

生物信息学在畜禽基因组研究中的应用

李莱娥^{1,2}, 马月辉¹, 叶绍辉²

(1. 中国农业科学院北京畜牧兽医研究所, 北京 100094; 2. 云南农业大学动物科学技术学院, 昆明 650201)

摘要:生物信息学已成为整个生命科学发展的的重要组成部分,是生命科学研究的前沿领域。作者简述了畜禽基因组学和畜禽生物信息学的研究概况,着重阐述了生物信息学在畜禽基因组研究中的应用现状,并展望其发展前景。

关键词:生物信息学;畜禽基因组学;结构基因组学;功能基因组学;比较基因组学

中图分类号:Q811.4

文献标识码:A

文章编号:1671-7236(2007)02-0079-03

人类基因组计划的开展,带来了一门新兴学科的产生——生物信息学。同时,其他物种的基因组计划也相继开展起来,随着人类基因组计划的启动和成功完成推动了动物结构基因组学、功能基因组学及比较基因组学和蛋白质组学的产生和蓬勃发展,为动物遗传改良奠定了遗传学基础。畜牧业是农业的重要组成部分,畜牧业产值占世界农业总产值的30%~40%。在美国,畜禽销售产值超过了700亿美元(Roberts, 2001)。因此从畜禽基因组学和生物信息学角度系统深入地开展品质性状遗传改良和多产性状的研究对于我国畜牧业的发展尤为重要。

1 畜禽基因组学及生物信息学的研究概况

1.1 畜禽基因组学的研究概况

基因组学(genomics),指对一个物种的所有基因进行基因组作图(包括遗传图、物理图谱、转录图谱)、核苷酸序列分析、基因定位和基因功能分析的一门学科。基因组学研究主要包括以下几方面的内容:以全基因组测序为目标的结构基因组学(structural genomics)和以基因功能鉴定为目标的功能基因组学(functional genomics)以及以基因组比较为基础,研究生物进化为目标的比较基因组学(comparative genomics)。结构基因组学代表基因组分析的早期阶段,以建立生物体高分辨率遗传、物理和转录图谱为主。比较基因组学和功能基因组学代表基因分析的新阶段。功能基因组学是利用结构基因组学提供的信息系统地研究基因功能,它以高通量、大规模试验方法以及统计与计算机分析为特征。比较基因组学也是在结构基因组学基础上发展起来的,以不同物种或同一物种基因组的比较来研究生物进化和品种多样性等问题。

随着基因组计划的不断深入,基因组学在生物学、医药、农业、环境科学等中的作用也越来越重要。畜禽的基因组学已通过一条非典型的途径建立起来,它经历了4个发展阶段:绘制一张大区域遗传图谱(约20厘摩),它包含有多态信息标记(微卫星)和进化保守基因标记;利用信息标记进行重要经济数量性状位点(QTLs)定位,这一工作需要复合家系或表型和基因型有分离的品系间的杂交;从信息标记深入到数量性状位点(QTLs),通过复合家系和回交试验直接识别性状基因本身,或利用保守标记在基因丰富的畜种中筛选出性状候选基因位点;结合基因组学,通过生理学至性状的表型距离,对性状基因的功能进行分析(丁冷水,1998)。

今后畜禽基因组学研究的主要内容是:畜禽的功能基因组学和比较基因组学,定位影响畜禽重要经济性状的主效基因;非孟德尔遗传基因或QTL的鉴别及育种应用;研究分子标记辅助育种的技术体系;建立畜禽各种组织cDNA文库;建立芯片技术平台,进行成批基因表达、高通量基因型等方面的研究。

