Impact of commensal microbiota on murine gastrointestinal tract gene ontologies

David M. Mutch,1,2 Rainer Simmering,1 Dominique Donnica,1 Grigorios Fotopoulos,1 James A. Holzwarth,1 Gary Williamson,1 and Irène Corthésy-Theulaz1
1Nestlé Research Center, Vers-Chex-les-Blanc, CH-1000 Lausanne 26; and 2Center for Integrative Genomics, Université de Lausanne, CH-1015 Lausanne, Switzerland

Submitted 3 May 2004; accepted in final form 24 June 2004

The gastrointestinal tract (GIT) comprises a highly dynamic environment, where nutritional, medicinal, hormonal, and endogenously synthesized molecules are utilized and metabolized by both the host and microbiota alike. However, several lines of evidence indicate that one should not consider gut microbiota as simply autonomous residents of the host GIT, but rather important contributors in defining the gut ecosystem (17). Previous research has implicated nonpathogenic bacteria as factors regulating mucosal inflammation (11, 27) and host defense (13). Enteric bacteria have also been demonstrated to supply energy substrates to epithelial cells, such as short-chain fatty acids produced through bacterial fermentation (11). Indeed, an essential partnership has evolved between gastrointestinal epithelial cells, immune cells, and resident bacteria for which the normal function and activity of each is dependent on the other two components (23). Although the focus herein will be on commensal microbiota, i.e., those microbial species present in the healthy gut, emerging evidence suggests that many microbial species (pathogenic, symbiotic, and commensal) rely on and/or exploit this partnership to survive.

Colonization of the mammalian GIT begins as the fetus passes through the birth canal, where the establishment and maintenance of the microbial community is influenced by the method of birth, health status, the organism’s age, and the diet (13, 23). The controlled exposure to foods during an organism’s growth stage leads to separate phases during the colonization process, i.e., aerobes and facultative bacteria followed by anaerobes. This process is shared by both humans and rodents (33). However, it is poorly understood to what extent these microbial species can influence the development and functions of the host GIT. With over 400 different species thought to inhabit the adult gut, of which only a small percentage have been successfully cultured in vitro, it is clear that the complete understanding for the presence and function of all bacterial species is far from ascertained (14, 41).

The development of germ-free animals have yielded in vivo models with which the scientific community can begin to decipher the physiological consequences of pathogenic, symbiotic, and resident bacteria. Wostmann (42) has thoroughly reviewed the various biological differences between conventional (i.e., mice with a normal, nonpathogenic gut microbiota) and germ-free rodents, which are by far the most widespread and practical germ-free models used. In addition to differences in intestinal morphology and immune functions, the presence/absence of microbiota modulates water absorption, luminal pH, the concentration of deconjugated bile acids, and the oxidation-reduction potentials of the gut (42). Furthermore, significant evidence indicates that the gut microbiota plays an important role in the metabolism and bioavailability of nutritional compounds, such as flavonoids (36), phytoestrogens (32, 39), and carbohydrates (5). However, identifying the underlying molecular players governing these biological functions was difficult prior to the complete characterization of the human and mouse genomes. The advent of global gene expression platforms has heralded an era in which the interspecies molecular cross talk between bacteria and mammalian cells can now be explored (31).

Two important studies have emerged in which the effects of nonpathogenic bacteria on the host gut transcriptome have been analyzed (11, 15). Both studies provide first glimpses of those genes modulated by the commensal bacteria; however, neither study compared the expression profiles between germ-free mice and conventional mice in multiple regions of the GIT and related these gene expression changes to significantly modified gene ontologies. As hundreds of different microbial
species may be present at any given time in the GIT, it has yet to be described how this complex bacterial community interacts with the host to modulate function along the length of the GIT. Furthermore, neither study utilized robust statistical models for the identification of differentially expressed genes, which is critically important for minimizing the number of false-positive and false-negative results in order to yield a biologically accurate interpretation of the microarray results.

