

На правах рукописи

Чаплин
Андрей Викторович

**Сравнительная геномика штаммов *Bifidobacterium longum*,
выделенных из кишечника здоровых людей**

03.02.03 – микробиология

Автореферат
диссертации на соискание ученой степени
кандидата медицинских наук

Москва - 2016

Работа выполнена в Государственном бюджетном образовательном учреждении высшего профессионального образования «Российский национальный исследовательский медицинский университет имени Н.И. Пирогова» Минздрава России

Научный руководитель:

доктор медицинских наук, профессор

Ефимов Борис Алексеевич

Официальные оппоненты:

Припутневич Татьяна Валерьевна – доктор медицинских наук, заведующая отделом микробиологии и клинической фармакологии Федерального государственного бюджетного учреждения «Научный центр акушерства, гинекологии и перинатологии имени академика В.И. Кулакова» Министерства здравоохранения Российской Федерации

Николаева Елена Николаевна – доктор медицинских наук, профессор кафедры микробиологии, вирусологии, иммунологии государственного бюджетного образовательного учреждения высшего профессионального образования «Московский государственный медико-стоматологический университет им. А.И. Евдокимова» Министерства здравоохранения Российской Федерации

Ведущая организация:

Государственное бюджетное образовательное учреждение высшего профессионального образования Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства здравоохранения Российской Федерации

Защита диссертации состоится «__» _____ 20__ года в __ часов на заседании диссертационного совета Д 208.046.01 при Федеральном бюджетном учреждении науки «Московский научно-исследовательский институт эпидемиологии и микробиологии им. Н.Н. Габричевского» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека по адресу: 125212, Москва, ул. Адмирала Макарова, 10. С диссертацией можно ознакомиться в библиотеке и на сайте Федерального бюджетного учреждения науки «Московский научно-исследовательский институт эпидемиологии и микробиологии им. Н.Н. Габричевского» Федеральной службы по надзору в сфере защиты прав потребителей и благополучия человека по адресу: 125212, г. Москва, ул. Адмирала Макарова, 10, <http://www.gabrich.ru>

Автореферат разослан «__» _____ 2015 г.

Ученый секретарь диссертационного совета
доктор медицинских наук

Борисова Ольга Юрьевна

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность исследования:

Микробное сообщество кишечника человека является сложной системой, включающей сотни видов микроорганизмов, которая участвует во множестве взаимодействий с человеческим организмом, в том числе, в формировании колонизационной резистентности и модулировании местного иммунитета. Микрофлора характеризуется чрезвычайным разнообразием, находясь в зависимости от возраста, стиля жизни и индивидуальных особенностей. Знания о ней длительное время оставались неполными и оценочными вследствие естественных ограничений методов культивирования и биохимической идентификации. В настоящее время, благодаря появлению технологий высокопроизводительного секвенирования и развитию метагеномики и сравнительной геномики, стал возможен значительный прогресс в исследовании видового и внутривидового разнообразия бактерий интестинальной микробиоты (Blaser M.J., 2014).

Бифидобактерии являются неотъемлемым компонентом микрофлоры кишечника человека. В детском возрасте род *Bifidobacterium* является одним из численно доминирующих в кишечном микробиоценозе (Turroni F. и др., 2012; Ringel-Kulka T. и др., 2013). Среди основных групп облигатных анаэробов интестинальной микрофлоры бифидобактерии наиболее легко культивируются, что определяет большое количество информации, накопленной о данной группе микроорганизмов. Также значительную роль играет практически отсутствующая патогенность кишечных бифидобактерий, вследствие чего они активно используются в составе пробиотических препаратов. Наибольшее внимание при этом привлекают виды, характерные для нормальной микрофлоры человека. В связи с этим необходимо отметить вид *B. longum*, которому свойственна широкая распространенность среди как детей, так и взрослых (Шкопоров А.Н. и др., 2006; Джюев Ю.П. и др., 2011; Turroni F. и др., 2012). Ряд исследований постулировал высокий уровень генетической изменчивости данного вида бактерий, затрагивающий как общие характеристики их генома, так и присутствие отдельных генов, играющих роль в колонизации и персистенции в микрофлоре кишечника человека (Lee J.-H. и др., 2008; Sela D.A. и др., 2008). Однако систематического анализа изменчивости *B. longum*, основанного на большом числе полных геномных последовательностей, до настоящего времени не проводилось.

Остается неохарактеризованной долговременная персистенция штаммов бифидобактерий в кишечнике человека. Сопровождается ли изменение видового состава и численности бифидобактерий в кишечнике человека при переходе от детского типа микрофлоры к взрослому также и сменой штаммового состава? В настоящее время стало возможно получение данных об этом уровне вариабельности микрофлоры благодаря появлению методов высокопроизводительного секвенирования.

Таким образом, появление новых технологий, подходов и концепций в области

получения и анализа генетической информации позволяет проводить более глубокое изучение бактерий нормальной микрофлоры, что делает данную работу, посвященную исследованию представителей вида *B. longum*, актуальной в свете современного развития микробиологии.

Степень разработанности темы исследования

Видовой состав бифидофлоры детей и взрослых описан в большом числе исследований, однако, результаты в значительной степени варьируются в зависимости от используемых методов (Шкопоров А.Н. и др., 2006; Turrone F. и др., 2009; Джигоев Ю.П. и др., 2011; Жиленкова О.Г. и др., 2014), ставя необходимость проведения крупных унифицированных исследований на больших выборках.

Опубликовано значительное количество последовательностей геномов штаммов *B. longum* (Schell M.A. и др., 2002; Hao Y. и др., 2011; Nam J.-S. и др., 2011). Охарактеризованы различные генные островки, кодирующие свойства, предположительно способствующие колонизации интестинального тракта: адгезию к кишечному эпителию, синтез бактериоцинов, катаболизм олигосахаридов человеческого молока, производство экзополисахаридов (Lee J.-H., O'Sullivan D.J., 2010). Однако оценка размера пан-генома и исследование представленности различных функциональных категорий генов проводились только для рода в целом, без учета особенностей отдельных видов (Bottacini F. и др., 2010; Lukjancenko O. и др., 2012). Значительная роль горизонтального переноса генов у представителей данного вида постулировалась лишь на анализе последовательности только одного генома, при отсутствии сравнительного подхода (Schell M.A. и др., 2002).

Наблюдения отдельных случаев долговременной персистенции бифидобактерий в ходе взросления проводились только на уровне ДНК-типирования без привлечения геномных данных (Shkoporov A.N. и др., 2008).

Цель исследования:

Провести анализ внутривидовой геномной изменчивости и способности к долговременной персистенции штаммов *Bifidobacterium longum* микрофлоры кишечника здоровых людей.

Задачи исследования:

1. Исследовать видовое разнообразие представителей рода *Bifidobacterium* в микрофлоре кишечника здоровых людей, провести полногеномное секвенирование и сборку геномов отдельных представителей вида *B. longum*.
2. Провести реконструкцию внутривидовой филогении *B. longum* с использованием различных биоинформатических подходов.
3. Оценить вклад механизмов формирования геномной изменчивости *B. longum* и провести исследование структуры пан-генома наиболее генетически охарактеризованно-

го подвида *B. longum* subsp. *longum*.

4. Исследовать набор генов гликозил-гидролаз, обеспечивающих утилизацию углеводов, среди представителей вида *B. longum*.
5. Провести сравнение штаммов *B. longum*, выделенных от одних и тех же обследуемых лиц через длительные интервалы времени.

Научная новизна

Проведенное исследование видового разнообразия бифидобактерий микрофлоры кишечника с использованием комплекса молекулярных методов, включающего MALDI-TOF MS и анализ нуклеотидной последовательности гена 16S рПНК, показало статистически значимое снижение встречаемости видов *B. bifidum* и *B. breve*, повышение встречаемости вида *B. adolescentis* с возрастом, и высокую частоту обнаружения *B. longum* вне зависимости от возраста.

Впервые были охарактеризованы открытая структура пан-генома и высокая активность внутривидового и межвидового горизонтального переноса генов у представителей подвида *B. longum* subsp. *longum*. Филогенетический анализ, проведенный на основании нуклеотидных последовательностей консервативных генов, полного набора генов и средней нуклеотидной идентичности, подтвердил монофилетичность подвидов *longum*, *infantis* и *suis*, и позволил уточнить таксономическую принадлежность ранее описанных штаммов *B. longum*. Впервые был проведен анализ полного набора генов гликозил-гидролаз, который раскрыл метаболические особенности подвидов *longum*, *suis* и *infantis* вида *B. longum*, отражающие различные экологические ниши, занимаемые подвидами.

В штаммах, геномные последовательности которых были просеквенированы в рамках данного исследования, были выявлены два ранее не описанных семейства плазмид размером 3,9 тыс. п. о. и 7 тыс. п. о., а также системы CRISPR-Cas с нетипичной структурой: 2 локуса CRISPR, функционирующих за счёт одного набора генов *cas*.

Сравнительный анализ генетических последовательностей штаммов бифидобактерий, выделенных от одних и тех же обследуемых лиц через длинные периоды времени, впервые показал возможность долговременной персистенции близкородственных внутривидовых групп *B. longum*.

Теоретическая значимость работы

В ходе исследования сформированы представления о масштабах внутривидового генетического разнообразия *B. longum* и доказана важная роль горизонтального переноса генов в его формировании.

