



OPEN

## Characterization, genome analysis and genetic tractability studies of a new nanocellulose producing *Komagataeibacter intermedius* isolate

Pietro Cannazza<sup>1</sup>, Antti J. Rissanen<sup>2</sup>, Essi Sarlin<sup>2</sup>, Dieval Guizelini<sup>3</sup>, Carlotta Minardi<sup>2</sup>, Pauli Losoi<sup>2</sup>, Francesco Molinari<sup>1</sup>, Diego Romano<sup>1</sup> & Rahul Mangayil<sup>2,4</sup>✉

Bacterial nanocellulose (BC) is a highly versatile biopolymer currently pursued as a material of choice in varied themes of biomedical and material science research fields. With the aim to extend the biotechnological applications, the genetic tractability of the BC producers within the *Komagataeibacter* genus and its potential as an alternative host chassis in synthetic biology have been extensively studied. However, such studies have been largely focused on the model *Komagataeibacter* spp. Here, we present a novel *K. intermedius* strain capable of utilizing glucose, and glycerol sources for biomass and BC synthesis. Genome assembly identified one bacterial cellulose synthetase (*bcs*) operon containing the complete gene set encoding the BC biogenesis machinery (*bcsI*) and three additional copies (*bcsII–IV*). Investigations on the genetic tractability confirmed plasmid transformation, propagation of vectors with pBBR1 and p15A origin of replications and constitutive and inducible induction of recombinant protein in *K. intermedius* ENS15. This study provides the first report on the genetic tractability of *K. intermedius*, serving as starting point towards future genetic engineering of this strain.

Bacterial nanocellulose (BC) is a versatile biopolymer predominantly synthesized by species of *Acetobacteraceae* family and is widely studied in *Komagataeibacter* genera. In *Komagataeibacter* spp., BC synthesis is controlled by cyclic dimeric guanosine monophosphate (c-di-GMP) and occurs via the catalytic activity of bacterial cellulose synthetase (*bcs*) operon encoding BcsABCD proteins<sup>1</sup>. Briefly, the process involves c-di-GMP activation of BcsAB and formation of glucan units employing UDP-glucose, generated via the enzymatic activities of glucokinase, phosphoglucomutase and UDP-glucose pyrophosphorylase. The glucan units are then crystallized into nanocellulose fibrils by BcsD and secreted through the outer membrane pore formed by BcsC, which is later bundled extracellularly to form BC pellicle<sup>2</sup>. Unlike plant-based cellulose, BC offers unique properties such as high purity (i.e., devoid of lignin, hemicellulose, or pectin), biocompatibility, high water retention, superior mechanical and optical properties which has been utilized commercially in medical wound-dressings, health foods and presently researched as a sustainable material in electrical, sensing and biomedical applications<sup>2</sup>.

Among the *Komagataeibacter* spp., *K. xylinus*, *K. rhaeticus* and *K. hansenii* are considered as the model organisms for studies on nanocellulose production<sup>3,4</sup>. Despite the bioprocess and engineering advancements in the model organisms<sup>5,6</sup>, low production titers have encouraged the researchers to isolate novel nanocellulose producing bacteria<sup>7–10</sup>. Although the studies include whole genome analysis and characterization of the isolates' capacity to valorize agro- and industrial waste effluents for BC biogenesis<sup>9,11</sup>, studies on the genetic tractability of *Komagataeibacter* spp. have been largely constrained to the model nanocellulose producers<sup>6</sup>. Here we report the isolation and characterization of a novel *K. intermedius* isolate from Kombucha SCOBY (symbiotic culture of bacteria and yeast). Genome sequencing and whole-genome analysis identified new genes associated with

<sup>1</sup>Department of Food, Environmental and Nutritional Sciences (DeFENS), University of Milan, Via Celoria 2, 20133 Milan, Italy. <sup>2</sup>Faculty of Engineering and Natural Sciences, Tampere University, Tampere, Finland. <sup>3</sup>Graduate Program in Bioinformatics, Sector of Professional and Technological Education, Federal University of Parana (UFPR), Curitiba, PR, Brazil. <sup>4</sup>Department of Bioproducts and Biosystems, Aalto University, Espoo, Finland. ✉email: rahul.mangayil@tuni.fi; rahul.mangayil@aalto.fi

nanocellulose synthesis and putative genes involved in pentose bioconversion. Furthermore, the work studied the applicability of transformation protocols established for related *Komagataeibacter* spp. and identified suitable plasmid backbones, constitutive and inducible promoter systems.

## Materials and methods

**Chemicals and materials.** Kombucha SCOBY, chemicals, GeneJET Genomic DNA Purification Kit and crude glycerol was in this study were obtained from sources mentioned in Mangayil et al.<sup>9</sup>.

**Isolation, characterization, and culturing of BC-producing strains.** Isolation and biochemical characterization of the bacterial cellulose producing isolate was conducted as described in Mangayil et al.<sup>9</sup>. For pre-inoculum preparation, 10 µl from the glycerol stock were inoculated into 5 ml HS-glucose medium (g/L, 5 peptone, 5 yeast extract, 2.7 Na<sub>2</sub>HPO<sub>4</sub>, 1.15 citric acid and 20 g/L glucose) and incubated statically for 5 days at 30 °C. The BC pellicles were lysed overnight (O/N) with 1% cellulase and the released cells washed thrice with 1X Phosphate Buffered Saline (PBS; g/L, g/L, 8 NaCl, 0.2 KCl, 1.44 Na<sub>2</sub>HPO<sub>4</sub>, and 0.24 KH<sub>2</sub>PO<sub>4</sub>; pH 7.4) were used as the pre-inoculum in subsequent experiments. The BC production experiments were conducted through static cultivations in 6-well culture plates (Argos Technologies, Cole-Parmer, US) containing HS medium individually supplemented with 20 g/L of studied carbon sources<sup>9</sup>.

**BC film preparation and material characterization.** BC film preparation and dry weight measurements were conducted as mentioned in Mangayil et al.<sup>12</sup>. Structural and chemical properties of the BC sheets synthesized from glucose, pure glycerol and crude glycerol were analysed using a scanning electron microscope (SEM; Zeiss ULTRAPlus, Germany) and X-ray diffractometer (XRD; Empyrean multipurpose diffractometer, PANalytical B.V, US), respectively, as described in Mangayil<sup>12</sup>. The thermal properties were analysed under N<sub>2</sub> atmosphere using a Thermogravimetric analyser (TGA; TG 209 F3 Tarsus, Netzsch-Gerätebau GmbH, Germany<sup>9</sup>).

**Identification and phylogenetic classification of the BC producing bacterium.** For strain identification, the isolate was statically cultured in loosely capped 50 ml Corning tubes at 30 °C in 10 ml HS-glucose medium for 5 days. The BC pellicle was lysed, and the cells were washed as previously described. The genomic DNA (gDNA) was extracted using the GeneJET Genomic DNA Purification Kit (Thermo Scientific, USA) as per the manufacturer's instructions. The 16S rRNA gene sequence amplified using the identification service from Macrogen (primer pairs 27F, 5'-AGAGTTTGATCMTGGCTCAG-3' and 1492R, 5'-TACGGYTACCTT GTTACGACTT-3') and sequencing (primer pairs 785F, 5'-GGATTAGATACCCTGGTA-3' and 907R, 5'-CCG TCAATTCMTTTRAGTTT-3') can be found in the NCBI GenBank database under the accession number MT094082. Homology comparisons of the 16S rRNA gene were conducted using the nucleotide BLAST (blastn)<sup>13</sup> tool against the NCBI GenBank 16S rRNA gene sequence repository for *Komagataeibacter* (taxid:1434011). Using the MEGA X<sup>14</sup> integrated tool, alignments against the 16S rRNA gene sequences of *Komagataeibacter* type strains were conducted using ClustalW<sup>15</sup> and the evolutionary relationship was inferred using the Neighbor-Joining method<sup>16</sup> with Kimura 2-parameter model<sup>17</sup> and bootstrapping (1000 times) at uniform rates.

**Genome sequencing, assembly and bioinformatics.** *K. intermedius* ENS15 gDNA was sequenced at Novogene Europe (Cambridge, UK) using the Illumina Novaseq 6000 system. The raw reads were quality controlled using FastQC<sup>18</sup>, trimmed using Trimmomatic<sup>19</sup> and de-novo assembled using SPAdes<sup>20</sup>. The misassemblies were identified using QUAST (5.0.2)<sup>21</sup>. Contig reordering, misassembly correction and gap filling was conducted with GFinisher<sup>22</sup> tool using *K. intermedius* AF2 genome (GenBank accession: GCA\_000817255.2) as the reference. The scaffolds were manually linked, the genome completeness was analysed by CheckM tool<sup>23</sup>, and annotated using Prokka<sup>24</sup>. Phylogenetic relatedness and genome-based taxonomy was performed in Type Strain Genome Server (TYGS)<sup>25</sup> (<https://tygs.dsmz.de/>, accessed on 23.07.2021). The plasmid sequences in the raw reads and draft genome were identified using plasmidSPAdes<sup>26</sup> and Recycler<sup>27</sup>, respectively. The genome assembly can be found in the NCBI GenBank under the accession number GCA\_000964425.1.

