Pangenome Flux Balance Analysis Toward Panphenomes

Charles J. Norsigian, Xin Fang, Bernhard O. Palsson,
and Jonathan M. Monk

Abstract Studies of the pangenome have been empowered by an exponentially increasing amount of strain-specific genome sequencing data. With this data deluge comes a need for new tools to contextualize, analyze, and interpret such a vast amount of information. Network reconstructions, genome-scale metabolic models (GEMs), and the corresponding computational analysis frameworks such as flux balance analysis (FBA) have been proven useful toward this end. Network reconstructions can be used to interpret genomic variation not just from a single strain but for an entire species. By applying these approaches at the pangenome scale, it becomes possible to systematically evaluate phenotypic properties for an entire species thus enabling the study of a panphenome directly from a pangenome. Applying insights gained from analysis of the panphenome has diverse implications with applications ranging from human health to metabolic engineering. The future of pangenomics will include panphenomic analyses, thus supplementing traditional pangenomic analyses and helping to address the Big-data-to-knowledge grand challenge of analyzing thousands of genomic sequences.

Keywords Flux balance analysis · Genome-scale modeling · Panphenome · Multi-strain · Comparative systems biology

1 Introduction

Studying differences between strains of a species using the construct of a pangenome revolutionized the field of comparative genomics for bacteria (Tettelin et al. 2005; Medini et al. 2005). This framework allowed scientists to overcome issues related to species with high genomic variability and lack of a reference genome. The pangenome alone cannot be used to quantify the phenotypic effects

C. J. Norsigian · X. Fang · B. O. Palsson · J. M. Monk (✉)
Department of Bioengineering, University of California San Diego, La Jolla, CA, USA
e-mail: jmonk@ucsd.edu

of genetic variability. Over the past decade, network reconstructions have become an indispensable tool in molecular systems biology because of their ability to provide a mechanistic link between experimental studies and computational analyses (Bordbar et al. 2014). Thus, genome-scale network reconstructions provide an avenue for extending the power of the pangenome toward evaluating the phenotypic capabilities of a species or the panphenome. High-quality reconstructions can be expanded through bioinformatic techniques to map information from a reference strain to additional strains of the target organism. This chapter describes how reconstructions and genome-scale models have been applied to study the pangenome by predicting all possible phenotypes for strains in a species. Using these tools, large-scale genomic data sets combined with experimental phenotypes can now be integrated and queried to systematically probe the diversity of strains within a species. Genome-scale metabolic network reconstructions can delineate conserved and unique metabolic capabilities across the strains of a species. These differences and designations can be used to define the metabolic potential of a species often informative of lifestyle diversity. In this chapter, we detail the following elements toward true panphenomic analysis: (1) The foundation of reconstructions and flux balance analysis; (2) The extension of these tools using a “multi-strain” approach to calculate metabolic panphenomes for several bacterial species; and (3) A future perspective on the multi-strain approach: moving beyond metabolism for a full calculation of the panphenome.

2 Network Reconstructions and Flux Balance Analysis

The growing collections of sequences that have been used to study pangenomes are laden with valuable information, however, strings of nucleotide bases alone do not make this information easily accessible or immediately apparent. Thus, there is a critical need for tools that can be used to interrogate this massive amount of data to generate new knowledge. Genome-scale network reconstructions in concert with flux balance analysis (FBA) provide such a tool. This section describes the process of reconstruction as well as mathematical approaches that can be used to query and compute with reconstruction, in particular, FBA.

2.1 Network Reconstructions Structure Biological Knowledge

Genome-scale reconstructions are organism-specific knowledge bases. They are built systematically using a quality-controlled bottom-up workflow that incorporates genome annotation, omics data sets, and legacy knowledge. The literature detailing the construction and analysis of network reconstructions is extensive (O’Brien et al. 2015; Thiele and Palsson 2010; Herrgård et al. 2008). In brief, these tools organize knowledge by linking genes, gene products, and cellular components (Fig. 1a).

