

## ARTICLE OPEN

Corrected: Author Correction

# Mechanistic insights into bacterial metabolic reprogramming from omics-integrated genome-scale models

Noushin Hadadi<sup>1\*</sup>, Vikash Pandey<sup>2</sup>, Anush Chiappino-Pepe<sup>2</sup>, Marian Morales<sup>1</sup>, Hector Gallart-Ayala<sup>3</sup>, Florence Mehl<sup>3</sup>, Julijana Ivanisevic<sup>3</sup>, Vladimir Sentchilo<sup>1</sup> and Jan R. van der Meer<sup>1</sup>

Understanding the adaptive responses of individual bacterial strains is crucial for microbiome engineering approaches that introduce new functionalities into complex microbiomes, such as xenobiotic compound metabolism for soil bioremediation. Adaptation requires metabolic reprogramming of the cell, which can be captured by multi-omics, but this data remains formidably challenging to interpret and predict. Here we present a new approach that combines genome-scale metabolic modeling with transcriptomics and exometabolomics, both of which are common tools for studying dynamic population behavior. As a realistic demonstration, we developed a genome-scale model of *Pseudomonas veronii* 1YdBTEX2, a candidate bioaugmentation agent for accelerated metabolism of mono-aromatic compounds in soil microbiomes, while simultaneously collecting experimental data of *P. veronii* metabolism during growth phase transitions. Predictions of the *P. veronii* growth rates and specific metabolic processes from the integrated model closely matched experimental observations. We conclude that integrative and network-based analysis can help build predictive models that accurately capture bacterial adaptation responses. Further development and testing of such models may considerably improve the successful establishment of bacterial inoculants in more complex systems.

npj Systems Biology and Applications (2020)6:1; <https://doi.org/10.1038/s41540-019-0121-4>

## INTRODUCTION

Microbiome engineering is an upcoming discipline that aims to manipulate, complement or restore the functionality of existing damaged communities, e.g., contaminated soils, by adding specific new metabolic capabilities.<sup>1</sup> A rational engineering approach requires a detailed understanding of general principles of the functioning of the microbial community and its physiological adaptations to perturbations, but such understanding is currently lacking and fragmentary.<sup>1,2</sup> The technically most simple way to provide new metabolic capacities to an existing microbial community is by strain addition (what one could call an  $N + 1$  scenario).<sup>3,4</sup> After an initial screening of the existing capacity of the microbial community, one or more preselected and well-characterized strains with the intended complementation could be prepared, formulated and inoculated into the community. Depending on the aims, such inoculants should maintain and reproduce for longer-term inside the resident community or only deploy their metabolic capacity transiently.<sup>3,4</sup>

Inoculation of preselected strains has been widely practised for pollutant bioaugmentation, using bacteria with particular metabolic capabilities that enable them to efficiently degrade and grow on common pollutants such as toxic aromatic compounds.<sup>5</sup> However, even the simplest inoculations and  $N + 1$ -strategies are rarely effective because it is insufficiently understood what inoculants need to establish successfully within a (new) existing community, and how they need to adjust their physiology to meet the requirements of the new environment and degrade the desired toxic compound(s). Modeling strategies based on the integration of a variety of (nowadays more easily) accessible condition-specific omics data, would help to better understand and predict how cellular regulation and physiology at different growth conditions and environments interplay. However, the impact and advantage of such integrative analysis are not yet

explored to its full extent.<sup>6</sup> We propose and demonstrate here that combining comprehensive genome-wide transcriptomics, exometabolomics and metabolic modeling can better predict physiological adaptation.

Metabolic modeling has largely advanced through the development of GENome-scale Metabolic models (GEMs) and constraint-based modeling techniques such as Flux Balance Analysis (FBA). GEMs can be built from the annotated genomes and they describe an organism's metabolism as completely as possible, linking genotype to metabolic phenotypes.<sup>7</sup> GEMs encompass metabolites, metabolic reactions, and genes coding for the enzymes catalyzing the reactions. Together with FBA, GEMs predict steady-state fluxes,<sup>8,9</sup> and therefore, they can predict cellular physiology. While the genome specifies the complete set of biochemical reactions which the cell can potentially carry out, the actual enzymatic capacity at each physiological condition is orchestrated by regulatory networks in the cell. GEMs do not explicitly consider regulation, whose effects are better reflected in the global transcriptome and the metabolome.<sup>10–14</sup> FBA approaches have been extended with RNAseq and metabolomics data to capture cell regulation and more accurately describe cellular metabolic behavior.<sup>15</sup> For example, transcriptional regulation of gene expression has been linked to GEMs, either by taking into account the absolute expression values, scoring genes and subsequently reaction fluxes as active or non-active based on their expression,<sup>16–18</sup> or by incorporating relative gene-expression.<sup>14,18</sup> Use of relative gene expression is assuming that the relative changes between two conditions correlate with the resulting differential flux profiles. Both approaches can lead to condition-specific GEMs that are more effective for inferring the actual biochemical activity and the observed physiology of the microorganism.

<sup>1</sup>Department of Fundamental Microbiology, University of Lausanne, 1015 Lausanne, Switzerland. <sup>2</sup>Laboratory of Computational Systems Biotechnology, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. <sup>3</sup>Metabolomics Platform, University of Lausanne, 1015 Lausanne, Switzerland. \*email: Noushin.hadadi@unige.ch

As a study system for predicting physiology from an integrated GEM-transcriptome-metabolome approach, we here use *Pseudomonas veronii* 1YdBTEX2. Strain 1YdBTEX2 is capable of degrading a variety of mono-aromatic hydrocarbons such as benzene, toluene, ethylbenzene, and *m*- and *p*-xylene (BTEX).<sup>20–22</sup> The ability of *P. veronii* 1YdBTEX2 to grow in contaminated environments makes it a promising candidate for rational complementation of microbial communities in contaminated soils.<sup>23</sup> Based on an available manually curated high-quality genome,<sup>22</sup> we reconstructed the first GEM for *P. veronii* 1YdBTEX2 (iPsvr). Genome-wide transcription changes and exometabolome compounds were measured during growth of *P. veronii* on toluene, in exponential and in stationary phase. Transcriptome and exometabolome data were integrated into the iPsvr using the recently developed tool *REMI* (Relative Expression and Metabolomics Integrations).<sup>19</sup> *REMI* integrates gene expression, thermodynamic, and metabolite abundance data into metabolic models, and then maximizes the consistency between differential gene expressions, metabolite concentrations and levels of metabolic fluxes for any given pair of conditions. Therefore, by building a context-specific model for each condition, it allows elucidating metabolic states and predicting cellular phenotypes and flux distributions of condition-specific differences e.g., wild type vs mutation. Two obtained metabolic models representing exponential and stationary physiologies were then used to evaluate growth rates and the production of biomass precursors, and model predictions were compared to the experimentally observed values. Although the temporal variations of the growth rate cannot be predicted using GEMs,<sup>24</sup> we showed that introducing the additional regulatory information from gene expression and metabolomics data into GEMs allows for consistently estimating growth rates at different growth phases. Finally, we incorporated into iPsvr previously published transcriptomics data of *P. veronii* transits from liquid culture to sand<sup>22</sup> to understand its physiological adaptation in soil. Our work shows strong consistency of model outputs with the experimental data, manifesting that integration of condition-specific omics data into a curated GEM constitutes a major improvement for prediction of metabolic reprogramming during adaptation.

## RESULTS

### Developing an integrated genomic-transcriptomic-metabolomic workflow

To develop a pipeline that integrates genomics with transcriptomic and metabolomic data we advanced in three stages: (1) Quantify the cellular states at each phenotype, by genome-wide transcriptomics and exometabolomic data from spent media composition (Fig. 1a); (2) Construct a GEM for *P. veronii* strain 1YdBTEX2 (iPsvr), gap-fill missing parts of the metabolism (compounds and reactions), complement genome annotation using the transcriptomics and exometabolomic data and estimate the steady-state growth rate using FBA (Fig. 1b); and (3) Link the interrelationships between growth phases and the differentially expressed genes and metabolite abundances by statistical inference and by *REMI*.

In this work, we used the pipeline to: (i) generate two growth-phase-specific models of *P. veronii* growing on toluene, iPsvr-EXPO and iPsvr-STAT, which correspond to the exponential and the stationary phase, respectively, and employ them to predict quantitative and dynamic readouts of *P. veronii* metabolism in the two conditions (Fig. 1c); and (ii) build two context specific models, iPsvr-LIQ and iPsvr\_SAND, for the growth on liquid medium and in a sand environment, respectively, by integrating previously published gene expression data<sup>22</sup> into iPsvr. The models were then used to predict the physiology in a liquid-to-sand transition.

