

Использование биоинформатических алгоритмов для анализа генетических данных массового параллельного секвенирования (NGS-технологии) для поиска полиморфных вариантов (SNP), имеющих практическое значение для оценки генетического разнообразия особей дикого кабана в Республике Беларусь

Кипень В.Н.¹, Цыбовский И.С.¹, Мельнов С.Б.²

¹ГУ «Научно-практический центр Государственного комитета судебных экспертиз Республики Беларусь», Минск, Республика Беларусь

²РУП «БелНИЦ «Экология», Минск, Республика Беларусь

Одной из основных причин современных изменений в генетической структуре кабана европейского (*Sus scrofa scrofa*) является антропогенное воздействие. Изменение генофонда происходит путем истребления особей в ходе охоты или гибридизации диких животных с домашними во время свободного выпаса скота. Гибридизация диких кабанов с домашней свиньей как фактор, который вносит существенные изменения в генетический состав популяции, носит спорный характер (Karlsson A., 2007; Lin Y., 2014).

Изучение генетического разнообразия вида *Sus scrofa scrofa* на территории Республики Беларусь, а также оценка уровня интрогрессии аллелей домашних животных *Sus scrofa domesticus* в геном дикого кабана может иметь существенное значение для решения природоохранных задач. Оценка генетического разнообразия возможна путем исследования полиморфизма мтДНК, STR или SNP-локусов. Исследователи все чаще обращаются к анализу SNP. Это связано, в первую очередь, с наличием коммерческих чипов средней и высокой плотности для исследования генетических изменений в контексте всего генома. Одним из предложений на рынке полногеномных исследований является чип PorcineSNP60 DNA Analysis Kit v2 компании Illumina[©] с количеством анализируемых SNP более 60 тысяч. Однако стоимость анализа для одного образца все еще остается весьма высокой. В этом разрезе актуальной становится задача по созданию целевых таргетных панелей с ограниченным количеством SNP со значением $MAF \geq 10-15\%$ (MAF – Minor allele frequency) – порядка 100-150. Как было первоначально показано в исследовании Ramos et al. (2011) и впоследствии подтверждено нами (Кипень В.Н., 2016) с помощью биоинформатического анализа данных полногеномных проектов по секвенированию образцов вида *Sus scrofa*, существует достаточно большое количество высокополиморфных SNP, пригодных для решения данной задачи.

Цель и задачи. С учетом ранее полученных результатов (Ramos et al., 2011; Кипень В.Н. и др., 2016) по анализу полиморфных вариантов, представленных в чипе Pig_Illu60K (Illumina), для полногеномных сиквесных проектов по анализу генома дикого кабана из Европы (24 шт.) и домашней свиньи (породы крупная белая, ландрас, дюрок, пьетрен, мейшань – 91 шт.), оценить дифференцирующий потенциал 193 SNP, описанных Ramos et al. (2011). Анализ был выполнен с помощью алгоритма SRA Nucleotide BLAST и программы BioEdit v.7.2.5. Количество включенных в анализ SNP – 193. Отдельно дополнительное внимание было уделено участку X-хромосомы, на котором нами ранее было выявлено два полиморфных локуса с предположительно высоким дифференцирующим потенциалом для различения особей дикого кабана от домашней свиньи – было исследовано генотипическое разнообразие для 25 SNP в диапазоне X:58911107-X:60656889 (rs80801780, rs80817425, rs81473227, rs80903635, rs80891217, rs321949235, rs80924261, rs80985865, rs80987186, rs81473228, rs330095107, rs80936726, rs81473257, rs80803330, rs80981028, rs80998473, rs80979282, rs321294367, rs80925206, rs339458440, rs331950236, rs80980235, rs80843461, rs80862185, rs81473281). Число полногеномных прочтений для дикого кабана – 24, для коммерческих пород домашней свиньи – 91 (крупная белая – 19, ландрас – 23, пьетрен – 6, дюрок – 28, мейшань – 15). Общее количество проанализированных сиквенса >16,0 млрд.

Из 193 проанализированных SNP только для 16 была показана строгая специфичность в отношении дикого кабана. Для всех коммерческих пород в рамках исследования 16 SNP

было показано наличие только одного генотипа. Для двух локусов, расположенных на X-хромосоме, частота специфического аллеля составила более 80%. Дополнительный анализ с использованием алгоритма MDR (Multifactor dimensionality reduction, (Moore J., 2006)), направленный на оценку возможного совместного наследования выявил наличие строгой сцепленности между локусами (вероятность рекомбинации локусов в мейозе составляет менее 2 %). Один из этих локусов – полиморфный вариант rs81473228 (H3GA0051811), – расположен в интронной области гена *HEPH* (Gene ID: 100512938). Функция ортолога *HEPH* у человека сходна с функцией церулоплазмينا, белок гефестин содержит гомологичный церулоплазмину активный центр и участки связывания меди. Гефестин участвует в метаболизме железа, активизируя окисление Fe^{2+} в Fe^{3+} , является ферментом ферроксидазой и содержит 6 ионов меди.

Используя современные алгоритмы анализа «сырых» данных (SRA) проектов полногеномного секвенирования, нам удалось выявить SNP, которые отличаются строгой специфичностью в отношении особей дикого кабана: общее количество SNP – 16, локализация – 4, 5, 6, 9, 10, 12, 13, 14, 15, X-хромосомы; частота специфического аллеля – 4,2-87,5%. Полученные результаты предстоит проверить на практике ввиду возможного наличия межпопуляционных различий по частоте распространенности того или иного SNP. На основании полученных данных представляется возможным формирование таргетной панели для оценки генетического разнообразия особей дикого кабана на территории Республики Беларусь на современном этапе его эволюционного развития.