

УДК 633/635:575.1:57.011

ТЕОРИЯ ОШИБОК ИДЕНТИФИКАЦИИ ГЕНОТИПОВ ОТДЕЛЬНЫХ РАСТЕНИЙ ПО ИХ ФЕНОТИПАМ ПО КОЛИЧЕСТВЕННЫМ ПРИЗНАКАМ В РАСЩЕПЛЯЮЩИХСЯ ПОПУЛЯЦИЯХ НА РАННИХ ЭТАПАХ СЕЛЕКЦИИ**Н.В. КОЧЕРИНА**

Рассматривается проблема селекционной идентификации лучших генотипов растений по их фенотипам. Представлено обоснование принципа фоновых признаков для эффективной работы селекции. Анализировали теорию ошибок линии регрессии с целью применения в качестве теории ошибок идентификации генотипов растений по их фенотипам. Охарактеризована возможность использования достоверности средних величин для подбора родительских пар при гибридизации. Оценивали эффективность применения принципа фоновых признаков в решении задач идентификации в F_2 уникального растения — родоначальника будущего сорта.

До создания принципа фоновых признаков (1, 2) сама возможность разработки методов быстрой оценки (без испытания по потомству) индивидуальных генотипов по их фенотипам по количественным признакам категорически отрицалась на основании утверждений ряда исследователей. Так, Кренке отмечал: «Начиная от первых стадий развития, не существует константно особого фенотипического выражения для модификаций и для наследственных признаков» (3). По мнению Плохинского: «Для одной особи бессмысленно определять, какая часть ее фенотипа обусловлена наследственностью и какая условиями жизни. Генетическая информация, полученная одним индивидуумом, реализуется в таком взаимодействии с условиями жизни, при котором обе причины неотделимы друг от друга» (4). Никоро с соавт. писали: «Для оценки генотипа необходимо знать величину генотипического отклонения данной особи от средней величины популяции (g_i), однако нет способа отделить g_i от e_i — экологического отклонения той же особи» (5). Уильямс утверждал: «В продуктивности одного организма невозможно разделить генетические и внешние воздействия на признаки со слабой наследуемостью и отбор в F_2 ненадежен» (6). По Рокицкому, «Фенотип особей — это единый целостный организм. О генотипе особей непосредственно, без анализа их потомства, судить невозможно» (7). По Савченко, «Вычленив непосредственно для каждого организма влияние на развитие признака генотипа и среды не представляется возможным» (8).

Эти положения, возможно, верны, но лишь в том случае, если отслеживать один признак. По данным Драгавцева, если одновременно следить за двумя признаками, то вышеперечисленные постулаты становятся неточными (1, 2). Ауэрбах подчеркивала, что самым важным для генетики был третий закон Менделя — свободного комбинирования аллелей (9). Расщепление по каждой паре аллелей происходит независимо, если гены родительских форм находятся в разных парах хромосом. Этот закон Мендель открыл, как известно, только потому, что одновременно следил за расщеплением двух признаков семян гороха.

Если в любой популяции растений (дикой или F_2) найти количественный признак с нулевой (или близкой к нулю) генетической дисперсией, то варьирование этого признака будет только экологическим. Тогда особь, имеющая плюсовое отклонение по этому признаку от средней по-

пуляционной — есть модификация, попавшая в лучшую микроэкологическую нишу (этот признак был назван фоновым, ФП). Если другой, селекционный, признак (СП) сдвинут у этой особи в плюс-сторону соответственно фоновому, то это — модификация, и отбирать эту особь не следует. Если у другой особи ФП выражен на уровне средней популяционной величины, а СП сдвинут в плюс-сторону, то это — рекомбинация (или мутация), необходимая для продуктивной селекционной работы.

Принцип ФП даже на этом упрощенном способе хорошо работает в системах парных скрещиваний, которые широко применяют в селекции. Для этого достаточно у обеих родительских форм найти одинаковый по среднему уровню количественный признак, формирующийся в те же фазы онтогенеза, что и СП. Дадим определение ФП для случая, когда $\sigma_{g\text{ФП}}^2 \approx 0$.

