Expressive Analysis of Gut Microbiota in Pre- and Post- Solid Organ Transplantation Using Bayesian Topic Models

Luigi Santacroce[1,2(✉)], Sara Mavaddati[2], Javad Hamedi[3], Bahman Zeinali[4], Andrea Ballini[5], and Massimo Bilancia[1]

[1] Ionian Department (DJSGEM), University of Bari Aldo Moro, Taranto, Italy
[2] Microbiology and Virology Laboratory, Policlinico University Hospital of Bari, Bari, Italy
luigi.santacroce@uniba.it
[3] Department of Microbial Biotechnology, School of Biology and Center of Excellence in Phylogeny Living Organisms, College of Science, University of Tehran, Tehran, Iran
[4] Department of Biology, Faculty of Basic Sciences, Tehran University, Tehran, Iran
[5] Department of Biosciences, Biotechnologies and Biopharmaceutics, University of Bari Aldo Moro, Bari, Italy

Abstract. There is a growing evidence that variation in gut microbial communities has important associations with overall host health, and that the diversity and the richness of such communities is helpful in distinguishing patients at high risk of life-threatening post-transplantation conditions. The aim of our paper is to provide an expressive and highly interpretable characterization of microbiome alterations, with the goal of achieving more effective transplantations characterized by a rejection rate as low as possible, and to avoid more severe complications by treating patients at risk in a timely and effective way. For this purpose, we propose using topic models to identify those bacterial species that have the most important weight under the two different experimental conditions (healthy and transplanted patients, or patients whose fecal microbiota has been sampled both in pre- and post-transplantation phases). Topic models are Bayesian statistical models that are not affected by data scarcity, because conclusions we can draw borrow strength across sparse gut microbiome samples. By exploiting this property, we show that topic models are expressive methods for dimensionality reduction which can help analyze variation and diversity in gut microbial communities. With topic models the analysis can be carried out at a level close to natural language, as the output can be easily interpreted by clinicians, since most abundant species are automatically selected and the microbial dynamics can be tracked and followed over time.

Keywords: Solid-organ transplantation · Liver Transplantation · Gut microbiota · Topic models · Latent Dirichlet Allocation · Translational medicine

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12252, pp. 150–165, 2020.
https://doi.org/10.1007/978-3-030-58811-3_11

1 Introduction

One of the most reliable but extreme treatment options for the end stage disease (ESD) patients is solid organ transplantation (SOT)[1,2]. It is well known that immunosuppression is a major treatment to prevent acute and chronic rejection, which unfortunately can lead to serious toxic side effects [3]. Similarly, infections [4], graft-versus-host disease (GVHD) and other post transplantation complications are frequent adverse events in the clinical practice [5,6]. In this context, there is a growing evidence that variation in gut microbial communities has important associations with overall host health, and that the diversity and the richness of such communities is helpful in distinguishing patients at high risk of life-threatening post-transplantation conditions, such as infections by multidrug-resistant bacteria (MDRB), from patients with less severe side effects [7].

It is well known that gut microbiota plays an essential role in promoting intestinal inflammation through the 'intestinal microbiota-immunity' axis, impacting on distal immune response and affecting other gut-organ axes, as well as in modulating disease in distant tissues by releasing metabolites such as short-chain fatty acids (SCFAs), tryptophan, phenylalanine and tyrosine [8–10]. Based on those physiological pathways, intestinal microbiota alterations in patients after transplantation have been recently investigated, demonstrating a significant shift in the intestinal microbiome composition and diversity compared to pre-transplantation condition [11,12]. However, it would be useful to have a broader view, in order to provide accurate and personalized host trait prediction based on microbiome alterations, with the goal of achieving a more effective transplantation characterized by a rejection rate as low as possible, treating patients at risk in a timely and effective way in order to avoid more severe complications.

If we limit ourselves to analyze the relationship between the microbiome and human disease, the wealth of available high-throughput 16 S ribosomal RNA data has allowed the publication of several excellent meta-analyses, which are likely to play a significant role in clinical practice over the next years [13]. However, if our objective is that of analyzing the dynamics of microbiota before or post transplantation, having in mind to compare gut bacterial populations between healthy patients and ESD patients candidate for transplantation or, for example, to study the associations existing between gut microbiota alterations and the onset of potentially life-threatening complications, difficulties arise because we do not have, at present, a large amount of published data. To give an idea, we searched PubMed using the MeSH thesaurus and a suitable query, with which we retrieved 276 relevant papers dealing with microbiota and SOT. However, when we re-analyzed this collection only two papers, both of them dealing with liver transplantation (LT), had abundance data published and readily available in their supplementary sections. With such a scarcity of data a classical meta-analysis would be impossible to conduct. However, we postulate that an expressive analysis of composition, diversity and richness of microbial communities is still feasible using suitable Bayesian topic models that have been introduced for

dealing with automatic text analysis and classification, but that are well suited to microbiome analysis and have potential for providing clinically meaningful summaries.

To be more specific, the class of Bayesian topic models has been introduced for discovering the abstract 'topics' that occur in a collection of documents [14,15], even though the class of such statistical models can be readily adapted to microbiome data in a simple and meaningful way. Using topic models, we are able to identify those bacterial species that have the most important weights under different experimental conditions. Most importantly, Bayesian topic models are not affected by data scarcity, because conclusions we can draw borrow strength across sparse microbiome samples, allowing expressive comparison with a baseline or a pre-transplantation condition.

2 Related Work

The details of correspondence between microbiome data and text data have been explained by [16], which review a few statistical modeling techniques used in text categorization and explore how to transfer them to the microbiome. A similar approach is followed by [17], where the microbial community structure across multiple samples is inferred based on topic models. Similarly, [18] show how each microbial sample can be considered as a 'document', which has a mixture of functional groups in the human gut microbiome, while each functional group can be seen as a latent topic, that is a weighted mixture of taxonomic levels. A somewhat more technical introduction is reported by [19], which explores a difficult arising from the fact that the number of topics (functional groups) of the topic models must be pre-specified, and if the pre-specified number of topics is changed then the number of topics to be interpreted also changes. As a solution to the problem a Bayesian nonparametric model is proposed, by developing a hierarchical specification that can mitigate the parametric assumptions underlying a traditional Bayesian topic model, and can automatically decide the number of topics from the data.

