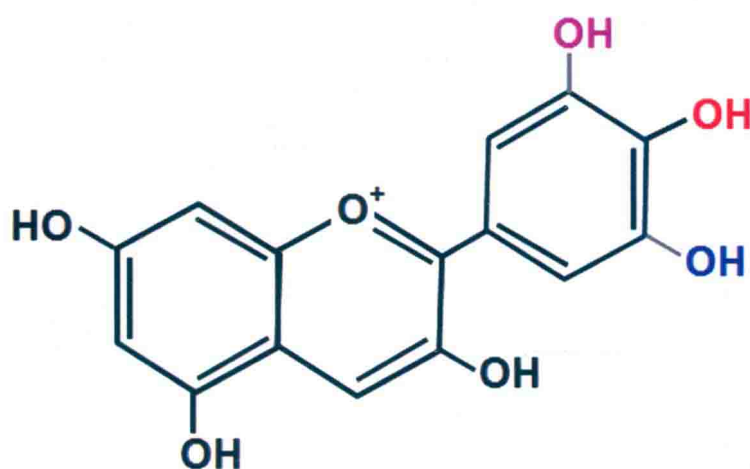


海南省自然科学基金项目 (No.317012)

国家自然科学基金地区科学基金项目(No.31660074)

花青素代谢途径调控复合体及其顺式元件的分析

◎ 祝志欣 著



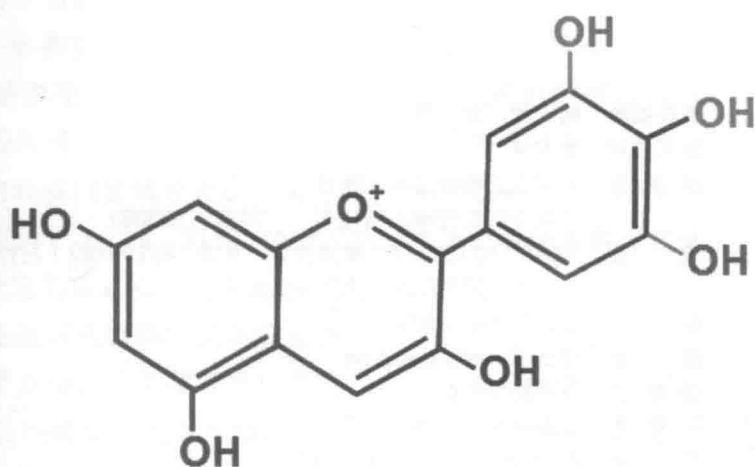
中国农业科学技术出版社

海南省自然科学基金项目 (No.317012)

国家自然科学基金地区科学基金项目(No.31660074)

花青素代谢途径调控复合体及其顺式元件的分析

◎ 祝志欣 著



中国农业科学技术出版社

图书在版编目 (CIP) 数据

花青素代谢途径调控复合体及其顺式元件的分析 / 祝志欣著. —北京:
中国农业科学技术出版社, 2018. 12

ISBN 978-7-5116-3936-3

I. ①花… II. ①祝… III. ①花青素—中间代谢—研究 IV. ①Q942.6

中国版本图书馆 CIP 数据核字 (2018) 第 287379 号

责任编辑 褚 怡 李 华
责任校对 李向荣
出 版 者 中国农业科学技术出版社
北京市中关村南大街12号 邮编: 100081
电 话 (010) 82109708 (编辑室) (010) 82109702 (发行部)
(010) 82109709 (读者服务部)
传 真 (010) 82106650
网 址 <http://www.castp.cn>
经 销 者 各地新华书店
印 刷 者 北京建宏印刷有限公司
开 本 710mm × 1 000mm 1/16
印 张 9.25 彩插12面
字 数 164千字
版 次 2018年12月第1版 2018年12月第1次印刷
定 价 58.00元

前 言

花青素 (anthocyanin) 是大部分被子植物花朵颜色的决定性色素, 可使花朵呈现出粉、红、紫、蓝等一系列颜色。同时, 它与叶绿素 (chlorophyll) 和类胡萝卜素 (carotenoid) 配合, 也可以使植物的果实和叶片呈现出五彩斑斓的着色模式。

