

# 浙江省生物信息学学会

---

## 2016 年首届“学会杯”生物信息学竞赛

浙江省生物信息学学会  
Bioinformatics Society of Zhejiang Province

# 解码生命的奥秘

Bioinformatics: Decoding the mysteries of life

首届“学会杯”生物信息学大赛



### 一、大赛背景

21 世纪是生物信息大数据的时代，生物信息学已是自然科学的核心领域之一。为了培养汇聚更多的生物信息学人才，激发师生和相关专业人员创新创业实战能力，加快生物信息学在科研及应用领域的普及。本学会联合浙江省各高校，科研单位及生物信息相关企业举办首届“学会杯”生物信息学大赛。

### 二、大赛描述

大数据分析、精准医疗、胎儿健康早筛、突发传染病监测、超级细菌致病基因查找等诸多领域都离不开生物信息学的运用。此次大赛主要面向高校在校生（包括硕士生、博士生）和各界社会相关领域人才，旨在培养学生的数据分析、大数据动手能力，调动学生的学习兴趣；挖掘高校大数据人才，发现大数据优质项目，为企业输送数据分析与大数据人才，为学生创造就业机会，鼓励高校大数据创新创业。希望通过大赛培养发掘更多的生物信息人才，让参赛队员接触科研和社会实践中的生物信息学问题，为科研和生物企业的发展培养提供更精锐的人才。

### 三、参赛规则

- (1) 组委会命题：各参赛队选择其中一题，通过优化的方案设计和程序运用解决问题。
- (2) 优秀自主项目提交、征集。
- (3) 参赛团队需由 3-5 人组成，设队长一名。不限制学科背景，不限地域。
- (4) 大赛评审委员会将根据参赛作品挑选 10 个队伍进入初赛，评定其中三个最强队进入决赛。
- (5) 参赛者提交的参赛作品必须是由参赛者独立完成的原创作品，不得抄袭，不得违反任何相关法律法规。
- (6) 大赛组委会所提供的数据集和平台仅限于本次大赛使用，不得用于其他任何目的，如若因此给数据提供方或平台提供方造成损失，由参赛选手负责。
- (7) 参赛作品的知识产权等相关权益归参赛队伍所有，但获奖作品赛后须在开源网站上公开发布，最终解释权归学会所有。

### 四、大赛题目

大赛设置八个备选题目，参赛团队可选择一个题目或多个题目，需使用自己擅长的软件及算法，提交源代码及报告并给出解决方案；经竞赛组委会专家组判定，以最优方式解决问题者获胜。

#### (1) 掌纹识别

本学会向参赛者提供 CASIA 掌纹数据库，共 301 个样本，5239 张照片。参赛者需要提交两个可用 Win32 控制台控制的应用程序。两个程序都需要能接受命令行参数，并输出文本文件。

第一个程序名为‘ENROLL\_[team name]’，可以录入一个掌纹图片，并输出一个临时文件记录其特征。命令行格式为：

ENROLL\_[team name] imageFilePathtemplateFilePathoutputFilePath

[team name]	参赛队名，全英文
imageFilePath	录入图片文件路径，格式同 CASIA 数据库
templateFilePath	临时文件输出路径
outputFilePath	一个文本文件，在文件末尾添加一行空格分隔的记录： imageFilePathtemplateFilePath result。其中‘result’标记为‘OK’或‘FAIL’，代表此程序是否成功运行

第二个程序名为‘MATCH\_[team name]’，可以读取一个用于匹配的掌纹图片，利用 ENROLL 程序生成的临时文件，计算并输出一个 0 到 1 之间的实数，代表与录入掌纹的匹配程度，数值越大表示越匹配。命令行格式为：

MATCH\_[team name] imageFilePathtemplateFilePathoutputFilePath

[team name]	参赛队名，全英文
imageFilePath	匹配图片文件路径，格式同 CASIA 数据库
templateFilePath	ENROLL 程序生成的临时文件路径

outputFilePath	一个文本文件，在文件末尾添加一行空格分隔的记录： imageFilePathtemplateFilePath result similarity。其中 'result' 标记为 'OK' 或 'FAIL'，代表此程序是否成功运行；similarity 为一个 0 到 1 之间的实数，代表与录入掌纹的匹配程度，数值越大表示越匹配
----------------	---

两个程序都只能使用当前输入数据进行计算，不能利用先前输入的数据或生成的临时文件。不能调用网络资源。

组委将根据平均 EER、ENROLL 程序失效率、MATCH 程序失效率、ENROLL 程序最长运行时间、MATCH 程序最长运行时间、ENROLL 程序最大使用内存、MATCH 程序最大使用内存、最大临时文件大小进行综合评定。

CASIA 保留对数据的所有权利

附现今常用掌纹数据库

CASIA Palmprint Database

Size: 301 subjects, 5,239 images

All images are captured all by common web cameras in a contact-less palmprint recognition system, which is firstly developed by CASIA.