Fig. 1 (a) Reconstructions consist of layered information connecting annotated genes on the genome sequence to their encoded biological products (e.g., RNA, protein) and how those components interact with other biological components (e.g., protein metabolite, in the case of a metabolic reaction/transformation). Figure reprinted from Reed et al. (2006). (b) Genome-scale models exist for species across the tree of life that are being made for new species and constantly improving. Reprint from Monk et al. (2014). (c) Reconstructions can be converted to a mathematical format by account for use of biological components (e.g., consumption/production). This allows for molecular accounting and enforcement of constraints. (d) Enforcement of constraints (e.g., media updates) and applying an objective (e.g., production of biomass, e.g., growth) allows for simulation of biological phenotypes from the genotype. Panel c and d reproduced from O’Brien et al. (2015). Reprint from O’Brien et al.)

Reconstructions can be made for several cellular processes including transcriptional regulation (Gianchandani et al. 2006, 2009), expression (Thiele et al. 2009) and metabolism (Feist et al. 2009). The reconstruction approach is iterative and thus all reconstructions are continually improving as new knowledge is generated. Thus, reconstructions serve as a valuable resource to integrate and reconcile biochemical data allowing researchers to collaborate, test, and readily share new hypotheses about functions in a target organism (Monk et al. 2014).

Reconstructions of cellular metabolism have been the most developed and extensively used type thus far (Bordbar et al. 2014). Metabolic network reconstructions are composed of all known metabolic genes, their encoded proteins and catalyzed reactions. This information is synthesized by aggregating organism-specific databases, high-throughput data, and primary literature (Thiele and Palsson 2010).

Advancements have allowed for partial automation of this process (Henry et al. 2010; Agren et al. 2013). Reactions are organized into pathways, pathways into subsystems, and ultimately into genome-scale networks; thus, representing biological processes at multiple scales. The resulting network reconstruction is a unification of the information available for an organism with a genetic basis. Today, there exist collections of genome-scale reconstructions for a number of target organisms across the tree of life (Oberhardt et al. 2011; Monk et al. 2014) (Fig. 1b). For example, as of 2018, there are 178 available, curated reconstructions spanning the tree of life (<http://systemsbiology.ucsd.edu/InSilicoOrganisms/OtherOrganisms>). While this coverage is impressive, several other phyla remain devoid of any reconstruction initiative. To fully extend the study of panphenomes to all sequenced organisms, new reconstruction efforts must be initiated (Monk et al. 2014).

2.2 *Flux Balance Analysis Enables Computation of Phenotype from Genotype*

Reconstructions alone are static, and unable to be used for predictions. A major value of the metabolic reconstructions emerges when they are converted into a mathematical format, enabling computational interrogation using a variety of methods (Orth et al. 2010; Lewis et al. 2012). This conversion translates the biochemical reactions of a reconstructed network via tabulation of reaction stoichiometry into a chemically accurate mathematical format that becomes the basis for a genome-scale model (GEM) (Fig. 1c). The flow of metabolites through the network is constrained by these stoichiometries represented as balances or inequalities for bounds (Reed 2012). Further constraints can be added to a network such as thermodynamic reversibility constraints and limitations to nutrient uptake or by-product secretion. Computationally predicted network states consistent with imposed constraints are potential physiological states of the target organism within a defined condition.

Flux balance analysis (FBA) can be applied to these models for prediction of an organism's phenotype. This mathematical approach for analyzing the flow of metabolites through a metabolic network is the original constraints-based method (Orth et al. 2010). This approach relies on an assumption of steady-state growth and mass balance. FBA uses the stated objective (for example, biomass production, e.g., growth) to find the solution(s) using linear programming that optimize an objective function (O'Brien et al. 2015). In a defined environment (defined inputs), GEMs can be used to compute network outputs (Fig. 1d) FBA allows for computational tracing of balanced reaction states beginning with defined inputs to produce output metabolites. Biomass synthesis is computed using FBA by computing the balanced reactions states that produce all the required metabolites for growth simultaneously. Additionally, the model accounts for the energetic, redox, and chemical balances that must also be maintained (O'Brien et al. 2015).

