

UC Davis

UC Davis Electronic Theses and Dissertations

Title

Dietary Carbohydrate Analysis by Rapid-Throughput Liquid Chromatography-Mass Spectrometry Methods

Permalink

<https://escholarship.org/uc/item/82d7r4gz>

Author

Couture, Garret Anders

Publication Date

2022

Peer reviewed|Thesis/dissertation

Dietary Carbohydrate Analysis by Rapid-Throughput Liquid Chromatography-Mass Spectrometry Methods

By

GARRET COUTURE
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Chemistry

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Carlito B. Lebrilla, Chair

Matthew Augustine

Selina Wang

Committee in Charge

2022

DEDICATION

I dedicate this thesis to my mother, Denise Couture, and to my cousin, Amy Holt. Thank you for all that you have given me in love, support, and cherished memories.



ACKNOWLEDGEMENTS

I would first like to thank my mother, Denise Couture, for raising me on her own, for working so incredibly hard to provide me with anything I could ever want or need, for never giving up on me, and for teaching me so much about life and the pursuit of happiness. I would like to thank my cousin, Amy Holt, for creating some of my most cherished childhood memories and for always being there when it mattered most. I would like to thank my close friends, Jeffrey Le and Kellin Evans, who have been brothers to me for as long as I can remember and have always been a wellspring of love and support. I would like to thank Jeanine Leoncio-Bacani for encouraging me to pursue my passion, for her all her love and support, and for sticking by me through even the darkest of times. I would like to acknowledge the city I grew up in, Houston, Texas, for allowing me to see and experience the diversity of human life and for being the backdrop to my origin story. I would also like to thank all my friends and family that have come into and been a part of my life for shaping me into who I am today.

I would like to thank all the teachers and mentors I've had through the years. I would like to thank my 3rd grade teacher, Mrs. Reyes, for encouraging me and showing me the fun in learning. I would like to thank my high school principal, Dr. James McSwain, for providing me with a second chance. I would like to thank Raymond Gayle for his mentorship and for believing in me when no one else did. I would like to thank my organic chemistry professor, Dr. Steven Dessens, for instilling in me a passion for chemistry. I would also like to thank my undergraduate research advisor, Dr. Loi Do, for his mentorship inside and outside the laboratory, without which I would never have applied for graduate school.

I would like to thank my graduate research advisor, Dr. Carlito B. Lebrilla, for the opportunity to join his lab. I have grown, experienced, and learned so much about science and about life under his tutelage. None of this would have been possible without his mentorship, understanding, and support. There were moments where time was at a standstill during my tenure in the lab, but when I look back at it all, everything seems a blur and I want nothing more than to relive it exactly as it was. I wish there was a way to know you were in the good old days while you were still actually in them.

I must also thank all of my colleagues and lab mates that have helped me in so many ways throughout my graduate career. I thank previous lab members Dr. Matthew Amicucci, Dr. Eshani Nandita, Dr. Ace Galermo, Thai-Thanh Vo, and Diane Tu for introducing me to the lab and helping me in the beginning. I would like to give special thanks to my friend and colleague, Dr. Juan Jose Castillo, for his mentorship, leadership, and for teaching me so much about the value of hard work and determination. I would also like to thank Nikita (Nikko) Bacalzo, Ye (Winnie) Chen, Cheng-Yu (Charlie) Weng, Aaron Stacy, Yean (Shawn) Cheang, Chris Suarez, Sophia Jiang, Jessica Chao, Matthew Sujanto, Akshaya Karthikeyan, and Nicholas Cheng for their help and hard work on the projects included in this dissertation. Lastly, I thank the entirety of the Lebrilla League that I have met and worked with: Dr. Axe Xie, Dr. Jennyfer Tena, Dr. Qing (Dave) Zhou, Dr. Ying Sheng, Christopher Ranque, Siyu (Cathy) Chen, Anita Vinjamuri, Ryan Schindler, Armin Oloumi, Yasmine Bouchibti, and Xavier Holmes. It has truly been an honor to meet them all and work alongside them.

Last but not least, I would like to thank my dissertation committee members Dr. Matthew Augustine and Dr. Selina Wang for their help and for making my qualifying exam a positive experience.

Dietary Carbohydrate Analysis by Rapid-Throughput Liquid Chromatography-Mass Spectrometry Methods

ABSTRACT

Carbohydrates comprise the largest fraction of most diets and play an integral role in human health. They exhibit immense structural diversity and have important biological functions that are dictated by these structures. However, current methods for the analysis of dietary carbohydrates belie these complex structure-function relationships. Typically, total carbohydrates are not even measured and those components that are measured—like dietary fiber and sugars, rely on crude gravimetric or liquid chromatography analyses that provide limited information. While methods based on gas chromatography-mass spectrometry (GC-MS) are commonly used for structural analysis of isolated plant cell walls, their limited throughput, sensitivity, and selectivity have prevented their widespread use for food analysis. This dissertation describes the development and application of liquid chromatography-tandem mass spectrometry (LC-MS/MS) based methods for food carbohydrate quantification and structural elucidation that directly address the inherent limitations of the GC-MS approach. Chapter 1 provides background on food carbohydrate structures and their importance in the context of the gut microbiome and host health. Chapter 2 provides a detailed protocol describing the details of these recently developed LC-MS/MS methods for the analysis of food and fecal biospecimens. Chapter 3 presents the application of these methods to create a detailed glycomic map of the maize plant that provides insight towards greater utilization of the entire plant. Chapter 4 describes the application of a rapid-throughput and quantitative monosaccharide analysis to determine the total monosaccharide compositions of over 800 foods to develop an open-access food carbohydrate database. Chapter 5 presents a novel workflow that applies a multi-glycomic analysis to isolated

fiber fractions from existing methods for the determination of dietary fiber. The work presented in this dissertation highlights the importance and utility of these methods towards understanding how dietary carbohydrate structure impacts health through interaction with the gut microbiome.

PUBLICATIONS

Couture, G., Luthria, D. L., Chen, Y., Bacalzo, N. P., Tareq, F. S., Harnly, J., Phillips, K. M., Pehrsson, P. R., McKillop, K., Fukagawa, N. K., and Lebrilla, C. B. Multi-glycomic characterization of fiber from AOAC methods defines the carbohydrate structures. *Journal of Agricultural and Food Chemistry* **2022**.

Couture, G., Castillo, J. J., Bacalzo, N. P., Chen, Y., Chin, E. L., Blecksmith, S. E., Bouzid, Y. Y., Vainberg, Y., Masarweh, C., Zhou, Q., Smilowitz, J. T., German, J. B., Mills, D. A., Lemay, D. G., and Lebrilla, C. B. The development of the Davis Food Glycopedia—a glycan encyclopedia of food. *Nutrients* **2022**, 14, 1639–1656.

Delannoy-Bruno, O., Desai, C., Castillo, J. J., Couture, G., Barve, R. A., Lombard, V., Henrissat, B., Cheng, J., Han, N., Hayashi, D. K., Meynier, A., Vinoy, S., Lebrilla, C. B., Marion, S., Heath, A. C., Barratt, M. J., and Gordon, J. I. An approach for evaluating the effects of dietary fiber polysaccharides on the human gut microbiome and plasma proteome. *Proceedings of the National Academy of Sciences* **2022**, 119.

Couture, G., Vo, T.-T. T., Castillo, J. J., Mills, D. A., German, J. B., Maverakis, E., and Lebrilla, C. B. Glycomic mapping of the maize plant points to greater utilization of the entire plant. *ACS Food Science & Technology*. **2021**, 1, 2117–2126.

Castillo, J. J.; Galermo, A. G.; Amicucci, M. J.; Nandita, E.; Couture, G.; Bacalzo, N.; Chen, Y.; Lebrilla, C. B. A multidimensional mass spectrometry-based workflow for de novo structural elucidation of oligosaccharides from polysaccharides. *Journal of the American Society for Mass Spectrometry*. **2021**, 32 (8), 2175-2185.

Phillips, K. M., McGinty, R. C., Couture, G., Pehrsson, P. R., McKillop, K., and Fukagawa, N. K. Dietary fiber, starch, and sugars in bananas at different stages of ripeness in the retail market. *PLOS One* **2021**, 16.

Delannoy-Bruno, O.; Desai, C.; Raman, A.S.; Chen, R.Y.; Hibberd, M.C.; Cheng, J.; Han, N.; Castillo, J.J.; Couture, G.; Lebrilla, C.B.; Brave, R.A.; Lombard, V.; Henrissat, B.; Leyn, S.A.; Rodionov, D.A.; Osterman, A.L.; Hayashi D.K.; Meynier, A.; Vinoy, S.; Kirbach, K.; Wilmot, T.; Heath, A.C.; Klein, S.; Barret, M.J.; Gordon, J.I. Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature*. **2021**, 595, 91–95.

Patnode, M. L.; Guruge, J. L.; Castillo, J. J.; Couture, G. A.; Lombard, V.; Terrapon, N.; Henrissat, B.; Lebrilla, C. B.; Gordon, J. I. Strain-level functional variation in the human gut microbiota based on bacterial binding to artificial food particles. *Cell Host & Microbe*. **2021**, 29 (4), 664-673.

TABLE OF CONTENTS

Chapter 1: An Introduction to Food Carbohydrates, Their Interactions with the Gut

Microbiome, and Their Analysis.....	1
Overview.....	2
Carbohydrate Structure.....	3
Plant Carbohydrates and Food.....	8
Dietary Fiber, The Gut Microbiome, and Host Health.....	13
Food Carbohydrate Analysis.....	19
Conclusion.....	25
References.....	26

Chapter 2: A Multi-Glycomic Platform for the Analysis of Food Carbohydrates.....33

Abstract.....	34
Introduction.....	35
Development of the Protocol.....	36
Comparisons to Alternative Methods.....	38
Experimental Design.....	41
Limitations of the Protocol.....	51
Future Applications.....	52
Materials.....	53

Procedure.....	65
Anticipated Results.....	91
Supplementary Materials.....	99
References.....	110

Chapter 3: Glycomic Mapping of the Maize Plant Points to Greater Utilization of the Entire Plant.....114

Abstract.....	115
Introduction.....	116
Materials and Methods.....	118
Results and Discussion.....	123
Supplementary Materials.....	140
References.....	147

Chapter 4: The Development of the Davis Food Glycopedia—A Glycan Encyclopedia of Food.....150

Abstract.....	151
Introduction.....	152
Materials and Methods.....	154
Results.....	159
Discussion.....	168

Conclusion.....	176
Supplementary Materials.....	178
References.....	183

Chapter 5: Multi-glycomic characterization of fiber from AOAC methods defines the carbohydrate structures.....185

Abstract.....	186
Introduction.....	187
Materials and Methods.....	190
Results.....	196
Discussion.....	207
Conclusion.....	212
Supplementary Materials.....	214
References.....	222

Chapter 1

An Introduction to Food Carbohydrates, Their Interactions with the Gut Microbiome, and Their Analysis

OVERVIEW

Our diets play an integral role in our overall health. Considerable effort has been made towards identifying and quantifying the thousands of micronutrients and small molecules in the foods we eat, but these compounds represent only a small fraction of the total picture. By far the largest components of food are macronutrients: proteins, lipids, and carbohydrates. However, our understanding of the structures of these major components is severely limited. This is due to a lack of appropriate analytical methods for their elucidation and quantification. Carbohydrates are perhaps these least understood macronutrient as total carbohydrates are not even typically measured. Instead, they are calculated by subtracting a food's moisture, protein, fat, and ash from the total mass. Analytical methods for the components of total carbohydrates that *are* measured—such as dietary fiber, are also crude, outdated, and provide little to no structural information. Recent efforts towards understanding how dietary fiber impacts host health through its interaction with the gut microbiome has necessitated the advent of analytical methods capable of providing quantitative structural information.

This chapter first provides a primer on general carbohydrate structure as well as those structures found in the foods we consume. The interaction between these structures, the gut microbiome, its metabolomic output, and the impact on host health is then summarized. Lastly, the analytical approaches previously and currently used for food carbohydrate analysis are described. Together, these sections provide background and frame the work presented in the subsequent chapters.

CARBOHYDRATE STRUCTURE

Monosaccharides

Carbohydrates are naturally abundant chemical compounds comprised of carbon, oxygen, hydrogen, and in some cases, nitrogen. Just as the building blocks of proteins and nucleic acids are amino acids and nucleotides, respectively, monosaccharides are the monomeric building blocks of carbohydrates. All monosaccharides consist of several chiral hydroxymethylene groups terminated by an alcohol group on one end and either an aldehyde or ketone on the other end. In this way, monosaccharides can be broadly classified as “aldoses” or “ketoses,” respectively.¹ While all monosaccharides contain 3-9 carbon atoms, only those with 5, 6, or 9 carbons are commonly observed constituents of plant and animal carbohydrates.

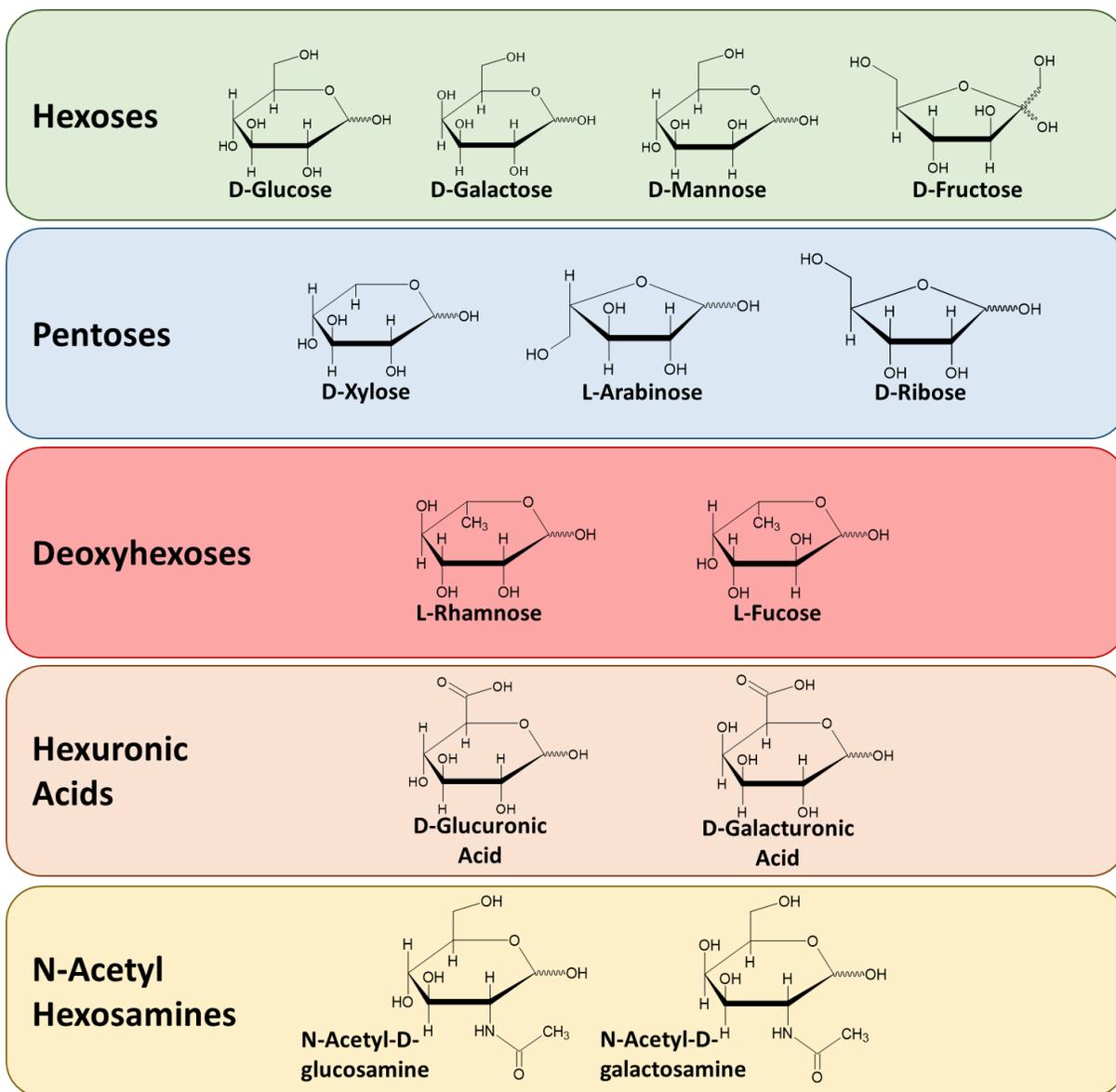


Figure 1.1 Structures of the 13 most abundant monosaccharides found in plants and food.

Commonly observed monosaccharides in plants and food include glucose, fructose, galactose, mannose, xylose, arabinose, ribose, rhamnose, fucose, glucuronic acid, galacturonic acid, N-acetylglucosamine, and N-acetylgalactosamine.² **Figure 1.1** illustrates their structures while **Table 1.1** summarizes their common abbreviations and symbols. Monosaccharides can further be classified by the number of carbon atoms they contain. For example, a 6-carbon aldose like glucose is called an “aldohexose” while a 5-carbon aldose like xylose is called an “aldopentose.”

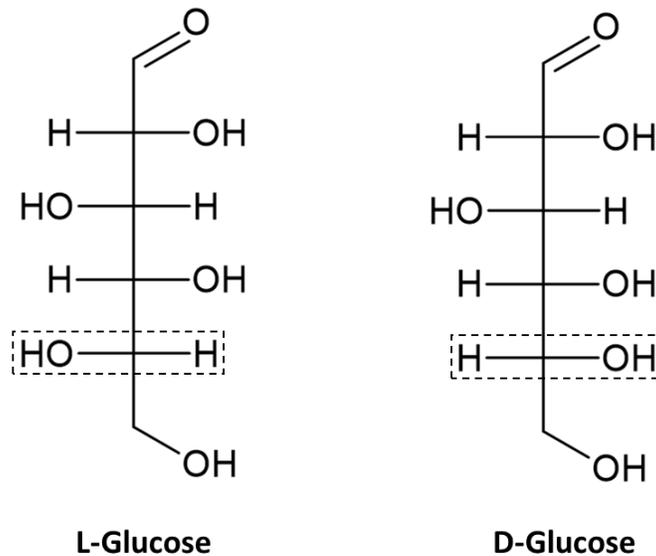


Figure 1.2 The enantiomers of glucose. The dashed boxes indicate the hydroxymethylene groups that determine the L- or D-configuration.

Because of the chiral nature of monosaccharides, several isomers can exist for each. Enantiomers (or mirror images) of the same monosaccharide are designated a L- or D- based on the configuration of the stereocenter furthest from the carbonyl carbon in its Fischer projection. If the -OH is on the left, the configuration is L-, if on the right it is D-.¹ (**Figure 1.2**) Many monosaccharides are naturally observed in the D- configuration (e.g. glucose, fructose, xylose, etc) while others are observed with their L- configuration (e.g. arabinose, rhamnose, and fucose). Monosaccharides that differ only in the configuration at one chiral center are called epimers.¹ For example, glucose and galactose are C4 epimers because they differ in the stereochemistry only at carbon number four.

Table 1.1 Common and abundant monosaccharides observed in plants and food with abbreviations and symbols.

Name	Abbreviation	Symbol	Name	Abbreviation	Symbol
D-glucose	Glc		L-rhamnose	Rha	
D-fructose	Fru		L-fucose	Fuc	
D-galactose	Gal		D-glucuronic acid	GlcA	
D-mannose	Man		D-galacturonic acid	GalA	
D-xylose	Xyl		N-acetyl-D-glucosamine	GlcNAc	
L-arabinose	Ara		N-acetyl-D-galactosamine	GalNAc	
D-ribose	Rib				

Although depicted in cyclic form in **Figure 1.1**, monosaccharides in solution exist in equilibrium with their acyclic forms and interchange rapidly. Where this equilibrium lies depends largely on the structure of the monosaccharide. For hexoses and pentoses, 5- and 6-membered rings tend to be the most chemically stable due to bond angles and sterics. These 5- and 6-membered rings resemble the structures of furans and pyrans, and are thus named furanoses and pyranoses, respectively.¹ Hexoses are most commonly observed in pyranose form (with the exception of fructose) while pentoses are commonly found in both forms. For example, arabinose is depicted in **Figure 1.1** as a furanose and is referred to as arabinofuranose (or *f*-arabinose for short). However, arabinose may also exist as a pyranose called arabinopyranose (or *p*-arabinose).

Upon cyclization, the -OH group on C4 or C5 form a bond with the aldehyde or ketone carbon C1. When this occurs, a hemiacetal is formed and C1 becomes a new chiral center called

the anomeric carbon. This gives rise to two special isomers called anomers that are interchangeable in solution. When the absolute configuration of C1 and the stereocenter furthest from C1 (C4 in pentoses, C5 in hexoses) are the same, this is called the α anomer. When the configurations are different, it is the β anomer.¹

Disaccharides, Oligosaccharides, and Polysaccharides

When the hemiacetal at the anomeric carbon of one monosaccharide reacts with the hydroxy group of another monosaccharide to form an acetal, a glycosidic bond is formed. This bond formed at the anomeric carbon gives rise to α and β linkages. Further, glycosidic bonds can be formed between the anomeric carbon and any of the hydroxy groups on another monosaccharide.³ Thus, even a seemingly simple dimer of glucose has ten possible isomers because the bond connecting the two can be formed on C1, C2, C3, C4, or C6 with either α or β stereochemistry. Unlike amino acids and nucleotides, monosaccharides can also participate in multiple linkages resulting in branching, furthering the structural diversity of carbohydrate oligomers and polymers. Saccharides comprised of two monosaccharides are called disaccharides while saccharides with degrees of polymerization (DP) from ~3-20 are called oligosaccharides. Larger carbohydrate structures are called polysaccharides. Chains of monosaccharides have two termini: the reducing and non-reducing end. This nomenclature historically arises from the ability of the free aldehyde to be oxidized to a carboxylic acid. There is only one reducing end and it contains the free aldehyde or ketone. There may be several non-reducing termini depending on the degree of branching and each has its anomeric carbon locked into a glycosidic linkage. Therefore, oligo- and polysaccharides are synthesized by adding monosaccharides to the non-reducing end.³

Free monosaccharides—except for glucose and fructose, are not abundant in nature outside of their use as energy for living organisms. Rather, they more often exist as part of larger oligo- and polysaccharide structures which give rise to myriad structures and functions. For example, oligosaccharides are commonly added to proteins as a post-translational modification in a process called glycosylation. The structures of these oligosaccharides modulate protein function and are implicated in many cancers, autoimmune diseases, and infections.⁴⁻⁶ This, however, is an extensive field of study in and of itself and is beyond the scope of this dissertation. Oligosaccharides are also used to store chemical energy in plants and are an important component of mammalian milk.⁷⁻¹⁰ Polysaccharides are an immensely diverse class of molecules present in all forms of life. In plants, starch polysaccharides are used to store energy while the cell wall is comprised of many different polymers including cellulose, hemicelluloses, and pectins.^{11, 12}

PLANT CARBOHYDRATES AND FOOD

The plant material we consume as part of our diet is our largest source of carbohydrates. These structures span in size from simple monosaccharides all the way to polysaccharides containing hundreds of thousands of monomers. Within this spectrum of molecular weights, there is immense diversity in the monosaccharide and linkage compositions of oligo- and polysaccharide structures. Only monosaccharides can be absorbed directly in the GI tract. All di-, oligo-, and polysaccharides must first be digested to be utilized. However, humans only possess the ability to digest and subsequently utilize the disaccharides sucrose, lactose, and maltose; starch polysaccharides; and starch-derived oligosaccharides.¹³⁻¹⁵ The vast remainder of the carbohydrate structures we consume cannot be digested endogenously and instead serve as a

substrate for the trillions of microbes that inhabit our colons.¹⁶⁻¹⁸ These carbohydrates are mostly derived from the cell walls of plants we consume and are collectively referred to as “dietary fiber.”¹⁹⁻²¹

Digestible Carbohydrates

Simple sugars and starch make up the majority of the carbohydrates we consume by mass and provide a large fraction of our total energy intake next to fats and protein.²² Sucrose, fructose, and glucose are by far the most abundant sugars in plant-derived foods while lactose is prevalent in dairy. Monosaccharides such as glucose and fructose are absorbed directly into circulation in the small intestine while all larger structures like the disaccharides sucrose and lactose must first be hydrolyzed by sucrase and lactase, respectively.^{13, 14}

Starch is a collective term used to describe two very similar, but distinct polysaccharides amylose and amylopectin. Both are homopolymers of glucose, but amylose is a linear chain of $\alpha(1\rightarrow4)$ glucose while amylopectin exhibits $\alpha(1\rightarrow6)$ branching approximately every 20 monomers.²³ **Figure 1.3a** depicts the general structure of these polysaccharides. Starch is the only dietary polysaccharide humans are capable of digesting and using as energy. In plants, starch is used to store the carbohydrates produced by photosynthesis and is arranged into semi-crystalline granules in organelles called amyloplasts.¹¹ In edible plants, the majority of a plant’s starch reserves are found in the fruit, nut/seed, grain, rhizome, or tuber.²⁴ Some examples of high-starch foods include potato, rice, and pulses. The enzyme α -amylase is secreted by human salivary glands and begins to breakdown starch during mastication. The remaining digestion occurs in the small intestine via pancreatic amylase.¹⁵ The resulting glucose monomers are then absorbed. While starch itself is readily digestible, it’s physical location in the food or its conformation may prevent a fraction of it from being accessed and broken down by amylases.

This fraction--termed “resistant starch,” functions as a dietary fiber as it reaches the colon and can be consumed by the gut microbiota.^{25, 26}

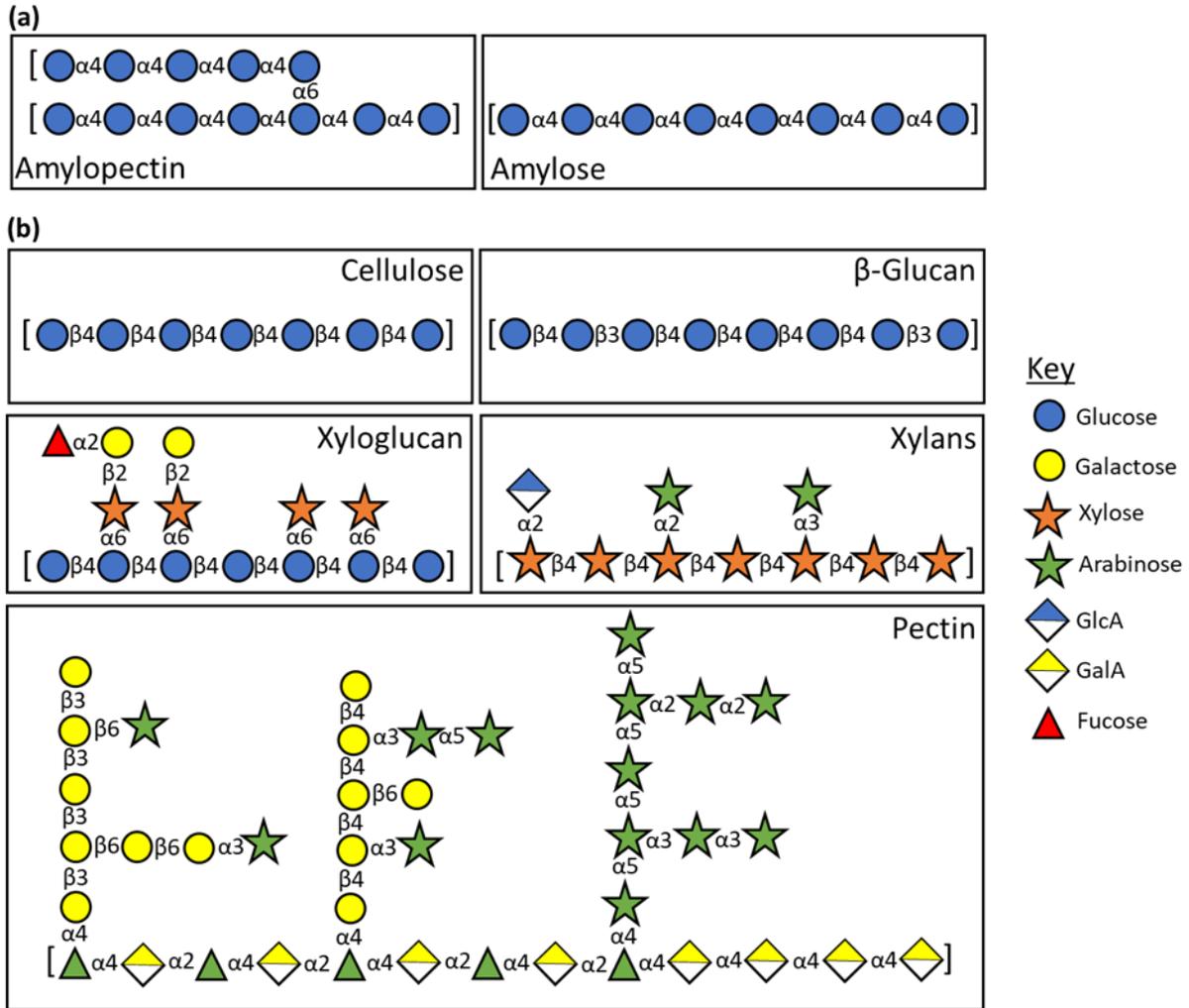


Figure 1.3a, b (a) General structure of the starch polysaccharides amylose and amylopectin. (b) Structures of various plant cell wall polysaccharides including cellulose, β -glucan, xyloglucan, xylans, and pectin.

Indigestible Carbohydrates

All dietary carbohydrates aside from those mentioned above are indigestible by humans. The complexity and diversity of these remaining structures far surpasses that of digestible carbohydrates in molecular weight, monosaccharide, and glycosidic linkage composition. Most

of these structures are plant cell wall polysaccharides which provide much of the dietary fiber in our diets. The plant cell wall is an intricate network of intertwined polysaccharides interspersed with cell wall proteins and lignin. **Figure 1.4** provides a representation of the cell wall of *Arabidopsis*.

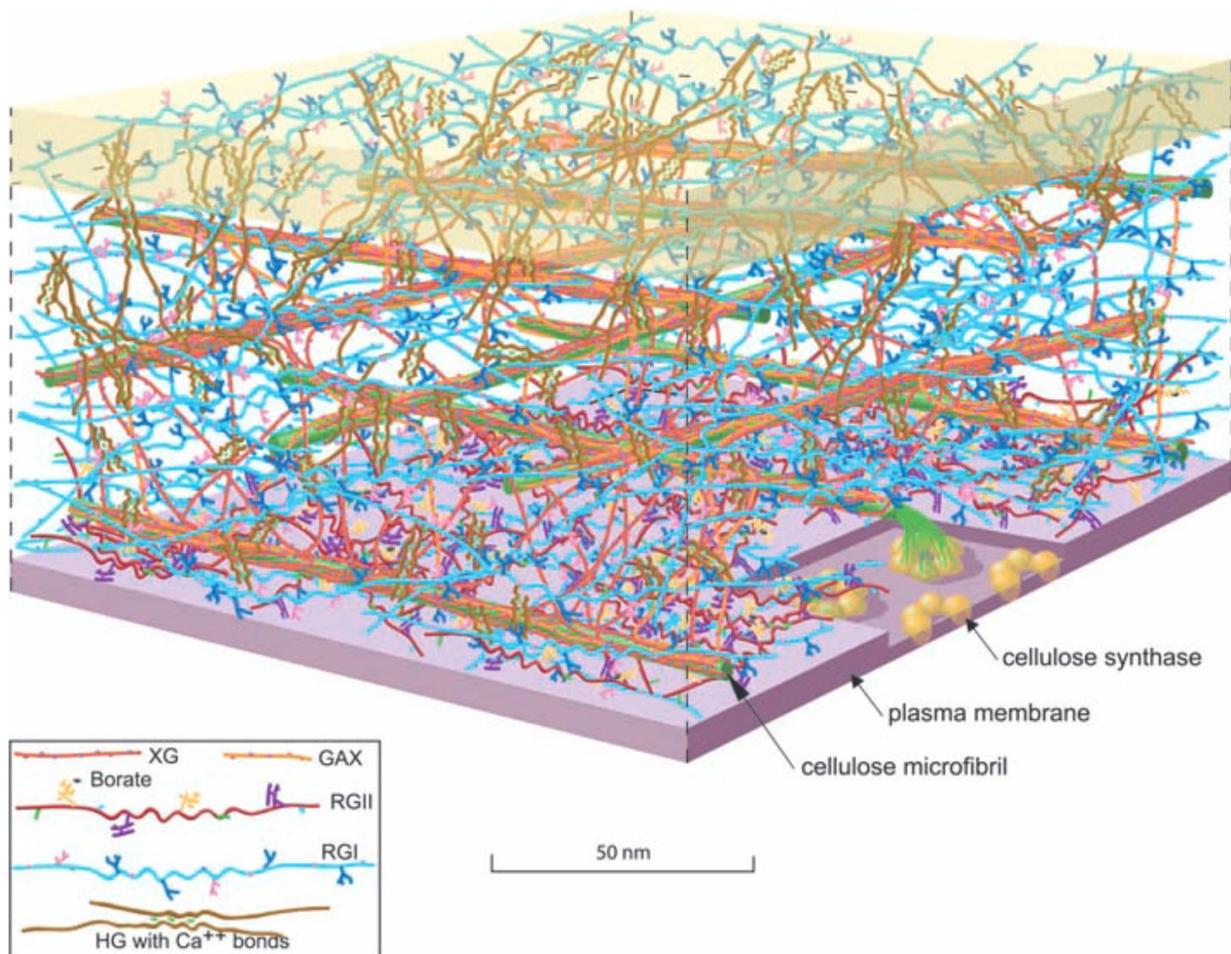


Figure 1.4 A diagram of the plant cell wall from *Arabidopsis*. Cellulose is represented as large, linear microfibrils with the hemicelluloses (xyloglucan [XG], glucuronoxylan [GAX]) and pectins (homogalacturonan [HG] and rhamnogalacturonan I/II [RGI/II]) intertwined between them. Reprinted from *Science*.²⁷

Cellulose is the primary cell wall polysaccharide, comprising a large fraction of the cell wall by dry weight, and is arranged into semi-crystalline microfibrils that add structural rigidity

to the plant cell.²⁸ Depending on the plant, tissue, and stage of development, hemicelluloses are the next most abundant polysaccharides in the cell wall. The term “hemicellulose” refers to a diverse group of polysaccharides composed of xylans (including arabinoxylan and glucuronoxylan), xyloglucan, β -mannans (including glucomannan and galactomannan), and mixed linkage β -glucan. Hemicelluloses are found non-covalently bound to the cellulose framework through hydrogen bonding. The nature of these interactions and the identity and fine structure of the hemicellulose play important roles in rigidity and cell elongation.^{12, 29} The last class of cell wall polysaccharides are called pectins and are also the most structurally diverse. Pectins can be broadly divided into three distinct polysaccharides: homogalacturonan, rhamnogalacturonan I, and rhamnogalacturonan. Homogalacturonan or polygalacturonic acid is a linear homopolymer of $\alpha(1\rightarrow4)$ GalA, but may also be substituted with xylose or apiose residues. Rhamnogalacturonan I possesses a backbone of the repeating disaccharide unit $[2)\text{-}\alpha\text{-Rha}(1\rightarrow4)\text{-}\alpha\text{-GalA}(1\rightarrow)]$. The rhamnose residues can be substituted at the 3- or 4- position with the polysaccharides galactan, arabinan, or arabinogalactan.^{30, 31} Pectic galactan is composed of a $\beta(1\rightarrow4)$ galactose backbone that may be substituted at the 6-position by galactose or arabinose residues in the case of arabinogalactan I.^{32, 33} Arabinan is a branched homopolymer of arabinose comprised of an $\alpha(1\rightarrow5)$ backbone branched at the 2- or 3-position with $\alpha(1\rightarrow2)$, $\alpha(1\rightarrow3)$, or $\alpha(1\rightarrow5)$ sidechains.^{34, 35} Rhamnogalacturonan II is arguably the most diverse polysaccharide known possessing a backbone of $\alpha(1\rightarrow4)$ GalA substituted at the 2- or 3-position with complex sidechains that may contain rhamnose, apiose, galactose, arabinose, xylose, fucose, and GalA linked to together by up to 21 different linkages.^{36, 37} Some of these structures are depicted in **Figure 1.3b**.

The presence and/or abundance of each cell wall polysaccharide depends largely on the plant species, the specific tissue within the same plant, and the stage of development.¹² Furthermore, the same polysaccharide from one plant may have a different structure than another. For example, by definition xyloglucan is composed of a $\beta(1 \rightarrow 4)$ glucose substituted with $\alpha(1 \rightarrow 6)$ xylose sidechains. However, in grasses some of the xylose residues may be further decorated with $\beta(1 \rightarrow 2)$ galactose while in some dicots this galactose may be further extended with an $\alpha(1 \rightarrow 2)$ fucose.^{38, 39}

DIETARY FIBER, THE GUT MICROBIOME, AND HOST HEALTH

The human gut is home to trillions of microorganisms that include bacteria, fungi, viruses, and protozoa with bacteria being the most abundant. Approximately 500-1000 species of bacteria inhabit the distal colon and collectively these microbes possess the millions of genes that comprise the gut microbiome, producing thousands of metabolites that modulate host health.⁴⁰⁻⁴⁴ There is a large amount of heterogeneity in the composition of the gut microbiota at the individual level that is influenced largely by age, geographical location, and diet.⁴⁵⁻⁴⁷ However, the gut microbiota is typically dominated by the Bacteroidetes, Firmicutes, Actinobacteria, Proteobacteria, and Verrucomicrobia phyla.^{45, 48} As commensal organisms, the gut microbiota rely largely on their host for sustenance, most of which comes from undigested and components of the host diet. This includes proteins, fats, carbohydrates, polyphenols, and various other small molecules that go undigested and unabsorbed by the upper GI tract. The largest fraction of these components—namely dietary fiber, escape digestion and arrive in the colon to serve as substrates

for the gut microbiota, thereby modulating their metabolic function and influencing host health.^{18, 49-54}

The constituent members of the gut microbiota dedicate a large portion of their genomes to the degradation of dietary and host glycans.^{17, 18, 55} These carbohydrate-active enzymes, or CAZymes, are classified into five functional groups: (1) glycoside hydrolases (GHs), (2) polysaccharide lyases (PLs), (3) carbohydrate esterases (CEs), (4) carbohydrate-binding modules (CBMs), and (5) glycosyltransferases (GTs). The functions of these groups, the number of families each group contains, and their percentage of the total CAZy database are summarized in **Table 1.2**. CAZymes are then further divided into families and subfamilies according to their reported or predicted functions.⁵⁵⁻⁵⁸ GTs are used for the biosynthesis of larger oligo- and polysaccharides from nucleotide sugars. The remaining groups are used in concert to catabolize host and dietary glycans for their ultimate use in fermentation.^{17, 55}

Table 1.2 CAZyme classes, their functions, the number of their constituent families, and their percentage of the total CAZy database.

CAZyme Class	Number of Families	Percentage of Total CAZymes	Function
Glycoside Hydrolases (GHs)	113	47%	Hydrolysis of glycosidic bonds
Polysaccharide Lyases (PLs)	19	1.5%	Cleavage of glycosidic bonds adjacent to uronic acids by β -elimination
Carbohydrate Esterases (CEs)	15	5%	Removal of ester modifications from uronic acids
Carbohydrate-Binding Modules (CBMs)	52	7%	Facilitate binding of catalytic domains to glycan substrates
Glycosyltransferases (GTs)	90	41%	Biosynthesis of glycans from phosphor-activated sugar donors

Microbes differ greatly in the genes they possess for carbohydrate utilization from the phylum all the way down to the strain level. For example, the genus *Bacteroides* tend to possess CAZymes for the degradation of many different plant- and host-derived oligo- and polysaccharides while *Bifidobacterium* tend to be more specialized towards oligosaccharides.^{16, 59, 60} Further, *Bacteroides thetaiotaomicron* is a prolific eater capable of utilizing a variety of plant polysaccharides (pectins, hemicelluloses, and starch) as well as host-derived glycans such as glycosaminoglycans and mucins while *Bacteroides ovatus* are more specialized towards the utilization of plant-derived glycans.⁶¹⁻⁶⁴ The microbiota collectively employ many different GHs and PLs the breakdown of dietary glycans. While there is some promiscuity, the majority of these CAZymes are specific at the monosaccharide, linkage, and spatial level necessitating an entire suite of enzymes for the systematic deconstruction of incoming glycans. For example, one of the most voracious and versatile saccharolytic phyla, Bacteroidetes, arranges these genes into polysaccharide utilization loci (PULs) that contain all the requisite GHs and PLs necessary for degrading specific polysaccharides.^{65, 66} **Figure 1.5** depicts several common fiber polysaccharides, their constituent glycosidic linkages, and several of the CAZymes used to break them down.

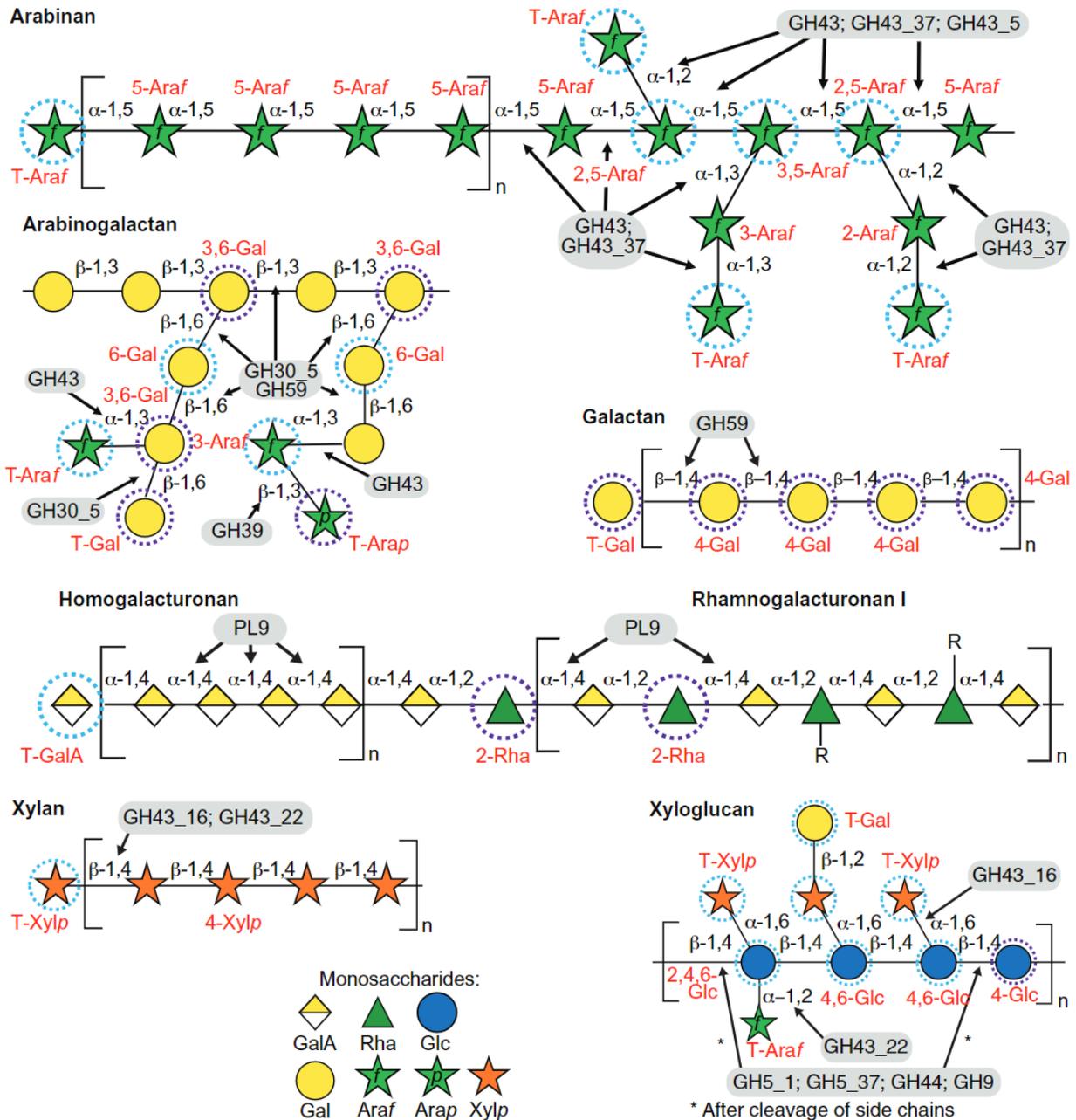


Figure 1.5 Structures of several common fiber polysaccharides that are abundant in food and some CAZymes (GH or PL) used to break them down. The designations in red font indicate linkages that can be monitored by LC-MS analysis. Adapted from *PNAS*.⁶⁷

Once complex dietary glycans are broken down and internalized by the gut microbiota, the liberated monosaccharides are converted into energy through the process of fermentation.¹⁸

49, 68-72 The primary end products of this process are gases (hydrogen, carbon dioxide, methane), alcohols (propanediol, ethanol, propanol), and short-chain fatty acids or SCFAs (acetate, propionate, butyrate, lactate, succinate). The major products produced depends on the organism and the substrate it is utilizing. However, SCFAs are the most prominent metabolite produced.¹⁸

49, 68, 73 **Figure 1.6** illustrates a summary of these pathways and **Table 1.3** provides the pathways utilized by some common gut microbes.

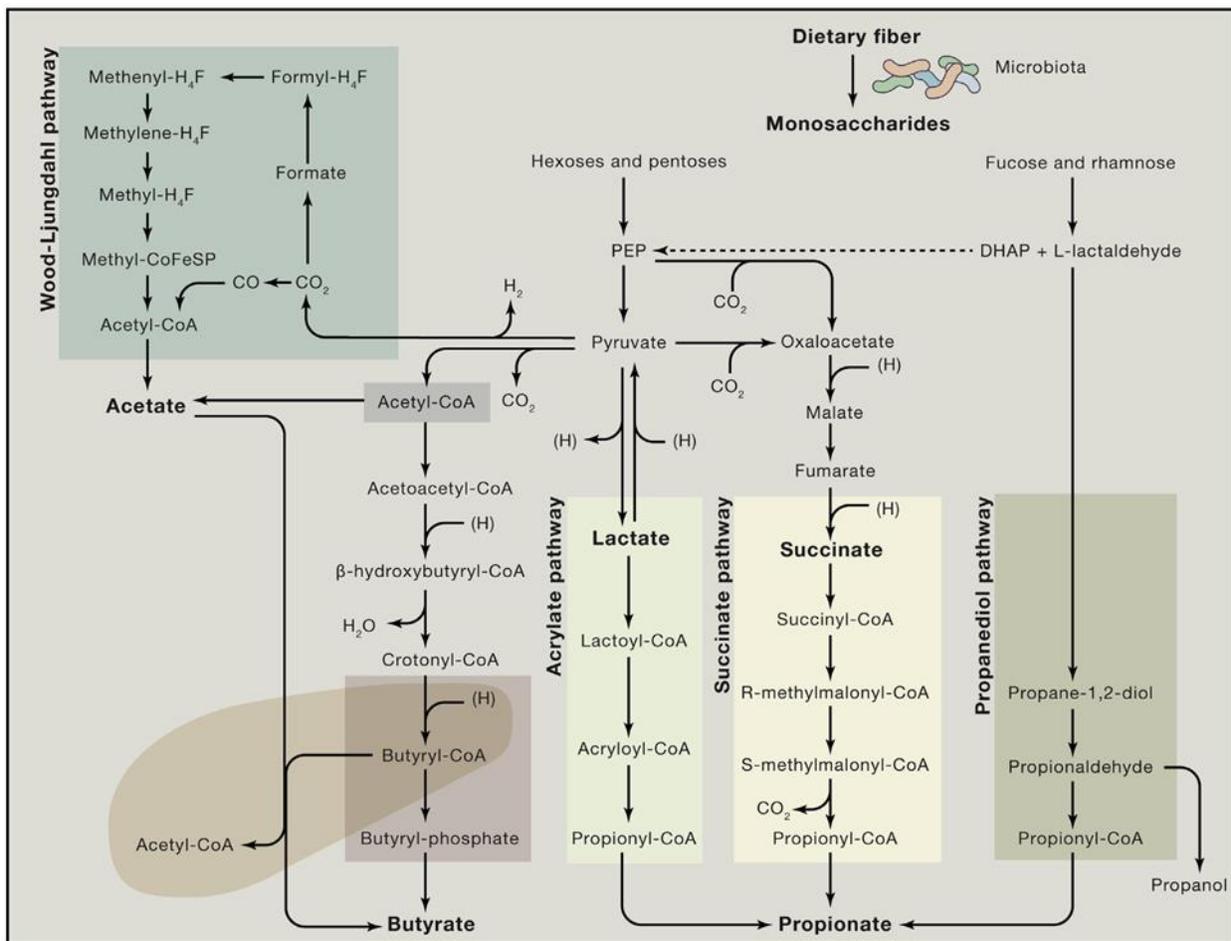


Figure 1.6 Known pathways for the biosynthesis of SCFAs by the gut microbiota from carbohydrate fermentation. Reprinted from *Cell*.⁴⁹

SCFAs also play a pivotal role in shaping gut microbial ecology. Diets high in fiber have been shown to increase the production of SCFAs relative to low fiber diets thus lowering the pH of the

gut and discouraging the growth of pathogenic microbes.⁷⁴⁻⁷⁶ Additionally, many microbial residents of the gut are able to utilize organic acids produced by their neighbors for energy, creating crosstalk between commensal species.⁷⁷

Table 1.3 SCFA Production by Microbes in the Gut. Adapted from *Cell*.⁴⁹

SCFAs	Pathways/Reactions	Producers
Acetate	from pyruvate via acetyl-CoA	most of the enteric bacteria, e.g., <i>Akkermansia muciniphila</i> , <i>Bacteroides</i> spp., <i>Bifidobacterium</i> spp., <i>Prevotella</i> spp., <i>Ruminococcus</i> spp.
	Wood-Ljungdahl pathway	<i>Blautia hydrogenotrophica</i> , <i>Clostridium</i> spp., <i>Streptococcus</i> spp.
Propionate	succinate pathway	<i>Bacteroides</i> spp., <i>Phascolarctobacterium succinatutens</i> , <i>Dialister</i> spp., <i>Veillonella</i> spp.
	acrylate pathway	<i>Megasphaera elsdenii</i> , <i>Coprococcus catus</i>
	propanediol pathway	<i>Salmonella</i> spp., <i>Roseburia inulinivorans</i> , <i>Ruminococcus obeum</i>
Butyrate	phosphotransbutyrylase/butyrate kinase route	<i>Coprococcus comes</i> , <i>Coprococcus eutactus</i>
	butyryl-CoA:acetate CoA-transferase route	<i>Anaerostipes</i> spp. (A, L), <i>Coprococcus catus</i> (A), <i>Eubacterium rectale</i> (A), <i>Eubacterium hallii</i> (A, L), <i>Faecalibacterium prausnitzii</i> (A), <i>Roseburia</i> spp. (A)

A, acetate is the substrate for producing butyrate; L, lactate is the substrate for producing butyrate.

SCFAs also strongly influence both gut epithelial health and systemic health through various pathways. Acetate, propionate, and especially butyrate are the preferred energy sources for colonocytes lining the intestinal epithelium.^{78, 79} SCFAs have also been shown to decrease epithelial permeability and stimulate the production of epithelial mucins thereby providing protection from pathogenic bacteria and local inflammation implicated in inflammatory bowel diseases.⁷⁹⁻⁸¹ Once absorbed and entered into circulation, SCFAs are largely metabolized in the liver and other peripheral tissues for the synthesis of glucose, cholesterol, long-chain fatty acids,

and glutamate.^{73, 82} Additionally, SCFAs are known to act on G-protein coupled receptors on immune cells and inhibiting histone deacetylases (HDACs) thereby reducing systemic inflammation and altering gene expression.^{79, 83, 84}

FOOD CARBOHYDRATE ANALYSIS

Classical Methods

Methods for the analysis of carbohydrates have existed since the 19th century. The earliest assays were based on the reduction of copper(II) to copper(I) by reducing sugars under strongly basic conditions (Fehling test). Since then, many methods have been developed and used including thin-layer chromatography (TLC), gas chromatography (GC) equipped with either a flame ionization detector (FID) or a mass spectrometer (MS), high performance anion exchange chromatography with pulsed amperometric detection (HPAEC-PAD), high performance liquid chromatography (HPLC), and nuclear magnetic resonance (NMR).⁸⁵⁻⁸⁹ GC-MS has become perhaps the most popular method for plant polysaccharide analysis and-- although it is obviously advanced greatly since the 19th century, instrumental methods have gone relatively unchanged for the last several decades.

In the analysis of food carbohydrates, HPLC with refractive index (RI) detector is commonly employed for the quantification of sugars and some oligosaccharides while UV-vis is employed for the quantification of starch following enzymatic digestion to glucose.⁹⁰ However, these methods are quite limited and have not been generally applied to dietary fiber. Currently, dietary fiber is measured using an enzymatic-gravimetric approach wherein proteins and starch are digested and removed and the resulting digestate is filtered, dried, and weighed to determine

the soluble, insoluble, and total fiber content.⁹¹ **Figure 1.7** provides the outline of AOAC 991.43, a commonly employed method for total dietary analysis.

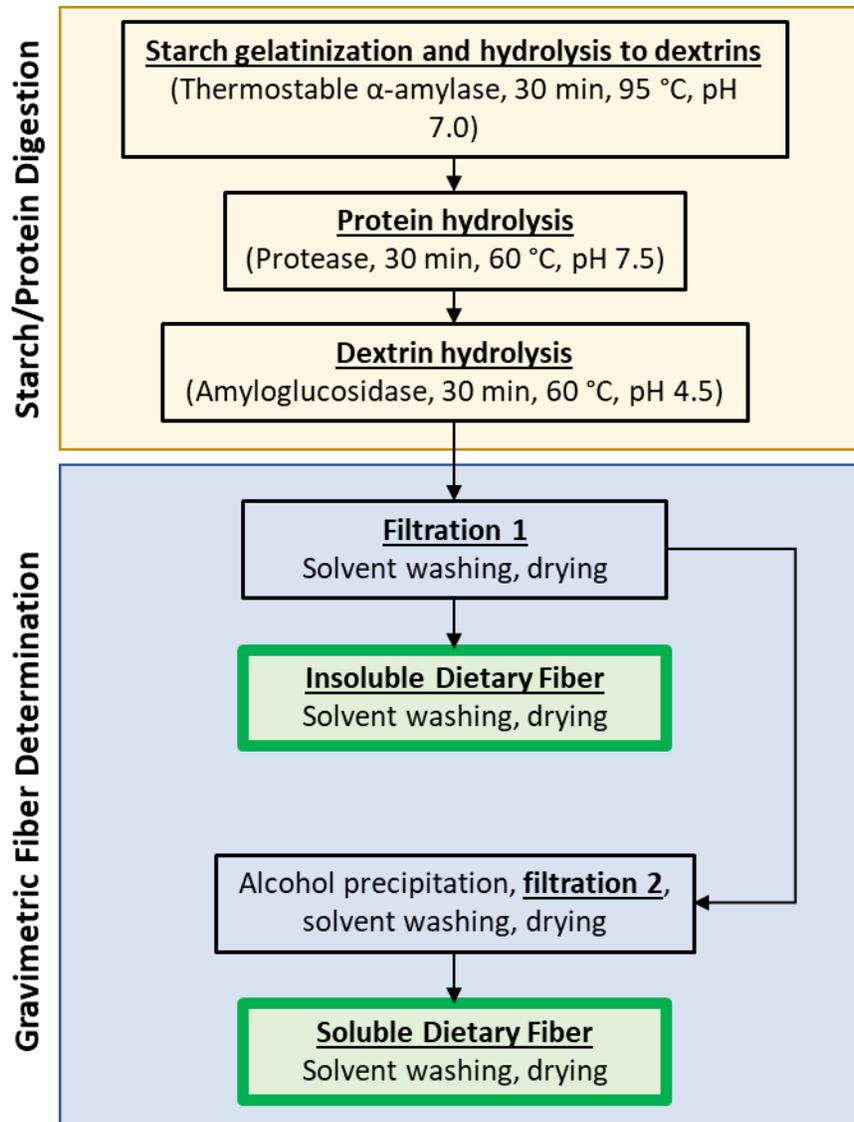


Figure 1.7 Overview of AOAC 991.43 for the determination of insoluble and soluble dietary fiber. These values are summed to calculate the total dietary fiber.

These methods are low-throughput and provide no structural information on the oligo- and polysaccharide components. Because the composition and metabolic output of the gut

microbiome is so heavily influenced by the specific structures of dietary fiber, analytical methods to quantify and elucidate these structures are needed to understand these interactions.

Rapid Throughput LC-MS Methods

LC-MS has recently gained attention as a powerful alternative to GC-MS for carbohydrate analysis. Mono-, di-, and oligosaccharides are commonly separated natively using hydrophilic interaction chromatography (HILIC) or porous graphitized carbon (PGC) and analyzed via quadrupole, time of flight (TOF), or tandem mass spectrometers.⁹²⁻⁹⁵ However, polysaccharides can be prohibitively large, reaching millions of Daltons in molecular weight. While size exclusion chromatography (SEC) can be sometimes be used to separate whole polysaccharides, this approach is severely limited by the general insolubility of polysaccharides, the resolution of the separation, poor ionization efficiency, and the limited structural information obtained.^{96, 97} Instead, the most common and informative approaches involve breaking these large polysaccharide structures down into their monomeric units with some chemical modification. These strategies can be used to provide quantitative monosaccharide compositions and the glycosidic linkage composition of the parent polysaccharides.⁹⁸ This has typically been accomplished using a GC-MS based approach which has several major limitations. The sample preparation is on a per-sample basis and the instrumental analysis requires long chromatographic run times and has limited sensitivity.⁹⁸⁻¹⁰⁰ The recent development of LC-MS based methods has directly addressed these limitations.

One recently developed approach for determining the monosaccharide composition of food polysaccharides involves the quantitative hydrolysis of the glycosidic bonds with trifluoroacetic acid (TFA) at elevated temperatures and subsequent derivatization with 1-phenyl-3-methyl-5-pyrazolone (PMP), allowing for their ultra-high performance liquid chromatography

(UHPLC) separation using a C18 column while simultaneously boosting their ionization and fragmentation efficiencies for analysis on a triple quadrupole mass spectrometer (QqQ MS).

Figure 1.8 outlines the sample preparation and provides an example of a chromatogram showing the separation of 14 monosaccharides in an 11-minute dynamic multiple reaction monitoring (dMRM) run. This method is adaptable to a 96-well plate format and is significantly higher-throughput than the GC-MS approach. Further, the instrument method provides a linear range over six orders of magnitude and limits of detection down to the femtomole level. External calibration with a serially diluted pool of monosaccharide standards allows for the absolute quantitation of carbohydrates in a sample.^{99, 100}

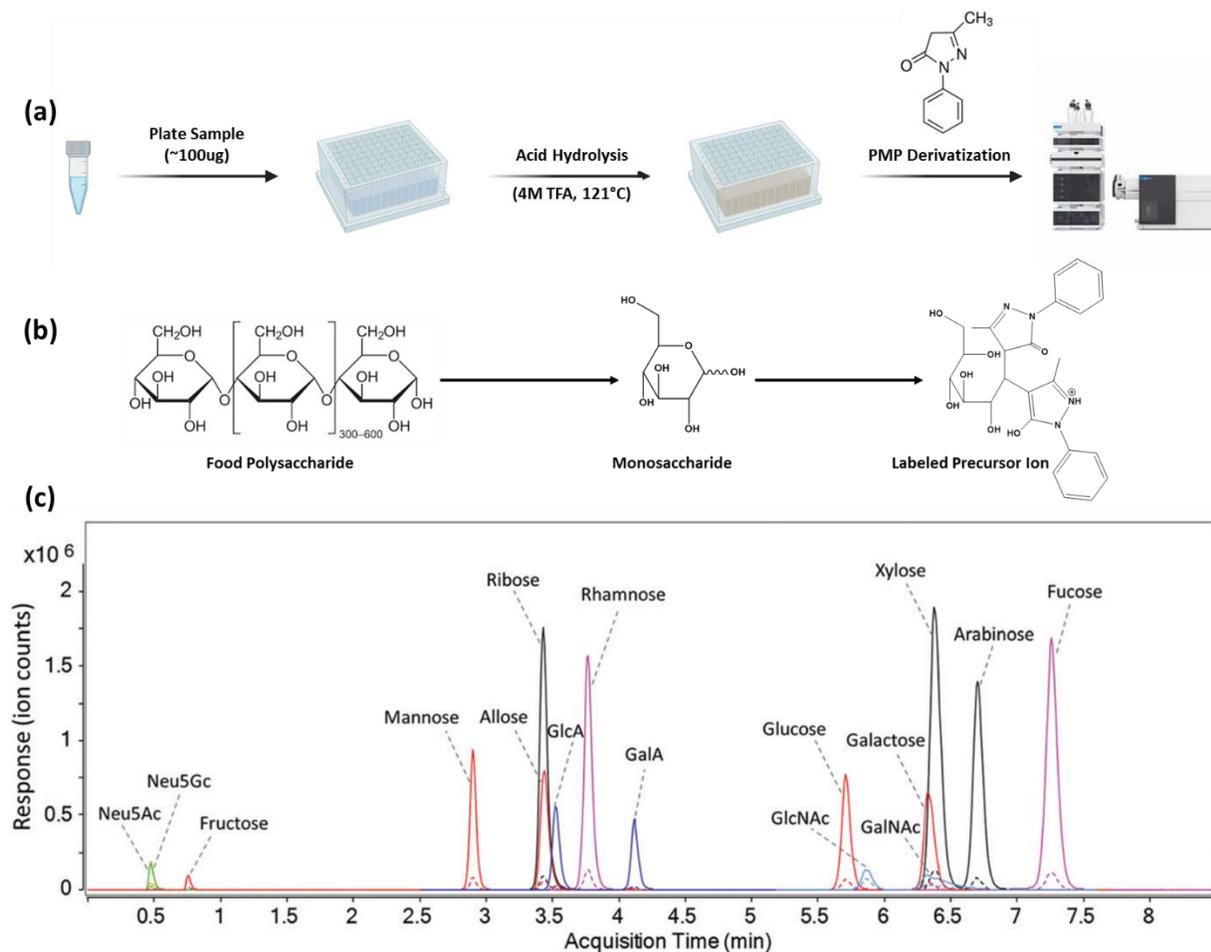


Figure 1.8a-c (a) Overview of the sample preparation steps for monosaccharide analysis. (b) Polysaccharides are hydrolyzed to their constituent monosaccharides and derivatized with PMP to give the labeled precursor ion shown. (c) Example MRM chromatogram of a pool containing the 16 monosaccharides that can be quantified with the method.

