

doi:10.3969/j.issn.1672-5565.2011.03.016

Codon Usage Database 自动寻错平台 CUDer 的组建与应用

余劲聪¹, 方柏山^{1,2*}

(1. 福建省高校工业生物技术重点实验室(华侨大学), 厦门, 361021;
2. 厦门大学化学化工学院化学工程与生物工程系, 厦门, 361005)

摘要: 密码子用法数据库(CUD)是密码子用法与密码子优化研究领域的一个重要的在线服务,为了找出该数据库中潜在的两种不再适用的记录,即已过期的陈旧记录和在遗传密码类型上实际无法有效支持的记录(简称不支持记录),本文通过结合前期自主研发的两个软件 BestCodon 与 CUDassist,组建了 CUD 自动寻错平台 CUDer,并应用于上述问题的研究。结果发现,CUD 中存在 317 条陈旧记录与 4 条不支持记录。对于陈旧记录,这些记录的物种分类号在 NCBI 中已发生变更,研究者应避免使用这些记录的相关数据;对于不支持记录,本文借助 CUDer 计算得到这些记录正确的密码子用法表(CUT),弥补了当前 CUD 的不足。此外,该平台也为面向 CUD 的自动化数据处理,提供了一个新的软件框架,具有一定的借鉴意义。

关键词: 密码子用法; 密码子用法表; 密码子用法数据库; CUDer 平台; 生物信息学

中图分类号: TP31; Q81 文献标识码: A 文章编号: 1672-5565(2011)-03-242-06

Construction and application of the auto - finding errors platform named CUDer for Codon Usage Database

YU Jin-cong¹, FANG Bai-shan^{1,2*}

(1. The Key Laboratory for Industrial Biotechnology of Fujian Higher Education Huaqiao University, Xiamen 361021, China; 2. Department of Chemical and Biochemical Engineering, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China)

Abstract: Codon Usage Database (CUD) is an important online server for researching the codon usage and codon optimization. This study aimed to locate two kinds of inapplicable records from the database, including both expired time-out records and unsupportable records, genetic code types of which were not supportable for CUD. The above work was accomplished by the auto-finding errors platform named CUDer, which mainly consisted of earlier developed softwares BestCodon and CUDassist. As a result, there were 317 time-out records and 4 unsupportable records in CUD. For the time-out records, compared with related data in NCBI, their Taxonomy ID had been changed were in the database, so researchers should be careful to use these records. For the unsupportable records, their codon usage tables (CUTs) were accurately available in this paper. In addition, the new software framework of CUDer could be referred to CUD-oriented automatic data processing.

Keywords: codon usage; codon usage table; Codon Usage Database; CUDer platform; Bioinformatics

在生物信息学领域,专业数据库和专业软件(包括在线服务)是两个不可或缺的基石。密码子

用法数据库^[1](codon usage database, CUD) (<http://www.kazusa.or.jp/codon>) 是一个专门提供密码子用

收稿日期: 2010-10-12; 修回日期: 2010-12-09.

基金项目: 国家自然科学基金资助项目(21076172), 国家自然科学基金资助项目(30770059), 高等学校博士学科点专项科研基金(20070385001), 福建省高校产学研合作科技重大项目(2010H6023)。

作者简介: 余劲聪, 博士研究生, 主要从事生物信息学与合成生物学研究, E-mail: yjincong@hqu.edu.cn.

* 通讯作者: 方柏山, 教授, 博士生导师, Tel: 0592-2185869, E-mail: fbs@xmu.edu.cn.

