

# UC Davis

## UC Davis Previously Published Works

### Title

Engineered plants provide a photosynthetic platform for the production of diverse human milk oligosaccharides.

### Permalink

<https://escholarship.org/uc/item/0m3378b9>

### Journal

Nature Food, 5(6)

### Authors

Barnum, Collin

Paviani, Bruna

Couture, Garret

et al.

### Publication Date

2024-06-01

### DOI

10.1038/s43016-024-00996-x

Peer reviewed

# Engineered plants provide a photosynthetic platform for the production of diverse human milk oligosaccharides

Received: 24 August 2023

Accepted: 13 May 2024

Published online: 13 June 2024

 Check for updates

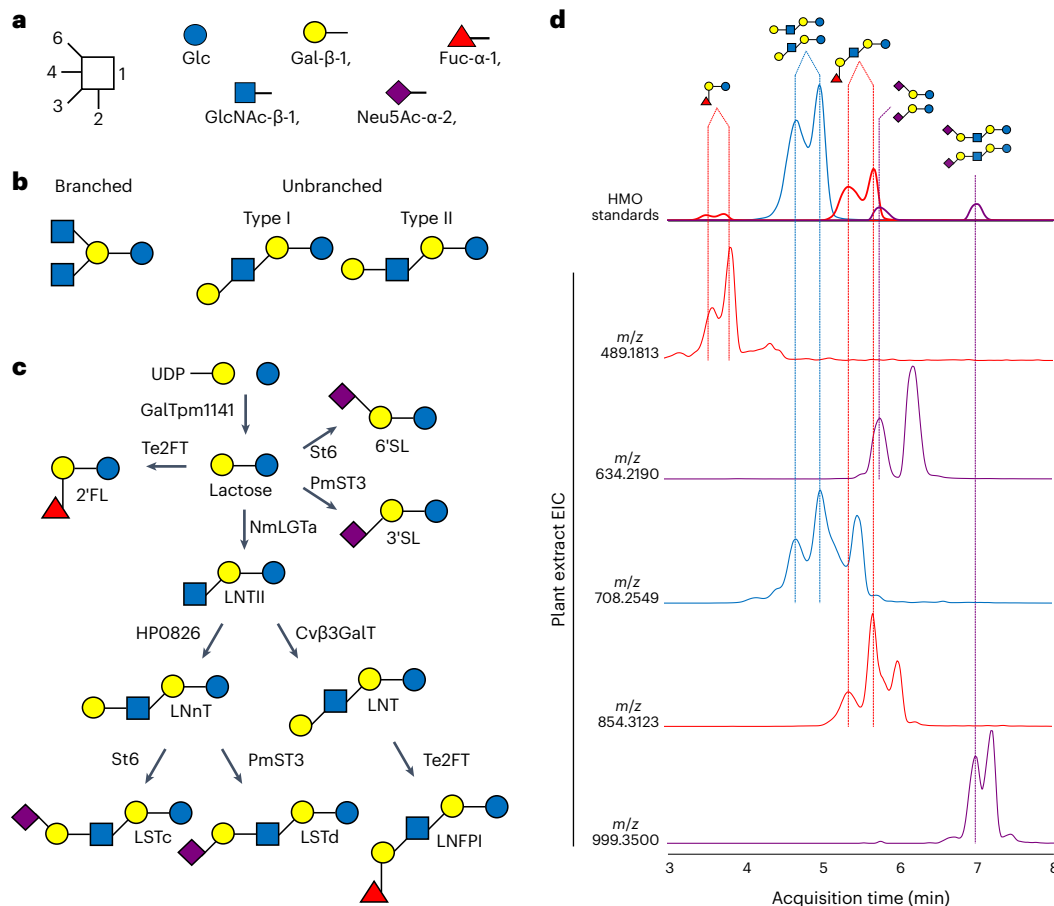
Collin R. Barnum<sup>1,2,3</sup>, Bruna Paviani<sup>4,5</sup>, Garret Couture<sup>5,6</sup>, Chad Masarweh<sup>4,5</sup>, Ye Chen<sup>5,6</sup>, Yu-Ping Huang<sup>4,5</sup>, Kasey Markel<sup>1,3,7</sup>, David A. Mills<sup>4,5,8</sup>, Carlito B. Lebrilla<sup>5,6</sup>, Daniela Barile<sup>4,5</sup>, Minliang Yang<sup>9</sup> & Patrick M. Shih<sup>1,3,7,10</sup> ✉

Human milk oligosaccharides (HMOs) are a diverse class of carbohydrates which support the health and development of infants. The vast health benefits of HMOs have made them a commercial target for microbial production; however, producing the approximately 200 structurally diverse HMOs at scale has proved difficult. Here we produce a diversity of HMOs by leveraging the robust carbohydrate anabolism of plants. This diversity includes high-value and complex HMOs, such as lacto-*N*-fucopentaose I. HMOs produced in transgenic plants provided strong bifidogenic properties, indicating their ability to serve as a prebiotic supplement with potential applications in adult and infant health. Technoeconomic analyses demonstrate that producing HMOs in plants provides a path to the large-scale production of specific HMOs at lower prices than microbial production platforms. Our work demonstrates the promise in leveraging plants for the low-cost and sustainable production of HMOs.

Human milk is a complete and comprehensive food evolved to nourish and protect infants. A key component to the distinct bioactive properties of human milk is the presence of a wide diversity of human milk oligosaccharides (HMOs) which are well documented in establishing the nascent gut microbiota of infants to prevent diseases and ensure healthy development<sup>1–4</sup>. While 75% of infants are supplemented with or exclusively fed infant formula in the first 6 months of life, current infant formulas are either devoid of HMOs or only contain one to two of the ~200 HMOs found in human milk, limiting the health outcomes of formula-fed infants<sup>3,5,6</sup>. In addition to their use for infant health, HMOs are being studied for their beneficial roles in adult health as a prebiotic to improve intestinal barrier function, lower gastrointestinal

inflammation and treat irritable bowel diseases<sup>7–11</sup>; however, the study of HMO benefits in adults has been limited to a small subset of HMOs. Currently, commercial HMO production relies on microbial fermentation but, to date, microbial fermentation is only able to commercially produce two to five simple HMOs of the ~200 HMOs found in human milk at a scale suitable to supplement food products<sup>6,12,13</sup>. While five simple HMOs constitute a large portion of HMO mass in human milk, diverse HMOs with a range of linkages and degrees of polymerization enable the growth of beneficial gut microbes which have preferences for specific HMOs<sup>14,15</sup>. Thus, there is a need to develop biological platforms to produce a wider diversity of HMOs found in human milk, enabling the supplementation of food products for both infants and adults.

<sup>1</sup>Department of Plant and Microbial Biology, University of California, Berkeley, Berkeley, CA, USA. <sup>2</sup>Department of Plant Biology, University of California, Davis, Davis, CA, USA. <sup>3</sup>Feedstocks Division, Joint Bioenergy Institute, Emeryville, CA, USA. <sup>4</sup>Department of Food Science and Technology, University of California, Davis, Davis, CA, USA. <sup>5</sup>Foods for Health Institute, University of California, Davis, Davis, CA, USA. <sup>6</sup>Department of Chemistry, University of California, Davis, Davis, CA, USA. <sup>7</sup>Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>8</sup>Department of Viticulture and Enology, University of California, Davis, Davis, CA, USA. <sup>9</sup>Department of Food, Bioprocessing and Nutrition Sciences, North Carolina State University, Raleigh, NC, USA. <sup>10</sup>Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA. ✉e-mail: [pmshih@berkeley.edu](mailto:pmshih@berkeley.edu)



**Fig. 1 | Production of all three HMO classes in planta.** **a**, HMOs are composed of Glc, Gal, Fuc, GlcNAc and/or Neu5Ac connected via Gal- $\beta$ -1,3/4, GlcNAc- $\beta$ -1,3/6, Fuc- $\alpha$ -1,2/3/4 or Neu5Ac- $\alpha$ -2,3/6 glycosidic linkages. **b**, HMOs can be divided into branched, unbranched, type I and/or type II HMOs. **c**, HMO biosynthetic pathways used in this study for HMO production in planta. **d**, Extracted ion chromatograms (EIC) showing the identification of 2'FL, 3'SL, 6'SL, LNFPI,

LSTa, LSTc and LNT/LNnT in extracts of individual plant leaves using LC-MS/MS (Q Exactive, Thermo Fisher Scientific). Red, blue and purple colouring denote fucosylated, neutral and acidic HMOs, respectively. Some further peaks are present due to in-source fragmentation of larger oligosaccharides with no available standards.

The combinatorial nature of glycosidic linkages, nucleotide sugar donors and oligosaccharide acceptor molecules enables the large diversity of HMOs found in human milk<sup>16</sup>. HMOs are composed of five distinct sugars—D-glucose (Glc), D-galactose (Gal), N-acetylglucosamine (GlcNAc), L-fucose (Fuc) and N-acetylneuraminic acid (Neu5Ac)—connected through various glycosidic linkages to generate a diverse range of molecular structures (Fig. 1a). HMO biosynthesis begins with the production of lactose which can be decorated with fucose or Neu5Ac to form a variety of trisaccharides and tetrasaccharides. Lactose can also be extended by glycosyltransferases through the addition of a GlcNAc- $\beta$ -1,3 or GlcNAc- $\beta$ -1,4 to form unbranched HMOs. Unbranched HMOs can be further extended by glycosyltransferases through the addition of a GlcNAc- $\beta$ -1,6 to produce branched HMOs (Fig. 1b). GlcNAc can undergo subsequent addition of Gal- $\beta$ -1,3 or Gal- $\beta$ -1,4 to form type I and type II HMOs, respectively (Fig. 1b). While HMOs can consist of all five monosaccharides, they are generally classified in three broad HMO groups on the basis of their composition: (1) neutral HMOs contain Glc, Gal and GlcNAc, (2) fucosylated HMOs contain a neutral core with one or more Fuc additions and (3) acidic HMOs contain a neutral core with one or more additions of Neu5Ac. Owing to the need for high amounts of nucleotide sugars and glycosylation potential in HMO biosynthesis, a suitable host must have robust sugar metabolism capable of managing the metabolic burden of HMO production.