The genes encoding the Bcs machinery, extracellular matrix formation, and carbohydrate uptake and catabolism were identified through manual search within the Prokka annotated GenBank file using Unipro UGENE software (v. 33.0)<sup>28</sup> and confirmed using protein BLAST (blastP)<sup>13</sup> against the NCBI non-redundant database. The phylogenetic relatedness and amino acid identities of *bcs* operon proteins, BcsAB and BcsC, were identified using MUSCLE using the default settings<sup>29</sup>. The phylogenetic tree was built in MEGA X integrated tool<sup>14</sup> using the Neighbor-Joining method and the evolutionary distances were computed using the Poisson correction method<sup>30</sup>. The domain architecture predictions were conducted through the NCBI conserved domain database (NCBI CDD)<sup>31</sup> and InterProScan<sup>32</sup> searches.

**Preparation of electrocompetent *K. intermedius* ENS15 cells and plasmid transformation.** The electrocompetent cells were prepared as described by Mangayil et al.<sup>12</sup>. To determine the appropriate antibiotic concentration for plasmid transformation, single aliquots (100 µl) of *K. intermedius* ENS15 competent cells were plated, in triplicates, on to HS-glucose agar supplemented with kanamycin (Kan; 400 mg/L, 500 mg/L, 600 mg/L, 700 mg/L) and chloramphenicol (Cm; 200 mg/L, 300 mg/L, 340 mg/L, 400 mg/L, 500 mg/L, 550 mg/L, 600 mg/L, 700 mg/L). The colony forming units (CFUs) were calculated after six days of static incubation at 30 °C. A blank cultivation, HS-glucose agar plates devoid of antibiotic supplementation and cells, was included as control.

Plasmids with seven different origins of replication (Table S1) obtained from Standard European Vector Architecture (SEVA) repository (<http://seva-plasmids.com/>) were tested for its replication in *K. intermedius* ENS15. As kanamycin was identified as an unsuitable selection marker, the native antibiotic resistance gene in the SEVA vectors with kanamycin resistance gene (pSEVA211, pSEVA241, pSEVA261, pSEVA271, pSEVA281 and pSEVA291) were replaced with chloramphenicol resistance (CmR) gene from pSEVA331 via SmaI/BoxI restriction and one-step cold phosphorylation-ligation using T4 poly nucleotide kinase and T4 DNA ligase. The resulting plasmids, named according to the SEVA guidelines as pSEVA311, pSEVA341, pSEVA361, pSEVA371, pSEVA381 and pSEVA391, were verified using PCR [PS5\_SEVA and PS6\_SEVA (targeting the origin of replication) and Fwd\_CmR and Rev\_CmR (targeting the CmR gene) primers (Table S1)] and sequencing.

Plasmid DNA (~200 ng) mixed with 100 µl of electrocompetent *K. intermedius* ENS15 cells were transferred to 1 mm electroporation cuvettes, pulsed (MicroPulser, BioRad) at 2.5 kV and immediately recovered with 1 ml of prewarmed HS medium containing 5% cellulase. The recovered cells were transferred to 15 ml culture tubes and incubated O/N at 30 °C and 220 rpm. After the incubation period, the transformants were plated on individual HS-glucose agar plates containing appropriate antibiotic concentrations and grown statically at 30 °C for 5–7 days. The transformation efficiency (%) was calculated by dividing CFU with the transformed DNA amount.

**Characterization of constitutive promoters.** To identify a suitable promoter for constitutive gene expression in *K. intermedius* ENS15, pSB1C3 vector harboring eight constitutive promoters (J23104, J23105, J23106, J23107, J23100, J23111, J23113, J23116, J23118) were selected from the *E. coli* Anderson promoter collection (iGEM Parts Registry, Spring 2017 Distribution). Promoters J23104 and J23105 that have been previously demonstrated to cater the strongest and weakest gene expression, respectively, in *Komagataeibacter* spp. were included as experimental controls<sup>4,6</sup>. To confirm plasmid replication and reliable gene expression, the genetic elements (promoter, RBS, mRFP1 and double terminator) in pSB1C3 were restricted with XbaI/PstI and ligated to similarly digested pSEVA331 vector. The resulting plasmids (Table S1) were transformed into *K. intermedius* ENS15 and plated onto HS-glucose agar containing 320 mg/L Cm and 1% v/v cellulase. Single colonies were individually picked and cultivated in 5 mL HS-Glu containing 32 µg/ml Cm and 1% v/v cellulase at 30 °C at 300 rpm for 3 days. Cells were collected (8000 g for 5 min), washed twice with 1X PBS and resuspended in 1 ml PBS. The promoter strength was evaluated by analysing the mRFP1 fluorescence, normalized per unit optical density at wavelength 600 nm ( $OD_{600nm}$ ), using Spark Multimode Microplate Reader (TECAN, Switzerland) at excitation wavelength 580 nm and emission wavelength 610 nm. For constitutive mRFP1 expression in pellicles, 50 µL of the recombinant *K. intermedius* cells harbouring the selected genetic elements were inoculated in 6-well culture plates and statically grown in 10 ml of HS-glucose medium containing 32 mg/L CmR at 30 °C for 6 days.

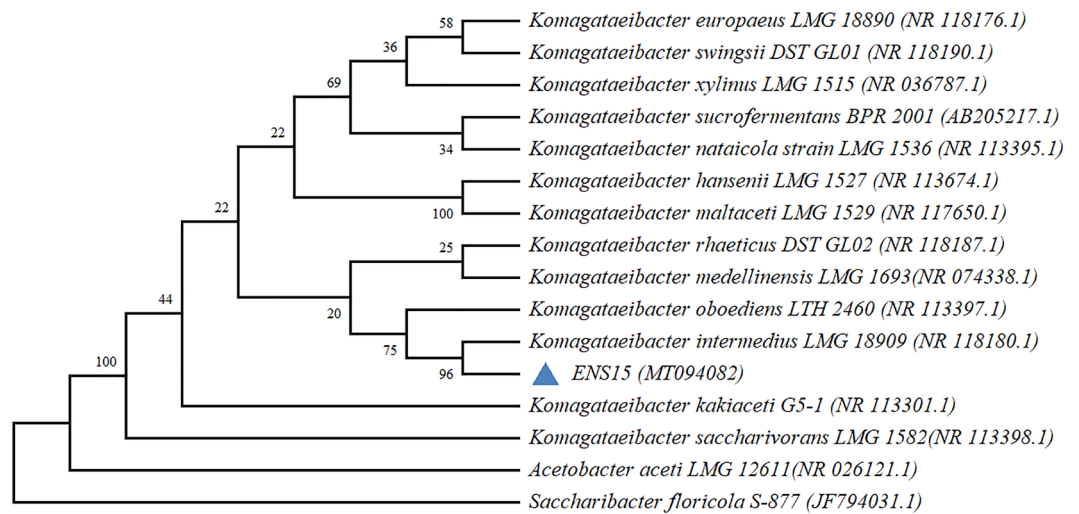
**Characterization of inducible gene expression constructs.** The suitability of pLux-LuxR (Addgene #78281) and pTet-TetR (Addgene #78283), inducible systems that has been reported to function in *K. xylinus* and *K. rhaeticus*<sup>4,6</sup> strains were tested for its functionality in *K. intermedius* ENS15. Furthermore, mRFP1 gene induction via a Cumate inducible system (Addgene #119872) was studied in *K. intermedius* ENS15<sup>33</sup>. To ensure functionality of the Cumate inducible system in *K. intermedius* ENS15, the induction cassette (comprising of CuO/CymR, RBS and sfGFP) was amplified using EcoRI\_Cumate and xbaI\_Cumate primers from pCT5-bac2.0 (Table S2), restricted with EcoRI/XbaI enzymes and ligated with similarly digested pSEVA331 vector, resulting in pSEVA331Cumate\_sfGFP plasmid. To avoid any discrepancy due to varying reporter gene and RBS among the studied vectors, the sfGFP and RBS in pSEVA331Cumate\_sfGFP plasmid was replaced with the RBS and mRFP1 from pSEVA J23104 mRFP1. The pSEVA 331\_Cumate\_sfGFP plasmid backbone (amplified using Vector.FOR\_Cumate and Vector.REV\_Cumate primers to exclude the RBS and sfGFP gene) was assembled with the PCR product (containing RBS and mRFP1 from J23104-mRFP1-331Bb) amplified using Fragment.FOR\_Cumate and Fragment.REV\_Cumate primers via NEBuilder HiFi DNA Assembly Master Mix, resulting in pCumate plasmid. Plasmid transformation, propagation of recombinant *K. intermedius* ENS15 cells and fluorescence tests were conducted as described in Section “Characterization of constitutive promoters”.

## Result and discussion

**Isolation, characterization and classification of BC-producing strain.** Isolation of single clones from CaCO<sub>3</sub> halo zones in Glucose-Yeast Extract-Calcium carbonate agar and iterated subculturing in HS-Glu agar resulted in enrichment of an isolate with beige-coloured, smooth-edged and umbonate shaped colonies characteristics (Fig. S1A). The isolate is hereafter called ‘ENS15’. Under 100X magnification, the isolate appeared as rod-shaped cells either singularly, in pairs or in chains, having cell sizes in the range of 2.6–5.0 µm \* 0.6–0.7 µm (Fig. S1B). Phylogenetic positioning with the 16S rRNA gene sequences of *Komagataeibacter* type strains affiliated the isolate to *K. intermedius* LMG 18909<sup>T</sup> (Fig. 1).