In this paper we considered useful to avoid unnecessary complexity, trying to to keep the analysis pipeline structure as simple as possible, favoring therefore the interpretability of the results. Based upon these principle, we used Latent Dirichlet Allocation (LDA) to collapse together taxonomic levels that had the same expression. LDA was first introduced in [14] as a form of dimension reduction that uses a probabilistic model to find the co-occurrence patterns of terms that correspond to semantic topics in a collection of documents. Latent topics have a simple interpretation in terms of microbial communities harbored by the human gut, as described in Subsect. 3.2. To decide on a suitable number of topics, the measure traditionally used for topic models is the perplexity of held-out documents, which is a decreasing function of the log-likelihood of the unseen documents; the lower the perplexity, the better the model [20]. However, we considered multiple indices to attenuate the effect of sampling variation, calculating all metrics at once over multiple LDA models with varying number of topics.

3 Materials and Methods

3.1 Datasets

The first dataset we analyzed in this paper is reported in [7]. The study recruited 177 adult patients undergoing LT. In total $c = 723$ fecal samples spread across the sampling time-points (pre-LT, peri-LT – weeks 1, 2, and 3 after LT, and post LT – months 1, 2, 3, 6, 9, and 12) were sequenced (16S V3-V4 rRNA, with a median of 4 samples per patient). After quality filtering the dataset included 703 samples from 175 patients (spread across the sampling time points). We considered pre-LT data (83 samples on 83 patients) and post-LT data, from months 6 to 12 after transplantation (207 samples on 118 unique patients). For each sample, most of identified operational taxonomic units (OTUs) were classified at the Genus-Species level, while some OTUs were classified at the Genus species only, and a few OTUs were neither classified at the Genus or Species Level, nor assigned to any known taxon. For both pre- and post-LT data, we had 878 OTUs in the taxonomy table. Therefore, for pre-LT the dimension of the abundance matrix was 878×83, and the post-LT matrix had dimension 878×207. Each cell of the abundance matrices contains the count of OTU abundances. For shortness, from this point on this dataset will be referred to as DATASET1.

The second dataset is contained in [21]. In the treatment group were enrolled 90 LT recipients, versus 61 healthy controls in the control sample, with no significant difference existing between the two groups in terms of age, sex distribution and body mass index. For each patient and healthy control 588 OTUs were clustered from fecal samples (1 fecal sample for each patient and 1 sample for each healthy control). Hence, the abundance matrices had dimension 588×90 and 588×61, respectively. Most of OTUs were classified at the Family and Genus levels, whereas some OTUs were assigned at the Genus level only. From this point on this dataset will be referred to as DATASET2.

3.2 Statistical Analysis

Latent Dirichlet Allocation (LDA) model discovers the different topics that the documents represent and how much of each topic is present in a given document. We now briefly introduce the necessary notation to LDA, and we provide an interpretation in terms of microbiome data. The generative model of LDA can be graphically represented as the following oriented graph (Fig. 1):

In a more descriptive way (for notational simplicity, the indicator d is suppressed):

- A document d is a stream of N terms (or words), $d = (w_1, \ldots, w_N)$. Using superscripts to denote components, the νth term in $|V|$ is represented as a unit-basis vector w such that $w^\nu = 1$ and $w^u = 0$ for $u \neq \nu$.
- For each document we have K underlying semantic themes (topics), $\beta_{1:K} = (\beta_1, \ldots, \beta_K)$, where each β_k is a $|V|$-dimensional vector of probabilities over the elements of V, for $k = 1, \ldots, K$.

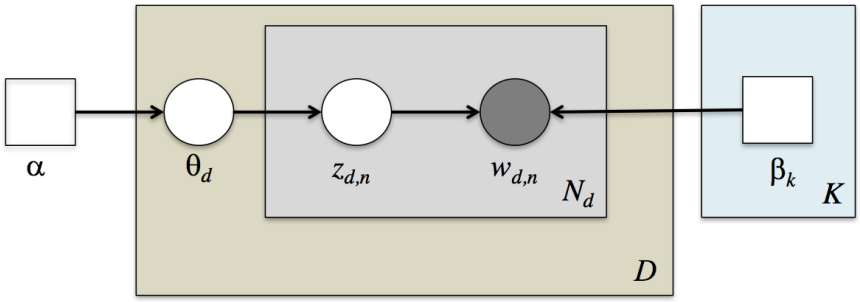


Fig. 1. A graphical model representation of Latent Dirichlet Allocation (LDA). Parameters α and $\beta_{1:K}$ are treated as unknown hyper-parameters to be estimated.

– $z_{1:N} = (z_1, \ldots, z_N)$ is a vector of K-dimensional vectors indicating, for $n = 1, \ldots, N$, the topic which has generated term w_n in document d. The indicator z of the k-th topic is represented as a K-dimensional unit-basis vector such that $z^k = 1$ and $z^j = 0$ for $j \neq k$.

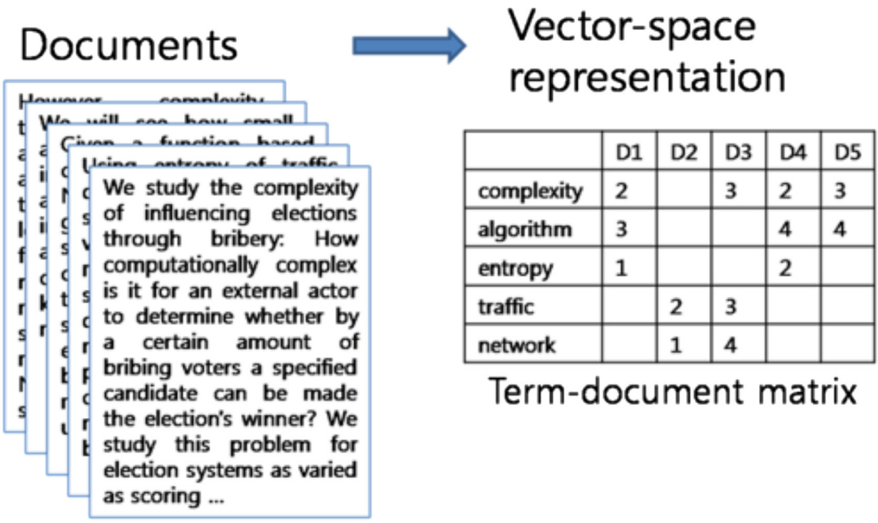
The topic indicator uniquely selects a probability distribution in $\beta_{1:K}$, as $\beta_{z_n} \equiv \beta_k$ when $z_n^k = 1$. Independently of each other, the generative process of each pre-processed document d is the following:

1. Draw topic proportions from a symmetric Dirichlet distribution over the K-dimensional simplex, $\theta|\alpha \sim \mathsf{Dirichlet}_K(\alpha)$.
2. For each term w_n, $n = 1, \ldots, N$, and independently of each other:
   (a) Choose a topic from a Multinomial distribution with probabilities θ, $z_n|\theta \sim \mathsf{Multinomial}_K(\theta)$.
   (b) Choose a term from a Multinomial distribution with probabilities dependent on z_n and $\beta_{1:K}$, $w_n|z_n, \beta_{1:K} \sim \mathsf{Multinomial}_{|V|}(\beta_{z_n})$.

The main input to LDA is the term-document matrix. The term-document matrix describes the frequency of terms that occur in a collection of documents (a corpus). In a term-document matrix, rows correspond to terms in the collection and columns correspond to documents (Fig. 2). The data matrix containing OTUs abundances has essentially the same structure, provided that we interpret rows and columns in the following way [16,17]:

– OTUs → rows → terms.
– Samples → columns → documents. In this particular case, each document has the same number of terms.
– Corpus → the set of all fecal samples associated with the environment (gut). The dimension c of the corpus (number of columns) represents the number of fecal samples.

Each column (fecal sample) contains some microbial species and does not contain others (those OTUs with null abundance in that fecal sample). A topic can be thought as a community of bacteria that share similar roles and biological



	D1	D2	D3	D4	D5
complexity	2		3	2	3
algorithm	3			4	4
entropy	1			2	
traffic		2	3		
network		1	4		

Fig. 2. The term-document matrix describing the frequency of terms that occur in a collection of documents. In a term-document matrix, rows correspond to terms in the collection and columns correspond to documents.

effects. Each fecal sample contains a mixture of topics (a mixture of bacterial communities) that LDA may contribute to disentangle. Parameters of the model were estimated using an empirical Bayes approach, following variational inference methods reliably implemented implemented in the `topicmodels` package [22], under R 3.6.1.

3.3 Choosing the Number of Topics

Setting the number of topics is the most computationally intensive part. We used three methods to compare the goodness-of-fit of the LDA model fit with varying number of topics K (K ranging between 2 and 60 with a step of 2 – 2, 4, 6, ..., and so on). All the computations were implemented in R 3.6.1 using specialized routines [23]. A sensible interpretation of the number of topics used for describing the gut microbial community is still unclear, an issue we discuss further in next section.

3.4 Assigning Terms to Topics

First, topics were sorted according their probabilities in the entire collection of samples. Alternatively, topics were sorted by counting how often a topic appeared as a primary topic. This alternative method is also known as Rank-1 [24]. Subsequently, for each of the 5-top topics, we plotted the 10-top terms (OTUs) in terms of their probability of occurrence. In this way, bacterial community associated with each topic could be further analyzed and characterized at a very

high expressivity level, close to natural language, far different from the aggregated results which are commonly obtained through traditional meta-analyses (for example, traditional meta-analysis provides the number of genera which are significantly different in the two experimental conditions).

4 Results

4.1 Determining the Optimal Number of Topics

The optimal number of topics has been determined for both DATASET1 and DATASET2. The results are shown in Fig. 3 and Fig. 4, respectively. For DATASET1 we found $K = 15$ for pre-LT and $K = 30$ for post-LT data. Similarly, for DATASET2 the optimal values were $K = 15$ for healthy controls and $K = 35$ for transplanted patients.

It is interesting to note that, in both cases, the two curves presented a qualitatively similar aspect, although they were obtained under different experimental designs (in fact, in the second case, we do not have partially matched pre-LT data, but healthy control subjects). In addition, both for ESD patients in the liver recipient waiting list and for healthy control subjects, the optimal number of topics is much lower than that determined for transplanted patients. If we interpret a topic as a microbial community sharing similar roles and biological effects in terms of overall host health, it is tempting to argue that LT always increases the diversity and richness of microbial communities. This hypothesis is only partially confirmed by computing classical diversity measures; for example, using alpha-diversity, for DATASET1 we had that the median alpha was equal to 2.94 for pre-LT data and 3.33 for post-LT, whereas for DATASET2 we had 3.37 for healthy subjects and 3.08 for transplanted patients.

4.2 Determining the 5-Top Topics

The 5-top topics for DATASET1 were labeled (13, 7, 6, 15, 3) for pre-LT data, while the corresponding Rank-1 5-top topics were (13, 15, 3, 7, 1), with four concordances (Fig. 5). The 5-top topics for DATASET1 for post-LT data were (13, 14, 29, 16, 19) and (13, 14, 19, 22, 1) using Rank-1, with three concordances. Similarly, the 5-top topics for DATASET2 were (5, 7, 2, 3, 9) and (3, 5, 2, 7, 9) for healthy subjects, with five concordances, using the two methods respectively (Fig. 6). For post-transplanted patients we had (11, 21, 33, 10, 6) and (11, 21, 12, 25, 6), respectively, with two concordances. For example, relative frequencies associated with the 5-top topics (13, 7, 6, 15, 3) (DATASET1 with the standard method) were respectively: $13 \rightarrow 18.21\%$, $7 \rightarrow 9.04\%$, $6 \rightarrow 8.04\%$, $15 \rightarrow 7.44\%$, $3 \rightarrow 6.97\%$. In other words, topic (microbial community) labeled as '13' is prevalent in about 18% of fecal samples, topic (microbial community) labeled as '7' is prevalent in about 9% of samples, and so on. It is also worth noting that, at this level, the analysis has a low grade of expressivity, as topics (microbial communities) are nothing but 'labels'.

