

# Метаболомный подход к оценке сортовой специфичности семян *Brassica napus* L.

Г.Н. Смоликова<sup>1</sup>, А.А. Шаварда<sup>1,2</sup>, И.В. Алексейчук<sup>1</sup>, В.В. Чанцева<sup>1</sup>, С.С. Медведев<sup>1</sup>

<sup>1</sup> Санкт-Петербургский государственный университет, Санкт-Петербург, Россия; <sup>2</sup> Ботанический институт им. В.Л. Комарова Российской академии наук, Санкт-Петербург, Россия

Современные молекулярно-биологические исследования характеризуются применением совокупности методологических платформ, позволяющих изучать организмы как на геномном и протеомном уровнях, так и на уровне метаболомов. Однако для обоснования возможности применения метаболомного анализа в селекции необходимо проведение исследований на широком спектре видов и сортов растений. В данной работе с использованием метода газовой хроматографии, сопряженной с масс-спектрометрией, проведена оценка содержания низкомолекулярных метаболитов в семенах разных сортов рапса, которые относились к одной репродукции. В каждом метаболомном профиле было аннотировано по 168 соединений, из которых 52 идентифицировано. Идентифицированные соединения включали аминокислоты, органические и жирные кислоты, токоферолы, фитостеролы. Обработка полученных данных осуществлялась методами мультивариантной статистики: методом главных компонент (МГК), методом дискриминантного анализа проекций на латентные структуры (ПЛС-ДА) и методом множественного регрессионного анализа проекций на латентные структуры. Созданные МГК и ПЛС-ДА модели демонстрировали достоверные различия между метаболомами исследованных сортов рапса. Наиболее значимый вклад в формирование моделей вносили аминокислоты и органические кислоты. Суммарный процент объясненной информации для МГК и ПЛС-ДА моделей составил в среднем 65%. Достоверность ПЛС-ДА модели, согласно перекрестной проверке по методу «венецианские жалюзи», составила 91,67%. Таким образом, показано, что метаболомный подход может служить эффективным инструментом для идентификации сортовой принадлежности семян. Необходимым условием при этом является создание постоянно обновляемой базы метаболомных профилей, характерных для конкретных сортов. Применение дискриминантного анализа проекций на латентные структуры позволит сравнивать метаболомы неизвестных образцов семян с имеющимися в базе данных метаболомными профилями и на этой основе классифицировать новые образцы семян.

Ключевые слова: *Brassica napus* L., семена, газовая хроматография, масс-спектрометрия, метаболомика, метод главных компонент, дискриминантный анализ проекций на латентные структуры.

## КАК ЦИТИРОВАТЬ ЭТУ СТАТЬЮ?

Смоликова Г.Н., Шаварда А.А., Алексейчук И.В., Чанцева В.В., Медведев С.С. Метаболомный подход к оценке сортовой специфичности семян *Brassica napus* L. Вавиловский журнал генетики и селекции. 2015;19(1):121-127. DOI 10.18699/VJ15.015

## HOW TO CITE THIS ARTICLE?

Smolikova G.N., Shavarda A.A., Alekseychuk I.V., Chantseva V.V., Medvedev S.S. The metabolomic approach to the assessment of cultivar specificity of *Brassica napus* L. seeds. *Vavilovskii Zhurnal Genetiki i Seleksii* – Vavilov Journal of Genetics and Breeding. 2015;19(1):121-127. DOI 10.18699/VJ15.015

DOI 10.18699/VJ15.015  
УДК 577.121:631.527.11:631.53.011.2  
Поступила в редакцию 01.12.2014 г.  
Принята к публикации 27.02.2015 г.  
© АВТОРЫ, 2015

 e-mail: galina.smolikova@gmail.com

## The metabolomic approach to the assessment of cultivar specificity of *Brassica napus* L. seeds

G.N. Smolikova<sup>1</sup>, A.L. Shavarda<sup>1,2</sup>,  
I.V. Alekseychuk<sup>1</sup>, V.V. Chantseva<sup>1</sup>, S.S. Medvedev<sup>1</sup>

<sup>1</sup> Saint-Petersburg State University, St.-Petersburg, Russia;  
<sup>2</sup> V.L. Komarov Botanical Institute of the RAS, St.-Petersburg, Russia

Recent biomolecular studies tend to involve combinations of different methods and approaches that allow analyzing organisms on the genomic and proteomic levels, as well as on the level of metabolomics. However, in order to justify the use of the metabolomics techniques in plant breeding, it is important to perform comprehensive analysis of a broad range of species and varieties. In this study, we evaluated the contents of low-molecular-weight substances in seeds of different rapeseed cultivars by the gas chromatography–mass spectrometry (GC-MS) technique. For every metabolomic profile, we estimated 168 target substances, and 52 of them were unambiguously identified. These compounds included amino acids, organic and fatty acids, tocopherols, and phytosterols. In order to keep the data assay within the context of multivariate statistics, we used principal component analysis (PCA), partial least square discriminant analysis (PLS-DA), and partial least square regression (PLS-R). Subsequent analysis revealed a significant difference between the metabolomic profiles of the investigated rapeseed cultivars, with the primary role of the amino acids and organic acids. Noticeably, the PLS-DA model showed 65% of the explained variance and, according to the Venetian blinds cross-validation test, 91.67% of the accuracy. Thus, we demonstrate the effectiveness of the metabolomics approach to the varietal identification of seeds. This strategy can be further improved with a continuously updated database of the metabolomic profiles of different species and cultivars. Application of the PLS-DA method will allow comparison of the metabolites of unknown samples with the existing profiles and, subsequently, identification of new seed samples.

Key words: *Brassica napus* L., seeds, gas chromatography, mass spectrometry, metabolomics, principal component analysis, projection to latent structures – discriminant analysis.

Термин «метабономика» впервые был предложен Оливером в 1998 г. как метод функциональной геномики, позволяющий проследить поток информации от гена к функции (Oliver et al., 1998). **Метаболом** представляет собой совокупность низкомолекулярных метаболитов, содержание которых, с одной стороны, обусловлено генетически, а с другой стороны, зависит от адаптации организма к внешним факторам. Таким образом, метаболом можно рассматривать как биохимическую реализацию генотипа. Современные молекулярно-биологические исследования характеризуются применением различных методологий, позволяющих изучать растительные организмы не только на геномном и протеомном уровнях, но также и на уровне метаболизма (Hollywood et al., 2006; Shulaev, 2006; Harrigan et al., 2007). **Только совместное применение** этих подходов позволяет комплексно подойти к изучению процессов, протекающих в живых системах, и является основой системной биологии (Афонников, Миронова, 2014).

В последние годы метаболомный подход начинает активно использоваться при создании новых и оценке уже существующих сортов сельскохозяйственно важных растений. При этом используются такие методы разделения веществ, как газовая хроматография, высокоэффективная жидкостная хроматография и капиллярный электрофорез. Для идентификации метаболитов используются масс-спектрометрия и ядерный магнитный резонанс. Сравнительный анализ метаболомных профилей был проведен для семян разных сортов кукурузы (Röhlig et al., 2009), клецвины (Pigott et al., 2011), нигеллы (Farag et al., 2014). Однако для обоснования возможности применения метаболомного анализа в селекции необходимо проведение подобных исследований на гораздо более широком спектре видов и сортов. При этом требуется также отработка методологии их сравнительной оценки с использованием мультивариантной статистики.

В данной работе проведен метаболомный анализ семян разных сортов рапса, основанный на газовой хроматографии, сопряженной с масс-спектрометрией. Статистическая обработка полученных данных осуществлялась с использованием метода главных компонент, дискриминантного анализа проекций на латентные структуры и множественного регрессионного анализа проекций на латентные структуры.

## Материалы и методы

Объектами исследования являлись семена рапса (*Brassica napus* L.) разных сортов из коллекции Всероссийского института растениеводства им. Н.И. Вавилова: Юбилейный (к-5285), Sielecki (к-4371), Русская Кудья (к-336) и Ордеж-4 (к-5273). Семена всех сортов относились к одной репродукции – Тамбовская область, урожай 2012 г.

## Оценка всхожести семян и развития проростков

Семена проращивали в чашках Петри на фильтровальной бумаге в термостате при 2 °С в течение 7 сут. (по 50 семян в 4 повторностях). Всхожесть учитывали согласно ГОСТ 12038-84. На 7-е сутки прорастания подсчитывали количество нормально развитых проростков, ненормально развитых проростков и непроросших семян.

## Проведение метаболомного анализа

Экстракцию метаболитов проводили в течение 7 дней в метаноле в герметически закрытых микропробирках типа эппендорф объемом 1,5 мл (SSI, США). Перед экстракцией семена взвешивали в количестве 10 семян на пробу, помещали в эппендорфы, растирали с добавлением небольшого количества метанола и далее доводили объем метанола до 1 мл. После экстракции эппендорфы центрифугировали при 3 000 g и переносили супернатант в виалы объемом 2 мл. Супернатант в виалах выпаривали до состояния тонкой пленки в вакуумном ротационном испарителе Eppendorf Concentrator Plus («Eppendorf», Германия).

Дериватизацию проводили методом исчерпывающего силилирования. Для этого выпаренные под вакуумом сухие метаболиты растворяли в пиридине, содержащем трикозан (nC<sub>23</sub>) в качестве внутреннего стандарта. Силилирование проводили при помощи бис-триметилсилилтрифторацетамида, содержащего 1 % триметилхлоросилана (BSTFA+TMCS, «Supelco», США) в сухом термоблоке при 100 °С в течение 15 мин.

## Газовая хроматография, сопряженная с масс-спектрометрией (ГХ-МС)

Определение концентраций метаболитов проводили методом ГХ-МС на газовом хроматографе Agilent 6850 GC, оснащенный квадрупольным масс-селективным детектором Agilent 5975B VL MSD («Agilent Technologies», США). Использовали неполярную капиллярную колонку DB-5HT (5 % фенилметилсилоксан, длина 30 м, внутренний диаметр 250 мкм, толщина пленки 0,1 мкм, J&W Scientific, США). **Пробу вводили** с помощью программируемого автоматического пробоотборника Agilent G4513A в режиме инжекции с делением потока 20 : 1 (Split); температура испарителя – 330 °С, давление – 60,3 кПа, скорость потока газа-носителя через испаритель – 23,9 мл/мин. Ионизацию веществ в камере масс-спектрометра осуществляли посредством электронного удара (70 эВ) при температуре ионизационной камеры 230 °С и температуре квадрупольа 150 °С.

## Идентификация метаболитов

Индексы удерживания веществ рассчитывали с помощью программы AMDIS (G. Mallard, «NIH») на основании данных по времени выхода нормальных углеводов

Посевные качества семян и развитие семидневных проростков разных сортов рапса

Названия сортов	Масса 1000 семян, г	Всхожесть семян, %	% от всхожести	
			нормально развитые проростки	ненормально развитые проростки
Юбилейный	2,74 ± 0,04	98 ± 2	90 ± 3	8 ± 4
Sielecki	2,59 ± 0,17	96 ± 2	92 ± 2	4 ± 2
Русская Кудья	2,65 ± 0,06	94 ± 3	82 ± 2	12 ± 3
Оредеж-4	2,66 ± 0,12	92 ± 2	73 ± 5	19 ± 6

C10–C35. Вещества идентифицировали с использованием баз данных NIST/EPA/NIH 11 Mass Spectral Library (<http://www.nist.gov/srd>) и пользовательской библиотеки AMDIS, созданной на кафедре физиологии и биохимии растений СПбГУ. Разметку пиков хроматограмм выполняли в программе UniChrom 5.09.1034 ([www.unichrom.com](http://www.unichrom.com)). Всего было размечено 168 пиков, из них идентифицировано 52. Неидентифицированные метаболиты размечали по индексу удерживания. Все 168 метаболитов были использованы для создания метаболомного профиля, представленного в виде матрицы содержания веществ (в мкг/г сырой массы).

Расчет содержания метаболитов проводили с использованием внутреннего стандарта ( $nC_{23}$ ) без учета коэффициентов чувствительности.

**Статистическая обработка данных и программное обеспечение**

Все измерения проводили в 3 биологических повторностях. Статистическую обработку показателей всхожести семян проводили в MS Microsoft Excel 2007 с использованием стандартного пакета анализа данных. На графиках и в таблицах представлены средние арифметические значения величин и ошибка среднего арифметического.

Матрица данных для анализа основывалась на концентрациях метаболитов, которые предварительно были приведены к единичной дисперсии с помощью шкалирования по стандартным отклонениям.

Метаболомный профиль оценивался с использованием метода главных компонент (МГК), дискриминантного анализа проекций на латентные структуры (ПЛС-ДА, Projection to Latent Structures–Discriminant Analysis) и множественного регрессионного анализа проекций на латентные структуры (ПЛС-R).

Обработка данных метаболомного профиля осуществлялась в среде MS Microsoft Excel 2007 с использованием макроса Multibase (<http://numericaldynamics.com>) и в среде Matlab 8.14 (MathWorks) с использованием статистического пакета.

Достоверность модели ПЛС-ДА оценивалась согласно перекрестной проверке (Cross-Validation Test) по методу «венецианские жалюзи» (Ballabio, Consonni, 2013). Данный метод позволяет в автоматическом режиме рассчитать множество моделей ПЛС-ДА по аналогии с исходной, но без использования части данных. Практически это выражалось в том, что в каждой модели не было включено по одной повторности каждого сорта. Неиспользованная повторность являлась контролем, который далее был классифицирован на основании созданной модели. Точность классификаций была усреднена по всем моделям, общее число которых было максимизировано для имеющегося объема данных. Среднее значение точности классификаций характеризует способность ПЛС-ДА модели достоверно идентифицировать неизвестный метаболомный профиль по заданным параметрам.

Выбор сортов рапса для метаболомного анализа преследовал цель минимизировать вклад качества семян в содержание низкомолекулярных метаболитов с тем, чтобы акцентироваться только на сортовой специфичности. В связи с этим были отобраны сорта рапса одной репродукции, семена которых незначительно различались по всхожести. Как видно из таблицы, масса 1 000 семян у всех исследованных сортов находилась в пределах 2,6–2,7 г, а общее количество проросших семян варьировало от 92 до 98 %.

**Результаты и обсуждение**

**Оценка всхожести семян и развития проростков**

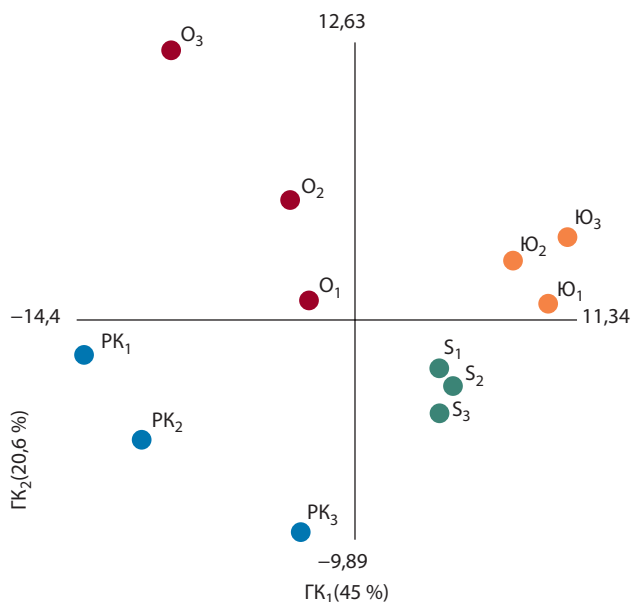
Однако более подробный анализ позволил установить, что отобранные сорта можно разделить на две группы. В первую группу были включены сорта Юбилейный и Sielecki, всхожесть семян у которых была равна 98 и 96 %, а количество нормально развитых проростков составляло 90 и 92 % соответственно. Во вторую группу были включены сорта Русская Кудья и Оредеж-4, всхожесть семян у которых была равна 94 и 92 %, а количество нормально развитых проростков составляло 82 и 73 % соответственно.

ГХ-МС анализ содержания низкомолекулярных метаболитов был проведен в экстрактах, полученных из целых семян рапса (Дополнительные материалы 1<sup>1</sup>).

**ГХ-МС анализ содержания низкомолекулярных метаболитов**

В каждой хроматограмме было размечено по 168 компонентов, из которых по 52 компонента было идентифицировано.

<sup>1</sup> Дополнительные материалы см. в Приложении 3 по адресу: <http://www.bionet.nsc.ru/vogis/download/pict-2015-04/appx3.pdf>



**Рис. 1.** Распределение семян разных сортов рапса на плоскости в координатах 1-й и 2-й главной компонент (ГК<sub>1</sub> и ГК<sub>2</sub> соответственно), рассчитанное по методу главных компонент.

PK<sub>1</sub>–PK<sub>3</sub> – сорт Русская Кудья; O<sub>1</sub>–O<sub>3</sub> – сорт Оредеж-4; S<sub>1</sub>–S<sub>3</sub> – сорт Sielecki; Ю<sub>1</sub>–Ю<sub>3</sub> – сорт Юбилейный. Модель построена на основе анализа концентраций 168 метаболитов. ГК<sub>1</sub> – 45 % объясненной дисперсии; ГК<sub>2</sub> – 20,6 % объясненной дисперсии.

цировано. Идентифицированные соединения включали такие метаболиты, как аминокислоты, органические и жирные кислоты, токоферолы и фитостеролы. Неидентифицированные компоненты хроматограмм аннотировали с указанием их индекса удерживания. В эту группу были включены сахара, поскольку применяемая методология не позволяет осуществить их точную идентификацию. Исключение составила сахароза, присутствующая в семенах в больших количествах.

Для метаболомного анализа создавали матрицу, в которой в качестве наблюдений выступали варианты (по 3 повторности каждого сорта семян рапса), а в качестве параметров – названия аннотированных метаболитов. Таким образом, профиль включал 12 наблюдений, 168 параметров и 2 016 концентраций метаболитов. Далее полученный метаболомный профиль был проанализирован методом главных компонент и методами дискриминантного и регрессионного анализа проекций на латентные структуры.

Метод главных компонент (МГК) относится к алгоритмам уменьшения размерности данных, которые позволяют отобразить многомерные данные на 2–3-мерной поверхности и тем самым делают их доступными для восприятия (Pearson, 1901). Метод заключается в том, что для каждого наблюдения все параметры располагают в Гильбертовом пространстве таким образом, чтобы параметры стали координатами, а каждое наблюдение

было представлено единственной точкой в пространстве. В нашем эксперименте было создано многомерное пространство из 168 координат, на котором представлено 12 точек. При этом расположение каждой точки в системе координат зависело от концентраций всех содержащихся в образцах метаболитов. Далее прокладывалась первая главная компонента (ГК<sub>1</sub>) с таким условием, чтобы сумма квадратов расстояний от нее до всех точек была минимальна. Вторая главная компонента (ГК<sub>2</sub>) была отложена перпендикулярно первой с тем же условием минимизации расстояний. Таким образом, все данные нашей матрицы, состоящей из показателей концентраций 2 016 метаболитов, оказались отображены на 2-мерной плоскости (рис. 1).

Алгебраически модель основана на следующем уравнении:

$$X = T \times P^t + E = \sum_{u=1}^A t_u \times p_u^t + E,$$

где  $X$  – матрица преобразованных (нормированных и центрованных) метаболомных данных размерностью  $I \times J$ , каждая строка ( $I$ ) – наблюдение/образец и каждый столбец ( $J$ ) – параметр/метаболит;  $T$  – матрица счетов (координат в новом пространстве) размерностью  $I \times A$ ;  $P$  – матрица нагрузок (коэффициентов метаболитов) размерностью  $J \times A$ ;  $E$  – матрица остатков/ошибок;  $A$  – число рассчитываемых главных компонент;  $t$  – главные компоненты,  $p$  – параметры главных компонент.

Особенностью МГК является то, что он относится к методам анализа данных «без учителя»: возможные различия по степени важности между отдельными метаболитами не принимаются во внимание и все параметры учитываются в равной степени.

На основании проведенного анализа можно видеть, что точки, являющиеся отражением метаболомов семян, объединились в 4 класса, соответствующие 4 исследуемым сортам. На рис. 1 видно, что точки, соответствующие одному классу, не перекрываются и находятся в разных областях модели относительно 1-й и 2-й главных компонент. Это свидетельствует о достоверных различиях между сортами. Суммарный процент объясненной информации для МГК модели составил 65,6 % (45 % для ГК<sub>1</sub> и 20,6 % для ГК<sub>2</sub>).

Как было сказано выше (см. табл. 1), семена исследуемых сортов по посевным качествам были разделены на две группы: 1) сорт Юбилейный и сорт Sielecki (всхожесть – 98 и 96 % соответственно); 2) сорт Русская Кудья и сорт Оредеж-4 (всхожесть – 94 и 92 % соответственно). Осуществив проекцию точек, соответствующих вышеуказанным сортам, на горизонтальную ось (ГК<sub>1</sub>), можно видеть, что 1-я группа расположилась в положительной области оси, а 2-я группа – в отрицательной области. При этом проекции точек относительно горизонтальной оси разместились справа налево в следующем порядке: Русская Кудья и Оредеж-4 (94 и 92 %), Sielecki (96 %),



Юбилейный (98 %). Это позволяет нам высказать предположение, что ГК<sub>1</sub> является компонентой, отражающей динамику всхожести семян.

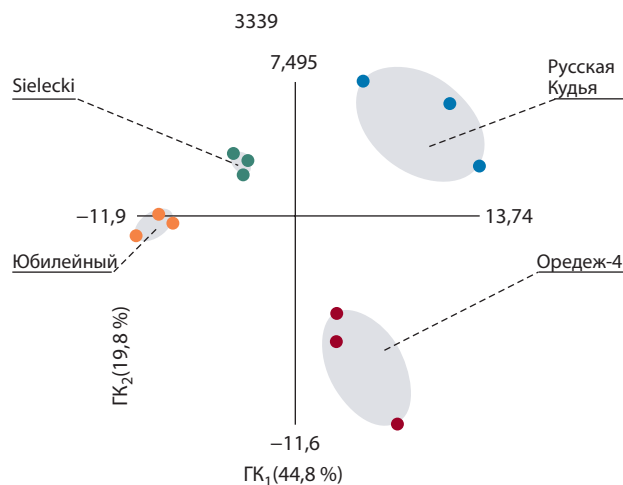
В Дополнительных материалах 2 показаны коэффициенты значимости идентифицированных метаболитов в формировании осей ГК<sub>1</sub> и ГК<sub>2</sub> и расположении точек описанной выше модели. Указанные коэффициенты могут иметь положительное или отрицательное значение в зависимости от их вклада в положительную или отрицательную области модели, при этом их значимость не зависит от знака и определяется только отклонением от нуля. Важно отметить, что положительная и отрицательная области пространства во многом условны и имеют, скорее, математическое, чем биологическое значение.

Так, например, можно видеть, что в расположение точек относительно положительной области ГК<sub>1</sub> наиболее значимый вклад внесли аланин, синаповая кислота, линолевая кислота, а также ряд органических кислот (фумаровая, малеиновая, яблочная, лимонная). Их коэффициент значимости составил 0,10. В расположение категорий в отрицательной области оси ГК<sub>1</sub> наиболее значимый вклад внесли мио-инозитол с коэффициентом значимости –0,11, а также этаноламин и фенилаланин с коэффициентами значимости –0,09. Как было сказано выше, ГК<sub>1</sub> можно характеризовать как компоненту, отражающую динамику всхожести семян. Поэтому можно ожидать, что содержание указанных соединений находится во взаимозависимости от всхожести семян.

Метод дискриминантного анализа проекций на латентные структуры (ПЛС-ДА), как и описанный выше метод главных компонент, является алгоритмом уменьшения размерности многомерных данных. Принципиальное отличие ПЛС-ДА от МГК состоит в том, что компоненты рассчитываются с учетом заранее определенных (дискриминированных) классов (Barker, Rayers, 2003). В связи с этим ПЛС-ДА относится к методам классификации данных «с учителем»: алгоритму заранее сообщается о принадлежности наблюдений к определенному классу; в результате теряется возможность «слепой» классификации данных, но появляется достоверность анализа значимости параметров относительно заданных классов. Модель ПЛС-ДА создается таким образом, чтобы разброс данных внутри каждого класса был минимальным, а разброс между классами – максимальным.

Математически ПЛС-ДА метод ищет оптимальное представление линейных отношений между независимыми переменными (в нашем случае концентрациями метаболитов) и зависимыми переменными (сортами). Модель основана на следующей системе уравнений:

$$\begin{cases} X = T \times P^t + E \\ Y = U \times Q^t + F \\ Cov(T, U) \rightarrow \max, \end{cases}$$



**Рис. 2.** Распределение семян рапса разных сортов на плоскости в координатах 1-й и 2-й главной компонент (ГК<sub>1</sub> и ГК<sub>2</sub> соответственно), рассчитанное по методу дискриминантного анализа проекций на латентные структуры.

Модель построена на основе анализа концентраций 168 метаболитов. ГК<sub>1</sub> – 44,8 % объясненной дисперсии; ГК<sub>2</sub> – 19,8 % объясненной дисперсии.

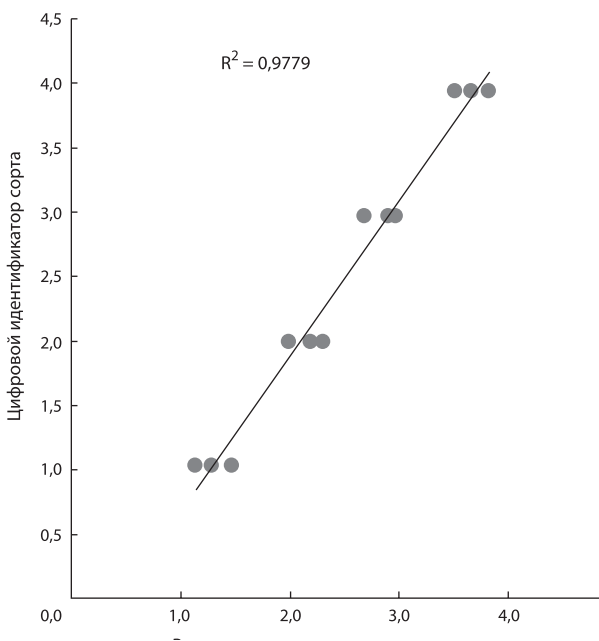
где  $X$  – матрица преобработанных (нормированных и центрированных) метаболомных данных размерностью  $I \times J$ , каждая строка ( $I$ ) – наблюдение/образец и каждый столбец ( $J$ ) – параметр/метаболит;  $Y$  – матрица ответов, описывающая принадлежность к классам размерностью  $I \times L$ ;  $T$  и  $U$  – матрица проекций  $X$  и  $Y$  соответственно размерностью  $I \times A$ ;  $P$  и  $Q$  – матрица нагрузок размерностью  $J \times A$  и  $L \times A$  соответственно;  $E$  и  $F$  – матрица остатков/ошибок. В ходе декомпозиции  $X$  и  $Y$  ковариация  $T$  и  $U$  максимизируется.

Для анализа данных методом ПЛС-ДА мы использовали тот же метаболомный профиль, что и для МГК анализа, но с одним отличием. Наблюдения предварительно были разделены на классы, соответствующие исследуемым сортам (Юбилейный, Sielecki, Русская Кудья, Оредеж-4). Таким образом, анализировались 4 класса, каждый из которых включал по 3 наблюдения (повторности).

Распределение классов на плоскости в координатах ГК<sub>1</sub> и ГК<sub>2</sub>, рассчитанное по методу ПЛС-ДА, представлено на рис. 2. Можно видеть, что так же, как и в МГК модели, классы не перекрывались и расположились в разных областях двумерной плоскости, что свидетельствует о достоверных различиях между ними. Суммарный процент объясненной информации практически не отличался от МГК модели и составил 64,6 % (44,8 % для ГК<sub>1</sub> и 19,8 % для ГК<sub>2</sub>).

Достоверность ПЛС-ДА модели, оцененная по методу перекрестной проверки, составила 91,67 %.

Осуществив проекцию классов на горизонтальную ось ГК<sub>1</sub>, можно видеть, что сорта разместились справа налево в следующем порядке: Юбилейный (98 %), Sielecki (96 %), Русская Кудья и Оредеж-4 (94 % и 92 %). То есть



**Рис. 3.** Множественная регрессионная модель зависимости сортовой принадлежности семян рапса от совокупности концентраций низкомолекулярных метаболитов. Цифровой идентификатор сорта: 1 – Юбилейный, 2 – Sielecki, 3 – Русская Кудья, 4 – Оредеж-4. Отношения реальных и рассчитанных значений представлены на графике. Модель построена на основе анализа концентраций 168 низкомолекулярных метаболитов, полученных методом ГХ-МС.  $R^2$  – коэффициент зависимости.

так же, как и в МГК модели,  $GK_1$  проявила себя как компонента, отражающая динамику всхожести семян.

Если МГК метод используется для выявления общих сходств или различий между объектами по всей совокупности их параметров, то ПЛС-ДА метод позволяет более эффективно выявлять соединения, которые связаны с конкретными различиями и тем самым могут являться маркерами дискриминированных классов (Jonsson et al., 2004). В Дополнительных материалах 3) приведены коэффициенты значимости идентифицированных соединений, на основании которых была построена ПЛС-ДА модель. Оказалось, что метаболиты, внесшие наибольший вклад в построение ПЛС-ДА модели, в значительной степени совпадали с результатами МГК модели, но с изменившимся знаком и более высокими коэффициентами значимости. Как уже было отмечено, знак коэффициента значимости метаболита в модели не имеет биологического значения, в то время как возросшее отклонение от нуля однозначно свидетельствует о более высокой чувствительности применяемого метода анализа.

В список наиболее значимых метаболитов можно включить аланин, синаповую, стеариновую, фумаровую кислоты (коэффициент значимости  $-0,11$ ), малеиновую,

лимонную и яблочную кислоты (коэффициент значимости  $-0,10$ ), мио-инозитол и этаноламин (коэффициенты значимости  $0,10$  и  $0,9$ ).

### Множественный регрессионный анализ

С целью обоснования возможности применения метаболомного анализа для оценки сортовой специфичности семян по алгоритму ПЛС дополнительно была рассчитана множественная регрессионная модель (рис. 3). Для этого каждому сорту был присвоен цифровой идентификатор от 1 до 4 и далее рассчитана множественная линейная зависимость между сортовой принадлежностью семян и содержащейся в них совокупностью концентраций низкомолекулярных метаболитов. Коэффициент детерминации реальных и рассчитанных значений сортовой принадлежности семян ( $R^2$ ) составил  $0,9779$ , что говорит о высокой степени линейной зависимости между анализируемыми факторами.

Таким образом, метаболомный подход может служить эффективным инструментом для идентификации сортовой принадлежности семян. Необходимым условием при этом является создание постоянно обновляемой базы метаболомных профилей, характерных для конкретных сортов семян.

Дальнейшее развитие данного подхода позволит использовать его для предсказания всхожести семян с неизвестной историей и оценки вклада метаболитов в формирование сортовых различий и всхожести.

Применение дискриминантного анализа проекций на латентные структуры позволяет сравнивать метаболомы неизвестных образцов семян с метаболомными профилями, существующими в базе данных, и на этой основе классифицировать новые образцы семян.

### Благодарности

Авторы выражают благодарность заведующей лабораторией масличных культур А.В. Гавриловой и куратору коллекции рапса А.Г. Дубовской (Всероссийский институт растениеводства им. Н.И. Вавилова) за предоставление семян рапса. Работа выполнена с использованием оборудования ресурсного центра СПбГУ «Развитие молекулярных и клеточных технологий» и поддержана грантами СПбГУ № 1.38.233.2014 и РФФИ № 14-04-01-624.

Авторы заявляют об отсутствии конфликта интересов.

### Список литературы

- Афонников Д.А., Миронова В.В. Системная биология. Вавиловский журнал генетики и селекции. 2014;18(1):175-192.  
Ballabio D., Consonni V. Classification tools in chemistry. Part 1: Linear models. PLS-DA. Analytical Methods. 2013;5:3790-3798. DOI: 10.1039/c3ay40582f  
Barker M., Rayers W. Partial Least Squares for Discrimination. J. Chemometrics. 2003;17:166-173. DOI: 10.1002/cem.785  
Farak M.A., Gad H.A., Heiss A.G., Wessjohann L.A. Metabolomics

- driven analysis of six *Nigella* species seeds via UPLC-qTOF-MS and GC-MS coupled to chemometrics. *Food Chem.* 2014;151: 333-342. DOI: 10.1016/j.foodchem.2013.11.032
- Harrigan G.G., Martino-Catt S., Glenn K.C. Metabolomics, metabolic diversity and genetic variation in crops. *Metabolomics.* 2007;3(3):259-272. DOI 10.1007/s11306-007-0076-0
- Hollywood K., Brison D.R., Goodacre R. **Metabolomics: current technologies and future trends.** *Proteomics.* 2006;6(17):4716-4723. DOI: 10.1002/pmic.200600106
- Jonsson P., Gullberg J., Nordstrom A., Kusano M., Kowalczyk M., Sjoström M., Moritz T. **A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS.** *Anal. Chem.* 2004;76:1738-1745. DOI: 10.1021/ac0352427
- Oliver S.G., Winson M.K., Kell D.B., Baganz R. **Systematic functional analysis of the yeast genome.** *Trends Biotechnol.* 1998;16: 373-378.
- Pearson K. **On lines and planes of closest fit to systems of points in space.** *Philos. Mag.* 1901;2:559-572.
- Pigott E.J., Roberts W., Ovenden S.P.B., Rochfort S., Bourne D.J. **Metabolomic investigations of *Ricinus communis* for cultivar and provenance determination.** *Metabolomics.* 2011. DOI: 10.1007/s11306-011-0355-7
- Röhlig R.M., Eder J., Engel K.-H. **Metabolite profiling of maize grain: differentiation due to genetics and environment.** *Metabolomics.* 2009;5(4):459-477. DOI: 10.1007/s11306-009-0171-5
- Shulaev V. **Metabolomics technology and bioinformatics.** *Briefings in bioinformatics.* 2006;7(2):128-139. DOI: 10.1093/bib/bbl012