Using this technique, a variety of phenotypes such as the effect of gene knock-outs, metabolite secretion, and growth capabilities on different substrates can be predicted rapidly and compared to experimental results to verify their accuracy (Monk and Palsson 2014). Some of the best models have accuracies >90% in agreement with experimental data (Monk et al. 2017; Brunk et al. 2018). In this way, GEMs provide a way to bridge the genotype to phenotype gap by providing a robust platform for analyzing the integrated mechanisms of gene products to produce unique phenotypic states. The utility of a highly curated GEM and the corresponding computational analyses is increased by the format's scalability. Through this methodology, phenotypes for the plethora of sequenced strains within a species become readily calculable. In the next section, we will highlight how high-quality reconstructions for a single strain can be extrapolated onto several strains of the same species to study the phenotypic potential of the pangenome and to gain insight into strain-specific metabolic capabilities.

3 The Multi-Strain Approach: Extending Genome-Scale Models to Robustly Explore the Pangenome Phenotypic Space

Once a high-quality reconstruction and genome-scale model exist, its contents (e.g., genes, metabolites, and reactions) can be mapped onto other, closely related strains in a species. Following this multi-strain approach, tools from comparative genomics (Monk and Bosi 2018) can be integrated with genome-scale modeling to identify genetic determinants underlying variability of phenotypes. Such a task is crucial to understand the evolutionary trajectories of a bacterial species. Strain-specific metabolic diversity has been illuminated through the use of genome-scale metabolic models. Prediction of unique metabolic capabilities and auxotrophies can be used to study species lifestyle diversity. This approach is scalable to the pangenome level and in turn enables panphenome analysis, thus empowering species-wide comparative systems biology. This multi-strain approach has been applied to several species in a variety of studies and we provide a brief overview of the key insights here.

3.1 *Genesis of the Multi-Strain Approach: Studying Escherichia coli*

The first instance of the multi-strain approach as described here was executed by Monk et al. where the authors leveraged a curated genome-scale model of *E. coli* K-12 MG1655 that has been continually updated over 15 years to construct genome-scale models of 55 other fully sequenced *E. coli* strains (Monk et al. 2013). Using FBA on all 55 of these models, the authors were able to extensively investigate the

Fig. 2 (a) Genome-scale models can be used to predict growth capabilities in different environments and nutritional niches. This figure represents growth predictions for 55 different strain-specific models of *E. coli* and *Shigella* on over 300 different carbon, nitrogen, phosphorus and sulfur sources. Strains, for the most part, clustered according to their isolated niches (e.g., extra versus intestinal). Reproduced from Monk et al. (2013). (b) Using these growth predictions allows for the classification of strains and their potential isolation site (e.g., bladder versus intestine). Decision trees could reliably separate ExPEC from InPEC strains. Left panel reproduced from Croxen and Finlay (2010). Right panel reproduced from Monk et al. (2013)

predicted metabolic capabilities of all the strains (Fig. 2a). The authors delineated strain-specific auxotrophies and substrate preferences among the set of strains. It is important to note that these predictions and insights were gained from sequence alone. Further, this study demonstrated the possibility of applying this approach to understand cases of patho-adaptation to a given environment and evaluate a given strain's infectious niche.

Further work scaled up the effort to include 1200 strains of *E. coli* and demonstrated a large amount of variability within the species both in gene content and consequent variability of gene products (Monk et al. 2017). It also utilized the differences across the 1200 strains to construct a robust classification tree for determination between extra-intestinal and intra-intestinal pathogens using predicted metabolic phenotypes (Fig. 2b). This type of classification schema opens the door to investigating how strain-specific traits impact the microbiome. An in-depth example of such analyses came in a study by Fang et al. into the metabolic capabilities of inflammatory bowel disease (IBD)-associated *E. coli* strains in the B2 clade (Fang et al. 2018). The authors found these strains have advantages in catabolizing sugars derived from mucus glycans. The interesting and novel outcomes of these *E. coli* studies clearly demonstrated the value of the approach, and the natural next step was to apply the methodology to other species.

3.2 Expanding the Reach of Multi-Strain Approach Across the Phylogenetic Tree

Numerous studies followed the first *E. coli* studies that focused on various organisms. Fouts et al. applied the multi-strain approach, broadened to examine various species of *Leptospira* known to have ranging levels of pathogenicity (Fouts et al. 2016). They demonstrated that the ability to synthesize vitamin B12 is limited to pathogenic species of *Leptospira* and may give them a survival advantage in a human host where B12 is sequestered by the body. This valuable distinguishing metabolic capability was captured by being able to leverage the base reconstruction across multiple species in the genus.

