

不同物种 *TYRP2* 基因完整编码区生物信息学分析

刘伟兰, 李祥龙*, 周荣艳, 李兰会, 杨清芳
(河北农业大学 动物科技学院, 河北 保定 071001)

摘要: 应用生物信息学和比较基因组学方法, 比较分析了人、毛猩猩、绵羊、牛、猪、马、犬、斑马鱼、大熊猫、小家鼠、褐家鼠、鸭嘴兽、非洲爪蟾、原鸡、绿头鸭、斑胸草雀、黑猩猩、恒河猕猴、狢、灰短尾负鼠和兔 21 个物种 *TYRP2* 基因完整编码区(CDS), 并对其遗传多样性、分子系统发育、编码蛋白的氨基酸序列、信号肽、跨膜结构域、疏水性/亲水性、二级结构等进行了预测。结果表明, 从 21 个物种的 42 条基因序列检测到 741 个多态位点, 共生成 26 种单倍型。物种间 *TYRP2* 基因序列编码区存在较丰富的遗传多样性, 物种内该基因序列编码区则存在较强的保守性。*TYRP2* 基因密码子有较强的偏爱性。根据核苷酸歧异度和遗传分化系数对物种的亲缘关系进行分析, 结果表明, 人与黑猩猩亲缘关系最近, 小家鼠和褐家鼠亲缘关系最近。*TYRP2* 基因多肽存在信号肽, 有跨膜结构域, 表现为疏水性, 二级结构的主要组成元件为无规卷曲和 α 螺旋。

关键词: *TYRP2* 基因; 遗传多样性; 系统发育; 多肽信息; 动物毛色

中图分类号: Q78 **文献标识码:** A **文章编号:** 1004-3268(2011)10-0144-05

Bioinformatics Analysis of Complete Coding Regions of *TYRP2* Gene among 21 Species

LIU Wei lan, LI Xiang long*, ZHOU Rong yan, LI Lan hui, YANG Qing fang
(College of Animal Science and Technology, Agricultural University of Hebei, Baoding 071001, China)

Abstract: In this study, the CDS of the *TYRP2* gene sequences of *Homo sapiens*, *Pongo abelii*, *Ovis aries*, *Bos taurus*, *Sus scrofa*, *Equus caballus*, *Canis familiaris*, *Danio rerio*, *Ailuropoda melanoleuca*, *Mus musculus*, *Rattus norvegicus*, *Ornithorhynchus anatinus*, *Xenopus*, *Gallus gallus*, *Anas platyrhynchos*, *Taeniopygia guttata*, *Pan troglodytes*, *Macaca mulatta*, *Callithrix jacchus*, *Monodelphis domestica* and *Oryctolagus cuniculus* was analyzed using the method of bioinformatics. The genetic diversity, amino acid sequences, signal peptide, trans membrane topological structure, hydrophobicity or hydrophilicity, and secondary structure were also analyzed. The results showed that a total of 741 polymorphic sites were detected from 42 sequences of 21 species, from which 26 haplotypes were sorted. There was rich genetic diversity of the CDS of the *TYRP2* gene among species, but it was conservative within species. The results also showed that *TYRP2* was trans membrane protein with signal peptide, and the polypeptide presented hydrophobicity, the main motifs of predicted secondary of *TYRP2* were random coil and alpha helix.

Key words: *TYRP2* gene; Genetic diversity; Phylogenesis; Polypeptide information; Animal coat color

