

REVIEW

Open Access



# Current status and applications of genome-scale metabolic models

Changdai Gu<sup>1†</sup>, Gi Bae Kim<sup>1†</sup>, Won Jun Kim<sup>1</sup>, Hyun Uk Kim<sup>2,3,4\*</sup> and Sang Yup Lee<sup>1,3,4\*</sup>

## Abstract

Genome-scale metabolic models (GEMs) computationally describe gene-protein-reaction associations for entire metabolic genes in an organism, and can be simulated to predict metabolic fluxes for various systems-level metabolic studies. Since the first GEM for *Haemophilus influenzae* was reported in 1999, advances have been made to develop and simulate GEMs for an increasing number of organisms across bacteria, archaea, and eukarya. Here, we review current reconstructed GEMs and discuss their applications, including strain development for chemicals and materials production, drug targeting in pathogens, prediction of enzyme functions, pan-reactome analysis, modeling interactions among multiple cells or organisms, and understanding human diseases.

## Introduction

Since the first genome-scale metabolic model (GEM) of *Haemophilus influenzae* RD was reported in 1999 [1], GEM reconstruction has been established as one of the major modeling approaches for systems-level metabolic studies. A GEM computationally describes a whole set of stoichiometry-based, mass-balanced metabolic reactions in an organism using gene-protein-reaction (GPR) associations that are formulated on the basis of genome annotation data and experimentally obtained information [2, 3]. Importantly, the GEM allows the prediction of metabolic flux values for an entire set of metabolic reactions using optimization techniques, such as flux balance analysis

(FBA), which uses linear programming [4]. GEM also serves as a platform for the integration and analysis of various types of data such as omics and kinetic data [5–7]. As the techniques for genome sequencing and relevant omics analyses continue to evolve, the quality and application scopes of GEMs have also expanded accordingly, and together they have contributed to better understanding of metabolism in various organisms. Starting with GEMs of model organisms, including *Escherichia coli* [8] and *Saccharomyces cerevisiae* [9], GEMs of various microorganisms and also multicellular organisms, such as humans [10] and plant cells [11], have been reconstructed.

Such progress in the reconstruction of GEMs has made it possible to construct a wide range of metabolic studies by generating model-driven hypotheses and implementing various context-specific simulations [12]. Relevant applications that have benefited from advances in the use of GEMs include, but are not limited to, strain development for the production of bio-based chemicals and materials, drug targeting in pathogens, the prediction of enzyme functions, pan-reactome analysis, modeling interactions among multiple cells or organisms, and understanding human diseases. Applications of GEMs are expected to expand further in coming years. To this end, we comprehensively review the current status and applications of GEMs reconstructed for diverse organisms belonging to bacteria, archaea, and eukarya (Fig. 1, Additional file 1, and Additional file 2). By discussing a broad range of studies involving GEMs, we show how GEMs can help us to gain novel biological insights beyond those provided by genome sciences and how they can help to develop biotechnological applications.

## Current status of reconstructed genome-scale metabolic models

As of February 2019, GEMs have been reconstructed for 6239 organisms (5897 bacteria, 127 archaea, and 215 eukaryotes), either manually or by using automatic GEM reconstruction tools that are discussed below (Fig. 1, Additional file 1, and Additional file 2). A total of 183 organisms (113 bacteria, 10 archaea, and 60 eukaryotes)

\* Correspondence: [ehukim@kaist.ac.kr](mailto:ehukim@kaist.ac.kr); [leesy@kaist.ac.kr](mailto:leesy@kaist.ac.kr)

† Changdai Gu and Gi Bae Kim contributed equally to this work.

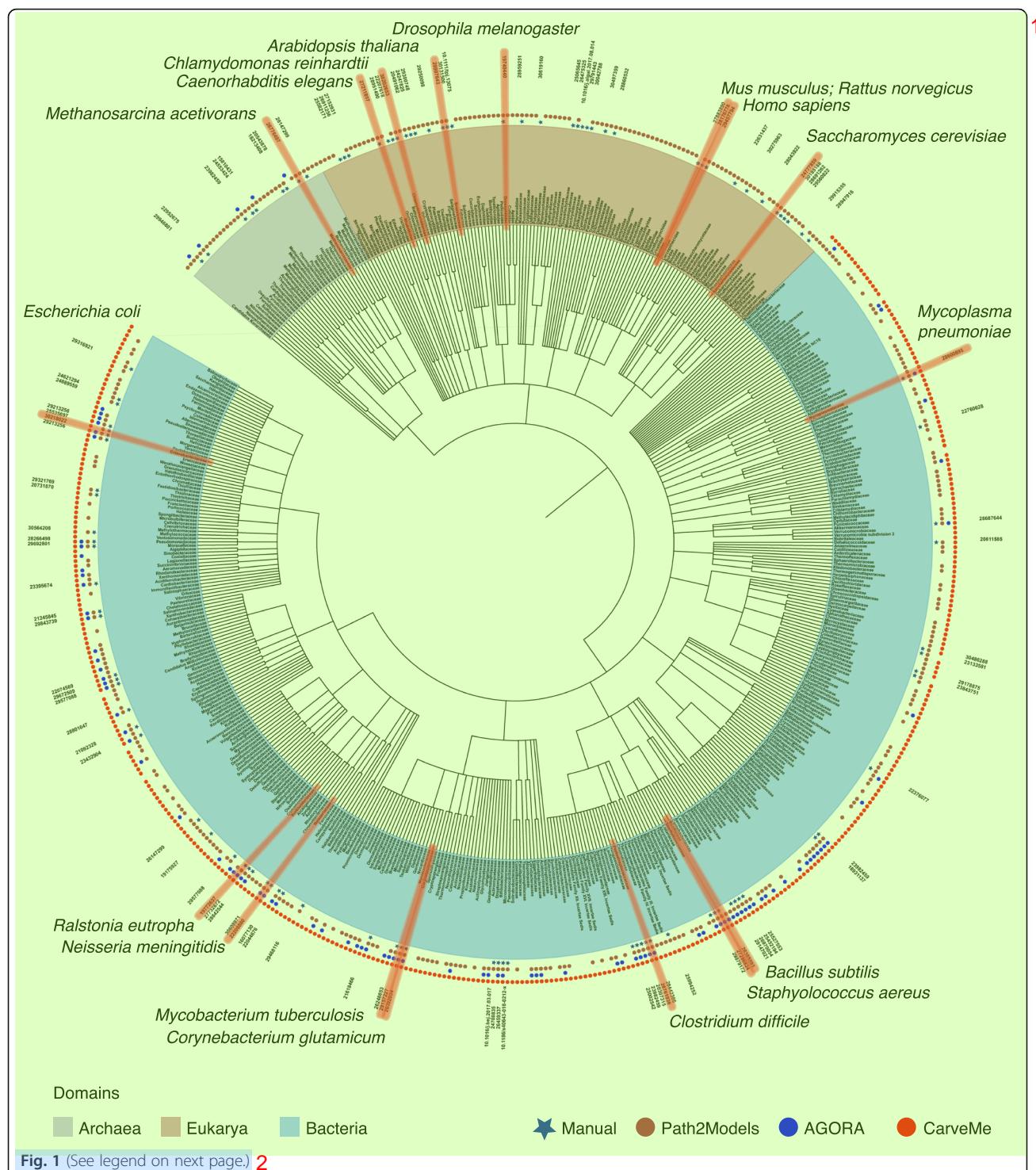
<sup>2</sup>Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Systems Biology and Medicine Laboratory, KAIST, Daejeon 34141, Republic of Korea

<sup>1</sup>Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Metabolic and Biomolecular Engineering National Research Laboratory, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea

Full list of author information is available at the end of the article



© The Author(s). 2019 **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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 Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.



(See figure on previous page.)

**Fig. 1** A phylogenetic tree of all of the GEMs reconstructed to date at the family level. GEMs for 434, 40, and 117 taxonomic families of bacteria (*light blue*), archaea (*light purple*), and eukarya (*pink*), respectively, are marked in the phylogenetic tree. Organism names are labeled with circles of different colors outside the circular phylogenetic tree, depending on the development methods used: manual, Path2Models [13], AGORA [14], and CarveMe [15]. For manually reconstructed GEMs, the relevant PubMed identifier (PMID) or digital object identifier (DOI) for the latest GEM version is additionally indicated. The phylogenetic tree was prepared as follows. First, organism names were collected from BioModels for the GEMs from Path2Models, from Virtual Metabolic Human (VMH) for AGORA models, and from a GitHub repository ([https://github.com/cdanielmachado/embl\\_gems/blob/master/model\\_list.tsv](https://github.com/cdanielmachado/embl_gems/blob/master/model_list.tsv)) for CarveMe. Next, National Center for Biotechnology Information (NCBI) taxids at the species level and taxonomic lineages for all of the organisms subjected to the GEM reconstruction were obtained from a dataset available (as of May 14, 2019) at the NCBI Taxonomy FTP (<ftp://ftp.ncbi.nih.gov/pub/taxonomy/taxdump.tar.gz>). Finally, a Newick file for all of the organisms with taxids was subsequently generated using an in-house Python script at the family level, and this file was used to create a phylogenetic tree using iTOL (<https://itol.embl.de/>) [16]. A phylogenetic tree of GEMs at the species level is available as Additional file 1. A full list of organisms subjected to the GEM reconstruction and preparation of phylogenetic trees is available as Additional file 2

have been subjected to manual GEM reconstruction. We 1 first discuss the current status of GEMs built for model organisms that are scientifically, industrially, and/or medically important, and then cover computational resources for GEM reconstruction.

### High-quality GEMs for model organisms 2

The GEMs for model organisms that have high scientific, 3 industrial, and/or medical values have been updated several times since their initial reconstruction as more relevant biological information became available over the years. GEMs are often updated by adopting up-to-date experimental information on GPR associations and cell growth under various conditions (such as in gene knockouts or when different carbon sources are used), and by resolving issues such as incorrect GPR associations and different database identifiers for the same metabolite. The resulting GEMs serve as an excellent knowledgebase for studying the metabolism of the target organisms and are capable of predicting the organism's various biological capabilities. As a result, high-quality GEMs of several model organisms reveal the history, rationale behind, and future directions of GEM development. Furthermore, they serve as good reference models for developing GEMs of other related organisms.

### Bacteria 4

**Escherichia coli** Being a model organism for bacterial genetics, the Gram-negative bacterium *Escherichia coli* has been subjected to genome-scale metabolic reconstruction campaigns for almost two decades. The first *E. coli* GEM, iJE660 [8], was reported in 2000 soon after the first release of the genome sequence of *E. coli* K-12 MG1655 [17]. The iJE660 model has subsequently been updated in terms of the coverage of GPR associations and prediction capacities, officially at least four times [18–21]. The most recent version, iML1515, contains information on 1515 open reading frames, twice the number of open reading frames incorporated in the original iJE660 model. The iML1515 model shows 93.4% accuracy for gene

essentiality simulation under minimal media containing 6 16 different carbon sources such as glucose, xylose, and acetic acid. Importantly, iML1515 was tailored in various ways to extract the most relevant knowledge from a large volume of biological data. For example, iML1515-ROS has additional reactions associated with the generation of reactive oxygen species (ROS) and is useful for antibiotics design; iML976, a subset of iML1515, only contains information on metabolic genes that are shared by over 1000 *E. coli* strains and provides understanding of the core and accessory metabolic capacities of *E. coli* strains, especially clinical ones; and context-specific GEMs, by using proteome data from cells grown under specific conditions, reduce the false-positive predictions [21]. As a 'model' GEM, iML1515 shows how a GEM can accurately predict cellular metabolic and physiological states, and is expected to evolve further as new data become available.