In 2016, Bosi et al. applied the workflow to 64 strains of *Staphylococcus aureus*. Beyond reconstructing metabolic capabilities, the approach was extended to identify virulence factors in the set of 64 strains (Bosi et al. 2016). By using a combination of predicted metabolic capabilities linked to virulence factors, they were able to stratify the strains by host type. This study added an additional layer to the promise of the multi-strain approach by showing that metabolic capabilities could be analyzed in concert with other components of the pangenome, namely virulence factors (toxins, adhesins, etc.), and that this combination held predictive power about a strain's host. This study also included explicit calculation of the core- and pangenome content of *S. aureus*, a metric of genomic diversity among strains in a species.

The multi-strain approach has also been applied to other pathogens such as *Acinetobacter baumannii* and *Salmonella*. In a study by Norsigian et al., a highly curated base GEM was used to create models for 75 different *A. baumannii* strains (Norsigian et al. 2018). These strain-specific models demonstrated major differences in metabolism between strains indicating that a classification scheme may be possible from sequence alone. Seif et al. built strain-specific models for 450 *Salmonella* strains from various serovars to show that metabolic capabilities can be used to

distinguish these serovars (Seif et al. 2018). This study indicates that the host-range may be limited by metabolic capabilities of different strains.

3.3 *Extending the Multi-Strain Approach to Investigate Additional Biological Qualities*

The multi-strain framework provides an inherently efficient means of interrogating the properties of many strains and a few studies have utilized this organizational efficiency to gain insight into properties outside of direct metabolic capabilities. For example, Choudhary et al. examined the agr type of 400 *S. aureus* strains to examine the structure of genes within the genome (Choudhary et al. 2018). The authors found that genomic virulence factor profiles are highly correlated with agr type. They also identified that divergence in histidine kinase protein confers signal specificity with clear differences in protein structural properties based on agr types. Another example of additional properties is the investigation of reactive oxygen species (ROS) tolerance. By leveraging the multi-strain approach in conjunction with 3D structures Mih et al. was able to simulate ROS production levels to demonstrate that antioxidant properties are exhibited in the structural proteome (Mih et al. 2018). A third example was conducted by Kavvas et al., who took a deeper level of resolution within the genome by looking at the unique alleles present within *Mycobacterium tuberculosis* genomes (Kavvas et al. 2018). Through machine learning techniques on the pangenome they were able to associate certain alleles potentially responsible for antimicrobial resistance. The results hint at metabolic rewiring at the allelic level required for adaptation to antibiotic resistance. The success of the multi-strain approach in all these various studies suggests that explicit calculation of the panphenome will provide novel insights.

4 Future Perspectives: Moving Beyond Metabolism: A Multi-Scale Approach to Calculating Full Panphenomes

This chapter details a computational approach (network reconstruction and FBA) to systematically calculate metabolic phenotypes for multiple strains in a species. Beyond calculation of metabolic phenotypes, new methods, both experimental and computational, offer exciting new avenues for research into the pangenome. These approaches can be applied at multiple different scales. At the lowest level, single nucleotide variants (SNV) can be compared across strains using sequence mapping toolkits like breseq and gatk (Deatherage and Barrick 2014; McKenna et al. 2010). These approaches can be scaled up from single base changes to full gene sequences to compare orthologous ORFs across genomes by comparing sequence-specific alleles across strains in a species (Fig. 3a). As described here, the presence/absence of given

Fig. 3 (a) Detailed view of amino acid polymorphisms (allele frequency) for this D gene among 1200 diverse *E. coli* strains. Phylogenetic tree illustrating the relatedness between the unique alleles. (b) Comparison of orthologous gene expression between three different strains of *E. coli* (K-12 W3110, K-12 MG1655

enzyme-encoding metabolic genes can be used to build strain-specific metabolic reconstructions that compute metabolic panphenomes. While most of the applications described here are applied to pathogens with relevance to human health, it is important to note that the pangenome can also be studied for use in metabolic engineering applications. For example, the pangenome can be mined to search for enzymes of interest to industrial microbiology (Moscatello and Pfeifer 2018).

