УДК 51-76, 519.233, 575.162

APPROACHES TO ANALYSIS OF GENOTYPE AND PHENOTYPE RELATION WITH QTL METHODS

E.Yu. Furta[a], I.M. Shabalina[a]

[a]Petrozavodsk State University, Petrozavodsk, 185910, Russian Federation
Corresponding author: lena-furta@rambler.ru

Abstract
The paper describes methods of QTL-analysis in the research of the genotype influence. QTL-analysis is a statistical method that connects phenotypic data with genotype, enables the determination of the exact localization and gene influence power. The QTL mapping idea is to phenotype observation and identification of the genome region on which the genotype is associated with the phenotype. With the help of molecular-genetic markers, molecular maps of individual chromosomes and genomes are made, genes and QTLs mapping are performed on them. Thus, genes with the greatest connectivity to phenotype were identified. The correlation between genotype and phenotype is studied across the full genome of the individual. The data provided by the Laboratory of Molecular Genetics of the innate immunity of Petrozavodsk State University were the initial data in this research. The essence of the project, carried out jointly with the laboratory, is to study the arrays of genetic information to identify and model the relationships between the genotype and the phenotype of biological organisms. The second-generation mice hybrids of lines C57BL/6 and MOLF were involved in the experiment. Genotyping and phenotyping were conducted based on sequencing data (determination of amino acid and nucleotide sequence) of matrix RNA. The practical result of the work is the identification of chains of activated genes, under the influence of which the cells of the tissues and organs under research die (apoptosis). The result of the study is a technique for analyzing the relationship between a phenotype and a genotype, through which groups of significant phenotypes were identified.

Keywords
QTL, molecular maps, genotype, phenotype, biological pathways

ПОДХОДЫ К АНАЛИЗУ ВЗАИМОСВЯЗИ ГЕНОТИПА И ФЕНОТИПА ПРИ ПОМОЩИ QTL-АНАЛИЗА

Е.Ю. Фурта[a], И.М. Шабалина[a]

[a] Петрозаводский государственный университет, Петрозаводск, 185910, Российская Федерация
Адрес для переписки: lena-furta@rambler.ru

Аннотация
Описаны методы QTL-анализа в исследовании влияния генотипа на фенотип. QTL-анализ основан на внутривидовой изменчивости, которая приводит к количественным изменениям исследуемого признака. С его помощью можно выявлять участки хромосом, где расположены гены или группы тесно сцепленных генов, обнаруживающих значительное количественное влияние на признак, и оценить такое влияние. Необходимым условием проведения QTL-анализа является построение карты сцепления. Карты сцепления (молекулярно-генетические карты, linkage-карты) отражают позицию маркеров и относительные генетические расстояния между маркерами вдоль хромосом. Исходные данные для исследования получены лабораторией молекулярной генетики врожденного иммунитета Петрозаводского государственного университета. Суть проекта, проводимого совместно с лабораторией, заключалась в исследовании массивов данных генетической информации для выявления и моделирования взаимосвязей между

генотипом и фенотипом биологических организмов. В эксперименте использованы гибриды мышей второго поколения линий C57BL/6 и MOLF. Генотипирование и фенотипирование проводилось на основе данных секвенирования (определение аминокислотной и нуклеотидной последовательности) матричной РНК. Практическим результатом работы является выявление цепочек активированных генов, под влиянием которых происходит отмирание клеток исследуемых тканей и органов (апоптоз). Для поиска таких генов использован метод генетического анализа, т.е. скрещивались особи противоположного фенотипа и анализировалось получаемое потомство. Использованы методы прикладной и математической статистики. Результатом исследования стала разработанная методика анализа связи фенотипа и генотипа, с помощью которой выявлены группы значимых фенотипов.

Introduction

Nowadays genetic studies play a huge role in medicine. Many genetic labs develop methods that will reliably link the manifestations of various diseases with specific areas of the genome. The genotype manifestation characteristics in the organism individual development process under the environment influence is called a phenotype. Most phenotypic traits are quantitative by nature. Quantitative features are associated with DNA regions, either containing genes responsible for the expression of the quantitative trait or linked to them. Such regions of DNA are called quantitative trait Loci (QTLs). Quantitative characteristics refer to characteristics that differ in their degree of expression and can be attributed to polygenic effects, so they are the product of two or more genes [1]. Loci of quantitative traits provide continuous variability of the trait in the population.

