What Pediatricians Need to Know about the Analysis of the Gut Microbiota

Gerald W. Tannock

Key Points

- The biodiverse, individualistic bowel microbiota of humans is acquired by a sequential process that produces a characteristic biological succession.
- Bifidobacteria predominate in the infant gut and may have a long-lasting impact on the physiology of the child (biological Freudianism).
- The mechanisms that regulate the composition and activities of the bowel community likely involve competitive exclusion and efficient regulation of microbial physiology.
- Nucleic acid-based methods of analysis are widely used to determine and monitor the composition of the bowel microbiota. This is because, currently, a large proportion of the members of the bacterial community cannot be cultured under laboratory conditions by traditional bacteriological methods.

Key Words: Microbiota, bowel, bifidobacteria, molecular analysis, infants.

Over evolutionary time, humans have developed an equilibrium with the microbial world, which consists of cloaking the body inside and out with microorganisms that are likelier to be friends than enemies. (Abigail Salyers and Dixie Whitt [microbiologists])

INTRODUCTION

Fabrication of the metaphorical cloak of microbes begins within hours of birth when bacteria from the environment, the mother and other humans who interact intimately with the baby inoculate the sterile skin and accessible body cavities of the infant. The baby is exposed to a diversity of microbial life during the ensuing days, weeks and months yet only certain bacterial types find the body to be “fertile soil” that provides their required carbon and energy sources and other physicochemical factors conducive to their life. It is easy to imagine that the infant body is assailed from every direction by microbes, and studies of the temporal acquisition of the gut microbiota, for example,
show that the composition of the microbial community of the bowel, essentially composed
in modern times of bacteria, progresses in the form of a characteristic biological succe-
sion [1–8]. The first heterogeneous collection of bacterial species quickly simplifies and
facultative anaerobes (Escherichia coli and enterococci) assume numerical dominance.
Obligately, anaerobic bacteria belonging to the genus Bifidobacterium are also numerous
and, by the age of three months, are the predominant members of the bowel microbiota,
even in formula-fed babies [1, 6, 9–11]. To all intents and purposes, the neonate during
the first few months of life is almost, in gnotobiotic terms, a monoassociated animal,
such is the simplicity in composition of the microbiota and the numerical predominance
of bifidobacteria.

Gnotobiotic mouse experiments have demonstrated that, at least in the short term, the
presence of bacteria in the bowel influences the physiology of the animal host. Particularly
striking have been the reports of the effects on murine gene expression in relation to the
epithelial barrier and the induction of angiogenesis in the ileum, and the accumulation
of body fat [12–15]. If these reported effects on mice are applicable to humans, then the
biological successions occurring in early life become pivotal in the development of the
child. Colonization of the bowel by certain bacterial species may have lifelong conse-
quences. Dubos et al. [16] aptly referred to this enduring influence of early microbial
associations as “biological Freudianism.” Microbial influences replace, in this view,
environmental and social factors that mould the “unconscious.” We are not conscious of
the impact of early microbial associations, but their long-lasting influences shape our
adult physiology – the microbiological past is alive in the physiological present.

Immune deviation, the process in which the bias toward a Th2 response by the
immune system of the infant is removed, may be an important consequence of expo-
sures to microbes in early life [17,18]. Many affluent countries have experienced an
increase in the prevalence of allergic diseases in recent decades, including atopic der-
matitis, asthma, and atopic rhinitis [19]. With reference to “biological Freudianism,”
colonization of the infant bowel by specific species of bacteria might be important in
the initial regulation of the developing immune system. Modern lifestyles and environ-
ments may alter qualitative exposure of infants to bowel commensals and this might
influence the risk of atopic diseases [20,21]. Members of the genus Bifidobacterium are
likely to be important bacteria in this respect because, as indicated above, they form a
major portion of the bowel community in early life.