In the future, processes beyond metabolism will also be reconstructed allowing for true panphenome calculations. For example, reconstructions of protein expression mechanisms already exist (Thiele et al. 2009) and have been integrated with models of metabolism (ME models) (O'Brien et al. 2013). These models account for the transcription and translation processes and molecular constituents required to express enzymes catalyzing metabolic reactions in the metabolic network. It is further possible to use the ME model framework to reconstruct proteostatic mechanisms and investigate the structural integrity of the proteome (Chen et al. 2017). In the future, multiple ME models of strains in a species will further expand the scope of computation possible on contents of the pangenome.

Beyond metabolism and expression, regulatory networks are another aspect of the pangenome that differ between strains and have been reconstructed for individual strains (Gianchandani et al. 2006, 2009). Understanding how certain strains regulate the same set of genes (core-genome), as well as diverse sets of genes, will further expand our understanding of the structure and function of the pangenome. A small-scale study of seven *E. coli* strains and their RNA-seq expression profiles in aerobic and anaerobic environments showed remarkably different expression levels even for shared genes of the core-genome (Monk et al. 2016) (Fig. 3b). Studying differentially expressed genes and the transcription factors known to regulate them may lead to the discovery of alternative regulatory strategies between strains of a species.

Just as sequence databases have grown tremendously in recent years, 3D crystal structures for the encoded genes have also grown dramatically (Brunk et al. 2016). The protein data bank (Berman et al. 2000) (PDB) is a repository of protein structures and these structures can now be integrated with genome-scale models (GEM-PRO) (Chang et al. 2013). Building multi-strain models with associated protein structures is another way to compare strains across a species. Using these tools, sequence diversity can be examined at the 3D level to see how mutations line up in 3D space, a level of analysis not possible at the sequence level. Furthermore, mutations in specific regions of the protein can be tabulated (Fig. 3c) and compared across strains (Mih et al. 2018).

Finally, a multi-strain approach should prove useful for studies of the microbiome. Multiple genome-scale models for species found in the microbiome

Fig. 3 (continued) and BL21). Overall the K-12 strains have a much higher correlation between their transcriptional profiles than did BL21. Reproduced from Monk et al. (2016) (c) Expanding analysis of sequence similarity by incorporating 3D structural information. The inclusion of structures mapped to sequences allows the visualization of how differences in sequences manifest in 3D space. (d) Expanding study of strains to the microbiome using metagenomics and strain-level resolution. Panels a, c, and d reproduced from Monk et al. (2017)

already exist (Magnúsdóttir et al. 2017), and GEM studies were proven effective in studying the impact of diet (Shoaie et al. 2015) and interactions between microbes (Shoaie et al. 2013). Expanding the multi-strain approach to study diverse strains in these species may lead to a deeper level understanding of the gut microbiome composition. Indeed, strain-level metagenomics is coming (Scholz et al. 2016) and expanding the study of the pangenome to the microbiome will have fruitful applications in the near future (Fig. 3d).

In closing, we must list some caveats and risks to the multi-strain approach. First, all of these approaches require high-quality sequence data connected to high quality, QC/QA data generation. The success of reliable and maximally effective future panphenomics rests on ensuring this quality. There must be a continued effort to ensure that sequencing projects are of quality not only quantity. Additionally, an interesting question pertaining to the concept of closed pangenomes is, how will the law of diminishing returns be exhibited in these sequence deposits? Will a point be reached where additional sequences provide no novel information? Further, the vision of the panphenome and its implications to understanding how microbial pathogens impact human health will rely on both the availability of metadata and the deposition of strains. Metadata on these strains will only deepen the possible questions to be asked of both pangenomes and panphenomes. A centralized repository of strains will also greatly expedite the experimental verification needed for such large computational predictions. The future of the panphenome is apparent and with it further explanations at the center of biological causality.

5 Conclusions

Significant advancements in DNA sequencing technology have led to an exponential increase in the number of sequenced strains. This creates a need for new ways to integrate and analyze this ever-increasing amount of sequence information. This need will only intensify as the number of sequenced strains within a species continues to grow exponentially. This chapter demonstrates how the pangenome is evolving from a theoretical concept to a queryable construct.

