

Компьютерное моделирование биотехнологических показателей аквакультуры на примере предсказания выживаемости сибирского осетра

Котов И.И.^{1*}, Байрамкулов Д.Д.¹, Орлов Ю.Л.^{1, 2**}

¹ Первый МГМУ им. И.М. Сеченова Минздрава России (Сеченовский Университет), Москва, Россия

² Институт цитологии и генетики СО РАН, Новосибирск, Россия

* mr_lorr_40t@mail.ru; ** orlov@bionet.nsc.ru

Ключевые слова: медицинская информатика; медицинская генетика; заболевания глаз; миопия; базы данных; генные онтологии

Мотивация и цель: Глобальные вызовы для биотехнологии связаны с возрастающей численностью населения планеты, что требует искать пути повышения производительности сельского хозяйства, улучшать сорта растений и породы животных, в том числе и рыб. Однако традиционные методы селекции дают ограниченный ответ на решение проблем, связанных с продовольственной безопасностью. По расчетам исследователей для обеспечения нормального уровня питания населения планеты в 2025 г. необходимо будет увеличить объем пищевой продукции в 2 раза [1], а вылов рыб в океанах – в 7 раз. В настоящее время аквакультура является самым быстрорастущим продовольственным сектором в мире – уже сейчас она поставляет почти половину съедобной рыбы, обеспечивая около 17 кг на душу населения. В мировой аквакультуре выращивают свыше 230 видов рыб, что намного больше по сравнению с сельскохозяйственными животными. Значительный вклад в развитие аквакультуры вносят методы современной биотехнологии. В нашей работе мы рассматриваем математические модели роста сибирского осетра при разведении в искусственных водоемах, в зависимости от структуры корма. Практическая цель данной работы – увеличить процентный состав выживаемости молодняка, используя общедоступные данные секвенирования следующего поколения вместе с метаболомными данными, для минимизации влияния фактор окружающей среды, а также внутренних болезней, на рост и развитие молодняка осетра сибирского.

Цели компьютерного исследования заключались в том, чтобы: установить ассоциации между различными модальными данными; построить модель прогнозирования признаков и оценить их точность; оценить вклад каждого информационного уровня, в прогнозирование различных признаков; получить новое представление о механизмах, лежащих в основе эффективности кормов.

Методы и алгоритмы: Определялись ключевые показатели, определяющие влияние выбора корма на рост и развитие мышечной массы. Для этого были рассмотрены 14 разнообразных слоёв мультиномики и клинических ковариат, чтобы оценить эффективность использования кормов (FE) и связанные с ней эффективности производительности рыбы *Oncorhynchus tshawytscha* [2, 3].

Осетр Сибирский играет важную составляющую в создании различных продуктов питания. Однако средний возраст фертильности женской особи составляет более пяти, а то и десяти лет, в то же время выживаемость мальков составляет не более

двадцати процентов от первоначального количества. В связи с этим актуальной задачей является увеличения выживаемости и прироста молодняка, в искусственно созданных «Осетровых фермах».

Показатель кормовой эффективности (FE) – показатель, позволяющий выявить эффективность использования корма, для продуктивных потребностей организма. Потребности могут выражаться по-разному: рост, связанная с этим метаболическая активность-направленная на увеличение мышечной массы. Также показатель показывает соотношение между потреблением пищи и приростом мышечной массы, что в свою очередь является и выражается как коэффициент конверсии корма (FCR). Важно отметить инверсионность коэффициентов: низкие значения FCR, соответствуют высоким значениям FE, в то же время высокие значения FCR соответствуют низким значениям FE. Важно правильно оценивать значения FCR, поскольку на практике, в промышленных системах, зачастую уровень затрат на рыбные корма составляет 50–60 %, от общего бюджета, что в свою очередь может превышать требуемый минимум, достаточный для требуемых значения FE.

Результаты: Собрана научная литература по генам роста осетровых. Представлена математическая модель роста рыбы в зависимости от структуры питания. FE является достаточно сложным показателем, зависящим от различных связанных факторов, такие как социальное поведение, потребление корма, его переваривание а также ассимиляция питательных веществ, метаболическая активность и генетика. Выявлено что виды и высоким значениям FE, питаются маленькими порциями, тем самым эффективнее расходуя полученный белок, углеводы и липиды, что в свою очередь приводит к наилучшему конечному результату-высоким показателям мышечной массы, в отношении потребляемого корма.

