



# 湖南省自然科学奖提名书

(2019年度)

## 一、项目基本情况

提名单位(专家)		邵阳市科技局		
项目名称		非线性方法在分子生物数据分析中的应用研究		
主要完成人		黄国华, 喻祖国		
主要完成单位		邵阳学院, 湘潭大学		
省财政资金拨款单位		邵阳学院		
学科分类 名称	1	模式识别	代码	5202040
	2	应用数学	代码	11087
	3		代码	
所属科学技术领域		计算机跨学科应用		
任务来源		省、市、自治区计划, 国家自然科学基金		
具体计划、基金的名称和编号： 1. 国家自然科学基金面上项目：分形及相关方法在分子进化与蛋白质研究中的应用，项目编号：11371016 2. 教育部“长江学者和创新团队发展计划”创新团队：微分方程数值方法与生物计算，编号：IRT1179 3. 湖南省科技计划项目：基于正样本学习理论的miRNA靶基因识别研究，项目编号：2014FJ3013				
已呈交的科技报告编号：				
项目起止时间		起始：2010年1月1日	完成：2017年12月31日	

湖南省科学技术奖励工作办公室制



## 二、提名意见

提名单位	邵阳市科技局		
通讯地址	湖南邵阳大祥区城北路16号	邮政编码	422000
联系人	刘佑林	联系电话	0739-5360552
电子邮箱	249729375@QQ.COM	传 真	0739-5363522

**提名意见：**

我单位认真审阅了该项目推荐书及附件材料，确认全部材料真实有效，相关栏目均符合湖南省科学技术奖励工作办公室的填写要求。

该项目将机器学习、模式识别和复杂网络等非线性方法运用到序列比较、蛋白质功能、药物靶点和蛋白质翻译后修饰位点识别等分子生物数据分析中，具体有：（1）建立了基于离散隐马尔科夫模型的组蛋白赖氨酸巴豆酰化修饰位点识别的计算方法；（2）建立了融合非直接作用蛋白质功能信息的蛋白质功能预测方法，建立全局与局部网络性质保持的药物靶点计算预测方法；（3）提出了多个基于图形的生物序列特征提取方法，为序列比较奠定了理论基础；相关研究结果发表在Journal of Theoretical Biology, Match-Communications in Mathematical and in Computer Chemistry和Molecular Phylogenetics and Evolution 等SCI刊物上，并得到了相关领域国际顶级期刊及同行专家的广泛引用和好评。

提名该项目为湖南省自然科学奖二等奖。

声明：本单位遵守《湖南省科学技术奖励办法》及其实施细则的有关规定，承诺遵守评审工作纪律，所提供的提名材料真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。如有材料虚假或违纪行为，愿意承担相应责任并接受相应处理。如产生争议，保证积极调查处理。

提名单位（盖章）

年 月 日

### 提名项目等级（请在相应栏打“ ”进行选择）

一等奖		第一完成人签字：   年 月 日
二等奖		
三等奖		

说明：实行“定标定额”评审，一等奖评审落选项目不再降格参评二、三等奖，二等奖评审落选项目不再降格参评三等奖。评审公示后不受理对评审结果中有关提名等级的异议。请认真对照省自然科学奖授奖条件，填写提名等级。



### 三、项目简介

本项目在国家自然科学基金、教育部“长江学者和创新团队发展计划”创新团队、湖南省科技厅计划和湖南省教育厅优秀青年等项目资助下，将非线性理论和方法（模式识别、复杂网络和机器学习）应用到分子生物数据分析中，主要创新如下：

(1) 建立了基于隐马尔科夫模型组蛋白赖氨酸巴豆酰化修饰位点的计算方法，也是第一个计算预测组蛋白赖氨酸巴豆酰化修饰位点的方法和工具，为组蛋白赖氨酸巴豆酰化修饰位点快速筛选提供了理论和方法支持。

