Food Chemistry: Molecular Sciences 2 (2021) 100023

Contents lists available at ScienceDirect

Food Chemistry: Molecular Sciences

journal homepage: www.elsevier.com/locate/fochms

A shotgun metagenomics approach to detect and characterize unauthorized genetically modified microorganisms in microbial fermentation products

Florence E. Buytaers [a,b], Marie-Alice Fraiture [a], Bas Berbers [a,b], Els Vandermassen [a], Stefan Hoffman [a], Nina Papazova [a], Kevin Vanneste [a], Kathleen Marchal [b,c], Nancy H.C. Roosens [a,1], Sigrid C.J. De Keersmaecker [a,1,*]

[a] *Transversal Activities in Applied Genomics, Sciensano, Brussels, Belgium*
[b] *Department of Plant Biotechnology and Bioinformatics, Ghent University, Belgium*
[c] *Department of Information Technology, IDlab, IMEC, Ghent University, Ghent, Belgium*

ARTICLE INFO

Keywords:
Shotgun metagenomics
Genetically modified microorganism
AMR
Long and short read sequencing
Microbial fermentation products
Identification

ABSTRACT

The presence of a genetically modified microorganism (GMM) or its DNA, often harboring antimicrobial resistance (AMR) genes, in microbial fermentation products on the market is prohibited by European regulations. GMMs are currently screened for through qPCR assays targeting AMR genes and vectors, and then confirmed by targeting known specific GM constructs/events. However, when the GMM was not previously characterized and an isolate cannot be obtained, its presence cannot be proven. We present a metagenomics approach capable of delivering the proof of presence of a GMM in a microbial fermentation product, with characterization based on the detection of AMR genes and vectors, species and unnatural associations in the GMM genome. In our proof-of-concept study, this approach was performed on a case with a previously isolated and sequenced GMM, an unresolved case for which no isolate was obtained, and a non-GMM-contaminated sample, all representative for the possible scenarios to occur in routine setting. Both short and long read sequencing were used. This workflow paves the way for a strategy to detect and characterize unknown GMMs by enforcement laboratories.

1 Introduction

A GMO (genetically modified organism) is defined as “an organism in which the genetic material has been altered in a way that does not occur naturally by mating and/or natural recombination” (Article 2 of Directive 90/220/EEC). GMOs include transgenic animals, plants and genetically modified micro-organisms (GMM). Food and feed products including enzymes or additives (e.g. vitamins) are often produced through the use of GMMs to replace chemical synthesis methods, as this is more practical and requires less resources (Deckers, Deforce et al., 2020). A specific microbial species is chosen for its suitability and ease of cultivation, and is then genetically modified to produce, for instance, the required compound in large quantities through microbial fermentation (Deckers, Deforce et al., 2020). Selection of the genetically modified (GM) strain(s) is often conducted based on selective growth with antibiotics. To this end, antimicrobial resistance (AMR) genes are often inserted in the modified strain, e.g. in the expression vector used for introduction of the new characteristic. However, the introduction of full-length AMR genes in microorganisms that might end up in food/feed products, poses a potential public health risk. Indeed, these can be easily transferred to other species, including pathogens, thereby leading to treatment failure (WHO, 2018). In Europe, the GMM’s authorization process for food enzymes and food additives (such as vitamins) falls under regulation EC 1331/2008, EC 1332/2008 and EC 1333/2008 (European Parliament and the Council of the European Union, 2008a, 2008b, 2008c). To allow that such microbial fermentation products can be produced with GMM in a contained environment, a confidential dossier must be submitted to the European commission, and EFSA performs a safety assessment. Moreover, these GMMs should be used only to produce microbial fermentation products and, in contrast to GM crop

* Corresponding author.
E-mail addresses: Florence.Buytaers@sciensano.be (F.E. Buytaers), Marie-Alice.Fraiture@sciensano.be (M.-A. Fraiture), Bas.Berbers@sciensano.be (B. Berbers), Els.Vandermassen@sciensano.be (E. Vandermassen), Stefan.Hoffman@sciensano.be (S. Hoffman), Nina.Papazova@sciensano.be (N. Papazova), Kevin.Vanneste@sciensano.be (K. Vanneste), Kathleen.Marchal@ugent.be (K. Marchal), Nancy.Roosens@sciensano.be (N.H.C. Roosens), Sigrid.DeKeersmaecker@sciensano.be (S.C.J. De Keersmaecker).

[1] Equally contributed.

https://doi.org/10.1016/j.fochms.2021.100023
Received 12 January 2021; Revised 8 April 2021; Accepted 17 April 2021
Available online 21 April 2021
2666-5662/© 2021 The Author(s). Published by Elsevier Ltd.

plants, are not intended for human and animal consumption. Therefore, these GMMs do not fall under EU regulations 1829/2003 and 1830/2003 for GMOs, and no dossier has to be submitted to the EU for authorization of the commercialization of GM food and feed for a specific GMM (European Parliament and the Council of the European Union, 2003a, 2003b). Consequently, the detection of an unexpected contamination by such a GMM in food and feed is per se unauthorized. This means that zero tolerance, including for its associated recombinant DNA, must be applied, i.e. neither viable cells nor DNA from the producer strain can be detected in the final commercialized product (Silano et al., 2019). Additionally, the companies producing these GMMs do not have to provide to the EU a method to identify them. This is in contrast with what is foreseen by regulations 1829/2003 and 1830/2003 for GM crop plants intended for the food and feed chain. Therefore, no detection/identification method for the GMM is available for enforcement laboratories. As several fermentation products were already shown to be contaminated by a living GMM or its DNA (Barbau-Piednoir et al., 2015; Fraiture et al., 2020; Paracchini et al., 2017), this calls for a proper control by enforcement laboratories.

Real-time polymerase chain reaction (qPCR) assays are the mandatory method for enforcement laboratories to screen and identify GM organisms (GMOs) in EU legislation 1829/2003 and 1830/2003. The aim is to ensure freedom of choice for the consumer by detecting unlabeled GMOs as well as the safety of food and feed. These qPCR assays target specific GM events, i.e. the insertion of the GM element(s) in the host genome which leads to unnatural associations. However, for GMM that are not intended for food and feed consumption, until recently no official methods were available that allowed their detection and characterization, as elaborated above. A novel strategy to detect and identify GMM in food and feed products was only recently developed. First, a qPCR screening is performed based on the detection of AMR genes and expression/shuttle vector carrying AMR genes. Hereto, a variety of qPCR tests have been developed, targeting the *cat*, (Fraiture, Deckers et al., 2020b), *aadD* (Fraiture, Deckers et al., 2020a) and *tet* (Fraiture, Deckers et al., 2020c) genes, conferring a chloramphenicol, kanamycin and tetracycline resistance respectively, and targeting the shuttle vector pUB110 carrying *aadD* (Fraiture, Bogaerts et al., 2020). These qPCR tests can be complemented with conventional PCR methods followed by Sanger sequencing. This will allow obtaining additional information on the presence of microbial DNA and their species/genus identification, using for example 16S rRNA or ITS-based methods (Deckers, Vanneste, Winand, Hendrickx et al., 2020; Deckers, Vanneste, Winand, Keersmaecker et al., 2020), as well as on the presence of full-length AMR genes (Fraiture et al., 2021). Demonstrating the presence of the full-length AMR gene is valuable information for risk assessment on the potential spread of this gene to other microorganisms in the environment including the human/animal gut after ingestion. If the screening based on AMR genes and/or vector is positive, thereby raising a strong suspicion of the presence of a GMM, a second line of analysis should be performed. This has the goal to both target unnatural associations (construct- or event-specific methods) and identify the GMM, thereby proving its presence in the microbial fermentation product (Barbau-Piednoir et al., 2015; Fraiture, Bogaerts, et al., 2020; Fraiture, Papazova, & Roosens, 2020; Paracchini et al., 2017). However, unlike it is the case for authorized GMOs as requested by regulations 1829/2003 and 1830/2003, no event/construct-specific method has been provided *a priori* by the producing companies to identify these GMMs.

If no second line qPCR analysis is available, the proof of the presence of a GMM can be obtained through whole genome sequencing (WGS) of a microbial isolate obtained from the fermentation product. This allows the identification of the unnatural association (Fraiture, Bogaerts et al., 2020). The knowledge of the DNA sequence can then lead to the future development and validation of a targeted GM-specific qPCR assay to be used in the identification step. This was the case for the identification method targeting a GM *Bacillus* overproducing vitamin B2 (RASFF2014. 1249) and one overproducing a protease (RASFF2019.3332) (Barbau-Piednoir et al., 2015; Fraiture et al., 2020). The GMM isolation process can however be arduous as the species and therefore the culture conditions are unknown. Moreover, isolation is not always possible, if the GMM is non-viable or non-culturable. Genetic modifications requiring the presence of a growth factor in order to culture the GMM are often encountered and unknown to the enforcement laboratories. Multiple species can also be present, and one of them can be missed by culturing. In other cases, only DNA of the GMM is present in the fermentation product. When no isolate can be obtained, a culture independent strategy has to be performed. For instance, a DNA walking method, as a targeted sequencing approach, can be used to detect unnatural associations. However, in order to apply this strategy, a minimum of knowledge is required. Indeed, the DNA walking strategy needs to anchor on a known sequence, like ARM genes and vector detected via the first line qPCR screening, in order to be able to characterize unknown flanking regions. Moreover, the DNA walking strategy can be time-consuming when regions of several kbps need to be covered, requiring successive DNA walking assays of each approximatively 1 or 2 kbps. A DNA walking strategy anchored on the pUB110 vector was previously used to identify the GM *Bacillus* overproducing alpha-amylase (RASFF2019.3332) (Fraiture, Papazova et al., 2020). Similarly as for the WGS approach, this can then subsequently lead to the design of new event/construct-specific methods (Fraiture, Papazova et al., 2020). Until now, WGS on isolates and DNA walking have enabled the development of qPCR methods allowing to identify 3 GMM constructs. However, in all the other scenarios, no fast and universal method is available to detect the presence of a GMM in a sample. This constitutes a major bottleneck for current GMM control, as many applications involving GMM are submitted to the European commission (for example, over a hundred dossiers for food enzymes mention the use of GMMs (Deckers, Deforce et al., 2020)).