**Bacillus subtilis** *Bacillus subtilis* is a representative 7 Gram-positive bacterium that has value to industrial biotechnology in the production of various enzymes and proteins [22, 23]. GEMs reconstructed for *B. subtilis* include iYO844 [24], a GEM by Goelzer et al. [25], iBsU1103 [26], iBsU1103V2 [27], iBsU1147 [28], and iBsU1144 [29]. The latest version, iBsU1144, built on the basis of the re-annotated genome information [30], was developed by incorporating thermodynamic information on the standard molar Gibbs free energy change for each reaction, in order to improve the accuracy and consistency of the reversibility of intracellular reactions. In one application study, iBsU1144 was employed to identify the effects of oxygen transfer rates (i.e., low, medium, and high oxygen transfer rates) on the production of serine alkaline protease and recombinant proteins using *B. subtilis* in silico. The *B. subtilis* GEMs will serve as important reference models for other Gram-positive bacteria.

**Mycobacterium tuberculosis** In the fight against micro- 8bial pathogens, understanding their condition-specific metabolism (e.g., metabolism at a specific lifecycle point) at a systems level is very important for the identification of

effective drug targets [31]. *Mycobacterium tuberculosis*, a bacterial pathogen that causes tuberculosis in humans, is one of the microbial pathogens that have been studied the most over the past 10 years using GEMs [32–38]. The most recent version of the GEM iEK1101 of *M. tuberculosis*, H37Rv, was developed by standardizing and combining biological information from previously released GEMs [38]. Upon reconstruction, iEK1101 was used to provide understanding of this pathogen's metabolic status under an in vivo hypoxic condition, which replicates a pathogenic state, and also in an in vitro drug-testing condition. Comparing the predicted metabolic flux distributions in the two conditions allowed evaluation of the pathogen's metabolic responses to antibiotic pressures [38]. Besides developing independent GEMs, a GEM of *M. tuberculosis* was integrated with a GEM of human alveolar macrophages to study host-pathogen interactions [39]. Development of systematic drug-targeting methods using GEMs of *M. tuberculosis* continues to be an active research area.

## Archaea 2

**Methanosarcina acetivorans** GEM reconstruction studies have mainly focused on prokaryotes and eukaryotes, but GEMs have also been built for a few archaea, such as *Methanosarcina acetivorans*, a methanogenic archaeal species which lives in a marine environment [40–43]. GEMs of *M. acetivorans* contain information on the methanogenesis pathway, the most representative metabolism of methanogens [44]. The iMAC868 model, the latest version of a *M. acetivorans* GEM [42], was established by integrating two previous models, iVS941 [40] and iMB745 [41]. iMAC868 was curated to represent a thermodynamically feasible methanogenesis reversal pathway that co-utilizes methane and bicarbonate, among other updates made in the GEM [42]. Another recent GEM of *M. acetivorans*, iST807 [43], was also updated on the basis of the previous version, iMB745 [41], in order to consider the effects of regulators on metabolic pathways in media containing different substrates. For example, in addition to the newly added metabolic pathways, tRNA-charging was explicitly incorporated into iST807, thereby allowing characterization of the effects of the differential expression of tRNA genes on metabolic fluxes. The GEMs for archaea will serve as a useful resource for metabolic studies on unusual characteristics of archaea in a wide range of habitats, including extreme environments and the human gut.

## Eukarya 4

**Saccharomyces cerevisiae** As the most representative eukaryotic microorganism, *S. cerevisiae* was the first eukaryotic organism to have its genome sequenced [45]. In addition, it was the first eukaryote for which a GEM

was reconstructed [9]. Since the first *S. cerevisiae* GEM was released in 2003 [9], the GEMs for this microorganism have been updated by several different research groups [46–51]. However, the resulting different versions of *S. cerevisiae* GEMs were found to have inconsistent annotations, which hampered their comparison and further GEM upgrades. To address this inconsistency problem, a consensus metabolic network, Yeast 1, was reconstructed through an international collaborative effort [52]. Although Yeast 1 was a comprehensive metabolic network, it could not be simulated for flux predictions; constraint-based simulation became possible with a later version, Yeast 4 [53]. Yeast 1 has been upgraded to the latest version, Yeast 7, in the past few years by incorporating new biological information and by correcting critical modeling errors, such as the removal of thermodynamically infeasible reactions [53–56]. Very recently, Yeast 7.Fe [57] was extended from Yeast 7.6 [56] by including additional information on iron metabolism, which had not been properly considered in the previous GEMs. The Yeast 7.Fe now allows estimation of the optimal turnover rate of iron cofactors and more rigorous examination of metabolism. The GEMs of *S. cerevisiae* will continue to serve as reference models for eukaryotic microorganisms.

**Chlamydomonas reinhardtii** The green microalgae *Chlamydomonas reinhardtii* has served as a model organism in studies of photosynthesis, phototaxis, cell motility, and bioenergy production [58, 59]. Owing to its biological importance, there has been much interest in reconstructing a GEM for *C. reinhardtii*. A total of six GEMs have been developed for *C. reinhardtii* to examine microalgal behaviors at the systems level, including iAM303 [60], iRC1080 [61], AlgaGEM [62], iCre1355 [63], a GEM by Winck et al. [64], and a GEM by Mora Salguero et al. [65]. The latest version by Mora Salguero et al. [65] allows dynamic simulations by including kinetic information on the effects of acetate as a nutrient on the growth rate at varying CO<sub>2</sub> levels. Overall, the dynamic simulation of the *C. reinhardtii* GEM was able to predict cellular responses to the environmental changes accurately. GEMs for a greater number of microalgal species have been reconstructed for applications in biotechnology, including the production of lipids and secondary metabolites [66].

**Caenorhabditis elegans** The nematode *Caenorhabditis elegans* has been employed as an established eukaryotic model organism in various studies, including work on aging [67], molecular and developmental biology [68], and nutrition [69]. The biological importance of *C. elegans* has led to multiple GEM reconstructions, including iCEL1273 [70], ElegCyc [71], and CeCon [72]. In 2017, the

WormJam Community was founded to develop a consensus GEM of *C. elegans* by merging and reconciling the existing GEMs that had been developed by different research groups [73]. The three different *C. elegans* GEMs were merged to give a draft consensus GEM, and manual curation of the GEM was conducted to incorporate additional metabolic characteristics of *C. elegans*. The manual curation process, which was based on biological insights and over 40 metabolomics studies, led to the correction of errors in multiple metabolic pathways (i.e., glycogen metabolism, sphingolipid metabolism, and the biosynthesis and degradation of fatty acids, such as branched chain fatty acids) and the addition of new metabolic pathways (i.e., biosynthesis of maradolipids and ascaroside). The WormJam consensus GEM, the most accurate *C. elegans* GEM to date, now provides better biological insights into *C. elegans* physiology.

**Arabidopsis thaliana** *Arabidopsis thaliana*, a model organism for plants, has also been an attractive target for extensive metabolic reconstruction studies, resulting in the development of four different GEMs: a GEM by Poolman et al. [74], AraGEM [11], a GEM by Mintz-Oron et al. [75], and a GEM by Cheung et al. [76]. Among these four GEMs, the latest version by Cheung et al. [76] was reconstructed to predict more accurate fluxes in the heterotrophic metabolism of *A. thaliana*, in particular by considering the transport costs associated with nutrient uptake and protein translocation between organelles and the maintenance costs for ATP and NADPH. By including information on these energy costs in the model, the metabolic flux distributions calculated using the updated GEM became more consistent with those obtained by <sup>13</sup>C-metabolic flux analysis. Despite greater biological complexity, plant GEMs are beginning to be used more extensively to understand the metabolism of plants [77].

**Homo sapiens** The availability of human GEMs has contributed to a better understanding of the biological mechanisms behind various diseases and to the design of appropriate disease treatments. Since the release of Recon 1, the first generic human GEM in 2007 [10], the Recon series has gone through several important updates, including the incorporation of additional biological information and the correction of various modeling errors [78–81]. Recon 2 M.2 is the version in which a framework for gene-transcript-protein-reaction associations (GeTPRA) was deployed to generate metabolic reactions by considering the effects of alternative splicing of metabolic genes (i.e., both principal and non-principal transcripts) [80]. Recon3D is the latest version and contains the most extensive human GPR associations and structural information on metabolites and enzymes. Recon3D can be used as

a resource for many biomedical applications, including the characterization of disease-associated mutations and metabolic responses to drugs [81].

The Human Metabolic Reaction (HMR) series [82] is another generic human GEM series, which contains information on subcellular localization and tissue-specific gene expression, both mainly from the Human Protein Atlas database [83, 84]. In comparison with the Recon series, the HMR series has more comprehensive information on fatty acid metabolism that has been manually curated. The HMR series led to the generation of several cell-type-specific GEMs, such as *iAdipocytes1809* [82], *iHepatocytes2322* [85], and *iMyocyte2419* [86], which were used to study obesity, non-alcoholic fatty liver disease (NAFLD), and diabetes, respectively. Both the Recon and the HMR series will serve as useful resources for various biomedical studies, as discussed below.

## Computational resources for automatic GEM reconstruction

Manual reconstruction of GEMs is a time-consuming procedure [3], in which a large number of GPR associations and many other sources of data and information must be considered. To address this challenge, several software programs for automatic GEM reconstruction have been developed (Table 1). A significant part of the GEM reconstruction procedure has now been automated, including, but not limited to, the annotation of genome sequence, the generation of a set of GPR associations unique to a target organism, the prediction of reaction reversibility on the basis of thermodynamics, and enzyme localization. Three independent studies involving the high-throughput generation of GEMs, namely Path2Models [13], AGORA [14], and CarveMe [15], led to the reconstruction of more than 6000 GEMs (Table 2). CarveMe led to the generation of the greatest number of GEMs, especially for bacteria, followed by Path2Models and AGORA (Fig. 1). CarveMe is a computational pipeline that automatically converts a manually curated reaction dataset (known as the ‘universal model’) into a target-organism-specific GEM. CarveMe was used to reconstruct 5587 bacterial GEMs, which corresponds to 91.8% of the bacterial GEMs reconstructed (Fig. 1). Path2Models, the first large-scale GEM reconstruction project, allowed reconstruction of GEMs for 2606 organisms, which are accessible at the BioModels database [98]. GEMs developed through Path2Models cover 20.1, 98.4, and 88.8% of all the GEMs reconstructed for bacteria, archaea, and eukaryotes, respectively (Fig. 1). AGORA models include the GEMs of 773 members of the human gut microbiota, which were prepared in a semi-automatic manner using the online GEM reconstruction platforms Model SEED [93] and KBase [104]. As of February 2019, 818 AGORA models are accessible at the Virtual Metabolic

**Table 1** Software programs for the reconstruction of GEMs 1

Tool	Language	Graphical user interface (GUI) available?	Source database for metabolic reactions	Use as a reference model?	Gap-filling	Eukaryote modeling	Simulation ready	Reference <span style="color:red">2</span>
AuReMe	Python	No	KEGG, BiGG Models, MetaCyc	No	Yes	Yes	Yes	[87]
AutoKEGGRec	Matlab	No	KEGG	No	No	No	No	[88]
CarveMe	Python	No	BiGG Models	Yes (universal model)	Yes	No	Yes	[15]
CoReCo	Python	No	KEGG	No	Yes	Yes	Yes	[89]
FAME	Python	Yes	KEGG	Yes	No	No	Yes	[90]
merlin	Java	Yes	KEGG, MetaCyc, UniProtKB, TCDB	No	No	Yes	No	[91]
MetaDraft	Python	Yes	BiGG Models	Yes	No	No	Yes	[92]
Model SEED	Web	Yes	In-house reaction database	Yes	Yes	Yes (only plants)	Yes	[93]
Pathway Tools	Python, Lisp	Yes	Pathway/Genome Database (PGDB), MetaCyc	No	Yes	Yes	Yes	[94]
RAVEN 2.0	Matlab	No	KEGG, MetaCyc	Yes	Yes	Yes	Yes	[95]
SuBliMinal Toolbox	Java	No	KEGG, MetaCyc	No	No	No	Yes	[96]

Human database [103]. Both manually curated high-quality 3 GEMs and automatically reconstructed GEMs are now available at several databases, as summarized in Table 2.

