

УДК 631.527.3:633.11

## ПРОГНОЗ ЭФФЕКТИВНОСТИ ОТБОРА В ПОПУЛЯЦИЯХ ПОТОМСТВА ПО КОСВЕННЫМ ОЦЕНКАМ ГЕНЕТИЧЕСКОЙ ДИВЕРГЕНЦИИ РОДИТЕЛЕЙ НА ПРИМЕРЕ МЯГКОЙ ЯРОВОЙ ПШЕНИЦЫ\*

А.В. СМИРЯЕВ, М.Г. ДИВАШУК, Т.И. ХУПАЦАРИЯ, С.С. БАЖЕНОВА, Т.Т. НГУЕН

(РГАУ-МСХА имени К.А. Тимирязева)

*Проведен анализ надежности нескольких косвенных количественных оценок генетической дивергенции родительских пар для прогноза эффективности отбора в популяциях их потомства. Проверяется упрощенное представление о положительных корреляционных связях генетического несходства родителей с доступной для отбора изменчивостью количественных признаков в потомстве. На основе результатов полевых экспериментов с мягкой яровой пшеницей, а также с помощью биометрико-генетического анализа выявлены возможные причины и ситуации вариабельности и разнонаправленности этих корреляционных связей.*

*Ключевые слова: подбор родительских пар, метрики дивергенции по фенотипам и SSR маркерам, прогноз изменчивости потомства, причины ненадежности прогнозов, расчетные количественные признаки.*

Генетическая детерминация изменчивости количественных признаков в популяциях сельскохозяйственных растений, как правило, неизвестна и лабильна при изменении условий выращивания, а также в зависимости от генотипов родительских форм [1]. Поэтому в последние десятилетия для подбора родительских сортов, нацеленного на увеличения полезной для отбора изменчивости по количественным признакам у потомства, используют различные оценки генетической дивергенции родителей, т.е. несходства их нуклеотидного или аллельного состава.

В частности, для косвенной оценки дивергенции обобщают в одно числометрику фенотипическое несходство в проявлении комплекса количественных признаков у пары родительских генотипов с помощью метрики Махаланобиса, евклидова расстояния, несходства формы реакции и др. [11]. Используют также метрики несходства, основанные на родословных родителей [5], их биохимических [21] и молекулярных [18, 19] маркеров.

При этом упрощенно предполагается: чем более несходны генетически родительские формы, тем шире полиморфизм популяции потомства по генам, опреде-

\* Работа по детекции SSR маркеров выполнена при финансовой поддержке Министерства образования и науки Российской Федерации в рамках выполнения ФЦП «Исследования и разработки по приоритетным направлениям развития научно-технологического комплекса России на 2007-2012 годы» ГК № 16.552.11.7032 от 29 апреля 2011 г. на оборудовании ЦКП «ВНИИСБ».

ляющим изменчивость количественных признаков, в т.ч. хозяйственно ценных. В результате в популяции потомства за счет рекомбинации возникнет больше трансгрессивных вариантов, увеличатся среднеквадратические отклонения признаков, вследствие чего эффективность отбора новых удачных генотипов в потомстве должна повыситься.

Однако надежность подобных методов подбора пар не всегда высока. В частности, нарушения подобных упрощенных представлений были выявлены при проверке эффективности биометрико-генетического метода подбора родительских пар у мягкой яровой пшеницы, основанного на метрике так называемого несходства формы реакции родителей [6].

### **Проблема количественного сравнения родительских пар по генетическому несходству в связи с изменчивостью их потомства**

Прежде всего следует уточнить смысл выражения «степень несходства» генотипов пары родителей в связи с уровнем изменчивости признаков, ожидаемой в расщепляющейся популяции их потомства.

Предположим, что в идеале полностью известен нуклеотидный состав каждой родительской формы — чистой линии одного вида. Достаточно ли опираться на общее генетическое несходство двух родителей, точнее на их так называемые генетические расстояния, определяемое с помощью методов биоинформатики [13]? Можно ли утверждать, например, для самоопылителей, что одна пара чистых линий «более несходна» по сравнению со второй, если у первой пары больше несходных нуклеотидов? Ожидается ли большее доступное отбору разнообразие по количественным признакам в позднем поколении — расщепленной популяции потомства от скрещивания первой пары?

По-видимому, не обязательно. Во-первых, потому что большая часть генома у многих видов состоит из не транскрибируемой ДНК (по [17] от 30 до 95%). Возможно, несходство родителей в основном касается именно «молчащей» части генома. Во-вторых, даже если известно о большем несходстве нуклеотидных цепочек экспрессирующихся генов, нет гарантии, что большее генетическое расстояние родителей вызовет дополнительную изменчивость конкретных признаков в популяции потомства. Действительно, это несходство у двух пар родителей может касаться разных участков одних или разных генов. Кроме того, генетическая детерминация изменчивости количественных признаков в любой популяции потомства, как известно, существенно меняется по условиям выращивания [1].

Далее следует учесть, что часто при косвенной оценке дивергенции учитываются различия родительских форм лишь по некоторым «удобным для изучения» маркерным локусам или нуклеотидным последовательностям, составляющим незначительную часть генома. Остается надеяться на положительную корреляционную связь подобных косвенных оценок общего несходства родителей (например, по маркерам на QTL или микросателлитным сайтам) с несходством «важных» генов, определяющих изменчивость хозяйственно ценных признаков.

Другой косвенный подход к подбору пар по степени дивергенции основан, как отмечено выше, на несходстве фенотипов родителей. Точнее, предполагается положительная корреляционная связь какой-либо метрики фенотипического несходства родителей с их несходством по «важным» генам ценных признаков.

Цель работы — исследовать возможные причины невысокой надежности прогнозов, основанных на косвенных оценках генетической дивергенции родителей, для эффективного отбора по количественным признакам в популяциях их потомства.