Previous research was carried out to develop approaches to genetic information analysis as applied to septic shock sensitivity investigation on the basis of a mice model. The genetic analysis method can be applied to search the gene responsible for tumor necrosis factor (TNF) sensitivity using two mice lines with opposite TNF sensitivity. The purpose of the investigation was to find the marker or the group of markers responsible for sensitivity or resistance to TNF. The search is based on the hypothesis that a phenotype corresponds to a genotype. Correlation between a genotype and sensitivity is a complex problem which cannot be solved by a single marker effect. We used two types of patterns building methods in the study. The obtained results reveal some combination of markers which might significantly affect the genotype. At the first stage the markers with significant positive correlations with resistance to TNF were defined. Such correlations were detected for the 1st, 8th and 11th chromosome markers. We found that the developed patterns have a great variety. Implementation of the hierarchic patterns building method showed that even patterns with high occurrence frequency matched to a few objects. The results provided evidence of significant correspondence of the 1st and 11th chromosome markers with resistance to TNF. The same results were demonstrated by the patterns built according to the graph method [2].

The data provided by the Laboratory of Molecular Genetics of the innate immunity of Petrozavodsk State University were the initial data in this research. This data contains information about 50 individuals. At the same time there are 16186 values describing phenotype (quantitative characteristic of each individual). These values were obtained in the genetic laboratory by sequencing RNA. The genotype values have been obtained from 129 polymorphic markers. These markers are short chromosome sections, markers divide chromosome into loci. The length of the marker corresponds to the genotype of the individual and it was determined by polymerase chain reaction and electrophoresis with special equipment in labs. All data was obtained during the experiment and processed by genetic scientists at the stage of preliminary data preparation.

The study uses information contained on 20 chromosomes. The genotype can take the following values at each marker:
− H is a heterozygous genotype;
− M is a resistant genotype;
− B is a sensitive genotype.

It is assumed that the genes responsible for manifesting the phenotypic trait are in loci for which the highest LOD (Logarithm of Odds) values were obtained using QTL-analysis methods.

QTL-analysis is a statistical method that connects phenotypic data with genotypic data and gives the possibility to determine the exact localization, quantity, effect and interaction of QTLs. The main idea of QTL mapping is to observe manifestations of a phenotype, and the identification of the genome region in which the genotype is associated with the phenotype. With the help of molecular-genetic markers, molecular maps of individual chromosomes and genomes are made, genes and quantitative traits loci mapping is performed on them [3].

The most obvious way to separate phenotypes into groups is the use of cluster analysis. Unfortunately, when it comes to genetic research, the work is done with a huge amount of data where classical methods of cluster analysis, such as k-means or hierarchical clustering, do not work on large data. In this connection, it is necessary to transform somehow the obtained data before using the methods of cluster analysis.

In this research biological pathways were used. Most common types of biological pathways:
1. metabolic pathway;

2. genetic pathway;
3. signal transduction pathway.

A biological pathway is a series of interactions among molecules in a cell that leads to a certain product or a change in a cell. Such pathway can trigger the assembly of new molecules, such as fat or protein. Pathways can also turn genes on and off or spur a cell to move. Some of the most common biological pathways are involved in metabolism, the regulation of gene expression and the transmission of signals. Pathways play key role in advanced studies of genomics [4].

