MINI-REVIEW

Moving beyond descriptions of diversity: clinical and research implications of bacterial imbalance in chronic rhinosinusitis*

Brett Wagner Mackenzie1, David W. Waite2, Michael Hoggard1, Michael W. Taylor3,4, Kristi Biswas1, Richard G. Douglas1

1 School of Medicine, Department of Surgery, The University of Auckland, Auckland, New Zealand
2 Australian Centre for Ecogenomics, School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Australia
3 School of Biological Sciences, The University of Auckland, Auckland, New Zealand
4 Maurice Wilkins Centre for Molecular Biodiscovery, The University of Auckland, Auckland, New Zealand

Rhinology 55: 291-297, 2017
https://doi.org/10.4193/Rhino17.135

*Received for publication: June 21, 2017
Accepted: August 27, 2017

Abstract
Chronic rhinosinusitis (CRS) is a debilitating disease which affects 5-16% of the general population and involves long-term inflammation of the sinonasal cavity. While microbial involvement in the pathogenesis of CRS has long been suspected, the exact role of microbes remains unclear. Recent application of cultivation-independent, molecular methods has provided much new information, taking advantage of developments in both laboratory- and bioinformatics-based analyses. The aim of this mini-review is to present a variety of available bioinformatics approaches, such as data classification techniques and network analyses, with proven applications in other aspects of human microbiome health and disease research. The uses of molecular techniques in the clinical setting are still in its infancy, but these tools can further our understanding of microbial imbalance during chronic disease and help guide effective patient treatment. The mini-review emphasises ways in which CRS bacterial gene-targeted sequencing data can progress beyond descriptive summaries and toward unlocking the mechanisms by which bacterial communities can be markers for sinus health.

Key words: human microbiome, chronic rhinosinusitis, microbial network, microbial co-occurrence, fragmentation

Introduction
Recent advances in sequencing technologies, and the bioinformatics tools used to analyse sequence data, have offered unprecedented insights into microbial involvement in human health and disease. The relative affordability, ease of sample preparation, and ability to multiplex hundreds of samples into a single sequencing run have made human microbiome research accessible and common. However, such studies are not without their challenges and truly informative studies are the result of crucial a priori decisions about study design. Studies focusing on the role of microbes in the pathogenesis of chronic disease will need to move beyond abundance descriptions (“who’s there?”) and towards characterising the alterations in microbial community structure and function that initiate, promote, and sustain chronic disease. This article aims to highlight the variety of available strategies for analysing the most common type of sequencing data (gene-targeted amplicon data) and the clinical and research implications gained from more sophisticated analysis of these datasets.

Chronic rhinosinusitis (CRS) is a debilitating disease, which affects approximately 5-16% of the global population, and presents a massive financial burden estimated at $13 billion in the United States per year and more than $30 million per year in revision surgeries, alone, in the UK (1-6). CRS is a complex group of diseases resulting from interactions between host genetics, immune system, and microbiome. CRS is defined as inflammation of the sinonasal mucosa lasting longer than 12 weeks and is characterized by nasal congestion or discharge, facial pain or pressure, loss of sense of smell, with endoscopic findings of polyposis and/or mucopurulent discharge (7). In patients where current treatment strategies, including corticosteroids, nasal saline irrigation, and short- or long-term antibiotics fail to remediate disease, functional endoscopic sinus surgery (FESS) is recommended to clear the sinonasal cavity of mucus and inflamed tissue and open obstructed drainage pathways. However, despite initial improvements in symptoms and high
success rates, approximately 10-15% of patients will be required to undergo revision surgery (8).

Several hypotheses exist regarding the role of the sinonasal bacterial community in CRS (9). One hypothesis, which reflects findings from other areas of human microbiome research, proposes that an imbalance (dysbiosis) of the bacterial community contributes to CRS pathogenesis and/or exacerbation of disease (9–11). The causality of dysbiosis in CRS pathogenesis remains to be determined. An aberrant host immune response combined with sinonasal inflammation may trigger dysbiosis. Alternatively, a change in the structure and function of the bacterial community (as a response to interactions among or between environmental processes, available resources, bacterial, viral, fungal or archaeal members and/or the host) may provoke an aberrant immune response which incites mucosal inflammation. In either case, understanding the mechanisms by which dysbiosis occurs, and the effects of the altered structure and function on the bacterial community itself, as well as the host health status, are pertinent to elucidating the role of microbes in CRS.