## Метрики оценки дивергенции

Для оценки дивергенции 6 сортообразцов мягкой яровой пшеницы в работе использованы 2 метрики несходства, оцененные по микросателлитным локусам (см. ниже), а также 3 метрики, основанные на фенотипических проявлениях количественных признаков у родителей. Это известные метрики Махаланобиса и евклидова расстояния [11], а также метрика несходства формы реакции генотипов.

Последнее понятие и метод оценки этой метрики несходства родительских сортообразцов были предложены ранее [8, 9]. В основе экспериментальной оценки такого несходства — многократные одновременные измерения в онтогенезе и в разных условиях совместного испытания родителей какого-либо одного ростового признака, который выступает в качестве индикатора несходства.

В каждом условии испытания получают кривую относительного роста этого признака для каждой пары из набора сравниваемых родительских сортообразцов. Обобщают такие данные из разных условий выращивания пары сортообразцов — аппроксимируют данные единой квадратичной регрессией относительной изменчивости признака в процессе роста. Количественно метрика парного несходства двух родителей равна среднему квадрату отклонений всех значений измеренного признака двух родителей от этой регрессии [12].

Предполагается, что каждое существенное изменение условий выращивания в полевом эксперименте вызывает отклонения скорости роста количественного признака, причем различия таких реакций двух сравниваемых сортообразцов в онтогенезе существенно зависят от большого числа полиморфных локусов. Позже происходит частичная компенсация отклонений физиологических процессов, в т.ч. ростовых, но также по-разному в зависимости от аллельного состава сортообразцов [15]. Включение в общую метрику несходства повторных оценок ростового признака по каждому году совместного испытания сортообразцов увеличивает информацию об их генетическом несходстве по все большему числу локусов. Это следствие так называемого переопределения генетической формулы количественного признака [1]. Таким образом, названная метрика для конкретного ростового признака (например, для многократно измеренной высоты растений) «накапливает» информацию о генетическом несходстве сравниваемых родительских форм.

Примеры оценки такой метрики у сортообразцов, где в качестве признака-индикатора использовалась высота растений, представлены в работах с озимой [10] и яровой пшеницей [12].

Микросателлитный анализ 6 родительских сортообразцов проводили с помощью 34 SSR-маркеров, протестированных в предыдущих работах, по методикам, описанным ранее [3]. Полученные длины микросателлитных сайтов использовали для косвенной оценки генетической дивергенции 15 пар 6 родителей. Основное предположение: чем более несходны длины сайтов с одинаковой локализацией у двух родителей, тем выше степень их общей генетической дивергенции. Несходство длин суммируется по 34 сайтам.

Принимая во внимание возможные различия частоты мутаций разных микросателлитных сайтов, в частности, зависимость скорости изменения числа нуклеотидов от длины сайта [22], в работе была проверена надежность двух метрик генетической дивергенции 6 родителей. Первая — манхеттенское расстояние пары родителей:

$$S_M = \sum_{n=1}^{\max} nm_n,$$

где  $m_n$  — число микросателлитных сайтов с одинаковой локализацией у двух родительских генотипов, но длиной, отличающейся на  $n$  нуклеотидов ( $n = 1..max$ ). Эта метрика не учитывает названные различия сайтов по скорости мутирования, что может вызвать преувеличенное влияние на  $S_m$  тех сайтов, которые в процессе эволюции и селекции меняют длину быстрее остальных.

Вторая метрика нивелирует эти различия скорости мутаций и оценивает лишь общее число микросателлитных сайтов с одинаковой локализацией у двух родителей, но разной длиной:

$$S_{01} = \sum_{n=1}^{\max} m_n .$$

Метрика  $S_{01}$  также имеет ограничение. В процессе создания любого сорта с использованием рекомбинационной селекции в его геноме закрепляются участки хромосом от разных сортообразцов того же вида, входящих в родословную сорта. Эти сортообразцы в разной степени несходны между собой. В результате при оценке генетической дивергенции любой пары сортов с использованием микросателлитных сайтов их разная длина косвенно отражает несходство аллелей участков хромосом, дивергентных в разной степени. Метрика  $S_{01}$  не учитывает этот фактор, так как игнорирует количественные различия длин микросателлитных сайтов с одинаковой локализацией.

Далее значения метрики Махаланобиса обозначены  $M^*$ , евклидово расстояния —  $E$ , несходства по форме реакции —  $H$ , по микросателлитам  $S_m$   $S_{01}$ . Возможность использования пяти названных метрик парного несходства родителей для прогноза разнообразия, доступного для отбора по количественным признакам в популяциях потомства, следовало проверить на основе экспериментальных данных.

### Экспериментальный материал

В качестве материала использовали 6 родительских сортообразцов яровой мягкой пшеницы из коллекции системы ВИР им. Н.И. Вавилова и 15 их гибридных популяций  $F_3$  (2009),  $F_4$  (2010) и  $F_5$  (2011) без отбора, полученных после попарного скрещивания этих сортообразцов [6].

Опыты проводили на полях лаборатории селекции и семеноводства полевых культур РГАУ-МСХА имени К.А. Тимирязева. Ежегодно посев 6 сортообразцов и 15 гибридных популяций в 3 рандомизированных блоках проводили вручную. Каждая делянка состояла из пяти учетных и двух защитных рядков по 20 зерен в каждом. Размер делянки: 1 м x 1,26 м. Посев проводили поперек полосы шириной 1 м с междурядьями 18 см, расстояния между соседними делянками 36 см, межполосная дорожка — 50 см. Уборку проводили вручную в фазе полной спелости, защитные и учетные рядки убирали отдельно.

Общее количество растений в учетных рядках из-за негативного влияния погодных условий разных лет колебалось в пределах от 70 до 100. На каждой делянке со всех пяти учетных рядков было использовано для измерений по 30 случайно выбранных растений. После уборки на них измеряли 7 признаков — элементов структуры урожая. Это высота растения (признак номер 1), длина колосового стержня главного колоса (2), число продуктивных колосков главного колоса (3), число зерен главного колоса (4), масса зерна главного колоса (5), число зерен боковых колосьев (6), масса зерна боковых колосьев (7). Кроме того, анализировали

\* обозначение метрик здесь и далее выделено авторами.