The ENS15 strain was able to grow in MA/9 medium supplemented with 30% glucose but grew poorly when cultured in medium containing 3% ethanol, 0.35% acetic acid, and 3% ethanol and 4% acetic acid as the sole carbon sources. Therefore, the remaining studies were conducted in PY medium (Fig. S2). The isolate did not require acetic acid for growth, demonstrated acetate and lactate oxidization, and acetic acid overoxidation (Fig. S3).

**Whole-genome analysis of *K. intermedius* ENS15.** To gain valuable insights on carbohydrate uptake and metabolism and BC biogenesis machinery, *K. intermedius* ENS15 genome was sequenced to a coverage of 400X and assembled de novo using SPAdes. Due to the low genome quality of the *K. intermedius* type strain LMG 18909<sup>T</sup> (Strain TF2, GenBank assembly accession: GCA\_000964425.1), scaffolding with GFinisher was conducted using *K. intermedius* AF2 genome (GenBank assembly accession: GCA\_000817255.2) as the reference. The assembly statistics are presented in Table S3.



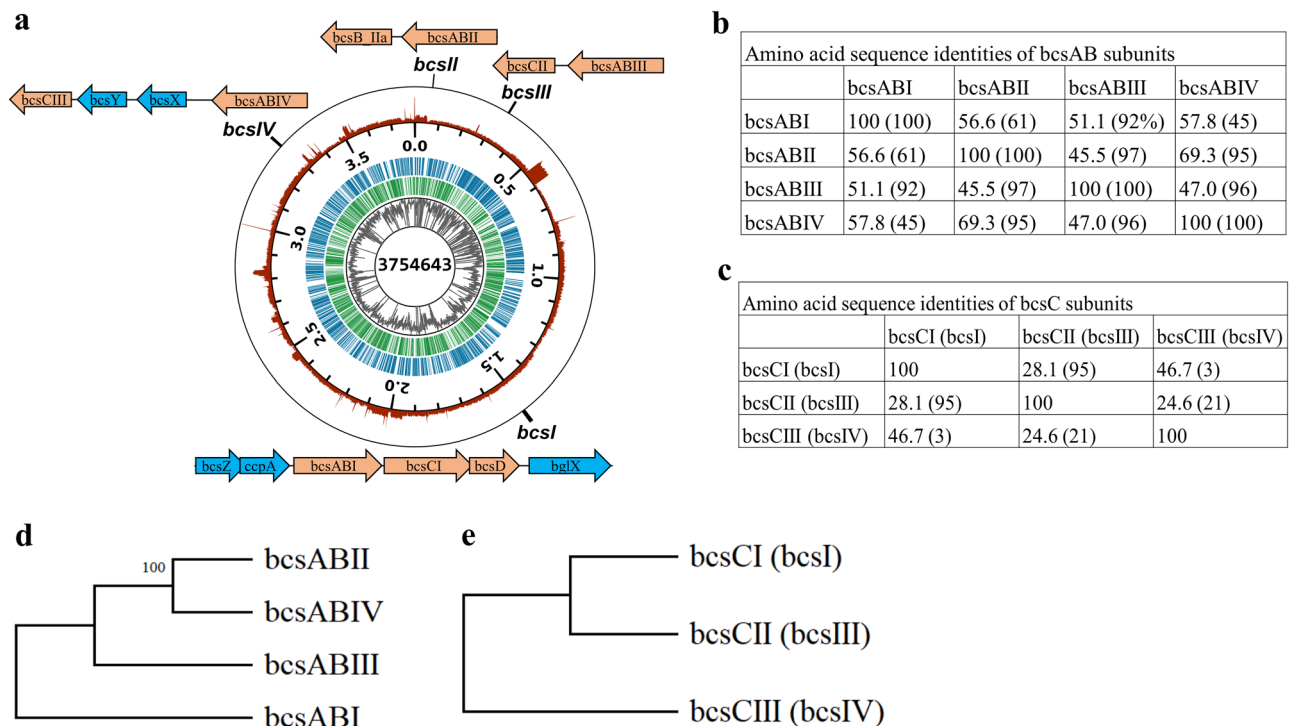
**Figure 1.** 16S rRNA gene-based phylogenetic relationship of ENS15 and *Komagataeibacter* type strains (NCBI accession numbers in parenthesis). The tree, drawn using MEGA X integrated tool<sup>14</sup> (<https://www.megasoftware.net/>), version 10.2.4, was rooted using *Saccharibacter floricola* JCM 12116. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches.

To confirm the strain's phylogenetic positioning among the *Komagataeibacter* sp., a genome-based validation using the TYGS database was conducted. The server calculates the digital DNA–DNA hybridization (dddH) values between the query and subject genomes based on three Genome BLAST distance phylogeny formulas: Genome to genome distance calculator formula 1 (GGDH;  $d_0$ ) calculates the length of all high-scoring segment pair (HSP) divided by total genome length, GGDH formula 2 ( $d_4$ ) calculates the sum of all identities found in HSPs divided by overall HSP length and GGDH formula 3 ( $d_6$ ) calculates the sum of all identities found in HSPs divided by total genome length. As the formula  $d_4$  is independent of genome length and thus robust against the use of incomplete draft, it is used for validation. The pairwise comparison results confirmed the isolate's affiliation with *K. intermedius* type strain LMG 18909<sup>T</sup> and *K. intermedius* AF2 with a digital DNA–DNA hybridization (distance formula  $d_4$ ; sequence level similarity; subspecies clustering threshold of 70%) score of 99% and 97%, respectively (Table S4).

The genome data indicated that the *K. intermedius* ENS15 has one chromosome of 3,754,643 bp (GC content, 61.8%) (Fig. 2) comprising of 3467 coding domain regions (CDSs), 120 repeat DNA, 19 non-coding RNA, 46 tRNAs and 1 rRNA (Table S3). Through manual search along the Prokka annotated files, blastP searches, and domain predictions (NCBI CDD, InterproScan), the genes encoding the proteins for bacterial cellulose synthesis, carbohydrate uptake and metabolism and gluconeogenesis were identified from the genome. Amino acid sequences of the aforementioned enzymes are presented in the Supplementary material.

At least one putative plasmid sequences, pKi (~80 Kb), containing genes encoding DNA replication, recombinant and repair, relaxases, DNA-binding transcriptional regulator, type IV bacterial secretion system, transposases and toxin-antitoxin systems were identified with plasmidSPAdes. Although DNA replication machineries could not be determined, two additional putative plasmid sequences pKi2 (22 Kb) and pKi3 (13Kbp), mostly containing genes encoding hypothetical proteins and transposases were identified. Recycler indicated that the draft genome did not contain any plasmid sequences. The presence of plasmid DNA within other *Komagataeibacter* spp. genomes have been reported. For instance, an abundance of plasmid DNA, ranging from four to seven have been identified from *K. rhaeticus* ENS9a, *K. xylinus* E25, *K. medellinensis* NBRC 3288 and *K. nataicola* RZS01 genomes, suggests the phenotypic diversity among the bacterial cellulose producers<sup>36</sup>. Blastn alignments of pKi1 and pKi2 sequences with the NCBI non-redundant nucleotide collection of *Komagataeibacter* taxid (id, 1,434,011) indicated similarities to plasmids present in *K. medellinensis* NBRC 3288 (pGXY010, 96% identity and 81% coverage) and *K. xylinus* E25 (pGX5, 98% identity and 88% coverage) strains, respectively.