法表(codon usage table, CUT) 浏览和查阅的在线服务, 当前已收集了高达 35 799 条记录的 CUT 数据, 共涉及 3 027 973 条完整的蛋白质编码基因。该在线服务为相关研究^[2-5] 提供了极大的便利。同时, CUD 也成为其它在线服务的数据源, 如图形化密码子用法分析器^[6] (GCUA) (<http://gcu.schoedl.de>), 该在线服务基于 CUD 的 CUT, 为研究者提供密码子相对适应度(relative adaptiveness, RA) 参数值的计算, 以及两个密码子用法表的比对图形等, 是密码子用法可视化方面的一个典型应用。此外, CUD 在美国国家生物技术信息中心(NCBI) (<http://www.ncbi.nlm.nih.gov>)、瑞士生物信息学研究所(SIB) 的 ExPASy (<http://www.expasy.ch>) 等均设有专门的链接, 由此也可见其重要性。在我们以往的研究中发现, 该在线服务仍存在问题, 并初步提出了解决方案。在本文中, 我们将对两个问题进行探讨, 一是以静态数据库作为基础的 CUD, 其是否存在“过期”记录, 即以 NCBI 定期快速更新的核酸序列数据库为准, CUD 中是否有一些陈旧记录缺乏更新, 导致其信息上的不完整性或失效; 二是 CUD 仅支持 13 种遗传密码的类型, 但从 NCBI 可知, 目前已有 17 种类型, 因此 CUD 数据库中是否存在超出其支持范围的记录, 在实际上无法提供给研究者正确的 CUT 等。为了探讨这些问题, 我们在先前的研究基础^[7] 上, 组建了 CUD 自动寻错平台“CUDer” 现报道如下。

1 平台的组建

1.1 总体设计

以获取“陈旧记录索引表”(Index for time-out records) 和“遗传密码不支持型记录索引表”(简称“不支持记录索引表”(Index for genetic-code-un-supportable records) 为主要目的, 进行平台设计。首先是预处理, 从 NCBI 获取“物种分类号-遗传密码号(线粒体遗传密码号)”列表(TaxID-GCID(MGCID)) 和“遗传密码号”列表(GCID_ncbi), 再从 CUD 获取“物种分类号-序列类型”列表(TaxID-SeqType) 和“遗传密码号”列表(GCID_cud), 作为后续处理的支持性数据。其次是检索寻错, 通过比较两个或多个列表文件中, 关键字段数据的异同, 生成新的“物种分类号-遗传密码号”列表(TaxID-GCID) 文件, 以及得到最终的两个结果文件。因此, 本平台共涉及 2 个在线服务的数据库, 5 个过程文件, 及最终获得的 2 个主要结果。这些数据库及文件之间的关系, 如图 1 所示。此外, 在处理过程中数

据的自动获取、汇总和参数计算等, 由两个自主研发的软件 CUDassist 与 BestCodon^[7], 并结合辅助性的小程序来完成, 这些程序共同组成了 CUDer 平台。

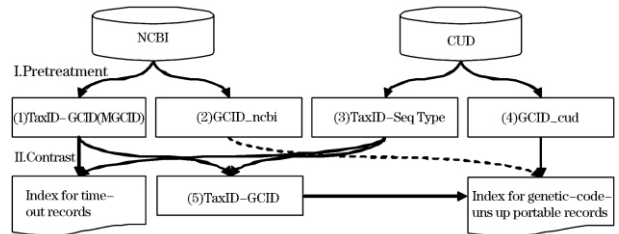


图1 密码子用法数据库自动寻错的处理流程

Fig. 1 The process flow for auto-finding errors in Codon Usage Database

1.2 预处理

在 NCBI 物种分类系统(Taxonomy) 的 FTP 分站点(<ftp://ftp.ncbi.nih.gov/pub/taxonomy>) 提供“物种分类号-遗传密码号(线粒体遗传密码号)”的对应关系列表, 其中收录了高达 658 423 个物种的数据, 将这一列表本地化后, 创建“TaxID-GCID(MGCID)”文件。值得注意的是, 该文件仅能提供一个物种或其线粒体与遗传密码类型的对应关系, 对于叶绿体等其它细胞器或基因组, 仍缺乏大规模的对应关系数据。这种情况限制了本文, 暂时只能对上述两种序列类型进行分析。另外, 该站点也提供已知的所有遗传密码类型的列表文件, 同样建立本地映射“GCID_ncbi”, 从中可知, 当前共有 17 种不同的遗传密码类型, 其遗传密码号分别为 1、2、3、4、5、6、9、10、11、12、13、14、15、16、21、22、23, 其中, 标准遗传密码的编号为 1。另一方面, 在 CUD 的 FTP 站点(<ftp://ftp.kazusa.or.jp/pub/codon/current>), 可根据相应的文件, 建立本地的“物种分类号-序列类型”列表文件(TaxID-SeqType), 从中可知, 当前 CUD 共收录了 35 792 条记录的数据。另外, 从 CUD 的 CUT 提供网页中, 可知当前该在线服务共支持 13 种遗传密码的类型(GCID_cud), 与“GCID_ncbi”比较, 缺少编号为 16、21、22、23 的四种遗传密码类型, 这是 CUD 中可能存在不支持记录的根本原因。

