RJ10527 (ALM1502-010) February 20, 2015
Other

IBM Research Report

Defining the Food Microbiome for Authentication, Food Safety, and Probiotic Validation

James H. Kaufman, David Chambliss
IBM Research Division
Almaden Research Center
650 Harry Road
San Jose, CA 95120-6099
USA

Bart C. Weimer, Dylan Storey
University of California, Davis
Davis, CA 95616
USA

Research Division
Almaden – Austin – Beijing – Brazil – Cambridge – Dublin – Haifa – India – Kenya – Melbourne – T.J. Watson – Tokyo – Zurich

Defining the food microbiome for Authentication, Food Safety, and Probiotic Validation

James H. Kaufman and David Chambliss
IBM Almaden Research Center
650 Harry Rd., San Jose CA 95125 USA
jhkauf@us.ibm.com, chamb@us.ibm.com

Bart C. Weimer and Dylan Storey
University of California, Davis
School of Veterinary Medicine
Davis, CA 95616 USA
bcweimer@ucdavis.edu, dstorey@ucdavis.edu

Abstract:
In light of recent revelations surrounding the veracity of claims regarding supplements and probiotic additives and foods, it is clear that culture independent methods for the validation and authentication of these products is needed. In this whitepaper we propose and outline a strategy that leverages high throughput sequencing (HTS) technologies, curated and annotated reference whole genome sequences, bacterial strain banks, previously described and novel informatics techniques, and a scalable computing platform to provide a solution that will be robust and extendable to encompass the supply chain. Through this work we propose to:

- Measure a baseline microbiome for raw products that commonly have probiotic cultures added to them during process, which will serve as a reference database of microbiomes.

- Measure the shift in the microbiome after the addition of probiotics to raw material. We propose carrying out experiments to generate data required to create predictive models that capture the effects of time, environment, and dose in order to significantly change the foundational microbiome.

- Characterize the resulting, steady state, or time varying microbiome in finished products. These measurements will be crucial in providing input data for developing assays for food authentication, validation of live probiotic organisms, and determine the robustness of the predictions for finished products.

Introduction:
According the Food and Agriculture Organization [1], by 2050 the world's food supply must grow by 70% to meet the increase in the world population [2]. That growth is projected to occur primarily in tropical zones, where agriculture is largely developing and contains many opportunities for fraud and addition of bacteria that are not normally part of the food supply today. Today, our food supply is global and complex. This complexity will only increase in the future with a further globalization. The use of probiotic bacteria has been projected to increase with an increasing population and globalization of food. It is common to add probiotic to many types of foods that result in microbiome shifts via live and active strains [3].

Our ability to ensure that food and food ingredients are authentic throughout our food supply chain is paramount if we wish to guarantee the veracity of supplier claims and to ensure that our food supply is both safe and well regulated. Use of advanced techniques based in genomics is becoming widely

available around the world. Culture independent methods are gaining acceptance but remain unvalidated and not widely implemented in the food supply chain.

The combination of food authenticity, safety, and probiotic addition brings a new challenge to food safety that has long been contemplated, but until recently technologies required were not sufficiently developed to provide a scientifically sound and integrated solution. To bring this analytical challenge into focus, fundamentally new technologies must be brought to bear that can be used routinely for global food surveillance. By assuring food safety up and down the supply chain, these technologies should help food producers, ingredient distributors, processors and manufacturers ensure that they comply with international standards to provide confidence that enables trade. By providing the means to quickly identify unsafe ingredients before they are delivered up the supply chain, billions of dollars of revenue can be saved while further minimizing risks to consumers while ensuring added beneficial microbes meet claims on the product label.

Both health and nutritional benefits of probiotics have been claimed [4]. Barzegari, et al. suggest that probiotic additives should be designed with respect to the native microbiome for different ingredients[4]. Ganesan et al. (2014) found that probiotics will survive in food ingredients for extended periods, but they may not be culturable yet actively producing bioactive compounds[3]. Together, these findings indicate that the baseline or foundational microbiome of a food can shift as beneficial bacteria are added that may be predictive of addition and safety. This concept opens the door for addition of bacteria to create a stable and disease resistant microbiome in the supply chain that ultimately protects the food from disease outbreaks.

