

文章编号: 1004-0374(2010)09-0952-07

## 纠错编码理论在生物信息处理领域的应用

刘 晓<sup>1\*</sup>, 田逢春<sup>1</sup>, 李素芳<sup>2</sup>

(1 重庆大学通信工程学院, 重庆 400044; 2 中国计量学院生命科学学院, 杭州 310018)

**摘 要:** 根据生物系统中存在信息传递并且需要高度稳定性的特性, 现代通信工程中的纠错编码理论被应用于生物遗传序列的研究和测试系统的设计。该文介绍其分析基础和现有应用并对未来研究展望。

**关键词:** 纠错编码; 遗传信息; 信息处理

**中图分类号:** Q-0; Q61; Q7 **文献标识码:** A

## Application of error-correction coding theory in the study of bioinformation

LIU Xiao<sup>1\*</sup>, TIAN Feng-chun<sup>1</sup>, LI Su-fang<sup>2</sup>

(1 College of Communication Engineering, Chong Qing University, Chongqing 400044, China;

2 College of Life Sciences, China Jiliang University, Hangzhou 310018, China)

**Abstract:** Based on the feature that information is stored and transmitted in living systems with request of high stability, error-correction coding theory of modern communication engineering is employed for study of genetic sequences and design of biological test systems. The basis, the current applications and the perspectives of the method are presented in this article.

**Key words:** error-correction coding; genetic information, information processing

随着信号处理技术, 如傅立叶变换、小波变换、频谱分析、分形、混沌动力学等的发展, 信息处理技术在生物学研究中的应用不断取得进步, 发现了DNA序列中的长程相关性, 并用于序列周期分析与基因预测、蛋白质结构预测、序列相似性分析等<sup>[1-3]</sup>。

在信息理论中, 纠错编码理论与方法是重要的组成部分。码元之间具有监督校验关系, 编码具有检查、纠正错误的能力是纠错编码理论区别于其他信息处理方法的特点。自Yockey<sup>[4]</sup>提出将纠错编码应用到遗传信息分析中, 研究人员已将纠错编码理论应用到生物学研究的多个方面, 取得了有意义的成果。

### 1 纠错编码理论与生物信息分析相结合的基础

#### 1.1 编码通信系统与生物信息系统共通性

数字通信方式在现代通信系统中应用已经非常广泛, 大大提高了通信系统的抗干扰能力和信息传递灵活性。数字通信系统中(图1), 信息编码通常

包括信源编码和信道编码两部分, 其后将经过调制的信号通过信道传送。信源编码将连续的消息变换为数字信号, 以便于处理。信道编码则使数字信号与传输媒质匹配, 提高传输的可靠性或有效性。信道编码可以赋予传输的数据一定检错与纠错能力, 也称为纠错编码。纠错编码基本原理是: 在发送端被传输的信息序列上附加一些码元(称为监督码元), 监督码元与信息码元之间以某种确定的规则相互关联(约束); 接收端按照既定的规则检验信息码元与监督码元之间的关系, 如果在传输过程中发生差错, 则信息码元与监督码元之间的关系将被破坏。从而接收端可以发现错误, 乃至纠正错误。编码理论在现代通信系统中应用非常广泛。

收稿日期: 2010-02-01; 修回日期: 2010-02-25

基金项目: 教育部高等学校博士学科点专项科研基金(20050611022); 重庆市自然科学基金(重点)项目(2009BA2021)