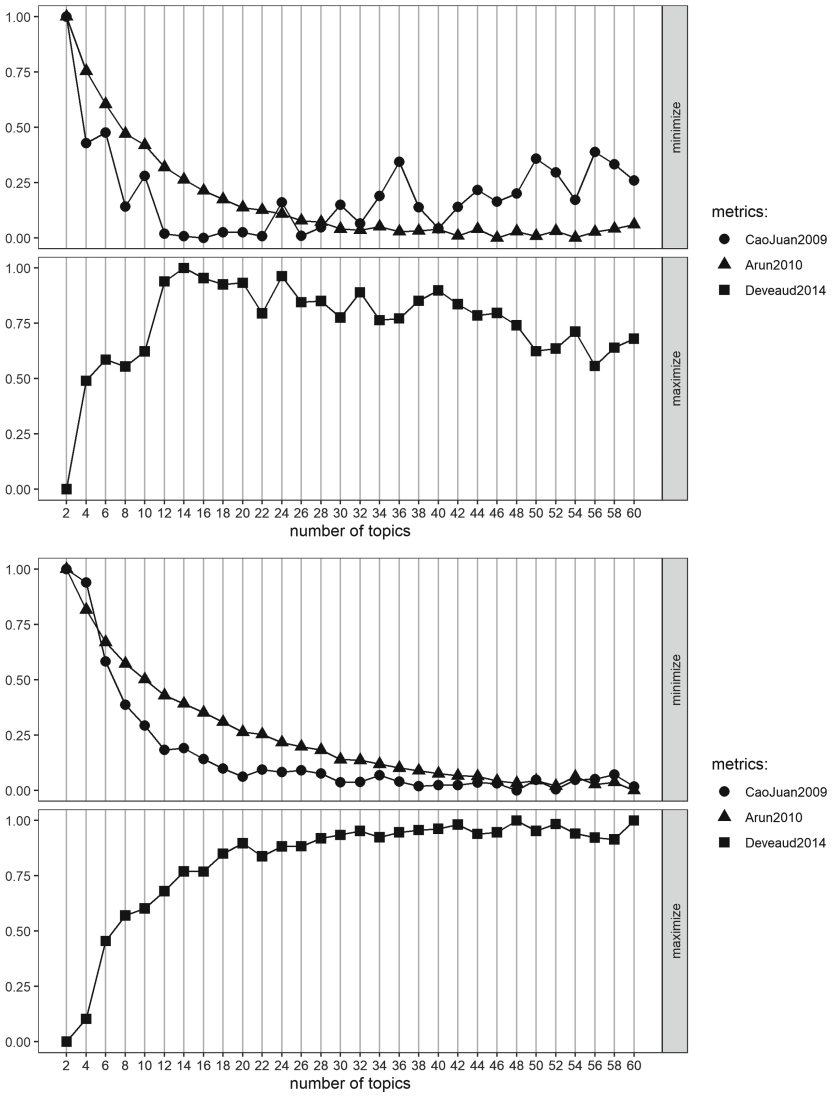


Fig. 3. Choosing the number of topics for DATASET1. The optimal choice was $K = 15$ for pre-LT (upper panel) and $K = 30$ for post-LT data (lower panel). We used three methods to compare the goodness-of-fit of the LDA model fit with varying number of topics K, with K ranging between 2 and 60 with a step of 2 (2, 4, 6, . . ., and so on).

4.3 Determining the 10-Top OTUs for Each Top-Topic

An expressive analysis could be obtained by plotting, for each top-topic, the 10-top terms (OTUs) ranked according to their probability of occurrence. Figure 7 shows the 10-top OTUs of DATASET1, for both pre-LT and post-LT abundances