The standard for probiotic addition includes addition of live and active cultures. Standard plating techniques are unlikely to provide accurate profiles of the probiotic microbiome in the food ingredients. Interestingly, consumption of single probiotic bacteria can shift the food microbiota [5], as can bioactive ingredients that alter the microbial community using small molecules [6-10]. This body of work creates an interesting analytical challenge that must be solved using multi-omic integration.

Measuring the addition of specific bacteria can be done using metagenomic tools, but since we are interested in live bacteria, advanced tools for metaRNAseq can be used, not as a tool for gene expression, but rather as a tool to reveal genes and gene sets that modulate during microbiome shifts. This approach may be useful analytically by measuring the probiotic with genetic tools directly and measuring the resulting microbiome of the food containing the probiotic culture so as to define hundreds of evidence lines that may be predictive of specific organisms in the community and how they shift the microbiome in constructive or ‘healthy’ ways. Constructive shifts may include resistance to spoilage and/or protective from pathogenic incursion. Conversely, undesirable shifts could lead to increase risk of disease and growth of pathogens to a sufficient level and activity that it may lead to possible human health risk.

With use of advanced high throughput sequencing (HTS) technologies it is possible and practical to follow the microbiota of foods and food ingredients using metaRNAseq to determine the live and active organisms. This technique can provide conclusive evidence of live and active bacteria in the food as well as quantitative estimates for monitoring probiotic mixtures and validating ingredient claims as well as an initial set of data to determine if baseline microbiota can shift predictably with the addition of specific organisms. Parallel controlled studies of the metabolome in the same samples can provide evidence for the composition, diversity, and metabolic bi-products of probiotic additives as well as defining new fast and inexpensive small molecule monitors and tests that can be easily implemented in a manufacturing environment as surrogate markers that can be an early warning sign to use more advanced and definitive tests, such as HTS to confirm the initial results.

We propose that a multicomponent program is required to properly overcome challenges associated with studying complex communities. He et al. recently demonstrated that there is a predictable and microbe specific response in vitro [11]. They combined metaRNAseq and NMR metabolomics to determine that E. coli strains have different small molecule signatures during infection and association. We propose that by extending this concept using HTS metaRNAseq and metabolomics directly in food, we can establish the baseline microbial community for food ingredients and also measure how the inclusion of new organisms shifts the microbiome. By leveraging a systems biology approach in an environmental setting we can define these samples in terms of both taxonomy and function. To ensure that this is done in a timely fashion it is paramount that flexible and scalable informatics solutions are developed and used throughout the proposed research.