\*通讯作者: E-mail: liuxiao@cqu.edu.cn

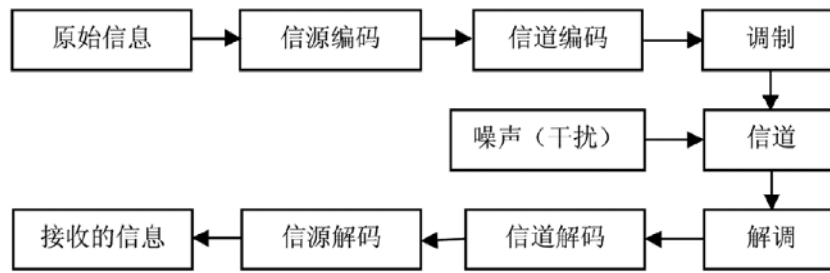


图1 数字通信信号传输模型

在生物系统中，同样存在准确传递信息的需求。为了克服遗传信息传递过程中各种干扰、突变对信息的影响，遗传信息系统本身具有多种自我维护/修复的方式。比如DNA修复系统包括错配修复、碱基切除修复、核苷酸切除修复等功能。又如，密码子具有简并性，由变异引起密码子中一个核苷酸的改变，其结果可能也只是变成了同一个氨基酸的另一个密码子，合成出与原来没有区别的蛋白质。简并密码的存在，起到了防错的作用，有助于维持物种的稳定性。[5, 6]认为，校验可以避免复制中出现的错误，但是无法纠正进行复制的原始信息中出现的错误。只有借助于纠错机制才能在受到温度、辐射及宇宙射线影响的环境中实现可靠的信息传递。否则，一方面难以实现DNA复制过程 $10^{-9} \sim 10^{-10}$ 数量级的错误率[7]；另一方面在相同的背景环境下，复制错误率应该是一个常数，而实际上不同物种有所不同。据此，可以认为生物系统必然使用了某种(复杂的)纠错机制，并且不同物种之间应该有所不同，因此生物学研究应该充分考虑纠错编码理论的应用。

Endy[8]将生物系统比做复制机器，不过在复制过程中会出差错。因此，从差错控制的角度来看，通信工程中的编码系统与生物遗传系统是可以相互联系相互对照的[9]，如图2。

### 1.2 冗余数据——编码冗余性

根据已经认识到的生物纠错机制，必须认识到获得纠错能力应该付出的对应“代价”——冗余性。即为了传递 $k$ 个符号的信息，必须使用 $n$  ( $n > k$ )个符号对这 $k$ 个符号进行无信息损失的取代，然后进行传送。生物体中的冗余性体现在很多方面，比如内含子可能是外显子的冗余、辅助信息的体现；DNA长程相关性表示的相关部分可视为信息的多次不准确重复(multiple unfaithful repetition)，它们经过生物体总体综合则可以得到正确的信息结果[6]；

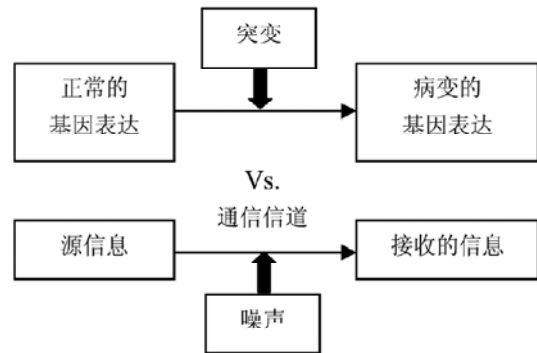


图2 生物系统与通信系统对比[9]

密码子蕴含的冗余性，64个密码子只对应20种氨基酸[10]，等等。最近，Nicolas等[11]在研究中发现斑马鱼的基因组拷贝了一些关键的基因，一旦有突变破坏了这些基因，生物体还留有一个备份，这为基因在遗传过程中提供了更好的抗干扰能力(或健壮性)。

### 1.3 碱基/密码子之间的相互关联——编码监督关系

纠错编码的监督码元与信息码元之间存在相互关联，在遗传序列中，遗传信息单元与相邻单元之间同样存在相互联系与影响，包括碱基之间、密码子之间、开放阅读框之间以至各个基因之间的相互关联和影响。这些在序列中由AGCT的重复而形成的结构/功能是相互关联的，其中涉及到两种关联特性，长程相关性与短程关联优势，分别是DNA尺度和较近碱基尺度上呈现的相关性。

Shpaer[12]在研究中发现某些密码子的翻译效率会受到密码子上下文关联的影响。张静等[13]在分析同义密码子携带的蛋白质结构信息如何表达时也提出，研究同义密码子携带的结构信息，不能只孤立地考虑同义密码子本身，而应该从mRNA序列水平上考虑，亦即研究同义密码子的上下文关系，在对可变翻译的研究中也发现了翻译的效率往往要受到起始密码子的上下文的影响[14]。赵辉等[15]在水稻和