Фоновый признак — это такой экологически подвижный признак, который не имеет в конкретных популяциях и среде генотипической изменчивости (или она крайне низка) и положительно экологически коррелирует с СП. Вскоре после создания и испытания принципа ФП было открыто явление разнонаправленности сдвигов количественного признака отдельной особи под влиянием генетических и средовых причин. Это событие практически исключает требование нулевого генетического разнообразия для ФП. Благодаря феномену разнонаправленности ФП появилась возможность оценить генетическую собственную дисперсию, которая не мешает идентификации генетических сдвигов СП, идущих почти в ортогональном направлении (10, 11).

Применительно к отдельной особи явление разнонаправленности сдвигов количественных признаков формулируется следующим образом: существует возможность количественно оценить, в какой мере отклонение признака особи от популяционной средней определяется ее генотипом и в какой — факторами среды (оценка проводится экспрессно без смены поколений). Суть явления разнонаправленности заключается в том, что генетические и средовые отклонения количественных признаков индивидуальных организмов популяции становятся разнонаправленными при рассмотрении их в определенных двумерных системах координат, построенных на основании СП и ФП.

Теперь охарактеризуем ФП для случая, когда $\sigma_{g\text{ФП}}^2 > 0$, и идентификацию проводят с учетом разнонаправленности генетических и экологических сдвигов признаков.

Фоновый признак — это такой признак, который чувствителен к N-причинам изменчивости, вызывающим сдвиги СП, кроме одной генетической причины, которая влияет на СП, но к которой ФП либо не чувствителен, либо смещается в противоположном СП направлении.

Рассмотрим модель Фишера-Райта-Лаша, описывающую генетическую организацию количественного признака: $\varphi_i = \mu + \gamma_i + \pi_i$, где φ_i — фенотипическое значение признака у i -й особи популяции, μ — средняя величина признака всех особей популяции, γ_i — генотипическое отклонение от популяционной средней, π_i — средовое отклонение от популяционной средней.

Разнонаправленность сдвигов количественного признака индивидуального организма под влиянием генетических и средовых факторов свидетельствует о более сложной структуре организации количественного признака у растений. Это отражено в следующей модели Драгавцева-Дьякова: $\varphi_i = \mu + \gamma_{atr\ i} + \gamma_{ad\ i} + \gamma_{imm\ i} + \gamma_{mic\ i} + \gamma_{ef\ i} + \gamma_{tol\ i} + \gamma_{ont\ i} + \pi_i + \varepsilon_{com\ i} +$

$+\gamma_{com\ i} + \pi_{ont\ i}$, где γ — генетические, π и ε — экологические отклонения признака, которые определяются следующими системами: γ_{attr} — «всасывания» (аттракции) продуктов фотосинтеза из стеблей и листьев в колос (корзину); γ_{mic} — аттрагированной пластики между зернами и мякиной в колосе (между ядром и лузгой у подсолнечника); γ_{ad} — адаптивности (общей адаптивности к конкретным полю и зоне испытания или адаптивности в случае организации провокационного фона — засуха, холод, жара, засоление, рН почвы и т.д.); γ_{imm} — горизонтальной устойчивости (горизонтальный иммунитет); γ_{ef} — «оплата» лимитирующего фактора почвенного питания (N, P, K и т.п.); γ_{tol} — толерантности к загущению; γ_{ont} — генетической вариабельности продолжительности фаз онтогенеза.

Вышеперечисленные генотипические параметры при их использовании в селекционной работе способствуют генетическому повышению урожая. Какими именно «рычагами» будет пользоваться селекционер — это зависит от типичной динамики лимитирующих факторов среды в конкретной зоне селекции и наличия достаточного генетического полиморфизма по всем или некоторым из перечисленных семи физиолого-генетических систем.