Определены подходы к выявлению внутривидовой таксономической принадлежности штаммов *B. longum* с использованием генетических последовательностей и обнаружены тенденции генетических различий их углеводного метаболизма.

Расширены знания о неконсервативных генетических элементах генома *B. longum* и встречаемости генов, кодирующих белки с предположительными пробиотическими свойствами.

Практическая значимость работы

В результате исследования создана и поддерживается библиотека штаммов бифидобактерий, включая штаммы с просеквенированными генетическими последовательностями, которая может быть использована для исследования молекулярно-генетических механизмов колонизации желудочно-кишечного тракта человека.

Нуклеотидные последовательности геномов, собранные до уровня контигов, и полные последовательности плазмид депонированы в международной базе данных GenBank (идентификаторы доступа JNWC00000000, JNVX00000000, JNVZ00000000, JNWA00000000, JNVY00000000, JNWC00000000, JNWD00000000, JRWN00000000, KP691640, KP691636, KP691641, KP691639, KP691635, KP691633, KP691637, KP691638). Разработана компьютерная база данных, характеризующая представленность различных групп ортологов в геномах *B. longum*.

Описанные новые генетические элементы и данные о встречаемости локусов, кодирующие белки с известными пробиотическими свойствами, могут быть применены в генетической инженерии бифидобактерий с целью повышения их пробиотического потенциала и использованы при отборе перспективных штаммов бифидобактерий.

Полученные в работе последовательности геномов, выложенные в открытом доступе, на данный момент уже используются для проведения отечественных и зарубежных научных исследований (Zakharevich N.V. и др., 2014; Bottacini F. и др., 2014; Zakharevich N.V. и др., 2015). Результаты исследования включены в материалы лекций кафедры микробиологии и вирусологии ГБОУ ВПО РНИМУ им. Н. И. Пирогова для студентов лечебного и педиатрического факультетов, аспирантов, ординаторов хирургического профиля, врачей-специалистов перинатальной медицины и педиатров, разработанная база данных используется для филогенетических исследований бактерий нормальной микрофлоры кишечника человека в научной работе кафедры микробиологии и вирусологии ГБОУ ВПО РНИМУ им. Н. И. Пирогова (акт внедрения от 07.12.15). Биоинформатические подходы применяются для сравнительно-геномного исследования молекулярно-генетических особенностей возбудителей дифтерии и коклюша в референс-центре по мониторингу за возбудителями дифтерии и коклюша ФБУН МНИИЭМ им Г.Н. Габричевского Роспотребнадзора (акт внедрения от 03.12.15).

Методология исследования

Методологической основой исследования послужили работы отечественных и зарубежных учёных в области классической микробиологии, генетики бактерий и биоинформатики. Предметом работы стало изучение бифидофлоры кишечника клинически

здоровых людей в возрасте от 1 месяца до 57 лет, полногеномное секвенирование набора штаммов *B. longum* и сравнительный анализ геномных последовательностей. Протокол исследования одобрен Этическим Комитетом РНИМУ им. Н. И. Пирогова (протокол заседания №122 от 17 декабря 2012 года).

Научная литература, посвящённая изучению микрофлоры человека, генетических характеристик бифидобактерий и применению сравнительно-геномных методов, была проанализирована формально-логическими методами. В работе использованы микробиологические, масс-спектрометрические, молекулярно-генетические, биоинформатические и статистические методы исследования.

Материалы и методы исследования

Культивирование и видовая идентификация бифидобактерий. Материалом для исследования служили фекалии здоровых людей, которые собирались стерильным шпателем и помещались в транспортный контейнер. В бактериологической лаборатории НИЛ микробиологии и биологической безопасности ГБОУ ВПО РНИМУ им. Н. И. Пирогова из исследуемого материала готовились серийные разведения в физиологическом растворе и высевались в объеме 0,1 мл из 10^5 и 10^7 разведений исследуемого материала раз на среду *Bifidobacterium agar* (HiMedia Labs Inc., Индия). Чашки с посевами инкубировались в микроанаэроаэратах (Oxoid, Великобритания), заполненных газовой смесью (85% N₂, 10% H₂, 5% CO₂) при 37°C в течение 48 часов. Первичный скрининг штаммов осуществлялся бактериоскопическим методом с использованием окраски по Граму. Штаммы, характеризовавшиеся типичной морфологией бифидобактерий, лиофилизировались в растворе сахарозы (10%) и желатина (1%) в лиофильной сушилке SB1 (Chemlab, Бельгия), последующие исследования чистых культур бифидобактерий проводились после выделения культур из лиофилизированных образцов. Идентификация бифидобактерий осуществлялась с использованием MALDI-TOF MS (Салгина А.В. и др., 2014) с использованием прибора Microflex LT (Bruker Daltonics, Германия). Анализ спектров выполнялся с использованием программного обеспечения MALDI Biotyper 3.0 (Bruker Daltonics, Германия). Для внутривидового типирования штаммов *B. longum* использовался анализ длин локусов VNTR 12, 23 и 25 (Matamoros S. и др., 2011). Штаммы бифидобактерий, видовую принадлежность которых не удавалось установить при помощи метода MALDI-TOF масс-спектрометрии, идентифицировались путем секвенирования фрагмента гена 16S рРНК. Амплификация проводилась с использованием праймеров UF1 (5'-AGAGTTTGATCCTGGCTCAG-3') и UR1 (5'-GCTGGCACGTTAGTCC-3'), секвенирование по Сэнгеру осуществлялось в ООО «Евроген», Москва.

Полногеномное секвенирование. Выделение ДНК для полногеномного секвенирования осуществлялось по модифицированному методу Stahl и соавторов (Shkorogov A.N. и др., 2008; Stahl M. и др., 1990). Секвенирование геномов штаммов 44В, 1-6В, 35В и 2-2В осуществлялось на платформе 454 FLX (Roche, Швейцария) в Институте Дж.

Крейга Вентера (JCVI), США, сборка проводилась с помощью программного обеспечения GS De Novo Assembler2.6 (<http://www.454.com/products/analysis-software/>). Секвенирование геномов остальных штаммов производилось на платформе Illumina HiSeq (Illumina, США) в ООО «Генотек», Москва, сборка осуществлялась с использованием CLC Genomics Workbench (<http://www.clcbio.com/products/clc-genomics-workbench/>).

Первичный биоинформатический анализ. Геномы бактерий аннотировались при помощи RAST (<http://rast.nmpdr.org/>). Полногеномное выравнивание последовательностей геномов проводилось с использованием progressiveMauve (<http://darlinglab.org/mauve/mauve.html>). Подсчет средней нуклеотидной идентичности осуществлялся с помощью JSpecies (<http://imedea.uib-csic.es/jspecies/>).

Реконструкция филогении. Для построения филогенетических деревьев из исследуемых геномов были взяты последовательности 43 генов, предположительно отвечающих за необходимые для жизнедеятельности клетки функции. Нуклеотидные последовательности генов выравнивались с помощью MUSCLE (<http://www.drive5.com/muscle/>). Филогенетическое дерево без учета рекомбинаций было построено на базе конкатенированных последовательностей генов по алгоритму Neighbor Joining с использованием модели замен Tamura-Nei в программном обеспечении MEGA (<http://www.megasoftware.net/>). Уровень достоверности был вычислен с помощью подхода bootstrap. Филогенетическое дерево с учетом рекомбинаций строилось в программе ClonalFrame (<http://www.xavierdidelot.xtreemhost.com/clonalframe.htm>).

Анализ групп ортологов. Кластеризация белок-кодирующих генов в группы ортологов выполнялась при помощи программного обеспечения OrthoMCL 2.0 (<http://www.orthomcl.org/common/downloads/software/v2.0/>) с использованием значений параметров, рекомендованных разработчиками: порог $e\text{-value } 10^{-5}$ и индекс расширения $I=1,5$ (Li L. и др., 2003). Функциональная группировка групп ортологов осуществлялась с использованием RPS-BLAST на базе данных COG (<http://weizhonglab.ucsd.edu/metagenomic-analysis/server/cog/>). Результаты кластеризации и функциональной группировки объединялись в базе данных, созданной на платформе MySQL (<http://mysql.com/>). Динамика изменения размера кор-генома и пан-генома *B. longum* subsp. *longum* вычислялась с помощью 10000 случайных перестановок порядка включения штаммов в анализ. Тип пан-генома описывался с помощью модели степенного закона (Tettelin H. и др., 2008).

Исследование неконсервативных элементов генома. IS-элементы были аннотированы путём поиска по базе данных ISFinder с помощью ISSaga (<http://issaga.biotoul.fr/>). Поиск плазмид производился на основе сходства последовательности с известными плазмидами *B. longum*, а также анализа покрытия контигов. Аннотация генов гликозилгидролаз проводилась с использованием CAZymes Annotation Toolkit (<http://mothra.ornl.gov/cgi-bin/cat/cat.cgi>). Системы рестрикции-модификации выявлялись путём поиска по базе данных REBASE (

lmb.cam.ac.uk/genomes/madanm/pres/rebase1.htm). Поиск локусов CRISPR проводился с использованием CRISPRFinder (<http://crispr.u-psud.fr/Server/>), фланкирующие области найденных регионов исследовались в поисках генов *cas*. Ортологичные локусы CRISPR системы CRISPR-Cas №1 амплифицировались с использованием праймеров 5'-CCSTATGGATGGTGGAAATCAG-3' и 5'-CAATTAGTGCCGGTCTATGGAG-3'. Для амплификации локусов CRISPR системы CRISPR-Cas №2 применялся метод вложенной ПЦР. Последовательности внешних праймеров были GTTGCTCGACATGGGATATGG-3' и 5'-TGGATCTGGTACAGGGTGAC-3', последовательности внутренних праймеров - 5'-GTATACGCCTACGCAATCGG-3' и 5'-AACATCCGCCGATAAACAGTC-3. Подтверждение интеграции мобильного генетического элемента в геном *B. longum* subsp. *longum* ЕКЗ проводилось с использованием ПЦР с праймерами 5'-ССАСТТСТССАГССГГАТГТТ-3' и 5'-ТГАССГСАГААГГТГТСТС-3.