植物美丽的色彩不仅有观赏价值, 更对植物的生活有直接的作用。鲜艳的花朵和果实可以帮助植物吸引传粉者和果实传播者, 有利于植物的繁殖。红色的叶片则可以保护植物对抗紫外线、病虫害以及食草动物的进食等。对人类而言, 花青素是一种天然强效的自由基清除剂, 具有抗氧化、抗突变、预防心脑血管疾病、保护肝脏、抑制肿瘤等多种保健功能。在心血管疾病和癌症频发的现代社会, 各种富含花青素的食物品种也受到人们的青睐。

花青素属于植物的次生代谢产物, 由一系列结构基因编码的酶催化合成。这一系列结构基因构成了花青素合成途径, 是被子植物中较为保守的一个途径, 可以被众多的环境因素诱导表达。多个物种的研究表明, 由MYB、bHLH和WDR三类调控因子 (TFs) 组成的MBW复合体调控着花青素代谢途径上众多结构基因, 让它们得以协同地表达。

基因的表达调控是基因功能得以实现的重要过程, 基因间的协同表达是基因组整体功能的重要方面。关于花青素代谢途径中结构基因的表达和调控基因的功能已有较广泛的研究, 但对于MBW复合体如何协同调控整个途径众多基因的表达还不太清楚。以上疑问的解决, 关键在于结构基因和调控基因之间的联结纽带——顺式调控元件。深入比较花青素代谢途径上众多靶基因对应于MBW复合体的顺式调控元件, 对于人们从整体的角度了解基因组内代谢途径的调控具有重要的意义。

本书以花色模式植物圆叶牵牛 (*Ipomoea purpurea*) 的花青素代谢途径

为主要研究对象,深入解析其MBW调控复合体在相关靶基因启动子上对应的顺式元件,直观展现了花青素代谢途径多个基因启动子上顺式元件的排列组合,这些顺式元件的分布差异直接影响了花青素代谢途径上基因的表达水平。随后本研究将考察对象扩展到其他植物类群,试图从转录因子和顺式调控元件两个方面探寻花青素代谢途径在表达调控上的特点及可能存在的进化信号。

对于调控花青素代谢途径的MBW复合体,研究发现,MYB的表达是花青素积累的关键:①MBW复合体的靶基因众多。一方面包括了复合体本身的两个成员**bHLH**和**WDR**,另一方面也包括了途径上的几乎所有核心结构基因,以及下游的部分修饰、转运基因如**3GGT**等。②MYB单独对以上靶基因具有弱激活功能,使得植物在信号诱导下产生特异的MYB后,可以反馈性地完成复合体中另两个成员**bHLH**和**WDR**的积累。③由MYB、**bHLH**和**WDR**形成的MBW复合体对靶基因的激活能力相对于单独MYB有大幅的提高,可以有效地开启花青素合成、修饰和转运相关基因的表达。

对于MBW复合体中MYB和**bHLH**对应的顺式元件,研究发现它们均有较大的变异范围,分别总结为MRE [MYB识别序列, **ANCNN (C/A) C**] 和BRE [**bHLH**识别序列, **CACNN (G/T)**]。这可以允许MBW复合体识别多个靶基因。此外,MRE和BRE在靶基因启动子上的组合排列可以直接影响该靶基因的转录激活水平。

紫玉兰 (*Magnolia liliflora*)、玉米 (*Zea mays*)、拟南芥 (*Arabidopsis thaliana*)、非洲菊 (*Gerbera hybrida*) 和圆叶牵牛代表了被子植物主要类群的物种。对它们MBW调控复合体成员的比较表明,花青素代谢途径并非是普遍保守的。在被子植物内,不论是调控因子的调节强度还是启动子的反应性都经历着缓慢而深刻的进化。就代谢途径上的各基因而言,启动子的不同可以引起基因间表达的差异,这种差异大于调控因子分化带来的影响;对物种间同源的结构基因而言,转录因子间的分化则是基因表达强度变化的一个重要因素。

本书在多年课题研究的基础上总结撰写而成,在撰写过程中,得到了鲁迎青研究员的悉心指导;肖大洲老师对本书的出版也做了不少辅助工作,在此一并表示感谢!本书引用了大量的相关文献,其中大多数已在书中注明出处,但难免有所疏漏。在此,向有关作者和专家表示感谢,并对没有注明出