**Predicted BC biogenesis machineries.** The BC synthetase complex, spanning the outer and inner membranes involved in BC synthesis and export to extracellular milieu, and the adjoining gene clusters present in the *K. intermedius* ENS15 genome is illustrated in Fig. 2a. Four *bcs* operons were annotated in *K. intermedius* ENS15. As previously reported for *Komagataeibacter* spp.<sup>4,5,9,10,37</sup>, only one operon contained the complete genes set encoding the BC biogenesis machinery (*bcsAI*, *bcsBI*, *bcsCI* and *bcsD*) was found in *K. intermedius* ENS15 genome (genomic position 1,511,250:1,520,353 bp). Though annotated as individual genes, a manual search identified a 14 bp overlap within *bcsAI* and *bcsBI* genes (*bcsABI*) and 1 bp overlap between *bcsCI* and *bcsD*. The genes encoding the BC biogenesis accessory proteins endoglucanase (BcsZ) and cellulose complementing factor A (CcpA), observed as gene fusions, and beta-glucosidase (BglX) were identified at proximal upstream and downstream regions of *bcsI* at genomic positions 1,508,964:1,511,059 bp and 1,520,578:1,522,779 bp, respec-



**Figure 2.** *K. intermedius* ENS15 genome. **(a)** Genome map. The genome contains one chromosome of 3,754,643 bp (GC content 61.8%) with a predicted total of 3467 genes, 120 repeat DNA, 47 tRNAs and 19 non-coding RNA. The displayed data from perimeter to center are, coverage (red), chromosome position (major ticks 500 kbp, minor ticks 100 kbp), forward strand CDSs (in the middle, blue), reverse strand CDSs (in the middle, green), GC-percentage (gray), and chromosome size (bp). The GC-percentage ranged from 17.6 to 69.8% and coverage (number of reads mapping to the locus) from 653 to 14,700. Both GC-percentage and coverage were calculated with 1 kbp window size. Coverage is displayed with logarithmic scaling. The genome map was prepared by matplotlib<sup>34</sup> (3.3.4; <https://matplotlib.org/3.3.4>), pysam 0.15.3 (<https://github.com/pysam-dev/oper/pysam>), and biopython<sup>35</sup> (1.78; <https://biopython.org/docs/1.78>) libraries using python language (3.7.11; <https://www.python.org/>). The bcs operon and the surrounding genes are indicated by manually curated figures directing to its position in the genome. **(b,c)** and **(d,e)** respectively indicates the amino acid sequence identities (MUSCLE program<sup>29</sup>) and the phylogenetic relationship between the predicted bcsAB and bcsC genes present in genome. The phylogenetic tree was build using MEGA X integrated tool<sup>14</sup> (<https://www.megasoftware.net/>, version 10.2.4).

tively. Furthermore, a standalone copy of *bcsZ* gene (*bcsZII*; 28% amino acid sequence identity to BcsZ with a 23% coverage) was identified at position 2,042,498:2,043,523 bp.

Three additional copies of *bcs* operon were present in complement in the genome. The second *bcs* operon (*bcsII*) found at genomic position 60,840:64,661 bp, contained fused *bcsAII* and *bcsBII* genes (*bcsABII*). In silico analysis in NCBI conserved domain database, identified BcsABII to contain domain architectures to glycosyltransferases and BcsB superfamily, confirming the annotation. An interesting observation was the presence of a gene annotated to encode cyclic di-GMP binding protein 31 bps upstream of *bcsABII* (genomic position 60,041:60,808; Fig. S4), which has not yet reported in *Komagataeibacter* spp. BlastP search of the gene against *Komagataeibacter* taxid (id:1434011) resulted in close alignments with *K. intermedius* AF2 cation tolerance protein (100% similarity; accession number KPH88887.1) and *K. oboediens* cellulose synthase catalytic subunit (96% similarity, Accession Number GCE79538.1). NCBI CDD and InterProScan searches predicted the amino acid sequence to contain domain hits for BcsB super family and cellulose synthase, subunit B (IPR003920), allowing to annotate the predicted gene as *bcsB\_IIa*. Nevertheless, further validations are required as the annotation could be attributed to the artefacts in the prediction tools which maintains the old annotation of BcsB with regards to c-di-GMP binding. A translated blast search of BcsII\_a against the nucleotide sequences of homologous *bcsB* gene from *Komagataeibacter* spp. demonstrating 39–41% sequence similarity (89% query coverage; Fig. S5), supports the hypothesis. Alternatively, such annotations could be related to incorrectly predicted CDS caused by a mutation within the downstream *bcsAB* reading frame. This is crucial to validate as such mutations can lead to premature termination at the transcript-level.

The third *bcs* operon (*bcsIII*) contained fused *bcsAB* genes (*bcsABIII*) at genomic position 335,426:339,913 bp. A second copy of *bcsC* gene (*bcsCII*) was found upstream of the *bcsABIII* gene (position 331,628:335,332 bp). The fourth *bcs* operon (*bcsIV*) comprising of fused *bcsAB* (*bcsABIV*), genes encoding for SGNH/GDSL hydrolase family protein (BcsX), acyltransferase (BcsY) and BcsC (*bcsCIII*) was found at genomic position 3,293,852:3,304,158 bp. Though the exact function of BcsX and BcsY proteins have not been yet characterized, presence of genes encoding for the stated proteins have been reported in *Komagataeibacter* spp<sup>9,38</sup>.

Variable copy numbers of *bcs* operon genes have been reported among *Komagataeibacter* spp. For instance, four copies of *bcs* operon genes have been identified in *K. intermedius* AF2 (draft genome, GenBank: GCA\_000817255.2)<sup>36</sup>, *K. rhaeticus* (NZ\_CP050139.1, NZ\_LT575493.1)<sup>5,9</sup> and *K. xylinus* (NZ\_CP024644.1) strains, whereas *K. hansenii*, a high yield bacterial cellulose producer, contains three *bcs* operons<sup>4</sup>. In general, apart from the ancillary *bcsB\_Ila* gene found at *bcsII*, *K. intermedius* ENS15 shares similar arrangement and composition of *bcs* operon genes as reported for *Komagataeibacter* spp. isolates in the literature. Nevertheless, absence of a complete operon, i.e., *bcsD*, *bcsZ*, *ccpA* and *bgIX* in *bcsII–bcsIV*, suggests that the three additional operon copies might have risen during evolution from duplications and deletion events of *bcsI* operon. Phylogenetic and amino acid sequence identities of BcsAB and BcsC proteins present in the four *bcs* operons are presented in Fig. 2b–e. Phylogenetic analysis of BcsAB proteins suggest *bcsII* and *bcsIV* to be closely related in comparison to the full length *bcsI* (similarity, 56–59%; coverage, 45–61%) and *bcsIII* (similarity, 45–47%; coverage, 96–97%). Whereas *bcsI* and *bcsIII* demonstrated a 51.1% sequence similarity from a coverage of 92%. Similar observation was made from BcsC phylogeny, wherein the highest similarity of 28.1% (coverage 95%) was found among *bcsI* and *bcsIII* operons.

**Insights on carbohydrate metabolism and BC production by *K. intermedius* ENS15.** Glucose is an excellent and widely used carbon source for BC production in *Komagataeibacter* spp. Prokka annotations, NCBI CDD domain predictions and blastP searches against the NCBI non-redundant database identified an operon containing the genes encoding phosphoenolpyruvate phosphotransferase (PtsI) system (genomic positions 3,146,276:3,150,394) involved in glucose uptake and utilization<sup>36</sup>. The operon contained, in the order, the genes encoding phosphoenolpyruvate-protein phosphotransferase and the phosphocarrier protein, PTS fructose transporter subunit IIA, and a bifunctional kinase/phosphorylase (involved in catabolite repression and phosphorylation of phosphocarrier protein). Similar to other *Komagataeibacter* spp., *K. intermedius* ENS15 genome did not indicate an annotation for the gene encoding phosphofructokinase (catalysing the conversion of fructose 6 phosphate to fructose 1,6-biphosphate) resulting in an incomplete Embden-Meyerhof-Parnas (EMP or glycolytic) pathway<sup>9,10</sup>. In *Komagataeibacter* spp., glucose oxidation takes two routes, either via glucose-6-phosphate (entering into pentose phosphate pathway, PPP) or through gluconate generation (through the catalytic activity of glucose dehydrogenase). The generated gluconate is either phosphorylated to 6-phosphogluconate (entering into central metabolic pathways for biomass formation) or oxidized to ketogluconate (exported to the cultivation medium). Corroborating to the genome information, after a 10-day cultivation period, the *K. intermedius* ENS15 completely utilized glucose generating gluconic acid as the major liquid end-metabolite (6.2 ± 1.8 g/L gluconic acid, corresponding to 31% of the initial glucose concentration), resulting in a final medium pH of 3.2. In addition to the pH drop, gluconate synthesis reduces the glucose availability for BC biosynthesis, as the enzymes phosphoglucomutase and UDP-glucose pyrophosphorylase provides the precursor UDP-glucose moieties to the BC machinery. After the cultivation period, *K. intermedius* ENS15 strain synthesized a BC titer of 1.1 ± 0.1 g/L (yield, 0.7 mg/g<sub>substrate</sub>; BC characterization, Fig. S6 and Table S5). The BC titer obtained from *K. intermedius* ENS15 are comparable to that reported from *K. intermedius* isolates grown in glucose. For instance, Fernández' et al. (2019) reported a titer of 1.2 g/L BC using *K. intermedius* JF2 isolate and observed a marginal improvement (1.6 g/L) when cultivated in HS medium containing glucose (10 g/L) and mannitol (20 g/L)<sup>39</sup>. In another study, using a base-resistant *K. intermedius* FST213-1 strain, isolated from fermented fruit juice, a BC titer of 2.3 g/L was reported after a 9-day static cultivation in HS medium containing 20 g/L glucose<sup>40</sup>. Table 1 compares the BC titers obtained from this study to that reported for *Komagataeibacter* spp. grown in HS medium containing glucose.