The rules and regulations that govern the bacterial community resident in the bowel
must be legislated anew each time a neonate is colonized. This is because the bacterial
collection is not the same in every human [22,23]. Probably fundamental to the homeos-
tasis of ecosystems associated with the human body is the phenomenon of “competitive
exclusion,” which is particularly well known in the case of the bowel community [24].
Long ago demonstrated in gnotobiotic experiments, the presence of the microbiota
enhances nonspecific resistance to infection. For example, germfree mice can be infected
by the oral route by a dose of Salmonella typhimurium as small as ten cells; the infectious
dose for conventional mice is about 10⁹ cells. The self-regulated, homeostatic community
already established in the conventional bowel provides a hurdle that only large numbers
(a high dose) of pathogenic cells can surmount [25]. We are not exactly sure how com-
petitive exclusion is mediated mechanistically but may best be summarized in the “niche
exclusion principle”: two species cannot simultaneously occupy the same ecological
niche [26]. Only the better adapted will be successful. This can easily be envisaged in the case of nutritional competition because the species/strain that best binds and transports a source of energy and carbon into its cells will out-compete a biochemically less capable organism. Not to be forgotten, moreover, is the production of antimicrobial molecules that could give a competitive edge by altering the chemical environment, making it unsuitable for growth of other species. Short chain fatty acids can be invoked in this respect, as can hydrogen sulfide, and perhaps “bacteriocins” [25,27]. Multiple mechanisms must participate synergistically in the control of populations within complex communities. It is extremely difficult, however, to define competitive mechanisms even using experimental animal models. The order in which bacterial strains are introduced into the experimental system can ordain which of the two organisms eventually dominates the ecosystem numerically [28]. The diet fed to experimental animals can also influence the outcome of competition experiments [29,30]. By changing the diet, the number and types of available ecological niches are changed. The bowel community is composed of hundreds of species, which in turn implies that this is the number of ecological niches in the ecosystem – the more niches, the more biodiversity. The diversity of bacterial types in the human bowel reflects, therefore, the intensely competitive nature of this ecosystem in which mutations and horizontal gene transfer have permitted adaptation of bacteria to perform diverse functions in the bowel community.

Clearly, investigations of the impacts of commensal bacteria on the infant in early life require analytical procedures that can be used in the laboratory to monitor the composition and activities of the bacterial community of the bowel. Much of the bacteriological information of the bowel community has been generated through the application of nucleic acid-based methodologies, most of which rely on the nucleotide base sequence of small ribosomal subunit RNA (16S rRNA in the case of bacteria) which provides a cornerstone of microbial taxonomy. Nucleic acid-based methods of detection suggest that about 50% of the bacterial cells seen microscopically in feces of adults cannot yet be cultured in the laboratory, even when accounting for the fact that some of the bacteria are dead [31]. This phenomenon, also manifested even more dramatically in terrestrial and aquatic ecosystems was, based on traditional bacteriological experience, totally unexpected and has been called “the great plate count anomaly” [32]. Operational taxonomic units (OTU; molecular species) never encountered in culture-based bacteriology are detectable by the molecular methods, revealing a new world remaining to be investigated by bacteriologists of the future. Although, in the case of infants, the microbiota is dominated in early life by the bifidobacteria, which are relatively easily cultured under laboratory conditions, nucleic acid-based analytical methods provide evidence of the increasing complexity of community composition as babies grow and develop.

The starting point for nucleic acid-based, analytical methods is the extraction of bacterial DNA or RNA directly from the fecal or other sample of interest, avoiding the need to cultivate any members of the bacterial community. 16S rRNA or the gene that encodes it has become a cornerstone of bacterial classification because it contains regions of nucleotide base sequence that are highly conserved across the bacterial world and that are interspersed with variable regions (V regions). These variable regions contain the “signatures” of phylogenetic groups and even species [33]. For this reason, variable regions of 16S rRNA (or 16S rRNA gene sequences) are the basis of the analytical methods. Bacterial DNA or RNA extracted from the samples (in theory nucleic acid from all of the bacterial
types in the sample will be represented in the extracts) and polymerase chain reaction (PCR) amplification (reverse transcription-PCR in the case of RNA extracts) of the 16S rRNA gene in part or complete, is carried out. Clone libraries of the 16S rRNA genes can be made and sequenced, or emulsion-based high-throughput PCR sequencing can be carried out, producing a catalog of the bacterial constituents of the ecosystem [23,35–37]. It is unfortunate that as much as 5% of the 16S rRNA gene sequences in databanks are inaccurate. This is because the results of high-throughput sequencing projects have polluted the DNA databanks with unreliable 16S rRNA gene sequences making meaningful analysis of catalogs of bacterial inhabitants difficult to achieve [38,39].

