СЕЛЬСКОХОЗЯЙСТВЕННАЯ БИОЛОГИЯ, 2015, том 50, № 5, с. 571-578

Биоинформатика и математическая статистика

УДК 575.17:575.118.5:575.162:57.087.1 doi: 10.15389/agrobiology.2015.5.571rus

ОЦЕНКА МЕРЫ ИНФОРМАЦИОННОГО ПОЛИМОРФИЗМА ГЕНЕТИЧЕСКОГО РАЗНООБРАЗИЯ*

Ю.В. ЧЕСНОКОВ, А.М. АРТЕМЬЕВА

Одна из основных целей генетики растений и животных — это идентификация и картирование генов. При определении генетического сцепления обычно стараются установить, какие маркерные локусы (маркеры) имеют аллели, косегрегирующие с аллелями желаемого локуса. Пригодность маркера для указанных целей зависит от числа аллелей, которые имеет этот маркер, и их соответствующих относительных частот. Количественно степень полиморфизма обычно измеряется двумя различными величинами, или показателями (мерами), — гетерозиготностью (heterozygosity, H), для которой объективный алгоритм оценки и формула изменчивости хорошо известны (M. Nei с соавт., 1974; M. Nei с соавт., 1979), и величиной информационного полиморфизма (polymorphism information content, PIC) (D. Botstein с соавт., 1980). Исходя из этого в работе на основе данных литературы описаны статистические подходы, применяемые для анализа информационного полиморфизма. Рассмотрены меры информационного полиморфизма, гетерозиготности и некоторых сопутствующих величин, определяемых при оценке генетического разнообразия как на межвидовом, так и на внутривидовом популяционном уровне. Мера, или величина, информационного полиморфизма (PIC) определяется способностью маркера устанавливать полиморфизм популяции в зависимости от числа обнаруживаемых аллелей и распределения их частот (D. Botstein с соавт., 1980). Таким образом, PIC выявляет дискриминационную способность маркера, фактически зависит от числа известных (устанавливаемых) аллелей и распределения их частот и тем самым эквивалентна генному разнообразию. Для доминантных маркеров максимальное значение PIC составляет 0,5. Следует отметить, что в случае маркеров с равным распределением частот внутри популяции величина PIC выше. Маркеры с множественными аллелями имеют еще большие значения этого показателя, однако при этом величина PIC также зависит от распределения частот аллелей. С помощью 21 пары SSR (simple sequence repeats) и 12 пар S-SAP (sequence specific amplified polymorphism) праймеров у 96 образцов *Brassica rapa* L. из стержневой коллекции ВИР мы обнаружили 135 SSR и 123 S-SAP полиморфных маркера. Среднее значение PIC для обоих типов маркеров — 0,316, тогда как для микросателлитных маркеров — 0,257, для S-SAP маркеров — 0,379, то есть в среднем на 50 % выше. Ожидаемую (H_E) гетерозиготность обычно определяют, когда описывают генетическое разнообразие, поскольку она менее чувствительна к размеру выборки, чем наблюдаемая гетерозиготность (H_O). Если H_O и H_E схожи (достоверно не различаются), скрещивание в популяции происходит практически случайно. При $H_O < H_E$, популяция инбредная. Если $H_O > H_E$, то в популяции система случайного скрещивания преобладает над инбридингом. Эффективное мультиплексное отношение (effective multiplex ratio, EMR) определяют как произведение общего числа полиморфных локусов (на праймер) и доли полиморфных локусов от их общего числа (W. Powell с соавт., 1996; J. Nagaraju с соавт., 2001). Маркерный индекс (marker index, MI) — статистическая величина, используемая для оценки суммарной пригодности маркерной системы (чем выше значение MI для методики, тем она лучше) (W. Powell с соавт., 1996; J. Nagaraju с соавт., 2001). Чтобы отразить способность сочетания «праймер—применяемая методика» устанавливать различия между большим числом генотипов, используют показатель разрешающей способности (resolving power, Rp) (J.E. Gilbert с соавт., 1999; A. Prevost с соавт., 1999). Представлена информация о некоторых продуктах программного обеспечения, которое может быть использовано для расчета величины информационного полиморфизма и гетерозиготности. Приведены формулы для установления эффективного мультиплексного отношения, маркерного индекса и показателя разрешающей способности комбинации «праймер—применяемая методика».