While monosaccharide analysis provides the quantitative monosaccharide composition, the glycosidic linkage structure is lost upon hydrolysis of the sample. To determine how the monosaccharides were linked and to infer the polysaccharide structures present, a separate linkage analysis is required. Previously, this was done by GC-MS methods that required extensive preparation on a per-sample basis and long chromatographic run times.^{88,98} However, a recently developed LC-MS approach has enabled the monitoring of over 100 glycosidic linkages within a 16-minute method utilizing similar sample preparation procedures as the monosaccharide analysis. Samples are first permethylated in DMSO using saturated sodium hydroxide (NaOH) and iodomethane (ICH₃) before acid hydrolysis and PMP derivatization. The permethylated glycosides are then separated by UHPLC equipped with a C18 column and analyzed using a QqQ MS operated in MRM mode.^{101, 102} **Figure 1.9** depicts a summary of the sample preparation and an example chromatogram of the linkages detected in a pool of oligosaccharide standards.

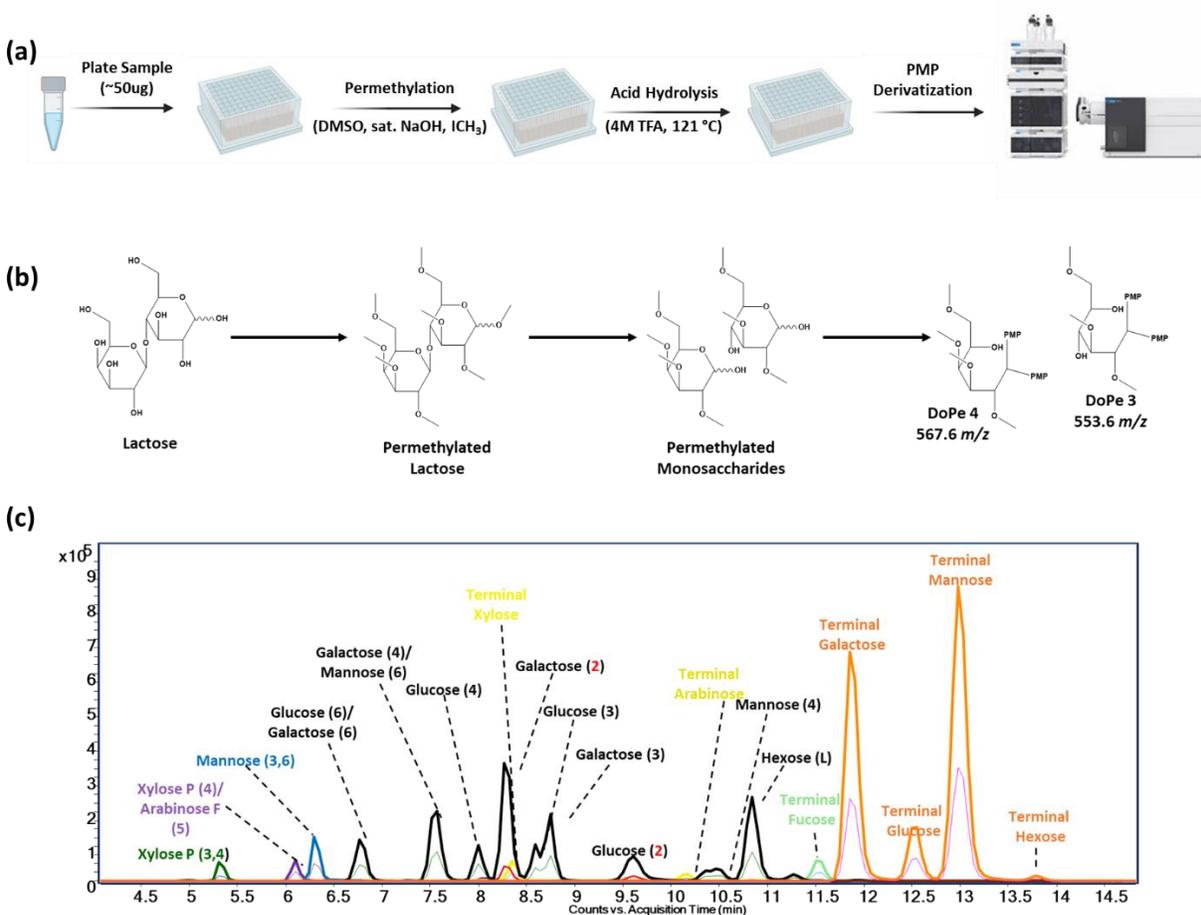


Figure 1.9a-c (a) Overview of the sample preparation steps for glycosidic linkage analysis. (b) Process of linkage analysis using lactose as an example. Saccharides are permethylated, hydrolyzed, and derivatized to produce glycosides with different Degrees of Permethylation (DoPe) that can be separated in both LC and MS dimensions. (c) Example MRM chromatogram of a pool of oligosaccharide standards.

Together the monosaccharide and glycosidic linkage analyses provide detailed and quantitative structural information on food carbohydrates in a rapid-throughput and highly sensitive format, allowing for the analysis of large sample sets with minimal material for the first time. These tools will be integral in furthering our understanding of how food carbohydrates impact the gut microbiome and host health.

CONCLUSION

Carbohydrates are the most abundant biomolecule on Earth, comprise most of the dry mass of plants, and are a major component of our diets. In plants, large polysaccharides provide structural support and function as energy storage molecules while smaller saccharides provide energy for many cellular processes. In the animals that consume these plants, carbohydrates are an important source of energy and can also function as dietary fiber, providing food for the gut microbiome and modulating host health. The biological functions of carbohydrates are dictated by their structures. It is thus necessary that these structures be known to understand how they influence health through their interaction with the gut microbiome. However, current analytical methods for food carbohydrate analysis are limited by the information they can provide, their throughput, and their sensitivity. Recently developed LC-MS based methods directly address these limitations, providing significantly more structural information with enhanced sensitivity in a rapid-throughput, 96-well plate format. This dissertation focuses on the improvement and application of these methods to expand our knowledge of food carbohydrate structures in the context of the gut microbiome. **Chapter 2** provides a detailed protocol integrating these methods into a “multi-glycomic” analysis of food carbohydrates. **Chapter 3** illustrates the use of these methods to obtain a glycomic “map” of the maize plant to inform the spatial distributions of the carbohydrates and identify possible opportunities for improved utilization of the entire plant. In **Chapter 4**, the monosaccharide analysis is used to create the Davis Food Glycopedia, an open-access database of food monosaccharide compositions with over 800 entries. In **Chapter 5**, these methods are built upon and coupled to enzymatic-gravimetric AOAC methods for dietary fiber analysis to quantify and elucidate the structures present in the isolated fiber fractions.

REFERENCES

1. Seeberger, P. H., Monosaccharide Diversity. In *Essentials of Glycobiology*, th; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor (NY), 2022; pp 21-32.
2. Castillo, J. J.; Couture, G.; Bacalzo, N. P., Jr.; Chen, Y.; Chin, E. L.; Blecksmith, S. E.; Bouzid, Y. Y.; Vainberg, Y.; Masarweh, C.; Zhou, Q.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lemay, D. G.; Lebrilla, C. B., The Development of the Davis Food Glycopedia-A Glycan Encyclopedia of Food. *Nutrients* **2022**, *14* (8).
3. Lebrilla, C. B.; Liu, J.; Widmalm, G.; Prestegard, J. H., Oligosaccharides and Polysaccharides. In *Essentials of Glycobiology*, th; Varki, A.; Cummings, R. D.; Esko, J. D.; Stanley, P.; Hart, G. W.; Aebi, M.; Mohnen, D.; Kinoshita, T.; Packer, N. H.; Prestegard, J. H.; Schnaar, R. L.; Seeberger, P. H., Eds. Cold Spring Harbor (NY), 2022; pp 33-42.
4. Pinho, S. S.; Reis, C. A., Glycosylation in cancer: mechanisms and clinical implications. *Nat Rev Cancer* **2015**, *15* (9), 540-55.
5. Lowe, J. B., Glycosylation, immunity, and autoimmunity. *Cell* **2001**, *104* (6), 809-812.
6. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J., Glycosylation in health and disease. *Nat Rev Nephrol* **2019**, *15* (6), 346-366.
7. Van Laere, A.; Van den Ende, W., Inulin metabolism in dicots: chicory as a model system. *Plant Cell Environ* **2002**, *25* (6), 803-813.
8. Ruhaak, L. R.; Lebrilla, C. B., Advances in Analysis of Human Milk Oligosaccharides. *Advances in Nutrition* **2012**, *3* (3), 406s-414s.
9. Wu, S. A.; Tao, N. N.; German, J. B.; Grimm, R.; Lebrilla, C. B., Development of an Annotated Library of Neutral Human Milk Oligosaccharides. *J Proteome Res* **2010**, *9* (8), 4138-4151.
10. Zivkovic, A. M.; German, J. B.; Lebrilla, C. B.; Mills, D. A., Human milk glycobiome and its impact on the infant gastrointestinal microbiota. *P Natl Acad Sci USA* **2011**, *108*, 4653-4658.
11. Pfister, B.; Zeeman, S. C., Formation of starch in plant cells. *Cell Mol Life Sci* **2016**, *73* (14), 2781-807.
12. Zhang, B. C.; Gao, Y. H.; Zhang, L. J.; Zhou, Y. H., The plant cell wall: Biosynthesis, construction, and functions. *J Integr Plant Biol* **2021**, *63* (1), 251-272.
13. Gudmand-Hoyer, E.; Skovbjerg, H., Disaccharide digestion and maldigestion. *Scand J Gastroenterol Suppl* **1996**, *216*, 111-21.
14. Wright, E. M.; Martin, M. G.; Turk, E., Intestinal absorption in health and disease--sugars. *Best Pract Res Clin Gastroenterol* **2003**, *17* (6), 943-56.
15. Brownlee, I. A.; Gill, S.; Wilcox, M. D.; Pearson, J. P.; Chater, P. I., Starch digestion in the upper gastrointestinal tract of humans. *Starch-Starke* **2018**, *70* (9-10).
16. Flint, H. J.; Scott, K. P.; Duncan, S. H.; Louis, P.; Forano, E., Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **2012**, *3* (4), 289-306.
17. Wardman, J. F.; Bains, R. K.; Rahfeld, P.; Withers, S. G., Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nat Rev Microbiol* **2022**, *20* (9), 542-556.

18. Cronin, P.; Joyce, S. A.; O'Toole, P. W.; O'Connor, E. M., Dietary Fibre Modulates the Gut Microbiota. *Nutrients* **2021**, *13* (5).
19. Dhingra, D.; Michael, M.; Rajput, H.; Patil, R. T., Dietary fibre in foods: a review. *J Food Sci Technol* **2012**, *49* (3), 255-66.
20. Gill, S. K.; Rossi, M.; Bajka, B.; Whelan, K., Dietary fibre in gastrointestinal health and disease. *Nat Rev Gastroenterol Hepatol* **2021**, *18* (2), 101-116.
21. O'Grady, J.; O'Connor, E. M.; Shanahan, F., Review article: dietary fibre in the era of microbiome science. *Aliment Pharm Ther* **2019**, *49* (5), 506-515.
22. U.S. Department of Agriculture, U. S. D. o. H. a. H. S., *Dietary Guidelines for Americans, 2020-2025*. 9 ed.; 2020.
23. Junejo, S. A.; Flanagan, B. M.; Zhang, B.; Dhital, S., Starch structure and nutritional functionality - Past revelations and future prospects. *Carbohydr Polym* **2022**, *277*, 118837.
24. MacNeill, G. J.; Mehrpouyan, S.; Minow, M. A. A.; Patterson, J. A.; Tetlow, I. J.; Emes, M. J., Starch as a source, starch as a sink: the bifunctional role of starch in carbon allocation. *Journal of Experimental Botany* **2017**, *68* (16), 4433-4453.
25. Birt, D. F.; Boylston, T.; Hendrich, S.; Jane, J. L.; Hollis, J.; Li, L.; McClelland, J.; Moore, S.; Phillips, G. J.; Rowling, M.; Schalinske, K.; Scott, M. P.; Whitley, E. M., Resistant starch: promise for improving human health. *Adv Nutr* **2013**, *4* (6), 587-601.
26. Warren, F. J.; Fukuma, N. M.; Mikkelsen, D.; Flanagan, B. M.; Williams, B. A.; Lisle, A. T.; Cui, P. O.; Morrison, M.; Gidley, M. J., Food Starch Structure Impacts Gut Microbiome Composition. *Mosphere* **2018**, *3* (3).
27. Somerville, C.; Bauer, S.; Brininstool, G.; Facette, M.; Hamann, T.; Milne, J.; Osborne, E.; Paredes, A.; Persson, S.; Raab, T.; Vorwerk, S.; Youngs, H., Toward a systems approach to understanding plant cell walls. *Science* **2004**, *306* (5705), 2206-11.
28. Taylor, N. G., Cellulose biosynthesis and deposition in higher plants. *New Phytol* **2008**, *178* (2), 239-252.
29. Scheller, H. V.; Ulvskov, P., Hemicelluloses. *Annu Rev Plant Biol* **2010**, *61*, 263-89.
30. Ridley, B. L.; O'Neill, M. A.; Mohnen, D. A., Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry* **2001**, *57* (6), 929-967.
31. Mohnen, D., Pectin structure and biosynthesis. *Current Opinion in Plant Biology* **2008**, *11* (3), 266-277.
32. Torode, T. A.; O'Neill, R.; Marcus, S. E.; Cornuault, V.; Pose, S.; Lauder, R. P.; Kracun, S. K.; Rydahl, M. G.; Andersen, M. C. F.; Willats, W. G. T.; Braybrook, S. A.; Townsend, B. J.; Clausen, M. H.; Knox, J. P., Branched Pectic Galactan in Phloem-Sieve-Element Cell Walls: Implications for Cell Mechanics. *Plant Physiology* **2018**, *176* (2), 1547-1558.
33. Moneo-Sanchez, M.; Vaquero-Rodriguez, A.; Hernandez-Nistal, J.; Albornos, L.; Knox, P.; Dopico, B.; Labrador, E.; Martin, I., Pectic galactan affects cell wall architecture during secondary cell wall deposition. *Planta* **2020**, *251* (5).
34. Verhertbruggen, Y.; Marcus, S. E.; Haeger, A.; Verhoef, R.; Schols, H. A.; McCleary, B. V.; Mckee, L.; Gilbert, H. J.; Knox, J. P., Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. *Plant J* **2009**, *59* (3), 413-425.
35. Wefers, D.; Florchinger, R.; Bunzel, M., Detailed Structural Characterization of Arabinans and Galactans of 14 Apple Cultivars Before and After Cold Storage. *Front Plant Sci* **2018**, *9*.

36. Bar-Peled, M.; Urbanowicz, B. R.; O'Neill, M. A., The synthesis and origin of the pectic polysaccharide rhamnogalacturonan II - insights from nucleotide sugar formation and diversity. *Front Plant Sci* **2012**, *3*.
37. O'Neill, M. A.; Ishii, T.; Albersheim, P.; Darvill, A. G., Rhamnogalacturonan II: Structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annual Review of Plant Biology* **2004**, *55*, 109-139.
38. Schultink, A.; Liu, L.; Zhu, L.; Pauly, M., Structural Diversity and Function of Xyloglucan Sidechain Substituents. *Plants (Basel)* **2014**, *3* (4), 526-42.
39. Fry, S. C., The Structure and Functions of Xyloglucan. *Journal of Experimental Botany* **1989**, *40* (210), 1-11.
40. Flint, H. J.; Scott, K. P.; Louis, P.; Duncan, S. H., The role of the gut microbiota in nutrition and health. *Nat Rev Gastroenterol Hepatol* **2012**, *9* (10), 577-89.
41. Valdes, A. M.; Walter, J.; Segal, E.; Spector, T. D., Role of the gut microbiota in nutrition and health. *BMJ* **2018**, *361*, k2179.
42. Fan, Y.; Pedersen, O., Gut microbiota in human metabolic health and disease. *Nat Rev Microbiol* **2021**, *19* (1), 55-71.
43. Johnson, A. J.; Zheng, J. J.; Kang, J. W.; Saboe, A.; Knights, D.; Zivkovic, A. M., A Guide to Diet-Microbiome Study Design. *Front Nutr* **2020**, *7*, 79.
44. Han, S.; Van Treuren, W.; Fischer, C. R.; Merrill, B. D.; DeFelice, B. C.; Sanchez, J. M.; Higginbottom, S. K.; Guthrie, L.; Fall, L. A.; Dodd, D.; Fischbach, M. A.; Sonnenburg, J. L., A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* **2021**, *595* (7867), 415-420.
45. Rinninella, E.; Raoul, P.; Cintoni, M.; Franceschi, F.; Miggiano, G. A. D.; Gasbarrini, A.; Mele, M. C., What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **2019**, *7* (1).
46. Yatsunenko, T.; Rey, F. E.; Manary, M. J.; Trehan, I.; Dominguez-Bello, M. G.; Contreras, M.; Magris, M.; Hidalgo, G.; Baldassano, R. N.; Anokhin, A. P.; Heath, A. C.; Warner, B.; Reeder, J.; Kuczynski, J.; Caporaso, J. G.; Lozupone, C. A.; Lauber, C.; Clemente, J. C.; Knights, D.; Knight, R.; Gordon, J. I., Human gut microbiome viewed across age and geography. *Nature* **2012**, *486* (7402), 222-7.
47. David, L. A.; Maurice, C. F.; Carmody, R. N.; Gootenberg, D. B.; Button, J. E.; Wolfe, B. E.; Ling, A. V.; Devlin, A. S.; Varma, Y.; Fischbach, M. A.; Biddinger, S. B.; Dutton, R. J.; Turnbaugh, P. J., Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014**, *505* (7484), 559-63.
48. Human Microbiome Project, C., Structure, function and diversity of the healthy human microbiome. *Nature* **2012**, *486* (7402), 207-14.
49. Koh, A.; De Vadder, F.; Kovatcheva-Datchary, P.; Backhed, F., From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **2016**, *165* (6), 1332-1345.
50. Zhao, J.; Zhang, X.; Liu, H.; Brown, M. A.; Qiao, S., Dietary Protein and Gut Microbiota Composition and Function. *Curr Protein Pept Sci* **2019**, *20* (2), 145-154.
51. Huyan, Z. Y.; Pellegrini, N.; Steegenga, W.; Capuano, E., Insights into gut microbiota metabolism of dietary lipids: the case of linoleic acid. *Food Funct* **2022**, *13* (8), 4513-4526.
52. Schoeler, M.; Caesar, R., Dietary lipids, gut microbiota and lipid metabolism. *Rev Endocr Metab Dis* **2019**, *20* (4), 461-472.

53. Kumar Singh, A.; Cabral, C.; Kumar, R.; Ganguly, R.; Kumar Rana, H.; Gupta, A.; Rosaria Lauro, M.; Carbone, C.; Reis, F.; Pandey, A. K., Beneficial Effects of Dietary Polyphenols on Gut Microbiota and Strategies to Improve Delivery Efficiency. *Nutrients* **2019**, *11* (9).
54. Rodriguez-Daza, M. C.; Pulido-Mateos, E. C.; Lupien-Meilleur, J.; Guyonnet, D.; Desjardins, Y.; Roy, D., Polyphenol-Mediated Gut Microbiota Modulation: Toward Prebiotics and Further. *Front Nutr* **2021**, *8*.
55. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **2009**, *37* (Database issue), D233-8.
56. Aspeborg, H.; Coutinho, P. M.; Wang, Y.; Brumer, H., 3rd; Henrissat, B., Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* **2012**, *12*, 186.
57. Mewis, K.; Lenfant, N.; Lombard, V.; Henrissat, B., Dividing the Large Glycoside Hydrolase Family 43 into Subfamilies: a Motivation for Detailed Enzyme Characterization. *Appl Environ Microbiol* **2016**, *82* (6), 1686-1692.
58. Stam, M. R.; Danchin, E. G.; Rancurel, C.; Coutinho, P. M.; Henrissat, B., Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng Des Sel* **2006**, *19* (12), 555-62.
59. O'Callaghan, A.; van Sinderen, D., Bifidobacteria and Their Role as Members of the Human Gut Microbiota. *Front Microbiol* **2016**, *7*.
60. Sela, D. A.; Chapman, J.; Adeuya, A.; Kim, J. H.; Chen, F.; Whitehead, T. R.; Lapidus, A.; Rokhsar, D. S.; Lebrilla, C. B.; German, J. B.; Price, N. P.; Richardson, P. M.; Mills, D. A., The genome sequence of *Bifidobacterium longum* subsp. *infantis* reveals adaptations for milk utilization within the infant microbiome. *Proc Natl Acad Sci U S A* **2008**, *105* (48), 18964-9.
61. Marcobal, A.; Barboza, M.; Sonnenburg, E. D.; Pudlo, N.; Martens, E. C.; Desai, P.; Lebrilla, C. B.; Weimer, B. C.; Mills, D. A.; German, J. B.; Sonnenburg, J. L., Bacteroides in the infant gut consume milk oligosaccharides via mucus-utilization pathways. *Cell Host Microbe* **2011**, *10* (5), 507-14.
62. Martens, E. C.; Chiang, H. C.; Gordon, J. I., Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **2008**, *4* (5), 447-57.
63. Salyers, A. A.; Vercellotti, J. R.; West, S. E.; Wilkins, T. D., Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Appl Environ Microbiol* **1977**, *33* (2), 319-22.
64. Martens, E. C.; Lowe, E. C.; Chiang, H.; Pudlo, N. A.; Wu, M.; McNulty, N. P.; Abbott, D. W.; Henrissat, B.; Gilbert, H. J.; Bolam, D. N.; Gordon, J. I., Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **2011**, *9* (12), e1001221.
65. Grondin, J. M.; Tamura, K.; Dejean, G.; Abbott, D. W.; Brumer, H., Polysaccharide Utilization Loci: Fueling Microbial Communities. *J Bacteriol* **2017**, *199* (15).
66. Bjursell, M. K.; Martens, E. C.; Gordon, J. I., Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem* **2006**, *281* (47), 36269-79.

67. Delannoy-Bruno, O.; Desai, C.; Castillo, J. J.; Couture, G.; Barve, R. A.; Lombard, V.; Henrissat, B.; Cheng, J.; Han, N.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Lebrilla, C. B.; Marion, S.; Heath, A. C.; Barratt, M. J.; Gordon, J. I., An approach for evaluating the effects of dietary fiber polysaccharides on the human gut microbiome and plasma proteome. *Proc Natl Acad Sci U S A* **2022**, *119* (20), e2123411119.
68. Fischbach, M. A.; Sonnenburg, J. L., Eating for two: how metabolism establishes interspecies interactions in the gut. *Cell Host Microbe* **2011**, *10* (4), 336-47.
69. Duncan, S. H.; Barcenilla, A.; Stewart, C. S.; Pryde, S. E.; Flint, H. J., Acetate utilization and butyryl coenzyme A (CoA): acetate-CoA transferase in butyrate-producing bacteria from the human large intestine. *Appl Environ Microb* **2002**, *68* (10), 5186-5190.
70. Louis, P.; Duncan, S. H.; McCrae, S. I.; Millar, J.; Jackson, M. S.; Flint, H. J., Restricted distribution of the butyrate kinase pathway among butyrate-producing bacteria from the human colon. *J Bacteriol* **2004**, *186* (7), 2099-2106.
71. Ragsdale, S. W.; Pierce, E., Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Bba-Proteins Proteom* **2008**, *1784* (12), 1873-1898.
72. Scott, K. P.; Martin, J. C.; Campbell, G.; Mayer, C. D.; Flint, H. J., Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium "Roseburia inulinivorans". *J Bacteriol* **2006**, *188* (12), 4340-4349.
73. Morrison, D. J.; Preston, T., Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **2016**, *7* (3), 189-200.
74. Naaeder, S. B.; Evans, D. F.; Archampong, E. Q., Effect of acute dietary fibre supplementation on colonic pH in healthy volunteers. *West Afr J Med* **1998**, *17* (3), 153-6.
75. Zeng, H.; Lazarova, D. L.; Bordonaro, M., Mechanisms linking dietary fiber, gut microbiota and colon cancer prevention. *World J Gastrointest Oncol* **2014**, *6* (2), 41-51.
76. Kamada, N.; Chen, G. Y.; Inohara, N.; Nunez, G., Control of pathogens and pathobionts by the gut microbiota. *Nat Immunol* **2013**, *14* (7), 685-90.
77. Duncan, S. H.; Louis, P.; Flint, H. J., Lactate-utilizing bacteria, isolated from human feces, that produce butyrate as a major fermentation product. *Appl Environ Microbiol* **2004**, *70* (10), 5810-7.
78. Comalada, M.; Bailon, E.; de Haro, O.; Lara-Villoslada, F.; Xaus, J.; Zarzuelo, A.; Galvez, J., The effects of short-chain fatty acids on colon epithelial proliferation and survival depend on the cellular phenotype. *J Cancer Res Clin Oncol* **2006**, *132* (8), 487-97.
79. van der Hee, B.; Wells, J. M., Microbial Regulation of Host Physiology by Short-chain Fatty Acids. *Trends Microbiol* **2021**, *29* (8), 700-712.
80. Suzuki, T.; Yoshida, S.; Hara, H., Physiological concentrations of short-chain fatty acids immediately suppress colonic epithelial permeability. *Br J Nutr* **2008**, *100* (2), 297-305.
81. Willemsen, L. E.; Koetsier, M. A.; van Deventer, S. J.; van Tol, E. A., Short chain fatty acids stimulate epithelial mucin 2 expression through differential effects on prostaglandin E(1) and E(2) production by intestinal myofibroblasts. *Gut* **2003**, *52* (10), 1442-7.
82. den Besten, G.; van Eunen, K.; Groen, A. K.; Venema, K.; Reijngoud, D. J.; Bakker, B. M., The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J Lipid Res* **2013**, *54* (9), 2325-2340.
83. Sanderson, I. R., Short chain fatty acid regulation of signaling genes expressed by the intestinal epithelium. *J Nutr* **2004**, *134* (9), 2450S-2454S.
84. Parada Venegas, D.; De la Fuente, M. K.; Landskron, G.; Gonzalez, M. J.; Quera, R.; Dijkstra, G.; Harmsen, H. J. M.; Faber, K. N.; Hermoso, M. A., Short Chain Fatty Acids

- (SCFAs)-Mediated Gut Epithelial and Immune Regulation and Its Relevance for Inflammatory Bowel Diseases. *Front Immunol* **2019**, *10*, 277.
85. Blakeney, A. B.; Harris, P. J.; Henry, R. J.; Stone, B. A., A Simple and Rapid Preparation of Alditol Acetates for Monosaccharide Analysis. *Carbohydr Res* **1983**, *113* (2), 291-299.
 86. Sawardeker, J. S.; Sloneker, J. H.; Jeanes, A., Quantitative Determination of Monosaccharides as Their Alditol Acetates by Gas Liquid Chromatography. *Anal Chem* **1965**, *37* (12), 1602-+.
 87. Lee, Y. C., High-Performance Anion-Exchange Chromatography for Carbohydrate Analysis. *Anal Biochem* **1990**, *189* (2), 151-162.
 88. Anumula, K. R.; Taylor, P. B., A Comprehensive Procedure for Preparation of Partially Methylated Alditol Acetates from Glycoprotein Carbohydrates. *Anal Biochem* **1992**, *203* (1), 101-108.
 89. Duus, J.; Gotfredsen, C. H.; Bock, K., Carbohydrate structural determination by NMR spectroscopy: modern methods and limitations. *Chem Rev* **2000**, *100* (12), 4589-614.
 90. McCleary, B. V.; Charmier, L. M. J.; McKie, V. A., Measurement of Starch: Critical Evaluation of Current Methodology. *Starch-Starke* **2019**, *71* (1-2).
 91. McCleary, B. V.; DeVries, J. W.; Rader, J. I.; Cohen, G.; Prosky, L.; Mugford, D. C.; Okuma, K., Determination of insoluble, soluble, and total dietary fiber (CODEX definition) by enzymatic-gravimetric method and liquid chromatography: collaborative study. *JAOAC Int* **2012**, *95* (3), 824-44.
 92. Fu, Q.; Liang, T.; Li, Z. Y.; Xu, X. Y.; Ke, Y. X.; Jin, Y.; Liang, X. M., Separation of carbohydrates using hydrophilic interaction liquid chromatography. *Carbohydr Res* **2013**, *379*, 13-17.
 93. Zaia, J., Mass spectrometry and glycomics. *OMICS* **2010**, *14* (4), 401-18.
 94. Ruhaak, L. R.; Lebrilla, C. B., Advances in analysis of human milk oligosaccharides. *Adv Nutr* **2012**, *3* (3), 406S-14S.
 95. Li, Q.; Xie, Y.; Wong, M.; Barboza, M.; Lebrilla, C. B., Comprehensive structural glycomic characterization of the glycocalyxes of cells and tissues. *Nat Protoc* **2020**, *15* (8), 2668-2704.
 96. Deery, M. J.; Stimson, E.; Chappell, C. G., Size exclusion chromatography/mass spectrometry applied to the analysis of polysaccharides. *Rapid Commun Mass Sp* **2001**, *15* (23), 2273-2283.
 97. Corona, A.; Rollings, J. E., Polysaccharide Characterization by Aqueous Size Exclusion Chromatography and Low-Angle Laser-Light Scattering. *Separ Sci Technol* **1988**, *23* (8-9), 855-874.
 98. Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A., Determining the polysaccharide composition of plant cell walls. *Nat Protoc* **2012**, *7* (9), 1590-1607.
 99. Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T. T. T.; Liu, Y. Y.; Lee, M.; Xu, G. G.; Lebrilla, C. B., A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *International Journal of Mass Spectrometry* **2019**, *438*, 22-28.
 100. Xu, G. G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis - quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2018**, *143* (1), 200-207.

101. Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B., Development of an Extensive Linkage Library for Characterization of Carbohydrates. *Anal Chem* **2019**, *91* (20), 13022-13031.
102. Galermo, A. G.; Nandita, E.; Barboza, M.; Arnicucci, M. J.; Vo, T. T. T.; Lebrilla, C. B., Liquid Chromatography-Tandem Mass Spectrometry Approach for Determining Glycosidic Linkages. *Anal Chem* **2018**, *90* (21), 13073-13080.

Chapter 2

A Multi-Glycomic Platform for the Analysis of Food Carbohydrates

ABSTRACT

Carbohydrates comprise the largest fraction of most diets and exert a profound impact on health. Components such as simple sugars and starch supply energy while indigestible components—deemed dietary fiber, reach the colon to provide food for the trillions of microbes that make up the gut microbiota. The interactions between dietary carbohydrates, our gastrointestinal tracts, the gut microbiome, and host health are dictated by their structures. However, current methods for food carbohydrate analysis lack the sensitivity, specificity, and throughput to quantify and elucidate these structures. This protocol describes a multi-glycomic approach to food carbohydrate analysis employing rapid-throughput liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods. A quantitative monosaccharide compositional analysis, comprehensive glycosidic linkage analysis, and unique polysaccharide analysis are performed from the same sample in 96-well plate format to reduce sample size and enhance throughput. Detailed stepwise processes for sample preparation, LC-MS/MS, and data analysis are provided. We illustrate the application of the protocol in the analysis of the carbohydrates in a diverse set of foods as well as different mushroom species and apple cultivars. Furthermore, we show the utility of these methods towards elucidating glycan-microbe interactions in germ-free and colonized feces from mice. These methods provide the structure to elucidate the relationships between dietary fiber, the gut microbiome, and ultimately human health. These structures will further guide nutritional and clinical feeding studies that enhance our understanding of the role of diet in nutrition and health.

INTRODUCTION

Carbohydrates are one of the largest components of the standard human diet. Humans have a limited capacity to digest these biomolecules, possessing a small number of carbohydrate-active enzymes (CAZymes) along the GI tract that target only a limited number of glycosidic linkages found primarily in starch and simple sugars.¹ However, food carbohydrates particularly those from plants, provide large diversities of structures that are not readily digested by human enzymes. Compounds containing these structures are commonly termed dietary fiber as they are not digested by endogenous saccharolytic enzymes and ultimately enter the large intestine where they feed the trillions of microbes that collectively comprise the gut microbiome.²⁻⁵ Dietary fiber is a major determinant of gut microbial ecology where it modulates microbial populations that in turn produce the broad spectrum of metabolites providing the host both short-term and long-term gains.⁶⁻¹⁰ The fiber-microbiome paradigm is a key mediator of general morbidity influencing various factors such as the incidence of certain cancers, the probability of metabolic diseases, the growth trajectory of children, and even the effectiveness of immunotherapy for cancer.¹¹⁻¹⁶

Despite their importance in common human diets, there remains a paucity of analytical methods for measuring and determining the structures of carbohydrates in food.^{17, 18} The most important characteristics of carbohydrates are their monosaccharide compositions and the linkages that bind the monosaccharides together. Human and bacterial glycosidases are specific towards these monosaccharide and linkage compositions. To address the current limitations, we have developed rapid-throughput platforms based on LC-MS/MS that identify and quantify the monosaccharide compositions of carbohydrates.^{19, 20} We also developed methods that quantify nearly one hundred different linkages.^{21, 22} Finally, we developed a chemical method that produces unique oligosaccharides from precursor polysaccharides providing what was previously

unimaginable, the simultaneous identification and quantification of the polysaccharides in food.^{23, 24}

DEVELOPMENT OF THE PROTOCOL

The development of these methods derived from our extensive experience in characterizing glycans and glycoconjugates such as human milk oligosaccharides (HMOs) in milk and glycoproteins and glycolipids in human and animal tissues.²⁵⁻³¹ The structural complexity of dietary carbohydrates can readily exceed those of mammalian glycans. We first dissociated the carbohydrates to their monosaccharide compositions. For monosaccharide analysis, carbohydrates are dissociated through a rigorous acid hydrolysis followed by labeling of the reducing carbon by 1-phenyl-3-methyl-5-pyrazolone (PMP). For linkage analysis, carbohydrates are first permethylated before subsequent acid hydrolysis and PMP labeling. The number and orientation of the added methyl groups allows for the identification of the linkage composition of the parent structure. The resulting glycosides from both monosaccharide and linkage analysis are then analyzed separately using ultra-high performance liquid chromatography-triple quadrupole mass spectrometry (UHPLC-QqQ MS). Polysaccharide through FITDOG analysis (**F**enton's **I**nitiation **T**owards **D**efined **O**ligosaccharide **G**roups) utilizes Fenton's chemistry to breakdown polysaccharides into characteristic oligosaccharides which are then reduced to their corresponding alditols before analysis by high performance liquid chromatography-quadrupole time of flight mass spectrometry (HPLC-qToF). The observed oligosaccharide signals then provide identification and quantification of the parent polysaccharides. **Figure 2.1** provides an overview of the three methods that comprise the multi-glycomic analysis.

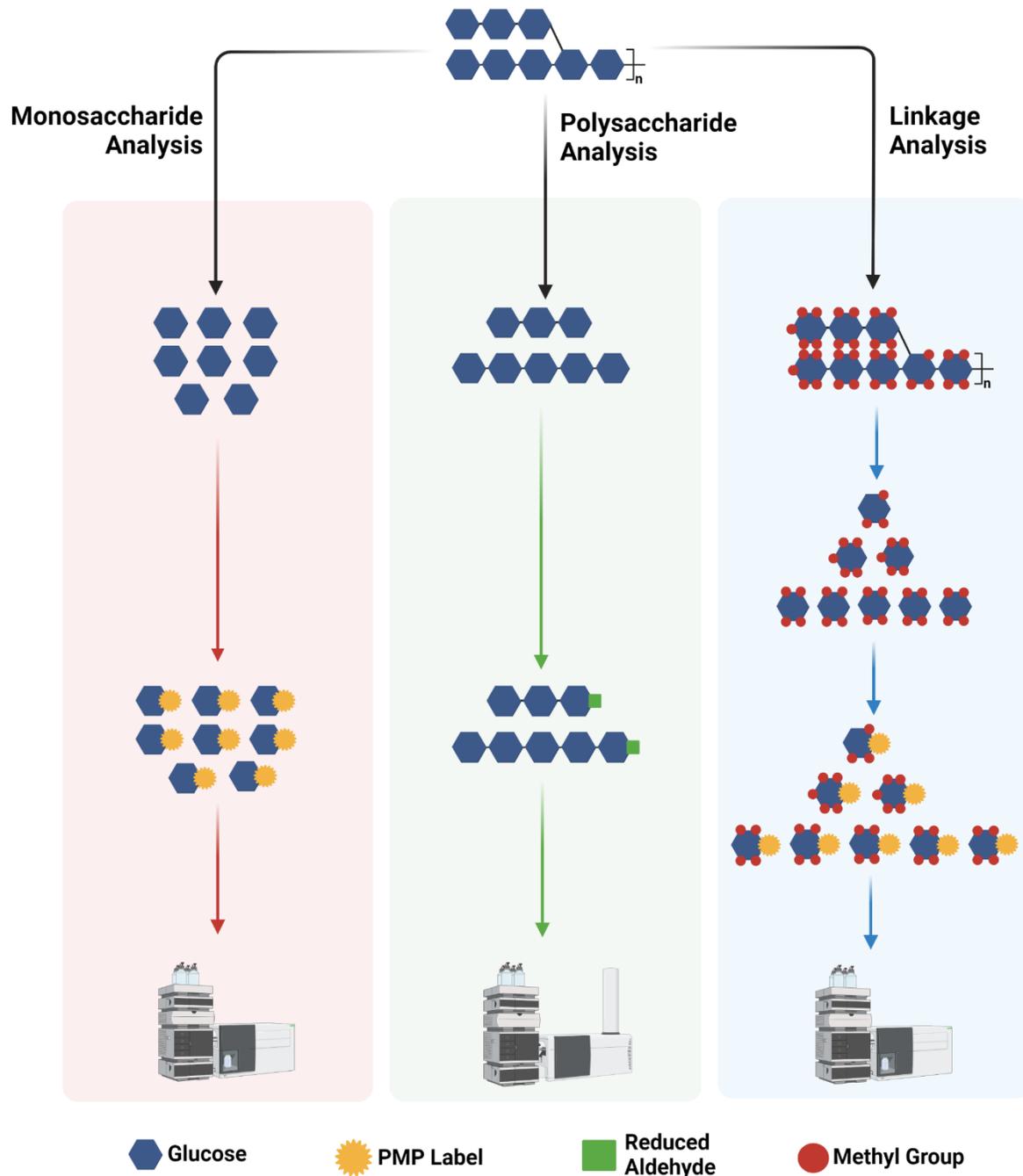


Figure 2.1. Representation of the multi-glycomic workflow. Food carbohydrates are quantified and structurally elucidated by monosaccharide compositional analysis (left panel), polysaccharide analysis (center), and glycosidic linkage analysis (right) using rapid throughput chemical and LC-MS/MS methods.

Throughput and sensitivity have been key considerations throughout the development of this protocol. All of the methods are in a 96-well plate format, and UHPLC is utilized, when possible, to increase throughput and decrease sample run time. Additionally, the entire protocol can be performed from just a few milligrams of sample, allowing for the determination of monosaccharide composition, glycosidic linkages, and amounts of polysaccharide with minimal amounts of material (sub-microgram, sub-microgram, and sub-milligram, respectively).

COMPARISONS TO ALTERNATIVE METHODS

The most commonly employed method for the characterization and quantification of carbohydrates in plants and food by monosaccharide and linkage analysis utilizes gas chromatography-mass spectrometry (GC-MS).³² GC-MS instruments are significantly cheaper than LC-MS/MS instruments and offer great chromatographic resolution. The ability to use electron impact (EI) ionization also makes identification of monosaccharides and their permethylated derivatives by mass spectral matching an attractive feature. For monosaccharide analysis, carbohydrates must be hydrolyzed, acetylated, and reduced to their corresponding alditol acetates (AAs) prior to injection.^{32, 33} For linkage analysis, an initial permethylation is needed to obtain corresponding permethylated alditol acetates (PMAAs).^{32, 34, 35} However, this analytical approach has significant limitations compared to the current methods. Instrument run times are typically longer for monosaccharide analysis (20 min vs. 5 min), and for linkage analysis (60 min vs. 16 min). Polysaccharide analysis analogous to FITDOG is not possible with GC-MS. Additionally, long run times and the need for samples to be in highly volatile solvents for GC analysis makes the adaptation of these methods to 96-well plates difficult. Thus, samples

are prepared one vial at a time severely limiting sample throughput. Additionally, because only a single quadrupole instrument is typically used, GC-MS chromatograms can be prone to high noise and relatively low sensitivity (milligram vs. picogram for monosaccharide analysis) especially if further upstream purifications are not performed.^{19, 32} Lastly, the coverage of glycosidic linkages by GC-MS is also limited compared to the current LC-MS/MS approach.

High-performance anion exchange chromatography coupled to pulsed amperometric detection (HPAEC-PAD) is also commonly employed for monosaccharide and oligosaccharide analysis. HPAEC-PAD is particularly attractive for carbohydrate analysis because it is highly selective, sensitive, and does not require derivatization prior to analysis.³⁶ However, isomer separation can be challenging and often requires long method run times.^{36, 37} Furthermore, linkage analysis by HPAEC appears not to be possible.

The LC-MS/MS methods presented here directly address many of the disadvantages of GC-MS and HPAEC-PAD. Namely, the chromatographic separation time of PMP-labeled glycosides on a C18 stationary phase is drastically reduced relative to AAs and PMAAs in GC-MS analysis and to underivatized carbohydrates in HPAEC-PAD. For example, the monosaccharide method described here separates 14 monosaccharides in 4.6 min while a similar separation using GC-MS may require upwards of 20 min (8 h vs. 48 h for 96 samples).^{19, 32, 33, 38} Because most of the sample preparation for this protocol is in aqueous solvent and significantly less prone to evaporative effects, it is also more adaptable to a 96-well plate format, unlike GC-MS. Further, the use of MS/MS improves the signal-to-noise ratio particularly for monosaccharide and linkage analyses and allows for fewer sample preparation steps than is more typical for GC-MS. Additionally, the LC-MS/MS approach expands greatly the limited linear

and dynamic ranges of the GC and HPAEC approaches, offering up to 6 orders of magnitude relative to the 2 to 3 provided by the latter.¹⁹

Polysaccharide analysis by FITDOG is a very recent method for carbohydrate analysis yielding the direct identification and quantification of polysaccharides.^{23, 39, 40} The method is highly specific because characteristic oligosaccharides are produced from each polysaccharide differentiating each polymer. Direct identification of polysaccharides in food has historically required extensive sample preparation and extraction techniques to isolate individual polysaccharides for GC-MS, HPAEC-PAD, and/or nuclear magnetic resonance (NMR) analysis.^{32, 36, 41-43} Quantification could only be performed gravimetrically. FITDOG analysis can be performed on mixtures of polysaccharides such as those present in food and feces with high throughput.

The concept of the polysaccharide analysis is similar to other bottom-up approaches used in genomics and proteomics where the polysaccharides are first broken down into smaller oligosaccharide fragments. The matching of the resultant oligosaccharide compounds to their parent polysaccharide structure is based on an oligosaccharide fingerprinting library obtained by reacting commercially-available polysaccharide standards. It is recommended to employ an ethanol precipitation step prior to polysaccharide analysis to ensure that endogenous oligosaccharides present in the samples are not attributed to polysaccharide structures. Our lab has developed a non-enzymatic and reproducible reaction to depolymerize common plant polysaccharides using the Fenton's reaction.^{23, 24, 39} Specifically, catalytic amount of Fe^{3+} and an excess of H_2O_2 are added to the reaction mixture to produce reactive oxygen radicals which facilitate the cleavage of glycosidic bonds in the polysaccharides. The resulting oligosaccharides from the depolymerization reaction are then reduced using NaBH_4 to prevent anomer separation

that we observe on some oligosaccharides separated on a PGC column. The oligosaccharide reduction helps reduce the complexity in matching retention times (RT) and accurate masses of the oligosaccharides to the fingerprinting library. Finally, reduced oligosaccharides are cleaned-up and enriched using C18 and PGC SPE. Oligosaccharide profiles are then analyzed using HPLC-qToF in data-dependent mode.

EXPERIMENTAL DESIGN

Overview of the protocol

The overall workflow is provided in **Figure 2.2**. Food, fecal, or plant tissue samples are first lyophilized before homogenization into a powder using a bead mill or a mortar and pestle. For rigid samples, a coffee grinder may first be used. Milligram quantities of the homogenized sample are then weighed into 1.5 mL screw-cap tubes and further homogenized after addition of water by a bead mill and incubation at 100 °C. An ethanol (EtOH) precipitation may also be used at this step to separate high and low molecular weight carbohydrates if separate analyses are needed for soluble and insoluble fractions. Each fraction may then be analyzed separately. If no EtOH precipitation is performed, the resulting suspensions are directly aliquoted into individual 96-well plates, one for monosaccharide, another for glycosidic linkage, and yet another for polysaccharide (FITDOG) analysis. If an EtOH precipitation is performed, the supernatant is removed, and the pellet washed and subsequently homogenized according to the aforementioned steps. For quantitative monosaccharide compositional analysis, sample aliquots are hydrolyzed with trifluoroacetic acid (TFA), derivatized with 1-phenyl-3-methyl-5-pyrazolone (PMP), extracted with chloroform (CHCl₃) to remove excess PMP, separated using UHPLC on a C18

column, and analyzed on a QqQ MS operated in dynamic multiple reaction monitoring (dMRM) mode. Absolute quantification is achieved using an external calibration curve. However, isotopically labeled internal standards may also be used. Glycosidic linkage analysis employs the same steps with the addition of a permethylation step prior to acid hydrolysis. Linkages are assigned using an in-house library of retention times and MRM transitions. Polysaccharide analysis is performed by using the FITDOG reaction in which polysaccharides in an EtOH precipitated sample aliquot are oxidatively cleaved using Fenton's chemistry into characteristic oligosaccharide fragments. The resulting oligosaccharides are then reduced with sodium borohydride (NaBH_4) and subjected to a solid phase extraction (SPE) cleanup with both C18 and porous graphitized carbon (PGC) before analysis on an HPLC-qToF equipped with a PGC column. This protocol provides a means to obtain three levels of information on the carbohydrates in food, feces, and plant tissues.

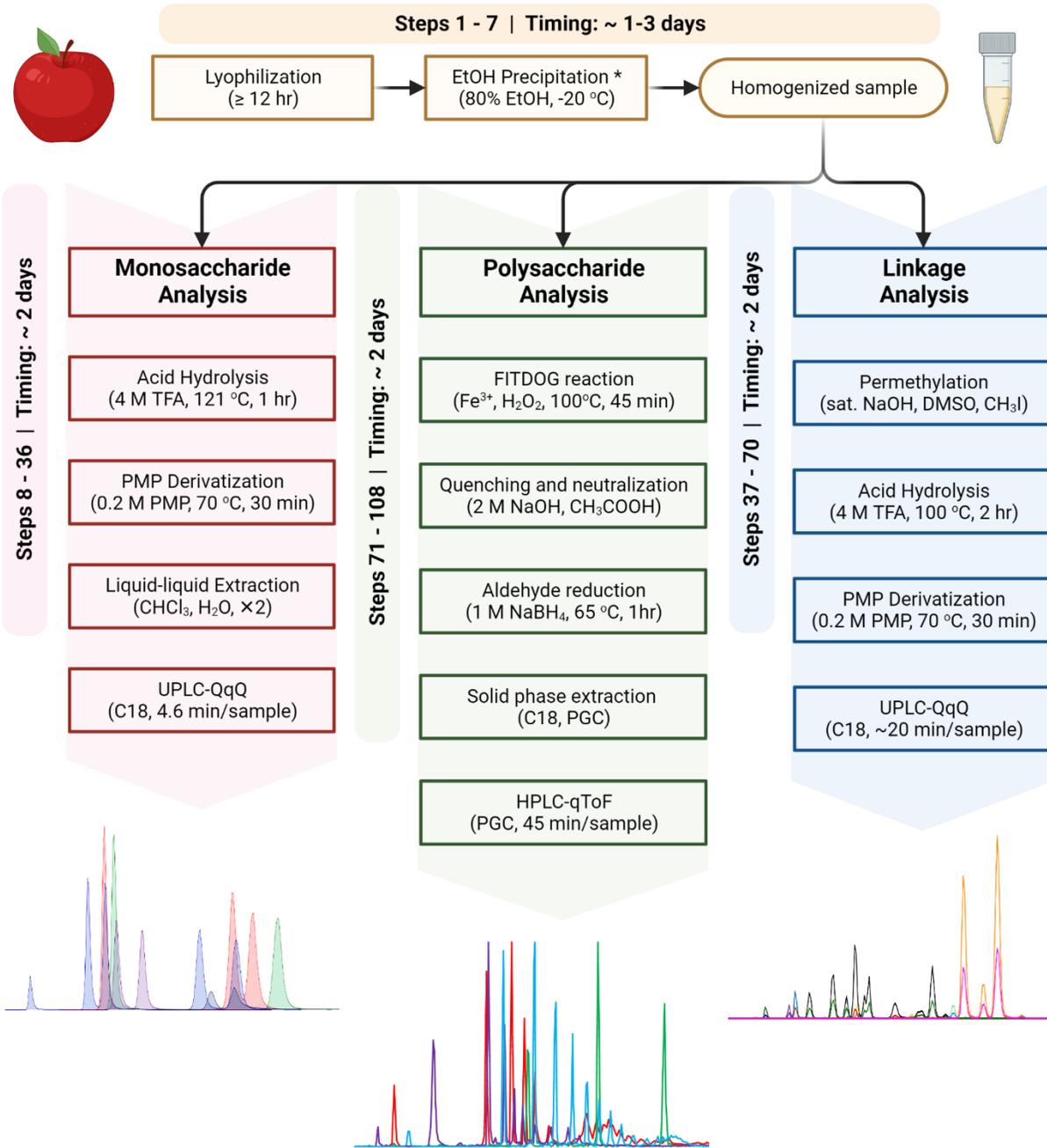


Figure 2.2. Summary of the workflow for comprehensive analysis of carbohydrates in food and feces. The major steps for each of the three analyses are summarized. Monosaccharide

analysis is shown in the left flowchart, polysaccharide FITDOG analysis in the center, and linkage analysis in the right. Representative chromatograms are depicted below each panel.

Sample preparation and homogenization (Steps 1-7)

Because the protocol uses such small aliquot volumes and masses, it is paramount that samples be thoroughly homogenized prior to entering the analytical workflow. We have found that sampling from dry material improves the precision and sensitivity of the overall protocol by normalizing moisture content to a minimum, thereby concentrating the carbohydrates in the sample. Lyophilization (or “freeze-drying”) has worked very well for this purpose. Samples should be flash-frozen at $-80\text{ }^{\circ}\text{C}$ for a minimum of 3 hr to prepare them for lyophilization. Once frozen, samples can be dried effectively. The length of time required to reach a minimum moisture content depends on several factors including the freeze-dryer being used, the original moisture content of the sample, and the sample matrix itself. In general, we have found that about 3 days ensures adequate drying of all sample types. If moisture content data is desired, weighing the sample vessel before and after drying is a necessary step. Some food samples will retain moisture (very hygroscopic samples) even after a few days in the lyophilizer. If total moisture content and absolutely dry basis values are needed, a separate method (e.g. Karl-Fischer titration) may be used. However, this additional step is not presented in this protocol.

Once samples are dried, they must be homogenized into a fine powder to allow precise aliquoting. Additional instruments may be needed for homogenization. For hard samples such as seeds, a coffee grinder is highly recommended. For hygroscopic and high-fat samples such as dried fruit and nuts, we have found that flash-freezing with liquid nitrogen and grinding in a

mortar and pestle is best. Many samples such as feces, vegetables, legumes, meat, grains, and leaf tissues are readily homogenized by bead milling alone using stainless-steel beads.

After drying, samples are weighed so that solutions can be produced for analysis. Samples should be weighed into screw-cap tubes that are compatible with the bead mill homogenizer for the subsequent homogenization steps. We typically weigh aliquots of about 10 mg and add 1 mL of water to create a solution of 10 mg/mL for ease of measurement, throughput, and storage. However, if samples are limited, as little as 1 mg may be used. The limitation is not with the sensitivity of the methods, but rather with the accuracy and precision of the analytical balance and the homogeneity of the starting sample. If the bulk sample is to be analyzed, then 1 mL of water and stainless-steel beads are added to the sample tube. Aqueous solutions are then subjected to bead-mill homogenization, incubation at 100 °C for 1 hr, followed by an additional round of bead-milling.

The polysaccharides may be separated from the smaller components (mono-, di-, and oligosaccharides) and both fractions analyzed separately by first adding 80% EtOH to the weighed unprocessed sample. The samples are vortexed and centrifuged. The soluble fractions are removed by pipette and analyzed separately. However, analysis of the soluble fraction is not included in this protocol, and we instead focus on the insoluble polysaccharide-containing fraction. After the supernatant is removed, the pellet is washed with two volumes of 80% EtOH and dried with vacuum centrifugation. Once dried, the Eppendorf tube containing the pellet is subjected to homogenization with the addition of water and stainless-steel beads. After homogenization, the Eppendorf tube is heated to 100 °C for 1 hr, followed by additional bead-milling. Sample solutions may be stored at –20 °C or –80 °C prior to analysis. If the samples are to be stored, they must be thawed and homogenized again via bead-milling prior to analysis.

Monosaccharide composition: hydrolysis, derivatization, extraction, and UHPLC-QqQ MS analysis (Steps 8-36)

The homogenized solutions are then aliquoted either to a polypropylene 96-well plate or to individual 1.5 mL screw-cap Eppendorf vials. The homogenized solutions appear as suspensions and must be vortexed well prior to pipetting. For the hydrolysis, water and TFA are added to the suspension and the mixture heated to 121 °C for 1 hr. This condition provides quantitative hydrolysis for most polysaccharide components while minimizing degradation of the liberated monosaccharides. However, some optimization may be required for other samples such as those containing fructans as fructose degrades at these conditions. Thus, fructose may be underrepresented under these conditions. If more accurate analysis of fructose is required, gentler hydrolysis conditions are recommended (100 °C for 1 hr). For large-scale analysis, a 96-well plate that can withstand the high temperature, a plate lid, and a clamp to seal the plate are necessary. The 96-well plate and plate lids are commercially available, and the clamp can be machined as shown in **Supplementary Figure 2.1**. Once the hydrolysis is complete, ice-cold water is added.

After hydrolysis, the resulting monosaccharides are labeled with PMP. Additionally, a pooled set of external standards, containing the 14 most common monosaccharides in food, are prepared to produce an external calibration curve. Aliquots of the hydrolysate are transferred to another 96-well plate where a methanolic PMP and ammonia solution are added for labeling. The reaction is carried out at 70 °C, thus the plate must be clamped. Once the derivatization is complete, the samples are dried by vacuum centrifugation. A programmable vacuum centrifuge is recommended to prevent solvent bumping. The dried and labeled monosaccharides are then

extracted with CHCl_3 to remove excess PMP. The aqueous layer containing the analytes is then removed and analyzed with a UHPLC-QqQ MS equipped with a C18 column. Peak areas for each monosaccharide obtained from samples are compared to the external calibration curve for quantitation. A chromatogram of a pooled monosaccharide standard solution is depicted in **Figure 2.3a**.

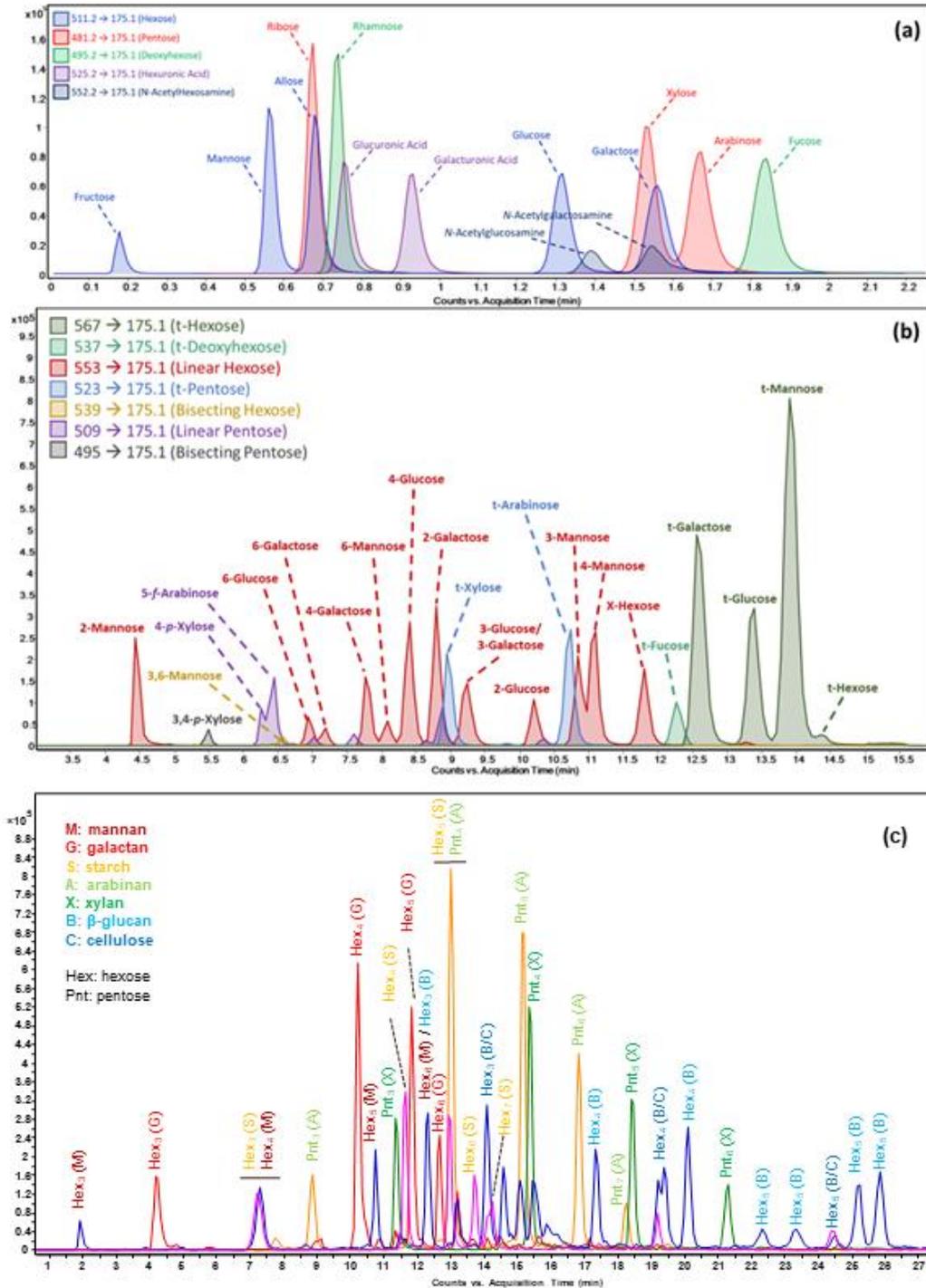


Figure 2.3. dMRM transitions and extracted ion chromatograms (EICs) of pooled monosaccharides, linkages, and polysaccharides. (a) UHPLC-dMRM quantifier ion transitions for the 14 monosaccharides monitored by the method. (b) UHPLC-dMRM quantifier transitions

for a pool of oligosaccharide standards containing the preponderant linkages found in food. (c) EICs of oligosaccharides generated by FITDOG from a pool containing common food polysaccharides.

Glycosidic linkage compositions: permethylation, hydrolysis, derivatization, extraction, and UHPLC-QqQ MS analysis (Steps 37-70)

The procedure for linkage analysis is similar to the monosaccharide analysis with the addition of a permethylation step prior to acid hydrolysis and labeling. Once samples are aliquoted into a 96-well plate, they are dried by vacuum centrifugation. Saturated NaOH is added followed by DMSO to solubilize and activate the carbohydrates. Iodomethane (CH_3I) is then added to permethylate the samples, and the reaction is quenched by the addition of water. Dichloromethane (DCM) is added to solubilize the permethylated products. The DCM layer is then extracted with five volumes of water to remove salts and DMSO. The organic layer is dried by vacuum centrifugation before being subjected to acid hydrolysis with 4 M TFA (100 °C for 2 hr). After hydrolysis, samples are dried again before PMP derivatization. The resulting permethylated and labeled glycosides are dried a final time before reconstitution in 70% MeOH prior to UHPLC-QqQ MS analysis. An in-house library containing the MRM transitions and retention times of the most commonly observed unique linkages in food is used to identify the observed linkages. While this library consists of the preponderant 47 linkages observed in food, up to 96 can be monitored in each sample using an expanded library.²¹ An example chromatogram from a carbohydrate standard pool is shown in **Figure 2.3b**.

**Polysaccharide composition: FITDOG, reduction, purification, and HPLC-qToF analysis
(Steps 71-108)**

It is recommended that samples be EtOH precipitated using the procedure previously detailed prior to FITDOG analysis to remove existing mono-, di, and oligosaccharides. Homogenized samples are then aliquoted to a 96-well plate where solutions of iron (III) sulfate and hydrogen peroxide are added to produce reactive oxygen radicals and facilitate glycosidic bond cleavage thereby converting polysaccharides to oligosaccharides. The reaction is quenched by the addition of NaOH and subsequently neutralized with acetic acid. To avoid volume limitations, aliquots of the depolymerized samples are transferred to another 96-well plate where they are reduced with NaBH₄. The reduced oligosaccharides are enriched using C18 and PGC SPE.

Samples are analyzed via an HPLC-qToF equipped with a PGC column. Oligosaccharide peaks are annotated based on accurate precursor mass and their observed tandem mass spectra. Oligosaccharides are then assigned to their parent polysaccharide structures by matching RT values and compositions to the fingerprint library. It is recommended that several standards be run in parallel with the samples to ensure proper RT matching. Starch and cellulose are relatively cheap and commercially-available and can be used for quality control purposes. Peaks from these standards are used to account for RT shifts when matching compounds to the fingerprint library. The resulting LC-MS chromatogram is provided in **Figure 2.3c** using a standard polysaccharide pool.

