

УДК 519.876.5

## ГЕНЕТИЧЕСКАЯ ИДЕНТИФИКАЦИЯ СОРТОВ СОИ С ИСПОЛЬЗОВАНИЕМ ОДНОНУКЛЕОТИДНЫХ ПОЛИМОРФИЗМОВ

© 2020 г. М.В. Грецова, М.Г. Самсонова

Санкт-Петербургский политехнический университет Петра Великого,  
195251, Санкт-Петербург, Политехническая ул., 29

E-mail: m.samsonova@spbstu.ru

Поступила в редакцию 18.11.2019 г.

После доработки 18.11.2019 г.

Принята к публикации 22.11.2019 г.

Разработан метод паспортизации сортов сои на основе однонуклеотидных полиморфизмов. Генетический алгоритм был использован для выбора оптимального набора, состоящего из 28 однонуклеотидных полиморфизмов, который затем был применен для паспортизации 243 сортов. Для всех сортов получены уникальные сигнатуры однонуклеотидных полиморфизмов, отличающиеся от сигнатур любого другого сорта хотя бы в трех позициях.

*Ключевые слова:* генетическая идентификация, однонуклеотидные полиморфизмы, соя.

DOI: 10.31857/S000630292001024X

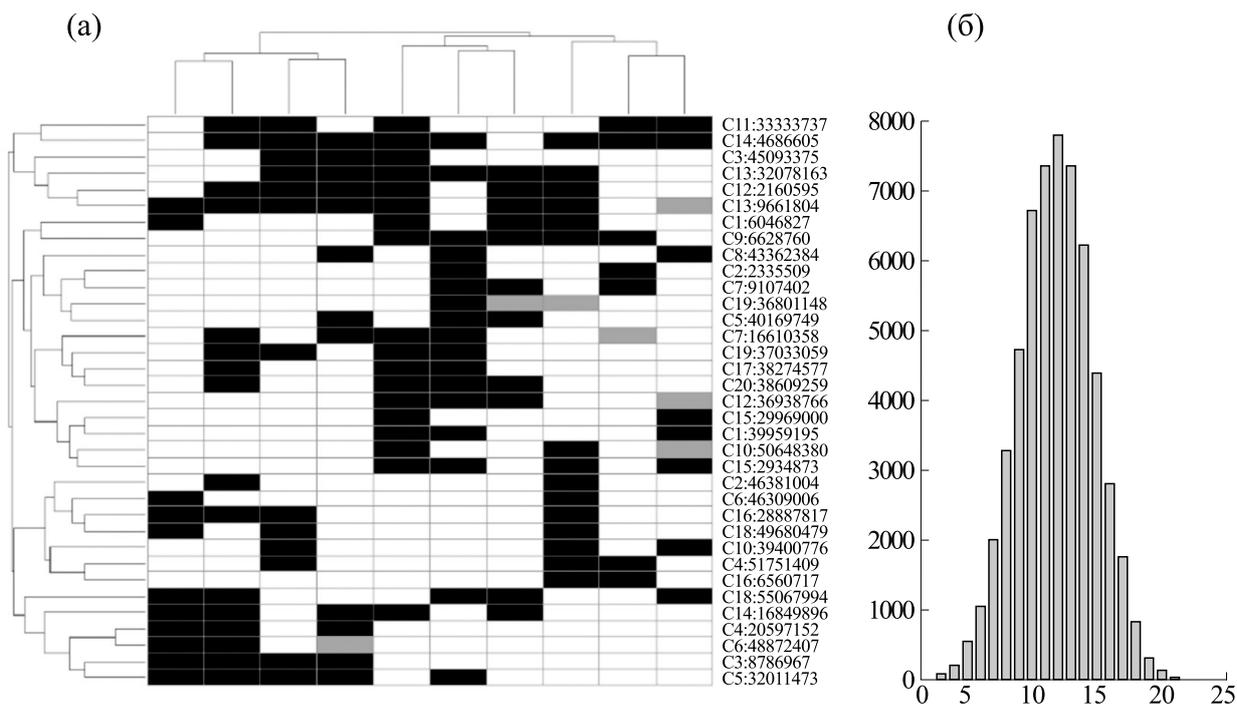
Генетическая паспортизация является востребованным инструментом для защиты прав селекционеров при регистрации, сертификации и коммерческом распространении сортов [1, 2]. На сегодняшний день наиболее перспективными представляются методы паспортизации, использующие ДНК-сигнатуры на основе молекулярных маркеров, таких как однонуклеотидные полиморфизмы (ОНП), микросателлитные последовательности и другие [3–5]. Преимуществами ОНП по сравнению с остальными типами ДНК-маркеров являются их высокая частота встречаемости в популяции, плотность и эволюционная стабильность, а также наличие широкого спектра подходов для их выявления [6]. Сочетание значений ОНП, характерное для некоторого сорта, называется ОНП-сигатурой этого сорта. Организмы разных сортов отличаются друг от друга в большом количестве позиций ДНК, поэтому использовать все найденные ОНП для идентификации сорта неудобно [6]. Так возникает задача поиска такого небольшого набора ОНП, с помощью которого можно идентифицировать сорта. Как правило, набор ОНП для построения сигнатуры находится в результате фильтрации исходного набора ОНП, в соответствии с некоторым набором критериев. При фильтрации необходимо учитывать степень гомозиготности ОНП. Кроме того, поскольку ОНП искомого набора не должны сильно коррелировать друг с другом, пред-

ставляется разумным для каждого гапблока рассматривать не более одного ОНП.