处的作者表示歉意。

本书的出版由海南省自然科学基金项目（No.317012）、国家自然科学基金项目（No.31660074）、海南大学科研启动基金（No.kyqd1840）共同资助完成，在此表示诚挚的感谢。

本书对花青素代谢途径进行了直观的描述，研究结果对认识花青素代谢途径的调控具有良好的促进作用。对植物花色、叶色和果色感兴趣的植物爱好者、观赏植物的育种工作者以及研究基因调控的科研工作者，本书可作为参考书籍。

尽管本书在撰写过程中力求逻辑严谨，内容充实，但由于著者水平有限，难免有所疏漏，恳请各位专家、同仁和广大的读者批评指正，以便今后完善。

著者

2018年10月于海南大学

目 录

1	绪 论	1
1.1	基因的转录调控	1
1.2	花青素代谢途径及其调控	11
1.3	花青素代谢途径改变导致的颜色变异	22
1.4	花青素代谢途径调控研究中存在的问题和本研究的目的	27
2	圆叶牵牛花青素代谢途径基因启动子克隆及序列分析	29
2.1	引 言	29
2.2	材料和方法	29
2.3	结 果	34
2.4	讨 论	44
3	圆叶牵牛花青素代谢途径MBW调控复合体的功能研究 及其靶基因的确定	46
3.1	引 言	46
3.2	材料和方法	47
3.3	结 果	56
3.4	讨 论	61
4	圆叶牵牛花青素代谢途径基因启动子MBW复合体 相关顺式元件的解析	64
4.1	引 言	64
4.2	材料和方法	64
4.3	结 果	68
4.4	讨 论	87

5	被子植物MBW复合体的进化初探	90
5.1	引 言	90
5.2	材料和方法	92
5.3	结 果	96
5.4	讨 论	102
6	结论与展望	105
6.1	结 论	105
6.2	展 望	108
	附 录	110
	参考文献	121

1 绪 论

1.1 基因的转录调控

基因的表达调控决定了生物体内各基因的表达时间、空间和丰度，是影响生物多样性的的重要因素。基因调控网络的机制复杂，可以分为表观遗传水平、转录水平、mRNA加工和成熟水平、翻译水平及翻译后水平等多个层次（Wray et al, 2003; Chen and Rajewsky, 2007; Shi et al, 2015）。其中，转录处于这些调控层次的上层，是基因表达所必需的，在这一水平上的调控也相对经济有效。转录调控是基因表达调控的重要方面（Wray et al, 2003; Wray, 2007; Watson et al, 2013）。转录后调控普遍存在于动物基因中（Filipowicz et al, 2008），但大部分植物基因主要还是依赖于转录水平的调控（Huang et al, 2007）。全面认识植物基因的转录调控有利于我们揭示植物体复杂生命活动的秘密，从而更合理地利用植物资源。

1.1.1 基因转录调控的研究概况

基因的转录调控主要由两方面的因素决定：转录因子（也称为反式调控因子，*trans factors*）和顺式元件（*cis elements*）。顺式元件是具有特定序列的DNA片段，常常为几个到十几个碱基，散布于基因的启动子上，它们与转录因子的互动可以活化RNA聚合酶，招募转录复合体，决定相应基因表达的时间、空间以及强度（Wray et al, 2003）。

根据特异性的不同，转录因子和顺式元件可以被简单地分为通用型和特异型两大类。通用型转录因子即转录复合体中的成员，是转录所必需的，在同一类型的基因间（如蛋白质编码基因，rRNA编码基因等）不具特异性，

它们所结合的DNA序列即通用顺式元件，如TATA框（TATA box）。特异转录因子即常规意义上的转录因子，决定了每个基因不同于其他基因的表达方式，而它们所结合的DNA序列即特异型顺式元件（Watson et al, 2013）。在AGRIS（*Arabidopsis* Gene Regulatory Information Server, <http://arabidopsis.med.ohio-state.edu>）数据库中，拟南芥基因组中已知的1 500多个特异转录因子被分为64个家族，每个家族都对应着各自的特异型顺式元件。其中，WRKY、MYB和MADS等家族呈现了植物特异的大规模扩增，是植物适应性和多样性的重要原因（Riechmann et al, 2000）。