Статистическая обработка данных. Статистическая обработка осуществлялась с использованием программного пакета R (<https://cran.rstudio.com/>). Во всех случаях различия считались достоверными при скорректированном $p < 0,05$. Для анализа встречаемости видов использовался критерий Манна-Уитни как эквивалент теста на нулевую ранговую бисериальную корреляцию (Willson V.L., 1976) с поправкой Холма-Бонферрони. Анализ численности бифидобактерий проводился с использованием регрессионного анализа, при этом проверки на нормальность распределения остатков и гомоскедастичность проводились с помощью визуальной оценки квантиль-квантильных графиков и графиков остаток-предсказание, соответственно. Сравнение частоты функциональных классов в разных частях пан-генома осуществлялось с использованием точного теста Фишера с контролирующей процедурой Бенджамини-Хохберга. Для снижения размерности данных о численностях представителей семейств гликозил-гидролаз в штаммах *B. longum* использовался метод главных компонент. Неравномерность распределения числа представителей семейств гликозил-гидролаз между подвидами выявлялась с использованием критерия Краскела-Уоллиса с поправкой Холма-Бонферрони.

Степень достоверности и апробация результатов исследования

О достоверности результатов свидетельствует достаточный объем выборки при анализе штаммов бифидобактерий (220 штаммов, выделенных от 93 лиц) и при анализе геномных последовательностей (28 штаммов). Все количественные данные обрабатывались с использованием общепринятых статистических подходов. Достоверность результатов реконструкции филогении на уровнях разделения вида на подвида и близкородственные группы подтверждается совпадением результатов, полученных с использованием методов, основанных на последовательностях консервативных генов, на общем наборе генов и на средней нуклеотидной идентичности. Депонированные генетические последовательности прошли проверку специалистами NCBI.

Диссертация апробирована на научно-практической конференции кафедры микробиологии и вирусологии ГБОУ ВПО РНИМУ им. Н.И. Пирогова Минздрава России, протокол №12 от 24.06.2015 г. Основные результаты диссертационной работы представлены, доложены и обсуждены на VIII Международной Пироговской научной медицинской конференции студентов и молодых ученых в Москве, 2013 г. и XXI Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2014» в Москве, 2014 г.

Личное участие автора в получении результатов

Автором выполнены все этапы молекулярно-генетических исследований, биоинформатического анализа полученных данных, статистической обработки и теоретического обобщения полученных результатов. Бактериологическое исследование, а также выделение ДНК проводились совместно с сотрудниками НИЛ микробиологии и биологической безопасности ГБОУ ВПО РНИМУ им. Н. И. Пирогова: старшим научным сотрудником Шкопоровым А. Н. и лаборантом Бржозовским А. Г, а также с сотрудниками лаборатории молекулярной генетики микроорганизмов ФГБУН НИИ ФХМ ФМБА России Парфёновой Т. В. и Ильиной Е. Н.

Основные положения, выносимые на защиту

1. Подвиды *longum*, *infantis* и *suis* вида *B. longum* являются монофилетическими группами, различимыми при анализе полных геномных последовательностей.
2. Внутривидовой и межвидовой горизонтальный перенос генов вносят значительный вклад в формирование генетического разнообразия представителей вида *B. longum* на уровне как последовательностей генов, так и общего их набора.
3. Набор генов гликозил-гидролаз в геномных последовательностях *B. longum* обладает высокой вариабельностью и отражает различные экологические ниши, занимаемые подвидами.

Публикации

Основное содержание диссертации отражено в 5 печатных работах, в том числе, 3 статьи опубликовано в рецензируемых изданиях, 2 – в материалах конференций.

Структура и объем диссертации.

Основной текст диссертации изложен на 123 страницах машинописного текста, состоит из введения, одной главы с обзором литературы, трех глав собственных исследований, обсуждения полученных результатов, выводов, практических рекомендаций, перспективы дальнейшей разработки, списка сокращений и списка литературы. Диссертация иллюстрирована 5 таблицами и 23 рисунками. Список литературы содержит 205 источников, в том числе, 13 - отечественных и 192 – зарубежных.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

1. Исследование разнообразия представителей рода *Bifidobacterium* в кишечной микрофлоре людей различного возраста

Изучение качественного и количественного состава бифидобактерий толстого кишечника проводилось у 93 практически здоровых людей обоего пола (30 мужского и 63 женского). Разброс возраста составлял от 1 месяца до 57 лет, при этом дети до двух лет составляли 33% от общей выборки.

В результате данного этапа исследования было выделено 260 штаммов бактерий, из которых для 220 была подтверждена принадлежность к роду *Bifidobacterium*. Если в ходе исследования оказывалось, что два или более штамма, выделенных от одного обследуемого, относятся к одному виду, то они рассматривались как один. Окончательное число штаммов бифидобактерий в работе составило 187. Среди них было идентифицировано 44 штаммов *B. adolescentis*, 63 - *B. longum*, 31 - *B. bifidum*, 13 - *B. catenulatum*, 7 - *B. pseudocatenulatum*, 2 - *B. angulatum*, 7 - *B. breve*, 13 - *B. animalis* и 5 - *B. dentium*. Таким образом, самым распространенным из видов являлся *B. longum* (обнаруживался в 63 случаях из 93), а наиболее редкими – виды *B. dentium* (5 случаев из 93) и *B. angulatum* (2 случая из 93). Последний был исключен из анализа встречаемости вследствие крайне малой представленности. С использованием MALDI-TOF MS было успешно идентифицировано 93% выделенных штаммов бифидобактерий. В 7% случаев результат масс-спектрометрической идентификации определялся как ненадежный (рейтинг < 1,7), в этих случаях проводилась идентификация с использованием секвенирования гена 16S рРНК.

Суммарная концентрация бифидобактерий у обследуемых в среднем составила $9,3 \pm 1,0 \log$ КОЕ/г, средние концентрации отдельных видов незначимо варьировали от 8,5 до 9,7 \log КОЕ/г. С использованием регрессионного анализа было показано, что суммарная концентрация бифидобактерий в среднем снижается с возрастом ($p=1,4 \cdot 10^{-7}$). Также снижение концентрации было выявлено для отдельных видов - *B. longum* ($p=1,7 \cdot 10^{-6}$), *B. catenulatum* ($p=0,014$) и *B. animalis* ($p=0,036$). Ни для одного из видов бифидобактерий не было выявлено статистически значимое повышение абсолютной концентрации с возрастом.

Для некоторых видов было показано, что их встречаемость (в концентрации более 10^6 КОЕ/г) зависит от возраста (Рисунок 1). Так, бактерии вида *B. breve* ($p=0,0005$) встречались только в детском возрасте, максимальный возраст их носителя в данном исследовании составил 2 года. Также преимущественно от обследуемых лиц детского возраста выделялись представители вида *B. bifidum* ($p=0,020$). Напротив, бактерии *B. adolescentis* ($p=0,0003$) обнаруживались только с возраста 1 год и более. Для остальных видов не было выявлено значимой зависимости частоты их выделения от возраста.

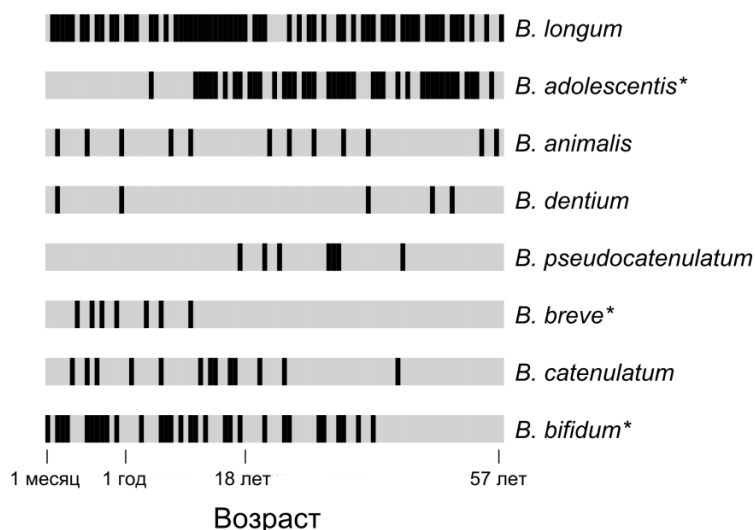


Рисунок 1 - Встречаемость видов бифидобактерий в зависимости от возраста.