A method not requiring prior isolation nor prior knowledge on the sequences, detecting all genes, including potential unnatural associations and potential species identification at once in a sample, would pave the way towards an open approach of generalized detection and characterization of unknown GMMs in microbial fermentation products. A shotgun metagenomics approach, i.e. sequencing all DNA from a sample, allows detection of any gene of interest as well as the detection of the species. It can also potentially reconstruct (partially) the genome of the strain(s) present, allowing to identify unnatural associations. This technology has been previously described for the successful characterization of food-borne pathogens at the strain level after a culture-based enrichment (Buytaers et al., 2020; Leonard et al., 2016). However, although the application to the field of GMM characterization is linked to a lower complexity of the microbial communities, some bottlenecks need to be addressed. First, not performing any enrichment is preferred to avoid the issue of species-specific growth conditions and non-viable GMM. Therefore, the shotgun metagenomics sequencing should be done in sufficient depth to allow for data analysis. This puts constraints on the cost-efficiency of the approach. Second, as the output of the metagenomics sequencing is a mix of reads representing all DNA present in the sample, putting the puzzle together to the species' genomic level, including detecting the unnatural association, is not straightforward. Short-read Illumina sequencing (max 300 bp reads), already described for metagenomics approaches in food using reference genome databases (Buytaers et al., 2020; Leonard et al., 2015), might not be sufficient for GMM, where this sequence information is largely missing. Long read sequencing such as offered by Oxford Nanopore Technologies (ONT) might facilitate the reconstruction of the genome (Somerville et al., 2018), especially with unnatural constructions such as GMMs. Moreover, it can help to obtain the unnatural associations or the full-length AMR gene, which are usually longer than 300 bp, on a single read. Several flow cells are currently on the market for this technology, e.g. the conventional MinION flow cell, and the newly released lower-output but also lower-cost Flongle flow cell, requiring half of the starting DNA material. As the metagenomics approach is still very expensive, the use of cheaper sequencing consumables such as the Flongle might contribute to reducing the price of the analysis while keeping a sufficient level of information (Grädel et al., 2019). However, ONT has been described to have a higher error rate as compared to short read sequencing, which could affect the results (Kono & Arakawa, 2019). This metagenomics approach has not yet been applied within the GMM field.

In this study, we present the first attempt to develop a strategy based on shotgun metagenomics for the general detection and characterization of GMMs in microbial fermentation products. Hereby, we envisaged to determine if and which AMR genes and shuttle vectors are present, and simultaneously provide information to identify the species present in the sample and to unequivocally prove the presence of the GMM by characterizing unnatural associations in its genome. To deliver a proof-of-concept of our approach, we have selected three samples, representative of the possible scenarios to occur in a routine setting, i.e. a previously analyzed sample containing a GMM *Bacillus subtilis* overproducing vitamin B2 (riboflavin), isolated and fully characterized at that time (RASFF 2014.1249) (Barbau-Piednoir et al., 2015; Berbers et al., 2020; Paracchini et al., 2017), a sample positive for some qPCR markers but for which no isolate could be obtained and a sample with no GMM contamination. The short and long read sequencing technologies were compared for their performances, including the newly released Flongle, as a smaller and cost-effective alternative. The most appropriate data analysis workflow was considered, depending on the sample type and applied sequencing technology.

2. Hypothesis

A shotgun metagenomics approach using short or long read sequencing is capable of detecting and (partially) characterizing unauthorized genetically modified microorganisms present in microbial fermentation products.

3. Materials and methods

3.1 DNA extraction and qPCR

Three samples of vitamin B2 (riboflavin) were investigated: one sample from 2014 containing a living GMM *Bacillus* strain (GMM14, RASFF 2014.1249), one sample containing DNA but negative to the previously developed qPCR methods (see paragraph below and Table 1) targeting AMR markers typical of GMMs and event-specific targets of the 2014 strain (GMMneg) and one sample from 2016 containing DNA corresponding to features of the 2014 strain (GMM16), but for which no strain could be isolated.

DNA was directly extracted from the vitamin powders without culture-based enrichment. Briefly, 200 mg of the sample was used

Table 1

Characterization of GMM samples and bacterial isolates. A: DNA concentration and integrity, qPCR and PCR results B: detection results (AMR genes, pUB110) after isolate (168 and 3557) or metagenomics sequencing.

A

Sample	DNA concentration (Qubit, ng/µl)	DNA integrity number	Cq qPCR RASFF2014			Cq qPCR *cat*	Cq qPCR *aadD*	Cq qPCR *tet-L*	PCR *tet-L* full gene
			vitB2-UGM	558	Cq qPCR 693				
***B. subtilis* 168**	820	9.8	nd	38,02	nd	nd	nd	nd	nt
***B. subtilis* isolate 3557 (RASFF2014)**	412	8.9	16.62	18.64	24.5	22.94	21.3	16.46	+
GMM14	104	1.9	18.55	19.1	25.39	23.8	22.62	17.94	+
GMM16	13.3	1	23.69	23.38	nd	28.15	28.8	32.72	-
GMMneg	too low to detect	0	nd	nd	nd	nd	nd	nd	-

B

Description	**pUB110**	**ampR1¨**	**ampR2¨**	**bleoR¨**	**cmR1¨**	**EryR-1¨**	**kanR1¨**	**tetR1¨**
gene		***bla***	***bla***	***ble***	***cat***	***erm B***	***aadD***	***tet-L***
Sequenced sample	target coverage (%)	target coverage (%)	target coverage (%)	target coverage (%)	target coverage (%)	target coverage (%)	target coverage (%)	target coverage (%)
B. subtilis 168*	0	0	0	0	0	0	0	0
B. subtilis isolate 3557 (RASFF2014) chromosome*	44	100	100	80	100	0	100	0
B. subtilis isolate 3557 (RASFF2014) plasmid*	0	100	100	0	0	100	0	100
GMM14 Illumina	44	100	100	80	100	100	100	100
GMM14 MinION	44	100	100	80	99	100	100	92
GMM14 flongle	0	68	68	78	52	75	100	57
GMM16 Illumina	44	100	100	80	100	100	100	0
GMM16 MinION	44	51	51	83	60	83	58	0
GMMneg Illumina	0	0	0	0	0	0	0	0
GMMneg MinION	0	0	0	0	0	0	0	0

Color scale	0	25	50	75	100

A: DNA concentration (measured with Qubit); DNA Integrity Number of the DNA extracts (determined with Tapestation); qPCR detection of two junction sites specific to a GMM *B. subtilis* from RASFF 2014.1249 (VitB2_UGM and 558), a specific site in the plasmid (693) and three AMR genes (nd: not detected after 40 cycles); PCR of the full *tet-L* gene (located on the pGMrib plasmid, nt: not tested), B: Shuttle vector and AMR genes detection (¨, description based on list of common AMR genes detected in GMM from Fraiture, Deckers et al. (2020a)) in WGS data of the isolate of wild type *B. subtilis* strain 168 and GMM strain 3557 linked to RASFF2014 (*, based on sequences from Berbers et al. (2020)) and in the assemblies from metagenomics sequencing using Illumina and MinION technologies, and reads from metagenomics Flongle sequencing of sample GMM14.

for DNA extraction using the Nucleospin Food kit (Macherey-Nagel, Düren, Germany). The protocol was followed according to the manufacturer's instructions. qPCRs and PCR were performed on the DNA extracts as well as on DNA extracts from isolates of *B. subtilis* strains 3557 (GM) and 168 (wild-type), obtained during a previous study (Berbers et al., 2020) as described in Supplementary Materials 2.

Quality and quantity of all DNA extracts were evaluated using the Nanodrop 2000 (Thermo Fisher Scientific, Waltham, MA, USA), Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and 4200 TapeStation (Agilent, Santa Clara, CA, USA). The latter results in the determination of a DIN (DNA Integrity Number) value representing the genomic DNA integrity based on fragment length (value between 1 and 10, with 10 reflecting the highest integrity).

3.2. Illumina (MiSEQ) shotgun metagenomics sequencing

The three DNA extracts were further processed using the Nextera XT library preparation kit (Illumina, San Diego, CA, USA) before sequencing on the Illumina MiSeq, generating paired-end 250-bp reads with the reagent kit v3 according to the manufacturer's instructions. The samples were sequenced in one run of 4 libraries (including another sample not belonging to this study), generating 2,895,502, 2,314,885 and 957 reads for GMM14, GMM16 and GMMneg, respectively.

3.3. Oxford nanopore technologies (MinION) shotgun metagenomics sequencing

The DNA libraries were prepared with the Genomic DNA by Ligation protocol (SQK-LSK109; Oxford Nanopore Technologies, Oxford, UK). The preparation was performed according to the recommendations for MinION sequencing, with one sample on one MinION flow cell. When the DNA concentration of the sample was too low to obtain 1 µg in 48 µl as recommended (samples GMM16 and GMMneg), 48 µl of the available DNA were used as input DNA. The prepared library was then loaded on a primed flow cell (R9.4.1) and a 48-hours sequencing run was started. The resulting fast5 files were basecalled using Guppy on fast mode (version 3.2.1, Oxford Nanopore Technologies), generating 971,569 reads with a median length of 302 bp for GMM14, 1,247,825 reads with a median length of 215 bp for GMM16 and 2,187 reads with a median length of 214 bp for GMMneg. The statistics were obtained using NanoPlot version 1.28.0 (De Coster et al., 2018) (full statistics in Supplementary Materials 1).

3.4. Illumina data analysis

The reads were trimmed using Trimmomatic version 0.38.0 with a sliding window approach requiring an average Phred score of 20 evaluated on a window of 4 bases (Bolger et al., 2014). Taxonomic classification of the reads was conducted using Kraken2 version 2.0.7 (Wood et al., 2019) as described in Buytaers et al. (Buytaers et al., 2020). The Illumina reads were assembled using SPAdes version 3.13.0 (Bankevich et al., 2012) with –meta mode. The presence of AMR genes in the contigs was detected using Blastn 2.7.1 on the ResFinder database (Kleinheinz et al., 2014) and on a set of the most common AMR genes detected in bacterial GMMs described in Fraiture, Deckers et al. (2020a). For shuttle vectors detection, the database UniVec was first tested but this did not give satisfying results (no detection of the expected vectors, e.g. pUB110, described to be present in the isolate by Berbers et al. (Berbers et al., 2020), maybe due to incorrect metadata). Therefore, a Blast was only performed on the reference sequence of the pUB110 shuttle vector (GenBank: M19465.1) as it is well documented and described to be used in several GMMs including RASFF2014.1249 (Berbers et al., 2020; Fraiture, Papazova et al., 2020). This pUB110 vector is linked to the presence of the *aadD* AMR gene (Fraiture, Papazova et al., 2020). The presence of the riboflavin producing genes (*rib* operon of *B. subtilis* and *Bacillus amyloliquefaciens* as annotated with Prokka from the reference sequence from Berbers et al. (Berbers et al., 2020)) was also investigated with Blast. All Blast analyses were conducted with default parameters. Contigs that had a hit for AMR, pUB110 or *rib* genes were extracted and annotated using Prokka version 1.11 (Seemann, 2014) and then blasted online to the nucleotide database (https://blast.ncbi.nlm.nih.gov/Blast.cgi) to determine the species detected on the contigs. This was visualized using SeqBuilder Pro 15 v15.3.0 (DNASTAR Lasergene). Finally, the reads were mapped to the reference genome of the isolate from RASFF 2014.1249 (accession chromosome: NZ_CP045672.1 and plasmid: NZ_CP045673.1), using BWA MEM version 0.7.17 with default settings (Li & Durbin, 2010). The breadth of coverage to the full genome, chromosome and plasmid was calculated using SAMtools version 1.9 (Li et al., 2009) and awk (using the command: samtools depth -a alignment.sorted.bam | awk '{c++; if ($3>0) total+=1}END{print (total/c)*100}') and the mapping was evaluated using QualiMap version 2.2.1 (Okonechnikov et al., 2015). The mapping was also visualized on IGV (Robinson et al., 2011) and the plasmid was manually annotated for the sites of the *tet-L* gene, the qPCR VitB2_UGM and the qPCR 693 following annotations from Berbers et al. (2020).