LIMITATIONS OF THE PROTOCOL

One limitation of this protocol is the requirement of LC-MS/MS instruments, which are generally large and expensive. The current standard carbohydrate methods employ GC-MS and HPAEC-PAD, which are also large and expensive albeit somewhat cheaper than LC-MS instruments. Furthermore, LC-MS instruments have become more available and even essential in central facilities of most research and academic institutions.

Another limitation in the current iteration of the protocol is the limited coverage of acidic polysaccharides such as galacturonan and rhamnogalacturonan. These structures are still represented by the current protocol, but glycosidic bonds adjacent to uronic acids are recalcitrant to acid hydrolysis. These polysaccharides require enzymatic digestion or reduction of the carboxylic acid prior to hydrolysis for more quantitative coverage. One or both steps may be added to future protocols if better coverage of these polysaccharides is required.

Additionally, data analysis can be cumbersome and time-consuming particularly in the linkage and FITDOG analyses. Due to the nature of glycan structures, there can be a significant number of isomeric species co-eluting throughout the LC-MS run. It then becomes difficult for these compounds to be quantified properly. Automated software is used to annotate and quantitate the compounds however manual confirmation is often necessary when overlapping peaks are present. Recent advances in machine learning for automated peak integration may soon solve this limitation.⁴⁴

FUTURE APPLICATIONS

Carbohydrates are the most abundant component of many diets, and they play profound roles in our overall health through direct utility and their interactions with the gut microbiome. However, the carbohydrate structures and abundances in the foods we consume are still poorly defined. The methods in this protocol can be applied to any food for the quantification and structural elucidation of dietary carbohydrates and fiber. They can also be incorporated into existing fiber analysis protocols like AOAC 991.43 or 2017.16 to determine the compositions of isolated insoluble and soluble dietary fiber fractions.

This protocol can also be used to explore the utilization of fiber by gut microbes both *in vitro* and *in vivo*. A current major challenge in gut microbiology is connecting measured carbohydrate-active enzyme gene and transcript abundances from metagenomic and metatranscriptomic experiments to an explicit functional outcome. By applying the described protocol to the feces of study participants or from small-scale bioreactors, fecal glycan abundances and structures can be determined thus providing mechanistic insight towards food composition, and microbiome and CAZyme function.^{13, 14, 45, 46} The data can also inform differences in the responses of study participants to dietary interventions.^{13, 14}

MATERIALS

Reagents

- Acetonitrile (ACN, Honeywell, cat. no. 34967)

! CAUTION Acetonitrile is a flammable liquid and vapor. Wear personal protective equipment and use only in a chemical fume hood.

- Acetic acid (Glacial, Supelco; MilliporeSigma, cat. no. AX0073)

! CAUTION Acetic acid is a flammable liquid and causes severe skin irritation and eye damage. Wear personal protective equipment when handling.

- Ammonium hydroxide solution (NH₄OH, 28-30%, NH₃ basis, MilliporeSigma, cat. no. 221228)

! CAUTION Ammonium hydroxide solution is corrosive and causes severe skin burns and eye damage. Wear personal protective equipment and use only in a chemical fume hood.

- Ammonium acetate (NH₄Ac, 99.999% trace metals basis, MilliporeSigma, cat. no. 372331)

! CAUTION Ammonium acetate is a combustible solid. Wear personal protective equipment when handling.

- Chloroform (CHCl₃, MilliporeSigma, cat. no. 34854)

! CAUTION Chloroform is a flammable liquid and vapor, acutely toxic, and causes eye and skin irritation. Wear personal protective equipment when handling and use only in a chemical fume hood.

- Ethanol (EtOH, MilliporeSigma, cat. No. E7023)

! CAUTION Ethanol is a flammable liquid and vapor, acutely toxic, and causes eye and skin irritation. Wear personal protective equipment when handling and use only in a chemical fume hood.

- Formic Acid (FA, Optima™ LC/MS Grade; Fisher Chemical™, cat. no. A117)

! CAUTION Formic acid is a flammable liquid and vapor and causes severe skin burns and eye damage. Wear personal protective equipment when handling and use only in a chemical fume hood.

- Hydrogen peroxide solution (H₂O₂, 30 wt. % in H₂O, MilliporeSigma, cat.no. 216763)

! CAUTION Hydrogen peroxide is a strong oxidizer and causes severe skin burns and eye damage. Wear personal protective equipment and use only in a chemical fume hood.

- Methanol (MeOH, MilliporeSigma, cat. no. 34860)

! CAUTION Methanol is a flammable liquid and vapor. It is toxic by inhalation and contact with skin and eyes. Wear personal protective equipment and use only in a chemical fume hood.

- Iron (III) sulfate pentahydrate (Fe₂(SO₄)₃·5H₂O, 97%, Thermo Scientific, cat.no. AC345235000)

! CAUTION Iron (III) sulfate pentahydrate causes skin irritation and eye damage. Wear personal protective equipment when handling.

- Sodium acetate (NaCH₃CO₂, MilliporeSigma, cat. no. 79714)
- Sodium hydroxide (NaOH, 99.99% trace metals basis; MilliporeSigma, cat. no. 306576)

! CAUTION Sodium hydroxide causes severe skin burns and eye damage. Wear personal protective equipment when handling.

- Sodium hydroxide (NaOH, ACS reagent $\geq 97.0\%$; MilliporeSigma, cat. 795429)

! CAUTION Sodium hydroxide causes severe skin burns and eye damage. Wear personal protective equipment when handling.

- Sodium borohydride (NaBH₄, MilliporeSigma, cat.no. 806373)

! CAUTION Sodium borohydride causes skin corrosion, eye damage, and reproductive toxicity. Contact with water also releases flammable gas. Wear personal protective equipment when handling.

- Trifluoroacetic acid (TFA, Optima™ LC/MS Grade; Fisher Chemical™, cat. no. A116)

! CAUTION Trifluoroacetic acid causes severe skin burns and eye damage. Wear personal protective equipment when handling. Use only in a chemical fume hood.

- 3-Methyl-1-phenyl-2-pyrazoline-5-one (PMP, MilliporeSigma, cat.no. M70800)

! CAUTION PMP is toxic and causes eye irritation. Wear personal protective equipment when handling.

Monosaccharide Analysis Standards

- D-(+)-Glucose (MilliporeSigma, cat. no. G8270)
- D-(+)-Galactose (MilliporeSigma, cat. no. G0750)
- D-(-)-Fructose (MilliporeSigma, cat. no. F2793)
- D-(+)-Mannose (MilliporeSigma, cat. no. 92683)

- D-(+)-Allose (MilliporeSigma, cat. no. 285005)
- D-(-)-Ribose (MilliporeSigma, cat. no. R7500)
- D-(+)-Xylose (MilliporeSigma, cat. no. X1500)
- L-(+)-Arabinose (MilliporeSigma, cat. no. A3256)
- L-(+)-Rhamnose monohydrate (MilliporeSigma, cat. no. 41651)
- L-(-)-Fucose (MilliporeSigma, cat. no. 93183)
- D-(+)-Glucuronic acid (MilliporeSigma, cat. no. G5269)
- D-(+)-Galacturonic acid (MilliporeSigma, cat. no. 92478)
- N-Acetyl-D-Glucosamine (MilliporeSigma, cat. no. A8625)
- N-Acetyl-D-Galactosamine (MilliporeSigma, cat. no. A2795)

Linkage Analysis Standards

- 2-O-(α -D-Mannopyranosyl)-D-mannopyranose (Biosynth, cat. no. OM05906)
- 1,3- α -1,6- α -D-Mannotriose (Biosynth, cat. no. OM05762)
- 1,4- β -D-Mannotriose (Biosynth, cat. no. OM31999)
- 1,4- β -D-Xylobiose (Biosynth, cat. no. OX05190)
- 1,5- α -L-Arabinotriose (Biosynth, cat. no. OA32462)
- 3³- α -L-Arabinofuranosyl-xylotetraose (Megazyme, cat. no. O-XA3XX)
- Isomaltotriose (Biosynth, cat. no. OI05352)
- Maltohexaose (Biosynth, cat. no. OM06869)
- Nigerose (Biosynth, cat. no. ON06975)
- Sophorose (Biosynth, cat. no.)
- Amylopectin (Biosynth, cat. no. YA39745)

- 3-O-(β -D-Galactopyranosyl)-D-galactopyranose (Biosynth, cat. no. OG10186)
- 4-O-(β -D-Galactopyranosyl)-D-galactopyranose (Biosynth, cat. no. OG04727)
- Lactose (Biosynth, cat. no. OL04771)
- 2'-Fucosyllactose (Biosynth, cat. no. OF06739)
- Sophorose monohydrate (Biosynth, cat. no. OS06893)

Polysaccharide (FITDOG) Analysis Standards

- Starch (corn, analytical grade, MilliporeSigma, cat. no. S5296)
- Chitin (shrimp shells, MilliporeSigma, cat. no. C9752)
- Cellulose (microcrystalline powder, extra pure, ACROS Organics)
- Linear arabinan (sugar beet pulp, Megazyme, cat. no. P-LARB)
- Mannan (ivory nut seeds, Megazyme, cat. no. P-MANIV)
- Galactan (potato, Megazyme, cat. no. P-GALPOT)
- Xylan (beechwood, Megazyme, cat. no. P-XYLNBE)
- Xyloglucan (tamarind seeds, Megazyme, cat. no. P-XYGLN)
- β -glucan (barley, Megazyme, cat. no. P-BGBM)

Equipment

- Falcon centrifuge tubes (50 mL; Corning, cat.no. 352070)
- Falcon centrifuge tubes (15 mL; Corning, cat.no. 352196)
- Freeze Dryer (SP Scientific, cat.no. BTP-3ESE0W)
- Pipettes (Gilson, cat. no. FA10003M, FA10005M, FA10006M)

- Multichannel Pipettes (USA Scientific, cat. no. 7112-0510, 7112-1100, 7112-3000; Eppendorf, cat. no. 3125000222)
- Bead Mill Homogenizer (OMNI International, cat.no. 19-042E)
- Analytical Balance (METTLER TOLEDO, cat.no. XS105)
- 30 mL Tubes (OMNI International, cat.no. 19-6635)
- 96-Well Polypropylene DeepWell Plate (Thermo Scientific, cat.no. 95040452)
- Sealing Lid for 96-Well Plate (Thermo Scientific, cat.no. AB-0675)
- Plate shaker (Scientific Industries, model no. SI-4000)
- Centrifugal Vacuum Concentrator (SP Scientific, model no. QUC-12060-C00)
- Pipette tips (USA Scientific, cat. no. 1111-3800, 1110-9800, 1112-1820)
- 1.5 mL screw cap tube (Sarstedt, cat.no. 72.692.005)
- C18 SPE PLATE (40 μ L filter plate, C-18, Glysci, cat. FNESC18)
- Graphitized carbon (PGC) SPE plate (40 μ L filter plate, Carbon (Hypercarb), Glysci, cat. FNESCAR)
- Poroshell HPH C18 UHPLC Column (1.9 μ m, 2.1 \times 50 mm, Agilent Technologies, cat. no. 699675-702)
- Poroshell HPH C18 UHPLC guard cartridges (1.9 μ m, 2.1 \times 5 mm, Agilent Technologies, cat. no. 821725-945)
- ZORBAX RRHD Eclipse Plus C18 UHPLC Column (1.8 μ m, 2.1 \times 150 mm, Agilent Technologies, cat. no. 959759-902)
- ZORBAX RRHD Eclipse Plus C18 UHPLC guard cartridges (1.8 μ m, 2.1 \times 5 mm, Agilent Technologies, cat. no. 821725-901)

- Analytical PGC (Hypercarb™) HPLC Column (5 µm, 1 × 150 mm, Thermo Scientific, cat. no. 35005-151030)
- Hypercarb Guard column (5 µm, 1 × 10 mm, Thermo Scientific, cat. no. 35005-011001) with Universal Uniguard Holder (1.0 mm i.d., Thermo Scientific, cat. no. 851-00)
- 96-Well twin.tec PCR plates (Eppendorf; Thermo Fisher Scientific, cat. no. E951020401)
- Polypropylene vial (250 µL; Agilent, cat. no. 5188-2788)
- Crimp/snap-top vials and caps (2 mL; Agilent Technologies, cat. no. 5182-0541)
- E-Pure water purification system (Thermo Fisher Scientific, cat. no. D4631)
- Ultra-high performance liquid chromatography (UHPLC) system (1290 Infinity II LC system, Agilent Technologies)
- Triple Quadrupole LC/MS (Agilent Technologies, model no. 6495A)
- High-performance liquid chromatography (HPLC) system (1260 Infinity II; Agilent Technologies)
- Accurate-mass Q-TOF LC/MS system (Agilent Technologies, model no. 6530)
- Incubator/Oven (Jeio Tech, model no. OF-01E)
- Centrifuge (Eppendorf, model no. 5811F)
- Metal clamp for 96-well plates (machined using pictures and dimensions in

Supplementary Figure 2.1)

Software

All required software can be run on a standard personal computer equipped with a Windows operating system.

- Agilent MassHunter Workstation for LC/QQQ (B.08.00; Agilent Technologies)

- Agilent MassHunter Workstation for LC/TOF and LC/Q-TOF (B.08.00; Agilent Technologies)
- Agilent MassHunter Workstation Qualitative Analysis (B.08.00 Agilent Technologies)
- Agilent MassHunter Workstation Quantitative Analysis (B.08.00; Agilent Technologies)

Reagent setup

CRITICAL: Milli-Q water generated from the E-pure water purification system is used for reagent step unless other types of water are specified

1-phenyl-3-methyl-5-pyrazolone (PMP) solution for monosaccharide and linkage analysis

This solution is a mix of equal parts methanolic 0.2 M PMP and ammonia solution (28-30 % w/v). To prepare enough for one 96-well plate of samples, weigh 522.6 mg into a 50 mL Falcon tube. Dissolve completely in 15 mL of methanol. Add 15 mL of ammonia solution and vortex well.

! CAUTION: Methanol is flammable and toxic. Ammonia is a skin and respiratory irritant.

Prepare this solution in a fume hood.

Saturated NaOH solution for linkage analysis

Weigh 6.3 g of NaOH into a 15 mL Falcon tube. Add 5 mL of water and vortex until completely dissolved. This may be scaled down proportionately if less solution is needed.

! CAUTION: NaOH is corrosive and a significant amount of heat is generated from preparing the saturated solution. Ensure the lid is securely tightened before vortexing. Periodically crack the lid to relieve excess vapor pressure.

- **CRITICAL:** This solution should be used immediately after preparation. Allowing it to cool will result in a solid or slurry.

Monosaccharide analysis LC solvent A

This solvent is 5% (vol/vol) LC-MS-grade ACN in water and 25 mM ammonium acetate with a pH at 8.2. To a 1-liter volumetric flask, add 50 mL ACN and add water to a final volume of 1 L. Weigh out approx. 1.927 g of ammonium acetate and dissolve with the prepared 1 L 5% ACN mixture. Add approximately 150 μ L ammonia solution to adjust the pH to 8.2. This solution can be used for up to one week.

- **CRITICAL:** This solution should be prepared fresh before running a batch of samples.

Monosaccharide analysis LC solvent B

This solvent is 95% (vol/vol) LC-MS-grade ACN in water. To a 1-liter volumetric flask, add 50 mL Milli-Q water and add ACN to fill in a final volume of 1 L. Transfer to an LC solvent container. This solution can be used for up to one week.

- **CRITICAL:** This solution should be prepared fresh right before running a batch of samples.

Glycosidic linkage analysis LC solvent A

This solvent is 5% (vol/vol) LC-MS-grade ACN in water and 25 mM ammonium acetate with a pH at 7.7. To a 1-liter volumetric flask, add 50 mL ACN and add water to a final volume of 1 L. Weigh out 1.927 g of ammonium acetate and dissolve with the prepared 1 L 5% ACN mixture. Add approximately 60 μ L ammonium hydroxide to adjust the pH to 7.7. Transfer to an LC solvent container. This solution can be used for up to one week.

- **CRITICAL:** This solution should be prepared fresh right before running a batch of samples.

Glycosidic linkage analysis LC solvent B

This solvent is 95% (vol/vol) LC-MS-grade ACN in water. To a 1-liter volumetric flask, add 50 mL Milli-Q water and add ACN to fill in a final volume of 1 L. Transfer to an LC solvent container.

This solution can be used for up to one week.

- **CRITICAL:** This solution should be prepared fresh right before running a batch of samples.

Polysaccharide (FITDOG) analysis LC solvent A

This solvent is 3% (v/v) of ACN in water with 0.1% (v/v) FA. In a 1 L volumetric flask, add 30 mL of ACN and 1 mL of FA, fill to mark with Milli-Q water, and mix thoroughly. Fill to mark again with Milli-Q water if needed. Transfer to an LC solvent container. This solution can be used for up to one week.

- **CRITICAL:** This solution should be prepared fresh right before running a batch of samples.

Polysaccharide (FITDOG) analysis LC solvent B

This solvent is 90% (v/v) of ACN in water with 0.1% (v/v) FA. In a 1 L volumetric flask, add 90 mL of Milli-Q and 1 mL of FA, fill to mark with ACN, and mix thoroughly. Fill to mark again with ACN if needed. Transfer to an amber glass LC solvent container. This solution can be used for up to one week.

- **CRITICAL:** This solution should be prepared fresh right before running a batch of samples.

Sodium acetate buffer for FITDOG reaction

This solution is 44 mM NaCH₃COO in water at pH 5.2. Prepare 44 mM sodium acetate solution. Adjust pH to 5.2 by adding glacial acetic acid. This solution can be stored at 4 °C for several months.

FITDOG reaction mixture

This solution has 41.8 mM NaCH₃COO, 65 μM Fe(III), and 1.5% (w/v) H₂O₂. Mix 95 mL of sodium acetate buffer and 5 mL of 30% (w/v) H₂O₂. Weigh 3.56 mg of Fe₂(SO₄)₃•5H₂O and dissolve it in the prepared solution.

- **CRITICAL:** This solution should be prepared freshly each time, right before the experiment.

Quenching solution for FITDOG reaction

This solution is 2 M NaOH in water. Weigh 4.0 g of NaOH and dissolve it in 50 mL water.

- **CRITICAL:** This solution should be prepared freshly each time, right before the experiment.

Reducing solution for FITDOG analysis

This solution is 1 M NaBH₄ solution. Weigh 1.89 g of NaBH₄ and dissolve it in 50 mL water.

- **CAUTION:** Dissolving NaBH₄ in water forms H₂ gas. Depressurize container a few times while dissolving NaBH₄.

- **CRITICAL:** This solution should be prepared freshly each time, right before the experiment.

PGC Priming Solution

This solution is 80% ACN with 0.1% TFA. To prepare 1 L of solution, mix 800 mL ACN, 1 mL TFA, and 199 mL of water. This solution can be stored at 4 °C for several months.

PGC Elution Solution

This solution is 40% ACN with 0.05% TFA. To prepare 1 L of solution, mix 400 mL ACN, 0.5 mL TFA, and 599.5 mL water. This solution can be stored at 4 °C for several months.

Pooled calibration standards for quantitative monosaccharide analysis

Weigh 14 mg of each of the 14 monosaccharide standards into separate 1.5 mL tubes. Add water to prepare solutions of exactly 14 mg/mL. Pool equal aliquots together to prepare a 1 mg/mL pooled stock solution. Serially dilute the 1 mg/mL pool according to the calibration levels in **Supplementary Table 2.1**. Once prepared, the stock solution and calibration standards may be stored at -20 °C and kept for several months.

Pooled oligosaccharide standards for linkage analysis

Weigh 10 mg of each of the linkage standards into separate 1.5 mL tubes. Add water to prepare stock solutions of 10 mg/mL. Combine equal aliquots of each to create a pooled stock solution to be used directly for linkage analysis. Once prepared, the stock solutions may be stored at -20 °C and kept for several months.

Calibration standards for polysaccharide (FITDOG) analysis

Several polysaccharide standards may be pooled together to reduce the number of samples. A recommended pooling scheme is summarized in **Supplementary Table 2.2**. Weigh polysaccharides into 2-mL screwcap tube and add 1 mL of water, incubate at 100 °C for 1 hr, and then homogenize with stainless-steel beads. If needed, calibrator stock mixtures may be stored in –20 °C for several weeks. Serial dilution should be carried out right before the experiment.

PROCEDURE

Sample Preparation

Lyophilization (freeze-drying) and dry homogenization of samples *Timing: 16-72 h, depending on the moisture content of the samples.

1. Collect the sample, place it in an appropriately-sized screw cap tube, and freeze in a –80 °C freezer for at least 3 h. If moisture content is needed, record the mass of the samples and the tubes.
2. Start the freeze-dryer by turning on the condenser. It will take several minutes for the temperature of the condenser tray to reach the appropriate temperature (–60 °C). Once the condenser tray is cold enough, start the vacuum.
3. Remove the frozen sample tubes from the freezer, remove the lids, and place them gently back on top of their respective tubes.
 - **CRITICAL:** Ensure the cap is completely free from the thread on the tube as the loss of pressure may cause the tube to seal.
4. Gently place the frozen sample tubes in the glass freeze-dryer jars, fasten them to the vacuum manifold, and lyophilize until a minimum moisture content is achieved. This can

be done by weighing the sample tube in intervals. However, we have found that three days is sufficient to dry nearly all sample types without the necessity of weighing. If moisture content is needed, record the mass of the dried sample.

5. If the samples are powdered after drying, they may be homogenized directly with a bead mill (step 6). If samples are oily (like nut butters) or appear hygroscopic (like many fruits), they must be flash-frozen with liquid nitrogen and homogenized with a mortar and pestle.
6. Transfer the lyophilized samples into separate 30 mL screw-cap tubes and grind the samples using a bead mill homogenizer with 5 mm stainless-steel beads for 2 mins at 4 m/s.
 - **CRITICAL:** Ensure samples are a homogenous powder before moving forward.

Preparation of sample suspensions *Timing: 2 h

7. If removal of low-molecular weight saccharides (mono-, di-, and oligosaccharides) before compositional analysis is desired, follow option A. Otherwise, follow option B.

- **A. Removal of low-molecular weight saccharides by ethanol precipitation**

- ***Timing: 3 h**

- i. Weigh out 10 mg (± 0.5 mg) of samples into 1.5 mL screw-cap tubes using an analytical balance and record the mass. Add 1 mL of 80 % EtOH.
- ii. Homogenize the samples on a bead mill at 4 m/s for 1 min. Centrifuge at 10,000 g for 10 min. Carefully remove the supernatant without disturbing the pellet using a pipette, add another 1 mL of 80% EtOH, homogenize, and centrifuge again to wash the pellet. Repeat the wash once more.

- iii. Remove the supernatant from the final wash and dry the resulting pellets completely in a centrifugal vacuum evaporator. This takes approximately 1 h, depending on how much of the supernatant was successfully removed.
 - PAUSE POINT: The dried pellet may be stored at $-20\text{ }^{\circ}\text{C}$ until further preparation
 - iv. Add 1 mL of water and 2-mm stainless-steel beads to the sample pellets.
 - v. Homogenize the samples on a bead mill at 4 m/s for 2 min. Incubate the suspended samples at $100\text{ }^{\circ}\text{C}$ for 1 h before bead milling once more with the same settings.
 - PAUSE POINT: The suspended sample stocks may be stored at $-20\text{ }^{\circ}\text{C}$ before further analysis. If frozen, samples should be homogenized again via bead mill after thawing.
- **B. Preparation of sample stock suspensions without EtOH precipitation**
 - *Timing: 2 h**
 - i. Weigh out 10 mg (± 0.5 mg) of the samples into 1.5 mL screw-cap tubes using an analytical balance and record the mass. Add 1 mL of water and 2 mm stainless-steel beads.
 - ii. Homogenize the samples on a bead mill at 4 m/s for 2 min. Incubate the suspended samples at $100\text{ }^{\circ}\text{C}$ for 1 h before homogenizing once more with the same settings.
 - PAUSE POINT: The suspended sample stocks may be stored at $-20\text{ }^{\circ}\text{C}$ before further analysis. If frozen, samples should be homogenized again via bead mill after thawing.

Quantitative Monosaccharide Analysis: Acid Hydrolysis * Timing: 1.5 h

8. Aliquot 10 μL from the sample stock suspension into each well of a 96-well plate or into 1.5 mL screw cap tubes. Add 90 μL water to each sample.
 - **CRITICAL:** Stock suspensions should be vortexed before pipetting to ensure homogeneity.
9. Add 44.5 μL of TFA to each well/tube. Seal the plate or tubes immediately to avoid evaporation. Vortex lightly using a plate shaker/vortex mixer for 1 min and centrifuge for 30 s at 300 g.
 - **CAUTION:** TFA is a corrosive chemical. Use only in a fume hood.
10. After centrifuging, incubate the 96-well plate/tubes at 121 $^{\circ}\text{C}$ for 1 h.
 - **CRITICAL:** If using a 96-well plate, the plates must be sealed well using a clamp. These can be machined quite easily (see pictures and dimensions in **Supplementary Figure 2.1**).
11. Once the incubation is complete, remove from the oven, and allow to cool to room temperature. Once cooled, remove from the clamp and centrifuge at 300 g for 30 s.
 - **CAUTION:** The samples and clamps are extremely hot at after incubation. Handle only with heat-resistant gloves.
12. Add 855.5 μL of ice-cold water. Centrifuge at 300 g for 30 s.

Quantitative Monosaccharide Analysis: PMP Derivatization * Timing: 45 min

13. Transfer 10 μL of the hydrolyzed sample solution to another 96-well plate or screw cap tube. Transfer 50 μL of each level (L1-L10) of the monosaccharide calibration curve.
14. Transfer 200 μL of the PMP solution to each well or tube.

15. Seal the plate or tubes, vortex lightly for 1 min and centrifuge at 300 g. Incubate at 70 °C for 30 min.

- **Critical Step:** 96-well plates must be sealed and incubated with the clamp during the reaction.

16. Once the incubation is complete, remove the samples from the incubator, and allow to cool to room temperature. Centrifuge at 300 g for 30 s.

- **CAUTION:** The samples and clamps are hot at after incubation. Handle only with heat-resistant gloves.
- **CRITICAL:** 96-well plates should remain securely sealed while cooling and should be centrifuged before removing the lid.

17. Once cooled, remove the lids, place the samples in a centrifugal vacuum evaporator fitted with a pressure programmer and dry completely. This will take at least overnight to dry.

- **CRITICAL:** The dryer should be programmed for the stepwise evaporation of the MeOH/ammonia/water solution to avoid solvent bumping. We have optimized this to include an initial pressure drop from atmosphere to 200 mbar followed by a gradient drop to 10 mbar over 4 h.
- **PAUSE POINT:** The dried and derivatized samples may be stored at -20 °C for several weeks before subsequent steps.

Quantitative Monosaccharide Analysis: Chloroform Extraction * Timing: 30 min

18. Add 250 µL chloroform into each tube or well of the 96-well plate containing sample and calibration standard. Vortex on a plate shaker until the pellet is nearly completely dissolved. Add 250 µL water, vortex for 1 min on a plate shaker, and centrifuge at 300 g for 1 min.

- CAUTION: Chloroform is flammable, toxic, and will dissolve most disposable solvent reservoirs. Perform these steps in a fume hood with a glass solvent reservoir.

19. Remove and discard 150 μ L of the chloroform layer (bottom) from each well. Add another 250 μ L chloroform into each well. Vortex for 1 min and centrifuge at 300 g.

- CAUTION: Perform this extraction in a fume hood and discard waste in an appropriate hazardous waste vessel.

20. Transfer 100 μ L of the top aqueous layer from each well into a 96-well analysis plate or autosampler vials compatible with the autosampler to be used. Centrifuge before injection.

- CRITICAL: Ensure that only the aqueous layer is transferred.

Quantitative Monosaccharide Analysis: UHPLC-QqQ MS analysis *Timing: Depends on batch size (~5 min per sample)

21. Start the MassHunter Acquisition software on the UHPLC-QqQ MS and place the 96-well plate or vials into the autosampler compartment.

- CRITICAL: At the beginning of the batch, run at least 3-5 blanks and a mid-level calibration standard to equilibrate the LC system. For instrument quality control (QC), a mid-level calibration standard should be injected every 12 samples along with a blank sample.

Troubleshooting: (Make sure to the correct assign well plate in the autosampler compartment. When loading the method section, the needle position needs to adjust to be able inject sample. Check the vial sensing tab.)

22. Make a worklist using the “Worklist” Tab or load a worklist created in the Offline Worklist Editor.

23. Load the freshly made solvent A and B for monosaccharide analysis onto the pump. Update the solvent level in “Bottle Filling” section in the software. Purge both solvents for at least 5 min at a flowrate of 5 mL/min. A detailed list of the LC-MS parameters is included in **Table 2.1**.

24. Install an Agilent Poroshell HPH C18 column (2.1 mm × 50 mm, 1.8 µm particle size) equipped with an Agilent Poroshell HPH C18 guard cartridge (2.1 mm × 5 mm, 1.8 µm particle size) into the column compartment. After purging, turn on the pump and allow to equilibrate for at least for 10 min.

- **CRITICAL:** Remember to update the column position after installing it. During the conditioning, start from a lower flowrate like 0.2 mL/min and increase stepwise to match the operating flowrate.

25. Start the worklist after purging, conditioning, and monitoring the first QC using MassHunter Qualitative B.08.00 software. Extract the transitions for the 14 monosaccharides and assign their retention time using **Supplementary Table 2.3**. Adjust the retention time of the dMRM transitions in the Acquisition tab, if needed.

Troubleshooting: If no signal is observed or peaks are cut off, expand the retention time windows in the dMRM table to capture all peaks and then narrow them down afterwards.

- **CRITICAL:** Observe the QCs throughout the run to see if they are reproducible in terms of retention time and abundance.
- **PAUSE POINT:** After finishing the batch, the plate or tubes can be stored at –20 °C. Samples can be stored for up to 2 weeks for re-injection. Wrap well with aluminum foil before doing so.

Quantitative Monosaccharide Analysis: Data Analysis * Timing: 30 min – 60 min

26. Start the MassHunter Quantitative Analysis B.08.00 software and open a new batch under the data file folder. In the window “Add Samples,” select monosaccharide calibration curve standards and samples and click “OK”.
27. After loading samples, label calibration standards with their Level Name as listed in **Supplementary Table 2.1** and update their type as “Cal” for calibration.
 - CRITICAL: If the columns for “Level” and “Type” are not shown, right click to “Sample” column to add. Sample type can be modified in the worklist as well.
28. Click ‘Method’ --> “New”--> “New Method from Acquired MRM Data” and select one of the mid-level calibration standards from this batch. Click the “MRM Compound Setup” tab and a list of compound names, transition and RT will be shown. Update the list of 14 monosaccharide compounds.
29. Assign 14 monosaccharides based on the elution order of isomers shown in **Supplementary Table 2.3**. For isomers that are monitored using the same dMRM transitions, select the compound and right click to “Duplicate compound,” update the compound name, and assign the retention time.
30. Click “Qualifier Setup” tab, make sure the quantifier and qualifier transitions match with **Supplementary Table 2.3**. For each compound, quantifier has a product ion of 175.2 m/z and a qualifier ion of 217.1 m/z (GlcNAc/GalNAc are the exception with qualifier ions of 258.1 m/z . The precursor ion should match for both. If any compound does not have a qualifier, right-click the compound and add “New Qualifier” manually.
31. Click “Concentration Setup,” select a compound to add “New Calibration Level”. Add 10 calibration levels and update the “Level” and “Conc.” sections as shown in **Supplementary Table 2.1**. Apply the calibration level to all compounds. After setting the

calibration curve for one compound, right click and select “Copy Calibration Level to...” and select all compounds.

32. Save as a new method and click exit to “Analyze” the batch. This method can be saved for future analysis requiring that the user need only update the retention times of the compounds.
33. Go to “View” and click “Compounds-at-a-glance.” A window displaying the integrated peaks for each compound in each sample will come up. Check the integrations of each compound and correct if needed.
34. The calibration curve for each monosaccharide is displayed in the “Calibration Curve” box. Exclude high-end data points, if necessary, based on the r^2 value. Retain at least 6 points. Click “Analyze Batch” again to apply any corrected integration in the calibration curves to the samples.

Troubleshooting: Calibration curves should have an r^2 of 0.99 or greater. High concentration data points may be omitted but ensure that at least 6 data points are used. If any of the calibration curves fall well below this value, the samples (along with a new calibration curve) may need to be derivatized and run again.

35. Go to “File” and export the table as an Excel (.xlsx) file.
 - **CRITICAL:** Ensure the “Final Conc.” for each compound is exported. To do this, ensure “Multiple Sample/Compound View” is displayed in the software. Click “Add/Remove Columns” and only include the “Final Conc.” column for export.
36. After exporting the table, convert the sample concentration to mg of monosaccharide per milligram of dried material (mg/mg). The default sample concentration unit expressed in the software is ng/mL. First, convert the sample units to mg/mL by dividing by 10^6 . During

the acid hydrolysis, each sample is 100-fold diluted. During the PMP derivatization, the sample is 5-fold diluted relative to the standard calibration curve. Therefore, a dilution factor of 500 is needed to calculate the sample concentration. Lastly, the sample concentration is divided by the mass weighed to create its stock solution to arrive at a unit of mg/mg.

Glycosidic Linkage Analysis: Permethylation * Timing: 4 h

37. Pipette 5 μ L from the 10 mg/mL sample stock suspension (approx. 50 μ g sample) and the pooled linkage standards to a 96-well plate or 1.5 mL screw cap tube. Dry completely under vacuum centrifugation.

- **CRITICAL:** Stock suspensions should be vortexed before aliquoting to ensure homogeneity.

38. Add 5 μ L of the saturated NaOH solution to the 96-well plate or 1.5 mL tubes containing the samples. Replace the plate/tube lid and centrifuge at 300 g for 1 min. Place the 96-well plate/tubes on a plate shaker for 30 min.

- **CRITICAL:** Ensure the 5 μ L of NaOH solution is centrifuged down to the bottom of the well/tube to redissolve the sample.

39. Stop the shaker after 30 min and purge the 96-well plate/tubes with argon in the vacuum chamber.

40. Purge the DMSO bottle with argon and transfer to a glass solvent reservoir using a needle and syringe. Add 150 μ L of argon-purged DMSO into each well. Centrifuge at 300 g and vortex for 30 min in the argon-purged chamber.

- **CRITICAL:** DMSO can be purged using a Schlenk line setup. Remove air in the bottle by vacuum and purge with argon. Use a polypropylene syringe and clean cannula to transfer purged DMSO in a glass reservoir.
- **CAUTION:** DMSO is toxic and flammable. Use only in the fume hood.

Troubleshooting (permethylation is not working, significant underpermethylation is observed)

41. Add 40 μL iodomethane into each well/tube. Centrifuge at 300 g for 30 s and vortex for 50 mins on a plate shaker.

- **CAUTION:** Iodomethane is toxic and an oxidizer. Use only in a fume hood.

42. Quench the reaction with 700 μL ice-cold water and 300 μL dichloromethane. Vortex and centrifuge at 300g for 1 min.

43. Remove and discard 700 μL of the water layer (top) and add another 700 μL fresh water to each well. Vortex and centrifuge at 300g for 1 min Repeat this step 2 times.

- **CRITICAL:** Graduated 1-mL pipette tips help to visualize alignment during the water removal step when using a plate and multi-channel pipette.

44. After 3 extractions, remove as much of the water layer as possible without disturbing the DCM layer and dry the plate using vacuum centrifugation.

Glycosidic Linkage Analysis: Acid Hydrolysis * Timing: 2.5 h

45. Reconstitute the dried, permethylated sample with 60.5 μL water and vortex the samples for several minutes.

46. Add 30.5 μL of TFA and seal the plate/tubes immediately to avoid evaporation. Vortex for 1 min and centrifuge at 300g.

- CAUTION: TFA is a corrosive chemical. Use only in a fume hood.
- CRITICAL: If using a 96-well plate, the plates must be sealed well using a clamp. These can be machined quite easily (see pictures and dimensions in **Supplementary Figure 2.1**) or purchased from labware manufacturers such as the Artic White SecureClamp.

47. Incubate the 96 well plate/tubes at 100 °C for 2 h.

- CRITICAL: If using 96-well format, the plate must be tightly sealed with the clamp to avoid evaporation.

48. Once the incubation is complete, remove from the oven, and allow to cool to room temperature. Once cooled, centrifuge at 300 g for 30 s and place the samples in a centrifugal vacuum evaporator. It usually takes 2-4 h to dry completely.

- CAUTION: The samples and clamps are extremely hot at after incubation. Handle only with heat-resistant gloves.
- CRITICAL: The lids (if using 96-well plate) should be securely sealed during the cool down process and the plates should be centrifuged prior to opening.

Glycosidic Linkage Analysis: PMP Derivatization * Timing: 45 min

49. Transfer 200 µL of PMP solution to each well for derivatization.

- CAUTION: MeOH is flammable and NH₄OH is corrosive. Please add the solution in a fume hood.

50. Seal the plate or tubes and vortex for 1 min and centrifuge at 300 g. Incubate at 70 °C for 30 min.

- **CRITICAL:** 96-well plates must be sealed and incubated with the clamp during the reaction.

51. Once the incubation is complete, remove the samples from the incubator, and allow to cool to room temperature. Centrifuge at 300 g for 30 s.

- **CAUTION:** The samples and clamps are hot at after incubation. Handle only with heat-resistant gloves.

52. Once cooled, place the samples in a centrifugal vacuum evaporator fitted with a pressure programmer (more detailed can be found in Step 17) and dry completely. This will take at least overnight.

- **PAUSE POINT:** The dried and derivatized samples can be stored at -20 °C for several weeks until instrument analysis.

Glycosidic Linkage Analysis: UHPLC-QqQ-MS analysis * Timing: 16 min per sample

53. Add 70 μ L MeOH and 30 μ L Nanopure water into each well/tube. Vortex and centrifuge at 300 g for 1 min at RT.

- **CRITICAL:** Ensure methanol dissolves dried pellet before adding Nanopure water.

54. Transfer 70 μ L from each sample into an injection 96 well plate/injection vial. Centrifuge before injection.

55. Start the Acquisition software and place 96-well injection plate into the autosampler compartment and close the door probably.

- **CRITICAL:** At the beginning of the batch, run at least 3-5 blanks and a calibration standard point to equilibrate the LC system. For QC, an oligosaccharide/polysaccharide standard could be injected every 12 samples along with a blank sample.

Troubleshooting (Make sure to assign well plates in the autosampler compartment. When loading the method section, the needle z-axis position may need to be adjusted to sense the bottom of the well.)

56. Make a worklist using the “Worklist” Tab or load the worklist created by the “Offline Worklist Editor”

57. Load the freshly made solvent A and B for monosaccharide analysis into the pump.

Update the solvent level in “Bottle Filling” section in the software. Purge both solvents for at least 5 mins at 5 mL/min. The detailed glycosidic linkage analysis LC-MS method is included in **Table 2.1**.

58. Load the column with guard column in the LC compartment. After the purging, turn on the pump and condition the column for 10 min.

- **CRITICAL:** Update the column position after loading component. If using a new column, condition the column using low flow rate. Start from a lower flowrate like 0.2 mL/min and increase stepwise to match the desired pressure (approx. 450 bar).

59. Start the worklist after purge and condition and monitor the first QC using MassHunter Qualitative B.08.00 software.

Troubleshooting: If retention times have shifted significantly since last run or peaks are coeluting, change the flowrate and make sure the solvent is fresh and the pH is around 7.7.

60. Observe the QCs throughout the run to see if they are reproducible.

61. After finishing the batch, change the lid and save the injection plate with aluminum wrap in -20°C . The samples can be stored for 1 week for re-injection.

Glycosidic Linkage Analysis: Data analysis for glycosidic linkage * Timing: 1h, depending on the size of the batch

62. Start the MassHunter Qualitative Analysis B.08.00 software and open the data files.

Assign the retention time based on **Figure 2.3b** and **Supplementary Table 2.4** by extracting each transition.

- **CRITICAL:** Make sure QCs are reproducible and retention time shifts are minimal. Each glycosidic linkage has 2-3 transitions: if the product ion 217.2 is more abundant than 231.2, this linkage is most likely 2-linked; if the 231.2 is more abundant than 217.2, this linkage is most likely not 2-linked.

63. Start MassHunter Quantitative Analysis B.08.00 software and open a new batch under the data file folder.

64. When a window “Add Samples” pop out, select samples and click “OK”.

65. Click “Method” --> “New” --> “New Method from Acquired MRM Data” and select one of the data files from this batch.

66. Click “MRM Compound Setup” tab and a list of compound names, transition and RT will be shown. Update the compound list from the previous assignment.

- **CRITICAL:** Assign the retention time for each glycosidic linkage based on the elution order of isomers shown in **Supplementary Table 2.4**.
67. Click “Qualifier Setup” tab, make sure the quantifier and qualifier match with **Supplementary Table 2.4**.
- **CRITICAL:** For 2-linked glycosidic linkages, make sure the qualifier product ion is 217.1 m/z. Otherwise, the product ion for qualifier should be 231.2 m/z. The precursor ion should match with each other. If the compound does not have a qualifier, right-click the compound and add “New Qualifier” manually.
68. Save as a new method and click exit to “Analyze” the batch. This method could be saved for future analysis and only update the retention time.
69. Go to “View” and click “Compounds-at-a-glance”. After a window popping out, check the integrations of each compound. Go to “File” and export Table as excel file.
- **CRITICAL:** The “Area” for each compound should be exported. Enable the “Multiple Sample/Compound View” display in the software. Click “Add/Remove Columns” and only include “Area” column for export.
70. After exporting the table, rearrange the table order as needed. Relative composition of glycosidic linkage by peak area of each sample could be graphed using Excel.

Polysaccharide (FITDOG) Analysis: Depolymerization reaction * Timing: 2 h

71. Pipette 100 μ L (approx. 1.0 mg sample) from the 10 mg/mL stock suspension of the sample to a 96-well plate or 1.5 mL screw cap tube. For calibrator standards, the same volume, 100 μ L, is used for analysis.

- **CRITICAL:** Stock suspensions should be vortexed before aliquoting to ensure homogeneity.

72. Add 900 μ L of freshly made FITDOG Reaction Mixture to each well/tube. Mix the reaction mixture using pipette (for plate), or vortex (if using tubes).

73. Seal the plate or tubes and incubate at 100 $^{\circ}$ C for 45 min.

- **CRITICAL:** If using 96-well format, the plate must be tightly sealed with the clamp to avoid evaporation.

74. Remove samples from the oven and allow to cool to room temperature for around 10 min.

- **CAUTION:** The samples and clamps are extremely hot at after incubation. Handle only with heat-resistant gloves.

75. Slowly add 500 μ L of the quenching solution (2 M NaOH, freshly made). Slowly mix with repeated pipetting.

- **CRITICAL:** Some bubbles may form so the quenching solution should be added as slow as possible to prevent cross well contamination.

76. Slowly add 61 μ L of glacial acetic acid. Slowly mix with repeated pipetting.

- **CRITICAL:** Some bubbles may form so the acetic acid should be added as slow as possible to prevent cross well contamination.
- **CRITICAL:** Glacial acetic acid is corrosive. Use only in a fume hood.

77. Transfer 400 μ L into a new clean 96-well plate, or tubes, for the reduction.

Polysaccharide (FITDOG) Analysis: Oligosaccharide reduction * Timing: 2 h

78. Slowly add 400 μL of the reducing solution (1 M NaBH_4 , freshly made) into each well/tube.

- **CRITICAL:** Bubbles will form so the reducing solution should be added as slow as possible to prevent cross well contamination. Cooling down the samples and the reducing solution with ice bath prior to mixing will minimize the formation of bubbles.

79. Loosely place the plate lid or tube cap. Do not seal the lids as gas is formed during the reaction.

80. Incubate in the oven at 65 $^{\circ}\text{C}$ for 1 hr.

81. After oven incubation, remove from oven and cool down to room temperature.

- **CAUTION:** The samples will be hot at after incubation. Handle only with heat-resistant gloves.

Polysaccharide (FITDOG) Analysis: Solid phase extraction (SPE) * Timing: 2 h

All the subsequent centrifugation steps should be at 1000 g for 1 min, unless otherwise stated.

82. Prime the C18 SPE plate by adding 250 μL ACN to each well and then centrifuge.

Repeat once. Discard the washings.

83. Condition the C18 SPE with 250 μL water and then centrifuge. Repeat 3 more times.

Discard the washings.

84. Transfer the C18 SPE plate to a clean collection plate. Load 400 μL of the reduced oligosaccharide sample. Centrifuge and collect the flow-through.

- **CRITICAL:** Make sure that the SPE sits on a clean collection plate before loading the sample.

85. Load the remainder of the reduced sample, centrifuge, and collect flow-through. This is the end of the C18 SPE.
86. Prime the PGC SPE plate with 400 μ L of each solution in the following order: (1) water, (2) PGC priming solution (80% ACN/ 0.1% TFA), (3) water. Centrifuge between each step and discard washings.
87. Load 400 μ L of the C18-cleaned sample. Centrifuge and discard flow-through. Load the remainder of the sample.
88. Wash the bound oligosaccharides in the PGC plate with 400 μ L water. Centrifuge and discard washing. Repeat this step 5 more times.

Troubleshooting: Some of the sample or water is left in the SPE wells after centrifugation. If this happens, centrifuge for the second time at higher speed (1300 *g*).

89. Change the collection plate to a clean one. Elute the oligosaccharides with 400 μ L of the PGC elution solution (40% ACN/ 0.05% TFA) and centrifuge at 1000 *g* for 2 min.
- **CRITICAL:** Make sure that the SPE sits on a clean collection plate before adding the elution solution.
90. Place the samples in a centrifugal vacuum evaporator to dry completely. The drying will take at least 12 hr.
- **PAUSE POINT:** The dried samples may be stored at -20 °C for several weeks before the subsequent steps.

Polysaccharide (FITDOG) Analysis: HPLC-qToF MS analysis *Timing: 45 min per sample

91. Reconstitute samples with 50 μ L water and vortex mix for 15 min. Centrifuge at 1000 *g* for 1 min.

92. Transfer 50 μ L into LC plate or vials. Centrifuge at 1000 *g* for 1 min.
93. Start the MassHunter Acquisition software on the HPLC-qToF MS and place the sample plate or vials in the autosampler compartment.
94. Make a worklist sequence using the “Worklist” tab in the acquisition software or in the Offline Worklist Editor software.
- **CRITICAL:** It is recommended to run a sequence of blanks (at least 2) first, then a standard to check the performance of the instrument. A mixture of starch and cellulose can be used as a workflow QC. Calibrator standards are also recommended to be injected first before the samples. Inject instrument QC every 10-12 samples.
95. Load the freshly made solvent A and B for polysaccharide (FITDOG) analysis. Make sure to update the solvent level in the “Bottle Filling” section in the software. Purge pump for at least 10 min at a flow rate of 5 mL/min at 50% A/ 50% B composition.
96. Install a Thermo Scientific Hypercarb PGC column (1 mm \times 150 mm, 5 μ m particle size) equipped with guard column (Hypercarb, 1 mm \times 10 mm, 5 μ m particle size) in the column compartment.
- **CRITICAL:** New columns should be conditioned first, starting at a lower flow rate and increasing stepwise to match the operating flowrate (0.132 mL/min).
97. Start the worklist sequence after purging and conditioning the LC column. The complete parameters for the method are summarized in **Table 2.1**. Monitor the backpressure during the first run and check the oligosaccharide peaks from an initial QC run using MassHunter Qualitative software.

Troubleshooting: LC pump pressure is abnormally high or is fluctuating. This can be caused by a clog in the LC tubings or in the column. Identify the source of the clog starting from the tubes from the pump to the column. Guard column may also be replaced.

98. Throughout the run, monitor QC signals in terms of retention times and ion count abundances.

Troubleshooting: RT shifts of > 5 min or inconsistent RT shifts may complicate the peak area integration and alignment. Flush the column with multiple alternating rounds of high aqueous and high organic compositions.

Troubleshooting: Fluctuating or decreasing ion signals throughout a batch may indicate problems with the ESI source or the QTOF instrument. ESI may be visually checked and cleaned to minimize background noise signal. Check tune can be done to quickly assess the performance of the instrument and to re-calibrate the m/z axis.

Polysaccharide (FITDOG) Analysis: Data analysis *Timing: depends on the batch size

99. Open MassHunter Qualitative Analysis software and load the LC-MS/MS files.

100. Using the calibrator standards, extract ion chromatograms (XIC) of each relevant oligosaccharide precursor m/z values and take note of the retention times. Refer to **Supplementary Table 2.5** for the complete list of oligosaccharide library.

- CAUTION: Confirm monosaccharide class compositions from the tandem mass spectra.

101. Open MassHunter Q-TOF Quantitative Analysis software and open a new batch in the same directory as the files. In the window “Add Samples”, include all relevant sample files and calibrator standards.
102. Edit the method (“Method” --> “Edit”, or F10) and update the compound and retention time tables.
103. Save as a new method and click “Exit” and “Analyze” the batch.
 - CRITICAL: This method can be saved for future analysis requiring that the user need only update the retention times of the compounds.
104. In the main window, click “View” and then “Compounds-at-a-glance” to view all XICs. Check the integrations of each compound and sample and correct if needed.
105. After doing the peak integrations, close the “Compounds-at-a-glance” window, save the batch file, and then export the peak area table as .xlsx or .csv file.
106. Using Excel or other spreadsheet software, quantitation can be done on the peak areas. For each polysaccharide, identify the top 3 most abundant oligosaccharides, get the average peak area of these 3 oligosaccharides, and plot it against the initial concentration ($\mu\text{g/mL}$ or mg/mL) used in the workflow.
 - CRITICAL: At least 5 points should be used for the calibration curve. Linear or quadratic fit can be used to generate the calibration curve for each polysaccharide.
107. Apply the top 3 averaging method to the samples and use the calibration curves to interpolate the quantity of polysaccharides present in the sample. Concentrations can be converted into mg/mg units by dividing by the weighed mass of the sample in mg .
108. Convert mg/mg dry basis to fresh weight basis by using the moisture content of the sample

Table 2.1. LC-MS/MS data acquisition parameters for monosaccharide, linkage, and polysaccharide analysis.

	Polysaccharide (FITDOG)	Monosaccharide	Linkage
LC parameters			
Column packing material	PGC	C18	C18
Typical injection volume (uL)	10	2	2
Solvent A (vol/vol)	0.1% FA, 3% ACN, 96.9% H ₂ O	25 mM ammonium acetate in 5% ACN (pH 8.2)	25 mM ammonium acetate in 5% ACN (pH 7.7)
Solvent B (vol/vol)	0.1% FA, 90% ACN, 9.9% H ₂ O	95% ACN, 5% H ₂ O	95% ACN, 5% H ₂ O
Flow rate (mL/min)	0.132	1.05	
Gradient (%B)	0-15 min: 3%-25%	0-1.9 min: 11%	0-5 min: 21%
	15-18 min: 25%	1.9-2.2 min: 11%-99%	5-9 min: 21%-22%
	18-30 min: 25%-99%	2.2-3.8 min: 99%	9-11 min: 22%
	30-32 min: 99%	3.8-4.6 min: 11%	11-13.6 min: 22-24.5%
	32-34 min: 99%-3%		13.6-13.8 min: 99%
	34-45 min: 3%		13.8-16 min: 21%
ESI source parameters			
Polarity	Positive	Positive	Positive
Drying gas temperature (°C)	100	290	290
Drying gas flow (L/min)	9	11	11
Sheath gas temperature (°C)	150	300	300
Sheath gas flow (L/min)	11	12	12
Nebulizer (psi)	20	30	30
Capillary voltage (V)	1800	1800	1800
Nozzle voltage (V)	1500	1500	1500
Fragmentor (V)	65		
Skimmer (V)	50		
Oct 1 RF Vpp (V)	500		
High pressure RF (V)		150	150
Low pressure RF (V)		60	60
MS parameters		See MRM table in Supplementary Table 3	See MRM table in Supplementary Table 4

m/z range	(MS) 250-3000, (MS/MS) 50-2000
Cycle time	4.86 s
Acquisition mode	Auto MS/MS (DDA)
MS scan rate	1 spectra/s (1000 ms/spectrum)
MS/MS scan rate	1.33 spectra/s (752 ms/spectrum)
MS threshold	Absolute threshold, 50; relative threshold, 0.01%
MS/MS threshold	Absolute threshold, 5; relative threshold, 0.01%
Calibrant ion	922.009798 (\pm 35 ppm)
Activation type	CID
Activation energy	CE = $1.45 \times (m/z) / 100 - 3.5$
Max. precursors per cycle	5
Precursor selection threshold	Absolute threshold, 50; relative threshold, 0.01%
Precursor target (counts/spectrum)	25,000
Dynamic exclusion	Excluded after 2 spectra, released after 0.5 min
Precursor charge state preference	2, 1, unknown, 3, >3
Isotope model	Common organic molecules

Troubleshooting Table

Step	Problem	Possible Reason	Solution
25	No signal observed or peaks appear cut-off in monosaccharide analysis chromatograms	Retention time shifting has caused compounds to fall outside the dMRM windows or autosampler is not set-up properly	Adjust retention time windows to reflect updated retention times in the instrument method. Ensure the sample plates are oriented correctly and the autosampler is set to draw from the bottom of the sample well
34	Calibration curves in monosaccharide analysis are not linear	Standards/samples were not derivatized properly	Re-aliquot hydrolyzed samples along with standards, derivatize, and run again
40	Permethylation is not working or significant underpermethylation is observed (high trisecting linkage abundances)	Samples were not adequately dissolved or the permethylation reaction was not carried out properly	Prepare and perform linkage analysis on a fresh set of samples

55	No signal observed in the linkage analysis chromatograms	Sample plates are not oriented properly in the autosampler correctly or the autosampler is not set-up properly	Ensure the sample plates are oriented correctly and the autosampler is set to draw from the bottom of the sample well
60	Retention times have significantly changed since last run or peaks are coeluting	The flowrate in the method may need to be changed or the LC solvent is old or not at the correct pH	Adjust the flowrate in the method if needed. Ensure the LC solvent is freshly prepared and the pH is 7.7.
88	Some of the sample or water is left in the SPE wells after centrifugation.	This sometimes happen due to the hydrophobic nature of the PGC.	Centrifuge for the second time at higher speed (1300 g).
97	LC pump pressure is abnormally high or is fluctuating.	This can be caused by a clog in the LC tubings or in the column.	Identify the source of the clog starting from the tubes from the pump to the column. Guard column may also be replaced.
98	Retention times are shifting.	Possibly caused by clogged, fouled, or old guard or main column.	Flush the column with multiple alternating rounds of high aqueous and high organic compositions. Guard and/or the main column may be replaced.
98	Ion signals are fluctuating or decreasing throughout the batch run.	This may indicate problems with the ESI source or the QTOF instrument.	ESI may be visually checked and cleaned to minimize background noise signal. Check tune can be done to quickly assess the performance of the instrument and to re-calibrate the m/z axis.

Timing

- Step 1-6, lyophilization and dry homogenization of samples: 16-72 h, depending on moisture content of samples
- Step 7A, removal of low-molecular weight saccharides by ethanol precipitation and preparation of sample stock suspensions: 3 h, depending on the size of the batch
- Step 7B, preparation of sample stock suspensions without precipitation: 2 h, depending on the size of the batch
- Step 8-12, monosaccharide acid hydrolysis: 1.5 h
- Step 13-17, monosaccharide PMP derivatization: 45 min
- Step 18-20, monosaccharide chloroform extraction: 30 min
- Step 21-25, monosaccharide UHPLC-QqQ MS analysis, 5 min per sample
- Step 26-36, monosaccharide data analysis, 30-60 min
- Step 37-44, linkage permethylation: 4 h
- Step 45-48, linkage acid hydrolysis: 2.5 h
- Step 49-52, linkage PMP derivatization: 45 min
- Step 53-61, linkage UHPLC-QqQ MS analysis: 16 min per sample
- Step 62-70, linkage data analysis: 1 h, depending on the size of the batch
- Step 71-77, FITDOG depolymerization reaction: 2 h
- Step 78-81, FITDOG oligosaccharide reduction: 2 h
- Step 82-90, FITDOG SPE clean-up: 2 h
- Step 91-98, FITDOG HPLC-qTOF analysis: 45 min per sample
- Step 99-108 FITDOG data analysis: 12-16 h per 96-well plate

ANTICIPATED RESULTS

Dietary fiber consumption of human microbiota-colonized mice

Extensive research on the human gut microbiome have been enabled by using microbiota-colonized mouse models. In this approach, germ-free mice are colonized by a consortium of bacteria (from cultures, or fecal donor). By applying this protocol to these kinds of studies, we were able to monitor the consumption of the fibers by the gut microbes.^{13, 14, 45, 46} Analysis of fecal samples of germ-free mice showed that the linkages reflected the compositions of the fibers fed to the mice (**Figure 2.4a**). However, comparison of microbe-colonized and germ-free mice showed specific linkages were drastically decreased in the presence of gut microbes (**Figure 2.4b**). Specifically, total arabinose (**Figure 2.4c**) and the *t-f*-arabinose linkage (**Figure 2.4d**) were decreased with microbe-colonized mice in each of the fiber supplement types. The glycomic data were further integrated with metagenomic and metatranscriptomic analysis to provide specificity for the glycosyl hydrolases in the microbes thereby elucidating specific fiber-microbe-host interactions.

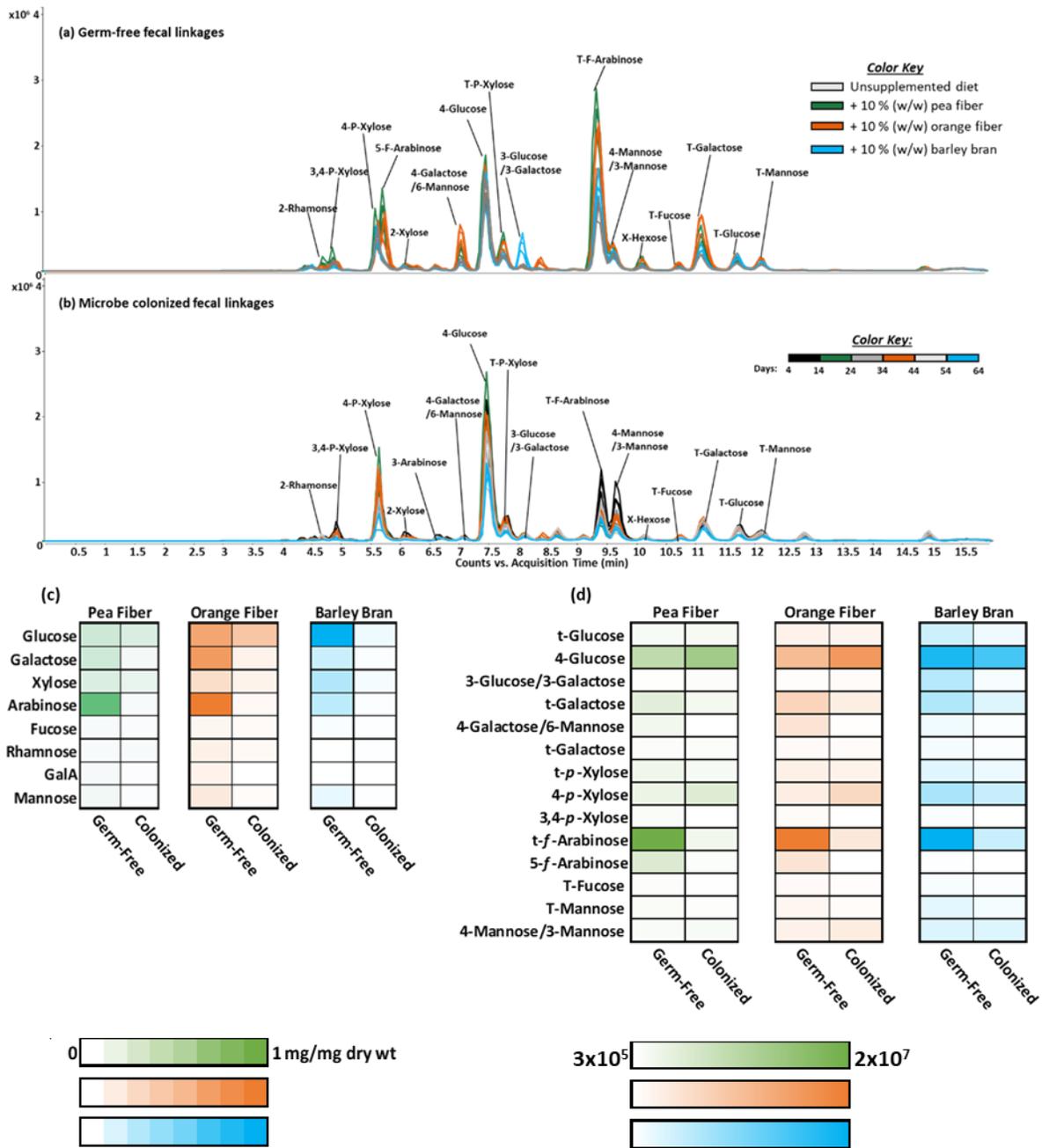


Figure 2.1. Monosaccharide and linkage profiles of fecal glycans obtained from germ-free and human microbiota colonized mice. (a) Chromatogram of fecal glycosidic linkages from germ-free mice fed diets supplemented with 10 % (w/w) pea fiber, orange fiber, or barley bran. (b) Chromatogram of fecal linkages from inoculated mice fed the same diets over the course of 64 days. Replicate traces indicate technical replicates. (c) Heat maps illustrating the differences

in the concentration of the most abundant fecal monosaccharide residues between germ-free and colonized mice expressed in mg/mg dry weight. (d) Differences in the observed peak area abundances of selected linkages between germ-free and colonized mice.