Примечание: Диаграмма разделена на 93 столбца, каждый из которых соответствует одному обследуемому, они расположены в порядке от наименьшего возраста к наибольшему. Черный цвет означает наличие данного вида среди штаммов, выделенных от данного обследуемого, серый – отсутствие. Звездочками отмечены виды, встречаемость которых значимо зависит от возраста ($p < 0,05$).

2. Выбор штаммов для анализа полных геномных последовательностей

Большая часть штаммов *B. longum*, выбранных для полногеномного секвенирования в данном исследовании (10 из 12), были выделены как численно преобладающие бифидобактерии в нескольких временных точках в ходе 11-летнего обсервационного исследования кишечной микрофлоры троих здоровых детей (Рисунок 2).

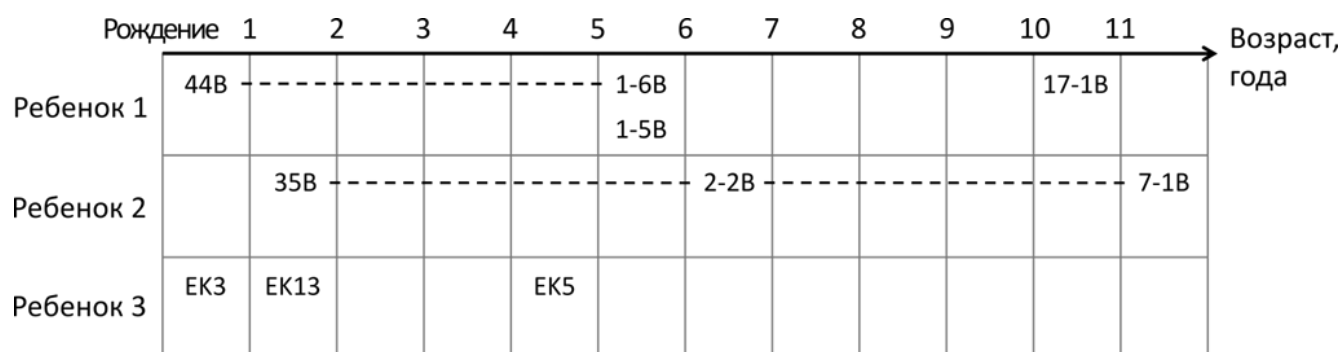


Рисунок 2 - Штаммы, выделенные в результате длительного обсервационного исследования и выбранные для полногеномного секвенирования.

Примечание: горизонтальная ось отображает возраст обследуемого лица, надписи в ячейках соответствуют названиям выделенных штаммов. Пунктирные линии соединяют штаммы, сходные по последовательностям VNTR.

Для всех штаммов в данном наблюдательном исследовании, которые были идентифицированы как *B. longum*, был проведен анализ длин локусов VNTR. В результате было обнаружено, что некоторые штаммы, выделенные от одних и тех же людей в разные временные точки, схожи по набору длин локусов VNTR, что может указывать на их близкое родство. Один из таких случаев включал в себя штамм 44В (выделенный от ребенка 1 в возрасте 9 месяцев) и штамм 1-6В (выделенный от того же ребенка в возрасте 5 лет). Второй случай включал штаммы от ребенка 2: 35В (выделенный в 11 месяцев), 2-2В (выделенный в 6 лет) и 7-1В (выделенный в 11 лет). Все эти штаммы были подвергнуты полногеномному секвенированию. Чтобы оценить общее генетическое разнообразие *B. longum*, было осуществлено секвенирование геномов штаммов 1-5В и 17-1В, которые также были выделены от ребенка 1, но не обладали схожестью по длинам локусов VNTR с другими штаммами в исследовании.

Помимо этого были прочтены геномные последовательности двух дополнительных штаммов, выделенных ранее из кишечной микрофлоры детей и описанных в предыдущих работах: 72В и VMKB44 (Shkorporov A.N. и др., 2006, 2008). Итого в данном исследовании секвенированы последовательности геномов 12 штаммов (Таблица 1), из которых 11 принадлежали к подвиду *longum*, и 1 штамм относился к подвиду *infantis*.

Таблица 1 - Геномные последовательности штаммов *B. longum*, секвенированные в ходе данного исследования.

Штамм	Идентификатор доступа GenBank	Статус генома	Размер, млн. п. о.	Подвид
ЕК3	JNWC000000000	45 контигов	2,56	<i>infantis</i>
1-5В	JNVX000000000	28 контигов	2,37	<i>longum</i>
17-1В	JNVZ000000000	23 контига	2,47	<i>longum</i>
1-6В	AJTF000000000	171 контиг	2,69	<i>longum</i>
2-2В	AJTJ000000000	141 контиг	2,63	<i>longum</i>
35В	AJTI000000000	131 контиг	2,51	<i>longum</i>
44В	AJTM000000000	62 контига	2,56	<i>longum</i>
72В	JNWA000000000	39 контигов	2,37	<i>longum</i>
7-1В	JNVY000000000	37 контигов	2,41	<i>longum</i>
ЕК5	JNWC000000000	35 контигов	2,23	<i>longum</i>
ЕК13	JNWD000000000	48 контигов	2,47	<i>longum</i>
VMKB44	JRWN000000000	36 контигов	2,51	<i>longum</i>

Также для анализа были использованы финализированные либо собранные до уровня контигов последовательности геномов 16 штаммов, просеквенированных в других исследованиях и публично доступных в базе данных GenBank (на момент сентября 2014 г). Финальное число объектов исследования составило 28 штаммов.

3. Внутривидовая филогения *B. longum*

Для филогенетического анализа был выбран набор из 43 консервативных генов из каждого доступного генома *B. longum*, а также их гомологи из *B. breve* UCC2003. Данные последовательности были выровнены, выравнивания подверглись соединению (конкатенации), после чего была осуществлена реконструкция филогении (Рисунок 3). В результате штаммы *B. longum* образовали три крупных кластера. Каждый из них содержал типовой штамм, относящийся к одному из трех подвидов, вследствие чего данные кластеры были ассоциированы с тремя подвидами *B. longum*.

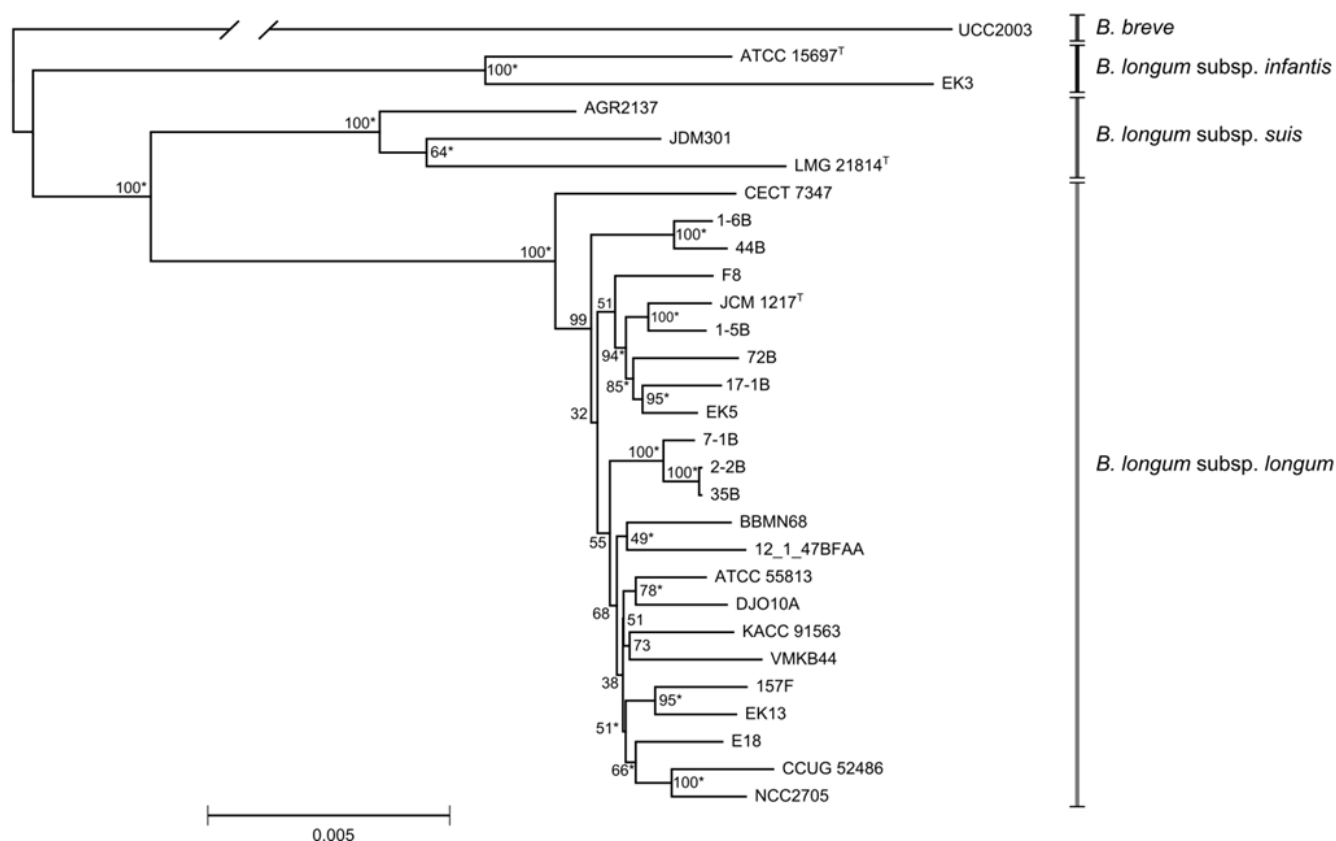


Рисунок 3 - Филогенетическое дерево вида *B. longum*, основанное на объединенных последовательностях 43 консервативных генов.