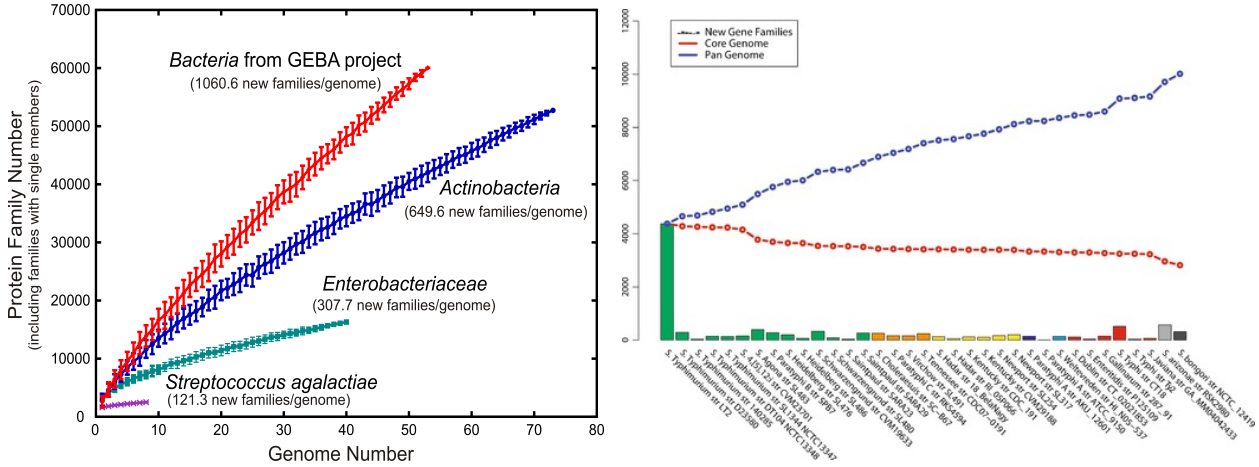

Figure 1: Increase in new gene/protein families with increasing genome sequences.
Panel A represents microbe families (Wu et al. 2009).
Panel B represents *Salmonella* enterica (Jacobsen et al., 2011).

Implementation of an NGS-based approach for authentication of probiotic microbiota in food requires a reference database of whole genomes (WGD) and metagenomes (MGD). WGD are needed to ensure that added bacteria are classified correctly in the community and captures the genomic diversity of specific organisms. The food microbiome database is also needed to determine what is 'normal' and how it shifts with and without probiotic bacteria supplementation. The *absence* of such a standard database is a significant hindrance to global food safety and public health in broadly implementing genomics-based methods. Use of projects like GenomeTrakr and the 100K Pathogen Genome Project attempt to solve this for pathogenic bacteria, but there is not yet a coordinated effort to address this problem or to validate the completeness of candidate references in the probiotic arena. Validation and authentication of probiotic composition, by definition, must detect not only organisms by names, but also the active genes in a probiotic microbiome. The Genomic Encyclopedia of Bacteria and Archaea (GEBA) sequenced 3500 genomes are specific phylogenetic branches with the intent of increasing the breadth of phylogenomic-based assignments from an ecology perspective. With as few as 50 new genomes spread over the tree of life, this study identified approximately 1060 new protein families (Figure 1A). Figure 1B represents a complementary approach where 34 very closely related *Salmonella* genomes were sequenced revealing an apparent linear increase in new protein families. The Lactic Acid Bacteria Genome Consortium also used this approach in the early 2000's where the genomes of closely related lactic acid bacteria were sequenced. It resulted in reclassification of the entire group of organisms [12]. Since then additional genomes from this group of probiotic bacteria have been sequenced and technologies have advanced so that closed genomes can be produced in a short time.

***Table 1**. Genome sequence availability of probiotic bacteria from various public databases (date). Genome Sequence availability through NCBI. SRA submissions include both DNA and RNA sequencing samples.*

Organism	Assembled	Submitted to SRA
Lactobacillus	400	1594
Lactococcus	56	66
Bifidobacterium	202	207
E. coli (probiotic)	2	8142
Bacillus	688	825
Yeast	82	20,458

Genomic sequences availability for bacteria are certainly a resource for forming reference data sets for analysis, but these resources are uneven, poorly curated, of questionable quality, and do not adequately represent the required organism diversity (Table 1). Recently it was found that the SRA has a large proportion of sequences from bacterial sources that are not of sufficient quality to be used as a reference because they cannot be assembled into a meaningful sequence [14]. For example, the diversity of *Salmonella* genomes (Figure 2) suggests that a very large number of genome sequences may be required to provide an adequate reference database to pin point specific epidemiology questions and to begin tracking virulence traits, such as antibiotic resistance and effector molecule conservation that are known to be directly involved in disease risk and outbreaks. Figure 2 shows the cumulate rate of observation of "new" putative genes as a function of number of isolates sampled. The data was obtained by de-novo assembly of over 4000 individual isolates from 100K pathogen project. A subset of 792 samples were then selected at random in separate (jackknife) trials and the cumulative gene count was determined for each trial. The approximately linear behavior observed on the log-log plot suggests a power law behavior with the rate of cumulative gene discovery increasing with the number of isolates – as observed above with de novo assembly data.

