

# 王凯

电话: +86 13426113372 E-mail: jiang0112006c@163.com

## 个人信息

姓名: 王凯	生日: 1986. 11. 8	籍贯: 北京
性别: 男	政治面貌: 共青团员	民族: 汉

## 教育背景

- |                   |        |         |    |
|-------------------|--------|---------|----|
| ● 2009.9 – 2012.7 | 北京化工大学 | 化学工程与技术 | 硕士 |
| ● 2005.9 – 2009.7 | 北京化工大学 | 生物工程    | 学士 |

## 专业技能

- 熟练掌握支持向量机, 神经网络等数据挖掘算法和分类器应用建立定量、定性模型; 在 kaggle 上应用 xgboost, SVM, RandomForestClassifier 等参与竞赛项目。Kaggle 主页 <https://www.kaggle.com/kaiwang0112006>
- 熟练掌握 python 和 R 语言, 用以应用实现数据统计分析方法, 作图展示和建模预测 (numpy, scipy, pandas, sklearn, matplotlib)。具有 HTML, CSS, Javascript 编程能力, 有在 Django 和 Pyramid 框架下搭建网站的经历;
- 掌握 Mysql 数据库使用; 熟练使用 MongoDB 数据库的开发使用
- 熟练使用 github 完成代码版本控制, 分享和其他服务。  
Pelican+github 实现静态页面个人博客 <http://kaiwang0112006.github.io/>  
Jupyter+Mathjax+github 完成统计报告 [http://kaiwang0112006.github.io/simplejupyter/stroop\\_git.html](http://kaiwang0112006.github.io/simplejupyter/stroop_git.html)  
github 主页 <https://github.com/kaiwang0112006>
- 具有二代测序数据分析能力, 包括序列的比对和拼接等; 熟悉基因组拼接软件: ABySS, SOAP, Celera Assembler 等, 转录组拼接软件: Trinity, Oases 等
- 掌握 Linux 操作系统使用, 熟练使用 shell, 熟悉 pbs 作业调度系统
- 拥有较好的英语听说读写能力, 通过大学英语六级考试 (500 分), 雅思(6.5)

## 科研经历

- 基于乙酰胆碱酯酶的分类预测模型研究;  
课题以乙酰胆碱酯酶为靶标设计具有高活性的抑制剂分子。从文献和数据库中收集和整合已知的化合物数据信息, 利用支持向量机和自组织神经网络方法建立分类和预测模型。
- 针对极光激酶靶标和已有的极光激酶分子抑制剂建立分类预测模型;  
在课题后期加入, 完成模型评价部分的方法实现, 以验证模型的可靠性和鲁棒性。
- HCV 聚合酶抑制剂的分类模型研究;

实现了结合 Random Forest, F-score 等方法进行特征值的选择优化, 以及模型评价方法。

## 工作经历

- 2012.8.1-至今 在中科院遗传与发育生物学研究所工作 (助理工程师);
  1. 2012.8.1-2013.4 主要工作包括二代测序数据分析, 主要完成短序列拼接以及序列比对分析, 并通过序列比对和统计分析等手段, 完成对各种作物的表型调控的基因水平研究。  
完成项目: 橡胶树转录组组装 (组装结果分析已完成文章发表); 柠条转录组组装。
  2. 2013.5-至今, 基因型与表型数据的整合数据展示平台开发工作。数组库中包含水稻种质资源信息, 以及基因型和表型信息, 大部分数据来自 3000 株水稻基因组项目 (The 3000 Rice Genomes Project)。完成水稻种质, 表型和基因型的展示和检索功能, 包括品系名称, 表型名称, 基因以及 snp 的检索等功能。同时还在进行挖掘工具的开发, 目前主要针对从水稻表型数据的值域关系定位相关水稻品系。(www.mbkbase.org)  
参与项目: 数据库建设及新算法开发 (科学院战略性先导科技专项)
- 2011.6-2011.9 在保诺科技 (北京) 有限公司 计算化学部实习。收集大型化合物数据库中的数据 (DrugBank, NCI, Zinc), 建立化合物结构数据库。使用 Pipeline Pilot 软件开发供公司内部使用的化合物结构检索, 结构性质计算和数据库枚举的数据展示界面;

## 发表文章

- Dejun Li, Rizhong Zeng, Yan Li, Manman Zhao, Jinquan Chao, Yu Li, **Kai Wang**, Lihuang Zhu, Wei-Min Tian, Chengzhi Liang, Gene expression analysis and SNP/InDel discovery to investigate yield heterosis of two rubber tree F1 hybrids, **Scientific Reports**, 2016. 6:24984 (IF= 5.578)
- **Kai Wang**, Xiaoying Hu, Zhi Wang, Aixia Yan\*, Classification of Acetylcholinesterase Inhibitors and Decoys by a Support Vector Machine, **Combinatorial Chemistry & High Throughput Screening**, 2012. 15(6), 492-502 (SCI 论文 IF=2.62)
- Aixia Yan\*, **Kai Wang**, Quantitative structure and bioactivity relationship study on human acetylcholinesterase inhibitors, **Bioorganic & Medicinal Chemistry Letters**, 2012, 22(9), 3336-3342 (SCI 论文 IF=2.661)
- Maolin Wang, **Kai Wang**, Aixia Yan\*, Changyuan Yu\*. Classification of HCV NS5B polymerase inhibitors using Support Vector Machine. **Int.J.Mol.Sci**, 2012, 13(4), 4033-4047. (SCI 论文 IF=2.279)
- Aixia Yan\*, Xianglei Nie, **Kai Wang**, Maolin Wang, Classification of Aurora Kinase inhibitors by Self-Organizing Map (SOM) and Support Vector Machine (SVM), **Eur. J. Med. Chem**, 2012 (accepted, SCI 论文 IF=3.193)
- Aixia Yan\*, Yang Chong, Liyu Wang, Xiaoying Hu, **Kai Wang**, Prediction of Biological Activity of Aurora-A Kinase Inhibitors by Multilinear Regression Analysis and Support Vector Machine, **Bioorganic & Medicinal Chemistry Letters**, 2011,21(8), 2238-43. (SCI 论文 IF=2.661)
- **王凯**, 阎爱侠\*. 基于因特网的常用化合物活性数据库简介. 计算机与应用化学, 2010, 27(12), 1629-1632.