(2) 发现了水平可视网络存在多分形特性，分形维数是 Hurst 指数的线性函数；对于拉普拉斯算子，水平可视网络的次小特征值、第三小特征值和最大特征值的平均对数近似 Hurst 指数的线性函数，平均拉普拉斯能量近似 Hurst 指数的二次函数；对于规范的拉普拉斯算子，水平可视网络的次小特征值、第三小特征值平均对数近似 Hurst 指数的线性函数，而最大特征值的平均对数和平均拉普拉斯能量分别是 Hurst 指数的四次和三次多项式。以直接作用的蛋白质功能为基础，融合非直接作用的蛋白质功能，在蛋白质相互作用网络中建立了蛋白质功能识别的统计预测方法；以融合多方面关联知识的化合物-化合物关系构建化合物-化合物网络，建立了全局与局部网络性质保持一致的药物靶点的预测方法。为药物设计提供了方法和理论支持。

(3) 提出了免比对序列比较的框架：将序列转化为图或曲线，并从中提出特征，以特征形成的距离代替原始序列进行比较；提出了一种基于曲线的核苷酸波峰、波谷和中心点等明显几何特征的特征提取方法；提出了基于图的蛋白质序列特征提取方法，该特征提取方法满足平移、旋转不变性和尺度不变性；提出了序列无向图的内部分离族 DNA 和蛋白质序列特征提取方法。这些方法为进行快速序列比较提出了理论和方法支持。

本项目发表的 8 篇代表性论文 SCI 他引 34 次，总他引 xx 次。上述成果得到了 Genome Biology, Molecular Biology and Evolution 和 Genomics, Proteomics & Bioinformatics 等国际 SCI 期刊引用和好评。



## 五、客观评价

该项目的成果在分子生物数据处理等相关的交叉领域产生了很好的国际影响，8 篇代表性论文 SCI 他引 34 次，总他引 xx 次。成果得到了 Genome Biology, Molecular Biology and Evolution 和 Genomics, Proteomics & Bioinformatics 等国际 SCI 著名期刊引用。具体如下：

### 1. 对重要科学发现点 1 的评价：

SCI 期刊 Artificial Intelligence in Medicine(人工智能学科 1 区)发表的论文【Identify and analysis crotonylation sites in histone by using support vector machines, 2017, 83: 75-81】和 SCI 期刊 Journal of Molecular Graphics and Modelling (计算机图形与计算机辅助药物设计学科 2 区)发表的论文【Prediction of lysine crotonylation sites by incorporating the composition of k-spaced amino acid pairs into Chou's general PseAAC, 2017, 77: 200-204】对代表性论文 1 的首创性都给予了肯定“ In a pioneer work, Huang et al. used a discrete hidden markov model to detect histone crotonyllysine sites “和” Huang and Zeng proposed the first predictor, named CrotPred, for the prediction of histone crotonylation sites”.

### 2. 对重要科学发现点 2 的评价：

SCI 期刊 Physica A: Statistical Mechanics and its Applications 发表的论文【Spectral analysis for weighted tree-like fractals, 2018, 492: 1892-1900】对代表性论文 5 给予正面评价，说决定与分形相联系的规范拉普拉斯矩阵的特征值和特征向量取得了很大进步“ In determining the eigenvalues and eigenvectors of the normalized Laplacian matrix associated to their fractal representations, there has been great important progress.”.

SCI 期刊 IEEE transactions on nanobioscience 发表的论文【A new method for identifying essential proteins by measuring co-expression and functional similarity, 2016, 15(8): 939-945】对代表性论文 6 给予了积极评价“ Develop computational methods to predict the key features underlying mechanisms and functions based on molecular networks have become the major focus of many researchers”.