拟南芥的研究中也发现单核苷酸多态性(single nucleotide polymorphism, SNP)位点变化的情况也受到其前后碱基组分的影响。

卢欣等<sup>[16]</sup>用非线性的方法分析了细菌的DNA序列,发现在细菌DNA序列中普遍存在长程相关现象;赵小杰等<sup>[17]</sup>在利用功率谱分析编码DNA序列基础上,用Hurst指数对人的DNA序列的自相似性进行描述,结果也反映出人DNA序列中的长程相似性。

Gonzalez等<sup>[10]</sup>与Luo等<sup>[18]</sup>均指出了编码序列中存在核苷的关联具有短程优势:对DNA编码区序列具有普遍存在的很强的紧邻和次紧邻碱基关联;序列中的核苷有着众多的长程作用和相互影响,但在此复杂背景中存在一定的简单性,即很强的短程关联。同时还证明了碱基紧邻关联随着进化增加,并且两密码子间的紧邻碱基关联具有最强进化相关性,因此碱基关联具有生物学重要性。

通过对比纠错编码系统与生物遗传系统存在的相似性与共通性,说明两个研究领域存在交叉。遗传序列经过数学表达,即可转化为便于工程分析的数值序列,若我们将遗传序列视为经过信道编码而获得的具有某种纠错特性的编码序列,则我们可将目前比较成熟的纠错编码理论引入对生物系统的分析,或者反之用生物系统可能蕴含的编码方式对现有纠错编码技术进行改进。

## 2 基于纠错编码的生物信息分析模型

基于纠错编码思想,研究人员先后提出了若干模型用以描述生物系统的遗传信息传递过程。

Gatlin将DNA视为经过源编码的信息序列,并且具有差错控制能力(DNA中的冗余碱基可以用于差错控制),氨基酸序列与蛋白质是最终接收到的信息<sup>[19,20]</sup>。相关的系统模型如图3所示。

Yockey<sup>[4]</sup>从信息理论和编码理论的数学基础入手,开展了理论生物学的讨论和研究。Yockey提出的生物信息模型如图4所示,其中DNA仍然是系统的输入信息,蛋白质是系统的输出信息。此模型中,Yockey将所有噪声表现在mRNA信道中。他认为遗传密码子到氨基酸的映射涉及分组码,并将其看作是差错保护机制的一部分。

May等<sup>[20]</sup>认为Gatlin模型只考虑了转录和翻译过程产生的错误,而没有提及复制过程产生的错误。由于DNA的复制过程也会产生差错,因此基于纠错编码原理提出了遗传中心法则的编码模型,如图5所示。此模型中特别强调了DNA复制过程是一个引入差错的过程,并根据Battail的思想假设信息处理过程具有嵌套编码(nested encoding)的特征。模型中的DNA序列是包含了冗余信息,具有纠错能力的遗传序列。