1.3 检索寻错

首先, 以列表 (3) TaxID-SeqType”中的物种分类号作为检索项, 逐一在列表 (1) TaxID-GCID(MGCID)”中进行检索, 如果能找到相同的分类号, 则说明列表(3)中该记录为非陈旧记录, 反之, 则为陈旧记录。对于非陈旧记录, 在序列类型为核基因组与线粒体的情况下, 可获取该记录对应的遗传密码号(或线粒体遗传密码号, 下文不再特别说明), 从而得到列表 (5) TaxID-GCID”列表(5)包含了

当前 CUD 收录的非陈旧记录,且序列类型为核基因组与线粒体的物种分类号和遗传密码号数据。对于陈旧记录,建立“陈旧记录索引表”。其次,以列表(5)中的遗传密码号作为检索项,逐一在列表(4)“GCID_cud”中进行检索,如果能找到相同的遗传密码号,则说明列表(5)中该记录为支持记录,反之,则为非支持记录。对于非支持记录,建立“非支持记录索引表”,而且对该索引表中每一条记录,逐一生成对应的 CUT。此外,以列表(2)“GCID_ncbi”中的遗传密码号作为检索项,逐一在列表(4)中进行检索,如果能找到相同的遗传密码号,则说明列表(2)中该记录为支持的遗传密码类型,反之,则为不支持的遗传密码类型。

2 平台的应用

2.1 CUD 的记录组成

在 CUD 中缺少对数据库中各种记录类型的说明,为了增加 CUD 的可利用性,并且为后续分析提供必要的参考数据,首先对 CUD 的记录组成进行了解析。结果表明,虽然 CUD 的公告记录为 35 799 条,但其实际记录要略少,为 35 792 条。在本文中涉及计算记录类型比例时,都以实际记录数为准。如表 1 所示,根据序列类型划分,数据库中包括 Nucleus、Mitochondrion、Chloroplast 等,共计 12 种类型的记录。其中,上述三种类型占据了记录集的绝大

多数份额,合计约 98.6%。可是,因为仅能从 NCBI 中获取 Nucleus 与 Mitochondrion 两种类型与其对应的物种分类号对应关系,所以在文本的后续分析中,只能对这二者,共计 27 073 条记录进行处理。此外,如表 2 所示,根据遗传密码类型划分,数据库中共涉及 16 种遗传密码。其中,“Standard Code”、“Vertebrate Mitochondrial Code”与“Bacterial, Archaeal and Plant Plastid Code”三种类型,各自所占的比例均超过 15%,合计约 70%。除了存在 317 条陈旧记录,无法辨别其遗传密码类型外,在已知的 17 种类型中,CUD 仅不涉及“Trematode Mitochondrial Code”一种遗传密码,可见其记录类型的多样性仍较为丰富。

表 1 根据序列类型划分的密码子用法数据库的记录组成

Table 1 Records constituted Codon Usage Database divided by types of sequences

Type of sequence	Number of record	Percentage %	Cumulative %
Nucleus	16 803	46.9	46.9
Mitochondrion	10 270	28.7	75.6
Chloroplast	8 233	23.0	98.6
Plastid	437	1.22	99.9
Kinetoplast	18	0.050 3	99.9
Chromoplast	12	0.033 5	99.9
Nucleomorph	9	0.025 2	100.00
Leucoplast	3	0.008 38	100.00
Apicoplast	2	0.005 59	100.00
Cyanelle	2	0.005 59	100.00
Proplastid	2	0.005 59	100.00
Secondary	1	0.002 79	100.00