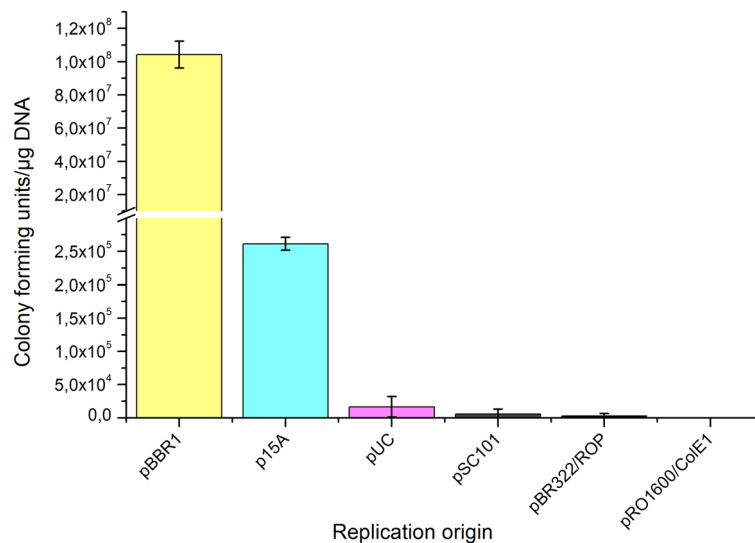
As opposed to glucose, glycerol has been regarded as an ideal substrate for BC biogenesis due to the substrate's capacity to bypass gluconate generation and improve process sustainability<sup>9,10</sup>. While an annotation specific for glycerol uptake facilitator protein was not present, three gene copies encoding aquaporin Z was identified from the genome. NCBI CDD searches confirmed that among the three copies, two genes encoded aquaporin Z (Genomic Positions 611,611:612,540 bp and 3,112,617:3,113,501 bp) and one showed domain architectures similar to glycerol uptake facilitator (Genomic Position 2,538,382:259,365 bp). Manual search along the predicted *glpF* locations identified genes encoding glycerol kinase (position 2,541,272:2,542,771 bp) and Glycerol-3-phosphate regulon repressor (position 2,544,983:2,545,741), and glycerol dehydrogenase (position 2,412,749:2,414,956 bp) downstream and upstream to *glpF*, respectively. Putative gene encoding dihydroxyacetone kinase was present at genomic location 612,665:614,299 bp, downstream of aquaporin Z\_I gene. Furthermore, multiple copies of genes encoding oxaloacetate-decarboxylating malate dehydrogenase (catalysing oxaloacetate to pyruvate reaction) and pyruvate phosphate dikinase (catalysing the transformation of pyruvate to phosphoenol pyruvate) were identified from the genome, hinting possible gluconeogenic function from the central metabolic pathways<sup>9</sup>. Although literature indicates several reports on BC production from glycerol sources using *Komagataeibacter* spp. (Table 1), to date BC production from unrefined crude glycerol have not been reported using *K. intermedius* strains. In comparison to glucose cultivations, *K. intermedius* ENS15 completely utilized the carbon source, synthesizing 1.3 ± 0.0 g/L BC (yield, 0.7 mg/g<sub>substrate</sub>; BC characterization, Fig. S6 and Table S5). Subsequently, the strain's ability to utilize crude glycerol, a by-product from biodiesel production process, and synthesize BC was investigated. Replacing pure glycerol with crude glycerol as the carbon source resulted in a 31% drop in BC production titer (Table 1). Crude glycerol, an unrefined glycerol source, obtained from biodiesel production plants often contains impurities such as soaps, fatty acids, ash and other organic compounds, negatively influencing bacterial metabolic processes<sup>9,52</sup>. Thus, the drop in BC production from *K. intermedius* ENS15 could be attributed towards the impurities present in the supplemented substrate. Alternatively, such could be a strain related feature. Although Kose et al. (2013) have reported a BC titer of 3.4 g/L using *Gluconacetobacter intermedius* NEDO-01 strain, in the study a pre-treated crude glycerol fraction was employed for BC biogenesis to eliminate the effects caused

Bacterial strains	Medium supplements <sup>a</sup>	Production titer (g/L)	References
<b>Glucose</b>			
<i>Komagataeibacter</i> sp. W1	–	1.2	41
<i>K. rhaeticus</i> AF1	2% Ethanol	4.1	3
<i>K. rhaeticus</i> PG2	–	~ 4	42
<i>K. rhaeticus</i> ENS9a	–	1.4	9
<i>K. rhaeticus</i> ENS9b	–	0.6	10
<i>G. xylinus</i>	–	1.3	43
<i>K. xylinus</i> B-12068	3% Eth	2.2	44
<i>G. xylinus</i> CGMCC 2955	–	~ 3.3	45
<i>K. hansenii</i> JR-02	2% Eth	~ 1.2	46
<i>K. hansenii</i> B22	–	1.6	7
<i>K. intermedius</i> FST213-1	–	1.2	40
<i>K. intermedius</i> JF2	–	1.2	39
<i>K. intermedius</i> ENS15	–	1.1 ± 0.1	This study
<b>Pure glycerol</b>			
<i>Komagataeibacter</i> sp. W1	–	1.2	41
<i>K. rhaeticus</i> PG2	–	6.9	42
<i>K. rhaeticus</i> ENS9a	–	2.6	9
<i>K. rhaeticus</i> ENS9b	–	3.0	10
<i>Komagataeibacter</i> sp. nov. CGMCC 17276	30 g/L CSL, 0.3% AA, 1.5% Eth	4.5	38
<i>K. xylinus</i> B-12068	3% Eth	23.2	
<i>K. hansenii</i> JR-02	–	~ 2.4	46
<i>K. intermedius</i> ENS15	–	1.3 ± 0.0	This study
<b>Crude glycerol</b>			
<i>G. sacchari</i>	–	0.1	47
<i>Komagataeibacter</i> sp. nov. CGMCC 17276 <sup>b</sup>	30 g/L CSL, 0.3% AA, 1.5% Eth	6.0	38
<i>A. xylinum</i>	–	~ 1.5	48
<i>G. xylinus</i> KCCM 41431 <sup>c</sup>	–	7.0	49
<i>G. xylinus</i>	–	~ 1.5	50
<i>K. intermedius</i> ENS15	–	0.8 ± 0.1	This study
<b>Xylose</b>			
<i>A. hansenii</i> ATCC 10821	–	40 mg/L	51
<i>A. xylinus</i> IFO 15606	–	100 mg/L	51
<i>A. pasteurianus</i> IFO 14814	–	70 mg/L	51
<i>K. intermedius</i> ENS15	–	400 ± 0.0 mg/L	This study

**Table 1.** BC production metrics from *Komagataeibacter* spp. grown statically in growth medium containing glucose, pure glycerol, crude glycerol, and xylose reported in literature. CSL corn steep liquor, AA acetic acid, Eth ethanol. <sup>a</sup>Additives to HS medium (5 g/L yeast extract, 5 g/L peptone) reported in respective literature, <sup>b</sup>BC production using optimized cultivation conditions, <sup>c</sup>HS medium concentration 9 g/L YE, 9 g/L Pep.

by the inhibitory compounds<sup>53</sup>. The study by Carreira et al. (2011) reports crude glycerol toxicity on *G. sacchari*, wherein the toxicity effects were elevated by using a 25–50 fold diluted substrate resulting in a BC titer of 0.1 g/L<sup>47</sup>.

Attempts in employing lignocellulosic biomass as the feedstock for BC production has been hindered by inefficient xylose utilization capabilities of *Komagataeibacter* spp. Early studies on BC biogenesis from xylose was reported by Ishihara et al. (2002), wherein minimal BC titers (ranging from 0 to 100 mg/L) were reported from *K. xylinus* and *K. hansenii* strains (Table 1)<sup>51</sup>. The authors reported that the restrictions were slightly elevated in culture medium containing xylose/xylulose mix resulting in a BC titer of 3 g/L. We identified that *K. intermedius* ENS15 grown in HS medium containing 20 g/L xylose synthesized BC, by far the highest, titer of 400 mg/L. The yeast extract and tryptone present in the HS medium have been reported to influence the BC production among *Komagataeibacter* spp.<sup>9,10</sup>. Nevertheless, the low background titers (0.01 ± 0.0 g/L) obtained from HS medium devoid of xylose, encouraged to identify the genes involved in xylose transport and metabolism in *K. intermedius* ENS15. The genome contained genes encoding proteins involved in xylose transport (xylose-proton symporter, genomic position 2,339,051:2,340,538 bp and xylose transporter, genomic position 2,049,024:2,050,391 bp), xylulose kinase (catalyzing the conversion of xylulose to xylulose 5-phosphate, an intermediate in pentose phosphate pathway, genomic positions 2,051,664:2,053,133 bp and 2,673,868:2,675,346 bp) and ribulose-phosphate 3-epimerase. Although blastP searches with amino acid annotations against *E. coli* xylose isomerase did not provide any significant hits, manual searches around the genes encoding xylose transport and xylulose metabolism proteins were conducted. At 9 Kbp upstream of xylulose kinase (2,673,868:2,675,346 bp) a putative gene, annotated by Prokka as a hypothetical protein, was identified to contain domain architectures affiliated to sugar



**Figure 3.** Analysis of variable Ori in *K. intermedius* ENS15. The replication origins pBBR1, p15A, pUC, pSC101, pBR322/ROP and pRO1600/ColE1 are present in pSEVA331, pSEVA361, pSEVA381, pSEVA371, pSEVA391 and pSEVA341, respectively. The presented data includes the mean values and standard deviations from duplicate cultivations in HS-glucose agar containing 340 µg/ml cm.