在同一生物体或物种中，顺式元件的存在可以说是恒定的，主要通过染色体的结构变动和碱基修饰来调节其显露与否（Jaenisch and Bird, 2003；Fraser and Bickmore, 2007；Shi et al, 2015）；而反式调控因子则根据其所处的发育阶段以及外界环境信号而呈现动态变化。对于机体的同一个细胞或组织，基因组内的所有基因均处于同样的转录因子环境下，各基因启动子上的顺式元件排列组合的差异是它们差异表达的重要原因（Gompel et al, 2005；Ni et al, 2009）。转录因子与顺式元件一起决定了基因组内基因的表达状况，是生物进化的重要方面（Hoekstra and Coyne, 2007；Wray, 2007）。

细胞内的所有过程都由多个基因的表达产物协同完成，因此基因组中很多基因的表达都是被协同调控的，如位于同一个代谢途径上的多个结构基因。多基因间的协同调控在原核生物和真核生物之间存在着相当大的差异。研究基因组内功能相关的多个基因如何协同表达是我们从整体角度认识基因组转录调控的一个良好切入点。

1.1.1.1 原核生物基因的协同调控

1961年，法国巴斯德研究所著名的科学家Jacob和Monod在解析大肠杆菌乳糖操纵子的基础上，发表了《蛋白质合成中的遗传调节机制》一文，提出操纵子学说，开创了基因调控的研究（Jacob and Monod, 1961）。该学说现已成为原核生物基因调控的主要理论之一。

原核生物属于单细胞生物，每个细胞都直接与外界环境直接接触，基因间的协同表达常常依靠操纵子（operon）来进行（Watson et al, 2013）。所谓操纵子，是指功能上相关的多个结构基因前后相连成串，由一个共同

的启动子进行转录调控,实现这些基因间的协同表达(Jacob and Monod, 1961; Watson et al, 2013)。如大肠杆菌乳糖代谢调控需要3种酶参加:①LacZ,即 β -半乳糖酶,可将乳糖分解成半乳糖和葡萄糖;②LacY,即渗透酶,可以增加糖的渗透,易于摄取乳糖和半乳糖;③LacA,转乙酰酶,将 β -半乳糖转变成乙酰半乳糖。当环境中没有葡萄糖而存在大量乳糖时,LacZ、LacY、LacA的数量急剧增加,几分钟可达到千倍以上;乳糖消耗完时,这3种酶的合成也同时停止(Jacob and Monod, 1961)。如此良好的协同性来自它们同属一个操纵子,受到同样顺式元件和转录因子的调控(彩插图1-1A)。原核生物其他的操纵子如阿拉伯糖操纵子、组氨酸操纵子、色氨酸操纵子等(Watson et al, 2013)。

1.1.1.2 真核生物基因的协同调控

不同于原核基因,真核生物的基因大都被独立调控。一个基因对应于一个启动子,而每个启动子常常拥有多种顺式元件,用于应对不同的诱导因素(Watson et al, 2013)。这样对于真核生物应该说是具有适应性的,可以使同一种基因产物被应用到不同的条件中,丰富了基因间的搭配使用,如MADS类蛋白的搭配使用而决定花器官的形成(Theissen et al, 2000)。同时,各个基因在表达量上的精确调控也有赖于其启动子中顺式元件的独特组合排列(Pearson et al, 2005)。

同源异型基因(*Hox genes*)调控着动物的形体,其编码蛋白的结构和功能在动物界高度保守,它们的突变往往会导致身体的畸形,如果蝇附肢的增加、减少或错位(Wittkopp and Kalay, 2012)。*Hox*基因在染色体上通常成簇存在,呈共线性分布,其排列顺序直接对应于其作用顺序和作用位置(Carroll, 1995; 2008)。这种共线性表现出了真核生物在关键基因的协同表达上还是部分地沿用了原核生物的策略。但*Hox*基因不同成员具有各自独立的启动子,使它们的表达相互独立,以决定不同器官的发生。这个过程依赖于*Hox*基因启动子上复杂的顺式元件组合。顺式元件层面上的变化被认为是*Hox*基因进化的重要原因(Wittkopp and Kalay, 2012)。

近年来,随着全基因组水平工作的增加,发现植物中一些次生代谢基因也具有成簇分布的现象,这些基因一般通过主开关基因直接调控它们的协同表达(Field and Osbourn, 2008)。如黄瓜6号染色体存在一个由5个