Ключевые слова: гетерозиготность, величина информационного полиморфизма, эффективное мультиплексное отношение, маркерный индекс, показатель разрешающей способности, программное обеспечение.

Одна из основных целей генетики растений и животных — идентификация и картирование генов, определяющих проявление интересующих исследователя признаков. Существует большое число маркерных локусов, визуализируемых с помощью различных маркерных систем (кратко

* Работа выполнена при частичной поддержке РФФИ (грант № 13-04-00128-а).

их называют маркерами), чьи позиции и порядок, в котором они располагаются на хромосоме, хорошо известны. При определении генетического сцепления обычно стараются установить, какие маркерные локусы (маркеры) имеют аллели, косегрегирующие с аллелями желаемого локуса. Пригодность маркера для указанных целей зависит от числа аллелей, которые имеет этот маркер, и их соответствующих относительных частот. Качественно маркер характеризуется как полиморфный, если он представлен, как минимум, двумя аллелями и его наиболее часто встречающийся аллель в популяции имеет частоту не менее 99 %. Количественно степень полиморфизма обычно измеряется двумя различными величинами, или показателями (мерами). Одна известна как гетерозиготность (heterozygosity, H) и ее объективный алгоритм оценивания и формула изменчивости хорошо известны (1, 2). Другой единицей измерения служит величина информационного полиморфизма (polymorphism information content, PIC) (3).

Молекулярные маркеры стали эффективным инструментом и средством, с помощью которого оценивают и характеризуют как внутри-, так и межвидовое генетическое разнообразие. Маркерные системы различают по мере (то есть величине) их информативности, что, в свою очередь, зависит от степени их полиморфизма. Концепцию полиморфизма используют для определения генетической изменчивости в популяции, что в последние десятилетия стало предметом интенсивного изучения в различных научных дисциплинах — генетике, экологии, ботанике, зоологии и некоторых других. Примеры этому многочисленны и очевидны (4-10). Однако при планировании применения молекулярных маркеров для какого-либо исследования или для практического использования в селекционных программах неизбежно возникают вопросы, на которые исследователям зачастую приходится искать ответ. Насколько трудно будет найти пригодные для планируемой работы полиморфные локусы? Как много маркеров необходимо будет задействовать? Насколько полиморфным должен быть каждый подобранный маркер? На все эти вопросы можно найти ответ, оценив меру информативности маркеров. Двумя основными параметрами, определяемыми для этого, являются гетерозиготность (H) и величина информационного полиморфизма (PIC). В дополнение к ним существуют некоторые сопутствующие показатели, с помощью которых можно также установить эффективность выбранной системы «праймер—маркер» и(или) избранного методического подхода.

В настоящей работе кратко обобщены описанные в научной литературе статистические подходы, используемые для установления меры информационного полиморфизма, гетерозиготности и некоторых сопутствующих величин, определяемых при оценке генетического разнообразия как на межвидовом, так и на внутривидовом популяционном уровнях.

Гетерозиготность (H). Гетерозиготность локуса, которая определяется как вероятность того, что в популяции особь гетерозиготна по этому локусу (11), может быть рассчитана по формуле:

$$H = 1 - \sum_{i=1}^{l} P_i^2, \qquad [1]$$

где P_i — частота i-го аллеля среди общего числа l аллелей. Иными словами, геретозиготность может быть рассмотрена как средняя порция локусов с двумя различными аллелями в одном локусе у одной особи. Обычно это распространяется на всю популяцию или какую-то ее часть и подразделяется на наблюдаемую и ожидаемую гетерозиготность. Ожидаемая (expected) гетерозиготность (H_E), или генное разнообразие по M. Nei (1), — это ожидаемая вероятность того, что особь будет гетерозиготна по соответст-