Nevertheless, from this sequence information, DNA probes can be designed. DNA (oligonucleotide) probes that target specific rRNA sequences (16S or 23S) within ribosomes, to which they hybridize, are used in this method [40–45]. The probes are 5' labeled with a fluorescent dye, which permits both detection and quantification of specific bacterial populations (fluorescence in situ hybridization [FISH]). Bacterial cells within which hybridization with a probe has occurred fluoresce and hence can be detected and counted. Permeabilization of the bacterial cells is required in order to standardize intracellular access of DNA probes to their targets [46]. Within the cell, the secondary structure of rRNA molecules and their molecular interactions within the ribosome may hinder the access of the probes to their target sites. In situ accessibility influences the amount of fluorescence generated from the probe [47]. A high degree of in situ accessibility would facilitate the binding of the probe to its target site and therefore permit the probe to emit a bright fluorescence signal. The determination of the brightness of fluorescence (probe relative fluorescence) conferred by a probe is a means of evaluating its in situ accessibility [46,47]. Modeling of the secondary structure of 16S rRNA molecules allows in silico investigation of the in situ accessibility of the entire molecule [47–49] and the target site can be assessed in terms of accessibility. If the target region is located in a poor or nonaccessible site, helper probes (unlabeled oligonucleotides) can be derived that are complementary to regions adjacent to the probe’s target site, promoting the binding of the probe and therefore amplifying the fluorescence signal [49–51]. Other important technical considerations include (i) the physiological state of the bacterial cells because the number of ribosomes per bacterial cell is greater the higher their metabolic activity. Therefore, bacterial cells in a quiescent state have weak fluorescence and may not be detected. (ii) the degree of hybridization stringency, which depends on three factors: temperature, salt concentration, and formamide concentration of the hybridization solution. The manipulation of these factors influences the specificity of hybridisation and hence detection and quantification. (iii) a 16S rRNA gene database that is rapidly increasing in size. More than 600,000 16S rRNA sequences are available from the Ribosomal Database (http://rdp.cme.msu.edu), permitting the in silico development and validation of a large panel of probes targeting different phylotypes resident in the human bowel. Several of the currently used probes were designed and tested using older versions of the Ribosomal Database so continual reassessment of specificity and coverage of these probes is essential in order to update and confirm their continuing reliability. (iv) epifluorescence microscopic detection and quantification of bacterial populations was used originally to enumerate fluorescent cells but, because of the laborious and time-consuming nature of this work, automated systems have been developed [52], culminating in the use of flow cytometry to count
the fluorescent cells. Rapid and easy to set up, flow cytometry combines quantitative and multifactor analysis (size, internal granularity, fluorescence signal).

Deriving a catalog of bowel inhabitants, or enumerating groups of bacteria using DNA probes for every sample that needs to be investigated is a daunting task. A relatively simple, semiquantitative screening method to compare the bacterial composition of multiple samples is provided by PCR combined with denaturing gradient gel electrophoresis (DGGE) or temporal temperature gradient gel electrophoresis (TTGE). This approach, which provides comparative “snapshots” of microbiota compositions, has been demonstrated to have enormous utility in a number of bowel microbiota studies [53–56]. DNA or RNA is extracted directly from intestinal or faecal samples. Then a variable 16S rRNA gene sequence is amplified using PCR primers that anneal with conserved sequences that span the selected V region. One of the PCR primers has a GC-rich 5′ end (GC clamp) to prevent complete denaturation of the DNA fragments during gradient gel electrophoresis. To separate the 16S fragments amplified from different types of bacteria and present in the PCR product, a polyacrylamide gel is used. The double-stranded 16S fragments migrate through a polyacrylamide gel containing a chemical or thermal gradient until they are partially denatured by the chemical or temperature conditions. The fragments do not completely denature because of the GC clamp, and migration is radically slowed when partial denaturation occurs. Because of the variation in the 16S sequences of different bacterial species, chemical and thermal stability is also different; therefore different 16S “species” can be differentiated by this electrophoretic method. 16S rRNA gene fragments from different bacteria have different migration distances in the electrophoretic gel and a profile of the numerically predominant members of the microbiota is thereby generated [57, 58]. Individual fragments of DNA can be cut from electrophoretic gels, further amplified and cloned, then sequenced. The sequence can be compared to those in gene databanks in order to obtain identification of the bacterium from which the 16S sequence originated. In a further development of this methodology, PCR primers specific for bacterial groups can be derived. These primers generate an electrophoretic profile of the species comprising a specific bacterial genus, for example bifidobacterial species, within the bacterial community [59–61].

