

过敏原综合数据库 COMPARE 的介绍和应用

赵 岚¹, 高中山^{1,2*}

(1. 浙江大学农业与生物技术学院, 杭州 310058; 2. 浙江大学医学院过敏研究中心, 杭州 310058)

摘 要: 随着分子生物学和基因组、蛋白质组技术的高速发展, 越来越多的过敏原蛋白得到鉴定。为了提高对过敏原的研究及其在食品安全上的应用, 很多机构都建立了不同类型的过敏原数据库。本文详细介绍了由美国健康与环境科学研究所(Health and Environmental Sciences Institute, HESI)协调组织国际合作团队建立的一个过敏原综合数据库 COMPARE (COMprehensive Protein Allergen REsource), 该数据库通过高通量序列筛选算法结合专家人工审核方法, 从各个公共蛋白质数据库、其他过敏原数据库以及相关文献中筛选过敏原, 每年更新一次。此外还开发了基于 FASTA 算法的序列比对工具 COMPASS (COMPare Analysis of Sequences with Software), 允许用户进行实时序列比对。该数据库已广泛应用于新过敏原蛋白的鉴定以及蛋白质潜在致敏性的评价, 对于食品安全管理和保障有重要参考价值。本文通过对 COMPARE 数据库的全面介绍, 旨在提高该数据库在我国的应用价值, 推动过敏原分子生物信息学的研究以及食品安全的发展。

关键词: 过敏原; 数据库; COMPARE; 生物信息学; 序列比对; 交叉反应; 食品安全

Introduction and application of the COMPARE allergen database

ZHAO Lan¹, GAO Zhong-Shan^{1,2*}

(1. College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China;
2. Allergy Research Center, School of Medicine, Zhejiang University, Hangzhou 310058, China)

ABSTRACT: With the development of molecular biology and genomes and proteomic technologies, more and more allergen components have been identified. In order to improve the efficiency of the research and the application in the food safety, several institutions have established various types of allergen databases. This study introduced a comprehensive allergen database named COMPARE (COMprehensive Protein Allergen REsource) which was developed by an international collaborative team coordinated by the Health and Environmental Sciences Institute (HESI), USA. The allergens listed in the COMPARE database were identified via a high-throughput, automated sequence sorting algorithm and manual curation of the annotated sequence entries retrieved from public protein sequence databases, other allergen databases, and the relevant scientific literatures on a yearly basis. And COMPASS (COMPare Analysis of Sequences with Software), a built-in FASTA-based sequence comparison tool, was also developed, supporting real-time sequence comparisons. This database had been widely used in the identification of new allergen components and assessment of the potential protein allergenicity, contributing to food safety assurance and management. Through the comprehensive introduction of COMPARE database, this study aimed to improve the application value of COMPARE database in China, and promoted its utilization in bioinformatics of allergen

基金项目: 国家重点研发计划项目(2019YFE0106600)

Fund: Supported by the National Key Research and Development Plan of China (2019YFE0106600)

***通信作者:** 高中山, 博士, 教授, 主要研究方向为过敏原分子基础和诊断应用。E-mail: gaozhongshan@zju.edu.cn

***Corresponding author:** GAO Zhong-Shan, Ph.D, Professor, College of Agriculture and Biotechnology, Zhejiang University, Hangzhou 310058, China. E-mail: gaozhongshan@zju.edu.cn

molecules research and food safety.

KEY WORDS: allergen; database; COMprehensive Protein Allergen REsource; bioinformatics; sequence alignment; cross-reactivity; food safety

