulatory indicator — concordance is essentially perfect (CCC = 1.000; MAE = 0.002 $\log_{10}$).

3. The systematic discrepancy in selective indicators (coliforms, Enterobacteriaceae, sulfite-reducing clostridia) — where DELOREAN detects 10^3–10^7 CFUeq/g but culture reports < LOD — constitutes quantitative evidence of sublethally damaged or VBNC populations in double-treated products (pasteurization + HPP). These populations retain membrane integrity but lose culturability on selective media.

 doi:10.20944/preprints202605.0325.v1

4. The double treatment generates a microbial succession qualitatively distinct from that described for HPP alone, dominated by spore-formers (*Clostridium* spp., *Sporosarcina*), thermotolerant LAB (*Carnobacterium*, *Enterococcus*), and recontamination psychrotrophs (*Psychrobacter*, *Acinetobacter*), with near-complete absence of the mesophilic LAB typical of HPP-only products.

5. A critical shelf-life window is identified between days 45 and 75 post-HPP in which microbiological control is lost simultaneously across all 8 evaluated matrices, with a 10^2–10^3-fold biomass increase.

6. DELOREAN detects pathogen signals (*C. perfringens*, *Salmonella*, *Cronobacter*, *S. aureus*) in samples with negative culture. These detections are managed as alerts requiring confirmation by reference methods, positioning DELOREAN as a screening tool complementary to culture.

7. The K calibration factors employed are empirical priors derived from an independent pilot study on cod flakes and applied without modification to the validation dataset (eight diverse matrices), so that the results constitute a legitimate out-of-distribution test. Extension of the framework to other product categories (meat, dairy, vegetables) will require specific calibrations with dedicated paired data.

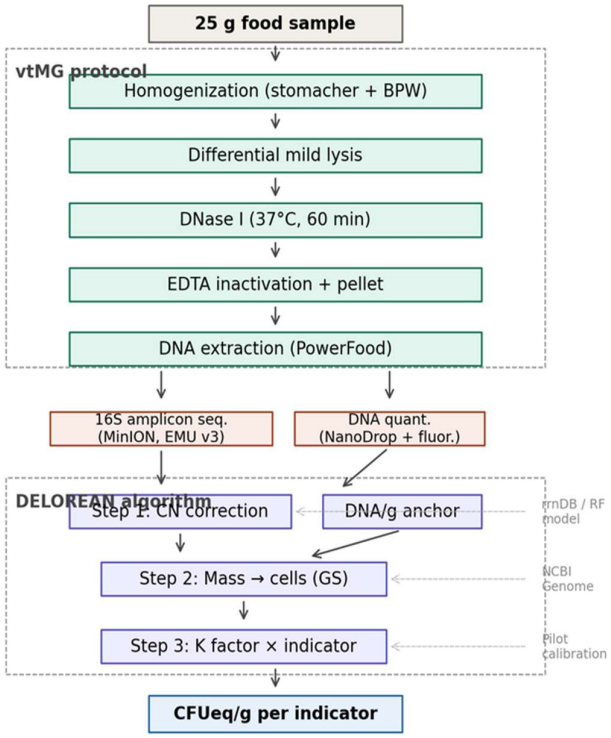

Figure 5. vtMG + DELOREAN pipeline. Green: vtMG (differential lysis → DNase → extraction). Purple: DELOREAN algorithm (CN correction → mass-to-cell → K calibration). Gray: external data sources.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. - Table S1: Global metrics (n = 40) with stratification explanation - Table S2: Calibrated K factors by indicator × matrix - Table S3: Random Forest model parameters and metrics - Table S4: Complete taxonomic composition (all samples, all taxa) - Table S5: DNA quantification data (spectrophotometric + fluorimetric) - Figure S1: Rarefaction curves - Figure S2: Principal coordinates analysis (PCoA) - Figure S3: Bland-Altman by individual indicator group

Author Contributions: Javier Gamboa: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Writing — Original Draft, Visualization **Marco Ferreira:** Resources, Data Curation, Project Administration, Writing — Review & Editing. All authors have read and agreed to the published version of the manuscript.

 doi:10.20944/preprints202605.0325.v1

Funding: This research received no external funding. Data obtained in the context of an industrial SLS (Brasmar Group SGPS, S.A.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Raw 16S rRNA amplicon sequencing data have been deposited in the NCBI Sequence Read Archive under BioProject PRJNA1454730. The DELOREAN algorithm is available at https://github.com/jgamboa-biotecno/DELOREAN (DOI: 10.5281/zenodo.19614013) under GPL-3.0 license. Complete culture data are provided in the manuscript Tables and supplementary material.