### Genome-wide gene expression and metabolite formation over time

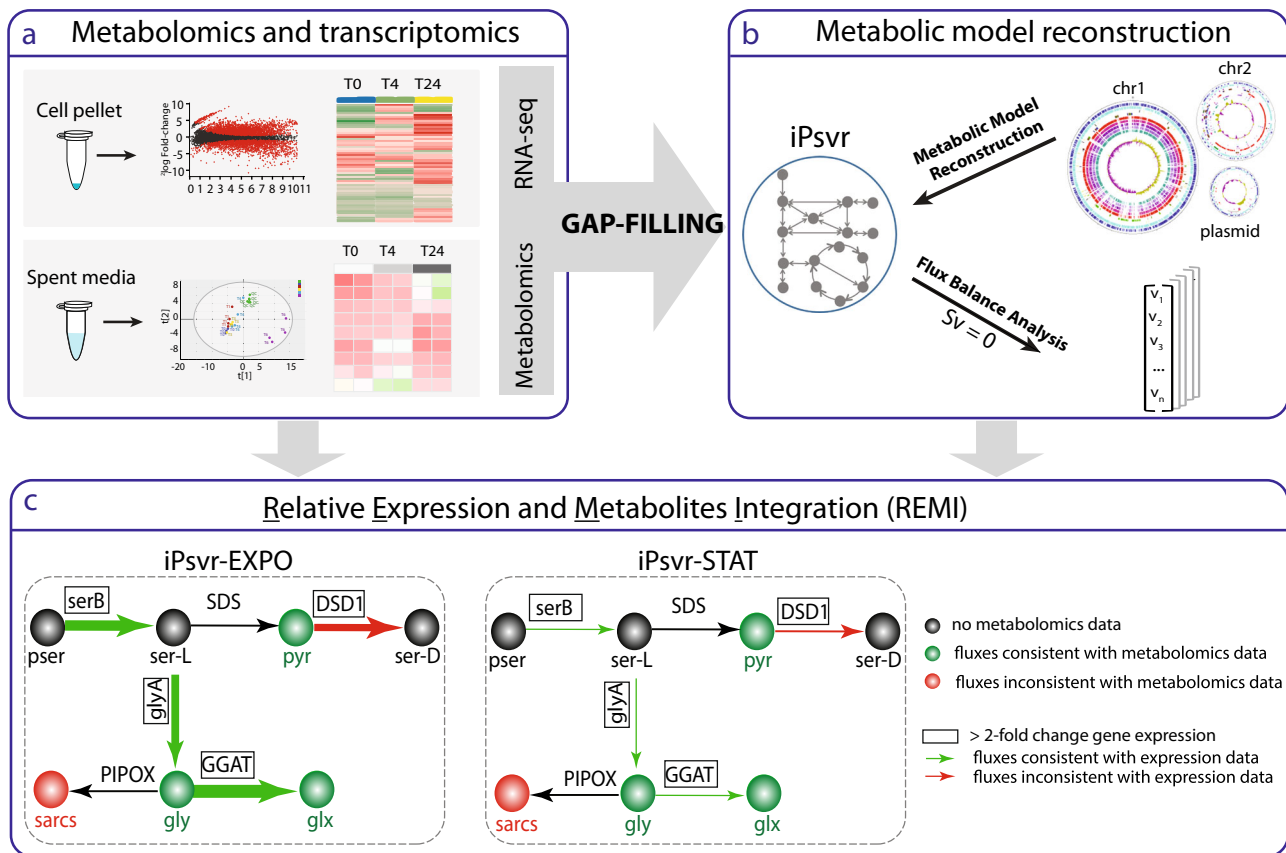
Whole-genome gene expression profiles and metabolite formation in the spent medium were analyzed in *P. veronii* cultures growing in liquid minimal medium with toluene as sole carbon and energy source, sampled at 0 h (T0h), 4 h (T4h, EXPO) and 24 h (T24 h, STAT) after inoculation.

Based on growth kinetics of *P. veronii* 1YdBTEX2 on toluene, we identified 4h and 24 h to be representative time points for *P. veronii* 1YdBTEX2 metabolic activity at the early exponential growth and late stationary phase, respectively. Genome-wide gene expression was quantified by mapping Illumina 100 nucleotide long single-end sequencing reads from deeply sequenced cDNA libraries to the protein coding genes in *P. veronii* genome (read numbers indicated in Table S1). For each sampling time point, four replicates clustered closely together, with slightly higher variability observed among the T24h replicates (Fig. S1A). A pair-wise comparison of expression levels showed that 1458 (818 up-regulated and 640 down-regulated) out of the total 6943 genes (21%) were significantly differentially expressed between EXPO and STAT phase cells, with at least 2-log fold-change induction (false discovery rate [FDR] < 0.05) (Fig. 2a, b).

*A priori*, based on the genome annotation, a subset of 1241 “metabolic genes” were used in the GEM reconstruction (iPsvr). Out of these 1241, 300 (21%) were significantly differentially expressed in EXPO vs. STAT phase cells (FDR < 0.05) (Table S2). The transition to STAT phase in bacteria is characterized by growth arrest in response to several factors, such as carbon and nutrient depletion, the accumulation of toxic compounds and environmental stress, which decrease ribosomal activity and therefore protein synthesis. This is consistent with toluene becoming depleted in STAT phase. As anticipated, enriched GO terms for the category “Biological Process” among the differentially expressed genes between EXPO and STAT included “protein folding” (GO:0006457), “tRNA aminoacylation for protein translation” (GO:0006418), “intracellular protein transmembrane transport” (GO:0065002) and “regulation of transcription, DNA-templated” (GO:0006355) (Table S3), thus indicating cells to be more active in EXPO phase, as expected. Consistent with nutrients becoming depleted in STAT phase, the terms “benzoate catabolic process via hydroxylation” (GO:0043640) and “tricarboxylic acid cycle” (GO:0006099) (Table S3), important for aromatic compound catabolism, were under-represented in the STAT phase transcriptome.

The untargeted metabolomic analyses of the spent medium detected 1630 (positively charged) and 3509 (negatively charged) distinct ion species or metabolite features. Unsupervised principal component analysis yielded three distinct clusters indicating metabolic phenotype differentiation over time, from inoculum to stationary phase (Fig. S1B). Similar to the transcriptomics data, a greater variability was observed among the T24h replicates.

Temporal patterns of annotated metabolites by HMDB database (accurate mass) matching<sup>25</sup> showed a significant increase in the spent media over time of the majority of the metabolites implicated in the toluene and benzene degradation pathways and central carbon pathways, including glycolysis, purine and pyrimidine metabolism and amino acid metabolism (Fig. 2c). This implies their production by the bacteria and progressive release into the media. One specific group of metabolites, including citrate (C00158), glutamate (C00025), glutamine (C00064), and aspartate (C00049), accumulated significantly in EXPO phase (T4h) in comparison to T0h, followed by lower levels in STAT phase (Fig. 2c). This suggests their excretion in EXPO phase and subsequent re-consumption when other nutrients became limiting (Fig. 2c).



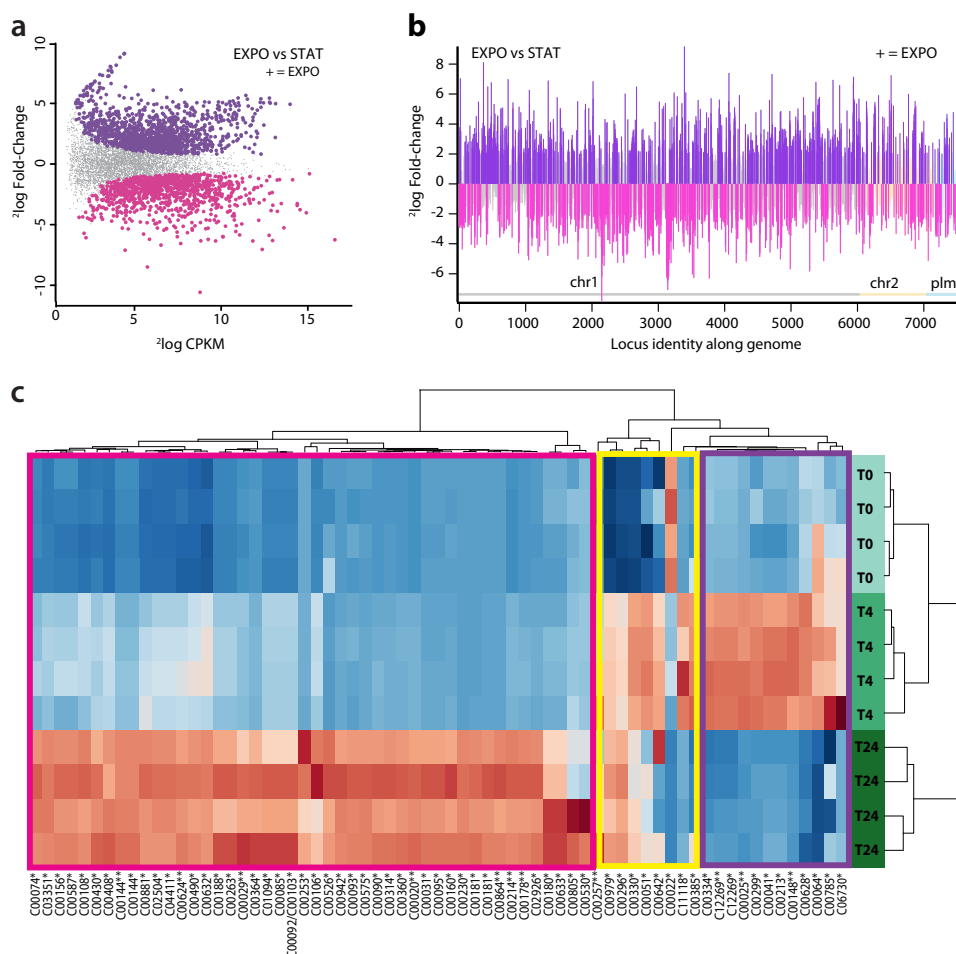
**Fig. 1 Schematic overview of the integrated genomic-transcriptomic-metabolomic pipeline applied in this study.** **a** Stage 1: Relative gene expression and exometabolomic data were determined and analyzed, and these data were used to gap-fill iPsvr at blocked reactions. **b** Stage 2: A genome-scale metabolic model (GEM) of *P. veronii* strain 1YdBTEX2, iPsvr, was reconstructed and flux-balance analysis (FBA) was performed to simulate the growth of the cell. **c** Stage 3: Additional relative differential gene expression and metabolite abundance data were integrated into the metabolic model with REMI and physiology-specific models were built. Here, the REMI methodology is illustrated on a section of inferred iPsvr glycine, serine and threonine metabolism and iPsvr-EXPO (exponential phase) and iPsvr-STAT (stationary phase) as the two physiology-specific models. Significantly differentially expressed genes (here, *serB*, *glyA*, *DSD1* and *GGAT*) are outlined in boxes. The thickness of arrows designates the fold-change in estimated fluxes, where green arrows indicate consistency with the gene-expression fold-change values, and the red ones inconsistencies. Measured metabolite concentrations (here: pyr, sarcs, gly and glx) are indicated in green if the values are consistent with estimated fluxes and in red otherwise. Phosphoserine phosphatase, SerB; Serine hydroxymethyltransferase, GlyA; D-serine dehydratase, DSD1; Glyoxylate aminotransferase, GGAT; Pyruvate, pyr; Sarcosine, sarcs; Glycine, gly; Glyoxylate, glx.

