

# 一、新一代植物遗传学

自然变异（Natural variation）是现代生物学研究的一大基础问题。随着测序技术的进步，人们将会获得越来越多的物种个体基因组序列信息。但是，究竟生物体基因型变异是如何转化为表型变异的呢？植物是除了人体组织外解决这一问题的最理想的研究对象。

在分析生物如何产生复杂表型机制时，植物一直都是最佳的突破口。例如，人们以谷物为研究对象，率先发现是多基因座分离（segregation at multiple loci）现象导致表型连续分布（continuous distribution of phenotypes）；通过豌豆中的个体分子标志物发现了数量遗传性状位点（Quantitative trait loci, QTL）等；借助传统遗传学技术，人们现在可以将多个与数量遗传性状相关的基因克隆入植物基因组；而拟南芥（*Arabidopsis thaliana*）则不仅仅在植物生物学领域起到了非常重要的“旗舰”作用，同时也是其它生物学领域常用的一种重要模式生物。此外，人们早期取得的一些生物学研究成果都来自于农作物，比如玉米、水稻和西红柿。

现在，各种生物的遗传图谱被逐一揭示，物种基因型的分析费用也在大幅降低。因此，使用连锁遗传作图法（小词典1）寻找、发现、以及确认相异品系（divergent strains）的基因（或QTL），甚至寻找能引发特定表型效应的单核苷酸位点都已成为科研工作中的一项“常规项目”。

值得关注的是，尽管人们已经对能引起各种不同表型的拟南芥基因突变进行了大规模的筛查工作，但植物遗传学家使用连锁遗传作图法还是能够不断发现大量新的对拟南芥表型有影响的QTL基因。

由于要绘制一张高质量的连锁遗传图谱既费时又费力，加上它并非唯一一种能有效识别和完成基因定位的方法，因此，遗传学家们逐渐将注意力转移至全基因组关联作图（genome-wide association (GWA) mapping）技术。简单地说，该技术就是指从人类全基因组范围内的序列变异（单核苷酸多态，SNP）中，筛选出那些与疾病性状有关联的SNP。

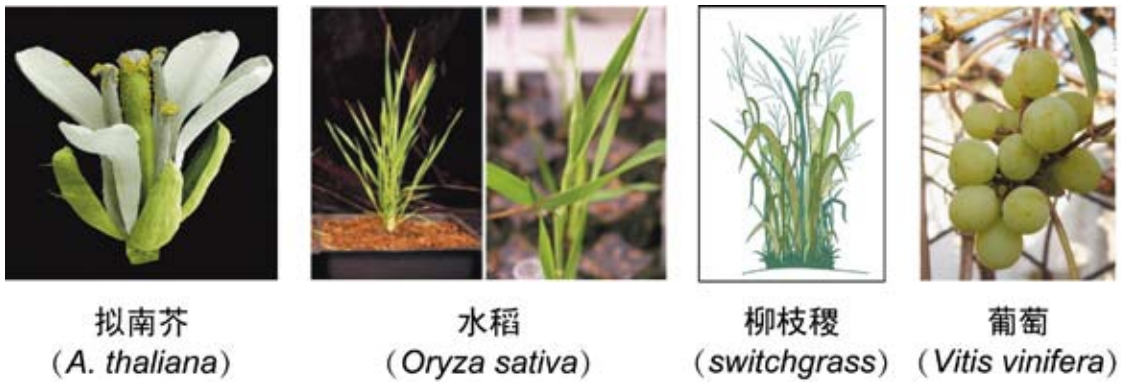
采用GWA技术可以找出个体特定表型与其基因组中DNA序列变异之间的关联信息，这与通过成千上万的SNP位点来寻找特定的表型相似。GWA技术的精确度比连锁遗传图谱技术要高，因为它的研究对象是普通的种群，而非杂交的后代。当然，这些自然种群中的相互关联性更能反映真实的历史重组事件。

下文将介绍GWA技术最近取得的一些重大成果、GWA技术在植物研究中的巨大应用潜力以及对GWA技术未来发展方向进行的预测。

## 1. 采用GWA作图需要考虑的问题

GWA技术适用于研究多种植物，尤其是拟南芥（*A. thaliana*）和水稻（*Oryza sativa*）等自花授粉物种，或者柳枝稷（*switchgrass*）、葡萄（*Vitis vinifera*）等无性繁殖物种。