Conflicts of Interest: J.G. is an independent researcher affiliated with Biotecno.org; M.F. is an employee of the Quality Department of Brasmar Group SGPS, S.A. Analysis and interpretation were performed independently.

References

1. Afzal, M.I., et al. (2010). *Carnobacterium maltaromaticum*: identification, isolation tools, ecology and technological aspects in dairy products. *Food Microbiology*, 27(5), 573-579.
2. Angly, F.E., et al. (2014). CopyRighter: a rapid tool for improving the accuracy of microbial community profiles through lineage-specific gene copy number correction. *Microbiome*, 2, 11.
3. Bokulich, N.A., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10(1), 57-59.
4. Bowman, J.S., & Ducklow, H.W. (2015). Microbial communities can be described by metabolic structure: a general framework and application to a seasonally variable, depth-stratified microbial community from the coastal West Antarctic Peninsula. PLoS ONE, 10(8), e0135868.
5. Bozoglu, F., et al. (2004). Injury recovery of foodborne pathogens in high hydrostatic pressure treated milk during storage. *FEMS Immunology & Medical Microbiology*, 40(3), 243-247.
6. Chakravorty, S., et al. (2007). A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *Journal of Microbiological Methods*, 69(2), 330-339.
7. Charalampous, T., et al. (2019). Nanopore metagenomics enables rapid clinical diagnosis of bacterial lower respiratory infection. *Nature Biotechnology*, 37, 783-792.
8. Considine, K.M., et al. (2008). High-pressure processing — effects on microbial food safety and food quality. *FEMS Microbiology Letters*, 281(1), 1-9.
9. Curry, K.D., et al. (2022). Emu: species-level microbial community profiling of full-length 16S rRNA Oxford Nanopore sequencing data. *Nature Methods*, 19(7), 845-853.
10. De Coster, W. & Rademakers, R. (2023). NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*, 39(5), btad311.
11. De Filippis, F., et al. (2018). Recent past, present, and future of the food microbiome. *Annual Review of Food Science and Technology*, 9, 589-608.
12. Emerson, J.B., et al. (2017). Schrödinger's microbes: tools for distinguishing the living from the dead in microbial ecosystems. *Microbiome*, 5, 86.
13. Ercolini, D. (2013). High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Applied and Environmental Microbiology*, 79(10), 3148-3155.
14. Fittipaldi, M., et al. (2012). Progress in understanding preferential detection of live cells using viability dyes in combination with DNA amplification. *Journal of Microbiological Methods*, 91(2), 276-289.
15. Gayán, E., Govers, S.K., & Aertsen, A. (2017). Impact of high hydrostatic pressure on bacterial proteostasis. Biophysical Chemistry, 231, 3-9.
16. Hasan, M.R., et al. (2016). Depletion of human DNA in spiked clinical specimens for improvement of sensitivity of pathogen detection by next-generation sequencing. *Journal of Clinical Microbiology*, 54(4), 919-927.
17. Huang, H.-W., et al. (2014). Responses of microorganisms to high hydrostatic pressure processing. *Food Control*, 40, 250-259.