Research regarding the microbial role in the pathogenesis of CRS has shifted from a single-pathogen hypothesis to a community change approach that favours molecular techniques over traditional microbial culture. Molecular microbiological methods encompass both standard gene-targeted sequencing as well as the group of ‘meta–omics’ techniques. These techniques encompass whole-environment analysis of DNA (meta-genomics), expressed RNA (meta-transcriptomics), translated proteins (meta-proteomics) or metabolite profiles (meta-metabolomics). The major difference between gene-targeted and meta-omics techniques is that the latter does not involve amplification of a specific gene in sample preparation prior to sequencing.

Gene-targeted sequencing of variable regions in the bacterial 16S rRNA gene is by far the most common type of sequencing applied for studying the bacterial role in CRS, and has revealed microbial communities in both healthy and CRS-affected individuals that are characterized by high inter-personal variation (10,12). While the number of published articles focusing on the role of the microbiome in CRS has increased in recent years, a consensus on how microbial community diversity differs between diseased versus healthy sinuses has not yet been reached. This could be due to a combination of factors and challenges associated with using sequencing technologies to study CRS such as statistically underpowered studies, differences in DNA extraction methods and sample preparation, choice of PCR primers and thermocycling protocols, sequencing technologies, and bioinformatics analyses. The aim of this mini-review is to present a variety of available bioinformatics techniques, with proven applications in other aspects of human microbiome health and disease research, that offer ways in which CRS bacterial gene-targeted sequencing data can progress beyond descriptive summaries and toward unlocking the mechanisms by which microbial communities can be markers for sinus health.

**Moving beyond descriptions of diversity**

The number of bioinformatics programs available for analysing high throughput-sequencing-based data (>5000 tools available) far surpasses that of any other type of biological data analysis (approximately 2400 tools, including mass spectrometry, PCR, bioimaging, microarray, NMR spectroscopy, flow cytometry and mass cytometry, DNA fingerprinting, and nCounter system data) (https://omictools.com). Many programs specific to bacterial 16S rRNA gene sequencing data exist for initial quality filtering of
Table 1. Summary of the different microbial community analyses discussed in this review and examples of their application.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Citation</th>
<th>Application</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supervised classification</td>
<td>Knights et al., 2010</td>
<td>Identify discriminative taxa &amp; classify unlabelled data</td>
<td>Yan et al., 2013</td>
</tr>
<tr>
<td>Linear discriminant analysis (LDA) effect size (LEfSe)</td>
<td>Segata et al., 2011</td>
<td>Biomarker discovery</td>
<td>Rooks et al., 2014; Bourrier et al., 2016</td>
</tr>
<tr>
<td>Co-occurrence Network inference (CoNet)</td>
<td>Faust et al., 2012</td>
<td>Detects non-random patterns of co-occurrence and exclusion</td>
<td>Wang et al., 2016; Soffer et al., 2015; Welsh et al., 2015</td>
</tr>
<tr>
<td>Fragmentation analysis</td>
<td>Widder et al., 2014</td>
<td>Identify “gatekeeper” bacteria</td>
<td>Wagner Mackenzie et al., 2016</td>
</tr>
<tr>
<td>Functional prediction (PICRUSt, Tax4Fun)</td>
<td>Langille et al., 2013</td>
<td>Functional profile prediction of bacterial communities</td>
<td>Goodrich et al., 2014; Clemente et al., 2015</td>
</tr>
</tbody>
</table>

Supervised learning techniques to classify data
Supervised classification is a type of machine learning adapted from microarray and macro-ecology datasets, and is a powerful tool for predicting patterns in highly complex datasets (21). The major advantage with supervised learning is that once a user-defined model is built, it can be used to accurately classify unlabelled data. A training dataset of samples with taxon abundances is initially used to create a model that incorporates good discrimination of samples based on treatment, good generalization, and with an optimized expected prediction error (the known error of the model on future predictions). Several filtering methods can be applied to remove uninformative features and restrict the number of predictive taxa. In the case of CRS-related microbial research, supervised learning classifiers could, for example, be used to predict progression of disease from a group of acute rhinosinusitis patients that may be susceptible to developing CRS, or identifying features that help determine FESS recovery in severe CRS cases. In one study of *Staphylococcus aureus* carriage in sinonasal sites, a random forests type of supervised learning approach was applied to identify bacteria that were associated with carriage status (22). Several different types of classification methods are freely available through the statistical software package R and many in-depth summaries are available to help guide researchers in their decision making when applying supervised learning techniques to their data (21,23–25).