Dawy等<sup>[21]</sup>则提出了基因表达过程的通信模

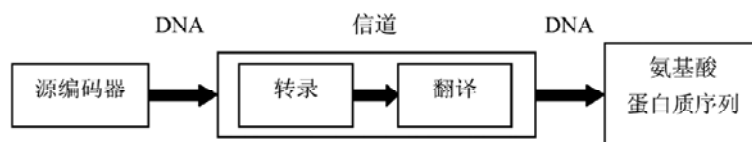


图3 基于通信理论观点的遗传系统模型<sup>[19]</sup>

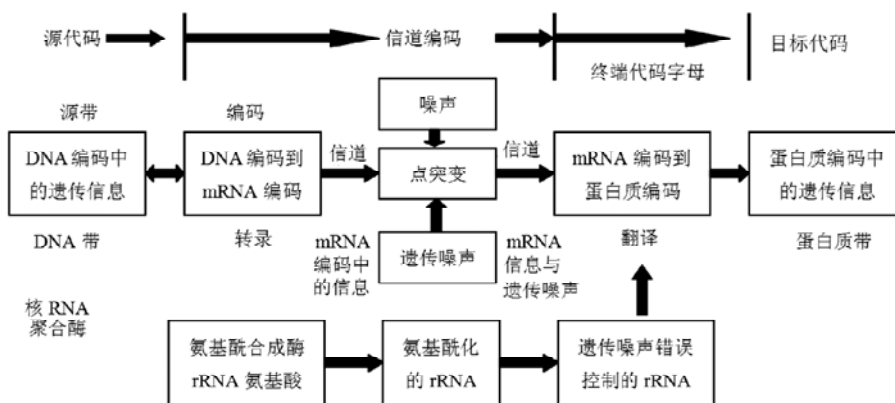


图4 DNA-mRNA-蛋白质通信系统模型<sup>[4]</sup>

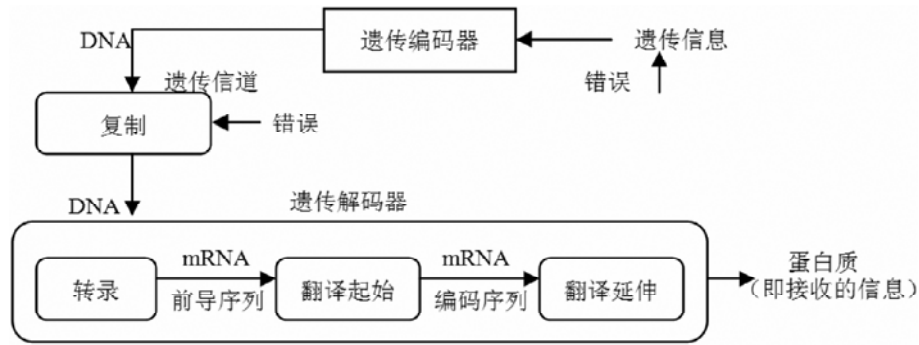


图5 遗传中心法则编码模型<sup>[20]</sup>

型，如图6。该模型中包含两级解码：一是RNA聚合酶，将DNA转录为mRNA；二是核糖体，将mRNA翻译为蛋白质。Dawy等<sup>[22]</sup>随后表示将DNA转录为mRNA过程作为解码(而不是编码)的原因在于他们认为mRNA相对更短，是去除了冗余的表现。

不同于前述模型中将DNA序列作为信道输入信号并将氨基酸序列作为输出信号，Bouaynaya等<sup>[23]</sup>提出了蛋白质之间的遗传信息通信模型，如图7。这是一个抽象分析模型，将细胞定义为蛋白质的集合，细胞分裂则定义为蛋白质集合间的信息通信模型；化学诱变剂和辐射在DNA存储和复制过程中引发错误，转录和翻译过程则实现信息的解码。

### 3 纠错编码理论在生物学研究中的应用

目前，纠错编码理论在生物学研究中的应用涉

及到信号表达方法的设计、翻译起始特性分析、微阵列分析设计、生物传感器的设计和序列相似性分析等多个方面。

Mac Donail<sup>[24]</sup>提出,核苷酸的复制本质上是信息传递过程,进化压力所形成的字母表的成分不应该仅仅限于物理化学观点,差错编码应该扮演强迫性的进化角色,在结合核苷酸氢键化学结构的基础上提出了核苷酸的数字表达方法。

Wang<sup>[9]</sup>将纠错编码应用于微阵列数据分析,对基因表达中的基因改变进行检测和定位,以进行基因突变检测和相应的疾病预后;Khan等<sup>[25]</sup>则利用纠错编码原理改进微阵列的测量可靠性和容错性,当阵列中有一个或多个点被破坏时就将其对应数据丢弃,利用其余测试点计算表达水平。