вующему локусу в мультилокусных системах (для всех анализируемых локусов). Другими словами, это установленная фракция всех особей, которые были бы гетерозиготны по любому случайно выбранному локусу. Ее часто высчитывают, основываясь на установлении квадратного корня из частоты нуль-аллеля (рецессивного), следующим образом (подобно уравнению [1]): $H_E = 1 - \sum_i^n p_i^2$, где p_i — частота i-го аллеля, n_i — общее число аллелей во всех локусах. Наблюдаемая (observed) гетерозиготность (H_O) — это часть генов, которые в популяции гетерозиготны. Она рассчитывается для каждого локуса как общее число гетерозигот, поделенное на размер выборки. Значения для H_E и H_O варьируют от 0 (нет гетерозиготности) до практически 1 (большое число аллелей с равной частотой встречаемости). Ожидаемую гетерозиготность обычно определяют, когда описывают генетическое разнообразие, поскольку она менее чувствительна к размеру выборки, чем наблюдаемая гетерозиготность. Если H_O и H_E схожи (достоверно не различаются), то скрещивание в популяции происходит практически случайно. При $H_O < H_E$, популяция инбредная. Если $H_O > H_E$, то в популяции система случайного скрещивания преобладает над инбридингом.

Мера информационного полиморфизма (PIC). Мера, или величина, информационного полиморфизма (polymorphism information content — PIC) определяется способностью маркера устанавливать полиморфизм в популяции в зависимости от числа обнаруживаемых аллелей и распределения их частот (3). Таким образом, PIC выявляет дискриминационную способность маркера, фактически зависит от числа известных (устанавливаемых) аллелей и распределения их частот и тем самым эквивалентна генному разнообразию. В самой простой форме величина PIC может быть рассчитана подобно гетерозиготности (см. уравнение [1]):

$$\mathrm{PIC}_j = 1 - \sum_{i=1}^{n} P_i^2, \qquad [2]$$

где i — i-й аллель j-го маркера, n — число аллелей j-го маркера, P — частота аллелей. Примеры расчетов значения PIC для биаллельного и мультиаллельного маркеров представлены в таблице 1. В то же время для кодоминантных маркеров уравнение [2] может быть представлено следующим образом (12):

$$\mathrm{PIC} = 1 - \left(\sum^{k}{}_{i=1} P_i^2\right) - \sum^{k-1}{}_{i=1} \sum^{k}{}_{j=1} 2\, P_i^2\, P_j^2, \qquad [3]$$

где k — число аллелей, P_i и P_j — частота соответственно i-го и j-го аллеля в популяции. Для доминантных маркеров величину PIC рассчитывают согласно описанию (13):

$$\mathrm{PIC} = 1 - [f2 + (1 - f)2], \qquad [4]$$

где f — частота маркера в наборе данных. Для доминантных маркеров максимальное значение PIC составляет 0,5. Следует отметить, что в случае маркеров с равным распределением частот внутри популяции величина PIC выше. Маркеры же с множественными аллелями имеют еще большие значения этого показателя, однако при этом величина значения PIC также зависит от распределения частот аллелей (табл. 1).

1. Примеры расчета PIC для биаллельного и мультиаллельного маркеров

Частота аллелей	Формула расчета по уравнению [2]	Значение PIC
Биаллельный маркер		
$P_1 = 0{,}5$; $P_2 = 0{,}5$	$1 - (0{,}5^2 + 0{,}5^2)$	0,50
$P_1 = 0{,}4$; $P_2 = 0{,}6$	$1 - (0{,}4^2 + 0{,}6^2)$	0,48
$P_1 = 0{,}3$; $P_2 = 0{,}7$	$1 - (0{,}3^2 + 0{,}7^2)$	0,42
$P_1 = 0{,}2$; $P_2 = 0{,}8$	$1 - (0{,}2^2 + 0{,}8^2)$	0,32
$P_1 = 0{,}1$; $P_2 = 0{,}9$	$1 - (0{,}1^2 + 0{,}9^2)$	0,18