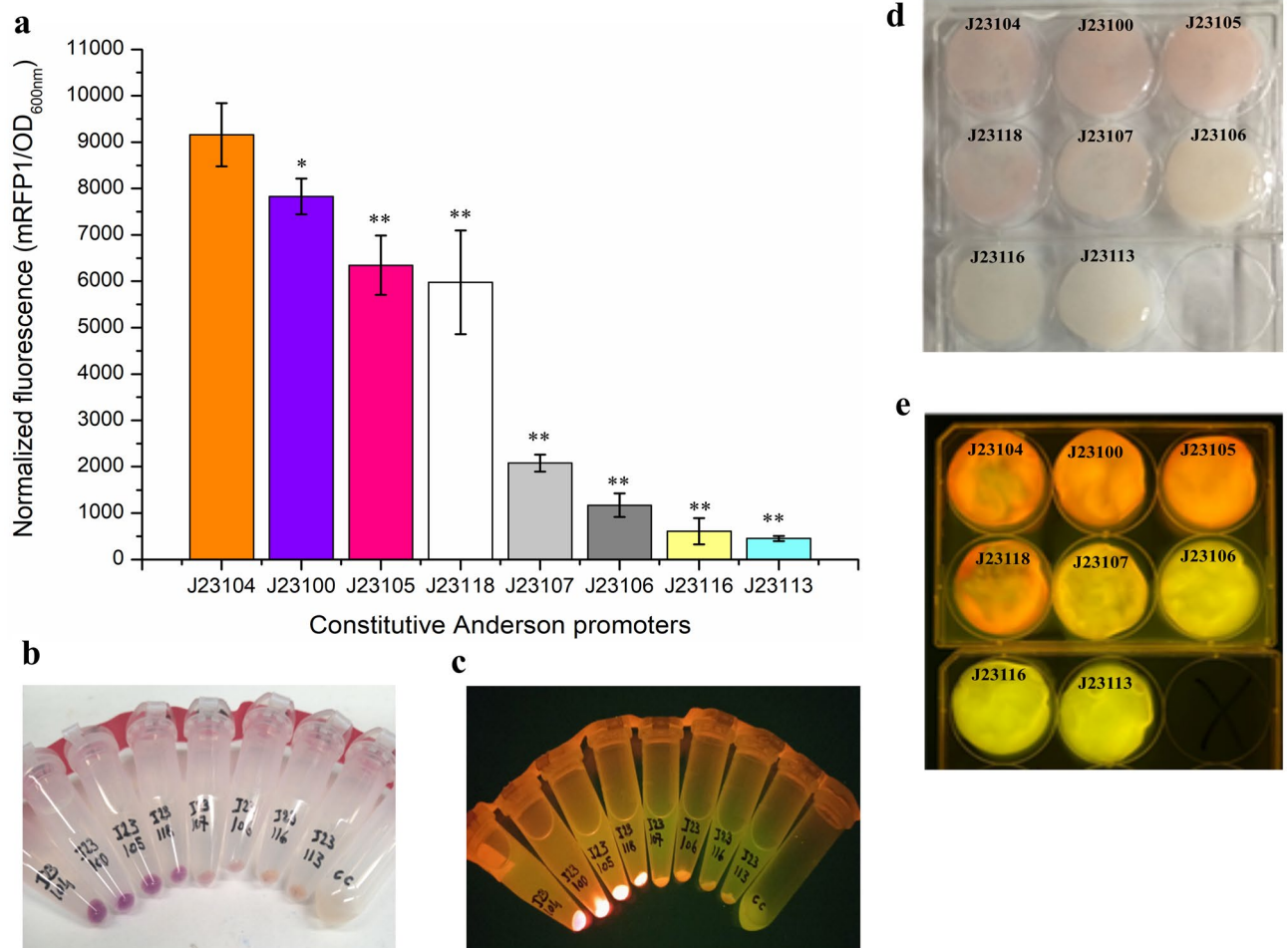
phosphate isomerase/epimerase and xylose isomerase-like superfamily. However, as these domain architectures are shared among several proteins, in depth experimental investigations are required to confirm the hypothesis.

**Plasmid transformation in *K. intermedius* ENS15.** Plasmid transformation and characterization of genetic engineering toolsets have been studied mostly in *K. xylinus*, *K. hansenii* and *K. rhaeticus* strains<sup>5,6</sup>. Among the plasmids tested, pSEVA321 (RK2 Ori), pSEVA331 (pBBR1 Ori), pSEVA351 (RFS1010 Ori), pBla-Vhb-122 (pBBR1 Ori) and pBAV1C (RepA, modified pWV01 Ori) have been reported to replicate in *Komagataeibacter* spp.<sup>4,12</sup>. Here, the genetic tractability of *K. intermedius* ENS15 was studied using SEVA plasmids. Based on the previous studies, pSEVA331 (containing pBBR1 Ori and CmR) was selected as the positive control. SEVA plasmids, pSEVA 211, pSEVA 241, pSEVA 261, pSEVA 271, pSEVA 281 and pSEVA 291, containing KanR and replication origins that have not yet tested in *Komagataeibacter* spp. were chosen in this study (Table S1). In presence of chloramphenicol, wild type *K. intermedius* ENS15 cells was able to grow on HS-glucose agar plates containing 200 µg/ml of the antibiotic and was completely inhibited at 300 µg/ml (Fig. S7A). Since growth inhibition was not observed even at high kanamycin concentrations (700 µg/ml; Fig. S7B), the KanR gene was replaced with CmR from pSEVA331. We identified that *K. intermedius* ENS15 can be transformed via electroporation. Among the tested SEVA plasmids, pSEVA331 (pBBR1 Ori) and pSEVA361 (p15A Ori) showed replication in *K. intermedius* ENS15 (Fig. 3).

**Characterization of constitutive and inducible promoters.** Constitutive synthetic *E. coli* promoters from Anderson collection have been characterized in *K. rhaeticus* iGEM, *K. hansenii* ATCC 53582 and *K. xylinus* ATCC 700178<sup>5,6</sup>. Constitutive promoters, J23100, J23106, J23111, J23113, J23116, J23107 and J23118, that have not yet characterized in *Komagataeibacter* spp. were selected in this study. mRFP1 was chosen as reporter to quantify the promoter strength. Based on the data reported by Teh et al.<sup>6</sup> and Florea et al.<sup>4,5</sup> J23104 and J23105 promoters reported to confer strong and weak strengths in *Komagataeibacter* spp., were included as positive and negative controls, respectively. For successful propagation in *K. intermedius* ENS15, regulatory elements (promoters, ribosome binding site, mRFP1 and the double terminator) were cloned into pSEVA331 vector. Similar to the reports from *K. rhaeticus* iGEM and *K. hansenii* ATCC 53582, J23104 promoter was identified as the strongest constitutive promoter in *K. intermedius* ENS15 (Fig. 4a). Surprisingly J23105, a weak promoter reported in *K. rhaeticus* iGEM, *K. hansenii* ATCC 53582 and *K. xylinus* ATCC 700178 strains, was found to confer strong fluorescence signal in *K. intermedius* ENS15. The studied promoters showed a similar trend in the mRFP1 signal in both *Escherichia coli* XL1 and *K. intermedius* ENS15 strains with J23104, J23100, J23105, J23118 promoters as the strong promoters (Fig. 4 and Fig. S8). However, a 2-sample t test with J23104 indicated that, except for J23100, the remaining promoters demonstrated a statistically significant drop in the fluorescence signal.