这是由于这些自然突变率低的物种基因组被测序之后，植物遗传学家就能根据它们的序列资料对各种不同的表型进行了分析。而且，GWA在研究复杂遗传背景下的表型，尤其是表型受



环境影响时，还可以依靠统计学技术进行分析。

但是，值得关注的是，在使用GWA作图技术时，所选择的种群结构是一个非常关键的因素（详见后文），尤其是在分析与环境适应相关的表型时更是如此。幸而，我们可以借助统计学手段来解决这个问题。但应明确的是，统计学方法并非每次都能起作用，当它不起作用时，就需要借助杂交对照组的连锁遗传作图分析技术了。因此，我们也许应该将连锁遗传作图技术作为GWA作图技术的一个必需的补充性组成部分来看待（图1）。

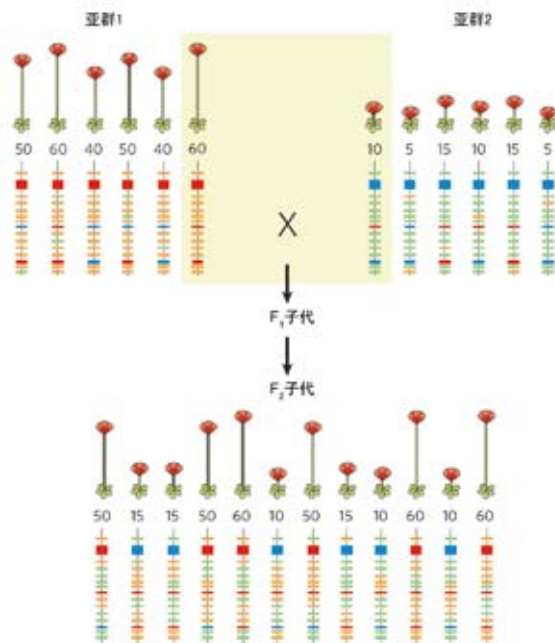


图1 如果两个待研究亚群间的遗传差异非常明显，那么就不适合采用GWA技术分析。如图所示，两个不同亚群的植株高矮差异明显（图中所示的数字即代表植株的高度），而且基因型也不相同。携带红色等位基因的植株较高，而携带蓝色等位基因的植株较矮。其它的背景标志物（绿色和橙色）主要和亚群特征相关，但也能影响植株高度，不过这些标志物不能起到决定性作用。通过杂交得到了F<sub>2</sub>子代，这样，任何背景标志物和主要性状标志物之间的连锁不平衡就会被打破。此时哪怕分辨率不高，也能很容易获得主要性状基因的遗传图谱了。

另一个需要考虑的问题，就是这两种作图技术（GWA作图和连锁作图）在研究性状遗传结构如何影响统计方法上也有所不同（图2）。在一个种群中，某个等位基因对表型的影响力取决于它在种群中的出现频率和其表型效应的大小。因此，GWA作图法在发现稀有等位基因时（哪怕这些基因的表型效应非常大）就显得无能为力了（图2c）。与此相反，对随机选择出的两个亲本杂交出的子代植株通过作图法发现的等位基因可能表型效应非常大，但却没什么意义。因为，从进化的角度来说，这些等位基因都属于罕见基因。换句话说，就是使用杂交体对QTL作图很有可能会将研究人员带往少见的，甚至在很多时候是有害的等位基因。这些等位基因的表型效应也许很大，但它们与自然界中的大部分的表型多样性都没有什么关系。