Примечание: узлы дерева, подтвержденные алгоритмом ClonalFrame отмечены звездочками (*), цифры в узлах отображают уровень поддержки, полученный с помощью метода bootstrap.

При анализе выравниваний отмечались многочисленные события горизонтального переноса генов между представителями разных подвидов, предположительно произошедшие по механизму гомологичной рекомбинации с геной конверсией. Вследствие этого был применен алгоритм ClonalFrame, разработанный для филогенетического анализа с учетом рекомбинационных событий. Подсчитанное отношение частоты рекомбинаций к частоте однонуклеотидных замен составило 0,047, а отношение вероятности, что нуклеотид будет изменен рекомбинацией, к вероятности однонуклеотидной мутации -

0,72, что отражает относительно высокий вклад горизонтального переноса генов в формирование изменчивости *B. longum*. Как минимум одно рекомбинационное событие затронуло ген 16S рРНК штамма *B. longum* subsp. *suis* JDM301, что ставит под сомнение достоверность разделения подвидов с применением последовательностей только данного гена.

Штаммы 1-6В и 44В, выделенные от ребенка 1, сформировали тесный кластер, также подобная степень сходства была обнаружена между штаммами 2-2В, 35В и 7-1В, выделенными от ребенка 2. Однако полностью идентичных штаммов найдено не было. В то время как штаммы 2-2В и 35В обладали очень высокой степенью сходства последовательностей консервативных генов (только 6 из 65310 п. о., использованных для выравнивания, отличались), дистанции между остальными штаммами, выделенными от одних и тех же детей в разные моменты времени, были значительно больше. Например, наблюдались затрагивающие множество нуклеотидов различия между последовательностями генов, кодирующих аспартил-тРНКсинтазу (*aspS*), UDP-N-ацетилмурамоилаланил-D-глутамат 2,6-диаминопимелат лигазу (*murE*), пролил-тРНК синтазу (*proS*), аденилосукцинатлиазу (*purB*) и главную сигма-субъединицу РНК-полимеразы (*rpoD*), между штаммами 1-6В и 44В, что указывает на события горизонтального переноса генов, произошедшие после расхождения филогенетических линий, ведущих к данным штаммам.

Также было произведено вычисление средней нуклеотидной идентичности (ANI) между всеми парами штаммов *B. longum*, чтобы осуществить кластеризацию штаммов на основании альтернативного подхода. В результате получены точно такие же результаты по поводу разделения штаммов на три подвида. ANI между двумя штаммами подвида *infantis* составила 98,28%, средняя ANI между штаммами подвида *suis* составила 98,35%, и средняя ANI между штаммами подвида *longum* была 98,89% (98,88% после исключения схожих штаммов 1-6В, 44В, 2-2В, 35В и 7-1В). Также было обнаружено, что ANI между подвидами *longum* и *infantis* варьировала между 95% и 96% (в среднем 95,66%), рядом с порогом, определяющим отнесение бактерий к одному виду или к разным.

4. Общая характеристика геномных последовательностей *B. longum*

Среди исследуемого набора штаммов размер генома был унимодально распределен со средним значением 2,39 млн. п. о. Наибольший геном (2,83 млн. п. о.) среди проанализированных штаммов принадлежал ATCC 15697^T, типовому штамму *B. longum* subsp. *infantis*. Было обнаружено, что штаммы подвида *infantis* в общем обладали более крупными геномами, чем штаммы подвида *longum* (критерий Манна-Уитни, $p=0,02$).

GC-состав генома *B. longum* был равен $59,97 \pm 0,23\%$, у большинства штаммов он варьировал в пределах от 59,63% до 60,33%, за исключением штамма ЕК3, у которого GC-состав был значительно ниже и равнялся 59,38%. Структурные характеристики генома изучались на основании 8 финализированных последовательностей геномов *B. longum*. Полногеномное выравнивание показало высокий уровень синтении, только одна

большая перестройка была отмечена в геноме штамма *B. longum* subsp. *infantis* ATCC 15697^T. Однако, было обнаружено значительное количество участков, подвергавшихся инсерционно-делеционным событиям, что определяло значительную вариабельность размера генома.

Для анализа генного состава штаммов *B. longum* осуществлялась аннотация всех геномных последовательностей, включенных в анализ, и формирование групп ортологов. Всего было получено 5324 группы ортологичных генов, в том числе, 287 групп, содержащих несколько паралогичных генов как минимум в одном геноме, и 1527 групп, представители которых находятся только в одной из геномных последовательностей. Среднее число групп ортологов в геноме составляло 2049 ± 163 , то есть, менее чем половину от общего числа групп ортологов, встречающихся среди *B. longum*.

Попарное сравнение числа общих групп ортологов между парой штаммов было использовано для построения иерархической кластеризации. В результате получено разделение штаммов *B. longum* на те же подвиды, что и в случае подходов на основе последовательностей консервативных генов и средней нуклеотидной идентичности.

5. Пан-геном *B. longum* subsp. *longum*

Неравномерность в выборке штаммов может оказать значительное влияние на анализ пан-генома, независимо от модели, используемой для интерпретации данных. На сегодняшний день, подвид *longum* является значительно более изученным, чем остальные подвиды *B. longum*, и 23 из 28 штаммов, включенных в данное исследование, принадлежали к данному таксону. Подобная гетерогенность выборки может внести погрешности в оценку параметров пан-генома, поэтому было принято решение выполнить данный анализ только для представителей подвида *longum*.

Размер пан-генома, полученного из последовательностей 23 геномных последовательностей *B. longum* subsp. *longum* составил 4169 групп ортологов, что составляет примерно 78% от общего числа групп ортологов, присутствующих во всех трех подвидах *B. longum* (5324 группы). Это значение было в два раза больше, чем среднее число групп ортологов в штамме *B. longum* subsp. *longum* (2027 ± 126). Распределение числа групп ортологов в зависимости от числа штаммов, в которых они содержатся, имело типичную форму U-образной кривой, то есть, большинство генов были очень редкими либо, напротив, представленными во всех штаммах (Рисунок 4).

На основании данного распределения все группы ортологов в пан-геноме *B. longum* subsp. *longum* были разделены на три множества: консервативные гены (присутствуют во всех 23 штаммах и формируют кор-геном), умеренно распространенные гены и редкие гены (присутствуют только в одном штамме). Функциональная аннотация с использованием базы данных COG показала неравномерное распределение нескольких функциональных классов среди данных множеств.

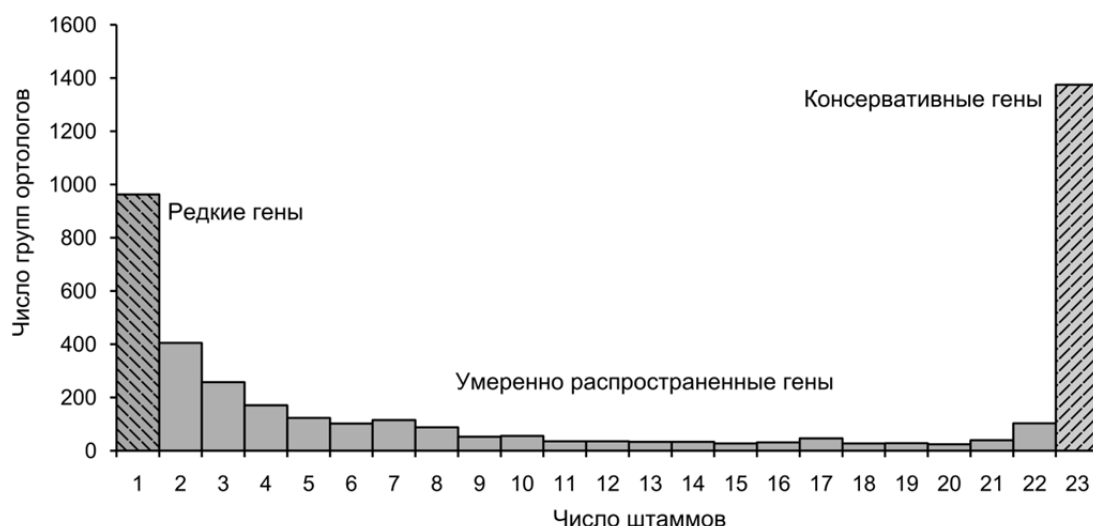


Рисунок 4 - Распределение числа групп ортологов в штаммах *B. longum* subsp. *longum* в зависимости от числа штаммов, в которых они присутствуют.

Множество консервативных генов было обогащено генами, вовлеченными в трансляцию и посттрансляционные модификации, продукцию и конверсию энергии, а также генами метаболизма и транспорта аминокислот, коферментов и нуклеотидов. Напротив, гены синтеза поверхностных структур были преимущественно представлены среди умеренно распространенных и редких генов, отражая широкое разнообразие в пан-геноме *B. longum* сортаз, гликозилтрансфераз и белков модификации углеводов, определяющих штаммоспецифичность поверхностных молекул данных бактерий. Группа «Репликация, рекомбинация и репарация» была перепредставлена среди редких групп ортологов, отражая значительное содержание в пан-геноме *B. longum* генов мобильных генетических элементов (транспозаз и сайт-специфических рекомбиназ), распределение которых является практически штаммоспецифическим.