3.5. MinION data analysis

The fastq was converted to fasta format using Seqtk version 1.3 (https://github.com/lh3/seqtk/blob/master/README.md). The fasta file was used for taxonomic classification and species identification using Kraken2 and the same database and parameters as for Illumina reads. A more detailed species identification was conducted with Megablast using Blastn 2.7.1 (Camacho et al., 2009) to the regions V3-V4 of 16S rRNA sequences from Deckers, Vanneste, Winand, and Keersmaecker et al. (2020) combined to the 16S rRNA database available on NCBI, as well as to the NCBI nucleotide database (Sayers et al., 2019) with max_target-seqs set to 1. The fastq was used for assembly using Canu version 1.8 (Koren et al., 2017) modifying the parameters stopOnLowCoverage to 1, minReadLength to 200 and minOverlapLength to 100 in order to fit the relatively short reads obtained in the sequencing runs. Gene detection was conducted directly on the contigs using Blastn with the same parameters and on the same databases as for Illumina assembled reads. The contigs that had a hit were annotated using Prokka and blasted online to the nucleotide database to determine the species from which the sequences originated from. The mapping on the reference genome linked to RASFF 2014.1249 and calculation of the breadth of coverage were performed in the same way as for the Illumina data analysis (using -x ont2d command to use BWA MEM on ONT reads).

3.6. Flongle sequencing and data analysis

The DNA extract from sample GMM14 was also sequenced on a Flongle flow cell, after library preparation with the Genomic DNA by Ligation protocol (SQK-LSK109; Oxford Nanopore Technologies, Oxford, UK) following the manufacturer's recommendations. The library was loaded on a Flongle presenting 60 available pores at the start of a run of 24 h. The resulting fast5 files were basecalled using Guppy on fast mode (version 3.2.1, Oxford Nanopore Technologies) generating 60,093 reads with a median length of 306 bp.

The Flongle data analysis was similar to the MinION data analysis except for the assembly that could not be achieved due to low coverage. Therefore, the gene detection was conducted directly on the reads. As no contig was obtained, the search for unnatural associations was done as follows. The reads that had a hit for the presence of AMR genes were compared to the result of the Blast to the nucleotide database of the same read in order to obtain the species from which the sequence originated.

3.7. Data availability

All sequencing data is publicly available at NCBI SRA under project PRJNA686880.

4. Results

4.1. Development of a shotgun metagenomics-based approach for the characterization of a GMM

4.1.1. Sample selection and preparation

For the development of the shotgun metagenomics-based method for the characterization of GMM in microbial fermentation products, we selected a sample that was known to contain a GMM, and for which this GMM had been previously characterized after isolation. We selected the sample linked to the RASFF 2014.1249 (GMM14), containing a *B. subtilis* overproducing vitamin B2 (riboflavin) and therefore positive for the GM-events VitB2_UGM (Barbau-Piednoir et al., 2015) and 558 (Paracchini et al., 2017). This GMM had been isolated and fully sequenced before (Berbers et al., 2020). As a negative control sample (GMMneg), we included a vitamin B2 sample from which DNA could be extracted but without detection of the GM-events specific to vitamin B2 overproduction (i.e. VitB2_UGM and 558). Therefore this sample was considered as 'probably not containing vitamin B2 overproducing GMM'.

To verify the samples, we performed a qPCR detection of event-specific markers as well as the presence of 3 AMR genes (*cat, aadD tet*) on the DNA extracts (Table 1.A). The markers specific to the GMM strain from RASFF 2014.1249 were detected in sample GMM14 (as expected), confirming the result of the initial screening. The 3 AMR genes were detected in GMM14, as expected based on the full genome sequence of the GMM *B. subtilis* that was previously isolated from this sample (Berbers et al., 2020). The same qPCR markers were detected in the DNA extracted from the isolate *B. subtilis* strain 3557 previously described in the context of the RASFF 2014.1249. These markers were not detected in DNA from the wild-type *B. subtilis* strain 168. Similarly, none of the genetic markers were detected in GMMneg.

The extracted DNA was then used for library preparation for shotgun metagenomics sequencing. We investigated both short read as well as long read sequencing technologies, especially as we assumed that long read sequencing would facilitate downstream data analysis to identify unnatural associations as well as full-length AMR genes. However, long read sequencing requires highly concentrated and high molecular weight DNA (Nanopore Protocol, 2019). The DNA concentration in sample GMMneg was too low according to the standards of MinION sequencing (i.e. need for 1 µg starting material in 48 µl). Moreover, the integrity of the extracted DNA was very low in all samples, as determined based on the obtained DIN value (less than 2, Table 1.A). The samples were nevertheless included in the downstream sequencing analysis.

4.1.2. Gene detection in assemblies from shotgun metagenomics

After sequencing, we looked for the presence of pUB110 and AMR genes in the contigs (Table 1.B, Supplementary Materials 2, Supplementary Materials 3), both from the short read as well as from the long read MinION sequencing output.

A part of the pUB110 shuttle vector (corresponding to the same portion as detected in the isolate from RASFF 2014.1249) as well as the genes linked to resistance to ampicillin (*bla*), bleomycin (*ble*, not present in the ResFinder database), chloramphenicol (*cat*), erythromycin (*ermB*), kanamycin & neomycin (*aadD*) and tetracyclin (*tet-L*) were detected in GMM14 after both short and long read metagenomics sequencing (Table 1.B). The shuttle vector and all these AMR genes have been previously described to be present on the chromosome and pGMrib plasmid of the isolate mentioned in RASFF 2014.1249 (Table 1.B, (Berbers et al., 2020)). Moreover, pUB110 is harboring the AMR genes *aadD* and *ble*. The assemblies of Illumina and MinION reads both allowed a detection of almost all genes with a coverage of more than 90%, except *ble (*with a coverage of 80%). The *ble* gene was already covered only at 80% in the chromosome of the isolate previously sequenced (Berbers et al., 2020), meaning that the recovery of what is expected to be present is 100%. The AMR genes were detected to cover the full-length of the reference genes on the contigs. However, the full-length genes were not detected on single reads, as the reads sequenced with MinION sequencing were shorter than the average length of an AMR gene. Nevertheless, the high coverage of the gene from the contigs is a strong indication that the full-length AMR gene was present in the sample.

As the GMM14 sample is a riboflavin feed additive, the presence of genes linked to riboflavin production (vitamin B2) was also investigated (results presented in Supplementary Materials 2). The *rib* operon from *B. subtilis* and from *B. amyloliquefaciens* origin were both detected in the metagenomics sequencing of GMM14 with a coverage higher than 80% for the two sequencing instruments.

As expected, none of the most common AMR genes reported in GMMs were detected in GMMneg (Table 1, Supplementary Materials 2), confirming the negative results obtained with qPCR (Table 1). The riboflavin producing genes were also not detected in this sample after Illumina or MinION sequencing (Supplementary Materials 2).

4.1.3. Species identification via shotgun metagenomics

Next, we used the sequenced reads to identify the species present in the samples, to see whether these correspond to known microbial species used for known GMM based on patent information (Fraiture, Deckers et al., 2020b). The sequenced reads from the two sequencing devices were classified per genus using Kraken (Fig. 1.A). *Bacillus* (green, Fig. 1.A.) was the main organism for the two sequencing methods in sample GMM14, although 35 to 40% of the reads could not be classified (light grey, Fig. 1.A.). More taxa were detected with Illumina sequencing but the small proportions might represent false positive classifications of some short reads. Two alternative methods aiming at obtaining more accurate information on the present species, were tested on the longer reads sequenced with the MinION: a Blast to a 16S rRNA database as shown in Fig. 1.B and a Blast to the NCBI nucleotide database as shown in Fig. 1.C. 786 of the MinION reads of GMM14 had 16S rRNA hits. With the two methods, *B. subtilis* could be detected as the main species in the sample (green, Fig. 1.B and C). However, other *Bacillus* species were sometimes detected with the 16S rRNA method (light green, Fig. 1.B). This could be expected as the 16S rRNA genes are very similar for these species and the method has been reported to be unable to differentiate efficiently between *B. subtilis* and *B. licheniformis* (Deckers, Vanneste, Winand, & Keersmaecker et al., 2020). The classification using the NCBI nucleotide database covers the full genome of each species, and thereby allows for more genomic markers to be used to attain species resolution. 31% of the reads could be classified to genus *Bacillus* sp. and 55% of all classified reads were detected as *B. subtilis* without ambiguity. The small proportion of *Escherichia* (yellow, Fig. 1.A, 1.B, 1.C) detected in both sequencing runs is partly explained by a misclassification (i.e. 12% of the reads classified as *E. coli* are mapping to the *B. subtilis* GM reference defined by Berbers et al. (2020)) but could also indicate the presence of DNA of this species in the sample. In conclusion, *B. subtilis* was detected in high proportions, corresponding to the GMM species that was previously isolated from the GMM14 sample.