три расчетных признака: число зерен с растения (8), массу зерна с растения (9), массу 1000 зерен (10).

Для посева семян любой гибридной популяции следующего поколения случайно выбирали одинаковое число семян каждого из 30 растений того же повторения предыдущего поколения. Благодаря этому влияние дрейфа генов было минимальным.

С целью экспериментальной оценки изменчивости 10 признаков в каждой из 15 популяций-семей  $F_3$ ,  $F_4$ ,  $F_5$  отдельно вычисляли показатель  $\sqrt{D}$  — среднеквадратическое отклонение для каждого признака. Этот показатель оценивает реальное, доступное для отбора разнообразие признака в популяции. Кроме того, оценивали  $m$  — среднее значения признака в каждой популяции.

Метрику Махаланобиса  $M$  парного несходства 6 родительских сортообразцов (всего 15 значений) оценивали по величинам всех их 10 признаков, каждый из которых был предварительно усреднен по данным за 4 года испытания. Для расчета 15 значений метрики  $M$  по этим усредненным данным использовали программу из пакета AGROS. Поскольку евклидово расстояние в отличие от метрики Махаланобиса не учитывает корреляционных связей между признаками, 15 значений парного евклидова расстояния 6 сортообразцов рассчитывали по той же программе, но все ковариации признаков задавали равными нулю.

Были оценены ранее в четырехлетнем эксперименте [12] 15 значений метрики  $H$  — парного несходства формы реакции 6 родителей.

Для оценки надежности прогнозов изменчивости признаков в потомстве по пяти метрикам-прогнозам ( $M$ ,  $E$ ,  $H$ ,  $S_M$ ,  $S_{01}$ ) вычисляли  $r$  — коэффициенты корреляции между значениями этих метрик для 15 пар родителей и 15 экспериментальными значениями  $\sqrt{D}$  по каждому из 10 признаков в их потомстве ( $F_3$ , , или  $F_5$ ).

### Результаты и их обсуждение

Полученные по одним четырехлетним данным о 6 родительских сортообразцах 15 значений метрики Махаланобиса  $M$  и 15 значений евклидова расстояния  $E$ , имели между собой значимую, но невысокую корреляцию  $r(E, M) = 0,70$ . Это можно объяснить тем, что в евклидовом расстоянии ковариации между 10 признаками считаются равными нулю. Отметим, что оценки этих ковариаций в метрике Махаланобиса, по-видимому, несут существенные статистические ошибки: слишком мала выборка родительских форм (6), по которой оценивались ковариации. Всего оценок ковариаций  $10 \cdot 9/2 = 45$ . Ясно, что случайная ошибка 15 значений  $M$ , в расчете каждого из которых участвуют все 45 параметров одновременно, достаточно велика.

Метрики  $E$  и  $M$  слабо коррелировали с остальными тремя метриками несходства:  $r(H; M) = 0,39$ ,  $r(S_M; M) = 0,18$ ,  $r(S_{01}; M) = 0,52$ ,  $r(H; E) = 0,34$ ,  $r(S_M; E) = 0,20$ ,  $r(S_{01}; E) = 0,50$ . Оценки корреляций остальных 3 метрик между собой в среднем выше:  $r(S_{01}; S_M) = 0,82$ ,  $r(H; S_M) = 0,48$ ,  $r(H; S_{01}) = 0,46$ , хотя две последние корреляции немного ниже критического уровня, равного 0,51 при  $df = 13$ .

Далее вычисляли коэффициенты корреляции ( $r \cdot 100$ ) метрик-прогнозов ( $M$ ,  $E$ ,  $H$ ,  $S_M$  и  $S_{01}$ ), косвенно оценивающих парное генетическое несходство 6 сортообразцов, с  $\sqrt{D}$  — наблюдаемой изменчивостью 10 признаков в их 15 гибридных популяциях:  $F_3$  (2009),  $F_4$  (2010),  $F_5$  - (2011). В соответствии с обсуждаемым упрощенным представлением ожидали получить положительные корреляции. Однако результаты расчетов оказались неожиданными (см. таблицу).

**Коэффициенты корреляции (г-100)**  
(смысл параметра  $\max = H_y \cdot 100$  пояснен в следующем разделе)

Номер** признака	1	2	3	4	5	6	7	8	9	10	
$F_3$ (2009)	<i>M</i>	12	-25	25	-46	-31	-60*	-62*	-61*	-65*	-19
	<i>E</i>	14	26	45	-29	-14	-79*	-66*	-79*	-69*	8
	<i>H</i>	-2	-32	-1	-4	-17	-23	-28	-4	-64*	-25
	$S_M$	65*	26	-2	08	-02	-40	-45	-36	-46	18
	$S_{01}$	56*	31	31	-23	-13	-67*	-73*	-67*	-74*	11
	max	94	79	82	65	40	70	70	75	62	86
$F_4$ (2010)	<i>M</i>	27	5	26	3	4	-4	-26	-38	-24	12
	<i>E</i>	10	17	29	-2	0	-31	0	-28	-1	31
	<i>H</i>	-30	9	-1	-4	-52*	-53*	-52*	-57*	-63*	52*
	$S_M$	46	84*	44	20	7	-14	0	-16	-5	54*
	$S_{01}$	47	71*	51*	-1	5	-32	-11	-31	-13	43
	max	95	77	86	77	76	93	85	92	85	88
$F_5$ (2011)	<i>M</i>	20	13	10	-9	-17	-76*	-64*	-76*	-63*	10
	<i>E</i>	28	43	7	20	9	-62*	-44	-61*	-41	24
	<i>H</i>	-15	19	19	21	-5	-41	-49	-34	-43	42
	$S_M$	40	47	-18	0	-30	-41	-42	-39	-42	55*
	$S_{01}$	65*	57*	1	1	-26	-72*	-70*	-73*	-69*	56*
	max	73	74	82	40	46	78	63	76	66	72

\* значимые коэффициенты корреляции при  $p < 0,05$ ; \*\* номера признаков — из раздела «Экспериментальный материал».