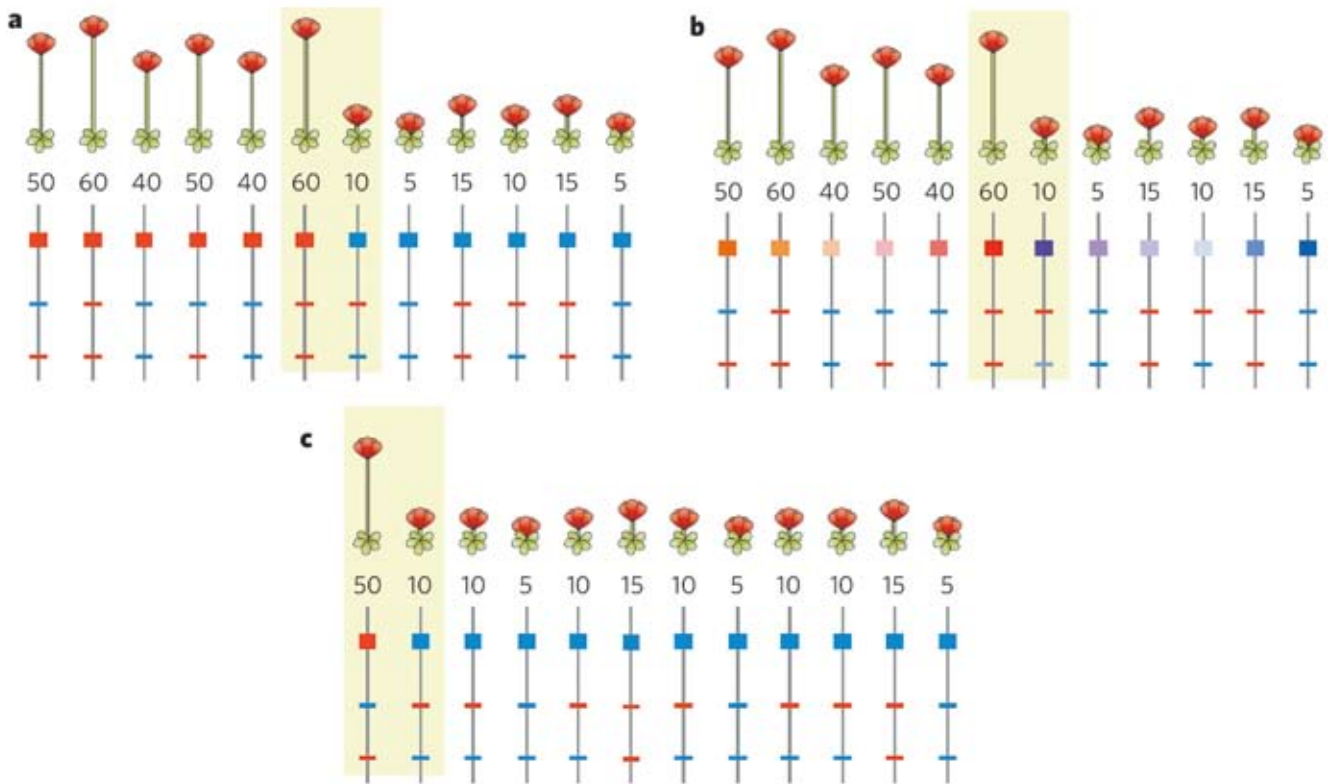


图2 GWA作图法的效果很大程度上取决于性状的遗传结构。与图1一样，图中携带红色等位基因为高植株，携带蓝色等位基因为矮植株。a. 种群结构中没有致混淆因素，所以可以很容易地使用GWA作图法发现主要性状基因，否则，就需要采用图1的方法了。b. 虽然也有主要性状基因，但同时还存在许多其它不同的会影响植株高矮性状的基因，如图中彩色所示，因此若只使用GWA法就起不到什么作用。c. GWA法无法发现少见的但表型效应很大的基因（顶部位点），只能发现常见的但表型效应很小的基因（中部位点）。在这三个例子（a-c）中，不使用GWA法而是通过连锁作图法对杂交系植株进行分析就可以成功找出主要性状基因。但这样找到的基因在种群中只是罕见的基因，因此对于种群中大多数的个体来说没有意义，也就无法代表该种群的特征。

## 2. GWA作图法的可用资源

随着国际人类基因组单体型图计划（International HapMap Project，该计划旨在全世界范围内发现人类的SNP的不断发展，拟南芥基因多态性的参考资源也不断丰富。现在，这些资源已经达到了可以和人类同类资源相媲美的地步。为了更好地利用这些数据，人们很早就开始了一项前沿研究。该研究的目的是对拟南芥多态性的基础模式，包括变异水平、基因连锁不平衡的程度、种群结构等进行研究。它一共对96个拟南芥样品的共计1,400个基因座进行了测序。接着，人们又使用高密度芯片从这96个样品中再选出20个样品进行再次测序。该研究发现SNP数据帮助Affymetrix公司设计并制造出含有250,000条探针的用于对拟南芥进行GWA分析的基因型芯片（genotyping microarray）。现在，人们已经使用该芯片对1,300个自然近交系品种进行了基因型分析，读者可登陆<http://walnut.usc.edu/2010>了解更多拟南芥基因多态性数据。该芯片还将用于更多的表型研究项目。该研究项目的操作模式有望成为研究其它物种的参考模