基因组成的基因簇，可能参与苦味物质葫芦素的合成（彩插图1-1B）。开关基因*Bt*协同调控这些基因的表达，并决定了黄瓜植株中叶片和果实的苦味（Shang et al, 2014）。这样的成簇分布似乎有利于保持这些基因在表观遗传修饰水平以及转录调控上的一致性，便于它们的协同表达（Nutzmann and Osbourn, 2015）。

真核生物其他需要协同表达的基因常常散布于基因组中，如拟南芥类黄酮代谢途径上的一系列基因（彩插图1-1C）（数据统计自拟南芥信息资源中心，The *Arabidopsis* Information Resource, TAIR）。目前，关于植物是如何协同表达这些基因的研究还相当稀少，而这个问题的解答非常有助于人们从全局的角度认识植物基因的转录调控，揭示植物体复杂生命活动的秘密。

1.1.2 启动子的结构

启动子是基因的重要组成部分，直接对启动子进行功能研究是揭示其转录调控的最直接手段。但相对于基因编码区，启动子序列并不显现于最后的功能蛋白质或RNA中。其中的顺式元件序列较短，分散分布，很难通过序列保守性或分子钟等方法对其进行研究。长期以来，启动子的结构特点给研究带来了很大障碍，对启动子的研究远远落后于蛋白质或RNA的研究。虽然如此，大部分启动子还是具有以下结构上的共性。

启动子常常位于基因5'端上游，根据功能分成两部分：一部分包含了通用顺式元件，与通用转录因子结合，被称为核心启动子区，在位置上一般与基因的转录区紧紧相连；另一部分则与特异转录因子结合，决定基因的特异表达，一般位于核心启动子区5'上游，其上的特异顺式元件在分布上有近有远，使得启动子在长度上没有明确的界限（Gurley et al, 2006）。近些年的研究表明，特异顺式元件也可以位于5'UTR及内含子（Andres and Coupland, 2012），或者很远处的增强子或沉默子中（Wray et al, 2003）。

从结构上，广义的启动子常常包括了以下基本元素。

（1）转录起始位点。转录起始位点是RNA聚合酶结合模板链并起始转录的位点，对应于新生RNA链的第一个核苷酸，此位点通常为嘌呤（A或G），并以腺嘌呤A为主（Bucher, 1990）。核心启动子区一般位于转录起始点前后40bp的范围内，含有像TATA框之类的通用顺式元件（Wray et al,

2003; Watson et al, 2013)。

(2) 通用顺式元件。TATA框是最普遍的通用顺式元件, 可被RNA聚合酶所识别, 一般与转录起始点之间有特定的距离, 以保证转录的精确起始。原核生物中, 转录起始位点上游-10nt左右一般含有TATA框(TATAAT), 在-35bp左右存在TTGACA区, 两者最佳距离为17bp, 用于结合RNA转录复合体的起始亚基(Watson et al, 2013)。真核生物中, TATA框常位于转录起始位点上游-32bp $\pm$ 7bp处(Lodish et al, 2004)(彩插图1-2)。植物TATA框保守核心序列为TATA(A/T)A(A/T), 是大多数植物启动子正确表达所必需的, 其序列的变化反映了TATA框在效率上的差异(Lee and Young, 2000)。对于没有TATA框的基因, 其表达过程中可能存在着某些替代机制, 涉及特异转录因子与启动子中其他顺式元件的相互作用, 如Inr元件(initiator, 起始子)位置与转录起始点重叠, 在一些基因中代替TATA框起始转录(Ohler and Niemann, 2001)。而含有多个TATA框的基因则可能对应着可变的转录起始点(Goodyer et al, 2001)。

CAAT框(CAAT box)是真核生物中另一个常见的通用顺式元件, 其一致序列为(T/C)(A/G)(A/G)CCAATCA(Bucher, 1990), 位于转录起始点上游约-75bp处, 可以提高基因转录的效率, 但似乎与启动子的特异性无关, 一般不存在于看家基因中(Romier et al, 2003)。

虽然通用顺式元件存在于很大一部分比例的启动子中, 但也有许多启动子上缺乏这些通用顺式元件, 特别是一些表达量总是很低的基因(Wray et al, 2003; Gurley et al, 2006), 不过这些通用顺式元件的信息还是可以为人们分析启动子提供基本的参考。