Например, для метрики *H* в  $F_3$  единственный значимый коэффициент корреляции  $r$  проявился с изменчивостью продуктивности гибридных растений (9-й признак), причем  $r$  был отрицателен (-0,64\*). В следующем поколении  $F_4$  он также достоверно отрицателен (-0,63\*), а корреляции  $r$  с несколькими другими признаками увеличились и также стали значимо отрицательными; т.е. для них коэффициенты  $r$  имеют знак, противоположный по сравнению с ожидаемым по прогнозу *H*. Подобные отрицательные корреляции получены для остальных четырех метрик, хотя не все они значимы. Исключением стал признак 10 — масса 1000 зерен: в  $F_4$  для него корреляция с *H* достоверно положительна. Положительны корреляции для 10-го признака и с остальными четырьмя метриками в хотя по трем из них корреляции незначимы по уровню.



Отмечена также особая тенденция изменения значимости корреляций показателя  $\sqrt{D}$  у признаков 6-9 с разными метриками в поколениях. Например, в  $F_3$  и  $F_5$  все эти корреляции с  $H$  незначимы (хотя и сходны с  $F_4$ ). В тех же двух поколениях  $F_3$  и  $F_5$  корреляции  $\sqrt{D}$  тех же признаков с метриками  $M$  и  $S_{01}$  значимы. И, наоборот, в  $F_4$  для  $H$  эти корреляции значимы, а для  $M$  и  $S_{01}$  — незначимы.

Следует изучить причины изменчивости корреляций  $r$ , проявившейся в расчетах с реальными экспериментальными данными. При этом можно выделить, как минимум, три типа изменчивости:

- 1) изменчивость по 3 годам корреляции  $r$  одной метрики при ее сопоставлении с показателем  $\sqrt{D}$  одного признака;
- 2) изменчивость корреляции по 10 признакам (вплоть до изменения знака показателя  $r$ ) для одной метрики в пределах одного года;
- 3) изменчивость корреляции значений  $\sqrt{D}$  одного признака для данных одного года с оценками генетической дивергенции 6 родительских сортообразцов на основе 5 разных метрик.

### Возможные причины изменчивости первого типа

Одной из основных причин вариабельности коэффициента корреляции по трем годам является сильное взаимодействие генотип-среда по каждому признаку. Из примера на рис. 1 следует, что по годам меняются не только значения, но даже ранги показателя  $\sqrt{D}$ .

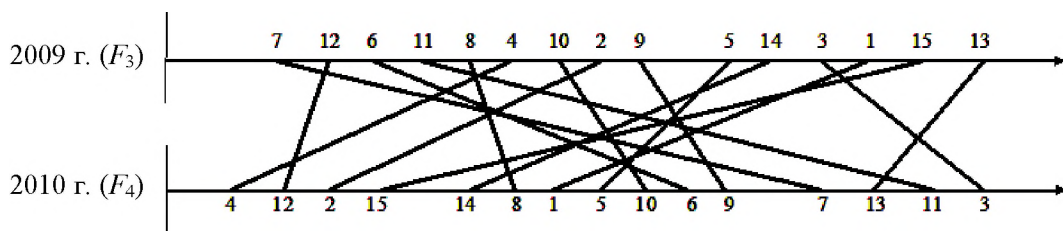


Рис. 1. Пример изменения показателя  $\sqrt{D}$  15 гибридных популяций по признаку 9 — масса зерна с растения за 2 года. Приведено к одному масштабу [6]

Напомним, что в отличие от показателя  $\sqrt{D}$ , значения метрики  $H$  для пар родительских форм были оценены ранее на основе 4-летних данных [12] и считаются постоянными как в поколениях, так и по признакам. То же относится к метрикам  $E$  и  $M$ . Метрики  $S_M$ ,  $S_{01}$ , естественно, вообще не изменяются по годам. Поэтому наблюдаемая изменчивость рангов показателя  $\sqrt{D}$  15 гибридных популяций в поколениях затрудняет экспериментальную оценку по  $r$  эффективности подбора пар с помощью любой метрики, отражающей генетическое несходство родительских форм.

Второй причиной отсутствия значимых коэффициентов корреляции какого-либо признака, особенно в поколении  $F_3$  (см. таблицу), может выступать разрушение сформированных удачных генетических блоков несходных родительских генотипов вследствие их гибридизации, сопровождающейся рекомбинацией. Это приводит к различию приспособленностей и естественному отбору гибридных генотипов в семьях — ранних поколениях самоопыления [2, 16].

Третья возможная причина — снижение уровня остаточной гетерозиготности в 15 популяциях-семьях потомства, особенно при переходе от  $F_3$  к  $F_4$ .

В качестве еще одной причины вариабельности  $r$  первого типа следует назвать изменчивость коэффициента наследуемости в широком смысле для показателя  $\sqrt{D}$  любого признака по годам. Это связано с модификационной изменчивостью признака и с разной степенью пестроты почвенного плодородия экспериментального поля по годам. Суть в следующем.

Корреляции  $r$ , по сути, являются фенотипическими (обозначаются  $r_p$ ), так как для их оценки использованы значения  $\sqrt{D}$ , усредненные по трем повторениям опыта. В биометрической генетике известна формула для  $r_p$  двух признаков —  $x$  (в данном случае значения какой-либо метрики несходства родителей) и  $y$  (в данном случае  $\sqrt{D}$  признака у потомства) [7]:

$$r_p = H_x H_y r_g + e_x e_y r_e,$$

где  $r_g, r_e$  — генотипическая и средовая корреляции признаков;  $H = \sqrt{H^2}$ , т.е. корень из коэффициента наследуемости в широком смысле;  $e = \sqrt{1 - H^2}$ .