Differences in the glycomics profiles of apple varieties

Most of the produce that we commonly consume comes in different varieties or cultivars. The multi-glycomics workflow was used to characterize the carbohydrates present in different varieties of apple (**Figure 2.5**). Five different varieties were included, and for each variety, eight retail samples were analyzed. From the monosaccharide analysis, Granny Smith had a significantly higher amount of arabinose and galactose compared to the other four varieties (**Figure 2.5a, d**). The results were corroborated with the polysaccharide analysis where galactan was found higher in abundance in Granny Smith apples (**Figure 2.5c**). Arabinan was also found to be high in Granny Smith but the differences did not reach statistical significance (**Figure 2.5f**). Additionally, linkage analysis confirmed both findings from the monosaccharide and polysaccharide results. Namely, 4-galactose which is present primarily in galactans was significantly higher in Granny Smith (**Figure 2.5b**).

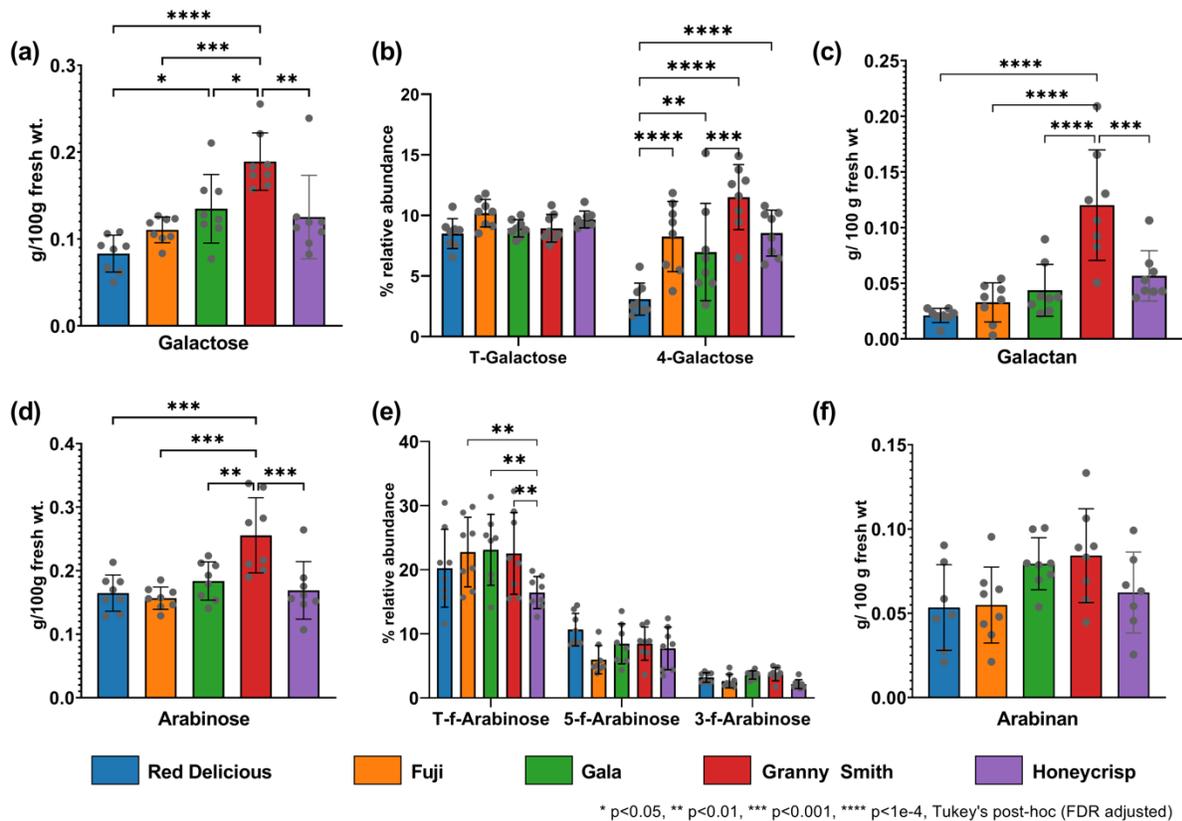


Figure 2.5. Monosaccharide, linkage and polysaccharide (FITDOG) composition analyses of different varieties of apples. Each bar represents mean while error bars represent standard deviation (n = 8 retail samples each variety). (a) Absolute galactose content measured using monosaccharide analysis. (b) Relative abundance of galactose linkages. (c) Absolute galactan content measured using FITDOG analysis. (d) Absolute arabinose content measured using monosaccharide analysis. (e) Relative abundance of arabinose linkages. (f) Absolute arabinan content measured using FITDOG analysis.

Carbohydrate-centric food database (Glycopedia)

One of the major attributes of this workflow is its increased throughput. This protocol is amenable to a 96-well plate format, enabling the parallel analysis of large batches of samples. Recently, we have published a carbohydrate-focused food composition database (Davis Food Glycopedia).³⁸ Over 800 food samples were analyzed for their monosaccharide compositions. Foods were categorized and clustered based on their monosaccharide profiles. For example, grain products had significantly higher amounts of glucose. Conversely, fruits and vegetables had greater monosaccharide diversity (**Figure 2.6**). A more powerful utility of this database is in formulating diets and menus that can be tailored towards specific monosaccharide compositions. For example, consuming more whole grain products (vs. highly refined and processed grain products), will result in higher consumption of arabinose and xylose. We are continuously expanding this database in terms of both the number of food entries, as well as the depth of analysis.

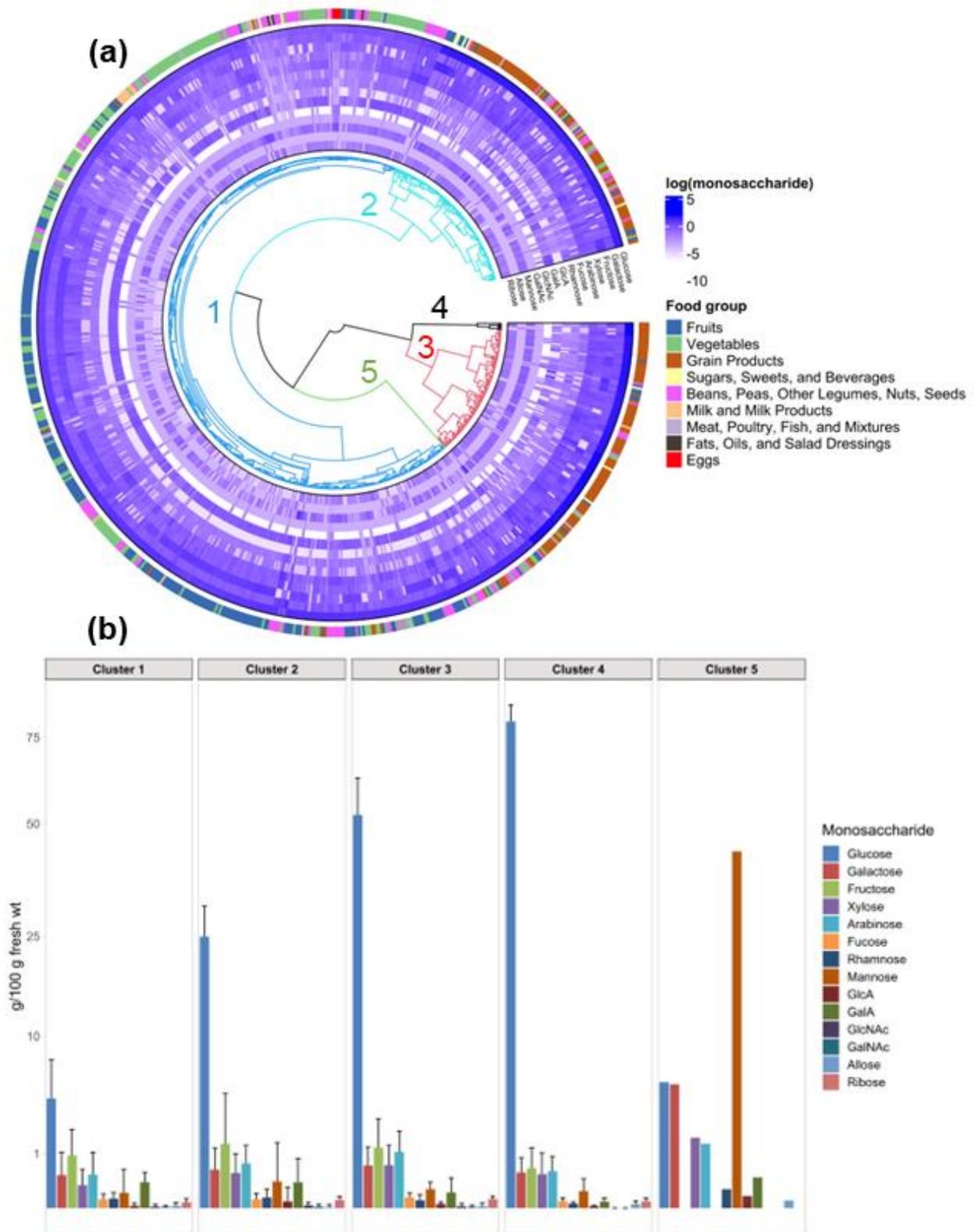


Figure 2.6. Clustering analysis of the Davis Food Glycopedia. (a) Circular heatmap and

dendrogram of food samples based on their total monosaccharide composition profiles. Heatmap values are log-transformed. Outermost heatmap track corresponds to assigned food group for each sample. (b) Cluster-averaged absolute monosaccharide compositions. Cluster numbers are indicated in (a) as shown.

A subset of the foods used in the Glycopedia was additionally analyzed using linkage and polysaccharide (FITDOG) analyses (**Figure 2.7**). As expected, grain products which are higher in glucose contained mainly starch corresponding to >80% of the polysaccharides based on FITDOG analysis (**Figure 2.7a**). The data was further corroborated with the abundance of 4-glucose in the linkage analysis (**Figure 2.7b**). Fruits and vegetables were found to contain diverse carbohydrate profiles consisting of the monosaccharide glucose, galactose, xylose, arabinose, and mannose (**Figure 2.7c**). The linkage and FITDOG analysis identified the polysaccharide structures as cellulose, xyloglucan, galactan, arabinoxylan, and mannans, respectively. Furthermore, fine variations in the linkage profile evince the presence of fine structures in specific polysaccharides. For example, the arabinan found in beans and peas is a linear structure comprised nearly exclusively of 5-linked and terminal (*t*-) arabinose, while the arabinan found in fruits and vegetables is a branched structure containing 2- and 3-linked arabinose (**Figure 2.7b**). The level of information obtained from this protocol is therefore unprecedented in terms of structural depth while providing enhanced throughput.

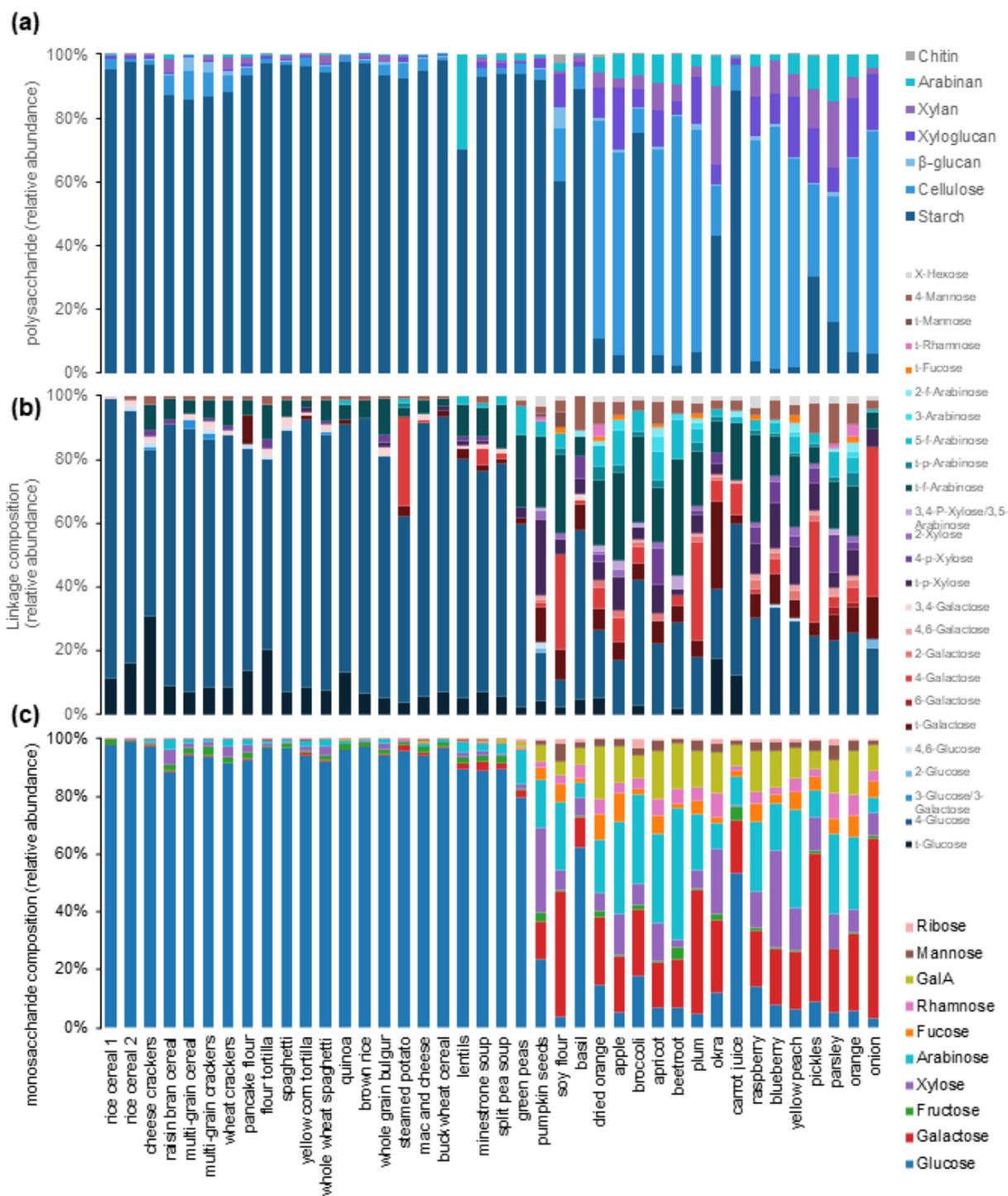
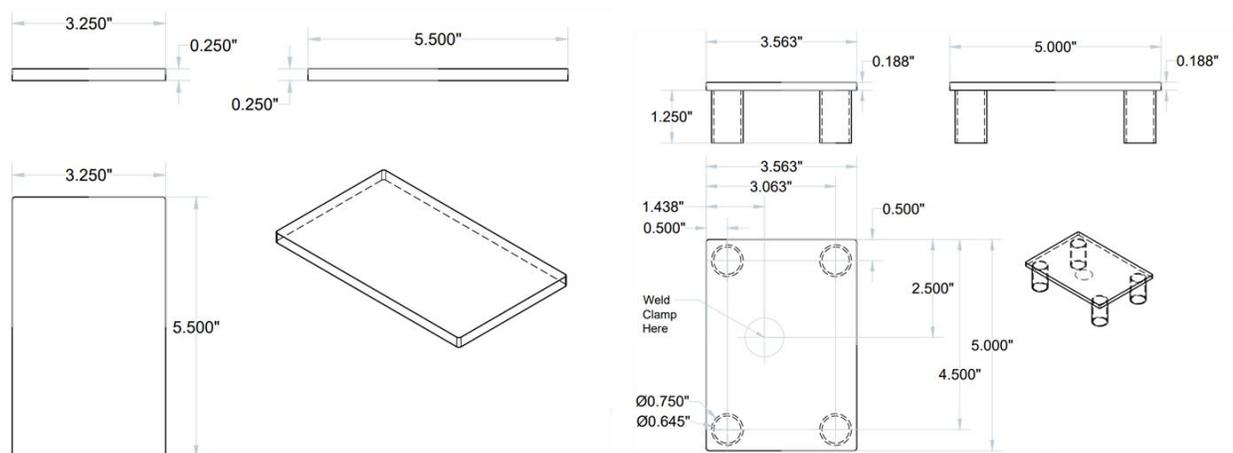


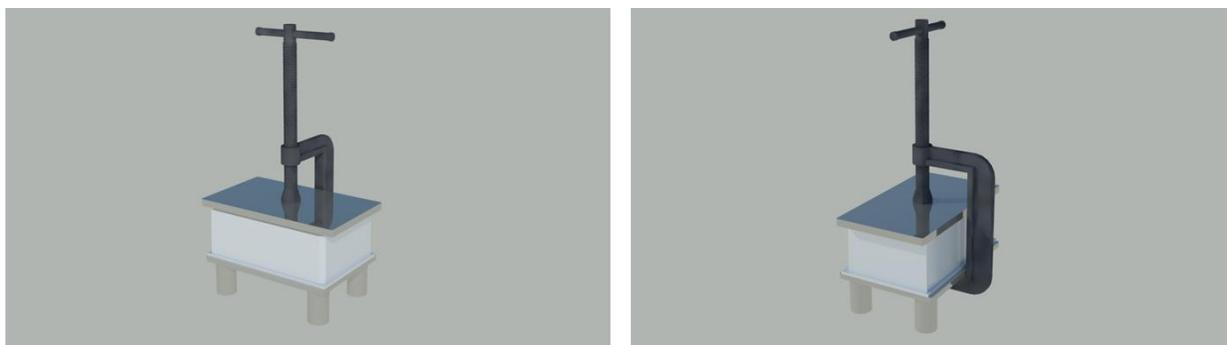
Figure 2.7. Multi-glycomics analysis of food. (a) Polysaccharide (FITDOG) composition, (b) glycosidic linkage composition, (c) monosaccharide composition.

SUPPLEMENTARY MATERIALS

(a)



(b)



Supplementary Figure 2.1. Schematic diagram of the stainless-steel plate clamps used for monosaccharide and linkage analysis. (a) Schematic of top and bottom plates with dimensions compatible with the 96-well plates described in the protocol. (b) Illustrations of finished clamps with C-clamps included. C-clamp is welded to the bottom of the bottom plate.

Supplementary Table 2.1. Calibrators used for monosaccharide analysis.

Level	Conc. (ng/mL)
L1	1
L2	10
L3	100

Level	Conc. (ng/mL)
L4	1000
L5	5000
L6	10000
L7	25000
L8	50000
L9	75000
L10	100000

Supplementary Table 2.2. Recommended pooling scheme of polysaccharide standards for FITOG analysis.

Pool 1	Pool 2	Pool 3	Pool 4
Arabinan	Xylan	Starch	Arabinoxylan
Galactan	Mannan	Cellulose	Galactomannan
Xyloglucan	β -glucan		
	Chitin		

Supplementary Table 2.3. Dynamic MRM transition list for monosaccharide analysis.

Name	TS	Transition	Scan	Type	RT	Left RT Delta	Right RT Delta	RT Delta Unit	Ion Polarity
Fructose	1	511.2 -> 175.1	MRM	Target	0.269	0.1	0.1	Minutes	Positive
Mannose	1	511.2 -> 175.1	MRM	Target	0.753	0.1	0.1	Minutes	Positive
Allose	1	511.2 -> 175.1	MRM	Target	0.899	0.1	0.1	Minutes	Positive
Ribose	1	481.2 -> 175.1	MRM	Target	0.901	0.2	0.2	Minutes	Positive
GlcA	1	525.2 -> 175.1	MRM	Target	0.915	0.1	0.1	Minutes	Positive
Rhamnose	1	495.2 -> 175.1	MRM	Target	0.986	0.2	0.2	Minutes	Positive
GalA	1	525.2 -> 175.1	MRM	Target	1.051	0.1	0.1	Minutes	Positive
Glucose	1	511.2 -> 175.1	MRM	Target	1.491	0.2	0.2	Minutes	Positive
GlcNAc	1	552.2 -> 175.1	MRM	Target	1.503	0.2	0.2	Minutes	Positive
GalNAc	1	552.2 -> 175.1	MRM	Target	1.657	0.2	0.2	Minutes	Positive
Galactose	1	511.2 -> 175.1	MRM	Target	1.695	0.2	0.2	Minutes	Positive
Xylose	1	481.2 -> 175.1	MRM	Target	1.723	0.2	0.2	Minutes	Positive
Arabinose	1	481.2 -> 175.1	MRM	Target	1.816	0.2	0.2	Minutes	Positive
Fucose	1	495.2 -> 175.1	MRM	Target	2.035	0.2	0.2	Minutes	Positive

Supplementary Table 2.4. MRM transition list for glycosidic linkage analysis.

Name	TS	Transition	Scan	Type	Precursor Ion	Product Ion	RT	Ion Polarity
2,4,6-Glucose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.135	Positive
2,3,6-Glucose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.263	Positive
2-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	4.436	Positive
2,4,6-Galactose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.455	Positive
3,4,6-Glucose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	4.647	Positive
2-Rhamnose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	4.716	Positive
2,4-Xylose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	4.918	Positive
3,4,6-Galactose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	5.031	Positive
2,3-Arabinose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	5.238	Positive
4,6-Glucose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	5.279	Positive
3,4- <i>p</i> -Xylose/3,5-Arabinose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	5.431	Positive
3,6-Galactose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	5.471	Positive
3,4,6-Mannose	2	525.2 -> 175.1	MRM	Target	525.2	175.1	5.543	Positive
2,5-Arabinose	2	495.2 -> 175.1	MRM	Target	495.2	175.1	5.751	Positive
2,4-Glucose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.111	Positive
4- <i>p</i> -Xylose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	6.258	Positive
5- <i>f</i> -Arabinose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	6.386	Positive
3,4-Galactose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.495	Positive
3,6-Mannose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.559	Positive
3,4-Glucose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	6.815	Positive
6-Glucose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	6.933	Positive
2-Xylose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	7.026	Positive
4,6-Galactose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	7.072	Positive
6-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	7.125	Positive
3-Arabinose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	7.538	Positive
4,6-Mannose	2	539.2 -> 175.1	MRM	Target	539.2	175.1	7.648	Positive
4-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	7.766	Positive
6-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	8.086	Positive
4-Glucose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	8.406	Positive
2- <i>f</i> -Arabinose	2	509.2 -> 175.1	MRM	Target	509.2	175.1	8.755	Positive
2-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	8.79	Positive
<i>t-p</i> -Xylose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	8.942	Positive
<i>t</i> -Glucuronic acid	2	581.2 -> 175.1	MRM	Target	581.2	175.1	8.965	Positive
3-Glucose/3-Galactose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	9.174	Positive
<i>t-p</i> -Arabinose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	9.744	Positive
2-Glucose	2	553.6 -> 175.1	MRM	Target	553.6	175.1	10.199	Positive

Name	TS	Transition	Scan	Type	Precursor Ion	Product Ion	RT	Ion Polarity
t-f-Arabinose	2	523.2 -> 175.1	MRM	Target	523.2	175.1	10.67	Positive
t-Galacturonic acid	2	581.2 -> 175.1	MRM	Target	581.2	175.1	10.694	Positive
3-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	10.839	Positive
4-Mannose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	11.095	Positive
x-Hexose	2	553.6 -> 175.1	MRM	Target	553.2	175.1	11.799	Positive
t-Fucose	2	537.2 -> 175.1	MRM	Target	537.2	175.1	12.263	Positive
t-Galactose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	12.559	Positive
t-Rhamnose	2	537.2 -> 175.1	MRM	Target	537.2	175.1	12.903	Positive
t-Glucose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	13.391	Positive
t-Mannose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	13.904	Positive
t-Hexose	2	567.2 -> 175.1	MRM	Target	567.2	175.1	14.352	Positive

Supplementary Table 2.5. List of oligosaccharides used for FITDOG analysis.

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Starch	Hex3	C18 H34 O16	3.40	No
Starch	Hex4	C24 H44 O21	11.75	Yes
Starch	Hex5	C30 H54 O26	13.70	Yes
Starch	Hex6	C36 H64 O31	14.59	Yes
Starch	Hex7	C42 H74 O36	15.09	No
Starch	Hex8	C48 H84 O41	15.59	No
Starch	Hex9	C54 H94 O46	16.26	No
Starch	Hex10	C60 H104 O51	17.26	No
Starch	Hex11	C66 H114 O56	18.49	No
Starch	Hex12	C72 H124 O61	19.49	No
Starch	Hex13	C78 H134 O66	20.10	No
Starch	Hex14	C84 H144 O71	20.55	No
Starch	Hex15	C90 H154 O76	20.88	No
Starch	Hex16	C96 H164 O81	21.27	No
Starch	Hex17	C102 H174 O86	22.11	No
Starch	Hex18	C108 H184 O91	23.28	No
Starch	Hex19	C114 H194 O96	24.40	No
Starch	Hex20	C120 H204 O101	24.77	No
Starch	Hex21	C126 H214 O106	24.89	No
Starch	Hex22	C132 H224 O111	25.03	No
Cellulose	Hex3	C18 H34 O16	14.75	Yes
Cellulose	Hex4	C24 H44 O21	20.21	Yes
Cellulose	Hex5	C30 H54 O26	25.73	Yes
Mannan	Hex3	C18 H34 O16	1.67	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Mannan	Hex4	C24 H44 O21	3.40	Yes
Mannan	Hex5	C30 H54 O26	9.97	Yes
Mannan	Hex6	C36 H64 O31	12.93	Yes
Mannan	Hex7	C42 H74 O36	13.93	No
Mannan	Hex8	C48 H84 O41	14.77	No
Mannan	Hex9	C54 H94 O46	15.43	No
Mannan	Hex10	C60 H104 O51	15.94	No
Mannan	Hex11	C66 H114 O56	16.49	No
Mannan	Hex12	C72 H124 O61	16.83	No
Mannan	Hex13	C78 H134 O66	17.27	No
Mannan	Hex14	C84 H144 O71	17.72	No
b-Glucan	Hex3	C18 H34 O16	12.59	No
b-Glucan	Hex3	C18 H34 O16	14.75	No
b-Glucan	Hex3	C18 H34 O16	16.27	No
b-Glucan	Hex4	C24 H44 O21	18.33	No
b-Glucan	Hex4	C24 H44 O21	20.39	No
b-Glucan	Hex4	C24 H44 O21	21.23	Yes
b-Glucan	Hex5	C30 H54 O26	23.46	No
b-Glucan	Hex5	C30 H54 O26	24.57	No
b-Glucan	Hex5	C30 H54 O26	25.74	No
b-Glucan	Hex5	C30 H54 O26	26.52	Yes
b-Glucan	Hex5	C30 H54 O26	27.30	Yes
Xylan	Pnt3	C15 H28 O13	10.87	Yes
Xylan	Pnt4	C20 H36 O17	16.21	Yes
Xylan	Pnt5	C25 H44 O21	19.45	Yes
Xylan	Pnt6	C30 H52 O25	22.40	No
Xylan	Pnt7	C35 H60 O29	25.96	No
Xylan	Pnt8	C40 H68 O33	28.08	No
Chitin	HexNAc3	C16 H30 O11 N2	10.60	Yes
Chitin	HexNAc4	C24 H43 O16 N3	14.80	No
Chitin	HexNAc5	C32 H56 O21 N4	16.30	Yes
Chitin	HexNAc6	C40 H69 O26 N5	18.20	Yes
Chitin	HexNAc7	C48 H82 O31 N6	19.10	No
Chitin	HexNAc8	C56 H95 O36 N7	20.40	No
Chitin	HexNAc9	C64 H108 O41 N8	21.12	No
Arabinan, linear	Pnt3	C15 H28 O13	5.18	Yes
Arabinan, linear	Pnt4	C20 H36 O17	13.42	Yes
Arabinan, linear	Pnt5	C25 H44 O21	15.92	Yes
Arabinan, linear	Pnt6	C30 H52 O25	17.76	No
Arabinan, linear	Pnt7	C35 H60 O29	19.26	No
Arabinan, linear	Pnt8	C40 H68 O33	20.66	No
Arabinan, linear	Pnt9	C50 H84 O41	21.88	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Galactan	Hex3	C12 H24 O11	2.28	Yes
Galactan	Hex4	C18 H34 O16	7.35	Yes
Galactan	Hex5	C24 H44 O21	12.20	Yes
Galactan	Hex6	C30 H54 O26	13.31	No
Galactan	Hex7	C36 H64 O31	13.92	No
Galactan	Hex8	C42 H74 O36	14.47	No
Galactan	Hex9	C48 H84 O41	15.37	No
Galactan	Hex10	C54 H94 O46	17.00	No
Galactan	Hex11	C60 H104 O51	18.40	No
Galactan	Hex12	C66 H114 O56	19.40	No
Galactan	Hex13	C72 H124 O61	20.00	No
Galactan	Hex14	C78 H134 O66	20.50	No
Xyloglucan	Hex2:Pnt1	C17 H32 O15	12.10	Yes
Xyloglucan	Hex2:Pnt2	C22 H40 O19	14.60	Yes
Xyloglucan	Hex3:Pnt1	C23 H42 O20	14.00	Yes
Xyloglucan	Hex3:Pnt1	C23 H42 O20	16.50	No
Xyloglucan	Hex3:Pnt1	C23 H42 O20	18.10	No
Xyloglucan	Hex3:Pnt2	C28 H50 O24	18.80	No
Arabinogalactan	Hex3	C18 H34 O16	1.73	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	2.18	No
Arabinogalactan	Hex3	C18 H34 O16	2.3	No
Arabinogalactan	Hex4	C24 H44 O21	2.85	No
Arabinogalactan	Hex1:Pnt3	C21 H38 O18	3.21	No
Arabinogalactan	Hex4	C24 H44 O21	3.37	No
Arabinogalactan	Hex2:Pnt2	C22 H40 O19	3.82	No
Arabinogalactan	Pnt3	C15 H28 O13	4	No
Arabinogalactan	Hex4	C24 H44 O21	4.68	No
Arabinogalactan	Hex3	C18 H34 O16	4.83	No
Arabinogalactan	Hex5	C30 H54 O26	6.59	No
Arabinogalactan	Hex4	C24 H44 O21	6.95	No
Arabinogalactan	Hex5	C30 H54 O26	7.28	No
Arabinogalactan	Hex3:Pnt1	C23 H42 O20	7.28	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	9.22	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	11.21	No
Arabinogalactan	Hex6	C36 H64 O31	11.36	No
Arabinogalactan	Hex5	C30 H54 O26	11.5	No
Arabinogalactan	Hex2:Pnt1	C17 H32 O15	11.66	No
Arabinogalactan	Hex4	C24 H44 O21	12.14	No
Arabinogalactan	Hex3:Pnt1	C23 H42 O20	12.26	No
Arabinogalactan	Hex5	C30 H54 O26	12.8	No
Arabinogalactan	Hex4	C24 H44 O21	12.97	No
Arabinogalactan	Hex3:Pnt1	C23 H42 O20	13.31	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Arabinogalactan	Hex2:Pnt2	C22 H40 O19	13.31	No
Arabinogalactan	Hex5	C30 H54 O26	13.48	No
Arabinogalactan	Hex6	C36 H64 O31	13.63	No
Arabinogalactan	Hex7	C42 H74 O36	14.3	No
Arabinogalactan	Hex6	C36 H64 O31	15	No
Arabinogalactan	Hex5:Pnt1	C35 H62 O30	15	No
Arabinogalactan	Hex6	C36 H64 O31	15.44	No
Arabinogalactan	Hex7	C42 H74 O36	16.25	No
Arabinogalactan	Hex6	C36 H64 O31	16.37	No
Arabinogalactan	Pnt4	C20 H36 O17	16.51	No
Arabinogalactan	Hex7	C42 H74 O36	16.69	No
Arabinogalactan	Hex8	C48 H84 O41	17	No
Arabinogalactan	Hex4:Pnt1	C29 H52 O25	17.19	No
Arabinogalactan	Hex7	C42 H74 O36	17.54	No
Arabinogalactan	Hex8	C48 H84 O41	18.1	No
Arabinogalactan	Hex8	C48 H84 O41	18.44	No
Arabinogalactan	Hex5	C30 H54 O26	18.81	No
Arabinoxylan	Pnt3	C15 H28 O13	2.04	No
Arabinoxylan	Pnt3	C15 H28 O13	2.67	No
Arabinoxylan	Pnt3	C15 H28 O13	4.23	No
Arabinoxylan	Pnt4	C20 H36 O17	4.74	No
Arabinoxylan	Pnt4	C20 H36 O17	7.48	No
Arabinoxylan	Pnt4	C20 H36 O17	8.23	No
Arabinoxylan	Pnt4	C20 H36 O17	10.76	No
Arabinoxylan	Pnt4	C20 H36 O17	12.07	No
Arabinoxylan	Pnt4	C20 H36 O17	12.38	No
Arabinoxylan	Pnt4	C20 H36 O17	13.15	No
Arabinoxylan	Pnt5	C25 H44 O21	13.41	No
Arabinoxylan	Pnt5	C25 H44 O21	13.69	No
Arabinoxylan	Pnt5	C25 H44 O21	14.25	No
Arabinoxylan	Pnt5	C25 H44 O21	14.55	No
Arabinoxylan	Pnt5	C25 H44 O21	14.91	No
Arabinoxylan	Pnt5	C25 H44 O21	15.59	No
Arabinoxylan	Pnt5	C25 H44 O21	16.11	No
Arabinoxylan	Pnt3	C15 H28 O13	16.47	No
Arabinoxylan	Pnt6	C30 H52 O25	16.6	No
Arabinoxylan	Pnt5	C25 H44 O21	17.05	No
Arabinoxylan	Pnt6	C30 H52 O25	17.3	No
Arabinoxylan	Pnt5	C25 H44 O21	18.03	No
Arabinoxylan	Pnt4	C20 H36 O17	18.45	No
Arabinoxylan	Pnt5	C25 H44 O21	18.53	No
Arabinoxylan	Pnt6	C30 H52 O25	18.77	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Arabinoxylan	Pnt4	C20 H36 O17	19.13	No
Arabinoxylan	Pnt6	C30 H52 O25	19.65	No
Arabinoxylan	Pnt4	C20 H36 O17	19.75	No
Arabinoxylan	Pnt6	C30 H52 O25	20.24	No
Arabinoxylan	Pnt6	C30 H52 O25	21.16	No
Arabinoxylan	Pnt6	C30 H52 O25	21.39	No
Arabinoxylan	Pnt7	C35 H60 O29	21.67	No
Arabinoxylan	Pnt6	C30 H52 O25	21.94	No
Arabinoxylan	Pnt5	C25 H44 O21	22.06	No
Arabinoxylan	Pnt7	C35 H60 O29	22.59	No
Arabinoxylan	Pnt6	C30 H52 O25	22.71	No
Arabinoxylan	Pnt6	C30 H52 O25	23.14	No
Arabinoxylan	Pnt5	C25 H44 O21	24.41	No
Arabinoxylan	Pnt7	C35 H60 O29	24.55	No
Arabinoxylan	Pnt7	C35 H60 O29	25.37	No
Arabinoxylan	Pnt6	C30 H52 O25	26.06	No
Arabinoxylan	Pnt8	C40 H68 O33	27.62	No
Arabinoxylan	Pnt7	C35 H60 O29	28.46	No
Curdlan	Hex3	C18 H34 O16	13.7	No
Curdlan	Hex4	C24 H44 O21	20.72	No
Curdlan	Hex5	C30 H54 O26	27.96	No
Galactomannan	Hex3	C18 H34 O16	2.58	No
Galactomannan	Hex4	C24 H44 O21	3.88	No
Galactomannan	Hex4	C24 H44 O21	4.25	No
Galactomannan	Hex4	C24 H44 O21	4.67	No
Galactomannan	Hex4	C24 H44 O21	8.18	No
Galactomannan	Hex5	C30 H54 O26	9.99	No
Galactomannan	Hex5	C30 H54 O26	11.02	No
Galactomannan	Hex5	C30 H54 O26	11.34	No
Galactomannan	Hex4	C24 H44 O21	12.14	No
Galactomannan	Hex5	C30 H54 O26	12.26	No
Galactomannan	Hex6	C36 H64 O31	13.03	No
Galactomannan	Hex5	C30 H54 O26	13.15	No
Galactomannan	Hex6	C36 H64 O31	13.57	No
Galactomannan	Hex6	C36 H64 O31	13.79	No
Galactomannan	Hex6	C36 H64 O31	14.1	No
Galactomannan	Hex7	C42 H74 O36	14.63	No
Galactomannan	Hex6	C36 H64 O31	14.85	No
Galactomannan	Hex4	C24 H44 O21	15.21	No
Galactomannan	Hex7	C42 H74 O36	15.35	No
Galactomannan	Hex8	C48 H84 O41	15.9	No
Galactomannan	Hex6	C36 H64 O31	16.48	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Galactomannan	Hex7	C42 H74 O36	16.63	No
Galactomannan	Hex8	C48 H84 O41	16.93	No
Glucomannan	Hex3	C18 H34 O16	1.92	No
Glucomannan	Hex4	C24 H44 O21	5.62	No
Glucomannan	Hex3	C18 H34 O16	6.55	No
Glucomannan	Hex3	C18 H34 O16	8.73	No
Glucomannan	Hex4	C24 H44 O21	10.62	No
Glucomannan	Hex5	C30 H54 O26	11.45	No
Glucomannan	Hex4	C24 H44 O21	11.6	No
Glucomannan	Hex3	C18 H34 O16	11.84	No
Glucomannan	Hex4	C24 H44 O21	12.27	No
Glucomannan	Hex4	C24 H44 O21	12.92	No
Glucomannan	Hex6	C36 H64 O31	12.93	No
Glucomannan	Hex5	C30 H54 O26	13.2	No
Glucomannan	Hex6	C36 H64 O31	13.56	No
Glucomannan	Hex3	C18 H34 O16	14.02	No
Glucomannan	Hex6	C36 H64 O31	14.31	No
Glucomannan	Hex7	C42 H74 O36	15.02	No
Glucomannan	Hex5	C30 H54 O26	15.26	No
Glucomannan	Hex6	C36 H64 O31	15.52	No
Glucomannan	Hex8	C48 H84 O41	15.82	No
Glucomannan	Hex5	C30 H54 O26	15.94	No
Glucomannan	Hex4	C24 H44 O21	16.23	No
Glucomannan	Hex6	C36 H64 O31	16.46	No
Glucomannan	Hex5	C30 H54 O26	16.62	No
Glucomannan	Hex4	C24 H44 O21	16.77	No
Glucomannan	Hex7	C42 H74 O36	17.02	No
Glucomannan	Hex5	C30 H54 O26	17.14	No
Glucomannan	Hex6	C36 H64 O31	17.3	No
Glucomannan	Hex5	C30 H54 O26	17.62	No
Glucomannan	Hex6	C36 H64 O31	17.74	No
Glucomannan	Hex5	C30 H54 O26	17.86	No
Glucomannan	Hex4	C24 H44 O21	18.01	No
Glucomannan	Hex5	C30 H54 O26	18.26	No
Glucomannan	Hex7	C42 H74 O36	18.37	No
Glucomannan	Hex6	C36 H64 O31	18.51	No
Glucomannan	Hex7	C42 H74 O36	18.79	No
Glucomannan	Hex5	C30 H54 O26	19.06	No
Glucomannan	Hex6	C36 H64 O31	19.14	No
Glucomannan	Hex7	C42 H74 O36	19.24	No
Glucomannan	Hex6	C36 H64 O31	19.4	No
Glucomannan	Hex7	C42 H74 O36	20.03	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Glucomannan	Hex7	C42 H74 O36	20.31	No
Glucomannan	Hex8	C48 H84 O41	20.83	No
Glucomannan	Hex4	C24 H44 O21	21.11	No
Glucomannan	Hex7	C42 H74 O36	21.54	No
Glucomannan	Hex5	C30 H54 O26	21.85	No
Glucomannan	Hex7	C42 H74 O36	21.99	No
Glucomannan	Hex5	C30 H54 O26	22.27	No
Glucomannan	Hex6	C36 H64 O31	22.76	No
Glucomannan	Hex7	C42 H74 O36	23.22	No
Glucomannan	Hex7	C42 H74 O36	23.88	No
Glucomannan	Hex6	C36 H64 O31	24.72	No
Glucomannan	Hex5	C30 H54 O26	25.42	No
Glucomannan	Hex6	C36 H64 O31	25.77	No
Glucomannan	Hex8	C48 H84 O41	27	No
Glucomannan	Hex9	C54 H94 O46	27.68	No
Lichenan	Hex3	C18 H34 O16	2.04	No
Lichenan	Hex4	C24 H44 O21	2.88	No
Lichenan	Hex3	C18 H34 O16	3.06	No
Lichenan	Hex3	C18 H34 O16	3.34	No
Lichenan	Hex3	C18 H34 O16	5.85	No
Lichenan	Hex5	C30 H54 O26	6.14	No
Lichenan	Hex5	C30 H54 O26	7.59	No
Lichenan	Hex4	C24 H44 O21	9.07	No
Lichenan	Hex6	C36 H64 O31	10.1	No
Lichenan	Hex6	C36 H64 O31	10.78	No
Lichenan	Hex5	C30 H54 O26	11.08	No
Lichenan	Hex7	C42 H74 O36	12.13	No
Lichenan	Hex3	C18 H34 O16	12.59	No
Lichenan	Hex6	C36 H64 O31	12.89	No
Lichenan	Hex5	C30 H54 O26	13.42	No
Lichenan	Hex5	C30 H54 O26	14.29	No
Lichenan	Hex6	C36 H64 O31	14.75	No
Lichenan	Hex3	C18 H34 O16	15.02	No
Lichenan	Hex6	C36 H64 O31	15.34	No
Lichenan	Hex5	C30 H54 O26	16.53	No
Lichenan	Hex7	C42 H74 O36	16.62	No
Lichenan	Hex6	C36 H64 O31	17.23	No
Lichenan	Hex6	C36 H64 O31	17.62	No
Lichenan	Hex7	C42 H74 O36	18.1	No
Lichenan	Hex7	C42 H74 O36	18.52	No
Lichenan	Hex8	C48 H84 O41	18.96	No
Lichenan	Hex7	C42 H74 O36	20.39	No

Polysaccharide	Oligosaccharide	Formula (reduced)	RT(min)	Used as quantifier?
Lichenan	Hex8	C48 H84 O41	20.99	No
Lichenan	Hex9	C54 H94 O46	21.45	No
Lichenan	Hex4	C24 H44 O21	22.22	No
Lichenan	Hex5	C30 H54 O26	24.88	No
Lichenan	Hex5	C30 H54 O26	25.82	No
Lichenan	Hex6	C36 H64 O31	27.31	No
Lichenan	Hex6	C36 H64 O31	31.06	No

REFERENCES

1. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZY): an expert resource for Glycogenomics. *Nucleic Acids Res* **2009**, *37* (Database issue), D233-8.
2. Flint, H. J.; Scott, K. P.; Duncan, S. H.; Louis, P.; Forano, E., Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **2012**, *3* (4), 289-306.
3. Wardman, J. F.; Bains, R. K.; Rahfeld, P.; Withers, S. G., Carbohydrate-active enzymes (CAZymes) in the gut microbiome. *Nat Rev Microbiol* **2022**, *20* (9), 542-556.
4. Cronin, P.; Joyce, S. A.; O'Toole, P. W.; O'Connor, E. M., Dietary Fibre Modulates the Gut Microbiota. *Nutrients* **2021**, *13* (5).
5. Dhingra, D.; Michael, M.; Rajput, H.; Patil, R. T., Dietary fibre in foods: a review. *J Food Sci Technol* **2012**, *49* (3), 255-66.
6. Han, S.; Van Treuren, W.; Fischer, C. R.; Merrill, B. D.; DeFelice, B. C.; Sanchez, J. M.; Higginbottom, S. K.; Guthrie, L.; Fall, L. A.; Dodd, D.; Fischbach, M. A.; Sonnenburg, J. L., A metabolomics pipeline for the mechanistic interrogation of the gut microbiome. *Nature* **2021**, *595* (7867), 415-420.
7. David, L. A.; Maurice, C. F.; Carmody, R. N.; Gootenberg, D. B.; Button, J. E.; Wolfe, B. E.; Ling, A. V.; Devlin, A. S.; Varma, Y.; Fischbach, M. A.; Biddinger, S. B.; Dutton, R. J.; Turnbaugh, P. J., Diet rapidly and reproducibly alters the human gut microbiome. *Nature* **2014**, *505* (7484), 559-63.
8. Koh, A.; De Vadder, F.; Kovatcheva-Datchary, P.; Backhed, F., From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. *Cell* **2016**, *165* (6), 1332-1345.
9. Morrison, D. J.; Preston, T., Formation of short chain fatty acids by the gut microbiota and their impact on human metabolism. *Gut Microbes* **2016**, *7* (3), 189-200.
10. Fischbach, M. A.; Sonnenburg, J. L., Eating For Two: How Metabolism Establishes Interspecies Interactions in the Gut. *Cell Host Microbe* **2011**, *10* (4), 336-347.
11. Spencer, C. N.; McQuade, J. L.; Gopalakrishnan, V.; McCulloch, J. A.; Vetizou, M.; Cogdill, A. P.; Khan, M. A. W.; Zhang, X.; White, M. G.; Peterson, C. B.; Wong, M. C.; Morad, G.; Rodgers, T.; Badger, J. H.; Helmink, B. A.; Andrews, M. C.; Rodrigues, R. R.; Morgun, A.; Kim, Y. S.; Roszik, J.; Hoffman, K. L.; Zheng, J.; Zhou, Y.; Medik, Y. B.; Kahn, L. M.; Johnson, S.; Hudgens, C. W.; Wani, K.; Gaudreau, P. O.; Harris, A. L.; Jamal, M. A.; Baruch, E. N.; Perez-Guijarro, E.; Day, C. P.; Merlino, G.; Pazdrak, B.; Lochmann, B. S.; Szczepaniak-Sloane, R. A.; Arora, R.; Anderson, J.; Zobniw, C. M.; Posada, E.; Sirmans, E.; Simon, J.; Haydu, L. E.; Burton, E. M.; Wang, L.; Dang, M.; Clise-Dwyer, K.; Schneider, S.; Chapman, T.; Anang, N. A. S.; Duncan, S.; Toker, J.; Malke, J. C.; Glitza, I. C.; Amaria, R. N.; Tawbi, H. A.; Diab, A.; Wong, M. K.; Patel, S. P.; Woodman, S. E.; Davies, M. A.; Ross, M. I.; Gershenwald, J. E.; Lee, J. E.; Hwu, P.; Jensen, V.; Samuels, Y.; Straussman, R.; Ajami, N. J.; Nelson, K. C.; Nezi, L.; Petrosino, J. F.; Futreal, P. A.; Lazar, A. J.; Hu, J.; Jenq, R. R.; Tetzlaff, M. T.; Yan, Y.; Garrett, W. S.; Huttenhower, C.; Sharma, P.; Watowich, S. S.; Allison, J. P.; Cohen, L.; Trinchieri, G.; Daniel, C. R.; Wargo, J. A., Dietary fiber and probiotics influence the gut microbiome and melanoma immunotherapy response. *Science* **2021**, *374* (6575), 1632-1640.
12. Gehrig, J. L.; Venkatesh, S.; Chang, H. W.; Hibberd, M. C.; Kung, V. L.; Cheng, J.; Chen, R. Y.; Subramanian, S.; Cowardin, C. A.; Meier, M. F.; O'Donnell, D.; Talcott, M.;

- Spears, L. D.; Semenkovich, C. F.; Henrissat, B.; Giannone, R. J.; Hettich, R. L.; Ilkayeva, O.; Muehlbauer, M.; Newgard, C. B.; Sawyer, C.; Head, R. D.; Rodionov, D. A.; Arzamasov, A. A.; Leyn, S. A.; Osterman, A. L.; Hossain, M. I.; Islam, M.; Choudhury, N.; Sarker, S. A.; Huq, S.; Mahmud, I.; Mostafa, I.; Mahfuz, M.; Barratt, M. J.; Ahmed, T.; Gordon, J. I., Effects of microbiota-directed foods in gnotobiotic animals and undernourished children. *Science* **2019**, *365* (6449).
13. Delannoy-Bruno, O.; Desai, C.; Castillo, J. J.; Couture, G.; Barve, R. A.; Lombard, V.; Henrissat, B.; Cheng, J.; Han, N.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Lebrilla, C. B.; Marion, S.; Heath, A. C.; Barratt, M. J.; Gordon, J. I., An approach for evaluating the effects of dietary fiber polysaccharides on the human gut microbiome and plasma proteome. *Proc Natl Acad Sci U S A* **2022**, *119* (20), e2123411119.
14. Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B.; Barve, R. A.; Lombard, V.; Henrissat, B.; Leyn, S. A.; Rodionov, D. A.; Osterman, A. L.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Kirbach, K.; Wilmot, T.; Heath, A. C.; Klein, S.; Barratt, M. J.; Gordon, J. I., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021**, *595* (7865), 91-95.
15. O'Grady, J.; O'Connor, E. M.; Shanahan, F., Review article: dietary fibre in the era of microbiome science. *Aliment Pharmacol Ther* **2019**, *49* (5), 506-515.
16. Gill, S. K.; Rossi, M.; Bajka, B.; Whelan, K., Dietary fibre in gastrointestinal health and disease. *Nat Rev Gastroenterol Hepatol* **2021**, *18* (2), 101-116.
17. Barratt, M. J.; Lebrilla, C.; Shapiro, H. Y.; Gordon, J. I., The Gut Microbiota, Food Science, and Human Nutrition: A Timely Marriage. *Cell Host Microbe* **2017**, *22* (2), 134-141.
18. Amicucci, M. J.; Nandita, E.; Lebrilla, C. B., Function without Structures: The Need for In-Depth Analysis of Dietary Carbohydrates. *J Agr Food Chem* **2019**, *67* (16), 4418-4424.
19. Xu, G. G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis - quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2018**, *143* (1), 200-207.
20. Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T. T. T.; Liu, Y. Y.; Lee, M.; Xu, G. G.; Lebrilla, C. B., A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *Int J Mass Spectrom* **2019**, *438*, 22-28.
21. Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B., Development of an Extensive Linkage Library for Characterization of Carbohydrates. *Analytical Chemistry* **2019**, *91* (20), 13022-13031.
22. Galermo, A. G.; Nandita, E.; Barboza, M.; Amicucci, M. J.; Vo, T. T. T.; Lebrilla, C. B., Liquid Chromatography-Tandem Mass Spectrometry Approach for Determining Glycosidic Linkages. *Analytical Chemistry* **2018**, *90* (21), 13073-13080.
23. Amicucci, M. J.; Nandita, E.; Galermo, A. G.; Castillo, J. J.; Chen, S.; Park, D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B., A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis. *Nat Commun* **2020**, *11* (1), 3963.
24. Nandita, E.; Bacalzo, N. P.; Ranque, C. L.; Amicucci, M. J.; Galermo, A.; Lebrilla, C. B., Polysaccharide identification through oligosaccharide fingerprinting. *Carbohydr Polym* **2021**, *257*.

25. Wong, M.; Xu, G. G.; Park, D.; Barboza, M.; Lebrilla, C. B., Intact glycosphingolipidomic analysis of the cell membrane during differentiation yields extensive glycan and lipid changes. *Sci Rep-Uk* **2018**, *8*.
26. Park, D. D.; Xu, G.; Wong, M.; Phoomak, C.; Liu, M. Q.; Haigh, N. E.; Wongkham, S.; Yang, P. Y.; Maverakis, E.; Lebrilla, C. B., Membrane glycomics reveal heterogeneity and quantitative distribution of cell surface sialylation. *Chem Sci* **2018**, *9* (29), 6271-6285.
27. Chu, C. S.; Ninonuevo, M. R.; Clowers, B. H.; Perkins, P. D.; An, H. J.; Yin, H. F.; Killeen, K.; Miyamoto, S.; Grimm, R.; Lebrilla, C. B., Profile of native N-linked glycan structures from human serum using high performance liquid chromatography on a microfluidic chip and time-of-flight mass spectrometry. *Proteomics* **2009**, *9* (7), 1939-1951.
28. Barboza, M.; Pinzon, J.; Wickramasinghe, S.; Froehlich, J. W.; Moeller, I.; Smilowitz, J. T.; Ruhaak, L. R.; Huang, J.; Lonnerdal, B.; German, J. B.; Medrano, J. F.; Weimer, B. C.; Lebrilla, C. B., Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria-host interactions. *Mol Cell Proteomics* **2012**, *11* (6), M111 015248.
29. Ninonuevo, M. R.; Park, Y.; Yin, H.; Zhang, J.; Ward, R. E.; Clowers, B. H.; German, J. B.; Freeman, S. L.; Killeen, K.; Grimm, R.; Lebrilla, C. B., A strategy for annotating the human milk glycome. *J Agric Food Chem* **2006**, *54* (20), 7471-80.
30. Wu, S.; Tao, N.; German, J. B.; Grimm, R.; Lebrilla, C. B., Development of an annotated library of neutral human milk oligosaccharides. *J Proteome Res* **2010**, *9* (8), 4138-51.
31. Li, Q. Y.; Xie, Y. X.; Wong, M. R.; Barboza, M.; Lebrilla, C. B., Comprehensive structural glycomic characterization of the glycocalyxes of cells and tissues. *Nature Protocols* **2020**, *15* (8), 2668-2704.
32. Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A., Determining the polysaccharide composition of plant cell walls. *Nat Protoc* **2012**, *7* (9), 1590-607.
33. Blakeney, A. B.; Harris, P. J.; Henry, R. J.; Stone, B. A., A Simple and Rapid Preparation of Alditol Acetates for Monosaccharide Analysis. *Carbohyd Res* **1983**, *113* (2), 291-299.
34. Doares, S. H.; Albersheim, P.; Darvill, A. G., An Improved Method for the Preparation of Standards for Glycosyl-Linkage Analysis of Complex Carbohydrates. *Carbohyd Res* **1991**, *210*, 311-317.
35. Anumula, K. R.; Taylor, P. B., A Comprehensive Procedure for Preparation of Partially Methylated Alditol Acetates from Glycoprotein Carbohydrates. *Anal Biochem* **1992**, *203* (1), 101-108.
36. Rohrer, J. S., High-performance anion-exchange chromatography with pulsed amperometric detection for the determination of oligosaccharides in foods and agricultural products. *Acs Sym Ser* **2003**, *849*, 16-31.
37. Hanko, V. P.; Rohrer, J. S., Determination of carbohydrates, sugar alcohols, and glycols in cell cultures and fermentation broths using high-performance anion-exchange chromatography with pulsed amperometric detection. *Anal Biochem* **2000**, *283* (2), 192-199.
38. Castillo, J. J.; Couture, G.; Bacalzo, N. P., Jr.; Chen, Y.; Chin, E. L.; Blecksmith, S. E.; Bouzid, Y. Y.; Vainberg, Y.; Masarweh, C.; Zhou, Q.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lemay, D. G.; Lebrilla, C. B., The Development of the Davis Food Glycopedia-A Glycan Encyclopedia of Food. *Nutrients* **2022**, *14* (8).
39. Castillo, J. J.; Galermo, A. G.; Amicucci, M. J.; Nandita, E.; Couture, G.; Bacalzo, N.; Chen, Y.; Lebrilla, C. B., A Multidimensional Mass Spectrometry-Based Workflow for De Novo

Structural Elucidation of Oligosaccharides from Polysaccharides. *J Am Soc Mass Spectrom* **2021**, 32 (8), 2175-2185.

40. Nandita, E.; Bacalzo, N. P., Jr.; Ranque, C. L.; Amicucci, M. J.; Galermo, A.; Lebrilla, C. B., Polysaccharide identification through oligosaccharide fingerprinting. *Carbohydr Polym* **2021**, 257, 117570.

41. Ndukwe, I. E.; Black, I.; Heiss, C.; Azadi, P., Evaluating the Utility of Permethyated Polysaccharide Solution NMR Data for Characterization of Insoluble Plant Cell Wall Polysaccharides. *Anal Chem* **2020**, 92 (19), 13221-13228.

42. Perez Garcia, M.; Zhang, Y.; Hayes, J.; Salazar, A.; Zabolina, O. A.; Hong, M., Structure and interactions of plant cell-wall polysaccharides by two- and three-dimensional magic-angle-spinning solid-state NMR. *Biochemistry* **2011**, 50 (6), 989-1000.

43. Zhao, W. C.; Fernando, L. D.; Kirui, A.; Deligey, F.; Wang, T., Solid-state NMR of plant and fungal cell walls: A critical review. *Solid State Nucl Mag* **2020**, 107.

44. Wu, Z. Q.; Serie, D.; Xu, G. G.; Zou, J., PB-Net: Automatic peak integration by sequential deep learning for multiple reaction monitoring. *J Proteomics* **2020**, 223.

45. Ranque, C. L.; Stroble, C.; Amicucci, M. J.; Tu, D.; Diana, A.; Rahmannia, S.; Suryanto, A. H.; Gibson, R. S.; Sheng, Y.; Tena, J.; Houghton, L. A.; Lebrilla, C. B., Examination of Carbohydrate Products in Feces Reveals Potential Biomarkers Distinguishing Exclusive and Nonexclusive Breastfeeding Practices in Infants. *J Nutr* **2020**, 150 (5), 1051-1057.

46. Patnode, M. L.; Guruge, J. L.; Castillo, J. J.; Couture, G. A.; Lombard, V.; Terrapon, N.; Henrissat, B.; Lebrilla, C. B.; Gordon, J. I., Strain-level functional variation in the human gut microbiota based on bacterial binding to artificial food particles. *Cell Host Microbe* **2021**, 29 (4), 664-673 e5.

Chapter 3

Glycomic Mapping of the Maize Plant Points to Greater Utilization of the Entire Plant

ABSTRACT

The goal of food sustainability is possible if greater utilization of plants are achieved. In corn only the kernels are currently used for human consumption, however edible carbohydrates that may function as dietary fiber are present throughout the plant. A glycomic map of the maize plant was obtained providing a broad structural view of the carbohydrate distribution revealing that non-cellulosic material was present throughout. Newly developed rapid throughput liquid chromatography tandem mass spectrometry (LC-MS/MS) based methods for analyzing monosaccharide and linkage compositions show unique structural features in the respective segments and parts of the plants from the roots to the tassel. The most abundant monosaccharides of the 14 that were monitored included glucose, xylose and arabinose. Additionally, galactose, fructose, rhamnose, mannose, galacturonic acid, and glucuronic acid were found in lower abundances. The relative abundances of each monosaccharide varied with the parts of the plants. Linkage compositions also varied and provided further structural information that included the presence of polysaccharides such as xylans, starch, pectins, xyloglucans, arabinans, galactans and β -glucans. The nonstructural carbohydrate components including the free mono- and disaccharides were also measured to provide a unique geographical map of their abundances. The glycomic map of corn would guide traditional plant breeding methods and new genome editing tools toward tissue specific enhancements of carbohydrate polymers that have unique and specific functional utility.

INTRODUCTION

The world population is estimated to reach nearly 10 billion by the year 2050.¹ Thus, an increased demand on existing food and energy supply chains is expected, especially in the setting of climate change-associated food shortages. Not surprisingly, there is a growing interest in designing multi-purpose crops.²⁻⁶ Carbohydrates, being the most abundant biomolecule class on earth and a component of every food, will play a central role in addressing these issues. In this context, maize constitutes an important opportunity as one of the most widely and abundantly cultivated grains worldwide. Consumption of corn varies by region with the highest among developing countries. Together with wheat, they represent about 80% of cereal requirements in these regions.⁷ As food, maize has been bred over millennia to maximize the yield primarily of the grain.⁸ There have been enormous resources dedicated to breeding programs that ensure genetic diversity in the collections as well as in the broader global effort to assemble, document, and utilize the resulting efforts. Despite world production reaching 1.1 billion tons in 2020,⁷ only a small fraction of that crop is destined for human consumption, the remainder of its dry mass is largely relegated to use as livestock feed and biofuel production.⁹ The remaining vegetative tissue after harvest of the grain, the stover, has carbohydrate as its principal component.¹⁰⁻¹² However, current usage of corn stover pays little attention to the potential of this major product as human food. While a large fraction of these carbohydrates is in the form of the linear homopolysaccharide cellulose,¹¹ there are similarly abundant polysaccharides that are bioactive with functions that include modulating the gut microbiome and can collectively be classified as dietary fiber.¹³⁻¹⁶ These carbohydrates could serve as important additional sources of bioenergy and nutrition for animals and humans.

Quantitation of cellulose, the primary component of grass cell walls, requires hydrolysis with sulfuric acid and is readily achieved with well-established protocols.^{10, 17, 18} In contrast, the primary objective in this work was to identify in maize the distributions of more easily fermentable, non-cellulosic polysaccharides, which have potential as alternative sources of human nutrition. The first hurdle to do so was to develop analytic methodologies to detect these non-cellulosic polysaccharides, as the existing methods for analysis of diverse cell wall components are outdated, relatively unchanged for decades. These older methodologies involve lengthy sample preparation, long chromatographic run times, and relatively low sensitivity, which impede their use for large sample sets.¹⁷ To address the need for high-throughput analytical methods capable of providing an in-depth structural understanding of plant carbohydrates, we developed a robust LC-MS/MS-based platform for the analysis of plant glycomes. These methods allowed us to quantify the total monosaccharides, free saccharides, and glycosidic linkages within each component of the maize plant. The plant was divided into 213 tissue samples, each subjected to three analyses. The resulting 639 analyses were used to construct an in-depth glycomic map of maize.

Determining carbohydrate components and their spatial distributions throughout the plant in greater detail is the primary focus of this work. These findings provided a much higher resolution picture of corn stover carbohydrate composition than what was previously possible, highlighting more possible nutritional value of this abundant agricultural byproduct. The analytical methodologies developed and employed here are a platform for rapid-throughput and robust glycomic mapping of other plants, which will aid in their development as multi-purpose crops and ultimately lead to improved food security. The foundational knowledge and the analytical tools employed could guide selective breeding and genetic modifications to increase the abundances of

specific polysaccharides ultimately increasing the crops' sustainability and nutritional and economical value.

MATERIALS AND METHODS

Materials. Sodium acetate, trifluoroacetic acid (TFA), chloroform (HPLC grade), ammonium acetate, ammonium hydroxide solution (NH₄OH) (28-30%), sodium hydroxide pellets (semiconductor grade, 99.99% trace metals basis), dichloromethane, anhydrous dimethyl sulfoxide (DMSO), iodomethane, 3-methyl-1-phenyl-2-pyrazoline-5-one (PMP), methanol (MeOH, HPLC grade), fructose, ribose, rhamnose, mannose, allose, glucuronic acid, galacturonic acid, glucose, galactose, N-acetylglucosamine, N-acetylgalactosamine, xylose, arabinose, fucose, maltose, and sucrose were purchased from Sigma-Aldrich (St. Louis, MO). Acetonitrile (HPLC grade) was purchased from Honeywell (Muskegon, MI). Viscozyme® was provided by Novozyme (Davis, CA). Nanopure water was used for all experiments.