In this paper the expression profiles of the corpus, jejunum, descending colon, and rectum of germ-free and conventionalized mice were examined. Using a novel method of clustering based on Gene Ontology (GO) annotations, we have identified those biological functions that are significantly modified by the commensal microflora. Certain differentially regulated genes contribute directly to significantly modified functional biological functions displaying regional specificity; however, we find that the most significant biological processes altered in the host GIT by the presence of microbiota are immune response and water transport. Furthermore, immune response and water transport are consistently modulated in all GIT regions examined, reinforcing the concept that the normal healthy gut is, in part, dependent on the molecular relationship between the host and commensal microflora.

MATERIALS AND METHODS

Animals. Male C3H mice (Nestlé Research Centre, Lausanne, Switzerland) with an average age of 6 wk were placed in Macrolon cages (5 mice per cage) and housed in a room with a 12:12-h light/dark cycle and at a temperature of 22 °C. All mice used in this study were born under germ-free conditions. Some of the axenic mice were transferred immediately after birth into Macrolon cages with two adult females and housed under conventional conditions to constitute the conventional group. To decrease the time of conventionalization, fecal pellets from conventional mice were placed into the cage. The bacterial status of mice was checked by: 1) culturing intestinal samples with selective agar plates and 2) analyzing these samples by FISH technology using group-specific probes (data not shown). The cages of the axenic mice were arranged in sterile isolators equipped with a sterile water supply. All mice were given access to a sterilized standard diet (UAR 03-40; UAR, Villemoisson, France) and water ad libitum.

Tissue dissection. Mice were euthanized using 3% isoflurane at an age of 9 wk. Whole tissues corresponding to the corpus, jejunum, descending colon, and rectum were dissected from each mouse, washed with ice-cold PBS, placed into 2 ml of RNA-Later (Ambion, Cambridgeshire, UK) and frozen in liquid nitrogen. The corpus was defined as the part of the stomach different from the antrum (distinguished by a difference in color). The jejunum was defined as the tissue between the first 4 cm and the last 2 cm of the small intestinal tube. The descending colon and the rectum were defined as the last 3 cm of the digestive tube, where the first 2 cm was the descending colon and the last centimeter was the rectum. The handling and euthanization of animals was compliant with federal, state, and local laws and regulations and was approved by the State Veterinary Office.

Microarray analysis. Total RNA, corresponding to the four aforementioned whole tissues, was extracted from both conventional (n = 3) and germ-free mice (n = 3) using the Tripure Isolation reagent (Roche Diagnostics, Rotkreuz, Switzerland). RNA was then repurified using the Nucleospin kit, and contaminating genomic DNA was removed by DNase I treatment (Macherey-Nagel, Oensingen, Switzerland). All samples were monitored with the Agilent 2100 Bioanalyzer (Agilent Biotechnologies, Boeblingen, Germany) and consistently demonstrated high-quality RNA (28S/18S ratio ~2, but always less than 3). The cRNA synthesis, hybridization, and scanning procedures were performed as previously described (26). Samples were hybridized to the Affymetrix Mu74Av2 GeneChip (Affymetrix UK, High Wycombe, UK), which represented all sequences in the murine UniGene database at the time this study was performed (~6,000 unique genetic elements) and 6,000 expressed sequences tags (ESTs). The complete data set is publicly available in the NCBI Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) through accession number GSE1392.

Identifying differentially expressed genes. Differentially expressed genes were identified using a global error assessment (GEA) method of analysis with a conservative alpha value (α ≤ 0.001). For a detailed description and general protocol of the GEA model, see Mansourian et al. (22).

Gene Ontology clustering. GO stems from a common effort to organize genes (or proteins) hierarchically to exploit known relationships between these entities (1, 38). GO is organized as a set of three inverted trees (molecular function, biological process, and cellular compartment), with the root being at the top of the tree and the leaves at the bottom. The root is the most general category (herein referred to as GO node), whereas the leaves are the most specific (http://www.geneontology.org/). This leads to the concept of depth (i.e., how far down the tree a given GO node is present). The lower down the tree a GO node is, the more specific is the description and vice versa.