In the GMMneg sample, 62% of the reads were unclassified after Kraken analysis of the Illumina sequencing while *Bosea, Sphingomonas* and *Rhizobium* were detected as the main genera (Fig. 1.A). The latter two genera are known as common contaminants of Illumina sequencing (Winand et al., 2020). For MinION sequencing, more than 93% of the reads could not be classified, while *Ochrobactrum* and

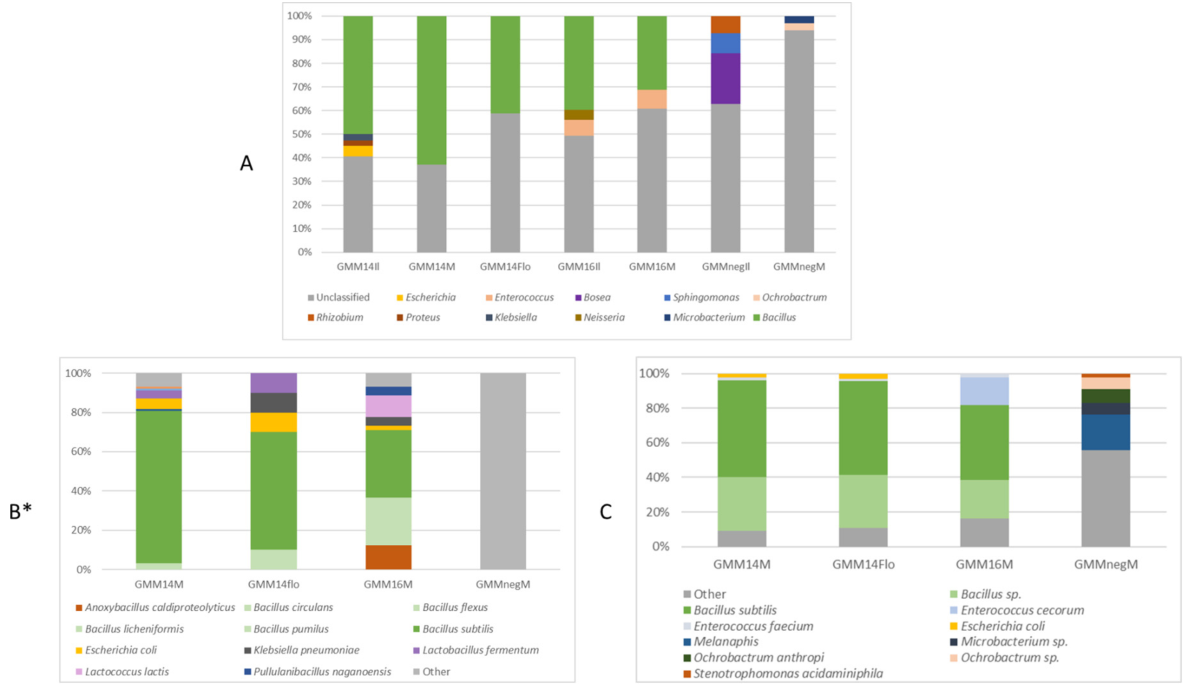

Fig. 1. Species identification in the different samples. A: Kraken taxonomic classification results for Illumina ('Il'), MinION ('M') and Flongle ('Flo') reads. Taxa representing < 2% of the reads are counted in unclassified. B: Blast to 16S rRNA database results for MinION ('M') and Flongle ('Flo') reads. "Other" (grey): species representing < 2% of the reads with hits (or for GMMneg: no hit obtained) *: Results presented to species level, as output from workflow described in Materials and methods section, however it was reported that 16S rRNA analysis is limited to genus level (Winand et al., 2020) C: Blast to nucleotide database results for MinION ('M') and Flongle ('Flo') reads. "Other" (grey): Species representing less than 2% of the reads with hits (e.g. *Streptococcus pyogenes*).

Microbacterium each represented 3% of the reads. *Bosea, Rhizobium* and *Ochrobactrum* are all part of the order Rhizobiales. The presence of these genera was not confirmed with the Blast to the 16S rRNA or nucleotide database (Fig. 1.B and 1.C). Indeed, no 16S rRNA hit was obtained in the MinION sequencing of the sample, with the database used, while *Melanaphis*, *Microbacterium* sp., *Ochrobactrum anthropic* or sp. and *Stenotrophomonas acidaminiphila* were detected with the NCBI nucleotide database. The very low concentration of DNA in the sample (Table 1) led to a very low quantity of reads after sequencing, which could explain the inconsistency in genus identification for both sequencing methods. None of these genera are known as previously reported GMM (Fraiture, Deckers et al., 2020b), and most probably represent a contamination. It should be noted that most likely in routine analysis, based on the negative results of the first line screening, no additional analysis would be performed. In our study, the sample was only used as a negative control for the metagenomics approach.

4.1.4. Detection of unnatural associations in the assembled metagenomic reads

The contigs containing AMR genes obtained for sample GMM14 were further investigated to determine if some unnatural associations (i.e. presence of parts of sequences belonging to different species or vector(s) in the same genome) were present. Given the nature of the sample, the same was done for contigs containing genes linked to riboflavin production (in this case the *rib* operon). As *B. subtilis* was detected as the main species, contigs harboring *B. subtilis* genome and parts of genomes from other species were investigated as probable unnatural association linked to the GMM in the sample (Fig. 2). This was done for the Illumina and MinION assemblies. Notably, several similar hits for genome, plasmid or vector identity could be obtained with the same confidence for the contigs investigated. However, only one hit was shown per region in the figure, in order to illustrate the unnatural association without aiming at identifying the exact origin of this segment of genome. An insertion of the chloramphenicol resistance gene (*cat*) in the genome of *B. subtilis* was detected with both sequencing technologies, interrupting the sequence of the *recA* gene (Fig. 2.A and 2.D). This corresponds to the 558 qPCR assay specific GM-event previously described for the RASFF2014 strain (Berbers et al., 2020; Paracchini et al., 2017). Another contig in the Illumina-based assembly contained 2 genes of the *rib* operon in the *B. subtilis* genome adjacent to a plasmid sequence originating from *Streptococcus pyogenes* (Fig. 2.B). Moreover, a part of the *B. subtilis* genome carrying *ribA* from the *rib* operon was linked to a part of an expression vector and the pUB110 plasmid sequence from *Staphylococcus aureus*, harboring 2 AMR genes (*ble* and *aadD*) in a contig from MinION sequencing (Fig. 2.C). The same pattern was observed in the chromosome of the GMM isolate described by Berbers et al. (Berbers et al., 2020). These sequences prove an unnatural association in the genome of *B. subtilis*, detected as the main species in the sample, and hence the presence of a GMM in sample GMM14.

As no AMR or *rib* genes were detected in GMMneg, this analysis was not conducted for this sample. This sample is considered not to contain a GMM strain.

4.1.5. Validation of method: Mapping of metagenomics reads to a previously characterized GMM reference genome

As a validation step to demonstrate that our metagenomics analysis detected the GMM previously characterized as an isolate from the same sample, we mapped the sequenced reads to the reference genome of this rapid alert (GCA_009914705.1 (Berbers et al., 2020)). We could show that the reference genome is fully covered with our metagenomics reads. The breadth of coverage was calculated as 100% for the chromosome and pGMrib plasmid with the two sequencing technologies, with a mean coverage of 57 on the chromosome sequence and 317 on the plasmid sequence for the MinION sequencing, and a mean coverage of 119 on the chromosome sequence and 492 on the plasmid sequence for the Illumina sequencing (Supplementary Materials 4). This additional validation step proves that the GMM detected with the metagenomics

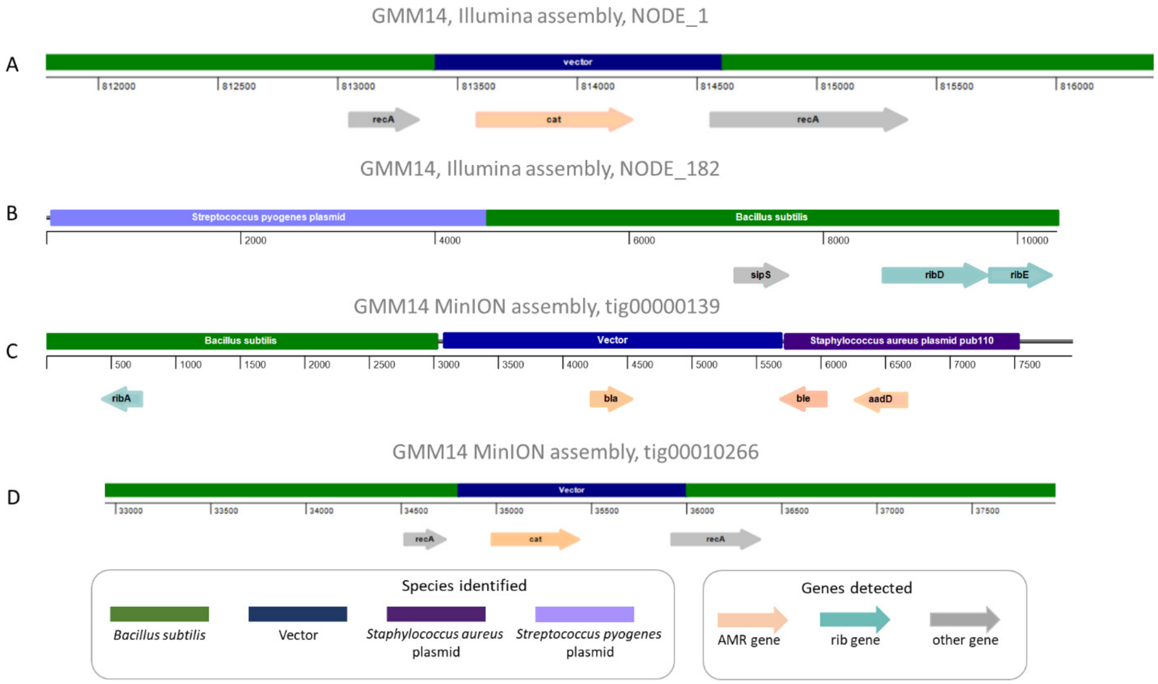

Fig. 2. Detection of species and genes on contigs of GMM14 sequenced with different technologies representing unnatural associations in the genome. A-B: contigs from Illumina assembly. C-D: Contigs from MinION assembly.

approach is indeed similar in genome structure to this previously sequenced isolate. This was expected since the sample analyzed originated from the same riboflavin product. Moreover, the mapping to the pGMrib plasmid was visualized (Fig. 3.A and B) with tags annotating the positions of the *tet-L* resistance gene (Berbers et al., 2020) and the positions for the qPCR VitB2_UGM (Barbau-Piednoir et al., 2015) as well as the qPCR 693 (Paracchini et al., 2017) assay. These were all covered as expected from the qPCR results (Table 1).