随着人类基因组计划的进行而开展的动物基因组计划尽管晚于农作物基因组的研究,但其作为一种理想的动物模型具有其他生物不可替代的作用,尤其是对于人类医学研究。家畜由于几千年的选择培育产生了丰富的表型变异,这使得家畜成为很好的用来研究表型变异遗传基础的素材(Andersson等,2004)。家畜基因组的研究可以帮助提高人类健康水平和家畜的生产能力(Hamernik等,2003)。现在,欧盟和美国都启动了动植物基因组计划,尤其是研究经济意义较大的生物基因组,如动物中的猪、羊、牛、马、狗、鸡等都已开始(Gellin等,2000)。牛基因组测定排序计划领导小组已经宣布:世界首例牛的基因组排序目前已经完成,并将免费放入公共数据库供全世界的生物学者和农业研究者使用。研

收稿日期:2006-02-28

作者简介:李莱娥(1978-),女,云南人,硕士生,研究方向:动物遗传育种学。

通讯作者:马月辉,研究员,博导。E-mail:yuehui.ma@263.net

究畜禽基因组对畜禽品质和生产能力提高以及揭示人类基因组的结构和功能、人类疾病分子机理和生物进化规律具有重要作用。

1.2 畜禽生物信息学研究概况 生物信息学(bioinformatics)是由生物科学、计算机科学、信息科学、应用数学、统计学等多门学科相互交叉而形成的一门新兴学科。它以计算机为主要工具,开发各种软件,对日益增长的DNA和蛋白质的序列和结构等相关信息进行收集、储存、发行、提取、加工、分析和研究,同时建立理论模型,指导试验研究。它由数据库、计算机网络和应用软件3大部分构成(陈力学,2003),包含2个方面:对海量数据的收集、整理与服务;从中发现新的规律。它不仅具有重要的学术价值,还有很大的商业价值,有着远大的发展前景。随着后基因组时代的到来,它将发挥越来越不可替代的作用。

畜禽生物信息学研究的工作主要包括:收集分析国内外基因库数据,建立与动物良种繁育相关的基因组数据库;运用先进有效的农业生物信息学研究手段,结合我国丰富的特有遗传资源,开展中国优良家畜资源的单核苷酸多态性(SNP)和插入缺失多态性的研究;分离、克隆有自主知识产权的重要经济价值的新基因及重要的基因表达调控元件,发现控制优良性状基因的分子标记。

2 生物信息学在畜禽基因组研究中的应用

生物信息学是基因组学研究的基础,对基因、基因结构、基因产物的功能分析都是必不可少的技术手段。动物功能性基因及重要经济性状的基因定位研究;功能性基因组和比较功能性基因组研究;肉、奶产量和品质性状的主基因定位、分离、克隆、测序和表达调控;明确畜禽基因组中影响重要经济性状的遗传座位和这些座位之间的相互关系,以及每个座位上等位基因调控表达的分子机理;充分发掘我国动物品种的基因资源,获得多项基因专利,奠定未来依据基因组进行动物品种设计和改良的基础,在这些研究工作中都少不了生物信息学的帮助。生物信息学在畜禽基因组研究中的应用主要体现在以下几方面。

2.1 结构基因组学中的应用 结构基因组学是基因组学的一个重要组成部分和研究领域,它是一门通过基因作图、核苷酸序列分析确定基因组成、基因定位的科学(段民孝,2001)。生物信息学在结构基因组学中的应用主要在于:基因组作图、核苷酸序列分析、QTLs及基因定位、新基因的发现和定位等方面。

2.1.1 基因组作图

2.1.1.1 遗传图谱 通过遗传重组所得到的基因在具体染色体上线性排列图称为遗传连锁图。它是

通过计算连锁的遗传标记之间的重组频率,确定他们的相对距离,一般用厘摩(cM,即每次减数分裂的重组频率为1%)表示。绘制遗传连锁图的方法很多,如早期的RFLP(限制性酶切片段长度多态性)、AFLP(随机扩增片段长度多态性)及后来发展起来的STR(微卫星)、SNP(单核苷酸多态性)等。