A background distribution of annotations for all known murine protein sequences was calculated as a reference using the GO annotations generated by the European Bioinformatics Institute (EBI) for Swiss-Prot/TrEMBL (http://www.geneontology.org/GO.current.annotations.shtml). The annotations are calculated for each GO node in each of the three trees. Using these annotations, we determined the proportion mapping to each GO node and used this as the background binomial distribution. The background distribution of genes yields two values, defined as a and t, where a refers to the number of genes annotated to a specific GO node and t refers to the total number of genes assigned to all GO nodes. Therefore, the probability of a randomly selected gene being annotated to a specific GO node is defined as

\[
p_i = \frac{a}{t} \tag{1}
\]

An α-value is calculated for each GO node by comparing the number of IDs in the list of differentially expressed genes in proportion to the size of the list with the background distribution, i.e., the murine genome. This α-value indicates whether a given GO node is significantly overrepresented within the list of genes identified as differentially regulated. More specifically, the α-value is calculated as the probability that x of n genes are annotated to a given GO node, where x represents the genes annotated to a specific GO node and x represents the total number of genes considered differentially regulated. This is determined using the following equation:

\[
\sum_{j=x}^{n} \left( \frac{n!}{j!(n-j)!} \right) \times p^j \times (1-p)^{n-j} \tag{2}
\]

where p refers to the value obtained with Eq. 1.

Equation 2 was based on the ProToGO server calculation (http://www.protonet.cs.huji.ac.il/ProToGO/Introduction.html) and was calculated for each gene in our lists of differentially expressed genes. The final results were then visualized using GraphViz software (12).
Translated from German

RESULTS

Transcriptomic analysis of germ-free and conventional mice. To concentrate on only those genes that are differentially regulated between conventional and germ-free mice, a conservative \( \alpha \)-value was set \((\alpha = 0.001)\) for the GEA analysis. Despite this conservative \( \alpha \)-value, a large number of genes were identified as differentially regulated within each of the four tissues; however, not all of the genes have been annotated. Furthermore, as more than one probe set may correspond to a single gene, Ensembl identifiers for each probe set were obtained and used to determine the number of unique genes modulated within each tissue. The numbers of differentially regulated probe sets and the number of those probe sets annotated in each tissue were as follows: the corpus, 207 probe sets (140 annotated corresponding to 139 unique genes); the jejunum, 267 probe sets (179 annotated corresponding to 174 unique genes); the descending colon, 204 probe sets (129 annotated corresponding to 126 unique genes); and the rectum, 137 probe sets (73 annotated corresponding to 70 unique genes). Figure 1 illustrates the number of annotated genes in each GIT region; however, additional information is comprised within Fig. 1. First, the direction of all differentially regulated genes is indicated, i.e., up- or downregulated by the presence of conventional bacteria (see Supplemental Table S1, for a detailed list of these genes, available online at the Physiological Genomics web site).\(^1\) In all four tissues, the majority of genes (>50%) had increased expression levels in conventionalized mice. Although fewer genes were differentially regulated in the rectum, nearly 70% of these genes were upregulated in the presence of the microbiota. Second, both the number of genes unique to each region and the number of genes present in at least one other GIT region are indicated. In all tissues examined, the majority of differentially regulated genes were not found to be region specific. Finally, those genes that are highly regulated (greater than 3-fold) are depicted and serve as a means to demonstrate that few genes are highly modulated by the conventional microbiota. Despite the fact that the many genes are differentially expressed within each of the four tissues, GO clustering provided a means to identify, group together, and concentrate on only those genes that contribute to specific biological functions. Therefore, the following results will discuss only those significant GO molecular functions and biological processes that were identified via GO clustering.

Molecular functions modulated by the intestinal microbiota. Interestingly, examining the data at a global, unbiased level, i.e., not focusing on any single gene or biological function, revealed that relatively few molecular functions are affected by the microbiota of the digestive tract. Molecular function has been defined by the GO Consortium as the tasks performed by individual gene products without taking into account where and when the specific biochemical activity actually occurs (e.g., enzyme or transporter) (1, 38). In the current study, the most significant molecular function affected along the GIT by the presence of commensal bacteria was ‘defense/immunity protein activity’ (GO depth 1), which has been recently redefined as “reactions, triggered in response to the presence of a foreign body or the occurrence of an injury, which result in restriction of damage to the organism attacked or prevention/recovery from the infection caused by the attack” (http://www.geneontology.org). This molecular function was significantly modulated in all four tissues: corpus, \(5.6 \times 10^{-6}\); jejunum, \(1.1 \times 10^{-4}\); descending colon, \(3.0 \times 10^{-6}\); and rectum, \(1.8 \times 10^{-5}\). As illustrated in Fig. 2, all four tissues had genes belonging to this GO category and the genes contained in this GO node were consistently upregulated (~2- to 3-fold) in the presence of a gut microbiota. The specific genes comprised within these four tissues are listed in Table 1, where two genes had two corresponding probe sets. As can be seen in Fig. 2, the tissues did not have an equivalent number of genes in this GO node; however, several of the genes were differentially expressed in all four tissues (Table 1). Previous findings have revealed that germ-free animals have a modulated immunoglobulin (Ig)