As these tools and strategies for GEM reconstruction 4 are advancing rapidly, an increasing number of GEMs for many different organisms, including those of research interest, have become available. Nevertheless, challenges remain with the automatic GEM reconstruction. The most urgent challenges are to evaluate the quality of an automatically generated draft GEM and to automate the refinement procedure [105]. A draft GEM usually includes inaccurate reactions and GPR associations; for example, the biomass generation reaction is not tailored for a target organism, and reactions are often incorrectly constrained (e.g., there are no constraints or incorrect reaction reversibility in the draft GEM). Solutions exist to meet this

challenge. The quality of the draft GEMs can be evaluated by using a set of task functions that are relevant to a target organism; this feature is now addressed by the software program memote (which stands for ‘metabolic model tests’) [106], but in an organism-independent manner. Some established algorithms for refining draft GEMs, such as gap-filling [107, 108] and the integration of experimental data (such as those from cultivation experiments under various conditions) [5], can be integrated with the GEM reconstruction tools to allow (semi-)automatic refinement of the draft GEMs. For complex problems such as the formulation of an organism-specific biomass generation reaction, manual refinement assistance could be provided in the GEM reconstruction tool, for example, by providing the biomass generation equations of biologically related organisms. In the future, it is expected that this step will 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761</span

also be automated, for example, by automatically extracting information on a target organism-specific biomass composition and even GPR associations from the literature using text mining techniques.

## Applications of GEMs 2

GEMs of various organisms have been widely employed in scientific discovery as well as in various industrial and medical applications [7, 109, 110]. Importantly, the development of omics data integration methods for GEMs resulted in the expansion of the application scope of GEMs [111] by tailoring a GEM according to specific conditions of interest. Relevant omics data integration algorithms [5, 112, 113] include GIMME [114], iMAT [115], MBA [116], INIT [117], mCADRE [118], tINIT [119], CORDA [120], and TIMBR [121]. The integration of omics data with GEMs is particularly important for modeling multicellular organisms, such as human and plants, because the generic GEMs that are available for these organisms need to be transformed into context-specific GEMs. Generic GEMs do not address condition-specific metabolism because they have information on all metabolic genes regardless of their expression levels in a specific tissue or cell type. Relevant studies involving context-specific GEMs include the prediction of condition-specific (e.g., specific to life cycle stage or cultivation environment) drug targets in pathogens, the prediction of host-pathogen metabolic interactions, and the characterization of the reprogrammed metabolism of liver cancer stem cells (LCSCs) and the endothelium of sepsis patients, which are all discussed below. Further details of various omics data integration methods have been thoroughly discussed elsewhere [5, 112, 113].

## Production of chemicals and materials 4

GEMs have long been used to predict targets for effective gene manipulation (by knockout or through the up- or downregulation of gene expression, for example) for the enhanced microbial production of chemicals and materials. Notably, GEMs have been applied to redesign aspects of the metabolism of both bacteria and eukaryotes in order to produce an increasing number of chemicals and materials. A recent example is the enhanced production of aromatic polymers involving D-phenyllactic acid as a monomer (e.g., poly(3-hydroxybutyrate-co-D-phenyllactate)) using metabolically engineered *E. coli* strains (Fig. 2a) [122]. Direct production of D-phenyllactic acid from glucose was first attempted by implementing flux response analysis of the *E. coli* GEM iJO1366 [20] to examine the effects of engineering central and aromatic amino acid biosynthetic reactions on D-phenyllactic acid production. Additional knockouts of *tyrB* and *aspC* genes in an engineered *E. coli* base strain (XB201T) producing 0.55 g/L of D-phenyllactic acid successfully increased D-phenyllactic acid production to 1.62 g/L. Fed-batch fermentation of the

final strain produced 13.9 g/L of poly(61.9 mol% 3-hydroxybutyrate-co-38.1 mol% D-phenyllactate).

In another example involving *Yarrowia lipolytica*, an eukaryotic microorganism known to accumulate large amounts of lipids [130], its GEM was used to improve the production of dodecanedioic acid [123] (Fig. 2b). First, a GEM of *Y. lipolytica*, iYLI647, was newly reconstructed and employed to find target reactions that can lead to the enhanced production of dodecanedioic acid [123]. For this, several *in silico* strain design methods were implemented using iYLI647, including (1) flux activity analysis (a method examining the effects of changes in individual reaction fluxes on a target chemical production rate) [131] and the transcriptomics-based strain optimization tool (tSOT) [132], both of which identify overexpression targets; (2) genetic design by local search (GDLS) [133], which is used to identify knockout targets; and (3) cofactor modification analysis (CMA) [134], which identifies cofactor modification targets. Application of algorithms such as these to redesign a microbial strain's metabolism allows the identification of more robust gene manipulation targets, and is becoming an essential practice in metabolic engineering.

## Drug targeting in pathogens 8

Another important application of GEMs is to predict the viability of an organism under a given condition. This simulation approach has been utilized to suggest metabolic drug targets whose inhibition can effectively kill a pathogen. The GEM of a target pathogen can be used to predict essential genes (or reactions) [32, 135] and essential metabolites [136, 137], each of which can lead to a different drug discovery strategy. A recent study using GEMs suggested drug targets in *Plasmodium falciparum* that are specific to the life cycle stage of the malaria-causing pathogen [124]. *P. falciparum* goes through a complex life cycle to reproduce itself [138]. Because each life cycle stage has a different metabolic network structure, it is likely that different drug targets can be found for each stage. Thus, the stage-specific GEMs of *P. falciparum* were reconstructed [124]. Integration of a generic GEM with stage-specific transcriptome and physiology data, such as stage-specific growth rates and metabolite secretion rates, led to five stage-specific models that represent the trophozoite, schizont, early gametocyte, late gametocyte, and ookinete (Fig. 2c). Gene essentiality analysis of the stage-specific GEMs showed 71.2% accuracy in comparison with experimentally characterized drug targets (42 out of 59 drug targets). The prediction outcome indicates that the quality of the *P. falciparum* GEM needs to be improved further. In addition, new drug targets beyond these 59 targets need to be identified, especially novel targets that are effective in the proliferative and late gametocyte stages. Life-cycle-stage-specific modeling and simulation

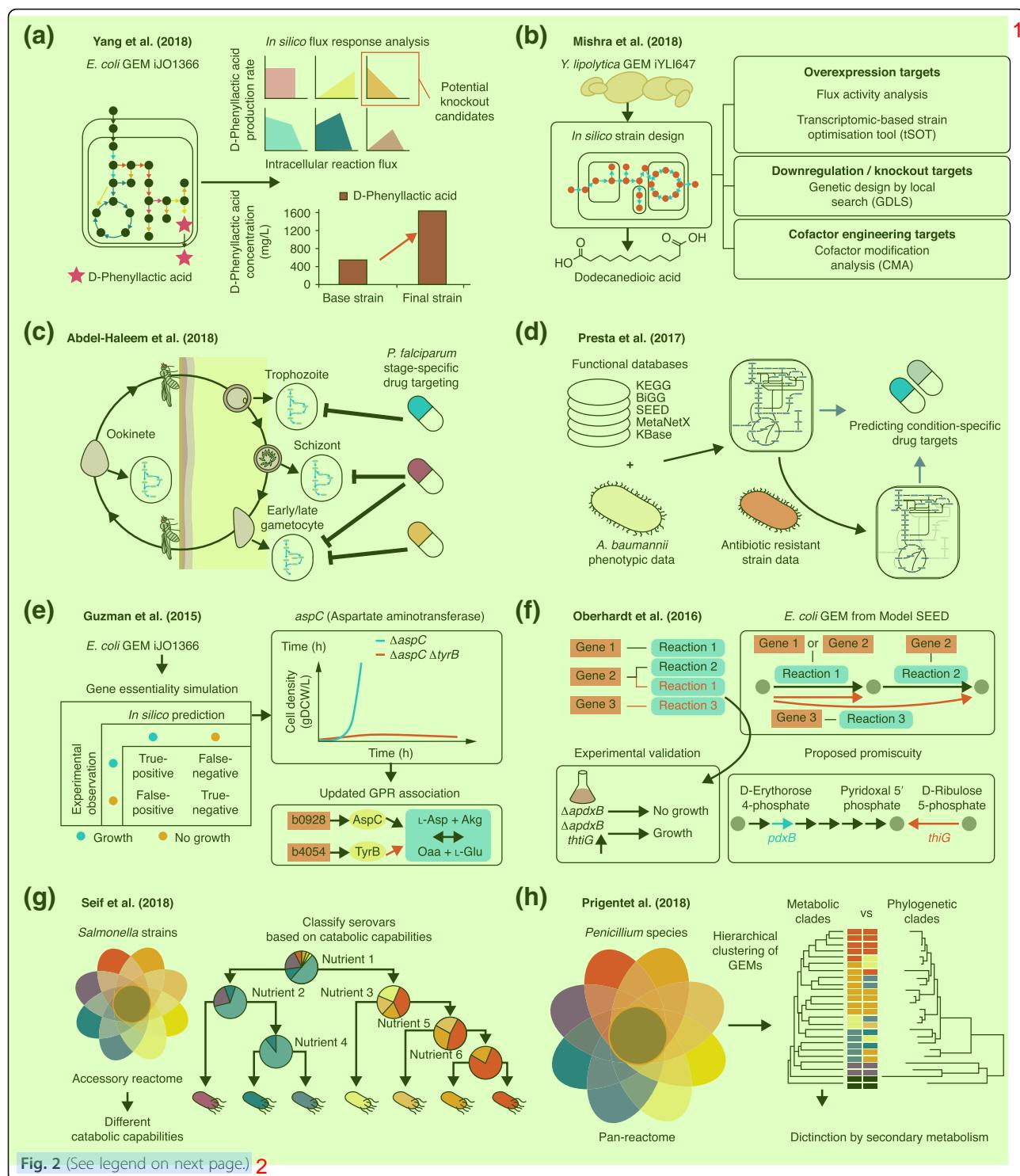


Fig. 2 (See legend on next page.) 2

(See figure on previous page.)

**Fig. 2** Applications of GEMs for the production of chemicals and materials, drug targeting in pathogens, the prediction of enzyme functions, and pan-reactome analysis. **a**, Flux-response analysis using the *Escherichia coli* GEM iJO1366 [20] has been used to identify gene manipulation targets for the enhanced production of a monomer of an aromatic polymer D-phenyllactic acid in *E. coli* [122]. The final strain has two additional genes (*tyrB* and *aspC*) knocked out from an *E. coli* base strain XB201T that expresses AroG<sup>fbr</sup>, PheA<sup>fbr</sup>, and FldH. The final strain produced 1.62 g/L of D-phenyllactic acid, much more than the 0.55 g/L produced by the base strain. **b**, Reconstruction of the *Yarrowia lipolytica* GEM iYLI647 and its application for the prediction of reaction engineering targets using four different in silico strain-design strategies [123]. **c**, Identification of stage-specific antimalarial drug targets for *Plasmodium falciparum* using stage-specific GEMs that represent five different life cycle stages [124]. **d**, Reconstruction of a GEM for *Acinetobacter baumannii* using several databases and its application for the prediction of condition-specific drug targets to combat antibiotic-resistant *A. baumannii* [125]. **e**, Discovery of new isozyme functions for genes that have been shown to be nonessential in experiments but that were predicted to be essential in a gene essentiality simulation of the *E. coli* iJO1366 GEM (i.e., false-negative prediction for the *aspC* gene) [126]. It was found that tyrosine aminotransferase, which is encoded by *tyrB* (red line), can compensate for the loss of aspartate aminotransferase, encoded by *aspC*, which catalyzes the conversion of L-aspartate (L-Asp) and α-ketoglutarate (Akg) to oxaloacetate (Oaa) and L-glutamate (L-Glu). **gDCW/L** grams dry cell weight per liter. **f**, The PROmiscuity PrEdictOR (PROPER) method identifies promiscuous enzymes at a genome-scale in a target organism [127]. Promiscuous functions for all of the genes in the target organism (*E. coli*) were predicted using the PROPER method and an *E. coli* GEM from the Model SEED, which identified 98 alternative routes for the biosynthesis of various metabolites. For example, the product of the *thiG* gene in *E. coli* was newly found to biosynthesize pyridoxal 5'-phosphate, which is also known to be biosynthesized by the product of the *pdxB* gene. **g**, Analysis of the pan-reactome and accessory reactome of 410 *Salmonella* strains spanning 64 serovars using their respective GEMs [128]. Simulation of the GEMs under various nutrient conditions revealed the different catabolic capabilities of the different strains as well as their preferred growth environments. **h**, Analysis of the pan-reactome and accessory reactome of 24 *Penicillium* species by using their respective GEMs [129]. Hierarchical clustering of the 24 GEMs revealed additional insights into the biosynthetic pathways of secondary metabolites, which successfully differentiated the metabolic clades

approaches such as this will be important for drug targeting in other pathogens that exhibit different life stages, but this approach requires the acquisition of stage-specific omics data.