Specific strains of bacteria can be differentiated by producing their DNA fingerprint, much as is done in human forensics. Chromosomal DNA is extracted from pure cultures of bacteria. The DNA is digested by a restriction endonuclease chosen on the basis of the mol % G + C content of the DNA of the bacterial species and on the recognition sequence of the enzyme. An endonuclease that will cut the DNA rarely is desired so that a relatively small number of DNA fragments result from the digestion and a relatively simple pattern will be generated in the electrophoretic gel. The digestion generates large fragments of DNA that would not separate by the usual agarose gel electrophoresis that is based on molecular sieving. Therefore, pulsed field gel electrophoresis (PFGE) is used in which the mixture of fragments in the DNA digest are exposed to alternating electrical fields that force the fragments to change orientation rather than to migrate through the agarose gel immediately after the electrical field is changed from one direction to another. The rate of reorientation is size dependent, so larger molecules change direction more slowly than smaller ones. The pulse time (the time spent in a field of particular direction) is varied and this dictates the DNA class size that spends most of the time reorientating rather than migrating. The DNA fragments are thus separated
by the retardation of net movement rather than by sieving. The pattern of fragments generated in the gel represents the genetic fingerprint of the bacterial culture and is characteristic of that strain of bacteria [62]. A particularly useful application of PFGE of DNA digests is to provide a means by which a probiotic bacterial strain can be tracked during the course of a probiotic study. The bacterial group of interest can be selectively cultured and colonies are randomly picked to obtain pure cultures. The genetic fingerprint of these isolates is then determined by PFGE of DNA digests, and compared with that of the probiotic bacterial strain. The presence or absence of the probiotic strain in fecal samples can be determined by this method. The advantage of PFGE of DNA digests is that a specific strain can be tracked during studies aimed at determining the persistence of the probiotic strain in the gut. Disadvantages include the requirement for bacteriological culture and the immense logistical effort required to genetically fingerprint hundreds or thousands of bacterial isolates.

An example of the use of genetic fingerprinting in a probiotic study is provided by the work of Tannock et al. [63] who analyzed the composition of the *Lactobacillus* populations present in the feces. The composition of the fecal bacterial community of ten human subjects was monitored before (control period of six months), during (test period of six months) and after (posttest period of three months) the administration of a milk product containing *Lactobacillus rhamnosus* DR20 (daily dose of $1.6 \times 10^9$ lactobacilli). The composition of the *Lactobacillus* population of each subject was analyzed by PFGE of bacterial DNA digests in order to differentiate between DR20 and other strains present in the fecal samples. Consumption of the probiotic transiently altered the composition of the *Lactobacillus* populations of the subjects, but to varying degrees. The detection of DR20 among the numerically predominant strains was related to the presence or absence of a stable autochthonous population of lactobacilli during the control period. The probiotic strain did not predominate in samples collected from subjects with *Lactobacillus* populations of stable composition.

What predicates which bacterial species or strain will establish in the infant bowel? The short answer is probably: only bacteria that can utilize the substrates provided by the diet and the particular human host. Bifidobacteria, *E. coli* and enterococci can utilize a wide range of monosaccharides and oligosaccharides, which would be provided by the diet. Genomic analysis of bifidobacterial species shows that these bacteria are indeed endowed well with enzymic capacity to hydrolyze and ferment oligosaccharides [64, 65]. After weaning, however, the range of fermentable substrates available to the bacteria changes from monosaccharides and oligosaccharides to complex plant polymers (dietary fiber) that pass undigested through the small bowel and hence become one of the principle sources of carbon and energy for bowel bacteria [66]. The other major source of complex carbohydrates is provided by the mucins (constituents of mucus) that are continually secreted into the bowel from goblet cells present in the mucosal lining [67]. Assessment of obligately anaerobic inhabitants of the adult bowel through analysis of a few representative fully sequenced genomes, as well as by studies of community genetics [68–70] show that they are superbly adapted to digesting complex polysaccharides. The bacteria produce numerous hydrolytic enzymes and can regulate their use according to the kinds of substrates that they sense in their environment. Strict regulation of catabolic pathways must be an extremely important attribute in a habitat where the nutritional profile will vary from day to day according to the omnivorous and varied dietary preferences.
Chapter 2 / What Pediatricians Need to Know about the Analysis of the Gut Microbiota

of the human host, and helps to explain the remarkable consistency in biochemistry and biodiversity of the human bowel [71].

Increasingly, it becomes clear that, because of the individualistic compositions of the microbiota, phylogenetic analysis of the bowel community of infants may not offer useful information, beyond that which has already been accumulated [6]. New methodologies are required to explore the impact of the maturing bowel microbiota of the developing child. “Who is there?” needs to be replaced by “What are they doing, and how are they doing it?” We need to define the molecular webs that interconnect bacteria-bowel milieu-infant mucosa. In this view, the identity of the phylogenetic entities inhabiting the bowel is minimized and their biochemistry is emphasized. Because much of the microbiota has not yet been cultivated, new culture-independent methodologies must be invented and applied to investigations of the bowel.