Reif和Labean<sup>[26]</sup>采用编码理论对生物工艺进行

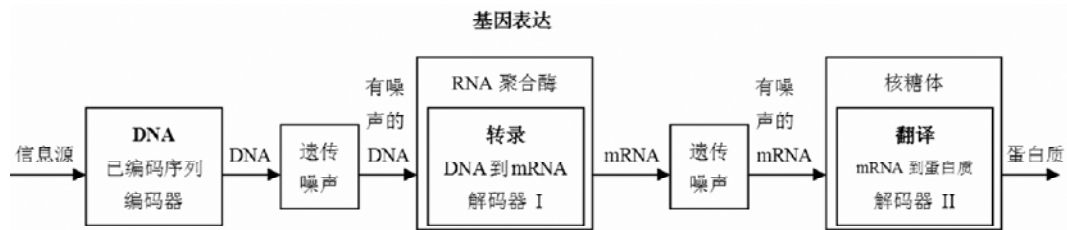


图6 基因表达过程的通信模型<sup>[21]</sup>

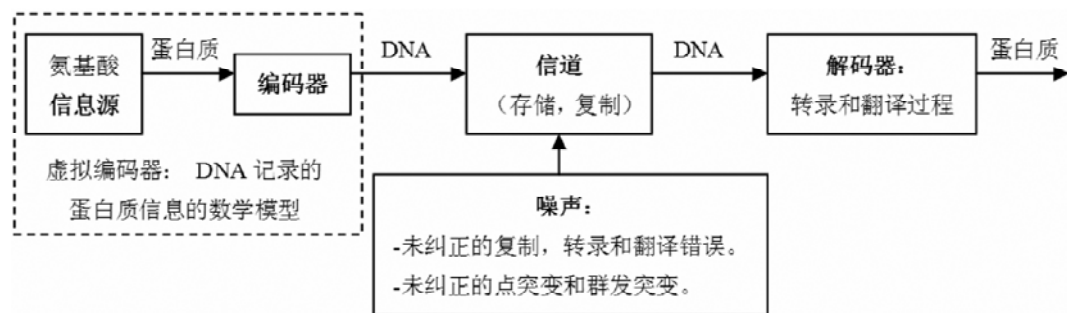


图7 蛋白质通信系统模型<sup>[23]</sup>

了尝试性的应用,包括使用纠错编码减少DNA芯片中的错误率,采用矩阵量化编码计算在有限差错情况下提高DNA芯片读取(写入)电子数据的速率,利用矩阵量化编码计算对某些数据分级以提升在DNA数据库中的相关搜索。

May等<sup>[27]</sup>以大肠埃希菌*E. coli K-12*作为分析对象,使用分组码和卷积码模型进行了蛋白质翻译起始建模及分析。利用*E. coli K-12*的16S rRNA的最后13个碱基作为模板,生成奇偶校验位并构建码字集,实现翻译起始特性的分析。Ponnala等<sup>[28]</sup>同样以此模型为基础,对卷积码模式分析方法的生成矩阵算法进行了改进,也完成了相同的功能。Dawy等<sup>[22]</sup>同样利用16S rRNA作为模板,但通过求解碱基间自由能的方式分析了翻译的起始和结束。

May等<sup>[27]</sup>采用方法的共同点是以分析对象的某个子序列(16S rRNA的最后13个碱基)作为处理信号码集的基础,该方法在处理不同分析对象时则显得复杂,缺乏通用性。另外,从他们对翻译起始的分析结果来看,卷积码分析效果不够理想。而卷积码是一种具有较好性能的信道编码方式,理论和实际上均已证明卷积码的性能至少不比分组码差,应该可以寻找较好的卷积编码模型来分析DNA序列的编码特性。作者所在研究小组分别基于分组码与卷积码提出了新的分析方法,根据密码子的简并性、短程关联占优的普适特性设计了独立于分析对象的编码矩阵,为同批次分析不同对象提供了可行的方法,并利用卷积码的码元之间的相关性反映密码子上下文关联特性,寻求了几种效果较好的卷积码,将其用于序列相似性分析,通过对11个物种的 $\beta$ -球蛋白第一个外显子编码序列的相似性/不相似性分析验证了该方法的有效性<sup>[29-31]</sup>。此外还提出了将分形编码、纠错编码与小波变换相结合对生物系统健壮性进行分析的方法<sup>[32]</sup>。

