

В высших приматах присутствует отбор в пользу динуклеотидов CG**Научный руководитель – Алексеевский Андрей Владимирович****Белоусов Владислав Александрович***Студент (бакалавр)*

Национальный исследовательский университет «Высшая школа экономики», Факультет математики, Москва, Россия

E-mail: belv3395@gmail.com

В настоящей работе проведена оценка давления отбора на частоты динуклеотидов в геноме человека. Для этого на основании выравниваний геномов шимпанзе, карликового шимпанзе (бонобо) и человека были вычислены вероятности всех типов мутаций во всех трехбуквенных контекстах. Например, вероятность мутации из ACG в ATG за половину времени от общего предка равна 2.83×10^{-2} . Вычислялась эта вероятность следующим образом:

Была предложена математическая модель эволюционного процесса в условиях следующего упрощения: вероятность мутации одного нуклеотида в другой зависит лишь от трехбуквенного контекста и временного промежутка, за который эта мутация происходит.

Использовалось выравнивание участков геномов человека (сборка hg19), шимпанзе (сборка panTro4) и бонобо (сборка <http://www.ebi.ac.uk/ena/data/view/AJFE01000000>), полученное ранее в дипломной работе В. Голикова.

Для данного трехбуквенного контекста *тройкой* будем называть три нуклеотида, стоящие в средней колонке выравнивания. С помощью предложенной модели выводится функция зависимости количества каждой из троек в данном контексте от вероятностей мутаций. Методом наибольшего правдоподобия подбираются наиболее подходящие вероятности мутаций отдельно для каждого контекста. Произведена оценка изменения количества всех нуклеотидов и динуклеотидов на основании рассчитанных вероятностей.

Поскольку ближайший общий предок этих организмов жил около 6 млн лет тому назад, вычисленные вероятности отражают только мутации, прошедшие отбор. Полученные вероятности мы сравнили с частотами мутаций в контекстах, вычисленными по данным о мутациях *de novo* в NNN тройках папа, мама, ребенок (ссылка).

<i>Word</i>	<i>Overall</i>	<i>de novo</i>			Mutations over 3 million years		
		<i>Gain</i>	<i>Loss</i>	<i>Effect</i>	<i>Gain</i>	<i>Loss</i>	<i>Effect</i>
<i>A</i>	338334882	1394	993	$1.19E - 06$	864878	779269	$2.53E - 04$
<i>C</i>	225174414	1077	1427	$-1.55E - 06$	808894	895455	$-3.84E - 04$
<i>G</i>	225379336	1034	1466	$-1.92E - 06$	811038	897133	$-3.82E - 04$
<i>T</i>	338888878	1376	995	$1.12E - 06$	864073	777026	$2.57E - 04$
<i>AT</i>	89587945	821	813	$8.93E - 08$	556136	580471	$-2.72E - 04$
<i>AG</i>	79550996	555	525	$3.77E - 07$	422055	399250	$2.87E - 04$
<i>CG</i>	9269677	571	922	$-3.79E - 05$	432494	466742	$-3.69E - 03$

Пусть число нуклеотидов C и G изменяется в $(1+c)$ и $(1+g)$ раз соответственно. Тогда при условии отсутствия отбора на число динуклеотидов CG, мы ожидаем изменения их количества в $(1+c+g)$ раз в первом приближении. Заметим, что и в *de novo* мутациях и в мутациях, прошедших отбор, ожидаемый и наблюдаемый процент убыли динуклеотидов CG различаются в 10 и 5 раз соответственно.

Таким образом, мы наблюдаем отбор в пользу сохранения динуклеотида CG.

Источники и литература

- 1) Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. (2012) Rate of de novo mutations and the importance of father's age to disease risk. Nature 488: 471-475.