Unlike many microbes used in commercial fermentation, plants have evolved to create a wide range of glycans which encompass a diversity of

nucleotide sugars from photosynthetically fixed CO<sub>2</sub>. As masters of sugar anabolism, plants are able to create vast amounts of complex oligosaccharides and polysaccharides<sup>17</sup>. This has led to commercial operations for the production of prebiotic oligosaccharides from plant biomass, such as  $\beta$ -glucan, xylooligosaccharides, inulin or soy oligosaccharides<sup>18–21</sup>. Many of these products can either be purified or directly consumed as a food, providing an easy means of ingestion. Additionally, plants can be grown in open fields, requiring minimal inputs, limiting the need for expensive substrates and axenic conditions<sup>22</sup>. The robust sugar metabolism of plants and ability to be grown at agricultural scales make plants an ideal platform for the large-scale production of HMOs.

Owing to the unique advantages of plants as a platform for carbohydrate production, we tested their ability to produce a range of HMOs using both transient and stable expression in *Nicotiana benthamiana*. Here, we report in planta production of neutral, fucosylated and acidic HMOs showcasing the intrinsic advantages of a plant-based production platform. Furthermore, we show that plant-produced HMOs provide selective growth of key bifidobacteria, indicating their potential prebiotic efficacy. Finally, we assess the economic viability of HMO production in planta compared to current microbial platforms.

## Results

### Production of all three HMO classes in planta

Production of HMOs requires the expression of glycosyltransferases capable of creating specific glycosidic linkages. While HMO

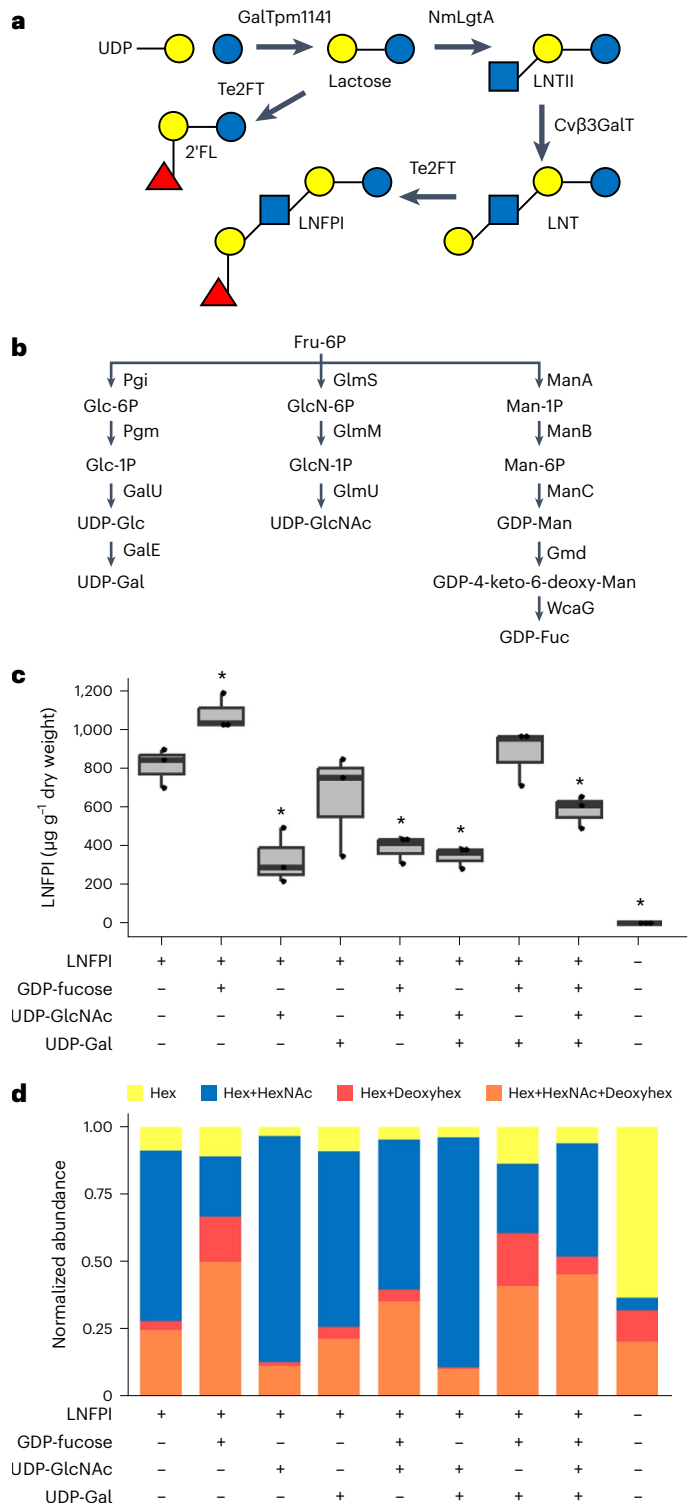
biosynthesis in humans takes place in the Golgi by means of a largely unknown pathway<sup>23</sup>, various microbial enzymes are capable of generating HMOs in the cytosol. To produce HMOs in plants, we localized bacterial HMO biosynthetic enzymes to the cytosol for the production of neutral, fucosylated and acidic HMOs (Fig. 1c). We tested these pathways through transient expression in *N. benthamiana*. Transient expression permits relatively high throughput screening of biosynthetic pathways in planta by injecting strains of *Agrobacterium tumefaciens* into plant leaves, allowing the *Agrobacterium* to insert genes for HMO biosynthetic genes into the plant cell<sup>24</sup>. Leaves transiently expressing HMO biosynthetic pathways were subjected to liquid-liquid extraction, C18 solid-phase extraction (SPE) and porous graphitic carbon (PGC) SPE before characterization by mass spectrometry (MS) (Supplementary Fig. 1).

Neutral HMOs function as the core scaffolds of other more complex HMOs (fucosylated and acidic); thus, we first targeted the type I and type II neutral HMO core structures, lacto-*N*-tetraose (LNT) and lacto-*N*-neotetraose (LNnT). Expression of a neutral HMO biosynthetic pathway using two  $\beta$ -1,4-galactosyltransferases (*GalTPM1141* (ref. 25), *Hp0826* (ref. 26)), one  $\beta$ -1,3-galactosyltransferase (*Cv $\beta$ 3GalT*<sup>27</sup>) and one  $\beta$ -1,3-*N*-acetylglucosaminyltransferase (*NmLgtA*<sup>28</sup>) resulted in the production of lactose and various neutral HMOs with degrees of polymerization ranging from three to seven. Notably, *N. benthamiana* transiently expressing this pathway produced the tetrasaccharides LNT (*m/z* 708.2559) and LNnT (*m/z* 708.2559), which represent principal type I and type II HMOs in human milk, respectively<sup>29</sup> (Fig. 1d and Supplementary Table 1). Additionally, we identified the production of larger neutral oligosaccharides with varying degrees of polymerization using tandem mass spectrometry (MS/MS) fragmentation to determine the number of hexose and HexNAc sugars (Supplementary Table 1). This included several neutral isomers of pentasaccharides and heptasaccharides (Supplementary Table 1). Our findings demonstrate that plants have the ability to produce previously inaccessible oligosaccharides with various degrees of polymerization which may expand HMO functional bioactivity.

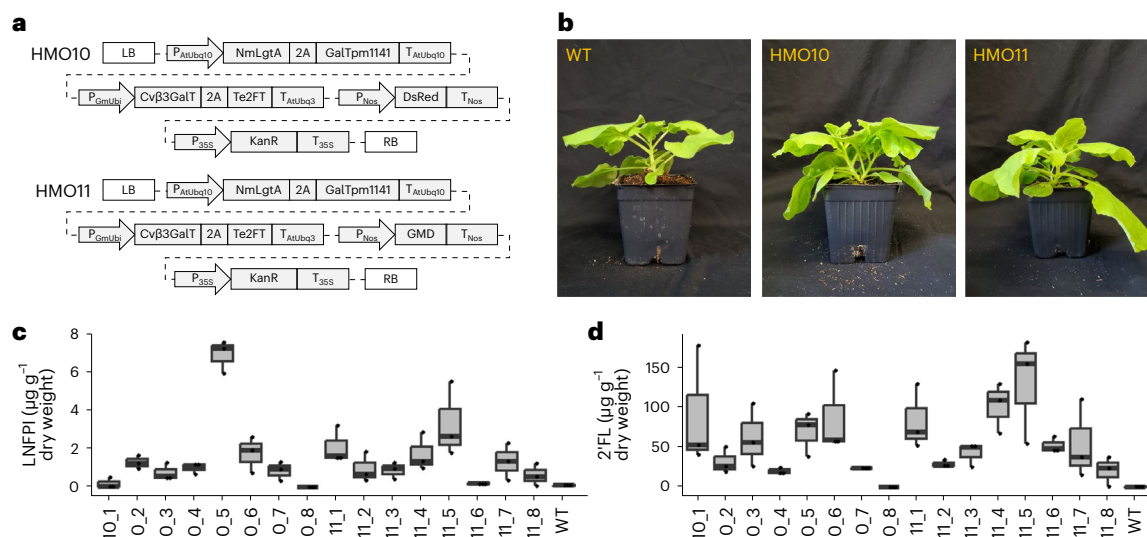
Following the success of generating type I and type II neutral HMOs, we examined the ability of plants to decorate neutral HMOs with fucose, as fucosylated HMOs are the most abundant class of HMO in human milk<sup>15</sup>. We transiently expressed an  $\alpha$ -1,2-fucosyltransferase (*Te2FT*<sup>30</sup>) alongside the neutral HMO biosynthetic pathway to produce the most abundant fucosylated HMOs in human milk: 2'-fucosyllactose (2'FL) (*m/z* 489.1819) and lacto-*N*-fucopentaose I (LNFPI) (*m/z* 854.3136) (Fig. 1d and Supplementary Table 1). Additionally, several fucosylated hexasaccharide isomers were identified by *m/z* and MS/MS fragmentation (Supplementary Table 1). While the structure of each isomer could not be determined, each is composed of four hexoses, one HexNAc and

one deoxyhexose, indicating that either LNFPI can be further decorated with hexose sugars or pentasaccharide neutral HMOs can be decorated with additional fucose.