4.1.6. Evaluation of Flongle sequencing

Based on the results described above, a workflow using either short or long read sequencing seems to enable the characterization of a GMM in a microbial fermentation product. However, in the short read sequencing run, more than one sample was included, to make it cost-effective. This might not be desirable in a routine set-up, where samples are often arriving, and hence need to be analyzed, on a one-by-one basis. The long read sequencing included one sample per flow cell, thereby rendering the cost per analyzed sample more expensive than the short read sequencing. Therefore, a Flongle sequencing was carried out on the GMM14 sample, as a less expensive, more flexible (one sample, with less input material required than MinION) and fast (24 h) sequencing alternative. The same analysis steps were performed as for the Illumina and MinION sequencing. This was done to evaluate whether the same information could be obtained using a long read sequencing device with a lower output (6% of the amount of reads for the same sample compared to the MinION sequencing, Supplementary Materials 1). All expected genes could be detected in the Flongle reads, but with a coverage starting from 52%, which would be filtered out of most classical analyses, and generally lower sequence similarity (80–90%) to the reference sequence compared to the results obtained with the MinION contigs (>90%, Table 1, Supplementary Materials 2). The lower coverage is explained by the use of reads instead of an assembly. The use of short sequences could also explain why the shuttle vector pUB110 could not be detected with a coverage higher than 0.2% while 44% was present in the reference genome from the isolate. Nevertheless, the AMR genes present on this plasmid (*ble* and *aadD*) were correctly covered. The genus *Bacillus* and the species *B. subtilis* were detected as the main microorganism in the sample, in the same proportions as for the Illumina and MinION sequencing (Fig. 1.A, B and C; 16S rRNA classification based on 31 reads with hits). As no assembly was obtained, the unnatural associations could not be detected as such. Nevertheless, an analysis with Blast to the nucleotide database of the reads that had a hit to an AMR gene confirmed that these AMR genes were detected in species or synthetic constructs other than *B. subtilis* (Supplementary Materials 5), the main species detected in the sample. This raises the suspicion about a possible alteration of the genome of *B. subtilis* to add AMR genes naturally present in other species. A mapping was performed to the reference genome of the isolate from RASFF 2014.1249 (Berbers et al., 2020) to confirm that it has a similar genome structure, as the same AMR and *rib* genes were detected. The reference genome was not fully mapped (92.6% breadth of coverage to the chromosome and 100% to the plasmid). A mean coverage of 3 was determined to the chromosome sequence and 15 to the plasmid sequence (Supplementary Materials 4). The low coverage, linked to the lower output of the Flongle, might explain the loss in breadth of coverage compared to the Illumina and MinION sequencing of the same sample. However, the obtained results indicated that the GMM detected using shotgun metagenomics Flongle sequencing of the GMM14 sample is similar in genome content to the previously sequenced isolate from RASFF 2014.1249. The mapping to the plasmid reference sequence was visualized as well (Fig. 3.C), indicating that the *tet-L* gene as well as the qPCR VitB2_UGM and qPCR 693 sites were covered, which corresponds the qPCR results (Table 1).

4.2. Applicability of the method: sample positive for GMM B. subtilis qPCR markers but without isolated bacterium

A vitamin B2 sample received for routine analysis in 2016 (GMM16), which tested positive for the GM-associated junctions VitB2_UGM and 558 by qPCR, but for which no living bacterium could be isolated, was used to demonstrate the applicability of our developed workflow. The re-extracted DNA gave, as expected, a positive qPCR signal for the vitamin B2 specific GM-events (VitB2_UGM and 558) and also for the 3 AMR genes (*cat*, *aadD* and *tet*) (Table 1). *tet-L*, which

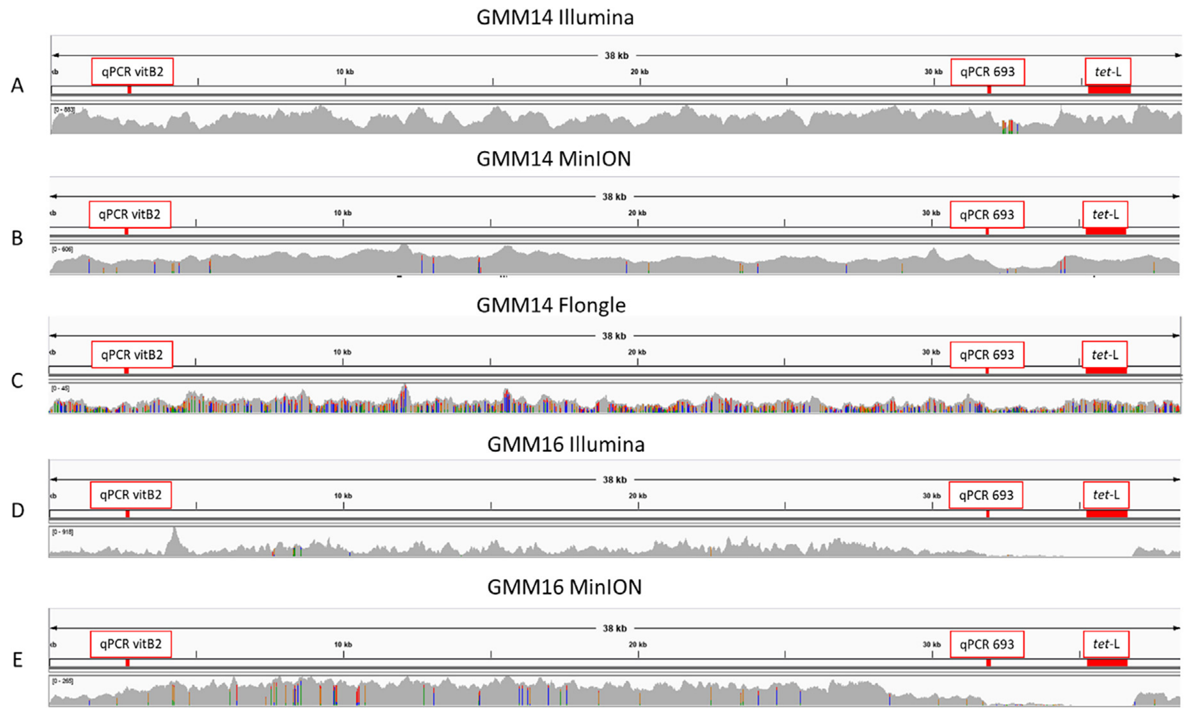

Fig. 3. Coverage of the pGMrib plasmid from *B. subtilis* strain 3557 (RASFF 2014) with annotation of the qPCR 693, qPCR VitB2 site and the *tet*-L gene. Colored bars: deviations from the reference. A: GMM14 Illumina sequencing. B: GMM14 MinION sequencing. C: GMM14 Flongle sequencing. D: GMM16 Illumina sequencing. E: GMM16 MinION sequencing.

is known to be present on the pGMrib plasmid of the previously described GM *B. subtilis* (Berbers et al., 2020)*,* was detected with a higher Cq of 32.7 compared to the 2 other AMR genes (*cat* and *aadD*), present on the chromosome of the same GM strain. This Cq was also higher compared to another qPCR marker that should be present on the pGMrib plasmid of the reference, the VitB2_UGM (Cq of 23.69, Table 1). The qPCR 693 assay (Paracchini et al., 2017), targeting the junction of pGMBsub03 to pUC19 located on the pGMrib plasmid in the GM *B. subtilis* isolate (Berbers et al., 2020; Paracchini et al., 2017), was not detected in this sample after 40 cycles of the assay. As this was a sign of difference with the previously described isolate from RASFF 2014.1249, a PCR of the *tet* gene (Fraiture, Deckers et al., 2020c) was then performed on all samples to verify the presence of the full *tet* gene. This PCR was negative for GMM16 (Table 1). The DIN value of the obtained DNA extract was very low, indicating the presence of degraded DNA. This and the low concentration of the DNA (Table 1) were not optimal according to Oxford Nanopore's guidelines for MinION sequencing. Nevertheless, the DNA was used for short (Illumina) and long read (MinION) sequencing. MinION was selected over Flongle sequencing to account for the higher Cq value obtained for the detection of the *tet* marker, and hence anticipating a need for higher coverage/output.

4.2.1. Gene detection in assemblies from shotgun metagenomics sequencing

After gene detection in the assemblies from Illumina and MinION sequencing, the shuttle vector pUB110 and the resistance genes *bla, ble, cat, ermB* and *aadD* could be detected (Table 1.B). The shuttle vector was covered at 44% as observed for sample GMM14 and for the isolate from that sample. Most AMR genes were detected in full-length (100% target coverage) in the Illumina assembly except for *ble*, but the same percentage was covered as previously observed for sample GMM14 and the associated isolate. The assembly of the MinION reads allowed the detection of the same genes albeit with a lower coverage (the lowest being 51%) and a lower identity (see Supplementary Materials 2). Again, the genes were fully covered in the contigs, but the full-length genes could not be detected directly in the reads due to their limited length.

The genes detected were the same as the genes present on the previously characterized GMM14 isolate (and metagenomics GMM14 sample), except for the absence of the tetracycline resistance gene (*tet-L*) in the contigs, for which a higher Cq was obtained with qPCR.

Genes linked to riboflavin production were detected in assemblies from both types of sequencing reads for this sample (Supplementary Materials 2), i.e. genes from the *rib* operon from *B. subtilis* and *B. amyloliquefaciens,* confirming that most probably the DNA sequenced belonged to the organism producing the substrate. The coverage and identity of the detected genes was higher for the Illumina contigs than for the MinION sequencing, for which some genes were detected with a coverage lower than 50%.

4.2.2. Species identification via shotgun metagenomics

After taxonomic classification with Kraken (Fig. 1.A), more than 50% of the reads from the GMM16 sample could not be classified for both the Illumina and ONT data. The majority of the classified reads was attributed to the *Bacillus* genus after Illumina or MinION sequencing. This genus is listed as one of the most commonly referenced GMMs (Fraiture, Deckers et al., 2020b), especially for the production of riboflavin. *Enterococcus* and *Neisseria* were detected in smaller proportions with the two sequencing technologies. Identification to a higher resolution was attempted with a Blast to the 16S rRNA database (Fig. 1.B, based on 596 reads with hits) and nucleotide database (Fig. 1.C). *B. subtilis* was detected at 34 and 43% with those methods, while other *Bacillus* species (for the 16S rRNA) or *Bacillus* sp. (for nucleotide) covered a remaining 20%. The other detected species were not consistent between the methods, and the presence of *Neisseria* was not confirmed.

4.2.3. Detection of unnatural associations in the assembled metagenomic reads

We looked for unnatural associations in contigs containing AMR genes and genes linked to riboflavin production (Fig. 4). A *cat* insertion in the sequence of the *recA* gene of *B. subtilis* was detected in the Illu-