Если упрощенно считать, что значения любой метрики определены без ошибок, то  $H_x = 1, e_x = 0$ ; тогда  $r_p = H_y r_g$ . Следовательно, модуль корреляции  $r_p$  для любого признака, даже если модуль истинной генотипической корреляции ( $r_g$ ) близок к единице, не может быть больше значения  $H_y$ . Для любого года и признака величину  $H_y$  можно оценить из обычного дисперсионного анализа показателя  $\sqrt{D}$  в повторениях опыта, причем  $H_y$  не постоянно по годам. Значения  $\text{max} = H_y \cdot 100$  приведены в таблице.

Следует при этом отметить, что оценка  $H_y$  для каждого года и признака, а также сами оценки корреляций  $r \cdot 100$ , несут статистические ошибки из-за ограниченного объема выборки показателя  $\sqrt{D}$  (15 гибридов в трех повторениях) и неполного соответствия экспериментальных данных предположениям дисперсионного анализа [14]. Поэтому модули некоторых значений  $r \cdot 100$  в таблице оказались немного больше соответствующих оценок  $\text{max} = H_y \cdot 100$ .

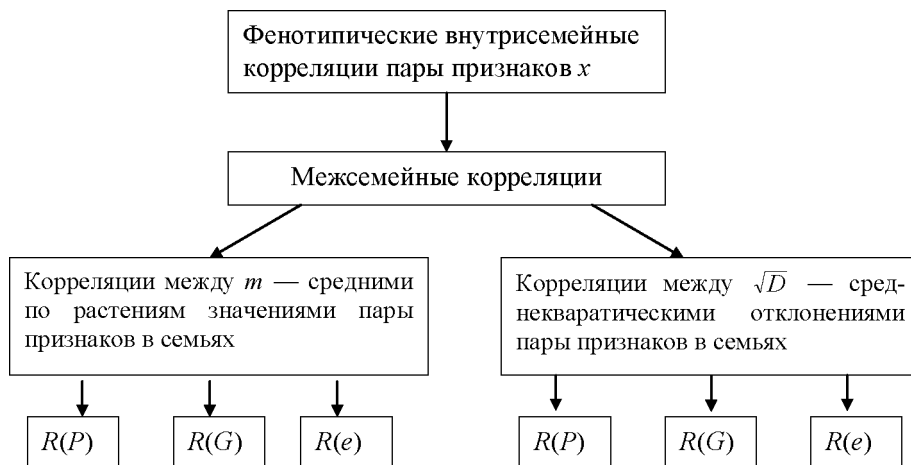
### Возможные причины изменчивости второго типа

Названные причины вариабельности коэффициентов корреляции  $r$  первого типа известны из литературных данных. Однако они не дают объяснения разнонаправленности корреляций отдельных метрик с показателем  $\sqrt{D}$  некоторых признаков в каком-то конкретном поколении. Например, почему метрика  $H$  в  $F_4$  достоверно положительно коррелирует с последним признаком, но достоверно отрицательно — с несколькими предыдущими (см. таблицу)? Это нельзя считать чистой случайностью, так как аналогичная ситуация повторяется для метрики  $S_{01}$  в  $F_5$ .

Рассмотрим влияние на коэффициенты  $r$  эффектов генов, вызывающих скоррелированную изменчивость для пар самих количественных признаков. Эти эффекты генов отражены в коэффициентах их корреляций. Можно выделить два основных уровня парных корреляционных связей изучаемых признаков (рис. 2): внутрисемейный и межсемейный.

Первые (внутрисемейные) объясняются плейотропными эффектами генов, определяющих изменчивость признаков и, возможно, конкуренцией растений на делянке каждой популяции потомства. Для любой пары признаков в каждой





**Рис. 2.** Структурная схема парных корреляций признаков.  $R(P)$ ,  $R(G)$ ,  $R(e)$  — межсемейные коэффициенты корреляции: фенотипические, генотипические, средовые

из 15 семей коэффициенты корреляции оценивались по измерениям 30 растений в трех повторениях, т.е. по 90 растениям.

Вторые (межсемейные) оценивают на основе обобщенных показателей, полученных внутри каждой из 15 семей-популяций в трех повторениях:  $m$  — среднего значения признака в семье или  $\sqrt{D}$  — среднеквадратического отклонения признака. Плейотропные эффекты в этом случае влияют на межсемейные корреляции  $m$  или  $\sqrt{D}$  опосредовано.

Имея данные в повторениях, каждую парную межсемейную фенотипическую ковариацию как между 15 значениями  $m$ , так и  $\sqrt{D}$ , можно подразделить на генотипическую и средовую. После чего оценить коэффициенты корреляции:  $R(P)$ ,  $R(G)$ ,  $R(e)$ .

Анализ межсемейных корреляций, например, в  $F_4$  (2010) выявил плеяду четырех тесно коррелирующих по  $\sqrt{D}$  признаков с номерами 6, 7, 8, 9: оценки  $R(P)$  — парных фенотипических корреляций менялись в пределах от 0,85 до 0,98 [6]. Причем  $R(G)$  — коэффициенты генотипические корреляции были даже несколько выше. То же относилось к корреляции этих 4 признаков по  $m$  — средним значениям в семьях (от 0,73 до 0,98). Фактически названная группа из 4 признаков выступает в каждой из 15 семей как один признак, т.е. их изменчивость в основном определяется одними генами. Ситуация в 2011 г. аналогична, но корреляции несколько ниже.

