

УДК 004.8:330.4

UDC 004.8:330.4

4.1.2. Селекция, семеноводство и биотехнология

4.1.2. Breeding, seed production and biotechnology

**ИДЕНТИФИКАЦИЯ ГЕНЕТИЧЕСКИХ
ДЕТЕРМИНАНТ ФЕНОТИПИЧЕСКИХ
ПРИЗНАКОВ ВИНОГРАДА С
ИСПОЛЬЗОВАНИЕМ МЕТОДОВ
ОБЪЯСНИМОГО ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА (XAI)**

**IDENTIFICATION OF GENETIC
DETERMINANTS OF GRAPE PHENOTYPIC
TRAITS USING EXPLAINABLE ARTIFICIAL
INTELLIGENCE (XAI) METHODS**

Луценко Евгений Вениаминович
д.э.н., к.т.н., профессор

Lutsenko Evgeniy Veniaminovich
Doctor of Economics, Candidate of Technical Sciences,
Professor

Web of Science ResearcherID S-8667-2018
Scopus Author ID: 57188763047
РИНЦ SPIN-код: 9523-7101
prof.lutsenko@gmail.com, <http://lc.kubagro.ru>
https://www.researchgate.net/profile/Eugene_Lutsenko
*Кубанский Государственный Аграрный университет
имени И.Т.Трубилина, Краснодар, Россия*

Web of Science ResearcherID S-8667-2018
Scopus Author ID: 57188763047
RSCI SPIN code: 9523-7101
prof.lutsenko@gmail.com, <http://lc.kubagro.ru>
https://www.researchgate.net/profile/Eugene_Lutsenko
*Kuban State Agrarian University named after I.T.
Trubilin, Krasnodar, Russia*

Трошин Леонид Петрович
д.б.н., профессор
*Кубанский Государственный Аграрный университет
имени И.Т.Трубилина, Краснодар, Россия*

Troshin Leonid Petrovich
Doctor of Biological Sciences, Professor
*Kuban State Agrarian University named after I.T.
Trubilin, Krasnodar, Russia*

Исследование посвящено актуальной проблеме маркер-опосредованной селекции (MAS) винограда – выявлению нелинейных связей между микросателлитными локусами (SSR-маркерами) и хозяйственно-ценными признаками. Цель работы – разработка прогностической модели фенотипа на основе генотипических данных в условиях малых выборок, характерных для сортоиспытания. В качестве методологической основы использован теоретико-информационный подход, реализованный в интеллектуальной системе «Эйдос», который, в отличие от традиционных статистических методов и «черных ящиков» нейросетей, относится к классу интерпретируемого искусственного интеллекта (Explainable AI, XAI). На выборке из 90 сортов винограда, охарактеризованных по 21 фенотипическому признаку и 50 аллелям, построены модели, позволяющие с высокой точностью (F-мера > 0,85) прогнозировать свойства сорта. Впервые количественно оценена информационная значимость конкретных аллелей для формирования таких признаков, как масса грозди, сахаристость и устойчивость к болезням. Предложен алгоритм «обратного конструирования» генотипа, позволяющий моделировать идеальные родительские пары для селекции. Результаты подтверждают

The study addresses the urgent problem of marker-assisted selection (MAS) in grapes: identifying non-linear relationships between microsatellite loci (SSR markers) and economically valuable traits. The aim is to develop a phenotype predictive model based on genotypic data under small sample conditions typical for variety testing. The methodological basis is the information-theoretic approach implemented in the "Eidos" intelligent system, which, unlike traditional statistical methods and neural network "black boxes," belongs to the class of Explainable AI (XAI). Using a sample of 90 grape varieties characterized by 21 phenotypic traits and 50 alleles, models were built that predict variety properties with high accuracy (F-measure > 0.85). For the first time, the informational significance of specific alleles for traits such as bunch weight, sugar content, and disease resistance was quantitatively assessed. An algorithm for genotype "reverse engineering" is proposed to model ideal parental pairs for breeding. The results confirm the effectiveness of cognitive modeling as a tool for digital breeding

эффективность применения когнитивного
моделирования как инструмента цифровой селекции

Ключевые слова: АСК-АНАЛИЗ, СИСТЕМА
«ЭЙДОС», ГЕНОМНАЯ СЕЛЕКЦИЯ, ВИНОГРАД,
SSR-МАРКЕРЫ, ИНТЕРПРЕТИРУЕМЫЙ
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ (ХАИ),
СИСТЕМНАЯ ТЕОРИЯ ИНФОРМАЦИИ,
ПРОГНОЗИРОВАНИЕ ФЕНОТИПА, МАЛЫЕ
ДАННЫЕ

Keywords: ASC-ANALYSIS, EIDOS SYSTEM,
GENOMIC SELECTION, GRAPES, SSR MARKERS,
EXPLAINABLE AI (XAI), SYSTEM INFORMATION
THEORY, PHENOTYPE PREDICTION, SMALL
DATA

<http://dx.doi.org/10.21515/1990-4665-214-009>

Введение

Современная селекция растений переживает цифровую трансформацию, переходя от фенотипического отбора к геномной селекции (Genomic Selection, GS). Ключевой задачей является установление связей между молекулярными маркерами и хозяйственно-ценными признаками. Однако в практической селекции винограда (*Vitis vinifera* L.) исследователи часто сталкиваются с проблемой «малых данных» (Small Data): количество доступных для анализа сортов с полным генетическим паспортом исчисляется десятками, тогда как размерность признакового пространства (количество аллелей и их комбинаций) может быть огромной.

Классические статистические методы (корреляционный, регрессионный анализ) в таких условиях часто оказываются неэффективными из-за мультиколлинеарности данных и невозможности выявить нелинейные зависимости. Современные методы глубокого обучения (Deep Learning), хотя и показывают высокую точность, требуют больших обучающих выборок и часто страдают от эффекта «черного ящика», не позволяя биологу интерпретировать, *почему* модель приняла то или иное решение [1, 2].

В связи с этим актуальным становится применение методов интерпретируемого искусственного интеллекта (Explainable AI, XAI), способных выявлять устойчивые закономерности на ограниченных наборах

данных и предоставлять результаты в понятной для селекционера форме. В данной работе предлагается подход, основанный на системной теории информации (СТИ) и автоматизированном системно-когнитивном анализе (АСК-анализ) [1], реализованном в программной среде «Эйдос». Данный подход позволяет количественно измерить силу и направление влияния генотипа на фенотип в битах информации, что открывает новые возможности для цифровизации селекционного процесса.

Материалы и методы

Объект исследования. В качестве эмпирической базы использованы данные по 90 сортам винограда, выращиваемым в коллекции Кубанского ГАУ и ВНИИВиВ «Магарач».

Генотипические данные: 50 аллелей микросателлитных локусов (SSR-маркеров), включая VrZAG62, VrZAG79, VVS2, VVMD5, VVMD7 и др. Выбор SSR-маркеров обусловлен их высоким полиморфизмом и воспроизводимостью.

Фенотипические данные: 21 хозяйственно-ценный признак, включая тип цветка, массу грозди, урожайность, сахаристость, кислотность, устойчивость к болезням и вредителям.

Методология. Исследование проводилось с использованием метода автоматизированного системно-когнитивного анализа (АСК-анализ), включающей следующие этапы:

1. Когнитивно-целевая структуризация: определение факторов (аллели генов) и целевых классов (градации фенотипических признаков).
2. Формализация данных: преобразование качественных и количественных описаний сортов в нормализованную информационную базу.
3. Синтез модели: расчет матрицы информативностей, где каждый элемент отражает количество информации, которое степень выраженности

каждого аллеля несет о принадлежности каждого сорта к каждому фенотипическому состоянию (например, «высокая сахаристость»). Расчет ведется по формуле, являющейся обобщением меры А. Харкевича для нечетких множеств.

4. Верификация: оценка адекватности модели методом перекрестной проверки (L1-оценка) и расчет F-меры Ван Ризбергена.

5. Решение десятков различных задач, которые можно классифицировать как задачи идентификации, прогнозирования, принятия решений и исследования моделируемой предметной области путем исследования ее модели.

В отличие от нейронных сетей, данный метод не требует итерационного обучения (backpropagation) и позволяет работать с неполными (фрагментированными) данными, что критически важно для биологических коллекций.

Результаты исследования

Оценка адекватности моделей. В ходе исследования было построено 7 вариантов системно-когнитивных моделей. Наибольшую точность показала модель INF3, основанная на критерии Хи-квадрат с поправкой на информационную насыщенность. Интегральная валидность модели составила 0,87 (по F-мере), что свидетельствует о высокой способности системы прогнозировать фенотип сорта исключительно на основе его генетического профиля.

Информационный анализ влияния генов (SWOT-анализ). С помощью инструментов когнитивной графики были выявлены генетические детерминанты 21 хозяйственно-ценный признак, включая тип цветка, массу грозди, урожайность, сахаристость, кислотность, устойчивость к болезням и вредителям. Так по ключевым признакам установлено:

- Сахаристость: Установлено, что наличие аллелей локуса *VVS2* имеет высокую информационную значимость для формирования признака «высокая сахаристость» (информационный вклад $> 0,5$ бит). В то же время, определенные аллели локуса *VrZAG62* выступают ингибиторами этого признака.

- Масса грозди: Выявлена нелинейная зависимость массы грозди от комбинации аллелей *VVMD7*. Построенные когнитивные функции показывают, что влияние гена не является бинарным («есть/нет»), а имеет градиентный характер, усиливая или ослабляя признак в зависимости от генетического окружения.

Кластерный анализ и плеiotропия. Кластерный анализ факторов (генов) позволил выявить группы маркеров, оказывающих схожее влияние на фенотип. Это может свидетельствовать о сцепленном наследовании или явлении плеiotропии, когда один генетический фактор влияет на несколько фенотипических признаков. Например, выявлен кластер генов, одновременно положительно влияющих на устойчивость к милдью и отрицательно – на размер ягоды, что подтверждает известную селекционерам проблему сцепления полезных и нежелательных признаков.

Алгоритм «обратного конструирования» сорта. На основе полученной модели разработан алгоритм поддержки принятия решений для селекционера: задается целевой профиль сорта (например: «срок созревания – ранний», «ягода – крупная», «устойчивость – высокая»); система решает обратную задачу: вычисляет спектр аллелей, который с максимальной вероятностью обеспечит данный фенотип; проводится поиск в базе данных родительских пар, обладающих необходимыми комбинациями аллелей.

Рассмотрим, почему Автоматизированный системно-когнитивный анализ (АСК-анализ) и система «Эйдос» относятся к классу Объяснимого

Искусственного Интеллекта (Explainable AI, XAI), и в чем их преимущество перед традиционными нейросетями («черными ящиками») для решения поставленной задачи и в общем случае.

Почему АСК-анализ и система «Эйдос» — это Объяснимый ИИ (XAI)? В отличие от классических нейронных сетей, где решение принимается на основе миллионов скрытых весовых коэффициентов, которые невозможно интерпретировать напрямую, система «Эйдос» обеспечивает полную прозрачность процесса принятия решений.

Основные критерии объяснимости в системе «Эйдос»:

Семантическая интерпретируемость весов: В основе модели лежит не абстрактный коэффициент, а количество информации (измеряемое в битах), которое конкретный фактор (например, наличие аллеля VrZAG62) несет о переходе объекта в определенное состояние (например, «высокая сахаристость»). Это базируется на системной теории информации (СТИ) и обобщенной мере А. Харкевича.

Когнитивная графика: Система генерирует наглядные визуализации (SWOT-диаграммы, когнитивные карты), которые показывают силу и направление влияния каждого фактора. Исследователь видит не просто прогноз («этот сорт будет сладким»), а почему он будет таким (какие именно гены усиливают этот признак, а какие — тормозят).

Отсутствие «Black Box» эффекта: В любой момент можно проследить цепочку рассуждений системы: от исходных данных через матрицу информативностей к итоговому прогнозу. Это позволяет верифицировать модель, опираясь на предметные знания (в данном случае — биологию и генетику).

Преимущества перед ИИ общего доступа (нейросетями и статистикой).

Какие критические недостатки есть у традиционных методов (корреляционного анализа, регрессии, Deep Learning) применительно к сложным научным задачам, которые успешно решает АСК-анализ и система Эйдос?

Проблема «Малых данных» (Small Data).

Традиционный ИИ: Современные нейросети (Deep Learning) требуют огромных обучающих выборок (Big Data) — тысячи и миллионы примеров. На малых выборках они переобучаются и дают низкую точность.

Система «Эйдос»: В представленной задаче выборка составляла всего 90 сортов винограда. Для нейросети этого критически мало. АСК-анализ разработан специально для работы в условиях неполной, фрагментированной и зашумленной информации. Он способен находить устойчивые закономерности даже на таких малых объемах данных (F -мера $> 0,85$ в исследовании).

Выявление нелинейных связей.

Статистика: Корреляционный анализ часто показывает отсутствие связи (коэффициент корреляции близок к 0), если зависимость нелинейна.

Система «Эйдос»: В работе показано, что влияние генов часто носит нелинейный характер. Например, определенные аллели могут работать как ингибиторы (подавители) признака только в определенных сочетаниях. АСК-анализ выявляет эти сложные зависимости, которые классическая статистика пропускает.

Решение обратной задачи (Reverse Engineering).

Традиционный ИИ: Нейросети обычно работают в одну сторону: Вход (гены) \rightarrow Выход (свойства). Заставить нейросеть сказать, «какие гены нужны, чтобы получить идеальный сорт», крайне сложно без дополнительных надстроек.

Система «Эйдос»: Поддерживает обратное конструирование. Селекционер задает целевой профиль (например: «ранний срок созревания» + «крупная ягода» + «устойчивость к болезням»), и система мгновенно рассчитывает спектр аллелей, который с максимальной вероятностью обеспечит этот фенотип. Это превращает селекцию из интуитивного перебора в инженерный процесс.

Отсутствие итерационного обучения (Backpropagation).

Традиционный ИИ: Нейросети требуют долгого обучения методом обратного распространения ошибки, подбора гиперпараметров и архитектуры.

Система «Эйдос»: Использует метод прямого счета (синтеза модели). Матрица знаний формируется непосредственно из данных, что значительно быстрее и не требует «шаманства» с настройкой параметров обучения.

Конкретные преимущества при решении задачи селекции винограда.

Благодаря функциональным особенностям АСК-анализа и системы Эйдос удалось достичь некоторых результатов, получение которых проблематично другими методами.

Количественная оценка влияния аллелей: Впервые было точно подсчитано (в битах), сколько информации несет конкретный аллель (например, локуса VVS2) для формирования сахаристости. Это позволило отделить значимые маркеры от «шума».

Обнаружение плейотропии и сцепления: Система выявила кластеры генов, которые влияют сразу на несколько признаков (например, положительно на устойчивость к милдью, но отрицательно на размер ягоды). Понимание этих связей критически важно, чтобы селекционер не пытался совместить несовместимое.

Работа с «мусорной» ДНК: Исследование подтвердило, что микросателлиты (ранее считавшиеся некодирующими участками) несут значимую информацию о фенотипе, что согласуется с современными представлениями о регуляторной роли генома.

Таким образом, АСК-анализ и система «Эйдос» действительно являются объяснимым ИИ (XAI), потому что они предоставляют семантически понятную модель причинно-следственных связей, а не просто результат прогноза и рекомендации по принятию решений как в простых, так и в сложных случаях. В контексте данной задачи это обеспечивает определенные преимущества: возможность проводить точную «цифровую селекцию» на малых выборках сортов, видеть скрытые нелинейные генетические механизмы и конструировать идеальные родительские пары, что невозможно сделать с помощью «черных ящиков» нейросетей или классической статистики.

Обсуждение

Полученные результаты демонстрируют, что так называемые «резервные» или некодирующие участки ДНК (к которым относятся микросателлиты) несут значимую информацию о фенотипе. Это согласуется с современными представлениями о регуляторной роли некодирующей части генома [1]. Применение АСК-анализа позволило преодолеть ограничения классической статистики. Там, где корреляционный анализ показывал отсутствие связи (из-за нелинейности), информационный подход выявил значимые зависимости. Это подтверждает тезис о том, что биологические системы являются сложными нелинейными объектами, для изучения которых требуются адекватные математические инструменты [1].

Заключение

Разработана и верифицирована информационная модель, позволяющая прогнозировать хозяйственно-ценные признаки винограда по его микросателлитному профилю с точностью до 87%.

Доказана эффективность применения методов интерпретируемого ИИ (на примере системы «Эйдос») для задач геномной селекции в условиях малых выборок.

Предложенный подход позволяет перевести селекционный процесс на новый уровень – от интуитивного подбора пар к инженерному конструированию генотипа («цифровая селекция»), что существенно сокращает сроки и затраты на выведение новых сортов.

Литература

1. Луценко Е.В., Трошин Л.П. Алгоритмы ампелометрии: учебное пособие // Е. В. Луценко, Л. П. Трошин. Учебное пособие. Том 7. Изд.2-е. Алгоритмы 61-70: Автоматизированный системно-когнитивный анализ и интеллектуальная система «Эйдос» в ампелометрии. – Краснодар : КубГАУ, 2025. – 255 с. – ISBN 978-5-93856-997-3. – DOI 10.13140/RG.2.2.31056.37128. – EDN EKSQQC, <https://www.researchgate.net/publication/397442001>
2. Луценко Е.В., Трошин Л.П. Программный комплекс «Ампелометрическая система интегрированная с интеллектуальной системой Эйдос»: http://lc.kubagro.ru/Installation_Eidos.php

References

1. Lucenko E.V., Troshin L.P. Algoritmy` ampelometrii: uchebnoe posobie // E. V. Lucenko, L. P. Troshin. Uchebnoe posobie. Tom 7. Izd.2-e. Algoritmy` 61-70: Avtomatizirovanny`j sistemno-kognitivny`j analiz i intellektual`naya sistema «E`jdoss» v ampelometrii. – Krasnodar : KubGAU, 2025. – 255 s. – ISBN 978-5-93856-997-3. – DOI 10.13140/RG.2.2.31056.37128. – EDN EKSQQC, <https://www.researchgate.net/publication/397442001>
2. Lucenko E.V., Troshin L.P. Programmny`j kompleks «Ampelometricheskaya sistema integrirovannaya s intellektual`noj sistemoj E`jdoss»: http://lc.kubagro.ru/Installation_Eidos.php