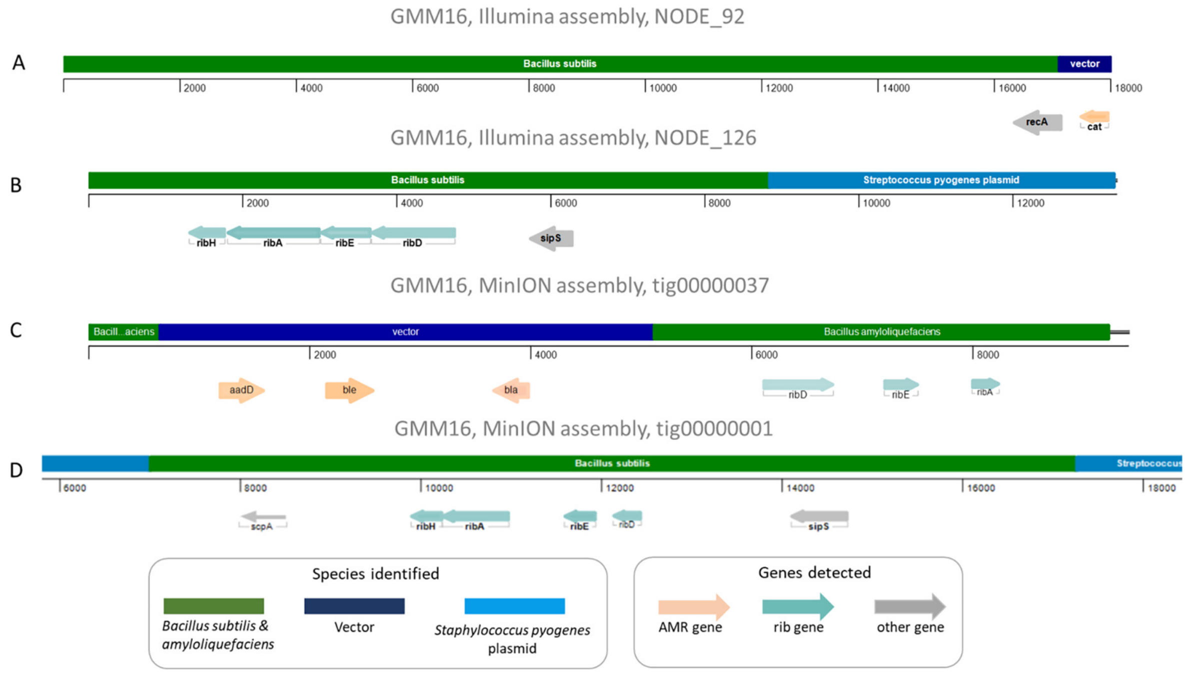

Fig. 4. Detection of species and genes on contigs and reads of GMM16 sequenced with different technologies to represent the presence of unnatural associations in the genome. A-B: Contigs from Illumina assembly. C-D: Contigs from MinION assembly.

mina assembly of the sample (Fig. 4.A). The same insertion was described in the GMM linked to RASFF 2014.1249 (558 junction) (Berbers et al., 2020). Other unnatural associations of the genome of *B. subtilis*, harboring genes from the *rib* operon with plasmids from other species, were also detected in the Illumina and MinION assemblies (Fig. 4.B and 4.D). Moreover, an association of the *B. subtilis* genome, a vector containing two AMR genes (*aadD* and *ble*), and the *B. amyloliquefaciens* genome harboring the *rib* operon, was also detected in the MinION assembly (Fig. 4.C). The presence of these unnatural associations in the genome of *B. subtilis*, detected as the main species in the sample, proved the presence of a GMM in sample GMM16.

4.2.4. Mapping to a previously characterized GMM reference genome

Following the high similarity of the information detected in GMM16 with the previously characterized isolate from RASFF 2014.1249, except for the absence of the *tet*-L gene previously described to be present on the pGMrib plasmid of the GMM (Berbers et al., 2020), we conducted a mapping of the GMM16 metagenomics reads to the reference genome obtained for the isolate linked to RASFF 2014.1249 (Berbers et al., 2020). This resulted in a full mapping of the chromosome (100% breadth of coverage for the MinION sequencing and 99.9% for Illumina, Supplementary Materials 4) with a mean coverage of 24 after MinION sequencing and 39 after Illumina sequencing, but a partial mapping of the plasmid (99.9% breadth of coverage for the MinION sequencing and 97.5% for Illumina) with a mean coverage of 138 after MinION sequencing and 194 after Illumina sequencing (Supplementary Materials 4). A visualization of the mapping to the plasmid sequence showed the absence of reads mapping to the region of the *tet*-L gene (position 35241-36617 (Berbers et al., 2020)) in the metagenomics reads (Fig. 3.D and 3.E), in contrast to the metagenomics reads obtained for sample GMM14 (Fig. 3.A, 3.B and 3.C). This corroborates the absence of amplification of the full *tet* gene with PCR (Table 1). When zooming in (not represented in the figure), the region of qPCR 693 is also missing, confirming the result obtained with qPCR as well, while the rest of the pGMsub03, the region in which the qPCR 693 and the *tet-L* gene are described to be present in the reference strain from RASFF 2014.1249 (Berbers et al., 2020), is covered with very few reads, and not covered anymore if filtering for reads that map uniquely. All other regions of the pGMrib plasmid, however, were covered with reads. Therefore, we can conclude that the GMM present in the GMM16 sample, for which no isolate could be obtained, is similar in genomic content at least for the chromosome to the isolate from RASFF 2014.1249. The plasmid might be different or have been modified but the region of the qPCR Vitb2_UGM as well as the erythromycin resistance gene were still detected with qPCR and/or after sequencing.

5. Discussion

GMMs are commonly used to produce microbial fermentation products. According to European regulation, the viable GMM or its DNA, often containing AMR genes, cannot be present in the final product of commercialized genetically modified food and feed (Deckers, Deforce et al., 2020). It is important for enforcement laboratories within Europe to have access to methods allowing the detection and characterization of such GMMs or their DNA. Construct/event-specific qPCRs have been previously developed on a case-by-case basis after WGS of an isolate (Barbau-Piednoir et al., 2015; Fraiture et al., 2020; Paracchini et al., 2017). These methods have been used as a second-line analysis after detection of AMR genes and a shuttle vector for which first line qPCR assays have been developed based on publicly available patent information (Fraiture, Deckers et al., 2020a). The development of such construct/event-specific methods, however, requires prior isolation of the contaminant for its characterization, and each test is only specific to one GMM. An alternative targeted approach, based on DNA walking, has been proposed and does not rely on obtaining an isolate (Fraiture et al., 2021). However, it still requires prior knowledge to design primers and can be very laborious as the unnatural association might only be obtained after several consecutive reactions that each have to be carefully designed. The DNA walking approach has led to the design of an additional event-specific marker (Fraiture, Papazova et al., 2020). Using WGS or DNA walking methods, until now only 3 GMMs have been characterized and can be identified using qPCR. In this study, we propose an alternative open approach based on shotgun metagenomics to potentially allow untargeted identification of GMMs. This does not require isolation and allows detecting any AMR gene present in the DNA, identify the species present in the sample and expose the presence of unnatural associations of sequences in the genome. Our workflow was established with the aim to be usable in the future by the European enforcement laboratories as an alternative or addition to their current investigation tools.

Our results deliver a proof-of-concept for a shotgun metagenomics approach as a viable alternative to detect and characterize a GMM present in microbial fermentation products without the need for isolation or enrichment. In our workflow, the prediction of the presence of a GMM was based on the simultaneous detection of AMR genes or vectors in species previously described as common GMM producers (Fraiture, Deckers et al., 2020b), and the encounter of unnatural associations in the genome (Fig. 5).

Altogether, our method allowed to achieve the same information as obtained with the currently used standard methods (detection of AMR genes or vectors with qPCR and detection of unnatural associations with WGS or with event-specific qPCRs). Moreover, it can potentially replace additional testing such as the detection of the genus/species with 16S rRNA-based methods. However, our method is able to perform all these analyses at once, thereby saving time. It even extends the characterization of the GMM, such as detecting the presence of AMR genes for which no qPCR methods have yet been developed (in this case *bla, ble, erm*), and identify to species level, even when multiple species are present, which is not always possible with the 16S rRNA method (Yang et al., 2016). This method also allows to describe previously unknown unnatural associations that could lead to the development of new event-specific qPCR methods.

We compared two sequencing technologies producing short reads or long reads. The results obtained with Illumina and MinION sequencing were equally satisfying, leading to the detection of all genes of interest and unnatural associations, with equal breadth of coverage after mapping to a reference genome. A shuttle vector and several unnatural AMR genes could be detected in the assemblies. This is a strong indication that the full-length gene is present in the samples. The identification of the reads to the species level was only obtained with the MinION sequencing with a Blast of the reads to the NCBI nucleotide database. This could be expected since the use of 16 s rRNA genes was previously described as insufficient to obtain species resolution (Winand et al., 2020). Moreover, the error rate of MinION sequencing is higher and might lead to a misclassification on the short and highly similar 16S rRNA region. This analysis could not be conducted on the short Illumina reads, however, illustrating the advantage of long read sequencing for species identification. The classification of contigs was not feasible for this application due to difficult or even dangerous interpretation of the results as by nature of the sample, these contigs represent an association of several species. Flongle sequencing yielded 6% of the amount of reads obtained from MinION sequencing, with a lower cost. These reads were of similar median length as the reads obtained with the classical MinION flow cell, but with a generally lower read quality, and allowed species identification and detection of the genes of interest. However, as no assembly could be performed, genes were detected with a lower coverage that might not pass classical thresholds of analysis. Although no thresholds for metagenomics analyses have been established yet, EFSA recently published a statement on the requirements for WGS analysis of isolated microorganisms intentionally used in the food chain (EFSA, 2021). They advised query sequence hits with at least 70% length of the subject sequence to be reported when submitting a char-

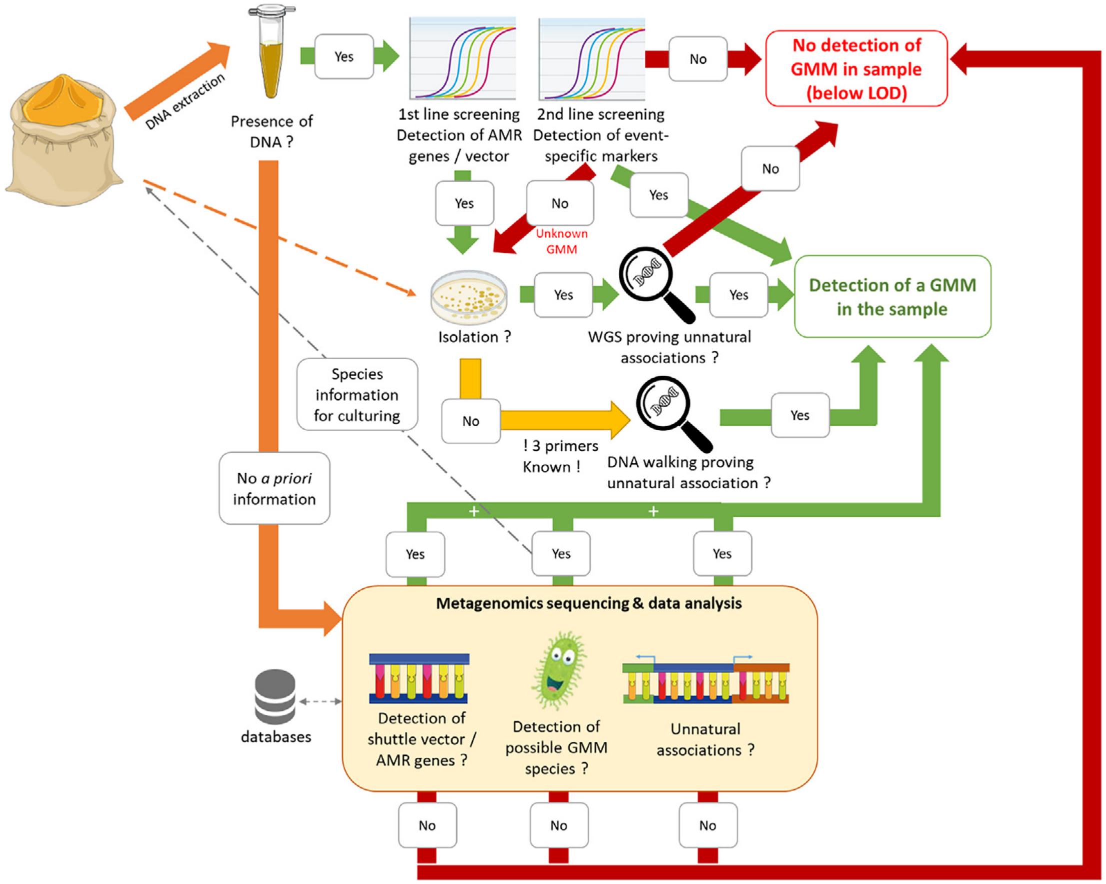
DNA extraction
Presence of DNA ?
Yes
1st line screening
Detection of AMR genes / vector
2nd line screening
Detection of event-specific markers
No
No detection of GMM in sample (below LOD)
Yes
No
Unknown GMM
Yes
No
Isolation ?
Yes
WGS proving unnatural associations ?
Yes
Detection of a GMM in the sample
Species information for culturing
No
! 3 primers Known !
DNA walking proving unnatural association ?
Yes
No *a priori* information
+
+
Yes
Yes
Yes
Metagenomics sequencing & data analysis
databases
Detection of shuttle vector / AMR genes ?
Detection of possible GMM species ?
Unnatural associations ?
No
No
No