Genome-scale metabolic model (GEM) of *P. veronii* strain 1YdBTEX2 (iPsvr)

A draft GEM was generated from the curated *P. veronii* genome<sup>22</sup> by using the RAVEN toolbox<sup>26</sup> (Fig. 1b). The draft GEM was gap-filled by iterative manual curation until we obtained a model able to carry non-zero flux through the biomass reaction at steady state. This signified cell ‘growth’, and indicated that the model was performing as expected for a biological system. The cell biomass composition and compartment information were derived from two available models of other Pseudomonads species: *P. putida*<sup>27–29</sup> and *P. stutzeri*<sup>30</sup> (see Methods).

We used Biolog growth data of *P. veronii* 1YdBTEX2 incubated with 26 different carbon substrates to further curate and refine the model. In comparison to a no-carbon control, Biolog indicated growth of strain 1YdBTEX2 on 15 out of the tested 26 substrates as the sole carbon source (Table S4). We confirmed the experimental observations using iPsvr for 20 substrates, whereas the model could not grow on the remaining six substrates (D-Xylose, D-Galacturonic Acid, L-Phenylalanine, D-Glucosaminic Acid, Itaconic Acid, and Putrescine). We hence gap-filled the model to include these additional metabolic tasks (Table S4). This gap-filling improved the metabolic network connectivity, and it resulted in a more versatile model that can be used for studies on an extended range of substrates.

The scope of iPsvr GEM was further widened by restoring the connectivity of the remaining ‘blocked’ reactions, i.e., isolated reactions that carry zero flux at any condition. To this end, we explicitly considered the empirical gene-expression and exometabolomics data. We first used a graph-based algorithm (see Methods) to decompose the iPsvr metabolic network into its main subnetworks of 1370 reactions and 23 blocked reactions/pathways of different lengths, with the longest blocked pathway consisting of seven reaction steps (Table S5). Out of 191 blocked reactions/pathways, we identified those associated with differentially expressed genes between the two growth conditions and the ones whose participating metabolites were present in the exometabolomic data. The identified reactions/pathways were next unblocked by gap-filling as described in the Methods section. Interestingly, we identified gap-filling reactions that had been annotated to *P. veronii* genes with RAVEN but had a lower score than the ones chosen as a baseline for the draft reconstruction of iPsvr. The gap-filling algorithm introduced in total 71 new metabolic reactions together with their corresponding 28 genes to iPsvr (Table 1 and Table S6). The reconstructed and gap-filled iPsvr accounted for 1243 genes, 1610 metabolic reactions and 1681 metabolites localized within two intracellular compartments, the cytosol and periplasm, and the extracellular environment (Table 1).



**Fig. 2** Genome-wide gene-expression and metabolite abundance differences in *P. veronii* 1YdBTEX in exponential (EXPO) relative to stationary (STAT) phase. **a** Smear-plot of global gene expression intensity ( $2\log$  counts per kilobase per million, CPKM) versus expression changes ( $2\log$  fold change) in EXPO vs. STAT. In gray, genes not statistically differentially expressed ( $\log_{FC} < 1$ ,  $FDR > 0.05$ ,  $P > 0.01$ ); magenta, genes with lower, and dark purple, genes with higher expression in EXPO (-). **b** Differential expression per gene between EXPO and STAT plotted as a function of genomic location (chromosome 1, chr1; chromosome 2, chr2; and plasmid, plm; organized according to locus\_tag number). Bars indicate  $2\log$ -fold change, dark purple and pink denote statistically significant higher and lower expressed genes in EXPO, respectively. **c** Heatmap of 164 metabolites annotated with their KEGG IDs showing clustering at different culture sampling time points. Metabolites are clustered (x-axis) in three major groups (left to right): (i) accumulation at T24h (STAT) (pink box), (ii) accumulation at T4 continued to T24 (yellow box), and (iii) accumulation at T4h (EXPO) followed at depletion at T24h (STAT) (purple box).

The experimentally determined maximal specific growth rate of *P. veronii* strain 1YdBTEX2 in minimal medium with toluene as the sole carbon source was in the range of  $0.25\text{ h}^{-1}$  to  $0.35\text{ h}^{-1}$ . In absence of proper empirical measurements of toluene uptake rate, we estimated a maximal uptake rate of toluene from a constrained maximal specific growth rate to  $0.35\text{ h}^{-1}$  in the thermodynamics-based flux balance analysis (TFA)<sup>8,31,32</sup> We computed that a toluene uptake rate of  $0.45\text{ mmol gDW}^{-1}\text{ h}^{-1}$  is required to achieve this growth rate.

Omics-based curation and gap-filling in iPsvr for toluene degradation and phenylalanine metabolism

Given the importance of toluene degradation by *P. veronii* 1YdBTEX2, we manually curated the predicted toluene (Fig. 3a) and phenylalanine metabolic pathways (Fig. 3b) to ensure they were fully functional in iPsvr. In *P. veronii*, toluene is converted via (1S, 2R)-3-methylcyclohexa-3,5-diene-1,2-diol to 3-methylcatechol, which is further degraded, according to the KEGG pathway database, through two pathways until the central carbon metabolism compounds are reached (Fig. 3a and Fig. S2). One of these pathways, which produces pyruvate and acetaldehyde in

four reaction steps (1.13.11.2, 3.7.1.-, 4.2.1.8 and 4.1.3.39 in Fig. S2), has been experimentally shown to be the main toluene degradation route based on gene expression data.<sup>22</sup> The iPsvr growth simulation on minimal media with toluene as the sole carbon source showed that this pathway was functional. In contrast, the second functional pathway (Fig. 3a),<sup>22</sup> was blocked in iPsvr because it lacked the enzyme 2.8.3.6 (highlighted in red in Fig. 3a), and it was therefore disconnected from the main metabolic subnetwork. We found that three out of the six genes in this pathway (PVR\_r1g5041, PVR\_r1g5042, PVR\_r1g1440) were more than 2-fold differentially expressed between the two growth conditions (highlighted in green in Fig. 3a), which further suggested that this pathway is indeed active in *P. veronii*. Homology-based BLAST searches<sup>33</sup> of the gene sequences of 2.8.3.6 against the *P. veronii* genome identified the corresponding gene for catalyzing this reaction (PVE\_r1g3867, e-value of  $10^{-20}$ , to *scoA*). Note that the default e-value in RAVEN is  $10^{-50}$ , which is why the reaction was not initially captured in the model from the genome annotation. Therefore, the missing reaction (2.8.3.6) was added to iPsvr, and the toluene pathway (Fig. 3a) was connected to the rest of the metabolic network through the Krebs cycle and carried flux.

The phenylalanine metabolic pathway was gap-filled using the exometabolomic data (Fig. 3b). Three out of the eight reaction steps of this pathway (4.1.1.28, 1.14.14.54 and 1.14.13.-) were initially missing in iPsvr, leading to a dead-end pathway without flux. Four metabolites of phenylalanine metabolism were detected in the exometabolomics data (colored in blue in Fig. 2b), one of which, 2-hydroxyphenylacetate, was absent in the iPsvr GEM. This suggested that phenylalanine metabolism should proceed via 2-hydroxyphenylacetate in *P. veronii* (Fig. 3b). The enzyme producing 2-hydroxyphenylacetate from phenylacetaldehyde (1.14.14.54) was found by BLAST (PVE\_r1g94, e-value of  $10^{-20}$  to CYP504), added to iPsvr and could thus link both metabolites. The next step (1.14.13.-), yielding homogentisate (Fig. 3b) was also added, but no protein sequence could be assigned, as it corresponds to an orphan reaction (KEGG reaction ID: R05450). The first enzyme (4.1.1.28), which decarboxylates L-phenylalanine to phenethylamine, must be present, given the detection of both compounds in the exometabolomics data, although no *P. veronii* protein sequence could be identified through BLAST search.