$$N_{genes} \sim \eta_{isolates}^{0.465}$$

This type of scale free or power law behavior is reminiscent of the MacArthur-Wilson scaling law well established in insular biogeography and other population ecology studies [15]. Considering that horizontal gene transfer is common in the environment and among many types of bacteria it suggests that each and every isolate is like a "sample" from a different microbial community and the diversity of that community scales with the number of isolates just as the biodiversity in any ecological habitat scales with the area (or diversity) of the underlying habitat. Understanding the underlying scaling behavior will allow prediction of the number of samples required for gene discovery and can reveal the universality of observed exponents. This perspective is new to food microbiology and enables one to move to a population-based understanding of the isolates to treat them as individuals rather than a homogenous type of organism, e.g. *Salmonella*. The concept can also be expanded to metagenomes or metaRNAseq data sets where individual samples have ecologies that are unique within a type of environment (e.g. a factory, a field, or a food source) Over time, as the necessary date is collected,

one can foresee a new field of molecular risk assessment evolving that will bring about a new perspective on public health benefits and risks that enables a new type of large scale genomic-based microbiology to solve long standing questions that traditional isolate and growth microbiology cannot address. **For these reasons, the development of a probiotic reference culture and metagenomic databases is critical for avoiding errors in unintended conclusions such as concluding presence of an organism that is not actually in the sample.**

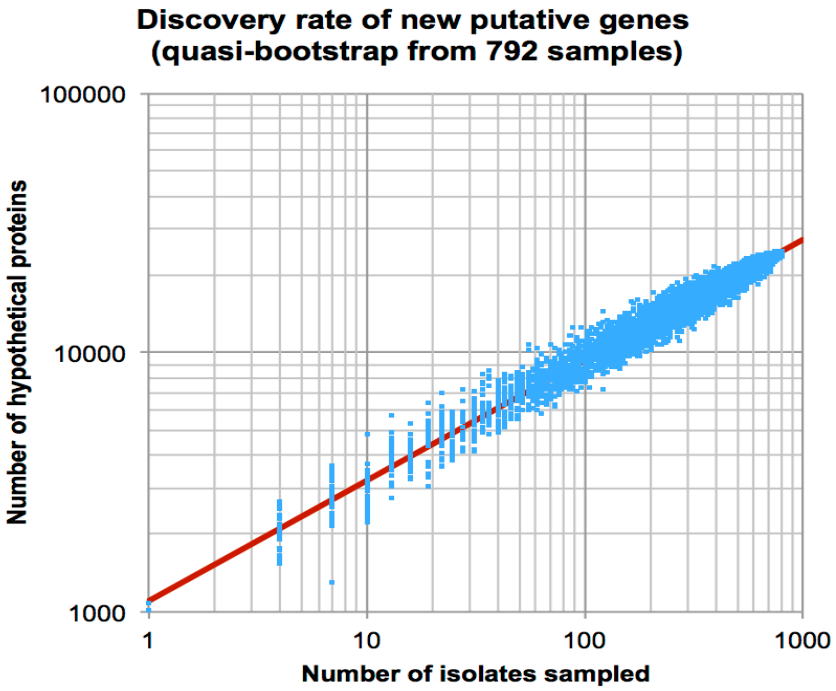

Figure 2 The cumulative rate of discovery of putative "new genes" as a function of the number of isolates of salmonella increases as a power law (linear on a log-log plot).

Microbial communities are dynamic systems that change when specific members, such as probiotics, enter the food chain. Traditionally, 16s rDNA sequences have been used to define the community membership. However, this measure is uneven for phylogenetic estimation and often leaves one wanting more information about the state of the community. Using metagenomics, the community membership and function can be used to define the microbiome in foods, with and without the addition of probiotic additives. This is predicated on the hypothesis that the background microbiome inherent to the food is stable enough to be predictive of both the food type (source) and shifts in response to addition of specific components that drive the biochemistry within the food to shift the community. Viewed as a biochemical dynamic system, a food or food ingredient provides an environment for a predictable set or genes (i.e. organisms and metabolic pathways) restrained by the *active genes* in the microbiome (i.e. expressed RNA). For this reason it is not sufficient to study microbiome DNA (most of which may be derived from the host ingredient itself). Rather, applying metaRNAseq methods it will identify both normal and abnormal microbial communities in terms of the active genes derived by live cells (host and microbiome) that cooperate to create a novel environment. The resulting data will be a diverse set of transcripts that are not intended to study the fundamental character of the food, but rather be used to discover patterns that predict changes to the microbiome using function and organism identity as a function of the classification statistics. **Measurement of microbiomes in food products will enable robust classification and predictions that can be used bioinformaticaly. It can also support definition of new biomarkers for surveillance assays**

to routinely test the fidelity of products in the food supply chain.