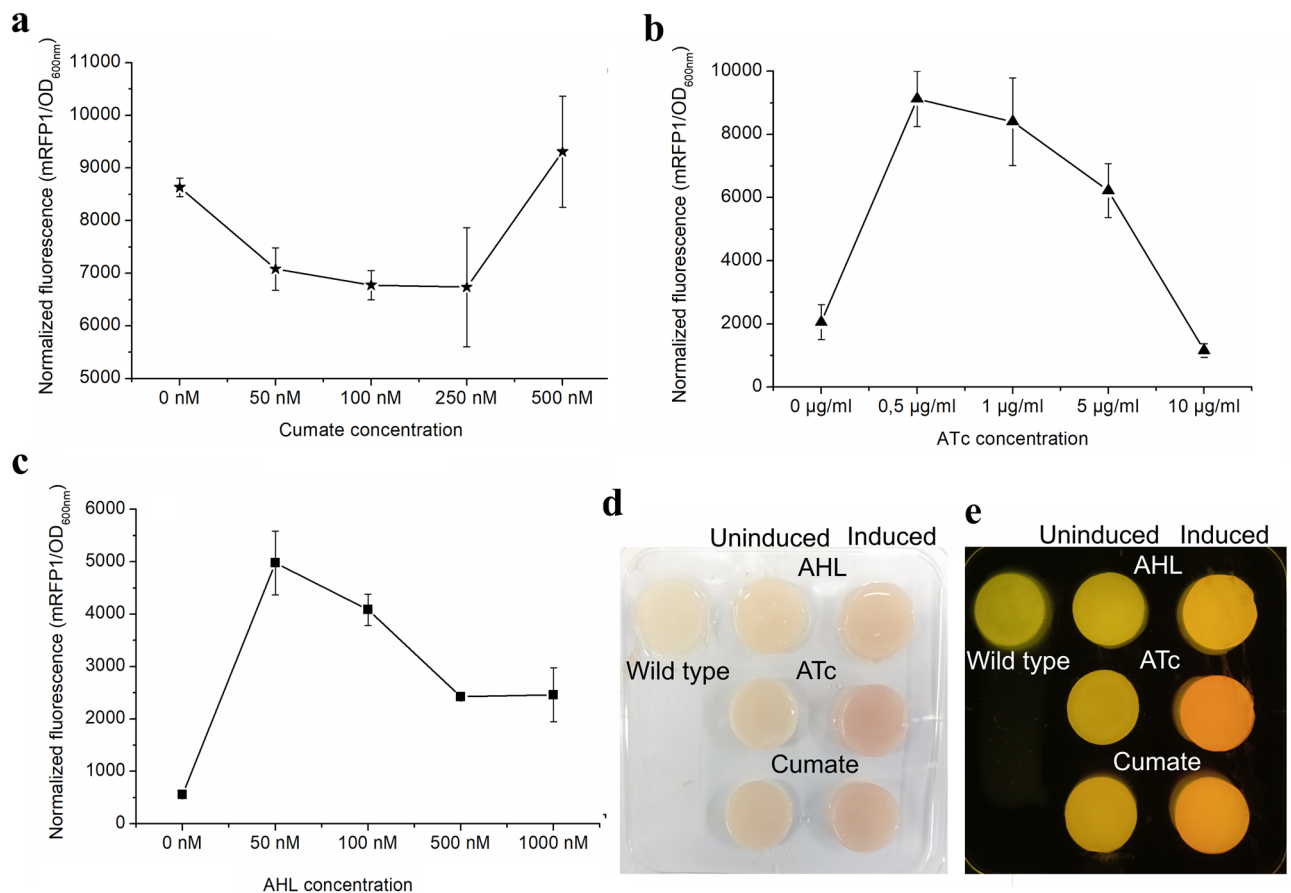
AHL-, ATc-, IPTG- and arabinose-inducible systems, under the control of pLux, pTet, pLac and pBAD promoters, have been previously characterized in *Komagataeibacter* spp.<sup>5,6</sup>. In the studies, the pLux promoters demonstrated highest mRFP1 expression and less leakiness than the tested counterparts. Due to the lack of lac repressors in *Komagataeibacter* genomes, pLac promoter was employed as a constitutive promoter by Teh et al.<sup>6</sup>. Furthermore, in the study, constructs carrying arabinose-inducible pBAD promoter was studied in *Komagataeibacter* spp. In comparison to pLux promoter, the authors identified that the arabinose-inducible construct





**Figure 4.** Characterization of constitutive Anderson promoters in *K. intermedius* ENS15. **(a)** strengths of Anderson promoters as measured by normalized fluorescence (mRFP1 fluorescence/OD<sub>600nm</sub>). The promoter strengths were quantified by culturing the recombinant *K. intermedius* ENS15 strains in HS-glucose medium containing chloramphenicol and cellulase for 3 days at 30 °C. The cells were pelleted by centrifugation, washed and resuspended in 1X PBS. The data points represent the mean experimental results and standard deviations from triplicate experimental and technical repeats (n = 6). Statistical significance of the drop in the fluorescence intensity in relation to the data from J23104 was analysed using Two sample t-test in Minitab 19. p < 0.05 and p > 0.05 are represented by \* and \*\*, respectively. **(b,c)** Fluorescence from recombinant *K. intermedius* ENS15 cells under normal and blue light, respectively. A control cell (CC) which is the wildtype *K. intermedius* ENS15 is placed for visual comparison in subfigures **(b,c)**. **(d,e)** BC sheets produced by recombinant *K. intermedius* ENS15 strain under normal and blue light, respectively. The images were cropped, and brightness and contrast were adjusted to improve clarity.

imparted low mRFP1 expression and identified that the presence of arabinose in the growth medium negatively affected nanocellulose production. Contemplating the literature reports, we investigated mRFP1 expression in *K. intermedius* ENS15 using previously characterized AHL (pLux)- and ATc (pTet)-inducible systems, and a Cumate-inducible expression system that has never been characterized in *Komagataeibacter* spp. (Fig. 5). The Cumate switch consisting of a *cymR* repressor gene and *CuO* repressor-binding operator under strong constitutive promoters have been employed for tight expression of recombinant genes in *E. coli* strains and *Bacillus* spp.<sup>33,54</sup>. However, we noted that the Cumate-inducible system was quite leaky in *K. intermedius* ENS15 (Fig. 5a) and the inducer negatively impacted the cell growth (Table S6). Such effect on the cell growth was not observed in *K. intermedius* ENS15 cells harboring AHL- and ATc-inducible constructs, except in the highest ATc concentration (10 µg/ml) which could be attributed towards the tetracycline toxicity (Table S6). In the case of mRFP1 induction using AHL- and ATc-inducible systems in *K. intermedius* ENS15, contradictory results to that published for *K. rhaeticus* iGEM strain was observed. Although AHL-inducible construct imparted low leakiness, a higher induction was observed using ATc-inducible construct with an optimal inducer concentration of 0.5 µg/ml (Fig. 5b). Furthermore, robust induction of mRFP1 expression while encased in the pellicle indicates that, despite entrapped within nanocellulose, the *K. intermedius* ENS15 cells can effectively receive signals from their environment enabling its applicability in applications requiring long-term cell survival and tolerance to toxic compounds (Fig. 5d,e).



**Figure 5.** Characterization of inducible constructs in *K. intermedius* ENS15. (a–c) mRFP1 expression using AHL-, ATc- and Cumate-inducible constructs as measured by normalized fluorescence (mRFP1 fluorescence/OD<sub>600nm</sub>), respectively. The promoter strengths were quantified by culturing the recombinant *K. intermedius* ENS15 strains in HS-glucose medium containing chloramphenicol, cellulase and varying inducer concentrations for 3 days at 30 °C. The cells were pelleted by centrifugation, washed and resuspended in 1X PBS. The data points represent the mean experimental results and standard deviations from triplicate experimental and technical repeats (n = 6). (d,e) BC sheets produced by recombinant *K. intermedius* ENS15 strain under normal and blue light, respectively. For induction within the BC pellicle, single colonies of wild type and recombinant *K. intermedius* ENS15 cells harbouring AHL-, ATc- and Cumate-inducible constructs were precultivated in HS-glucose medium containing 32 µg/mL CmR. The precultivated cells were inoculated, in duplicates, in to 6-well culture plates containing 10 ml of HS-glucose medium containing 32 mg/L CmR and optimal inducer concentration (AHL, 50 nM; ATc, 500 ng/mL; Cumate, 500 nM) and statically grown at 30 °C for 10 days. The images were cropped, and brightness and contrast were adjusted to improve clarity.

## Conclusions

A novel nanocellulose producing bacterium, affiliated to *K. intermedius* LMG 18909<sup>T</sup>, was isolated from Kombucha SCOBY. The *bcs* operons and accessory genes involved in BC biogenesis and predicted putative genes involved in carbohydrate metabolism was supported by the experimental observation that the isolate metabolized glucose, pure and crude glycerol for biomass formation and BC production. Although BC synthesized from xylose is the highest titer reported from *Komagataeibacter* spp., modest production metrics attributes to a non-efficient route undertaken by the isolate for pentose conversion to gluconeogenesis pathway. The results on successful propagation of SEVA plasmids carrying varying origin of replications and recombinant mRFP1 expression via constitutive and inducible promoter systems indicate that *K. intermedius* strains are genetically tractable. Coupling the obtained empirical data hypothesizes the possibilities of overexpressed native/heterologous genetic systems for improvements in pentose bioconversion to BC. Future studies will involve metabolic engineering attempts using the optimal expression vectors to engineer *K. intermedius* ENS15 strain for efficient xylose bioconversion for BC production.

## Data availability

All data generated or analysed during this study are included in this published article and in the accompanying supplementary file. The 16S rRNA gene sequence can be found in the NCBI GenBank database under the accession number MT094082. The genome assembly can be found in the NCBI GenBank under the accession

number GCA\_021555195.1. The protein sequences mentioned in this study can be accessed through the protein FASTA (.fna) file within the GenBank source database. *K. intermedius* ENS15 strain is available from the corresponding author upon request.