Далее, γ_{com} относится к «шумам» (как и π_i , ε_{com} , π_{ont}), поэтому должно быть элиминировано при идентификации генотипа, хотя является генотипической компонентой. Рассмотрим эти «шумы», мешающие идентификации генотипа конкретной особи по ее фенотипу: π_i — средовое отклонение, вызываемое вариациями индивидуальных микроэкологических ниш; ε_{com} — отклонение негенетической (стартовой) конкуренции, возникающей за счет неодинакового стартового роста проростков одного генотипа (разные глубина заделки семян и ориентация последних в почве); γ_{com} — отклонение генетической конкуренции, вызываемое разной генетической агрессивностью растений в борьбе за влагу, азот, свет и т.д.; π_{ont} — отклонение, вызванное сменой лимитирующих факторов в онтогенезе в период закладки и развития признака.

Теперь рассмотрим положения теории регрессионного анализа, так как этот математический инструмент поможет дать ответ на интересующий нас вопрос теории ошибок линий регрессии (12).

Корреляционную зависимость между признаками можно описывать различными способами. В частности, любая форма связи может быть выражена уравнением общего вида $y = f(x)$, где признак y — зависимая переменная, или функция от независимой переменной x . Известно, что изменение функции в зависимости от изменений одного или нескольких аргументов называется регрессией. Для выражения регрессии служат эмпирические и теоретические ряды, их графики (линии регрессии), а также корреляционные уравнения (уравнения регрессии) и коэффициент линейной регрессии.

Отметим, что показатели регрессии отражают корреляционную связь двусторонне (исключение — односторонняя связь рядов динамики), учитывая изменение среднего показателя \bar{y}_x признака Y при изменении значений x_i признака X , и наоборот, и показывают изменение среднего показателя \bar{x}_y признака X по измененным значениям y_i признака Y . В общем случае признак Y рассматривается как функция многих аргументов, однако на практике учитываются не все, в простейшем случае (как при описании линейной регрессии) всего один: $y = a + bx$, где a — свободный член, в системе координат Oxy (графически он представляет отрезок ор-

динаты), b — коэффициент регрессии.

Коэффициент регрессии можно вычислить, минуя расчет средних квадратических отклонений и коэффициента корреляции, по формулам

$$b_{y_x} = \frac{(y_i - \bar{y})(x_i - \bar{x})}{(x_i - \bar{x})^2}; \quad b_{x_y} = \frac{\Sigma(y_i - \bar{y})(x_i - \bar{x})}{\Sigma(y_i - \bar{y})^2}.$$

Если говорить о связи между коэффициентами регрессии и корреляции, то известно, что коэффициент корреляции равен средней геометрической коэффициентов регрессии — $r = \sqrt{b_{y_x} b_{x_y}}$. Заметим, как и коэффициент корреляции, коэффициент регрессии характеризует только линейную связь и сопровождается знаком (+) при положительной и знаком (–) при отрицательной связи.

Следует отметить, что параметры линейной регрессии можно определить способом наименьших квадратов, основанным на требовании, чтобы сумма квадратов отклонений вариант от линии регрессии была наименьшей, то есть этому условию удовлетворяет следующая система нормальных уравнений: $an + b\Sigma x = \Sigma y$; $a\Sigma x + b\Sigma x^2 = \Sigma xy$.

Эмпирические показатели регрессии служат оценками соответствующих генеральных параметров. При этом возникают статические ошибки, так как значения этих показателей являются случайными. Из теории математической статистики известно, что ошибка выборочного коэффициента регрессии выражается в виде следующих формул:

$$S_{b_{y_x}} = \sqrt{\frac{(y_i - \bar{y})^2 - \frac{[\Sigma(y_i - \bar{y})(x_i - \bar{x})]^2}{\Sigma(x_i - \bar{x})^2}}{(n - 2)\Sigma(x_i - \bar{x})^2}};$$

$$S_{b_{x_y}} = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 - \frac{[\Sigma(y_i - \bar{y})(x_i - \bar{x})]^2}{\Sigma(y_i - \bar{y})^2}}{(n - 2)\Sigma(y_i - \bar{y})^2}},$$

где $S_{b_{y_x}}$ — ошибка коэффициента регрессии Y по X ; $S_{b_{x_y}}$ — ошибка коэффициента регрессии X по Y ; \bar{x} и \bar{y} — средние арифметические суммы всех вариант выборки.