式，例如正在进行中的水稻研究项目（参见<http://irfgc.irri.org>国际水稻功能基因组学协会）就借鉴了拟南芥研究中的经验。

与人类HapMap项目情况相同，随着科技的进步，拟南芥和水稻研究中使用的多种技术已经过时了。今后，人们将使用新一代的测序技术，例如Illumina公司最新的测序仪和Applied Biosystems公司的SOLiD系统来发现更多的SNP。构建第一代拟南芥、水稻单倍体型图（haplotype map）时使用的芯片杂交技术（microarray-hybridization），由于花费过高，并且准确性不高而放弃使用。而且测序费用的大幅下降意味着在寻找发现对基因型最有价值的SNP时，精妙的试验设计将不再重要。目前，研究人员已经发现了140,000个标签SNP（小词典2），这足以覆盖整个拟南芥基因组。现阶段，已经不会再因为经济原因而在对全部250,000个已知的高质量非单态SNP进行基因型研究面前却步了。这类非单态SNP只在少数个体中有发现，因此我们对用这些SNP来预测其它SNP的能力有多大不得而知。

### 3. 种群结构的重要性

那么，使用GWA方法在寻找、发现基因这一方面的应用前景在哪里？众所周知，种群统计学（demography）会影响连锁不平衡（linkage disequilibrium）。其中一个例子，就是欧洲人群中的连锁不平衡现象要比非洲人多，这也印证了人类的非洲起源论。

另一个例子则来自野生的拟南芥。北美洲的野生拟南芥中连锁不平衡现象要多于欧洲的野生拟南芥，这也和拟南芥是由欧洲移民带到美洲这一事实相吻合。那么为什么会出现这种现象呢？究其原因，可能是因为种群在“殖民”时存在一个“瓶颈”，还没有足够的时间通过遗传重组来减少等位基因的连锁不平衡现象。

有这样一种可能还没有被人们广泛接受的观点，即由于存在种群结构，因此一个样品植物性状的遗传结构取决于这个样品的出处。比如，使用GWA作图法很快就发现了FRIGIDA基因对来自欧洲大陆西北部的拟南芥开花时间有影响，但对来自亚洲中部的拟南芥则不起作用。如果性状改变是由一个基因的多个等位基因（和数量少但出现频率高的等位基因情况相反）引起的，那么研究人员使用来自世界各地的样品进行GWA研究，就很有可能得出该性状没有主要位点的错误结论（图2）。这也是前面提到的种群结构带来的问题。一个等位基因的重要性总是取决于它所在的群体，但从进化的角度来说，我们完全不清楚哪一个群体最重要。

在关联研究中，种群结构（population structure）是一个很重要的致混淆因素，因此需要给予足够的重视。尤其是在研究和局部适应性相关的性状时，如植物的花期或者人类的肤色等更是如此。对玉米和拟南芥的研究已经处在了解决这个问题的关键时刻，人们将有望使用统计学方法来解决这个问题。现在，研究人员已经发现了控制玉米维生素A原（provitamin A）含量的基因位点，这对于饮食结构单一的人来说无疑是天大的福音。

### 4. 将GWA作图法和连锁作图法结合起来

其实，要解决上述种群结构问题很简单，只需要将针对天然种群的GWA作图法和针对试验种群的连锁作图法结合起来就行了。这样就能将GWA法的高分辨率和连锁作图法的抗



干扰性结合起来。目前，使用该联合策略，研究人员已经在拟南芥的研究中取得了成功。不过，由于研究人员在研究人类遗传学时不可能获得杂交对照组，所以得依靠传递不平衡分析（transmission-disequilibrium test, TDT）方法的帮助。该方法是通过分析等位基因从亲代传到子代的情况来判断连锁现象的。