Поэтому, в частности, каждый из 4 названных признаков в 2010 г. показал сходную по величине отрицательную корреляцию  $\sqrt{D}$  с  $H$  — метрикой парного генотипического несходства родителей. Корреляция  $H$  с  $\sqrt{D}$  в  $F_4$  для 10-го признака (масса 1000 зерен), напротив, значимо положительная (см. таблицу). Причем 10-й признак не коррелирует ни с одним из остальных 9. Напомним, что этот признак является расчетным по отношению к паре других признаков. Его значение  $X_{10}$  для каждого растения внутри семьи равно

$$X_{10} = X_9 / X_8. \quad (1)$$

Внутрисемейные корреляции значений  $x_9$  9-го признака (масса зерна с растения) с  $x_8$  8-го (число зерен с растения) в пределах каждой из 15 семей также весьма высоки: от 0,83 до 0,96. Следовательно, в любой семье можно достаточно надежно выразить признак  $x_9$  через признак  $x_8$ :

$$x_9 = a + bx_8, \quad (2)$$

где  $a, b$  — коэффициенты линейной регрессии.

Подставляя уравнение (2) в (1) и деля числитель в правой части на знаменатель, получаем:

$$x_{10} = a/x_8 + b. \quad (3)$$

Из справочника Э. Ллойда и У. Ледермана [4] с учетом уравнения (3) получаем приблизительную оценку связи показателя  $\sqrt{D}$  для 8-го и 10-го признаков в пределах каждой семьи:

$$\sqrt{D_{10}} \approx |a| \cdot \sqrt{D_8} / m_8^2, \quad (4)$$

где  $m_8$  — оценка математического ожидания 8-го признака в семье.

Учитывая соотношение (4) для 15 семей в предположении слабых колебаний коэффициента  $a$  по семьям (что справедливо при тесной межсемейной корреляции между  $\sqrt{D_8}$  и  $\sqrt{D_9}$ ), можно сделать следующий вывод. Если при изменении  $\sqrt{D_8}$  по семьям знаменатель ( $m_8^2$ ) в уравнении (4) меняется скоррелированно с  $\sqrt{D_8}$ , но растет быстрее, чем  $\sqrt{D_8}$ , то уменьшение  $\sqrt{D_8}$  по семьям сопровождается увеличением  $\sqrt{D_{10}}$  в них. Этих условий достаточно для возникновения положительной корреляции 15 значений  $H$  с  $D$  при отрицательной корреляции  $H$  с  $\sqrt{D_8}$ . Если же знаменатель в (4) растет приблизительно синхронно с числителем, то, несмотря на корреляцию  $\sqrt{D_8}$  с  $H$ , значимой корреляции  $\sqrt{D_{10}}$  с  $H$  не будет, т.е. связь  $\sqrt{D_{10}}$  в семьях с изменением генетической дивергенцией родительских сортообразцов не проявится ( $r \approx 0$ ). Этот вывод справедлив не только для метрики  $H$ , но и для любой другой метрики генетической дивергенции родителей.

Отметим, что из уравнения (1) следует: изменчивость 10-го признака, определяется теми же генами, что 8-го и 9-го. Напомним, что два последних признака относятся к одной плеяде: они тесно коррелируют между собой как в каждой семье, так и по 15 семьям, т.е. изменчивость ( $\sqrt{D}$ ) всех трех признаков (8, 9, 10), хотя и определяется практически одними генами, но их экспрессия различается по признакам. То же относится к показателю  $m_8$  — значению признака, среднему по растениям семьи. В пределах каждой семьи  $m_8$  вычисляется по тем же экспериментальным данным (измерениям  $x_8$ ), что и  $\sqrt{D_8}$  — среднеквадратическое отклонение признака. Поэтому изменчивость обоих показателей зависит от одних генов, но проявляющихся по-разному в  $m_8$  и  $\sqrt{D_8}$ .

С помощью олигогенных биометрико-генетических моделей несложно показать, что в подобных ситуациях схемы наследования трех признаков (8, 9, 10) внутри семей различаются. Например, если в какой-либо семье схема наследования для 8-го и 9-го признаков аддитивно-доминантная, то в схеме 10-го признака почти наверняка проявится эпистаз. Аналогична ситуация для межсемейной изменчивости показателей  $m$  и  $\sqrt{D}$  любого признака.

В более общем случае, без предположения о тесной корреляции по семьям между  $\sqrt{D_9}$ ,  $\sqrt{D_8}$  и  $m_8$ , с использованием оценок из справочника [4] можно показать:

$$\sqrt{D_{10}} \approx |V_9 - V_8| \cdot m_9 / m_8,$$

где  $|V_9 - V_8|$  — модуль разности коэффициентов вариации 9-го и 8-го признаков.

При определенных вариантах совместного варьирования показателей 9-го и 8-го признаков положительная корреляционная связь  $\sqrt{D_{10}}$  в семьях потомства с любой метрикой генетической дивергенции родителей приведет к отрицательной связи последней с показателями  $\sqrt{D_8}$ ,  $\sqrt{D_9}$  или к отсутствию корреляционной связи с ними. Все зависит от различий экспрессии генов, определяющих изменчивость показателей  $m$ ,  $\sqrt{D}$ , 8-го и 9-го признаков.

Существуют и другие формы связи между исходными и расчетными признаками. Например,  $x_0$  — общую массу растения можно вычислить как сумму  $x_2$  — массы полезной части, например колоса, и  $x_1$  — остальной:

$$X_0 = X_1 + X_2.$$

Если внутрисемейные коэффициенты корреляции между  $X_1$  и  $X_2$  близки к -1 (например, из-за внутрисемейного полиморфизма по аттракции пластических веществ из соломы в колос), то в любой семье показатель  $\sqrt{D_0}$  будет связан с дисперсиями  $D_1$  и  $D_2$  следующим образом:

$$\sqrt{D_0} \approx |\sqrt{D_1} - \sqrt{D_2}|.$$

Кроме того, положительная межсемейная регрессионная связь  $\sqrt{D_1}$  и  $\sqrt{D_2}$  при их увеличении может привести к снижению модуля разности, что будет означать снижение  $\sqrt{D_0}$  в семьях. Коэффициенты корреляции любой метрики генетической дивергенции родителей с  $\sqrt{D_1}$  и  $\sqrt{D_2}$  в семьях потомства снова будут иметь противоположный знак по сравнению с  $\sqrt{D_0}$ .