Если мы знаем коэффициент корреляции между признаками X и Y (r_{xy}) и средние квадратические отклонения S_x и S_y , то ошибку коэффициента регрессии можно определить по формулам

$$S_{b_{y_x}} = \frac{S_{y_i y}}{S_{x_i x}} \sqrt{\frac{1 - r^2}{n - 2}}; \quad S_{b_{x_y}} = \frac{S_{x_i x}}{S_{y_i y}} \sqrt{\frac{1 - r^2}{n - 2}},$$

где i — порядковый номер варианты.

Достоверность выборочного коэффициента регрессии b проверяется с помощью t -критерия с числом степеней свободы $k = n - 2$ и принятым уровнем значимости a . Нулевая гипотеза сводится к предположению, что в генеральной совокупности коэффициент регрессии равен нулю.

Статистические ошибки регрессии, которые возникают при решении эмпирических уравнений линейной регрессии, S_{yx} и S_{xy} , вычисляются по формулам $S_{yx} = S_y \sqrt{1 - r^2}$ и $S_{xy} = S_x \sqrt{1 - r^2}$, а также

$$S_{yx} = \sqrt{\frac{(y_i - \bar{y})^2}{n - 2}}; \quad S_{xy} = \sqrt{\frac{(x_i - \bar{x}_y)^2}{n - 2}},$$

где \bar{y}_x и \bar{x}_y — средние значения зависимых переменных соответственно Y и X , рассчитанные по уравнениям регрессии.

Важность определения ошибки линий регрессии обусловлена тем, что она отражает величину возможных отклонений эмпирически найденных значений переменных Y и X от линий регрессии, построенной по значениям \bar{y}_x или \bar{x}_y . Статистическая ошибка линейной регрессии позволяет с той или иной вероятностью устанавливать границу, содержащую определенную часть всех эмпирически найденных значений переменных величин. Для этого используются следующие уравнения: $\bar{y}_x = (a + b_{y_x}) \pm tS_{yx}$; $\bar{x}_y = (a + b_{x_y}) \pm tS_{xy}$. Если оценке подвергают не отдельные индивиды, а целые выборочные группы с их средними характеристиками, границы доверительного интервала линейной регрессии устанавливаются по уравнению

$$\bar{y}_x = (a + b_{y_x}) \pm \frac{tS}{\sqrt{n}}, \text{ где } n \text{ — численность выборки.}$$

Обращаем внимание на то, что применение метода линейной регрессии к анализу оценки значений переменных Y и X применимо лишь в тех ситуациях, когда исходные данные распределяются нормально или их распределение не очень сильно отклоняется от нормального закона. Это позволяет нам использовать эту теорию в качестве теории ошибок идентификации генотипов растений по их фенотипам. Если окажется, что генеральная совокупность, из которой взята выборка, распределяется явно асимметрично, что бывает крайне редко, то этот подход не сработает и должен быть заменен соответствующим непараметрическим методом.

Графики разнонаправленности, подробно описанные Драгавцевым (13), могут быть использованы для решения двух задач: идентификации генотипов отдельных растений в расщепляющих популяциях по их фенотипам, когда на графики наносят показатели СП и ФП отдельных растений; подбора родительских пар при гибридизации, когда на графики наносят средние показатели СП и ФП сортов из рабочей (стартовой) коллекции селекционера. В первом случае теория ошибок идентификации рассматривается как теория ошибок линии регрессии, выраженной зоной ошибок, ограниченной двумя параболой, идущими вдоль линий регрессии. Точки, выходящие за пределы параболы, достоверно являются надежными генетическими сдвигами по одной из физиологических систем (аттракции, микрораспределений, адаптивности, полигенного иммунитета, «оплаты» лимитирующего фактора почвенного питания, толерантности к загущению и генетической вариабельности продолжительности фаз онтогенеза) в зависимости от СП и ФП, выбранных для построения графика в двумерных координатах. Во втором случае, когда на графике СП—ФП нанесены средние показатели признаков сортов, очень хорошо работает из-

вестная формула достоверности различий средних величин:

$$D = \bar{x}_1 - \bar{x}_2; m_D = \pm \sqrt{m_1^2 + m_2^2}; t = \frac{D}{m_D} > 3,$$

где D — разность двух средних величин (\bar{x}_1 и \bar{x}_2), m_D — ошибка разности

средних, m_1 и m_2 — ошибки средних арифметических, где $m = \pm \frac{\sigma}{\sqrt{n}}$ (σ —

среднее квадратическое отклонение, n — число растений при расчете среднего показателя) (14).

Рассмотрим эффективность применения принципа фоновых признаков в решении задач идентификации в F_2 (или M_2) уникального растения — родоначальника будущего нового сорта. По Allard, в среднем для количественных признаков растений около 80 % изменчивости — экологическая и лишь ≈ 20 % — генотипическая (15). Как видим, превышение лучшего генотипа в лучшей нише над худшим генотипом в лучшей нише составит примерно 20 %. Превышение же худшего генотипа в лучшей нише над лучшим генотипом в худшей равно $80\% - 20\% = 60\%$. Ясно, что в отбор по фенотипам попадут в основном генетически средние образцы, а генетически уникальное растение будет утеряно.

Для безошибочного отбора по фенотипу лучших генотипов необходимо, чтобы часть из них попала в лучшие микроэкологические ниши. Вероятность попадания любого лучшего генотипа, лежащего за пределами $+2\sigma$ генотипического распределения, то есть любого из 2,5 % лучших генотипов в любую нишу, составит $p_1 = 2,5:100 = 0,025$. Вероятность встречи в любой точке делянки лучшего генотипа с лучшей (одной из 2,5 %) микронишей составит $p_1 = 0,025 \cdot 0,025 = 0,000625$ (16, 17). Отсюда видим, что только 6 из 10000 растений с лучшими фенотипами будут генетически лучшими, то есть важными для эффективной селекции.

Отобрав лучшее по фенотипу растение, мы отберем лучший уникальный генотип с вероятностью 6:10000. Поскольку вероятность попадания лучшего генотипа в среднюю микронишу составляет $0,95 \cdot 0,025 \approx 0,023$ (в 38,3 раза выше, чем в лучшую нишу), то при «вылавливании» уникального генотипа случайный отбор часто может быть эффективнее отбора по «плюсовым» фенотипам.

Этот пример с генетической точки зрения рассматривает количественный признак, определяемый примерно тремя локусами, так как в F_2 одна полная гомозигота по доминантным положительным аллелям трехлокусного признака встречается один раз среди 64 растений. Среди 128 растений будет две доминантно полных гомозиготы, что составляет около 2 % от всех особей популяции (в нашем примере мы хотим отобрать 2,5 % особей, лежащих за пределами $+2\sigma_e$).

Если система идентификации генотипов, основанная на принципе фоновых признаков работает идеально (а она работает так при наличии идеального ФП и всегда при ортогональной идентификации), то ее эффективность можно оценить следующим образом:

$$E_{\text{ФП}} = \frac{10000_{\text{раст.}}}{6_{\text{раст.}}} \approx 1000,$$

то есть она работает эффективнее интуитивно-глазомерной в 1000 раз при оценке каждого поколения популяции.

Система фоновых признаков может быть наложена не только на F_2

(или M_2), но и на F_3 , F_4 и т.д., и тогда ее общая эффективность возрастает еще в 2 или 3 раза.

Заметим, что эта эффективность рассчитана для трехлокусного признака. Однако генотипические различия большинства количественных признаков определяются 6-20 и более локусами. Для 6 локусов одна гомозигота должна встречаться на 4096 растений (на 64 в трехлокусном случае), то есть мы должны отобрать не 2,5 % особей, а всего лишь 0,024 %. При этом эффективность работы принципа ФП возрастает еще в 2,5:0,024 \approx 100 раз.

Л И Т Е Р А Т У Р А

1. Драгавцев В.А. Метод оценки роли наследственности и среды в развитии признаков древесных растений, не требующей смены поколений. Бот. журн., 1966, 7: 939-946.
2. Драгавцев В.А. О возможности элиминации межиндивидуальной средовой компоненты дисперсии при оценке коэффициента повторяемости у растений. Генетика, 1969, 5, 2: 30-35.
3. Кренке Н.П. Феногенетическая изменчивость. М., 1933-1935, 1: 167.
4. Плохинский Н.А. Наследуемость. Новосибирск. 1964: 5.
5. Никоро З.С., Харитонов З.Н., Решетникова Н.Ф. Различные способы определения племенной ценности животных. В кн.: Теоретические основы селекции животных. М., 1968: 300.
6. Уильямс У. Генетические основы и селекция растений. М., 1968: 350.
7. Рокитский П.Ф. Введение в статическую генетику. Минск, 1974: 200.
8. Савченко В.К. Генетический анализ в сетевых пробных скрещиваниях. Минск, 1984: 19.
9. Ауэрбах Ш. Генетика. М., 1966.
10. Дьяков А.Б., Драгавцев В.А., Бехтер А.Г. Конкуренциоспособность растений в связи с селекцией. Новый принцип анализа дисперсии продуктивности. В сб.: Теория отбора в популяциях растений. Новосибирск, 1976: 237-251.
11. Драгавцев В.А., Дьяков А.Б. Разнонаправленность сдвигов количественного признака индивидуального организма под влиянием генетических и средовых причин в двумерных системах признаковых координат. В кн.: Эколого-генетический скрининг генофонда и методы конструирования сортов сельскохозяйственных растений по урожайности, устойчивости и качеству. СПб, 1998.
12. Лакин Г.Ф. Биометрия. М., 1980: 181-200.
13. Драгавцев В.А. К проблеме генетического анализа полигенных количественных признаков растений. СПб, 2003.
14. Федоров А.И. Методы математической статистики в биологии и в опытном деле. Алма-Ата, 1967.
15. Allard R.W. Principles of plant breeding. N-Y-London, 1960.
16. Гнеденко Б.В. Курс теории вероятностей. М., 1954.
17. Феллер В. Введение в теорию вероятностей и ее приложения. М., 1967, 1-2.

ГНЦ Всероссийский НИИ растениеводства
им. Н.И. Вавилова, 190000, С.-Петербург,
ул. Большая Морская, 44; e-mail: kns@lsi.ru

Поступила в редакцию
19 июня 2006 года

THEORY OF ERRORS OF GENOTYPES IDENTIFICATION OF INDIVIDUAL PLANTS ON THEIR PHENOTYPES ON QUANTITATIVE DETERMINANTS IN SEGREGATED POPULATIONS AT THE EARLY STAGES OF SELECTION

N.V. Kocherina

S u m m a r y

The author considers the problem of selection identification of the best plant genotypes on their phenotypes. The basis of background determinants principle for productive selective work was presented. The theory of errors of regression line with the purpose of use as the theory of errors of identification of plant genotypes on their phenotypes was analyzed. The possibility of use of reliability of average for picking out of parent pairs during hybridization was characterized. The efficiency of background determinants principle was estimated for identification in F_2 unique plant — ancestor of future variety.