(3) 特异顺式元件。特异顺式元件即被特异转录因子结合的DNA序列, 是基因实现特异表达的基础。转录因子根据其结合DNA的结构域被分类, 如MYB家族即含有MYB结构域的蛋白, 而bHLH家族成员均含有bHLH结构域(Riechmann et al, 2000)。同一家族的转录因子通常具有类似的DNA结合序列, 最著名的特异顺式元件如G框(G-box), 其序列为5'-CACGTG-3', 是高度保守的特异顺式元件之一, 可以被bHLH家族的转录因子所识别(Li et al, 2006)。而bZIP家族的识别序列含有5'-ACGT-3'核心序列, 与G框的序列部分重叠(Franco-Zorrilla et al, 2014)。考虑到定义

以及研究上的方便, 狭义的启动子仅指位于转录起始点5'端的包含特异顺式元件的DNA片段, 又称为近端启动子 (proximal promoter)。

另外, 许多真核基因具有超远距离的顺式元件, 超出了狭义启动子的范围, 如增强子和沉默子, 它们可以与特异转录因子结合, 增强或抑制基因的转录。与近端启动子上的顺式元件相比, 它们需要通过DNA弯曲 (DNA looping) 来实现对转录的调控, 这导致了它们在调控上的几大特点: ①无方向性 (orientation-independent), 即不管位于基因的上游或下游, 均能影响相应启动子。②无距离性 (position-independent), 即它们对基因的作用, 与两者间的距离关系不大 (Dyanan, 1989; Wray et al, 2003)。增强子和沉默子的这些特性虽然方便了基因组内的基因表达调控, 但不利于对它们的确认和功能研究。

1.1.3 启动子研究所面临的问题

启动子上顺式元件的组成和排列在很大程度上决定了基因的表达谱。作为基因组中功能如此重要的序列类型, 大量研究者致力于利用数据库与生物信息学手段对其中的基序 (motif) 进行预测, 但各类基序在活体真核细胞中的功能仍然需要由分子生物学试验手段来验证其真实和准确性。理论上说, 解析基因的启动子就能解析基因的表达, 但在实际的启动子试验性研究过程中, 很多困难摆在人们面前。

首先, 启动子边界难以界定, 其序列扩增相对于基因编码区难度较大。近些年基因组测序技术的发展, 使很多模式物种在启动子扩增上的难度大大降低, 但对于没有参考基因组序列的物种, 启动子扩增仍然比较困难。

其次, 在得到了基因的启动子序列后, 鉴定其中真正的转录因子结合位点并不容易。一个启动子上一般有多个顺式元件, 它们序列较短、散在而不均匀地分布 (Wray et al, 2003), 加大了对它们进行试验鉴定的难度。启动子上顺式元件的种类和数目因基因而异。通过对一些研究的较好的启动子的总结, 可以看出很多真核基因启动子上有10~50个顺式元件, 对应于5~15种不同的转录因子 (Gibson and Honeycutt, 2002)。

转录因子的结合序列也常常具有可变性, 可以容忍一个到多个的碱基变异, 而不严重影响其结合 (Wang et al, 2013), 它们对应的顺式元件常以位置权重矩阵 (position weight matrix) 的形式被呈现出来。另一方面, 序

列合适的位点也会因为各种原因在活体中并没有功能,如表观遗传水平的修饰、位置的不合适以及其他未知的原因(Shi et al, 2015)。启动子上顺式元件以外的序列部分,常常因为碱基组成影响DNA的构象,并最终影响转录的效果(Wray et al, 2003)。即使在研究的很好的启动子上,也很难确定其中所有的顺式元件均被找出。对于从序列上推断的疑似结合位点,一般需要试验数据进行功能上的验证。

随着功能基因组研究的深入,对启动子的研究和解析越来越显示出必要性。相信随着分子生物学、生物信息学等学科的发展,更多的启动子序列会得到分析,更多的顺式元件的功能会逐渐明确,并加速人们对各种特异顺式元件的应用,从而有力地推动基因工程的发展。

