Scholastic Medical Sciences

Research Article

Volume 3 Issue 8 August 2026

A Scoping Review of the Development of Genome-Scale Models Prior to 2026

Jullean Louise Dizon Baylon[1,2], Shamsia Bte Ali[1,2], Shnan Harish Modankap[1,2], Lavanya Dhingra[1,2], Nurhaliza Mohamed Ali[1,2], Parkavi Subramanian[1,2], Maurice Han Tong Ling[2,3,4]

[1]*Department of Applied Sciences, Northumbria University, United Kingdom*

[2]*Management Development Institute of Singapore, Singapore*

[3]*Newcastle Australia Institute of Higher Education, University of Newcastle, Australia*

[4]*HOHY PTE LTD, Singapore*

Corresponding Author: Maurice Han Tong Ling, Management Development Institute of Singapore, Singapore,

Received: July 04, 2026**; Published:** July 24, 2026

Abstract

Genome-scale models (GSMs) have been an important tool in systems biology and metabolic engineering. This scoping review examined the development and current advancement in genome-scale models published prior to 2026 in PubMed. The included studies were categories into four themes: Reconstructed Models covering the transition from M-models to ME-models and automated pipelines. New Models, spanning prokaryotic, eukaryotic, and human GSMs. Refined Models, covering gap correction, pathway addition, and improved gene annotation. Usage of GSMs, spanning biotechnology, medical, and other applications. The finding shows that GSMs have grown substantially in scope and complexity, extending beyond microbial systems into human and tissue-specific applications, while increasingly supporting strain engineering, and therapeutic target discovery, demonstrating their versatility for fundamental research and biotechnological innovation, with future development to incorporate artificial intelligent and cross-species comparative analysis.

Introduction

Genome-scale models (GSMs) are comprehensive computational representations of an organism's metabolic network that integrates genomic, biochemical, and physiological information into a single mathematical framework. GSMs have become indispensable tools in systems biology because they enable researchers to investigate metabolism at the systems level rather than focusing on individual genes or isolated biochemical pathways [1, 2]. By incorporating all known metabolic reactions within an organism, GSMs provide a holistic view of cellular metabolism and allow researchers to simulate metabolic behaviour under different environmental and genetic conditions. GSMs have been also applied across multiple research domains, from cells and tissues to whole body systems, to investigate disease mechanisms, identify potential therapeutic targets, support biomarker discovery, and advance precision medicine [3–5].

Over the past two decades, the applications of GSMs have expanded considerably beyond microbial metabolism they are now widely used across diverse biological systems. Although several review articles have discussed specific aspects of genome-scale modelling, few have systematically classified studies according to the nature of their contribution. Therefore, there is limited literature providing a comprehensive review of how GSMs have evolved across different stages of development and their expanding application. The aim of this scoping review is to provide researchers with a comprehensive overview of research efforts that have been conducted into the continuous development and functionality of genome-scale models in biology, biotechnology, and other field of study. Earlier published studies have tended to focus on specific GSM updates rather than examining distinct categories such as new, refined, reconstructed, and application of GSMs. This review aims to expand the scope of earlier efforts. Which will also serve as a useful historical record of documenting evolution of genome-scale models across various organisms and field of study.

Methods

A systematic literature search was conducted using the PubMed database with the following search term "genome-scale model"[tiab] OR "genome-scale metabolic model"[tiab], for publication till 31

December 2025. The search URL was https://pubmed.ncbi.nlm.nih.gov/?term=”genome-scale+model”[tiab]OR+”genome-scale+metabolic+model”[tiab]&filter=dates.1000/1/1-2025/12/31. Following this initial search the retrieved articles were screened using three inclusion criteria; namely, (a) English-language publications, (b) full-text availability, and (c) primary research articles.

Results and Discussion

A total of 1079 articles were initially identified through a PubMed search. After exclusions, 923 studies were included for this review (Figure 1). These 923 studies were themed (Table 1) into (a) Reconstructed GSMs (n = 41), (b) New GSMs (n = 290), (c) Refined GSMs (n = 77) and, (d) Usage of GSMs (n = 515). Due to the large number of included studies, representative references were cited with the remaining references listed as Supplementary materials.

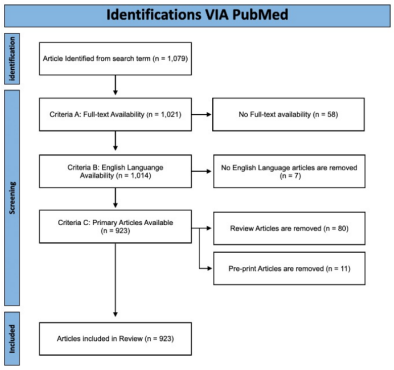

Figure 1: PRISMA Diagram of literature search and study selection process.