近来,纠错编码理论的应用还出现在生物传感器的设计中。Liu等<sup>[33]</sup>利用前向纠错编码构建生物传感器系统模型,给出了提高病菌检测率和检测灵敏度的方法。May等<sup>[34]</sup>基于纠错编码理论提出将普通生物检查转换为基于脱氧核酶的算法,并在计算生物传感器系统对H5N1地域性菌株进行验证性检测与分类,特别指出该系统在军事领域(比如应对生化武器威胁)所具有的应用潜力。

在分类算法中有一种称为纠错输出编码(error correcting output coding, ECOC)的算法,是应用

二分类学习算法解决多分类监督学习算法的方法,其中利用了纠错编码的原理提高其判断准确性。ECOC与支持向量机(support vector machine, SVM)相结合在癌症诊断<sup>[35]</sup>、转录因子分类分析<sup>[36]</sup>等领域应用并给出了较好的结果,被认为有很好的应用前景。

## 4 研究展望

目前,非编码序列的研究已经在生物信息研究领域引起广泛的关注,大量研究发现非编码序列是可以表达的,表达产物是非编码RNA,研究已证明非编码RNA具有重要的生物功能,在动植物生长发育中起着重要作用<sup>[37, 38]</sup>。那么非编码区的序列按何种方式存储信息,能否用编码理论挖掘埋藏在非编码序列中的隐性调控基因,这是应该关注的研究内容。

在现有分析模型中,对来自序列外部的影响并未充分考虑,例如横向基因转移(horizontal gene transfer, HGT),是不同生物个体之间或者单个细胞内部不同细胞器之间的基因交换,在原核生物中较为普遍。基因组序列研究表明:在域内或在域间存在广泛的基因水平转移,真核生物拥有来自细菌或古生菌的基因,而两个原核域的生物也有频繁的基因交换,甚至一些细菌也获得过真核生物的基因<sup>[39, 40]</sup>。这表明基因的垂直传递不是影响细胞生物进化的惟一过程,横向基因转移也可能深刻地影响生物的进化过程。DNA甲基化、组蛋白共价修饰等表观遗传修饰现象同样说明基因表达的改变并不一定是由DNA序列改变引起,在DNA序列之外还有许多因素影响基因的表达<sup>[37]</sup>。如何反映外部因素的影响,应该是在明确外部影响形式和方法的基础上,结合系统生物学的观念来建立更完善的分析模型。考虑到生物系统的复杂性与非线性,对其功能表达完整的模型应该是非常复杂的<sup>[38, 41, 42]</sup>,需要考虑通过若干简单线性子模块的组合来实现。

还需要提到的是量子生物学,这是运用量子力学的理论、概念和方法研究生命物质和过程的一门学科。如果可以确定生物分子本身的化学结构或各级结构,就有可能研究和这种分子相联系的生物学活性的本质,或者它们之间的相互作用。量子生物学所研究的问题实际上涉及分子生物学的全部内容,量子信息技术(包括量子纠错编码)可以突破现有信息技术的物理极限,为信息科学的发展提供新

的原理和方法, 有可能为生物学研究提供特别的分析和解决能力<sup>[43, 44]</sup>, 这是值得关注的研究方法。

系统生物学的应用与研究也应特别关注。系统生物学是研究一个生物系统中所有组成成分(基因、mRNA、蛋白质等)的构成, 以及在特定条件下这些组分间的相互关系的学科, 其特点是以整体性研究为特征的一门新科学<sup>[45]</sup>。系统生物学的研究, 启发、要求研究人员应该从系统的角度对生物系统进行分析, 特别是对生命现象这样复杂的存在形式, 其控制、受控因素必然相当丰富。利用系统的分析方法, 可能综合多个功能模块、多种方式的影响。对于纠错编码理论, 不仅可研究多种编码方式的结合, 还可与其他多种理论分析方法相结合, 以更准确、完整地反映生命过程。

## 5 总结

由于生物系统具有信息存储和信息传递的特点, 并要求在存在干扰的环境(过程)中保证极高稳定性(极高纠错能力), 因此具有相同应用要求的信息理论、编码理论有理由成为生物研究分析方法中的重要理论和方法。生物系统的复杂性迫使我们多种研究理论相结合对其进行分析, 同时也启发我们借鉴它所蕴含的信息处理方式用于对通信编码的研究和改进, 这都需要开展更多更深入的研究工作。