0 引言

过去二十年间,随着经济的发展、环境的变化和人们生活方式、饮食结构等的改变,各种食物过敏问题呈现增加态势^[1]。常见的过敏反应是由免疫球蛋白 E (immunoglobulin E, IgE) 抗体介导的一种获得性免疫疾病,又称 I 型超敏反应,指机体对过敏原(大多是蛋白质)的过度免疫应答而引发的一系列皮肤、呼吸道、胃肠道、心血管、神经系统等症状,通常在接触过敏原的几分钟到两小时内出现^[2-4]。过敏原是诱发过敏反应的关键物质,一般分布在少数几个蛋白家族中。同一蛋白家族的过敏原具有相似的结构和功能^[5],不同来源的过敏原之间存在着广泛的交叉反应,即已经对某一过敏原过敏的个体,在接触相似过敏原时而诱发的过敏反应。其发生的分子基础是不同来源的过敏原在氨基酸序列或立体结构上具有相似性,可以被同种 IgE 抗体识别并结合,进而激发机体产生过敏症状^[6]。因此可通过序列和立体结构的比对,对某些蛋白的潜在致敏性进行预测。根据世界卫生组织和粮食及农业组织(World Health Organization and Food and Agriculture Organization, WHO/FAO)的指南,在连续 80 个氨基酸中,与已知过敏原序列等同度超过 35%,可认为该蛋白具有潜在的致敏性^[7]。随着致敏物质来源的增加和蛋白组学技术的快速发展,越来越多的过敏原得到鉴定。此外,需要进行潜在致敏性评价的物质,如新型食品、转基因作物等种类也在快速增长。为了提高过敏原研究的效率,并让研究

成果更好地服务于社会,不同类型的过敏原数据库应运而生^[8-9]。本文主要介绍了过敏原综合数据库 COMPARE (COMprehensive Protein Allergen REsource)的发展历程、数据收录标准和使用方法,旨在提高该数据库在过敏原分子生物信息中的研究及食品安全领域的应用。

1 过敏原数据库概况

目前可以使用的公共过敏原数据库较多,其中最常用的包括由世界卫生组织下属的国际免疫学联合会所建立的过敏原系统命名数据库 WHO/IUIS (International Union of Immunological Societies, <http://www.allergen.org/>)^[10],由内布拉斯加大学食物过敏研究和资源项目 FARRP (Food Allergy Research and Resource Program)所支持建立的 Allergen Online 数据库(<http://www.allergenonline.org/>)^[11],由德克萨斯大学医学分部所建立的过敏原结构数据库 SDAP (Structural Database of Allergenic Proteins, <https://fermi.utmb.edu/>)^[12],还有 Allergome (<http://www.allergome.org/>)^[8,13],过敏原家族数据库 AllFam (<http://www.meduniwien.ac.at/allfam/>)^[5],过敏原表位数据库 IEDB (Immune Epitope Database, <http://www.iedb.org/>)^[14],以及由我国广州医科大学附属第二医院的陶爱林建立的过敏原查询数据库 Allergenica (<https://allergenica.gzhmu.edu.cn/>)^[15]等,这些过敏原数据库所收录的数据有一定程度的重叠,但其面对的目标用户有差异,且由于功能、过敏原收录标准和网站维护更新频率的不同而存在较大的差异^[9]。各个过敏原数据库之间的对比如表 1 所示。

表 1 不同过敏原数据库的比较

Table 1 Comparison of the different allergen databases

数据库名称	主要特征	更新频率	运营与维护机构	创建时间
WHO/IUIS	过敏原分子的医学官方命名数据库,包含过敏原序列及致敏性相关信息,有严格的收录标准;用户自主提交鉴定的过敏原	持续更新	世界卫生组织下属的国际免疫学联合会	2000 年
Allergen Online	收录过敏原需要通过专家审查;主要包含过敏原序列及相关文献信息;数据库可下载,但部分过敏原缺乏致敏性信息	年度更新	内布拉斯加大学食物过敏研究和资源项目	2005 年
Allergome	包含过敏原序列及相关文献,但没有经过严格审查,且部分过敏原命名与 IUIS 不一致	持续更新	意大利过敏原资源实验室	2003 年
SDAP	主要收录过敏原分子结构相关信息,也具有序列比对工具	2013 年之后无更新	德克萨斯大学医学部	2002 年
IEDB	主要收录过敏原分子的表位信息,也可以对过敏原的表位进行预测,但用户界面相对复杂	季度更新	美国国家过敏和传染病研究所	2006 年
COMPARE	过敏原综合数据库,包含过敏原序列、相关文献、致敏性等信息,具有 COMPASS 在线序列比对工具;数据收录有较为严格的流程和标准	年度更新	美国健康与环境科学研究所	2016 年