Ed Buckler小组在研究玉米时，创造性地采取了一种被称作巢式联合作图（nested association mapping）的新方法。要使用GWA法来研究玉米估计要好几年之后，这是由于玉米的基因组比人类的要大得多，也更富多态性，而且连锁不平衡现象不广泛，使用GWA法进行相关研究，费用非常昂贵。研究人员使用普通的标准基因型杂交得到了5,000种重组自交系（recombination inbred line, RIL）群体。对最初的亲本玉米进行测序，而对RIL则进行基因型检测，最终得到足以覆盖每一个RIL的单倍体型图。因为在构建RIL的过程中杂交过程是受限的，所以只需使用不多的标志物就可以得到精确度相对较高的单倍体型图。这种策略和构建异质种群（heterogeneous stock, HS）小鼠以及遗传多样性项目（CC）（小词典3）小鼠类似。

因此，巢式联合作图法实际上只需要依赖试验杂交系来构建基因图谱，这就避免了种群结构带来的干扰。在这些遗传作图过程中，亲本中含有的等位基因可以被用来提高GWA图谱的分辨率。只要从众多的RIL中精心挑选出几个合适的样品，就可以发现该策略在拟南芥试验中的确很成功。

## 5. 未来的发展方向

过去两年间，研究人员发表了大量使用GWA对人类进行研究的成果。但是发现的众多等位基因中，只有很少一部分能够解释种群中的变异现象。在植物中进行类似研究的成果将会在短期内发表，届时大家将会知道是否表型效应大的等位基因在植物中比在人类中要多。

一直以来，我们对于植物进化了解得不多，这都是因为无法将亲缘关系很近的植物基因组进行比较。在酵母、新杆状线虫（*Caenorhabditis*）、果蝇（*Drosophila*）和灵长类动物的研究中，这种基因组比较都发挥了巨大的作用。



酵母



新杆状线虫（*Caenorhabditis*）



果蝇  
(*Drosophila*)



灵长类动物

比较最近公布的基因组不仅能得到保守的遗传元素图谱，还能发现谱系特异性的选择现象以及最近掺入基因组中的外来片段。而发现外来遗传片段对于植物研究来说最有价值，因为植物的种间杂交往往能得到更好的子代品种。我们所得到的基因组序列中与拟南芥亲缘关系最接近的物种是木瓜，它在约7,000万年前与拟南芥出现分化。这种缺乏合适基因组信息的现象实在是太讽刺了，因为多年来，植物的物种内多态性信息比其它非人类物种都要多。因此，植物遗传学界对于*Arabidopsis lyrata*和*Capsella rubella*这两种与拟南芥亲缘关系最近的植物（它们分别在500

万年前和1,000万年前与拟南芥出现分化)的基因组测序工作即将完成这一消息都激动不已。同样,对水稻的野生祖先基因组进行测序也将为人们了解人工种植的水稻提供更多有用的信息。

尽管大家都很热衷于研究种属间的差异,但实际上最有价值的还应该是种内序列的多样性。直到最近,大家的研究重点才开始转到研究个体间的细小差异,例如SNP、小的片段插入和缺失等等。人们越来越清楚,不论是玉米、拟南芥还是人类,个体间都会因为基因和基因组重组现象而有所不同。

这种深入研究物种内多样性的项目之中有一个就是最近刚刚启动的“千人基因组计划(‘1000 Genomes’ project, <http://1000genomes.org>)”。该计划希望能借助最新的测序技术,对全球范围内至少1,000个人的基因组进行测序,从而得到一份详尽的人类基因变异情况资料。

目前,有研究学者也在筹办一个补充性的拟南芥计划,也即“1001基因组(1001 Genomes, <http://1001genomes.org>)”计划。拟南芥是最适合这类研究的物种,因为它适合于种群遗传学(population genetics)分析,以及用来了解基因型和环境之间的相互作用(genotype-environment interactions)等等(表1)。

表1 拟南芥最适于进行“1001基因组”计划的具体原因

拟南芥的基因组大小还不到人类基因组的1/20,而且重复结构少,因而比人类基因组测序要简单、省钱得多。

大部分拟南芥品种的基因组都是天然自交系(naturally inbred)品种,而人类和大部分模式生物则都是杂合子。因此,使用全基因组鸟枪法对拟南芥进行基因组测序时就可以避免单倍型相(haplotype phase,即等位基因在同源染色体上的排列问题)这一问题了,这将大大简化后续分析工作。