Focusing initially on the community metaRNAseq signature (mRNAsig) has several additional advantages. It is agnostic to and robust against horizontal gene transfer and the transfer of plasmids. Over time, evolving organisms may respond to pressure by incorporating new genes. The mRNAsig will detect this and can be leveraged for use in detection, source tracking, and outbreak investigation. Since the RNA transcriptome will also provide Ribosomal 16S RNA, this approach also provides an opportunity to calibrate metaRNAseq method against more traditional 16S community analyses allowing for both species and gene identification.

As we move to molecular methods unintended observations will be made for hazards that were not previously observed in the food supply. It is likely that in depth sequencing will uncover previously unrecognized community members that were not previously detected. However, one must be cautious to ensure that the sequences do not falsely indicate organisms that are not present if short sequence comparisons are used, as was done in a recent environmental survey. Creation of a probiotic reference culture and metagenomic database will reduce the likelihood of false positives, determine sensitivity and selectivity of molecular protocols, and support proper calibration of molecular methods all required for effective a regulatory policies and action.

To support this research agenda, we propose in parallel to create and demonstrate a scalable software system. This system must be extendible and adaptable with the understanding that not all organisms, and not all genes, have been identified in food systems. Analysis of a variety of microbiome meta-data sets demonstrates that as many as 50-80% of metaRNAseq reads are unidentified [14]. However, these same unidentified reads are repeatable features of datasets from equivalent sources. For this reason a metagenomic software service for food safety must track both annotated and un-annotated transcripts, and compare both known and unknown components from sample to sample. This necessitates a well-designed highly structured database and associated operations for querying and adding data. Furthermore, extraction of salient features from the data for use as predictive markers will be necessary in order to deliver robust assays for use in field settings.

IBM Research has extensive experience in the design and implantation of large scale software systems [15,16], extensible databases[17,18], high performance computing platforms and architectures [20,21], and the application of these systems to real world problems [22]. In particular the public health research group at IBM Almaden created widely adopted open source tools that helped enable Healthcare Information Exchanges in the U.S. [23-25], and helped implement in open source the reference standards that integrated HIE's with public health (as specified in todays regulations for "meaningful use" regulations [26-29]. Furthermore IBM contributed open source tools to the eclipse foundation that support rapid development of new epidemiological models for human, veterinary, and foodborne disease [30-33]. **The computational methods required for this project will leverage algorithms that are solvable in exponential time or greater and have similar memory requirements. These challenges require modern information management system that is scalable and designed to handle the challenge.** As such, our proposed research program includes an explicit aim for the development and testing of such a system.

Proposed Experimental Outline:

Aim 1: Characterize and quantify the baseline microbiome.

We hypothesize that microbial communities are strongly and consistently influenced by their surrounding environment. We anticipate that between communities group metabolism will be patterned, consistent, and predictive so that actively expressed gene sets can be used as markers of

a communities. We will determine the normality gene sets as well as those that are highly variable between foods, which will fuel authentication estimates as well as shifts within a product that will predict specific communities that are indicative of environmental shifts.

Expected outcome. We expect that communities from different regions (but the same food product) may vary in taxonomic community measurements but will be conserved in actively expressed metabolic pathways. As such we will investigate and compare communities in both terms of phylogenies and metabolic function so as to define specific sets of genes that are predictive. This approach will also be informed by the food type and chemistry.

Challenges. Obtaining appropriate samples. One challenge is the ability to extract nucleic acids from many food types. While there are challenges in the extraction of DNA (or RNA) from complex food matrices these troubles have been largely been overcome using variations on established methods. It is largely a problem of highly emulsified foods, but it will be determined on a case-by-case basis for foods with varying fat composition. It is also possible that the RNA will be degraded so that traditional analysis methods will not suitable. In this case, alternate non-parametric methods will be used that are enabled by references, but do not entirely rely on references.