PolyU Palmprint Database 2nd

Size: 386 palms, 7,752 images

Around twenty samples from each of these palms were collected in two sessions, where around 10 samples were captured in the first session and the second session, respectively. The average interval between the first and the second collection was two months.

UST Hand Image Database

Size: 5,660 images of 283 subjects

All images contain complete hand regions, which reserve hand shape information precisely. All hand image are color images in JPEG format.

## (2) OMIM 的 API

题目描述：利用 OMIM 提供的 API (<http://www.omim.org/help/api>) 进行数据的下载与整理。

编程语言要求：不限（要求形成脚本或程序，预留 API 参数）

结果展示：请提供如下截图中所列的结果。

Section	Description
animalModel	Animal Model
biochemicalFeatures	Biochemical Features
clinicalFeatures	Clinical Features
clinicalManagement	Clinical Management
cloning	Cloning
cytogenetics	Cytogenetics
description	Description
diagnosis	Diagnosis
evolution	Evolution
geneFamily	Gene Family
geneFunction	Gene Function
geneStructure	Gene Structure
geneTherapy	Gene Therapy
geneticVariability	Genetic Variability
genotype	Genotype
genotypePhenotypeCorrelations	Genotype/Phenotype Correlations
heterogeneity	Heterogeneity
history	History
inheritance	Inheritance
mapping	Mapping
molecularGenetics	Molecular Genetics
nomenclature	Nomenclature
otherFeatures	Other Features
pathogenesis	Pathogenesis
phenotype	Phenotype
populationGenetics	Population Genetics
text	Text (unfielded text section at the start of the entry)

详细内容及方法请参考：<http://www.omim.org/help/api>

### (3) 23andMe 的 API

题目描述：利用 23andME 提供的 API (<https://api.23andme.com/docs/>) 进行数据的下载与整理。

编程语言要求：不限（要求形成脚本或程序，预留 API 参数）

结果展示：请提供 Genetics, Phenotypes, Ancestry 三大类里的数据下载结果展示。

参考链接如下：<https://api.23andme.com/docs/reference/>

(4) 编写应用程序，寻找一段或者多段 DNA 序列，其在细菌种内是保守的，在种间是特异的。即所找到的序列能特异的代表某一个菌种，不会和其他菌种混淆，可以是单独的一段序列，也可以是多段序列的组合。【Tip：当某一个菌种内不存在唯一的代表序列时，可以使用多段序列组合起来行使代表序列的功能】。

1. 可以是命令行程序，编程语言不限，计算时间可以忍受；
2. 每段序列大于等于 200bp；
3. 测试数据给了两个不同的细菌序列文件，每个文件就是一个细菌，每个文件中有该细菌不同株系的序列。文件之间叫“种间”，文件之内叫“种内”。
4. 测试数据链接：链接：<http://pan.baidu.com/s/1miClgBe> 密码: 3sun

(5) 测序数据文件（我们提供测序数据以及物种信息）是否存在污染？详细描述并实现判断方法和污染结论。 数据下载链接如下：

S03P1.R2.fastq.gz 链接: <http://pan.baidu.com/s/1qYPAgpe> 密码: fk4m  
S03P1.R1.fastq.gz 链接: <http://pan.baidu.com/s/1kUII3yv> 密码: e86s  
S02P1.R2.fastq.gz 链接: <http://pan.baidu.com/s/1dFG3e5F> 密码: 9hgs  
S02P1.R1.fastq.gz 链接: <http://pan.baidu.com/s/1c2NWHna> 密码: p5tr  
S01P1.R2.fastq.gz 链接: <http://pan.baidu.com/s/1geXJH9h> 密码: urm4  
S01P1.R1.fastq.gz 链接: <http://pan.baidu.com/s/1c2FjnZE> 密码: 7j4j  
md5.txt 链接: <http://pan.baidu.com/s/1nvw9YQX> 密码: 9pms

(6) 单细胞研究目前是比较热的领域，尤其是单细胞转录组研究。正在成为胚胎发育，癌症等疾病研究的利器。请查阅相关文献，从不同角度评估目前单细胞转录组表达定量的多种工具，选择出其中较优定量工具。建立一套从 Hiseq 原始数据处理到差异表达基因筛选的 pipeline。如果能针对单细胞的特性，再附带个性化应用研究的模块部分，可以加分。(参考资料“Software tools for single-cell tracking and quantification of cellular and molecular properties”，<https://research.cchmc.org/pbge/sincera.html#downloads>，利用已发表的单细胞转录组文章数据做测试)。