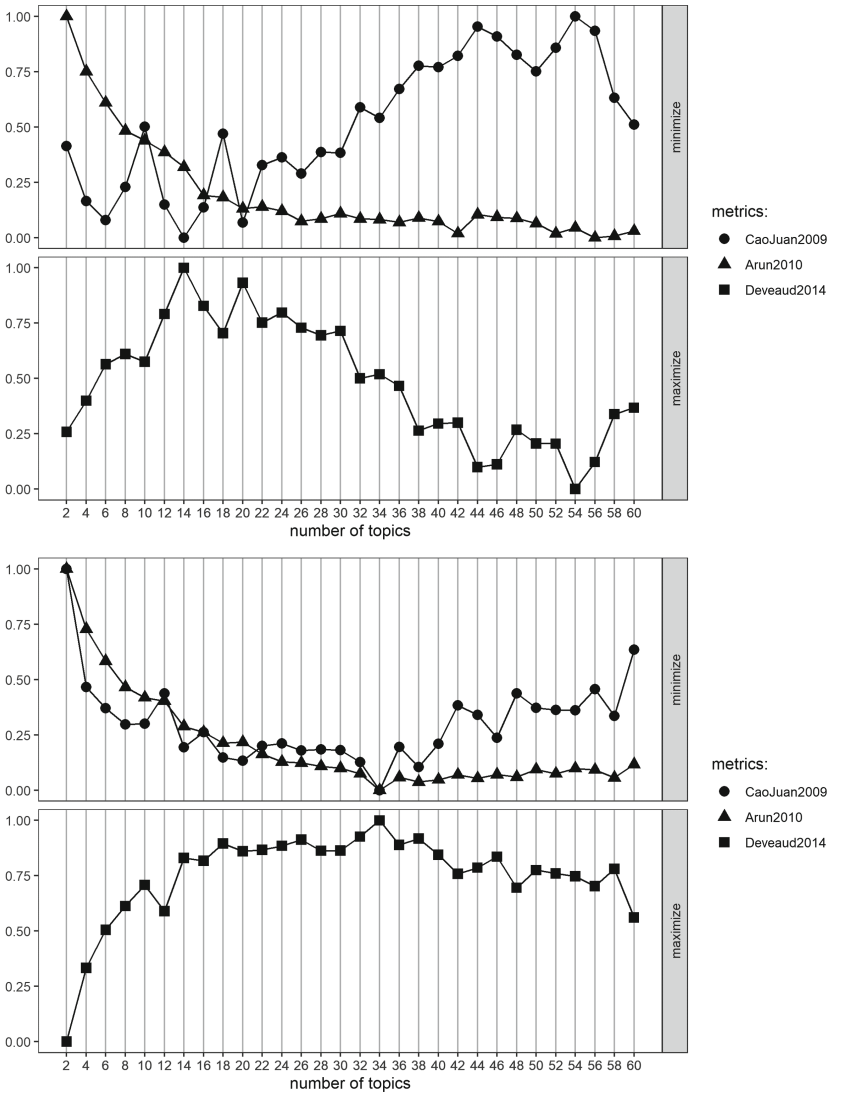


Fig. 4. Choosing the number of topics for DATASET2. The optimal choice was $K = 15$ for pre-LT (upper panel) and $K = 35$ for post-LT data (lower panel). We used three methods to compare the goodness-of-fit of the LDA model fit with varying number of topics K, with K ranging between 2 and 60 with a step of 2 (2, 4, 6, . . ., and so on).

respectively. In the same way, Fig. 8 contains the 10-top OTUs of DATASET2, for healthy control subjects and transplanted patients respectively. To make the analysis expressive, each OTU was mapped onto its Genus-Species taxon (when available), or its Genus or Family (when Species and/or Genus were not available). OTUs that could not be mapped onto any known taxon were denoted

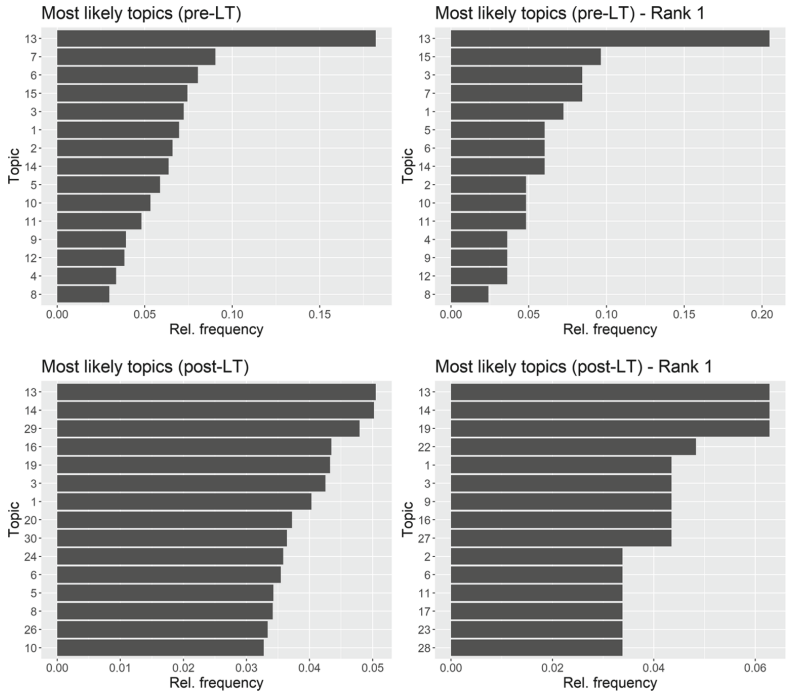


Fig. 5. Top-topics for DATASET1, using the two methods described in the text. With the first method, topics were sorted according their probabilities in the entire collection of samples. Using the second method, known as Rank-1, topics were sorted by counting how often a topic appeared as a primary topic.

as NA. It is worth noting that different OTUs may be assigned to the same Genus or Family (particularly when a precise Genus-Species classification was not determined). For example, for topic 15 in Fig. 7 (Rank-1 method, pre-LT) we had 6 OTUs that have been assigned to Genus *Bacteroides*. In this case, the probability that is plotted on the graph is collapsed at Genus level over the corresponding probabilities ($0.0596 + 0.0373 + 0.0335 + 0.0299 + 0.0288 + 0.0270 = 0.2161$).

5 Discussion and Conclusion

The importance and role of personalized medicine in a successful SOT cannot be neglected. Alteration of intestinal microbiome in patients after transplantation has been extensively investigated. Although there is specific microbiome in each of organ eligible for transplantation (for example liver, kidney, heart lung and pancreas), gut microbiome seems to play the most important role for overall host health through the various axes existing between solid organs and the gut microbiome itself. However, a broad view of associations between intestinal

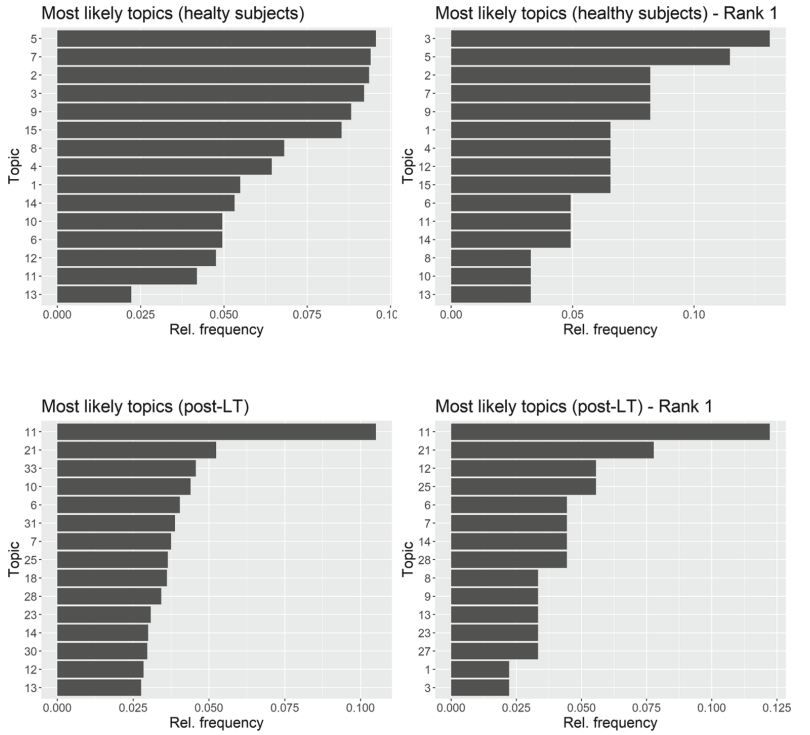


Fig. 6. Top-topics for DATASET2, using the two methods described in the text. With the first method, topics were sorted according their probabilities in the entire collection of samples. Using the second method, known as Rank-1, topics were sorted by counting how often a topic appeared as a primary topic.

microbiome and disease in the pre- and post-transplantation phases is obscured by many factors, such as the highly variable individual response to an invasive transplantation.