In this chapter, we describe how the foundational aspects of GEMs and FBA can be used to predict phenotypic states for multiple strains in a species. The multi-strain approach has proven useful in extending this utility in a number of studies providing evolutionary insights as well as practical applications. As the library of available sequences continues to grow, the possibility of scaling these techniques to the level of the pangenome is becoming a reality. The result, a species-wide panphenome, would create a deeper level of understanding than the collection of gene content within the pangenome alone.

The ability to systematically characterize an entire species' phenotypic capabilities will enhance the depth of pangenome analysis possible and pull valuable information inherent to genome sequences to the forefront (Fig. 4). The linkages and distinct features at the pangenome scale for a species offer obvious value for future knowledge generation, especially pertaining to human health and disease.

Fig. 4 The established assembly of the pangenome through the use of genome-scale reconstructions and corresponding computational analyses enables the calculation of panphenomes. The panphenome increases the depth of analysis possible by providing a framework in which to delineate strain-specific phenotypes. This stratification based on sequence similarity allows for the determination of which pieces of reconstructed networks are shared among various groups of strains in a species. This will continue to further inform the generation of evolutionary hypotheses

Further, the future potential applications outlined here such as inclusion of expression, regulation, and structures into these workflows will only further advance the scope of genome-scale science. Genome sequences are laden with critical information and the tools/workflows described in this chapter provide a means for extracting this information into actionable knowledge.

References

- Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J (2013) The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput Biol* 9(3):e1002980
- Berman HM, Westbrook J, Feng Z (2000) The protein data bank. *Nucleic Acids Res* 28 (1):235–242. <https://academic.oup.com/nar/article/doi/10.1093/nar/28.1.235/2384399/The-Protein-Data-Bank>
- Bordbar A, Monk JM, King ZA, Palsson BO (2014) Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet* 15(2):107–120
- Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BØ (2016) Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc Natl Acad Sci USA* 113(26):E3801–E3809
- Brunk E, Mih N, Monk J, Zhang Z, O'Brien EJ, Bliven SE, Chen K, Chang RL, Bourne PE, Palsson BO (2016) Systems biology of the structural proteome. *BMC Syst Biol* 10(March):26
- Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F et al (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* 36 (3):272–281
- Chang RL, Andrews K, Kim D, Li Z, Godzik A, Palsson BO (2013) Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. *Science* 340(6137):1220–1223. <https://doi.org/10.1126/science.1234012>. PMID: 23744946
- Chen K et al (2017) Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proc Natl Acad Sci* 114(43):11548–11553

- Choudhary KS, Mih N, Monk J, Kavvas E, Yurkovich JT, Sakoulas G, Palsson BO (2018) The *Staphylococcus aureus* two-component system AgrAC displays four distinct genomic arrangements that delineate genomic virulence factor signatures. *Front Microbiol* 9:1082
- Croxen MA, Finlay BB (2010) Molecular mechanisms of *Escherichia coli* pathogenicity. *Nat Rev Microbiol* 8(1):26–38
- Deatherage DE, Barrick JE (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using Breseq. *Methods Mol Biol* 1151:165–188
- Fang X, Monk JM, Mih N, Du B, Sastry AV, Kavvas E, Seif Y, Smarr L, Palsson BO (2018) *Escherichia coli* B2 strains prevalent in inflammatory bowel disease patients have distinct metabolic capabilities that enable colonization of intestinal mucosa. *BMC Syst Biol* 12(1):66
- Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7(2):129–143
- Fouts DE, Matthias MA, Adhikarla H, Adler B, Amorim-Santos L, Berg DE, Bulach D et al (2016) What makes a bacterial species pathogenic? Comparative genomic analysis of the genus *Leptospira*. *PLoS Negl Trop Dis* 10(2):e0004403
- Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO (2006) Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol* 2(8):e101
- Gianchandani EP, Joyce AR, Palsson BØ, Papin JA (2009) Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system. *PLoS Comput Biol* 5(6):e1000403
- Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol* 28(9):977–982
- Herrgård MJ, Thiele I, Reed JL, Palsson BØ (2008) Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7:129–143. <http://www.nature.com/nrmicro/journal/v7/n2/abs/nrmicro1949.html>
- Kavvas ES, Catoiu E, Mih N, Yurkovich JT, Seif Y, Dillon N, Heckmann D et al (2018) Machine learning and structural analysis of *Mycobacterium tuberculosis* pan-genome identifies genetic signatures of antibiotic resistance. *Nat Commun* 9(1):4306
- Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* 10(4):291–305
- Magnúsdóttir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, Greenhalgh K et al (2017) Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol* 35(1):81–89
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K et al (2010) The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20(9):1297–1303
- Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. *Curr Opin Genet Dev* 15(6):589–594
- Mih N, Brunk E, Chen K, Catoiu E, Sastry A, Kavvas E, Monk JM, Zhang Z, Palsson BO (2018) ssbio: a python framework for structural systems biology. *Bioinformatics* 34(12):2155–2157
- Monk J, Bosi E (2018) Integration of comparative genomics with genome-scale metabolic modeling to investigate strain-specific phenotypical differences. In: Fondi M (ed) *Metabolic network reconstruction and modeling: methods and protocols*. Springer, New York, pp 151–175
- Monk J, Palsson BO (2014) Genetics. Predicting microbial growth. *Science* 344(6191):1448–1449
- Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ (2013) Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci USA* 110(50):20338–20343
- Monk J, Nogales J, Palsson BO (2014) Optimizing genome-scale network reconstructions. *Nat Biotechnol* 32(5):447–452
- Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, Herrgård MJ, Feist AM (2016) Multi-Omics quantification of species variation of *Escherichia coli* links molecular features with strain phenotypes. *Cell Syst* 3(3):238–251. e12
- Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, Takeuchi R et al (2017) iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol* 35(10):904–908