Preparation of Samples. A 10-foot-tall dent corn plant (*Zea mays* var. *indentata*), also known as field corn, was acquired from a small farm located in Wheatland, California. The plant was harvested 10 weeks after planting and was grown under atmospheric conditions with no additional watering.

The harvested plant was left overnight in the dark before being prepared for sampling. First, the plant was thoroughly washed with nanopure water and staged using a method known within the corn agronomy in the U.S. based on the amount of present leaf collars denoted as discolored line between the leaf blade and leaf sheath with V1 indicating the first leaf closest to the roots and V(n) with (n) being the leaf closest to the tassels.¹⁹ There are two main growth stages,

vegetative (V) and reproductive (R) with stage V referring to anatomical parts indirectly related to the corn ear, the R segment. For the convenience of naming, this system was modified to accommodate for the specificity of sample collection. The subject is a V12 plant with 12 leaf collars, dividing the plant into segments called internodes bounded by the nodes above and below. Within each segment, sample collection consisted of a leaf collar, leaf, leaf midrib, stalk sheath, stalk node, and stalk. Each stalk segment was separated into rind and pith (outer and inner, respectively) components when possible. Each component was then further broken down into 3 or 4 more parts. For example, the samples “V2 Rind” refer to the outer stalk portions of the segment bounded by the V2 and V3 nodes with the “V2 Rind1” being closest to the V3 node and the “V2 Rind3” sample being closest to the V2 node. The leaves were collected in a similar fashion with “V2 Leaf Tip” being the upper most portion of the leaf blade and the “V2 Leaf Base” being the portion closest to the stalk.

The reproductive stage of this V12 plant has 2 corn regions denoted as R1 and R2. R1 is the first corn region closest to the roots at V4 segment (V4R1), containing 12 husks with “V4R1 Husk1” being the outermost layer and “V4R1 Husk12”) being the innermost layer and closest to the kernels. R2 locates between V5 and V6 and contains 6 husk layers. Each corn region was collected for their kernels, cobs, husks, inner and outer silk. The corn ear was processed in a similar fashion as described above where it was divided into 3-4 equal parts with the kernels carefully separated from their respective cobs.

Other important areas are the tassel and root system. Vt denotes tassel region with “Vt Stalk” being the stalk where the tassels sprouted, and tassel stems and flowers collected for analysis. The root system was collected for the nodules by themselves and with their nodes, the brace roots and seminal roots above and below ground level, as well as the propagating roots.

This highly thorough collection method allowed for a total of 213 samples where each sample was lyophilized to complete dryness before being pulverized into a fine powder using a Bead Ruptor Elite Bead Mill Homogenizer (Omni International, Kennesaw, GA). 10 mg/mL stock solutions of plant tissue were prepared in nanopure water and bullet blended with stainless steel beads at speed four for two minutes before being heated to 100°C for 1 hr. Once cooled, the samples were subjected to one more round of bullet blending.

Free Monosaccharide and Total Monosaccharide Analysis

Free Saccharide Analysis. Samples underwent a 10-fold dilution and were transferred to a 96-well plate for derivatization. A cereal quality control sample was also prepared and analyzed alongside the maize samples to assess reproducibility. The coefficients of variation (CVs) for the technical replicates are provided in **Supplementary Table 3.3b**. A procedure using PMP was adapted from Xu et al.²⁰ Briefly, a pooled standard solution of 14 monosaccharides was prepared and serially diluted to concentrations of 0.001 µg/mL to 100 µg/mL. Additionally, a standard solution containing the disaccharides sucrose and maltose was prepared and serially diluted to the same concentrations. A solution containing equal parts (v/v) of ammonium hydroxide solution (28-30% v/v) and 0.2 M PMP in methanol was prepared and added to each sample in a 96-well plate. The samples were heated to 70°C for 30 min followed by vacuum centrifugation to complete dryness. Next, the samples were reconstituted in 250 µL of nanopure water and washed twice with chloroform.

Total Monosaccharide Analysis. Aliquots of each stock solution underwent an enzymatic treatment with Viscozyme® (Novozyme, Davis, CA) at 50°C in 25 mM sodium acetate buffer (pH 5.0). Samples were then subjected to acid hydrolysis with 4 M TFA for 1 h at 121°C in 96-well

plates. Hydrolysis was quenched by the addition of cold nanopure water. The released monosaccharide residues were then derivatized and extracted according to the procedure previously described. A single sample was chosen and analyzed in triplicate to assess method reproducibility. The coefficients of variation (CVs) for the technical replicates are provided in **Supplementary. Table 3.1.**

Instrumental Analysis. Separation of the PMP-labeled monosaccharides was carried out on an Agilent 1290 Infinity II UHPLC system equipped with a 2-position/10-port switching valve in the column compartment. The LC stack was coupled to an additional binary pump and set up for automated column regeneration (ACR). Mass spectral analysis was carried out on an Agilent 6495A triple quadrupole (QqQ) mass spectrometer (Agilent Technologies, Santa Clara, CA). For analysis, 1 μ L of sample was injected onto one of two Agilent Poroshell HPH C18 column (2.1 mm \times 50 mm i.d., 1.8 μ m particle size) equipped with Agilent Poroshell HPH C18 guard cartridges (2.1 mm \times 5 mm i.d., 1.8 μ m particle size). Binary pump 1 was set to a 2.2 min isocratic gradient of 12% B with a constant flow rate of 1.050 mL/min for separation of compounds. Binary pump 2 was set to repeat a column regeneration and equilibration sequence with the following gradient: 0-0.1 min, 12% B; 0.1-0.2 min, 99% B; 0.2-1.4 min, 99% B; 1.4-1.5 min, 12% B; 1.5-2.2 min, 12% B. Mobile phase A consisted of 25mM ammonium acetate with pH adjusted to 8.2 with ammonium hydroxide solution in 5% (v/v) acetonitrile. Mobile phase B consisted of 95% (v/v) acetonitrile in water.

Linkage Analysis

Permethylation, hydrolysis and derivatization of plant tissue samples. A permethylation procedure was adapted from Galermo et al.^{21, 22} Aliquots containing 50 μ g of plant tissue were

transferred to a 96-well plate and permethylated using iodomethane in a solution of DMSO containing saturated NaOH. The samples were allowed to react on a shaker at room temperature for 50 min under argon before being quenched by the addition of cold water. A liquid–liquid extraction using DCM and cold water was performed and repeated five times to remove excess NaOH and DMSO. The upper aqueous layer was discarded while the bottom organic layer containing permethylated products was dried to completion by vacuum centrifugation. Permethylated samples were then subjected to acid hydrolysis at 100°C with 4 M TFA and subsequently dried by vacuum centrifugation. The released permethylated monosaccharide residues were derivatized with PMP following the previously described procedure. A single sample was chosen and analyzed in triplicate to assess method reproducibility. The coefficients of variation (CVs) for the technical replicates are provided in **Supplementary. Table 3.2**.

Instrumental Analysis. Separation and analysis of the permethylated glycosides were carried out on an Agilent 1290 Infinity II UHPLC system coupled to an Agilent 6495A triple quadrupole (QqQ) mass spectrometer (Agilent Technologies, Santa Clara, CA). For analysis, 1 μ L of sample was injected onto an Agilent Zorbax RRHD Eclipse Plus C18 column (2.1 mm \times 150 mm i.d., 1.8 μ m particle size) equipped with an Agilent Zorbax Eclipse Plus C18 guard cartridge (2.1 mm \times 5 mm i.d., 1.8 μ m particle size) and separated using a 15 min binary gradient with a constant flow rate of 0.22 mL/min. Mobile phase A consisted of 25 mM ammonium acetate with pH adjusted to 8.2 with ammonium hydroxide solution in 5% (v/v) acetonitrile. Mobile phase B consisted of 95% (v/v) acetonitrile in water. The following binary gradient was used: 0.00–5.00 min, 21.00% B; 5.00–9.00 min, 21.00–22.00% B; 9.00–11.00 min, 22.00% B; 11.00–13.60 min, 22.00–24.50% B; 13.60–13.61 min, 24.50– 99.00% B; 13.61–13.80 min, 99.00% B; 13.80–13.81 min,

99.00–21.00% B; 13.81–15.00 min, 21.00% B. Samples were introduced into the mass spectrometer using an electrospray ionization (ESI) source operated in the positive ion mode. Nitrogen drying and sheath gas temperatures were set at 290 and 300 °C, respectively. Drying and sheath gas flow rates were set at 11 and 12 L/min, respectively. The nebulizer pressure was set to 30 psi. Capillary and fragmentor voltages were set at 1800 and 380 V, respectively. Data was acquired using multiple reaction monitoring (MRM) mode. The collision energy was set to 35 eV. Data analysis was performed using Agilent MassHunter Quantitative Analysis software version B.08.00

RESULTS AND DISCUSSION

Overview of Analytical Workflow

A rapid-throughput, LC-MS-based workflow was developed and employed for the profiling of both structural and non-structural carbohydrates in plant tissues. A snapshot highlighting the utility of the methodology is provided in **Figure 3.1**. This workflow consisted of three separate analyses each applied to all 213 tissue samples, providing unique information towards the overall carbohydrate composition of the maize plant. A total monosaccharide compositional analysis was performed first. This analysis provided quantitative information on the total non-cellulosic carbohydrates present. A glycosidic linkage compositional analysis was then performed to determine how the observed monosaccharide residues were connected and provide structural information. The monosaccharide and linkage compositions provided general information on the polysaccharide compositions throughout the maize plant. Lastly, a free saccharide analysis was performed to understand how non-structural carbohydrates such as free fructose, glucose, and sucrose were distributed throughout the tissues of the plant.

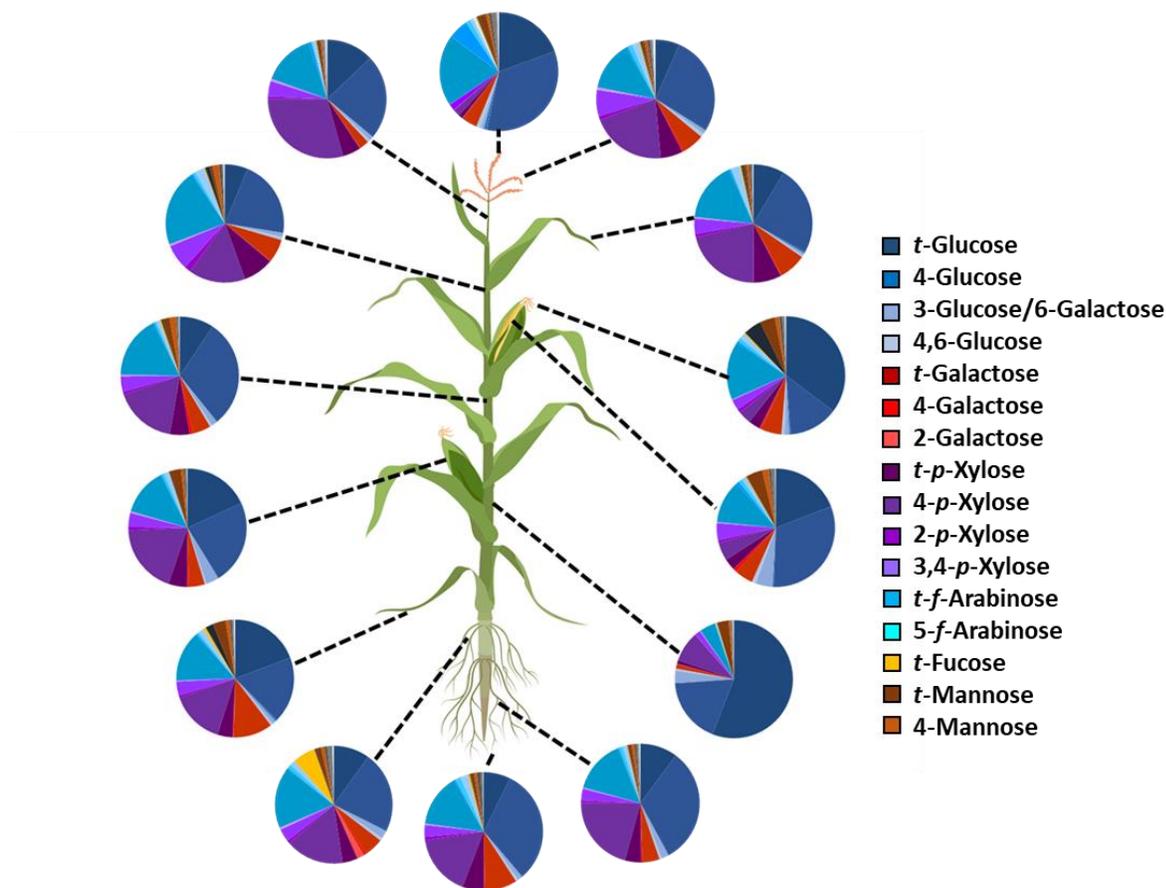


Figure 3.1: A snapshot of the structural and compositional information obtained by applying rapid throughput, whole-plant glycomic analyses. The pie charts depict the relative glycosidic linkage composition of the indicated tissues. The linkage composition was observed to be highly heterogeneous throughout the plant indicating that the carbohydrate structures and abundances differed in the various tissues.

Monosaccharide Composition Map of the Maize Plant

To obtain quantitative information on the non-cellulosic structural carbohydrates, we subjected each tissue sample to the optimized workflow for absolute quantitation of

monosaccharides. **Supplementary Figure 3.1** shows the extracted ion chromatograms (EICs) of all monosaccharides contained in the standard pool that were quantified in the maize plant.

The respective retention times, MRM transitions, and collision energies are detailed in **Supplementary. Table 3.4**. Of the 14 monosaccharides monitored in the method, glucose, xylose, and arabinose were the most abundant throughout the plant. Galactose, fructose, rhamnose, mannose, galacturonic acid, glucuronic acid, and fucose were also present with lower abundances (**Figure 3.2**).

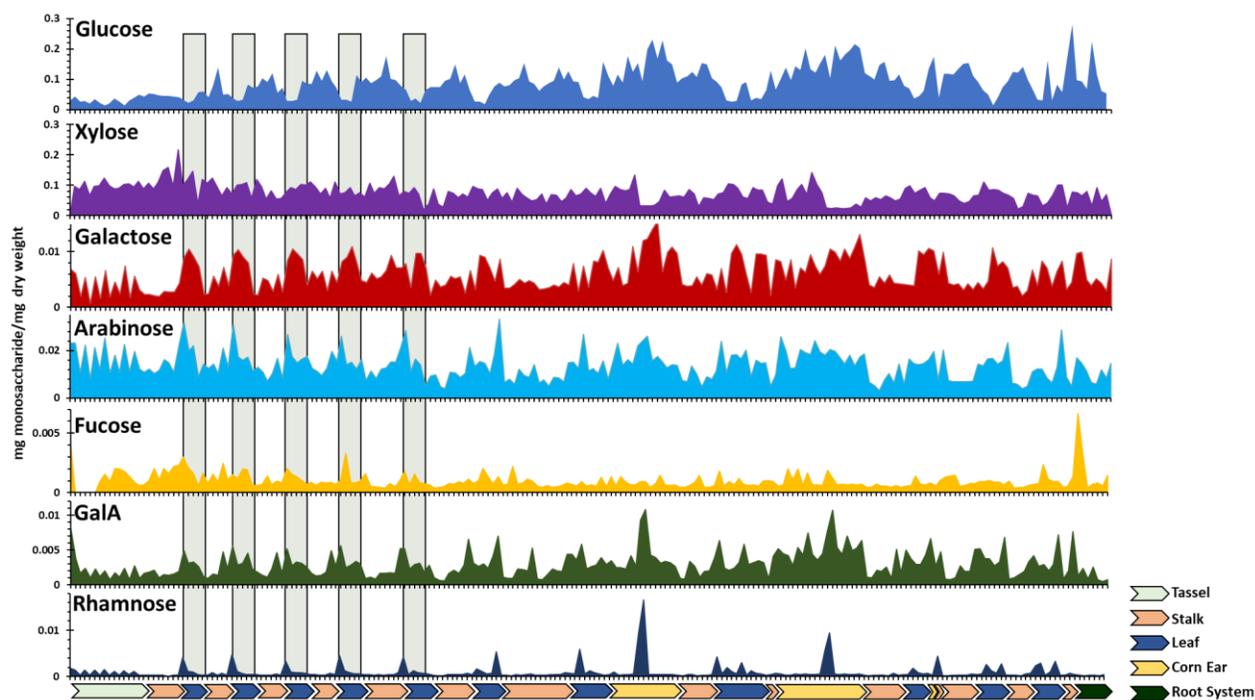


Figure 3.2. Total monosaccharide composition of the maize plant from top to bottom (left to right). The most abundant monosaccharides detected are shown (note the different scales of the axes). The colored bar along the x-axis and key indicate the tissue type. Monosaccharide compositions differed between tissues with the stalk containing more glucose than other portions

of the plant. The leaves consistently displayed more galactose, arabinose, and galacturonic acid (GalA) than the stalk especially towards the top of the plant (highlighted with shaded green boxes). The leaf collars are readily identified by their large rhamnose content while the brace roots their fucose content.

Large differences in total monosaccharide composition were observed across different tissues. The leaves showed relatively small amounts of glucose relative to the stalk while galactose, arabinose, and galacturonic acid were found to be significantly higher in abundances. The tassels also exhibited increased abundances of these monosaccharide residues relative to the stalk and roots. Galactose, arabinose, and galacturonic acid are constituents of pectins, suggesting that the pectin content of the leaves and tassels was significantly higher than in the stalk. Similarly, the abundances of galacturonic acid were significantly higher in the stalk pith relative to the stalk rind throughout the plant (**Supplementary. Figure 2.2**). The average galacturonic acid content was also found to be highest collectively in the corn ears, lesser in the leaves and even less in the stalk (**Supplementary. Figure 2.3**). Because galacturonic acid belongs principally to pectic polysaccharides, the results likely indicate higher pectin content in the corn ears. The leaf collars contained a significantly greater abundances of rhamnose compared to all other stalk tissues. The brace roots contained significantly more fucose than all other tissues. Coincidentally, secretions of aerial roots in a landrace maize from Sierra Mixe contain similarly high levels of fucose, which is commonly found in animal carbohydrates and more unusual in plants.²³ The corn husks, like the leaves, contained greater amounts of glucose, galactose, arabinose, and galacturonic acid compared to other tissues. The corn kernels and cobs, however, had the highest total glucose abundances and lowest non-glucose residues owing to their relatively high starch content (**Figure 2.2**).

Plotting the monosaccharide abundances between different monosaccharides yielded correlations pointing to specific polysaccharides (**Figure 3.3a, b**). The combination of galactose, arabinose, and galacturonic acid, when correlated point to the presence of pectin. Along the stalk, xylose and arabinose produced strong positive correlations, and both tended to increase towards the top of the plant (**Figure 3.3d, e**). The two monosaccharides make up arabinoxylans, which are constituent polysaccharides of several foods including rice and other grains.^{12, 24, 25} Xylose and arabinose were negatively correlated with glucose (primarily from starch, **Figure 3.3c**) suggesting that the presence of starch corresponds to the absence of arabinoxylans. Starch was therefore greater at the bottom of the plant and arabinoxylans were greater toward the top of the plant.

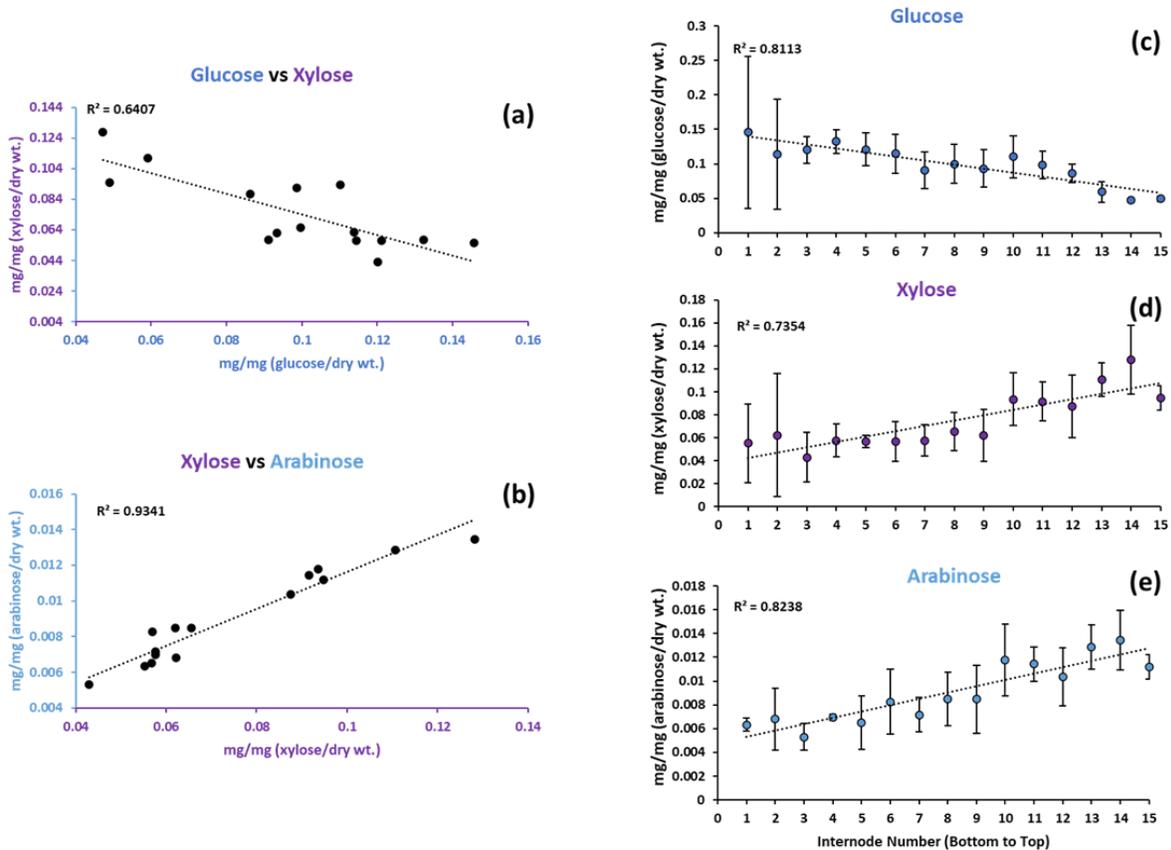


Figure 3.3. Relationship between the three most abundant monosaccharides (glucose, xylose, and arabinose) detected in maize stalk. Each point represents an internode for which all tissue samples in that internode were averaged. (a) Glucose and xylose displayed a strong negative correlation throughout the stalk. (b) Xylose and arabinose displayed a strong positive correlation throughout the stalk. (c) The concentration of glucose in maize internodes decreased from top to bottom of the plant. (d) The xylose and (e) arabinose concentration decreased in internodes from top to bottom of the plant. Error bars represent the standard deviation.

Glycosidic Linkage Distribution in Maize

The structural characteristics and relative abundances of the polysaccharides present in each sample were determined using glycosidic linkage analysis. Linkage analysis was performed using a recently developed rapid throughput method that yielded relative quantitation.^{21, 22} **Supplementary Figure 3.4** shows the EICs of each linkage represented in a pool of oligosaccharide standards containing the most common linkages encountered in plants. Retention times, MRM transitions, and collision energies are provided in **Supplementary Table 3.5**. Linkage analysis confirmed trends observed in the monosaccharide analysis and provided specific information on the structural features of the polysaccharides present in each tissue. The method monitors nearly 100 different linkages, however only 34 were observed in the sample. Based on this analysis, 4-linked glucose (4-Glc), 4-linked xylose (4-Xyl), and terminal arabinose (*t*-Ara) comprised the largest fraction (**Figure 3.4**) indicating that the majority of the carbohydrate content throughout the plant was comprised of cellulose and xylans. Due to the nature of the methods, linkage analysis samples the cellulose, while monosaccharide analysis does not. The two branching residues, 2,4- and 3,4-xylose, together with a large abundance of *t*-Ara also indicated

the presence of arabinoxylan throughout the plant.¹⁷ This arabinoxylan structure could be further described as being branched primarily through a 3,4-xylose linkage, which was the primary branching residue detected in the analysis.

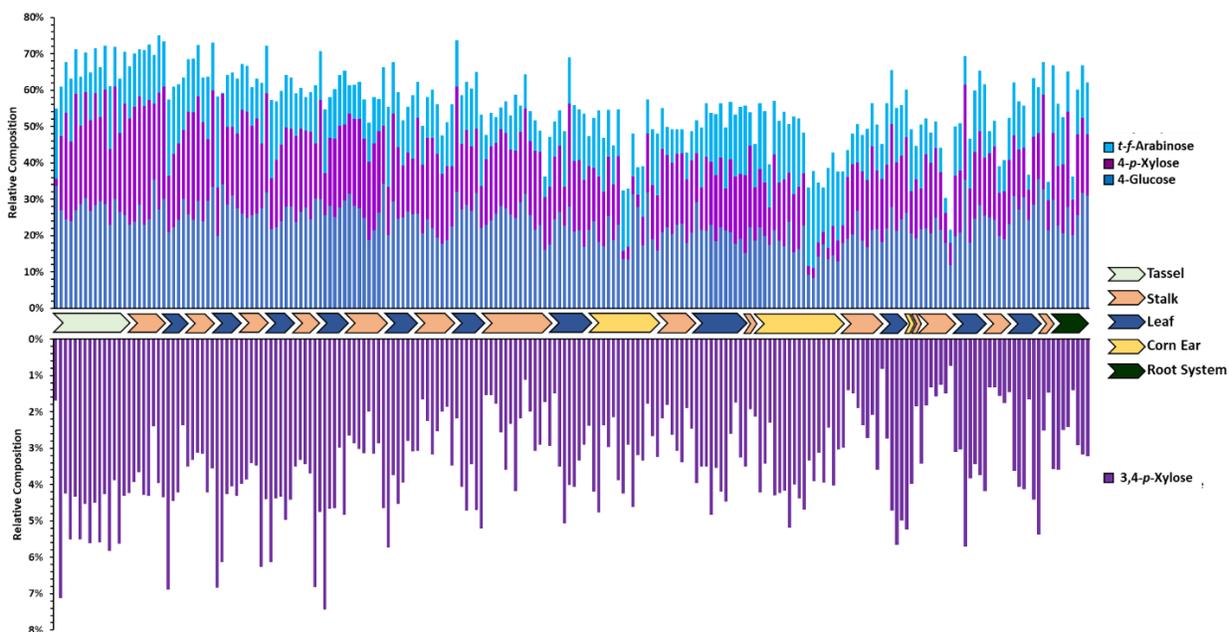


Figure 3.4. The most abundant glycosidic linkages detected throughout the maize plant. The large amounts of 4-glucose, 4-xylose, 3,4-xylose, and *t*-arabinose indicated that cellulose, xylan, and arabinoxylan were the most abundant polysaccharides present throughout the plant. However, their distributions were heterogeneous with the lower portions of the stalk containing less branching residues.

The next most abundant linkages were terminal xylose and terminal galactose, which were accompanied by increases in 2-xylose (**Supplementary Figure 3.5**). Together, this provided evidence that xyloglucan was the most abundant hemicellulose present after xylans, which was consistent with what was known of grass cell walls.^{26,27} Furthermore, these xyloglucan-associated

linkages were found to be higher in the tassel flower, leaf, husk, and root tissues compared to the stalk.

Linkage analysis also revealed structural information and relative abundances of the more minor hemicellulose components present in differing amounts throughout the plant. Terminal-, 4-, and 4,6-mannose residues were detected in all tissue samples indicating that the cell walls contained small amounts of mannans (**Supplementary Figure 3.6**). These residues had higher abundances specifically in the stalk nodes and leaf collars. The stalk regions below the corn ears also tended to have higher amounts of these linkages than the same tissues above. In addition to mannan-associated linkages, 3-linked glucose was also detected throughout the plant indicating the possible presence of β -glucan (**Supplementary Figure 3.7**). Overall, this linkage generally displayed the lowest abundances in the leaf blades, tassels, and roots with the highest abundances in the stalk nodes, stalk sheaths, and corn husks followed by the leaf midribs and stalk tissues. Below the ears of corn, the stalk tended to contain more of the 3-linked residue relative to stalk tissues above.

The increased arabinose content seen in the total monosaccharide analysis of the leaves were found to be due to relative increases in 2- and 3-arabinose, suggesting a corresponding increase in branched arabinans (**Supplementary Figure 3.8**).^{28, 29} On the other hand, absolute galactose content of the leaves was noted to correspond to similar increases in 4-, 6-, 3,6-linked and terminal galactose residues. The increases in 4-galactose, however, seemed to be associated only with the stalk node and stalk sheath (**Supplementary Figure 3.9**), while the 6- and 3,6-galactose were noted to be associated with the leaves themselves (**Supplementary Figure 3.10**). This finding suggested differences in the sidechain structures of the pectic polysaccharides within

these tissues. The 4-galactose indicated linear galactan while the 6- and 3,6- linkages are unique to arabinogalactan.¹⁷

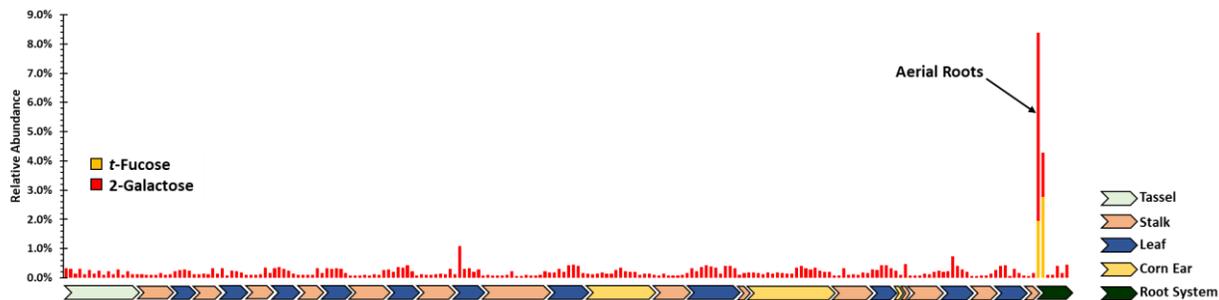


Figure 3.5. The brace roots and node displayed the highest abundance of *t*-fucose while 2-galactose was exclusively observed in these tissues. These linkages likely belong to a selectively fucosylated xyloglucan or mucilage polysaccharide structure.

Unique compositional differences were observed in the brace roots and pollen of the maize plant. The brace roots displayed the highest concentration of fucose compared to all other plant tissues, and this finding was confirmed by linkage analysis, which produced terminal fucose (*t*-Fuc). Furthermore, 2-galactose was observed only in these tissues (**Figure 3.5**). These observations together suggested the presence of a uniquely fucosylated polysaccharide such as a fucosylated xyloglucan or the remnants of a mucilage containing both *t*-Fuc and 2-Gal that was previously found in a landrace of maize from Sierra Mixe.^{23, 30}

The pollen had very little xylose and instead contained elevated amounts of galactose, fucose, rhamnose, galacturonic acid, and arabinose. Despite the sample size, the analytical platform was sufficiently sensitive to elucidate the structures of even small plant components. Linkage analysis revealed relatively higher abundances of 6-galactose, *t*-galactose, *t*-fucose, 5-

arabinose, 3-arabinose, and 2-arabinose with 2,5-arabinose being detected exclusively in the pollen. The total monosaccharide and linkage analysis of the pollen suggests the presence of a cell wall rich in pectic polysaccharides.^{28, 29}

Free Mono- and Disaccharide Distribution in the Maize Plant

The distribution of free mono- and disaccharides were quantitatively measured throughout the whole plant. Of the 16 compounds (monosaccharides and two disaccharides) monitored, glucose, fructose, and sucrose were found to be the most abundant free mono- and disaccharides (**Supplementary Table 3.3a**). The concentration of each were markedly different both across and within tissues (**Figure 3.6**). Total free sugar (fructose, glucose, and sucrose) concentrations increased from the top to bottom of the plant. However, sucrose was found to be sequestered in the stalk, while the highest concentrations of glucose and fructose were found in the tissues of the corn ears. Furthermore, leaves were observed to contain minimal mono- and disaccharides with the majority being found in the stalk.

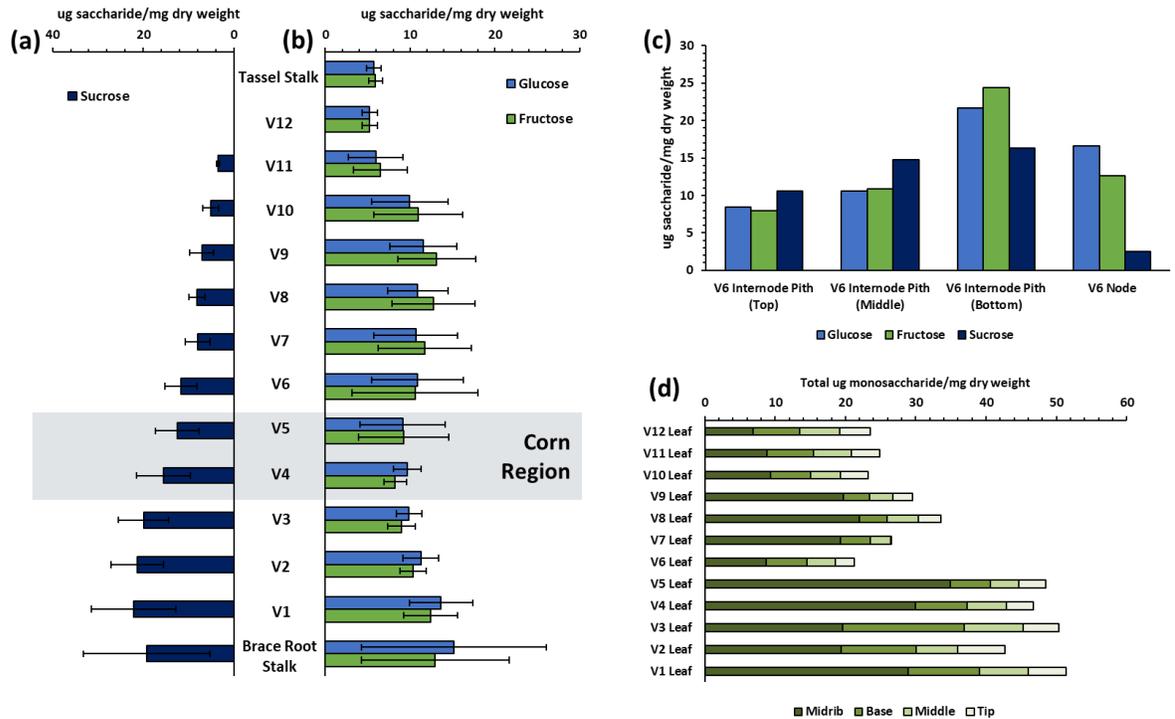


Figure 3.6. Distribution of free glucose, fructose, and sucrose in the stalk and leaf tissues. (a) and (b) Concentration of sucrose and glucose/fructose, respectively, in all stalk internodes from top to bottom of the plant. Bars indicate averages of all samples taken within each internode. Error bars represent the standard deviation. (c) Example of heterogeneous distribution of free sugars within internodes. The pith just above the node tended to contain the most sugars while the nodes were consistently low or void of sucrose. (d) Sum of fructose and glucose concentrations in each sample within each leaf. Sugar concentrations were observed to follow an increasing gradient from leaf tip to middle, base, and midrib. Leaves below the corn ears (V5 – V1) contained significantly more free sugars than those above the corn ears (Student’s t-test, $p < 3e^{-6}$).

In addition to differences between tissues, there were also variations within tissues. The average concentration of free sugars increased from the top to bottom of the stalk (**Figure 3.6a**,

b). Furthermore, the free sugar concentration tended to increase from top to bottom in the pith of each internode. The stalk nodes, however, were consistently low or void of sucrose (**Figure 3.6c**). The leaves showed a similar trend where the sugar concentrations tended to follow a gradient from the tip of each leaf to its base, while the midribs consistently contained the highest concentrations (**Figure 3.6d**). In addition, the leaves below the corn ears contained significantly higher free sugar concentrations than those found above.

A Full Glycomic View of the Maize Plant

Carbohydrates as human food provide energy and growth due to the presence of glucose, primarily as starch, free glucose, fructose, and sucrose. A map of glycans and polysaccharides in plants can provide parts of the plants that can be consumed and used particularly as fiber. Human digestive enzymes are limited to cleaving polysaccharides with $\alpha(1,4)$ -glucose bonds. Polysaccharides in food that are not cleaved by the host enzymes are called fiber, a broad general term with little structural specificity. The utility to digest fiber is obtained by recruiting microbes with a wide array of carbohydrate-active enzymes (CAZymes) that can cleave a wide range of glycosidic linkages including those from cellulose.¹³ CAZymes such as glycosyl hydrolases (GHs) and polysaccharide lyases (PLs) in the microbiome are highly selective and limited to specific linkages. For example, the major glycosidic linkages 4-Glc, 4-Xyl, and *t-f*-Ara from cellulose and (arabino)xylan observed here could be cleaved by any number of microbial β -glucanases, β -xylanases, and α -arabinosidases, respectively. Although there has been considerable effort in characterizing the microbiome and its role in health and nutrition, much less is known regarding the structures of the food that modulate the microbiome. However, the key to understanding the microbiome and in developing better synbiotics (the combination of pre- and probiotic) lies in

knowing the structures of food and the specificity of the bacterial enzymes. The cell wall polysaccharide components of maize can thus be used to target specific enzymatic functionalities in the gut microbiome with the goal of altering the composition of the bacterial community and host health. This concept has been demonstrated in pea and orange fiber preparations in which supplementation caused increases in the expression of GHs and PLs specific to the polysaccharides comprising the fibers.³¹ Likewise, maize fiber preparations with their primary constituents being cellulose, (arabino)xylan, and xyloglucan could be used target gut microbial species possessing β -glucanases (4-Glc), β -xylanases (4-Xyl), and α -arabinofuranosidases (*t-f-Ara*). Common plant breeding practices will soon be supplemented by gene-editing tools like CRISPR to provide powerful tools to answer global food supply, agricultural, and bioenergy demands.^{6, 32} While the focus of these efforts have been on the grain, modern agriculture should instead broaden it to produce a more broadly useable plant. However, the lack of structural targets in various tissues constrained the utility of these methods. Previous investigations of maize carbohydrate composition have characterized a limited number of polysaccharides from a relatively small number of tissues in the plant.^{10, 33-35} These earlier approaches lacked the sensitivity, quantitation, and sample throughput to provide the spatial distributions of the diverse structural and non-structural carbohydrates in crops. However, such information will help interpret the outcomes of gene-editing and breeding experiments at the whole organism level, accelerating the development of defined multi-purpose crops.

The carbohydrate map can also elucidate the fundamental mechanism of carbohydrate storage and function in plants. The cellulose microfibrils in grass cell walls are thought to be crosslinked and non-covalently bound by xylan and glucuronoxylan through hydrogen bonding interactions. This association is thought to add structural rigidity to the cell wall.^{36, 37} Substitution

of xylan with arabinosyl residues as in arabinoxylan decreases the extent to which hydrogen bonding can occur resulting in less cell wall rigidity. Thus, it is thought that grass cell elongation is positively correlated to xylan substitution with arabinose.³⁸ These conclusions are consistent with observations here. Both total xylose and 4-xylose abundance were consistent in leaf tissues. In the stalk, however, segments toward the top of the plant tended to display greater total and 4-xylose content than segments toward the bottom of the plant. In addition, the top segments of the stalk displayed greater amounts of 3,4-xylose and terminal arabinose indicative of a greater arabinoxylan content. The general decrease of xylan content towards the more mature, bottom portions of the stalk may also suggest that wall hardening in those tissues is due to other processes such as lignification rather than xylan crosslinking.²⁵ Lignification is a prevalent feature of the maize cell wall known to increase recalcitrance and decrease digestibility.² The analytical methodology employed here does not capture this feature, however lignin can be measured using several well-established protocols.³⁹ The leaves also tended to have a greater abundance of 3,4-xylose relative to the stalk indicating a greater degree of arabinose branching. If these hypotheses are true, the more branched xylan structures in the top portions of the plant and leaves could be more easily extracted or digested due to their weaker association with the cellulose framework thus also resulting in less cellulosic recalcitrance.² This could make certain portions of harvested plants more attractive and more efficient sources for a variety of uses including feed and biofuel production.

Although a minor hemicellulose component in grass cell walls,²⁴ xyloglucan may play a role in cell elongation, morphogenesis and as a component of root mucilage.^{30, 36, 40} We identified xyloglucan here by the *t*-Gal, *t*-Xyl, and 2-Xyl linkage, although *t*-Gal may also be present in pectic sidechains.¹⁷ In grasses, substitutions are most commonly observed not to extend past the *t*-Xyl

residue.⁴¹ However, structures exhibiting galactose have been reported and occurrence can even be tissue-dependent.⁴² Indeed, the 2-Xyl linkage was detected throughout the plant here and found to be less abundant in the stalk relative to leaf, husk, and root tissues. Most notably, each leaf collar exhibited significantly higher 2-Xyl levels than surrounding tissues. The abundance of these xyloglucan-associated linkages as a whole were found to be increased in the leaf tissues and tended to decrease from top to bottom portions of the stalk. This finding supports more recent evidence that xyloglucan comprises a larger fraction of the cell wall in meristematic cells likely present in the more immature tissues of the upper stalk and may play an important role in cell elongation and morphogenesis.^{24, 36, 40, 43} Further, these findings suggest that xyloglucan structure may dictate different physiological functions that are yet to be fully resolved. The xyloglucan-associated linkages were also found to increase in the propagating roots, supporting its putative role as a mucilage secreted into the rhizosphere.³⁰

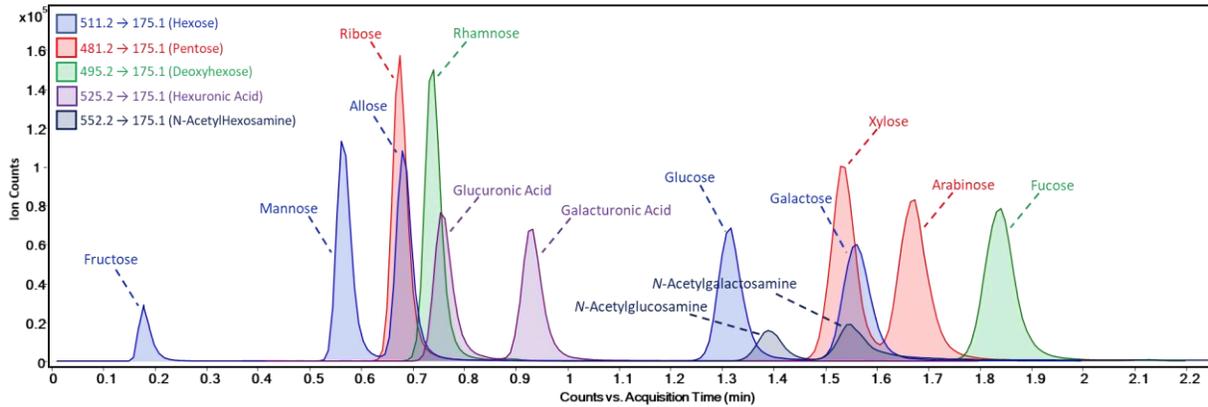
The presence of β -glucan in the cell wall is a feature quite unique to grasses. Its presence in grain has been proposed as an alternate storage polysaccharide during certain development stages and under some environmental conditions.^{44, 45} Early studies into its function in the cell wall also found its accumulation in the elongation phase of coleoptiles leading to the conclusion that β -glucan plays a role in this process.²⁴ However, more recent studies have also shown accumulation of β -glucan in mature stalk tissues of several grass species including maize suggesting a more complex role than previously thought.⁴⁶ The relative abundance of 3-glucose was used as a proxy to compare β -glucan content throughout the maize plant in the present study. An observation unique to this whole-plant glycomic view is the distribution of this 3-linked glucose residue follows the distribution of free glucose, free fructose, and sucrose throughout the plant. This supports the notion that β -glucan functions as an alternate storage polysaccharide not

only in the developing corn kernels of the primary sink but also the mature stalk tissues of the secondary sink.

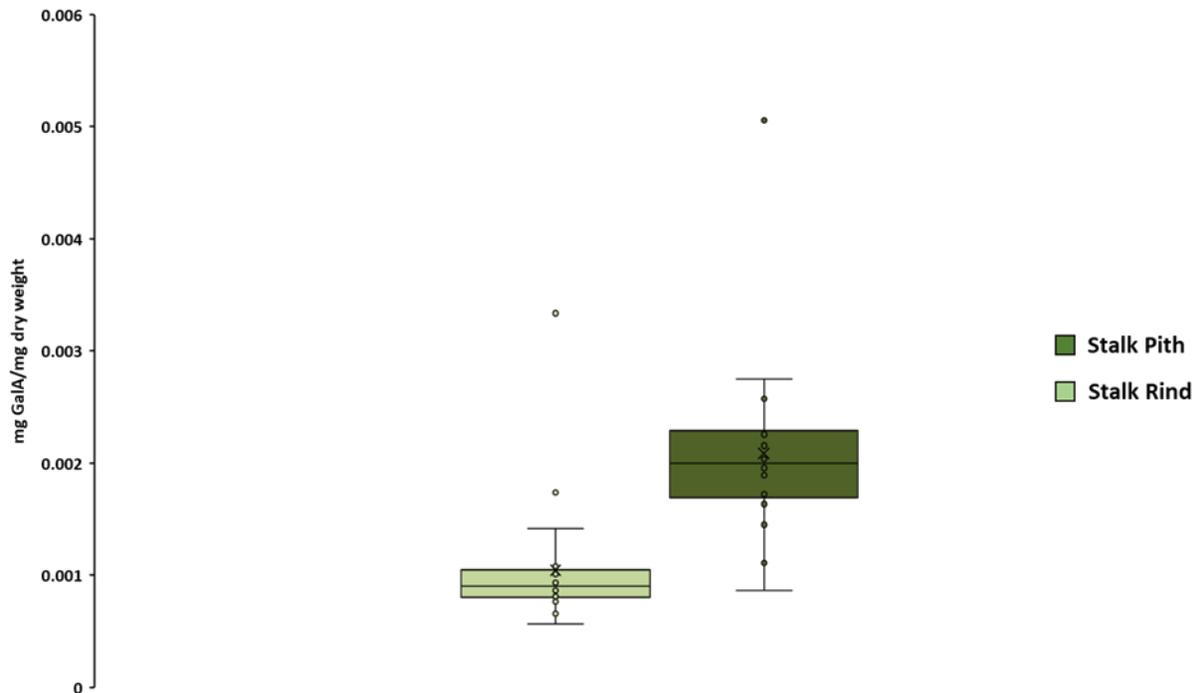
Grass cell walls are known to contain relatively little pectin, but this component nonetheless plays many important roles in the primary cell wall including wall strength, cell adhesion, and cell defense.⁴⁷ Pectins are the most structurally diverse group of cell wall polysaccharides and can be comprised of polygalacturonan and rhamnogalacturonan I/II in addition to extensive neutral sidechains such as arabinan, galactan, and arabinogalactan.^{28, 29, 47} Elucidation of these structures when isolated is a challenging task and analyzing whole tissue as done here only adds to the complexity. However, pectins contain several distinctive structural features that can provide information on how pectin content and structure changes throughout the plant. The pectin backbone can be comprised exclusively of galacturonic acid (GalA) as in polygalacturonan or include rhamnose as in rhamnogalacturonan.^{28, 29, 47} Because these monosaccharide residues belong principally to pectins, they were used to describe overall pectin content as well as rhamnogalacturonan content between tissues. GalA was generally found in greater abundance in the pollen, flowers, leaves and corn ears relative to the tissues of the stalk and root. Further, the inner stalk was found to have more GalA than the outer stalk throughout its length. However, the most significant GalA content was observed in the pollen, stalk nodes, leaf collars and silk indicating the highest pectin content in these tissues. Distribution of rhamnose followed a more pronounced trend with definitive spikes at the leaf collars perhaps indicating particularly large amounts of rhamnogalacturonan in these tissues. Further, rhamnose content followed a gradient in each leaf blade with the highest concentration observed in the base and the lowest at the tip of the leaf. A large fraction of the rhamnose residues in rhamnogalacturonan are substituted with neutral side chains like arabinan, galactan, and

arabinogalactan. The presence of these structures is thought to add flexibility to the cell wall by preventing the formation of cellulose microfibrils.³⁶ Here, the arabinan-associated linkages 2-, 3-, and 5-arabinose were found to have increased abundance in all leaf tissues and tassel flowers. Arabinogalactan-associated linkages 6- and 3,6-galactose were also found to have increased abundance. This finding, along with the greater abundance of the branching 3,4-xylose linkage in xylan, may provide a polysaccharide-based rationale for the decreased recalcitrance and increased flexibility of leaf tissues relative to the stalk. Together, these findings successfully illustrate a comprehensive map of the carbohydrates present in the maize plant. The methods employed to generate this map are rapid throughput, adaptable and can be utilized to elucidate functional features of carbohydrates in plants. The results of these experiments will provide a unique perspective towards genetic traits of the entire corn plant in particular, but in principle all agricultural food candidates making for better human and animal nourishment, crop efficiency, soil regeneration, and agricultural sustainability.

SUPPLEMENTARY MATERIALS

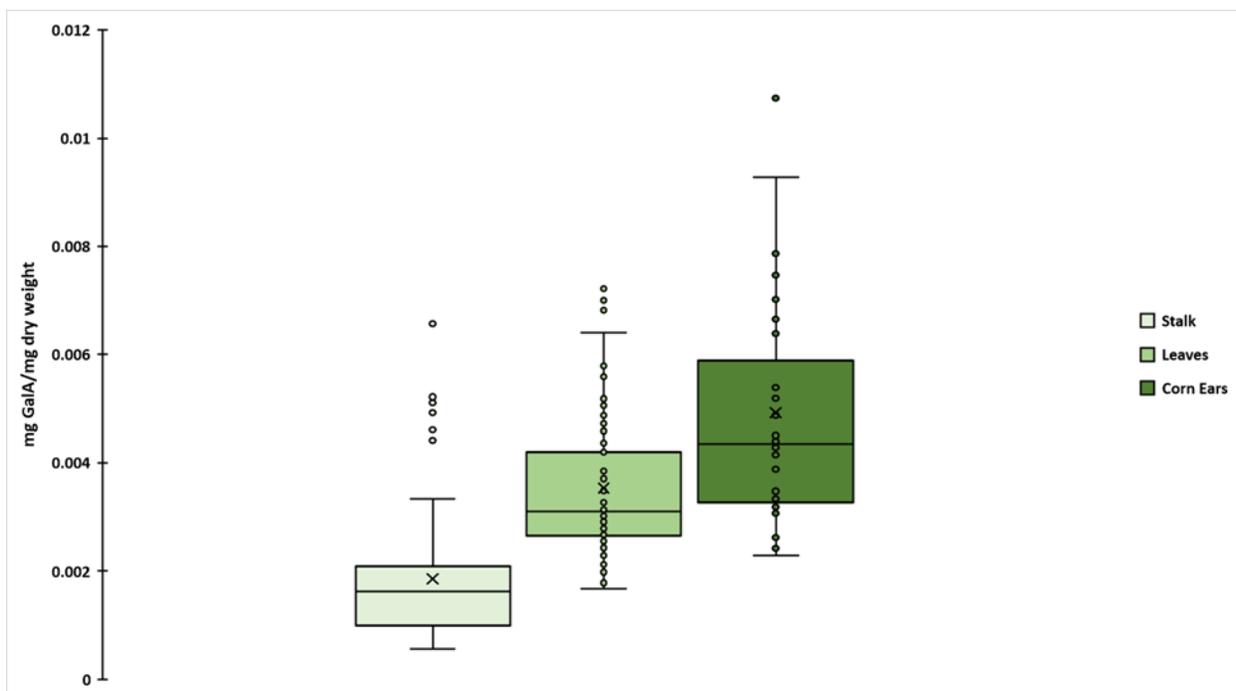


Supplementary Figure 3.1. Extracted ion chromatograms (EICs) of quantifier ion transitions in monosaccharide pool. All 14 monosaccharides monitored are nearly baseline separated within a 2.2 min isocratic elution using UHPLC-QqQ-MS operated in dMRM mode.

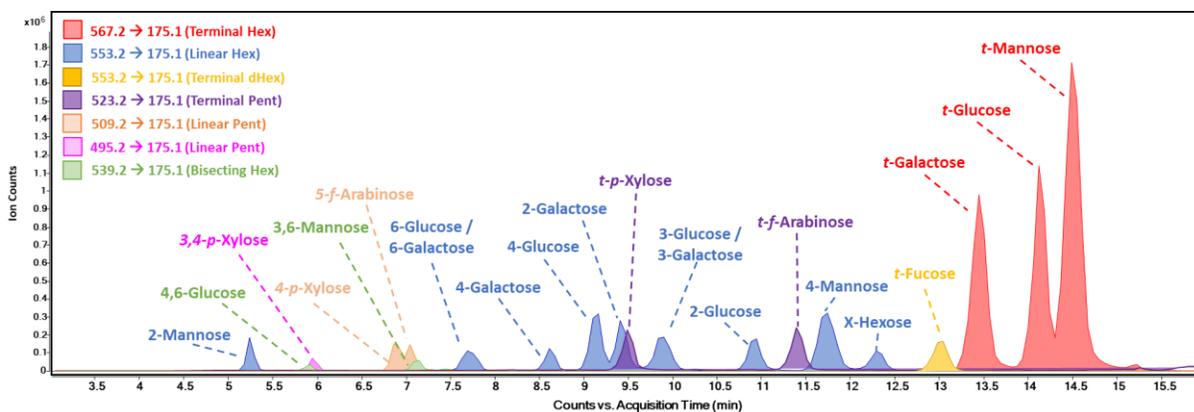


Supplementary Figure 3.2. The GalA content of the stalk pith was significantly greater than

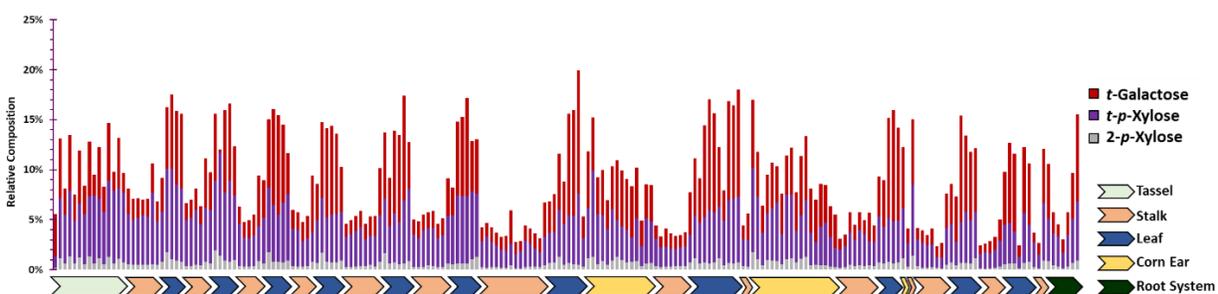
the rind. (Student's t-test, $p < 5e^{-6}$). Each point represents an inner or outer stalk tissue sample taken from each of 14 internodes. The GalA content was averaged across all inner stalk samples and all outer stalk samples.



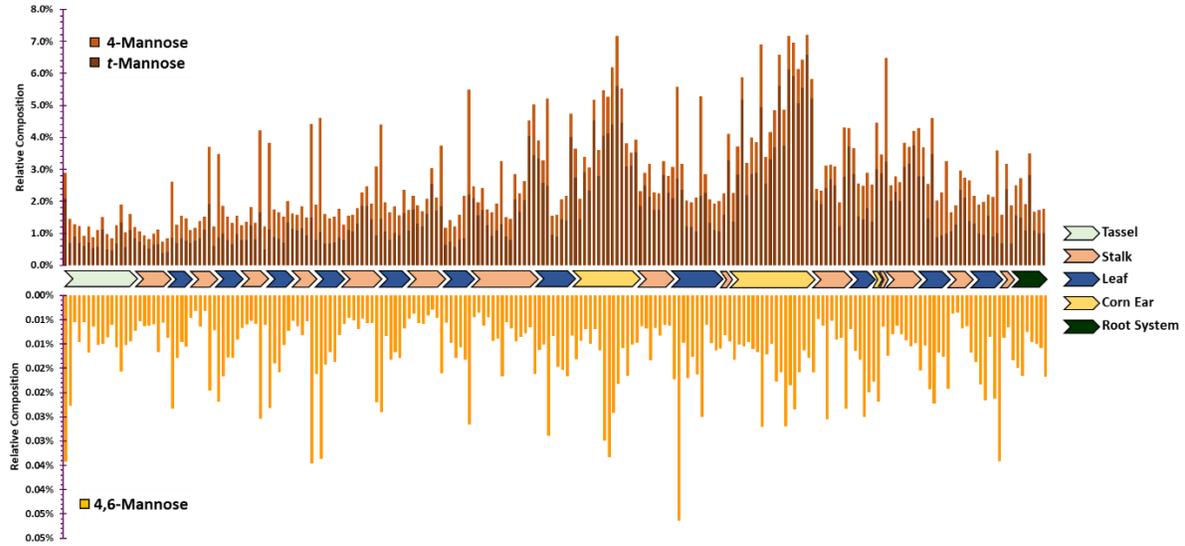
Supplementary Figure 3.3. The GalA content differed in tissues of the stalk, leaf, and corn ears. The collective tissues of the stalk were found to contain significantly less GalA than the leaves and the leaves less than the tissues of the corn ears (one-way ANOVA with Tukey's HSD, p -values < 0.0001).



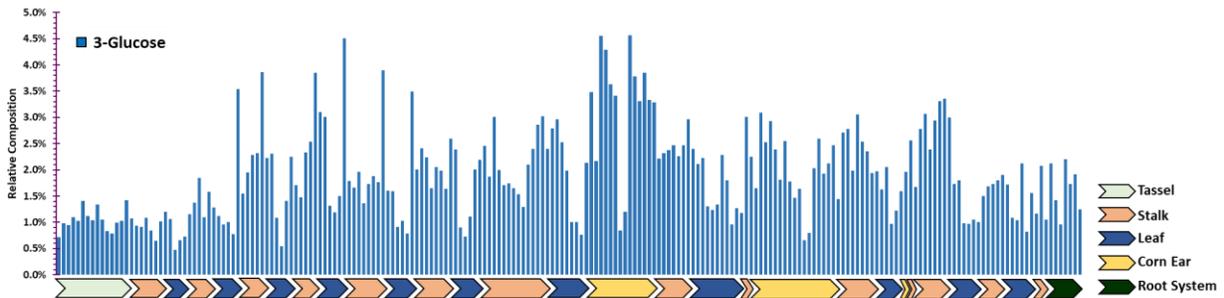
Supplementary Figure 3.4. Extracted ion chromatograms (EICs) of quantifier ion transitions in glycosidic linkage pool. Most compounds monitored are baseline separated within a 16 min gradient elution using UHPLC-QqQ-MS operated in MRM mode. The most abundant linkages are labeled.



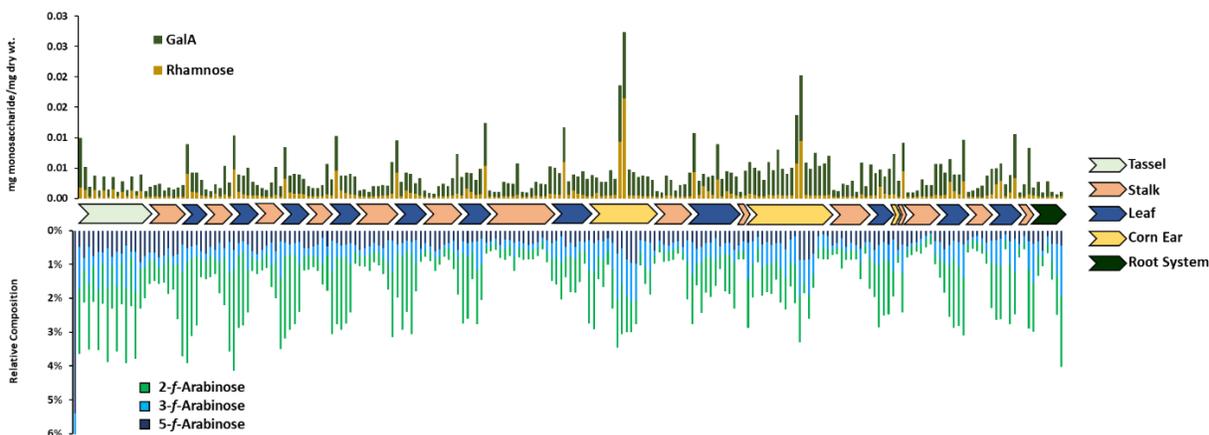
Supplementary Figure 3.5. The abundance of *t*-galactose, *t*-*p*-xylose, and 2-*p*-xylose indicated that xyloglucan was the next most abundant polysaccharide in the maize plant after xylans. In general, the tassels, leaves, and corn ears tended to have more xyloglucan-associated linkages than the stalk. In addition, the lower regions of the stalk displayed less than the upper regions.



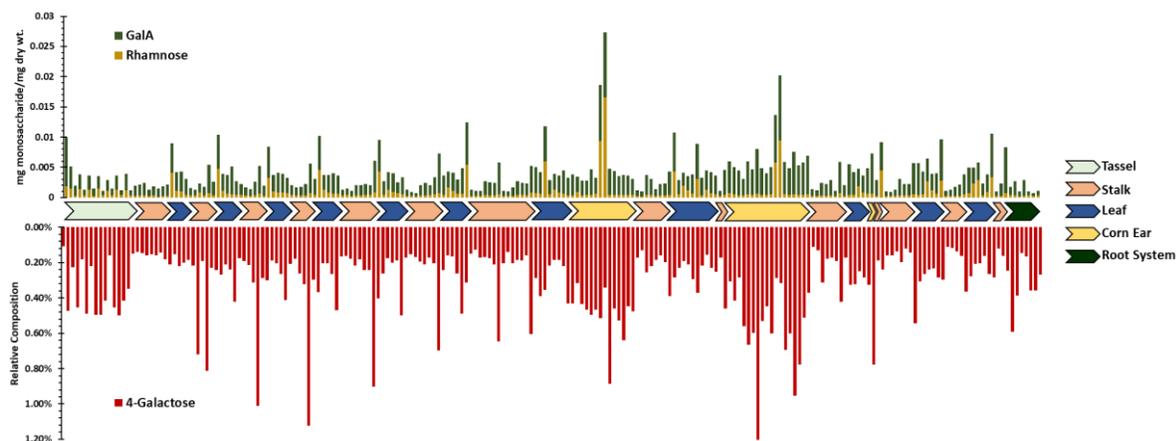
Supplementary Figure 3.6. The mannan-associated linkages 4-, 4,6-, and *t*-mannose were observed in low abundance throughout the plant. The stalk nodes, leaf collars, and corn ear tissues displayed larger relative abundances of these linkages. The lower regions of the stalk also contained more than the upper regions.