Для оценки размера кор-генома и пан-генома *B. longum* subsp. *longum* был построен график, отражающий число групп ортологов как функцию числа штаммов, включенных в исследование (Рисунок 5). Абсолютная разница между средними и медианными значениями в каждой точке была равна $5,7 \pm 2,1$ групп ортологов, показывая относительную гомогенность выборки штаммов. В связи с этим для дальнейшего анализа были использованы только средние значения.

Зависимость числа групп ортологов в кор-геноме от числа штаммов, включенных в анализ, аппроксимировалась суммой степенной функции и медленно убывающей линейной функции ($y=571,8x^{-1,167}-4,39x+1459,1$; $R^2=0,99998$). Добавление последних 3 штаммов в выборку приводило к уменьшению размера кор-генома на 3-5 групп ортологов. Однако данное снижение может быть результатом погрешности, возникающей за счёт использования нефинализированных геномов. Таким образом, нельзя исключить, что функция снижения размера кор-генома выходит на плато.

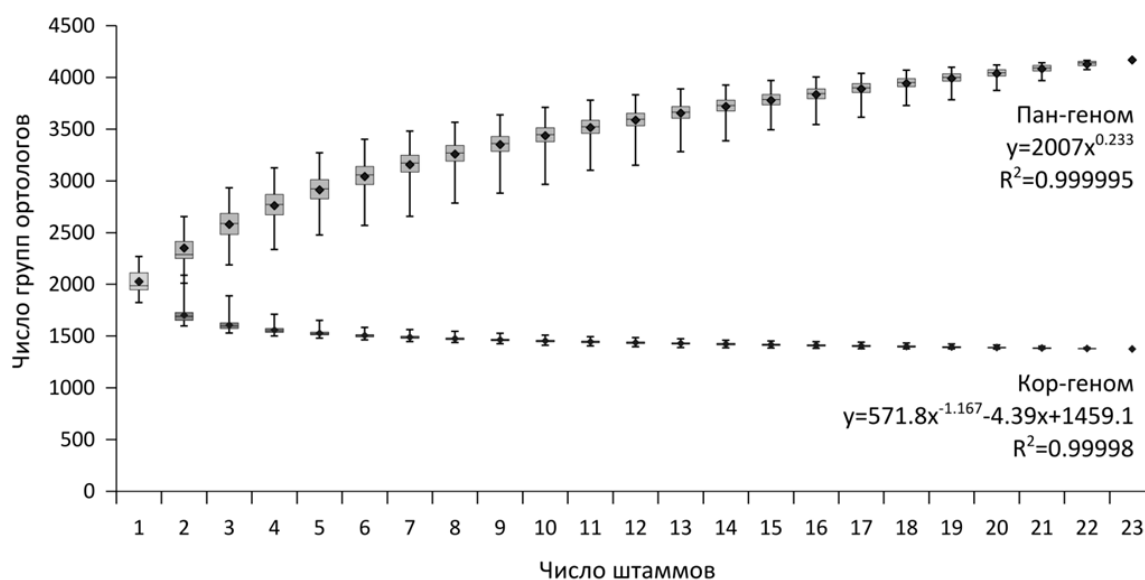


Рисунок 5 - Размеры пан-генома и кор-генома *B. longum* subsp. *longum* как функций числа последовательно добавляемых штаммов. Разброс данных вычислен с помощью 10000 случайных перестановок порядка включения штаммов.

Функция нарастания пан-генома аппроксимировалась степенной функцией $y=2007x^{0,233}$ ($R^2=0,999995$). Число новых групп ортологов с добавлением одного штамма в анализ аппроксимировалось уравнением $y=469,7x^{-0,76}$ ($R^2=0,998$). Коэффициент степенной функции в данном уравнении находился между -1 и 0, свидетельствуя о типичной открытой структуре пан-генома. Это означает, что прирост числа групп ортологов с добавлением новых штаммов в анализ не стремится к нулю, а график числа групп ортологов как функции числа штаммов не выходит на горизонтальное плато. Таким образом, новые гены будут постоянно обнаруживаться с нарастанием числа штаммов *B. longum* subsp. *longum* с просеквенированным геномом, и потребуется очень большое число штаммов для полного описания генетического разнообразия данной группы бактерий. В соответствии с аппроксимирующей функцией, более 3000 дополнительных штаммов, включенных в исследование, потребуется для того, чтобы снизить прирост размера пан-генома до 1 группы ортологов на добавляемый штамм.

6. Гены гликозил-гидролаз *B. longum* subsp. *longum*

Метаболизм бифидобактерий, в соответствии с общепринятыми взглядами, является специализированным на сбраживании широкого спектра углеводов. В данном исследовании было обнаружено, что финальные этапы углеводного катаболизма (галакто-N-биоизный/лакто-N-биоизный путь и фруктозо-6-фосфат-фосфокетолазный путь) присутствуют во всех исследованных геномах *B. longum*, но при этом спектр гликозил-гидролаз, обеспечивающих расщепление сложных углеводов, продемонстрировал значительную вариабельность.

С использованием CAZymes Analysis Toolkit была проведена аннотация генов гликозил-гидролаз в геномных последовательностях *B. longum*. Количества представителей семейств гликозил-гидролаз в каждом геноме были использованы для анализа с помощью метода главных компонент (Рисунок 6). Полученные результаты показали, что подвиды *B. longum* являются различимыми по спектру наличествующих генов гликозил-гидролаз. Представители подвидов *longum* и *infantis* резко отличались друг от друга, в то время как представители подвида *suis* формировали промежуточный «слой» между ними.

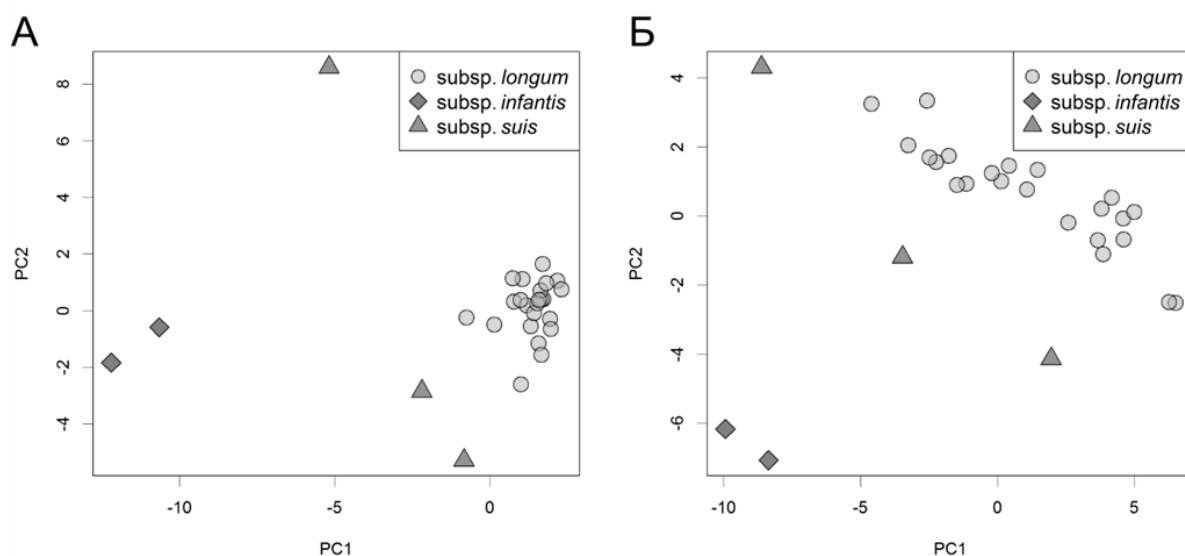


Рисунок 6 - Метод главных компонент в применении к количествам генов представителей семейств гликозил-гидролаз: (А) с предварительной нормировкой (Б) без нормировки.

Десять семейств генов гликозил-гидролаз были статистически значимо неравномерно распределены между подвидами. Данные гены могут быть объединены в три кластера. Первый содержит семейства, широко представленные в геномах штаммов подвида *infantis* и умеренно распределенные в остальных подвидах. Этот кластер содержит семейство GH92, представляющее собой экзо- α -маннозидазы, семейство GH20, включающее в себя экзо-N- β -ацетилглюкозаминидазы и экзо-лакто-N-биозидазы, а также семейство GH33, содержащее сиалидазы. Второй кластер представляет собой семейства, отсутствующие в подвиде *infantis*, но обнаруживаемые в других подвидах. Он включает в себя семейство GH27, представители которого проявляют различные активности, включая α -галактозидазную и β -L-арабинопиранозидазную, семейство GH101, содержащее эндо- α -N-ацетилгалактозаминидазы, семейство GH127, включающее в себя β -L-арабинопиранозидазы, и семейство GH31, содержащее множество ферментов с разными активностями, в том числе α -глюкозидазной и α -ксилозидазной. Третий кластер включает в себя семейства, которые, по-видимому, практически отсутствуют в представителях подвида *longum*: GH29 и GH95, содержащие α -фукозидазы, а также GH109, члены кото-

рого представляют собой α -N-ацетилглюкозаминидазы, действующие через NAD-зависимый механизм гидролиза.