\(^1\)The Supplementary Material for this article (Supplemental Table S1) is available online at http://physiolgenomics.physiology.org/cgi/content/full/00105.2004/DC1.
More specifically, germ-free rats were characterized as having equivalent levels of IgM as conventional mice, low levels of IgG, and a total absence of IgA in the serum; however, exposure to a conventional microbiota slowly increased circulating IgA levels (42). Furthermore, evidence exists suggesting that the interaction between microbiota and gut-associated lymph tissue is important for the diversification of antibodies (19). Therefore, the decreased expression of genes involved in defining both the variable (V) and constant (C) regions of IgGs suggests that the altered Ig levels previously reported may arise by transcriptional regulation. Interestingly, all of the genes are expressed in the small and large intestine, whereas only a subset is expressed in the corpus. This makes intuitive sense, as the number and quantity of bacteria present

Table 1. Description of differentially expressed genes in the defense/immunity protein activity molecular function cluster

<table>
<thead>
<tr>
<th>Gene Number</th>
<th>Probe Set ID</th>
<th>Description</th>
<th>Swiss-Prot ID</th>
<th>Tissue (C, J, D, R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1; 2</td>
<td>97564_f_at; 101656_f_at</td>
<td>Ig kappa chain V-II region 26-10</td>
<td>P01631</td>
<td>C, J, D, R</td>
</tr>
<tr>
<td>3</td>
<td>96972_f_at</td>
<td>Ig kappa chain V-IV region S107B precursor</td>
<td>P01680</td>
<td>C, J, D, R</td>
</tr>
<tr>
<td>4</td>
<td>100721_f_at</td>
<td>Ig heavy chain V region 3 precursor</td>
<td>P01749</td>
<td>C, J, D, R</td>
</tr>
<tr>
<td>5</td>
<td>100583_at</td>
<td>Ig gamma-1 chain C region, membrane-bound form</td>
<td>P01689</td>
<td>C, J, D, R</td>
</tr>
<tr>
<td>6</td>
<td>100362_f_at</td>
<td>Anti-myosin immunoglobulin heavy chain variable region</td>
<td>Q9L83</td>
<td>C, J, D, R</td>
</tr>
<tr>
<td>7</td>
<td>99369_f_at</td>
<td>Ig kappa chain V-V region MOPC 173</td>
<td>P01643</td>
<td>J, D, R</td>
</tr>
<tr>
<td>8</td>
<td>97720_at</td>
<td>Ig heavy chain V regions TEPC 15/S107/HPCM1/HPCM2/HPCM3</td>
<td>P01787</td>
<td>J, D, R</td>
</tr>
<tr>
<td>9</td>
<td>97008_f_at</td>
<td>Ig heavy chain V region MOPC 104E</td>
<td>P01756</td>
<td>J, D, R</td>
</tr>
<tr>
<td>10</td>
<td>93297_f_at</td>
<td>Ig heavy chain V region 108A precursor</td>
<td>P01758</td>
<td>J, D, R</td>
</tr>
<tr>
<td>11</td>
<td>92470_f_at</td>
<td>Ig heavy chain V region 23 precursor</td>
<td>P01748</td>
<td>J, D, R</td>
</tr>
<tr>
<td>12</td>
<td>92316_f_at</td>
<td>Ig lambda-2 chain C region</td>
<td>P01844</td>
<td>J, D, R</td>
</tr>
<tr>
<td>13; 14</td>
<td>102823_at; 102824_g_at</td>
<td>Ig gamma-2B chain C region, membrane-bound form</td>
<td>P01867</td>
<td>C, J, R</td>
</tr>
<tr>
<td>15</td>
<td>102157_f_at</td>
<td>Ig kappa chain V-V region HP R16.7</td>
<td>P01644</td>
<td>J, D, R</td>
</tr>
<tr>
<td>16</td>
<td>92741_g_at</td>
<td>Ig delta chain C region secreted form</td>
<td>P01881</td>
<td>J, D</td>
</tr>
</tbody>
</table>