2.1.1.2 物理图谱 物理图谱是利用限制性内切酶将染色体切成片段,再根据重叠序列确定片段之间连接顺序,以及遗传标志之间物理距离(bp、kb、或Mb)的图谱。Cheng等(2003)通过对鸡染色体的57091个BCA克隆进行限制性酶切绘制了鸡的染色体物理图谱。这张物理图谱包含了2331个BCA重叠群,其物理距离跨度估计为1510Mb。这些BCA重叠群都通过人工鉴定并用367个DNA标记进行了扫描。共有361个重叠群绑定在鸡的已知遗传图谱上了。这张图谱代表了鸡的第一张基于BCA的基因组范围内的物理图谱,这为鸡基因组许多领域的研究提供了一个强有力的平台,如靶标的发现、新基因及QTLs定位、位点克隆、鸟类基因组组成及进化分析、鸡和哺乳动物的比较基因组研究和鸡的大规模测序等。

2.1.1.3 转录图谱 利用EST作为标记所构建的分子遗传图谱被称为转录图谱。现在国际上几大生物数据库已经收集了包括猪、牛、鸡等很多家养动物的EST数据。这些EST不仅为基因组遗传图谱的构建提供了大量的分子标记,而且来自不同组织和器官的EST也为基因的功能研究提供了有价值的信息。截止2003年7月底,收录在DDBJ/EMBL/GenBank数据库中的猪的EST序列将近140000条。但是,通过常规的cDNA合成方法得到的这些EST很少是全长cDNA,而且,通过序列拼接得到的全长cDNA数量也有限。相比之下,对全长cDNA克隆转录的蛋白质产物进行分析却是很有价值的。于是,Hirohide等(2004)利用一种新的方法构建了猪的全长cDNA文库并将其中的5'EST进行拼接,建立了一个猪的PEDE(pig EST data explorer)。PEDE能为同源性搜索、单核苷酸多态性检测及从序列数据和cDNA克隆中获取有用信息提供方便。这一数据库能够加快对家畜性状的分析以及获取一些新的信息,从而使猪能成为一种医学研究的动物模型。

2.1.2 核苷酸序列信息分析 现在已经得到了大量畜禽的基因组或EST序列数据,如何将实验室中得到的生物信息转化为计算机能够处理的数字信息,是生物信息学中的一个重要课题,从测序仪的光密度采样与分析、碱基读出、载体识别与去除、拼接、

填补序列间隙到重复序列标识、可读框预测和基因标注的每一步都是紧密依赖基因组信息学的软件和数据库(陈力学,2003)。另外,大规模测序产生的海量数据的处理和分析都离不开生物信息学的帮助。

2.1.3 基因及 QTLs 定位 基因及 QTLs 定位的成功应用要依赖于保守基因更详尽的连锁图谱和比较完整的基因组数据库。

2.1.4 新基因和 SNP 的发现和鉴定 利用 EST 数据库(dbEST)可发现新基因。截止 2003 年 7 月底,收录在 DDBJ/EMBL/GenBank 数据库中的猪的 EST 序列将近 140000 条。因此,如何利用这些信息发现新基因和新的多核苷酸位点,其中很大一部分工作都需要生物信息学来进行。

2.2 比较基因组学中的应用 比较基因组学是在基因组图谱及测序的基础上,通过对系统发育中代表性物种之间全方位基因和基因家族的比较分析,构建系统发育的遗传图谱,揭示基因、基因家族的起源和功能及其在进化过程中复杂化和多样化的机制的学科。比较基因组学可用于生物进化及品种多样性、疾病相关基因、动物品种遗传资源改良等方面的研究(Roberts,2001;Georges,2001)。而生物信息学在比较基因组学研究中的工作主要包括:序列相似性比较:就是将待研究的序列与已知序列进行比较,用于确定该序列的生物属性。常用的程序有 BLAST 和 FASTA 等。序列同源性分析:将待研究的序列加入到一组与之同源、来自不同物种的序列中进行多序列同时比较,以确定该序列与其他序列间的同源性大小。程序上主要是 CLUSTAL。