黑色素是不溶于水和有机溶剂的无定形小颗粒, 主要存在于人和高等动物的皮肤、毛发和眼球组织中, 它作为一种突出的叫噪多聚物, 是脊椎动物表

层结构中主要的色素^[1-4]。黑色素的合成过程是一个由酪氨酸酶(TYR)催化体内酪氨酸(Tyr)羟化而启动的一系列生化反应的过程^[5]。酪氨酸酶是黑色

收稿日期: 2011-04-10

作者简介: 刘伟兰(1986), 女, 河北无极人, 在读硕士研究生, 研究方向: 遗传学。E-mail: liuweilan123@126.com

*通讯作者: 李祥龙(1963), 男, 河北丰南人, 教授, 博士, 博士生导师, 主要从事动物遗传育种研究。

E-mail: lixianglongcn@yahoo.com

©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

素合成过程中的关键酶,酪氨酸酶相关蛋白 1 (TYRP1)和酪氨酸酶相关蛋白 2 (TYRP2)各自催化特异的反应,三者均参与黑色素形成所涉及的多步生化反应途径,共同构成了“酪氨酸酶蛋白家族”^[1]。作为“酪氨酸酶蛋白家族”的重要成员,TYRP2 于 1992 年被确认为酪氨酸酶相关蛋白家族的第 3 个成员。其具有多巴色素异构酶(DT)活性,可催化多巴色素转变为 5,6 二羟基吲哚羧酸(DHICA),具有加速黑色素生成的作用^[5]。同时,黑色素细胞中的 TYRP2 控制着 5,6 二羟基吲哚羧酸与 5,6 二羟基吲哚的比例,是影响动物毛色的重要蛋白,推测 TYRP2 可能对黑色素合成早期阶段的酪氨酸酶催化活性有调节作用^[5]。鉴此,比较了 *TYRP2* 基因完整编码区(CDS)在 21 个物种间的差异,以期

为研究动物毛色的分子遗传机制和动物育种工作提供帮助。

1 材料和方法

1.1 *TYRP2* 基因序列来源

在 NCBI 网站的 GenBank 中寻找 *TYRP2* 的核苷酸序列,选择保守性较强的绵羊 *TYRP2* 基因的完整编码区(CDS)(EU760773.1)进行 BLAST 分析。下载相似度较高的(一般为 80%以上)基因的 CDS 及其对应的蛋白质序列。主要下载了人、毛猩猩、绵羊、牛、猪、马、狗、斑马鱼、大熊猫、小家鼠、褐家鼠、鸭嘴兽、非洲爪蟾、原鸡、绿头鸭等 21 个物种 42 条 *TYRP2* 基因的 CDS 及其对应的蛋白质序列进行分析,具体见表 1。

表 1 不同物种的 *TYRP2* 序列来源

物种	CDS 序列	蛋白质序列
绵羊(<i>Ovis aries</i>)	NM_001130024.1、EU760773.1	NP_001123496.1、ACF21683.1
牛(<i>Bos taurus</i>)	NM_001012666.1	NP_001012684.1
猪(<i>Sus scrofa</i>)	NM_001025227.1	NP_001020398.1
马(<i>Equus caballus</i>)	XM_001491619.2	XP_001491669.2
犬(<i>Canis familiaris</i>)	XM_542639.2	XP_542639.2
豺(<i>Callithrix jacchus</i>)	XM_002742475.1、XM_002742476.1	XP_002742475.1、XP_002742476.1
大熊猫(<i>Ailuropoda melanoleuca</i>)	XM_002918642.1	XP_002918688.1
恒河猕猴(<i>Macaca mulatta</i>)	XM_001083129.2、XM_002800798.1	XP_001083129.1、XP_002800844.1
毛猩猩(<i>Pongo abelii</i>)	XM_002824362.1、XM_002824361.1、NM_001922.3、BC028311.1、D17547.1、L18967.1	XP_002824408.1、XP_002824407.1、ABM82392.1、ABM85575.1、AAA20870.1、BAF85804.1
人(<i>Homo sapiens</i>)	DQ894649.2、S69231.1、AK293115.1、NM_001129889.1、DQ902581.1、A1139318.9、DQ891466.2	CAC19460.1、AAH28311.1、BAA04484.1、ABI73976.1、NP_001913.1、NP_001123361.1、AAC60627.1
兔(<i>Oryctolagus cuniculus</i>)	XM_002713001.1、NM_010024.3、BC067064.1、EU554632.1、BC082330.1、X63349.1	XP_002713047.1、AAH67064.1、AAH82330.1、ACB56630.1、NP_034154.1、CAA44951.1
褐家鼠(<i>Rattus norvegicus</i>)	XM_001078239.1、XM_224517.4	XP_224517.4、XP_002713047.1
灰短尾负鼠(<i>Monodelphis domestica</i>)	XM_001377599.1	XP_001377636.1
鸭嘴兽(<i>Ornithorhynchus anatinus</i>)	XM_001512063.1	XP_001512113.1
原鸡(<i>Gallus gallus</i>)	NM_204935.1	NP_990266.1
绿头鸭(<i>Anas platyrhynchos</i>)	HM775320.1	ADU15763.1
斑胸草雀(<i>Taeniopygia guttata</i>)	XM_02199349.1	XP_002199385.1
非洲爪蟾(<i>Xenopus</i>)	NM_001017161.2	NP_001017161.1
斑马鱼(<i>Danio rerio</i>)	NM_131555.1、BC164222.1、BC129260.1	AAI29261.1、AAI64222.1、NP_571630.1
黑猩猩(<i>Pan troglodytes</i>)	XR_022754.1	