Metabolic pathways were considered at this stage of work. In biochemistry, a metabolic pathway is a linked series of chemical reactions occurring within a cell [5]. These pathways can be of two types: anabolic pathways and catabolic pathway. Anabolic pathways synthesize molecules with the utilization of energy. Catabolic pathways break down of complex molecules by releasing energy in process [6]. Each metabolic pathway consists of a series of biochemical reactions that are connected by their intermediates: the products of one reaction are the substrates for subsequent reactions, and so on. Metabolic pathways are often considered to flow in one direction. Although all chemical reactions are technically reversible, conditions in the cell are often such that it is thermodynamically more favorable for flux to proceed in one direction of a reaction [7]. Metabolic pathways are regulated by signal pathways. The next step of this research is to find the signal pathways in customizing classification results. Signal transduction is the process by which a chemical or physical signal is transmitted through a cell as a series of molecular events, most commonly protein phosphorylation catalyzed by protein kinases, which ultimately results in a cellular response [8].

Methods of QTL-analysis in the research of the genotype influence on phenotype

In compiling this section, information was used from the [9]. QTL-analysis is based on intraspecies variability, which leads to quantitative changes in the trait under study.

The requirement for QTL-analysis is the construction of a linkage map. The linkage maps (molecular genetics maps) indicate the position of markers and the relative genetic distances between markers along the chromosomes. With the linkage maps the location of genes and QTLs associated with the feature of interest is established, therefore, such maps are also called QTL-maps. Since genes and markers are separated when chromosomes are recombined, they can be analyzed in progeny and compared with parental forms. The closely located genes and markers are more often transmitted linked than the genes and markers located farther apart. There is a mixture of recombinant genotypes in the population. Their frequency can be used to calculate the genetic distance between markers. Analyzing the separation of markers, the relative position and distance between the markers can be set: the lower the recombination frequency between the markers, the closer they are located on the chromosome and vice versa. Mapping is used to convert the frequency of recombination into units of genetic distances, called centimorgans (cM). Thus, 1 cM corresponds to recombination frequency of 1 %. Markers with recombination frequency of 50 % or more are considered unlinked, i.e., located on different chromosomes or far apart on one chromosome.

The linkage between the markers (or between the marker and QTL) is defined as the ratio of the probability of this linkage to the probability of lack of linkage. This ratio is usually expressed in logarithmic form and called LOD. The LOD score at a marker is calculated as follows. First, consider the null hypothesis of no QTL, in which case, with y_i denoting the phenotype for individual i, $y_i \sim N\left(\mu, \sigma^2\right)$ (i.e., the phenotypes follow a single normal distribution, independent of the genotypes). We consider the likelihood function $L_0\left(\mu, \sigma^2\right) = P\left(\text{data}|\text{no }QTL, \mu, \sigma^2\right) = \prod_i \varphi\left(y_i, \mu, \sigma^2\right)$, where φ is the density of the normal distribution. We take as estimates of μ and σ^2 the values for which the likelihood is maximized (such estimates are called the maximum likelihood estimates, MLEs). For this model, the MLE of μ is simply the phenotype average, $\bar{y}$, and the MLE of σ^2 is $\dfrac{RSS_0}{n}$, where $RSS_0 = \sum_i \left(y_i - \bar{y}\right)^2$ is the null residual sum of squares and n is the sample size. The $\log_{10}$ likelihood for the null hypothesis is obtained by plugging in the MLEs; with a bit of algebra, it reduces to $-\dfrac{n}{2}\log_{10} RSS_0$.

Under the alternative hypothesis, that there is a QTL at the marker under test, we assume that $y_i \mid g_i \sim N\left(\mu_{g_i}, \sigma^2\right)$, where g_i is the genotype of individual i at the marker, μ_{AB} and μ_{AA} (AA and AB are groups which were divided by genotype) are the phenotype averages for the two genotype groups, and σ^2 is the residual variance (assumed to be the same in the two groups). The likelihood function is $L_1\left(\mu_{AA}, \mu_{AB}, \sigma^2\right) = P\left(\text{data}|QTL \text{ at marker}, \mu_{AA}, \mu_{AB}, \sigma^2\right) = \prod_i \varphi\left(y_i, \mu_{g_i}, \sigma^2\right)$. We again estimate the parameters by

maximum likelihood: the values for which L_1 achieves its maximum. The MLEs for the μ_i are simply the phenotype averages within the two genotype groups. The MLE of σ^2 is the pooled estimate, $\dfrac{RSS_1}{n}$, where $RSS_1 = \sum_i \left(y_i - \mu_i \right)^2$ is the residual sum of squares under the alternative. The $\log_{10}$ likelihood for the alternative hypothesis is obtained by plugging in the MLEs; it reduces to $-\dfrac{n}{2}\log_{10} RSS_1$.