表 2 根据遗传密码类型划分的密码子用法数据库的记录组成

Table 2 Records constituted Codon Usage Database divided by types of genetic codes

Genetic code ID and name	Number of record	Percentage %	Cumulative %
1 Standard Code	11 140	31.100	31.1
2 Vertebrate Mitochondrial Code	8 126	22.700	53.0
11 Bacterial, Archaeal and Plant Plastid Code	5 730	16.000	69.8
5 Invertebrate Mitochondrial Code	1 158	3.240	73.1
4 Mold Mitochondrial; Protozoan Mitochondrial; Coelenterate Mitochondrial; Mycoplasma; Spiroplasma Code	358	1.000	74.1
9 Echinoderm Mitochondrial; Flatworm Mitochondrial Code	128	0.358	74.4
6 Ciliate Nuclear; Dasycladacean Nuclear; Hexamita Nuclear Code	68	0.190	74.6
3 Yeast Mitochondrial Code	31	0.086 6	74.7
12 Alternative Yeast Nuclear Code	14	0.039 1	74.7
10 Euplotid Nuclear Code	12	0.033 5	74.8
13 Ascidian Mitochondrial Code	4	0.011 2	74.8
15 Blepharisma Macronuclear Code	3	0.008 38	74.8
16 Chlorophycean Mitochondrial Code	2	0.005 59	74.8
14 Alternative Flatworm Mitochondrial Code	1	0.002 79	74.8
22 Trematode Mitochondrial Code	1	0.002 79	74.8
23 Scenedesmus obliquus mitochondrial Code	1	0.002 79	74.8
21 Trematode Mitochondrial Code	0	0	74.8
- Time-out records	317	0.886	75.7

2.2 陈旧记录索引表

如表 3 所示,陈旧记录中包括 Nucleus、Mito-

chondrion 与 Chloroplast 三种类型,其各自的数量分别为 158 条、138 条及 21 条,共计 317 条,占总记录数的比例约为 0.886% (317/35 792)。由于当前能够获取的支持性数据不足,对于这些陈旧记录具体的变更原因和情况,尚无法确认。但是,该索引表为相关研究者提供了必要的参考,从而可避开具有潜

在错误的 CUT 数据。值得注意的是,本文定义的陈旧记录,仅为由于物种分类号变更而导致的记录过期情况,而对于非陈旧记录中序列数据的增减情况,同样因为缺乏支持性数据,而无法自动辨别,这将是今后需要进一步研究的方向。