1.2 方法

首先利用 BioEdit 7.0 软件对来自不同物种的 42 条 *TYRP2* 基因的 CDS 序列进行比对,选取所有序列的共有片段(长度为 1200bp)。然后利用 DnaSP 5.0 软件对其进行遗传多样性分析,生成单倍型,计算平均核苷酸差异度(*k*)、核苷酸歧异度

(*Dxy*)、遗传分化系数(*Gst*)、同义替换位点数(*SS*)、非同义替换位点数(*NSS*)和密码子偏爱性。再利用 MEGA 4.0 软件的 UPGMA 方法进行聚类分析,构建聚类图。对蛋白质的分析,用在线软件对蛋白质序列是否存在信号肽(<http://www.cbs.dtu.dk/services/SignalP/>)、跨膜结构域的位置([http://](http://www.cnki.net)

www.cds.dtu.dk/services/TMHMM)、疏水/亲水性能(<http://us.expasy.org/cgi-bin/protscale>)、二级结构的主要组成元件(http://npsa.pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_hnn.html)进行预测与分析。

2 结果与分析

2.1 不同物种 *TYRP2* 基因核苷酸分析

2.1.1 多态位点、单倍型及其多样性 在所分析的共有片段为 1200bp 的 42 条序列中,发现 741 个多态位点,多态百分率为 61.75%,其中单一多态位点 124 个,多态百分率为 10.33%;简约多态位点 617 个,多态百分率为 51.42%,这些突变位点是否与不同物种毛发颜色的变化有关还需要进一步研究。不同物种内的多态位点数也各不相同,绵羊、狢、恒河猕猴、毛猩猩、人、小家鼠、褐家鼠、斑马鱼的多态位点数分别为 8、0、0、0、3、4、0、10(表 2),百分率分别为 0.67%、0、0、0、0.25%、0.33%、0、0.83%。根据序列比对,在 42 条 *TYRP2* 基因序列中发现 26 种单倍型。各物种的单倍型多样性大小不一,单倍型多样性表明种群间 *TYRP2* 基因遗传变异丰富。全部物种的平均核苷酸差异数(*k*)为 210.589,核苷酸多样性为 0.18233。其中,对含有 2 条以上序列的

物种进行种内的序列分析,绵羊的平均核苷酸差异数和核苷酸多样性最高,其次是斑马鱼、小家鼠和人,其他单倍型数都为 1,表明绵羊的 *TYRP2* 基因存在较丰富的遗传多样性。

表 2 不同物种 *TYRP2* 基因序列分析

物种	序列/条	多态位点 数/个	单倍型 数/个	单倍型 多样性	平均核苷酸 差异数目/个
绵羊	2	8	2	1.000	8.000
狢	2	0	1	0.000	0.000
恒河猕猴	2	0	1	0.000	0.000
毛猩猩	2	0	1	0.000	0.000
人	11	3	2	0.700	1.600
小家鼠	5	4	3	0.182	0.545
褐家鼠	2	0	1	0.000	0.000
斑马鱼	3	10	2	0.667	6.667

2.1.2 物种间核苷酸歧异度和遗传分化 由表 3 可见,含有 2 条以上序列的物种种群间核苷酸歧异度(*D_{xy}*) 在 0.013~0.420,遗传分化系数(*G_{st}*) 在 0.096~1.000。不同物种间核苷酸歧异度和遗传分化系数的变化范围都很大,其中狢、恒河猕猴、毛猩猩的核苷酸歧异度均较小,说明这几个物种的亲缘关系较近,其中绵羊与其他物种间的核苷酸歧异度均较大,这说明绵羊与本研究其他物种的亲缘关系较远。

