INTRODUCTION

The number of sequenced sequences is growing at a tremendous pace and requires efficient prediction of their function. In this case, an important factor in sequence annotation is the age of the gene, which is associated with the presence of its orthologs in a large number of taxa. For genes that have occurred recently during evolution (young or taxon-specific genes), orthologs are absent or poorly represented in other organisms. For them, function prediction is difficult and its accuracy is lower than for ancient genes, for which homologs are more widely represented

Aim: Development of a method for annotation of gene functions that would work with high accuracy regardless of the age of the genes
MATERIALS AND METHODS

1) Method *k*-nearest neighbors (KNN)

2) Method using orthogroups annotation (OG)

3) In addition, we used the combinations of the two previous methods (KNN+OG)

**Measures of algorithm evaluation:**

1. **Specificity** (*SP*): \[ SP = \frac{TP}{TP+FP} * 100 \]

2. **Sensitivity** (*SN*): \[ SN = \frac{TP}{TP+FN} * 100 \]

3. **Accuracy** (*AC*): \[ AC = \frac{SN+SP}{2} * 100 \]

4. **F1 Score** (*F1*): \[ F1 = 2 \frac{SP+SN}{SP+SN} * 100 \]
Combining the results of KNN and OG methods gives a fundamental improvement in the accuracy of function recognition for genes of all ages and at $k = 4$ allows to level out the effect of age of genes on the accuracy of their annotation regardless of age, the accuracy becomes approximately the same.
## COMPARISON WITH BLAST2GO

<table>
<thead>
<tr>
<th></th>
<th>KNN+OG</th>
<th>Blast2GO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SN</td>
<td>SP</td>
</tr>
<tr>
<td>Young genes</td>
<td>61.49</td>
<td>82.66</td>
</tr>
<tr>
<td>Middle genes</td>
<td>58.22</td>
<td>89.50</td>
</tr>
<tr>
<td>Old genes</td>
<td>58.94</td>
<td>92.75</td>
</tr>
</tbody>
</table>

On average, our new method is more accurate than Blast2GO by more than 14%
CONCLUSION

An OG method is proposed for predicting gene functions, taking into account information about the belonging of a gene to the OrthoDB orthological group, which, in combination with KNN results, gives a higher prediction accuracy and allows leveling the effect of gene age.

Comparison of the proposed method with the Blast2GO method showed that the accuracy of our approach exceeds that of Blast2GO by 18% or less.
ACKNOWLEDGEMENTS

Development of algorithms was funded by the Kurchatov Genome Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, no. 075-15-2019-1662

Thank you for your attention!