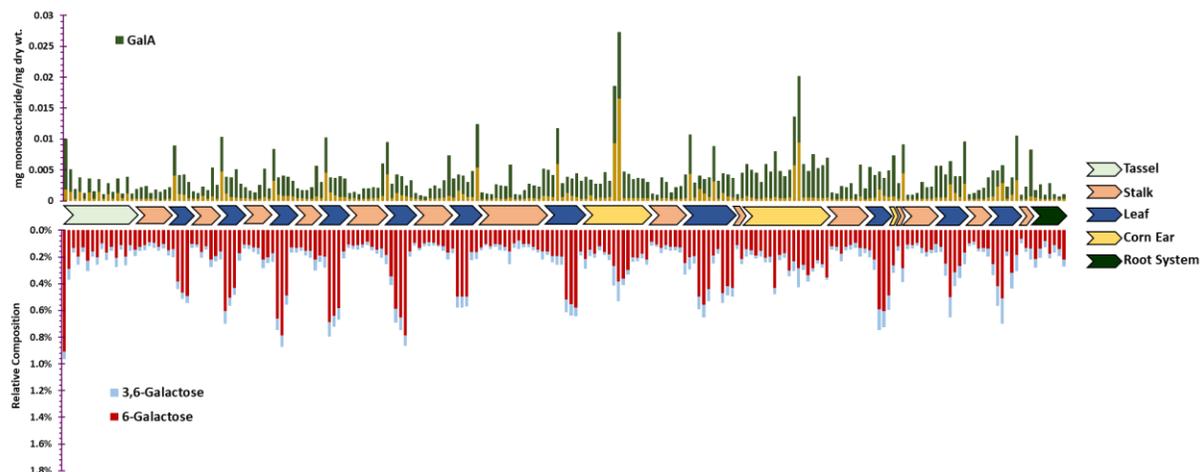
Supplementary Figure 3.7. The observation of 3-Glucose indicated that β -glucan was also a cell wall component of all maize tissues. This linkage tended to be highest in the stalk, particularly the stalk nodes and sheaths. The corn husks were also found to contain higher abundances while the leaf blades consistently contained the least.



Supplementary Figure 3.8. Increases in the pectic monosaccharide residues GalA and Rha corresponded to increases in linkages associated with branched arabinan. Leaf tissues, tassel flowers, corn husks, and the propagating roots were found to have increased arabinan linkages relative to the stalk tissues.



Supplementary Figure 3.9. The presence of 4-Galactose indicated galactan was a minor pectin component. The 4-galactose linkage was found to be largely increased in the stalk nodes and sheaths as well as tissues of the corn ears.



Supplementary Figure 3.10. The arabinogalactan-associated linkages 3,6-Galactose and 6-Galactose were also found in minor amounts throughout the plant. These were found to be particularly abundant in leaf tissues and pollen suggesting increased arabinogalactan content.

Supplementary Table 3.1. Monosaccharide composition and coefficients of variation (CV) of technical replicates. All values expressed in mg monosaccharide/mg dry tissue weight.

Sample Name	Glc	Gal	Fru	Xyl	Ara	Fuc	Rha	GlcA	GalA
V7 Leaf2 1	0.026	0.009	0.003	0.087	0.015	0.002	0.001	0.00039	0.003
V7 Leaf2 2	0.031	0.010	0.007	0.097	0.015	0.001	0.001	0.00026	0.003
V7 Leaf2 3	0.032	0.010	0.010	0.097	0.015	0.001	0.001	0.00026	0.003
Coefficient of Variation (CV, %)	10.6%	5.5%	52.3%	6.3%	2.5%	68.1%	1.6%	23.8%	5.8%

Supplementary Table 3.2. Glycosidic linkage composition and coefficients of variation (CV) of technical replicates. All values expressed as percent (%) relative dMRM chromatogram peak areas observed for each linkage. Only linkages with abundances $\geq 1\%$ are shown.

Sample Name	t-Glc	4-Glc	3-Glc/ 3-Gal	t-Gal	t-Xyl	4-Xyl	3,4-Xyl	t-f-Ara	t-Man
V8 Pith1	26.3	24.8	1.4	1.7	3.9	21.9	3.1	10.7	1.3
V8 Pith1	29.7	21.6	1.0	1.5	3.0	24.0	2.4	9.2	1.3
V8 Pith1	26.3	16.0	1.3	1.7	2.8	18.1	1.8	7.7	1.1
Coefficient of Variation (CV)	7.1	21.6	15.9	7.8	17.2	14.0	27.7	16.7	8.8

Supplementary Table 3.3. Free mono- and disaccharide content and coefficients of variation (CV) of technical replicates. All values expressed in ug saccharide/mg dry tissue weight.

Sample Name	Glc	Fru	Suc	Mal
Corn Flakes Cereal 1	1.85	1.9	3.34	1.19
Corn Flakes Cereal 2	1.85	2.16	2.92	1.15
Corn Flakes Cereal 3	1.87	2.07	4.45	1.12
Coefficient of Variation (CV, %)	0.6%	6.6%	22.2%	2.9%

REFERENCES

1. *World Population Prospects 2019: Highlights*; United Nations, Department of Economic and Social Affairs, Population Division: 2019.
2. Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D., Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* **2007**, *315* (5813), 804-7.
3. Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J., Jr.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T., The path forward for biofuels and biomaterials. *Science* **2006**, *311* (5760), 484-9.
4. Loque, D.; Scheller, H. V.; Pauly, M., Engineering of plant cell walls for enhanced biofuel production. *Curr Opin Plant Biol* **2015**, *25*, 151-61.
5. Slewinski, T. L., Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. *J Exp Bot* **2012**, *63* (13), 4647-70.
6. Khan, M. Z.; Zaidi, S. S.; Amin, I.; Mansoor, S., A CRISPR Way for Fast-Forward Crop Domestication. *Trends Plant Sci* **2019**, *24* (4), 293-296.
7. *Grain: World Markets and Trade*; United States Department of Agriculture: 2021.
8. Yang, C. J.; Samayoa, L. F.; Bradbury, P. J.; Olukolu, B. A.; Xue, W.; York, A. M.; Tuholski, M. R.; Wang, W.; Daskalska, L. L.; Neumeyer, M. A.; Sanchez-Gonzalez, J. J.; Romay, M. C.; Glaubitz, J. C.; Sun, Q.; Buckler, E. S.; Holland, J. B.; Doebley, J. F., The genetic architecture of teosinte catalyzed and constrained maize domestication. *Proc Natl Acad Sci U S A* **2019**, *116* (12), 5643-5652.
9. Michael McConnell, O. L., David Olson, Tom Capehart *Feed Outlook*; USDA ERS: 2021.
10. Santoro, N.; Cantu, S. L.; Tornqvist, C. E.; Falbel, T. G.; Bolivar, J. L.; Patterson, S. E.; Pauly, M.; Walton, J. D., A High-Throughput Platform for Screening Milligram Quantities of Plant Biomass for Lignocellulose Digestibility. *Bioenerg Res* **2010**, *3* (1), 93-102.
11. Penning, B. W.; Shiga, T. M.; Klimek, J. F.; SanMiguel, P. J.; Shreve, J.; Thimmapuram, J.; Sykes, R. W.; Davis, M. F.; McCann, M. C.; Carpita, N. C., Expression profiles of cell-wall related genes vary broadly between two common maize inbreds during stem development. *Bmc Genomics* **2019**, *20* (1).
12. Hatfield, R. D., Structural Polysaccharides in Forages and Their Degradability. *Agron J* **1989**, *81* (1), 39-46.
13. Martens, E. C.; Kelly, A. G.; Tauzin, A. S.; Brumer, H., The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *J Mol Biol* **2014**, *426* (23), 3851-3865.
14. Sonnenburg, E. D.; Sonnenburg, J. L., Starving our Microbial Self: The Deleterious Consequences of a Diet Deficient in Microbiota-Accessible Carbohydrates. *Cell Metab* **2014**, *20* (5), 779-786.
15. Flint, H. J.; Scott, K. P.; Duncan, S. H.; Louis, P.; Forano, E., Microbial degradation of complex carbohydrates in the gut. *Gut Microbes* **2012**, *3* (4), 289-306.
16. Flint, H. J.; Bayer, E. A.; Rincon, M. T.; Lamed, R.; White, B. A., Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. *Nat Rev Microbiol* **2008**, *6* (2), 121-31.

17. Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A., Determining the polysaccharide composition of plant cell walls. *Nat Protoc* **2012**, *7* (9), 1590-1607.
18. Rongpipi, S.; Ye, D.; Gomez, E. D.; Gomez, E. W., Progress and Opportunities in the Characterization of Cellulose - An Important Regulator of Cell Wall Growth and Mechanics. *Front Plant Sci* **2018**, *9*, 1894.
19. Lori J. Abendroth, R. W. E., Matthew J. Boyer, Stephanie K. Marlay, *Corn Growth and Development*. 2011.
20. Xu, G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis - quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2017**, *143* (1), 200-207.
21. Galermo, A. G.; Nandita, E.; Barboza, M.; Amicucci, M. J.; Vo, T. T.; Lebrilla, C. B., Liquid Chromatography-Tandem Mass Spectrometry Approach for Determining Glycosidic Linkages. *Anal Chem* **2018**, *90* (21), 13073-13080.
22. Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B., Development of an Extensive Linkage Library for Characterization of Carbohydrates. *Anal Chem* **2019**, *91* (20), 13022-13031.
23. Amicucci, M. J.; Galermo, A. G.; Guerrero, A.; Treves, G.; Nandita, E.; Kailemia, M. J.; Higdon, S. M.; Pozzo, T.; Labavitch, J. M.; Bennett, A. B.; Lebrilla, C. B., Strategy for Structural Elucidation of Polysaccharides: Elucidation of a Maize Mucilage that Harbors Diazotrophic Bacteria. *Anal Chem* **2019**, *91* (11), 7254-7265.
24. Carpita, N. C., Structure and Biogenesis of the Cell Walls of Grasses. *Annu Rev Plant Physiol Plant Mol Biol* **1996**, *47*, 445-476.
25. Mourtzinis, S.; Cantrell, K. B.; Arriaga, F. J.; Balkcom, K. S.; Novak, J. M.; Frederick, J. R.; Karlen, D. L., Distribution of Structural Carbohydrates in Corn Plants Across the Southeastern USA. *Bioenerg Res* **2014**, *7* (2), 551-558.
26. Burton, R. A.; Fincher, G. B., Current challenges in cell wall biology in the cereals and grasses. *Front Plant Sci* **2012**, *3*.
27. Park, Y. B.; Cosgrove, D. J., Xyloglucan and its Interactions with Other Components of the Growing Cell Wall. *Plant Cell Physiol* **2015**, *56* (2), 180-194.
28. Thomas, J. R.; Darvill, A. G.; Albersheim, P., Structure of Plant-Cell Walls .25. Rhamnogalacturonan-I, a Pectic Polysaccharide That Is a Component of Monocot Cell-Walls. *Carbohydr Res* **1989**, *185* (2), 279-305.
29. Zykwincka, A.; Thibault, J. F.; Ralet, M. C., Organization of pectic arabinan and galactan side chains in association with cellulose microfibrils in primary cell walls and related models envisaged. *Journal of Experimental Botany* **2007**, *58* (7), 1795-1802.
30. Galloway, A. F.; Pedersen, M. J.; Merry, B.; Marcus, S. E.; Blacker, J.; Benning, L. G.; Field, K. J.; Knox, J. P., Xyloglucan is released by plants and promotes soil particle aggregation. *New Phytol* **2018**, *217* (3), 1128-1136.
31. Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B.; Barve, R. A.; Lombard, V.; Henrissat, B.; Leyn, S. A.; Rodionov, D. A.; Osterman, A. L.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Kirbach, K.; Wilmot, T.; Heath, A. C.; Klein, S.; Barratt, M. J.; Gordon, J. I., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021**, *595* (7865), 91-95.
32. Zhang, Y.; Pribil, M.; Palmgren, M.; Gao, C. X., A CRISPR way for accelerating improvement of food crops. *Nat Food* **2020**, *1* (4), 200-205.

33. Zhang, Q.; Cheetamun, R.; Dhugga, K. S.; Rafalski, J. A.; Tingey, S. V.; Shirley, N. J.; Taylor, J.; Hayes, K.; Beatty, M.; Bacic, A.; Burton, R. A.; Fincher, G. B., Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. *BMC Plant Biol* **2014**, *14*, 27.
34. Arnaud, B.; Durand, S.; Fanuel, M.; Guillon, F.; Mechin, V.; Rogniaux, H., Imaging Study by Mass Spectrometry of the Spatial Variation of Cellulose and Hemicellulose Structures in Corn Stalks. *J Agric Food Chem* **2020**, *68* (13), 4042-4050.
35. Abedon, B. G.; Hatfield, R. D.; Tracy, W. F., Cell wall composition in juvenile and adult leaves of maize (*Zea mays* L.). *J Agric Food Chem* **2006**, *54* (11), 3896-900.
36. Cosgrove, D. J., Growth of the plant cell wall. *Nat Rev Mol Cell Biol* **2005**, *6* (11), 850-61.
37. Carpita, N. C.; McCann, M. C., The functions of cell wall polysaccharides in composition and architecture revealed through mutations. *Plant Soil* **2002**, *247* (1), 71-80.
38. Costa, T. H. F.; Vega-Sanchez, M. E.; Milagres, A. M. F.; Scheller, H. V.; Ferraz, A., Tissue-specific distribution of hemicelluloses in six different sugarcane hybrids as related to cell wall recalcitrance. *Biotechnol Biofuels* **2016**, *9*.
39. Faust, S.; Kaiser, K.; Wiedner, K.; Glaser, B.; Joergensen, R. G., Comparison of different methods for determining lignin concentration and quality in herbaceous and woody plant residues. *Plant Soil* **2018**, *433* (1-2), 7-18.
40. Zhao, F.; Chen, W.; Sechet, J.; Martin, M.; Bovio, S.; Lionnet, C.; Long, Y.; Battu, V.; Mouille, G.; Moneger, F.; Traas, J., Xyloglucans and Microtubules Synergistically Maintain Meristem Geometry and Phyllotaxis. *Plant Physiol* **2019**, *181* (3), 1191-1206.
41. Hsieh, Y. S. Y.; Harris, P. J., Xyloglucans of Monocotyledons Have Diverse Structures. *Mol Plant* **2009**, *2* (5), 943-965.
42. Hensel, A., [Xyloglucan--structure, genesis, and functions of a widely distributed substance group]. *Pharm Unserer Zeit* **1993**, *22* (4), 228-34.
43. Kozlova, L. V.; Nazipova, A. R.; Gorshkov, O. V.; Petrova, A. A.; Gorshkova, T. A., Elongating maize root: zone-specific combinations of polysaccharides from type I and type II primary cell walls. *Sci Rep* **2020**, *10* (1), 10956.
44. Roulin, S.; Feller, U., Reversible accumulation of (1-->3,1-->4)-beta-glucan endohydrolase in wheat leaves under sugar depletion. *J Exp Bot* **2001**, *52* (365), 2323-32.
45. Roulin, S.; Buchala, A. J.; Fincher, G. B., Induction of (1-->3,1-->4)-beta-D-glucan hydrolases in leaves of dark-incubated barley seedlings. *Planta* **2002**, *215* (1), 51-9.
46. Vega-Sanchez, M. E.; Verhertbruggen, Y.; Scheller, H. V.; Ronald, P. C., Abundance of mixed linkage glucan in mature tissues and secondary cell walls of grasses. *Plant Signal Behav* **2013**, *8* (2), e23143.
47. Willats, W. G.; McCartney, L.; Mackie, W.; Knox, J. P., Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol* **2001**, *47* (1-2), 9-27.

Chapter 4

The Development of the Davis Food Glycopedia—

A Glycan Encyclopedia of Food

ABSTRACT

The molecular complexity of the carbohydrates consumed by humans has been deceptively oversimplified due to a lack of analytical methods that possess the throughput, sensitivity, and resolution required to provide quantitative structural information. However, such information is becoming an integral part of understanding how specific glycan structures impact health through their interaction with the gut microbiome and host physiology. This work presents a detailed catalogue of the glycans present in complementary foods commonly consumed by toddlers during weaning and foods commonly consumed by American adults. The monosaccharide compositions of over 800 foods from diverse food groups including fruits, vegetables, grain products, beans/peas/legumes/nuts/seeds, sweets and beverages, animal products, and more were obtained and used to construct the “Davis Food Glyclopedia” (DFG), an open-access database that provides quantitative structural information on the carbohydrates in food. While many foods within the same group possessed similar compositions, hierarchical clustering analysis revealed similarities between different groups as well. Such a glyclopedia can be used to formulate diets rich in specific monosaccharide residues to provide a more targeted modulation of the gut microbiome, thereby opening the door for a new class of prophylactic or therapeutic diets.

INTRODUCTION

Carbohydrates make up the largest component of human diets, comprising up to 85% depending on geographic location and socioeconomic status.¹ These biomolecules play a profound role in shaping our gut microbial communities, the spectrum of microbial metabolites produced, and the resulting impacts on our health. While the human genome contains just 17 enzymes for the saccharolytic dissection of dietary carbohydrates, the microbial genomes of the gastrointestinal tract include thousands of such enzymes,² underscoring the importance of elucidating the structure of dietary carbohydrates in order to understand the specific relationship between dietary carbohydrates and the gut microbiome. For example, in early life, human milk oligosaccharides (HMOs) play a large role in feeding select *Bifidobacterium species*, thereby shaping the infant's microbial communities and providing health benefits such as priming the immune system, strengthening the gut barrier, and blocking pathogens.³ In adults, a high fat/high carbohydrate or "Western" diet has long been implicated in a variety of metabolic diseases such as cardiovascular diseases, type 2 diabetes, obesity, and gastrointestinal disorders.^{4,5} On the other hand, consumption of plant-based foods is associated with reducing the risks of those metabolic diseases.⁶

Recent research has emphasized the importance of the gut microbiome in this nutrition-health paradigm. Specifically, dietary carbohydrates modulate human health through their interaction with the gut microbiome.^{7,8} Despite their importance and being one of the most abundant components in foods, their structures, abundances, and functions are still poorly characterized due to a general lack of appropriate analytical methods.⁹ While the analysis of proteins and lipids have advanced greatly, the analysis of carbohydrates has been hindered by the inherent complexity of carbohydrate structures. Current groupings categorize carbohydrates into

the broad classifications of sugars, starch, and soluble/insoluble fiber, terms which provide little information on specific chemical or structural content therein. Indeed, the common term “fiber” offers no monosaccharide or structural specificity yet is regularly employed to represent various heterologous polysaccharides composed of differing sugar and linkage assemblages. Perhaps most revealing, total carbohydrates in most foods are currently measured indirectly by gravimetric mass difference of other macronutrients and micronutrients thereby depicting a form of nutritional “dark matter.”¹⁰ Such lack of chemical resolution impedes efforts to resolve the relationships between carbohydrates, the gut microbiome, and host health. There is thus a need for rapid throughput methods that are capable of characterizing carbohydrate structures and their microbiome interactions in large feeding studies.¹¹

Food carbohydrates are comprised of a diverse set of molecules ranging from free monosaccharides, disaccharides, oligosaccharides, and large polysaccharides. Additionally, each monosaccharide residue connects to another through numerous linkages (as many as 10 for each glycosidic linkage). Methods for oligosaccharide analysis using liquid chromatography-tandem mass spectrometry (LC-MS/MS) have been developed for structural elucidation, however the analyses remain very difficult and require a number of separation steps and structural elucidation techniques.⁹ Furthermore, even the most fundamental information, the monosaccharide composition is not known in most foods. The lack of this basic structural information inhibits our understanding of the role of the most abundant material in our diet. It prevents effective design of important clinical trials that could elucidate the specific roles of specific carbohydrate structures in food.

In this work, a recently developed workflow utilizing a rapid-throughput ultra-high performance liquid chromatography-triple quadrupole mass spectrometry (UPLC-QqQ

MS) method was employed to determine the monosaccharide compositions of over 800 food samples. Foods from diverse groups such as fruits, vegetables, fats, grains, dairy, beverages, and processed foods were subjected to monosaccharide analysis, and the resulting monosaccharide compositions were used to create a foundational collection of compositions in a resource here named the Davis Food Glycopedia (DFG). The method entailed the absolute quantitation of 14 naturally occurring monosaccharides separated on a five-minute UPLC-QqQ MS analysis in a 96-well plate format. The monosaccharide compositions of foods within and between food groups, as individual foods, and as part of a diet was revealed. This platform and the resulting glycopedia will allow us to formulate feeding trials where the diets may be highly enriched for specific monosaccharide compositions. Tailoring diets will enable future studies to better understand the role of food carbohydrates in shaping the gut microbiome in infants and adults. Furthermore, the presented findings will allow for dietary interventions that are more precisely formulated for modulating the gut microbiome and impacting human health.

MATERIALS AND METHODS

Selection of foods for inclusion in the DFG. Foods were initially selected for the DFG to design a feeding trial of arabinose-rich foods to selectively enrich beneficial gut microbiota in toddlers (12-36 months). The toddler foods selected included single foods recommended for toddlers according to the 2020-2025 Dietary Guidelines for Americans¹² which included a diverse group of vegetables with appreciable amounts of L-arabinose such as dark and green vegetables; red and orange vegetables; beans, peas, and lentils; in addition to lower arabinose-containing vegetables, fruits and starches. In addition to analyzing single foods, the DFG

includes food mixtures and snacks. Food selection for the DFG was then expanded greatly to cover foods that are commonly consumed by adults.

To determine foods commonly consumed by adults, three datasets were reviewed: (1) the Nutritional Phenotyping study (NutPheno),¹³ (2) What We Eat in America (WWEIA) 2017-2018,¹⁴ and (3) the Food and Nutrient Database for Dietary Studies Ingredients Database (FNDDS-Ing).¹⁵ The NutPheno study was a cross-sectional study that included healthy male and female adults, aged 18-66 years, living near Davis, CA. The NutPheno study included 393 adult subjects who reported dietary intake with up to 4 days of 24-hour recalls using the Automated Self-Administered 24-hour Dietary Assessment Tool (ASA24).¹⁶ WWEIA is the dietary component of the National Health and Nutrition Examination Survey (NHANES), a nationally-representative cross sectional study and consists of two 24-hour dietary recalls. Both the WWEIA dietary assessment and ASA24 use the Food and Nutrient Database for Dietary Studies (FNDDS), and the food descriptions and numeric identifiers (Food Code) come from FNDDS. The FNDDS-Ing database file describes the ingredients used to build FNDDS mixed foods (recipes), and therefore lists ingredients that might otherwise not be reported through WWEIA or ASA24 (in which subjects typically report final dishes instead of each individual ingredient in a food).

In the NutPheno study, a total of 2435 unique foods (corresponding to 2435 unique FNDDS Food Codes) were reported from a total of 1499 recalls. To identify candidate adult foods to add to the DFG from the NutPheno study, the frequency of each food reported in NutPheno was counted, and the 200 most frequently reported foods was manually cross-matched by searching the food description in the DFG for the closest match. Of the top 200 most frequently consumed foods, 135 did not have a matching DFG food. A second round of manual

curation was conducted on these 135 NutPheno foods to identify candidate foods to add to the DFG (e.g. would likely contribute to dietary glycan consumption and/or are typically consumed in large quantities or very frequently, n = 59).

The same process described above for NutPheno foods was used for FNDDS-Ing and WWEIA. A total of 2744 unique ingredients were identified in FNDDS-Ing. The frequency of an ingredient corresponds to the total number of times the ingredient is used in FNDDS recipes. Of the top 200 most frequently reported ingredients, 137 did not have matches to the DFG, 79 of which were considered as candidates to add. A total of 7083 foods were reported in WWEIA. Of the top 200 most frequently consumed foods, 130 had no DFG match, and 49 were considered as candidates to add.

Sources of materials. All foods and food products were purchased from local markets (Davis and Sacramento, CA) including Safeway, Trader Joe's, Davis Food Co-op, Whole Foods, Nugget Markets, Target, and online (Amazon). Trifluoroacetic acid (TFA, HPLC grade), 3-methyl-1-phenyl-2-pyrazoline-5-one (PMP), chloroform (HPLC grade), ammonium hydroxide solution (NH₄OH) (28-30%), ammonium acetate, sodium acetate, glacial acetic acid, methanol (HPLC grade), D-fructose, D-mannose, D-allose, D-glucose, D-galactose, L-rhamnose, L-fucose, D-ribose, D-xylose, L-arabinose, N-acetyl-D-glucosamine (GlcNAc), N-acetyl-D-galactosamine (GalNAc), D-glucuronic acid (GlcA), and D-galacturonic acid (GalA) were purchased from Sigma-Aldrich (St. Louis, MO). 96-well Nunc plates and lids were purchased from Thermo Scientific. Viscozyme was provided by Novozyme (Davis, CA). Acetonitrile (ACN) (HPLC grade) was purchased from Honeywell (Muskegon, MI). Nanopure water was used for all experiments.

Preparation of food samples. A total of 828 foods including fresh, frozen, commercial, and processed were purchased from local grocery stores in Davis, CA. Each food was documented with detailed descriptions prior to the sample preparation. For some raw and packed foods, samples were cooked, baked, or steamed as indicated on the package for cooking instructions. Foods were lyophilized to complete dryness and the moisture content was obtained. Samples underwent a dry bead blast or mortar and pestle for homogenization. A 10 mg aliquot of dried food sample was weighed into a 1.5-mL screw cap Eppendorf tube and reconstituted with water to make a stock solution of 10 mg/mL. The stock solution then underwent a bullet blending procedure followed by heat treatment (1 h at 100 °C) and another round of bullet blending prior to monosaccharide analysis.

Monosaccharide analysis of food samples. The monosaccharide analysis of foods was adapted from Xu et al.¹⁷ and Amicucci et al.¹⁸ with the following modifications. A 10 µL aliquot from the homogenized stock solution was subjected to incubation with Viscozyme treatment at 50 °C for 1 h in 390 µL of 25 mM acetate buffer (pH 5). A 100 µL aliquot from the enzyme digest was subjected to hard acid hydrolysis with 4 M TFA for 1 h at 121 °C and quenched with 855 µL of ice-cold water. A pool of monosaccharide standards consisting of D-fructose, D-mannose, D-allose, D-glucose, D-galactose, L-rhamnose, L-fucose, D-ribose, D-xylose, L-arabinose, GlcNAc, GalNAc, GlcA, and GalA were used to generate a calibration curve and were prepared in water ranging in concentration from 0.001 to 100 µg/mL. The released monosaccharides in samples and standards were then derivatized with 0.2 M PMP solution in methanol and 28% NH₄OH at 70 °C for 30 min. Samples were then dried to completeness by vacuum centrifugation. The excess PMP was removed by a chloroform extraction and a 1 µL aliquot of the derivatized monosaccharides were subjected to UPLC-QqQ MS analysis.

Mass spectrometry instrumental analysis. Derivatized glycosides were separated on an Agilent Poroshell HPH-C18 column (2.1 × 50 mm, 1.9 μm) and guard using an Agilent 1290 Infinity II UPLC system. A constant flow rate of 1.050 mL/min was employed on a 2 min isocratic elution at 12% solvent B followed by a 1.6 min flush at 99% solvent B and 0.79 min equilibration for a total run time of 4.6 min for the separation of compounds. Solvent A consisted of 25 mM ammonium acetate in 5% acetonitrile with pH adjusted to 8.2 using ammonia solution. Solvent B consisted of 95% acetonitrile in water. The separated glycosides were then detected on an Agilent 6495B triple-quadrupole mass spectrometer (QqQ MS) operated in positive ion mode using dynamic multiple reaction monitoring (dMRM).

Data analysis. Raw LC-MS files were analyzed using Agilent MassHunter Quantitative Analysis software (Version B 08.00). Chromatographic peaks were manually integrated and matched with standards. Monosaccharides were quantified by external calibration curve fitted with linear regression. Clustering analysis based on monosaccharide profiles were done with R using circlize library (v 0.4.13). Dendrograms and heatmaps were also generated using circlize. Enrichment of food groups in each cluster was determined using hypergeometric test and statistical significance was assigned based on FDR-adjusted p-values.

Assigning food groups to DFG foods. The DFG food groups are adapted from the FNDDS food groups that are defined by the first two digits of the FNDDS Food Code:¹⁹ (1) milk and milk products, (2) meat, poultry, fish, and mixtures, (3) eggs, (4) beans, peas, other legumes, nuts, and seeds, (5) grain products, (6) fruits, (7) vegetables, (8) fats, oils, and salad dressings, and (9) sugars, sweets and beverages (excluding juice and plant-based milks). Food groups were assigned based on a food's first ingredient; the second ingredient was used if water was the first ingredient. For example, both mango juice and fresh yellow mango are fruits, and orange

preserves (first ingredient is sugar) is in sugars, sweets, and beverages. When the ingredient labels for multi-ingredient foods could not be found online, the food group was assigned based on the product name and description.

RESULTS

We employed a recently developed LC-MS platform to quantitate the monosaccharide compositions of over 800 foods. The collection included whole and processed foods with emphasis on the earliest complementary (weaning) foods and common adult foods typical of diets among the US population. The resulting glycan compositions of the foods were used in a clustering analysis to identify food groups with common or similar monosaccharide characteristics. The DFG was further used to create an example meal enriched in specific monosaccharides to illustrate its utility in designing potential feeding trials that would probe host-microbe interactions.

Monosaccharide compositional analysis of foods

Foods purchased in local markets were documented with detailed descriptions and processed using preparation procedures that included cooking (where applicable), lyophilization, and homogenization. Moisture contents were determined with the lyophilization step. The samples were digested using pectinase enzyme and acid hydrolysis then labeled to enhance MS ionization and facilitate chromatographic separation in a five-minute UPLC-QqQ MS analysis. The absolute monosaccharide abundances were obtained using dynamic multiple reaction monitoring (dMRM) and standard monosaccharide solutions were analyzed to generate external calibration curves.

Foods are traditionally assigned to groups. In this study, the largest groups fall into those classified as fruits, vegetables, grain products, beans/peas/legumes/nuts/seeds (**Supplementary Figure 4.1**). Other groups including meat/poultry/fish and mixtures, eggs, milk/milk products, fats/oils/salad dressings, and sugars/sweets/beverages were also included but contain fewer entries. The 14 monosaccharides monitored included glucose, galactose, fructose, xylose, arabinose, fucose, rhamnose, mannose, GlcA, GalA, GlcNAc, GalNAc, allose, and ribose. All were found in measurable abundances except allose, GlcNAc, and GalNAc. Average abundances of monosaccharides were calculated for the food groups (**Figure 4.1a-j**). The most commonly found and often the most abundant monosaccharide was glucose likely from starch and/or sucrose. Other common and abundant monosaccharides particularly those from plant-based foods were fructose, xylose, arabinose, galactose, and GalA with all but fructose likely due to cell wall polysaccharides such as arabinoxylan and pectins.^{20, 21} Xylose and arabinose were most abundant in grains (**Figure 4.1e**) due to the presence of arabinoxylans in their cell walls, while GalA and rhamnose were common in beans, fruits, and vegetables (**Figure 4.1a, d, i**) likely due to the abundance of pectins.^{20, 21} Grain products had the highest overall measured carbohydrates by fresh weight (**Figure 4.1e**) due to high starch and low moisture content followed by beans, peas, and legumes, fruits, and vegetables (**Figure 4.1a**). Eggs and fats/oils/salad dressings (**Figure 4.1b and 1c**, respectively) were found to have the lowest carbohydrate content. The meat/poultry/fish/mixtures group contained significant amounts of glucose largely due to bread coatings of meats such as those found in breaded chicken and fish (**Figure 4.1f**). However, the analysis of unprocessed meat, poultry, and fish revealed very little glucose. Additionally, soups which are in this group also contain large amounts of glucose likely from the vegetable components of these products.

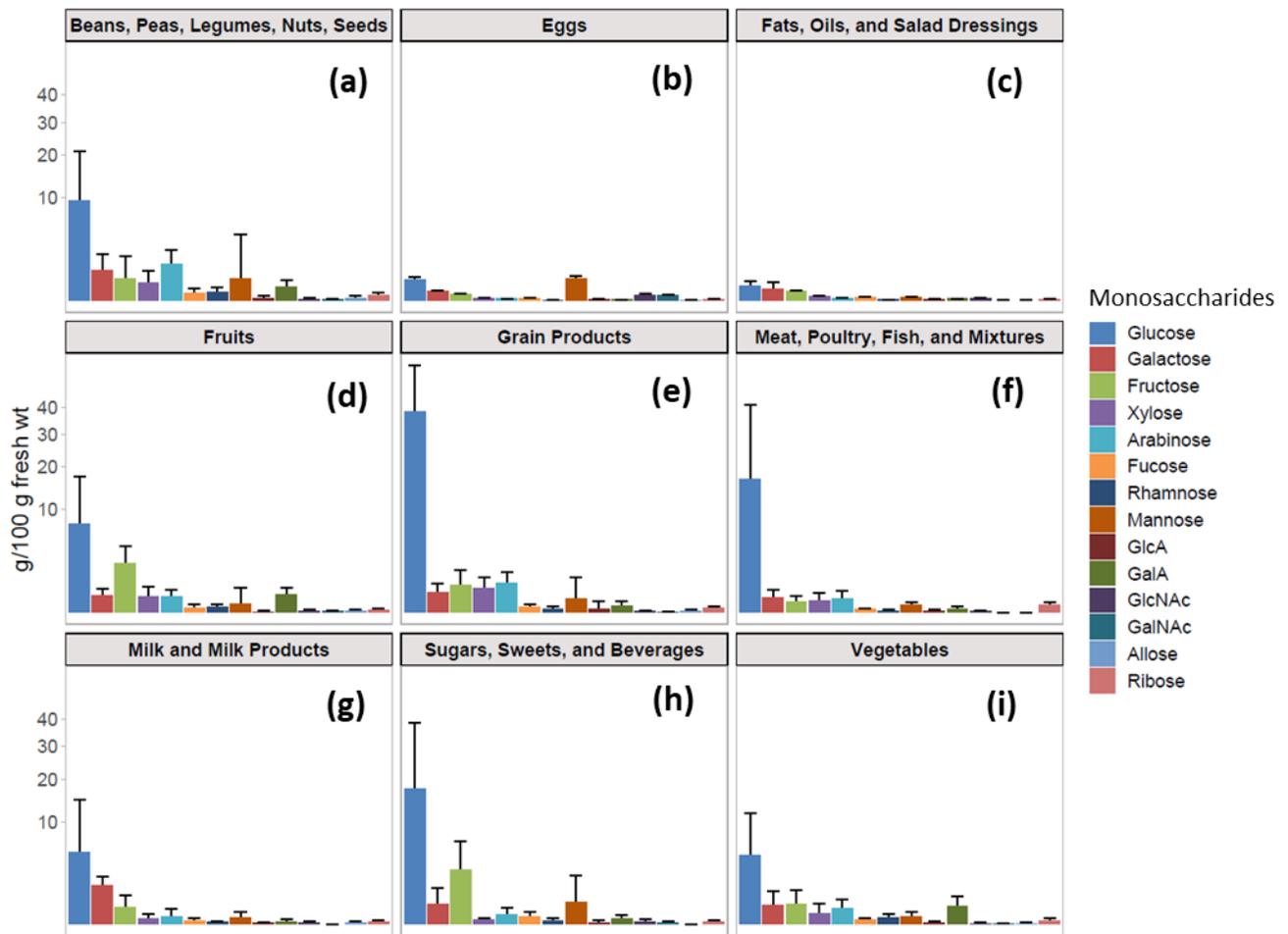


Figure 4.1a-j. Average monosaccharide compositions of all nine food groups. The y-axis follows a square root scale. Error bars represent the standard deviation.

The traditional method of grouping foods misrepresents the carbohydrate content. When foods from plant-based groups were analyzed, each entry had markedly different monosaccharide compositions. **Figure 4.2a-d** depicts the monosaccharide compositions of 20 representative foods from each plant-based food groups, which include fruits, grain products, vegetables, and beans/peas/legumes/nuts. Even the fruit group exhibited diverse monosaccharide compositions, although it tended to contain significantly more fructose than other groups, as

expected (**Figure 4.2a**). Grain products (**Figure 4.2b**) exhibited the highest glucose content from starch, but also contained xylose, arabinose, and galactose. In grains, “white” products like white bread, flour tortillas, and white rice tended to contain less non-glucose monosaccharides than their whole grain counterparts such as whole-grain bread, grains, and brown rice (**Figure 4.2b**, **Supplementary Figure 4.2**). Aside from potatoes and corn, which have high-starch contents, most vegetables (**Figure 4.2c**) had diverse monosaccharide compositions consisting of glucose, fructose, galactose, xylose, arabinose, GalA, and mannose and were more similar to fruits but with markedly less fructose. Beans, peas, legumes, and nuts (**Figure 4.2d**) contained relatively high amounts of arabinose. However, beans and peas had larger amounts of glucose than nuts due to a higher starch content.

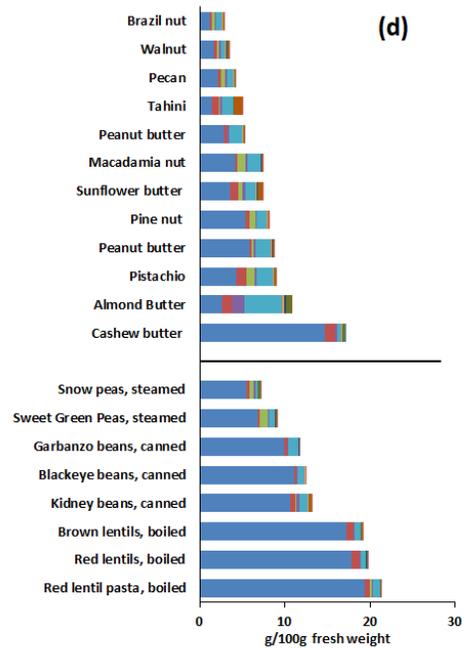
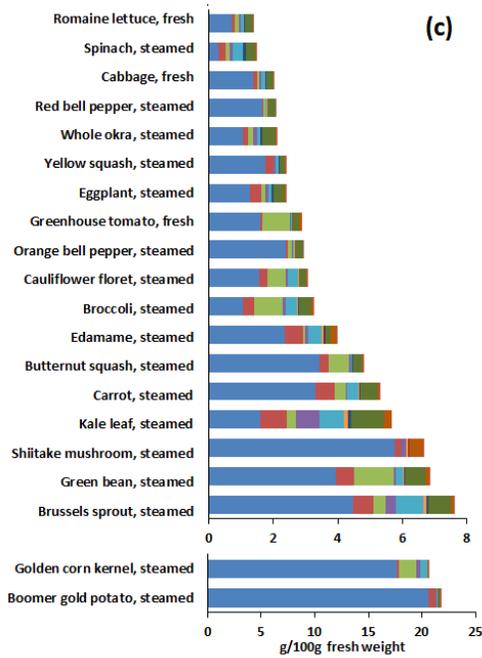
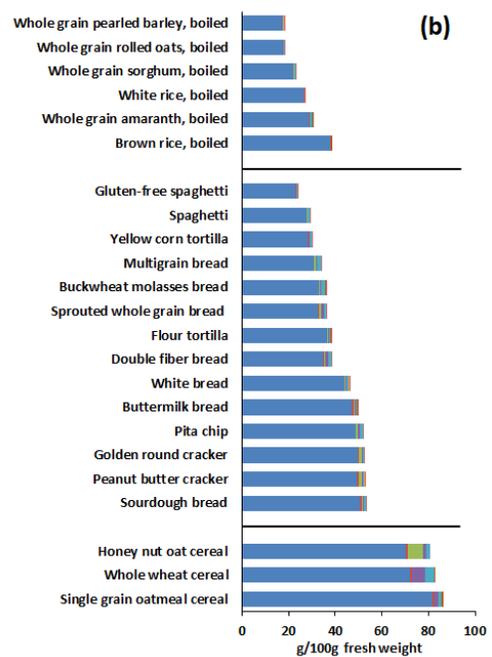
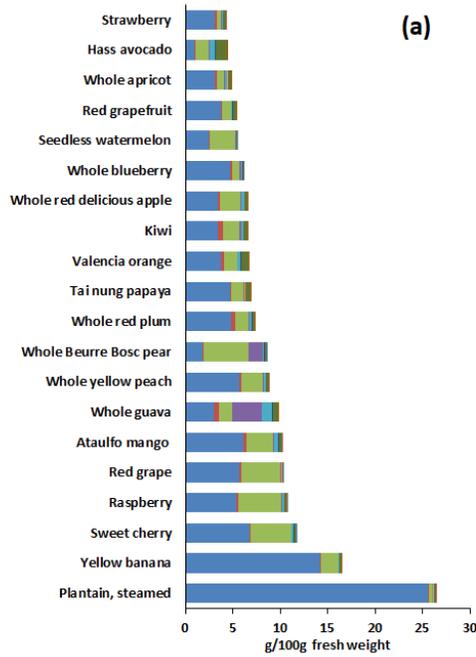


Figure 4.2a-d. Monosaccharide compositions of selected representative foods from each plant-based food group for (a) fruits, (b) grains products, (c) vegetables, and (d) beans, peas, legumes, and nuts.

Solanaceous foods (or nightshades) like tomatoes, eggplant, and bell peppers (**Figure 4.2c**) contained very little arabinose and non-glucose monosaccharides while members of the Brassicaceae family like brussels sprouts, broccoli, kale, and cauliflower (**Figure 4.2c**) yielded significantly larger quantities of arabinose and other monosaccharides such as GalA, galactose, rhamnose, and fucose. In nuts, almonds contained the largest amount of arabinose whereas tahini (made from sesame seeds) contained large amounts of mannose (**Figure 4.2d**). In fruits, pears and guava tended to contain more xylose than other fruits while berries were very low in non-glucose and non-fructose monosaccharides (**Figure 4.2a**).

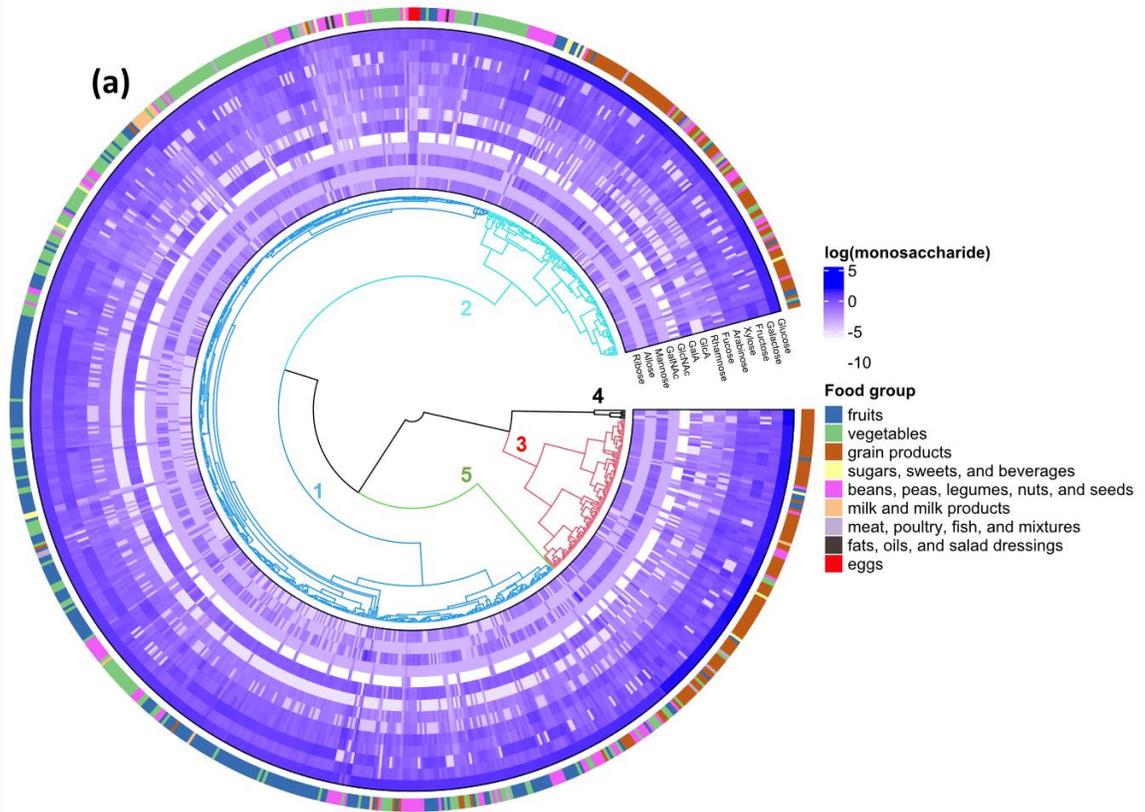
Clustering Analyses yields new combinations of foods

To visualize common features and differences among individual monosaccharide compositions irrespective of food group, an unsupervised hierarchical clustering analysis of the DFG was performed (**Figure 4.3a**). The average monosaccharide compositions of each cluster are depicted in **Figure 4.3b**. A total of 5 clusters were used to divide the 828 foods into clusters based on their total monosaccharide compositions (**Figure 4.3a**). The number of clusters was based on various clustering indices.²² Food belonging to the same groups (fruits, vegetables, and grain products) largely clustered together as defined by their monosaccharide compositions,

however these classifications were essentially dominated by the amount of glucose. Cluster enrichment factors were included in **Supplementary Figure 4.3**.

Glucose was the most abundant component in many of the samples, and hence was the major factor for the separation of the clusters. Cluster 1 was the largest, comprising of over half of the total foods surveyed. This cluster was significantly enriched in fruits, vegetables, and the beans, peas, legumes, nuts, and seeds group but also contained entries from all of the other food groups. We further separated Cluster 1 and obtained sub-clusters as shown in **Supplementary Figure 4.4**. Based on this sub-clustering analysis, specific groupings were observed such as Cluster 1A (intermediate glucose and fructose) with apples and toddler food products, Cluster 1B with stone fruits, tomatoes, berries, squash, fruit juices, and soups, and Cluster 1C (high glucose) with potatoes, bananas, and oats. Cluster 1F comprised of almond butters and flax seed with high arabinose and xylose values, while Clusters 1G (soy flour, roasted seaweed) and 1H (coffee grounds, dried coconut chips) are high in galactose and mannose, respectively. The average monosaccharide composition for Cluster 1 as a whole was most dissimilar to Clusters 2-5 (**Figure 4.3b**) and largely reflected the fruit and vegetable food groups. Specifically, Cluster 1 contained significantly lower amounts of glucose and a larger overall diversity than other clusters. Cluster 2 was significantly enriched in grain products, which were largely breads and cooked whole grains such as oats, barley, millet, quinoa, and rice. Additionally, Cluster 2 contained pastas, dried fruits, and plant-based meat products. Cluster 3 was significantly enriched in grain products, most of which were dried cereals and snacks. This cluster also contained foods from other groups such as fruits and vegetables, which were also mostly dried and snack products. Cluster 4 contained only six items, all of which were dried rice or corn products exhibiting the highest glucose and lowest non-glucose monosaccharides of all food

analyzed. Cluster 5 is a single-member group with coconut flour having the highest amount of mannose in the DFG.



(b)

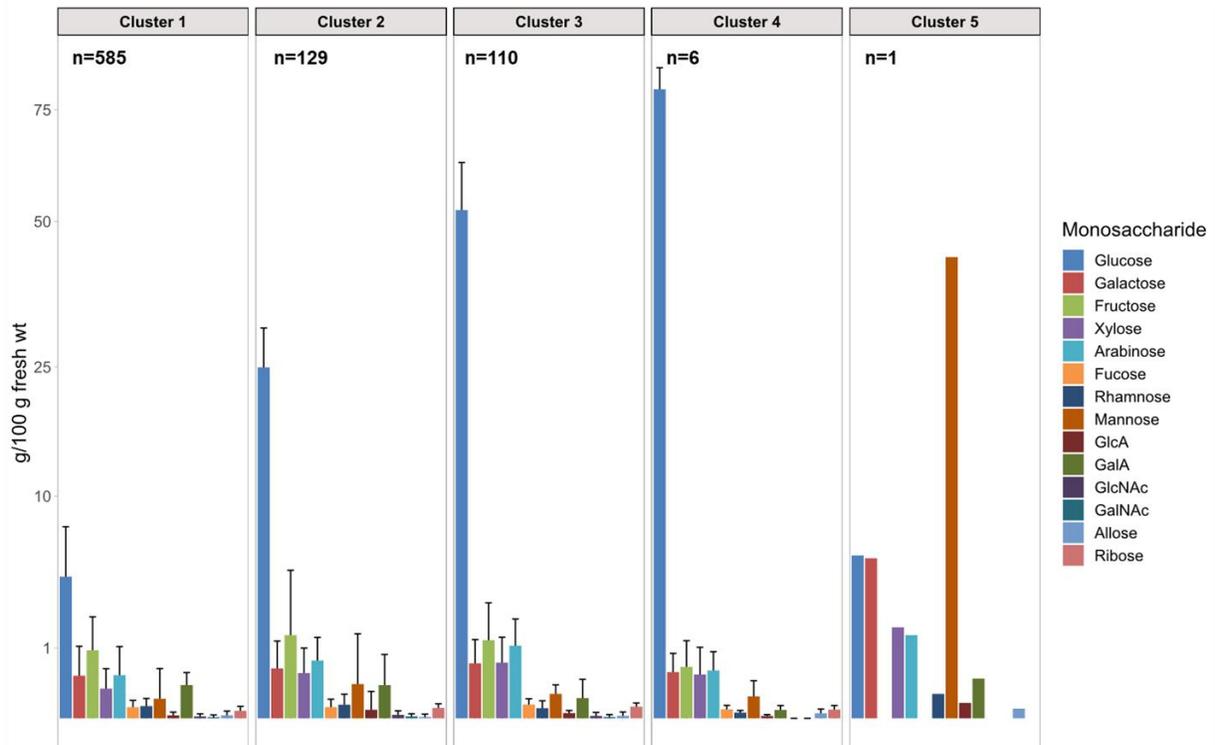


Figure 4.3. (a) Hierarchical cluster analysis of all 828 foods based on their absolute monosaccharide compositions, (b) Average monosaccharide composition of each cluster. The y-axis follows a square root scale. Error bars represent the standard deviation.

DISCUSSION

Current methods in dietary carbohydrate analysis are limited to quantifying sugars, starch, and fiber. Within the definition of fiber is an immense amount of structural complexity that can alter the gut microbiome and affect host health. Dietary recommendations emphasize the importance of consuming fiber. However, the term “fiber” makes no distinction of the monosaccharide composition nor the primary structure of the molecule. The reality is that food glycans are composed of a very large number of compounds, each with their unique structural variations and potentially specific activities both to the consuming host and their associated microbiome. Thus, the advice “eat more fiber,” is not meaningful because fiber from two different sources can have completely different monosaccharide compositions, glycosidic bond linkages, degree of polymerization, and in turn, biological functions. The analytical methods used to measure carbohydrates must be updated to match the evolving throughput and coverage of sequencing and metabolomic analyses. To address this need, we developed and utilized a rapid-throughput, LC-MS based monosaccharide analysis to determine the total monosaccharide composition of over 800 foods to create the Davis Food Glyclopedia (DFG) which will inform future feeding studies in infants transitioning to complementary diets, toddlers and adults. The total monosaccharide composition and quantitation provides more useful information on dietary carbohydrates than traditional gravimetric methods especially in the context of the gut

microbiome and infant nutrition. This comes with greatly increased sample throughput making the construction of large food glycan libraries possible.

The DFG revealed the most abundant monosaccharide in the foods was primarily glucose from simple sugars such as sucrose and from starch polysaccharides. From an evolutionary and agricultural perspective, humans have historically used innovative strategies to seek and cultivate sugar- and starch-dense foods and parts of foods as a source of energy as evidenced by the expansion of salivary amylase genes in humans.²³ While these energy-rich foods were once a necessity for survival, increasingly sedentary lifestyles and overconsumption of highly processed versions of these foods has contributed to a variety of metabolic disorders such as obesity, type 2 diabetes, and heart disease particularly in Western populations.⁵ The DFG provides information not only on digestible glucose content (i.e. starch), but also on non-glucose content corresponding to various dietary fiber structures. This information can be used to inform dietary choices to alleviate these metabolic disorders by reducing starch and sugar consumption and increasing the consumption of specific fiber types to shape the gut microbiome in a targeted manner. Clustering analysis of the DFG revealed that the food group does not necessarily inform a food's carbohydrate composition. For example, fruits and vegetables are two food groups that clustered together due to their similar monosaccharide compositions with higher average GalA and rhamnose, which reflected their pectin content. In contrast, grain products clustered away from fruits and vegetables due to their high glucose, xylose, and arabinose content which reflected their starch, β -glucan, and arabinoxylan polysaccharide constituents. Together, these results suggest that diets meant to target the gut microbiome should be informed by the carbohydrate composition rather than the food group alone.

Creation of diets based on monosaccharide compositions

Even in its current limited form, the DFG can be used to create fiber focused diets. The utility of this resource is that meals can be created with known amounts of carbohydrates based on monosaccharide compositions. According to the USDA Dietary Guidelines for Americans 2020-2025 and USDA MyPlate, it is recommended for adults to consume 2 cups of fruits, 2.5 cups of vegetables, 6 ounces of grains, 5.5 ounces of protein, and 3 cups of dairy in a day.¹² These recommendations are based on consuming 2,000 calories per day and have different food groups compared to the food groups described in this work. To generate a relative chart of each food group (**Supplementary Figure 4.5**), the recommended servings in an example meal were converted from cups and ounces to grams. The ingredients for the example dinner meal included 4 ounces of chicken breast, 0.5 cups of broccoli, 0.33 cups of carrots, 0.33 cups of summer squash, 0.75 cups of pasta, 1 tablespoon of oil, 1 cup of a navel orange, and 1 cup of milk. Based on the USDA Dietary Guidelines for Americans 2020-2025, the recommended food groups relative composition for the example meal yielded 30 % for vegetables, 26 % for fruits, 19 % for dairy, 13 % for grains, and 12 % for proteins.

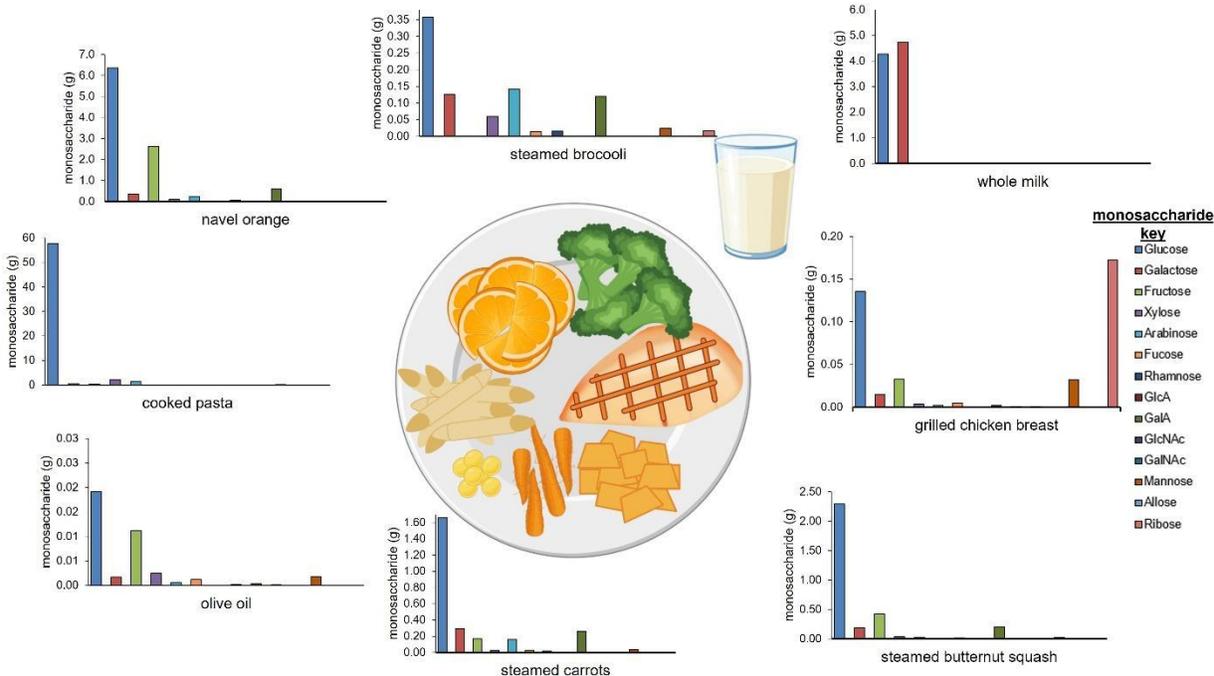


Figure 4.4. Example meal with quantitative monosaccharide bar graphs of each ingredient. The serving amounts are based on the USDA Dietary Guidelines for Americans 2020-2025.

The total dietary carbohydrate content in the example meal was determined using values from the DFG. Additionally, the monosaccharide concentrations and composition of each ingredient in the meal were determined (**Figure 4.4**). The calculated total carbohydrate content in the entire meal was 89.09 g (**Table 4.1**). The cooked penne pasta, navel orange, and glass of whole milk resulted in the highest total carbohydrate amounts (per ingredient and serving) with values of 62.4 g, 10.4 g, 9.1 g, respectively. As expected, olive oil, and grilled chicken breast had minimal carbohydrates (per ingredient and serving) with values 0.0 g, and 0.4 g, respectively. The cooked penne pasta had less relative monosaccharide diversity with glucose from starch as the most abundant. On the other hand, steamed broccoli, steamed carrots, navel orange and steamed butternut squash had the most (non-glucose) monosaccharide diversity with higher

amounts of galactose, fructose, xylose, arabinose, and galacturonic acid present. Whole milk contained glucose (4.27 g) and galactose (4.74 g) per cup which matches the known composition of lactose, the major disaccharide in milk.

Table 4.1. The absolute monosaccharide composition and amounts in an example dinner meal.

food	serving amount	amount (grams)	moisture (%)	monosaccharide (g)															
				Glc	Gal	Fruc	Xyl	Ara	Fuc	Rhm	GlcA	GalA	GlcNAc	GalNAc	Man	All	Rib	total	
grilled chicken breast	4 oz	113.4	62	0.14	0.01	0.03	0	0	0	0	0	0	0	0	0.03	0	0.17	0.4	
steamed broccoli	0.5 cups	38	89.6	0.36	0.13	0	0.06	0.14	0.01	0.02	0	0.12	0	0	0.02	0	0.02	0.9	
steamed carrots	0.33 cups	50	87.1	1.66	0.29	0.17	0.03	0.16	0.02	0.02	0	0.26	0	0	0.04	0	0	2.7	
steamed butternut squash	0.33 cups	66.7	89.6	2.29	0.19	0.42	0.03	0.03	0.01	0.01	0	0.2	0	0	0.02	0	0	3.2	
cooked pasta	0.75 cups	150	54.7	57.6	0.51	0.45	2.1	1.51	0	0.02	0.01	0.03	0	0	0.15	0	0.03	62.4	
olive oil	1 Tbsp	14.3	0.8	0.02	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	
navel orange	1 medium orange	165	87.3	6.37	0.37	2.64	0.12	0.24	0.04	0.06	0	0.6	0	0	0	0	0	10.4	
whole milk	1 cup of milk	245	89	4.27	4.74	0	0	0.01	0	0	0	0	0	0	0.03	0	0.01	9.1	
total	N/A	N/A	N/A	72.7	6.25	3.72	2.35	2.09	0.09	0.13	0.01	1.21	0	0	0.3	0	0.25	89.09	

In addition to determining the total carbohydrate content in each ingredient in a meal, the database was used to determine the total amount of each monosaccharide by adding the total monosaccharides from each ingredient. In the exemplified meal above, the glucose was the most abundant monosaccharide with a total of 72.70 g. The next most abundant monosaccharides were galactose and fructose with a total of 6.25 g and 3.72 g, respectively. Xylose (2.35 g) and arabinose (2.09 g) were similar in abundance, while fucose (0.09 g), rhamnose (0.13 g),

galacturonic acid (1.21 g), mannose (0.30 g), and ribose (0.25 g) were present in smaller amounts.

The DFG can be used not only to quantify and compare the monosaccharide compositions of different meals, but also to create personalized meals rich in specific monosaccharides and, by extension, fibers for altering and modulating the gut microbiome or other health endpoints. Arabinose is a prime example for this purpose as it is found commonly only in plants, is not digested or absorbed well endogenously in animal models, and has been shown to play an important role in shaping the gut microbiome.²⁴⁻²⁶ Arabinose is not abundant in foods as a free monomer, rather it is a part of ubiquitous cell wall polysaccharides such as arabinoxylan in grains and pectins in fruits and vegetables.^{20, 27} While this method does not differentiate from which polymer the arabinose originates, arabinose can nonetheless be quantitated to identify foods to maximize dietary levels of this monosaccharide. **Figure 4.5** provides the broad arabinose content of the individual food groups. The highest average arabinose content was observed in beans, peas, legumes, nuts, and seeds (1.24 g/100 g fresh weight) followed by grain products, vegetables, and fruits (0.8, 0.27, and 0.24 g/100 g fresh weight). In general, the highest arabinose concentrations were found in plant-based foods such as legumes, grains, vegetables, and fruits. However, the range of arabinose in each plant-based food group was large and dependent on the specific food and moisture content. For example, pear cultivars tended to have more arabinose than apple cultivars (**Supplementary Figure 4.6**). Relatively dry foods like cereals, nut butters, and dehydrated legume, vegetable, and fruits products consistently displayed the highest arabinose concentrations and total measured carbohydrate in each group. With the DFG resource, the monosaccharide profile of a meal can be altered by simply swapping an ingredient with another from the same food group with a higher

concentration of the desired monosaccharide. Continuing with arabinose in the example meal from **Figure 4.4**, the navel orange (0.15 g/100 g arabinose) and cooked pasta (1.51 g/100g arabinose) can be exchanged for a Bartlett pear (0.37 g/100 g arabinose) and sprouted wheat bread (2.11 g/100g arabinose), respectively, to significantly and selectively increase the arabinose content of the meal.