 doi:10.20944/preprints202605.0325.v1

18. Johnson, J.S., et al. (2019). Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. *Nature Communications*, 10, 5029.
19. Klindworth, A., et al. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41(1), e1.
20. Langille, M.G.I., et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, 31(9), 814-821.
21. Leisner, J.J., et al. (2007). *Carnobacterium*: positive and negative effects in the environment and in foods. *FEMS Microbiology Reviews*, 31(5), 592-613.
22. Li, H., Sun, X., Liao, X., & Gänzle, M. (2020). Control of pathogenic and spoilage bacteria in meat and meat products by high pressure: Challenges and future perspectives. Comprehensive Reviews in Food Science and Food Safety, 19(6), 3476-3500.
23. Louca, S., Doebeli, M., & Parfrey, L.W. (2018). Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome*, 6, 41.
24. Matsuo, Y., et al. (2021). Full-length 16S rRNA gene amplicon analysis of human gut microbiota using MinION nanopore sequencing confers species-level resolution. *BMC Microbiology*, 21, 35.
25. McBride, G.B. (2005). A proposal for strength-of-agreement criteria for Lin's concordance correlation coefficient. NIWA Client Report: HAM2005-062.
26. Menghi, M., et al. (2024). Saponin treatment for eukaryotic DNA depletion alters the microbial DNA profiles. *Scientific Reports*, 14, 4222.
27. Moussa, M., Espinasse, V., Perrier-Cornet, J.M., & Gervais, P. (2007). Damage in Escherichia coli cells treated with a combination of high hydrostatic pressure and subzero temperature. Applied and Environmental Microbiology, 73(20), 6508-6518.
28. Nkuipou-Kenfack, E., Engel, H., Fakih, S., & Nocker, A. (2013). Improving efficiency of viability-PCR for selective detection of live cells. Journal of Microbiological Methods, 93(1), 20-24.
29. Nocker, A., Sossa-Fernandez, P., Burr, M.D., & Camper, A.K. (2007). Use of propidium monoazide for live/dead distinction in microbial ecology. *Applied and Environmental Microbiology*, 73(16), 5111-5117.
30. Ritz, M., Tholozan, J.L., Federighi, M., & Pilet, M.F. (2001). Morphological and Physiological Characterization of *Listeria monocytogenes* Subjected to High Hydrostatic Pressure. *Applied and Environmental Microbiology*, 67(5), 2240-2247.
31. Schoch, C.L., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *PNAS*, 109(16), 6241-6246.
32. Stämmler, F., et al. (2016). Adjusting microbiome profiles for differences in microbial load by spike-in bacteria. *Microbiome*, 4, 28.
33. Starke, R., Pylro, V.S., & Morais, D.K. (2021). 16S rRNA gene copy number normalization does not provide more reliable conclusions in metataxonomic surveys. *Microbial Ecology*, 81(2), 535-539.
34. Stoddard, S.F., et al. (2015). rrnDB: improved tools for interpreting rRNA gene abundance in bacteria and archaea. *Nucleic Acids Research*, 43(D1), D593-D598.
35. Tandon, D., Dong, Y., & Hapfelmeier, S. (2023). Pipeline for species-resolved full-length 16S rRNA amplicon nanopore sequencing analysis of low-complexity bacterial microbiota. bioRxiv. doi:10.1101/2023.12.05.570138.
36. Tourlousse, D.M., et al. (2017). Synthetic spike-in standards for high-throughput 16S rRNA gene amplicon sequencing. *Nucleic Acids Research*, 45(4), e23.
37. Waechter, S., et al. (2023). Comparative analysis of full-length 16S ribosomal RNA genome sequencing in human fecal samples using primer sets with different degrees of degeneracy. *Frontiers in Genetics*, 14, 1213829.
38. Wu, V.C.H. (2008). A review of microbial injury and recovery methods in food. *Food Microbiology*, 25(6), 735-744.
39. Yang, D., Jiang, Z., Meng, Q., Wang, S., Pan, H., Rao, L., & Liao, X. (2023). Analyzing the pressure resistant, sublethal injury and resuscitable viable but non-culturable state population of *Escherichia coli*, *Staphylococcus aureus*, *Bacillus amyloliquefaciens* and *Lactiplantibacillus plantarum* under high pressure processing. *Food Research International*, 173(Pt 1), 113336.

 doi:10.20944/preprints202605.0325.v1