Aim 2: Characterize and quantify the shifting micro-biome after the addition of probiotics to its final steady state in finished food products.

The introduction of new members to a community (e.g. probiotic additives) will shift the community equilibrium to a new steady state. We anticipate that this shift will be measurable and consistent through changes in gene expression (through RNAseq) and metabolites, and that these shifts will occur in a predictive manner over time. The acquisition of this data in combination with the baseline active microbiome estimates of raw products will also allow for the development of predictive models and robust assays through data feature extraction.

Defining the mathematical relationship between probiotic communities in finished products as a function of the starting microbial communities, introduced bacteria, and environmental factors requires a saleable microbiological and computational effort.

Aim 3: Systematic creation of a reference database of probiotic genomes from well defined isolates

Common analytical methods associated with HTS are not exempt to type I and II errors, which may cause an interpretation of risk where there is none. It will be imperative that this work ensures that these risks are kept to a minimum by implementing good experimental design and rigorous statistical techniques. In order to further minimize these risks the systematic creation of a reference database of probiotic genomes from well-defined isolates is paramount. While other projects are doing this for pathogens no such project exists for probiotic bacteria or metaRNAseq microbiome estimates.

Aim 4: Develop and deploy a scalable software service for the analysis of microbiomes and development of assays.

Bioinformatic workflows are too often built and executed *ad hoc* lab specific workflows with documentation of the actual process rarely being well recorded or presented. In order for next gen sequencing to realize its full potential across a wide variety of food authentication, quality, and food

safety applications, there is a need to expose, formalize, and standardize *transparent* workflows for bioinformatics analysis, processing, and data management. The provenance of intermediate and final results data must be retained in a system of record both to support scientific validation and to provide auditability of externalized conclusions. Industrial organizations seeking to leverage genomic, metagenomic, and metabolomic data for food safety require that the informatics technologies be available as an inexpensive and scalable cloud computing and web based service. With such a service, all data sets and the workflows used to create them should be stored using well-designed data management systems leveraging industry standard software and architecture. Since the bioinformatics used to create the required evidence is critical, we propose that an extensible and modular informatics software service should be created, demonstrated, and used as an integral part of this program. As discussed above, standard libraries for authentication of probiotic composition (and safety) will improve as more genome sequences become available, providing a robust database to represent the global genome for use in metaRNAseq and other metagenomic applications. As such, a bio-informatic software service must index and track the history of all reference databases used in analyses by dataset. It should also track and organize the inevitable discovery of "unknown" proteins (genes) and gene families and indexing that data to aid in the refinement and growth of the reference database itself.

Conclusion:

We propose that a coordinated effort be initiated to apply high throughput sequencing (HTS) technologies to the study of probiotic additives in food with the goal of validating the veracity of claims regarding supplements and probiotic formulations and foods. This effort requires a number of Aims, some of which can be pursued in parallel. In particular we propose to:

- Characterize and quantify the baseline microbiome
- Characterize and quantify the shifting micro-biome after the addition of probiotics to its final steady state in finished food products
- Create a reference database of probiotic genomes from well defined isolates
- Develop and deploy a scalable software service for the analysis of microbiomes and development of assays

Each of these aims is essential to meeting the stated goals, but can yield important benefits and impacts to other related and important goals. We expect this work will identify new biomarkers for surveillance assays to routinely test the fidelity of products in the food supply chain. Furthermore, the informatics service required to achieve the goals of this project will have direct application to an even broader range of goals for food safety. FDA, USDA, CDC, and other would all benefit from an informatics facility that is reproducible, scalable, flexible (extendible) and high performance. It will drive end-to-end workflows on large collections of samples, so that both the labor cost and risk of error in production deployments can be minimized. The platform will accommodate users with different roles and expertise, be resource cost-efficient, and scalable through cloud other distributed techniques.