- Moscattello N, Pfeifer BA (2018) Constraint-based metabolic targets for the improved production of heterologous compounds across molecular classification. *AIChE J Am Inst Chem Eng* 9 (July):293
- Norsigian CJ, Kavvas E, Seif Y, Palsson BO, Monk JM (2018) iCN718, an updated and improved genome-scale metabolic network reconstruction of *Acinetobacter baumannii* AYE. *Front Genet* 9(April):121
- O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 9(1):693
- O'Brien EJ, Monk JM, Palsson BO (2015) Using genome-scale models to predict biological capabilities. *Cell* 161(5):971–987
- Oberhardt MA, Puchalka J, Martins dos Santos VAP, Papin JA (2011) Reconciliation of genome-scale metabolic reconstructions for comparative systems analysis. *PLoS Comput Biol* 7(3): e1001116
- Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nat Biotechnol* 28(3):245–248
- Reed JL (2012) Shrinking the metabolic solution space using experimental datasets. *PLoS Comput Biol* 8(8):e1002662
- Reed JL, Famili I, Thiele I, Palsson BO (2006) Towards multidimensional genome annotation. *Nat Rev Genet* 7(2):130–141
- Scholz M, Ward DV, Pasolli E, Tolio T, Zolfo M, Asnicar F, Truong DT, Tett A, Morrow AL, Segata N (2016) Strain-level microbial epidemiology and population genomics from shotgun metagenomics. *Nat Methods* 13(5):435–438
- Seif Y, Kavvas E, Lachance J-C, Yurkovich JT, Nuccio S-P, Fang X, Catoiu E, Raffatellu M, Palsson BO, Monk JM (2018) Genome-scale metabolic reconstructions of multiple salmonella strains reveal Serovar-specific metabolic traits. *Nat Commun* 9(1):3771
- Shoaie S, Karlsson F, Mardinoglu A, Nookaew I, Bordel S, Nielsen J (2013) Understanding the interactions between bacteria in the human gut through metabolic modeling. *Sci Rep* 3:2532
- Shoaie S, Ghaffari P, Kovatcheva-Datchary P, Mardinoglu A, Sen P, Pujos-Guillot E, de Wouters T et al (2015) Quantifying diet-induced metabolic changes of the human gut microbiome. *Cell Metab* 22(2):320–331
- Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV et al (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. *Proc Natl Acad Sci USA* 102(39):13950–13955
- Thiele I, Palsson BØ (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 5(1):93–121
- Thiele I, Jamshidi N, Fleming RMT, Palsson BØ (2009) Genome-scale reconstruction of *Escherichia coli*'s transcriptional and translational machinery: a knowledge base, its mathematical formulation, and its functional characterization. *PLoS Comput Biol* 5(3):e1000312

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.