表 3 密码子用法数据库中 317 条陈旧记录的索引表

Table 3 The index table for 317 time-out records in Codon Usage Database

Nucleus						Mitochondrion			Chloroplast		
591	30 555	65 017	111 610	241 707	352 473	3 053	63 175	126 868	227 170	325 334	3 053
592	30 604	69 272	113 374	258 443	362 393	3 058	67 897	126 869	228 695	331 343	49 629
601	31 233	70 540	117 552	264 654	371 647	4 916	70 609	126 870	228 910	331 345	78 476
602	32 017	72 150	119 024	269 278	382 253	4 928	73 413	127 257	238 858	338 845	83 837
1 315	32 555	72 204	129 363	273 302	393 047	7 751	75 364	135 628	240 476	340 057	87 242
3 053	33 549	74 495	129 818	280 359	395 879	8 836	75 367	137 063	243 320	342 379	100 155
4 916	34 521	74 567	134 640	283 163	399 127	9 156	75 444	137 254	247 392	342 389	100 166
5 662	35 136	75 444	136 995	284 686	399 128	9 457	79 950	144 758	259 055	369 601	101 620
6 352	36 232	75 912	137 787	288 957	399 365	9 610	80 433	144 760	259 393	374 628	121 786
6 353	36 347	76 762	138 953	290 521	400 106	9 862	80 436	147 057	260 625	374 629	142 343
6 644	36 428	77 791	149 049	292 565	411 359	10 174	86 216	154 809	261 078	374 633	175 768
6 988	36 429	78 450	152 465	292 817	415 528	27 589	90 063	166 752	265 732	381 123	184 704
7 605	36 431	78 881	160 458	298 551	415 961	27 729	90 706	169 829	268 764	388 984	217 843
7 692	38 315	79 888	164 673	298 966	418 142	30 492	95 117	169 830	279 963	388 995	247 890
8 159	38 437	81 421	176 094	298 972	435 896	30 555	101 873	171 946	284 519	388 996	270 118
8 209	41 255	81 933	177 186	299 449	440 076	30 604	101 875	172 529	287 708	388 998	306 378
8 564	42 065	82 078	180 465	299 728	447 324	32 514	107 230	176 094	287 715	389 000	362 393
9 457	46 155	83 837	182 091	303 481	447 398	32 555	112 598	176 940	297 821	389 008	393 532
9 594	47 231	84 641	184 118	326 330	36 235	115 591	177 142	297 822	389 013	401 073	
9 610	47 971	85 560	185 207	326 939	42 833	115 592	179 901	297 826	389 014	402 122	
10 110	48 740	86 172	189 832	331 626	42 836	115 596	179 913	297 829	389 022	402 142	
10 340	50 413	89 769	203 964	338 100	43 322	121 562	184 159	299 341	389 023		
10 504	57 478	90 706	210 165	338 104	52 219	122 226	187 422	303 484	389 025		
10 671	57 867	92 636	214 472	338 139	5 2437	122 227	194 936	311 353	449 315		
10 722	57 868	101 270	217 364	346 224	5 5540	122 228	210 165	311 357	449 316		
10 964	59 898	101 356	226 583	346 873	56 215	126 575	219 819	322 586	449 321		
11 275	60 887	105 504	226 891	348 774	56 223	126 614	220 549	322 587			
29 458	61 523	107 571	228 004	351 508	60 887	126 722	223 514	322 589			

2.3 不支持记录索引表

如表 4 所示,在 CUD 已收录的数据中,其在遗传密码上实际不支持的记录,共有 4 条,占总记录数的比例约为 0.011 1% (4/35 792)。显然,不支持记录所占的比例很小,一方面是因为 CUD 相对 NCBI 而言,其收录的记录总数较少(35 792/658 423),另一方面,CUD 不支持的 4 种遗传密码类型,通常其所对应的记录数也较少。但是,对于各种已知类型的遗传密码的支持,甚至对用户自定义遗传密码类型的良好支持,将是 CUD 这类在线服务应该具备的功能。这也为今后我们开发新的密码子用法数据库,提出了一个基本要求。本课题组在前期已自主研发了一款能支持各种遗传密码类型的软件 BestC-

odon^[7],同时结合密码用法数据库辅助工具 CUDassist,对上述 4 条不支持记录,通过计算给出正确转换后的 CUT,其结果如表 5 所示,这些 CUT 数据对当前的 CUD 是一个必要的补充。

表 4 密码子用法数据库中遗传密码不支持型记录的索引表

Table 4 The index table for four genetic-code-unsupportable records in Codon Usage Database

Taxonomy ID	Name of species	Type of sequence	Genetic code ID
3088	Scenedesmus obliquus	mitochondrion	22
42467	Thraustochytrium aureum	mitochondrion	23
60187	Rhizophyidium sp. 136	mitochondrion	16
109760	Spizellomyces punctatus	mitochondrion	16

表5 密码子用法数据库中4个遗传密码不支持型记录的密码子用法表

Table 5 The codon usage tables of 4 genetic - code - unsupported records in Codon Usage Database