From this point of view, meta-analysis is the best instrument to achieve a broader view and to arrive at sound conclusions. At present, the wealth of available high-throughput 16 S ribosomal RNA data has allowed the publication of a number of excellent meta-analyses analyzing the relationship between gut microbiome and human disease. However, in a broad collection of 276 relevant papers dealing with microbiota and SOT, only two papers had microbial abundance data published readily available in the supplementary section, both of them dealing with LT. Such a lack of data makes it impossible to construct ensemble-based analyses, or to carry out host-trait prediction based on microbial abundance. However, in this paper we suggest that an expressive comparative analysis of composition, diversity and richness of microbial communities is still feasible using suitable Bayesian models, that have been introduced for dealing

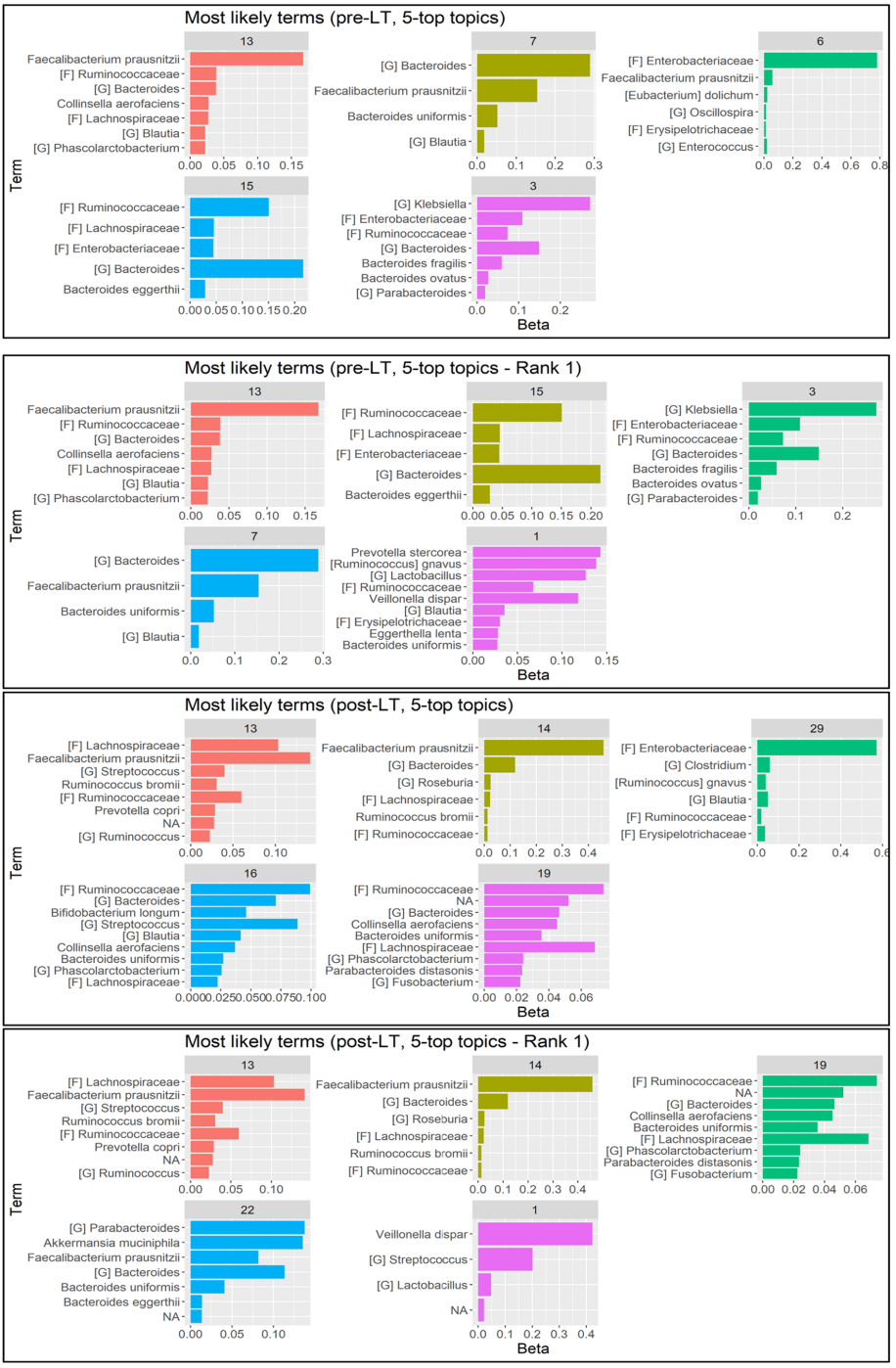


Fig. 7. Top-terms of the 5-top topics for DATASET1 (pre-LT and post-LT, both standard and Rank-1 methods) ranked according to their probability of occurrence.