Finally, the LOD score is the difference between the $\log_{10}$ likelihood under the alternative hypothesis and the $\log_{10}$ likelihood under the null hypothesis, and so we obtain the following:

$$\mathrm{LOD} = \frac{n}{2}\log_{10}\frac{RSS_0}{RSS_1}.$$

If values of LOD ≥ 3, they can be used for making linkage maps. LOD = 3 means that the linkage between markers is 1000 times more likely than its lack. The linked markers are grouped together into linkage groups that represent segments of chromosomes or whole chromosomes [9].

Three main methods are used to QTL detection: marker regression, interval mapping, and Haley-Knott regression.

LOD score is very similar to statistics of Analysis of variance (ANOVA). Of course, the simplest statistical method for QTL mapping is analysis of variance at marker loci. This approach suffers when there is appreciable missing marker genotype data and when the markers are widely spaced. Interval mapping, though more complicated and more computationally intensive, allows for missing genotype data [10].

Marker regression is the simplest method for detection of QTLs linked to a specific marker. The main drawback of this method is that the farther QTL is from the marker, the less likely it is to be detected, since there is a possibility of recombination between markers and QTL. A lot of DNA markers covering the entire genome at intervals of less than 15 cM can solve this problem [11].

The interval mapping analyzes the presence of QTL in the intervals between neighboring pairs of linked markers along all chromosomes. This method compensates for possible recombination between the marker and QTL and it is more advanced than the one-marker assay. For calculating LOD score in this case EM-algorithm is used [12].

The Haley–Knott regression combines interval mapping with linear regression. This is a more accurate and effective method compared to the first two methods. Especially its effectiveness is manifested when working with linked QTLs [13].

Approaches to the phenotypes classification

As above said, missed information is one of the most important QTL-analysis problems. When applying the marker regression method, individuals with lost information on markers are excluded from consideration. A graph can be plotted from the values obtained for each chromosome. This graph shows what areas of the chromosome genotype and phenotype are associated with maximum.

After plotting the graphs on the LOD values, the obtained curves can be used to predict the lost information. If different phenotypes graphs profiles are similar, they can be grouped. The most obvious way to break into groups of phenotypes is the use of cluster analysis. When it comes to genetic research, the work is done with a huge amount of data. Unfortunately, classical methods of cluster analysis, such as the k-means method or hierarchical clustering, do not work on large data. In this case, it is necessary to convert in some way the data before using cluster analysis methods.

Cluster analysis includes various classification algorithms. The purpose of cluster analysis is the distribution of analyzed objects over relatively homogeneous groups. Such groups are called clusters.

Before the start of the clustering procedure it is assumed that each sample object forms a separate cluster. The union of objects into clusters is made considering the distance between objects. The methods of cluster analysis are very sensitive to the selection of a measure of proximity between objects and between clusters.

A hierarchical approach to clustering implies the successive consolidation of smaller clusters into larger clusters. Convenience of such clustering lies in the fact that it is not necessary to determine the number of clusters in advance. When using hierarchical methods, the principles of the nearest or farthest neighbor are most often used. When visualizing hierarchical clustering results graphs are often used.

Hierarchical clustering "nearest neighbors" algorithm comes out of the matrix of distances between observations. The matrix distance between the clusters is defined by the "nearest neighbor" rule. At the first step of the algorithm, each observation is treated as a separate cluster. Next, at each step of the algorithm, the two closest clusters are combined, and the distance matrix is recalculated. The algorithm ends when all the original observations are combined into one cluster.

The "farthest neighbor" hierarchical clustering (or "complete-linkage clustering") differs only in the way

of calculating the distance between classes [14].