表 3 不同物种核苷酸歧异度和遗传分化系数

物种	绵羊	狢	恒河猕猴	毛猩猩	人	小家鼠	褐家鼠	斑马鱼
绵羊		0.333	0.333	0.333	0.294	0.099	0.333	0.096
狢	0.145		1.000	1.000	0.617	0.360	1.000	0.447
恒河猕猴	0.151	0.052		1.000	0.617	0.360	1.000	0.447
毛猩猩	0.150	0.050	0.033		0.617	0.360	1.000	0.447
人	0.152	0.052	0.032	0.013		0.415	0.617	0.416
小家鼠	0.189	0.161	0.163	0.165	0.169		0.360	0.183
褐家鼠	0.193	0.170	0.169	0.171	0.171	0.074		0.447
斑马鱼	0.330	0.333	0.338	0.340	0.420	0.332	0.333	

注:上三角为遗传分化系数(*G_{st}*),下三角为核苷酸歧异度(*D_{xy}*)

2.2 不同物种 *TYRP2* 基因核苷酸多样性

2.2.1 密码子偏爱性 在蛋白质翻译过程中,不同物种或同一物种的不同基因在密码子的使用上都存在明显的偏爱性^[6]。所筛选的不同物种 *TYRP2* 基因序列编码区中密码子有效值(*ENC*)为 54.657(<61),偏爱指标(*CBI*)为 0.295(>0),经 χ^2 检验,计算得到未校正的 χ^2 值为 0.230,说明 *TYRP2* 基因对密码子有较强的偏爱性^[7]。目前,对密码子的偏爱性及其在进化中的作用还没有合理的解释,有待今后进一步研究。

2.2.2 同义替换和非同义替换 不同物种 42 条

TYRP2 基因序列编码区中,同义替换平均位点数为 276.71 个,非同义替换平均位点数为 1857.29 个。含有 2 条序列以上的物种同义替换位点数(*SS*)为 281.42~302.17,同义替换核苷酸多样性均值($\pi(s)$)为 0.13416(表 4)。非同义替换位点数(*NSS*)为 864.56~885.58,非同义替换核苷酸多样性均值($\pi(a)$)为 0.19582。本研究中发现,所选物种 *TYRP2* 基因的非同义替换位点数均明显高于同义替换位点数,说明所分析的这些种群进化过程中可能受到正选择的影响^[8]。

表 4 不同物种 TYRP2 基因同义替换和非同义替换位点数

物种	同义替换位点数	非同义替换位点数
绵羊	302.17	864.83
狨	284.00	883.00
恒河猕猴	287.00	880.00
毛猩猩	281.83	885.53
人	281.42	885.58
小家鼠	281.47	885.53
褐家鼠	281.50	885.50
斑马鱼	290.44	864.56

2.2.3 不同物种 TYRP2 基因遗传分化分析 根据不同物种之间核苷酸歧异度 (D_{xy}), 用 MEGA 4.0 软件的 UPGMA 方法构建不同物种间的聚类图(图 1)。由图 1 可以看出, 人先和毛猩猩聚在一起, 然后依次跟恒河猕猴、狨聚在一起, 说明人与毛猩猩的亲缘关系最近; 小家鼠和褐家鼠聚在一起, 说明小家鼠与褐家鼠的亲缘关系很近, 这些都与动物学分类是相符合的。