构建系统进化树:根据序列同源性分析的结果,重建反映物种间进化关系的进化树。为完成这一工作已发展了许多种软件包,如 PLYPT 和 MEGA 等。

2.3 功能基因组学中的应用 功能基因组学是指在全基因组序列测定的基础上,从整体水平研究基因及其产物在不同时空、条件下的结构与功能关系及活动规律的学科(钱小红等,2003;贺光,2002)。其研究内容包括基因功能发现、基因表达分析及突变检测。基因的功能包括:生物学功能,如作为蛋白质激酶对特异蛋白质进行磷酸化修饰;细胞学功能,如参与细胞间和细胞内信号传递途径;发育上功能,如参与形态建成等。另外,由于对基因表达分析的重视使哺乳动物转基因成为后基因组时代的焦点。动物转基因技术的发展(如转基因学、基因打靶、核移植或动物克隆 3 大技术)增强了我们从机体整体水平上来研究基因表达及修饰基因表达的能力(Shashikant 等,2003)。

2.3.1 基因功能的发现

2.3.1.1 疾病相关基因的研究 利用功能基因组学研究可以对家畜及其它动物的生理反应过程和进化适应性的遗传基础进行研究。Powell(2003)通过功能基因组学来研究家畜等动物对环境氧水平变化的反应,来解释其生理反应和进化的遗传基础。Yao 等(2001)利用牛的 EST 和 cDNA 芯片来研究其免疫生物学反应,即研究生物机体对外界刺激物的反应。

2.3.1.2 胚胎发育相关基因的研究 利用 DNA 芯片、文库分析、数据库其它哺乳动物的序列比较分析以及 RNA 干涉等方法可研究动物卵母细胞及早期胚胎中特异性基因及其功能(Sirard 等,2003)。

2.3.2 基因表达分析 现在用于基因表达分析的方法很多,如基因表达系列分析(SAGE)、基因芯片(gene chip)等。这些方法都是一种高通量的表达分析方法,其产生的大量数据都需要用生物信息学的手段来分析。尤其是基因芯片,从确定基因芯片检测对象到基因芯片设计,从芯片检测结果分析到试验数据管理和信息挖掘,无不需要生物信息学的支持和帮助(孙啸,2001)。

3 小结

随着现代生物技术的迅速发展,以及生物信息学这一技术平台在畜禽基因组研究领域的广泛应用,畜禽基因组研究将会取得巨大进展并将给生物学、医学、农业等带来举足轻重的作用。在生物信息学发展的带动下,必将会带动畜禽基因组及相关学科的跨越式发展。

参 考 文 献

- 1 孙啸. 生物信息学在基因芯片中的应用. 生物物理学报, 2001, 17(1): 27~33.
- 2 陈力学. 生物信息学在基因组研究中的应用. 国外医学临床生物化学与检验学分册, 2003, 24(6): 339~340.
- 3 段民孝. 基因组学研究概念. 北京农业科学, 2001, 2: 6~10.
- 4 贺光. 生物信息学在蛋白质研究中的应用. 国外医学遗传学分册, 2002, 25(3): 156~158.
- 5 钱小红, 贺福初. 功能基因组与蛋白质组. 科学出版社, 2003, 1~23.
- 6 Andersson L, Georges M. Domestic animal genomics: deciphering the genetics of complex traits. Nat Rev Genet, 2004, 5(3): 202~212.
- 7 Cheng W R, Mi Kyung L, Bo Y, et al. ABAC-Based Physical Map of the Chicken Genome. Genome Research, 2003, 9(3): 2754~2758.
- 8 Gellin J, Brown S, Marshal G, et al. Comparative gene mapping workshop: progress in agriculturally important animals. Mamm Genome, 2000, 11(2): 140~144.
- 9 Georges M. Recent progress in livestock genomics and potential impact on breeding programs. Theriogenology, 2001, 55(1): 15

DNA 甲基化、组蛋白修饰与基因沉默

索效军,张年

(湖北省农业科学院畜牧兽医研究所,武汉 430209)