The k-means method is the simplest and often used method of cluster analysis. This method has a significant drawback: before the start of the clustering procedure, it is necessary to determine in advance the number of clusters k, to which the investigated objects will be divided. The clustering algorithm involves minimizing the total rms deviation of observations from the center of the cluster. By default, Euclidean distance is chosen as a measure of distance. The following steps are performed during the execution of the algorithm.

1. The initial distribution of objects on clusters. Randomly selected K observations, these observations are considered the centers of clusters.
2. For each observation, the closest cluster center is selected, initial clusters are formed.
3. Calculate the new means to be the centroids of the observations in the new clusters.

Steps 2 and 3 are iteratively repeated until the clusters stop changing [15].

As noted earlier, the classical methods of cluster analysis do not enable the work with large amounts of data. In this case, it is necessary to transform the original data in such a way that the clustering process is simplified, but the result of clustering can be suitable for further analysis. To perform this transformation, the obtained data were divided into groups on chromosomes: markers located on one chromosome were united and 20 groups were obtained. The values of LOD <3 were replaced by zeros for the stability of the classification. Next, in each group, the k-means clustering was started. Thus, the table was obtained with phenotypes in the rows and chromosome number in the columns. The cells of this table contained the cluster number into which the phenotype on this chromosome got. Then phenotypes with all cluster numbers equal to zero were excluded from the table, because these clusters are non-informative. Thereby the dimension of the original data table was reduced to 5823 rows (phenotypes) and 20 columns (chromosomes). Further the shortened table data was clustered by k-means, the number of clusters was set to five. These actions made it possible to identify the cluster with 261 elements. Most of these elements have phenotype profiles with peak on the seventh chromosome. You can see several examples of these profiles in Fig.1–4. (X – is the designation for the 20th chromosome)
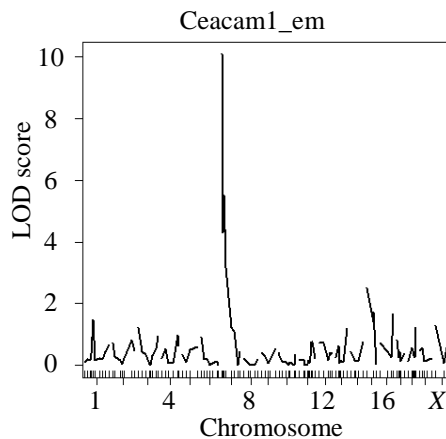

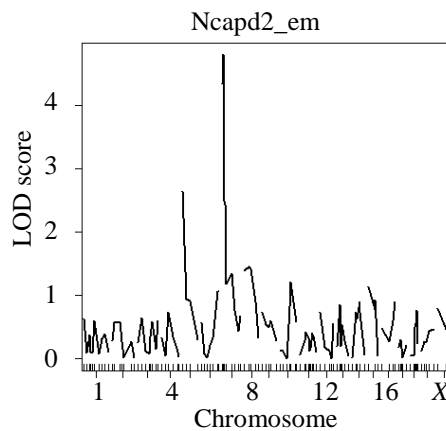Fig.1. Ceacam1 phenotype with peak on the seventh chromosome

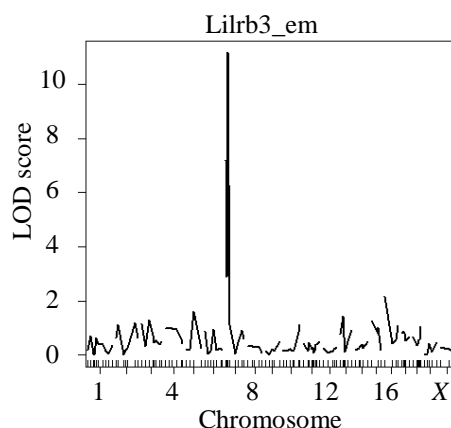Fig. 2. Ncapd2 phenotype with peak on the seventh chromosome