### Omic integration into iPsvr

REMI (Relative Expression and Metabolomics Integrations) was used to integrate the relative gene-expression and metabolite

<b>Genes</b>	
From annotation	1243
Added for gap-filing	26
<b>Reactions</b>	
Metabolic reactions	1610
From annotation	1539
Added for gap-filing	71
Transport reactions	365 <sup>a</sup>
Exchange reactions	173 <sup>a</sup>
<b>Metabolites</b>	
Intracellular compounds	1681
Extracellular compounds	173 <sup>a</sup>
Cellular building blocks	62

<sup>a</sup>iPsvr includes more extracellular compounds, transport and exchange reactions, which we kept blocked in this study. They are part of the model that will allow studying different media composition in the future metabolic studies of *P. veronii*

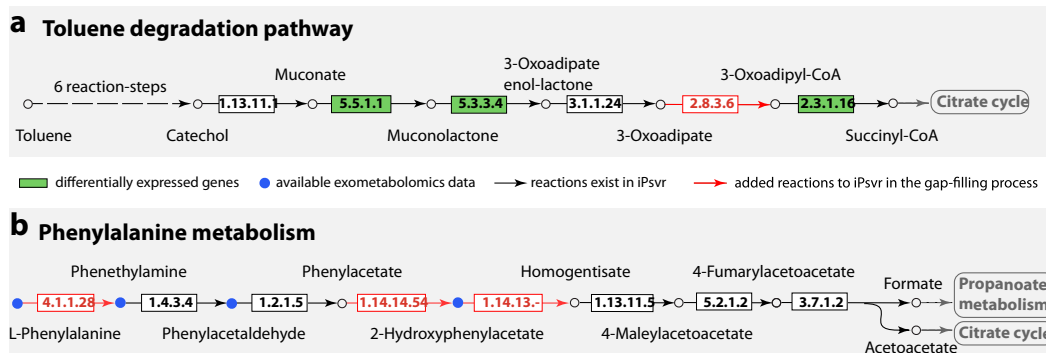
abundance data between EXPO and STAT into iPsvr. We explored three scenarios, depending on the type of data integrated into iPsvr: (i) REMI-TGex for the integration of relative gene-expression data only; (ii) REMI-TM for the integration of metabolomics data only; and (iii) REMI-TGexM for the simultaneous integration of both the gene-expression and metabolite abundance data into a thermodynamically curated model. The “T” in the three methods stands for the inclusion of thermodynamic constraints. Contextualized metabolic models, iPsvr-EXPO and iPsvr-STAT, coherent with omics data obtained at the exponential and stationary growth phase, respectively, were built, and the growth rates were again simulated (Table 2). 279 out of a total of 831 integratable gene-expression and metabolite abundance data points could be consistently included in REMI (Table 2, third row).

We analyzed the EXPO and STAT phase using iPsvr-EXPO and iPsvr-STAT models. To this end, we used two variants of the REMI method: REMI-TGex, where we integrate thermodynamics and gene expression data, and REMI-TGexM, where we integrate thermodynamics along with both gene-expression and metabolomics data. Both REMI-TGex and REMI-TGexM analyses show that the growth rates at the STAT phase are considerably lower than at the EXPO phase, as observed experimentally. However, the integration of exometabolomics data alone (REMI-TM) was insufficient to predict a reduction in growth rate at the STAT phase. Both REMI-TGex and REMI-TGexM correctly estimated near-zero growth rates of 0.0056 h<sup>-1</sup> and 0.0028 h<sup>-1</sup>, respectively, for the stationary phase (Table 2).

One of the distinguishable characteristics of REMI is the possibility of generating alternative solutions that correspond to the observed physiology (Methods). In this study, the alternative solutions allow investigating different flux distributions that result in the same simulated growth rate. Since REMI-TGexM can integrate the highest number of available experimental data as constraints (Table 2), we used this model in the study of alternative solutions.

REMI analysis revealed that there were 40 alternative solutions with the maximal consistency score (MCS) of 279 (Supplementary File. 1). We identified a set of 202 constraints, 162 from gene expression and 40 from exometabolomics data (Table S7), that were common amongst all alternative solutions. A total of 45 constraints varied amongst alternative outcomes (Table S7). Interestingly, the 40 constraints from exometabolomics data were common among all the 40 solutions and the difference (45 constraints) originated from the gene expression data.

We used the KEGG mapper tool to study those pathways that deregulate during the transition from exponential to stationary phase.<sup>34</sup> The common reactions and metabolites amongst all the alternative solutions generated by REMI were mapped to KEGG



**Fig. 3** Gap filling of two blocked pathways in iPsvr based on the gene expression and exometabolomics data. **a** One of the two toluene degradation pathways in *P. veronii* (from the KEGG pathway database<sup>53</sup>) involved one missing enzyme (highlighted in red), and three genes (highlighted in green) differentially expressed between the exponential and stationary growth conditions. **b** The phenylalanine metabolic pathway involved three missing enzymes (highlighted in red) and four metabolites (highlighted in blue) identified in the exometabolomic data.

**Table 2.** Summary of the REMI results of the relative integration of exometabolomics (REMI-TM), gene expression (REMI-TGex) and both datasets (REMI-TGexM) into iPsvr.

Method	Score		Growth rate (h <sup>-1</sup> )	
	TMCS	MCS	EXPO	STAT
<b>REMI-TM</b>	215	50	0.253	0.250
<b>REMI-TGex</b>	616	235	0.2567	0.0056
<b>REMI-TGexM</b>	831	279	0.25 ± 0.0004	0.0028 ± 0.0016
<b>Experimental</b>	—	—	0.31 ± 0.05	0

Theoretical maximum consistency score, TMCS; Maximum consistency score, MCS; Std, standard deviation

pathways using KEGG mapper<sup>34</sup> (Table S7). Biosynthesis of amino acids (KEGG map01230) with 41 deregulated reactions and metabolites scored as the most deregulated pathway. This was followed by Carbon metabolism (KEGG map01200) with 28 and Purine metabolism (KEGG map00230) with 21 deregulated reactions and metabolites. The complete list of the deregulated pathways is provided in Table S7. This analysis suggests that the deregulation of central carbon metabolism and amino acid synthesis plays a critical role in growth arrest in the transition from exponential to stationary phase.

Analysis of the biomass precursor production at different growth phases

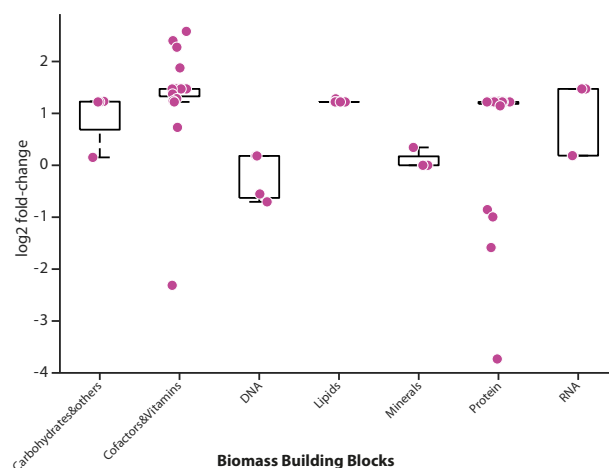
To further understand the underlying mechanisms of growth reduction when cells transit from EXPO to STAT, we identified the biomass precursors that may become limiting in the stationary phase, therefore leading to growth arrest. The 62 biomass precursors were grouped into seven groups of biomass building blocks (BBBs), namely carbohydrates, cofactors and vitamins, DNA nucleotides, lipids, minerals, amino acids, and RNA nucleotides. For each metabolite, we calculated the fold-change of their maximal production between EXPO and STAT using REMI-TGexM (Table S8).

For the BBB analysis, we used REMI-TGexM and investigated the variation of maximum possible BBB production across all the 40 alternative solutions (Supplementary Information 2), and through sampling, we identified the most representative solution (see Methods). Our results indicated large variations in the maximum production of BBBs between the two phases (Fig. 4), with the maximal relative change occurring in the production of cofactors and vitamins. The production of 52 out of the 62 BBBs (83%), was higher in iPsvr-EXPO than in iPsvr-STAT (Fig. 4, Table S8). The production of minerals remained unchanged, and rather surprisingly, it is possible to achieve higher production of 7 BBBs in STAT conditions (Fig. 4, Table S8). The maximum production of 19 out of the 20 precursors that were classified as cofactors and vitamins, was higher in iPsvr-EXPO than in iPsvr-STAT (Fig. 4, Table S6), and the only exception was Biotin. The maximum allowed production of four amino acids; L-Methionine, L-Cysteine, L-Leucine and L-Tyrosine was also higher in STAT (Fig. 4).

*P. veronii* adaptation from liquid culture to the soil as growth environment

We analyzed the physiology for *P. veronii* during adaptation to the soil environment, using a previously published genome-wide transcriptome dataset of cells exposed for 1 h to liquid medium or to sand, with either toluene or succinate as carbon substrate.<sup>22</sup>

Integrating previously obtained transcriptomic data from liquid versus sand exposure<sup>22</sup> into iPsvr using REMI-TGex, we produced two models, iPsvr-LIQUID for the liquid media and iPsvr-SAND for



**Fig. 4** Differential *in silico* production (mmol/gDW/h) of seven biomass building block groups between the exponential and stationary growth phase. The result of the integration of relative gene expression and metabolite abundance data (EXPO vs STAT) using REMI-TGexM. Each dot in the graph represents the individual biomass precursors within that category and all the dots represent statistically significant changes ( $p = \text{value} < 0.05$ ). For BBB group statistics, see Table S8.

the sand environment. While TMCS was 612 for this dataset, we obtained a MCS of 246 and 265 for toluene and succinate carbon sources, respectively. We then generated alternative solutions for the maximum consistency score and simulated growth for the two media (Supplementary Information 1). *In silico* growth rates of  $0.297 \pm 0.002 \text{ h}^{-1}$  and  $0.44 \pm 0.008 \text{ h}^{-1}$  were predicted for toluene and succinate in liquid medium, respectively. Remarkably, the model predicted strong reduction of growth rate upon transition to the sand, i.e., from  $0.297 \text{ h}^{-1}$  to  $0.161 \pm 0.0001 \text{ h}^{-1}$  (toluene) and  $0.44 \text{ h}^{-1}$  to  $0.161 \pm 0.002 \text{ h}^{-1}$  (succinate). This suggests that cells have to adapt to soil as their new environment and have to reprogram their physiology before resuming growth.