While neutral and fucosylated HMOs represent most HMOs in human milk, acidic HMOs constitute the last main class found in mammalian milks which provide unique bioactivities as a result of the presence of *N*-acetylneuraminic acid<sup>31</sup>. Plants do not natively produce the donor molecule for production of acidic HMOs, CMP-Neu5Ac. To produce acidic HMOs, we simultaneously expressed the neutral HMO biosynthetic pathway, sialyltransferases and a mammalian pathway for the production of CMP-Neu5Ac (Supplementary Fig. 2)<sup>32</sup>. Expression of



**Fig. 2 | Manipulation of nucleotide sugar biosynthetic pathways modulates HMO profiles in planta.** **a**, HMO biosynthetic pathway expressed for the production of LNFPI. **b**, Nucleotide sugar biosynthetic pathways expressed with LNFPI pathway. **c**, Quantification of LNFPI production through expression of LNFPI biosynthetic pathway alongside combinatorially expressed nucleotide sugar biosynthetic pathways using an internal calibration curve obtained with an Agilent 6530 Accurate-Mass Q-ToF MS. The middle bar represents the median. Upper and lower whiskers correspond to the largest and smallest values within  $1.5 \times$  the interquartile range, respectively. Upper and lower hinges represent the third and first quartiles, respectively. Statistical analysis was conducted using a heteroscedastic two-tailed Student's *t*-test with the LNFPI pathway expressed alone used as the reference group. \* $P < 0.05$ . *P* values are: LNFPI + fucose, 0.030; LNFPI + GlcNAc, 0.012; LNFPI + fucose + GlcNAc, 0.01; LNFPI + GlcNAc + Gal, 0.01; LNFPI + fucose + GlcNAc + Gal, 0.043. A sample size of three leaves was used for each experiment. **d**, Effect of nucleotide sugar biosynthetic pathway overexpression on HMO profile produced using the LNFPI pathway. Values reflect normalized peak area. Hexose, HexNAc, deoxyhexose (Deoxyhex) determined using *m/z* and MS/MS fragmentation. We performed mass spectral analysis on an Agilent Q-ToF MS.



**Fig. 3 | Production of HMOs in stably transformed plants.** **a**, Constructs used in creation of stable lines containing biosynthetic enzymes for the production of LNFPI. **b**, Photos of 4-week-old transgenic *N. benthamiana*. **c**, Concentration of LNFPI produced in leaves of each stable line. **d**, Concentration of 2'FL produced in leaves of each stable line. For quantification, three leaves from each plant were analysed separately. Quantification of LNFPI and 2'FL obtained with a Thermo

Fisher Scientific Q Exactive Mass spectrometer. LB, left border; 2A, 2A peptide; RB, right border;  $P_{35S}$ , promoter;  $T_{35S}$ , terminator; WT, wild type. The middle bar represents the median. Upper and lower whiskers correspond to the largest and smallest values within  $1.5 \times$  the interquartile range, respectively. Upper and lower hinges represent the third and first quartiles, respectively. A sample size of three leaves was used for each experiment.

an  $\alpha$ -2,6-sialyltransferase (*St6*; ref. 33) alongside the neutral HMO biosynthetic pathway produced the acidic trisaccharide, 6'-sialyllactose (6'SL) ( $m/z$  634.2191), of type II acidic HMO, sialyllacto-*N*-neotetraose c (LSTc) ( $m/z$  999.3505) (Fig. 1d and Supplementary Table 1). Expression of an  $\alpha$ -2,3-sialyltransferase (*PmST3*; ref. 34) with the neutral HMO biosynthetic pathway produced a myriad of acidic HMOs, such as the acidic trisaccharide, 3'-sialyllactose (3'SL) ( $m/z$  634.2187) and the acidic pentasaccharide, LSTd ( $m/z$  999.3510) (Fig. 1d and Supplementary Table 1). In addition to making several LST isomers in vivo, six isomers of acidic hexasaccharides were identified using  $m/z$  and MS/MS fragmentation. Each isomer was composed of four hexoses, one HexNAc and one Neu5Ac (Supplementary Table 1). Together, these results show the ability of plants to produce all three classes of HMOs combinatorially or simultaneously (Supplementary Fig. 3), including type I and type II structures, marking the greatest diversity of HMOs made in a single heterologous organism.

### Optimized production of complex fucosylated HMOs

Microbial production platforms suffer from an inability to produce HMOs with higher degrees of polymerization at large scales, leaving many larger, more complex HMOs understudied. LNFPI is a fucosylated pentasaccharide that is the second most abundant fucosylated HMO. Despite its high abundance in breast milk, LNFPI has remained recalcitrant to fermentative production in microbes, limiting efforts to study its potential health benefits. Therefore, we sought to optimize the production of LNFPI in planta by overexpressing the requisite nucleotide sugar biosynthetic pathways. We transiently expressed the biosynthetic pathway for LNFPI (Fig. 2a) alongside pathways for the production of UDP-galactose, UDP-*N*-acetylglucosamine (UDP-GlcNAc) and GDP-fucose and quantified LNFPI production (Fig. 2b). Expression of the LNFPI pathway with the GDP-fucose pathway increased production of LNFPI by 32.9% ( $1,075.03 \mu\text{g g}^{-1}$  of dry weight) compared to the expression of the LNFPI pathway alone ( $808.91 \mu\text{g g}^{-1}$  of dry weight), indicating that GDP-fucose is limiting in *N. benthamiana* (Fig. 2c). Surprisingly, overexpression of the GDP-fucose pathway also resulted in the production of lactodifucotetraose (LDFT) ( $m/z$  635.2394) and lacto-*N*-difuco-hexaose I (LNFdHI) ( $m/z$  1,000.3720) (Supplementary Table 1) despite not expressing an  $\alpha$ -1,3- or  $\alpha$ -1,4-fucosyltransferase,

indicating the presence of native plant fucosyltransferases capable of glycosylating HMOs. Overexpression of all other nucleotide sugar pathway combinations resulted in similar or lower levels of LNFPI production compared to expression of the LNFPI pathway alone.

The overall profile of oligosaccharides produced was altered by overexpressing nucleotide sugar biosynthetic pathways (Fig. 2d). The number of hexose (Glc, Gal), HexNAc (GlcNAc) and deoxyhexose (Fuc) sugars in each oligosaccharide identified was determined through identification by means of  $m/z$  and MS/MS fragmentation and peak area was normalized using a LNFPI calibration curve. Overexpression of the UDP-GlcNAc and LNFPI pathways increased the relative amount of neutral oligosaccharides containing a hexose and HexNAc compared to expression of the LNFPI pathway alone (Fig. 2d). Overexpression of the GDP-fucose and LNFPI pathways resulted in a shift in the overall oligosaccharide composition, favouring the production of oligosaccharides containing at least one deoxyhexose, indicating an increase in the level of fucosylated oligosaccharides (Fig. 2d). These results demonstrate that tailoring the availability of nucleotide sugars enables control over the ratio of HMOs produced.

Because scaling HMO production in plants requires the growth of stably transformed plants, we developed transgenic lines of *N. benthamiana* expressing the LNFPI biosynthetic pathway. We generated two constructs for the constitutive production of 2'FL and LNFPI in transgenic *N. benthamiana* (Fig. 3a). HMO10 contains genes that encode four enzymes required to produce lactulose, 2'FL, LNTII, LNT and LNFPI connected by means of 2A peptides<sup>35</sup> to allow several coding sequences to be driven by a single constitutive promoter. To explore the effects of overexpressing portions of the GDP-fucose pathway, we also generated stable lines expressing HMO11, which contains a GDP-D-mannose-4,6-dehydratase (*Gmd*<sup>25</sup>) from the GDP-fucose pathway. *Gmd* transiently expressed alongside the neutral HMO pathway altered the HMO profile of plants in a similar way to expression of the full GDP-fucose pathway (Supplementary Fig. 4).

Transgenic T0 HMO-producing stable lines were assessed for fucosylated HMO yield. Most transgenic plants showed no drastic phenotypes compared to wild-type plants (Fig. 3b) and we conducted quantitative PCR with reverse transcription (RT-qPCR) analysis to confirm the expression transgenes (Supplementary Fig. 5). LNFPI and

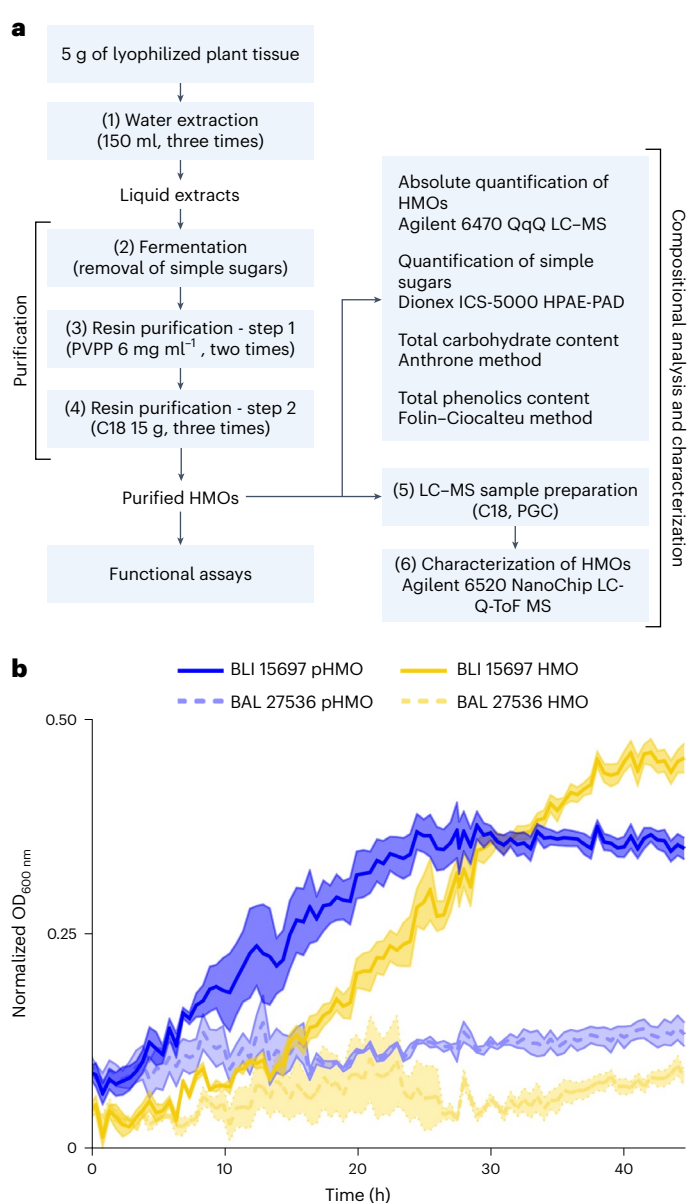
2'FL were detected in leaves of transgenic *N. benthamiana* expressing both HMO10 and HMO11. Highest producing LNFPI lines accumulated an average concentration of  $6.88 \mu\text{g g}^{-1}$  dry weight (Fig. 3c). Leaves from HMO11 no. 5 produced the highest concentration of 2'FL, reaching an average concentration of  $130.35 \mu\text{g g}^{-1}$  dry weight (Fig. 3d). The low abundance of LNFPI in stable lines compared to transient expression could indicate that the  $\beta$ -1,3-*N*-acetylglucosaminyltransferase and  $\beta$ -1,3-galactosyltransferase suffer from altered activity due to the presence of 2A peptides or lower expression. Together, these results show the ability to produce and optimize a diversity of HMOs from photosynthetically fixed  $\text{CO}_2$ , laying the foundation for future efforts to create high-HMO-yielding transgenic plants for commercial HMO production.