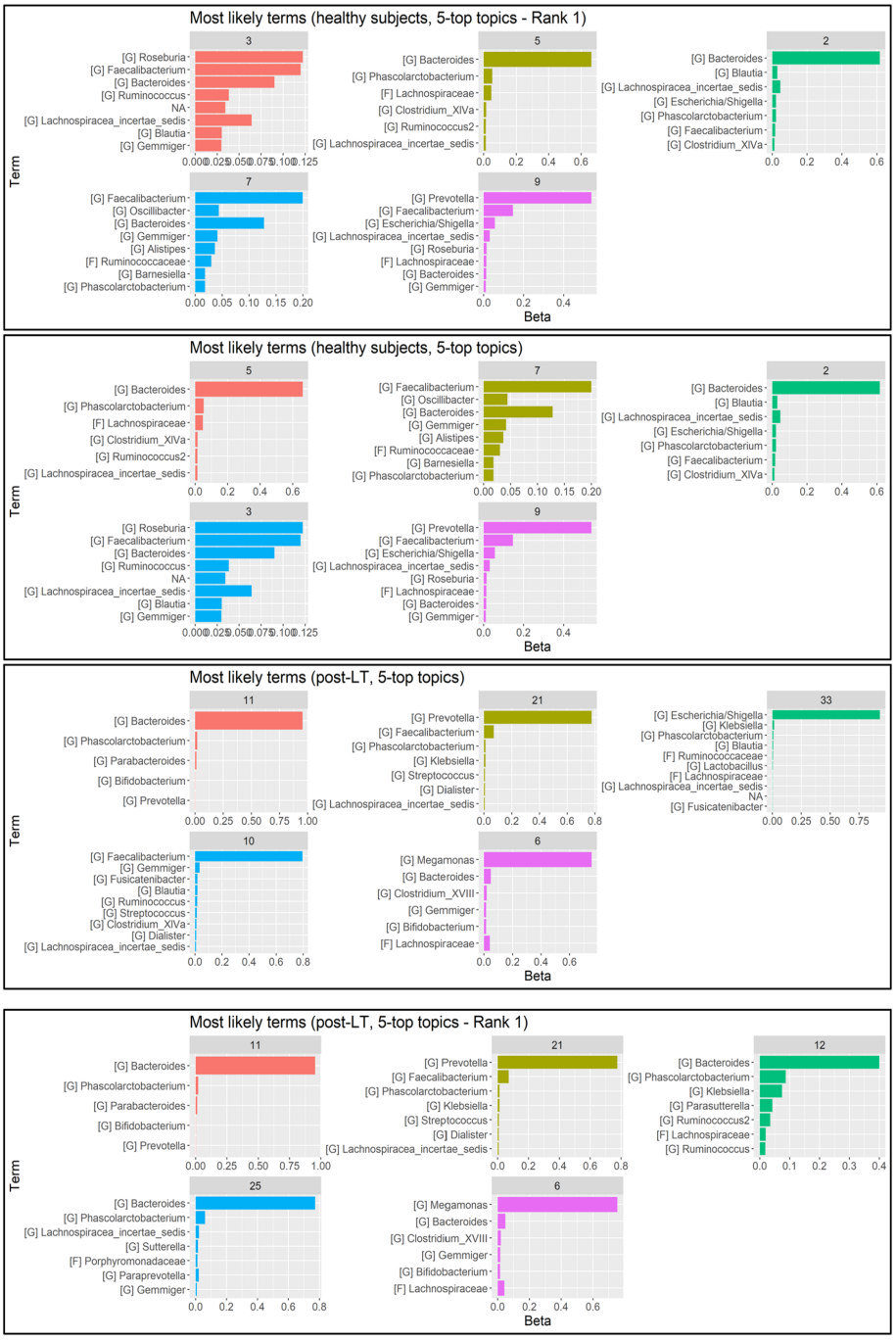


Fig. 8. Top-terms of the 5-top topics for DATASET2 (healthy and post-LT subjects, both standard and Rank-1 methods) ranked according to their probability of occurrence.

with automatic text analysis and classification, but that are well suited to microbiome analysis and have potential for providing clinically meaningful summaries.

The expressivity of topic models can lead to some interesting conclusions. For example, post-LT gut microbiomes in DATASET1 were rich in beneficial commensal bacteria such as *Faecalibacterium prausnitzi*, *Lachnospiraceae* and *Ruminococcacae*, whereas less abundance of *Ruminococcacae* and *Lachnospiraceae* was found in DATASET2, with a significant coexistence of opportunist pathogens such as *Streptococcus* and *Veillonella dispar*. In general, pre-LT ESD patients have a lower abundance of healthy commensal species in comparison to the post-LT phase. Moreover, after LT an increase in diversity is observed in most of the cases, due to a greater diversification in composition and abundance of both healthy and pathogen species. In DATASET1 there was a high abundance in pre-LT patients of both commensal healthy and pathogen bacteria, such as *Faecalibacterium prausnitzi*, *Bacteroides*, *Klebsiella*, *Enterobacteriaceae* and *Ruminococcacae*. This is consistent with previous studies, that have noted that patients with liver ESD have a decrease in non pathogenic bacteria, including *Lachnospiraceae*, *Ruminococcaceae*, and *Clostridium XIV* and a concomitant increase in pathogenic bacteria such as *Enterococcus*, *Enterobacteriaceae*, and *Bacteroidaceae* [25]. However, we realized a controversial result since pre-LT ESD patients exhibited high levels of *Ruminococcaceae* (using the rank-1 method). In addition, post-LT patients of DATASET1 had higher level of *Lachinospiraceae*, *Faecalibacterium prausnitzi*, *Ruminococcacae*, *Bacteroides*, *Enterobacteriacea* and *Streptococcus*. As long as DATASET2 is analyzed, healthy control subjects had high abundances of *Roseburia, Faecalibacterium, Bacteroides, Prevotella, Lachnospiracea*. Furthermore, whereas post-LT data demonstrated high abundances of *Bacteroides*, *Faecalibacterium*, *Prevotella* and *Megamonas* but lesser *Bacteroides*, *Lachnospiraceae*, *Blutia* and *Streptococcus*.

In conclusion, it seems even with the same SOT, there is a high variation in OTU levels. Topic models are expressive methods for dimensionality reduction, that can help to analyze variations in gut microbial communities. With topic models the analysis can be carried out at a level close to natural language that can be easily interpreted by clinicians, because most abundant species are automatically selected, and the microbial dynamics can be tracked and followed over time. This objective could be greatly facilitated by the availability of dynamic versions of topic-based models [26], which could help in monitoring post-SOT phase for prescribing the best treatment option. Future research could take these questions as the point of departure to narrow down the focus.

Author contributions. Conceptualization, LS, MB and SM; methodology, LS and MB; validation, JH, BZ and AB; investigation, LS, MB and SM; resources, LS; data curation SM; statistical analysis MB; visualization, JH, BZ and AB; supervision, MB; project administration, LS, MB and SM; writing–original draft preparation, LS, MB and SM; writing–review and editing, JH, BZ and AB. The authors declare that they have no conflicts of interest. No additional data are available.