The specific tissues in which these genes are expressed are indicated by C (corpus), J (jejunum), D (descending colon), and R (rectum). A graphical visualization is depicted in Fig. 2.
in the intestine is greater than that in the stomach, thereby resulting in a reduced transcriptional modulation of Igs in the corpus.

We extended our analysis and used GO clustering to identify region-specific molecular functions. As demonstrated in both this study and the study performed by Bates et al. (2), a vast number of genes are differentially regulated along the GIT. Indeed, the present study revealed over 2,500 genes are differentially regulated among the four GIT tissues examined. Additionally, it was previously demonstrated that the transcriptional profiles of the various GIT regions reflect region-specific functions of the gut (2), thereby reinforcing the importance of examining the effect of commensal bacteria within a specific section of the GIT.

In the corpus, the only other molecular function that was significantly modulated was “tissue kallikrein activity” (4.7e-07, GO depth 6), which refers to a process entailing the preferential cleavage of Arg-Xaa bonds (where Xaa represents any residue) in small molecule substrates. This node comprised three genes that were all similarly regulated in the presence of microbiota: glandular kallikrein K6 (KLK6), glandular kallikrein K26 (KLK26), and γ-renin, submandibular gland (KLK16) were downregulated 3.7-, 3.6-, and 3.0-fold relative to germ-free mice, respectively. Kallikreins are involved in the cleavage of kininogens, which result in a corresponding increase in kinins (37). In the intestine, kinins have been associated with visceral pain, motility, and electrolyte secretion and speculated to mediate intestinal inflammation; therefore, our findings suggest that the conventional microbiota modulate the production of these potential inflammatory molecules. This finding was confirmed when we clustered the unique genes that were differentially regulated solely in the corpus, i.e., these genes were among the 68 genes unique to the corpus.

The jejunum has several additional significant nodes (Fig. 3), corresponding to “oxidoreductase activity” (7.1e-08, GO depth 4), “carboxylesterase activity” (7.0e-06, GO depth 5), “protein binding” (6.4e-06, GO depth 2), and “apoptosis regulator activity” (7.9e-07, GO depth 1); however, when put into context with the highly significant node corresponding to “defense/immunity protein activity” (1.1e-44), one can quickly appreciate how the immune function is far more significantly modulated by the microbiota. Table 2 lists the genes contained in these additional nodes and demonstrates that, in most cases, the clusters provide clear messages regarding how these genes react in the presence of gut bacteria, i.e., oxidoreductase and carboxylesterase activity are increased in the presence of microbiota. Furthermore, an additional point regarding GO clustering is illustrated with these results. As GO depth increases, the interpretation of the molecular function node becomes increasingly clear. As can be seen in Table 2, the genes corresponding to the “protein binding” and “apoptotic regulator” nodes do not uniformly respond to the presence of microbiota, which is in contrast to carboxylesterase and oxidoreductase activities. Therefore, it can be expected that in GO nodes

---

**Fig. 3.** Clustering in the jejunum with GO molecular functions. Clustering the 267 differentially regulated probe sets in the jejunum produced this GO tree. Highly significant nodes are in black, less significant nodes are in gray, and open nodes are not significant. All significant nodes must have a minimum P value of ≤0.001.
with a low depth (i.e., GO depth 1 or 2), a mixture of gene expression profiles are present. Although this does not provide a clear-cut response similar to that observed with the defense/immunity protein activity nodes, it does indicate that this molecular function is significantly regulated with the experiment. When clustering only the 77 differentially regulated genes unique to the jejunum, both carboxylesterase and apoptotic regulator activity were identified as highly significant GO nodes; reinforcing that these molecular functions are unique to the jejunum.