### Purification and functional characterization of HMOs from plants

Mixtures of prebiotic sugars can have varying effects on the enrichment of beneficial gut microbes<sup>36</sup>. Therefore, we sought to assess the bifidogenic activity of extracts from HMO-producing plants; however, crude plant extracts can contain chemicals that interfere with bacterial growth assays, such as simple sugars (glucose, fructose and sucrose) and antimicrobial specialized metabolites. Therefore, we developed a method to extract and purify HMOs from *N. benthamiana* transiently expressing the biosynthetic pathways for LNFPI (Fig. 2a) and GDP-fucose (Fig. 2b) using an optimized extraction and purification process. Briefly, we performed a water extraction, yeast fermentation to remove simple sugars and a two-step resin adsorption with polyvinylpyrrolidone (PVPP) and C18 SPE. This resulted in an HMO-rich extract that contained negligible amounts of simple sugars and phenolic compounds (Supplementary Table 3). The HMO extract contained target fucosylated HMOs (Supplementary Fig. 6), including LNFPI, 2'FL and LNDFHI (Supplementary Table 4). The extract also contained a variety of additional oligosaccharides without assigned structures, which were composed of combinations of hexose, deoxyhexose and HexNAc sugars (Supplementary Fig. 6). These represent potentially non-natural oligosaccharide structures which could provide potential health benefits. Overall, these results show the ability to isolate HMOs from plants, improving their promise as an HMO production platform.

To assess the bifidogenic activity of plant-produced HMOs, we conducted growth assays to compare the effects of plant-produced HMOs to HMOs derived from human milk. We chose to assess the effects of plant-derived HMOs on *Bifidobacterium longum* subsp. *infantis* ATCC 15697 (BLI 15697) as it is a known HMO consumer<sup>37</sup>. We also included *Bifidobacterium animalis* subsp. *lactis* ATCC 27536 (BLAC 27536) as a negative control which does not consume HMOs<sup>38</sup> but will grow on simple sugars that could be present in plant extracts or human milk. BLI 15697 grown in media containing plant-derived HMOs showed increases in optical density  $\text{OD}_{600\text{nm}}$  similar to that of BLI 15697 in HMO isolated from human milk, demonstrating that plant-produced HMOs possess the same selective bifidogenic activity as HMOs isolated from human milk (Fig. 4b). As expected, BLAC 27536 showed no growth in either plant-derived HMOs or HMO isolated from human milk, indicating that both extracts contained a minimal amount of simple sugars present (Fig. 4b). Together, these results demonstrate the ability of purified, plant-produced HMOs to mimic the bifidogenic activity of HMOs produced in humans.

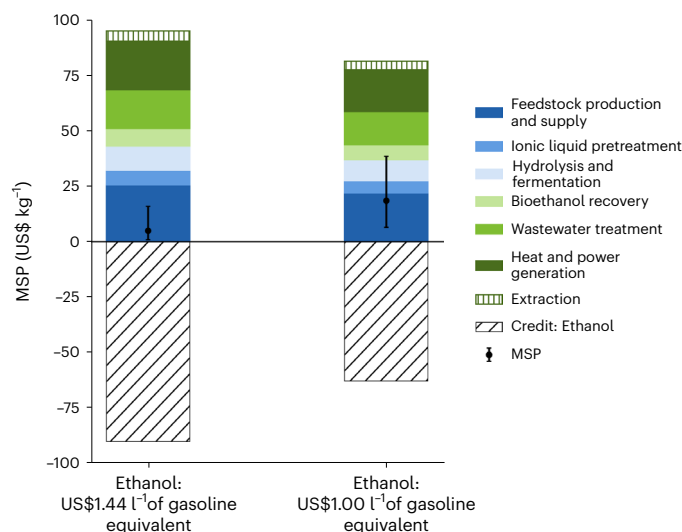
**Economic viability of plant-based HMO production.** Plant-based HMO production can be commercially viable if it demonstrates cost-competitive or cost-advantage with current state-of-the-art production routes. To assess the economic viability of HMO production in a commercially relevant crop, we developed process models and compared the cost of HMO production in plants and microbes. To do this, we performed techno-economic analysis (TEA) of the theoretical production of LNFPI. In the plant system, we adopted typical cellulosic



**Fig. 4 | Optimized purification protocol developed for functional analysis of plant-produced HMOs. a**, Workflow of extraction, purification and characterization of HMOs from *N. benthamiana* leaves. **b**, Growth curves of HMO-consuming (BLI 15697) and control (BAL 27536) strains in media supplemented with HMO isolated from breast milk (HMO) or HMOs isolated from plants (pHMO). Error bars represent standard deviation.

biorefinery design using biomass from sorghum to coproduce HMOs along with biofuel because coproducing value-added bioproducts in biorefineries is a promising approach to maximize the use of biomass and hence improve the economics of biorefineries<sup>39,40</sup>. We assume that biomass sorghum can accumulate 0.31% dry weight of LNFPI in the entire biomass, as this was our highest yield following purification of LNFPI (Supplementary Table 4). We also developed process models and conducted TEA for the HMOs in *Escherichia coli* using the established processes and the highest reported yields of LNFPI<sup>41</sup> from peer-reviewed papers at the time of conducting this TEA. The comparison between plant and microbial systems to produce the same product aims to provide in-depth understanding of the cost-benefits of the systems.

TEA results indicate that producing LNFPI from the plant system is economically favourable compared to the microbial system



**Fig. 5 | Plant-based platform improves the economics of producing the HMO, LNFPI.** Estimated MSP of LNFPI produced using biomass sorghum as a model production platform in two bioethanol price scenarios. Error bands represent final values calculated with  $\pm 20\%$  of input parameters.

(Supplementary Fig. 7). In the plant system, the minimum selling prices (MSPs) of LNFPI are US\$4.9 kg<sup>-1</sup> when selling ethanol at the cellulosic ethanol selling price and US\$18.4 kg<sup>-1</sup> when ethanol is sold at the target fuel price, respectively (Fig. 5). However, microbial-based LNFPI results in MSP of US\$45.0 kg<sup>-1</sup> with downstream recovery and purification being the largest cost contributor, followed by the cost of glucose (Supplementary Fig. 7b). The high cost obtained in the microbial system is largely due to its extremely low highest reported yields (0.48%) and recovery rate after bioconversion (62%). These results indicate that when microbial hosts are unable to produce HMOs at a comparable yield, plants may be a cost-effective bioprocess for high-value products. In addition to relative cost advantages of plant systems, using biomass as the feedstock to coproduce high-value compounds and biofuel offers considerable environmental benefits because biomass can absorb CO<sub>2</sub> from the atmosphere during its growth<sup>42</sup>. Although current HMO yields in stable lines of the model plant *N. benthamiana* (Fig. 3) are below the yields in transiently expressing tissue, high yields could be achieved by optimizing the HMO constructs used for the production of stable lines, finding an optimal crop species for HMO production and optimizing growing conditions.

## Discussion

Human milk oligosaccharides are major contributors to the bioactive properties of human milk<sup>1,2</sup>. While the unique bioactivities of few HMOs have been investigated<sup>3,8–11</sup>, over 100 HMOs remain to be studied owing to lack of access to material, representing a wealth of potential bioactive molecules. Here, we report the production of all three classes of HMOs in planta. Notably, expression of HMO biosynthetic pathways in planta created a variety of complex HMOs, including oligosaccharides that at present are not produced in microbial platforms, which indicates plants could serve as a platform for the production of a range of HMOs which are not currently producible in microbial hosts. Furthermore, we optimized the production of both specific HMOs and HMO classes by overexpressing nucleotide sugar biosynthetic pathways. The diversity of HMOs produced from inserting a relatively small number of genes shows the ability of plants to generate complex sugars. Transient expression serves as a viable platform for testing HMO biosynthetic genes and small-scale production of HMOs for functional validation. Optimized laboratory-scale purification methods enabled the isolation of oligosaccharides, including HMOs, from plant tissue, showing

the potential of plants as an industrial source of HMOs. HMOs purified from plants provided selective bifidogenic activity, indicating that they would serve as a potent prebiotic in vivo. Future pathway engineering could yield biosynthetic pathways for the production of all HMOs present in human milk, including complex, branched HMOs. This would enable the study and potential consumption of bioactive HMOs that are currently unavailable.

Despite the field of plant synthetic biology being in its nascent stages, plants have the potential to produce compounds of interest at lower costs than microbial platforms because of their intrinsic and unique metabolic capabilities<sup>39</sup>. Additionally, plants are capable of using atmospheric carbon dioxide during their growth cycle to produce target compounds and biofuels, improving the sustainability of target compound production. We demonstrate the ability of stably transformed plants to produce two HMOs that are abundant in breast milk of most mothers, 2'FL and LNFPI. While the feasibility of purifying HMOs from plants at industrial scales still needs to be validated, our TEA results indicate that HMO production in commercially relevant crops has the potential to be a more cost-effective platform than microbial production for complex HMOs.