1.1.4 PCR介导的启动子克隆方法

启动子作为基因的重要部分,其序列并不显现于mRNA或蛋白质中,需从基因组中获得。对启动子的克隆常常面临只知道基因的编码区,而对5'上游序列完全未知的情况。克隆启动子的方法大致分为两类:一类是利用探针从基因组文库中进行杂交得到;另一种是利用特殊的PCR技术克隆得到。由于文库杂交法试验难度大,而PCR介导的方法简便快捷,是获得启动子序列的重要方法(Tonooka and Fujishima, 2009)。

PCR介导的启动子克隆主要包括常规PCR和各种染色体步移PCR法。对于目标基因,首先利用其已知序列在NCBI等数据库上进行比对,查看近缘种的同源基因编码区上游序列的扩增情况,分析已扩增序列中的保守区,并尽量针对保守区设计引物。常规PCR法主要针对有参考序列的基因,如已被全基因组测序的物种的基因或近缘种同源基因的启动子序列已被克隆的基因。当常规PCR方法不适用时,可选用其他各种PCR介导的染色体步移法。

所谓染色体步移法,即由基因组中已知序列出发,逐步探知其旁侧序列的方法,如目前常用的TAIL PCR, inverse PCR, cassette PCR和two-step PCR等(Tonooka and Fujishima, 2009)。在这些PCR中,只有一个特异的PCR引物,并主要通过两类办法解决另一头引物的问题,一类是通过使用随机简并引物(如TAIL PCR和two-step PCR),另一类则在另一头引入已知序列(如inverse PCR和cassette PCR)。

1.1.4.1 TAIL-PCR (Thermal Asymmetric Inter Laced PCR, 热不对称交错式PCR)

TAIL-PCR是1995年由Liu和Whittier (1995) 提出, 该技术通过随机引物和3个嵌套的特异引物组合进行3轮连续的PCR。利用不同的退火温度选择性地扩增目标片段, 其基本原理及过程如图1-1所示。在TAIL-PCR中, 简并引物的序列会直接影响最后的扩增效率, 不同的物种所适用的简并引物因基因组的碱基序列组成结构的差异而有所不同。

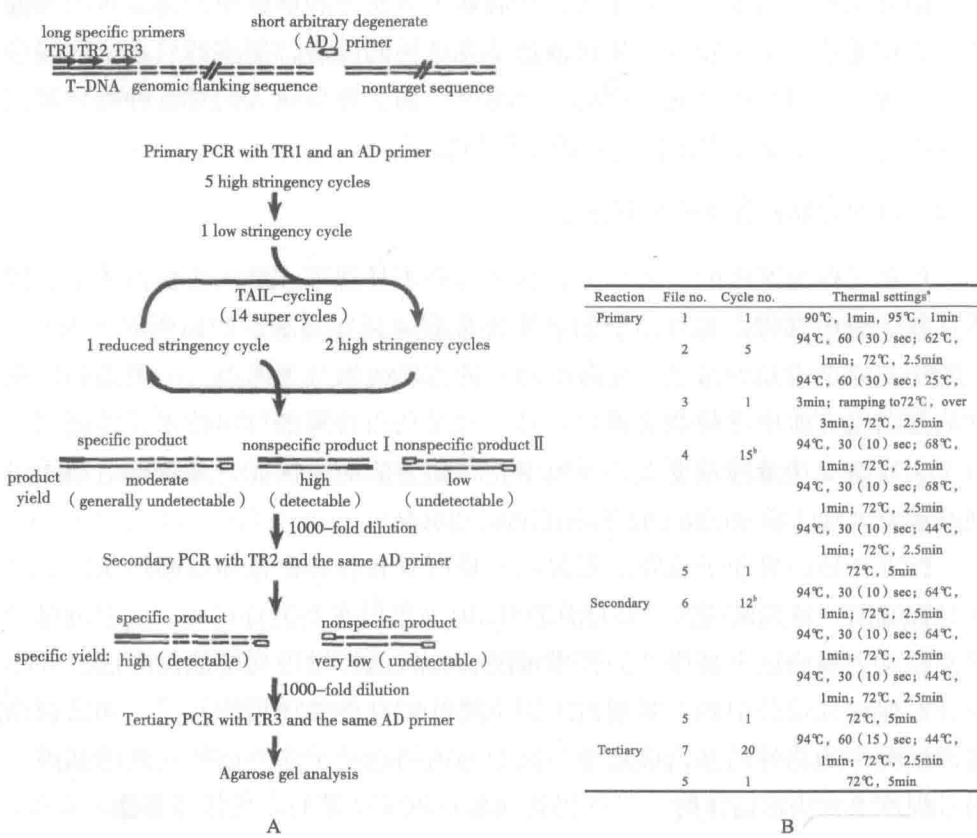


图1-1 TAIL-PCR的流程原理图及PCR程序 (Liu and Whittier, 1995)