Таким образом, полученные результаты подтверждают данные других исследований, что геномные последовательности подвидов *longum* и *infantis* отражают специализацию на катаболизме разных групп субстратов: подвид *longum* проявляет большой геномный потенциал для утилизации растительных полисахаридов, таких как арабиноксиланы, в то время как подвид *infantis* более приспособлен для деградации олигосахаридов человеческого молока, для которых наиболее характерными углеводными остатками являются фукоза и сиаловая кислота. *B. longum* subsp. *suis* по-видимому имеет «промежуточный» углеводный метаболизм между подвидами *longum* и *infantis*: ни одно семейство гликозил-гидролаз не было значимо недопредставлено или перепредставлено в этом семействе по сравнению с другими.

7. Неконсервативные элементы генома *B. longum*

В штаммах, секвенированных для данного исследования, был произведен поиск малых криптических плазмид. Соответствующие последовательности были найдены в 8 из 12 штаммов, все они были замкнуты в кольцо на основании повторов на концах контингов, продолженных с использованием картирования исходных прочтений (Таблица 2). Соотношения покрытия плазмиды и генома варьировало от 4,0 до 20,2 для всех исследуемых плазмид, за исключением возможно интегрированной в геном плазмиды штамма 35B.

Большинство обнаруженных плазмид имело длину 3624 п. о. и принадлежало к хорошо изученному семейству бифидобактериальных плазмид, прототипическими представителями которого считаются pKJ36 и pB44 (Shkорогов и др., 2008), последняя из них была выделена из штамма VMKB44, просеквенированного в рамках данной работы. Сравнение нуклеотидных последовательностей показало, что плазмиды p1-5B и p1-6B (обнаруженные в штаммах, которые были выделены от ребенка 1 в возрасте 5 лет) были идентичны, также как и плазмиды p35B и p2-2B (обнаруженные в штаммах, которые были выделены от ребенка 2 в разных временных точках).

Другие малые криптические плазмиды, обнаруженные в исследуемых штаммах, не имели гомологии с ранее описанными плазмидами на всём своем протяжении, но, предположительно, они произошли от представителей известных плазмидных семейств в результате рекомбинационных событий. Плазмиды p1-5B, p1-6B и p17-1B (обнаруженные в штаммах, которые были выделены от одного ребенка) имели 100% идентичную последовательность. Они обладали размером 3919 п. о. и имели в своем составе регионы, гомологичные ранее описанным плазмидам бифидобактерий pNAL8M (1068 п. о., 99% идентичность), pNAC1 (1957 п. о., 89% идентичность) и pFI2576 (719 п. о., 92% идентичность).

Таблица 2 - Малые криптические плазмиды в штаммах *B. longum*, просеквенированных в рамках данного исследования.

Плазида	Штамм	Отношение покрытия плазмиды к покрытию генома	Размер, п. о.	Идентификатор доступа Genbank
p1-5B1	1-5B	10,5	3919	KP691640
p1-6B1	1-6B	4,9	3919	KP691636
p17-1B	17-1B	13,2	3919	KP691641
p1-5B2	1-5B	6,6	3624	KP691639
p1-6B2	1-6B	4,0	3624	KP691635
p35B	35B	1,0	3624	KP691634
p2-2B	2-2B	5,4	3624	KP691633
pB44	VMKB44	20,2	3624	AY066026
p72B	72B	11,0	3624	KP691638
pEK13	EK13	12,7	7050	KP691637

Плазида pEK13 из штамма EK13 имела длину 7050 п. о. – больше, чем большинство известных криптических плазмид бифидобактерий. Она включает в себя практически полную последовательность плазмиды, схожей с pB80 *B. bifidum* (4978 п. о. , 99% идентичность), а также регион, гомологичный плазмиде pNAC1 (2885 п. о. 96% идентичность). Эти два региона имели два перекрытия с длинами 404 и 408 п. о., идентичные на 82,5%. Указанная структура дает основания полагать, что pEK13 произошла в результате коинтеграции двух предковых плазмид вследствие гомологичной рекомбинации по схожим участкам.

С использованием алгоритма ISsaga проводилось автоматическое предсказание позиций IS-элементов, простейших из прокариотических мобильных генетических элементов, в восьми финализированных геномах *B. longum*. В геномных последовательностях штаммов *longum* и *suis* содержалось 28-40 уникальных IS-элементов, в геноме штамма *B. longum* subsp. *infantis* ATCC 15697^T был обнаружен 61 локус. Некоторые семейства IS-элементов присутствовали во всех штаммах *B. longum* (IS3, ISL3, IS21, IS30, IS21), но большинство из них обнаруживалось в одном или нескольких геномах. Общее число IS-элементов в штаммах *B. longum* варьировало между 42 и 90.

При анализе геномной последовательности штамма EK3 обнаружено, что один из контигов (JNWB01000008.1, 3733 п. о.) обладает GC-составом 36.71%, значительно отличаясь от GC-богатых нуклеотидных последовательностей бифидобактерий. Он представляет собой мобильный генетический элемент, предположительно заимствованный от представителя рода *Streptococcus*, и содержащий открытую рамку считывания, кодирующую редуктазу ионов ртути. Анализ покрытия показал, что он находится в геноме примерно в 20 копиях, исходя из чего можно предположить его высокую активность.

В ходе исследования геномных последовательностей штаммов *B. longum* обнаружено 4 различающихся системы CRISPR-Cas. Система №1 (консенсусная последовательность повтора ANCTACCCCGCACACGCGGGGATAAACC), относящаяся к классу I, присутствует в штаммах 44В, 1-6В и 2-2В. Она имеет необычную структуру, включая в себя один набор генов *cas* и два локуса CRISPR с последовательностями повторов, отличающимися на один нуклеотид. Штаммы 44В и 1-6В, выделенные от ребенка 1, обладают идентичным набором спейсеров в данной системе, что подтверждает близкое родство этих штаммов. Однако, как было выявлено при исследовании последовательностей генома и подтверждено при помощи ПЦР, среди близкородственных штаммов от ребенка 2 система №1 присутствует в штамме 2-2В и отсутствует в штаммах 35В и 7-1В.

Система №2 (повтор CAAGCTTATCAAGAAGGGTGAATGСТААТТСССAGC), относящаяся к хорошо изученным и активно применяющимся в генной инженерии системам CRISPR-Cas II класса, широко распространена среди *B. longum* (обнаружена в 7 штаммах из 28). Локусы CRISPR штаммов 2-2В, 35В и 7-1В, выделенных от ребенка 2, идентичны за исключением одного спейсера (третьего с 3'-конца, отсутствующего в штамме 2-2В). Штаммы 44В и 1-6В также имели одинаковые спейсеры в собранных регионах CRISPR данной системы. Система №3 (повтор CTTGCATACGTCAAACGTATGCACTTCATTGAGGA) принадлежит к системам CRISPR-Cas типа *Dpsuc*. Она была обнаружена только в штамме 17-1В. Система №4 (повтор GTTTGCCCCCGCANGCGCGGGGATGATCCG) имеет типичную структуру систем CRISPR-Cas I класса. Она присутствует в штаммах ВВМN68 и ЕК3, однако, последовательность повторов в данных бактериях отличается на один нуклеотид.

Большой вклад в защиту бактериальных клеток от плазмид и бактериофагов также вносят системы рестрикции-модификации, которые производят метилирование специфических участков ДНК бактериальной клетки, и элиминацию молекул ДНК, содержащих неметилированные участки. С использованием поиска против базы данных REBASE и ручной курации результатов было найдено 12 ортологичных групп предположительных эндонуклеаз в геномных последовательностях штаммов *B. longum*, при этом в 8 из 12 случаев они были расположены рядом с генами ДНК-метиلاзы (и генами белков специфичности в случае систем I типа). Общее число эндонуклеаз рестрикции в геномах *B. longum* варьировало от 1 до 6 (с медианным значением 3). Штаммы подвида *infantis* имели только один ген предположительной эндонуклеазы рестрикции (BLIJ_0294 в штамме ATCC 15697^T), который содержал сдвиг рамки считывания в штамме ATCC 15697^T, но был, по-видимому, неповрежденным в штамме ЕК3. Кодиремый белок обладал 79% идентичности с эндонуклеазой рестрикции VbrU^{III}R из *B. breve* (Vbr_1118 в штамме UCC2003). Однако регион вокруг данного гена в штаммах *B. longum* subsp. *infantis* не содержал ДНК-метилазы, что указывает на то, что этот ген неактивен или сменил свою функцию.

8. Белки, принимающие участие в колонизации кишечника человека

Важной задачей в геномике бифидобактерий является поиск генов, влияющих на основные процессы их взаимодействия с макроорганизмом, а также с другими, в том числе, патогенными микроорганизмами.

Синтез лантионин-содержащих бактериоцинов (лантибиотиков), известных своим широким спектром активности (Lee J.-H. и др., 2011), может играть важную роль в антагонистической активности бифидобактерий против различных патогенов. Кластеры генов синтеза лантибиотиков были выявлены в геномах 4 из 28 штаммов бифидобактерий. При этом были отмечены оба подхода к посттрансляционной модификации: в штаммах *B. longum* subsp. *infantis* ATCC 15697^T и ЕК3 дегидратация аминокислотных остатков и циклизация может осуществляться белками LanB и LanC, а в *B. longum* subsp. *longum* DJO10A и KACC 91563 эти функции совмещены в одном белке LanM. В силу этого можно предположить, что и конечные продукты синтеза также могут иметь различные структуру и свойства, и нуждаются в отдельном экспериментальном изучении для каждого исследуемого штамма.