Из этих примеров следует, что биометрико-генетический анализ позволяет сформулировать различные условия, при которых полезная для отбора изменчивость расчетного количественного признака у потомков не будет коррелировать с метрикой генетического несходства родителей или будет иметь обратный знак корреляции по сравнению с исходными признаками. Возможно, именно эти причины, а не недостатки конкретных косвенных оценок дивергенции родителей, объясняют разнонаправленность корреляций, например, метрики  $H$  с изменчивостью ( $\sqrt{D}$ ) в семьях по некоторым признакам и отсутствие значимых корреляций  $H$  с  $\sqrt{D}$  для других признаков. Аналогичное соотношение корреляций  $z$  метрики-прогноза  $S_{01}$  в  $F_5$  с реальным показателем  $\sqrt{D}$  10-го признака и с двух предыдущих признаков (см. таблицу), по-видимому, объясняется теми же причинами.

Кроме того, сами понятия «расчетный» и «исходные признаки» условны. Можно рассматривать 9-й признак как расчетный, а 8-й и 10-й как исходные для него. Отметим также, что при желании практически для любого количественного признака можно подобрать новые «исходные» и он станет «расчетным».

Следует заключить, что увеличение генетической дивергенции родителей, т.е. общего несходства их аллельного состава, не гарантирует увеличения полезного для отбора разнообразия в потомстве по всем изучаемым количественным признакам. Для части признаков возможно отсутствие реакции или даже закономерное снижение ответной изменчивости в потомстве.

### Возможные причины изменчивости третьего типа

Следует подчеркнуть, что в работе сравнивалась эффективность метрик, основанных на качественно отличных данных о 6 родительских сортообразцах. Это  $S_M$ ,  $S_{O1}$  — несходство длин их микросателлитных сайтов,  $M$ ,  $E$  — несходство фенотипических значений 10 признаков сортообразцов на момент уборки и данных, полученных в онтогенезе по одному признаку — высоте растений (метрика  $H$ ). Несмотря на эти качественные отличия, ни разу не проявилась ситуация, когда среди пяти метрик для одного признака в одном поколении были как значимо положительные, так и значимо отрицательные значения  $r$ . Но величины корреляций менялись существенно. Максимальная вариабельность наблюдалась для 5-го признака в  $F_4$ : коэффициент корреляции  $r$  для пяти метрик менялся от 0,07 до -0,52\* (см. таблицу).

Наблюдаемые различия корреляций  $r$  разных метрик с показателем  $\sqrt{D}$  одного признака в один год естественны и вызваны причинами, аналогичными рассмотренными в предыдущих двух разделах. Прежде всего, такая вариабельность  $r$  третьего типа указывает на ассоциацию пяти метрик с разными полиморфными генами как сильно, так и слабо влияющими на изменчивость конкретного признака. Причем некоторые из этих ассоциаций наверняка имеют нелинейный характер.

Кроме того, лишь упрощенно можно считать, что величины исследуемых метрик несходства родителей определены без ошибки. На самом деле они имеют различные и трудно оцениваемые ошибки.

Так, существенную ошибку выборочности вызывает ограниченное число проанализированных локусов при оценке метрик, вычисленных по молекулярным маркерам [22]. Чем меньше таких маркеров, тем в среднем менее равномерно в метрике отражено несходство разных участков генома сортообразцов.

Отметим, что метрика  $S_M$ , в таблице имеет лишь 4 из 30 корреляций, значимо отличающихся от 0, тогда как для  $S_{O1}$  14 из 30 корреляций значимы. По-видимому, ошибка метрики  $S_M$ , при оценке генетического несходства 6 сортообразцов больше, чем у  $S_{O1}$ , что обусловлено их особенностями. Например, метрика  $S_M$ , как отмечалось выше, не учитывает различия микросателлитных сайтов по скорости мутирования. Эти различия отражены в структуре метрик  $S_M$ , и  $S_{O1}$ .

Метрики  $M$  и  $E$ , основанные на несходстве количественных признаков родительских сортов, не учитывают известного явления конвергентного влияния научной селекции [11]. А именно, сходство требований ко многим признакам сортов одной культуры (модель идеального сорта) часто приводит к их фенотипическому сходству, в частности, по признакам их габитуса. Но это далеко не всегда означает сходства сортов по генотипам. Например, короткостебельность сортов может быть обусловлена разными сочетаниями аллелей в локусах с разной локализацией. Метрики  $M$  и  $E$  этого не учитывают: в них близость родительских форм по величинам количественных признаков подразумевает сходство этих форм по аллельному составу.

Возможно, и по этим причинам возникает вариабельность корреляций  $r$  3-го типа, а многие корреляции в таблице незначимо отличаются от 0.

По-видимому, бессмысленно искать «самую лучшую» метрику, гарантирующую надежный прогноз разнообразия потомства по всем хозяйственно ценным количественным признакам в любых условиях выращивания. При выборе метрик для практического использования необходимо учитывать названные ограничения и недостатки.