Чтобы рассчитать значение FE для данных мультиомики и клинических ковариат, использовался вариант разновидности алгоритма дерева решений случайного леса (random forest – RF). RF-метод машинного обучения на основе ансамбля, способный выявлять сложные закономерности зависимости между фенотипическими признаками и ковариатами, благодаря чему широко используется для задач классификации и регрессии. Устойчивость к выбросам позволяет обрабатывать гетерогенные типы данных, без потребности к их нормализации, что в свою очередь делает его устойчивым к переобучению, а также позволяет распознавать нелинейные связи и вычислять их характеристики. Однако текущая модель RF, ограничивалась одной матрицей, при объединении данных мультиомики и не предполагается на распределения данных. Текущую проблему решает представленная модель RF[1], рандомизирую выбор слоев для каждого типа узла, а также сохраняя распределение отдельных уровней данных. одного и того же образца (мышечная метаболомика и мышечная протеомика). Кроме-того установлена значимая корреляция между биомаркёрами крови и физическим здоровьем. Были выявлены отрицательные значения FE, при наличие мышечного глицерина, что может указывать на высокие показатели скорости использования субстратов липолиза, в свою очередь более низкие показатели FE, могут свидетельствовать о использовании глицерина в качестве субстрата. Интересно что глицерин является наиболее распространённым компонентом кормовой добавки, вытесняя аминокислоты и перенаправляя метаболический

поток в мышцах в сторону глюконеогенеза, тем самым открывая возможность использования аминокислот в полезных для организма процессах.

Выводы: В результате оценки собранных слоёв, было обнаружено значительное соответствие между многими слоями. Наибольшее соответствие было выявлено между слоями, собранными с использованием одного подхода (метаболические плазменные показатели и мышечные), а также между слоями, полученными из одного и того же образца (мышечная метаболомика и мышечная протеомика). Кроме-того установлена значимая корреляция между биомаркерами крови и физическим здоровьем. Были выявлены отрицательные значения FE, при наличии мышечного глицерина, что может указывать на высокие показатели скорости использования субстратов липолиза, в свою очередь более низкие показатели FE, могут свидетельствовать о использовании глицерина в качестве субстрата. Интересно что глицерин является наиболее распространённым компонентом кормовой добавки, вытесняя аминокислоты и перенаправляя метаболический поток в мышцах в сторону глюконеогенеза, тем самым открывая возможность использования аминокислот в полезных для организма процессах.

Вывод текущего исследования строиться на важности индивидуального измерения показателей FE, что в свою очередь позволяет персонализировано оптимизировать затраты под выбранного рыбного представителя. Это позволит грамотно распределять выделенные ресурсы бюджета на кормовые добавки, тем самым контролируя производственных выход здорового и эффективного метаболитические поколения молодых осетров. Данный подход является достаточно дорогостоящим, поскольку требует персонализированного исследования, анализа омики, в том числе и метаболомики, а также проведение секвенирования следующего поколения (NGS).

Благодарности: Авторы благодарны сотрудникам аквафермы «Осетровые фермы», Романовское хозяйство, Калужская область, за организацию работы.

Computer modeling of biotechnological indicators of aquaculture on the example of predicting the survival rate of Siberian sturgeon

Kotov I.I.^{1*}, Bayramkulov D.D.¹, Orlov Y.L.^{1, 2**}

¹ Sechenov First Moscow State Medical University (Sechenov University), Moscow, Russia

² Institute of Cytology and Genetics SB RAS Novosibirsk, Russia

* mr_lorr_40t@mail.ru, ** orlov@bionet.nsc.ru

Key words: medical informatics; medical genetics; eye diseases; myopia; databases; gene ontologies

Motivation and Purpose: Global challenges for biotechnology are related to the increasing world population, which requires finding ways to increase agricultural productivity, improve plant varieties and animal breeds, including fish. However, traditional breeding methods provide a limited answer to solving problems related to food security. According to researchers' calculations, to ensure a normal level of nutrition of the world population in 2025, it will be necessary to increase the volume of food production by 2 times [1], and fish catch in the oceans – by 7 times. Aquaculture is currently the fastest growing food sector in the world – already supplying nearly half of

the world's edible fish, providing about 17 kg per capita. More than 230 species of fish are farmed in global aquaculture, a far greater number than farm animals. The methods of modern biotechnology make a significant contribution to the development of aquaculture. In our work we consider mathematical models of growth of Siberian sturgeon during breeding in artificial reservoirs, depending on the feed structure. The practical goal of this work is to increase the percentage of juvenile survival using publicly available next generation sequencing data together with metabolomic data to minimize the influence of environmental factors, as well as internal diseases, on the growth and development of juvenile Siberian sturgeon.

The objectives of the computerized study were to: establish associations between different modal data; build a trait prediction model and assess its accuracy; evaluate the contribution of each information level, to the prediction of different traits; and gain new insights into the mechanisms underlying feed efficiency.