Figure 1-1 Procedures of TAIL-PCR (Liu and Whittier, 1995)

A.TAIL-PCR三轮PCR的流程原理图; B.PCR程序设定

第一轮扩增中, 利用高温62℃退火线性扩增5个循环, 低温25℃退火扩增1个循环, 再进行2高 (68℃) 1低 (44℃) 退火扩增12个超级循环。第二轮套式扩增中, 将第一轮产物稀释500~1 000倍作为模板, 进行2高 (64℃) 1低 (44℃) 退火扩增10~14个超级循环。第三轮套式扩增为可选步骤, 将第二轮产物稀释500~1 000倍作为模板, 进行低温44℃退火扩增20个循环, 主要是用来富集特异的目标片段

TAIL-PCR与普通PCR相比有以下特点：①利用简并引物在靶序列上创造结合位点；②利用嵌套式PCR提高扩增产物特异性；③利用两轮高温、一轮低温交替的超级PCR循环提高特异产物的扩增，降低非特异产物的比例；④模板高倍数稀释以降低非特异产物在下一轮作模板的比例；⑤通过电泳检测第二轮和第三轮产物相差的大小来判断产物的特异性（Liu and Chen, 2007）。

1.1.4.2 两步基因步移法（Two-step gene walking）

两步基因步移法是2007年由Pilhofer等人提出，它利用已知序列设计2个特异引物：基因步移引物和测序引物。整个过程分为两步：第一步，利用基因步移引物进行特殊的PCR扩增程序；第二步，对之前得到的PCR产物直接利用测序引物进行测序。这个方法力图最大化的使用特异引物，而降低随机引物带来的随机扩增，其具体过程如图1-2所示。

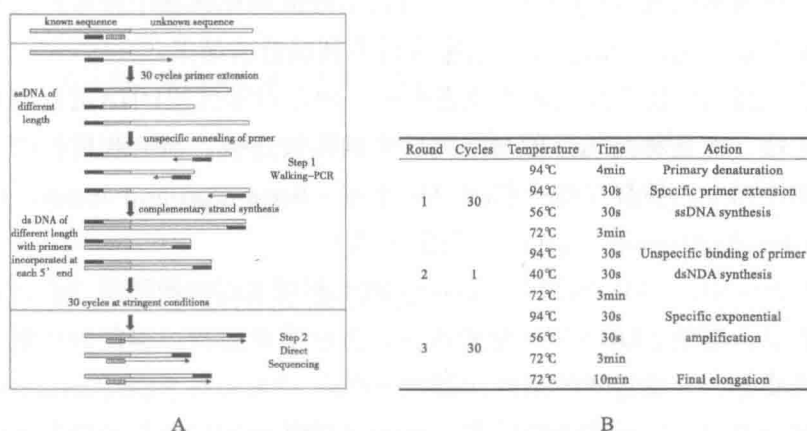


图1-2 两步基因步移法的流程原理图及PCR程序（Pilhofer et al, 2007）

Figure 1-2 Procedures of two-step gene walking（Pilhofer et al, 2007）

A. 两步基因步移法的流程原理图。已知的序列部分用灰色长条显示，白色长条代表未知序列。利用已知序列设计两个特异引物，基因步移引物如黑色短线所示，测序引物如斜纹短线所示。B. PCR程序设定。第一步的PCR扩增只加入基因步移引物，共61个循环。前30个循环退火温度较高（如56℃），为单方向的扩增，此引物会产生一组长度不等的单链DNA（由DNA聚合酶从模板分离的位置决定）；之后进行一个低退火温度（如40℃）的循环，此引物可在3'端随机结合使互补链得以扩增；随后再次进行30个循环的高退火温度（如56℃）反应，得到5'端相同，而3'端长度不同的双链DNA混合物。第二步，用内侧的测序引物直接对之前得到的长度不同的双链DNA混合物进行测序