To understand how the transition from liquid medium to sand impacts the intracellular metabolic fluxes and growth of *P. veronii*, we performed the analysis of alternative solutions. In the case of succinate, among the 195 alternative solutions, 152 reactions were in common and 88 reactions were varying between them, whereas for toluene, 134 reactions were common and 97 reactions varied across 84 different alternatives (Table S9). The KEGG mapper tool<sup>34</sup> was used to identify the most deregulated pathways within the transition from liquid to sand, for both succinate and toluene (Table S9). Similar to the previous pathway analysis of transition from EXPO to STAT, the two most deregulated pathways were Carbon metabolism (KEGG map01200) and Biosynthesis of amino acids (KEGG map01230), for both Toluene and Succinate (Table S9). Fatty acid metabolism (KEGG map01212) was the third most deregulated pathway (with 20 deregulated reactions) for toluene, and the fifth (with 8 deregulated reactions) for succinate. Interestingly, fatty acid metabolism was not amongst the 25 top ranked deregulated pathways in transition from EXPO to STAT. Another difference from the EXPO to STAT pathway analysis was the deregulation of 2-Oxocarboxylic acid metabolism (KEGG map01210) as the third and the sixth deregulated pathways in toluene and succinate sand exposures, respectively. This suggests that although lower growth is observed in both transitions, i.e., from EXPO to STAT and from liquid to sand, the deregulated pathways underlying the growth reduction are not necessarily the same.

We further investigated the production of BBBs in both growth environments (Table S10). In contrast to the EXPO to STAT

transition results, in the case of cells exposed to sand, the production of the majority of precursors was higher in sand compared to liquid media (Fig. S3a, b). This suggests that although cells grow much less in sand than in liquid, they still actively produce BBBs, and the maximum production can be even higher in sand than liquid media for certain BBBs (Table S10). The higher production of lipids and amino acids in sand for both toluene and succinate might indicate that cells reshape their membrane content and protein metabolism upon transition to sand.

## DISCUSSION

This work demonstrates how a metabolic-model-based (multi) omics data integration approach can accurately capture cell physiology and adaptation during growth or environmental transitions. Metabolic models are typically used for predicting optimal cell growth, but not for growth transitions and adaptation. We show here the usefulness of models with comprehensively integrated transcriptomics and exometabolomic to capture changes in growth rate during adaptation. As example, we studied transition of the toluene-degrading bacterium *P. veronii* 1YdBTEX2 from exponential growth to stationary phase, and during adaptation from liquid to a sand environment.

For the purpose of this study, we reconstructed and curated the first genome-scale metabolic model of *P. veronii*, iPsvr. iPsvr is useful for a wide range of applications in biological and biotechnological research, such as the integration of omics data, investigation of relationships among species in microbial communities, bioremediation strategies, design of metabolic engineering strategies, hypothesis-driven discovery, and analysis of metabolic network properties. As such, iPsvr may represent a valuable resource for the study of complex microbiomes and microbiome engineering.

GEMs usually contain inconsistencies that manifest as blocked reactions, i.e., the reactions that carry zero flux. Most of the available metabolic models are gap-filled to obtain a functional model that produces all the BBBs. Nevertheless, many reactions in GEMs are not directly connected to biomass formation and may remain blocked even after routine gap-filling. Roughly, 20–50% of the reactions of published GEMs are blocked reactions.<sup>35,36</sup> For example, in the most recently and best-curated published *E. coli* model, iML1515, 10% of the reactions are blocked.<sup>37</sup> We identified the blocked reactions/pathways in iPsvr and introduced a novel omics-based gap-filling approach, which targets the gap-filling of reactions/pathways by looking at differentially expressed genes or metabolites. We showed that the gap-filling step improved the metabolic model connectivity, thus increasing the number of functional pathways that can carry flux. The curation also increased the consistency of the metabolic model with the omics data, expressed as the number of genes/metabolites with available omics data that can be integrated into iPsvr. We illustrated through two gap-filling examples on toluene and phenylalanine metabolism how we could add the missing parts of metabolism and increase consistency. Gap-filling using metabolomics data not only improved the network connectivity, but also allowed integrating metabolites and uncovering missing reactions that must be part of the metabolic network but are missing due to unannotated and misannotated genes as well as promiscuous enzymes.<sup>38</sup>

Statistical analysis of gene expressions and metabolite levels revealed significant variations across the exponential and stationary phase, exemplifying the known distinct physiologies under these conditions. However, the system's level physiology and the underlying adaptation mechanisms remained unrevealed.

It should be highlighted that in FBA, the growth rate is evaluated under the assumption that all metabolic fluxes in the cell are geared towards the maximal production of biomass at each moment of the cell growth and under any environmental

condition. Nevertheless, inclusion of REMI,<sup>19</sup> a method for the integration of omics data into metabolic models and for building context-specific models, produced predictions of *P. veronii* growth rates under dynamic conditions. Integration of the transcriptomic data (REMI-TGex and -TGexM) proved to be vital for predictions of growth rate dynamics, i.e., higher growth rate at exponential versus almost a growth arrest at the stationary phase, whereas inclusion of metabolomic data alone (REMI-TM) was insufficient. This might be expected, because only a small subset of metabolites could be putatively identified, such that their impact on the predictions was comparatively smaller. Overall, these results clearly demonstrated that integrating multi-omics data in GEMs significantly increases the consistency of the model predictions with experimental observations.

The analysis of alternative outcomes generated by REMI allowed us to better characterize the intracellular state of *P. veronii* growth by identifying the metabolic pathways whose fluxes were consistently changed across all alternatives thus identifying where the fluxome was mostly rewired. Interestingly, although the same growth reduction phenotype was observed for both the transition from EXPO to STAT and liquid to sand, the deregulated pathways were not completely identical.

REMI analysis of iPsvr-EXPO and iPsvr-STAT suggested *P. veronii* adapts to STAT phases primarily by limiting the metabolic fluxes for the production of the majority of cofactors and vitamins required for growth, with metabolism reshuffling being regulated at the transcriptional level. In addition, an increased production of specific amino acids in the STAT phase suggests the cell is preparing for starvation and survival. In contrast to the EXPO to STAT transition, REMI analysis of iPsvr-LIQUID and iPsvr-SAND indicated that, surprisingly, the maximum production of most of precursors from the *cofactors and vitamins*, and the BBBs *amino acids* actually increased in cells transitioned to sand, consistent for both sand with toluene or succinate as the sole carbon source. We hypothesize that the production of these BBBs and the function of their associated metabolic pathways are rewired to adapt to the local sand conditions. This finding resonates with previous work of Morales et al.,<sup>22</sup> who concluded from Gene Ontology terminology that *P. veronii* cells inoculated in sand readjusted their metabolism during the first hour of contact.

Collectively, our results demonstrate the importance of integrating into metabolic models the contextualization of condition-specific gene-expression and metabolite-abundance data. This increases the value of growth rate predictions and improves the assessment of the relative changes of measurable metabolic phenotypes. The method as demonstrated here is thus an important advancement to explain, quantify or predict cellular responses to environmental or genetic perturbations, which is crucial for microbiome engineering.

## METHODS

Culture conditions for *P. veronii* transcriptome and exometabolomic studies

*P. veronii* 1YdBTEX2 was grown on solid 21C mineral medium<sup>39</sup> with toluene in the vapor phase at 30 °C for 3 days. Colonies were inoculated into three 100 ml screw-cap conical flasks containing 25 ml of the mineral medium amended with 0.5 ml of a 1:19 mixture of toluene:tetradecane (Sigma-Aldrich ref. 34866; Aldrich ref. 87140). Flasks were incubated at 30 °C and at 180 rpm on an orbital shaker until a culture turbidity (OD<sub>600</sub>) of 0.7 (mid-exponential growth phase). At this point, the three cultures were pooled, and the bacterial cells were harvested by centrifugation (swing-out rotor A-4-44, Eppendorf; 3220 g, 8 min, 30 °C). The cell pellet was resuspended in the mineral medium and centrifuged again, as described above, to remove any residual carbon. After that, the cells were resuspended in mineral medium, diluted to obtain a starting OD<sub>600</sub> of 0.16 and transferred into four replicate flasks, to which the toluene:tetradecane mix was added, as described above. The cultures were incubated at 30 °C and 180 rpm and regularly sampled for OD measurements. For

transcriptomics,  $\sim 1 \times 10^9$  cells were harvested from an appropriate volume of the culture by centrifugation at 3500 g, for 6 min at 30 °C; snap-frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$  until RNA extraction. Three quadruplicate sample sets were produced: the inoculum at time 0 (T0h), that is cells starved for  $\sim 30$  min at 25 °C in mineral medium without a carbon source; exponentially growing cells (at  $\text{OD}_{600} \approx 0.5$ , harvested after 4 h of incubation in the toluene-amended mineral medium); stationary phase cells (at  $\text{OD}_{600} \approx 1.8$ – $1.9$  harvested after 24 h incubation). To study the composition of the exometabolome (metabolites in the culture medium excreted or leaked out of cells), 1 ml of the culture was sampled at the same time points. Cell culture was transferred into a 1.5 ml polypropylene microcentrifuge tube, which was clarified by 20 min centrifugation at  $21,100 \times g$  and 4 °C, after which 0.5 ml of spent media was transferred to a new polypropylene tube, stored at  $-80^\circ\text{C}$  and shipped on dry ice to the metabolomics facility.