Продолжение таблицы 1

Мультиаллельный маркер		
$P_1 = 0,33$; $P_2 = 0,33$; $P_3 = 0,33$	1 – (0,332 + 0,332 + 0,332)	0,67
$P_1 = 0,4$; $P_2 = 0,3$; $P_3 = 0,3$	1 – (0,42 + 0,32 + 0,32)	0,66
$P_1 = 0,4$; $P_2 = 0,4$; $P_3 = 0,2$	1 – (0,42 + 0,42 + 0,22)	0,64
$P_1 = 0,5$; $P_2 = 0,3$; $P_3 = 0,2$	1 – (0,52 + 0,32 + 0,22)	0,62
$P_1 = 0,5$; $P_2 = 0,4$; $P_3 = 0,1$	1 – (0,52 + 0,42 + 0,12)	0,58
$P_1 = 0,6$; $P_2 = 0,2$; $P_3 = 0,2$	1 – (0,62 + 0,22 + 0,22)	0,56
$P_1 = 0,6$; $P_2 = 0,3$; $P_3 = 0,1$	1 – (0,62 + 0,32 + 0,12)	0,54
$P_1 = 0,7$; $P_2 = 0,2$; $P_3 = 0,1$	1 – (0,72 + 0,22 + 0,12)	0,46
$P_1 = 0,8$; $P_2 = 0,1$; $P_3 = 0,1$	1 – (0,82 + 0,12 + 0,12)	0,35

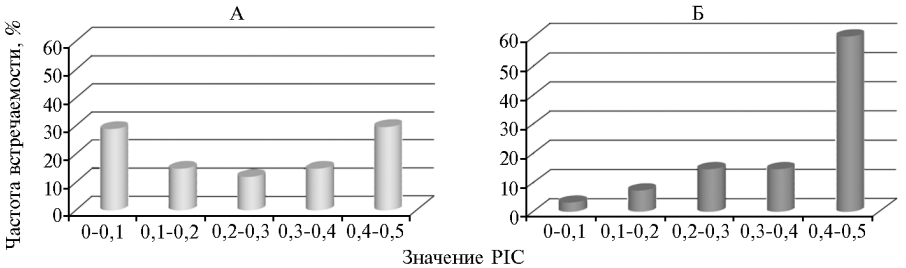
Примечание. PIC — polymorphism information content (информационный полиморфизм).

2. Выявленные аллели молекулярных маркеров SSR и S-SAP и их PIC при оценке стержневой коллекции *Brassica rapa* L., сохраняемой в ВИР (Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова)

Число аллелей	Доля аллелей, %	Значение PIC
	SSR маркеры	
39	28,9	0-0,1
20	14,8	0,1-0,2
16	11,9	0,2-0,3
20	14,8	0,3-0,4
40	29,6	0,4-0,5
Всего		
135	100	
	S-SAP маркеры	
4	3,2	0-0,1
9	7,3	0,1-0,2
18	14,6	0,2-0,3
18	14,6	0,3-0,4
74	60,0	0,4-0,5
Всего		
123	100	

Примечание. SSR — simple sequence repeats, S-SAP — sequence specific amplified polymorphism. PIC — polymorphism information content (информационный полиморфизм).

В нашей работе при оценке стержневой коллекции *Brassica rapa* L. (96 образцов), хранящейся в ВИР, с помощью 21 пары SSR (simple sequence repeats) и 12 пар S-SAP (sequence specific amplified polymorphism) праймеров найдено 135 SSR и 123 S-SAP полиморфных маркера, которые использовались для построения филогенетического древа (14). Для каждого маркера нами были подсчитаны значения PIC (табл. 2). Их распределение по частоте встречаемости для каждого типа маркеров представлены на диаграммах (рис.). Среднее значение PIC для обоих типов маркеров 0,316, тогда как для микросателлитных маркеров — 0,257, для S-SAP маркеров — 0,379, то есть в среднем на 50 % выше. Таким образом, в наших исследованиях по оценке генетического полиморфизма в коллекции *B. rapa* с использованием двух типов маркеров наибольшую информативность отмечали у S-SAP маркеров.



Распределение значений PIC (polymorphism information content) **для молекулярных маркеров SSR** (simple sequence repeats, А) **и S-SAP** (sequence specific amplified polymorphism, Б) **при оценке стержневой коллекции *Brassica rapa* L., сохраняемой в ВИР** (Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова).