### Библиографический список

1. Драгавцев В.А. К проблеме генетического анализа полигенных количественных признаков растений / ВИР. СПб, 2003. 35 с.
2. Жучелко А.А. Экологическая генетика культурных растений. Кишинев: Штиинца, 1980. 587 с.
3. Крутил П.Ю., Дивашук М.Г., Феселко П.А., Карлов Г.П. Адаптация микросателлитных SSR-маркеров пшеницы для анализа геномов пырея среднего, пырея удлинненного и пшенично-пырейных гибридов // Известия ТСХА. 2011. Вып. 3. С. 49-57.
4. Ллойд Э., Ледермал У. Справочник по прикладной статистике. Т. 1. М.: Финансы и статистика, 1989. 450 с.
5. Мартынов С.П. Генеалогический подход к анализу генетического разнообразия зерновых культур с помощью информационно-аналитической системы генетических ресурсов: методические указания / ВИР. СПб., 2006. 87с.
6. Нгуел Т. Т., Смиряев А.В., Бажелова С.С. Оценка эффективности методов подбора родительских пар на примере мягкой яровой пшеницы // Известия ТСХА. 2010. Вып. 6. С. 145-152.
7. Рокицкий П.Ф. Введение в статистическую генетику. Минск: Высшая школа, 1974. 447 с.
8. Смиряев А.В. Экологическая и онтогенетическая изменчивость количественного признака в серии сравнительных полевых испытаний // Доклады ВАСХНИЛ. 1985. №. 8. С. 9-12.
9. Смиряев А.В. Анализ сходства генотипов по эколого-онтогенетической изменчивости ростового признака//Генетика. 1997. Т. 33. С. 61-67.
10. Смиряев А.В., Пыльлев В.В., Тао Юл-Шел. Сопоставление параметров относительной изменчивости ростовых признаков и генетического сходства сортов озимой пшеницы // Генетика. 1997. Т. 33. С. 67-68.
11. Смиряев А.В., Кильчевский А.В. Генетика популяций и количественных признаков. М.: изд-во «КолосС», 2007. 270 с
12. Смиряев А.В., Пыльлев В.В. Биометрико-генетический анализ несходства генотипов в диаллельном скрещивании яровой пшеницы // Генетика. 2008. Т.44. С. 276-282.
13. Смиряев А.В., Палкила Л.К. Основы биоинформатики / РГАУ-МСХА имени К.А. Тимирязева. М., 2008. 103 с.
14. Смиряев А.В. Хунацария Т.П. , Бажелова С.С., Нгуел Т. Т. Совершенствование биометрического метода оптимизации выборки растений при экспериментальном сравнении гибридных популяций // Известия ТСХА. 2012. Вып. 2. С. 71-86.
15. Шведлуха В.С. Рост растений и его регуляция в онтогенезе. М.: Колос, 1992. 258 с.
16. Шмальгаузен П.П. Факторы эволюции. Теория стабилизирующего отбора. М.: Наука, 1968. 451 с.
17. Cavalier-Smith T., *fed*). The Evolution of Genome Size. Wiley, New York. 1985. 523 p.
18. Couvreur Le F., Faure S., Dubreuil P., Praud S. Analysis of genetic structure in a panel of elite wheat varieties and relevance for association mapping // Theor. Appl. Genet. 2011. Vol. 123. P. 715-727.
19. Hai L, Wagner C., Friedt W. Quantitative structure analysis of genetic diversity among spring bread wheats (*Triticum aestivum* L.) from different geographical regions // Genetica. 2007. Vol. 130. P. 213-225.
20. Van Hintum Th. J. L. Data resolution: a jackknife procedure for detennining the consistency of molecular marker datasets // TAG. 2007. Vol. 115. P. 343-349.



21. *Reif J.C., Melchinger A.E., Frisch M.* Genetic and mathematical properties of similarity and dissimilarity coefficients applied in plant breeding and seed bank management // *Crop Sci.* 2005. Vol.45. P. 1-7.

22. *Watkins J.C.* Microsatellite Evolution: Markov Transition Functions for a Suite of Models // *Theoretical population biology.* 2007. Vol. 71. P. 147-159.

## EFFECTIVENESS FORECAST OF OFFSPRING SELECTION IN POPULATIONS BASED ON DIFFERENT INDIRECT ESTIMATES OF PARENTS GENETIC DIVERGENCE ON THE EXAMPLE OF SOFT SPRING WHEAT

A.V. SMIRYAEV, M.G. DIVASHUK, T.I. KHUPATSARYA, S.S. BAZHENOVA, T.T. NGUEN

(Russian State Agrarian University — K.A. Timiryazev MAA)

*Reliability<sup>1</sup> analysis for a number of indirect quantitative estimates of genetic divergence of parental pairs was carried out. to predict the quantitative traits diversity in populations of their offspring. The simplistic view of the positive correlation between parental genetic dissimilarity and available variability of all quantitative traits in their offspring was verified. The results of field experiments with soft spring wheat, as well as using biometric-genetic analysis allowed identification of the possible causes and situations of variability and multi-directional correlations.*

*Key words: choice of parental pairs, phenotype metrics of divergence, SSR-marker metrics, forecast of offspring variability, causes of unreliable forecasts, calculated quantitative traits.*

**Смиряев Анатолий Владимирович** — д. б. н., проф. кафедры генетики и биотехнологии РГАУ-МСХА имени Тимирязева (127550, г. Москва ул. Тимирязевская, 49; тел.: (499) 976-08-94; e-mail: genetics V/ timacad.ru).

**Дивашук Михаил Георгиевич** — к.б.н., с.н.с. Центра молекулярной биотехнологии РГАУ-МСХА имени К.А. Тимирязева (127550, г. Москва, ул. Тимирязевская, 49; тел.: (499) 977-72-01; e-mail: divashuk(@gmail.com).

**Хуацария Титико Ипполитович** — к. б. н., проф. кафедры селекции и семеноводства полевых культур РГАУ-МСХА имени Тимирязева. Тел.: (499) 976-12-72.

**Баженова Светлана Сергеевна** — к. с.-х. н., доцент кафедры селекции и семеноводства полевых культур РГАУ-МСХА имени Тимирязева. E-mail: selection@timacad.ru.

**Нгуен Тхань Туан** (Вьетнам) — к. б. н., преподаватель кафедры селекции и генетики Ханойского аграрного университета (Ханой-234; тел.: 016-797-98-088; e-mail: thanglongmost@yahoo.com).