In the descending colon, only “oxidoreductase activity” (1.4e^-16), GO depth 4) is additionally modulated. The genes with this annotation corresponded to cytochromes P-450 2B10, 2D26, and 3A13, and these were all downregulated in conventional mice by 2.4-, 3.2-, and 1.9-fold, respectively. Interestingly, these same genes were also differentially expressed in the jejunum and, although more highly expressed, suggest that xenobiotic metabolism in the intestinal tract is modulated in the global gene expression profiles of the rectum. Although the precise intestinal functions of the aforementioned genes are known to various degrees, one can hypothesize that such physiological processes as fluid and electrolyte transport (uroguanylin) (34), lipid absorption (colipase) (43), and protection from digestive enzyme activity (α1-antitrypsin 1-1; antileukoproteinase 1) (35) are regulated by the rectal microbial community. However, these specific physiological processes were not identified by GO clustering, implying that the general class of “enzyme regulator activity” comprises genes implicated in multiple molecular functions that are nonsignificantly modulated at the gene transcription level by the microbiota. Little work has been reported regarding the global gene expression profiles in the rectum; indeed, this is the first report in which the gene expression profile of the rectum has been analyzed. Furthermore, even less work exists concerning the effect of conventional bacteria on the host rectum transcriptome. Our results suggest that the rectum plays as important a role in the host-bacterial relationship as do the small intestine and colon.

**Biological processes modulated the intestinal microbiota.** The GO Consortium defines biological process as “broad biological goals that are accomplished by ordered assemblies of molecular functions” (1, 38). When examining the significant GO biological processes stemming from this study, we find that, across all four tissues, “response to biotic stress” (corpus, 2.7e^-16; jejunum, 4.2e^-09) and/or “response to pest/pathogen/parasite” (descending colon, 4.6e^-04; rectum, 7.2e^-10) are the most significantly affected processes in the
presence of a conventional microbiota. Figure 4 demonstrates the GO clustering results for the corpus and illustrates the highly significant aforementioned biological process. For reasons of brevity, the GO trees for the other three tissues (which are very similar) are not shown.

The GO Consortium defines “response to biotic stress” as “a change in the state of an organism in the presence of a biotic stimulus, including response to bacteria, fungi, oxidative stress, pest/pathogen/parasite, starvation, sterol depletion, unfolded protein, and ER-overload” and defines “response to pest/pathogen/parasite” as “a change in the state of an organism in the presence of a pest or pathogen or parasite” (1, 38). The former is the parent GO term for the latter, which implies that genes responding to a multitude of biotic stimulus are affected in the upper GIT. Interestingly, the response to a biotic stimulus becomes more specific in the descending colon and rectum (i.e., only to pest/pathogen/parasite), indicating that the residental bacteria have less significant roles in the modulation of various biotic stimuli in the distal GIT. Nevertheless, we confirm that clustering both GO molecular functions and biological processes yields similar results, namely, that the microbiota strongly affect genes involved in immune-related functions.

In contrast to the molecular function analysis, which revealed the individual molecular players that respond to the presence of a gut microbiota, examining broad biological processes identified additional genes contributing to GIT function. Most interestingly, the residential bacteria affect water
homeostasis in the corpus, descending colon, and rectum. In all three tissues, examining the “water homeostasis” node revealed a consistent downregulation of aquaporin 4 (Aqp4). Indeed, two probe sets corresponding to Aqp4 (102703_s_at; 102704_at) were downregulated in the presence of microbiota (~2.5-fold). Aqp4 is the predominant aquaporin in the stomach and colon, whereas Aqp3/7 appear to be the major aquaporins of the human/rat small intestine (20). As no significant changes were observed for either Aqp3 or Aqp7 in the jejunal tissue, the specific regulation of Aqp4 reveals both a region- and family-member-specific regulation of GIT water homeostasis. Previous studies using immunohistochemical techniques have located Aqp4 in the basolateral compartment of stomach parietal cells (acid/liquid secretion) and the basolateral membrane of colonic epithelial cells (fecal dehydration) (10, 24). Additionally, our findings suggest that fecal dehydration occurs in the rectum; however, the cellular location of rectal Aqp4 has not yet been reported. Perhaps most importantly, we propose that Aqp4 is the molecular player responsible for the differences in large intestinal water absorption between conventional and germ-free mice reported by Wostmann (42).