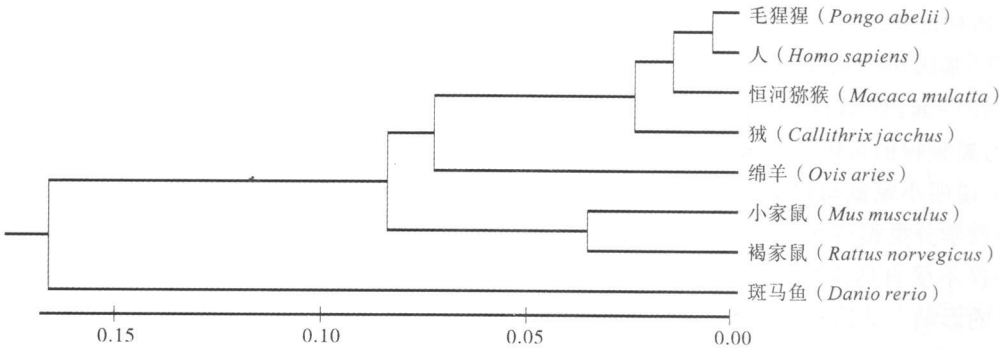


图 1 根据物种间的核苷酸歧义度构建的聚类分析结果

2.3 不同物种 TYRP2 氨基酸序列分析

2.3.1 信号肽的预测与分析 信号肽有指导分泌性蛋白到内质网膜上合成的作用, 在蛋白质合成结束之前被切除, 一般有 16~26 个氨基酸残基, 包括疏水核心区、信号肽的 C 端和 N 端^[9]。利用在线工具 SignalP 3.0 Server 对 41 条氨基酸序列进行在线分析。由表 5 可知, TYRP2 蛋白有信号肽的存在, 不同物种信号肽的裂解位点不同, 但亲缘关系较近的物种之间有很大的相同点。

表 5 不同物种 TYRP2 蛋白信号肽裂解位点

信号肽裂解位点	物种
19, 20 SEA - QF	斑马鱼
25, 26 VQA - QF	斑胸草雀
23, 24 ARA - QF	大熊猫、绵羊、牛
28, 29 TWA - QF	灰短尾负鼠
21, 22 VRA - QF	非洲爪蟾
54, 55 AWA - QF	犬
23, 24 AQG - QF	人
23, 24 ARA - QF	褐家鼠、小家鼠
23, 24 AQA - QF	毛猩猩、恒河猕猴、猪
23, 24 AWA - QF	兔、马、狨
25, 26 AEA - QF	绿头鸭、原鸡
29, 30 VRT - QF	鸭嘴兽

2.3.2 跨膜结构域的预测与分析 跨膜结构域常是通过跨膜蛋白的效应区域展现的, 一般由 20 个左右的疏水氨基酸残基组成, 主要形成的结构是 α 螺旋^[9]。利用在线工具 TMHMM 2.0 Server 对不同

物种 TYRP2 氨基酸序列进行预测分析, 发现大部分物种均具有一个跨膜结构域, 如绵羊 (NP_001123496.1) 跨膜的氨基酸位置为 496 - 491 位。但犬和灰短尾负鼠具有 2 个跨膜结构域, 如灰短尾负鼠 (XP_001377636.1) 其跨膜氨基酸位置为 7 - 29 位和 474 - 496 位; 而褐家鼠 (XP_224517.4) 不具有跨膜结构域, 蛋白位于细胞外侧。该结果说明 TYRP2 蛋白是膜相关蛋白, 大部分物种的跨膜结构域为 22 个氨基酸, 不同物种的跨膜结构域位置不同。

2.3.3 疏水性/亲水性的预测与分析 蛋白质亲水/疏水性氨基酸的组成是蛋白质折叠的主要驱动力。一般通过亲水性分布图来反映蛋白质的折叠情况。蛋白质折叠时形成疏水内核和亲水表面, 同时在潜在跨膜区出现高疏水值, 据此可以测定跨膜螺旋等结构和蛋白质表面氨基酸分布^[9]。利用在线工具 ProtScale 预测不同物种 TYRP2 蛋白序列的疏水性/亲水性, 该蛋白最低分值在 -3.5 ~ -2.7, 最高分值在 3.0 ~ 4.0, 物种内的最低分值和最高分值基本相同, 物种间的最低和最高分值存在差异。但是从各个物种的分析结果看, 氨基酸的疏水/亲水性基本相似, 整个多肽表现为疏水性。

2.3.4 二级结构的预测与分析 蛋白质二级结构是指氨基酸残基形成的 α 螺旋、 β 折叠、延伸链、无规卷曲以及基序等组件^[9]。用 PBIL LYON-GER LAND 信息库对不同物种 TYRP2 氨基酸序列的二

级结构进行预测, 结果表明, 各个物种的 TYRP2 蛋白二级结构的主要结构元件是无规卷曲(51.14%~59.84%)和 α 螺旋(24.76%~35.58%), 其次为延伸链(9.58%~15.45%)。