## [参 考 文 献]

- [1] Adami C. Information theory in molecular biology. *Phys Life Rev*, 2004, 1(1): 3-22
- [2] Hanus P, Goebel B, Dingel J, et al. Information and communication theory in molecular biology. *Electr Eng*, 2007, 90: 161-73
- [3] 刘宏德, 吴建盛, 谢建明, 等. 信号处理技术在生物分子序列分析中的应用. *生物信息学*, 2008, 6(4): 180-2
- [4] Yockey H. *Information theory and molecular biology*. New York: Cambridge University Press, 1992: 102
- [5] Battail G. Should genetics get an information-theoretic education? *IEEE Eng Med Biol*, 2006, 25(1): 34-45
- [6] Battail G. Can we explain the faithful communication of genetic information? DIMACS Series in Discrete Mathematics and Theoretical Computer Science [M]// Slogol PH, SoliunIn E, van Wijngoorden. *Advances in Information Recording*, NJ USA, 2008: 79-103
- [7] Radman M. Fidelity and infidelity. *Nature*, 2001, 413(6852): 115
- [8] Endy D. Foundations for engineering biology. *Nature*, 2005, 438(7067): 449-53
- [9] Wang XH, Istepanian RSH, Geake T. Error control coding in microarray data analysis [C]. *Workshop on Genomic Signal Processing and Statistics*, 2004
- [10] Gonzalez DL, Giannerini S, Rosa R. Strong short-range correlations and dichotomic codon classes in coding DNA sequences. *Phys Rev E*, 2008, 78(5): 051918
- [11] Nicolas R, Miklós Bercsényi, László Orbán, et al. Duplication of fgfr1 permits fgf signaling to serve as a target for selection during domestication. *Curr Biol*, 2009, 19(19): 1642-7
- [12] Shpaer EG. Constraints on codon context in *Escherichia coli* genes—their possible role in modulating the efficiency of translation. *J Mol Biol*, 1986, 188(4): 555-64
- [13] 张静, 顾宝洪, 石秀凡, 等. 大肠杆菌基因中密码子前后碱基的使用与蛋白质结构. *生物物理学报*, 2001, 17(1): 174-80
- [14] 张静, 蔡军, 李衍达. 人类可变翻译事件的研究. *中国科学 C辑: 生命科学*, 2007, 37(2): 198-203
- [15] 赵辉, 李启寨, 李俊, 等. 相邻碱基组分与产生SNP的转换或颠换在植物基因组中的研究. *中国科学 C辑: 生命科学*, 2006, 36(1): 1-8
- [16] 卢欣, 陈惠民, 李衍达. 细菌DNA序列中的长程相关性. *清华大学学报: 自然科学版*, 1999, 39(7): 98-102
- [17] 赵小杰, 刘勇, 姚力, 等. 基于非线性方法的DNA序列分析. *生物物理学报*, 2004, 20(1): 37-42
- [18] Luo LF, Lee WJ, Jia LJ, et al. Statistical correlation of nucleotides in a DNA sequence. *Phys Rev E*, 1998, 58(1): 861-71
- [19] Gatlin LL. *Information theory and the living system*[M]. New York, NY: Columbia University Press, 1972
- [20] May EE, Vouk MA, Bitzer DL, et al. An error-correcting code framework for genetic sequence analysis. *J Franklin I*, 2004, 341: 89-109
- [21] Dawy Z, Gonzalez, Faruck MG, et al. Modeling and analysis of gene expression mechanisms a communication theory approach[C]. 2005 IEEE International Conference, 2005: 815-9
- [22] Dawy Z, Morcos F, Weindl J, et al. Translation initiation modeling and mutational analysis based on the 3'-end of the *Escherichia coli* 16S rRNA sequence. *Biosystems*, 2009, 96(1): 58-64
- [23] Bouaynaya N, Schonfeld D. Protein communication system: evolution and genomic structure. *Algorithmica*, 2007, 48(4): 375-97
- [24] Mac Donall DA. Molecular informatics: hydrogen-bonding, error-coding, and genetic replication[C]. 43rd Annual Conference on Information Sciences and Systems, Baltimore MD, 2009: 853-8
- [25] Khan AH, Ossadtchi A, Leahy RM, et al. Error-correcting microarray design. *Genomics*, 2003, 81(2): 157-65
- [26] Reif JH, Labean TH. Computationally inspired biotechnologies: improved DNA synthesis and associative search using error-correcting codes and vector-quantization[C]. *DNA computing. 6th International Workshop on DNA-Based Computers*, DNA, 2000, 2000: 145-2
- [27] May EE, Vouk MA, Bitzer DL. Classification of *Escherichia coli* K-12 ribosome binding sites. *IEEE Eng Med Biol*, 2006, 25(1): 90-7
- [28] Ponnala L, Bitzer DL, Vouk MA. On finding convolutional code generators for translation initiation of *Escherichia coli* K-12[C]. *Engineering in Medicine and Biology Society* 2003,