As with the GO molecular function clustering, we have identified certain region-specific biological processes that are affected by the microbiota. However, the biological processes identified in the jejunum, e.g., “induction of programmed cell death,” reflect the molecular functions previously discussed, e.g., “apoptosis regulator activity.” Those significant molecular functions identified in the rectum were similarly reflected in the significant biological processes identified (e.g., “digestion” vs. “enzyme regulator activity”). One additional biological process, specific to the corpus, was identified through GO clustering and corresponded to “negative regulation of glucose import” (6.6e−04). Further examination of this GO node revealed that PEA-15 is upregulated 2.7-fold in conventional mice. Although not a transporter itself, PEA-15 has been implicated in the regulation of glucose transport by controlling both the content of Scl2a1/Glu1 glucose transporters on the plasma membrane and the insulin-dependent trafficking of Scl2a4/Glut4 from the cell interior to the surface (7).

**DISCUSSION**

The present work used microarray technology to examine the molecular relationship between the commensal microbiota and several distinct regions along the length of the mouse GIT. Using the Mu74Av2 GeneChip, we examined the 6,000 functionally characterized sequences present in the murine UniGene database and nearly 6,000 EST sequences, thereby corresponding to ~33% of the mouse genome. Assaying whole gut tissues revealed that the residential bacteria affect few biological processes (as determined by GO clustering with the most current GO annotations) in the digestive tract; however, we recognize that examining only the mucosal layer or other specific intestinal regions (e.g., Peyer’s patches) may reveal further, more subtle gene changes that have been masked by analyzing the whole tissue. Nevertheless, investigation of distinct regions of the gut led to the identification of region-specific biological processes modulated by the microbiota and suggests that bacterial species with regional patterns of colonization may be implicated in the molecular cross talk between host and microbiota.

The results presented in the current work were obtained with a novel method of clustering. Currently, several clustering methods exist to reveal groups of genes that behave similarly across any given number of experimental conditions (29); however, such unsupervised methods as hierarchical and k-means clustering do not use existing biological knowledge to determine clusters (18). Rather, these genes are first clustered based on similar expression patterns and then assigned biological functions. Therefore, extracting comprehensible information from a given cluster is not always evident, primarily due to gene products having multiple biological annotations. Furthermore, although unsupervised clustering can suggest that biological function is related to gene expression, several instances have been documented in which biologically related genes show inverse expression profiles (4, 8, 16). Therefore, we aimed to circumvent these potential complications by using a clustering method in which genes are grouped by biological function and not gene expression; however, this means of clustering is highly dependent on the accuracy and uniformity of the biological terminology established by the GO Consortium (38). As the complete sequencing of the mouse transcriptome and genome has only recently been described (9, 25), the next challenge lies in the functional annotation of each gene. The GO Consortium provides a database in which annotations are standardized between different species, thereby providing a common language within the scientific community. However, the functional characterizations of mammalian genomes are currently an ongoing process, and it is expected that the GO database is incomplete and that annotation errors are inevitable (6), e.g., as demonstrated by the revised definition for “defense/immunity protein activity” GO class. To keep pace with the advances made by the scientific community, the GO Consortium releases monthly updates of their database, with the goal of providing the most up-to-date and accurate annotations possible. Furthermore, the findings presented in this manuscript suggest that the GO database does not bias the annotation of specific biological classes, as demonstrated by the identification of GO clusters comprising genes underlying the previously noted disparate physiological differences between germ-free and conventional mice.

Using germ-free animal models, such functions as immune responses, water absorption, and oxidation-reduction potentials were previously attributed to the presence of residential bacteria (42). However, prior to the advent of microarray technology, identification of those genes involved in the communication between the microbiota and host was performed by examining either single or small subsets of genes. These studies yielded important findings revealing that the mouse microbiota affect genes involved in the mucosal defense and inflammatory response of the host (21, 27, 30). In contrast, the current generation of studies has exploited the power of microarrays and used this global technology to examine the role of conventional bacteria on the expression of thousands of genes in the mouse colon (11) and ileum (15). Genes involved in immune responses, xenobiotic metabolism, apoptosis, and postnatal intestinal maturation have been related to the presence of a nonpathogenic microbiota and therefore emphasize the importance of residential bacteria in defining a healthy gut ecosystem. Inasmuch as these studies have demonstrated the merit of examining the cross talk between the microbiota and host, important aspects of microarray data analysis have since
been better characterized, thereby leading to a more accurate biological interpretation of the results. For example, it is now accepted that the variability in microarray data is not homogeneously distributed, i.e., lowly expressed genes have an increased variability compared with highly expressed genes (26).