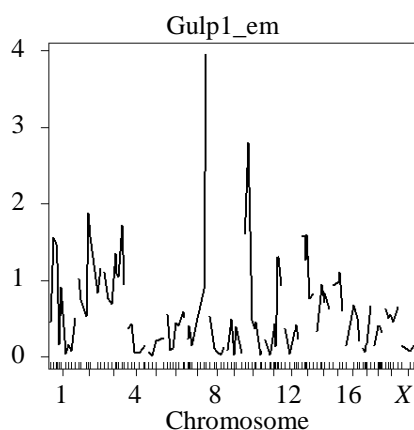Fig. 3. Lilrb3 phenotype with peak on the seventh chromosome

Fig. 4. Gulp1 phenotype with peak on the seventh chromosome

Investigation of metabolic pathways

These metabolic pathways were considered in more detail on the advice of geneticists:
− phospholipid biosynthesis I;
− ketolysis;
− ketogenesis;
− glycolysis III;
− serine biosynthesis;
− tRNA splicing.

Let us consider the stages of research on the pathway of synthesis of phospholipids (phospholipid biosynthesis I), which proved to be the most suitable of the studied ways.

The first step was to exclude all phenotypes that are not included in the pathway from the table with LOD values. Thus, the following phenotypes were remained in the new table: Gpam, Agpat9, Agpat1, Agpat4, Lclat1, Cds1, Cds2, Pgs1, Crls1, Ptdss1, Ptdss2, Pisd.

At the second stage it was necessary to track on which chromosomes the LOD values of selected phenotypes are significant. If there were none, then the phenotype was "noisy", where all LODs are insignificant. The following phenotypes from the previous stage were attributed to significant: Gpam, Lclat1, Agpat4, Cds1, Cds2, Crls1.

The third step was to compare the obtained significant phenotypes and studied metabolic pathway. The following observations can be made:
1. As can be seen from the Fig. 5 the pathway is branched because it describes synthesis process. There is at least one significant phenotype on each branch except the last right branch. Since all the components of this pathway are affected by the impact LPS it can be asserted about the involvement of the process of phospho-lipids biosynthesis in the response of cells to bacterial endotoxin which was injected to individuals at the beginning of the experiment.
2. All phenotypes, which we attributed to insignificant at this stage of the study, got into "noisy" clusters during the clustering stage. Thus, it can be argued that the modified clustering with pre-processing of data works correctly.
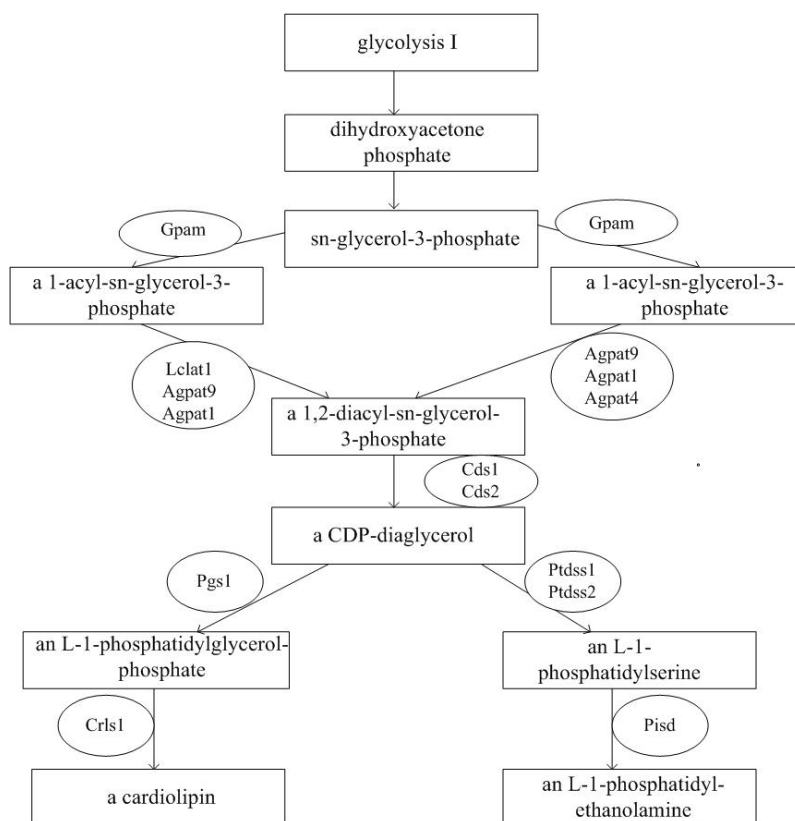
glycolysis I