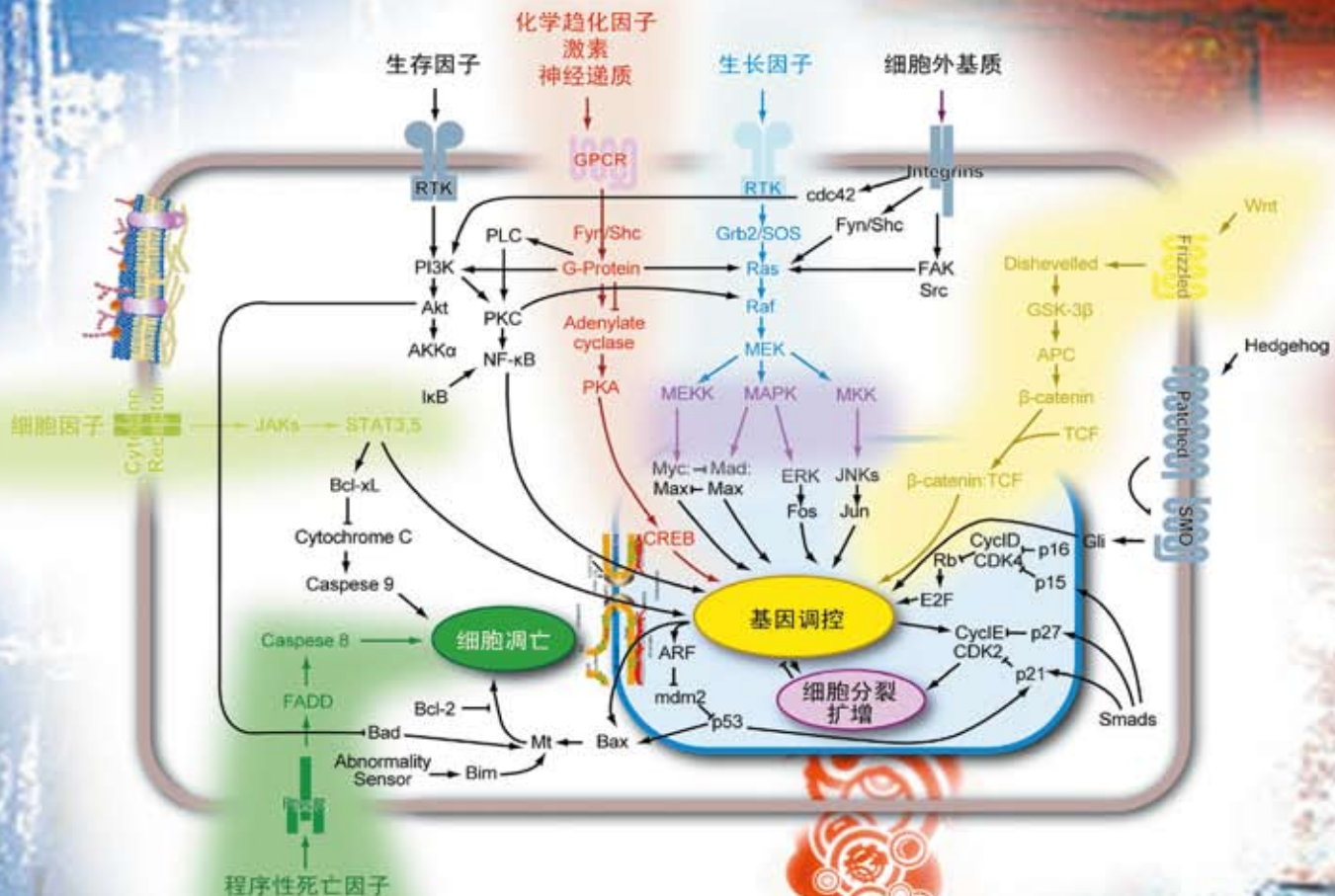
对拟南芥开展相关实验,可以避开道德伦理等问题。研究人员无法对参与“千人基因组”计划的人使用GWA法,除非对来自这1,000个人的细胞系进行表型分析。相反,对那些挑选出来用于基因组测序的自然条件下的自交系拟南芥,人们则可以制备足够多的种子。这样一来,具有相同表型的任何数量的植物都可以在很多不同环境下进行表型研究了。因此而得到的序列信息也能在很多层面上,包括生化水平、代谢水平、生理水平、形态学水平和整个植物水平上直接用于GWA分析。