Main Theme	Sub-Theme	Number of Articles with Representative Reference
1. Reconstructed GSMs	1.1. M-model to ME-model	35 [1, 6–8]
	1.2. New automated pipelines	6 [9–11]
2. New GSMs	2.1. Prokaryotic GSMs	179 [11–13]
	2.2. Eukaryotic GSMs	98 [5, 14]
	2.3. Human GSMs (global and tissue-specific)	13 [3, 15]
3. Refined GSMs	3.1. Correction of Gaps	26 [11, 16]
	3.2. Additions of pathways	30 [7, 14, 17]
	3.3. Improved gene annotations	21 [9, 10]
4. Usage of GSMs	4.1. Biotechnology and Industrial Applications	202 [18]
	4.2. Medical Applications	112 [12, 19–21]
	4.3. Other Applications	201 [22–24]

Table 1: Theming of 923 Included Studies.

Theme 1.1. Reconstructed Genome-scale Models: M-model to ME-model. M-models primarily focus on predicting metabolic fluxes using constraint-based modelling techniques such as Flux Balance Analysis. They have also been widely used to predict growth rates, identifying genes, simulate gene knockouts, and optimise engineering strategies [1, 6, 8]. However, M-models only represent metabolic reaction and lack the ability to account for the cellular resources required to synthesise enzymes or regulate gene expression. Additionally, M-models cannot accurately represent the metabolic costs associated with transcription, translation, protein synthesis, and enzyme assembly [6, 8]. To overcome limitations, Metabolism and expression models (ME-models) were developed to extend M-models’ capabilities by incorporating the processes of transcription, translation, protein assembly, and enzyme synthesis into the metabolic network. This advancement reflects a shift from modelling metabolic reactions alone to representing the broader biological process that influence cellular metabolism. Consequently, ME-models provide a more comprehensive framework for understanding genotype-phenotype relationships and predicting cellular behaviour under different environmental conditions. However, despite its advantages, ME-models required extensive biochemical and kinetic information, making them more difficult to construct, curate, and compute than M-models [6, 7].

Theme 1.2. Reconstructed Genome-scale Models: New automated pipelines. Traditionally, genome-scale models were constructed through extensive manual curation, requiring researchers to integrate genomic information, biochemical databases, metabolic reactions, and literature evidence into a single computational network. Although manually created models generally produce the highest quality reconstructions, they require substantial expertise and are both labour intensive and time consuming. The development of automated reconstruction tools such as ModelSEED, RAVEN, KBase, and other reconstruction pipelines has addressed these challenges [9–11]. These tools automate much of the reconstruction process by integrating genome annotation with curated metabolic database to rapidly generate draft models. As a result, reconstruction has become faster, more reproducible, and scalable, enabling researchers to develop GSMs for newly sequenced organism in a fraction of time required.

Theme 2.1. New Genome-scale Models: Prokaryotic Genome-scale Models. Prokaryotic organisms have played a fundamental role in the development of genome-scale models. Prokaryotes have been the focus of genome-scale model development because of their relatively simple cellular organisation and well-annotated genome [11]. Compared with more complex organisms, bacterial genomes are generally easier to annotate, allowing the researcher to construct much more accurate metabolic models. The abundance of experimental and genomic data has enabled the construction of high-quality prokaryotic GSMs, which have been widely applied to investigate microbial metabolism, identify essential genes, and even support antimicrobial drug discovery [12, 13].

Theme 2.2. New Genome-scale Models: Eukaryotic Genome-scale Models. The successful construction of microbial GSMs provided the foundation for extension genome-scale models to eukaryotic organisms. Unlike prokaryotes, eukaryotes possess greater biological complexity, including multiple intracellular organelles, compartmentalised metabolic pathways, larger genomes, and additional regulatory mechanisms. These characteristics make the construction of GSMs considerably more challenging [5, 14]. Despite these challenges, advances in genome sequencing, computational modelling and genome annotation have enable the development of GSMs for fungi, plants, algae, parasites, and animals. Overall, the development of eukaryotic GSMs demonstrates the increasing capability of GSMs to represent biologically complex systems.

Theme 2.3. New Genome-scale Models: Human Genome-scale Models. Human genome-scale models represent one of the most significant advances in systems biology, as they provide a comprehensive framework for studying human metabolism under both healthy and disease conditions. Unlike microbial GSMs which typically describe the metabolism of a single cell type, human GSMs must account for the metabolic diversity of multiple organs and tissues through the body [3, 15]. Overall, human GSMs have shown growing contribution to precision medicine and specific metabolic research.