As the demand for HMOs increases because of the growing infant formula and adult prebiotic markets, plants may emerge as a cost-competitive and sustainable platform for the production of diverse HMOs at agricultural scales. Furthermore, the diversity of plant-produced HMOs will provide researchers with access to HMOs that were previously inaccessible. Since the structure of an HMO determines its bioactivity, this could lead to the discovery of HMOs that treat various gastrointestinal illnesses. Additionally, production of HMOs in plants could permit direct consumption as food by directly ingesting the plant or products made from the plant. Such a product may serve as a consumable source of prebiotic HMOs for humans or be added to forage crops for animal consumption. Overall, the production of HMOs in planta provides the opportunity to simultaneously improve the scale of HMO production and expand the diversity of HMOs available to improve the gastrointestinal health of infants and adults.

## Methods

### Plant growth

*N. benthamiana* was grown from seed in 3.5 inch square pots in a controlled environment facility. Plants were grown with a 12 h/12 h day/night cycle at  $-700 \mu\text{mol}$  of photons per  $\text{m}^2 \text{s}^{-1}$ . Daytime growth chamber temperatures were kept at 26 °C. Night-time growth chamber temperatures were kept at 25 °C. Relative humidity in the growth chamber was kept between 60% and 75%.

### Plasmid construction and transient expression

For transient expression, the native sequences for candidate glycosyltransferases were PCR amplified and cloned into the binary vector, PMS057 (ref. 43), using Golden Gate assembly<sup>44</sup>, Gibson assembly<sup>45</sup> or restriction-ligation (see Supplementary Table 5 for sequences). XLI-blue *E. coli* cells were transformed with the assembled plasmids via heat shock<sup>46</sup>. Transformed cells were selected by plating cells on Lysogeny broth (LB) agar plates containing 50  $\mu\text{g ml}^{-1}$  of kanamycin. Plasmid assembly was confirmed by means of miniprep and Sanger sequencing (Azenta). *A. tumefaciens* str. GV3101 was transformed using sequence-verified plasmids by electroporation<sup>47</sup>. Transformed colonies were selected using LB agar plates containing 50  $\mu\text{g ml}^{-1}$  of kanamycin, 50  $\mu\text{g ml}^{-1}$  of rifampicin and 10  $\mu\text{g ml}^{-1}$  of gentamicin. *A. tumefaciens* str. GV3101 harbouring individual candidate glycosyltransferases were grown in LB overnight to OD<sub>600nm</sub> (VWR, V-1200) of 0.8–1.2. The cultures were centrifuged at 4,000g for 10 min and the supernatant was decanted. Cell pellets were resuspended in infiltration media (10 mM MES, 10 mM MgCl<sub>2</sub>, 500  $\mu\text{M}$  acetosyringone, pH 5.6) and incubated at room temperature for 1 h with gentle rocking (Thermolyne, VariMix). *A. tumefaciens* strains harbouring each glycosyltransferase were

mixed in equal amounts alongside a strain harbouring the p19 silencing suppressor<sup>48</sup> to reach a final OD<sub>600 nm</sub> of 0.5. *A. tumefaciens* mixtures were injected into the abaxial side of a leaf on a 4-week-old *N. benthamiana* using a needleless syringe. Each experiment was performed with three biological replicates.

For the production of stable lines, HMO10 and HMO11 constructs were generated through a multipart Golden Gate assembly containing subcloned transcriptional units. Assembled plasmids were transformed and sequence verified as described above. *N. benthamiana* was transformed using *A. tumefaciens* str. EHA105 harbouring HMO10 or HMO11 by the UC Davis Plant Transformation Facility.

**Quantitative PCR with reverse transcription.** Total messenger RNA was extracted using E.Z.N.A. plant RNA kit (Omega Bio-tek) following manufacturer's directions using the RB lysis buffer variation and on-column DNase digestion; complementary DNA synthesis was achieved with SSIV Vilo IV kit using random hexamers (Thermo Fisher Scientific). Quantitative PCR was performed using a CFX96 Real-Time thermocycler (Bio-Rad) programmed for detection of SYBR intercalating dye with the following temperature programming: 95 °C for 3 min, then 95 °C for 30 s, 60 °C for 45 s, repeated 34 times, then a gradual increase from 65 °C to 95 °C at 0.5 °C per minute to generate melt curves. Sso-Advanced Universal SYBR Green Supermix (Bio-Rad) was used for qPCR amplification. A previously validated primer set was used to amplify EF1 $\alpha$  for internal normalization, primers for target genes were designed with Benchling's qPCR primer design wizard and synthesized by IDT. One target gene from each transcriptional unit was chosen for both constructs (Supplementary Table 6). Melt curves for the product of all primer sets were unimodal and steep, suggesting only a single product was formed for each primer set. No reverse-transcriptase controls showed no amplification within the dynamic range of samples, confirming the efficacy of DNase treatment and no template controls instituted at the beginning of RNA extraction with no plant matter and kept in parallel with real samples throughout all molecular steps did not amplify, confirming lack of contamination with extraneous DNA. Normalized relative expression was calculated using the  $\Delta\Delta C_q$  method and normalized by setting the average level of amplification in the wild-type samples as 1.

### HMO extraction for identification of HMOs from individual leaves

*N. benthamiana* leaves transiently expressing HMO biosynthetic enzymes were harvested 5 days after infiltration. Three leaves of *N. benthamiana* stable lines transformed with HMO10 and HMO11 were harvested at 4 weeks old. Following harvest, vasculature was removed and leaves were frozen in liquid nitrogen and lyophilized (Labconco, Freezone 4.5) for 2 days. Lyophilized leaves were homogenized via a bead mill (Retsch, MM400) at 20 Hz for 10 min. Oligosaccharides were extracted from 20 mg of lyophilized leaf tissue by ethanol precipitation. To each sample, 1 ml of 80% ethanol was added before homogenization on a bead mill at 10 Hz for 1 min. Samples were then precipitated overnight at -20 °C and centrifuged at 10,000g for 15 min. The supernatant was transferred to a 2 ml screw-cap tube. The pellet was washed twice by adding 500  $\mu$ l of 80% ethanol, homogenizing via bead mill for 1 min and centrifuging at 10,000g for 15 min. The supernatant and washes were combined and dried in a vacuum centrifuge (Genevac EZ-2, SP Scientific). Dried supernatants were reconstituted in 200  $\mu$ l of water and subjected to both C18 and PGC SPE (Thermo Fisher Scientific) in 96-well plate format. C18 cartridges containing 25 mg of stationary phase were first conditioned by two additions of 250  $\mu$ l of acetonitrile (ACN) followed by four additions of 250  $\mu$ l of water. Samples were then loaded and eluted with two volumes of 200  $\mu$ l of water. PGC cartridges containing 40 mg of stationary phase were conditioned by addition of 400  $\mu$ l of water, 400  $\mu$ l of 80% (v/v) ACN and water, followed by two volumes of 400  $\mu$ l of water. The sample eluate from C18 SPE was

then loaded, washed thrice with 500  $\mu$ l of water and eluted using two volumes of 200  $\mu$ l of 40% (v/v) ACN and water. The purified extracts were dried in a vacuum centrifuge and reconstituted in 100  $\mu$ l of water before injecting 5  $\mu$ l for liquid chromatography mass spectrometry (LC-MS) analysis.

For quantification of LNFPI in Figs. 2 and 3, LNFPI at known concentrations was added to the extraction solution. The extraction solution was then used on wild-type *N. benthamiana*. This was done to ensure accuracy of HMO quantification by accounting for HMO losses in the extraction processes and ion suppression that could occur due to the plant metabolites present.

### LC-MS analysis of HMOs from individual leaves

For initial screening, chromatographic separation was carried out using a Thermo Scientific Vanquish UHPLC system equipped with a Waters BEH C18 Amide column (HILIC) (1.7  $\mu$ m, 100 mm  $\times$  2.1 mm). A 10 min binary gradient was used based on ref. 49: 0.0–4.0 min, 25–35% A; 4.0–8.50 min, 35–65% A; 8.50–8.70 min, 25% A. Mobile phase A consisted of 3% ACN (v/v) in water with 0.1% formic acid and mobile phase B consisted of 95% ACN (v/v) in water with 0.1% formic acid.

For identification of HMOs produced, we performed LC-MS analysis using a Thermo Scientific Vanquish 3000 UPLC system connected to Thermo Scientific Q Exactive mass spectrometer. Chromatographic separation was carried out using a Hypercarb PGC column (5  $\mu$ m, 150 mm  $\times$  1 mm, Thermo Scientific). A 40 min binary gradient using 3% ACN in water containing 0.1% formic acid (Solvent A) and 90% (v/v) ACN in water containing 0.1% formic acid was performed as follows: 100% A, 0–2.5 min; 100–84% A, 2.5–15 min; 84–42% A, 15–20 min; 42–0% A, 20–22 min; 0% A, 22–28 min; 0–100% A, 28–30 min; 100% A, 30–40 min.

For identification of HMOs, the Q Exactive mass spectrometer equipped with an electrospray ionization source was operated in positive ionization mode with the following parameters: scan range  $m/z$  133.4–2,000; spray voltage 2.5 kV, capillary temperature 320 °C, aux gas heater temperature 325 °C, sheath gas flow rate 25, aux gas flow rate 8, sweep gas flow rate 3. MS/MS analysis was performed using stepped collision energies of 20, 30, 40 eV. MsDIAL was used for data analysis<sup>50</sup>.

For quantification of LNFPI and HMO profiling, mass spectral analysis was carried out on an Agilent 6530 Accurate-Mass Q-ToF MS operated in positive mode using data-dependent acquisition. The gas temperatures were held at 150 °C. The fragmentor, skimmer, octopole and capillary were operated at 70, 55, 750 and 1,800 V, respectively. The collision energy was based on the empirically derived linear formula  $(1.8 \times (m/z/100) - 3.6)$ . The reference mass used for calibration was  $m/z$  922.009798. The Agilent MassHunter Qualitative software was used for data analysis. Oligosaccharides were identified using an inhouse library, their MS/MS spectra and comparison to either authenticated standards or a pool of HMOs of known composition.