Condition-specific drug targeting using GEM has also been conducted for *Acinetobacter baumannii* [125], which is one of the six ESKAPE pathogens (*Enterococcus faecium*, *Staphylococcus aureus*, *Klebsiella pneumoniae*, *A. baumannii*, *Pseudomonas aeruginosa*, and *Enterobacter* spp.) associated with antimicrobial resistance [139]. Specifically, an updated version of the *A. baumannii* GEM, iLP844, was reconstructed and transformed into condition-specific GEMs by integration with transcriptome data. The transcriptome data were obtained from cells treated and untreated with colistin, one of the last-resort antibiotics against multi-drug-resistant pathogens [125]. Condition-specific drug targets were obtained by predicting genes that are exclusively essential in colistin-treated cells, and not homologous to any human genes, so that possible side-effects in the human body are avoided (Fig. 2d). It should be noted that similar approaches have also been applied to predict drug targets for diseased human cells such as cancers [119, 140]. Once a condition-specific GEM is built, drug targets can be predicted relatively easily. More demanding challenges are to validate the targets experimentally and to identify drugs that can effectively inhibit the predicted drug targets.

#### Prediction of enzyme functions 4

Rigorous analysis of the simulation results from a GEM also allows the identification of previously unidentified reactions or enzyme functions. In this context, two representative studies demonstrate how GEMs can be used to

unveil additional functions of an enzyme. One study focused on a set of genes that were shown in experiments to be nonessential, but that were predicted to be essential in a gene essentiality simulation of the *E. coli* GEM iJO1366 (i.e., there was a false-negative prediction of cell growth; Fig. 2e) [126]. Such false-negative predictions were thought to be caused by the presence of previously unidentified reactions that made a nonessential gene essential upon its knockout in silico. Among the ‘false-negative’ genes, *aspC*, *argD*, and *gltA* were selected for experimental validation because sequence homology analysis identified high-confidence candidate isozymes. Knockout of the genes encoding the potential isozymes revealed that tyrosine aminotransferase, which is encoded by *tyrB*, can compensate for the loss of aspartate aminotransferase, which is encoded by *aspC* (Fig. 2e). The same knockout approach also identified potential isozymes that could serve as alternative reaction enzymes for those encoded by *argD* and *gltA*.

In another study, a new method called PROmiscuity PrEdictOR (PROPER) [127] was developed to identify promiscuous enzymes in a target organism at the genome scale. For the implementation of PROPER, gene-similarity trees were built for all of the genes in *E. coli* using Position-Specific Iterated (PSI) BLAST, which show their homologous genes from the SEED database. The gene-similarity trees were used to generate a matrix that presents the primary and potential promiscuous functions (i.e., metabolic reactions) of *E. coli* enzymes encoded by the corresponding genes. Finally, ‘replacer’ genes were identified in the matrix, which have a potential promiscuous function that is identical to the primary function of another conditionally essential gene (‘target’ gene) in *E. coli*. A potential promiscuous function of a replacer gene can be validated if expression of this gene

can prevent cell death upon knockout of the target gene.<sup>1</sup> Among the target-replacer gene pairs predicted using the PROPER method and an *E. coli* GEM from Model SEED [93], the *pdxB-thiG* gene pair was experimentally validated (Fig. 2f). The *pdxB* gene is a conditionally essential gene involved in the biosynthesis of pyridoxal 5'-phosphate in *E. coli*, and served as a target gene in this study. The *thiG* gene, a replacer gene in this study, encodes 1-deoxy-D-xylulose 5-phosphate:thiol sulfurtransferase, an enzyme in the thiazole biosynthetic pathway, which was shown to biosynthesize pyridoxal 5'-phosphate without involving the enzyme encoded by *pdxB*.

These two studies have demonstrated that a high-quality GEM of a target organism allows the prediction of new enzyme functions and enzyme promiscuity, which is extremely useful because not all enzyme functions are experimentally validated.<sup>2</sup>

### Pan-reactome analysis<sup>3</sup>

Computational resources for high-throughput GEM reconstruction are now allowing the metabolic analysis of multiple organisms, including multiple strains of a single species [141, 142] or multiple species of a single genus [128, 129]. Analysis of a pan-reactome, an entire set of reactions, of biologically related organisms using GEMs provides a better understanding of the metabolic traits and lifestyles of these organisms. This concept was applied to study the metabolic traits of 410 *Salmonella* strains, spanning 64 serovars, by reconstructing a GEM for each strain [128]. The constructed pan-reactome revealed that the metabolic differences among the strains come from the accessory reactome, a set of reactions that are present in only some strains. These reactions were largely involved in alternative carbon metabolism and in cell wall or membrane metabolism. In particular, the strains could be distinguished on the basis of their different catabolic capabilities by analyzing their growth under various nutrient conditions in silico (Fig. 2g). Further investigation of serovar-specific catabolic capabilities helped to reveal the growth environments that are preferred by the *Salmonella* serovars and provided information about their evolution. Automatic GEM reconstruction tools have definitely contributed to pan-reactome analysis, and will continue to be applied to various groups of biologically related organisms of high scientific, industrial, and/or medical importance.<sup>4</sup>

Along the same lines, a pan-reactome analysis was conducted to provide information on the metabolic features of 24 *Penicillium* species that are well-known for the production of secondary metabolites [129]. Analysis of the 24 reconstructed GEMs revealed that most of the reactions involved in primary metabolism were conserved among these species. Subsequent hierarchical clustering of the 24 GEMs showed that the biosynthetic pathways for

secondary metabolites were the most distinctive pathways<sup>6</sup> in differentiating the metabolic clades, and that these pathways contributed to the genomic diversity of the 24 *Penicillium* species (Fig. 2h). Comparison of the metabolic clades with the phylogenetically classified clades, which were based on entire protein sequences for each species, demonstrated that stratifying species solely by using the phylogenetic tree could not fully explain the metabolic differences among the species. These representative studies demonstrate that the use of GEMs can bring about additional biological insights into a group of biologically related organisms. In the near future, an automated GEM refinement procedure using experimental data will much improve the quality of pan-reactome analysis, which at present is mainly conducted using draft GEMs.

### Modeling interactions among multiple cells or organisms<sup>7</sup>

The modeling of metabolic interactions among multiple<sup>8</sup> cells or organisms is also an important application of GEMs. This approach has been used for various studies of intermicrobial interactions, including the cross-feeding of microorganisms (or the exchange of metabolites between microorganisms) [92, 143, 144] and the evolutionary trajectory of microbial communities [145]. A recent study using GEMs has revealed that the secretion of costless metabolites contributes to the better growth of other interacting microorganisms, and ultimately to a greater taxonomic diversity in nature (e.g., in a nutrient-poor environment) [144]. Costless metabolites were defined as those that do not negatively affect the producing organism's fitness cost (i.e., growth rate) upon secretion [144]. The pairwise growth of the 24 microbial species examined in this study was simulated under various environmental conditions, involving different carbon sources and varying availability of oxygen, in order to examine the effects of the paired microorganisms' cross-feeding on their growth (Fig. 3a). The number of media that allowed the growth of at least one of the two microorganisms substantially increased if the exchange of costless metabolites between the microorganisms was allowed in the simulation. Interestingly, more frequent bidirectional exchanges between the two microorganisms and a greater number of costless metabolites were observed under anaerobic conditions than under aerobic conditions. These carefully designed in silico simulations using GEMs allowed the identification of new biological insights into intermicrobial interactions at a scale that would be difficult to replicate experimentally.<sup>9</sup>

In a study involving type 2 diabetes patients treated with the drug metformin [146] (Fig. 3b), the metabolism of four representative gut microbiota species, *Escherichia* sp., *Akkermansia muciniphila*, *Subdoligranulum variabile*, and *Intestinibacter bartlettii*, was examined using their respective GEMs. The GEMs were obtained from the AGORA models. Upon metformin treatment,

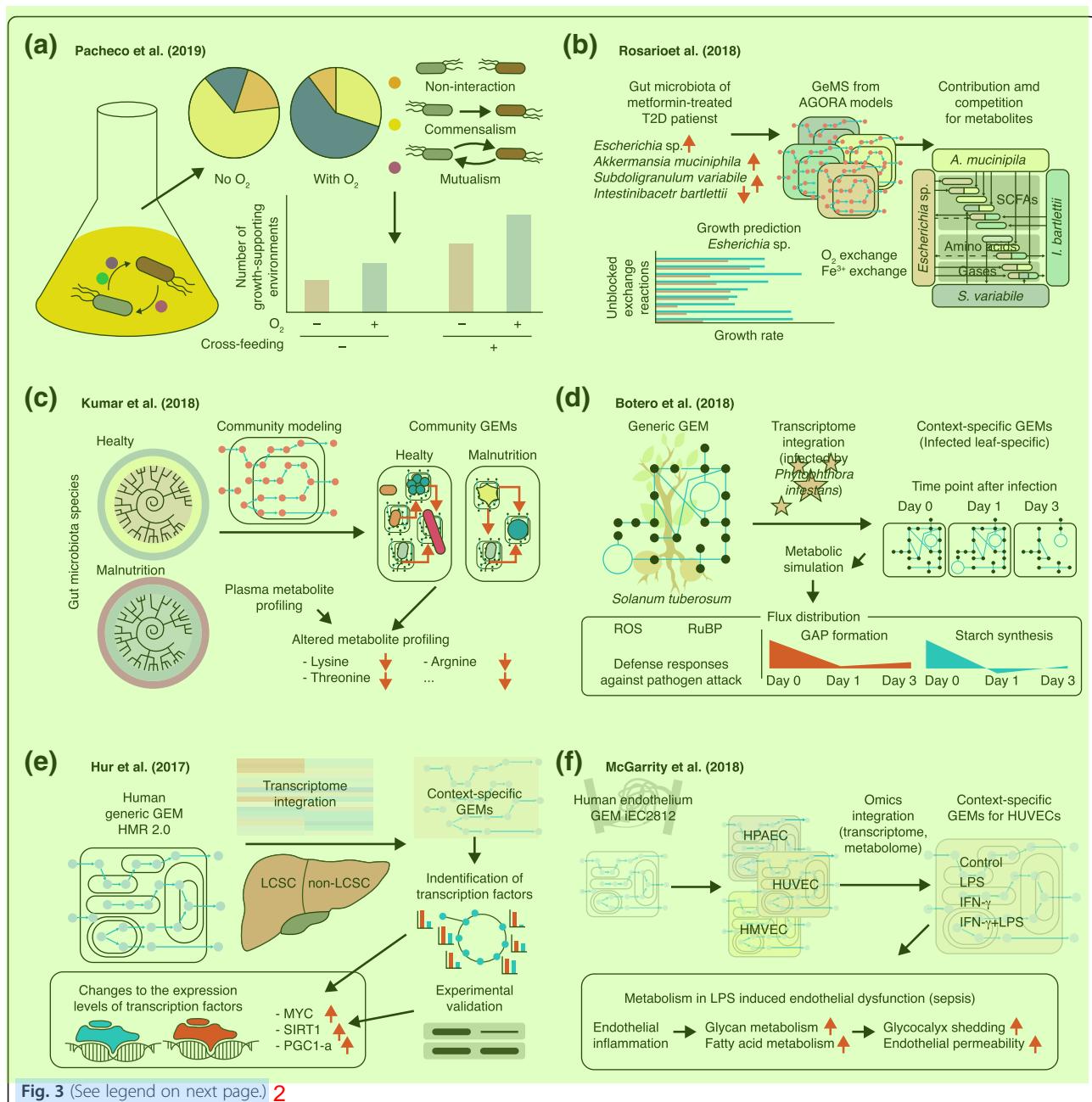


Fig. 3 (See legend on next page.) 2

(See figure on previous page.)