codon	3088		42467		60187		109760		codon	3088		42467		60187		109760	
	AA	Fraction	AA	Fraction	AA	Fraction	AA	Fraction		AA	Fraction	AA	Fraction	AA	Fraction	AA	Fraction
AAA	K	0.850	K	0.840	K	0.94	K	0.950	GAA	E	0.830	E	0.670	E	0.880	E	0.820
AAC	N	0.350	N	0.180	N	0.049	N	0.180	GAC	D	0.330	D	0.130	D	0.056	D	0.150
AAG	K	0.150	K	0.160	K	0.064	K	0.049	GAG	E	0.170	E	0.330	E	0.120	E	0.180
AAT	N	0.650	N	0.820	N	0.950	N	0.820	GAT	D	0.670	D	0.870	D	0.940	D	0.850
ACA	T	0.450	T	0.350	T	0.330	T	0.250	GCA	A	0.440	A	0.230	A	0.230	A	0.280
ACC	T	0.140	T	0.070	T	0.044	T	0.150	GCC	A	0.073	A	0.077	A	0.046	A	0.097
ACG	T	0.056	T	0.056	T	0.017	T	0.036	GCG	A	0.067	A	0.071	A	0.046	A	0.046
ACT	T	0.360	T	0.520	T	0.610	T	0.570	GCT	A	0.420	A	0.620	A	0.680	A	0.570
AGA	R	0.064	R	0.250	R	0.650	R	0.530	GGA	G	0.510	G	0.160	G	0.340	G	0.180
AGC	S	0.11	S	0.068	S	0.024	S	0.061	GGC	G	0.084	G	0.064	G	0.025	G	0.055
AGG	R	0.470	R	0.150	R	0.043	R	0.034	GGG	G	0.085	G	0.180	G	0.052	G	0.066
AGT	S	0.450	S	0.250	S	0.300	S	0.300	GGT	G	0.320	G	0.600	G	0.580	G	0.700
ATA	I	0.049	I	0.280	I	0.240	I	0.200	GTA	V	0.400	V	0.210	V	0.510	V	0.470
ATC	I	0.080	I	0.100	I	0.048	I	0.170	GTC	V	0.047	V	0.074	V	0.052	V	0.059
ATG	M	1.000	M	1.000	M	1.000	M	1.000	GTG	V	0.120	V	0.071	V	0.038	V	0.096
ATT	I	0.870	I	0.620	I	0.710	I	0.630	GTT	V	0.430	V	0.640	V	0.400	V	0.380
CAA	Q	0.890	Q	0.770	Q	0.980	Q	0.910	TAA	*	0.100	*	0.330	*	0.850	*	0.590
CAC	H	0.450	H	0.190	H	0.078	H	0.230	TAC	Y	0.210	Y	0.200	Y	0.075	Y	0.190
CAG	Q	0.11	Q	0.230	Q	0.016	Q	0.092	TAG	L	0.160	*	0.061	L	0.083	L	0.110
CAT	H	0.550	H	0.810	H	0.920	H	0.770	TAT	Y	0.790	Y	0.800	Y	0.930	Y	0.810
CCA	P	0.340	P	0.170	P	0.400	P	0.380	TCA	*	0.880	S	0.170	S	0.240	S	0.180
CCC	P	0.075	P	0.058	P	0.023	P	0.073	TCC	S	0.061	S	0.054	S	0.036	S	0.062
CCG	P	0.035	P	0.043	P	0.027	P	0.066	TCG	S	0.0023	S	0.046	S	0.011	S	0.025
CCT	P	0.550	P	0.730	P	0.550	P	0.480	TCT	S	0.370	S	0.410	S	0.390	S	0.370
CGA	R	0.130	R	0.110	R	0.080	R	0.030	TGA	*	0.025	*	0.061	*	0.150	*	0.410
CGC	R	0.032	R	0.058	R	0.011	R	0.068	TGC	C	0.210	C	0.150	C	0.096	C	0.130
CGG	R	0.0039	R	0.058	R	0.000	R	0.025	TGG	W	1.000	W	1.000	W	1.000	W	1.000
CGT	R	0.300	R	0.380	R	0.220	R	0.320	TGT	C	0.790	C	0.850	C	0.900	C	0.870
CTA	L	0.039	L	0.110	L	0.024	L	0.150	TTA	L	0.011	*	0.550	L	0.810	L	0.460
CTC	L	0.048	L	0.033	L	0.0019	L	0.025	TTC	F	0.240	F	0.100	F	0.190	F	0.400
CTG	L	0.016	L	0.031	L	0.0058	L	0.026	TTG	L	0.500	L	0.400	L	0.017	L	0.071
CTT	L	0.220	L	0.430	L	0.058	L	0.150	TTT	F	0.760	F	0.900	F	0.81	F	0.60

Note: AA amino acid, * terminal signal.