COMPARE (<https://comparedatabase.org/>)是由美国非营利性组织—健康与环境科学研究所(Health and Environmental Sciences Institute, HESI)于 2016 年协调组织建立的, 包含过敏原氨基酸序列、相关文献以及序列比对等信息和功能的过敏原综合数据库。该数据库能用于搜索查询已知的过敏原氨基酸序列及其相关文献, 分析不同过敏原的交叉反应, 评估新蛋白的潜在致敏性, 服务于科研、食品安全管理、农业和食品加工等产业^[16]。

相比于其他公共过敏原数据库, COMPARE 的优势在于: 1) 该数据库由非营利性组织 HESI 监督和协调, 致力于促进与人类健康相关的科学问题的解决, 并且聘请了过敏学临床、基础研究以及生物信息学等多个领域的知名专家, 以公开、透明的原则进行有关过敏原的数据审核、收录以及数据库功能的更新; 2) 利用专业生物信息学手段, 采取高通量、序列自动分选算法对呈现指数式增长的蛋白质序列进行筛选, 并在此基础上进行人工检查分析, 提高过敏原筛选、收录的效率; 3) 该过敏原数据每年更新维护一次, 不断提高信息的准确度、完整性以及序列比对工具的精确度; 4) 用户无需下载任何应用软件, 即可在网站中进行快速的序列比对和数据结果的下载。

2 COMPARE 数据库的建立与发展

COMPARE 数据库建立之初, 由生物信息学领域的专家设计了专门的算法, 以覆盖多个过敏相关词汇(Allergen, Allergy, Allergenicity 等)的‘Allerg’为查询字从美国国家生物技术信息中心 NCBI (National Center for Biotechnology Information)蛋白质序列数据库(该数据库

涵盖了从 GenBank、RefSeq 和 TPA 中编码序列翻译而来的蛋白序列, 以及其他数据库如 SwissProt、PIR、PRF、PDB 等)中搜索与过敏相关的动物、植物、真菌等的蛋白序列, 初步去除与过敏原不相关的序列后, 再进行人工审核, 然后将筛选到的过敏原与 AllergenOnline v.16 数据库进行了比对测试, 发现部分过敏原如抑制蛋白等, 由于其蛋白序列的关键词中不包含‘Allerg’而未被搜索到, 因此又根据此缺陷对该算法进行了补充。2016 年版的 COMPARE 中共包含 1970 个相关过敏原条目。之后, COMPARE 数据库保持每年更新一次的频率, 不断完善过敏原搜索算法, 并扩大了搜索范围, 从最初的公共蛋白质数据库逐步拓展到 IUIS、AllergenOnline 等过敏原数据库, 并从已发表的过敏原相关的文献中搜索新发现的过敏原。

在 2019 年版本的 COMPARE 中加入了在线序列比对功能, 新开发了 COMPASS (COMPare Analysis of Sequences with Software)工具。使用 FASTA 搜索算法来评估过敏原之间的序列相似度, 允许 3 种形式的实时序列比对, 即序列全长比对, 80 个氨基酸序列滑动窗口比对以及 8 个氨基酸滑动窗口比对(图 1)。

用户可以采用默认参数或手动设置参数, 将目标蛋白序列与 COMPARE 数据库中收录的过敏原进行实时在线序列比对, 并能快速得到并下载比对结果。2020 年 7 月, 该数据库又新开发了一个可视化的组件, 允许用户以图像的形式快速查看序列比对结果。2022 年 1 月更新的 COMPARE2022 是第六版, 包含了 2463 个过敏原及其信息(表 2)。

COMPARE
Comprehensive Protein Allergen Resource

Analysis of Sequences with Software (COMPASS)

About Updates Contact Visit COMPARE Database

COMPare Analysis of Sequences with Software

The COMprehensive Protein Allergen REsource (COMPARE) database is a peer-reviewed, publicly accessible, database of allergen amino acid sequences. A key feature of the COMPARE allergen database is its capability in aiding an assessment of potential allergy health hazards, by using comparative sequence software.