В то же время при определении генетического разнообразия у риса (*Oryza sativa* L.) среднее значение PIC было в 2 раза выше в случае использования SSR маркеров (0,66), чем в варианте, когда применяли RFLP (restriction fragment length polymorphism) маркеры (0,36) (15). При исследовании молекулярного генетического разнообразия у сладкой кукурузы (*Zea*

mays L.) с помощью RAPD (random amplified polymorphic DNA) и SSR маркеров получены результаты, указывающие на более высокую симилярность между изученными популяциями, чем внутри них (16). При этом авторы отмечают, что RAPD маркеры имели более низкие средние значения PIC (0,17), чем SSR маркеры (0,57). Абсолютные значения PIC для RAPD и SSR маркеров в описанном случае не подлежали сравнению из-за максимальной величины PIC 0,5 и 1,0 соответственно для RAPD и SSR локусов. RAPD и SSR маркеры также использовались для оценки генетической изменчивости и взаимоотношений у тунисских местных сортов ячменя (*Hordeum vulgare* L.) (17). Несмотря на то, что был выявлен высокий уровень полиморфизма как для RAPD, так и для SSR маркеров, а средние значения PIC составили 0,477 и 0,533 соответственно для RAPD и SSR маркеров, авторы заключают, что SSR маркеры все же лучше подходят для оценки генетического разнообразия ячменя, чем RAPD маркеры, поскольку SSR маркеры обладали более высоким полиморфизмом (90,7 %) по сравнению с RAPD маркерами (74,0 %).

И, наконец, еще один пример установления значения PIC, на который хотелось бы обратить внимание. Исследователи из Аргентины и США провели анализ генетического разнообразия аргентинских сортов мягкой пшеницы (*Triticum aestivum* L.), созданных в период с 1932 по 1995 год (18). Используя SSR и AFLP (amplified fragment length polymorphism) маркеры, они установили, что нет существенных различий по генетическому разнообразию между группой сортов, полученных до 1960 года, и группами, выпущенными в каждое из трех последующих десятилетий. Среднее значение разнообразия, установленное с помощью SSR маркеров, было практически идентичным для всех четырех временны́х периодов. Генетическое разнообразие, выявленное посредством AFLP маркеров, подтвердило отсутствие уменьшения генетического разнообразия во времени. Однако между сортами мягкой пшеницы, созданными в 1970-х (PIC = 0,28) и 1980-х (PIC = 0,34) годах были найдены достоверные различия (P = 0,01). В целом результаты, полученные по PIC, указывают на то, что аргентинские сорта мягкой пшеницы поддерживались практически на одном и том же уровне генетического разнообразия на протяжении более 60 лет, а их различия обусловлены преимущественно реализуемыми селекционными программами, но никак не степенью генетического разнообразия полученных сортов. Таким образом, мера информационного полиморфизма служит важным компонентом при составлении планов селекционных программ и одним из ключевых информационно-статистических показателей при их выполнении.

Программное обеспечение для расчета H и PIC. Для правильного составления плана генетических исследований и оценки полученных результатов зачастую приходится проводить расчеты величин гетерозиготности (H) и информационного полиморфизма (PIC) для описания информативности маркеров, но до последнего времени не было простых и общедоступных калькуляторов для таких расчетов. Для упрощения работ по маркерным исследованиям группа венгерских ученых в 2012 году предложила разработанную ими интерактивную Интернет-программу PICcalc (http://w3.georgikon.hu/pic/english/default.aspx) (19). Программа позволяет вычислять значения H и PIC по аллельным частотам при введении показателей в ручном режиме или с использованием специального файла, содержащего бинарную матрицу данных. Дополнительные опции дают возможность рассчитывать значения для определенного числа локусов, используя для этого простой текстовый файл, что гарантирует установление H и PIC для праймера или наборов праймеров, использованных для анализа различных генетических маркерных систем, имеющих дело с бинар-

ными данными. Для мультилокусных маркеров, например AFLP, ISSR (inter simple sequence repeats) или RAPD, теоретически предполагается, что фрагменты равной длины амплифицируются на соответствующих локусах хромосом и что они представляют одиночный доминантный локус с двумя возможными аллелями (наличие/отсутствие ампликона). Максимальное значение H и PIC для доминантных маркеров в этом случае будет равно 0,5, поскольку для такого типа маркеров допускается только два аллеля на локус и обе величины подвержены влиянию числа и частоты аллелей (13, 20, 21). С учетом этой особенности доминантных маркеров в программе специально предусмотрена возможность расчета H и PIC для них (19).

Ранее группа американских ученых также предложила компьютерную программу, которая позволяла проводить расчет показателей H и PIC (http://darwin.cwru.edu/pic) (22). Основное отличие от упомянутого выше Интернет-ресурса заключается в том, что авторы вывели единообразно распределяемое минимальное отклонение несмещенной оценки PIC в соответствии с его точным значением дисперсии. Для того чтобы установить это, они получили формулу для расчета любого многочлена в наборе переменных, распределенных мультиномиально.