### 3. 对重要科学发现点 3 的评价：

中科院生物学科大类 1 区 Top 期刊 Genomics, Proteomics & Bioinformatics 发表的研究论文【Similarity Estimation Between DNA Sequences Based on Local Pattern Histograms of Binary Images, 2016, 14(2):103-112】引用了代表性论文 2，介绍了 DNA



序列可视化方法“Every base type in a DNA sequence is replaced by an individual vector in a two-dimensional (2D)”，并为了与代表性论文 2 的结果进行比较选择了同样的物种；

中科院生物大类 1 区 Top 期刊 Molecular biology and evolution 在 2013 年的论文【AGP: a multimethods web server for alignment-free genome phylogeny, 2013, 30(5), 1032-1037】介绍了代表性论文 2 的主要思想“Deng et al. (2011) and Huang et al. (2011) used natural vector and a 10-dimensional statistical vector to characterize the two-dimensional (2D) graphical DNA curve, they were called two-dimensional natural vector (2DNV) and two-dimensional statistical vector (2DSV), respectively.”;

著名数学家、美国数学学会 Fellow (Fellow of the American Mathematical Society)、美国伊利诺伊大学杰出教授 (UIC Distinguished Professor)、电气电子工程师学会 Fellow (IEEE Fellow)、约翰·西蒙·古根海姆奖 (John Simon Guggenheim Fellowship)、艾尔弗雷德·P·斯隆奖 (Alfred P. Sloan Fellowship) 清华大学著名统计学家 Stephen S. T. Yau 教授发表在中科院生物大类 2 区的论文【Virus classification in 60-dimensional protein space, 2016, 99, 53-62】和中科院生物大类 3 区的论文【Phylogenetic analysis of protein sequences based on a novel k-mer natural vector method, 2018】都对代表性论文 3 给予了积极评价：“Even though it was suggested that using proteome sequences was better than using whole genome DNA sequences for genome-based phylogeny reconstruction, these studies were only based on a specific Baltimore class or certain families of viruses in the National Center for Biotechnology Information database (NCBI) “和“Therefore, using protein sequences to analyze the phylogeny of species makes more sense than using DNA sequences”;

中科院生物大类 1 区 Top 期刊 Genome Biology 发表的综述【Alignment-free sequence comparison: benefits, applications, and tool, 2017, 18(1):186】中引用了代表性论文 2;

中科院合肥智能所智能计算实验室主任，同济大学特聘教授，英国利物浦约翰摩尔大学教授，国际模式识别学会会士黄德双教授在论文【Novel graphical representation of genome sequence and its applications in similarity analysis, 2012, 391(23), 6128-6136】中对代表性论文 2 给予正面评价“Lately, there are several similar approach to dealing with the primary sequences. Applying the above methods, the researchers have analyzed the similarities and dissimilarities of DNA primary sequences”。



## 六、代表性论文专著目录

序号	论文专著名称/刊名 / 作者	影响因子	年卷页码 (xx年 xx卷 xx页)	发表时间 (年月日)	通讯作者 (含共同)	第一作者 (含共同)	国内作者	SCI 他引次数	他引总次数	知识产权是否归国内所有
1	A Discrete Hidden Markov Model for Detecting Histone Crotonyllysine Sites/ Match-Communications in Mathematical and in Computer Chemistry/Guohua Huang, Wenfei Zeng	2.4	2016年75卷 717页	2016年 03月 31日	Guo hua Huang	Guo hua Huang	黄国华, 曾文飞	4	4	是
2	Alignment-free comparison of genome sequences by a new numerical characterization/Journal of Theoretical Biology/Guohua Huang, Houqing Zhou, Yongfan Li, Lixin Xu	1.93	2011年 281卷107页	2011年 04月 28日	Guo hua Huang	Guo hua Huang	黄国华, 周后卿, 李勇帆, 徐立新	16	16	是
3	Whole-proteome based phylogenetic tree construction with inter-amino-acid distances and the conditional geometric distribution profiles/Molecular Phylogenetics and Evolution/Xian-Hua Xie, Zu-Guo Yu, Guo-Sheng Han, Wei-Feng Yang, Vo Anh	4.23	2015年89卷 37页	2015年 04月 13日	Zu-Guo Yu	Xian - Hua Xie	谢显华, 喻祖国, 韩国胜, 阳卫锋	2	2	是
4	A 2D Pattern Matching Algorithm for Comparing Primary Protein Sequences/Current Bioinformatics/Guohua Huang, Weiping Xie, Yongfan Li, Lixin Xu, Houqing Zhou	0.55	2014年9卷 210页	2014年 05月 15日	Guo hua Huang	Guo hua Huang	黄国华, 黄卫平, 谢文平, 李勇帆, 徐立新, 周后卿	2	2	是
5	A novel genome signature based on inter-nucleotide distances profiles for visualization of metagenomic data/Physica A: Statistical Mechanics and its Applications/Xian-Hua Xie, Zu-Guo Yu, Yuan-Lin Ma, Guo-Sheng Han, Vo Anh	2.28	2017年 482卷87页	2017年 04月 20日	Zu-Guo Yu	Xian - Hua Xie	谢显华, 喻祖国, 马元林, 韩国胜	0	0	是