COMPARE is now equipped with a comparative

NOTE: The COMPASS tool does NOT cache personal identification data or automatically store information. Users must download the appropriate reports/files in order to keep a record of work

- Full Length Sequence Search 序列全长比对
- 80-mer Sliding Window FASTA Search 80个氨基酸滑动比对
- 8-mer FASTA Search 8个氨基酸滑动比对

图 1 COMPASS 序列比对工具网络界面

Fig.1 COMPASS, a sequence comparison tool on the web page

表 2 COMPARE 数据库的建立与发展过程
Table 2 Establishment and development process of the COMPARE database

年份	数据库更新内容	数据库版本	释放日期	收录的过敏原数量
2016	起始阶段, 开发基础算法, 从 NCBI 的蛋白质数据库中搜索过敏原	COMPARE2017	2017.02.03	1970
2017	完善算法, 并扩大过敏原搜索范围, 从文献中筛选过敏原	COMPARE018	2018.02.16	2038
2018	开发了序列比对功能	COMPARE2019	2019.01.18	2081
2019	对历史数据进行审核	COMPARE2020	2020.01.29	2248
2020	序列比对结果可视化	COMPARE2021	2021.01.29	2348
2021	更新了已收录过敏原在 IUIS 数据库中的对应信息	COMPARE2022	2022.01.26	2463

3 COMPARE 数据库过敏原的审核与收录

COMPARE 数据库具有规范的数据审核流程和标准, 通过生物信息学手段从各种蛋白质数据库和 IUIS 等过敏原数据库以及已发表的相关文献中搜索到的准过敏原需要经过同行评审后才能确定是否可以被接收(图 2)。COMPARE 数据库的评审小组由过敏学相关的基础或临床研究领域中的国际知名专家组成, 各专家分别独立地完成对各个准过敏原审核流程, 经专家们达成共识的准过敏原才能被收录到数据库中。决定一个准过敏原是否被纳入 COMPARE 数据库的关键依据之一是该过敏原必须有明确的与患者血清中 IgE 抗体的结合数据, 此类证据可以来自已经发表的相关文献, 或未发表但已经过同行评审的文章等资料。对准过敏原的 IgE 结合能力进行审核时需要综合考虑所用实验方法, 患者血清以及过敏原分子来源(粗提物、天然或重组的过敏原单一组分)等条件。实验方法必须科学合理, 且包含阴性对照。所用患者血清需来源于对候选过敏原来源物种有明确过敏反应史的患者, 或对候选过敏原来源物种有潜在的交叉反应的患者, 例如, 若准过敏原来源于某一豆科植物, 但临床上没有收集到对该豆类过敏患者的血清, 使用对黄豆等近缘种过敏的患者血清进行 IgE 结合能力的测定也是可行的方法之一。如果准过敏原的 IgE 结合能力是通过粗提物与患者血清的免疫印迹检测得到的, 则该过敏原在免疫印迹图中必须能够完全与其他蛋白分开, 且包含完整的序列信息; 如果使用天然纯化的过敏原组分进行 IgE 结合能力的评价, 则需要对该蛋白的纯度进行说明; 此外, 通过质谱所鉴定到的多肽(包含 10 个以上的氨基酸)且有明确的辅助说明资料的, 可以作为单独的过敏原序列被收录。凡是在 IgE 结合能力检测中缺少相应的阴性对照血清, 或蛋白纯度达不到标准等的准过敏原都将会被剔除, 但未被 COMPARE 数据库收录的过敏原当拥有足够的相关数据后可重新被收录。