Сопутствующие величины. Эффективное мультиплексное отношение (effective multiplex ratio, EMR) определяют как произведение общего числа полиморфных локусов (на праймер) и доли полиморфных локусов от их общего числа (23, 24):

$$EMR = n_p(n_p/n), \qquad [5]$$

где n_p — число полиморфных локусов, n — общее число локусов. Чем выше значения EMR, тем эффективней система «праймер—маркер».

Маркерный индекс (marker index, MI) — статистическая величина, используемая для оценки суммарной пригодности маркерной системы. Маркерный индекс есть произведение величины информационного полиморфизма (или ожидаемой гетерозиготности, H_E) и эффективного мультиплексного отношения (23, 24):

$$MI = PIC \times EMR, \qquad [6]$$

Чем выше значение MI для методики, тем она лучше. В то же время чтобы отразить способность сочетания «праймер—применяемая методика» устанавливать различия между большим числом генотипов, используют показатель разрешающей способности (resolving power, Rp) (25, 26):

$$Rp = \sum I_b, \qquad [7]$$

где $I_b = 1 - (2 \times 0{,}5 - p)$ — информативность ампликона, p — доля особей, у которых выявлен ампликон I.

Таким образом, существует несколько подходов, позволяющих оценивать меру информационного полиморфизма и сопутствующие ей величины. ДНК-маркеры в настоящее время признаны довольно удобным и качественным инструментом оценки генетического разнообразия на молекулярном уровне. Однако перед использованием той или иной маркерной системы необходимо оценить техническую оснащенность лаборатории, потребность в применении выбранной маркерной системы и ее соответствие решаемым задачам, профессиональную подготовку персонала, а также предстоящие эксплуатационные расходы и доступные средства вспомогательного обслуживания. Требуемое программное обеспечение должно быть выбрано на основании расчета его пригодности для решения стоящих перед исследователем задач, в том числе задач по популяционной генетике, если речь идет об оценке информационного полиморфизма. Морфологические параметры весьма важны для интерпретации полученных результа-

тов. Установление статистически достоверных ассоциативных и корреляционных связей между морфологическими и молекулярно-генетическими показателями служит ключевым обстоятельством при принятии окончательных решений. И, конечно же, нельзя не учитывать биологические особенности изучаемых видов при оценке исследуемых генетических параметров, поскольку один и тот же параметр может формироваться у разных видов неодинаково не только в фило-, но и в онтогенезе. Последнее особенно важно для эффекта взаимодействия «генотип—среда».