References:

1. FAO; http://www.fao.org/news/story/en/item/35571/icode/
2. Fischer, R. A., Derek Byerlee, and Gregory O. Edmeades. "Can technology deliver on the yield challenge to 2050." *Expert Meeting on How to feed the World in*. Vol. 2050. 2009. *and* http://www.z-punkt.de/megatrends-uebersicht.html
3. Ganesan, B., Brothersen, C., & McMahon, D. J. (2014). Fortification of foods with omega-3

polyunsaturated fatty acids. *Critical reviews in food science and nutrition*, *54*(1), 98-114.
4. Barzegari, Abolfazl, and Amir Ata Saei. "Designing probiotics with respect to the native microbiome." *Future microbiology* 7.5 (2012): 571-575.
5. Ganesan and Weimer *in preparation*
6. Ferreyra, J., Katherine W. Bouley, D., **Weimer B.C.**, and Sonnenburg J. L., 2014. Gut microbiota-produced succinate promotes *Clostridum difficile* infection after antibiotics or motility disturbance. *Cell Host Microbe*. 16(6):770-7. (PMID: 25498344);
7. Marcobal A., Barboza M., Sonnenburg E.D., Martens E., Desai P., Lebrilla C., **Weimer B.C.,** Mills D.A., German B., Sonnenburg J.L. 2011. Bacteroides in the Infant Gut Consume Milk Oligosaccharides via Mucus-Utilization Pathways. *Cell Host Microbe*. 10:507-14. (PMID: 22036470);
8. Maga, E. A., **Weimer B.C.**, and Murray J.D., 2012. Lysozyme-rich milk can modulate gut microbiota much like human milk. *Gut Microbes (Enteric Pathogens Special Issue)* 12:4(2). (PMID: 23235404);
9. Maga, E. A., Desai, P., **Weimer B.C.**, Dao, N., Küeltz, D., and Murray J.D.. 2012. Consumption of lysozyme-rich milk can alter microbial fecal populations. *App. Environ. Microbiol*. 78:6153-60);
10. Barboza, M., Froehlich, J.W., Pinzon, J., Moeller, I., Lonnerdal, B., German, J.B., **Weimer B.C.**, and Lebrilla, C.B., 2012. Glycosylation of human milk lactoferrin exhibits dynamic changes during early lactation enhancing its role in pathogenic bacteria-host interactions. *Molecular and Cellular Proteomics*. 11(6):M111.015248. (PMID: 22261723).
11. He, Xuan, H., Mishchuk, D.O., Shah, J., **Weimer B.C.**, Slupsky. C.M., 2013. Cross-talk between two *E. coli* strains and a human colorectal adenocarcinoma-derived cell line. *Scientific Reports* 3:3416-3426. (PMCID: PMC3849634)
12. Makarova, K., Slesarev, A., Wolf, Y., Sorokin, A., Mirkin, B., Koonin, E., ... & Mills, D. (2006). Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences*, *103*(42), 15611-15616.
13. Land, M. L., Hyatt, D., Jun, S. R., Kora, G. H., Hauser, L. J., Lukjancenko, O., & Ussery, D. W. (2014). Quality scores for 32,000 genomes. *Standards in Genomic Sciences*, *9*(1), 20. MacArthur, R. H.; Wilson, E. O. (1967). *The Theory of Island Biogeography*. Princeton, New Jersey: Princeton University Press.
14. Afshinnekoo, E., Meydan, C., Chowdhury, S., Jaroudi, D., Boyer, C., Bernstein, N., ... & Mason, C. E. (2015). Geospatial Resolution of Human and Bacterial Diversity with City-Scale Metagenomics. Cell Systems
15. Ford, D. A., Ruvolo, J., Edlund, S., Myllymaki, J., **Kaufman**, J., Jackson, J., & Gerlach, M. (2001, October). Tempus Fugit: a system for making semantic connections. In Proceedings of the tenth international conference on Information and knowledge management (pp. 520-522). ACM.
16. Slaney, M., Ponceleon, D., **Kaufman**, J. (2001, October). Multimedia edges: finding hierarchy in all dimensions. In Proceedings of the ninth ACM international conference on Multimedia (pp. 29-40). ACM.
17. Myllymaki, J., & **Kaufman**, J. (2003, January). Dynamark: A benchmark for dynamic spatial indexing. In Mobile data management (pp. 92-105). Springer Berlin Heidelberg.
18. Myllymaki, J., & **Kaufman**, J. (2003, May). High-performance spatial indexing for location-based services. In Proceedings of the 12th international conference on World Wide Web (pp. 112-117). ACM.
19. **Kaufman**, J. H., Deen, G., Lehman, T. J., & Thomas, J. (2003). Grid computing made simple. Industrial Physicist, 9(4), 31-33. Reprinted in GRID Today. http://www.aip.org/tip/INPHFA/vol-9/iss-4/p31.html.
20. Lehman, T. J., & **Kaufman**, J. H. (2003, December). OptimalGrid: middleware for automatic deployment of distributed FEM problems on an Internet-based computing grid. In Cluster