Metagenomics is a facet of microbial ecology in which a microbial community is studied in terms of its collective genomes (community genetics), rather than focusing on the diversity of species and their individual genomes [32]. For functional studies, the metagenomic approach traditionally entails the cloning of large fragments of community genomic DNA that have been extracted directly from the ecosystem of choice. The cloned DNA fragments are large enough to encode operons and therefore might result in the expression, by a surrogate bacterial host, of several enzymes that could catalyze a relatively complex metabolic process, including the synthesis of secondary metabolites. Metagenomic libraries derived from microbial community genomes can be screened for heterologous phenotypic traits that include enzymes and other proteins that are essential to the functioning of the ecosystem. Hence they provide a means of accessing and assessing details of community biochemistry through its underpinning genetics.

Measurement of the impact of bacteria on the transcriptome of the bowel mucosa of the child, such has been achieved with experimental animals, would be ethically and technically difficult to achieve. A new approach to determine the impact of the microbiota on the infant’s tissues is therefore required. Metabolomics is the nontargeted, holistic, quantitative analysis of changes in the complete set of metabolites in the cell (the metabolome) in response to environmental or cellular changes [72]. Metabolites are low-molecular weight organic compounds (<1000 Da) that participate in general metabolic reactions, or, are required for maintenance, growth, and normal functioning of a cell. Changes in cellular physiology are amplified through transcription of genes and translation to proteins but, due to regulatory mechanisms and/or substrate availability, a tenfold increase in concentration of a transcript or enzyme is not necessarily reflected in a tenfold increase in a particular cell activity. Alterations in transcriptome or proteome do, however, have large effects on the concentrations of intermediary metabolites in the cell because they reflect the activities of metabolic pathways. Of particular importance is the ability of metabolomics to penetrate the mechanisms of intracellular signaling in which both concentrations of metabolites and their associated dynamics are important. Whereas knowledge of the intracellular metabolites (metabolic fingerprint, the endometabolome) is essential in this work, changes in the physiology of the bowel ecosystem could be more easily revealed by investigation of the exometabolome (metabolic footprint) represented by the extracellular milieu which contains metabolites secreted or consumed by bacteria in the bowel [73]. The metabolic footprint of the bowel bacteria is reflected in the metabolome of the animal host because bacterial metabolites are absorbed from the gut lumen into the lymph and
blood circulations. Hence, the body fluids (blood, lymph, bile, sweat, urine) of the host contain numerous bacterial products that may provide biomarkers of food–microbe–host interrelationships and possible indicators of health or disease. The host metabolome is the sum of the interacting metabolomes of the whole organism and thus represents the end product of genetic, environmental, and host–bacterial relationships. The study of microbiota and host metabolomes might therefore contribute to a full systems biology approach to understanding and maintaining bowel health of the infant [74]. Preparation of the blueprint of the interactive bowel networks will require a systems biology investigation encompassing a diversity of scientists from different disciplines. The primary aim of the research will be to understand how all of the heterogeneous parts (dietary components, bacterial consortia, host physiology, and development) are integrated in early life, with a supplementary aim of identifying biomarkers of health or disease. A fusing of biological and computational expertise will be required for success.

> From all points of view, the child is truly the father of the man, and for this reason we need to develop an experimental science that might be called biological Freudianism. Socially and individually the response of human beings to the conditions of the present is always conditioned by the biological remembrance of things past. (Rene Dubos, Dwayne Savage, and Russell Schaedler [microbiologists])

**SUMMARY AND CONCLUSION**

The infant bowel becomes colonized by a biodiverse collection of bacteria soon after birth. A regulated process (succession) can be recognized in which the proportions of different bacterial groups comprising the microbiota change during the first few years of life. The physiological impact of this bacterial succession may have long-lasting, physiological consequences. Much of the microbiota is uncultivable by traditional bacteriological methods, therefore nucleic acid-based analytical methods are widely used to appraise the state of the microbiota. The highly individualistic compositions of individual bowel communities, even in infants, confound comparisons of dietary and other environmental influences. Increased understanding in the future of bacteria–host interactions will probably result from the application of advanced chemical analyses.

**REFERENCES**


Chapter 2 / What Pediatricians Need to Know about the Analysis of the Gut Microbiota