Linear discriminant analysis (LDA) effect size (LEfSe)
Linear discriminant analysis (LDA) effect size (LEfSe) is a relatively new technique designed specifically for biomarker discovery from sequencing data (26). LEfSe analyses first perform a series of significance tests to identify important features that differentiate groups of samples, then the effect size of each feature is estimated. LEfSe is an easily applied technique that can

raw sequencing data, clustering of similar sequence reads into groups (termed operational taxonomic units, OTUs), taxonomic assignment of OTU clusters, and downstream processing of quality filtered, taxon-assigned data to analyse differences between sample groupings. The most popular bioinformatics programs, such as mothur and QIIME, offer user-friendly workflows and tutorials (13,14). However, the results generated from these generic workflows offer only a preliminary glimpse at the data and should serve as a starting point for subsequent analyses. The most popular bacterial community sequencing data analyses include alpha and beta diversity measures (Figure 1). Alpha diversity measures diversity within a sample, essentially how many species of bacteria are in a sample (richness) and how the various species are taxonomically distributed across that sample (evenness) (15,16). A variety of diversity estimates exist that make different assumptions about the data and place different emphasis on richness or evenness (17). Alpha rarefaction curves provide insights into the sampling effort and sequencing depth of a group of samples by interpolating collected data to predict the true diversity within the microbial community (Figure 1A) (16). Beta diversity measures differences in bacterial composition and abundance between samples or treatments. A variety of beta diversity measurements exist that weight community diversity and abundance differently. Most often, beta diversity measurements are used to investigate differences between treatments. These results are then visualised using plots to show how samples arrange relative to each other in space (Figure 1B). Several comprehensive reviews focussing on these basic analyses are available (18–20).

What follows next are several types of analyses which can be applied to gene-targeted data that move us beyond basic descriptions of community diversity. This is not a complete list of available bioinformatics analyses, nor do we suggest that all methods are appropriate for every dataset. Rather, we encourage researchers in CRS to consider seeking out and adapting bioinformatics tools from other human microbiome studies that, when used in conjunction with the familiar alpha and beta diversity tests, improve understanding. A summary of these techniques and their applications can be found in Table 1.
be used online through Galaxy (http://huttenhower.sph.harvard.edu/lefse/) or applied through the bioinformatics package mothur. LEfSe analysis parameters are modifiable, which means stringent or relaxed limits can be specified by the user in order to explore biomarkers at different effect sizes (levels of importance). Additionally, because of its ability to include primary and secondary phenotypic structures (disease state combined with other patient descriptors) in the model, LEfSe lends itself easily to CRS datasets where patients may have known comorbidities such as asthma. LEfSe has successfully been applied in several studies to differentiate disease states based on microbiome data.

**Co-occurrence Network inference (CoNet)**

Microbial network analyses aim to uncover the underlying ecological relationships, such as mutualism, parasitism, commensalism, and competition, among taxa in a microbial community. Reviews highlighting the strengths and weaknesses of network analyses used in microbial datasets have recently been published. Co-occurrence Network inference (CoNet) is one type of network analysis tool that identifies significant co-occurrence and exclusion patterns of bacterial taxa based on relative abundances. Many challenges exist when analysing relative abundance estimates, as an increase in one taxon means that by definition there must be a decrease in at least one other taxon, and this can result in false correlations between taxa. CoNet has mitigated these issues by developing an integrated approach to measure community organization. CoNet predicts association networks from several similarity and dissimilarity measures (Pearson Correlation, Spearman Correlation, Kullback-Leibler divergence, and Bray-Curtis dissimilarity) using a generalized boosted linear model that combines the individual metrics. CoNet is used through Cytoscape, open source software that is available from www.cytoscape.org.