Computing, 2003. Proceedings. 2003 IEEE International Conference on (pp. 164-171). IEEE.
21. Lehman, T. J., Deen, R. G., & **Kaufman**, J. H. (2006). An autonomic operating environment for large-scale distributed applications. Integrated Computer-Aided Engineering, 13(1), 81-99.
22. **Kaufman**, J. H., et al. "From regional healthcare information organizations to a national healthcare information infrastructure." Perspectives in health information management/AHIMA, American Health Information Management Association 2 (2005) from Keynote: Digitization of Healthcare Summit held in conjunction with “InnoTech.” Portland, OR, March 9, 2005. Sponsor: Governor, State of Oregon,
23. Centennial article: Bigus, J. P, et al.; , "Information technology for healthcare transformation," IBM Journal of Research and Development , vol.55, no.5, pp.6:1-6:14, Sept.-Oct. 2011. doi: 10.1147/JRD.2011.2160684
24. Eckman B et al. 2006. Varieties of interoperability in the transformation of the healthcare information infrastructure. IBM System Journal 2007, vol. 46, no1, pp. 19-41
25. Smith E, **Kaufman** JH. 2006. Lowering the barrier to a decentralized NHIN using the open healthcare framework. ICMCC (International Council on Medical & Care Compunetics), The Hague, 2006.
26. Renly S, Knoop S, **Kaufman** J, Ram R. 2008. Creating CDA R2 laboratory reports to meet public health surveillance requirements. IHIC2008 (Best case study)
27. Renly S, Knoop S, Ram R, Eshel T, Carmeli B, **Kaufman** JH. 2007. Public Health Affinity Domain (PHAD): A standards-based solution for laboratory-based surveillance systems. Proceedings of the Second NSF Workshop on Biosurveillance. Lecture Notes in Computer Science 4506:147-58. Springer: Berlin/Heidelberg, 2007.
28. Renly S, Knoop S, Ram R, Eshel T, Carmeli B, **Kaufman** JH. 2007. A standards-based solution for public health reporting and surveillance. Proceedings AMIA 2007, Nov 10-4, 2007, Chicago, IL.
29. See: http://www.healthit.gov/policy-researchers-implementers/meaningful-use-regulations
30. **Kaufman** J, Edlund S, Douglas J. 2009. Infectious disease modeling: creating a community to respond to biological threats. Statistical Communications in Infectious Diseases, Vol 1, Issue 1, Article 1. The Berkeley Electronic Press. (Featured Article in First Issue of New Journal)
31. Hu K, Thoens C, Bianco S, Edlund S, Davis M, Douglas J, **Kaufman** JH. "The effect of antibody-dependent enhancement, cross immunity, and vector population on the dynamics of dengue fever." Journal of Theoretical Biology, 319:62–74, 21 Feb 2013, doi:10.1016/j.jtbi.2012.11.021
32. Lessler J, **Kaufman** JH, Ford DA, Douglas JV. 2009. The cost of simplifying air travel when modeling disease spread. PLoS ONE 4(2): e4403. doi: 10.1371/journal.pone.004403.
33. **Kaufman** J, Lessler J, Harry A, Edlund S, Hu K, Douglas J, Thoens C, Appel B, Kasbohrer A, Filter M. "A likelihood-based approach to identifying contaminated food products using sales data: performance and challenges." PLoS Comput Biol 10(11): e 100399. doi: 10.1371/journal.pcbi.1003999.