## Biolog data

*P. veronii* 1YdBTEX2 was plated from  $-80^\circ\text{C}$  on LB agar and incubated at 30 °C for 16 h to obtain isolated colonies. A single colony was subsequently restreaked on a minimal medium agar plate, which was incubated in a closed glass jar with a vial of toluene. Single colonies grown on toluene were transferred to closed glass bottles with 20 ml minimal medium to which 20  $\mu\text{l}$  of toluene dissolved in 1 ml of tetradecane was added. Cultures were incubated with agitation at 30 °C and cells were harvested after 24 h by centrifugation, removing tetradecane carefully. Cells were resuspended in minimal medium to a density of  $10^6$  per 75  $\mu\text{l}$ . Wells of BIOLOG Ecoplates™ containing triplicates of 31 carbon sources and one no-carbon control were filled with 75  $\mu\text{l}$  minimal medium, after which 75  $\mu\text{l}$  of *P. veronii* cell suspension was added and mixed. In total five Ecoplates were inoculated, covered with a sterile seal and incubated in the dark for 16 days at room temperature. Absorbance at 590 nm, representative for the respiration activity of the cells, was measured by spectrophotometry every second day.

## RNA extraction, RNA-seq library preparation and sequencing

Total RNA from the frozen cell pellets was extracted using the RNA PowerSoil Total RNA Isolation Kit (MoBio Laboratories) as recommended by the manufacturer. Contaminating genomic DNA was removed by two cycles of TURBO DNase (Invitrogen) digestion and RNeasy MinElute Cleanup kit (QIAGEN) column purification. The quantity, purity and integrity of RNA samples were assessed using agarose gel electrophoresis, NanoDrop spectrophotometer (ThermoFisher Scientific) measurements and Agilent 2100 Bioanalyser (Agilent Technologies) profiling.

The completeness of DNA removal was verified by PCR using primers Pv\_chr2\_fw, ATCGGCTGTCCGACATCGGACG and Pv\_chr2\_rev, TCGAAGAGCTCCACCGAGAGCCGCC and 1 pg of genomic DNA as a positive control, as described previously.<sup>22</sup> Next, 4  $\mu\text{g}$  of each RNA sample were depleted from ribosomal RNAs, converted to the reverse-complement stranded Illumina sequencing library using the ScriptSeq Complete Kit (Bacteria, Illumina) and indexed with ScriptSeq™ Index PCR primers set 1 (Epicentre, Illumina) following the standard protocol. The resulting directional RNA-seq libraries were sequenced using single-end 100-nt read chemistry on an Illumina HiSeq 2500 platform (Illumina) at the Lausanne Genomic Technologies Facility.

## Untargeted LC-HRMS metabolomics

The spend media (100  $\mu\text{l}$ ) samples collected at different time points were extracted with 400  $\mu\text{l}$  of ice-cold methanol to quench the metabolism, precipitate proteins and extract a broad range of polar metabolites. The extracted media were analyzed by HILIC-HRMS using an electrospray ionization source operating in both positive and negative mode. Pooled QC samples (representative of the entire sample set) were analyzed periodically (every 4 samples) throughout the entire analytical run in order to assess the quality of the data, correct the signal intensity drift and remove the peaks with poor reproducibility ( $\text{CV} > 30\%$ ) that can be considered chemical or bioinformatic noise. Data were acquired using a 1290 UHPLC system (Agilent Technologies) interfaced with a 6550 iFunnel Q-TOF mass spectrometer operating in a full-scan MS mode. In addition, pooled QC samples were analyzed in auto MS/MS mode (i.e., Data Dependent Analysis [DDA]) to acquire the MS/MS data for metabolite identification. In positive mode, chromatographic separation was carried out using an Acquity BEH Amide, 1.7  $\mu\text{m}$ , 100 mm  $\times$  2.1 mm I.D. column (Waters, Massachusetts, US). The mobile phase was composed of

A = 20 mM ammonium formate and 0.1% formic acid in water and B = 0.1% formic acid in acetonitrile. A linear gradient elution from 95% B (0–1.5 min) down to 45% B (17–19 min) was applied followed by 5 min for column re-equilibration to the initial gradient conditions. The flow rate was 400  $\mu\text{L}/\text{min}$ , column temperature 30 °C and sample injection volume 2  $\mu\text{l}$ . ESI source conditions were set as follows: dry gas temperature 290 °C and flow 14 L/min, fragmentor voltage 380 V, sheath gas temperature 350 °C and flow 12 L/min, nozzle voltage 0 V, and capillary voltage 2000 V. In negative mode, a SeQuant ZIC-pHILIC (100 mm, 2.1 mm I.D. and 5  $\mu\text{m}$  particle size; Merck, Darmstadt, Germany) column was used. The mobile phase was composed of A = 20 mM ammonium acetate and 20 mM  $\text{NH}_4\text{OH}$  in water at pH 9.3 and B = 100% acetonitrile. The linear gradient elution ran from 90% (0–1.5 min) to 50% B (8–11 min) down to 45% B (12–15 min). Finally, the initial chromatographic conditions were established during a 9 min post-run for column re-equilibration. The flow rate was 300  $\mu\text{L}/\text{min}$ , column temperature 30 °C and sample injection volume 2  $\mu\text{l}$ . ESI source conditions were set as follows: dry gas temperature 290 °C and flow 14 L/min, fragmentor voltage 380 V, sheath gas temperature 350 °C, nebulizer 45 psi and flow 12 L/min, nozzle voltage 0 V, and capillary voltage 2000 V. The fragmentor voltage of 380 V has been optimized in both negative and positive using pure chemical standards in order to improve metabolite signal and minimize in-source fragmentation. Thus, this value is found as the best compromise between sensitivity and in-source fragmentation.

In the MS-only mode, the instrument was set to acquire over the  $m/z$  range 50–1200, with the MS acquisition rate of two spectra/s. Targeted MS/MS data for dysregulated metabolite features were acquired using the inclusion list with narrow isolation window ( $\approx 1.3 m/z$ ), MS acquisition rate of 500 ms, and MS/MS acquisition rate of 500 ms.

## Data processing and statistical analysis

**Transcriptomics.** Read mapping, sorting and formatting of the raw reads was done with Bowtie2<sup>40</sup> and Samtools,<sup>41</sup> using the finalized gapless *P. veronii* 1YdBTEX2 genome sequence as in Morales et al.<sup>22</sup> Mapped reads were counted with HTSeq,<sup>42</sup> then further processed and analyzed with edgeR.<sup>43</sup> Only reads counted more than once per million in at least three replicates were kept. After normalization of the counts, transcript abundances were compared in pairwise conditions in a modified Fischer exact test (as implemented in edgeR). Genes were called significantly differentially expressed between two EXPO and STAT when their false-discovery rate was  $< 0.05$  and their fold-change  $> 2$  and were subsequently interpreted using Gene Ontology (GO) analysis. GO terms of *P. veronii* genes were inferred using the program BLAST2GO.<sup>44</sup> The same software was then used to analyze GO datasets of significantly differentially expressed genes in each pair-wise comparison, under the TopGO “Weight” algorithm.

**Metabolomics.** Raw LC-MS data were converted to mzXML files using ProteoWizard MS Convert. mzXML files were uploaded to XCMS for data processing including peak detection, retention time correction, profile alignment, and isotope annotation. Data were processed as a multi-group experiment, and the parameter settings were as follows: centWave algorithm for feature detection ( $\Delta m/z = 20$  ppm, minimum peak width = 5 s and maximum peak width = 30 s, S/N threshold = 6,  $mzdiff = 0.01$ , integration method = 1, prefilter peaks = 3 prefilter intensity = 1000, noise filter = 0), obiwarp settings for retention time correction (profStep = 1), and parameters for chromatogram alignment, including  $mzwid = 0.015$ ,  $minfrac = 0.5$  and  $bw = 5.5$ . Preprocessed data (following the signal-intensity drift correction and noise removal with the “batchCorr” R package) were filtered according to the  $p$ -value ( $< 0.05$ ) and signal intensity ( $> 1000$  ion counts). The remaining table of metabolite features together with the most significant ion features selected from the loadings plot of the multivariate models in positive and negative ionization mode were subjected to metabolite identification as described below.