Fig. 5. GMM detection decision tree, presenting the conventional workflow currently performed in enforcement laboratories (qPCR screening, DNA walking or WGS on the isolate) and the proposed metagenomics alternative when no isolate can be obtained. If simultaneously detecting AMR gene(s) typically not naturally occurring, possible GMM species and unnatural associations in the genome, depending on the available databases, we can conclude that a GMM was detected in the sample.

acterization dossier. Not being able to assemble the reads also complicated detection of unnatural associations. The breadth of coverage of the mapping was also lower due to some missing information in the lower output. Nevertheless, all information needed to prove that a GMM was present in the sample and to characterize it was obtained. Therefore, this is an interesting rapid and low-cost alternative for enforcement laboratories to get an overview of the content of a sample for which no information can be obtained with the normal qPCR screening, when the DNA concentration and quality are sufficient.

We have studied a previously described sample (GMM14) and then used the developed method to characterize a sample in which some GMM-specific markers were positive (qPCR VitB2_UGM, 558, detection of AMR genes) but for which no isolate could be obtained (GMM16). After thorough analysis of the reads and contigs obtained from this sample, we were able to detect, with both sequencing technologies, more AMR genes than detected with the qPCR (presence of *bla, ble, ermB*). We were also able to identify the main species as *B. subtilis* and detect unnatural associations in the genome, confirming that it was indeed a GMM. These parameters led to a strong suspicion that a similar GMM as the one previously characterized from RASFF 2014.1249 was present in this sample. The *tet* gene described to be present in the pGMrib plasmid of that GMM was however not detected after shotgun metagenomics sequencing of GMM16. A high Cq was obtained in qPCR screening, targeting a part of the *tet-L* gene and we could not demonstrate the presence of the full-length *tet-L* gene in the sample by PCR. After mapping to the reference genome (*B. subtilis* 3557, GenBank GCA_009914705.1, (Berbers et al., 2020)), we could establish that that part of the pGMrib plasmid was missing while the rest of the chromosome and the plasmid were fully covered by the sequenced reads. This suggests that this sample contains a similar but different GMM, with a plasmid that does not harbor the *tet-L* gene.

Our study is rather explorative. It needs to be seen as a proof-of-concept for the use of metagenomics approaches for the detection and identification of GMM. We illustrated this potential using a selection of samples representative for the possible scenarios in routine. In the future, additional samples need to be investigated. Moreover, some challenges still have to be overcome to make our workflow easier to implement in enforcement laboratories. First, short and long read sequencing both independently delivered the required result, demonstrating the presence of a GMM. Nevertheless, long read sequencing has some advantages in terms of costs, flexibility and species identification. However, the long read sequencing was performed with DNA extracts that were of lower quantity and quality, resulting in rather short median read lengths. If high molecular weight DNA could be obtained from the food/feed samples, the long read sequencing method could be used without the possible bias of assembly that can create chimeras. Moreover, the possible unnatural associations as well as full-length AMR genes might be detected on one (a few) single read (s). This would represent unequivocal proof of the presence of an AMR gene in the sample, potentially transmissible. For some MinION sequencing runs, the amount of DNA used in this study was not sufficient. Increasing the amount of sample material as input for the extraction could be a solution. Another alternative could be the enrichment of the sample by culture, maybe driven by information on which culture conditions to apply based on prior 16S rRNA analysis, in order to increase the DNA yield, and hence the flow cell output. These improvements could also pave the way towards a broader use of the Flongle flow cell if sufficient DNA quality and quantity can be obtained. It should be highlighted that during our study, the demand for the Flongle exceeded production capacities, especially with the needs for the current SARS-Cov-2 pandemic, leading to long waiting times aggravated by short expiration times that currently cannot match routine lab operating times. Besides, the treatment of the food/feed sample to remove viable cells and DNA as required by EU regulations, might have led to short fragments of damaged DNA already before its extraction. This would impede the possibility of extraction of high molecular weight DNA or the GMM to be enriched anyway. Therefore, although an assembly-free long-read based data analysis workflow would be ideal for unbiased detection of AMR genes, vectors and unnatural associations, the nature of the sample might force the use of assembly-based methods to identify a GMM. Second, the data analysis methodology we proposed is based on easy to use and well-established bioinformatics tools (Kraken, Spades, Blast, etc.). However, the development of push-button bioinformatics pipelines would be needed to allow full implementation in enforcement laboratories. Indeed, although next or third generation sequencers could be present in official control labs, the bioinformatics expertise for the application of these analyses might be missing. In this context, Galaxy (Afgan et al., 2018) could offer the tools we used in this study, in a more user-friendly way, not requiring the use of the command line. Additionally, Galaxy allows to compile workflows which can be shared amongst laboratories, contributing to accessibility and reproducibility. The search for unnatural associations in the proposed workflow is still manual and time-consuming. Other approaches that can be automated could be developed. It needs to be investigated to which extent these could be incorporated into a universal Galaxy workflow, suited for all GMM samples. Moreover, a more extensive analysis, e.g. including SNP-based analysis, could be included to unequivocally prove that a strain detected in the metagenomics sample is identical to a previously characterized and sequenced GMM isolate. Given the current error rate of the long read sequencing, this would be more suited for the short read sequencing only. However, it was shown that for determination of GMM genomes, the long reads help to obtain a more contiguous *de novo* assembly (Berbers et al., 2020). Rapid advances in bioinformatics tools available for ONT data (e.g. basecalling, assembly, polishing) might decrease the error rate on the long reads, which affected the target coverage observed for some reads after Flongle or MinION sequencing of a sample with lower DNA concentration (GMM16). However, this also comes with a cost as developed analysis pipelines might have to be reviewed and updated often. Hybrid assembly, thereby combining the assembly advantage of long reads with the accuracy of short reads, could ameliorate this issue. Although theoretically possible, hybrid assembly was, however, not conducted in this study as it would currently still represent a very high cost to be used routinely by enforcement laboratories. This might change in the future. Moreover, our analysis was only conducted on samples which most probably only contained one species (*B. subtilis*), and it has not yet been tested on more complex samples, in which unnatural associations might be less obvious to detect and genomes even more challenging to assemble. The detection of distinct closely related species and unnatural associations in more complex samples would require further development of appropriate analysis tools and databases. Generally, this open approach can in the future be applied to other GMM used to produce fermentation products like food enzymes. This requires that the corresponding sequence data is available in public databases, as it is able to detect any species and AMR genes / vector present in a sample based on the condition that reference data to compare with is available. Consequently, we believe that, if GMMs cannot fall under the GMO regulation, thereby resulting in no identification method being available (European Parliament and the Council of the European Union, 2003b, 2003a), sharing of information from the industry on all used vectors and species and sequences of GMMs confidentially reported to EFSA with the enforcement laboratories and/or the competent authorities would greatly help in the development of new detection methods, including metagenomics. Indeed, this would increase the list of sequences of genes or shuttle vectors and of known GMM species to look for, thereby facilitating the open approach offered by metagenomics. Such a database would also allow investigating more closely whether specific species found in a sample using taxonomic methods are linked to misclassifications, contaminations or genetic introductions from other species. Also a database of previously sequenced GMM isolates should be constructed as this will also provide more GMM genomes to map the metagenomics reads to. In this study the reference genome most probably linked to the samples was known and therefore could be used as a final confirmation.

In conclusion, this proof-of-concept study delivered a novel way to detect GMMs in food/feed products using shotgun metagenomics, by uncovering unnatural associations linked to the presence of typically used AMR genes and identification of the species. This could all be achieved with the analysis of one sequencing reaction. This confirms the hypothesis of this work. Therefore, this approach would fit within the workflow used by enforcement laboratories when detection of DNA and qPCR screening led to the suspicion of the presence of an unknown GMM such as for sample GMM16, when no isolate can be obtained (i.e. no possibility to do WGS of the isolate to confirm the GMM) and a DNA walking strategy is too laborious and neither successful nor possible because the anchor is not known (Fig. 5). The proposed shotgun metagenomics approach allows the identification and characterization of GMMs. Theoretically, this method can replace the currently used qPCR first and second line analyses steps in the enforcements labs. This includes the detection of AMR genes or event-specific markers for which no qPCR method has been developed yet and the identification of the species, which is currently not a standard procedure. However, until the metagenomics approach is appropriately validated, currently it would rather be used by the enforcement laboratories as an orientation step, requiring confirmation of the findings by PCR and/or Sanger sequencing. With additional protocol optimization allowing longer read lengths in the future, MinION sequencing might allow the immediate detection of full-length AMR genes, thereby supporting risk assessment and a complete *de novo* assembly of the genetically modified strain. This will contribute to an open approach of generalized detection and characterization of unknown GMMs in microbial fermentation products.

Funding information

The research that yielded these results was funded by the Transversal activities in Applied Genomics (TAG) Service from Sciensano, by an internal Sciensano budget and as in-kind contribution of Sciensano within the context of the One Health EJP project JRP12-AMRSH5-FARMED.