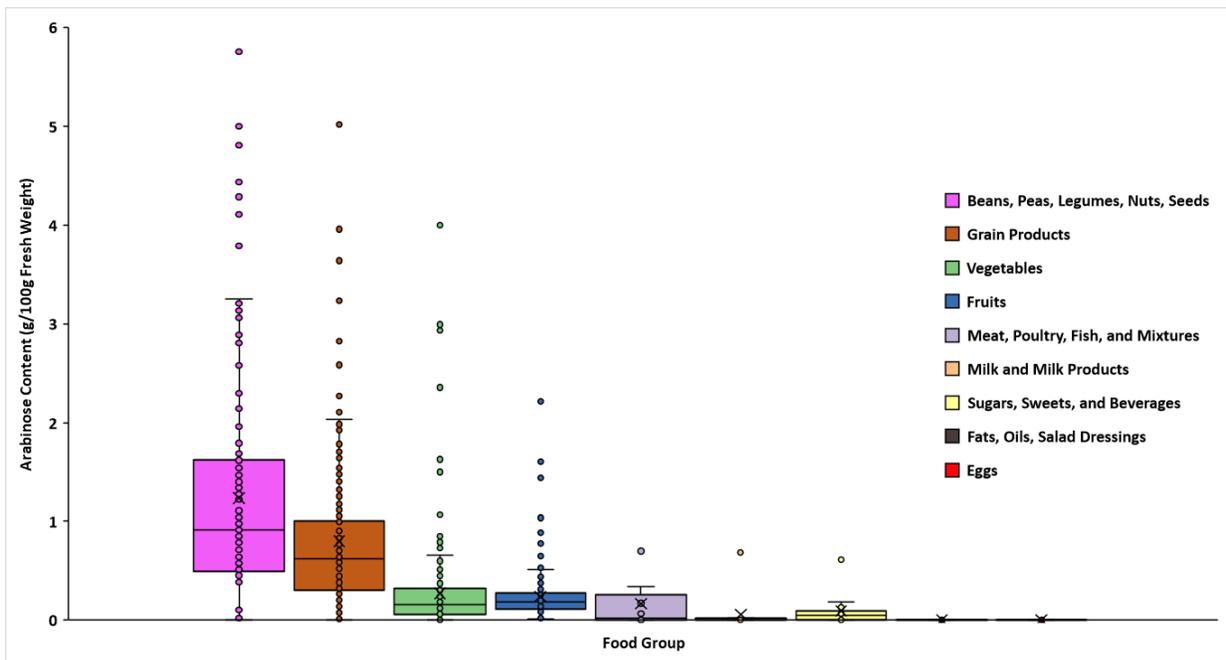


Figure 4.5. Average arabinose abundances in all food groups. Beans, peas, legumes, nuts, and seeds yielded the highest arabinose content while fats, oils, salad dressings, and eggs yielded the least.

Processed foods – monosaccharide composition in commercial complementary foods

To investigate the carbohydrate content in processed foods or foods containing multiple ingredients such as commercial complementary foods for toddlers, we compared the levels of arabinose in 23 products in a subset of a store name brand (Happy Family Brand foods). In processed foods, multiple ingredients were used to make the final product where the monosaccharide abundances of ingredients vary. Because the ingredients mostly contained raw food ingredients, the arabinose concentration (by fresh weight) for that whole food was used to generate the heat map in **Figure 4.6** for commercial complementary foods. For example, the raw ingredients for “Happy Tot Super Foods: pears, mangoes, spinach, super chia” included raw pears, mangoes, and spinach and were found to contain 0.27, 0.67, and 0.17 g of arabinose/100 g of fresh weight, respectively. Among this product, the arabinose content in mango (0.67 g/100 g fresh weight) was highest from all ingredients, while the total arabinose content of the complementary food product had lower amounts (0.47 g/100 g fresh weight). The rest of the complementary food for babies and toddlers had a varying range from 0.097 to 0.77 g of arabinose /100 g fresh weight.

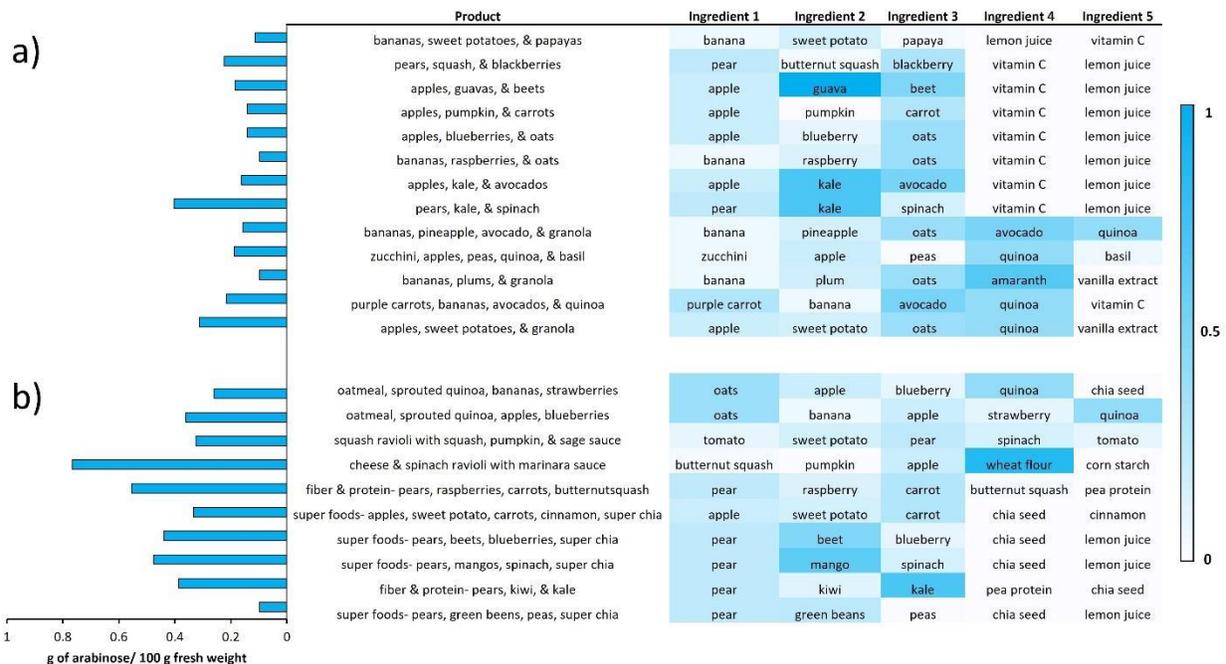


Figure 4.6. Heat map of 23 commercial complementary foods from Happy Family brand. The bar graphs on the left represents the total amount of arabinose found in the complementary food product along with the list of corresponding ingredients on the right with arabinose content in the ingredient whole food for babies (a) and toddlers (b).

The first whole food ingredient in processed foods contributes greatly to the monosaccharide composition. For example, when bananas were the first ingredients in the “Happy Family” infant products, the total arabinose content was low (less than 0.2 g of arabinose/100 g fresh weight). On the other hand, when pears were the first ingredients, the total arabinose had greater than 0.2 g of arabinose/100 g fresh weight with exception of the “Super Foods: pears, green beans, peas, super chia” product. The cheese & spinach ravioli with marinara sauce meal in the “Happy Family” toddler product yielded the highest arabinose content, likely due to the minimal moisture content.

CONCLUSION

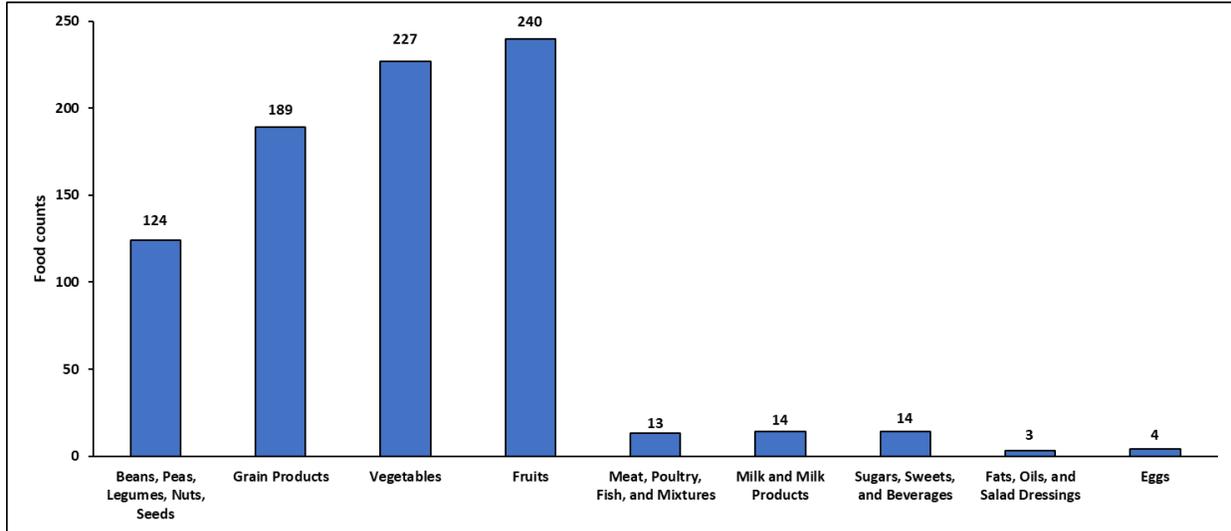
A novel rapid throughput UPLC-QqQ MS workflow was developed to determine the absolute quantitation of 14 monosaccharides in over 800 foods. Foods from diverse groups such as fruits, vegetables, fats, grains, dairy, beverages, and processed foods were subjected to monosaccharide analysis and the resulting monosaccharide compositions were used to create the “Davis Food Glyclopedia.” The results from the DFG can be used to determine specific monosaccharide amounts in foods and is capable of tailoring diets. Correlations were made within monosaccharides found in food. Clustering analysis was performed to visualize correlations in food groups based on quantitative monosaccharide compositions and amounts.

The resulting information provides an avenue for a more precise tailoring of diet to modulate the gut microbiome, which could lead to better health outcomes.

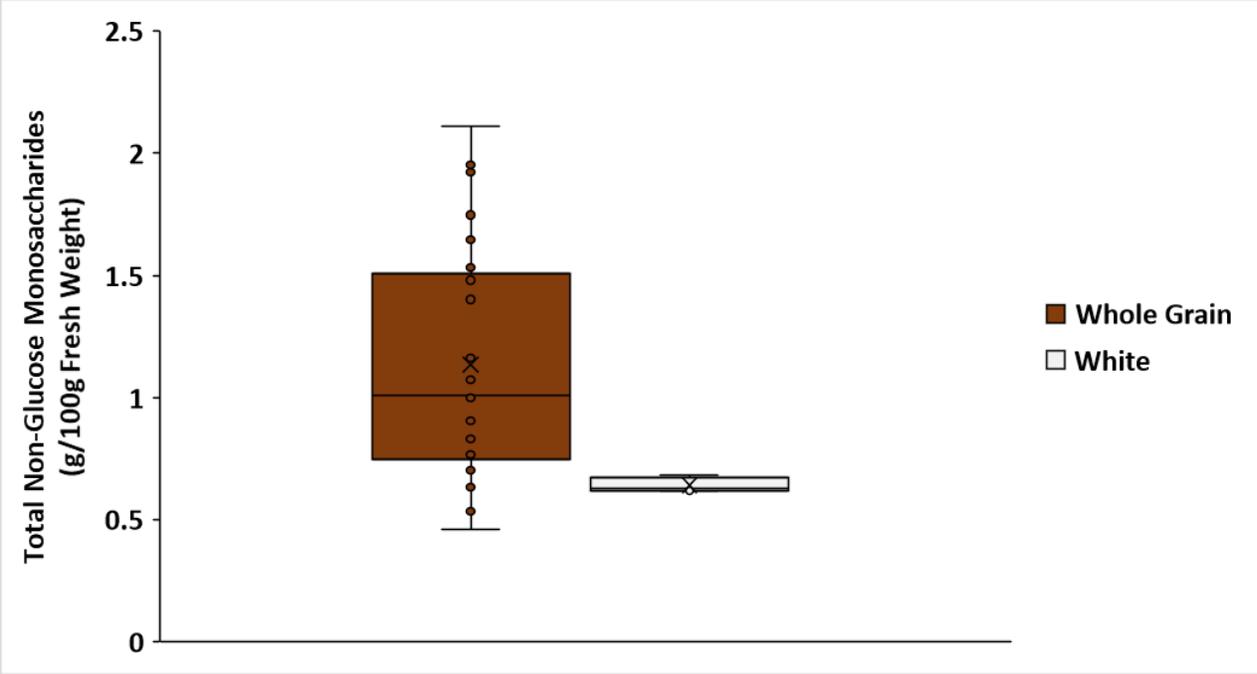
This research also demonstrated monosaccharide compositions can vary within food groups with several implications for nutrition research. For example, it will be necessary for nutrition studies to resolve dietary data at the individual food level, rather than summarizing servings at the food group level, if the intent is to study food-microbiome structure relationships. Mixed meals will need to be resolved at the ingredient level. Finally, the database will eventually need to be expanded to incorporate the full variety of plants consumed.

Even within the monosaccharide compositions determined, there lies an enormous amount of structural diversity as the polysaccharide and glycosidic linkage level information are not captured. Additionally, the methods utilized here did not employ sample preparation steps to separate free sugars and oligosaccharides from polysaccharides. Thus, for example, free fructose and glucose are not differentiated from inulin or starch, respectively. Future iterations of the DFG will utilize a rapid-throughput analytical workflow that will separate free saccharides from polysaccharides and provide linkage and polysaccharide level information on polysaccharides while free saccharides will be quantitated separately. These amendments will further provide a comprehensive and high-resolution picture of food carbohydrates.

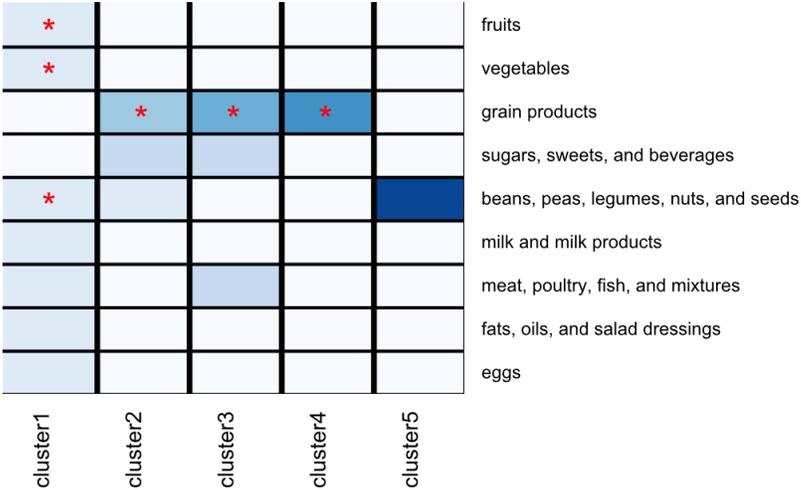
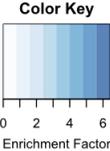
SUPPLEMENTARY MATERIALS



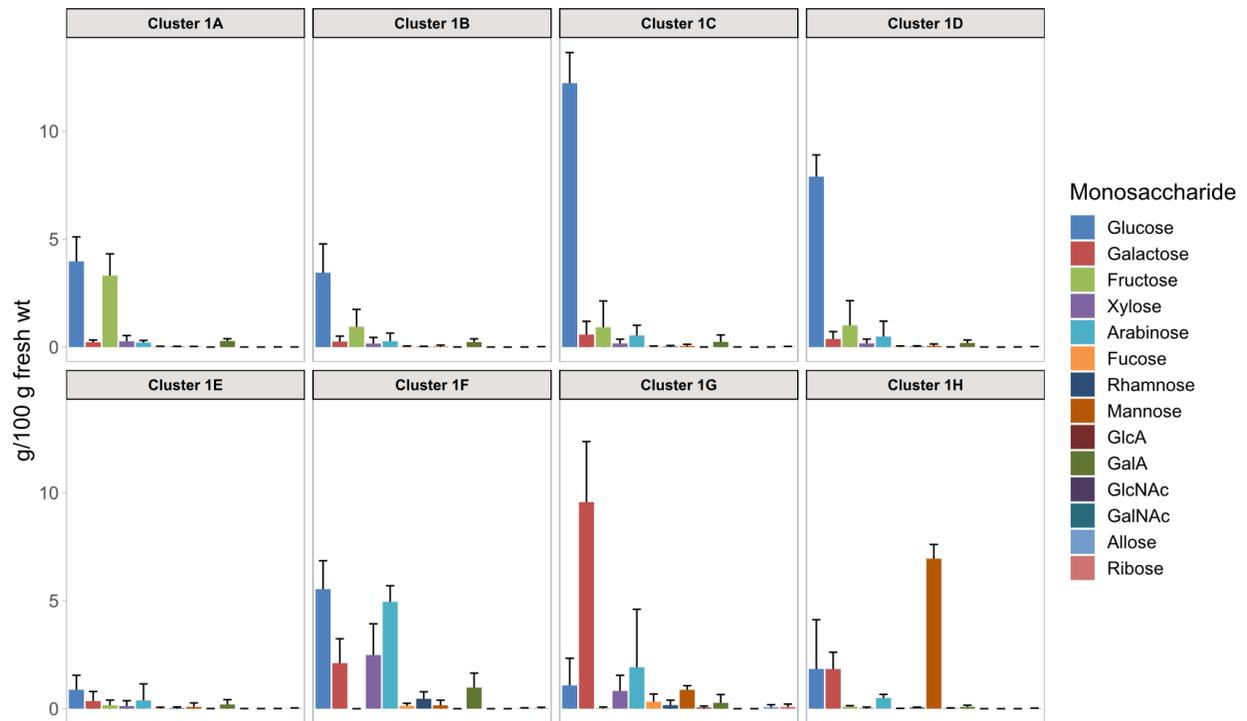
Supplementary Figure 4.1. Number of foods assigned to each of nine total food groups. Those groups containing plant-based foods contributed the largest number while animal-based and low carbohydrate groups contributed the least.



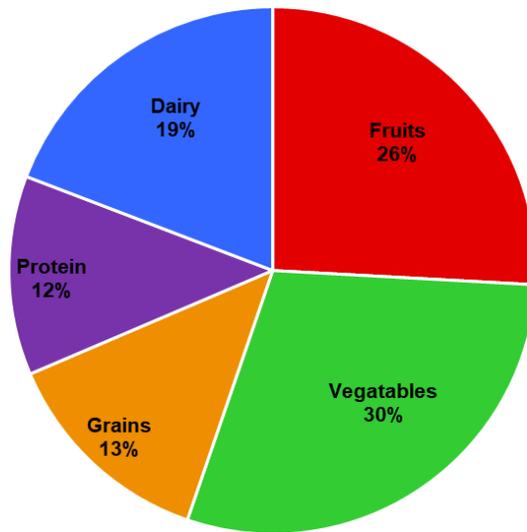
Supplementary Figure 4.2. Difference in total non-glucose monosaccharides between whole grain and “white” breads. Although not statistically significant, whole grain bread tended to contain more non-glucose residues (mostly from xylose, arabinose, and galactose).



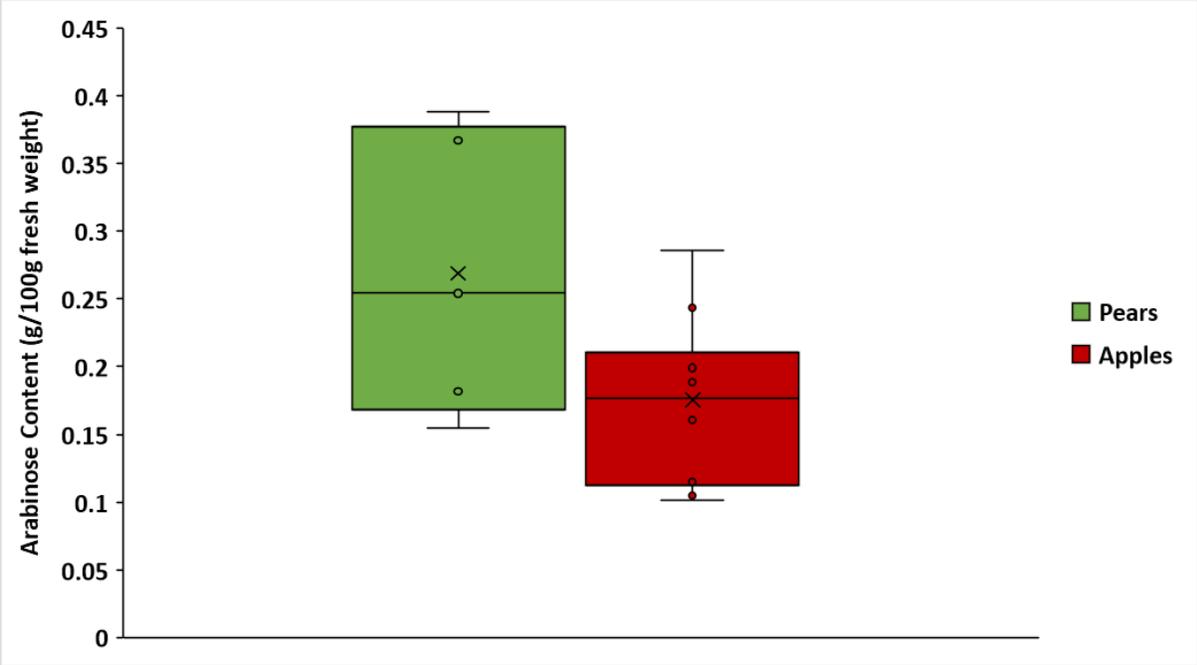
Supplementary Figure 4.3. Enrichment factor heatmap for each food group and cluster. Red asterisk indicates a significance of $p < 0.05$ using FDR adjusted p -values derived from a hypergeometric (over-representation) test.



Supplementary Figure 4.4. Average monosaccharide compositions of the sub-clusters A-H of Cluster 1. Error bars represent the standard deviation.



Supplementary Figure 4.5. Mass percentages of food groups used to generate an example meal based on recommendations in the USDA Dietary Guidelines for Americans 2020-2025.



Supplementary Figure 4.6. Although not statistically significant, pear cultivars tended to contain more arabinose content than apple cultivars on average.

REFERENCES

1. Le Couteur, D. G.; Solon-Biet, S.; Wahl, D.; Cogger, V. C.; Willcox, B. J.; Willcox, D. C.; Raubenheimer, D.; Simpson, S. J., New Horizons: Dietary protein, ageing and the Okinawan ratio. *Age Ageing* **2016**, *45* (4), 443-7.
2. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B., The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **2009**, *37* (Database issue), D233-8.
3. Henrick, B. M.; Rodriguez, L.; Lakshmikanth, T.; Pou, C.; Henckel, E.; Arzoomand, A.; Olin, A.; Wang, J.; Mikes, J.; Tan, Z.; Chen, Y.; Ehrlich, A. M.; Bernhardsson, A. K.; Mugabo, C. H.; Ambrosiani, Y.; Gustafsson, A.; Chew, S.; Brown, H. K.; Pramps, J.; Bohlin, K.; Mitchell, R. D.; Underwood, M. A.; Smilowitz, J. T.; German, J. B.; Frese, S. A.; Brodin, P., Bifidobacteria-mediated immune system imprinting early in life. *Cell* **2021**, *184* (15), 3884-3898 e11.
4. Collaborators, G. B. D. D., Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **2019**, *393* (10184), 1958-1972.
5. Cordain, L.; Eaton, S. B.; Sebastian, A.; Mann, N.; Lindeberg, S.; Watkins, B. A.; O'Keefe, J. H.; Brand-Miller, J., Origins and evolution of the Western diet: health implications for the 21st century. *Am J Clin Nutr* **2005**, *81* (2), 341-54.
6. Liu, L.; Wang, S.; Liu, J., Fiber consumption and all-cause, cardiovascular, and cancer mortalities: a systematic review and meta-analysis of cohort studies. *Mol Nutr Food Res* **2015**, *59* (1), 139-46.
7. Sonnenburg, J. L.; Backhed, F., Diet-microbiota interactions as moderators of human metabolism. *Nature* **2016**, *535* (7610), 56-64.
8. Makki, K.; Deehan, E. C.; Walter, J.; Backhed, F., The Impact of Dietary Fiber on Gut Microbiota in Host Health and Disease. *Cell Host Microbe* **2018**, *23* (6), 705-715.
9. Amicucci, M. J.; Nandita, E.; Galermo, A. G.; Castillo, J. J.; Chen, S.; Park, D.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lebrilla, C. B., A nonenzymatic method for cleaving polysaccharides to yield oligosaccharides for structural analysis. *Nat Commun* **2020**, *11* (1), 3963.
10. Phillips, K. M.; Haytowitz, D. B.; Pehrsson, P. R., Implications of two different methods for analyzing total dietary fiber in foods for food composition databases. *J Food Compos Anal* **2019**, *84*.
11. Barratt, M. J.; Lebrilla, C.; Shapiro, H. Y.; Gordon, J. I., The Gut Microbiota, Food Science, and Human Nutrition: A Timely Marriage. *Cell Host Microbe* **2017**, *22* (2), 134-141.
12. U.S. Department of Agriculture, U. S. D. o. H. a. H. S., Dietary Guidelines for Americans, 2020-2025. 9th Edition. **2020**.
13. Baldiviez, L. M.; Keim, N. L.; Laugero, K. D.; Hwang, D. H.; Huang, L.; Woodhouse, L. R.; Burnett, D. J.; Zerofsky, M. S.; Bonnel, E. L.; Allen, L. H.; Newman, J. W.; Stephensen, C. B., Design and implementation of a cross-sectional nutritional phenotyping study in healthy US adults. *BMC Nutr* **2017**, *3*, 79.
14. Service, U. A. R. What We Eat in America (WWEIA) Food Categories.
15. Service, U. A. R. Food and Nutrient Database for Dietary Studies (FNDDS) 2017-2018.
16. Sciences, N. D. o. C. C. a. P. Automated Self-Administered 24-Hour (ASA24) Dietary Assessment Tool

17. Xu, G.; Amicucci, M. J.; Cheng, Z.; Galermo, A. G.; Lebrilla, C. B., Revisiting monosaccharide analysis - quantitation of a comprehensive set of monosaccharides using dynamic multiple reaction monitoring. *Analyst* **2017**, *143* (1), 200-207.
18. Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T. T. T.; Liu, Y. Y.; Lee, M.; Xu, G. G.; Lebrilla, C. B., A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *International Journal of Mass Spectrometry* **2019**, *438*, 22-28.
19. Service, U. A. R. Food and Nutrient Database for Dietary Studies (FNDDS) 2015-2016.
20. Mohnen, D., Pectin structure and biosynthesis. *Curr Opin Plant Biol* **2008**, *11* (3), 266-77.
21. Burton, R. A.; Fincher, G. B., Evolution and development of cell walls in cereal grains. *Front Plant Sci* **2014**, *5*, 456.
22. Charrad, M.; Ghazzali, N.; Boiteau, V.; Niknafs, A., Nbclust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J Stat Softw* **2014**, *61* (6), 1-36.
23. Perry, G. H.; Dominy, N. J.; Claw, K. G.; Lee, A. S.; Fiegler, H.; Redon, R.; Werner, J.; Villanea, F. A.; Mountain, J. L.; Misra, R.; Carter, N. P.; Lee, C.; Stone, A. C., Diet and the evolution of human amylase gene copy number variation. *Nat Genet* **2007**, *39* (10), 1256-60.
24. Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J. Y.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B.; Barve, R. A.; Lombard, V.; Henrissat, B.; Leyn, S. A.; Rodionov, D. A.; Osterman, A. L.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Kirbach, K.; Wilmot, T.; Heath, A. C.; Klein, S.; Barratt, M. J.; Gordon, J. I., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021**, *595* (7865), 91-+.
25. Schutte, J. B.; Dejong, J.; Vanweerden, E. J.; Tamminga, S., Nutritional Implications of L-Arabinose in Pigs. *Brit J Nutr* **1992**, *68* (1), 195-207.
26. Osaki, S.; Kimura, T.; Sugimoto, T.; Hizukuri, S.; Iritani, N., L-arabinose feeding prevents increases due to dietary sucrose in lipogenic enzymes and triacylglycerol levels in rats. *J Nutr* **2001**, *131* (3), 796-799.
27. Marcotuli, I.; Hsieh, Y. S.; Lahnstein, J.; Yap, K.; Burton, R. A.; Blanco, A.; Fincher, G. B.; Gadaleta, A., Structural Variation and Content of Arabinoxylans in Endosperm and Bran of Durum Wheat (*Triticum turgidum* L.). *J Agric Food Chem* **2016**, *64* (14), 2883-92.

Chapter 5

Multi-glycomic characterization of fiber

from AOAC methods defines the carbohydrate

structures

ABSTRACT

Dietary fiber has long been known to be an essential component of a healthy diet and recent investigations into the gut microbiome-health paradigm have identified fiber as a prime determinant in this interaction. Further, fiber is now known to impact the gut microbiome in a structure-specific manner, conferring differential bioactivities to these specific structures. However, current analytical methods for food carbohydrate analysis do not capture this important structural information. To address this need, we utilized rapid-throughput LC-MS methods to develop a novel analytical pipeline to determine the structural composition of soluble and insoluble fiber fractions from two AOAC methods (991.43 and 2017.16) at the total monosaccharide, glycosidic linkage, and free saccharide level. Two foods were chosen for this proof-of-concept study: oats and potato starch. For oats, both AOAC methods gave similar results. Insoluble fiber was found to be comprised of linkages corresponding to β -glucan, arabinoxylan, xyloglucan, and mannan while soluble fiber was found to be mostly β -glucan with small amounts of arabinogalactan. For raw potato starch, each AOAC method gave markedly different results in the soluble fiber fractions. These observed differences are attributable to the resistant starch content of potato starch and the different starch digestion conditions used in each method. Together these tools are a means to obtain the complex structures present within dietary fiber while retaining “classical” determinations such as soluble and insoluble fiber. These efforts will provide an analytical framework to connect gravimetric fiber determinations with their constituent structures to better inform gut microbiome and clinical nutrition studies.

INTRODUCTION

The term “dietary fiber” was first introduced in the 1950s to define the portion of our diets that we cannot digest; namely cellulose, hemicellulose, and lignin.¹ This definition was iteratively refined through the 1970s as fiber’s purported health benefits were postulated and explored.²⁻⁵ In 2009, the World Health Organization and Codex Alimentarius arrived at an official definition to harmonize nomenclature efforts and health claims. As the definitions of dietary fiber evolved, so too did the analytical methods for measuring it. The determination of dietary fiber in food first began by applying methods for crude fiber analysis in animal feeds.⁶ These methods were then refined by the addition of digestive enzymes to measure digestible components such as starch separately from dietary fiber.⁷ Further expansion resulted in the delineation of soluble and insoluble fiber components as well as lignin.⁸ Since the 1970s, these analytical methodologies continued to be refined and harmonized into the methods defined by the Association of Official Analytical Chemists (AOAC) most commonly employed today. In 2015, the United States Food and Drug Administration (FDA) announced its definition of fiber as “non-digestible soluble and insoluble carbohydrates (with 3 or more monomeric units), and lignin that are intrinsic and intact in plants; isolated or synthetic non-digestible carbohydrates (with 3 or more monomeric units) determined by the FDA to have physiological effects that are beneficial to human health.” Within this definition, there lies an enormous amount of structural and functional diversity. Despite constant refinement of the methodologies, quantifying and characterizing these structures within dietary fiber remains a significant challenge.

Dietary fiber has garnered particular attention in recent years due to increased interest in the gut microbiome. Numerous reports have determined that fiber polysaccharides are a major driver of gut microbial ecology and heavily influence the plethora of microbial metabolites produced and

thus host health.⁹⁻¹² This interaction between fiber and the gut microbiota is dependent upon fiber structure. Specifically, gut microbes possess genes encoding for only certain glycosyl hydrolases and polysaccharide lyases that are specific for the degradation of particular monosaccharide and glycosidic linkage residues.^{13, 14} For example, the hemicellulose arabinoxylan (comprised of a $\beta 1 \rightarrow 4$ xylopyranose backbone with branches of $\alpha 1 \rightarrow 2,3$ arabinofuranose) would require a particular set of β -xylosidases and α -arabinosidases for degradation by gut microbes while xyloglucan (comprised of a $\beta 1 \rightarrow 4$ glucose backbone with branches of $\alpha 1 \rightarrow 6$ xylose $\beta 1 \rightarrow 2$ galactose, and others), would require β -glucosidases, α -xylosidases, β -galactosidases, and even α -fucosidases.^{15, 16} Thus, these hemicellulosic polysaccharides would potentially have completely different bioactivities. Further, the same polysaccharide from different sources contain unique monosaccharide and linkage profiles and these differences in fine structure also impact the gut microbiome differently.^{17, 18} Many other factors may also affect the structure of the polysaccharides in food such as time of harvest, post-harvest processing, and the method of cooking used.^{8, 19} Thus, analytical methods capable of characterizing and quantifying the carbohydrate structures in food are needed to understand the complex relationship between fiber and the gut microbiome. Rapid-throughput liquid chromatography-mass spectrometry (LC-MS) methods for monosaccharide and glycosidic linkage analysis have already proven essential in developing and understanding the mechanisms behind microbiota-directed therapeutic foods.^{20, 21} However, there are currently no analytical methods that utilize a multi-glycomic approach to integrate the free saccharide, total monosaccharide, and glycosidic linkage compositions of dietary fiber in food.

Despite efforts towards a generally applicable definition of dietary fiber, the term remains ambiguous and carries a different meaning to various stakeholders. To consumers and dieticians,

fiber is a necessary dietary component for optimal health. To many food companies, fiber may be an opportunity for marketing and product improvement. To plant scientists, fiber refers to the structures comprising the plant cell wall. To food scientists and chemists, fiber is a group of carbohydrates possessing glycosidic linkages preventing its digestion by human enzymes but allowing its fermentation by the gut microbiome. An ideal analytical framework for dietary fiber analysis would bridge these gaps and provide both “classical” enzymatic–gravimetric determinations as well as specific structural information utilizing monosaccharide and linkage analyses.

In this proof-of-concept study, two commonly employed AOAC methods (991.43 and 2017.16) were used to determine the total dietary fiber content of raw oats and potato starch utilizing a dietary fiber analyzer. The resulting soluble and insoluble fractions were determined, isolated, and subjected to comprehensive structural analysis employing three recently developed LC-MS-based methods to determine their total monosaccharide, glycosidic linkage, and free saccharide compositions. The products of each starch digestion were quantified and characterized, providing insight towards the applicability of each method. This analytical pipeline is a significant step forward in the information obtained from classical dietary fiber determinations and is unique in that it captures macroscopic, nutritional definitions such as “soluble fiber” and “insoluble fiber” while also quantifying and defining the structural composition of the oligo- and polysaccharide components of these fractions. This feature can effectively bridge the gaps between food chemistry, clinical science, and the gut microbiome, thus representing a path forward for dietary fiber analysis as well as its definition in the future.

MATERIALS AND METHODS

Trifluoroacetic acid (TFA), chloroform (HPLC grade), ammonium acetate, ammonium hydroxide solution (NH₄OH) (28-30%), sodium hydroxide pellets (semiconductor grade, 99.99% trace metals basis), dichloromethane, anhydrous dimethyl sulfoxide (DMSO), iodomethane, 3-methyl-1-phenyl-2-pyrazoline-5-one (PMP), methanol (MeOH, HPLC grade), fructose, ribose, rhamnose, mannose, allose, glucuronic acid (GlcA), galacturonic acid (GalA), glucose, galactose, N-acetylglucosamine (GlcNAc), N-acetylgalactosamine (GalNAc), xylose, arabinose, fucose, maltose, sucrose, and raffinose were purchased from Sigma-Aldrich (St. Louis, MO). Rye arabinoxylan, maltotetraose, maltopentaose, maltohexaose, kestose, stachyose, and verbascose were purchased from Megazyme (Bray, Ireland). Acetonitrile (HPLC grade) was purchased from Honeywell (Muskegon, MI). Sodium hydroxide (reagent grade), maleic acid, acetic acid (glacial), hydrochloric acid, D-sorbitol, calcium chloride (CaCl₂·2H₂O), deionized water, acetone (reagent grade), and ethanol (99%) were purchased from Fisher Scientific (Pittsburg, PA, USA). Dietary fiber analyzer bags (SDF bags (ANKOM #DF-S), IDF bags (ANKOM #DF-I)) and enzymes (α -amylase (ANKOM concentrate enzyme # TDF81), protease (ANKOM concentrate enzyme # TDF82), amyloglucosidase (AMG, ANKOM concentrate enzyme # TDF85)) were obtained from ANKOM Technology (Macedon, NY, USA). MES-Tris buffer solution (0.05 M, pH 8.2) was prepared using reagent MES (2-(N-morpholino)ethanesulfonic acid) and Tris (Tris(hydroxymethyl)aminomethane) obtained from Fisher Scientific (Pittsburgh, PA, USA). The sodium maleate buffer solution (50 mM, pH 6.0) was prepared using maleic acid, sodium hydroxide, and calcium chloride. The rapid integrated total dietary fiber assay kit (K-RINTDF) was obtained from the Neogen part of Megazyme (Lansing, MI, USA). Tris base was purchased from Sigma-Aldrich (St. Louis, MO, USA). Oat sample was purchased from the company

ANKOM Technology (Macedon, NY, USA) and raw potato starch (PenPure^R 10) was obtained from IngredionTM (Westchester, Illinois, USA).

Fiber content determination. Determination of insoluble (IDF) and soluble dietary fiber (SDF) was performed by AOAC 991.43 and 2017.16 methods using an automated ANKOM^{TDF} dietary fiber analyzer.^{1,2} In brief, samples (1 g) were weighed in IDF filter bags (5 replicates) and set on the serial Ankom fiber analyzer. In the sixth bag, the oat reference standard was analyzed to monitor the performance of the instrument. The fiber analyzer was programmed to deliver buffer and enzymes according to the AOAC methods.

For methods AOAC 991.43, digestion was sequentially performed with enzymes (α -amylase, protease, and amyloglucosidase (AMG)) under controlled temperature as previously reported.¹ In brief, the fiber analyzer was programmed to deliver 40 mL of MES-Tris buffer solution and 1 mL of enzymes (α -amylase, protease, and AMG) were sequentially added in each bag. The time and temperature were set to 30 min at 95 °C for α -amylase digestion, and 30 min at 60 °C for protease digestion. The hydrochloric acid was pumped to adjust the pH to 4.0-4.7 to terminate the digestion reaction. After completion of the digestions, the digested fiber materials filtered through the IDF bag to SDF bags. At this stage, the instrument was set to deliver 225 mL 95% ethanol at 60 °C to each digested sample for precipitation of soluble dietary fiber with ethanol. After 60 min, the solution was filtered through the SDF filter bags, and the filtrate was collected in the glass container. The precipitate was rinsed twice with 15 mL of 78% ethanol and 95% ethanol. The IDF and SDF bags were removed from the fiber analyzer and rinsed with acetone and air-dried for 30 min followed by drying in an oven at 100-102 °C for at least 90 min. The dried bags were weighed to determine the crude IDF and SDF contents in the samples.

For the AOAC 2017.16 method, the following modifications were done before fiber analysis.³ A 35 mL of sodium maleate buffer solution (50 mM, pH 6.0 and 2 mM CaCl₂) was used instead of 40 mL MES-Tris buffer (0.05 M, pH 8.2) in AOAC 991.43. In addition, 1 mL of 100 mg/mL of sorbitol, 2 mL of PAA/AMG enzyme (PAA (4 KU/5 mL) plus AMG (1.7 KU/5 mL)) were used in AOAC 2017.16 as compared to 1 mL of α -amylase (150 Ceralpha/mL; 10.8 U/mg in 50% glycerol + 0.09% sodium azide) in AOAC 991.43. The time and temperature were set to 4 h at 37 °C for the PAA/AMG digestion. Furthermore, the total incubation time was 5 h for AOAC 2017.16, whereas the total incubation time was 90 min for AOAC 991.43. After 4 h, the instrument delivered Tris base solution 3 mL (0.75 M) to adjust the pH 8.2 to terminate the reaction. The termination reaction incubation time was set to 30 min at 60 °C. The instrument was set to deliver 1 mL protease enzyme and the mixture was incubated for 30 min. The mixture pH was adjusted to 4.2 to terminate the reaction. After digestion, the soluble dietary fiber passed through the IDF filter bags to SDF filter bags. The SDF filter bags were pre-loaded with diatomaceous earth. The fiber analyzer was programmed to deliver 225 mL of 95% ethanol at 60 °C to each digested sample and the mixture was incubated for 60 min for precipitation of SDF. The precipitated was washed twice with 15 mL 78% ethanol and 95% ethanol and the samples were processed in the same way as reported above for AOAC 991.43. The filtrate considered at soluble dietary fiber soluble (SDFS) in ethanol was initially concentrated in a rotary evaporator followed by lyophilization. The dried IDF, SDF, and the concentrated filtrate residue were analyzed using mass spectrometric analysis. SDFS was determined via UHPLC-QqQ MS analysis by summing the concentrations of all oligosaccharides (saccharides with degree of polymerization [DP] > 2) measured by the method.

Determination of protein and ash content. Undigested protein content was determined using a combustion method on a rapid MAX N exceed Dumas analyzer from Elementar. Only one sample

was used to check the amount of undigested protein and duplicate analysis was done for ash content analysis. Initially, the bags were sealed from the top and the bottom. For protein analysis outside plastic covering was removed. For the determination of the ash content, the sealed bags were placed in a preweighed crucible and incinerated in the Barnstead thermolyne muffle furnace (Type 48000, Thermo Scientific, Walham, MA, USA) at 600 °C. Blank bags without samples were also separately incinerated in the same way to do the blank correction. For undigested protein determination, the sealed bags were further sealed at three positions. These were then cut into three small bags that were used for the determination of nitrogen content using Elementar Combustion Analyzer (Elementar Americas Inc., Ronkonkoma, NY, USA).⁴ The total protein of the sample was determined and summed to determine total nitrogen content in a bag. Only a single analysis for nitrogen and duplicated analysis for ash were carried out for each sample and the average of the two runs was used for ash corrections.

Preparation of Dietary Fiber Fractions for Glycomic Analysis. The oats, potato starch, and insoluble dietary fiber samples were homogenized by bullet-blending with an Omni Bead Ruptor Elite (Kennesaw, GA) before a 10 mg aliquot was weighed out into 1.5 mL screw-cap Eppendorf tubes. An arabinoxylan polysaccharide standard was also prepared in the same way and used as a control for monosaccharide and linkage analysis. The whole oat and potato starch samples were precipitated with 80% EtOH and the supernatant removed and saved for free saccharide analysis before further homogenization. A stock solution of 10 mg/mL was prepared by adding 1 mL of nanopure water. The stock solution then underwent a bullet blending procedure followed by heat treatment (1 h at 100°C) and another round of bullet blending to ensure homogeneity. Soluble dietary fiber and soluble dietary fiber precipitate samples were contained in a dry mixture with diatomaceous earth. These mixtures were transferred to 15 mL Falcon tubes and 10 mL of

nanopure water was added. The samples were then vortexed thoroughly and the resulting suspensions were incubated at 100 °C for 1 hr to solubilize the carbohydrates. The tubes were then centrifuged, and a 5 mL aliquot was diluted to 10 mg/mL by original dry weight for analysis. Waste and soluble dietary fiber supernatant samples were reconstituted in nanopure water to make 50 mg/mL stock solutions which were then diluted to 10 mg/mL for analysis.

Quantitative Monosaccharide Compositional Analysis. Methods were adapted from previous publications with some adjustments.²⁰⁻²² Aliquots of each stock solution and an arabinoxylan polysaccharide standard (used as a quality control) were subjected to acid hydrolysis with 4 M TFA for 1 h at 121 °C in 96-well plates. Hydrolysis was quenched by the addition of cold nanopure water. A 10 uL aliquot from each sample was then derivatized alongside an external calibration curve (0.001 to 100 µg/mL) containing 14 monosaccharides by adding 100 uL of 0.2 M PMP in methanol, 100 uL of ammonia solution (28-30 % w/v), and heating to 70 °C for 30 min. The derivatized glycosides were dried to completeness by vacuum centrifugation and extracted twice with chloroform to remove excess PMP. A 1 uL aliquot of the aqueous layer was injected into an Agilent 1290 Infinity II UHPLC system equipped with an Agilent Poroshell HPH C18 column (50 × 2.1 mm i.d., 1.8 µm particale size) and corresponding guard column. Separation of the PMP-labeled monosaccharides was achieved using a constant flow rate of 0.9 mL/min and a 2 min isocratic elution at 11 % B followed by a 1.6 min flush at 99% B, and 0.8 min equilibration for a total run time of 4.6 min. Solvent A consisted of 25mM ammonium acetate in 5% acetonitrile with pH adjusted to 8.2 using ammonia solution. Solvent B consisted of 95% acetonitrile in water. Mass spectral analysis was carried out on an Agilent 6495B triple quadrupole (QqQ) mass spectrometer (Agilent Technologies, Santa Clara, CA) operated in positive ion mode while using

dynamic multiple reaction monitoring (dMRM). The total monosaccharide content in each sample was determined by comparison to the external calibration curve.

Glycosidic Linkage Analysis. A permethylation procedure was adapted from Galermo et al.^{23,24} Aliquots of 5 uL were transferred from each sample stock solution to a 96-well plate and permethylated using iodomethane (40 uL) in a solution of DMSO (150 uL) containing saturated NaOH (5 uL). The samples were allowed to react on a shaker at room temperature for 50 min under argon before being quenched by the addition of cold water and then DCM. NaOH and DMSO were removed by repeated extraction with cold nanopure water. The upper aqueous layers were discarded while the bottom organic layer containing permethylated products was dried to completion by vacuum centrifugation. The permethylated samples were then subjected to acid hydrolysis at 100°C for 2 hr with 4 M TFA and subsequently dried by vacuum centrifugation. The released permethylated monosaccharide residues were derivatized with PMP following the previously described procedure. Once dried, the samples were then reconstituted in 100 uL of 70% aqueous methanol. Separation and analysis of the permethylated glycosides were carried out on an Agilent 1290 Infinity II UHPLC system equipped with an Agilent Zorbax RRHD Eclipse Plus C18 column (150 × 2.1 mm i.d., 1.8 μm particle size) and corresponding guard. Mass spectral analysis was also carried out on an Agilent 6495B triple quadrupole (QqQ) mass spectrometer (Agilent Technologies, Santa Clara, CA) operated in multiple reaction monitoring (MRM) mode. For analysis, 1 μL of sample was injected onto and separated using a 15 min binary gradient with a constant flow rate of 0.45 mL/min. Mobile phase A and B were the same as those used for monosaccharide analysis except the pH of mobile phase A was adjusted to 7.7. The following binary gradient was used: 0.00–5.00 min, 21.00% B; 5.00–9.00 min, 21.00–22.00% B; 9.00–11.00 min, 22.00% B; 11.00–13.60 min, 22.00–24.50% B; 13.60–13.61 min, 24.50– 99.00% B;

13.61–13.80 min, 99.00% B; 13.80–13.81 min, 99.00–21.00% B; 13.81–15.00 min, 21.00% B. Glycosidic linkages were identified by comparing their MRM transitions and retention times to an established library.

Free saccharide analysis. The 10 mg/mL stock solutions of each sample were diluted first with water and then into 75% ACN after centrifugation at 10,000 rpm for 5 min. A standard curve containing two monosaccharides, three disaccharides, and seven oligosaccharides ranging in concentration from 1 to 200 µg/mL was also prepared. A 5 µL aliquot was then injected onto the same Agilent UHPLC-QqQ MS instrument this time fitted with a Waters BEH Amide column (150 × 2.1 mm i.d., 2.5 µm particle size) and corresponding guard. Free saccharides were separated using a 25 min gradient elution: 0–4.0 min, 95% B; 4.0-5.5 min, 95 - 80% B; 5.5-16.0 min, 80-60% B; 16.0-19.0 min, 60% B; 20.0-25.0 min, 95% B. Mobile phase A consisted of 10 mM ammonium acetate in 10 % aqueous ACN while mobile phase B consisted of 10 mM ammonium acetate in 90 % ACN. The pH of both mobile phases was adjusted to 10.2 with aqueous ammonia. The 6495B QqQ MS was operated in negative ion mode and used single ion monitoring (SIM) for detection. Free saccharides in each sample were determined by comparison to the external calibration curve.

RESULTS

Two food samples (oats and potato starch) were analyzed by AOAC methods 991.43 and 2017.16 on an Ankom Dietary Fiber Analyzer (**Figure 5.1**). The resulting fractions were dried, weighed, and structurally elucidated using a comprehensive multi-glycomics workflow consisting of quantitative total monosaccharide and free saccharide analyses as well as a glycosidic linkage

analysis (**Figure 5.2**). The monosaccharide analysis monitored the prevalent monosaccharides in food and provided quantitative information on the carbohydrates present in each fiber fraction. The glycosidic linkage analysis monitored nearly 100 linkages and found over 50, comprising the preponderant linkages found in food. This analysis provided structural information on the saccharides found in the foods and their fiber fractions. Additionally, a quantitative free saccharide analysis was performed to capture free sugars (glucose, fructose, sucrose, maltose, and lactose) and low molecular weight oligosaccharides (raffinose, kestose, stachyose, verbascose, and maltooligosaccharides DP 4 to ~10) in each fraction. These results added a rich, integral layer of information capable of identifying, quantifying, and structurally elucidating the mono-, di-, oligo-, and polysaccharide components of each fraction in unprecedented detail.

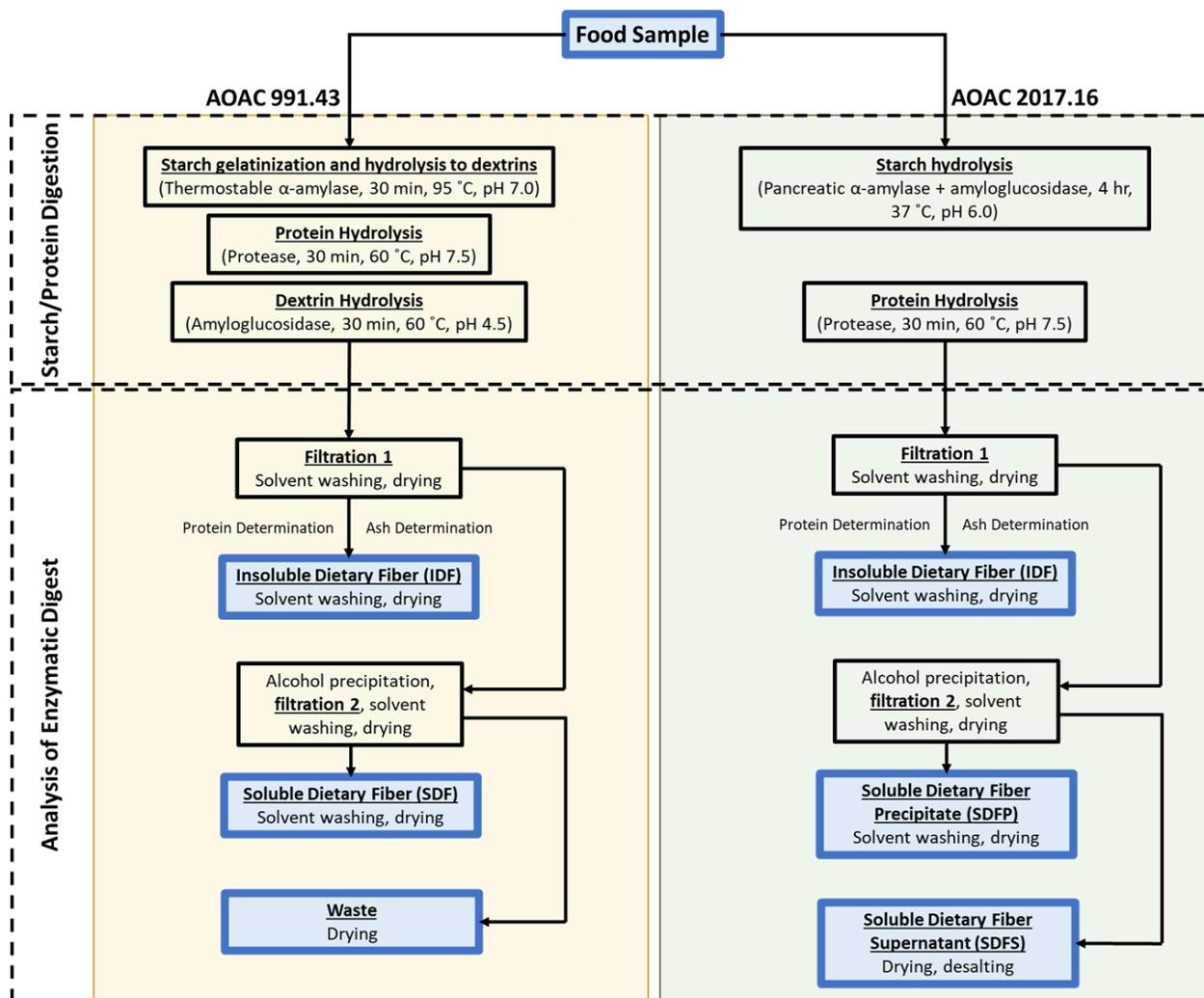


Figure 5.1. Overview of AOAC methods 991.43 and 2017.16. Fractions highlighted in blue were collected, weighed for gravimetric determinations, and subjected to a comprehensive glycomic analysis that provided total monosaccharide, linkage, and free saccharide compositions.

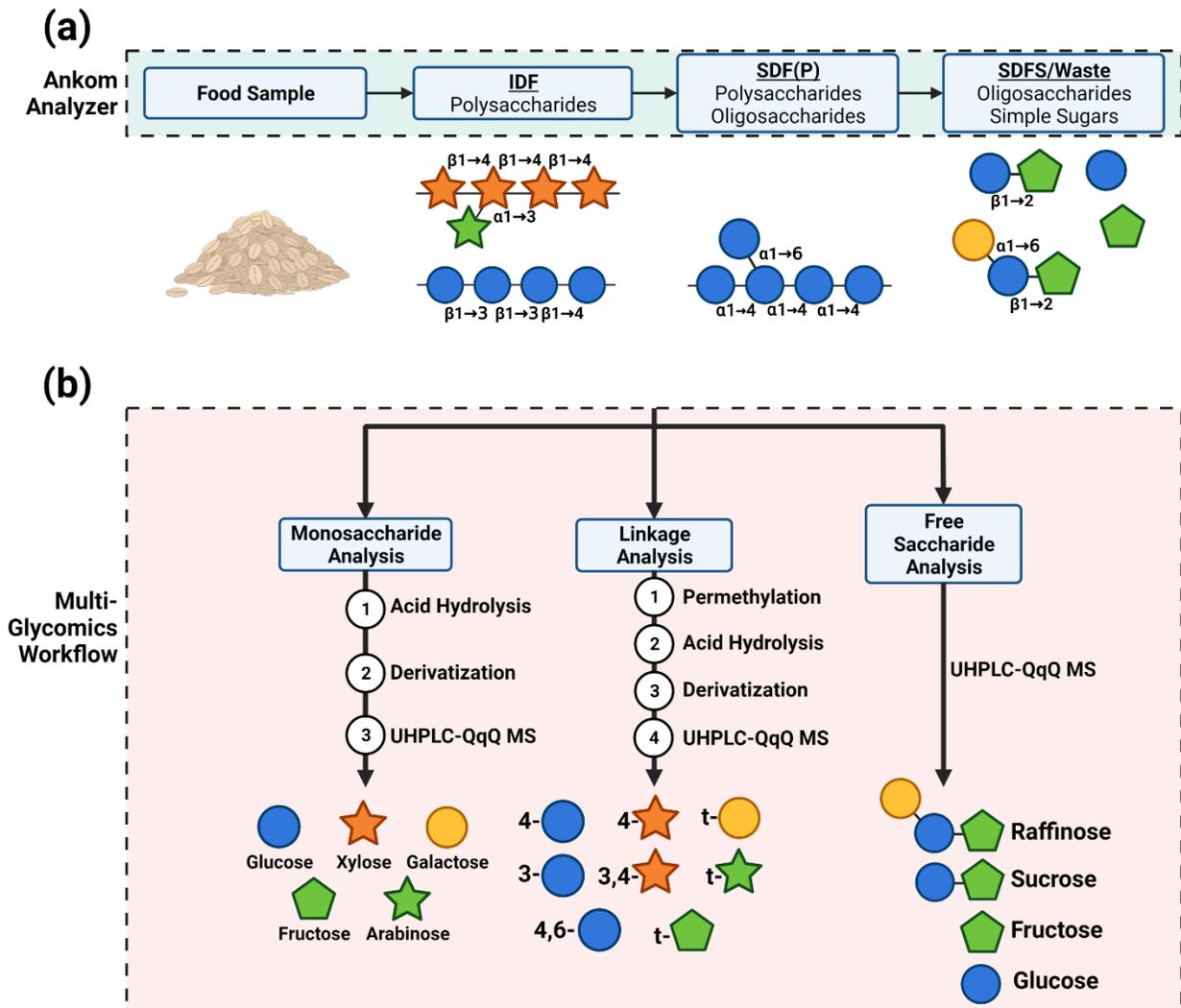


Figure 5.2. Overview of the multi-glycomics workflow used to analyze AOAC fiber fractions. (a) The enzymatic-gravimetric AOAC methods fractionate the native carbohydrate components from food into IDF, SDF(P), and SDFS. Select structures present in oat are shown as examples. (b) Each fraction was then subjected to quantitative monosaccharide compositional, glycosidic linkage, and free saccharide analyses.

Enzymatic-Gravimetric Determination of Dietary Fiber Fractions

The calculated average percentages of dietary fractions measured with both AOAC methods 991.43 and 2017.16 for oats and potato starch are provided in **Table 5.1**. Potato starch did not contain enough insoluble dietary fiber (IDF) to measure. AOAC method 991.43 does not include a soluble dietary fiber supernatant (SDFS) fraction.

Table 5.1. IDF (insoluble dietary fiber), SDF (soluble dietary fiber), SDFP (soluble dietary fiber precipitated with ethanol), SDFS (soluble dietary fiber soluble in ethanol) and, TDF (total dietary fiber) percent in oat and potato starch samples. Both samples were analyzed in triplicate.

Samples	(%) IDF (mean \pm SD)	(%) SDFP (mean \pm SD)	(%) SDFS (mean \pm SD)	(%) TDF (mean \pm SD)
AOAC 991.43				
Oat	11.91 \pm 0.42	8.64 \pm 0.51	NA	20.55 \pm 0.92
Potato Starch	NQ	0.35 \pm 0.02	NA	0.35 \pm 0.02
AOAC 2017.16				
Oat	10.25 \pm 0.07	8.32 \pm 0.53	2.84 \pm 0.23	21.41 \pm 0.23
Potato Starch	NQ	69.72 \pm 0.19	7.10 \pm 1.50	76.82 \pm 1.5
NA = Not applicable, NQ = Not quantifiable, and SD = standard deviation				

IDF is defined as fiber that is insoluble in water which encompasses many polysaccharide components with diverse structures as well as lignin. SDFP contains fiber that is soluble in water and insoluble in ethanol which may also be comprised of a large number of polysaccharide and

oligosaccharide components. SDFS is the component of fiber that is soluble in ethanol and is limited to small oligosaccharides.

Multi-glycomic Analysis of Oats

The monosaccharide and glycosidic linkage compositions of the oat sample and each of its dietary fiber fractions are provided in **Figure 5.3**. Monosaccharide analysis showed that the oat sample contained mostly glucose (580 $\mu\text{g}/\text{mg}$ by dry weight) with small amounts of xylose (16 $\mu\text{g}/\text{mg}$), arabinose (17 $\mu\text{g}/\text{mg}$), galactose (6 $\mu\text{g}/\text{mg}$), and mannose (1 $\mu\text{g}/\text{mg}$). The total carbohydrate content was 620 $\mu\text{g}/\text{mg}$ by dry weight (**Supplementary Table 5.1**). Glycosidic linkage analysis revealed that the majority of the glucose was 4-linked (66 %) with a smaller contribution from 3-linked glucose (3.5%). The large 4-linked component is consistent with a high amount of starch while the 3-linked glucose is more consistent with β -glucan as 3-linked is not found in starch. Linkage analysis also yielded xylose that were 4-linked (0.5%) and branched 3,4-linked (0.3%) with arabinose that were terminal (2.8%) consistent with the presence of arabinoxylan (**Supplementary Table 5.2**). The insoluble dietary fiber (IDF) fractions from both AOAC methods yielded similar carbohydrate abundances and compositions with glucose, xylose, and arabinose being the main constituents along with small amounts of mannose. Furthermore, linkage analysis of the IDF fractions produced nearly equal amounts of 3- and 4-glucose (suggesting β -glucan) and 4- and 3,4-xylose and t-arabinofuranose (suggesting arabinoxylan). The presence of 4-mannose further suggested the presence of β -mannan. Note that while we could not determine the polysaccharides unambiguously, we inferred the identities from the linkages as is commonly

done in the field.²⁵ Henceforth to facilitate the discussion, we provide polysaccharide identities that were inferred from the linkages.

Due to the high sensitivity of the analysis, xyloglucan linkages such as t-galactose, t-xylopyranose, and 4,6-glucose were also detected despite their low abundances. Although the IDF from both methods were similar in composition, 2017.16 gave significantly higher total glucose than 991.43. Furthermore, 4-, 4,6- and t-glucose were more abundant in the 2017.16 IDF (although only t- and 4,6-glucose reached statistical significance) suggesting the increased glucose relative to 991.43 was likely from starch (**Supplementary Figure 5.1a, b**), and likely resistant starch.