Поиск набора ОНП был реализован для выборки из 243 сортов сои, состоящей из 110 образцов коллекции Всероссийского института генетических ресурсов растений им. Н.И. Вавилова и 133 образцов коллекции компании «СоКо» (Краснодар). Ранее эти образцы были генотипированы по технологии GBS [7]. Были взяты следующие критерии фильтрации ОНП: 1) ОНП должны принадлежать гапблокам; 2) частота, с которой ОНП находится в каждом из гомозиготных состояний, должна быть не менее 0.2 – этот критерий задает степень полиморфизма позиции, в которой регистрируется ОНП; 3) результирующий набор не должен содержать двух ОНП, принадлежащих одному гапблоку. Для подбора оптимального набора ОНП использовался генетический алгоритм.

Среди всех найденных в результате генотипирования были выбраны 14745 ОНП, для которых частота встречаемости в выборке сортов составляла не менее 0.8. Поиск гапблоков для имеющегося набора ОНП был реализован отдельно для каждой хромосомы в программе Haploview [8]. Были найдены 12729 ОНП, принадлежащих 1947 гапблокам. После этого среди них были отобраны 5718 ОНП, удовлетворяющие второму критерию, – частота нахождения ОНП в каждом из гомозиготных состояний должна составлять не менее 0.2. Наконец, для каждого гапблока был выбран единственный ОНП, по которому разли-

*Сокращение:* ОНП – однонуклеотидные полиморфизмы.



Результат работы генетического алгоритма. (а) – Пример полученных ОНП-сигнатур для десяти сортов сои. Черным и белым цветом обозначены гомозиготные состояния локусов, серым цветом обозначены гетерозиготные состояния локусов. (б) – Распределение количества ОНП, по которым отличается пара сортов.

чается наибольшее количество сортов. Таким образом, в результате фильтрации был получен набор из 1332 ОНП.

Для поиска оптимального набора ОНП был запущен генетический алгоритм; в качестве индивидуумов в генетическом алгоритме были взяты наборы ОНП; каждый набор содержал от 20 до 40 ОНП так, что каждая хромосома сои была представлена в индивидууме одним или двумя ОНП. Начальная популяция, содержащая 1000 индивидуумов, была сгенерирована случайным образом. Отбор индивидуумов для формирования следующего поколения проводили в соответствии с функцией приспособленности индивидуума  $x$ :

$$f(x) = \min_{i,j \in S} \left( \frac{d_{ij}}{N} \right),$$

где  $d_{ij}$  – расстояние между сортами  $i$  и  $j$ , которое измеряется как количество ОНП в наборе  $x$ , по которым эти сорта различаются, т.е. находятся в противоположных гомозиготных состояниях,  $N$  – количество ОНП в наборе,  $S$  – множество сортов. Каждое следующее поколение популяции формировали из 500 наиболее приспособленных индивидуумов и их 500 потомков, полученных в результате скрещиваний так, чтобы в каждом потомке каждая хромосома была представлена одним или двумя ОНП.

Генетический алгоритм на 2000 поколений был запущен четыре раза. Для каждого сорта была найдена сигнатура, состоящая из 28 ОНП и отличающаяся от сигнатуры любого другого сорта хотя бы в трех позициях (см. рисунок).

### ФИНАНСИРОВАНИЕ РАБОТЫ

Работа выполнена в рамках Федеральной целевой программы (проект №14.575.21.0136 от 26.09.2017, уникальный идентификатор проекта RFMEFI57517X0136).

### БЛАГОДАРНОСТИ

Вычисления были проведены в суперкомпьютерном центре «Политехнический» СПбПУ и кластере Университета Южной Калифорнии.

Авторы выражают благодарность А.А. Иголкиной за ценные советы, рекомендации и помощь при выполнении исследования.

### КОНФЛИКТ ИНТЕРЕСОВ

Авторы заявляют об отсутствии конфликта интересов.

## СОБЛЮДЕНИЕ ЭТИЧЕСКИХ СТАНДАРТОВ

Настоящая работа не содержит описания каких-либо исследований с использованием людей и животных в качестве объектов.

## СПИСОК ЛИТЕРАТУРЫ

1. M. Agarwal, N. Shrivastava, and H. Padh., *Plant Cell Rep.* **27**, 617 (2008).
2. N. K. Korir, J. Han, L. Shangguan, et al., *Biotech.* **33**, 2 (2013).
3. M. W. Ganal, T. Altmann, and M. S. Röder, *Curr. Opin. Plant Biol.* **12**, 2 (2009).
4. D. L. Hyten, Y. Zhu, P. B. Cregan, et al., *Crop Sci.* **50**, 5 (2010).
5. D. I. Pacurar, M. L. Pacurar, N. Street, et al., *J. Exp. Bot.* **63**, 7 (2012).
6. Y. Lee, N. Jeong, J. H. Kim, et al., *Plant J.* **81**, 4 (2015).
7. А. Канапин, А. Самсонова, А. Соколкова и др., *Биофизика* (2020) (в печати).
8. J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, *Bioinformatics* **21**, 2 (2005).

## Genetic Identification of Soybean Varieties Using Single Nucleotide Polymorphism Markers

M.V. Gretsova and M.G. Samsonova

*Peter the Great St. Petersburg Polytechnic University, ul. Polytekhnicheskaya 29, St. Petersburg, 195251 Russia*

The seed certification method for soybean varieties has been developed based on single nucleotide polymorphisms. The genetic algorithm was employed to find optimal set consisting of 28 single nucleotide polymorphisms, which was then used to certify 243 varieties. For all varieties, unique signatures of single nucleotide polymorphisms were determined. The signatures which were different from those of any other varieties were considered to be unique if they differed at least in three positions.

*Keywords: genetic identification, single nucleotide polymorphisms, soybean*