CoNet has been applied to the Human Microbiome Project dataset to uncover microbial relationships across body areas. Those results suggest taxa that tend to significantly co-occur will have a cooperative relationship, whereas taxa that tend to negatively associate (exclude each other) have a competitive association. These relationships can be explained by phylotypic distance and functional similarities. For example, taxa that are distantly related but with similar functional roles tend to compete. CoNet has been successfully applied to study the HMP dataset, lung microbiome networks in patients with chronic obstructive pulmonary disease exacerbations, and environmental microbiology. However, only one study has been published applying CoNet to CRS datasets. CoNet results from this recent meta-analysis of CRS bacterial community data revealed bacteria that were associated with sinus health status, and potential cooperative or competitive interactions between taxa. CoNet, and other appropriate network analyses, can be used to further our understanding of dysbiosis and how it affects sinonasal microbial composition and health.

**Bacterial network fragmentation**

Network analyses, such as CoNet (discussed above), are useful for identifying bacteria that cooperate or potentially compete in a microbial community. However, understanding the organization of these networks is critical for tracking network persistence, stability, and how bacterial communities respond to disturbances. Fragmentation is a type of analysis used to describe the topology, or organization, of a bacterial network. Specifically, fragmentation measures the number of disconnected subgroups of taxa (Figure 2). Highly fragmented communities have an increased number of isolated subgroups, and increased fragmentation is associated with unstable bacterial communities. Measuring fragmentation is an extremely useful tool for studying how microbial communities organize themselves during health and disease. Computer modelling experiments can be conducted on abundance datasets to simulate the removal of highly connected taxa, and study the effects on the network. If
highly connected taxa are removed from the network, causing fragmentation scores to significantly increase, then the removed taxon can be termed a ‘gatekeeper’ \cite{40,41}. Gatekeeper bacteria are central to connecting many parts of the network to hold it together, as they appear to be critical to the structure and function of the bacterial community. Taken together, network co-occurrence conducted with fragmentation analyses can be used to discover gatekeeper bacteria, to elucidate meta-community dynamics, and predict overall bacterial community changes in network function when it’s disrupted (Figure 2C). Fragmentation analyses have been used to study environmental microbial communities \cite{42,43}. However, this analysis has had limited application to study human-associated microbial network construction \cite{44,45}. A recent meta-analysis of bacterial 16S rRNA gene sequencing data applied CoNet and fragmentation analyses to propose that CRS-associated bacterial communities have significantly increased fragmentation when compared to healthy sinonasal communities. Furthermore, the results suggest that Propionibacterium and Burkholderia are gatekeeper bacteria which may be critical for maintaining healthy sinonasal bacterial networks. Other key gatekeeper bacteria may exist in the sinus microbiome that have not yet been identified, but with good quality sequence datasets, this can be investigated in the future.

Functional prediction techniques
All the analyses discussed so far in this review aim to explore the composition and diversity of bacterial communities. A critical component for understanding the role of the microbiota in health and disease is to understand if there is a shift in community function, i.e. a change in the metabolic capacity of the bacteria which is a result of, or contributor to, the disease state. In order to discover in situ community functional capacity, total DNA from a sample needs to be sequenced, utilizing metagenomics techniques. However, metagenomics techniques are costly and, due to the overwhelming proportion of contaminating human DNA, very few samples can be sequenced at once (Wagner Mackenzie et al., submitted). Two programs, Tax4Fun and PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states) are currently available for functional prediction of microbial communities from gene-targeted sequencing data \cite{44,45}. The results from these programs provide a broad overview and prediction of possible genes present within a sample on the basis of previous observation about the detected taxa. They are not a replacement for metagenomic sequencing which provides direct, total, genetic information from all microbial components (viral, fungal, bacterial, and archaeal) in a sample. However, functional prediction of bacterial communities can offer insights into differences in overall community function, and a large portion of the data used for training PICRUSt was obtained from human-derived samples. Functional prediction is therefore a useful hypothesis-generating tool, and can help inform a decision to pursue metagenomic sequencing of these samples. Functional prediction of bacterial communities has successfully been applied to other human microbiome research \cite{46,47}, and can be easily implemented in QIIME or the statistical software package R.