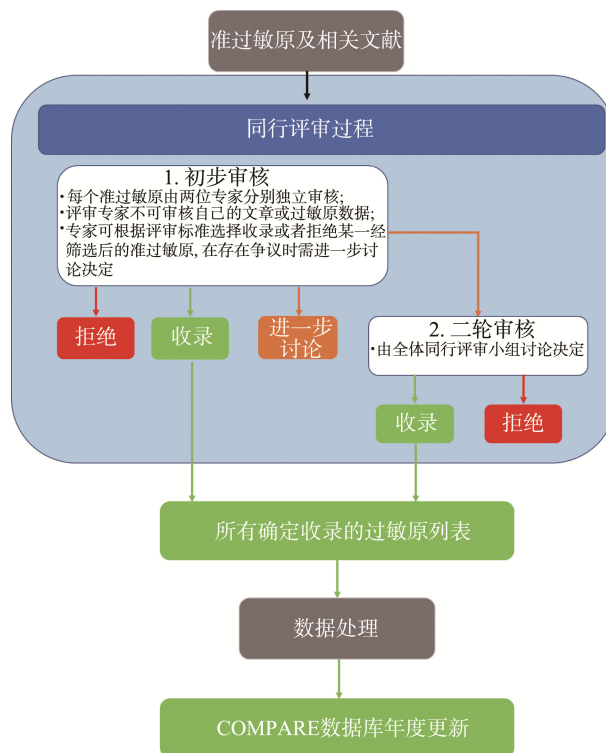


图 2 COMPARE 数据库中候选过敏原的同行评议流程
Fig.2 Process of COMPARE peer review cycle and workflow

评审小组每年对所有经筛选后的准过敏原进行审核的结果都以公开、透明的方式记录在 COMPARE 数据库的文件中, 所有用户均可以在数据库主页的“文档”选项卡中进行访问^[17]。

4 COMPARE 数据库的应用

经过 6 年的发展和改进, COMPARE 数据库现已成为集过敏原序列、文献检索及序列比对为一体的过敏原综合数据库。此数据库可为基础和临床研究人员、相关产品开发商、食品监管机构以及想要了解、学习过敏原相关知识的人提供服务。COMPARE 数据库的应用功能可分为 3 部分, 其一是过敏原的检索, 其二是过敏原序列比对, 最后是数据库的下载。

(1)过敏原检索

COMPARE 数据库主页中有一个“搜索”对话框, 用户可在该对话框中输入需要检索的过敏原名称、来源物种、蛋白名称、蛋白序列登录号等某一信息, 即可快速得到检索结果, 如在该对话框中输入蒿属花粉“Artemisia”, 一共得到了来自不同蒿属花粉中的 68 条过敏原记录, 每条记录中都包含过敏原来源物种及其通用名称, 蛋白质简要描述, 该过敏原在 IUIS 数据库中的名称, 序列号, 序列长度, 接收年份相关信息(图 3a)。点击需要进一步了解的过敏原记录后的“浏览”, 可看到该过敏原所对应的详细信息, 包括蛋白序列以及相关的文献(图 3b)。

COMPARE 数据库中过敏原分子的登录号大多来自于该分子在公共蛋白质数据库(如 SwissProt、PIR、PRF、PDB 等)中的登录号, 但也有一小部分尚未被任何公共蛋白质数据库收录的过敏原序列, 由 COMPARE 数据库赋予特定的登录号, 标识为“COMPARExxxxx”。每个过敏原分子都对应唯一的登录号, 用户可以用该登录号在公共蛋白质数据库或 COMPARE 数据库中对过敏原分子进行精确检索, 以获取更多相关信息。

(2)序列比对

COMPARE 数据库中的序列比对功能可执行 3 种形式的氨基酸序列比对。以最常用的序列全长比对工具为例, 在序列框中输入 FASTA 格式的询问序列, 数据库中默认的序列比对参数 E 值为 10, 字长为 2, 空位开放 (Gap open) 和扩展罚值(Gap extension penalties)分别为 10 和 2, 用户可以自行调整相应参数(图 4a), 序列比对结果文件中包含 3 部分内容, 第一部分是查询序列与 COMPARE 数据库中过敏原序列比对的概括信息总表, 包括匹配的过敏原序列号、简要描述、来源物种、序列相似度、E 值等(图 4b); 第二部分是 FASTA 搜索输出的原始序列比对格式(图 4c); 第三部分是比对结果可视化彩图(图 4d), 所有的结果数据都可以在该页面中下载为 PDF 文件。根据序列比对结果即可判段某个蛋白是否具有潜在的致敏性或与某一过敏原之间是否存在潜在的交叉反应性, 这对新型食品或转基因作物的安全性评价有着重要意义。据统计, 目前该数据库被来自 90 个国家地区的 4000 多个用户使用, 多数在经济合作与发展组织 (Organization for Economic Cooperation and Development, OECD)内, 也被 WHO/FAO 用于评价转基因食品安全评价。