3 总结与展望

密码子用法的研究是遗传密码与基因组的进化^[8-10]、基因的水平转移^[11]、基因表达的调控^[12]等领域不可或缺的内容,而密码子优化技术,在基因工程领域越来越多地得到应用,并已取得诸多成果^[13-16]。但是不论上述的何种研究方向,其共同的基础均是对CUT的解析或比较分析等。作为当前国际上重要的CUT提供者,CUD仍存在一些问題,本文针对陈旧记录与不支持记录两个问題,在以往的研究基础^[7]上提出了一个自动寻错平台CUDer,为面向CUD的自动化数据处理,提供了一个新的软件框架,具有一定的借鉴意义。研究结果发现,CUD中共存在317条陈旧记录与4条不支持记录,并给出了这些记录的物种分类号索引表,为相关研究者避开或慎用这些记录,提供了必要的参考。同时,对于CUD不支持的4条记录,本文中经计算给出其正

确的CUT数据,对当前的CUD来说,是一个必要的补充。随着测序技术的飞速发展,序列数据正在急剧地增加,从CUD与NCBI各自收录的物种数量之间的差距,可见以静态数据库为基础的CUD,暴露了其不可避免的问题。本文的研究为解决这些问题,提供了一个应急方案,但从长远来看,显然有必要建立一个以动态数据库为核心的、更为全面的“密码子用法数据库”,这也正是我们的下一个研究目标。

参考文献(References):

- [1] Nakamura Y, Gojobori T, Ikemura T. Codon usage tabulated from international DNA sequence databases: status for the year 2000 [J]. *Nucleic Acids Res*, 2000, 28(1): 292.
- [2] 尹春光, 杜立新, 赵桂平, 等. Mx 基因稀有密码子和 mRNA 结构及大肠杆菌表达优化[J]. *遗传*, 2009, 31(1): 75-82.
- [3] 廖登群, 张洪亮, 李自超, 等. 水稻 β -淀粉酶基因密码子使用特性分析[J]. *中国农业大学学报*, 2009, 14(5): 1-11.
- [4] 贺菽嘉, 肖绍文, 罗彬, 等. MAGE-D4a 融合蛋白的原核表达、纯化及鉴定[J]. *中国免疫学杂志*, 2010, 26(5): 432-435, 440.

(下转第249页)

在实验教学中,充分利用互联网、NewClass400多媒体网络教学系统的多媒体网络教室信息技术的教学管理功能,创造更为和谐的实验教学环境。教师可以利用多媒体网络教室的屏幕监视、语音监听等功能实现良好的教学管理。在实验教学过程中,教师可以在教师机通过屏幕监视和语音监听了解学生当前的学习状况,监督学生,并通过远程管理接管学生机,通过语音与学生进行单独交流,组织好课堂纪律。这种功能可使教师随时掌握课堂情况,并有针对性地单独进行管理,而不影响其它学生的正常学习状态。这种教学管理模式更有利于建立起和谐的师生关系,针对不同的学生个性特点因材施教。

3 结论与讨论

生物信息学实验教学是网络环境下生物教学的全新内容,是生物信息学教学的重要组成部分,关于生物信息学实验教学的研究还需要更多的实践探索与理论研究,闫晓红等探索了问题为导向的教学模式^[8]。我们经过多年的实验教学实践,通过优化实验教学内容,以综合性设计性实验内容为主,充分利用多媒体网络教室的网络化、开放性、交互性的优势初步构建了基于网络环境下的生物信息学实验教学环境,形成了以学生为主体的生态教学模式,极大地提高了教学效果,具有较强的可操作性和实践性。与此相适应的网络教学资源的建立则是更艰巨、长期的任务。生物信息学的实验教学依赖于发达的互