dihydroxyacetone phosphate

Gpam — sn-glycerol-3-phosphate — Gpam

a 1-acyl-sn-glycerol-3-phosphate

a 1-acyl-sn-glycerol-3-phosphate

Lclat1 Agpat9 Agpat1

Agpat9 Agpat1 Agpat4

a 1,2-diacyl-sn-glycerol-3-phosphate

Cds1 Cds2

a CDP-diaglycerol

Pgs1

Ptdss1 Ptdss2

an L-1-phosphatidylglycerol-phosphate

an L-1-phosphatidylserine

Crls1

Pisd

a cardiolipin

an L-1-phosphatidyl-ethanolamine

Fig. 5. Phospholipid biosynthesis I – metabolic pathway

Conclusion

Thus, a methodology for applying QTL analysis was developed to study the effect of the genotype on the transcript with the mouse model used as an example. Several groups of clusters were obtained on each chromosome, which are significant from a genetic point of view. These clusters are sent to geneticists for study. In further studies, it is planned to develop approaches and techniques that allow for analyzing the association of the phenotype with the entire genome, to assess the profile of the whole phenotype, and not on each chromosome.

Also, an approach to comparison of the significant values of LOD and metabolic pathways was developed. In the future, it is planned to expand the base of the investigated biological pathways, not only metabolic, but also signal ones. It is also planned to improve the approach for searching the biological pathway in the initial data.

Литература

1. Miles C.M., Wayne M. Quantitative trait locus (QTL) analysis // Nature Education. 2008. V. 1. N 1. P. 208.
2. Volkova T., Furta E., Dmitrieva O., Shabalina I. Pattern building methods in genetic data processing // Journal on Selected Topics in Nano Electronics and Computing. 2014. V. 2. N 1. P. 2–6. doi: 10.15393/j8.art.2014.3041
3. Хлесткина Е.К. Молекулярные маркеры в генетических исследованиях и в селекции // Вавиловский журнал генетики и селекции. 2013. Т. 17. Т 4-2. С. 1044–1054.
4. Biological Pathway Fact Sheet - National Human Genome Research Institute. URL: https://www.genome.gov/27530687 (accessed 21.09.18).
5. Nelson D.L., Cox M.M. Lehninger Principles of Biochemistry. 5th ed. New York: W.H. Freeman, 2008.
6. Reece J.B., Urry L., Cain M.L., Wasserman S.A. et al. Campbell Biology. 9th ed. Boston: Benjamin Cummings, 2011. 143 p.
7. Cornish-Bowden A., Cardenas M.L. Irreversible reactions in metabolic simulations: how reversible is irreversible? / In: Animating the Cellular Map. Stellenbosch, South Africa: Stellenbosch University Press, 2000. P. 65–71.
8. Bradshaw R.A., Dennis E.A. (eds.) Handbook of Cell Signaling.

References

1. Miles C.M., Wayne M. Quantitative trait locus (QTL) analysis. *Nature Education*, 2008, vol. 1, no. 1, p. 208.
2. Volkova T., Furta E., Dmitrieva O., Shabalina I. Pattern building methods in genetic data processing. *Journal on Selected Topics in Nano Electronics and Computing*, 2014, vol. 2, no. 1, pp. 2–6. doi: 10.15393/j8.art.2014.3041
3. Khlestkina E.K. Molecular markers in genetic studies and breeding. *Vavilov Journal of Genetics and Breeding*, 2013, vol. 17, no. 4-2, pp. 1044–1054. (in Russian)
4. *Biological Pathway Fact Sheet - National Human Genome Research Institute*. URL: https://www.genome.gov/27530687 (accessed 21.09.18).
5. Nelson D.L., Cox M.M. *Lehninger Principles of Biochemistry*. 5th ed. New York, W.H. Freeman, 2008.
6. Reece J.B., Urry L., Cain M.L., Wasserman S.A. et al. *Campbell Biology*. 9th ed. Boston, Benjamin Cummings, 2011, 143 p.
7. Cornish-Bowden A., Cardenas M.L. Irreversible reactions in metabolic simulations: how reversible is irreversible? In *Animating the Cellular Map*. Stellenbosch, South Africa, Stellenbosch University Press, 2000, pp. 65–71.
8. Bradshaw R.A., Dennis E.A. (eds.) *Handbook of Cell Signaling*. 2nd ed. Amsterdam, Netherlands, Academic Press, 2010, 2875 p.