### Extraction and purification of HMOs from pooled leaves

Five grams of lyophilized and ground *N. benthamiana* leaves transiently expressing the LNFPI and GDP-fucose biosynthetic pathways was mixed with 150 ml of water and agitated for 15 min at room temperature in a stirring plate. The mixture was centrifuged at 4,000g for 5 min and the supernatant was separated. The extraction was repeated two more times, combining the supernatant each time. The final supernatant was filtered using a 0.22  $\mu$ m Millipore Steritop vacuum filter. The extraction process was carried out in duplicate to ensure reproducibility.

Yeast fermentation was carried out to eliminate simple sugars (glucose, sucrose and fructose) from the extracts<sup>51</sup>. Briefly, autoclaved extracts were inoculated with 0.4 g l<sup>-1</sup> of commercial active dry yeast *Saccharomyces cerevisiae*<sup>52</sup> (UCD 522 Montrachet, Lallemant) at 30 °C, 150 rpm for 24 h (Excella E24 Incubator Shaker Series, New Brunswick Scientific). After 24 h, the samples were centrifuged at 4,000g for 5 min and filtered using a 0.22  $\mu$ m Millipore Steritop vacuum filter to remove the yeast. Samples were concentrated using a vacuum concentrator



(Genevac miVac Centrifugal Concentrator) at room temperature and frozen until their purification.

PVPP (Sigma-Aldrich) was used to bind phenolic compounds within the sample following a previous protocol<sup>52</sup>. Briefly, 3 g of PVPP was conditioned by mixing it with 100 ml of 12 M HCl at 100 °C for 30 min with constant stirring in a stirring plate. After cooling off, the slurry was centrifuged at 4,000g for 5 min and filtered using a 0.22 µm Millipore Steritop vacuum filter. Subsequently, the PVPP was washed with nanopure water until the flow-through reached pH 7. Activated PVPP was mixed with water to a final concentration of 20 mg of PVPP per ml.

PVPP suspension was added to the extracts at a concentration of 6 mg of PVPP per ml and agitated at room temperature for 15 min on a stirring plate. After the time had elapsed, the sample containing the plant extracts and the PVPP was centrifuged at 4,000g for 5 min and filtered using a 0.22 µm Millipore Steritop vacuum filter to separate the PVPP containing the bound phenolics. To eliminate residual phenolics from the extracts, more PVPP was added to the supernatant (6 mg of PVPP per ml) and the process was repeated. The filtrate containing the oligosaccharides was concentrated and frozen until further purification.

Two SPE columns of 60 ml were packed with 15 g of Bondesil-C18, 40 µm suspended in 20 ml of ACN. After the ACN was drained, a frit was added to the C18. Before loading the samples, the columns were conditioned with three volumes of ACN and three volumes of nanopure water. PVPP-treated extracts were loaded onto the conditioned C18 columns and the oligosaccharides were washed with 250 ml of nanopure water divided into four washes. To ensure the complete removal of interfering compounds, C18 SPE was repeated two more times. The purified HMO fractions were dried in a vacuum concentrator (Genevac miVac Centrifugal Concentrator) at room temperature and frozen.

### Compositional analysis of plant material

Total carbohydrate content was assessed by the anthrone method with modifications<sup>53</sup>. In a 96-well microplate, 40 µl of purified and diluted extracts were combined with 100 µl of anthrone reagent (2 mg ml<sup>-1</sup> (w/v) in cold 98% sulfuric acid) and mixed through pipette tip aspiration. The microplate was incubated for 3 min at 92 °C in a water bath followed by 5 min at a room temperature water bath and then 15 min in a 45 °C Thermolyne Benchtop muffle furnace (Thermo Fisher Scientific). The plate was cooled for 3 min at room temperature before measuring the absorbance with a SpectroMax M5UV/Vis spectrophotometer (Molecular Devices) at 630 nm. Total carbohydrate quantification calculations were based on a glucose standard curve. Each plant extract was prepared in duplicate and each sample was further analysed in duplicate.

Total phenolic content of the extracts was determined according to the Folin-Ciocalteu spectrophotometric method as described by ref. 54.

Simple sugars (glucose, sucrose and fructose) were quantified by high-performance anion exchange chromatography with pulsed amperometric detection on a Thermo Fisher Dionex ICS-5000+ HPAE-PAD system based on a method described by ref. 55 with modifications. Diluted extracts (1:100, v/v or 1:1,000 in nanopure water) were filtered through a 0.2 µm syringe filter (Agilent Captiva Econo Filter, PES, 13 mm, 0.2 µm) into 2 ml vials with septa. The samples (25 µl) were injected into a CarboPac PA200 guard column (3 × 50 mm) and a CarboPac PA200 analytical column (3 × 250 mm) and chromatographic separation was carried out with a 12 min gradient elution (from 0.6% to 25% B in 12 min), 0.5 ml min<sup>-1</sup> flow rate. The solvent system consisted of A: 100% water; and B: 200 mM sodium hydroxide. Calibration curves (correlation coefficient ≥ 0.999) were prepared using glucose, sucrose and fructose standards.

### Quantification of HMOs by QqQLC-MS

Detection and quantitation of HMOs were performed using an Agilent 6470 triple quadrupole LC-MS system (QqQLC-MS) equipped with an

Advance Bio Glycan Map column (2.1 mm × 150 mm, 2.7 µm, Agilent). The mobile phase consisted of 10 mM ammonium acetate in 3% ACN, 97% water (v/v, pH 4.5; A) and 10 mM ammonium acetate in 95% ACN, 5% water (v/v, pH 4.5; B). The chromatographic separation was carried out at 35 °C with gradient elution at a flow rate of 0.3 ml min<sup>-1</sup>. The MS analysis was conducted in positive ion mode with source parameters as follows: the gas temperature was 150 °C at a flow rate of 10 l min<sup>-1</sup>; the nebulizer was 45 psi; the sheath gas temperature was 250 °C at a flow rate of 7 l min<sup>-1</sup>; capillary voltage was 2,200 V. See Supplementary Table 2 for gradient and multiple reaction monitoring transitions.

### Characterization of HMOs by LC-QToF-MS

Oligosaccharides were purified by a two-step SPE using C18 (HyperSep C18-96, 50 mg bed weight; Thermo Fisher Scientific) and PGC (HyperSep Hypercarb-96, 25 mg bed weight; Thermo Fisher Scientific)<sup>56</sup>. The samples were filtered (Captiva Premium Syringe Filter PES membrane, 4 mm diameter, 0.2 µm pore size, LC/MS certified) into 200 µl vials.

Individual oligosaccharide compositions were analysed with an Agilent 6520 NanoChip LC-QToF mass spectrometer. Oligosaccharides separation was achieved with a microfluidic high-performance liquid chromatography PGC chip containing an enrichment (4 mm, 40 nl) and an analytical (75 µl × 43 mm) column as well as a nanoelectrospray tip, using a binary solvent gradient of solvent A (5 mM ammonium acetate in 3% ACN, 97% water (v/v)) and solvent B (5 mM ammonium acetate in 90% ACN, 10% water (v/v)) based on a previously optimized method<sup>55</sup>. The gradient was 0–16% B at 0–20 min, 16–44% B at 20–30 min, 44–100% B at 30–35 min, 100% B at 35–45 min and 100–0% B from 45 to 45.1 min, followed by a 15 min re-equilibration of 100% A<sup>57</sup>. The mass spectrometer was operated in positive ionization mode with a range of *m/z* 320–2,500 and an electrospray capillary voltage of 1,800–1,900 V. Reference masses of *m/z* 922.009 and 1,221.991 provided continuous internal calibration. All samples were analysed using MS/MS with tandem fragmented peaks selected by the automated precursor selection of the six ions with highest signal intensity with a medium isolation width. The Q-ToF MS had a ramped collision energy slope of 0.02 based on *m/z* values with an offset of –3.5 V. The acquisition rate of 1 spectrum per s was used for both MS and MS/MS. Each spectrum was manually examined and molecular masses were confirmed with Agilent MassHunter Qualitative Analysis B.07.00 software using the molecular feature extraction and a maximum tolerance of 20 ppm.

### Bacterial strains and growth conditions

*B. longum* subsp. *infantis* ATCC 15697 and *B. animalis* subsp. *lactis* ATCC 27536 were cultured at 37 °C in a Coy vinyl anaerobic bubble with an atmosphere of 2.5% H<sub>2</sub>, -5% CO<sub>2</sub> and balance N<sub>2</sub>. Routine culturing was done with Difco MRS + 0.05% L-cysteine HCl (MRSC) and carbohydrate-specific culturing was done with modified MRS (mMRSC), which was prepared per litre as follows: 10 g of Bacto proteose peptone no. 3, 10 g of Bacto casitone, 5 g of Bacto yeast extract, 2 g of triammonium citrate, 5 g of sodium acetate trihydrate, 200 mg of magnesium sulfate hexahydrate, 34 mg of manganese sulfate monohydrate, 0.5 g of L-cysteine HCl and 1.063 g of Tween-80. Normally, 2 g of anhydrous dipotassium phosphate would also be added but it was precipitated by the plant HMO preparation.

### Growth curves

One colony was used to inoculate 1 ml mMRSC + dipotassium phosphate + 1% lactose monohydrate and incubated for 24 h. The growth curve inoculum was cultured by diluting the 24 h culture 1:100 in mMRSC + dipotassium phosphate + 1% lactose, then incubating for 12–15 h. The inoculum was prepared by washing the cells twice with one volume of 1× PBS. Growth curve cultures (160 µl) were contained in flat-bottomed, optically clear, 96-well, lidded plates and they had a final inoculum and sugar concentration of 1% in mMRSC. Lactose was the growth control substrate, water was the no-growth control

substrate and pooled HMO<sup>58</sup> was the HMO-growth control substrate. Cultures were done in triplicate. Wells were overlaid with 40 µl of sterile mineral oil and incubated in a BMG SpectroStar Nano. The plate reader was set to read the OD<sub>600 nm</sub> of each well 30 times in a spiral pattern every half-hour with medium orbital shaking before each read. All media had uninoculated controls whose OD<sub>600 nm</sub> was subtracted from that of the inoculated medium.