6	Multifractality and Laplace spectrum of horizontal visibility graphs constructed from fractional Brownian motions/Journal of Statistical Mechanics-Theory and Experiment/Zu-Guo Yu, Huan Zhang, Da-Wen Huang, Yong Lina, Vo Anh	1.65	2016年 2016卷 033206页	2016年 03月 18日	Zu- Guo Yu	Zu- Guo Yu	喻祖国, 张欢, 黄达文, 林勇	5	5	是
7	A Novel Neighborhood Model to Predict Protein Function from Protein- Protein Interaction Data/Current Proteomics/Guohua Huang	0.54	2014年11卷 237页	2014年 12月 31日	Guo hua Hua ng	Guo hua Hua ng	黄国华	4	4	是
8	Large-Scale Prediction of Drug Targets Based on Local and Global Consistency of Chemical-Chemical Networks/Combinatorial Chemistry & High Throughput Screening/Guohua Huang, Kaiyan Feng, Xiaomei Li, Yan Peng	0.94	2016年19卷 121页	2016年 01月 05日	Guo hua Hua ng	Guo hua Hua ng	黄国华, 冯开颜, 李晓媚, 彭艳	1	1	是
合计								34	34	

补充说明（视情填写）：

承诺：上述论文专著用于提名湖南省自然科学奖的情况，已征得未列入项目主要完成人和主要完成单位的其他作者和其他署名单位的同意。知识产权归国内所有，且不存在争议。

第一完成人签名：



## 七、代表性论文专著被他人引用的情况

序号	被引代表性论文专著序号	引文题目/作者	引文刊名/影响因子	引文发表时间 (年月日)
1	1	Identify and analysis crotonylation sites in histone by using support vector machines/Wang-Ren Qiu, Bi-Qian Sun, Hua Tang, Jian Huang, Hao Lin	Artificial intelligence in medicine/3.2	2017年01月25日
2	2	Alignment-free sequence comparison:benefits, applications, and tool/Andrzej Zielezinski, Susana Vinga, Jonas Almeida, Wojciech M. Karlowski	Genome Biology/12.66	2017年10月03日
3	2	Similarity Estimation Between DNA Sequences Based on Local Pattern Histograms of Binary Images/Yusei Kobori, Satoshi Mizuta	Genomics, Proteomics & Bioinformatics/5.04	2016年04月27日
4	2	AGP: a multimethods web server for alignment-free genome phylogeny/Yongkun Li, Kun Tian, Changchuan Yin, Rong Lucy He, Stephen S.-T. Yau	Molecular Phylogenetics and Evolution/4.23	2016年03月15日
5	4	Quantitative Analysis of Genetic Relationship of Species Based on the Network for Protein-Coding DNA Sequences/Yan Peng, Yuewu Liu, Sichun Ruan, Yulin Wei	MATCH Communications in Mathematical and in Computer Chemistry/2.4	2016年12月31日
6	6	Spectral analysis for weighted tree-like fractals/Meifeng Dai, Yufei Chen, Xiaoqian Wang, Yu Sun, Weiyi Su	Physica A: Statistical Mechanics and its Applications/2.28	2017年12月05日
7	7	A new method for identifying essential proteins by measuring co-expression and functional similarity/Wei Zhang, Jia Xu, Xiong Li, Xiufen Zou	IEEE transactions on nanobioscience /2.46	2016年08月31日
8	3	Virus classification in 60-dimensional protein space/Yongkun Li, Kun Tian, Changchuan Yin, Rong Lucy He, Stephen S.-T. Yau	Molecular Phylogenetics and Evolution/4.23	2016年03月15日