CRediT authorship contribution statement

Florence E. Buytaers: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing - original draft, Writing - review & editing. **Marie-Alice Fraiture:** Investigation, Resources, Writing - review & editing. **Bas Berbers:** Investigation, Resources, Visualization, Writing - review & editing. **Els Vandermassen:** Resources, Writing - review & editing. **Stefan Hoffman:** Resources, Writing - review & editing. **Nina Papazova:** Resources, Writing - review & editing. **Kevin Vanneste:** Resources, Software, Writing - review & editing. **Kathleen Marchal:** Supervision, Writing - review & editing. **Nancy H.C. Roosens:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Resources, Validation, Writing - review & editing. **Sigrid C.J. De Keersmaecker:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing - original draft, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank the technicians from the Belgian enforcement laboratory (Transversal Activities in Applied Genomics, Sciensano) for the screening of the samples as well as the NGS sequencing platform from Transversal Activities in Applied Genomics (Sciensano). We also acknowledge Patrick Philipp and his team from Service Commun des Laboratoires (France) for access to the sample of GMM14 (RASFF2014.1249).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.fochms.2021.100023.

References

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B. A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., & Blankenberg, D. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research, 46*(W1), W537–W544. https://doi.org/10.1093/nar/gky379.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A., & Pevzner, P. A. (2012). SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology, 19*(5), 455–477. https://doi.org/10.1089/cmb.2012.0021.

Barbau-Piednoir, E., De Keersmaecker, S. C. J., Delvoye, M., Gau, C., Philipp, P., & Roosens, N. H. (2015). Use of next generation sequencing data to develop a qPCR method for specific detection of EU-unauthorized genetically modified Bacillus subtilis overproducing riboflavin. *BMC Biotechnology, 15*(1), 1–10. https://doi.org/10.1186/s12896-015-0216-y.

Berbers, B., Saltykova, A., Garcia-Graells, C., Philipp, P., Arella, F., Marchal, K., Winand, R., Vanneste, K., Roosens, N. H. C., & De Keersmaecker, S. C. J. (2020). Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified Bacillus. *Scientific Reports, 10*(1), 1–13. https://doi.org/10.1038/s41598-020-61158-0.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics, 30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Buytaers, F. E., Saltykova, A., Denayer, S., Verhaegen, B., Vanneste, K., Roosens, N. H. C., Piérard, D., Marchal, K., & De Keersmaecker, S. C. J. (2020). A Practical Method to Implement Strain-Level Metagenomics-Based Foodborne Outbreak Investigation and Source Tracking in Routine. *Microorganisms, 8*(8), 1191. https://doi.org/10.3390/microorganisms8081191.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics, 10*, 1–9. https://doi.org/10.1186/1471-2105-10-421.

Council Directive 90/220/EEC of 23 April 1990 on the deliberate release into the environment of genetically modified organisms. OJ L 117, 8.5.1990, p. 15–27.

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. *Bioinformatics, 34* (15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149.

Deckers, M., Deforce, D., Fraiture, M. A., & Roosens, N. H. C. (2020). Genetically modified micro-organisms for industrial food enzyme production: An overview. *Foods, 9*(3). https://doi.org/10.3390/foods9030326.

Deckers, M., Vanneste, K., Winand, R., Hendrickx, M., Becker, P., De Keersmaecker, S. C. J., Deforce, D., Marie-Alice, F., & Roosens, N. H. C. (2020). Screening strategy targeting the presence of food enzyme-producing fungi in food enzyme preparations. *Food Control, 117*(April). https://doi.org/10.1016/j.foodcont.2020.107295 107295.

Deckers, M., Vanneste, K., Winand, R., Keersmaecker, S. C. J. D., Denayer, S., Heyndrickx, M., Deforce, D., Fraiture, M. A., & Roosens, N. H. C. (2020). Strategy for the identification of micro-organisms producing food and feed products: Bacteria producing food enzymes as study case. Food Chemistry, 305(February 2019), 125431. https://doi.org/10.1016/j.foodchem.2019.125431.

EFSA (2021). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. *EFSA Journal, March*, 1–13.

European Parliament and the Council of the European Union. (2003a). Regulation (EC) No 1829/2003. Official Journal of the European Union. http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:32003R1829.

European Parliament and the Council of the European Union. (2003b). REGULATION (EC) No 1830/2003. Official Journal of the European Union.

European Parliament and the Council of the European Union. (2008a). Regulation (EC) No 1331/2008. Official Journal of the European Union.

European Parliament and the Council of the European Union. (2008b). Regulation (EC) No 1332/2008. Official Journal of the European Union.

European Parliament and the Council of the European Union. (2008c). Regulation (EC) No 1333/2008. Official Journal of the European Union.

Fraiture, M. A., Bogaerts, B., Winand, R., Deckers, M., Papazova, N., Vanneste, K., De Keersmaecker, S. C. J., & Roosens, N. H. C. (2020). Identification of an unauthorized genetically modified bacteria in food enzyme through whole-genome sequencing. *Scientific Reports, 10*(1), 1–12. https://doi.org/10.1038/s41598-020-63987-5.

Fraiture, M. A., Deckers, M., Papazova, N., & Roosens, N. H. C. (2020a). Are antimicrobial resistance genes key targets to detect genetically modified microorganisms in fermentation products?. *International Journal of Food Microbiology, 331*(February), 108749. https://doi.org/10.1016/j.ijfoodmicro.2020.108749.

Fraiture, M. A., Deckers, M., Papazova, N., & Roosens, N. H. C. (2020b). Detection strategy targeting a chloramphenicol resistance gene from genetically modified bacteria in food and feed products. Food Control, 108(September 2019), 106873. https://doi.org/10.1016/j.foodcont.2019.106873.

Fraiture, M. A., Deckers, M., Papazova, N., & Roosens, N. H. C. (2020c). Strategy to Detect Genetically Modified Bacteria Carrying Tetracycline Resistance Gene in Fermentation Products. *Food Analytical Methods*. https://doi.org/10.1007/s12161-020-01803-6.

Fraiture, M. A., Joly, L., Vandermassen, E., Delvoye, M., Van Geel, D., Michelet, J. Y., Van Hoeck, E., De Jaeger, N., Papazova, N., & Roosens, N. H. C. (2021). Retrospective survey of unauthorized genetically modified bacteria harbouring antimicrobial resistance genes in feed additive vitamin B2 commercialized in Belgium: Challenges and solutions. Food Control, 119(July 2020), 107476. https://doi.org/10.1016/j.foodcont.2020.107476.

Fraiture, M. A., Papazova, N., & Roosens, N. H. C. (2020). DNA walking strategy to identify unauthorized genetically modified bacteria in microbial fermentation products. *International Journal of Food Microbiology, 337*. https://doi.org/10.1016/j.ijfoodmicro.2020.108913 108913.

Grädel, C., Angel Terrazos Miani, M., Barbani, M. T., Leib, S. L., Franziska, S.-R., & Ramette, A. (2019). Rapid and Cost-Efficient Enterovirus Genotyping from Clinical Samples Using Flongle Flow Cells. Genes, 10(659).

Kleinheinz, K. A., Joensen, K. G., & Larsen, M. V. (2014). Applying the ResFinder and VirulenceFinder. December, 1–7.

Kono, N., & Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. *Development, Growth & Differentiation, 61*(5), 316–326. https://doi.org/10.1111/dgd.12608.

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive κ-mer weighting and repeat separation. *Genome Research, 27*(5), 722–736. https://doi.org/10.1101/gr.215087.116.

Leonard, S. R., Mammel, M. K., Lacher, D. W., & Elkins, C. A. (2015). Application of metagenomic sequencing to food safety: Detection of shiga toxin-producing Escherichia coli on fresh bagged spinach. *Applied and Environment Microbiology, 81*(23), 8183–8191. https://doi.org/10.1128/AEM.02601-15.

Leonard, S. R., Mammel, M. K., Lacher, D. W., & Elkins, C. A. (2016). Strain-level discrimination of shiga toxin-producing Escherichia coli in spinach using metagenomic sequencing. *PLoS ONE, 11*(12), 1–21. https://doi.org/10.1371/journal.pone.0167870.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Li, Heng, & Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics, 26*(5), 589–595. https://doi.org/10.1093/bioinformatics/btp698.

Nanopore Protocol. (2019). Genomic DNA by Ligation (SQK-LSK-109) version GDE_9063_v109_revW_14Aug2019.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2015). Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/btv566.

Paracchini, V., Petrillo, M., Reiting, R., Angers-Loustau, A., Wahler, D., Stolz, A., Schönig, B., Matthies, A., Bendiek, J., Meinel, D. M., Pecoraro, S., Busch, U., Patak, A., Kreysa, J., & Grohmann, L. (2017). Molecular characterization of an unauthorized genetically modified Bacillus subtilis production strain identified in a vitamin B 2 feed additive. *Food Chemistry, 230*, 681–689. https://doi.org/10.1016/j.foodchem.2017.03.042.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., & Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology, 29*(1), 24–26. https://doi.org/10.1038/nbt.1754.

Sayers, E. W., Agarwala, R., Bolton, E. E., Brister, J. R., Canese, K., Clark, K., ... Ostell, J. (2019). Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research, 47*(D1), D23–D28. https://doi.org/10.1093/nar/gky1069.

Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics, 30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Silano, V., Barat Baviera, J. M., Bolognesi, C., Brüschweiler, B. J., Cocconcelli, P. S., Crebelli, R., Gott, D. M., Grob, K., Lampi, E., Mortensen, A., Rivière, G., Steffensen, I. L., Tlustos, C., Van Loveren, H., Vernis, L., Zorn, H., Glandorf, B., Herman, L., Aguilera, J., & Chesson, A. (2019). Characterisation of microorganisms used for the production of food enzymes. *EFSA Journal, 17*(6), 1–13. https://doi.org/10.2903/j.efsa.2019.5741.

Somerville, V., Lutz, S., Schmid, M., Frei, D., Moser, A., Irmler, S., Frey, J. E., & Ahrens, C. H. (2018). Long read-based de novo assembly of low complex metagenome samples results in finished genomes and reveals insights into strain diversity and an active phage system. *BioRxiv, 476747*. https://doi.org/10.1101/476747.

WHO. (2018). Whole genome sequencing for foodborne disease surveillance.

Winand, R., Bogaerts, B., Hoffman, S., Lefevre, L., Delvoye, M., Van Braekel, J., Fu, Q., Roosens, N. H. C., De Keersmaecker, S. C. J., & Vanneste, K. (2020). Targeting the 16s rRNA gene for bacterial identification in complex mixed samples: Comparative evaluation of second (illumina) and third (oxford nanopore technologies) generation sequencing technologies. *International Journal of Molecular Sciences, 21*(1), 1–22. https://doi.org/10.3390/ijms21010298.

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *BioRxiv, 762302*. https://doi.org/10.1101/762302.

Yang, X., Noyes, N. R., Doster, E., Martin, J. N., Linke, L. M., Magnuson, R. J., Yang, H., Geornaras, I., Woerner, D. R., Jones, K. L., Ruiz, J., Boucher, C., Morley, P. S., & Belk, K. E. (2016). Use of metagenomic shotgun sequencing technology to detect foodborne pathogens within the microbiome of the beef production chain. *Applied and Environment Microbiology, 82*(8), 2433–2443. https://doi.org/10.1128/AEM.00078-16.