The use of fold-change cutoff values [as used by both Fukushima et al. (11) and Hooper et al. (15)] does not take this into account and assumes homogeneity, thereby increasing the risk of false positives and false negatives. To circumvent this, we have used a GEA model for the identification of differentially expressed genes, which moves beyond the fold-change cutoff and incorporates the concept of heterogeneity (22). Therefore, the use of a robust statistical model for the selection of differentially expressed genes in various regions of the digestive tract, in addition to a clustering method based on GO, yields a powerful method for defining the molecular relationship between host and microbiota.

Interestingly, the microbiota appear to minimally regulate gene expression in the GIT tissues examined. Indeed, when one compares the number of genes modulated along the GIT with the number of genes modulated in the jejunum, the difference is ~10-fold. This implies a rather minimal communication at the transcriptional level between the host and microbiota, as well as large differences between the distinct GIT regions. One possible explanation for this may arise due to the complexity of the gastrointestinal "environment," where differences in pH, motility, bile, and morphology exist along the length of the GIT and will undoubtedly influence gene expression and function in their own right.

Only one biological function was consistently modulated by the microbiota along the digestive tract ("defense/immunity protein activity"), and all genes with this annotation demonstrated a consistent upregulation in conventional mice. Although some of the genes appear to have regional specificity, it is interesting to note that this molecular function is both less represented and less significant in the corpus, thereby suggesting that the major influence of the microbial community on host immunity is in both the small and large intestines. The only other biological function modulated along the length of the GIT was "water homeostasis," for which a downregulation of Aqp4 was consistently observed. This transporter presents an attractive molecular candidate for the differences in water absorption between germ-free and conventional mice described by Westmann (42). Furthermore, this gene may yield the prebiotic/probiotic community with a molecular target to prevent gastrointestinal illnesses having modified water transport (e.g., diarrhea).

The various region-specific biological processes identified, such as "tissue kallikrein activity" in the corpus and "apoptosis regulator activity" in the jejunum, present intriguing findings that yield an exciting question. Can one hypothesize that these region-specific biological functions arise due to the region-specific microbial communities previously described (3)? Although there are more than 400 bacterial species thought to exist in the GIT, it is conceivable that a single bacterial species in the corpus, for example, is responsible for the coordination of "kallikrein activity" and its role on stomach inflammation. Alternatively, these results may imply that epithelial cells in distinct regions of the GIT are differentially receptive to the same bacteria (40). For example, finding that "apoptosis regulator activity" is, in large part, upregulated by the jejunal microbiota may suggest that small intestinal epithelial cells are more sensitive to the presence of bacteria than other regions of the gut and would therefore be important to consider with the current hypothesis that microbiota and apoptosis are related and influence the onset of inflammatory bowel disease (28).

Future, region-specific analyses will resolve this issue with more sensitivity for the continued effort in isolating and culturing the many bacterial species present in the gut. In conclusion, these findings corroborate the utility of clustering with biological annotations by yielding comprehensible and biologically cohesive results. Furthermore, this represents the first study describing the relationship between the microbiota and host in multiple regions of the GIT and, most importantly, suggests that the microbial community exerts potent, but limited, transcriptional responses along the length of the GIT in addition to distinct effects in specific regions of the host digestive tract.

ACKNOWLEDGMENTS

We thank Nadine Porta for assistance with total RNA extractions, Robert Mansourian for applying the GEA method of analysis to this microarray data set, and David Vilanova and Marie-Camille Zwahlen for assistance with the annotation of Affymetrix probe sets. We thank Elena M. Comelli and Christine Cherbut for the critical review of this manuscript.

Present addresses: R. Simmering, VTB-Hygiene Research, Henkel KGaA, 40191 Düsseldorf, Germany; and G. Fotopoulos, Global Marketing, Roche Consumer Health Ltd., 4303 Kaiseraugst, Switzerland.

REFERENCES