(7) 在基因研究的文章中我们常常需要对一个或者多个基因的结构进行展示，目前已有几款在线软件可以做结构图，但是效果都比较差，不能直接放在文章正文中。希望能开发一个好用的在线工具，帮忙做科研的老师做出比较漂亮的基因结构展示图。如下图所示结果：（输入文件最好能支持 gff3, bed 等常见基因坐标文件格式，支持同时展示多个基因的结果。可以添加结构外的特征图，比如进化树的拓扑结构，FPKM/RPKM 表达数据，）可以参考目前已有的在线工具，用他们的 example 数据作调试。<http://gsds.cbi.pku.edu.cn/> 美化效果之类的可以参考 <http://itol.embl.de/> <http://www.evolgenius.info/evolview/#login> 的风格。（如果开发的工具好用的话，可以发一篇工具类的文章，比如 Bioinformatics 杂志等。）

(8) 自拟题：参赛队伍自主提交的攻关题目，需经竞赛组委会评定。  
注:4,5 题用到的数据可以从”学会杯大赛” QQ 群下载。

## 五、大赛时间及报名

(1) 报名截止日期： 2016 年 8 月 1 日  
所有报名参赛队伍加入学会大赛 QQ 群下载报名表，需以 3-5 人为一小组填写报名表，统一发送至 [ssping@zju.edu.cn](mailto:ssping@zju.edu.cn) 斯老师； [huangfl@zju.edu.cn](mailto:huangfl@zju.edu.cn) 黄老师， QQ: 7950200。  
组委会初步审核通过即报名成功。



学会杯大赛 QQ 群  
219176940



微信公众号

(2) 方案提交截止日期：2016 年 9 月 20 日

参赛队伍需提交解决方案（软件源代码，PPT 讲解视频及其他支撑材料），由大赛组委会评审。

(3) 初赛评选日期：2016 年 10 月 6 日

筛选 10 支优秀参赛队伍，进入初赛答辩阶段，公平、公开、公正的公布初赛结果。  
初赛地点：浙江大学紫金港校区生命科学学院生物信息中心实验室。

(4) 总决赛及颁奖仪式：2016 年 10 月 15 日（学会年会）

答辩现场根据参赛队的算法原理、历史成绩、评委打分，角逐出冠亚季军 3 只团队，复赛中余下的团队获得创新鼓励奖等企业奖项并颁发获奖证书。届时还将有学术界专家教授评委点评作品。

## 六、大赛奖项

(1) 冠军	联川奖	奖励人民币 10000 元
亚军	北角奖	奖励人民币 8000 元
季军	晶佰奖	奖励人民币 5000 元
最具潜力奖	厚泽奖	奖励人民币 2000 元
最具创意奖	普天奖	奖励人民币 2000 元
最高效率奖	天科奖	奖励人民币 2000 元
最佳团队奖	可帮奖	奖励人民币 2000 元
最佳人气奖	中翰金诺奖	奖励人民币 2000 元
最佳应用奖	艾吉泰康奖	奖励人民币 2000 元
特别奖	基因帮奖	奖励人民币 2000 元

(2) 荣誉证书：大赛冠亚季军及其他奖项获得组委会颁发的荣誉证书

(3) 报名成功并提交解决方案者即可获得学会定制的精美纪念品一份。

## 七、大赛主办单位

主办单位：

浙江省生物信息学学会  
浙江大学生命科学学院

协办单位:

浙江理工大学  
湖州师范学院  
杭州联川生物技术有限公司  
浙江省微生物研究所（浙江天科）

## 八、大赛组委会联系方式：

学会官方网址：[www.zjbioinformatics.org](http://www.zjbioinformatics.org)

斯老师 0571-88206134 转 8210      13805789825

黄老师 0571-88206593                      13666652591（同微信号）      QQ: 7950200

## 九、大赛组委会成员

主任    : 陈铭

副主任 : 邵朝纲、朗秋蕾

成员   : 包家立、陈欢、丁维龙、杭兴宜、贺平安、胡骏、黄方亮、刘箴、樊龙江、  
          魏瑾、孙晋华、王振峰、向太和、徐程、徐清华、姚玉华、赵怀、郑玲燕

秘书处 : 徐程、斯舒平、沈小仙

## 十、赞助单位



---

浙江省生物信息学学会秘书处

2016年05月30日印发

---