摘要: DNA 甲基化、组蛋白去乙酰化和 H3 K9 甲基化是 3 种最典型的与染色质凝聚状态有关的共价修饰方式,它们都与基因的沉默存在着联系,近年来的研究主要集中于 DNA 甲基化与各种组蛋白修饰之间的关系,以及染色质阻遏状态的自身强化循环关系上。组蛋白去乙酰化和 H3 K9 甲基化可能有助于 DNA 甲基化模式的建立,此外,研究结果还显示 CpG 甲基化和其它组蛋白修饰存在相互作用。

关键词: DNA;组蛋白;甲基转移酶;甲基化;基因沉默

中图分类号: Q754

文献标识码: A

文章编号: 1671-7236(2007)02-0082-04

在哺乳动物中,DNA 甲基化、染色质结构和基因沉默存在某种内在的联系,这种现象多年前已被人们所发现,早期的研究结果显示高甲基化的 CpG 二核苷酸与异染色质形成有关,体外试验结果表明,甲基化 DNA 与染色质的凝聚状态有关,其它的研究结果也表明未甲基化或低甲基化的 CpG 岛富集于高乙酰化组蛋白上。为什么会存在这种机制?一直以来都不清楚,但最近染色质结构调节基因表达

的理论可能解释这种机制。

组蛋白翻译完成后,其氨基尾巴会发生多种共价修饰,如乙酰化、甲基化、磷酸化、泛素化、糖基化等,这些修饰方式共同构成了“组蛋白密码(histone code)”的假说,即在一个或多个组蛋白氨基尾端的多种修饰状态,可以互相联合或依次地被特定的蛋白(酶)或其它复合体等识别、结合而起作用,为发动或阻遏基因转录的染色质相关蛋白提供结合位点(Jenuwein 等,2001)。在基因沉默相关的各种组蛋白修饰中,最典型的是组蛋白去乙酰化、H3 K9 甲基化(Margueron 等,2005),在各种生物体中,这些修

收稿日期: 2006-07-20

作者简介: 索效军(1977 -),男,山西人,硕士生,研究方向:动物遗传学。

~ 21.

10 Hamernik D L, Lewin H A, Schok L B, et al. Allerton . Beyond livestock genomics. Anim Biotechnol, 2003, 14(1): 77 ~ 82.

11 Hirohide U, Tomoko E, Kohei S, et al. PEDE(Pig EST Data Explorer): construction of a database for ESTs derived from porcine full-length cDNA libraries. Nucleic Acids Research, 2004, 32: D484 ~ D488.

12 Powell F L. Functional genomics and the comparative physiology of hypoxia. Annu Rev Physiol, 2003, 65: 203 ~ 230.

13 Roberts R M. The place of farm animal species in the new ge-

nomics world of reproductive biology. Bio Reprod, 2001, 64 (2): 409 ~ 417.

14 Shashikant C S, Ruddle F H. Impact of transgenic technologies on functional genomics. Curr Issues Mol Bio, 2003, 5(3): 75 ~ 98.

15 Sirard M A, Dufort I, Coenen K, et al. The use of genomics and proteomics to understand oocyte and early embryo functions in farm animals. Reprod Supp, 2003, 61: 117 ~ 129.

16 Yao J, Burton J L, Saama P. Generation of EST and cDNA microarray resources for the study of bovine immunobiology. Acta Vet Scand, 2001, 42(3): 391 ~ 405.

The Application of Bioinformatics on Livestock and Poultry Genomics Study

LI Cai-e^{1,2}, MA Yue-hui¹, YE Shao-hui²

(1. Institute of Animal Science, Chinese Academy of Agricultural Science, Beijing 100094, China;

2. College of Animal Science and Technology, Yunnan Agricultural University, Kunming 650201, China)

Abstract: Bioinformatics is an important component of the whole life science, has become the advanced research domain of life science. This paper mainly discuss from the following aspects: livestock and poultry genomics, livestock and poultry bioinformatics and its application on livestock and poultry genome study and also project ist foreground.

Key words: bioinformatics; livestock and poultry genomics; structural genomics; functional genomics; comparative genomics