Received: 13 September 2022; Accepted: 18 November 2022

Published online: 28 November 2022

## References

- Römmling, U. & Galperin, M. Y. Bacterial cellulose biosynthesis: Diversity of operons, subunits, products and functions. *Trends Microbiol.* **23**, 545–557 (2015).
- Trache, D. *et al.* Nanocellulose: From fundamentals to advanced applications. *Front. Chem.* <https://doi.org/10.3389/fchem.2020.00392> (2020).
- Machado, R. T. A. *et al.* Komagataeibacter rhaeticus as an alternative bacteria for cellulose production. *Carbohydr. Polym.* **152**, 841–849 (2016).
- Florea, M., Reeve, B., Abbott, J., Freemont, P. S. & Ellis, T. Genome sequence and plasmid transformation of the model high-yield bacterial cellulose producer *Gluconacetobacter hansenii* ATCC 53582. *Sci. Rep.* **6**, 23635 (2016).
- Florea, M., Hagemann, H., Santosa, G., Abbott, J. & Micklem, C. N. Engineering control of bacterial cellulose production using a genetic toolkit and a new cellulose-producing strain. *Proc. Natl. Acad. Sci. U. S. A.* <https://doi.org/10.1073/pnas.1522985113> (2016).
- Teh, M. Y. *et al.* An expanded synthetic biology toolkit for gene expression control in acetobacteraceae. *ACS Synth. Biol.* **8**, 708–723 (2019).
- Semjonovs, P. *et al.* Cellulose synthesis by *Komagataeibacter rhaeticus* strain P 1463 isolated from Kombucha. *Appl. Microbiol. Biotechnol.* **101**, 1003–1012 (2017).
- Marić, L., Cleenwerck, I., Accetto, T., Vandamme, P. & Trček, J. Description of *komagataeibacter melaceti* sp. nov. and *komagataeibacter melomenus* sp. nov. isolated from apple cider vinegar. *Microorganisms* **8**, 1–15 (2020).
- Mangayil, R. *et al.* Characterization of a novel bacterial cellulose producer for the production of eco-friendly piezoelectric-responsive films from a minimal medium containing waste carbon. *Cellulose* **28**, 671–689 (2021).
- Cannazza, P. *et al.* Characterization of *Komagataeibacter* isolate reveals new prospects in waste stream valorization for bacterial cellulose production. *Microorganisms* **9**(11), 2230. <https://doi.org/10.3390/microorganisms9112230> (2021).
- Wu, J. M. & Liu, R. H. Cost-effective production of bacterial cellulose in static cultures using distillery wastewater. *J. Biosci. Bioeng.* **115**, 284–290 (2013).
- Mangayil, R. *et al.* Engineering and characterization of bacterial nanocellulose films as low cost and flexible sensor material. *ACS Appl. Mater. Interfaces* **9**, 19048–19056 (2017).
- Altschul, S. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
- Kumar, S., Stecher, G., Li, M., Nknyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/22.22.4673> (1994).
- Saitou, N. & Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
- Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1980).
- Wingett, S. W. & Andrews, S. FastQ screen: A tool for multi-genome mapping and quality control. *F1000Research* **7**, 1338 (2018).
- Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
- Bankevich, A. *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
- Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
- Guizelini, D. *et al.* GFinisher: A new strategy to refine and finish bacterial genome assemblies. *Sci. Rep.* **6**, 34963 (2016).
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* **25**, 1043–1055 (2015).
- Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
- Meier-Kolthoff, J. P. & Göker, M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* **10**, 2182 (2019).
- Antipov, D. *et al.* plasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* **32**, 3380–3387 (2016).
- Rozov, R. *et al.* Recycler: An algorithm for detecting plasmids from de novo assembly graphs. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btw651> (2016).
- Okonechnikov, K., Golosova, O. & Fursov, M. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **28**, 1166–1167 (2012).
- Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
- Zuckerandl, E. & Pauling, L. Evolutionary divergence and convergence in proteins. In *Evolving Genes and Proteins* (eds Bryson, V. & Vogel, H. J. B. T. -E. G. P.) 97–166 (Elsevier, 1965). <https://doi.org/10.1016/B978-1-4832-2734-4.50017-6>.
- Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43**, D222–D226 (2015).
- Zdobnov, E. M. & Apweiler, R. InterProScan—An integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**, 847–848 (2001).
- Seo, S.-O. & Schmidt-Dannert, C. Development of a synthetic cumate-inducible gene expression system for *Bacillus*. *Appl. Microbiol. Biotechnol.* **103**, 303–313 (2019).
- Hunter, J. D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
- Cock, P. J. A. *et al.* Biopython: Freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **25**, 1422–1423 (2009).
- Ryngajłło, M., Kubiak, K., Jędrzejczak-Krzepkowska, M., Jacek, P. & Bielecki, S. Comparative genomics of the *Komagataeibacter* strains—Efficient bionanocellulose producers. *Microbiologyopen* **8**, e00731 (2019).
- Gullo, M., La China, S., Petroni, G., Di Gregorio, S. & Giudici, P. Exploring K2G30 genome: A high bacterial cellulose producing strain in glucose and mannitol based media. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2019.00058> (2019).
- Lu, T. *et al.* Characterization and optimization of production of bacterial cellulose from strain CGMCC 17276 based on whole-genome analysis. *Carbohydr. Polym.* **232**, 115788 (2020).

39. Fernández, J. *et al.* Microbial cellulose from a Komagataeibacter intermedius strain isolated from commercial wine vinegar. *J. Polym. Environ.* **27**, 956–967 (2019).
40. Lin, S.-P. *et al.* Isolation and identification of cellulose-producing strain Komagataeibacter intermedius from fermented fruit juice. *Carbohydr. Polym.* **151**, 827–833 (2016).
41. Wang, S.-S.S. *et al.* Insights into bacterial cellulose biosynthesis from different carbon sources and the associated biochemical transformation pathways in Komagataeibacter sp. W1. *Polymers* **10**, 963 (2018).
42. Thorat, M. N. & Dastager, S. G. High yield production of cellulose by a Komagataeibacter rhaeticus PG2 strain isolated from pomegranate as a new host. *RSC Adv.* **8**, 29797–29805 (2018).
43. Kuo, C.-H., Chen, J.-H., Liou, B.-K. & Lee, C.-K. Utilization of acetate buffer to improve bacterial cellulose production by Gluconacetobacter xylinus. *Food Hydrocoll.* **53**, 98–103 (2016).
44. Volova, T. G., Prudnikova, S. V., Sukovatyi, A. G. & Shishatskaya, E. I. Production and properties of bacterial cellulose by the strain Komagataeibacter xylinus B-12068. *Appl. Microbiol. Biotechnol.* **102**, 7417–7428 (2018).
45. Liu, M. *et al.* Complete genome analysis of Gluconacetobacter xylinus CGMCC 2955 for elucidating bacterial cellulose biosynthesis and metabolic regulation. *Sci. Rep.* **8**, 6266 (2018).
46. Li, J. *et al.* Production of high crystallinity type-I cellulose from Komagataeibacter hansenii JR-02 isolated from Kombucha tea. *Biotechnol. Appl. Biochem.* **66**, 108–118 (2019).
47. Carreira, P. *et al.* Utilization of residues from agro-forest industries in the production of high value bacterial cellulose. *Bioresour. Technol.* **102**, 7354–7360 (2011).
48. Wu, M. *et al.* Valorizing kitchen waste through bacterial cellulose production towards a more sustainable biorefinery. *Sci. Total Environ.* **695**, 133898 (2019).
49. Ho Jin, Y., Lee, T., Kim, J. R., Choi, Y.-E. & Park, C. Improved production of bacterial cellulose from waste glycerol through investigation of inhibitory effects of crude glycerol-derived compounds by Gluconacetobacter xylinus. *J. Ind. Eng. Chem.* **75**, 158–163 (2019).
50. Dikshit, P. K. & Kim, B. S. Bacterial cellulose production from biodiesel-derived crude glycerol, magnetic functionalization, and its application as carrier for lipase immobilization. *Int. J. Biol. Macromol.* <https://doi.org/10.1016/j.ijbiomac.2020.03.047> (2020).
51. Ishihara, M., Matsunaga, M., Hayashi, N. & Tisler, V. Utilization of d-xylose as carbon source for production of bacterial cellulose. *Enzyme Microb. Technol.* **31**, 986–991 (2002).
52. Mangayil, R., Karp, M. & Santala, V. Bioconversion of crude glycerol from biodiesel production to hydrogen. *Int. J. Hydrogen Energy* **37**, 12198–12204 (2012).
53. Kose, R., Sunagawa, N., Yoshida, M. & Tajima, K. One-step production of nanofibrillated bacterial cellulose (NFBC) from waste glycerol using Gluconacetobacter intermedius NEDO-01. *Cellulose* **20**, 2971–2979 (2013).
54. Choi, Y. J. *et al.* Novel, versatile, and tightly regulated expression system for Escherichia coli strains. *Appl. Environ. Microbiol.* **76**, 5058–5066 (2010).

## Acknowledgements

R.M., A.J.R., and D.R. thanks Academy of Finland (Project No's. 323214 and 346983), Kone Foundation (Project No. 201803224) and Fondazione Cariplo (2020–1070), respectively, for supporting this study. R.M. would like to thank Mr. Rohith Krishnan for providing the Kombucha SCOBY. Open access funding provided by Tampere University, Finland.

## Author contributions

R.M. and P.C. designed and planned the experiments. R.M., P.C., and C.M. conducted strain isolation, characterization, BC production and molecular biology studies. R.M., A.J.R., D.G. and P.L. performed genome assembly and bioinformatics. E.S. conducted the material characterization studies. R.M. and P.C. interpreted the data and wrote the original draft. All authors reviewed the manuscript. R.M. supervised the work.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-24735-z>.

**Correspondence** and requests for materials should be addressed to R.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022