ЛИТЕРАТУРА

1. Nei M., Roychoudhury A.K. Sampling variances of heterozygosity and genetic distance. Genetics, 1974, 76: 379-390.
2. Nei M., Li W.H. Mathematical model for studying genetic variation in terms of restriction endonucleases. PNAS USA, 1979, 76: 5269-5273 (doi: 10.1073/pnas.76.10.5269).
3. Botstein D., White R.L., Skalnick M.H., Davies R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet., 1980, 32: 314-331.
4. Артемьева А.М., Чесноков Ю.В., Клоке Э. Морфолого-географический и молекулярно-генетический анализ коллекции белокочанной капусты ВИР. Доклады РАСХН, 2008, 5: 14-18.
5. Артемьева А.М., Чесноков Ю.В., Клоке Э. Генетическое разнообразие и внутривидовые филогенетические взаимоотношения культур вида *Brassica rapa* L. по результатам анализа микросателлитов. Информационный Вестник ВОГиС, 2008, 12(4): 608-619.
6. Артемьева А.М., Клоке Э., Чесноков Ю.В. Анализ филогенетических связей вида *Brassica oleracea* L. (капуста огородная). Информационный вестник ВОГиС, 2009, 13(4): 759-771.
7. Вишнякова М.А., Бурляева М.О., Алпатьева Н.В., Чесноков Ю.В. RAPD-анализ видового полиморфизма рода Чина (*Lathyrus* L.) сем. *Fabaceae* Lindl. Информационный вестник ВОГиС, 2008, 12(4): 595-607.
8. Артемьева А.М., Чесноков Ю.В. Коллекция капусты ВИР: этапы формирования и изучения. Вавиловский журнал генетики и селекции, 2012, 16(4/2): 1047-1060.
9. Иванов А.А., Буренин В.И., Чесноков Ю.В. Оценка филогенетических отношений видов рода *Beta* L. с помощью RAPD-маркеров. Доклады Россельхозакадемии, 2012, 3: 20-22.
10. Чесноков Ю.В., Буренин В.И., Иванов А.А. RAPD-анализ коллекционных образцов дикой и культурной свеклы (*Beta* L.). Сельскохозяйственная биология, 2013, 3: 28-36 (doi: 10.15389/agrobiology.2013.3.28rus, doi: 10.15389/agrobiology.2013.3.28eng).
11. Liu B.H. Statistical genomics: linkage, mapping and QTL analysis. CRC Press, Boca Raton, 1998.
12. Anderson J.A., Churchill G.A., Autrique J.E., Tanksley S.D., Sorrells M.E. Optimizing parental selection for genetic linkage maps. Genome, 1993, 36(1): 181-186 (doi: 10.1139/g93-024).
13. De Riek J., Calsyn E., Everaert I., Van Bockstaele E., De Loose M. AFLP-based alternatives for the assessment of distinctness, uniformity and stability of sugar beet varieties. Theor. Appl. Genet., 2001, 103: 1254-1265 (doi: 10.1007/s001220100710).
14. Артемьева А.М., Будан Х., Клоке Э., Чесноков Ю.В. Использование мобильных генетических элементов CACTA для уточнения филогенетических взаимоотношений внутри вида *Brassica rapa* L. Вавиловский журнал генетики и селекции, 2011, 15(2): 398-411.
15. Xu Y., Beachell H., McCouch S.R. A marker-based approach to broadening the genetic base of rice (*Oryza sativa* L.) in the US. Crop Sci., 2004, 44: 1947-1959 (doi: 10.2135/cropsci2004.1947).
16. Bered F., Freitas Terra T., Spellmeier M., Neto J.F.B Genetic variation among and within sweet corn populations detected by RAPD and SSR markers. Crop Breed. Appl. Biotechnol., 2005, 5: 418-425.
17. Karim K., Rawda A., Hatem C.-M. Genetic diversity in barley genetic diversity in local Tunisian barley based on RAPD and SSR analysis. Biological Diversity and Conservation, 2009, 2/1: 27-35 (doi: 10.1590/S0100-204X2015000200008).
18. Manifesto M.M., Schlatter A.R., Hopp H.E., Suarez E.Y., Dubcovsky J. Quantitative evaluation of genetic diversity in wheat germplasm using molecular markers. Crop Sci., 2001, 41: 682-690 (doi: 10.2135/cropsci2001.413682x).
19. Nagy S., Poczai P., Cernak I., Gorji A.M., Hegedus G, Taller J. PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. Biochem. Genet., 2012, 50: 670-672 (doi: 10.1007/s10528-012-9509-1).
20. Bolaric S., Barth S., Melchinger A.E., Posselt U.K. Genetic diversity in European perennial ryegrass cultivars investigated with RAPD markers. Plant Breed., 2005, 124: 161-166 (doi: 10.1111/j.1439-0523.2004.01032.x).
21. Henry R.J. Practical applications of plant molecular biology. Chapman and Hall, London, 1997.
22. Shete S., Tiwari H., Elston R.C. On estimating the heterozygosity and polymorphism information content value. Theor. Popul. Biol., 2000, 57: 265-271 (doi: 10.1006/tpbi.2000.1452).
23. Powell W., Morgante M., Andre C., Hanafey M., Vogel J., Tingey S., Rafal-

ski A. The utility of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol. Breed., 1996, 2: 225-238 (doi: 10.1007/BF00564200).
24. Nagaraju J., Damodar Reddy K., Nagaraja G.M., Sethuraman B.N. Comparison of multilocus RFLPs and PCR-based marker systems for genetic analysis of the silkworm, *Bombyx mori*. Heredity, 2001, 86: 588-597 (doi: 10.1046/j.1365-2540.2001.00861.x).
25. Gilbert J.E., Lewis R.V., Wilkinson M.J., Caligari P.D.S. Developing an appropriate strategy to assess genetic variability in plant germplasm collections. Theor. Appl. Genet., 1999, 98: 1125-1131 (doi: 10.1007/s001220051176).
26. Prevost A., Wilkinson M.J. A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor. Appl. Genet., 1999, 98: 107-112 (doi: 10.1007/s001220051046).