Theme 3.1. Refined GSMs: Correction of Gaps. As genome-scale models evolve, they undergo continuous refinement to address incomplete metabolic networks, incorrect reaction placements, or inconsistencies in gene-protein-reaction associations. These gaps can prevent the model from accurately simulating cellular metabolism. Gap-filling strategies address this limitation by incorporating reaction from curated metabolic databases [11, 16]. Consequently, gap correction has become an essential component of model refinement as it improves model completeness and enhances the ability of metabolic predictions. This allows GSM to generate more biologically accurate simulations.

Theme 3.2. Refined GSMs: Addition of New Pathways. Refinement of GSMs frequently involves the incorporation of newly identified metabolic reactions, transport processes, and biochemical pathways as knowledge of cellular metabolism expands. These additions improve the completeness of metabolic reconstructions which enables the models to better represent organism-specific metabolic capacities [14, 17]. The integration of new pathways not only improves metabolic flux predictions but also enables GSMs to capture organism specific metabolic capabilities more accurately [7]. Therefore, these refinements broaden the applicability of GSMs.

Theme 3.3. Refined GSMs: Improved Gene Annotation. Genome-scale models rely on accurate gene-protein-reaction or GPR associations to link genes with metabolic reactions [9, 10]. Furthermore, improved gene annotation strengthens the biological reliability of GSMs by reducing annotation errors and enhancing enzyme assignments. As genome annotation technologies continues to advance, accurate GPR association will remain fundamental to producing high quality and predictive metabolic models.

Theme 4.1. Usage of GSMs: Biotechnology and Industrial Application. Genome-scale models are extensively applied in biotechnology and industrial research to optimise microbial metabolism and improve the valuable products [18]. Constraint-based analyses, such as Flux Balance Analysis or FBA enables the identification of gene targets and pathway modification. Therefore, GSMs have become valuable tools for supporting sustainable and cost-effective biotechnological processes.

Theme 4.2. Usage of GSMs: Medical Applications. The applications of GSMs in the medical field are growing, particularly for investigating disease-associated metabolic alterations and identifying potential therapeutic targets. A major advantage of GSMs is their ability to integrate multi-omics datasets, including transcriptomic, proteomic, and metabolomic data, enabling a systems-level understanding of disease metabolism [20, 21]. In cancer research, GSMs have been used to identify tumour-specific metabolic vulnerabilities and metabolic dependencies, supporting the discovery of selective therapeutic targets while reducing the potential for off-targets effects in healthy tissues [20, 25]. In infectious disease research, pathogen-specific GSMs have been applied to predict essential genes and metabolic reactions that can serve as potential antimicrobial drug targets, providing a rational framework for antibiotic development [12, 19]. Collectively these studies demonstrate GSMs provides a versatile platform by applying it to metabolic disorders, including obesity and type 2 diabetes, where human gut microbiome and tissue-specific metabolic models have been used to identify altered metabolic pathways, predict disease-associated metabolic changes and discover potential biomarker for disease diagnosis and treatment [4, 21].

Theme 4.3. Usage of GSMs: Other Applications. Genome-scale models have demonstrated across a wide range of application beyond biotechnology and medicine. In environmental research, GSMs have been used to investigate microbial community metabolism, biogeochemical cycling and ecosystem functions by integrating meta transcriptomic data to characterise metabolic interaction within complex microbial communities [24]. In agriculture, GSMs have been applied to study plant metabolism, crop responses to environmental stress and host-microbe interactions. Context specific metabolic models have provided insights into plant adaptation to drought stress and other environmental conditions, supporting crop improvement and sustainable practices [22]. Beyond these field, GSMs have also been applied in comparative and translational biology to investigate metabolism across different animal models, facilitating cross-species comparisons and improving the understanding of conserved metabolic pathways [23]. These diverse applications highlight the growing importance of GSMs as versatile computational tools for address complex biological question across multiple scientific disciplines.

Future Considerations. Future research should focus on integrating artificial intelligence and machine learning with genome-scale modelling to improve automated reconstruction, gap filling, and model refinement [26, 27]. Greater integration of multi-omics datasets is also expected to improve the biological relevance and predictive accuracy of GSMs. As sequencing technologies and computational methods continue to advance, genome-scale models are likely to play an increasingly important role in systems biology, metabolic engineering, and precision medicine [28–30].

Conclusion

This scoping review provides a comprehensive overview of the development and advancement of genome-scale models (GSMs) from 2002 to 2025. Analysis of 923 primary research articles identified from four major themes: reconstructed, new, refined, and usage of genome-scale models. These findings show the rapid evolutions of genome scale modelling from basic metabolic reconstructions to computational platforms capable of investigating complex biological systems across diverse organisms.

Supplementary Materials

Supplementary materials can be downloaded from https://bit.ly/GSM_SR.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

The authors wish to thank the institute, Management Development Institute of Singapore, for its support towards this work. The cost of publication fees was borne by the authors.

References

1. Bernstein DB, Akkas B, Price MN, Arkin AP (2023). Evaluating E. coli genome-scale metabolic model accuracy with high-throughput mutant fitness data. Molecular Systems Biology 19(12):MSB202311566.
2. Chen C, Liao C, Liu Y-Y (2023). Teasing out missing reactions in genome-scale metabolic networks through hypergraph learning. Nature Communications 14(1):2375.
3. Robinson JL, Kocabaş P, Wang H, Cholley P-E, Cook D, Nilsson A, Anton M, Ferreira R, Domenzain I, Billa V, Limeta A, Hedin A, Gustafsson J, Kerkhoven EJ, Svensson LT, Palsson BO, Mardinoglu A, Hansson L, Uhlén M, Nielsen J (2020) An atlas of human metabolism. Science Signaling 13(624): eaaz1482.
4. Foguet C, Xu Y, Ritchie SC, Lambert SA, Persyn E, Nath AP, Davenport EE, Roberts DJ, Paul DS, Di Angelantonio E, Danesh J, Butterworth AS, Yau C, Inouye M (2022) Genetically personalised organ-specific metabolic models in health and disease. Nature Communications 13(1):7356.
5. Gustafsson J, Anton M, Roshanzamir F, Jörnsten R, Kerkhoven EJ, Robinson JL, Nielsen J (2023) Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data. Proceedings of the National Academy of Sciences 120(6):e2217868120.
6. Oftadeh O, Hatzimanikatis V (2024) Genome-scale models of metabolism and expression predict the metabolic burden of recombinant protein expression. Metabolic Engineering 84:109–116.
7. Li X, Walhout AJM, Yilmaz LS (2025) Enhanced flux potential analysis links changes in enzyme expression to metabolic flux. Molecular Systems Biology 21(4):413–445.
8. Tibocha-Bonilla JD, Gandhi V, Lieng C, Moyne O, Santibáñez-Palominos R, Zengler K (2025) Model of metabolism and gene expression predicts proteome allocation in Pseudomonas putida. npj Systems Biology and Applications 11(1):55.
9. Di Filippo M, Damiani C, Pescini D (2021) GPRuler: Metabolic gene-protein-reaction rules automatic reconstruction. PLOS Computational Biology 17(11):e1009550.
10. Seaver SMD, Liu F, Zhang Q, Jeffryes J, Faria JP, Edirisinghe JN, Mundy M, Chia N, Noor E, Beber ME, Best AA, DeJongh M, Kimbrel JA, D'haeseleer P, McCorkle SR, Bolton JR, Pearson E, Canon S, Wood-Charlson EM, Cottingham RW, Arkin AP, Henry CS (2021) The ModelSEED Biochemistry Database for the integration of metabolic annotations and the reconstruction, comparison and analysis of metabolic models for plants, fungi and microbes. Nucleic Acids Research 49(D1):D1555–D1555.
11. Zimmermann J, Kaleta C, Waschina S (2021) gapseq: informed prediction of bacterial metabolic pathways and reconstruction of accurate metabolic models. Genome Biology 22(1):81.
12. Pedram N, Rashedi H, Motamedian E (2020) A systematic strategy using a reconstructed genome-scale metabolic network for pathogen Streptococcus pneumoniae D39 to find novel potential drug targets. Pathogens and Disease 78(6): ftaa051.
13. Hirose Y, Zielinski DC, Poudel S, Rychel K, Baker JL, Toya Y, Yamaguchi M, Heinken A, Thiele I, Kawabata S, Palsson BO, Nizet V (2024) A genome-scale metabolic model of a globally disseminated hyperinvasive M1 strain of Streptococcus pyogenes. mSystems 9(9): e00736-24.
14. Zhang C, Sánchez BJ, Li F, Eiden CWQ, Scott WT, Liebal UW, Blank LM, Mengers HG, Anton M, Rangel AT, Mendoza SN, Zhang L, Nielsen J, Lu H, Kerkhoven EJ (2024) Yeast9: a consensus genome-scale metabolic model for S. cerevisiae curated by the community. Molecular Systems Biology 20(10):1134–1150.
15. Masid M, Ataman M, Hatzimanikatis V (2020) Analysis of human metabolism by reducing the complexity of the genome-scale models using redHUMAN. Nature Communications 11(1):2821.
16. Moyer DC, Reimertz J, Segrè D, Fuxman Bass JI (2025) MACAW: a method for semi-automatic detection of errors in genome-scale metabolic models. Genome Biology 26(1):79.
17. Pearcy N, Garavaglia M, Millat T, Gilbert JP, Song Y, Hartman H, Woods C, Tomi-Andrino C, Reddy Bommareddy R, Cho B-K, Fell DA, Poolman M, King JR, Winzer K, Twycross J, Minton NP (2022) A genome-scale metabolic model of Cupriavidus necator H16 integrated with TraDIS and transcriptomic data reveals metabolic insights for biotechnological applications. PLOS Computational Biology 18(5): e1010106.
18. Lázaro J, Jansen G, Yang Y, Torres-Acosta MA, Lye G, Oliver SG, Júlvez J (2022) Combination of Genome-Scale Models and Bioreactor Dynamics to Optimize the Production of Commodity Chemicals. Frontiers in Molecular Biosciences 9:855735.
19. Viana R, Dias O, Lagoa D, Galocha M, Rocha I, Teixeira MC (2020) Genome-Scale Metabolic Model of the Human Pathogen Candida albicans: A Promising Platform for Drug Target Prediction. Journal of Fungi 6(3):171.
20. Lewis JE, Kemp ML (2021) Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. Nature Communications 12(1):2700.
21. Proffitt C, Bidkhori G, Lee S, Tebani A, Mardinoglu A, Uhlen M, Moyes DL, Shoaie S (2022) Genome-scale metabolic modelling of the human gut microbiome reveals changes in the glyoxylate and dicarboxylate metabolism in metabolic disorders. iScience 25(7):104513.
22. Siriwach R, Matsuda F, Yano K, Hirai MY (2020) Drought Stress Responses in Context-Specific Genome-Scale Metabolic Models of Arabidopsis thaliana. Metabolites 10(4):159.
23. Wang H, Robinson JL, Kocabas P, Gustafsson J, Anton M, Cholley P-E, Huang S, Gobom J, Svensson T, Uhlen M, Zetterberg H, Nielsen J (2021) Genome-scale metabolic network reconstruction of model animals as a platform for translational research. Proceedings of the National Academy of Sciences 118(30):e2102344118.
24. Bjørnholm KD, Del Gaudio F, Li H, Li W, Vazquez-Liebanas E, Mäe MA, Lendahl U, Betsholtz C, Nilsson P, Karlström H, Vanlandewijck M (2023) A robust and efficient microvascular isolation method for multimodal characterization of the mouse brain vasculature. Cell Reports Methods 3(3):100431.

25. Ye L, Park JJ, Peng L, Yang Q, Chow RD, Dong MB, Lam SZ, Guo J, Tang E, Zhang Y, Wang G, Dai X, Du Y, Kim HR, Cao H, Errami Y, Clark P, Bersenev A, Montgomery RR, Chen S (2022) A genome-scale gain-of-function CRISPR screen in CD8 T cells identifies proline metabolism as a means to enhance CAR-T therapy. Cell Metabolism 34(4):595-614.e14.

26. Freischem LJ, Barahona M, Oyarzún DA (2022) Prediction of gene essentiality using machine learning and genome-scale metabolic models. IFAC-PapersOnLine 55(23):13–18.

27. Romano S, Wirbel J, Ansorge R, Schudoma C, Ducarmon QR, Narbad A, Zeller G (2025) Machine learning-based meta-analysis reveals gut microbiome alterations associated with Parkinson's disease. Nature Communications 16(1):4227.

28. Lachance J-C, Lloyd CJ, Monk JM, Yang L, Sastry AV, Seif Y, Palsson BO, Rodrigue S, Feist AM, King ZA, Jacques P-É (2019) BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. PLOS Computational Biology 15(4):e1006971.

29. Pio G, Mignone P, Magazzù G, Zampieri G, Ceci M, Angione C (2022) Integrating genome-scale metabolic modelling and transfer learning for human gene regulatory network reconstruction. Bioinformatics 38(2):487–493.

30. Hasibi R, Michoel T, Oyarzún DA (2024) Integration of graph neural networks and genome-scale metabolic models for predicting gene essentiality. npj Systems Biology and Applications 10(1):24.