In the first instance, putative metabolite identification was performed by accurate mass and retention time (AMRT) matching against an in-house database (containing information on 600 polar metabolites from the Mass Spectrometry Metabolite Library Supplied by IROA Technologies, Sigma-Aldrich, characterized under the same analysis conditions). For that, raw data files (.d) were processed using Profinder B.08.00 software (Agilent Technologies) with the following parameter settings: mass tolerance 10 ppm, retention time tolerance 0.2 min, height filter 1000 counts, and peak spectrum obtained as an average of scans at 10% of the peak. In parallel, the XCMS output table of significantly different metabolite features was



matched against the Human Metabolome Database (HMDB)<sup>25</sup> based on accurate mass (AM) with  $\Delta\text{ppm} = 10$ . The list of hits was further manually curated by taking into account the biological relevance of the hit (endogenous vs. exogenous metabolites) and a presence of the “true” peak shape (using the interactive XCMS Online interface<sup>45</sup>). Short listed ions of interest together with the metabolites identified by the in-house database were further subjected to targeted MS/MS validation. The metabolite identifications were validated by matching the MS/MS<sup>46</sup> data acquired in the pooled samples (for each experiment in each LC–MS analysis mode) against the in-house PCDL database, METLIN (<https://metlin.scripps.edu/>) standard metabolite database<sup>47</sup> or mzCloud (<https://www.mzcloud.org/>). Otherwise, if the MS/MS data quality did not allow for metabolite ID confirmation, the metabolite IDs remained putative, based only on accurate mass matching.

### Genome-scale model reconstruction

The *P. veronii* 1YdBTEX2 genome-scale metabolic reconstruction process combines the automated draft metabolic network process with several manual refinements and curation procedures, in total involving four main steps: (i) functional annotation of the genome: functional annotation of the genome is required prior to the reconstruction of the GEMs, and consequently, the quality of a GEM highly depends on the availability of a gapless genome. The protein sequences (FASTA files) of *P. veronii* 1YdBTEX2 were acquired from a previously published study<sup>22</sup> and were annotated to identify the associated reactions that enzymes catalyze to determine the stoichiometric matrix, done using the RAVEN Toolbox.<sup>26</sup> The generation of the draft metabolic network followed the protocol detailed previously,<sup>26,48,49</sup> and the output of this annotation process is summarized in Table S6. Version 1.07 of the RAVEN Toolbox and the version of KEGG database as of September 2017 were used. (ii) Compartmentalization, the definition of biomass reaction and uptakes and secretions: In the absence of information or experimental evidence about the compartmentalization and the cell content in *P. veronii* 1YdBTEX2, the cell compartment information (i.e., cytosol and periplasm), the transport mechanism between the compartments and the extracellular environment, the uptakes and secretions, and the biomass composition were obtained from two available models of closely related Pseudomonads: *Pseudomonas putida*<sup>27–29</sup> and *Pseudomonas stutzeri*.<sup>30</sup> The biomass reaction in a GEM designates the metabolic precursors that build the BBBs, i.e., DNA, RNA, amino acids, lipids and carbohydrates and their corresponding stoichiometric coefficients. (iii) Thermodynamic curation: In a thermodynamically curated model, the standard Gibb's free energy of a reaction and consequently the directionality of the reactions, i.e., reversible or irreversible, are associated with the reaction as additional constraints, which allow the performance of TFA. We followed the established protocol to thermodynamically curate metabolic models<sup>31,32</sup> to add thermodynamics constraints to iPsvr, where we could estimate the standard Gibb's free energy of formation of 76% metabolites, using the group contribution method (GCM)<sup>50</sup> and the standard Gibb's free energy of 84% of metabolic reactions in iPsvr. This allowed us to perform TFA, simulate the growth and determine whether iPsvr was functional, with further evaluations if it correctly predicted the expected growth-associate phenotypes. (iv) Gap-filling of iPsvr: The draft metabolic network of iPsvr did not contain all the necessary reactions for the production of all the biomass building blocks, and thus, the model did not show any growth. This is a very common observation in the reconstruction process of GEMs, since the (automatic) genome annotations are often incomplete or erroneous, with significant proportion of predicted proteins having no functions attributed.<sup>51</sup> Therefore, reactions without associated genes were included in order to obtain a functional model that simulates non-zero growth, done using a procedure called gap-filling. We followed the gap-filling procedure introduced previously,<sup>48</sup> wherein the production of each biomass precursor is defined as a metabolic task<sup>26</sup> and a mixed-integer linear programming (MILP) formulation is used to generate alternative groups of minimal number of reactions (borrowed from KEGG) that enable the production of the biomass precursor. The draft GEM was gap-filled by iterative manual curation until a model was obtained that was able to carry non-zero flux through the biomass reactions at steady state (i.e., signifying ‘growth’). Although gap-filling is a routine procedure in the curation of GEMs and most of the available GEMs are gap-filled to obtain a functional model that is able to grow under defined conditions, gaps in pathways not involved with biomass production are, however, mostly overlooked in GEM analysis. We performed a complementary second gap-filling step apart from the biomass reaction, where as a result, the consistency of iPsvr with the

obtained experimental data on gene expression and metabolomics was increased. Available GEMs usually contain a large well-connected subnetwork, which encompasses the most of the central carbon metabolism and a part of the secondary metabolism, and many isolated reactions (or sets of reactions) are probably disconnected from the rest of the network because of misannotations or insufficiently known pathways. Such isolated reactions/pathways are blocked, i.e., cannot carry flux under any condition, and therefore one or more reactions must be added to connect the blocked reactions with the rest of the metabolic network. To this end, the metabolic network structure of iPsvr was decomposed to the main component and to the isolated (disconnected) reactions/pathways using a MATLAB graph-based built-in function (conncomp). Using the same gap-filling approach, the blocked reactions/pathways that were associated with differentially expressed genes (in the pair-wise comparison of exponential- and stationary-phase datasets) or measured metabolites in exometabolomic data were gap-filled to become functional in the model (carry non-zero flux). (v) Quality control of iPsvr: To assure the quality of the model, especially reaction balances, we used Memote,<sup>52</sup> an automated genome-scale metabolic model test suite. Memote pinpointed reactions that were not balanced, and we manually balanced them, and commented the reasons for the reactions' imbalance (Table S11).

### Omics data integration

REMI<sup>19</sup> was used for the integration of transcriptomics and exometabolomics data into iPsvr. REMI assumes that reaction fluxes associated with genes that are significantly differentially expressed are deregulated. Moreover, REMI also considers that the *in vivo* metabolite abundance ratios between the two conditions, e.g., the two growth phases can be used to constrain reaction fluxes associated with the metabolites that are differentially regulated. Expression/abundance-based ratios between the two conditions are formulated as flux perturbations for each reaction and are imposed as constraints on individual fluxes. Based on the data used, it translates into three different methods: (i) REMI-TGex allows the integration of relative gene expression data into a thermodynamically curated GEM, (ii) REMI-TM allows the integration of metabolomics data and (iii) REMI-TGexM integrates simultaneously both the gene expression and metabolite abundance data as additional constraints into the metabolic model.

REMI aims to maximize consistency between differential expression and fluxes as well as differential metabolite concentrations and fluxes. To study condition-specific differences in metabolism between two conditions (perturbed vs reference), REMI considers a separate metabolic model for each condition. Then, to integrate differential expression, REMI enforces a higher flux through a reaction in the perturbed condition (perturbed model) as compared to a reference condition (reference model) if the genes of the reaction are upregulated. For downregulated reactions, REMI enforces a lower flux as compared to a reference. To integrate extracellular metabolite concentrations, REMI assumes that if a metabolite is upregulated, then the production of the metabolite is forced to be higher as compared to a reference condition, and similarly, a lower production of a metabolite is forced if a metabolite is found to be downregulated. Then, an optimization problem is formulated to maximize the number of constraints imposed by the relative gene expression and metabolite abundances that can be integrated into the model while preserving a growth phenotype. Two scores are calculated: a theoretical maximum consistency score (TMCS), representing the number of genes/metabolites with available omics data, and the maximum consistency score (MCS), representing the number of genes/metabolites whose relative omics data are consistent with relative network fluxes and therefore can be integrated in the model.

The MILP formulation enables enumerating alternative sets (size equal to MCS) from a given set of constraints. The most consistent models are built by activating constraints that are overlapping and consistent between all the alternatives. Alternative solutions were computed for the maximal consistency score with the MILP solver in IBM CPLEX 12.7.2, and the time limit for the optimization was set to 3 h.

We used the transcriptomics and/or exometabolomics datasets measured at the exponential and stationary phase and also at the sand versus liquid environment, to derive additional flux constraints for the TFA problem, which were applied using REMI.

## Thermodynamics-based flux analysis (TFA) and the analysis of the biomass building blocks

To determine fluxes and subsequently the growth rate at different growth and environmental conditions using iPsvr, we employed FBA, the most widely used constraint-based modeling technique for studying biochemical networks and cellular physiology. Previous studies show that the integration of appropriate thermodynamic constraints leads to more accurate metabolic model predictions and also a significant reduction in the ranges of the predicted fluxes (solution space).<sup>8,31</sup> We perform TFA for estimating the growth rate before and after the integration of omics data.

We further identified the biomass building blocks (BBB), wherein a low/zero production limits growth upon the transition of cells from the exponential to stationary phase or from liquid to the soil environment (two examples that were discussed in this work). Each BBB was tested by defining the TFA objective function as the maximum production of that metabolite under the defined media condition (the same for all the BBBs). Then, the BBB production was compared and the limiting BBBs for each case were identified.

## DATA AVAILABILITY

The RNAseq data that support the findings of this study have been deposited in the European Nucleotide Archive (ENA) under bioproject number PRJEB30057 and are accessible at <https://www.ebi.ac.uk/ena/data/search?query=PRJEB30057>. The processed metabolomics data is available on XCMS online server – under publically shared jobs.

All relevant data and scripts for reproducing the results are within the Supporting Information files, and on Zenodo at <https://zenodo.org/record/3479819#XcFsc-hKgaY>. The REMI code is available on GitHub at <https://github.com/EPFL-LCSB/remi>.