- Proceedings of the 25th Annual International Conference of the IEEE, 2003: 3854-7
- [29] 刘晓, 田逢春, 李素芳. 基于分组码模型的DNA序列编码特性分析. 生物信息学, 2009, 7(22): 28-31
- [30] 刘晓, 田逢春, 李素芳, 等. 一种利用卷积码模型分析遗传序列的方法. 西北农林科技大学学报: 自然科学版, 2010, 38(4): 207-14
- [31] Liu X, Tian FC, Wang SY. Analysis of similarity/dissimilarity of DNA sequences based on convolutional code model. Nucleos Nucleot Nucl, 2010, 29(2): 123-31
- [32] Liu YH, Tian FC, Liu X, et al. The robustness and fractal feature of the bio-genetics coding. Dynamics of Continuous, Discrete and Impulsive Systems, Series B: Application and Algorithms, 2007: 264-8
- [33] Liu Y, Chakrabartty S. Computer aided simulation and verification of forward error-correcting biosensors[C]. 2008 IEEE International Symposium on Circuits and Systems, Seattle, WA USA, 2008: 1826-9
- [34] May E, Lee MT, Dolan P, et al. Computational sensing and *in vitro* classification of GMOs and biomolecular events[C]. 26th Army Science Conference, 2008
- [35] Yukinawa N, Oba S, Kato K, et al. Optimal aggregation of binary classifiers for multiclass cancer diagnosis using gene expression profiles[C]. 4th International Symposium on Bioinformatics Research and Applications, Georgia State Univ, Atlanta, GA, 2009: 333-3
- [36] Zheng GY, Qian ZL, Yang Q, et al. The combination approach of SVM and ECOC for powerful identification and classification of transcription factor. BMC Bioinformatics, 2008, 9(1): 282-9
- [37] 曹更生, 柳爱莲, 李宁. 隐藏在基因组中的遗传信息. 遗传, 2004, 26(5): 714-20
- [38] 陈润生. 与生物信息学相关的两个前沿方向——非编码基因和复杂生物网络. 生物物理学报, 2007, 23(4): 290-5
- [39] Nogueira T, Rankin DJ, Touchon M, et al. Horizontal gene transfer of the secretome drives the evolution of bacterial cooperation and virulence. Curr Biol, 2009, 19(20): 1683-91
- [40] Morris PF, Schlosser LR, Onasch KD, et al. Multiple horizontal gene transfer events and domain fusions have created novel regulatory and metabolic networks in the oomycete genome. PLoS One, 2009, 4(7): e6133(1-12)
- [41] 徐自祥, 孙啸. 细胞代谢复杂网络研究进展. 生物信息学, 2009, 7(2): 120-32
- [42] 王青芸, 易慧, 刘来福, 等. 基因逻辑网络研究进展. 生物化学与生物物理进展, 2008, 35(11): 1239-46
- [43] Matsuno K, Paton RC. Is there a biology of quantum information? Biosystems, 2000, 55(1-3): 39-46
- [44] Schuster P. Free will, information, quantum mechanics, and biology. Complexity, 2009, 15(1): 8-10
- [45] Bennett MR, Hasty J. Systems biology - genome rewired. Nature, 2008, 452(7189): 824-5