联网,在教学中,为使信息技术与实验教学进行有效的整合,在教育生态观的指导下,充分发挥信息技术的教育潜能,构建了开放性、网络化的生物信息学实验教学环境,创造良好的学习环境,实现了学生与学习环境、教师等多方面的互动交流,信息技术真正成为学生学习的工具,真正有效提高学生实践能力。实验教学在人才培养中具有特殊意义,本研究将成为生物信息学及其相关课程实验教学改革的重要参考依据。

参考文献(References):

- [1] 戴凌燕,姜述君,高亚梅.《生物信息学》课程教学方法探索与实践[J].生物信息学,2009,7(4):311-313.
- [2] 高亚梅,韩毅强,余丽芸.生态视角下生物信息学课堂教学的实践研究[J].黑龙江高教研究(高教研究与评估),2010,2:87-89.
- [3] 高亚梅,韩毅强,余丽芸,戴凌燕等.应用多媒体网络教室构建生物信息学教学新模式[J].黑龙江教育学院学报,2007,26(11):27-28.
- [4] 姚启伦,冉景盛.生物信息学教学模式初探[J].安康学院学报,2008,20(3):115-116.
- [5] 韩毅强,高亚梅,王景伟,余丽芸,刘金文.生物类多学科综合实验教学的实践研究[J].高校实验室工作研究,2009,4:19-20.
- [6] 蒋彦,王小行,曹毅等.基础生物信息学及应用[M].北京:清华大学出版社,2003.
- [7] 杨冬梅.课堂教学生态观阐释[J].黑龙江高教研究,2006,(4):164-165.
- [8] 闫晓红,王宁,滕晓华.生物信息学实验课教学改革与实践-PBL教学模式在实验教学中的应用与探讨[J].东北农业大学学报,2009,7(5):99-101.
- [9] 赵薇,陈必链,王明兹,等.藻类 $\Delta 6$ 脂肪酸脱饱和酶基因密码子偏好性分析[J].热带海洋学报,2010,29(3):126-134.
- [10] Fuhrmann M, Hausherr A, Ferbitz L, et al. Monitoring dynamic expression of nuclear genes in *Chlamydomonas reinhardtii* by using a synthetic luciferase reporter gene [J]. *Plant Mol Biol*, 2004, 55(6): 869-881.
- [11] Yu J, Shui L, Fang B. BestCodon: Novel software for codon bias analysis [J]. *J Biotechnol*, 2008, 136(S1): 349-350.
- [12] Hershberg R, Petrov DA. Selection on codon bias [J]. *Annu Rev Genet*, 2008, 42: 287-299.
- [13] 王芳平,李宏,刘国庆,等.密码对的偏爱与基因组进化的相关性分析[J].生物信息学,2009,7(2):150-154,158.
- [14] 周祎,陈洪俊,李明福,等.基于相对密码子频率的茺菁花叶病毒基因组序列的进化分析[J].生物信息学,2010,8(2):142-146,149.
- [15] Chater KF, Chandra G. The use of the rare UUA codon to define "expression space" for genes involved in secondary metabolism, development and environmental adaptation in streptomyces [J]. *J Microbiol*, 2008, 46(1): 1-11.
- [16] Hatfield GW, Roth DA. Optimizing scaleup yield for protein production: Computationally Optimized DNA Assembly (CODA) and Translation Engineering [J]. *Biotechnol Annu Rev*, 2007, 13: 27-42.
- [17] Gustafsson C, Govindarajan S, Minshull J. Codon bias and heterologous protein expression [J]. *Trends Biotechnol*, 2004, 22(7): 346-353.
- [18] Oeser T, Wei R, Baumgarten T, et al. High level expression of a hydrophobic poly(ethylene terephthalate)-hydrolyzing carboxylesterase from *Thermobifida fusca* KW3 in *Escherichia coli* BL21 (DE3) [J]. *J Biotechnol*, 2010, 146(3): 100-104.
- [19] 栾世家,潘蔚琦,李婷,等. A 型流感病毒 NS1 基因密码子去优化改造引起病毒毒力减弱[J].生物工学报,2009,25(5):720-726.
- [20] 张正平,杨江科,徐莉,等.扩展青霉脂肪酶基因克隆、密码子优化及表达[J].微生物学报,2010,50(2):228-235.

(上接第246页)