2nd ed. Amsterdam, Netherlands: Academic Press, 2010. 2875 p.

9. Broman K.W., Sen S. A Guide to QTL Mapping with R/qtl. Springer, 2009. 396 p.

10. Broman K.W. Review of statistical methods for QTL mapping in experimental crosses // Lab Animal. 2001. V. 30. N 7. P. 44–52.

11. Кузнецов В.В., Романов Г.А. молекулярно-генетические биохимические методы современной биологии растений. М.: Бином. Лаборатория знаний, 2015. 487 с.

12. Chen Z. The full EM algorithm for the MLEs of QTL effects and positions and their estimated variances in multiple-interval mapping // Biometrics. 2005. V. 61. N 2. P. 474–480. doi: 10.1111/j.1541-0420.2005.00327.x

13. Haley C.S., Knott S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers // Heredity. 1992. V. 69. N 4. P. 315–324. doi: 10.1038/hdy.1992.131

14. Айвазян С.А., Бухштабер В.М., Енюков И.С., Мешалкин Л.Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989. 608 с.

15. Айвазян С.А. Теория вероятности и прикладная статистика. М.: Юнити, 2001. 641 с.

9. Broman K.W., Sen S. *A Guide to QTL Mapping with R/qtl*. Springer, 2009, 396 p.

10. Broman K.W. Review of statistical methods for QTL mapping in experimental crosses. *Lab Animal*, 2001, vol. 30, no. 7, pp. 44–52.

11. Kuznetsov V.V., Romanov G.A. *Molecular-Genetic and Biochemical Methods in Modern Plant Biology*. Moscow, BINOM. Laboratoriya znanij Publ., 2015, 487 p. (in Russian)

12. Chen Z. The full EM algorithm for the MLEs of QTL effects and positions and their estimated variances in multiple-interval mapping. *Biometrics*, 2005, vol. 61, no. 2, pp. 474–480. doi: 10.1111/j.1541-0420.2005.00327.x

13. Haley C.S., Knott S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 1992, vol. 69, no. 4, pp. 315–324. doi: 10.1038/hdy.1992.131

14. Ajvazyan S.A., Buhshtaber V.M., Enyukov I.S., Meshalkin L.D. *Applied Statistics. Dimension Reduction*. Moscow, Finance and Statistics Publ., 1989, 608 p. (in Russian)

15. Ajvazyan S.A. *The Probability Theory and Applied Statistics*. Moscow, Yuniti Publ., 2001, 641 p. (in Russian)

Авторы

Фурта Елена Юрьевна – соискатель, Петрозаводский государственный университет, Петрозаводск, 185910, Российская Федерация, ORCID ID: 0000-0001-8105-6556, lena-furta@rambler.ru

Шабалина Ирина Михайловна – кандидат технических наук, доцент, доцент, Петрозаводский государственный университет, Петрозаводск, 185910, Российская Федерация, ORCID ID: 0000-0001-9411-935X, irinashabalina09@gmail.com

Authors

Elena Yu. Furta – applicant, Petrozavodsk State University, Petrozavodsk, 185910, Russian Federation, ORCID ID: 0000-0001-8105-6556, lena-furta@rambler.ru

Irina M. Shabalina – PhD, Associate Professor, Associate Professor, Petrozavodsk State University, Petrozavodsk, 185910, Russian Federation, ORCID ID: 0000-0001-9411-935X, irinashabalina09@gmail.com