Received: 8 July 2019; Accepted: 28 November 2019;

Published online: 07 January 2020

## REFERENCES

- Sheth, R. U., Cabral, V., Chen, S. P. & Wang, H. H. Manipulating bacterial communities by in situ microbiome engineering. *Trends Genet.* **32**, 189–200 (2016).
- Stenuit, B. & Agathos, S. N. Deciphering microbial community robustness through synthetic ecology and molecular systems synecology. *Curr. Opin. Biotechnol.* **33**, 305–317 (2015).
- Buffe, C. G. et al. Precision microbiome reconstitution restores bile acid mediated resistance to *Clostridium difficile*. *Nature*. <https://doi.org/10.1038/nature13828> (2015).
- De Roy, K., Marzorati, M., Van den Abbeele, P., Van de Wiele, T. & Boon, N. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Environ. Microbiol.* <https://doi.org/10.1111/1462-2920.12343> (2014).
- Adams, G. O., Fufeyin, P. T., Okoro, S. E. & Ehinomen, I. Bioremediation, biostimulation and bioaugmentation: a review. *Int. J. Environ. Bioremed. Biodegrad.* **3**, 28–39 (2015).
- Robaina-Estévez, S. & Nikoloski, Z. On the effects of alternative optima in context-specific metabolic model predictions. *PLoS Comput. Biol.* **13**, e1005568 (2017).
- Lewis, N. E. et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* <https://doi.org/10.1038/msb.2010.47> (2010).
- Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805 (2007).
- Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat. Biotechnol.* **28**, 245–248 (2010).
- Johnson, C. H., Ivanisevic, J. & Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat. Rev. Mol. Cell Biol.* **17**, 451–459 (2016).
- Richelle, A., Joshi, C. & Lewis, N. E. Assessing key decisions for transcriptomic data integration in biochemical networks. <https://doi.org/10.1101/301945>.
- Töpfer, N., Kleessen, S. & Nikoloski, Z. Integration of metabolomics data into metabolic networks. *Front. Plant Sci.* **6**, 49 (2015).
- Pinu, F. R. & Villas-Boas, S. G. Extracellular microbial metabolomics: the state of the art. *Metabolites* **7**, 43 (2017).
- Pinu, F. R. et al. Metabolite secretion in microorganisms: the theory of metabolic overflow put to the test. *Metabolomics* **14**, 43 (2018).
- O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R. & Palsson, B. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* <https://doi.org/10.1038/msb.2013.52> (2013).
- Navid, A. & Almaas, E. Genome-level transcription data of *Yersinia pestis* analyzed with a new metabolic constraint-based approach. *BMC Syst. Biol.* **6**, 150 (2012).
- Töpfer, N., Jozefczuk, S. & Nikoloski, Z. Integration of time-resolved transcriptomics data with flux-based methods reveals stress-induced metabolic adaptation in *Escherichia coli*. *BMC Syst. Biol.* <https://doi.org/10.1186/1752-0509-6-148> (2012).
- Collins, S. B., Reznik, E. & Segrè, D. Temporal expression-based analysis of metabolism. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1002781> (2012).
- Pandey, V., Hadadi, N. & Hatzimanikatis, V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLOS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1007036> (2019).
- de Lima-Morales, D. et al. Draft Genome Sequence of *Pseudomonas veronii* Strain 1YdBTEX2. *Genome Announc.* <https://doi.org/10.1128/genomeA.00258-13> (2013).
- De Lima-Morales, D. et al. Degradation of benzene by *Pseudomonas veronii* 1YdBTEX2 and 1YB2 is catalyzed by enzymes encoded in distinct catabolism gene clusters. *Appl. Environ. Microbiol.* <https://doi.org/10.1128/AEM.03026-15> (2016).
- Morales, M. et al. The genome of the toluene-degrading *Pseudomonas veronii* strain 1YdBTEX2 and its differential gene expression in contaminated sand. *PLoS ONE* **11**, 1–21 (2016).
- Trigo, A., Valencia, A. & Cases, I. Systemic approaches to biodegradation. *FEMS Microbiol. Rev.* **33**, 98–108 (2009).
- Adadi, R., Volkmer, B., Milo, R., Heinemann, M. & Shlomi, T. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. *PLoS Comput. Biol.* **8**, e1002575 (2012).
- Wishart, D. S. et al. HMDB: The human metabolome database. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkl923> (2007).
- Agren, R. et al. The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1002980> (2013).
- Nogales, J., Palsson, B. & Thiele, I. A genome-scale metabolic reconstruction of *Pseudomonas putida* KT2440: iJN746 as a cell factory. *BMC Syst. Biol.* <https://doi.org/10.1186/1752-0509-2-79> (2008).
- Puchalka, J. et al. Genome-scale reconstruction and analysis of the *Pseudomonas putida* KT2440 metabolic network facilitates applications in biotechnology. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1000210> (2008).
- Yuan, Q. et al. Pathway-Consensus approach to metabolic network reconstruction for *Pseudomonas putida* KT2440 by systematic comparison of published models. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0169437> (2017).
- Babaei, P., Marashi, S.-A. & Asad, S. Genome-scale reconstruction of the metabolic network in *Pseudomonas stutzeri* A1501. *Mol. Biosyst.* **11**, 3022–3032 (2015).
- Ataman, M., Hatzimanikatis, V., Betenbaugh, M. J. & Bentley, W. E. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Curr. Opin. Biotechnol.* **36**, 176–182 (2015).
- Salvy, P. et al. pyTFA and matTFA: A Python package and a Matlab toolbox for thermodynamics-based flux analysis. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bty499> (2018).
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) (1990).
- Kanehisa, M. & Sato, Y. KEGG Mapper for inferring cellular functions from protein sequences. *Protein Sci.* <https://doi.org/10.1002/pro.3711> (2019).
- Ravikrishnan, A. & Raman, K. Critical assessment of genome-scale metabolic networks: The need for a unified standard. *Brief. Bioinform.* <https://doi.org/10.1093/bib/bbv003> (2015).
- Ponce-De-Leon, M., Calle-Espinosa, J., Peretó, J. & Montero, F. Consistency analysis of genome-scale models of bacterial metabolism: a metamodel approach. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0143626> (2015).
- Monk, J. M. et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.3956> (2017).
- Osterman, A. & Overbeek, R. Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.* [https://doi.org/10.1016/S1367-5931\(03\)00027-9](https://doi.org/10.1016/S1367-5931(03)00027-9) (2003).
- Gerhardt, P., Wood, W. A., Krieg, N. R., Murray, R. Methods for general and molecular bacteriology (American Society for Microbiology, 1994).
- Langmead and Steven L Salzberg. Bowtie2. *Nat. Methods* <https://doi.org/10.1038/nmeth.1923> (2013).
- Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btp352> (2009).
- Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btu638> (2015).

43. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btp616> (2009).
44. Conesa, A. et al. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bti610> (2005).
45. Gowda, H. et al. Interactive XCMS online: Simplifying advanced metabolomic data processing and subsequent statistical analyses. *Anal. Chem.* <https://doi.org/10.1021/ac500734c> (2014).
46. Mylonas, R. et al. X-rank: A robust algorithm for small molecule identification using tandem mass spectrometry. *Anal. Chem.* <https://doi.org/10.1021/ac900954d> (2009).
47. Tautenhahn, R. et al. An accelerated workflow for untargeted metabolomics using the METLIN database. *Nat. Biotechnol.* <https://doi.org/10.1038/nbt.2348> (2012).
48. Chiappino-Pepe, A., Tymoshenko, S., Ataman, M., Soldati-Favre, D. & Hatzimanikatis, V. Bioenergetics-based modeling of *Plasmodium falciparum* metabolism reveals its essential genes, nutritional requirements, and thermodynamic bottlenecks. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1005397> (2017).
49. Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat. Protoc.* **5**, 93–121 (2010).
50. Jankowski, M. D., Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* <https://doi.org/10.1529/biophysj.107.124784> (2008).
51. Schnoes, A. M., Brown, S. D., Dodevski, I. & Babbitt, P. C. Annotation error in public databases: Misannotation of molecular function in enzyme superfamilies. *PLoS Comput. Biol.* <https://doi.org/10.1371/journal.pcbi.1000605> (2009).
52. Lieven, C. et al. Memote: A community driven effort towards a standardized genome-scale metabolic model test suite. *Biorxiv* (2018).
53. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkw1092> (2017).

## ACKNOWLEDGEMENTS

This work was supported by SystemsX.ch, and evaluated by the Swiss National Science Foundation, within grant 2013/158 (Design and Systems Biology of Functional Microbial Landscapes 'MicroScapesX'). We thank Dr. Ljubisa Miskovic for constructive criticism of the manuscript and for his careful proof-reading, and Mr. Manupriyam Dubey and his student group in 'Experimental Design 2014' for providing us with the Biolog data.

## AUTHOR CONTRIBUTIONS

N.H. and J.V.M. conceived and developed the presented idea. N.H. and A.C.P. developed the metabolic model. N.H. and V.S. performed the cell culture experiments. V.S. and M.M. performed the RNAseq. H.G.A., F.M., and J.I. performed the metabolomics. N.H., M.M., H.G.A., and J.I. analyzed the omics data. N.H. and V.P. did the integration and performed computations. N.H. and J.V.M. wrote the original draft manuscript. All authors discussed the results and contributed to the final manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41540-019-0121-4>.

**Correspondence** and requests for materials should be addressed to N.H.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, 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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020