3 讨论

3.1 不同物种 TYRP2 的核苷酸分析

本研究结果表明, 不同物种间的核苷酸歧异度、净遗传距离和单倍型间的遗传距离均较大, 这说明不同物种间遗传分化已经十分明显。从不同物种的 42 条 TYRP2 基因序列构建的聚类图可看到, 人先和毛猩猩聚在一起, 然后再跟恒河猕猴、狨聚在一起, 说明人与黑猩猩的亲缘关系最近; 小家鼠和褐家鼠聚在一起, 说明小家鼠与褐家鼠的亲缘关系很近, 这些都与动物学分类相符合。

同义替换不受自然选择的控制, 非同义突变易受自然选择的影响。非同义的替代速率则远远低于同义替代, 且其发生的速率与基因密切相关, 这些都是净化选择的结果^[8]。本研究中发现, 所选物种 TYRP2 基因的非同义替换位点数均明显高于同义替换位点数, 说明这些种群进化过程中可能受到正选择的影响。在蛋白质编码过程中, 不同物种或同一物种的不同基因在密码子使用上均存在明显的偏爱性^[10]。本研究发现, TYRP2 基因的某些密码子使用表现出较强的偏爱性。目前, 对密码子的偏爱性以及其在进化中的作用还没有合理的解释, 有待今后进一步研究。

研究中发现, 物种间的 TYRP2 基因有 741 个突变位点, 这些突变位点是否与不同物种毛发颜色的变化有关还需要进一步研究, 在某些物种内还发现仅生成一种单倍型, 除了由于 TYRP2 基因编码区在物种内较保守的原因外, 可能还与这些物种样本含量太小有关。

3.2 不同物种 TYRP2 蛋白分析

对不同物种的信号肽分析结果表明, 这些物种的 TYRP2 蛋白有信号肽的存在, 可以推断该基因在游离的核糖体上起始合成后再进行蛋白转运从而发挥作用。跨膜结构域的预测表明, 这些物种的 TYRP2 存在跨膜结构域, 不同物种的跨膜结构域

存在着一定的差别, 大部分的物种是 1 个跨膜结构, 某些特殊的物种存在 2 个跨膜和无跨膜结构的特殊现象, 具体原因有待于进一步试验分析。

疏水/亲水性是蛋白质的一个重要的理化性质, 对不同物种的 TYRP2 蛋白的疏水/亲水性进行预测, 结果表明, 该蛋白在不同物种中均为疏水蛋白, 这与其存在跨膜结构域相吻合。对 TYRP2 进行二级结构预测, 结果表明, 该蛋白质在不同物种中有相似的二级结构, 无规卷曲和 α 螺旋是构成其二级结构的主要部件, 除此之外还存在少量的延伸链。

参考文献:

- [1] 刘薇. 丝羽乌骨鸡 BAC 文库的构建与黑色素相关基因 TYRP1 和 ID 的研究[D]. 北京: 中国农业大学, 2004.
- [2] Deng W, Tan Y, Wang X, *et al.* Molecular cloning, sequence characteristics and polymorphism analyses of the tyrosinase related protein2/DOPAchrome tautomerase gene of black boned sheep(*Ovis aries*)[J]. *Genome*, 2009, 52(12): 1001-1011.
- [3] Olivares C, Solano F. New insights into the active site structure and catalytic mechanism of tyrosinase and its related proteins[J]. *Pigment Cell Melanoma Res*, 2009, 22(6): 750-760.
- [4] Mohanty T R, Seo K S, Park K M, *et al.* Molecular variation in pigmentation genes contributing to coat colour in native Korean Hanwoo cattle[J]. *Anim Genet*, 2008, 39(5): 550-553.
- [5] 舒文, 毛华明. 黑色素的研究进展[J]. *国外畜牧学-猪与禽*, 2003, 23(2): 31-34.
- [6] Ghosh T. Studies on codon usage in *Entamoeba histolytica*[J]. *International Journal for Parasitology*, 2000, 30: 715-722.
- [7] Wright F. The "effective number of codons" used in a gene[J]. *Gene*, 1990, 87(1): 23-29.
- [8] 李易. 基因进化的同义与非同义替代计算及统计检验的比较分析[J]. *曲靖师范学院学报*, 2006, 25(6): 1-8.
- [9] 董娇, 周军, 辛培尧, 等. 不同植物 LDOX/ANS 基因的生物信息学分析[J]. *基因组学与应用生物学*, 2010, 29(5): 815-822.
- [10] Guo Z. Introduction to population genetics[M]. Beijing: Agricultural Press, 1993: 298-332.