### Technoeconomic analysis

In this study, we used SuperPro Designer v.12 to develop technoeconomic models of HMOs which can be produced from both plant and microbial systems. To do this, we first developed process models and then applied discount cash flow analysis of the theoretical production of LNFPI in plants and microbes. The simplified process flow diagram can be found in Supplementary Fig. 8. In the plant system, we adopted integrated cellulosic biorefinery design to coproduce HMO and ethanol to maximize the use of plant biomass. Biomass sorghum was used as the representative plant because its characteristics, such as high yields and drought tolerance, are ideal for biofuel production. Previous studies demonstrated that using biomass sorghum as the representative plant to accumulate value-added bioproducts could improve the economic performance of an integrated biorefinery<sup>39,40</sup>. Since ethanol is coproduced in the biorefinery, two ethanol selling prices were considered: (1) baseline cellulosic biofuel selling price of US\$1.44 l<sup>-1</sup> of gasoline equivalent and (2) target fuel price of US\$1.00 l<sup>-1</sup> of gasoline equivalent.

Briefly, biomass sorghum with 0.31% dry weight HMOs accumulated in the plant biomass are harvested and transported to the biorefinery gate for preprocessing and short-term onsite storage. HMO extraction, separation and recovery is then conducted on the basis of our laboratory processes and as described in previous texts. Water is used to extract HMO from the biomass (room temperature for 6 h) since the industrial extraction process will last longer than the laboratory-scale process given the large quantity of biomass being processed. The extraction efficiency is assumed at 90% based on previous studies<sup>39</sup>. After extraction, the slurry is first cooled down to room temperature and is sent to centrifugation to remove water. This water is sent to the wastewater treatment unit located in the biorefinery for recycling and reusing. Afterwards, multistage ultrafiltration is used to recover HMO from the extraction stream. Before transporting to HMO onsite storage, another centrifuge is applied to ensure the recovery and purity of the final product. In this plant system, the extracted HMOs are considered as the main product with the purity of >95%. The remaining biomass from biomass sorghum is routed to ionic liquid pretreatment for biomass deconstruction. After ionic liquid pretreatment, enzymatic hydrolysis and ethanol fermentation are conducted to produce ethanol, followed by distillation and molecular sieve to remove excess water. Wastewater from the overall process is routed to the wastewater treatment sector to produce reusable process water and biogas, which can be combusted in the onsite turbogenerator, along with other solids from biomass, to generate heat and electricity that can satisfy the facility's need. In the microbial system, HMOs were produced as the single product in the biorefinery. Unlike the plant system, pure sugar is used in the microbial system as the sole feedstock and no wastewater treatment sector and onsite combustion are designed for the microbial system. The downstream processing of microbial production is adopted from peer-reviewed publication<sup>41</sup>. After LNFPI is produced in the bioreactor under the bioconversion condition of 30 °C for 52 h, ultrafiltration is first applied to remove biomass and ion exchange chromatography to remove ions and other charged impurities. Multistage nanofiltration is further used to reduce the total volume and remove excess impurities. Gel filtration is used for the final purification of HMO from the microbial production process. The purity of the final product is >98% through this process.

After developing technoeconomic models in SuperPro Designer, we performed mass and energy balance and then applied discounted cash flow analysis to quantify the MSP of HMOs (US\$ kg<sup>-1</sup>). For both

systems, we assume the biorefinery can operate 24 h per day and 330 days per year for 30 years. The unit price of biomass sorghum is assumed at US\$95 per bone-dry tonne and we assume in the plant system the cellulosic biorefinery can intake 2,000 bone-dry tonnes of biomass sorghum per day. The unit price of ionic liquid is US\$2 kg<sup>-1</sup> with a range of US\$1–5 kg<sup>-1</sup>. Total capital investment includes installed equipment cost, piping costs, engineering costs, warehouse, site development, construction fees, contingency costs, land costs, startup and working capital. Annual operating costs include raw materials costs, utility costs, labour costs and facility-dependent costs such as insurance. These parameters are kept constant as per the 2011 National Renewable Energy Laboratory report<sup>39</sup>.

### Statistics and reproducibility

No statistical method was used to predetermine sample size as performing experiments in biological triplicate is standard of the field. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not blinded to allocation during experiments and outcome assessment. All experiments looking at individual leaves were conducted in biological triplicate. For the laboratory-scale purifications, extractions were completed in duplicate and each duplicate was measured with two technical replicates. Microbial growth assays were conducted in biological triplicate. For statistical analysis of LNFPI optimization, a heteroscedastic two-tailed Student's *t*-test with the LNFPI pathway expressed alone was used as the reference group in RStudio v.1.2.5033. GraphPad v.10 was used to plot the microbial growth assays.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

The main data supporting the findings of this study are available within the article and its Supplementary Information.

### References

- Walsh, C., Lane, J. A., van Sinderen, D. & Hickey, R. M. Human milk oligosaccharides: shaping the infant gut microbiota and supporting health. *J. Funct. Foods* **72**, 104074 (2020).
- Kirmiz, N., Robinson, R. C., Shah, I. M., Barile, D. & Mills, D. A. Milk glycans and their interaction with the infant-gut microbiota. *Annu. Rev. Food Sci. Technol.* **9**, 429–450 (2018).
- Vandenplas, Y. et al. Human milk oligosaccharides: 2'-fucosyllactose (2'-FL) and lacto-N-neotetraose (LNnT) in infant formula. *Nutrients* **10**, 1161 (2018).
- Wiciński, M., Sawicka, E., Gębalski, J., Kubiak, K. & Malinowski, B. Human milk oligosaccharides: health benefits, potential applications in infant formulas and pharmacology. *Nutrients* **12**, 266 (2020).
- 2022 Breastfeeding Report Card (CDC, 2022); <https://www.cdc.gov/breastfeeding/data/reportcard.htm>
- Urashima, T., Horiuchi, R., Sakanaka, M., Katayama, T. & Fukuda, K. Lactose or milk oligosaccharide: which is significant among mammals? *Anim. Front.* **13**, 14–23 (2023).
- Lee, S. et al. Human milk oligosaccharide 2'-fucosyllactose supplementation improves gut barrier function and signaling in the vagal afferent pathway in mice. *Food Funct.* **12**, 8507–8521 (2021).
- Šuligoj, T. et al. Effects of human milk oligosaccharides on the adult gut microbiota and barrier function. *Nutrients* **12**, 2808 (2020).
- Elison, E. et al. Oral supplementation of healthy adults with 2'-O-fucosyllactose and lacto-N-neotetraose is well tolerated and shifts the intestinal microbiota. *Br. J. Nutr.* **116**, 1356–1368 (2016).