将GWA方法和正向遗传学(小词典4)方法结合起来,最终会将基因型和表型联系起来,至少在拟南芥上是可以做到这一点的。而且,随着使用类似的方法对更多的物种进行研究,人们最终肯定会弄清楚进化过程中的分子遗传学机制。

原文检索: Magnus Nordborg & Detlef Weigel. (2008) Next-generation genetics in plants, *Nature* 456 (7223): 720-723.

 筱玥/编译





# 信号转导

- ◎ **Primer Array** : 检测样品中信号转导途径的激活情况。
- ◎ **OmicsLink™ ORF克隆**: 组成型表达蛋白, 激活信号转导途径。
- ◎ **OmicsLink™ shRNA克隆**: 敲减基因表达, 抑制级联信号途径。

**GeneCopoeia™**  
Expressway to Discovery

**GeneCopoeia, Inc.**  
Tel: 301-515-6982 (USA)  
Fax: 301-515-6983 (USA)  
Web: <http://www.genecopoeia.com>

**复能基因**  
**FulenGen**

广州复能基因有限公司  
电话: (020) 32052376、32052410、32290874  
传真: (020) 32052877  
公司网址: <http://www.fulengen.com>

### 1. 连锁遗传作图法 (linkage mapping)

连锁遗传图又叫遗传图谱 (genetic map)，它是具有遗传多态性 (在基因组的一个遗传位点上具有一个以上的等位基因，它在群体中的出现频率皆高于1%) 的遗传标志为“路标”，以遗传学距离 (在细胞减数分裂事件中两个位点之间进行交换，重组的百分率，1%的重组率称为1cM) 为图距的基因组图。图谱的建立为基因的识别和完成基因定位创造了条件。传统的遗传作图法认为，两个连锁基因距离越远，它们之间就越有可能发生重组现象。

### 2. 标签SNP (tag SNP)

一个染色体区域可以有很多SNP位点，能代表其它位点信息的SNP位点称为标签SNP；用少数几个标签SNP就能够提供该区域内大多数的遗传多态模式。利用标签SNP可极大提高关联分析的有效性。

### 3. 遗传多样性项目 (Collaborative Cross, CC)

Collaborative Cross项目是一项发展一组1000只新的、遗传多样性小鼠品系的计划。这些小鼠是由8个遗传分化的小鼠品系通过连续多代的同胞交配，最终获得了遗传多样且能控制的重组近交系。与常规采用双亲本衍生的群体相比，其遗传变异更丰富。Collaborative Cross能满足广泛的需求。

### 4. 正向遗传学 (forward-genetic)

研究突变表型以确定突变基因的传统遗传学方法。正向遗传学筛查是分子遗传学家最初可使用的一种方法，该技术意在检出产生特定表型的变异。为了提高变异的速度，常使用诱变剂来实现。而一旦分离出变异体，就可以鉴定出对应的突变基因。

## 二、小鼠复杂性状在行为遗传学研究中的应用

小鼠行为学遗传机制研究现正处于一个历史性的转折阶段。很快，人们就会探索出新的研究方法，并将其应用于小鼠体内以对小鼠行为学相关的所有基因、基因网络、mRNA和蛋白质等等进行研究。有了这些新技术，科研人员就可以进行更为深入的小鼠行为学研究，从而获取更多、更详细的资料。不过，要能够很好地分析这些资料，还必须将传统的以基因为基础的研究方法转变成网络式的研究方法。

早在分子生物学技术高速发展的20世纪90年代，研究人员就已经开始对小鼠行为学的遗传背景机制进行研究了。当时，他们主要采取两种研究策略 (表1)。

将以上两种策略结合在一起，就形成了一个发现与小鼠行为学相关基因的新方法。这种新策略首先对近交系杂交小鼠 (crosses between inbred mouse strains) 进行基因关联作图 (即上述第二种策略)，以确定影响特定行为表型的基因的大致位置。随后，进一步进行遗传学以