doi 10.18699/bgrs2024-2.1-17

Analysis of context-specific genome-scale models for metabolism of fast- and slow-growing chicken breeds

Kaplan V.S.[1]*, Kulyashov M.A.[1], Volyanskaya A.R.[2], Yevshin I.S.[1], Shagimardanova E.I.[3], Gusev O.A.[3, 4, 5], Kolpakov F.A.[1], Akberdin I.R.[1]

[1] *Department of Computational Biology, Scientific Center for Genetics and Life Sciences, Sirius University of Science and Technology, Sirius, Russia*

[2] *Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia*

[3] *Regulatory Genomics Research Center, Institute of Fundamental Medicine and Biology, Kazan Federal University, Kazan, Russia*

[4] *Life Improvement by Future Technologies (LIFT) Center, Moscow, Russia*

[5] *Intractable Disease Research Center, Juntendo University, Tokyo, Japan*

* *kaplanrimi@gmail.com*

Key words: chicken metabolism; genome-scale metabolic models; context-specific models; differential expression genes; differential expression fluxes; flux balance analysis

Motivation and Aim: Synthetic biology, a rapidly advancing field of modern biology, focuses on constructing new organisms with specified properties through genome synthesis or deep reorganization using genetic engineering, bioengineering, systems biology and bioinformatics.Systems biology, integrating omics data, detailed analysis of gene interactions and metabolic modeling, offers a promising strategy for enhancing biotechnological properties [1]. Genome-scale metabolic models (GEMs) have become a powerful systems biology tool for understanding the metabolism of organisms. Furthermore, context-specific genome-scale models enable the study of metabolism under specific environmental or physiological conditions. By integrating transcriptomic data into a metabolic model, we can pinpoint specific metabolic reactions and genes associated with growth performance providing valuable insights for targeted breeding and nutritional interventions. Such a model makes it possible to predict *in silico* genetic modifications (gene knockouts, promoter modifications) required to increase the growth rate of a cell culture or even whole organism, the yield of the target product when the cell culture grows under optimal conditions and depending on different substrates or nutrition if it is an animal organism. Despite significant advancements in genetic selection and feeding practices, the molecular mechanisms underlying growth differences among chicken breeds remain poorly understood [2].

The primary aim of this study is to investigate the metabolic variability between fast- and slow-growing chicken breeds through the analysis of context-specific genome-scale models (csGEMs). We assume that the differential growth rates observed between fast- and slow-growing chicken breeds are governed by distinct metabolic signatures, characterized by variations in metabolic pathways, enzyme activities, and regulatory mechanisms. Essentially, we propose csGEMs that provide an opportunity to uncover unique metabolic features associated with growth performance in chickens.

Methods and Algorithms: The F2 generation from crosses between Russian White and Cornish breeds has a high diversity in growth rate, reaching a live weight of 1.5 to 2.5 kg by the fourth week. 12 slow-growing and 12 fast-growing chickens from the F2 generation were selected. We examined the transcriptome profiling of these 24 chickens by

CAGE-seq in heart, liver, kidney, brain, breast and leg muscles. CAGE-seq allows one the identification of transcriptional start sites and their activity level. 4.5 billion paired-end reads of 150 nucleotides in length were sequenced by MGI sequencer. Then we performed quality control on the raw sequencing data to filter out low-quality reads and adapters. A chicken pan-genome representing a GRCg6a reference assembly and an additive of 159 million base pairs as short contigs was created, from which we analyzed CAGE-seq data at the next steps. The reads were mapped to the chicken pangenome. The bowtie2 [3] was used for mapping RNA-seq data to the genome. We then normalized transcript abundance using the Transcript Per Million kilobase (TPM) method. These expression profiles were integrated into the iES1300 metabolic model [4] using the RIPTiDe algorithm [5], resulting in the generation of four context-specific models tailored to the respective chicken breeds and tissue types. We employed Escher [6] to construct metabolic maps for the context-specific models. The flux distribution ratios of these context-specific models, comparing fast to slow-growing breeds for each tissue, were simulated. Key and significantly different values of the same flux between two types of breeds were identified as differential expressed fluxes (DEFs). DEFs are characterized by flux changes between fast and slow breeds exceeding 1.5 times, disappearance, appearance, or reverse direction. Five types of DEFs were distinguished: Upper Regulated (accelerated), Down Regulated (decreased), Off (disappeared), On (appeared), and Reversed. Furthermore, we performed a comparative analysis of metabolic maps for context-specific models built on two different genome assemblies by looking for the ratio between the difference in flux values of the fast-growing to slow-growing pangenomic GRCg6a-based model to the same difference for the GRCg6a-based model. We also searched for different expression genes (DEGs) in fast and slow growing chickens based on CAGE-seq data and filtered them by p-adj<0.05. For this purpose, the number of reads in the identified transcriptional starts was counted and differential transcriptional starts were found by DESeq2 method [7], to which known genes were then compared according to Refseq annotation.

We use a refined GEM of a chicken cell named iES1300 [4], consisting of 2427 reactions, 2569 metabolites, and 1300 genes. We modified the model by adding important missing pseudo-reactions, metabolite formulas in order to test mass and charge balancing. Such manipulations of the model took place using Python library CobraPy [8]. Then we integrated TPM in the model using a specific tool, RIPTiDe [5] in order to construct context-specific models. The flux distributions of these models were compared between each other and analyzed; these fluxes were compared to previously obtained DEGs.

Results: We enhanced the model by incorporating 513 demand reactions, formulating metabolite formulas for 2569 metabolites, creating reactions, and ensuring mass-charge balance for 866 reactions from 916 non-balanced reactions. Evaluation through Memote test-suite [9] revealed an increase in the consistency score from 18 % in the original iES1300 model to 36 % in the modified version. Mass balance of the modified model became equal to 99.0 %, charge balance 98.8 % and unbounded flux in default medium 74.0 %, whereas mass balance of original model was 0.0 %, charge balance 97.2 % and unbounded flux in default medium 69.8 %. No changes were found in growth rate between the original and modified models, but secretion and uptake rates had pronounced differences. The constructed csGEMs were tested for correlation with transcriptome data and changes in growth rates compared to the GEM (Table 1).

Table 1. Correlation of csGEMs with TPM and the difference between the growth rate of the csGEMs and GEM

	Changes in growth rate in comparison with the GEM	r, correlation with transcriptomic data	p-value
fast-leg csGEM	No change	0.173	0.002
slow-leg csGEM	to ~0.0111 from ~0.0342	0.229	<0.001
fast-liver csGEM	No change	0.106	0.05
slow-liver csGEM	to ~0.0127 from ~0.0342	0.208	0.001

Visual and quantitative analysis of the metabolic map with simulation results enabled us to identify the strongest changes in the distribution of fluxes which were observed in the glycolysis and Krebs cycle (Fig. 1).

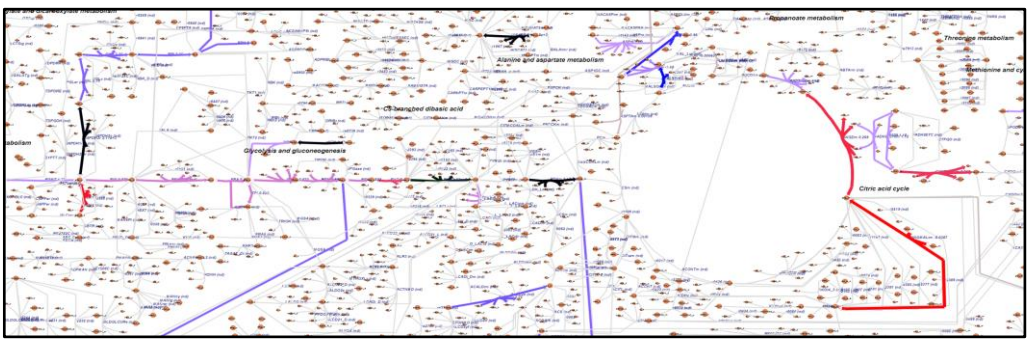

Fig. 1. A visualized part of the metabolic map with flux ratios for context-specific models based on GRCg6a and pangenomic GRCg6a assembled models

To check the quality of the FBA results we compared the results of DEFs and DEGs for leg muscle tissues. Additionally, the resulting DEFs for the GRCg6a- and pangenomic GRCg6a-based assemblies for leg muscle tissues were compared to each other. As a result, we found 28 out of 320 DEFs common in both types of context-specific models. Moreover, the resulting filtered DEGs were also compared to each other for GRCg6a- and pangenomic GRCg6a-based assemblies. It was identified from the analysis that 65 out of 131 common genes in both datasets are in the direction of DEGs for leg muscle tissues. We then compared DEGs with genes associated with DEF responses and found 67 of 250 genes are common in both types of the csGEMs. We also compared the DEFs between the original and modified models, and found 100 of 207 DEFs common in type. *Conclusion*: According to the Memote reports, we got a more qualitative model and made it more suitable for further biological interpretation. As we hypothesized, context-specific models of fast-growing muscle tissues significantly differ from model predictions for slow-growing tissues in terms of the growth rate as an objective function of flux balance analysis. Our pan-genomics assembly allowed us to refine the results of csGEMs, especially for crucial metabolic pathways such as glycolysis and the Krebs cycle. Our investigation delineated and clarified the key metabolic pathways and regulatory nodes contributing to the disparate growth rates observed in fast- and slow-growing chicken breeds.

All the results obtained have been make publicly available on gitlab: https://gitlab.sirius-web.org/collaboration/Chicken/gem_g.gallus2024

Funding: This study was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant No. 075-15-2021-1344).

References

1. Esvelt K.M., Wang H.H. Genome-scale engineering for systems and synthetic biology. *Mol Syst Biol*. 2013;9:641. doi 10.1038/msb.2012.66
2. Volyanskaya A.R., Akberdin I.R., Kulyashov M.A. et al. A bird's-eye overview of molecular mechanisms regulating feed intake in chickens-with mammalian comparisons. *Anim Nutr*. 2024;17:61-74. doi 10.1016/j.aninu.2024.01.008
3. Langmead B., Salzberg S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-9. doi 10.1038/nmeth.1923
4. Salehabadi E., Motamedian E., Shojaosadati S.A. Reconstruction of a generic genome-scale metabolic network for chicken: Investigating network connectivity and finding potential biomarkers. PLoS One. 2022;17(3):e0254270. doi 10.1371/journal.pone.0254270
5. Jenior M.L., Moutinho T.J. Jr., Dougherty B.V., Papin J.A. Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput Biol*. 2020;16(4):e1007099. doi 10.1371/journal.pcbi.1007099
6. King Z.A., Dräger A., Ebrahim A., Sonnenschein N., Lewis N.E., Palsson B.O. Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways. *PLoS Comput Biol*. 2015;11(8):e1004321. doi 10.1371/journal.pcbi.1004321
7. Love M.I., Huber W., Anders S. Differential analysis of count data – the DESeq2 package. *Genome Biol*. 2014;15:10-1186
8. Ebrahim A., Lerman J.A., Palsson B.O., Hyduke D.R. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol*. 2013;7:74. doi 10.1186/1752-0509-7-74
9. Lieven C., Beber M.E., Olivier B.G. et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol*. 2020;38(3):272-276. doi 10.1038/s41587-020-0446-y