**Fig. 3** Applications of GEMs for interspecies metabolic interactions and understanding human diseases. **a** GEM-based simulation of the effects of costless metabolites (i.e., metabolites that have no effects on the producing organism's growth rate) secreted by at least one of two paired microorganisms on their growth under anaerobic and aerobic conditions [144]. The number of growth-supporting environments was increased as a result of cross-feeding. **b** Prediction of the metabolites (e.g., short-chain fatty acids [SCFAs]) required or produced by four representative gut microbiota species, *Escherichia* sp., *Akkermansia muciniphila*, *Subdoligranulum variabile*, and *Intestinibacter bartletti*, which are known to be affected by the type 2 diabetes (T2D) drug metformin [146]. **c** Prediction of the metabolites produced by gut microbiota species from malnourished children using community GEMs that describe the metabolism of multiple gut microbiota species [147]. The prediction results were consistent with the children's plasma metabolite profiles. **d** Prediction of the suppressed photosynthesis of a potato plant (*Solanum tuberosum*) upon infection by the plant pathogen *Phytophthora infestans*, which triggers the plant's defense responses against pathogen attack through oxygenation of ribulose1,5-bisphosphate (RuBP) and subsequently increases in the intracellular levels of reactive oxygen species (ROS) [148]. Formation of glyceraldehyde-3-phosphate (GAP) and starch were also decreased as a result of the infection. **e** Identification of metabolic differences between liver cancer stem cells (LCSCs) and non-LCSCs, and of the transcription factors responsible for the metabolic changes, by using GEMs integrated with transcriptome data [149]. **f** Characterization of the reprogrammed metabolism of the endothelium cells of sepsis patients using a human endothelium GEM, iEC2812 [150]. Context-specific GEMs were created using transcriptome and metabolome data obtained from human umbilical vein endothelial cells (HUECs) treated with lipopolysaccharide (LPS) and/or interferon- $\gamma$  (IFN- $\gamma$ ). Simulation of the context-specific GEMs indicated that increased glycan and fatty acid metabolism led to increased glycocalyx shedding and endothelial permeability where there was endothelial inflammation. HPAEC human pulmonary artery endothelial cell, HMVEC human microvascular endothelial cell

*Escherichia* sp., *A. muciniphila*, and *S. variabile* are known to be enriched in the gut, while *I. bartletti* is reported to decrease. In the simulation studies, contributing or competing bacteria were predicted through simulation of the GEMs for short-chain fatty acids (e.g., acetic and butyric acids), amino acids, and gases (e.g., H<sub>2</sub>, H<sub>2</sub>S, and CH<sub>3</sub>SH), all of which play important roles in both intermicrobial metabolic interactions and the regulation of human metabolism. For example, *Escherichia* sp. and *S. variabile* were predicted to contribute to the production of short-chain fatty acids under aerobic and anaerobic conditions. In addition, *Escherichia* sp. was shown to be least affected by the availability of intestinal nutrients. Covering a greater range of microorganisms and the metabolites exchanged among them will add more scientific value to GEM-based studies of a specific microbiota.

In this regard, another recent study also deserves attention. Kumar et al. [147] examined the production of metabolites by gut microbiota in children with malnutrition using GEMs for 58 representative gut microbiota species (Fig. 3c) [147]. The GEMs for 58 representative gut microbiota species were reconstructed using Model SEED and then used to examine metabolic differences (i.e., common and unique reactions) among these microorganisms. Community metabolic models (CMMs) were also reconstructed by integrating the GEMs of individual microbiota species according to the composition of the gut microbiota; each CMM represents the entire gut microbiota species of each child. Simulation of the CMMs revealed that the production of essential amino acids by the gut microbiota of the malnourished children was decreased, which was consistent with the children's plasma metabolite profiles. The development of strategies for treating abnormal health conditions on the basis of findings from GEMs will be a major challenge for the near future.

The metabolic interaction between a host and a pathogen is another important type of interspecies interaction that can be studied using GEMs [151]. In one recent study, the effects of pathogen infection on the host plant's photosynthetic capacity were examined using GEMs [148]. A generic GEM of the leaf of a potato plant (*Solanum tuberosum*) was first reconstructed, and three context-specific GEMs were subsequently created by incorporating transcriptome data from the cells of plants that were infected with *Phytophthora infestans*, a plant pathogen that causes late blight, at days 0, 1, and 2 after infection (Fig. 3d). The three context-specific GEMs were subsequently used alone to infer the metabolic interactions without using the pathogen's GEM. Pathogen infection was predicted to affect Calvin cycle fluxes, and thus carbon fixation. In particular, at day 1 after infection, the carboxylase activity and oxygenase activities of ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO), the first enzyme committed to carbon fixation in the Calvin cycle, were predicted to be decreased and increased, respectively. Such changes are known to reduce photosynthesis, and subsequently to induce the production of ROS, which could be associated with a quick defense mechanism against pathogen attack. Moreover, the flux of glyceraldehyde 3-phosphate formation in the second part of the Calvin cycle, as well as starch biosynthesis flux (an indicator of plant health), was predicted to decrease drastically from day 0 to day 1, but to recover slightly at day 3. GEM-based studies of strategies to protect plants against pathogens by examining fluxes of Calvin cycle and other pathways, which indicate the health status of a plant, will be of interest.

Modeling interactions among multiple cells or organisms, especially microbiota, presents many technical challenges. First, the microbial species that constitute a specific microbiota are not fully elucidated in most, if not all, cases. This partly explains why the microbial communities covered by the studies described above were simplified by considering

only representative microbial species. Thus, the use of GEMs will become more powerful when it becomes possible to identify all (or at least most) microbial species in a given community. For example, key microorganisms or metabolites in a specific microbiota can be suggested more systematically by examining metabolic interactions among more varied combinations of microorganisms from the microbiota. Here various modeling and simulation algorithms beyond FBA can be developed, depending on the objective of study and the scale of the metabolic modeling to be examined. Second, it is extremely difficult to measure metabolites that are exchanged by microbial species *in vivo*. Metabolome analyses using food, stool, and/or serum have been the most frequently practiced approaches for characterizing the metabolism of microbiota species, but still have limitations in revealing metabolite exchange by microbial species *in vivo*. This issue has led to active discussions on the need to identify and apply accurate condition-specific constraints for the GEMs of gut microbiota species and to conduct community-level manual curation of the GEM of each gut microbiota species [152, 153]. Finally, it is very important to inform experimental microbiologists of how the GEMs of microbiota are reconstructed, how omics data are used to improve the GEMs, and what the GEMs of microbiota species can be used for. Taken together, the development of experimental and computational techniques for the accurate measurement of metabolites *in vivo* and proper communication with experimental microbiologists will allow better understanding of microbe–microbe and host–microbe interactions. This will be important because the GEMs of microbiota species tend to present flux predictions that are often distinct from those of model organisms [152, 153].

#### Understanding human diseases 4

Human diseases have also been studied using context-specific GEMs to elucidate metabolic malfunctions in cells that are under chronic or acute disease conditions, and to suggest effective therapeutic targets. Cancers, including cancers of the liver [140, 154–156], breast [157], prostate [158], lung [159], and colon-rectum [160], have been the most active target of context-specific GEMs. Chronic diseases, including NAFLD [161] and obesity [140], have also been examined using context-specific GEMs. In a study by Hur et al. [149], the metabolism of LCSCs that showed therapeutic resistance in hepatocellular carcinoma was investigated in comparison with non-LCSCs by building context-specific GEMs for both cell types using their transcriptome data [149] (Fig. 3e). Upon identifying reactions with fluxes that differed significantly between LCSCs and non-LCSCs, transcription factors that are known to be associated with those reactions were traced. As a result, MYC, a transcription factor that is important in cell proliferation, among other

transcription factors, was found to be heavily involved in the changed metabolism of the LCSCs. This prediction was experimentally validated, providing insights into the reprogrammed metabolism of LCSCs. This GEM-based comparative analysis, along with the use of relevant omics data, is also applicable to explaining the reprogrammed metabolism of other types of cancer cells or other abnormal cells that represent disease conditions.

Other studies have used GEMs to predict altered intracellular metabolic flux distributions in acute diseases such as sepsis [150] and viral infection [162]. In one study, the reprogrammed metabolism of endothelium in sepsis patients was investigated (Fig. 3f) [150]. A human endothelium GEM, iEC2812, was reconstructed and integrated with transcriptome data to represent the metabolism of three endothelial subtypes: human pulmonary artery endothelial cell (HPAEC), human umbilical vein endothelial cell (HUVEC), and human microvascular endothelial cell (HMVEC). The network structures of the three context-specific GEMs were compared with one another in order to identify metabolic differences among the three endothelial subtypes, which occurred mainly in nucleotide metabolism. Furthermore, context-specific GEMs for HUVECs were reconstructed using transcriptome and metabolome data, which were obtained from lipopolysaccharide (LPS) and/or interferon- $\gamma$  (IFN- $\gamma$ )-treated HUVECs. The treatment of endothelial cells with LPS and IFN- $\gamma$  triggers a cellular status similar to those seen under bacterial infection and during an immune response, respectively. These context-specific GEMs for the LPS- and IFN- $\gamma$ -treated HUVECs predicted elevated fluxes through glycan and fatty acid metabolism, which were found to increase glycocalyx shedding and endothelial permeability in the sepsis patients.

Human diseases are associated with highly complex cues and cascading of signals, and thus, the use of GEM-based simulations alone can provide only limited insights into disease. In the future, a number of important studies need to be performed to provide a better understanding of human diseases and to help in designing proper therapies. First, in addition to a metabolic network, regulatory and/or signaling networks should also be considered to allow a more accurate computational description of a diseased cell. These different types of biological network are connected with one another in a highly complex manner. Thus, it will ultimately be necessary to integrate metabolic, gene regulatory, and signaling networks in modeling and simulation. This will require an innovative computational framework that allows simultaneous simulation of material flow (metabolic network) and information flow (gene regulatory and signaling networks). Second, it is increasingly recognized that a number of human diseases are significantly affected by patients' lifestyles. Thus, it will be necessary to develop

strategies to integrate human GEMs with a framework of precision medicine that involves not only patient-specific omics data but also personal lifestyle data, such as dietary habits and patterns of various physical activities.