Methods and Algorithms: Key indicators determining the effect of feed selection on muscle mass growth and development were identified. For this purpose, 14 diverse layers of multomics and clinical covariates were considered to assess feed utilization efficiency (FE) and associated performance efficiencies of fish *Oncorhynchus tshawytscha* [2, 3].

Siberian sturgeon plays an important component in the development of various food products. However, the average age of fertility of female fish is more than five or even ten years, at the same time the survival rate of fry is not more than twenty percent of the original number. In this regard, the urgent task is to increase the survival rate and growth of young fish in artificially created “Sturgeon farms”

Feed efficiency (FE) is an indicator that reveals the efficiency of feed utilization for the productive needs of the organism. The needs can be expressed in different ways: growth, associated metabolic activity-directed to increase muscle mass. It also shows the relationship between food intake and muscle mass gain, which in turn is and is expressed as feed conversion ratio (FCR). It is important to note the inversion of the ratios: low FCR values correspond to high FE values, while high FCR values correspond to low FE values. It is important to correctly estimate FCR values, because in practice, in industrial systems, the level of fish feed costs is often 50–60 % of the total budget, which in turn may exceed the required minimum, sufficient for the required FE values.

Results: Scientific literature on sturgeon growth genes was collected. A mathematical model of fish growth in relation to feeding pattern is presented. FE is a rather complex index depending on various related factors such as social behavior, feed intake, feed digestion as well as nutrient assimilation, metabolic activity and genetics. Species with high FE values are found to eat smaller portions, thus more efficiently utilizing the protein, carbohydrates and lipids obtained, which in turn leads to the best end result-high muscle mass in relation to feed intake

To calculate the FE value for multivariate data and clinical covariates, a variant of the random forest (RF) decision tree algorithm was used.

An ensemble-based RF machine learning method capable of detecting complex patterns of dependence between phenotypic traits and covariates, making it widely used for classification and regression tasks. Its robustness to outliers allows it to handle heterogeneous data types without the need for normalization, which in turn makes it robust to overfitting, and also allows it to recognize nonlinear relationships and compute their characteristics. However, the current RF model, was limited to a single matrix, when combining multi-omics data and is not assumed on data distributions. The current

problem is solved by the presented RF model [1], by randomizing the choice of layers for each node type, and preserving the distribution of individual data layers. of the same sample (muscle metabolomics and muscle proteomics).

In addition, there was a significant correlation between blood biomarkers and physical health. Negative FE values were found in the presence of muscle glycerol, which may indicate high rates of lipolysis substrate utilization, while lower FE values may indicate the use of glycerol as a substrate. Interestingly, glycerol is the most common component of feed additives, displacing amino acids and redirecting the metabolic flux in muscle toward gluconeogenesis, thus opening up the possibility of utilizing amino acids in beneficial processes.

Conclusions: From the evaluation of the collected layers, significant concordance was found between many layers. The highest concordance was found between layers collected using the same approach (plasma metabolomic indices and muscle metabolomic indices) and between layers derived from the same sample (muscle metabolomics and muscle proteomics). In addition, a significant correlation was found between blood biomarkers and physical health. Negative FE values were found in the presence of muscle glycerol, which may indicate high rates of lipolysis substrate utilization, while lower FE values may indicate the use of glycerol as a substrate. Interestingly, glycerol is the most common component of feed additives, displacing amino acids and redirecting the metabolic flux in muscle towards gluconeogenesis, thus opening up the possibility of utilizing amino acids in beneficial processes. The conclusion of the current study is based on the importance of individual measurement of FE values, which in turn allows personalized cost optimization for the selected fish representative. This will allow a competent allocation of allocated budget resources for feed additives, thus controlling the production output of healthy and efficient metabolic generation of young sturgeons. This approach is quite expensive as it requires personalized research, omics analysis, including metabolomics, and next generation sequencing (NGS).

Acknowledgments: The authors are grateful to the staff of Sturgeon Farms Aquafarm, Romanovskoye farm. Tver region, for organizing the work.

Список литературы/References

1. Young T., Laroche O., Walker S.P. et al. Prediction of feed efficiency and performance-based traits in fish via integration of multiple omics and clinical covariates. *Biology*. 2023;12(8):1135. doi 10.3390/biology12081135
2. Hung S.S.O. Recent advances in sturgeon nutrition. *Anim Nutr*. 2017;3:191-204. doi 10.1016/j.aninu.2017.05.005
3. Zhu Y. et al. Metabolomics and gene expressions revealed the metabolic changes of lipid and amino acids and the related energetic mechanism in response to ovary development of chinese sturgeon (*Acipenser Sinensis*). *PLoS One*. 2020;15:e0235043. doi 10.1371/journal.pone.0235043