10. Iribarren, C. et al. Human milk oligosaccharide supplementation in irritable bowel syndrome patients: a parallel, randomized, double-blind, placebo-controlled study. *Neurogastroenterol. Motil.* **32**, e13920 (2020).
11. Iribarren, C. et al. The effects of human milk oligosaccharides on gut microbiota, metabolite profiles and host mucosal response in patients with irritable bowel syndrome. *Nutrients* **13**, 3836 (2021).
12. Bych, K. et al. Production of HMOs using microbial hosts—from cell engineering to large scale production. *Curr. Opin. Biotechnol.* **56**, 130–137 (2019).
13. Palur, D. S. K., Pressley, S. R. & Atsumi, S. Microbial production of human milk oligosaccharides. *Molecules* **28**, 1491 (2023).
14. Shani, G. et al. Fucosylated human milk oligosaccharide foraging within the species *Bifidobacterium pseudocatenulatum* is driven by glycosyl hydrolase content and specificity. *Appl. Environ. Microbiol.* **88**, e0170721 (2022).
15. Soyylmaz, B. et al. The mean of milk: a review of human milk oligosaccharide concentrations throughout lactation. *Nutrients* **13**, 2737 (2021).
16. Chen, X. in *Advances in Carbohydrate Chemistry and Biochemistry* (eds Baker, D. C. & Horton, D.) 113–190 (Academic Press, 2015).
17. Bar-Peled, M. & O'Neill, M. A. Plant nucleotide sugar formation, interconversion and salvage by sugar recycling. *Annu. Rev. Plant Biol.* **62**, 127–155 (2011).
18. Zhu, F., Du, B. & Xu, B. A critical review on production and industrial applications of beta-glucans. *Food Hydrocoll.* **52**, 275–288 (2016).
19. Palaniappan, A., Antony, U. & Emmambux, M. N. Current status of xylooligosaccharides: production, characterization, health benefits and food application. *Trends Food Sci. Technol.* **111**, 506–519 (2021).
20. Choct, M., Dersjant-Li, Y., McLeish, J. & Peisker, M. Soy oligosaccharides and soluble non-starch polysaccharides: a review of digestion, nutritive and anti-nutritive effects in pigs and poultry. *Asian-Australas. J. Anim. Sci.* **23**, 1386–1398 (2010).
21. Shoaib, M. et al. Inulin: properties, health benefits and food applications. *Carbohydr. Polym.* **147**, 444–454 (2016).
22. Barnum, C. R., Endelman, B. J. & Shih, P. M. Utilizing plant synthetic biology to improve human health and wellness. *Front. Plant Sci.* **12**, 1824 (2021).
23. Kellman, B. P. et al. Elucidating human milk oligosaccharide biosynthetic genes through network-based multi-omics integration. *Nat. Commun.* **13**, 2455 (2022).
24. Thompson, M. G. et al. *Agrobacterium tumefaciens*: a bacterium primed for synthetic biology. *BioDesign Res.* **2020**, 8189219 (2020).
25. Parschat, K., Schreiber, S., Wartenberg, D., Engels, B. & Jennewein, S. High-titer de novo biosynthesis of the predominant human milk oligosaccharide 2'-fucosyllactose from sucrose in *Escherichia coli*. *ACS Synth. Biol.* **9**, 2784–2796 (2020).
26. Ooi, K.-E., Zhang, X.-W., Kuo, C.-Y., Liu, Y.-J. & Yu, C.-C. Chemoenzymatic synthesis of asymmetrically branched human milk oligosaccharide lacto-*N*-hexaose. *Front. Chem.* **10**, 905105 (2022).
27. McArthur, J. B., Yu, H. & Chen, X. A bacterial  $\beta$ 1-3-galactosyltransferase enables multigram-scale synthesis of human milk lacto-*N*-tetraose (LNT) and its fucosides. *ACS Catal.* **9**, 10721–10726 (2019).
28. Li, Y. et al. Donor substrate promiscuity of bacterial  $\beta$ 1-3-*N*-acetylglucosaminyltransferases and acceptor substrate flexibility of  $\beta$ 1-4-galactosyltransferases. *Bioorg. Med. Chem.* **24**, 1696–1705 (2016).
29. Thurl, S., Munzert, M., Boehm, G., Matthews, C. & Stahl, B. Systematic review of the concentrations of oligosaccharides in human milk. *Nutr. Rev.* **75**, 920–933 (2017).
30. Chen, X., Zhao, C. & Yu, H. Te2ft enzyme for enzymatic synthesis of alpha1-2-fucosides. International patent application PCT/US2016/067601 (2017).
31. Hobbs, M., Jahan, M., Ghorashi, S. A. & Wang, B. Current perspective of sialylated milk oligosaccharides in mammalian milk: implications for brain and gut health of newborns. *Foods* **10**, 473 (2021).
32. Castilho, A. et al. Construction of a functional CMP-sialic acid biosynthesis pathway in *Arabidopsis*. *Plant Physiol.* **147**, 331–339 (2008).
33. Drouillard, S., Mine, T., Kajiwara, H., Yamamoto, T. & Samain, E. Efficient synthesis of 6'-sialyllactose, 6,6'-disialyllactose and 6'-KDO-lactose by metabolically engineered *E. coli* expressing a multifunctional sialyltransferase from the *Photobacterium* sp. JT-ISH-224. *Carbohydr. Res.* **345**, 1394–1399 (2010).
34. Chen, X., Thon, V. & Yu, H. PmST3 enzyme for chemoenzymatic synthesis of alpha-2-3-sialosides. US patent application US14/356,376 (2017).
35. Ryan, M. D. & Drew, J. Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. *EMBO J.* **13**, 928–933 (1994).
36. Ghodduji, H. B., Grandison, M. A., Grandison, A. S. & Tuohy, K. M. In vitro study on gas generation and prebiotic effects of some carbohydrates and their mixtures. *Anaerobe* **13**, 193–199 (2007).
37. Sela, D. A. et al. The genome sequence of *Bifidobacterium longum* subsp. infantis reveals adaptations for milk utilization within the infant microbiome. *Proc. Natl Acad. Sci. USA* **105**, 18964–18969 (2008).
38. Underwood, M. A. et al. A comparison of two probiotic strains of bifidobacteria in premature infants. *J. Pediatr.* **163**, 1585–1591 (2013).
39. Yang, M. et al. Accumulation of high-value bioproducts in planta can improve the economics of advanced biofuels. *Proc. Natl Acad. Sci. USA* **117**, 8639–8648 (2020).
40. Yang, M. et al. Comparing in planta accumulation with microbial routes to set targets for a cost-competitive bioeconomy. *Proc. Natl Acad. Sci. USA* **119**, e2122309119 (2022).
41. Derya, S. M. et al. Biotechnologically produced fucosylated oligosaccharides inhibit the binding of human noroviruses to their natural receptors. *J. Biotechnol.* **318**, 31–38 (2020).
42. Sandalow, D., Aines, R., Friedmann, J., McCormick, C. & Sanchez, D. *Biomass Carbon Removal and Storage (BiCRS) Roadmap* (US Department of Energy, 2021); <https://www.osti.gov/servlets/purl/1763937/>
43. Barnum, C. R., Endelman, B. J., Ornelas, I. J., Pignolet, R. M. & Shih, P. M. Optimization of heterologous glucoraphanin production in planta. *ACS Synth. Biol.* **11**, 1865–1873 (2022).
44. Engler, C., Gruetzner, R., Kandzia, R. & Marillonnet, S. Golden gate shuffling: a one-pot DNA shuffling method based on type IIs restriction enzymes. *PLoS ONE* **4**, e5553 (2009).
45. Gibson, G. R. et al. Expert consensus document: the International Scientific Association for Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of prebiotics. *Nat. Rev. Gastroenterol. Hepatol.* **14**, 491–502 (2017).
46. Froger, A. & Hall, J. E. Transformation of plasmid DNA into *E. coli* using the heat shock method. *J. Vis. Exp.* **2007**, 253 (2007).
47. Lin, J.-J. in *Electroporation Protocols for Microorganisms* (ed. Nickoloff, J. A.) 171–178 (Humana Press, 1995); <https://doi.org/10.1385/0-89603-310-4:171>
48. Lakatos, L., Szittyá, G., Silhavy, D. & Burgyán, J. Molecular mechanism of RNA silencing suppression mediated by p19 protein of tombusviruses. *EMBO J.* **23**, 876–884 (2004).
49. Xu, G. et al. Absolute quantitation of human milk oligosaccharides reveals phenotypic variations during lactation. *J. Nutr.* **147**, 117–124 (2017).

50. Tsugawa, H. et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. Methods* **12**, 523–526 (2015).
51. de Moura Bell, J. M. L. N. et al. An integrated bioprocess to recover bovine milk oligosaccharides from colostrum whey permeate. *J. Food Eng.* **216**, 27–35 (2018).
52. Magalhães, P. J. et al. Isolation of phenolic compounds from hop extracts using polyvinylpyrrolidone: characterization by high-performance liquid chromatography–diode array detection–electrospray tandem mass spectrometry. *J. Chromatogr. A* **1217**, 3258–3268 (2010).
53. Laurentin, A. & Edwards, C. A. A microtiter modification of the anthrone–sulfuric acid colorimetric assay for glucose-based carbohydrates. *Anal. Biochem.* **315**, 143–145 (2003).
54. Singleton, V. L., Orthofer, R. & Lamuela-Raventós, R. M. Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin–Ciocalteu reagent. *Methods Enzymol.* **299**, 152–178 (1999).
55. Huang, Y.-P., Paviani, B., Fukagawa, N. K., Phillips, K. M. & Barile, D. Comprehensive oligosaccharide profiling of commercial almond milk, soy milk and soy flour. *Food Chem.* **409**, 135267 (2023).
56. Huang, Y.-P., Robinson, R. C. & Barile, D. Food glycomics: dealing with unexpected degradation of oligosaccharides during sample preparation and analysis. *J. Food Drug Anal.* **30**, 62–76 (2022).
57. Bhattacharya, M., Salcedo, J., Robinson, R. C., Henrick, B. M. & Barile, D. Peptidomic and glycomic profiling of commercial dairy products: identification, quantification and potential bioactivities. *NPJ Sci. Food* **3**, 4 (2019).
58. Lee, H., Garrido, D., Mills, D. A. & Barile, D. Hydrolysis of milk gangliosides by infant-gut associated bifidobacteria determined by microfluidic chips and high-resolution mass spectrometry. *Electrophoresis* **35**, 1742–1750 (2014).
59. Humbird, D. et al. *Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover* (US Department of Energy, 2011); <https://www.osti.gov/servlets/purl/1013269/>

## Acknowledgements

C.R.B. acknowledges support from the National Institutes of Health NIGMS T32 Training Program and the US Department of Energy. C.R.B. and P.M.S. acknowledge support from the Joint BioEnergy Institute (<http://www.jbei.org>) supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy. P.M.S. acknowledges support from grant no. RO0AT009573 from the National Center for Complementary and Integrative Health at the National Institutes of Health. M.Y. acknowledges USDA NIFA HATCH NCO2948. D.A.M. acknowledges support of the P. J. Shields, Chair in Dairy Food Science. D.B. acknowledges the Hatch project

CA-D-FST-2744-H. We thank X. Chen for sharing plasmids with genes encoding several glycosyltransferases.

## Author contributions

C.R.B. performed the construct design and agrobacterium infiltrations. C.R.B., G.C. and Y.C. conducted the extractions and mass spectral analysis of single leaves. B.P. and Y.-P.H. conducted the laboratory-scale extractions and subsequent mass spectral analysis. C.M. conducted the microbial growth assays. M.Y. conducted the TEA. All authors contributed to writing the paper.

## Competing interests

D.A.M., D.B. and C.B.L. are cofounders of Infinit Health, a company focused on probiotic-based manipulation of the infant gut microbiota. D.A.M. and C.B.L. are cofounders of One.Bio, a company advancing new bioactive glycans. None of these companies had any role in the conceptualization, design, analysis or preparation of this manuscript. The other authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43016-024-00996-x>.

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

**Peer review information** *Nature Food* thanks Hao Du, Ted Johanson and Zhenhua Liu for their contribution to the peer review of this work.

**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) 2024

## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Mass spectrometry data was collected using Xcalibur 4.1.31.9 and MassHunter B.07.00. The microplate reader for growth assays was controlled using MARS.

Data analysis Mass spectrometry data was analyzed using a combination of MS-DIAL 4.90 and MassHunter B.07.00. Technoeconomic analysis was conducted and analyzed using SuperPro Designer v12. Bacterial growth was analyzed GraphPad Prism 10.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Raw mass spectrometry data will be deposited in a public repository.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="n/a"/>
Population characteristics	<input type="text" value="n/a"/>
Recruitment	<input type="text" value="n/a"/>
Ethics oversight	<input type="text" value="n/a"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No tests were conducted to determine sample size. Sample size was chosen to determine reproducibility based on experience of variability in these assays. For transient expression and bacterial growth experiments, at least three biological replicates were used. For 5 g purifications, extractions were done in duplicate, and each duplicate was measured with two technical replicates. Analysis of stably transformed lines measured metabolite quantity of three leaves on a single transgenic plant. This was the most appropriate way to measure T0 transgenic lines as they only have one plant available.
Data exclusions	No data was excluded for the analysis.
Replication	All transient expression experiments were done in biological triplicate to ensure that different plants still create the same metabolites. Where applicable, these independent measurements were used to generate averages and assess error. All biological replicates showed similar outcomes, indicating the data is reproducible. Furthermore, mass spectral analysis was conducted in multiple labs, with the data being reaffirmed by the analysis conducted in each lab by different personnel. For 5 g purification experiments, two extractions were performed side-by-side, which showed similar results, indicating the data was reproducible. Bacterial growth assays were completed in biological triplicate to ensure data was reproducible between replications.
Randomization	No randomization was done for these experiments. Randomization does not apply because experimental data are measurements of metabolite presence and quantity.
Blinding	There was no blinding as this was not applicable for our experiments. Blinding does not apply because experiment data are measurements of metabolite presence and quantity.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

- n/a  Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern

### Methods

- n/a  Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging