

УДК 004.45:575.116.4

ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ ГЕНЕТИЧЕСКОГО КАРТИРОВАНИЯ И АНАЛИЗА АССОЦИАЦИЙ МАРКЕР/ПРИЗНАК

Кочерина Н.В. – кандидат биологических наук,
с.н.с. лаборатории молекулярной и экологической генетики
Чесноков Ю.В. – доктор биологических наук,
заведующий лабораторией молекулярной и экологической генетики

ФГБНУ «Федеральный исследовательский центр Всероссийский институт генетических ресурсов растений имени Н.И. Вавилова»
ул. Большая Морская, д. 44, Санкт-Петербург, 190000
E-mail: alle007@mail.ru; yu.chesnokov@vir.nw.ru

В работе представлена информация об основных технических средствах и программном обеспечении, повсеместно используемых для проведения статистических расчетов и установления неравновесного сцепления маркер/признак, а также выявления ассоциаций между целевыми признаками/генами и молекулярными маркерами которые основываются на идентификации неравновесного сцепления. Кратко обсуждены предназначения основных компьютерных программ, наиболее широко используемых для проведения генетического картирования и установления ассоциаций маркер/признак. Приводятся Интернет-ссылки, позволяющие заинтересованным пользователям воспользоваться необходимым для них программным обеспечением.

Ключевые слова: генетическое картирование, неравновесное сцепление, установление ассоциаций маркер-признак, программное обеспечение

ВВЕДЕНИЕ

Большинство фенотипических признаков, представляющих интерес для селекции, являются количественными и, как правило, определяются аллельной структурой определенного числа генетических локусов. В связи с этим, как следует из накопленных к сегодняшнему дню данных, спектр генов, детерминирующих среднюю величину и генетическую дисперсию количественного признака, как правило, определяется лимитирующим фактором внешней среды (Драгавцев и др., 1984). Смена лимитирующего фактора влечет за собой также смену спектра генетических локусов, определяющих изменчивость признака (Драгавцев и др., 1984; Драгавцев, 2003; Чесноков и др., 2008). Тем не менее, было установлено существование отдельных ключевых генов/локусов хро-

мосом (Tanksley, 1993), которые при любых условиях вносят свой вклад в формирование определенного количественного признака, хотя мера этого вклада регламентируется внешней средой. Такие генетические локусы, названные локусами количественных признаков, QTL (quantitative trait loci), и составляют главный интерес современного молекулярного подхода к селекции полигенных признаков, в том числе маркерной помощи отбору, MAS (marker assisted selection – маркер-вспомогательная селекция) (Tanksley, 1993; Чесноков, 2013).

Для того, чтобы получить необходимую информацию о месторасположении QTL в геноме, проводят их молекулярно-генетическое картирование, в результате которого определяют не только их локализацию, но и то, от какого родителя получен тот или иной аллель QTL, процент фенотипической изменчивости, определяемый каждым из



выявленных QTL, и, конечно же, молекулярный маркер или маркеры с которыми сцеплены идентифицированный локус или локусы на группах сцепления. Одним из важнейших требований для картирования QTL является возможность производить обмеры и статистический анализ вариантов исследуемых признаков для установления достоверных корреляций между QTL и тем или иным маркером. Подробности применения таких подходов были описаны нами ранее (Чесноков, 2009; Кочерина и др., 2011).

Еще одним из необходимых требований, на основе которого строятся современные генетические карты – это наличие ассоциаций маркер-признак и неравновесного сцепления (linkage disequilibrium; LD) между ними (Чесноков, 2013). LD связано с исторически или эволюционно неравным и разнородным уровнем рекомбинации специфичных аллелей в различных локусах, определяющих проявление изучаемой генетической изменчивости в популяции. Неравновесное сцепление может быть установлено статистически и его выявление уже широко используется для картирования и даже клонирования генов, определяющих сложные признаки у человека, животных и растений (Чесноков, Артемьева, 2011). С изменением направления вектора усилий по идентификации и локализации генов/локусов хромосом от двуродительского скрещивания к естественным популяциям линий или образцов коллекций ГРП, и от традиционного QTL картирования к LD ассоциативному анализу, исследователи получили новый мощный инструмент по картированию интересующих их генов. Ассоциативное или LD картирование позволяет не только проводить анализ, классифицировать и группировать естественные и селекционные популяции (дикие виды, примитивные формы, сорта, линии), но и более эффективно использовать их в практической селекции с целью улучшения возделываемых культур.

Построение генетических карт, с использованием для этих целей молекулярных маркеров, и применение созданных карт для анализа ассоциаций маркер/признак, являются двумя предварительными условиями, требуемых для реализации маркер-вспомогательной селекции (MAS). Отличительной чертой проведения генетического картирования и установления ассоциаций маркер/признак, является оперирование большими объемами данных, полученных в результате экспериментальной оценки различных классов генетических маркеров и различного рода картирующих популяций. Для оперативного решения этой проблемы на сегодняшний день существует большое число статистических методологий и подходов, а также информационно-программных средств, подходящих для анализа различных типов популяций и маркеров. В настоящей работе приведено краткое описание предназначений основных компьютерных программ, наиболее широко используемых для проведения генетического кар-

тирования и установления ассоциаций маркер/признак. Без точной, статистически выверенной информации о такого рода ассоциациях невозможно проведение маркер-вспомогательной селекции, что в свою очередь, как ожидается, будет способствовать ускорению селекционного процесса, включая создание новых сортов с желаемыми характеристиками, в том числе для их применения в системе точного земледелия. В то же время мы надеемся, что в ближайшее время будут разработаны и использованы в практических целях новейшие программные обеспечения, которые позволят более эффективно на современном информационном уровне находить и оценивать ассоциации маркер/признак в новых типах популяций.

ПОСТРОЕНИЕ ГЕНЕТИЧЕСКОЙ КАРТЫ

Генетическая карта может быть построена посредством использования сегрегирующих популяций различных типов для видов с различным уровнем пloidности. Первым и наиболее часто используемым программным обеспечением для построения карт является MAPMAKER/EXP, которое было разработано в 1987 году (Lander et al., 1987). Почти все молекулярные карты, основанные на первом поколении молекулярных маркеров, – RFLP, – были построены с помощью этого программного продукта. Альтернативой является MAP MANAGER CLASSIC – графическая, интерактивная программа для картирования менделевских локусов. Она проводит все необходимые расчеты и построения на основе информации, полученной при межвидовых скрещиваниях с кодоминантными маркерами, беккроссных (backcross, BC) или рекомбинантных инбредных линий (recombinant inbred lines, RIL) у экспериментальных растений или животных (Manly, 1993; <http://www.mapmanager.org/mapmgr.html>).

Иногда для построения карт могут быть нужны некоторые специальные статистические модификации, при этом используют маркеры со строгим расхождением с сегрегацией в популяциях. MAPDISTO ([web/ftp: http://mapdisto.free.fr/](http://mapdisto.free.fr/)) – программа для картирования генетических маркеров в случае сегрегационного расхождения при использовании таких расщепляющихся экспериментальных популяций как BC, удвоенные гаплоиды (double haploid, DH) и RIL. Она может: (1) рассчитывать и рисовать генетические карты посредством графического интерфейса и (2) облегчать анализ маркерных данных, экспонируя сегрегационные расхождения благодаря различиям в жизнеспособности гамет или зигот.

Карты или данные от множественных популяций, полученных в результате различных скрещиваний, могут быть объединены в одну или консенсусную карту посредством совместного картирования (joint mapping). JOINMAP представляет собой программное обеспечение для построения карт генетического сцепления для нескольких типов картирующих популяций: BC₁, F₂, RIL, и DH, регенериро-

ванные из F_1 и F_2 поколений, а также семейств полных сибсов, полученных в результате скрещиваний неблизкородственных родителей (<http://www.kyazma.nl/index.php/mc.JoinMap/>). Она может объединять (соединять – join) данные, полученные из нескольких источников в интегрированную карту и реализовывать некоторые другие функции, включая определение групп сцепления, автоматическое определение фазы для семейств полных сибсов, полученных в результате скрещиваний неблизкородственных родителей, некоторые построения диагностических графиков и карт (van Ooijen, 2006).

Программным обеспечением со сравнительными функциями является CMAP, которое было разработано, как основанное на использовании Интернет-технологий, средство, позволяющее пользователям видеть сравнения генетических и физических карт. Пакет программы включает в себя инструменты для исправления данных карт (<http://www.gmod.org/cmap>; Ware et al., 2002).

КАРТИРОВАНИЕ QTL, ОСНОВАННОЕ НА СЦЕПЛЕНИИ

В экспериментах по картированию было продемонстрировано сцепление/ассоциации между целевыми признаками/генами и молекулярными маркерами которое основывается на генетическом сцеплении и неравновесном сцеплении (linkage disequilibrium, LD). Средства технической поддержки, необходимые для установления ассоциаций генотип-фенотип, обычно включают в себя: (1) статистические методы и программные средства для установления, оценки и сравнения ассоциаций генотип-фенотип посредством картирования сцепления, LD или ассоциативного картирования и картирования *in silico*, использующие для этого единообразные популяции, множественные популяции или все генетические ресурсы со всей доступной информацией обо всех разнообразных испытаниях по годам, временам года и местам их проведения; (2) статистические методы и программные средства для идентификации эффектов генетического фона или генетической среды окружения, аллелях локусов количественных признаков (quantitative trait loci, QTL) во множественных локусах и множественных аллелях в локусе; (3) программные средства, облегчающие процесс оценки от сцепленных маркеров до функциональных маркеров и генов-кандидатов; и (4) программные средства, облегчающие управление генетическими популяциями, картами и сопутствующими маркерными и фенотипическими данными.

Существует много коммерческих и находящихся в свободном доступе пакетов программного обеспечения предназначенных для установления ассоциаций между маркерными генотипами и фенотипами признаков. Наиболее распространенными и часто используемыми являются QTL Cartographer, MAPQTL, PLABQTL и QGENE. Все они оперируют только с двуродительскими популя-

циями, в то время как MCQTL (Jourjon et al., 2005) также проводит QTL картирование в многоаллельных ситуациях, включая двуродительские сегрегирующие популяции, а также наборы (серии, ряды) двуродительских или дуаллельных популяций. Наиболее часто используемым программным обеспечением в 1980-1990-х годах было MAP-MAKER/QTL, являющее сестринской программой MAPMAKER/EXP, разработанной Lander et al. (1987) (http://www-genome.wi.mit.edu/genome_software). Это ПО основывается на установлении максимального правдоподобия сцепления между маркером и фенотипом, используя для этого интервальное картирование, которое имеет дело с несложным QTL и несколькими стандартными популяциями. Другое раннее программное обеспечение MAPL (MAPping and QTL analysis; <http://lhm.ab.a.u-tokyo.ac.jp/software.html>; Ukai et al., 1995) позволяет пользователям получать результаты по уровню расщепления, тесту сцепления, значению рекомбинации, группировать маркеры, располагать маркеры в определенном порядке посредством метрического многомерного масштабирования, рисовать карту и графически изображать генотип и карту QTL посредством интервального картирования и дисперсионного анализа (analysis of variance, ANOVA).

Широко используемым для картирования QTL является программное обеспечение QTL Cartographer (<http://statgen.ncsu.edu/qtlcart/cartographer.html>), которое обеспечивает выполнение нескольких статистических методов, используя множество маркеров одновременно, включая составное интервальное картирование (composite interval mapping) и множественное составное интервальное картирование (multiple composite interval mapping). Взаимодействие между идентифицированными QTL также может быть установлено. PLABQTL использует составное интервальное картирование со многими функциями схожими с QTL Cartographer. QTL может быть локализовано и охарактеризовано в популяциях, полученных самоопылением потомков двуродительского скрещивания, или получением линий удвоенных гаплоидов (double haploid lines, DH). Простое и составное интервальное картирование производят, используя модель множественной регрессии. В качестве дополнительной функции ко многим другим пакетам ПО, может быть использован анализ взаимодействия QTL-окружающая среда (Utz, Melchinger, 1996).

Программа QGENE (<http://www.qgene.org/>) предназначена для проведения сравнительного анализа наборов данных по картированию QTL посредством вычислений, наиболее часто используемых пользователем. Кроме того, она написана вместе с дополнительным программным модулем построения (структурирования) программной системы, что позволяет ее расширять. QGENE создавалась в 1991 году как программа для картирования и моделирования популяций, в которую потом была добавлена возможность QTL анализа. Сравнительно недавно

QGENE была переписана на языке Java, что позволяет ее использование на любом компьютере с любой операционной системой. Она предлагает наиболее общепринятые методы QTL картирования и позволяет их параллельное (одновременное) сопоставление. Ее интерфейс по желанию может быть описан на языке пользователя, для этого необходимо самостоятельно написать трансляционный файл к программе. QGENE может быть использована для анализа признаков, QTL и пермутаций (математических перестановок), а также моделирования популяций и признаков.

Некоторые пакеты ПО могут быть использованы для построения карт сцепления у отдаленных скрещиваемых видов растений. ONEMAP представляет собой программу, предоставляющую такие возможности. Для этого используют семейства полных сибсов, полученных в результате скрещиваний неблизкородственных родителей (<http://www.ciagri.usp.br/~aafgarci/OneMap/>; Garcia et al., 2006). Другим ПО является MAPQTL – программное обеспечение, предназначенное для расчета позиций QTL (<http://www.mapqtl.nl>; van Ooijen, 2009), которое может быть использовано для нескольких типов картирующих популяций, включая BC₁, F₂, RIL, DH и семейства полных сибсов, полученных в результате скрещиваний неблизкородственных родителей. Она может быть применена для картирования QTL посредством интервального, композитного интервального и непараметрического картирования, с функцией для автоматического кофакторного отбора и проведения теста пермутации.

Несколько компьютерных программ для картирования рассматривают эпистаз (взаимодействия неаллельных генов) при картировании QTL. EPISTACY – это SAS программа, созданная для тестирования всех возможных двулокусных комбинаций для эпистатических взаимодействий или эффектов количественных признаков. Программа является настоящим SAS программным шаблоном, который пользователи должны настроить для того, чтобы она подходила к их данным. В простейших случаях пользователи могут изменить только названия файлов, содержащих их данные. Однако программа основана на методе наименьших квадратов и не позволяет проводить интервальное картирование (Holland, 1998).

Байесовскому картированию QTL в последние годы было уделено много внимания и, благодаря этому, на его основе были разработаны специальные компьютерные программы. Например, программное обеспечение BQTL (байесовское картирование локусов количественных признаков – Bayesian Quantitative Trait Locus mapping) было разработано для картирования генетических признаков в популяциях, полученных от скрещивания линий и в RIL (<http://hacuna.ucsd.edu/bqtl>; Borevitz et al., 2002). Оно выполняет: (1) установление максимального правдоподобия в мультигенных моделях; (2) байесовское оценивание

мультигенных моделей посредством аппроксимаций Лапласа; и (3) интервальное и композитное интервальное картирование генетических локусов. Программа BLADE (байесовское картирование неравновесного сцепления – Bayesian LinkAge DisEquilibrium mapping) была создана для байесовского анализа гаплотипов с целью неравновесного LD (linkage disequilibrium) картирования (<http://www.people.fas.harvard.edu/~junliu/TechRept/03folder/>; Liu et al., 2001; Lu, X. et al., 2003). MULTIMAPPER – программное обеспечение для байесовского картирования QTL при анализе BC, DH и данных поколения F₂, полученного в результате скрещивания инбредных линий (Martinez et al., 2005). Программа MULTIMAPPER/OUTBRED дополнила этот список популяциями, полученными от беспородных линий (<http://www.rni.helsinki.fi/~mjs/>).

Несколько пакетов ПО по картированию было разработано для картирования QTL, необходимых в особых специфических случаях. Программа MCQTL разработана для одновременного картирования QTL во множественных скрещиваниях и популяциях (<http://www.genoplante.com>; Jourjon et al., 2005). Она позволяет анализировать обычные популяции, полученные в результате скрещивания инбредных линий, и может соединять (связывать) семейства при допущении, что месторасположение QTL у всех у них одно и то же. Более того, возможно также диаллельное моделирование эффектов QTL, если используют множественные родственные семейства. Программа MAPPOP была предложена для селективного картирования и картирования при выбранном интервале значений случайной величины (bin mapping), посредством выбора хороших образцов из картирующих популяций, и для локализации новых маркеров на предсуществующих картах (Vision et al., 2000). В дополнение была разработана программа QTLNetwork, для картирования и визуализации генетической архитектуры, подчеркивающей сложные признаки, в экспериментальных популяциях, полученных в результате скрещивания двух инбредных линий (<http://ibi.zju.edu.cn/software/qtlnetwork>; Yang et al., 2008).

Как только базирующиеся на Интернет-технологиях программные средства стали все более и более востребованы, использующее Интернет-технологии QTL аналитическое программное обеспечение стало доступным. В качестве примера можно привести программу WEBQTL, созданная как интерактивная Интернет страница, полезная для изучения и анализа генетической модуляции тысяч фенотипов собранных за более чем 30-летний период сотнями исследователями, используя референсные панели рекомбинантных инбредных штаммов мышей (<http://www.webqtl.org/search.html>). WEBQTL включает плотные, проверенные на ошибки генетические карты, так же как и наборы данных экстенсивной экспрессии генов (фирмы Affymetrix), полученные в результате анализов более чем 35 штаммов мышей. Использующий Интернет-

технологии удобный для пользователя пакет QTL EXPRESS (<http://qtl.cap.ed.ac.uk>; Seaton et al., 2002), предназначенный для картирования QTL в беспородных популяциях, был разработан для скрещенных линий, семейств полусибс, а также ядерных семейств и сибсовых пар. Этот пакет предлагает две опции для тестов значимости QTL: тесты пермутации для определения эмпирических уровней значимости и метод ступенчатого перехода от малых значений к большим или наоборот (bootstrapping) для установления эмпирических доверительных интервалов локализации QTL. Фиксированные эффекты/ковариаты (независимые переменные) могут быть подогнаны, а модели могут включать одиночные или множественные QTL.

КАРТИРОВАНИЕ ПОСРЕДСТВОМ УСТАНОВЛЕНИЯ НЕРАВНОВЕСНОГО СЦЕПЛЕНИЯ

В последнее десятилетие ассоциативное или LD картирование стало очень популярным. Оно использует неструктурированные популяции, содержащие неродственных особей, образцы коллекций генетических ресурсов или случайным образом отобранные сорта (Чесноков, Артемьева, 2011; Чесноков, 2013). До ассоциативного картирования генотипированные единицы подвергаются статистическому анализу для удаления наиболее существенного фактора, – популяционной структуры, – которая может быть причиной ложноположительных ассоциаций, возникающих в большей степени из-за зависящих от обстоятельств, побочных корреляций, чем за счет настоящего сцепления. Для этих целей может быть использована программа STRUCTURE (<http://pritchardlab.stanford.edu/structure.html>; Pritchard et al., 2000a). Для ассоциативного картирования были разработаны некоторые пакеты ПО, в которые функциональные возможности анализа структуры популяций уже включены. STRAT, как программа, дополняющая STRUCTURE, использует структурированный ассоциативный метод для ассоциативного картирования разрешающий достоверные (действенные) методы изучения «случай-контроль» даже при наличии структуры у популяции (<http://pritchardlab.stanford.edu/software/STRAT.html>; Pritchard et al., 2000b).

Картирование QTL, основанное на LD

TASSEL – это универсальная программа для анализа признаков посредством установления ассоциаций, эволюционных отношений и сцепления, которая выполняет различные генетические анализы, включая ассоциативное картирование, установление разнообразия и расчет неравновесного сцепления (<http://sourceforge.net/projects/tassel/>; Zhang et al., 2006). Анализ LD между генотипами и фенотипами может быть выполнен либо посредством общей линейной модели, либо смешанной линейной модели. Общая линейная модель позволяет пользователям анализировать сложные (комплексные) полевые исследования, взаимодействия с окружающей сре-

дой и эпистаз. Смешанная модель специально разработана для обработки полигенных эффектов на множественных уровнях связанностей, включая информацию родословной. Эти анализы позволяют проводить установление и исследование LD у многих видов растений и животных.

Другие пакеты программного обеспечения включают в себя программу мультиаллельного межаллельного неравновесного анализа (Multiallelic Interallelic Disequilibrium Analysis Software, MIDAS), которая была разработана для анализа и визуализации межаллельного неравновесия между мультиаллельными маркерами (<http://www.genes.org.uk/software/midas>; Gaunt et al., 2006) и PEDGENIE (<http://bioinformatics.med.utah.edu/PedGenie/index.html>; Allen-Brady et al., 2006), которая создана как универсальное средство для анализа ассоциаций и перенос (распределение) неравновесия между генетическими маркерами и признаками в семьях произвольного размера и структуры. В эту программу может быть введена любая родословная любого размера от отдельных независимых особей до больших генеалогий. Анализ независимых особей и семей может проводиться совместно.

GENERECON (<http://www.daimi.au.dk/~mailund/GeneRecon/>) – другая программа, которая применяет для ассоциативного картирования теорию сращивания (слияния). Она основывается на методе Монте Карло по байесовской схеме цепи Маркова для тонкого ассоциативного картирования, используя при этом высокоплотные маркерные карты у животных. GENERECON предлагает недвусмысленные модели генеалогии образца в случае родственных хромосом по локусу, обуславливающему заболевание. Исследуемый случай и данные контроля представляются в виде информации о генотипе или гаплотипе, при этом программа производит оценку ряда параметров, наиболее важный из которых установление позиции локуса заболевания на хромосоме (Mailund et al., 2006).

Широко геномное ассоциативное картирование

Исследования по широко геномным ассоциациям (genome-wide association, GWA) на сегодня довольно часто применяют для нахождения связи между генетической изменчивостью и общими заболеваниями у человека, а также агрономически важных признаков у растений. В идеале хорошо оснащенные GWA исследования включают в себя измерения сотни тысяч полиморфизмов, выявляемых на уровне единичного нуклеотида (single nucleotide polymorphism; SNP) у тысяч особей. Чистый объем данных, генерируемый этими экспериментами, порождает очень высокие аналитические запросы. В ней есть ряд важных этапов, которые необходимо пройти в процессе анализа этих данных, многие из которых представляют собой критические моменты, ограничивающие производительность системы. Данные должны быть загружены и анализированы для выполнения изначального контроля качества до начала процедуры тестирования неравновесного сцепления. Оценка результатов может включать последующий статистический анализ, например, тест пермутаций, или после-

дующий контроль качества ассоциированных маркеров, например, проверка недоработанных интенсивностей генотипирования. Наконец, существенные ассоциации должны быть расположены в соответствии с приоритетом, используя для этого функциональные и биологические методы интерпретации, просмотр и редактирование имеющихся в распоряжении биологических примечаний и пояснений, информационных путей и модели неравновесного сцепления (Pettersson et al., 2008). GOLDSURFER2 (GS2) универсальная программа для анализа и визуализации GWA исследований была предложена (Pettersson et al., 2008; <http://www.well.ox.ac.uk/g2>). GS2 – это интерактивное, удобное для пользователя графическое приложение, которое может быть использовано на любых этапах GWA проекта от контроля качества исходных данных и анализа биологической оценки до проверки достоверности полученных результатов. Программа составлена на языке Java и может быть использована на всех платформах. С помощью GS2 очень большие наборы данных (например, 500000 маркеров и 5000 образцов) могут быть качественно оценены, быстро проанализированы и интегрированы в информацию о геномных последовательностях. При этом могут быть отобраны и функционально оценены кандидаты SNP.

Другие программные средства, предназначенные для исследований GWA, включают: GENOMIZER – не зависящая от типа операционной системы Java программа, определенная для анализа GWA экспериментов

(<http://www.ikmb.uni-kiel.de/genomizer>), PLINK – набор инструментальных средств программирования, разработанный для полногеномного анализа неравновесного сцепления (<http://pngu.mgh.harvard.edu/purcell/plink>; Purcell et al., 2007), MAPBUILDER – созданная для широкого хромосомного ассоциативного картирования (<http://bios.ugr.es/BMapBuilder>; Abad-Grau et al., 2006) и калькулятор расчета степени ассоциаций с двустадийной композицией (Calculator for Association with Two Stage design, CATS), который рассчитывает степень и другие необходимые величины для двустадийных GWA исследований (<http://www.sph.umich.edu/csg/abecasis/CaTS>; Skol et al., 2006).

Результаты масштабных GWA исследований все чаще и чаще размещаются в общедоступных базах данных. Но доступное программное обеспечение для анализа и интерпретации наборов GWA данных часто может быть затруднительным для использования (Buckingham, 2008). В настоящее время существует острая потребность в разработке усовершенствованных программных обеспечений для обработки наборов GWA данных, результаты которых было бы достаточно легко анализировать и исследователям проще обмениваться ими друг с другом, а также необходимо повысить качество таких программных продуктов, которые имели бы дело с увеличивающимися расчетными запросами, предъявляемыми этими наборами данных.

Литература

1. Драгавцев В.А. К проблеме генетического анализа полигенных количественных признаков растений. – 2003. СПб. ВИР. – 35 с.
2. Драгавцев В.А., Литун П.П., Шкель И.М., Нечипоренко Н.Н. Модель эколого-генетического контроля количественных признаков растений. // Доклады АН СССР. – 1984. – Т. 274, №3. – С. 720-723
3. Кочерина Н.В., Артемьева А.М., Чесноков Ю.В. Использование лод-оценки в картировании локусов количественных признаков у растений. // Доклады Российской академии сельскохозяйственных наук. – 2011. – №3. – С. 14-17
4. Чесноков Ю.В. Картирование локусов количественных признаков у растений. – 2009. СПб., ВИР. 100 с.
5. Чесноков Ю.В. Молекулярно-генетические маркеры и их использование в предселекционных исследованиях. – 2013. СПб: АФИ. 116 с.
6. Чесноков Ю.В., Артемьева А.М. Ассоциативное картирование у растений. // Сельскохозяйственная биология. – 2011. – №5. – С. 3-16
7. Чесноков Ю.В., Почепня Н.В., Бёрнер А., Ловассер У., Гончарова Э.А., Драгавцев В.А. Эколого-генетическая организация количественных признаков растений и картирование локусов, определяющих агрономически важные признаки у мягкой пшеницы. // Доклады Академии наук. – 2008. – Т. 418, №5. – С. 693-696
8. Abad-Grau M.M., Montes R., Sebastiani P. Building chromosome-wide LD maps. // Bioinformatics. – 2006. – V.22. – P. 1933-1934
9. Allen-Brady K., Wong J., Camp N.J. PedGenie: an analysis approach for genetic association testing in extended pedigrees and genealogies of arbitrary size. // BMC Bioinformatics. – 2006. V.7. – P. 209
10. Borevitz J.O., Maloof J.N., Lutes J., Dabi T., Redfern J.L., Trainer G.T., Werner J.D., Asami T., Berry C.C., Weigel D., Chory J. Quantitative trait loci

controlling light and hormone response in two accessions of *Arabidopsis thaliana*. // Genetics. – 2002. – V.160. – P. 683-696

11. Buckingham S.D. Scientific software: seeing the SNPs between us. // Nature Methods. – 2008. – V.5. – P. 903-908
12. Garcia A.A., Kido E.A., Meza A.N., Souza H.M., Pinto L.R., Pastina M.M., Leite C.S., Silva J.A., Ulian E.C., Figueira A., Souza A.P. Development of an integrated genetic map of a sugarcane (*Saccharum spp.*) commercial cross, based on a maximum-likelihood approach for estimation of linkage and linkage phases. // Theoretical and Applied Genetics. – 2006. – V.112. – P. 298-314
13. Gaunt T.R., Rodriguez S., Zapata C., Day I.N.M. MIDAS: software for analysis and visualization of interallelic disequilibrium between multiallelic markers. // BMC Bioinformatics. – 2006. – V.7. – P. 227
14. Holland J.B. EPISTACY: a SAS program for detecting two-locus epistasis interactions using genetic marker information. // Journal of Heredity. – 1998. – V.89. – P. 374-375
15. Jourjon M.F., Jasson S., Marcel J., Ngom B., Mangin B. MCQTL: multi-allelic QTL mapping in multi-cross design. // Bioinformatics. – 2005. – V.21. – P. 128-130
16. Lander E.S., Green P., Abrahamson J., Barlow A., Daly M.J., Lincoln S.E., Newburg L. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. // Genomics. – 1987. – V.1. – P. 174-181
17. Liu J.S., Sabatti C., Teng J., Keats B.J.B., Risch K. Bayesian analysis of haplotypes for linkage disequilibrium mapping. // Genome Research. – 2001. – V.11. – P. 1716-1724
18. Lu X., Niu T., Liu, J.S. Haplotype information and linkage disequilibrium mapping for single nucleotide polymorphisms. // Genome Research. – 2003. – V.13. – P. 2112-2117

ЗАКЛЮЧЕНИЕ

Таким образом, исходя из выше изложенного, можно заключить, что на сегодняшний день существует большое количество разнообразных и разноплановых программных обеспечений, позволяющих проводить не только статистическую обработку данных, строить генетические карты, но и довольно эффективно устанавливать, а также оценивать ассоциации маркер/признак. И хотя картирование с помощью молекулярных маркеров требует специализированной технической оснащенности (например, программного обеспечения, лабораторного оборудования), а также опыта, знаний и компетентности в понимании разрешающей способности того или иного метода, включая интерпретацию полученных результатов, совершенно очевидно, что в перспективе исследователи для того, чтобы прокартировать интересующий их ген или локусы хромосом, не будут нуждаться в том, чтобы начинать с нуля широкоштабный физиолого-генетический эксперимент по идентификации и анализу QTL или установлению ассоциаций маркер/признак, а смогут оперировать существующими, постоянно пополняемыми базами данных и современным программным обеспечением, позволяющими проводить статистически достоверный и оперативный анализ внутривидовых взаимодействий.

Благодарности. Работа выполнена при частичной финансовой поддержке грантов РФФИ 13-04-00128 и 16-04-00311-а.

SOFTWARE FOR GENETIC MAPPING
AND ANALYSIS OF MARKER/TRAIT ASSOCIATIONS

Kocherina N.V., Chesnokov Yu.V.

N.I. Vavilov Institute of Plant Genetic Resources (VIR)

190000, Russia, St.-Petersburg, Bolshaya Morskaya Str. 44

E-mail: alle007@mail.ru; yu.chesnokov@vir.nw.ru

Summary. The paper presents the information about technical tools and software wide used for statistic calculations and detection of linkage disequilibrium marker/trait as well as association between traits/genes and molecular markers based on identification of linkage disequilibrium. It is briefly discussed the destination of main computer programs most wide used for genetic mapping and association marker/trait. It is given Internet web-sites for downloading of necessary software.

Keywords: genetic mapping, linkage disequilibrium, identification association marker-trait, software.

19. Mailund T., Schierup M.H., Pedersen C.N.S., Madsen J.N., Hein J., Schauer L. GeneRecon – a coalescent based tool for fine-scale association mapping. // *Bioinformatics*. – 2006. – V.22. – P. 2317–2318
20. Manly K.F. A Macintosh program for storage and analysis of experimental genetic mapping data. // *Mammalian Genome*. – 1993. – V.4. – P.303–313
21. Martinez V., Thorgaard G., Robison B., Sillanpaa M.J. An application of Bayesian QTL mapping to early development in double haploid lines of rainbow trout including environmental effects. // *Genetical Research*. – 2005. – V.86. – P.209–221
22. Pettersson F., Morris A.P., Barnes M.R., Cardon L.R. Goldsurfer2 (Gs2): a comprehensive tool for the analysis and visualization of genome wide association studies. // *BMC Bioinformatics*. – 2008. – V.9. – P.138
23. Pritchard J.K., Stephens M., Donnelly P. Inference of population structure using multilocus genotype data. // *Genetics*. – 2000a. – V.155. – P.945–959
24. Pritchard J.K., Stephens M., Rosenberg N.A., Donnelly P. Association mapping in structured populations. // *American Journal of Human Genetics*. – 2000b. – V.67. – P.170–181
25. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A.R., Bender D., Maller J., de Bakker P.I.W., Daly M.J., Sham P.C. PLINK: a toolset for whole-genome association and population-based linkage analysis. // *American Journal of Human Genetics*. – 2007. – V.81. – P.559–575
26. Seaton G., Haley C.S., Knott S.A., Kearsey M., Visscher P.M. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. // *Bioinformatics*. – 2002. – V.18. – P.339–340
27. Skol A.D., Scott L.J., Abecasis G.R., Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. // *Nature Genetics*. – 2006. – V.38. – P.209–213
28. Tanksley S.D. Mapping polygenes. // *Annual Reviews of Genetics*. – 1993. – V.27. – P.205–233
29. Ukai Y., Osawa R., Saito A., Hayashi T. MAPL: a package of computer programs for construction of DNA polymorphism linkage maps and analysis of QTL. // *Breeding Science*. – 1995. – V.45. – P.139–142 (in Japanese)
30. Utz H.F., Melchinger A.E. PLABQTL: a program for composite interval mapping of QTL. *Journal of Agricultural Genomics*. – 1996. – <http://www.cabi-publishing.org/jag/papers96/paper196/indexp196.html> (доступно с 30 июня 2007 г.)
31. van Ooijen A.J. JoinMap 4.0: Software for the Calculation of Genetic Linkage Maps in Experimental Populations. *Plant Research International B.V & Kyazma B.V. Wageningen, Netherlands*. – 2006. – 45 p.
32. van Ooijen J.W. MapQTL 6. Software for the mapping of quantitative trait loci in experimental populations of diploid species. *Wageningen, Netherlands*. – 2009. – 60 p.
33. Vision T.J., Brown D.G., Shmoys D.B., Durrett R.T., Tanksley S.D. Selective mapping: a strategy for optimizing the construction of high-density linkage maps. // *Genetics*. – 2000. – V.155. – P.407–420
34. Ware D.H., Jaiswal P., Ni J., Yap I.V., Pan X., Clark K.Y., Teytelman L., Schmidt S.C., Zhao W., Chang K., Cartinhour S., Stein L.D., McCouch S.R. Gramene, a tool for grass genomics. // *Plant Physiology*. – 2002. – V.130. – P.1606–1613
35. Yang J., Hu C., Hu H., Yu R., Xia Z., Ye X., Zhu J. QTLNetwork: mapping and visualizing genetic architecture of complex traits in experimental populations. // *Bioinformatics*. – 2008. – V.24. – P.721–723
36. Zhang Z., Bradbury P.J., Kroon D.E., Casstevens T.M., Buckler E.S. TASSEL 2.0: a software package for association and diversity analyses in plants and animals. Poster presented at Plant and Animal Genomes XIV Conference, 14–18 January 2006. San Diego, California. – 2006.