## Conclusions

Having been established as one of the major modeling approaches for metabolic studies at systems level, the reconstruction and simulation of GEMs continue to be explored for an increasingly wider range of organisms and applications. Advances in the reconstruction and use of GEMs are largely attributed to the greater availability of biological data and information, and to the establishment of automatic GEM reconstruction tools. GEMs will continue to evolve by embracing a greater coverage of GPR associations, reconciling model inconsistencies, and developing novel mathematical modeling techniques that can be used in a high-throughput manner. GEMs will become more powerful by incorporating additional biochemical information that will provide explanations of cellular processes beyond metabolism [163]. Additional information that has been incorporated into GEMs successfully includes protein allocation [164–167], cellular macromolecular composition [168, 169], and protein structural information [21, 170–172]. Nevertheless, various biochemical properties, such as enzyme–substrate interactions, the structure of protein–protein complexes, and post-translational modification, still need to be considered further. It is expected that GEMs will find increasing applications in studying interactions among a greater number of cell types. These will include, for example, interactions among microorganisms in a given microbiota under spatiotemporally varying conditions, metabolic interactions between human (or animal or plant) cells and microbiota, and interactions among multiple human cells, to name a few. Although technical challenges remain to be overcome, GEMs will be applied to study an expanding and increasingly complex range of problems.

## Additional files

**Additional file 1:** **Figure S1.** A phylogenetic tree at the species level of all of the GEMs reconstructed to date. (PDF 9989 kb)

**Additional file 2:** **Table S1.** A full list of organisms subjected to the GEM reconstruction and used in the preparation of phylogenetic trees in Fig. 1 and Additional file 1. (XLSX 305 kb)

## Abbreviations

CMM: Community metabolic model; FBA: Flux balance analysis; GEM: Genome-scale metabolic model; GPR: Gene-protein-reaction; HMR: Human metabolic reaction; HUVEC: Human umbilical vein endothelial cell; IFN- $\gamma$ : Interferon- $\gamma$ ; LCSC: Liver cancer stem cell; LPS: Lipopolysaccharide; NAFLD: Non-alcoholic fatty liver disease; PROPER: PROmiscuity PrEdictoR; ROS: Reactive oxygen species

## Acknowledgments

We are grateful to Dr. Jae Yong Ryu and Woo Dae Jang for critical discussion.

## Authors' contributions

SYL and HUK conceived the contents of the manuscript. CDG, GBK, WJK, HUK, and SYL wrote the manuscript together and approved the final manuscript.

## Funding

This work was supported by the Technology Development Program to Solve Climate Changes on Systems Metabolic Engineering for Biorefineries (NRF-2012M1A2A2026556 and NRF-2012M1A2A2026557) and the Bio & Medical Technology Development Program (2018M3A9F3079664) from the Ministry of Science and ICT (MSIT) through the National Research Foundation (NRF) of Korea.

## Competing interests

The authors declare that they have no competing interests.

## Author details

<sup>1</sup>Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Metabolic and Biomolecular Engineering National Research Laboratory, Institute for the BioCentury, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 34141, Republic of Korea. <sup>2</sup>Department of Chemical and Biomolecular Engineering (BK21 Plus Program), Systems Biology and Medicine Laboratory, KAIST, Daejeon 34141, Republic of Korea.

<sup>3</sup>Systems Metabolic Engineering and Systems Healthcare Cross-Generation Collaborative Laboratory, KAIST, Daejeon 34141, Republic of Korea.  
<sup>4</sup>BioProcess Engineering Research Center and Bioinformatics Research Center, KAIST, Daejeon 34141, Republic of Korea.

Published online: 13 June 2019

## References

1. Edwards JS, Palsson BO. Systems properties of the *Haemophilus influenzae* Rd metabolic genotype. *J Biol Chem*. 1999;274:17410–6.
2. Kim HU, Kim TY, Lee SY. Metabolic flux analysis and metabolic engineering of microorganisms. *Mol BioSyst*. 2008;4:113–20.
3. Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*. 2010;5:93–121.
4. Orth JD, Thiele I, Palsson BO. What is flux balance analysis? *Nat Biotechnol*. 2010;28:245–8.
5. Ryu JY, Kim HU, Lee SY. Reconstruction of genome-scale human metabolic models using omics data. *Integr Biol (Camb)*. 2015;7:859–68.
6. Kim B, Kim WJ, Kim DI, Lee SY. Applications of genome-scale metabolic network model in metabolic engineering. *J Ind Microbiol Biotechnol*. 2015; 42:339–48.
7. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. *Cell*. 2015;161:971–87.
8. Edwards JS, Palsson BO. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proc Natl Acad Sci U S A*. 2000;97:5528–33.
9. Forster J, Famili I, Fu P, Palsson BO, Nielsen J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res*. 2003;13: 244–53.
10. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*. 2007;104:1777–82.
11. de Oliveira Dal'Molin CG, Quek LE, Palfreyman RW, Brumley SM, Nielsen LK. AraGEM, a genome-scale reconstruction of the primary metabolic network in *Arabidopsis*. *Plant Physiol*. 2010;152:579–89.
12. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods. *Nat Rev Microbiol*. 2012;10:291–305.
13. Buchel F, Rodriguez N, Swainston N, Wrzodek C, Czauderna T, Keller R, et al. Path2Models: large-scale generation of computational models from biochemical pathway maps. *BMC Syst Biol*. 2013;7:116.
14. Magnusdottir S, Heinken A, Kutt L, Ravcheev DA, Bauer E, Noronha A, et al. Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota. *Nat Biotechnol*. 2017;35:81–9.
15. Machado D, Andrejew S, Tramontano M, Patil KR. Fast automated reconstruction of genome-scale metabolic models for microbial species and communities. *Nucleic Acids Res*. 2018;46:7542–53.

16. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 2016;44:W242–5.
17. Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, et al. The complete genome sequence of *Escherichia coli* K-12. *Science*. 1997;277:1453–62.
18. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 2003;4:R54.
19. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol.* 2007;3:121.
20. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, Feist AM, et al. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol Syst Biol.* 2011;7:535.
21. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol.* 2017;35:904–8.
22. Zweers JC, Barak I, Becher D, Driessens AJ, Hecker M, Kontinen VP, et al. Towards the development of *Bacillus subtilis* as a cell factory for membrane proteins and protein complexes. *Microb Cell Factories*. 2008;7:10.
23. Harwood CR, Cranenburgh R. *Bacillus* protein secretion: an unfolding story. *Trends Microbiol.* 2008;16:73–9.
24. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem.* 2007;282:28791–9.
25. Goelzer A, Bekkai Brikci F, Martin-Verstraete I, Noirot P, Bessieres P, Aymerich S, et al. Reconstruction and analysis of the genetic and metabolic regulatory networks of the central metabolism of *Bacillus subtilis*. *BMC Syst Biol.* 2008;2:20.
26. Henry CS, Zinner JF, Cohoon MP, Stevens RL. iBSu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biol.* 2009;10:R69.
27. Tanaka K, Henry CS, Zinner JF, Jolivet E, Cohoon MP, Xia F, et al. Building the repertoire of dispensable chromosome regions in *Bacillus subtilis* entails major refinement of cognate large-scale metabolic model. *Nucleic Acids Res.* 2013;41:687–99.
28. Hao T, Han B, Ma H, Fu J, Wang H, Wang Z, et al. *In silico* metabolic engineering of *Bacillus subtilis* for improved production of riboflavin, Egl-237, (R,R)-2,3-butanediol and isobutanol. *Mol BioSyst.* 2013;9:2034–44.
29. Kocabas P, Calik P, Calik G, Ozdamar TH. Analyses of extracellular protein production in *Bacillus subtilis*—I: genome-scale metabolic model reconstruction based on updated gene-enzyme-reaction data. *Biochem Eng J.* 2017;127:229–41.
30. Belda E, Sekowska A, Le Fevre F, Morgat A, Mornico D, Ouzounis C, et al. An updated metabolic view of the *Bacillus subtilis* 168 genome. *Microbiology*. 2013;159:757–70.
31. Bose T, Das C, Dutta A, Mahamkali V, Sadhu S, Mande SS. Understanding the role of interactions between host and *Mycobacterium tuberculosis* under hypoxic condition: an *in silico* approach. *BMC Genomics*. 2018;19:555.
32. Beste DJ, Hooper T, Stewart G, Bonde B, Avignone-Rossa C, Bushell ME, et al. GSMN-TB: a web-based genome-scale network model of *Mycobacterium tuberculosis* metabolism. *Genome Biol.* 2007;8:R89.
33. Jamshidi N, Palsson BO. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ661 and proposing alternative drug targets. *BMC Syst Biol.* 2007;1:26.
34. Vashishtha R, Bhat AG, Kushwaha S, Bhardwaj A, Consortium O, Brahmachari SK. Systems level mapping of metabolic complexity in *Mycobacterium tuberculosis* to identify high-value drug targets. *J Transl Med.* 2014;12:263.
35. Rienksma RA, Suarez-Diez M, Spina L, Schaap PJ, Martins dos Santos VA. Systems-level modeling of mycobacterial metabolism for the identification of new (multi-)drug targets. *Semin Immunol.* 2014;26:610–22.
36. Garay CD, Dreyfuss JM, Galagan JE. Metabolic modeling predicts metabolite changes in *Mycobacterium tuberculosis*. *BMC Syst Biol.* 2015;9:57.
37. Ma S, Minch KJ, Rustad TR, Hobbs S, Zhou SL, Sherman DR, et al. Integrated modeling of gene regulatory and metabolic networks in *Mycobacterium tuberculosis*. *PLoS Comput Biol.* 2015;11:e1004543.
38. Kavvas ES, Seif Y, Yurkovich JT, Norsigian C, Poudel S, Greenwald WW, et al. Updated and standardized genome-scale reconstruction of *Mycobacterium tuberculosis* H37Rv, iEK1011, simulates flux states indicative of physiological conditions. *BMC Syst Biol.* 2018;12:25.
39. Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Mol Syst Biol.* 2010;6:422.
40. Satish Kumar V, Ferry JG, Maranas CD. Metabolic reconstruction of the archaeon methanogen *Methanosarcina Acetivorans*. *BMC Syst Biol.* 2011;5:28.
41. Benedict MN, Gonnerman MC, Metcalf WW, Price ND. Genome-scale metabolic reconstruction and hypothesis testing in the methanogenic archaeon *Methanosarcina acetivorans* C2A. *J Bacteriol.* 2012;194:855–65.
42. Nazem-Bokaee H, Gopalakrishnan S, Ferry JG, Wood TK, Maranas CD. Assessing methanotrophy and carbon fixation for biofuel production by *Methanosarcina acetivorans*. *Microb Cell Factories*. 2016;15:10.
43. Peterson JR, Thor S, Kohler L, Kohler PR, Metcalf WW, Luthey-Schulten Z. Genome-wide gene expression and RNA half-life measurements allow predictions of regulation and metabolic behavior in *Methanosarcina acetivorans*. *BMC Genomics*. 2016;17:924.
44. Thauer RK. Biochemistry of methanogenesis: a tribute to Marjory Stephenson. 1998 Marjory Stephenson prize lecture. *Microbiology*. 1998;144:2377–406.
45. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, et al. Life with 6000 genes. *Science*. 1996;274(546):563–7.
46. Duarte NC, Herrgard MJ, Palsson BO. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. *Genome Res.* 2004;14:1298–309.
47. Kuepfer L, Sauer U, Blank LM. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome Res.* 2005;15:1421–30.
48. Nookaew I, Jewett MC, Meechai A, Thammarrongtham C, Laoteng K, Cheevadhanarak S, et al. The genome-scale metabolic model iHN800 of *Saccharomyces cerevisiae* and its validation: a scaffold to query lipid metabolism. *BMC Syst Biol.* 2008;2:71.
49. Mo ML, Palsson BO, Herrgard MJ. Connecting extracellular metabolomic measurements to intracellular flux states in yeast. *BMC Syst Biol.* 2009;3:37.
50. Zomorrodi AR, Maranas CD. Improving the iMM904 *S. cerevisiae* metabolic model using essentiality and synthetic lethality data. *BMC Syst Biol.* 2010;4:178.
51. Osterlund T, Nookaew I, Bordel S, Nielsen J. Mapping condition-dependent regulation of metabolism in yeast through genome-scale modeling. *BMC Syst Biol.* 2013;7:36.
52. Herrgard MJ, Swainston N, Dobson P, Dunn WB, Argar KY, Arvas M, et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol.* 2008;26:1155–60.
53. Dobson PD, Smallbone K, Jameson D, Simeonidis E, Lanthaler K, Pir P, et al. Further developments towards a genome-scale metabolic model of yeast. *BMC Syst Biol.* 2010;4:145.
54. Heavner BD, Smallbone K, Barker B, Mendes P, Walker LP. Yeast 5—an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network. *BMC Syst Biol.* 2012;6:55.
55. Heavner BD, Smallbone K, Price ND, Walker LP. Version 6 of the consensus yeast metabolic network refines biochemical coverage and improves model performance. *Database (Oxford)*. 2013;2013:bat059.
56. Aung HW, Henry SA, Walker LP. Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. *Ind Biotechnol (New Rochelle N Y)*. 2013;9:215–28.
57. Dikicioglu D, Oliver SG. Extension of the yeast metabolic model to include iron metabolism and its use to estimate global levels of iron-recruiting enzyme abundance from cofactor requirements. *Biotechnol Bioeng*. 2019;116:610–21.
58. Scranton MA, Ostrand JT, Fields FJ, Mayfield SP. *Chlamydomonas* as a model for biofuels and bio-products production. *Plant J.* 2015;82:523–31.
59. Harris EH. *Chlamydomonas* as a model organism. *Annu Rev Plant Physiol Plant Mol Biol.* 2001;52:363–406.
60. Manichaikul A, Ghamsari L, Hom EF, Lin C, Murray RR, Chang RL, et al. Metabolic network analysis integrated with transcript verification for sequenced genomes. *Nat Methods*. 2009;6:589–92.
61. Chang RL, Ghamsari L, Manichaikul A, Hom EF, Balaji S, Fu W, et al. Metabolic network reconstruction of *Chlamydomonas* offers insight into light-driven algal metabolism. *Mol Syst Biol.* 2011;7:518.
62. Dal'Molin CG, Quek LE, Palfreyman RW, Nielsen LK. AlgaGEM—a genome-scale metabolic reconstruction of algae based on the *Chlamydomonas reinhardtii* genome. *BMC Genomics*. 2011;12(Suppl 4):S5.
63. Imam S, Schaeble S, Valenzuela J, Lopez Garcia de Lomana A, Carter W, Price ND, et al. A refined genome-scale reconstruction of *Chlamydomonas* metabolism provides a platform for systems-level analyses. *Plant J.* 2015;84:1239–56.
64. Winck FV, Melo DO, Riano-Pachon DM, Martins MC, Caldana C, Barrios AF. Analysis of sensitive CO<sub>2</sub> pathways and genes related to carbon uptake and accumulation in *Chlamydomonas reinhardtii* through genomic scale modeling and experimental validation. *Front Plant Sci.* 2016;7:43.