Гены пилей прочной адгезии типа IVb (O'Connell Motherway M. и др., 2011) и ген, кодирующий необходимую для их созревания препилин-пептидазу TadV, присутствовали во всех исследуемых штаммах *B. longum*, также как и ген ABC-транспортера, предположительно принимающего участие в адгезии (Shkorogov A.N. и др., 2008). Напротив, кластер генов, кодирующих сортаза-зависимые пили (Feroni E. и др., 2011), был найден только в 19 из 28 штаммов *B. longum*.

Секретируемый ингибитор сериновых протеаз (серпин), который может эффективно ингибировать панкреатическую эластазу и эластазу нейтрофилов (Ivanov D. и др., 2006), не является широко распространенным среди рода *Bifidobacterium* (Turroni F. и др., 2010), однако, кодирующий его ген, по-видимому, входит в кор-геном вида *B. longum*. Также во всех штаммах присутствовал ген, кодирующий гидролазу желчных солей (Kim G.-B. и др., 2005). На основании этого можно предположить, что эти ферменты являются жизненно важными для представителей вида *B. longum*, или, по крайней мере, значительно повышают их приспособленность. Поэтому изучение данной ферментативной активности при исследовании отдельных штаммов *B. longum* можно считать малоинформативным. Лocus синтеза ABC-транспортера фруктозы, экспрессия которого в бактериях микрофлоры кишечника способна препятствовать реализации токсического действия Шига-подобных токсинов (Fukuda S. и др., 2011), были обнаружены во всех геномах подвидов *longum* и *suis*, но не встречались у представителей подвида *infantis*.

9. Долговременная персистенция внутривидовых групп *B. longum*

Так как в данном исследовании по несколько штаммов было выделено от одних и тех же обследуемых лиц в разных возрастах, сравнение их геномов может предоставить информацию о долгосрочной штаммовой динамике в кишечной микрофлоре человека.

Как и было отмечено выше, сходство в группах штаммов 44В - 1-6В (выделенных от ребенка 1) и 35В- 22-В - 7-1В (выделенных от ребенка 2) было очень высоким. Их близкое родство может быть показано на основе нуклеотидных последовательностей консервативных генов, общего генного состава, средней нуклеотидной идентичности и таких переменных локусов, как CRISPR. Тем не менее, ни одна из пар штаммов внутри этих групп не была полностью идентична ни по одному из этих критериев. Можно заключить, что между всеми этими штаммами есть существенные генетические различия, даже с учетом поправок на нефинализированный характер сборки геномных последовательностей. По результатам анализа, выделенные от одного обследуемого ребенка штаммы 2-2В и 35В достаточно сходны, чтобы нельзя было исключить происхождение одного от другого. Однако выделенный позднее штамм 7-1В имеет более отдаленного общего предка с предком этих двух штаммов. Таким образом, данный набор бактерий представляет собой скорее случайную выборку из группы близкородственных штаммов, возможно, существующих одновременно в микрофлоре кишечника. Можно выдвинуть гипотезу, что штаммовый состав популяции *B. longum* в микрофлоре индивидуума не является случайным, и есть тенденция к персистенции узкой внутривидовой группы бактерий, наиболее приспособленной к конкретным особенностям иммунного ответа или генетически обусловленным свойствам кишечного эпителия.

ВЫВОДЫ

1. Бактерии вида *B. longum* являются наиболее часто встречающимися представителями рода *Bifidobacterium* в составе микрофлоры кишечника здоровых людей независимо от возраста. С возрастом уменьшается встречаемость видов *B. bifidum* ($p=0,020$) и *B. breve* ($p=0,0005$), а встречаемость вида *B. adolescentis* увеличивается ($p=0,0003$).
2. Подвиды *longum*, *infantis* и *suis* вида *B. longum* являются монофилетическими группами, различимыми по последовательностям консервативных генов, общему набору генов и средней нуклеотидной идентичности.
3. Внутривидовой и межвидовой горизонтальной перенос генов вносит значительный вклад в формирование генетического разнообразия представителей вида *B. longum*. Пан-геном подвида *B. longum* subsp. *longum* имеет открытую структуру, группа редко встречаемых генов обогащена генами мобильных генетических элементов и синтеза поверхностных структур бактериальных клеток.
4. Набор генов гликозил-гидролаз в геномных последовательностях *B. longum* отражает высокий генетический потенциал subsp. *longum* к утилизации растительных поли- и олигосахаридов, приспособленность subsp. *infantis* к сбраживанию углеводов животного происхождения и отсутствие выраженной специализации у представителей subsp. *suis*.

5. Установлена возможность долговременной персистенции близкородственных внутривидовых групп *B. longum* в периоде взросления продолжительностью до десяти лет.

ПРАКТИЧЕСКИЕ РЕКОМЕНДАЦИИ

1. При отборе штаммов *Bifidobacterium longum* для конструирования пробиотических препаратов рекомендуется проводить анализ геномной последовательности с целью установления присутствия генов, принимающих участие в колонизации кишечника человека и антагонистических взаимодействиях, обнаружения криптических плазмид и оценки количества мобильных генетических элементов в хромосоме.
2. Вследствие малого количества систем рестрикции-модификации в геномах *B. longum* subsp. *infantis* рекомендуется использование данных бактерий для геномной инженерии с целью создания штаммов с заданными свойствами.

ПЕРСПЕКТИВНЫЕ НАПРАВЛЕНИЯ ДАЛЬНЕЙШЕЙ РАЗРАБОТКИ ТЕМЫ

1. Важной задачей дальнейшей разработки темы является секвенирование геномных последовательностей большего числа штаммов представителей подвидов *infantis* и *suis* вида *B. longum*. Это позволит провести анализ пан-генома данных подвидов и изучение различий во встречаемости отдельных групп генов, плазмид и систем CRISPR-Cas.
2. Применение метагеномного секвенирования может предоставить более детальную информацию о долговременной персистенции внутривидовых групп *B. longum* на фоне смены видового состава бифидобактерий.

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ ПО ТЕМЕ ИССЛЕДОВАНИЯ

1. **Чаплин, А. В.** Горизонтальный перенос генов у *Bifidobacterium longum* / А. В. Чаплин // Вестник РГМУ. – 2013. – Специальный выпуск № 1. – С. 230.
2. **Чаплин, А. В.** Сравнительная геномика штаммов *Bifidobacterium longum*, колонизирующих кишечник человека / А. В. Чаплин // Ломоносов – 2014: XXI Международная конференция студентов, аспирантов и молодых ученых; секция «Биология»; 7-11 апреля 2014 г; Москва, МГУ имени М. В. Ломоносова, биологический факультет. Тезисы докладов. – Москва, 2014. – С. 196.
3. **Чаплин, А. В.** Изучение видового разнообразия бактерий рода *Bifidobacterium* кишечной микрофлоры с использованием метода MALDI-TOF масс-

спектрометрии / А.В. Чаплин, А.Г. Бржозовский, Т.В. Парфёнова, Л.И. Кафарская, Н.Н. Володин, А.Н. Шкопоров, Е.Н. Ильина, Б.А. Ефимов // Вестник РАМН. – 2015. –Т. 70– № 4 – С. 435-440.

doi:10.15690/vramn.v70.i4.1409

4. Shkoporov, A. N. Draft Genome Sequences of Two Pairs of Human Intestinal *Bifidobacterium longum* subsp. *longum* Strains, 44B and 1-6B and 35B and 2-2B, Consecutively Isolated from Two Children after a 5-Year Time Period / A. N. Shkoporov, B. A. Efimov, E. V. Khokhlova, A. V. Chaplin, L. I. Kafarskaya, A. S. Durkin, J. McCarrison, M. Torralba, M. Gillis, G. Sutton, D. B. Weibel, K. E. Nelson, V. V. Smeianov // Genome Announc. – 2013. – Т. 1 – № 3.
PMID: 23682142. doi: 10.1128/genomeA.00234-13
5. Chaplin, A. V. Intraspecies Genomic Diversity and Long-Term Persistence of *Bifidobacterium longum*. / A. V. Chaplin, B. A. Efimov, V. V.Smeianov, L. I. Kafarskaia, A. P. Pikina, A. N. Shkoporov // PLoS One. – 2015. – Т. 10 – № 8.
PMID: 26275230. doi: 10.1371/journal.pone.0135658

СПИСОК ИСПОЛЬЗУЕМЫХ СОКРАЩЕНИЙ

е. а. – единицы активности

КОЕ – колониеобразующие единицы

ПЦР – полимеразная цепная реакция

п. о. – пары оснований

ANI (англ. Average Nucleotide Identity) – средняя нуклеотидная идентичность

CRISPR (англ. Clustered Regularly Interspaced Short Palindromic Repeats) – сгруппированные регулярно расположенные короткие палиндромные повторы

MALDI-TOF MS (англ. Matrix Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry) – времяпролетная масс-спектрометрия с матрично-активированной лазерной десорбцией/ионизацией

subsp. (англ. subspecies) – подвид

VNTR (англ. Variable Number Tandem Repeats) – переменное число tandemных повторов