Implications for treatment and research
Medical treatment recommendations for CRS with polyps include topical and oral corticosteroids and antibiotics. There are few data that clearly support the efficacy of antibiotics for the management of CRS, with the exception of longer term courses of macrolides. In spite of the limited evidence of efficacy, antibiotics are still widely prescribed to patients with CRS \cite{49}. The effects of broad-spectrum, systemic antibiotics on the sinonasal microbiome, and their unintended consequences on bacterial communities like the gut microbiome, could potentially be significant since they affect a wide range of bacterial taxa. Some of the adverse follow-on effects of overtreatment with antibiotics could include: removal of keystone or gatekeeper bacterial species, an increase in antibiotic resistance, and loss of diversity which can contribute to disease predisposition and perpetuation of an aberrant bacterial community structure \cite{48}. Randomised controlled trials (RCTs) studying the efficacy of long-term antibiotic usage for CRS treatment are contradictory \cite{49–51}. However, RCTs are a critical next step for applying the results from microbiome studies to patient treatment, and understanding the effects of antibiotic treatment on the overall structure of the sinonasal microbial community. For example, bacterial co-occurrence and network fragmentation analyses can be used to identify the effect of antibiotic treatment on co-occurrence patterns of bacteria and community stability. Fragmentation analyses combined with alpha diversity measurements and patient symptom scores could be used to help guide the development and describe the effect of potential probiotic treatments that remediate a dysbiotic microbial community observed in CRS.

The different analyses presented in this review offer ways in which clinicians and researchers can better interpret microbial sequencing data in CRS. The clinical applications of these molecular methods are still in their infancy, and culture techniques are the standard for diagnosing bacterial pathogens. Currently, these molecular methods are economically disadvantageous to implement in clinic and diagnostic settings because they require the use of specialised equipment and trained staff. However, the future of microbiome research in CRS should be to develop predictors of disease severity and prognosis, as well as to facilitate the design of more effective and efficient alternatives to current treatment. To this end, application of molecular techniques that move beyond bacterial composition may help mitigate the needs for surgery and ongoing care for this chronic disease, ultimately reducing hospital stays, leading to faster diagnoses.
and earlier, more effective treatments (13). The development of prebiotics and probiotics, as a complementary treatment with corticosteroids and/or surgical intervention, is promising and should be explored (13).

Concluding remarks
A solid foundation of exploratory and preliminary studies describing bacterial community composition in CRS and healthy patients exists (7–12, 34–36). These studies have generated a plethora of questions regarding the mechanisms underlying the role of the microbiota in CRS. If sinonasal microbial dysbiosis is fundamental to the pathogenesis of CRS, then future studies utilizing gene-targeted sequencing must move beyond classic descriptions of composition, and incorporate larger sample sizes of well-characterised patient data. The restriction of short-length amplicon sequencing to genus-level taxonomic identification compels the addition of complementary data such as host immune, fungal, viral, or qPCR data. These bioinformatics analyses are examples of a few well-tested methods which have helped further our understanding of the role of the microbiome in other clinical areas, and hold promise for helping advance our knowledge of the potential role of the sinonasal microbiota in CRS pathogenesis. Given the frequency of prescribing of antibiotics for this condition, having a sound microbiological rationale guiding the use of these drugs is crucial.

Acknowledgements
The research in this study was supported by The Garnett Passe and Rodney Williams Memorial Foundation Charitable Trust Fund.

Authorship contribution
BWM constructed the original figures, and wrote and revised the work. All authors contributed to the conception of ideas expressed in this review and editing of the manuscript. KB, MT, and RD supervised the work.

Conflict of interest
The authors declare that this research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References
23. Holmes I, Harris K, Quinn C. DeepCult multimonial mixtures: Generative models for

Implications of bacterial imbalance in CRS

Richard G. Douglas, MD, FRACS
Department of Surgery, Level 12
Auckland City Hospital
Park Road
Grafton
Auckland 1023
New Zealand

E-mail: richard.douglas@auckland.ac.nz