## 八、主要完成人情况表

姓 名	黄国华	性 别	男	排 名	1	国 籍	中国
出生年月	1978年10月25日			出生地	湖南祁东	民 族	汉族
身份证号	430426197810257076			归国人员	否	归国时间	
技术职称	教授			最高学历	研究生	最高学位	博士
毕业学校	上海大学			毕业时间	2015年06月13日	所学专业	信息学与系统生物学
电子邮箱	guohuahhn@126.com			办公电话	15211971672	移动电话	15211971672
通讯地址	湖南邵阳市大祥区学院路邵阳学院信息工程学院					邮政编码	422000
工作单位	邵阳学院					行政职务	副院长
二级单位	信息工程学院					党 派	中国共产党
完成单位	邵阳学院					所在地	湖南、邵阳
						单位性质	学校
参加本项目的起止时间		自 2010年01月01日 至 2017年12月31日					
<p>对本项目主要学术贡献：            是该项目主要负责人，对发现点1、2、3均有重要贡献，建立了生物序列可视化方法和基于生物序列图形的特征提取方法，建立了基于蛋白质相互作用网络的非邻里蛋白质功能预测模型，建立了基于化合物相互作用网络的全局与局部特性保持的药物靶点预测方法，建立了基于隐马尔科夫模型的组蛋白赖氨酸巴豆酰化修饰位点预测方法，是代表性论文1、3、5、6、8的第一作者和通信作者。在该项目中投入的工作量达90%。</p>							
<p>曾获科技奖励情况：</p>							
<p>声明：本人同意主要完成人排名,遵守《湖南省科学技术奖励办法》及其实施细则的有关规定，承诺遵守评审工作纪律，保证所提供的有关材料真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。该项目是本人本年度被提名的唯一项目（团队）。如有材料虚假或违纪行为，愿意承担相应责任并接受相应处理。如产生争议，保证积极配合调查处理工作。</p> <p style="text-align: center;">本人签名：</p> <p style="text-align: center;">年 月 日</p>				<p>完成单位声明：本单位确认该完成人情况表内容真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。如产生争议，愿意积极配合调查处理工作。</p> <p>工作单位声明：本单位对该完成人被提名无异议。</p> <p style="text-align: center;">单位（盖章）</p> <p style="text-align: center;">年 月 日</p>			



姓 名	喻祖国	性 别	男	排 名	2	国 籍	中国
出生年月	1970年09月29日		出 生 地	湖南株洲		民 族	汉族
身份证号	43030319700929201X		归国人员	是		归国时间	2008年08月30日
技术职称	教授		最高学历	研究生		最高学位	博士
毕业学校	复旦大学		毕业时间	1997年06月30日		所学专业	基础数学
电子邮箱	yuzg@xtu.edu.cn		办公电话	0731-58293676		移动电话	18673428866
通讯地址	湖南省湘潭市湘潭大学数学与计算科学学院					邮政编码	411105
工作单位	湘潭大学					行政职务	教务处处长
二级单位	数学与计算学院					党 派	中国共产党
完成单位	湘潭大学					所 在 地	湖南
						单位性质	学校
参加本项目的起止时间	自 2010年01月01日 至 2017年12月31日						
<p>对本项目主要学术贡献：            对重要科学发现中第*项做出了创造性贡献。            主要负责时间序列、复杂网络多重分形算法研究，物种亲缘分析无序列比对算法的研究及软件开发，与第一申请人黄国华教授经常讨论生物信息问题研究。是代表作[2,4,7]的第一作者或通讯作者。在该项研究中的工作量占本人工作量的百分比为50%。</p>							
<p>曾获科技奖励情况：1. 2009年、2016年均获湖南省自然科学二等奖（均排名第1）；            2. 2009年获湖南青年科技创新奖杰出奖（单独）；            3. 2010年获批湖南省“芙蓉学者计划”特聘教授；            4. 2014年获批湖南省首届享受特殊津贴专家；            5. 2018年获批享受国务院政府特殊津贴专家、入选湖南省121人才工程第二层次人选。</p>							
<p>声明：本人同意主要完成人排名,遵守《湖南省科学技术奖励办法》及其实施细则的有关规定，承诺遵守评审工作纪律，保证所提供的有关材料真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。该项目是本人本年度被提名的唯一项目（团队）。如有材料虚假或违纪行为，愿意承担相应责任并接受相应处理。如产生争议，保证积极配合调查处理工作。</p> <p style="text-align: center;">本人签名：</p> <p style="text-align: center;">年 月 日</p>				<p>完成单位声明：本单位确认该完成人情况表内容真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。如产生争议，愿意积极配合调查处理工作。</p> <p>工作单位声明：本单位对该完成人被提名无异议。</p> <p style="text-align: center;">单位（盖章）</p> <p style="text-align: center;">年 月 日</p>			