ФГБНУ ФИЦ Всероссийский институт генетических ресурсов растений им. Н.И. Вавилова (ВИР),
190000 Россия, г. Санкт-Петербург, ул. Большая Морская, 42-44,
e-mail: yu.chesnokov@vir.nw.ru

Поступила в редакцию 26 февраля 2015 года

Sel'skokhozyaistvennaya biologiya [*Agricultural Biology*], 2015, V. 50, № 5, pp. 571-578

EVALUATION OF THE MEASURE OF POLYMORPHISM INFORMATION OF GENETIC DIVERSITY

Yu.V. Chesnokov, A.M. Artemyeva

Federal Research Center the N.I. Vavilov All-Russian Institute of Plant Genetic Resources, Federal Agency of Scientific Organizations, 42-44, ul. Bol'shaya Morskaya, St. Petersburg, 190000 Russia, e-mail yu.chesnokov@vir.nw.ru
Acknowledgements:
Supported in part by Russian Foundation for Basic Research (grant № 13-04-00128-a)
Received February 26, 2015 doi: 10.15389/agrobiology.2015.5.571eng

Abstract

Gene identification and mapping are one of the main goals of plant and animal genetics. Upon verifying genetic linkage it is usually found which marker loci (markers) possess alleles co-segregated with the alleles of the desired locus. Marker utility for these purposes depends on the number of alleles, which the marker possesses, and their relative rates. There are two indexes, or measures, usually used for the polymorphism degree evaluation. They are the heterozygosity (H) for which the evaluation method and variability formula are well known (M. Nei et al., 1974, 1979), and polymorphism information content (PIC) (D. Botstein et al., 1980). Based on published data, we described the statistical approaches which are used for analysis of polymorphism information. Herein, the value of polymorphism information content, heterozygosity and some associated values detected upon evaluation of genetic diversity on interspecific and intraspecific population levels are considered. PIC shows haw the marker can indicate the population polymorphism depending on the number and frequency of the alleles (D. Botstein et al., 1980). So the PIC reflects a discriminating ability of the marker and, in fact, depends on the number of known alleles and their frequency distribution, thus being equal to genetic diversity. PIC maximal value for dominant markers is 0.5. Note, that for the markers with equal distribution in the population the PIC values are higher. They are much higher for markers with multiple alleles, and, however, also depend on the frequency distribution of the alleles. Using 135 SSR (simple sequence repeats) and 123 S-SAP (sequence specific amplified polymorphism) primers, we found 135 SSR и 123 S-SAP polymorphic markers among 96 *Brassica rapa* L. samples from the VIR (N.I. Vavilov Institute of Plant Genetic Resources) core collection. The PIC values for both markers, SSR and S-SAP markers were 0.316, 0.257 and 0.379 (50 % higher on average), respectively. Expected heterozigosity (H_E) is usually used to describe the genetic diversity because it is less sensitive to the sample size compared to observed heterozigosity (H_O). **T**he crossings in the population are occasional, if H_O and H_E are similar (i.e., no reliable differences found). They are related as $H_O < H_E$ in inbred population, and as $H_O > H_E$ in case of occasional crossing prevailing compared with inbreeding. Effective multiplex ratio (EMR) is calculated as total number of polymorphic loci per primer multiplied by the rate of polymorphic loci from their total number (W. Powell с соавт., 1996; J. Nagaraju с соавт., 2001). Marker index (MI) is a statistical parameter used to estimate total utility of the maker system; the higher MI, the better method is used) (W. Powell et al., 1996; J. Nagaraju et al., 2001). Resolving power (Rp) is a parameter characterizing ability of the primer/marker combination to detect differences between large numbers of genotypes (J.E. Gilbert et al., 1999; A. Prevost et al., 1999). The information about some software which can be used for calculation of polymorphism information content value and heterozygosity is also summarized. The formula for effective multiplex ratio, marker index calculation, and resolving power calculation are shown.

Keywords: heterozygosity, polymorphism information content value, effective multiplex ratio, marker index, resolving power, software.