Analysis of the soluble dietary fiber (SDF, 991.43) and soluble dietary fiber precipitate (SDFP, 2017.16) revealed compositions composed of 3- and 4-glucose in ratios of about 1:3. Similar ratios are consistent with β -glucan.²⁶ Similarly, small amounts of 6-, 4,6- and t-galactose along with t-arabinose pointed to branched arabinogalactans. Small amounts of 4- and 3,4-xylose suggested the presence of arabinoxylan as a minor component. Note that the amylase and amyloglucosidase added for the digestion of starch in the original sample were not removed, and the monosaccharide and linkage compositions of their mannose-containing N-glycans were also visible. An abundance of 2-, 3-, and t-mannose further supports this notion. Mannose and the N-glycan associated linkages were also abundant in the blank samples from both methods with 2017.16 having more of these components than 991.43.

The monosaccharide analysis of the waste of 991.43 and soluble dietary fiber supernatant (SDFS) of 2017.16 fractions were found to contain nearly all glucose with the majority being terminal and minor components 4-linked. These findings pointed towards the presence of free glucose monomers along with small concentrations of maltooligosaccharides- derived from the

enzymatic starch digestions in each method. Enzyme was also present in these fractions as evidenced by the t- and 3-linked mannose residues.

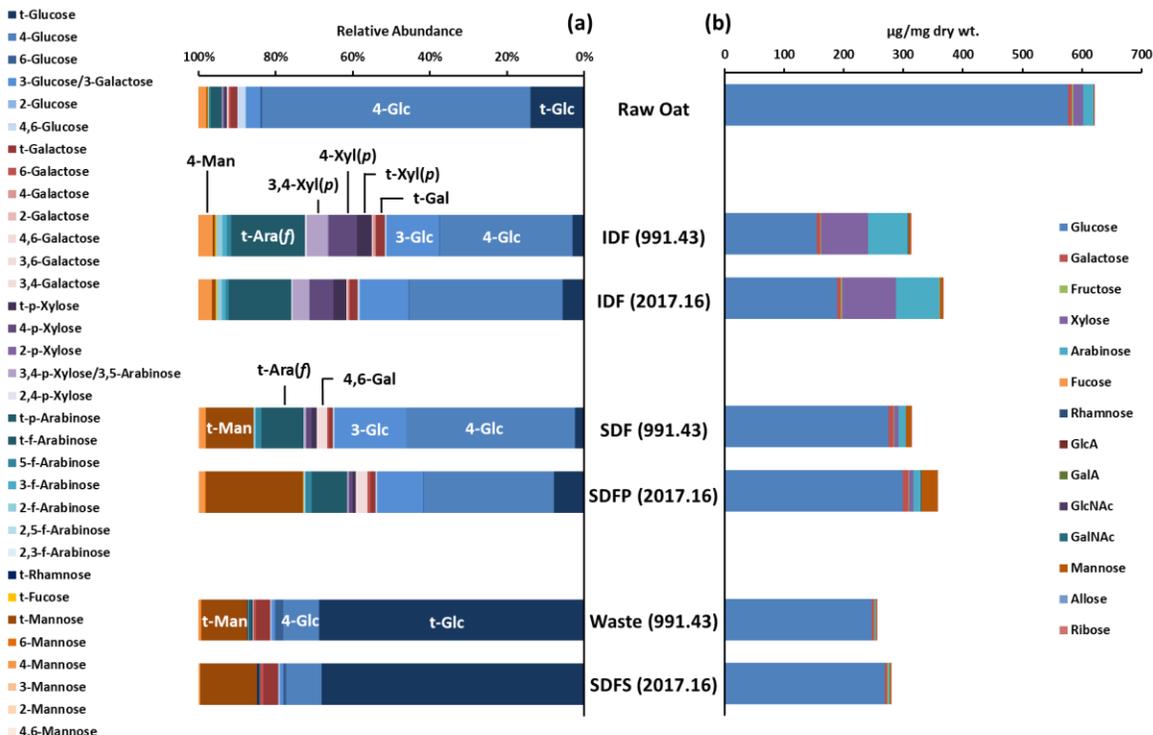


Figure 5.3. (a) Glycosidic linkage and (b) monosaccharide compositions of oat and its dietary fiber fractions from each AOAC method. Each bar represents an average of three technical replicates.

Free saccharide analysis of the waste and SDFS confirmed the majority of these fractions was free glucose (**Figure 5.4, Supplementary Table 5.3**) revealed differences in the profiles of starch breakdown products from the enzymatic digestions. Namely, the waste fraction from 991.43 contained more free saccharide from maltose and maltooligosaccharides. Lower abundances of these compounds were found in the SDFS fraction from 2017.16. Free saccharide

analysis of the SDF from 991.43 and SDFP from 2017.16 gave opposite results with more free saccharides found in the SDFP than in the SDF. Sucrose and fructooligosaccharides (stachyose and raffinose) were found in original oat samples. They were also found in the waste and SDFS fractions. Both the waste from 991.43 and SDFS from 2017.16 contained glucose, maltose, and maltooligosaccharides from the enzymatic digestions. However, the maltooligosaccharides from each method displayed different retention, suggesting distinct oligosaccharide products.

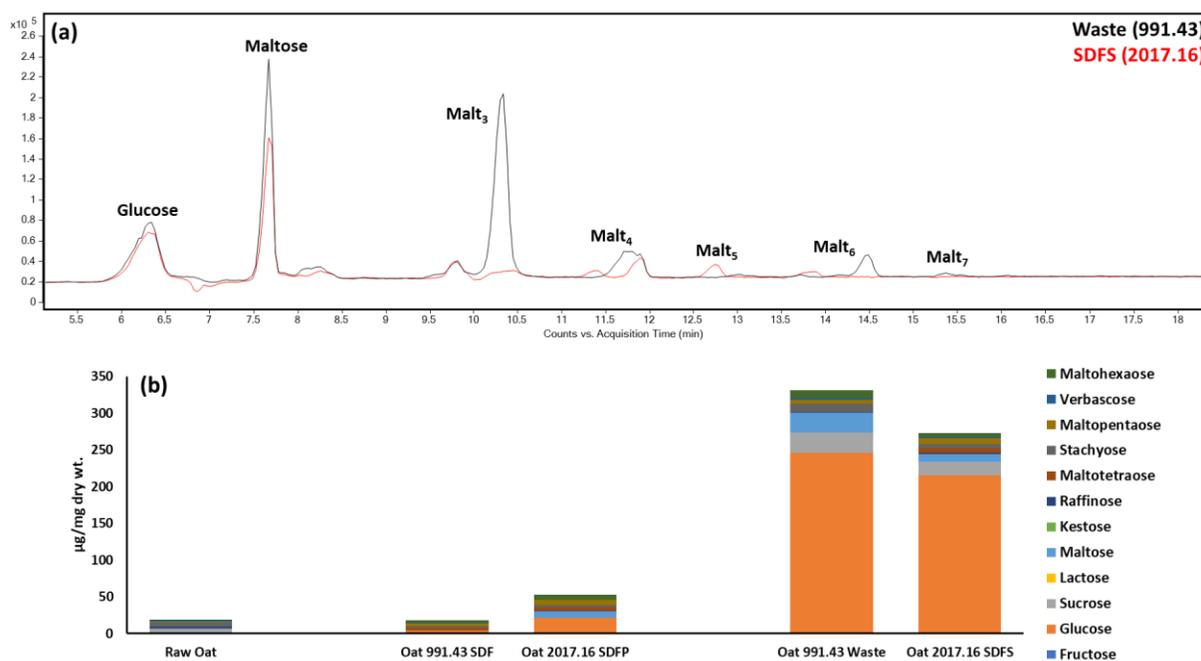


Figure 5.4a-b. (a) Total ion chromatogram depicting the free saccharides present in oat waste (991.43) and SDFS (2017.16). (b) Quantified results of free saccharides in the raw oat sample, SDF, SDFP, waste, and SDFS fractions.

Multi-glycomic Analysis of Potato Starch

The compositions of the potato starch and its AOAC fractions are illustrated in **Figure 5.5**. Potato starch was found to be nearly 100 % glucose by dry weight with 4-linked, t-, and 4,6-linked glucose being the predominant linkages present (**Supplementary Tables 5.4 and 5.5**) and consistent with starch. There was insufficient material collected from the IDF fractions of either method to perform glycomic analysis.

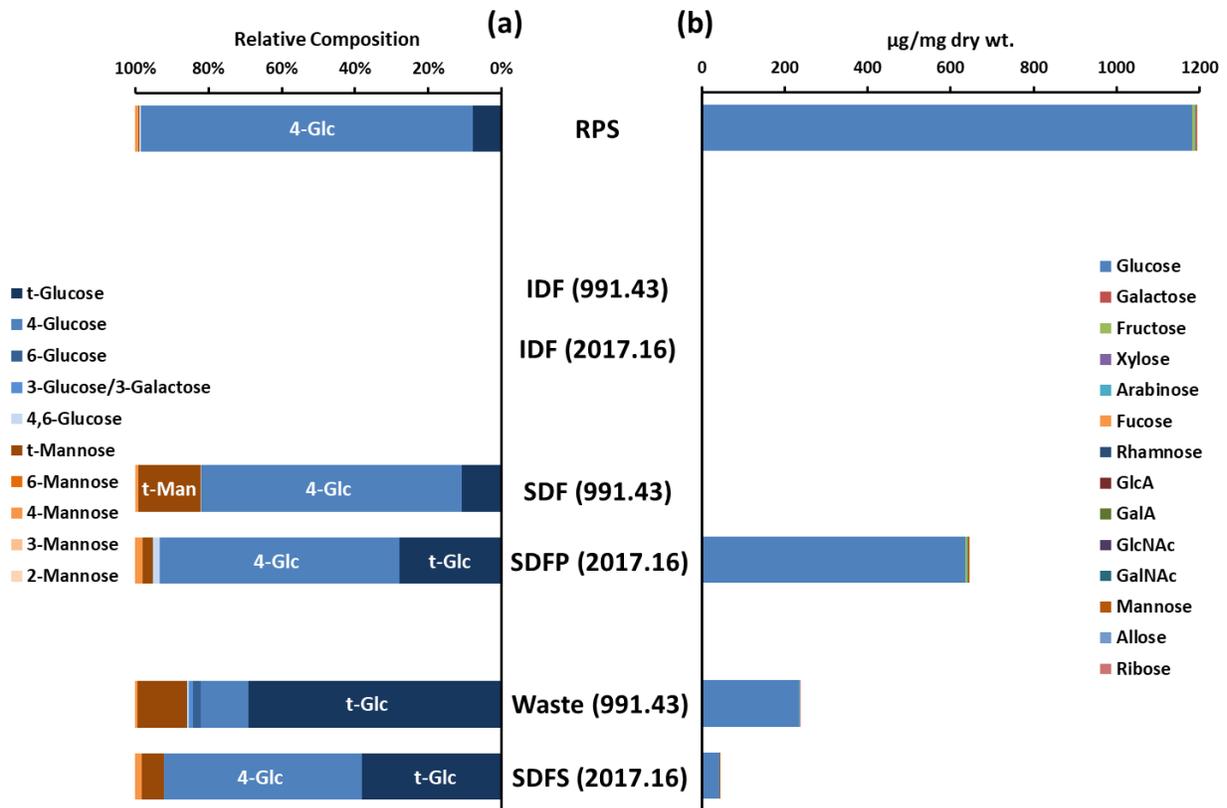


Figure 5.5a-b. (a) Glycosidic linkage and (b) monosaccharide compositions of potato starch and its dietary fiber fractions from each AOAC method. There was insufficient material in the IDF fraction from either method to perform glycomic analysis. Each bar represents an average of three technical replicates.

Monosaccharide analysis showed SDF contained only minute amounts of carbohydrates, while SDFP was found to possess the majority of the carbohydrates from the original sample, unlike the oat samples. The large discrepancy is likely due to the differences in the enzymatic starch digestion conditions in each method. AOAC 2017.16 utilizes a milder digestion than 991.43, resulting in large oligomers that precipitate out into SDFP. Linkage analysis confirmed that both SDF and SDFP contained maltodextrins arising from starch digestion. However, the procedure used in 2017.16 resulted in only partial digestion, leaving high molecular weight maltodextrins (DP>10, highest monitored in the method) in the SDFP fraction. Free saccharide analysis (**Figure 5.6, Supplementary Table 5.6**) further confirmed this finding. Potato starch contained small amounts of endogenous glucose, maltose, and maltooligosaccharides ranging from DP 3-10, while these free saccharides were absent in the SDF from 991.43, indicating efficient digestion in the latter. The SDFP from 2017.16, however, contained relatively large amounts of maltooligosaccharides.

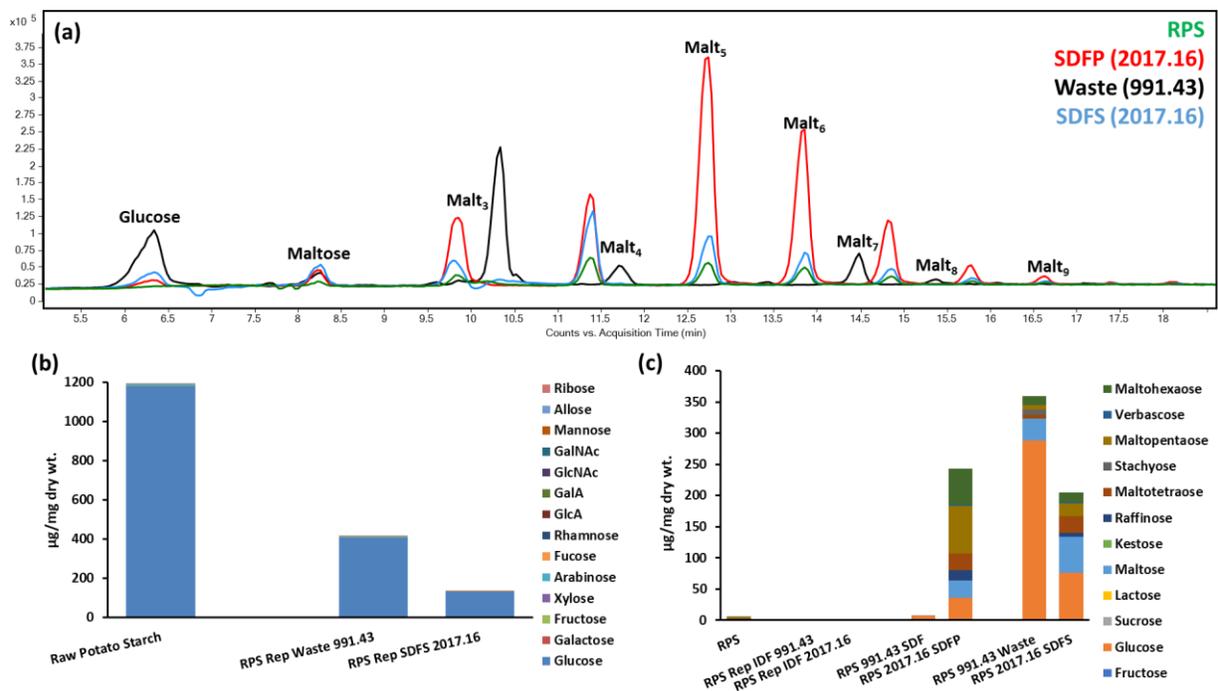


Figure 5.6a-c. (a) Total ion chromatogram depicting the free saccharides present in potato starch, SDFP (2017.16), waste (991.43) and SDFS (2017.16). (b) Total monosaccharide composition of potato starch, waste (991.43), and SDFS (2017.16). (c) Quantified free saccharide results for potato starch, SDFP, waste, and SDFS fractions.

The waste and SDFS fractions of potato starch from each method also exhibited major differences. The waste from 991.43 contained approximately three times the amount of total glucose than the SDFS fraction of 2017.16 (**Figure 5.6c**). Based on the ratio of 4- to t-glucose linkages, the average DP of the saccharides in the SDFS was greater than those in the waste fraction. Free saccharide analysis confirmed that SDFS indeed contained more maltooligosaccharides than the waste fraction. Additionally, the same maltooligosaccharides present in SDFP were found in SDFS. These compounds differed greatly from those found in the waste of 991.43. Specifically, the oligosaccharides in the fractions from 2017.16 followed the linear maltooligosaccharide ladder, while the oligosaccharides in the waste from 991.43 yielded different retention times suggesting different linkages. Indeed, closer inspection of the linkage data from the waste fraction revealed the presence of 6-glucose in addition to 4-glucose (**Figure 5.5, Supplementary Table 5.5**), suggesting that the starch digestions from each method produced distinct oligosaccharide products.

DISCUSSION

Dietary fiber is recognized as an essential component of a healthy diet.^{27, 28} Despite its importance, relatively little advancements have been made towards the analysis of fiber since its

earliest common definition nearly 50 years ago.¹ Currently, most methods for the determination of fiber are gravimetric measurements while total carbohydrates are not typically measured at all and are instead determined by difference after subtraction of moisture, protein, fat, and ash. Uncovering the structure-function relationships between dietary fiber, gut microbes, and human health is currently an active and important area of research, however current methods for fiber analysis are clearly not granular enough for this purpose.^{20, 21, 29-32} The role of dietary fibers cannot be understood without knowing structures particularly in a number of functions including modulating gut microbiome.³³⁻³⁶ The paucity of available methods for dietary fiber arises from the inherent difficulty of carbohydrate analysis as it requires a suite of historically low throughput and/or low sensitivity techniques and instrumentation.^{37, 38} However, advancements in methods utilizing rapid throughput, high sensitivity, liquid chromatography-mass spectrometry (LC-MS) have made it possible to provide quantitative structural information on the carbohydrates in hundreds of food samples rapidly using 96-well plate formats.^{20, 21, 31, 32, 39} These nascent methods make it feasible to perform in-depth carbohydrate analysis on large sample collections. Still, the output from these analyses often lack the explicit context that many in the field of nutrition and clinical science may recognize, replacing familiar terms like “total/soluble/insoluble fiber” with descriptions involving monosaccharides and glycosidic linkages. The latter descriptions have proven to be necessary details, but they can be difficult to reconcile with current dietary recommendations and classical determinations. Thus, it is important that these new approaches to carbohydrate analysis be coupled to classical methods involving isolation and gravimetric analysis of dietary fiber fractions such as IDF, SDF(P), and SDFS.

In the present work, two common foods (oats and potato starch) were analyzed on the Ankom Dietary Fiber Analyzer using two commonly employed AOAC methods: 991.43 and

2017.16. The resulting fiber fractions were recovered and subjected to a suite of rapid throughput LC-MS based methods for the comprehensive analysis of their carbohydrate contents at the monosaccharide, linkage, and free saccharide levels. This proof-of-concept study revealed an avenue to bridge the gap between classical definitions of fiber (IDF, SDF, etc.) and the higher resolution pictures required for gut microbial studies that are provided by LC-MS analyses. The resulting data can ultimately be used to make more meaningful and specific connections between dietary fiber, the gut microbiome, and host health.

From the analysis of the raw foods, oats were found to contain starch, β -glucan, arabinoxylan, mannan, and xyloglucan. Sucrose and the fructooligosaccharides stachyose and raffinose were also obtained from the same sample. As expected, potato starch was comprised solely of starch with small amounts of maltooligosaccharides. Separation of the carbohydrate fractions in the two AOAC methods 991.43 and 2017.16 followed by LC-MS analysis revealed the location of the respective components. In both AOAC methods, the IDF fraction from oats was found to be a diverse mix of polysaccharides with nearly equal abundances of arabinoxylan and β -glucan as well as smaller amounts of β -mannan and xyloglucan. Soluble oat β -glucan has garnered much attention for its role in ameliorating metabolic diseases and some cancers. However, the results suggest that β -glucan is only partially soluble with the largest component being in the soluble fractions of both methods.^{40,41} After β -glucan, arabinoxylan was also primarily in the IDF. Arabinoxylan is another important dietary fiber found in grains and has been shown to modulate the gut microbiome.^{42,43} Its fine structural detail is known to affect its function in modulation of metabolic functions.¹⁷ Less abundant components of the oats including β -mannan and xyloglucan were also detected through their specific and respective linkages (4-mannose for β -mannan and t-galactose, t-xylopyranose, and 4,6-glucose for xyloglucan) in the IDF fraction.⁴⁴ These minor

components are also known to be extensively utilized by gut microbes.^{16, 45} Both AOAC methods yielded similar IDF abundances and compositions, however the total glucose in IDF from 2017.16 (188.5 $\mu\text{g}/\text{mg}$) was higher than 991.43 (154.2 $\mu\text{g}/\text{mg}$) with statistical significance ($p = 0.005$, Student's t-test). Increased relative abundances of the starch-associated linkages 4-, 4,6-, and α -glucose in 2017.16 IDF suggested that starch was greater in the IDF of 2017.16 compared to 991.43. The former utilizes a much gentler starch digestion, and the additional starch was likely resistant starch that escaped the “softer” enzymatic digestion of 2017.16. One limitation of the methodology employed here is the inability to quantify cellulosic glucose which is likely a major component of IDF. However, this would require hydrolysis with sulfuric acid (H_2SO_4). TFA was used here to make the sample preparation more directly amenable to mass spectral analysis, but future developments could employ clean-up steps such as C18 solid phase extraction (SPE) of PMP-derivatized glycosides to remove the salts created from H_2SO_4 hydrolysis.

The SDF(P) fractions were comprised almost entirely of β -glucan. The differences in solubility between the measured β -glucans across fractions could arise from different molecular weights, linkage distributions, and their chemical interactions within the grain.^{44, 46} The ratio of 3- and 4-linked glucose to terminal glucose was significantly lower in the IDF (15.6) than the SDF(P) fractions (25), suggesting variations in degree of polymerization as a contributing factor.

Another challenging facet of defining dietary fiber arises from the concept of resistant starch (RS), which is defined analytically as the component of starch not digested after exposure to amylase and amyloglucosidase digestion. Four types of RS are defined (RS1, RS2, RS3, and RS4). RS1 is starch that is physically inaccessible to enzyme as in partially milled grains and seeds. RS2 is resistant to digestion due to its native conformation in raw foods that typically becomes digestible through cooking. RS3 is indigestible due to the process of retrogradation during which

amylose chains form double helices that resist gelatinization upon cooking. It commonly occurs in foods that were cooked and subsequently cooled. RS4 is starch that is chemically modified by cross-linking or derivatization, thereby impeding the activity of amylase on the substrate.⁴⁷ Potato starch was chosen as a model food to apply the current methodology towards understanding how RS behaves in AOAC methods 991.43 and 2017.16. As a raw and processed starch product, potato starch is expected to contain only RS2.⁴⁷ The AOAC methods used here differ greatly in the conditions used to digest starches present in foods. Starch digestion in 991.43 is carried out at elevated temperature with heat-stable amylase and is meant to hydrolyze all starch within a sample. The digestion in method 2017.16, however, is carried out under conditions mimicking physiological (37 °C, pH 7.2) and is meant to hydrolyze only digestible starch while defining undigested starch as “resistant.”⁴⁴ The biological meaning of this analytically resistant starch is a topic of some debate, but nonetheless provides a means of quantifying these components in foods.⁴⁸ By coupling these AOAC methods to the suite of presented LC-MS methods, it was determined that 991.43 was indeed effective at total starch digestion, leaving only a minute amount of bound glucose in the SDF fraction of potato starch. Most of the hydrolysate was found in the waste fraction, indicating that nearly all of the starch had been hydrolyzed to glucose, maltose, and small oligosaccharides. The small amount of bound glucose that was detected in the SDF could further be described as belonging to long chain maltodextrins as the ratio of 4-linked to terminal glucose was only slightly lower than that of the potato starch itself. This was contrary to 2017.16, where the majority of the bound glucose from potato starch was found in the SDFP rather than the SDFS. These results indicated that digestion hydrolyzed the starch into large oligosaccharides that were in turn soluble in water but insoluble in ethanol. Again, the ratio between 4-linked and terminal glucose in the linkage analysis suggested that the compounds escaping complete digestion

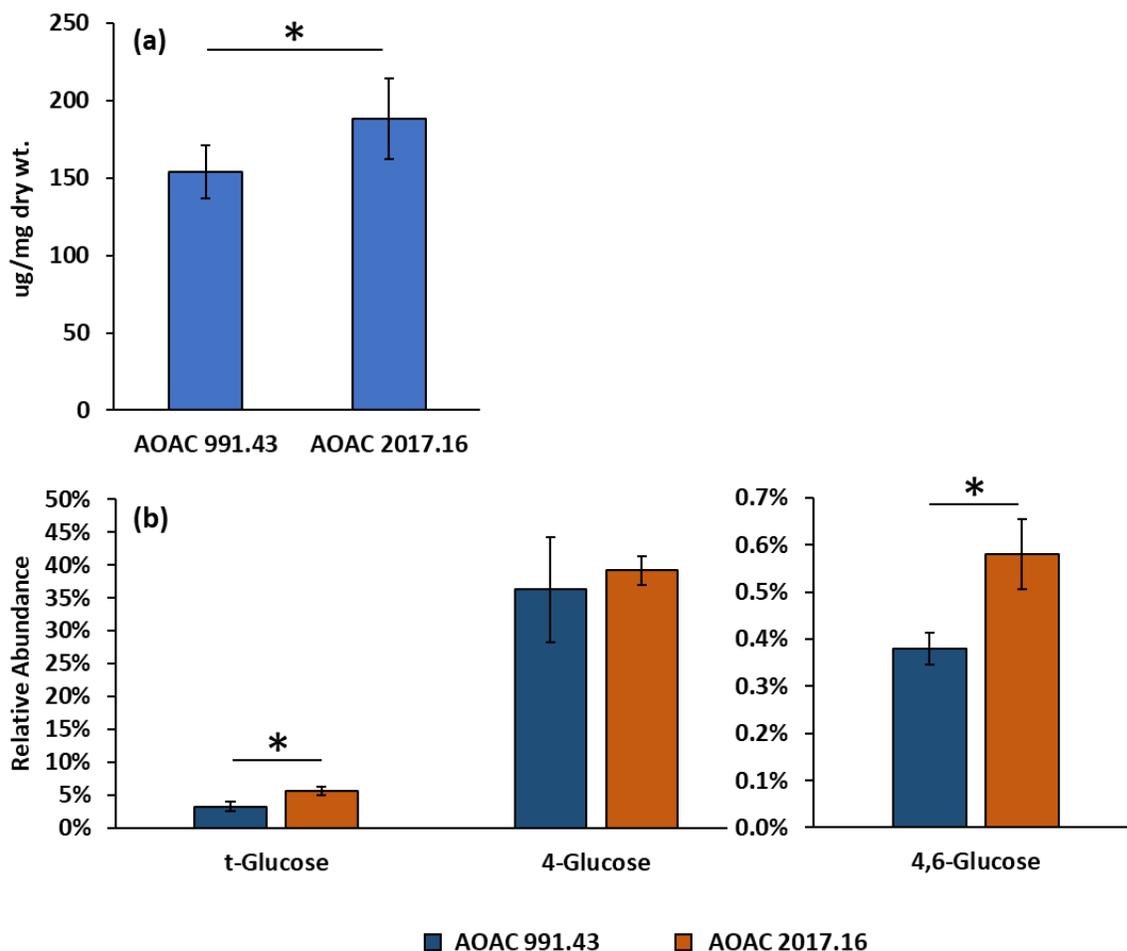
were maltodextrins. Furthermore, the free saccharide analysis revealed that a large fraction (~28%) of these maltodextrins exhibited DPs ranging from 3-10. The waste fraction from method 991.43 and SDFS fraction from 2017.16 also differed in their carbohydrate abundances and compositions. These fractions contained free glucose, maltose, and oligosaccharides resulting from the starch digestion. The waste fraction contained significantly more total carbohydrate than the SDFS fraction mostly due to increased free glucose.

CONCLUSIONS

Current methodologies for measuring dietary fiber and inclusion in food labels create a form of nutritional “dark matter” by providing only gravimetric determinations of inherently complex biomolecules. The oligo- and polysaccharide structures contained within the broad definitions of soluble and insoluble fiber vary widely between different foods, imbuing unique and structure-dependent bioactivities upon interaction with the gut microbiome. Quantifying and characterizing these structures in food will be integral in delineating the role of fiber-microbe interactions in human and animal health. The integrated methods described here provide a means to quantify the complex carbohydrate structures in food while retaining familiar and clinically relevant determinations such as “insoluble/soluble fiber.” Isolation of the dietary fiber fractions from oats and potato starch using two commonly employed AOAC methods (991.43 and 2017.16) on a commercial Fiber Analyzer allowed for their subsequent structural characterization at the monosaccharide, glycosidic linkage, and free saccharide levels using rapid-throughput LC-MS methods. The analysis of oats revealed that the non-cellulosic insoluble fiber from oats was composed of arabinoxylan and β -glucan (evidenced by 3- and 4-glucose) with small amounts of

xyloglucan and mannan while the soluble fiber fraction was chiefly composed of β -glucan and trace amounts of arabinogalactan as derived from both AOAC methods. For potato starch, the fractions obtained from each AOAC method were found to be markedly different mostly due to the nature of their starch digestions. The harsher digestion from 991.43 hydrolyzed potato starch produced mainly free glucose and maltose that were collected in the “waste” fraction. The milder digestion from 2017.16, meant to capture resistant starch, was found to hydrolyze potato starch mostly to higher molecular weight maltodextrins that were captured in the SDFP fraction, as well as glucose, maltose, and maltooligosaccharides collected in the SDFS fraction. Together, these findings provide new and comprehensive insight regarding the structures present as dietary fiber in these food products. In the future, LC-MS analyses that quantify and structurally define the glycans within dietary fiber will supersede existing gravimetric methods to provide the details necessary to understand the positive health effects of fiber. This proof-of-concept study will serve as a prelude towards more complex and diverse foods leading to more thorough understanding of dietary fiber and its effect on the gut microbiome. We propose that food labeling requirements include knowledge of specific carbohydrates and that the information be included in AOAC and CODEX tables. Both the LC-MS and McCleary method should be combined in future food analysis.

SUPPLEMENTARY MATERIALS



Supplementary Figure 5.1a-b. (a) Difference in total glucose content of the IDF fractions of oat obtained by AOAC 991.43 and 2017.16. (b) Differences in starch-associated glucose linkages of the IDF fractions of oat obtained by both AOAC methods. Asterisks denote statistical significance with a p-value < 0.05 by Student's t-test.

Supplementary Table 4.1. Monosaccharide compositions of oat and its dietary fiber fractions ($\mu\text{g}/\text{mg}$ dry weight). Values represent the mean of 3 technical replicates \pm the standard deviation (N.D. = not detected).

	Glc	Gal	Fru	Xyl	Ara	Fuc	Rha	GlcA	GalA	Man	Rib
Raw Oat	577 \pm 14.8	5.8 \pm 0.4	2.9 \pm 0.2	16.4 \pm 3	17.4 \pm 1.7	N.D.	<i>trace</i>	N.D.	N.D.	1.1 \pm 0.2	0.6 \pm 0
Oat IDF 991.43	154.2 \pm 17.1	6.7 \pm 1.4	0.9 \pm 0.2	78.5 \pm 23.3	67.5 \pm 18.4	0.1 \pm 0.2	0.3 \pm 0.1	N.D.	0.6 \pm 0.1	3.7 \pm 1	1.2 \pm 0.3
Oat SDF 991.43	275.7 \pm 38.2	7.5 \pm 1	1.2 \pm 0.2	7.6 \pm 0.7	12.1 \pm 2	N.D.	<i>trace</i>	N.D.	N.D.	9.7 \pm 1.6	1.3 \pm 0.2
Oat SDFP 991.43	252 \pm 27.8	5.3 \pm 0.9	1.3 \pm 0.2	29.7 \pm 0.6	29.3 \pm 1.1	N.D.	<i>trace</i>	N.D.	N.D.	3.7 \pm 3.4	0.8 \pm 0.1
Oat Waste 991.43	247.3 \pm 21	3.9 \pm 0.5	1.6 \pm 0.2	N.D.	1.9 \pm 0.3	N.D.	<i>trace</i>	N.D.	N.D.	0.5 \pm 0.1	<i>trace</i>
Arabinoxylan Quality Control	5.6 \pm 1.3	21 \pm 1.2	0.7 \pm 0.3	575.5 \pm 16.3	405.3 \pm 20.1	N.D.	0.4 \pm 0.1	N.D.	N.D.	2.4 \pm 0.3	0.9 \pm 0.1

Supplementary Table 4.2. Linkage compositions of oat and its dietary fiber fractions (% relative composition). Values represent the mean of 3 technical replicates \pm the standard deviation (N.D. = not detected). Raw oat was analyzed singly.

	Raw Oat	Oat IDF 991.43	Oat 1 IDF 2017.16	Oat SDF 991.43	Oat SDFP 2017.16	Oat Waste 991.43	Oat SDFS 2017.16	Arabinoxylan Quality Control
t-Glc	13.19	3.2 \pm 0.01	5.55 \pm 0.01	2.29 \pm 0	7.74 \pm 0.01	68.45 \pm 0.01	68.01 \pm 0	2.96 \pm 0.03
4-Glc	65.98	36.26 \pm 0.08	39.18 \pm 0.02	43.17 \pm 0.03	32.7 \pm 0.02	9.34 \pm 0	9.32 \pm 0.01	2.87 \pm 0.05
6-Glc	0.4	0.05 \pm 0	0.05 \pm 0	0.03 \pm 0	0.13 \pm 0	2.1 \pm 0	0.8 \pm 0	0.07 \pm 0
3-Glc/3-Gal	3.51	14.18 \pm 0.02	12.41 \pm 0.01	18.06 \pm 0.01	11.42 \pm 0.02	0.81 \pm 0	0.81 \pm 0	0.08 \pm 0
2-Glc	0.2	0.08 \pm 0	0.11 \pm 0	0.05 \pm 0	0.1 \pm 0	0.35 \pm 0	0.32 \pm 0	0.01 \pm 0
4,6-Glc	1.98	0.38 \pm 0	0.58 \pm 0	0.34 \pm 0	0.38 \pm 0	0.23 \pm 0	0.07 \pm 0	0.01 \pm 0
3,4-Glc	4.45	0.32 \pm 0	0.79 \pm 0	0.34 \pm 0	0.24 \pm 0	0.06 \pm 0	0.07 \pm 0	0.01 \pm 0
2,4-Glc	0.42	0.29 \pm 0	0.43 \pm 0	0.35 \pm 0	0.2 \pm 0	0.05 \pm 0	0.05 \pm 0	0.01 \pm 0
2,4,6-Glc	0.01	0.02 \pm 0	0.02 \pm 0	0.02 \pm 0	0.01 \pm 0	N.D.	N.D.	0 \pm 0
2,3,6-Glc	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
3,4,6-Glc	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

	Raw Oat	Oat IDF 991.43	Oat 1 IDF 2017.16	Oat SDF 991.43	Oat SDFP 2017.16	Oat Waste 991.43	Oat SDFS 2017.16	Arabinoxylan Quality Control
t-Gal	1.7	2.01±0.01	1.8±0	0.91±0	1.24±0	3.65±0	3.86±0	1.05±0
6-Gal	0.41	0.22±0	0.25±0	0.45±0	0.81±0	0.67±0	0.73±0	0.17±0
4-Gal	0.08	0.55±0	0.37±0	0.11±0	0.05±0	0.01±0	0.01±0	0.15±0
2-Gal	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
4,6-Gal	0.38	0.3±0	0.3±0	2.53±0	2.99±0	0.05±0	0.03±0	0.5±0
3,6-Gal	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
3,4-Gal	N.D.	0.03±0	0.03±0	0.03±0	0.02±0	N.D.	0.02±0	N.D.
3,4,6-Gal	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2,4,6-Gal	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
t-Fru	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
t-p-Xyl	0.41	3.47±0.01	3.19±0.01	1.19±0	0.76±0	N.D.	N.D.	3.36±0.01
4-p-Xyl	0.51	6.68±0.02	5.97±0	1.39±0	0.94±0	0.06±0	0.08±0	25.2±0.07
2-p-Xyl	0.02	0.26±0	0.21±0	0.05±0	0.04±0	0.01±0	0.01±0	0.19±0
3,4-p-Xyl/3,5-Ara	0.27	4.86±0.02	4.12±0.01	0.51±0	0.35±0	0.01±0	0.01±0	7.99±0.02
2,4-p-Xyl	0.02	0.4±0	0.31±0	0.08±0	0.05±0	N.D.	N.D.	0.72±0
t-p-Ara	N.D.	N.D.	0.23±0	0±0	0.41±0	1.02±0.01	0.36±0	N.D.
t-f-Ara	2.8	17.48±0.05	15.57±0.01	10.48±0.02	8.45±0	0±0	0±0	49.89±0.14
5-f-Ara	0.35	1.08±0	0.85±0	1.35±0	1.53±0	0.05±0	0.03±0	1.05±0
3-f-Ara	0.09	0.96±0	0.79±0	0.17±0	0.14±0	0.01±0	0.01±0	0.7±0
2-f-Ara	0.12	1.51±0	1.28±0	0.46±0	0.33±0	0.01±0	0.01±0	1.07±0
2,5-f-Ara	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2,3-f-Ara	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
t-Rha	0.03	0.07±0	0.07±0	0.06±0	0.05±0	0.19±0	0.22±0	0.02±0
4-Rha	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2-Rha	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
t-Fuc	0.03	0.21±0	0.21±0	0.05±0	0.17±0	0.02±0	0.04±0	0.05±0
t-GlcA	0.01	0.05±0	0.04±0	0.01±0	0.01±0	N.D.	N.D.	0.04±0

	Raw Oat	Oat IDF 991.43	Oat 1 IDF 2017.16	Oat SDF 991.43	Oat SDFP 2017.16	Oat Waste 991.43	Oat SDFS 2017.16	Arabinoxylan Quality Control
t-GalA	0.03	0.19±0	0.18±0	0.13±0	0.11±0	0.01±0	0.01±0	0.58±0
t-Man	0.39	0.77±0	1.06±0	11.95±0	24.89±0.03	12±0	14.68±0.01	0.73±0
6-Man	ND.	ND.	ND.	0.04±0	0.13±0	0.02±0	0.01±0	N.D.
4-Man	1.9	3.38±0.01	3.28±0.01	1.63±0	1.43±0	0.66±0	0.32±0	0.11±0
3-Man	N.D.	N.D.	N.D.	0.19±0	0.24±0	N.D.	0.06±0	N.D.
2-Man	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
4,6-Man	0.01	0.01±0	0.02±0	N.D.	N.D.	N.D.	N.D.	N.D.
3,4,6-Man	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
X-Hex	0.28	0.69±0	0.7±0	1.39±0	1.58±0	0.05±0	0.04±0	0.39±0
2-Hex	N.D.	0.01±0	0.01±0	0.18±0	0.36±0	N.D.	0.03±0	0.01±0
X,X-Hex	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2,X-Hex	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
2,X,X-Hex (I)	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

Hex = Unidentified Hexose

Supplementary Figure 4.3. Free saccharide composition of oat and its dietary fiber fractions ($\mu\text{g}/\text{mg}$ dry wt.). Values represent the mean of 3 technical replicates \pm the standard deviation (N.D. = not detected). Raw oat at was measured singly.

	Fru	Glc	Suc	Lac	Malt	Kes	Raf	Malt₄	Stach	Malt₅	Verb	Malt₆
Raw Oat	N.D.	0.8	6.4	N.D.	N.D.	N.D.	1.9	0.7	6	0.7	1.8	1.1
Oat 991.43 SDF	N.D.	4±0.4	N.D.	N.D.	N.D.	N.D.	N.D.	3.5±0.2	2.8±0.2	3.5±0.4	1±0.3	3.1±0.5
Oat 2017.16 SDFP	N.D.	21.9±0.5	N.D.	N.D.	8.8±2.2	N.D.	0.4±0.2	4.5±1	3.4±0.4	6.8±1.9	N.D.	7±1.1
Oat 991.43 Waste	N.D.	245.8±20.5	28.5±3.5	N.D.	26.8±2	N.D.	1.2±0.1	N.D.	10.8±0.6	5.7±0.5	1.9±0.4	11.2±0.1
Oat 2017.16 SDFS	N.D.	215.8±7.3	18.5±0.5	N.D.	10.2±1	N.D.	1.3±0.2	6.1±0.3	6.5±0.1	7.3±1.1	1.8±0	5.3±0.7

Suc = Sucrose, Lac = Lactose, Malt = Maltose, Kes = Kestose, Raf = Raffinose, Malt₄ = Maltotetraose, Stach = Stachyose, Malt₅ = Maltopentaose, Verb = Verbascose, Malt₆ = Maltohexaose

Supplementary Table 4.4. Monosaccharide compositions of raw potato starch (RPS) and its dietary fiber fractions ($\mu\text{g}/\text{mg}$ dry weight). Values represent the mean of 3 technical replicates \pm the standard deviation (N.D. = not detected).

	Glc	Gal	Fru	Xyl	Ara	Fuc	Rha	GlcA	GalA	GlcNAc	GalNAc	Man	All	Rib
Raw Potato Starch	1183 \pm 41	N.D.	<i>trace</i>	N.D.	<i>trace</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
RPS SDF 991.43	9 \pm 2	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	3 \pm 1	N.D.	N.D.
RPS SDFP 2017.16	863 \pm 31	N.D.	<i>trace</i>	N.D.	4 \pm 1	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	7 \pm 0	N.D.	N.D.
RPS Waste 991.43	409 \pm 51	N.D.	<i>trace</i>	N.D.	<i>trace</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
RPS SDFS 2017.16	133 \pm 16	N.D.	<i>trace</i>	N.D.	<i>trace</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	3 \pm 0	N.D.	N.D.
Arabinoxylan Quality Control	5 \pm 0	20 \pm 1	N.D.	521 \pm 19	383 \pm 13	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	2 \pm 0	N.D.	<i>trace</i>

Supplementary Table 4.5. Linkage compositions of raw potato starch (RPS) and its dietary fiber fractions (% relative composition). Values represent the mean of 3 technical replicates \pm the standard deviation (N.D. = not detected). Raw oat was analyzed singly.

	Raw Potato Starch	RPS SDF 991.43	RPS SDFP 2017.16	RPS Waste 991.43	RPS SDFS 2017.16
t-Glc	7.75	10.53 \pm 0.68	26.86 \pm 0.6	68.11 \pm 3.36	37.79 \pm 3.54
4-Glc	89.33	68.75 \pm 2.72	63.6 \pm 1.6	12.96 \pm 1.41	53.12 \pm 3.81
6-Glc	0.03	0.13 \pm 0.01	0.05 \pm 0	2.17 \pm 0.12	0.08 \pm 0.02
3-Glc/3-Gal	N.D.	N.D.	N.D.	1.07 \pm 0.88	0.14 \pm 0.03

	Raw Potato Starch	RPS SDF 991.43	RPS SDFP 2017.16	RPS Waste 991.43	RPS SDFS 2017.16
2-Glc	0.08	0.07±0.02	0.17±0.04	0.59±0.49	0.23±0.07
4,6-Glc	0.49	0.31±0.03	1.73±0.16	0.47±0.07	0.11±0.01
3,4-Glc	0.49	0.26±0.11	0.85±0.2	0.09±0.05	0.32±0.1
2,4-Glc	0.07	0.04±0.01	0.15±0.02	0.02±0.01	0.08±0.01
2,4,6-Glc	N.D.	N.D.	N.D.	N.D.	N.D.
2,3,6-Glc	N.D.	N.D.	N.D.	N.D.	N.D.
3,4,6-Glc	N.D.	N.D.	N.D.	N.D.	N.D.
t-Gal	0.03	0.16±0.03	0.02±0.01	0.32±0.04	0.07±0.02
6-Gal	N.D.	0.01±0	N.D.	N.D.	N.D.
4-Gal	N.D.	0.09±0.01	N.D.	0.01±0	N.D.
2-Gal	N.D.	N.D.	N.D.	N.D.	N.D.
4,6-Gal	N.D.	0.01±0	0.51±0.46	N.D.	N.D.
3,6-Gal	N.D.	N.D.	N.D.	N.D.	0.04±0.06
3,4-Gal	N.D.	0.02±0	N.D.	N.D.	0.02±0
3,4,6-Gal	N.D.	N.D.	N.D.	N.D.	N.D.
2,4,6-Gal	N.D.	N.D.	N.D.	N.D.	N.D.
t-Fru	N.D.	N.D.	N.D.	N.D.	N.D.
t-p-Xyl	N.D.	N.D.	N.D.	N.D.	N.D.
4-p-Xyl	0.19	0.26±0.25	0.03±0.02	0.08±0.1	0.15±0.21
2-p-Xyl	N.D.	0.01±0	N.D.	N.D.	N.D.
3,4-p-Xyl/3,5-Ara	N.D.	N.D.	N.D.	N.D.	0±0.01
2,4-p-Xyl	N.D.	N.D.	N.D.	N.D.	0±0.01
t-p-Ara	N.D.	1.79±1.36	0.37±0.29	0.01±0.02	N.D.
t-f-Ara	0.53	0.28±0.3	N.D.	N.D.	N.D.
5-f-Ara	N.D.	0.01±0	0.02±0.03	N.D.	0.13±0.22
3-f-Ara	0.01	0.01±0	0.01±0	0.01±0	0.01±0
2-f-Ara	N.D.	0±0.01	N.D.	N.D.	0.01±0
2,5-f-Ara	N.D.	N.D.	N.D.	N.D.	N.D.
2,3-f-Ara	N.D.	N.D.	N.D.	N.D.	N.D.

	Raw Potato Starch	RPS SDF 991.43	RPS SDFP 2017.16	RPS Waste 991.43	RPS SDFS 2017.16
t-Rha	0.01	0.02±0	N.D.	0.01±0	N.D.
4-Rha	N.D.	N.D.	N.D.	N.D.	N.D.
2-Rha	N.D.	N.D.	N.D.	N.D.	N.D.
t-Fuc	0.01	0.02±0	0.02±0	N.D.	0.02±0
t-GlcA	N.D.	N.D.	N.D.	N.D.	N.D.
t-GalA	0.01	N.D.	N.D.	N.D.	N.D.
t-Man	0.14	16.06±2.15	2.75±0.5	13.34±2.4	5.91±0.59
6-Man	N.D.	0.09±0.01	0.01±0	0.02±0	0±0
4-Man	0.83	0.6±0.04	1.96±0.25	0.69±0.11	1.75±0.09
3-Man	N.D.	0.19±0.03	N.D.	N.D.	N.D.
2-Man	N.D.	N.D.	N.D.	N.D.	N.D.
4,6-Man	N.D.	0.01±0	0.01±0	0.01±0	N.D.
3,4,6-Man	N.D.	N.D.	N.D.	N.D.	N.D.
X-Hex	N.D.	0.03±0.02	N.D.	N.D.	N.D.
2-Hex	N.D.	0.21±0.03	0.02±0	0±0	0.03±0.01
X,X-Hex	N.D.	N.D.	N.D.	N.D.	N.D.
2,X-Hex	N.D.	0.01±0	N.D.	N.D.	N.D.
2,X,X-Hex (I)	N.D.	N.D.	N.D.	N.D.	N.D.

Hex = Unidentified Hexose

Supplementary Figure 4.6. Free saccharide composition of RPS and its dietary fiber fractions ($\mu\text{g}/\text{mg}$ dry wt.). Values represent the mean of 3 technical replicates \pm the standard deviation (N.D. = not detected). RPS was measured singly.

	Fru	Glc	Suc	Lac	Malt	Kes	Raf	Malt₄	Stach	Malt₅	Verb	Malt₆
RPS	N.D.	0.4	N.D.	N.D.	1.2	N.D.	N.D.	1.5	N.D.	1.3	N.D.	1.2
RPS 991.43 SDF	N.D.	7 \pm 0.4	0.2 \pm 0.4	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
RPS 2017.16 SDFP	N.D.	35.4 \pm 8.7	N.D.	N.D.	28.7 \pm 9.8	N.D.	15.7 \pm 8	27.2 \pm 8.3	N.D.	76.2 \pm 31.5	2.2 \pm 0.5	57.2 \pm 23.7
RPS 991.43 Waste	N.D.	288.1 \pm 93.3	N.D.	N.D.	35.4 \pm 8.5	N.D.	N.D.	5.9 \pm 0.2	8.7 \pm 1.6	6.6 \pm 0.7	N.D.	15.1 \pm 3.7
RPS 2017.16 SDFS	N.D.	75.8 \pm 11	N.D.	N.D.	58.3 \pm 0.5	N.D.	5.8 \pm 0.8	26.8 \pm 4.8	N.D.	20.8 \pm 5.9	1.5 \pm 0.1	16.1 \pm 4.3

Suc = Sucrose, Lac = Lactose, Malt = Maltose, Kes = Kestose, Raf = Raffinose, Malt₄ = Maltotetraose, Stach = Stachyose, Malt₅ = Maltopentaose, Verb = Verbascose, Malt₆ = Maltohexaose

REFERENCES

1. Hipsley, E. H., Dietary "fibre" and pregnancy toxemia. *Br Med J* **1953**, 2 (4833), 420-2.
2. Trowell, H., The development of the concept of dietary fiber in human nutrition. *Am J Clin Nutr* **1978**, 31 (10 Suppl), S3-S11.
3. Trowell, H., Food and dietary fibre. *Nutr Rev* **1977**, 35 (3), 6-11.
4. Trowell, H., Ischemic heart disease and dietary fiber. *Am J Clin Nutr* **1972**, 25 (9), 926-32.
5. Cummings, J. H.; Englyst, A., Denis Burkitt and the origins of the dietary fibre hypothesis. *Nutr Res Rev* **2018**, 31 (1), 1-15.
6. Gill, S. K.; Rossi, M.; Bajka, B.; Whelan, K., Dietary fibre in gastrointestinal health and disease. *Nat Rev Gastroenterol Hepatol* **2021**, 18 (2), 101-116.
7. Hellendoorn, E. W.; Noordhoff, M. G.; Slagman, J., Enzymatic Determination of Indigestible Residue (Dietary Fiber) Content of Human Food. *J Sci Food Agr* **1975**, 26 (10), 1461-1468.
8. Dhingra, D.; Michael, M.; Rajput, H.; Patil, R. T., Dietary fibre in foods: a review. *J Food Sci Technol* **2012**, 49 (3), 255-66.
9. Kasahara, K.; Rey, F. E., The emerging role of gut microbial metabolism on cardiovascular disease. *Curr Opin Microbiol* **2019**, 50, 64-70.
10. Cani, P. D., Microbiota and metabolites in metabolic diseases. *Nat Rev Endocrinol* **2019**, 15 (2), 69-70.
11. Donia, M. S.; Fischbach, M. A., HUMAN MICROBIOTA. Small molecules from the human microbiota. *Science* **2015**, 349 (6246), 1254766.
12. Sonnenburg, J. L.; Backhed, F., Diet-microbiota interactions as moderators of human metabolism. *Nature* **2016**, 535 (7610), 56-64.
13. Martens, E. C.; Kelly, A. G.; Tauzin, A. S.; Brumer, H., The Devil Lies in the Details: How Variations in Polysaccharide Fine-Structure Impact the Physiology and Evolution of Gut Microbes. *J Mol Biol* **2014**, 426 (23), 3851-3865.
14. Hamaker, B. R.; Tuncil, Y. E., A Perspective on the Complexity of Dietary Fiber Structures and Their Potential Effect on the Gut Microbiota. *J Mol Biol* **2014**, 426 (23), 3838-3850.
15. Zhang, M. L.; Chekan, J. R.; Dodd, D.; Hong, P. Y.; Radlinski, L.; Revindran, V.; Nair, S. K.; Mackie, R. I.; Cann, I., Xylan utilization in human gut commensal bacteria is orchestrated by unique modular organization of polysaccharide-degrading enzymes. *P Natl Acad Sci USA* **2014**, 111 (35), E3708-E3717.
16. Tauzin, A. S.; Kwiatkowski, K. J.; Orlovsky, N. I.; Smith, C. J.; Creagh, A. L.; Haynes, C. A.; Wawrzak, Z.; Brumer, H.; Koropatkin, N. M., Molecular Dissection of Xyloglucan Recognition in a Prominent Human Gut Symbiont. *Mbio* **2016**, 7 (2).
17. Tuncil, Y. E.; Thakkar, R. D.; Arioglu-Tuncil, S.; Hamaker, B. R.; Lindemann, S. R., Subtle Variations in Dietary-Fiber Fine Structure Differentially Influence the Composition and Metabolic Function of Gut Microbiota. *Mosphere* **2020**, 5 (3).
18. Deehan, E. C.; Yang, C.; Perez-Munoz, M. E.; Nguyen, N. K.; Cheng, C. C.; Triador, L.; Zhang, Z. X.; Bakal, J. A.; Walter, J., Precision Microbiome Modulation with Discrete Dietary Fiber Structures Directs Short-Chain Fatty Acid Production. *Cell Host & Microbe* **2020**, 27 (3), 389-+.

19. Carmody, R. N.; Bisanz, J. E.; Bowen, B. P.; Maurice, C. F.; Lyalina, S.; Louie, K. B.; Treen, D.; Chadaideh, K. S.; Rekdal, V. M.; Bess, E. N.; Spanogiannopoulos, P.; Ang, Q. Y.; Bauer, K. C.; Balon, T. W.; Pollard, K. S.; Northen, T. R.; Turnbaugh, P. J., Cooking shapes the structure and function of the gut microbiome. *Nat Microbiol* **2019**, *4* (12), 2052-2063.
20. Delannoy-Bruno, O.; Desai, C.; Raman, A. S.; Chen, R. Y.; Hibberd, M. C.; Cheng, J. Y.; Han, N.; Castillo, J. J.; Couture, G.; Lebrilla, C. B.; Barve, R. A.; Lombard, V.; Henrissat, B.; Leyn, S. A.; Rodionov, D. A.; Osterman, A. L.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Kirbach, K.; Wilmot, T.; Heath, A. C.; Klein, S.; Barratt, M. J.; Gordon, J. I., Evaluating microbiome-directed fibre snacks in gnotobiotic mice and humans. *Nature* **2021**, *595* (7865), 91-+.
21. Delannoy-Bruno, O.; Desai, C.; Castillo, J. J.; Couture, G.; Barve, R. A.; Lombard, V.; Henrissat, B.; Cheng, J.; Han, N.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Lebrilla, C. B.; Marion, S.; Heath, A. C.; Barratt, M. J.; Gordon, J. I., An approach for evaluating the effects of dietary fiber polysaccharides on the human gut microbiome and plasma proteome. *Proc Natl Acad Sci U S A* **2022**, *119* (20), e2123411119.
22. Amicucci, M. J.; Galermo, A. G.; Nandita, E.; Vo, T. T. T.; Liu, Y. Y.; Lee, M.; Xu, G. G.; Lebrilla, C. B., A rapid-throughput adaptable method for determining the monosaccharide composition of polysaccharides. *International Journal of Mass Spectrometry* **2019**, *438*, 22-28.
23. Galermo, A. G.; Nandita, E.; Barboza, M.; Arnicucci, M. J.; Vo, T. T. T.; Lebrilla, C. B., Liquid Chromatography-Tandem Mass Spectrometry Approach for Determining Glycosidic Linkages. *Anal Chem* **2018**, *90* (21), 13073-13080.
24. Galermo, A. G.; Nandita, E.; Castillo, J. J.; Amicucci, M. J.; Lebrilla, C. B., Development of an Extensive Linkage Library for Characterization of Carbohydrates. *Anal Chem* **2019**, *91* (20), 13022-13031.
25. Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A., Determining the polysaccharide composition of plant cell walls. *Nat Protoc* **2012**, *7* (9), 1590-607.
26. Wang, Q.; Ellis, P. R., Oat beta-glucan: physico-chemical characteristics in relation to its blood-glucose and cholesterol-lowering properties. *Br J Nutr* **2014**, *112 Suppl 2*, S4-S13.
27. Anderson, J. W.; Baird, P.; Davis, R. H.; Ferreri, S.; Knudtson, M.; Koraym, A.; Waters, V.; Williams, C. L., Health benefits of dietary fiber. *Nutrition Reviews* **2009**, *67* (4), 188-205.
28. Wastyk, H. C.; Fragiadakis, G. K.; Perelman, D.; Dahan, D.; Merrill, B. D.; Yu, F. Q. B.; Topf, M.; Gonzalez, C. G.; Van Treuren, W.; Han, S.; Robinson, J. L.; Elias, J. E.; Sonnenburg, E. D.; Gardner, C. D.; Sonnenburg, J. L., Gut-microbiota-targeted diets modulate human immune status. *Cell* **2021**, *184* (16), 4137-+.
29. Romero Marcia, A. D.; Yao, T.; Chen, M. H.; Oles, R. E.; Lindemann, S. R., Fine Carbohydrate Structure of Dietary Resistant Glucans Governs the Structure and Function of Human Gut Microbiota. *Nutrients* **2021**, *13* (9).
30. Yao, T.; Chen, M. H.; Lindemann, S. R., Structurally complex carbohydrates maintain diversity in gut-derived microbial consortia under high dilution pressure. *FEMS Microbiol Ecol* **2020**, *96* (9).
31. Patnode, M. L.; Beller, Z. W.; Han, N. D.; Cheng, J.; Peters, S. L.; Terrapon, N.; Henrissat, B.; Le Gall, S.; Saulnier, L.; Hayashi, D. K.; Meynier, A.; Vinoy, S.; Giannone, R. J.; Hettich, R. L.; Gordon, J. I., Interspecies Competition Impacts Targeted Manipulation of Human Gut Bacteria by Fiber-Derived Glycans. *Cell* **2019**, *179* (1), 59-73 e13.

32. Patnode, M. L.; Guruge, J. L.; Castillo, J. J.; Couture, G. A.; Lombard, V.; Terrapon, N.; Henrissat, B.; Lebrilla, C. B.; Gordon, J. I., Strain-level functional variation in the human gut microbiota based on bacterial binding to artificial food particles. *Cell Host Microbe* **2021**, *29* (4), 664-673 e5.
33. Cronin, P.; Joyce, S. A.; O'Toole, P. W.; O'Connor, E. M., Dietary Fibre Modulates the Gut Microbiota. *Nutrients* **2021**, *13* (5).
34. Calatayud, M.; Van den Abbeele, P.; Ghyselinck, J.; Marzorati, M.; Rohs, E.; Birkett, A., Comparative Effect of 22 Dietary Sources of Fiber on Gut Microbiota of Healthy Humans in vitro. *Front Nutr* **2021**, *8*.
35. Myhrstad, M. C. W.; Tunsjo, H.; Charnock, C.; Telle-Hansen, V. H., Dietary Fiber, Gut Microbiota, and Metabolic Regulation-Current Status in Human Randomized Trials. *Nutrients* **2020**, *12* (3).
36. Oliver, A.; Chase, A. B.; Weihe, C.; Orchanian, S. B.; Riedel, S. F.; Hendrickson, C. L.; Lay, M.; Sewall, J. M.; Martiny, J. B. H.; Whiteson, K., High-Fiber, Whole-Food Dietary Intervention Alters the Human Gut Microbiome but Not Fecal Short-Chain Fatty Acids. *Msystems* **2021**, *6* (2).
37. Pettolino, F. A.; Walsh, C.; Fincher, G. B.; Bacic, A., Determining the polysaccharide composition of plant cell walls. *Nat Protoc* **2012**, *7* (9), 1590-1607.
38. Zhao, W. C.; Fernando, L. D.; Kirui, A.; Deligey, F.; Wang, T., Solid-state NMR of plant and fungal cell walls: A critical review. *Solid State Nucl Mag* **2020**, *107*.
39. Castillo, J. J.; Couture, G.; Bacalzo, N. P.; Chen, Y.; Chin, E. L.; Blecksmith, S. E.; Bouzid, Y. Y.; Vainberg, Y.; Masarweh, C.; Zhou, Q. W.; Smilowitz, J. T.; German, J. B.; Mills, D. A.; Lemay, D. G.; Lebrilla, C. B., The Development of the Davis Food Glycopedia-A Glycan Encyclopedia of Food. *Nutrients* **2022**, *14* (8).
40. El Khoury, D.; Cuda, C.; Luhovyy, B. L.; Anderson, G. H., Beta Glucan: Health Benefits in Obesity and Metabolic Syndrome. *J Nutr Metab* **2012**, *2012*.
41. Daou, C.; Zhang, H., Oat Beta-Glucan: Its Role in Health Promotion and Prevention of Diseases. *Compr Rev Food Sci F* **2012**, *11* (4), 355-365.
42. Mendis, M.; Leclerc, E.; Simsek, S., Arabinoxylans, gut microbiota and immunity. *Carbohydr Polym* **2016**, *139*, 159-166.
43. Pereira, G. V.; Abdel-Hamid, A. M.; Dutta, S.; D'Alessandro-Gabazza, C. N.; Wefers, D.; Farris, J. A.; Bajaj, S.; Wawrzak, Z.; Atomi, H.; Mackie, R. I.; Gabazza, E. C.; Shukla, D.; Koropatkin, N. M.; Cann, I., Degradation of complex arabinoxylans by human colonic Bacteroidetes. *Nature Communications* **2021**, *12* (1).
44. McCleary, B. V.; Cox, J.; Ry, R. I.; Delaney, E., Definition and Analysis of Dietary Fiber in Grain Products. *Food Chem Funct Anal* **2019**, *6*, 103-126.
45. La Rosa, S. L.; Leth, M. L.; Michalak, L.; Hansen, M. E.; Pudlo, N. A.; Glowacki, R.; Pereira, G.; Workman, C. T.; Arntzen, M. O.; Pope, P. B.; Martens, E. C.; Abou Hachem, M.; Westereng, B., The human gut Firmicute *Roseburia intestinalis* is a primary degrader of dietary beta-mannans. *Nature Communications* **2019**, *10*.
46. Du, B.; Meenu, M.; Liu, H. Z.; Xu, B. J., A Concise Review on the Molecular Structure and Function Relationship of beta-Glucan. *Int J Mol Sci* **2019**, *20* (16).
47. Sajilata, M. G.; Singhal, R. S.; Kulkarni, P. R., Resistant starch - A review. *Compr Rev Food Sci F* **2006**, *5* (1), 1-17.

48. Perera, A.; Meda, V.; Tyler, R. T., Resistant starch A review of analytical protocols for determining resistant starch and of factors affecting the resistant starch content of foods. *Food Res Int* **2010**, *43* (8), 1959-1974.