References

1. Black, C.K., Termanini, K.M., Aguirre, O., Hawksworth, J.S., Sosin, M.: Solid organ transplantation in the 21st century. Ann. Transl. Med. **6**(20), 409 (2018). https://doi.org/10.21037/atm.2018.09.68
2. Stehlik, J., Kobashigawa, J., Hunt, S.A., Reichenspurner, H., Kirklin, J.K.: Honoring 50 years of clinical heart transplantation in circulation. Circulation **137**(1), 71–87 (2018). https://doi.org/10.1161/CIRCULATIONAHA.117.029753
3. Gurley, K.E., Lowry, R.P., Clarke Forbes, R.D.: Immune mechanisms in organ allograft rejection: II. T helper cells, delayed-type hypersensitivity, and rejection of renal allografts. Transplantation. **36**(4), 401–405 (1983)
4. Fishman, J.A.: Infection in organ transplantation. Am. J. Transplant. **17**(4), 856–879 (2017). https://doi.org/10.1111/ajt.14208
5. Chong, A.S., Alegre, M.-L.: The impact of infection and tissue damage in solid-organ transplantation. Nat. Rev. Immunol. **12**(6), 459–471 (2012). https://doi.org/10.1038/nri3215
6. Mori, D.N., Kreisel, D., Fullerton, J.N., Gilroy, D.W., Goldstein, D.R.: Inflammatory triggers of acute rejection of organ allografts. Immunol. Rev. **258**(1), 132–144 (2014). https://doi.org/10.1111/imr.12146
7. Annavajhala, M.K., et al.: Colonizing multidrug-resistant bacteria and the longitudinal evolution of the intestinal microbiome after liver transplantation. Nat. Commun. **10**, 4715 (2019). https://doi.org/10.1038/s41467-019-12633-4
8. Maynard, C.L., Elson, C.O., Hatton, R.D., Weaver, C.T.: Reciprocal interactions of the intestinal microbiota and immune system. Nature **489**(7415), 231–241 (2012). https://doi.org/10.1038/nature11551
9. Atarashi, K., et al.: Induction of colonic regulatory T cells by indigenous Clostridium species. Science **331**(6015), 337–341 (2011). https://doi.org/10.1126/science.1198469
10. Dodd, D., et al.: A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. Nature **551**(7682), 648–652 (2017). https://doi.org/10.1038/nature24661
11. Bajaj, J.S., et al.: Liver transplant modulates gut microbial dysbiosis and cognitive function in cirrhosis. Liver Transpl. **23**(7), 907–914 (2017). https://doi.org/10.1002/lt.24754
12. Chalmers, J.D.: Microbial dysbiosis after lung transplantation. Am. J. Respir. Crit. Care Med. **194**(10), 1184–186 (2016). https://doi.org/10.1164/rccm.201606-1178ED
13. Duvallet, C., Gibbons, S.M., Gurry, T., Irizarry, R.A., Alm, E.J.: Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. Nat. Commun. **8**, 1784 (2017). https://doi.org/10.1038/s41467-017-01973-8
14. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. **3**, 993–1022 (2003)
15. Blei, D.M.: Probabilistic topic models. Commun. ACM. **55**(4), 77–84 (2012). https://doi.org/10.1145/2133806.2133826
16. Sankaran, K., Holmes, S.P.: Latent variable modeling for the microbiome. Biostatistics **20**(4), 599–614 (2019). https://doi.org/10.1093/biostatistics/kxy018
17. Yan, J., et al.: MetaTopics: an integration tool to analyze microbial community profile by topic model. BMC Genomics **18**(Suppl. 1), 962 (2017). https://doi.org/10.1186/s12864-016-3257-2

18. Chen, X., He, T., Hu, X., Zhou, Y., An, Y., Wu, X.: Estimating functional groups in human gut microbiome with probabilistic topic models. IEEE Trans. Nanobiosci. **11**, 203–215 (2012). https://doi.org/10.1109/TNB.2012.2212204
19. Okui, T.: A Bayesian nonparametric topic model for microbiome data using subject attributes. IPSJ Trans. Bioinform. **13**, 1–6 (2020). https://doi.org/10.2197/ipsjtbio.13.1
20. Wallach, H.M., Murray, I., Salakhutdinov, R., Mimno, D.: Evaluation methods for topic models. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML 2009), pp. 1–8. ACM Press, New York, USA (2009). https://doi.org/10.1145/1553374.1553515
21. Lu, H.-F., et al.: Fecal microbiome data distinguish liver recipients with normal and abnormal liver function from healthy controls. Front. Microbiol. **10**, 1518 (2019). https://doi.org/10.3389/fmicb.2019.01518
22. Grün, B., Hornik, K.: `topicmodels`: an R package for fitting topic models. J. Stat. Softw. **40**(13) (2011). https://doi.org/10.18637/jss.v040.i13
23. Nikita, M.: `ldatuning`: Tuning of the Latent Dirichlet Allocation Models Parameters (2019). https://cran.r-project.org/package=ldatuning
24. Boyd-Graber, J., Mimno, D., Newman, D.: Care and feeding of topic models: problems, diagnostics, and improvements. In: Airoldi, E.M., Blei, D., Erosheva, E.A., Fienberg, S.E. (eds.) Handbook of Mixed Membership Models and Their Applications, pp. 225–274. CRC Press, Boca Raton (2014)
25. Punzalan, C., Qamar, A.: Probiotics for the treatment of liver disease. In: Floch, M.H., Ringel, Y., Walker, W.A. (eds.) The Microbiota in Gastrointestinal Pathophysiology, pp. 373–381. Elsevier, Amsterdam (2017). https://doi.org/10.1016/B978-0-12-804024-9.00040-9
26. Jahnichen, P., Wenzel, F., Kloft, M., Mandt, S.: Scalable generalized dynamic topic models. In: Storkey, A., Perez-Cruz, F. (eds.) Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics PMLR 2018, vol. 84, pp. 1427–1435 (2018)