## 九、主要完成单位情况表

单位名称	邵阳学院				
排 名	1	法定代表人	彭希林	所 在 地	湖南
单位性质	学校	传 真	0739-5431794	邮政编码	422000
通讯地址	湖南邵阳市大祥区学院路				
联 系 人	毛颖玲	单位电话	0739-5431794	移动电话	13762880788
电子邮箱	hnsyxykjc2009@163.com				
对本项目科技创新和推广应用情况的贡献：					
<p>1. 本项目所发表的代表性论文[1,2,4,7,8]第一署名单位为邵阳学院。</p> <p>2. 本项目完成期间获得了邵阳学院人力、物力、时间等多方面支持，并部分获得经济上的资助和奖励。</p>					
<p>声明：本单位同意完成单位排名，遵守《湖南省科学技术奖励办法》及其实施细则的有关规定，承诺遵守评审工作纪律，保证所提供的有关材料真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。如有材料虚假或违纪行为，愿意承担相应责任并接受相应处理。如产生争议，保证积极配合调查处理工作。</p>					
<p>单位（盖章）</p> <p>年 月 日</p>					



单位名称	湘潭大学				
排 名	2	法定代表人	周益春	所 在 地	湖南
单位性质	学校	传 真	0731-58292061	邮政编码	411105
通讯地址	湖南省湘潭市湘潭大学科技处				
联 系 人	李朋	单位电话	0731-58292063	移动电话	13170327151
电子邮箱	dsr@xtu.edu.cn				

对本项目科技创新和推广应用情况的贡献：

1. 主要完成成员喻祖国在承担本项目时工作单位为湘潭大学。
2. 本项目所发表的代表作[2,4,7]的第一作者或通讯作者发表论文时签署了湘潭大学的名字。
3. 本项目完成期间获得了湘潭大学人力、物力、时间等方面的支持。并部分获得经济上的资助和奖励。

声明：本单位同意完成单位排名，遵守《湖南省科学技术奖励办法》及其实施细则的有关规定，承诺遵守评审工作纪律，保证所提供的有关材料真实有效，且不存在任何违反《中华人民共和国保守国家秘密法》和《科学技术保密规定》等相关法律法规及侵犯他人知识产权的情形。如有材料虚假或违纪行为，愿意承担相应责任并接受相应处理。如产生争议，保证积极配合调查处理工作。

单位（盖章）

年 月 日

## 主要完成人合作关系说明

本项目第一完成人黄国华在邵阳学院工作，第二完成人喻祖国在湘潭大学工作。他们科学研究方向一致，合作紧密。黄国华经常参与喻祖国课题组的讨论。此外，他们也经常通过邮件、电话或直接面对面讨论生物序列特征提取与蛋白质复杂网络特性等问题。

第一完成人签名：黄国华