- 1
65. Mora Salguero DA, Fernandez-Nino M, Serrano-Bermudez LM, Paez Melo DO, Winck FV, Caldana C, et al. Development of a *Chlamydomonas reinhardtii* metabolic network dynamic model to describe distinct phenotypes occurring at different CO<sub>2</sub> levels. *PeerJ*. 2018;6:e5528.
  66. Tibocha-Bonilla JD, Zuniga C, Godoy-Silva RD, Zengler K. Advances in metabolic modeling of oleaginous microalgae. *Biotechnol Biofuels*. 2018;11:241.
  67. Tissenbaum HA. Using *C. elegans* for aging research. *Invertebr Reprod Dev*. 2015;59:59–63.
  68. Hubbard EJ, Greenstein D. The *Caenorhabditis elegans* gonad: a test tube for cell and developmental biology. *Dev Dyn*. 2000;218:2–22.
  69. Shen P, Yue Y, Park Y. A living model for obesity and aging research: *Caenorhabditis elegans*. *Crit Rev Food Sci Nutr*. 2018;58:741–54.
  70. Yilmaz LS, Walhout AJ. A *Caenorhabditis elegans* genome-scale metabolic network model. *Cell Syst*. 2016;2:297–311.
  71. Gebauer J, Gentsch C, Mansfeld J, Schmeisser K, Waschnia S, Brandes S, et al. A genome-scale database and reconstruction of *Caenorhabditis elegans* metabolism. *Cell Syst*. 2016;2:312–22.
  72. Ma L, Chan AHC, Hattwell J, Ebert PR, Schirra HJ. Systems biology analysis using a genome-scale metabolic model shows that phosphine triggers global metabolic suppression in a resistant strain of *C. elegans*. *bioRxiv*. 2017; doi: <https://doi.org/10.1101/144386>.
  73. Witting M, Hastings J, Rodriguez N, Joshi CJ, Hattwell JPN, Ebert PR, et al. Modeling meets metabolomics—the WormJam consensus model as basis for metabolic studies in the model organism *Caenorhabditis elegans*. *Front Mol Biosci*. 2018;5:96.
  74. Poolman MG, Miguet L, Sweetlove LJ, Fell DA. A genome-scale metabolic model of *Arabidopsis* and some of its properties. *Plant Physiol*. 2009;151:1570–81.
  75. Mintz-Oron S, Meir S, Malitsky S, Ruppin E, Aharoni A, Shlomi T. Reconstruction of *Arabidopsis* metabolic network models accounting for subcellular compartmentalization and tissue-specificity. *Proc Natl Acad Sci U S A*. 2012;109:339–44.
  76. Cheung CY, Williams TC, Poolman MG, Fell DA, Ratcliffe RG, Sweetlove LJ. A method for accounting for maintenance costs in flux balance analysis improves the prediction of plant cell metabolic phenotypes under stress conditions. *Plant J*. 2013;75:1050–61.
  77. de Oliveira Dal'Molin CG, Nielsen LK. Plant genome-scale metabolic reconstruction and modelling. *Curr Opin Biotechnol*. 2013;24:271–7.
  78. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol*. 2013;31:419–25.
  79. Swainston N, Smallbone K, Hefzi H, Dobson PD, Brewer J, Hanscho M, et al. Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*. 2016;12:109.
  80. Ryu JY, Kim HU, Lee SY. Framework and resource for more than 11,000 gene-transcript-protein-reaction associations in human metabolism. *Proc Natl Acad Sci U S A*. 2017;114:E9740–9.
  81. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Drager A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol*. 2018;36:272–81.
  82. Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, et al. Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol Syst Biol*. 2013;9:649.
  83. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based human protein atlas. *Nat Biotechnol*. 2010;28:1248–50.
  84. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. *Science*. 2015;347(6220):1260419.
  85. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun*. 2014;5:3083.
  86. Varemo L, Scheele C, Broholm C, Mardinoglu A, Kampf C, Asplund A, et al. Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes. *Cell Rep*. 2015;11:921–33.
  87. Aite M, Chevallier M, Frioux C, Trottier C, Got J, Cortes MP, et al. Traceability, reproducibility and wiki-exploration for ‘a-la-carte’ reconstructions of genome-scale metabolic models. *PLoS Comput Biol*. 2018;14:e1006146.
  88. Karlsen E, Schulz C, Almaas E. Automated generation of genome-scale metabolic draft reconstructions based on KEGG. *BMC Bioinformatics*. 2018;19:467.
  89. Pitkanen E, Jouhten P, Hou J, Syed MF, Blomberg P, Kludas J, et al. Comparative genome-scale reconstruction of gapless metabolic networks for present and ancestral species. *PLoS Comput Biol*. 2014;10:e1003465.
  90. Boeleg J, Olivier BG, Teusink B. FAME, the flux analysis and modeling environment. *BMC Syst Biol*. 2012;6:8.
  91. Dias O, Rocha M, Ferreira EC, Rocha I. Reconstructing genome-scale metabolic models with *merlin*. *Nucleic Acids Res*. 2015;43:3899–910.
  92. Hanemaaijer M, Olivier BG, Rolling WF, Bruggeman FJ, Teusink B. Model-based quantification of metabolic interactions from dynamic microbial-community data. *PLoS One*. 2017;12:e0173183.
  93. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*. 2010;28:977–82.
  94. Karp PD, Latendresse M, Paley SM, Krummenacker M, Ong QD, Billington R, et al. Pathway tools version 19.0 update: software for pathway/genome informatics and systems biology. *Brief Bioinform*. 2016;17:877–90.
  95. Wang H, Marcisauskas S, Sanchez BJ, Domenzain I, Hermansson D, Agren R, et al. RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on *Streptomyces coelicolor*. *PLoS Comput Biol*. 2018;14:e1006541.
  96. Swainston N, Smallbone K, Mendes P, Kell D, Paton N. The SubIMInAl toolbox: automating steps in the reconstruction of metabolic networks. *J Integr Bioinform*. 2011;8:186.
  97. King ZA, Lu J, Drager A, Miller P, Federowicz S, Lerman JA, et al. BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res*. 2016;44:D515–22.
  98. Glont M, Nguyen TVN, Graesslin M, Halke R, Ali R, Schramm J, et al. BioModels: expanding horizons to include more modelling approaches and formats. *Nucleic Acids Res*. 2018;46:D1248–53.
  99. Purnputtapong N, Nookaew I, Nielsen J. Human metabolic atlas: an online resource for human metabolism. *Database (Oxford)*. 2015;2015:bav068.
  100. Pabinger S, Snajder R, Hardiman T, Willi M, Dander A, Trajanoski Z. MEMOSys 2.0: an update of the bioinformatics database for genome-scale models and genomic data. *Database (Oxford)*. 2014;2014:bau004.
  101. Moretti S, Martin O, Van Du Tran T, Bridge A, Morgat A, Pagni M. MetaNetX/MNXref—reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res*. 2016;44: D523–6.
  102. Seaver SM, Gerdes S, Frelin O, Lerma-Ortiz C, Bradbury LM, Zallot R, et al. High-throughput comparison, functional annotation, and metabolic modeling of plant genomes using the PlantSEED resource. *Proc Natl Acad Sci U S A*. 2014;111:9645–50.
  103. Noronha A, Modamio J, Jarosz Y, Guerard E, Sompairac N, Preciat G, et al. The virtual metabolic human database: integrating human and gut microbiome metabolism with nutrition and disease. *Nucleic Acids Res*. 2019; 47:D614–24.
  104. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, et al. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol*. 2018;36:566–9.
  105. Mendoza SN, Olivier BG, Molenaar D, Teusink B. A systematic assessment of current genome-scale metabolic reconstruction tools. *bioRxiv*. 2019; doi: <https://doi.org/10.1101/558411>.
  106. Lieven C, Beber ME, Olivier BG, Bergmann FT, Babaei P, Bartell JA, et al. Memote: a community-driven effort towards a standardized genome-scale metabolic model test suite. *bioRxiv*. 2018; doi: <https://doi.org/10.1101/350991>.
  107. Reed JL, Patel TR, Chen KH, Joyce AR, Applebee MK, Herring CD, et al. Systems approach to refining genome annotation. *Proc Natl Acad Sci U S A*. 2006;103:17480–4.
  108. Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling *in silico/in vivo* growth predictions. *PLoS Comput Biol*. 2009;5:e1000308.
  109. Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiol Rev*. 2009;33: 164–90.
  110. Kim WJ, Kim HU, Lee SY. Current state and applications of microbial genome-scale metabolic models. *Curr Opin Syst Biol*. 2017;2:10–8.
  111. Oberhardt MA, Palsson BO, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*. 2009;5:320.
  112. Machado D, Herrgard M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput Biol*. 2014;10:e1003580.
  113. Opdam S, Richelle A, Kellman B, Li S, Zielinski DC, Lewis NE. A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Syst*. 2017;4:318–29.
  114. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*. 2008;4:e1000082.
- 2

115. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. *Bioinformatics*. 2010;26:3140–2.
116. Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol*. 2010;6:401.
117. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaei I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol*. 2012;8:e1002518.
118. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol*. 2012;6:153.
119. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol*. 2014;10:721.
120. Schultz A, Qutub AA. Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput Biol*. 2016;12:e1004808.
121. Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, et al. Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun*. 2017;8:14250.
122. Yang JE, Park SJ, Kim WJ, Kim HJ, Kim BJ, Lee H, et al. One-step fermentative production of aromatic polyesters from glucose by metabolically engineered *Escherichia coli* strains. *Nat Commun*. 2018;9:79.
123. Mishra P, Lee NR, Lakshmanan M, Kim M, Kim BG, Lee DY. Genome-scale model-driven strain design for dicarboxylic acid production in *Yarrowia lipolytica*. *BMC Syst Biol*. 2018;12:12.
124. Abdel-Haleem AM, Hefzi H, Mineta K, Gao X, Gojobori T, Palsson BO, et al. Functional interrogation of *Plasmodium* genus metabolism identifies species- and stage-specific differences in nutrient essentiality and drug targeting. *PLoS Comput Biol*. 2018;14:e1005895.
125. Presta L, Bosi E, Mansouri L, Dijkshoorn L, Fani R, Fondi M. Constraint-based modeling identifies new putative targets to fight colistin-resistant *A baumannii* infections. *Sci Rep*. 2017;7:3706.
126. Guzman GI, Utrilla J, Nurk S, Brunk E, Monk JM, Ebrahim A, et al. Model-driven discovery of underground metabolic functions in *Escherichia coli*. *Proc Natl Acad Sci U S A*. 2015;112:929–34.
127. Oberhardt MA, Zarecki R, Reshef L, Xia F, Duran-Frigola M, Schreiber R, et al. Systems-wide prediction of enzyme promiscuity reveals a new underground alternative route for pyridoxal 5'-phosphate production in *E. coli*. *PLoS Comput Biol*. 2016;12:e1004705.
128. Seif Y, Kavvas E, Lachance JC, Yurkovich JT, Nuccio SP, Fang X, et al. Genome-scale metabolic reconstructions of multiple *Salmonella* strains reveal serovar-specific metabolic traits. *Nat Commun*. 2018;9:3771.
129. Prigent S, Nielsen JC, Frisvad JC, Nielsen J. Reconstruction of 24 *Penicillium* genome-scale metabolic models shows diversity based on their secondary metabolism. *Biotechnol Bioeng*. 2018;115:2604–12.
130. Li Q, Du W, Liu D. Perspectives of microbial oils for biodiesel production. *Appl Microbiol Biotechnol*. 2008;80:749–56.
131. Chung BK, Selvarasu S, Andrea C, Ryu J, Lee H, Ahn J, et al. Genome-scale metabolic reconstruction and *in silico* analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microb Cell Factories*. 2010;9:50.
132. Kim M, Yi JS, Lakshmanan M, Lee DY, Kim BG. Transcriptomics-based strain optimization tool for designing secondary metabolite overproducing strains of *Streptomyces coelicolor*. *Biotechnol Bioeng*. 2016;113:651–60.
133. Lun DS, Rockwell G, Guido NJ, Baym M, Kelner JA, Berger B, et al. Large-scale identification of genetic design strategies using local search. *Mol Syst Biol*. 2009;5:296.
134. Lakshmanan M, Chung BK, Liu C, Kim SW, Lee DY. Cofactor modification analysis: a computational framework to identify cofactor specificity engineering targets for strain improvement. *J Bioinform Comput Biol*. 2013;11:1343006.
135. Sigurdsson G, Fleming RM, Heinken A, Thiele I. A systems biology approach to drug targets in *Pseudomonas aeruginosa* biofilm. *PLoS One*. 2012;7:e34337.
136. Kim HU, Kim TY, Lee SY. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen *Acinetobacter baumannii* AYE. *Mol BioSyst*. 2010;6:339–48.
137. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, et al. Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol Syst Biol*. 2011;7:460.
138. Josling GA, Llinas M. Sexual development in *Plasmodium* parasites: knowing when it's time to commit. *Nat Rev Microbiol*. 2015;13:573–87.
139. Boucher HW, Talbot GH, Bradley JS, Edwards JE, Gilbert D, Rice LB, et al. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis*. 2009;48:1–12.
140. Bidkhori G, Benfeitas R, Elmas E, Kararoudi MN, Arif M, Uhlen M, et al. Metabolic network-based identification and prioritization of anticancer targets based on expression data in hepatocellular carcinoma. *Front Physiol*. 2018;9:916.
141. Bosi E, Monk JM, Aziz RK, Fondi M, Nizet V, Palsson BO. Comparative genome-scale modelling of *Staphylococcus aureus* strains identifies strain-specific metabolic capabilities linked to pathogenicity. *Proc Natl Acad Sci U S A*. 2016;113:E3801–9.
142. Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, et al. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. *Proc Natl Acad Sci U S A*. 2013;110:20338–43.
143. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol*. 2007;3:92.
144. Pacheco AR, Moel M, Segre D. Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nat Commun*. 2019;10:103.
145. McNally CP, Borenstein E. Metabolic model-based analysis of the emergence of bacterial cross-feeding via extensive gene loss. *BMC Syst Biol*. 2018;12:69.
146. Rosario D, Benfeitas R, Bidkhori G, Zhang C, Uhlen M, Shoae S, et al. Understanding the representative gut microbiota dysbiosis in metformin-treated type 2 diabetes patients using genome-scale metabolic modeling. *Front Physiol*. 2018;9:775.
147. Kumar M, Ji B, Babaei P, Das P, Lappa D, Ramakrishnan G, et al. Gut microbiota dysbiosis is associated with malnutrition and reduced plasma amino acid levels: lessons from genome-scale metabolic modeling. *Metab Eng*. 2018;49:128–42.
148. Botero K, Restrepo S, Pinzon A. A genome-scale metabolic model of potato late blight suggests a photosynthesis suppression mechanism. *BMC Genomics*. 2018;19:863.
149. Hur W, Ryu JY, Kim HU, Hong SW, Lee EB, Lee SY, et al. Systems approach to characterize the metabolism of liver cancer stem cells expressing CD133. *Sci Rep*. 2017;7:45557.
150. McGarry S, Anuforo O, Halldorsson H, Bergmann A, Halldorsson S, Palsson S, et al. Metabolic systems analysis of LPS induced endothelial dysfunction applied to sepsis patient stratification. *Sci Rep*. 2018;8:6811.
151. McCloskey D, Palsson BO, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of *Escherichia coli*. *Mol Syst Biol*. 2013;9:661.
152. Babaei P, Shoae S, Ji B, Nielsen J. Challenges in modeling the human gut microbiome. *Nat Biotechnol*. 2018;36:682–6.
153. Magnusdottir S, Heinken A, Fleming RMT, Thiele I. Reply to 'Challenges in modeling the human gut microbiome'. *Nat Biotechnol*. 2018;36:686–91.
154. Steenbergen R, Oti M, Ter Horst R, Tat W, Neufeldt C, Belovodskiy A, et al. Establishing normal metabolism and differentiation in hepatocellular carcinoma cells by culturing in adult human serum. *Sci Rep*. 2018;8:11685.
155. Wu HQ, Cheng ML, Lai JM, Wu HH, Chen MC, Liu WH, et al. Flux balance analysis predicts Warburg-like effects of mouse hepatocyte deficient in miR-122a. *PLoS Comput Biol*. 2017;13:e1005618.
156. Bjornson E, Mukhopadhyay B, Asplund A, Pristovsek N, Cinar R, Romeo S, et al. Stratification of hepatocellular carcinoma patients based on acetate utilization. *Cell Rep*. 2015;13:2014–26.
157. Gamez-Pozo A, Trilla-Fuertes L, Berbes-Soria J, Selevsek N, Lopez-Vacas R, Diaz-Almiron M, et al. Functional proteomics outlines the complexity of breast cancer molecular subtypes. *Sci Rep*. 2017;7:10100.
158. Marin de Mas I, Aguilar E, Zodda E, Balcells C, Marin S, Dallmann G, et al. Model-driven discovery of long-chain fatty acid metabolic reprogramming in heterogeneous prostate cancer cells. *PLoS Comput Biol*. 2018;14:e1005914.
159. Asgari Y, Khosravi P, Zabihiinpour Z, Habibi M. Exploring candidate biomarkers for lung and prostate cancers using gene expression and flux variability analysis. *Integr Biol (Camb)*. 2018;10:113–20.
160. Fuhr L, El-Athman R, Scrima R, Cela O, Carbone A, Knoop H, et al. The circadian clock regulates metabolic phenotype rewiring via HKDC1 and modulates tumor progression and drug response in colorectal cancer. *EBioMed*. 2018;33:105–21.
161. Shubham K, Vinay L, Vinod PK. Systems-level organization of non-alcoholic fatty liver disease progression network. *Mol BioSyst*. 2017;13:1898–911.
162. Aller S, Scott A, Sarkar-Tyson M, Soyer OS. Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against chikungunya, dengue and Zika viruses. *J R Soc Interface*. 2018;15. <https://doi.org/10.1098/rsif.2018.0125>.

163. King ZA, Lloyd CJ, Feist AM, Palsson BO. Next-generation genome-scale models for metabolic engineering. *Curr Opin Biotechnol.* 2015;35:23–9.
164. Sanchez BJ, Zhang C, Nilsson A, Lahtvee PJ, Kerkhoven EJ, Nielsen J. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol Syst Biol.* 2017;13:935.
165. Mori M, Hwa T, Martin OC, De Martino A, Marinari E. Constrained allocation flux balance analysis. *PLoS Comput Biol.* 2016;12:e1004913.
166. Goelzer A, Muntel J, Chubukov V, Jules M, Prestel E, Nolker R, et al. Quantitative prediction of genome-wide resource allocation in bacteria. *Metab Eng.* 2015;32:232–43.
167. Beg QK, Vazquez A, Ernst J, de Menezes MA, Bar-Joseph Z, Barabasi AL, et al. Intracellular crowding defines the mode and sequence of substrate uptake by *Escherichia coli* and constrains its metabolic activity. *Proc Natl Acad Sci U S A.* 2007;104:12663–8.
168. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol.* 2013;9:693.
169. Lerman JA, Hyduke DR, Latif H, Portnoy VA, Lewis NE, Orth JD, et al. *In silico* method for modelling metabolism and gene product expression at genome scale. *Nat Commun.* 2012;3:929.
170. Seif Y, Monk JM, Mih N, Tsunemoto H, Poudel S, Zuniga C, et al. A computational knowledge-base elucidates the response of *Staphylococcus aureus* to different media types. *PLoS Comput Biol.* 2019;15:e1006644.
171. Brunk E, Mih N, Monk J, Zhang Z, O'Brien EJ, Bliven SE, et al. Systems biology of the structural proteome. *BMC Syst Biol.* 2016;10:26.
172. Chang RL, Andrews K, Kim D, Li Z, Godzik A, Palsson BO. Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. *Science.* 2013;340:1220–3.

1

**Publisher's Note 2**

Springer Nature remains neutral with regard to jurisdictional claims in 3 published maps and institutional affiliations.

3

Ready to submit your research? Choose BMC and benefit from: 4

- 5
- fast, convenient online submission
  - thorough peer review by experienced researchers in your field
  - rapid publication on acceptance
  - support for research data, including large and complex data types
  - gold Open Access which fosters wider collaboration and increased citations
  - maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress. 6

Learn more [biomedcentral.com/submissions](http://biomedcentral.com/submissions) 7