图 3 COMPARE 数据库中蒿花粉过敏原的检索(a)及结果(b)

Fig.3 Searching the *Artemisia* pollen allergens in COMPARE database (a) and the result interface (b)

a Full Length Sequence Search ? 序列全长比对

Input Sequence: 输入比对序列

```

Mor-PR-10
MGVFTYETESTSIIPPAKLFKAFLLDADNLIIPKVASQAIAKTSEIVEGGPGTIKKITFG
YHTNGDQEIKEEQIKAGKEAAGLFRKAVESYLLANPEAYN
  
```

FASTA Algorithm Input Parameters

E-Value Threshold for Search: 10 (Default) E值
 Word Size: 2 (Default) 字长
 Gap Open & Gap Extension Penalties: 10/2 (Default) 空位开放和扩展罚值

Execute the FASTA Search of your Input Sequence against COMPARE ?

运行

b

Accession #	Description of COMPARE Allergen Alignment	Genus species	% Identity	Overlap	E-Value
ACJ23865.1	Allergen Fag s 1 pollen allergen [Fagus syl (160 aa) Show More	Fagus sylvatica	78.8%	160	7e-57
ACJ23864.1	Allergen Fag s 1 Show More	Fagus sylvatica	78.8%	160	2.3e-56
ACJ23866.1	Allergen Fag s 1 pollen allergen [Fagus syl (160 aa) Show More	Fagus sylvatica	78.1%	160	5.2e-56
APD76154.1	Allergen Jug r 5 Show More	Juglans regia	76.9%	160	3.9e-55
CAD10374.1	Putative ypr10 [Castanea sativa] (160 aa) Show More	Castanea sativa	77.5%	160	7.6e-55
AAB97141.1	Allergen Pru ar 1 Show More	Prunus armeniaca	73.1%	160	1.6e-52

c

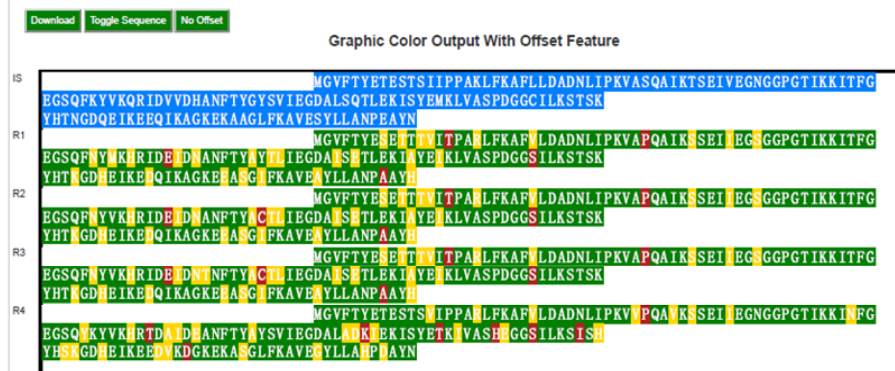
>>ACJ23864.1 Allergen Fag s 1; pathogenesis related prot (160 aa)
 initn: 867 initl: 840 opt: 840 Z-score: 1105.4 bits: 210.8 E(2463): 2.3e-56
 Smith-Waterman score: 840; 78.8% identity (96.2% similar) in 160 aa overlap (1-160:1-160)

```

      10      20      30      40      50      60
Mor-PR MGVFTYETESTSIIPPAKLFKAFLLDADNLIIPKVASQAIAKTSEIVEGGPGTIKKITFG
      .....
ACJ238 MGVFTYSEITTVITPARLFKAFVLDADNLIIPKVAPQAIAKSEIIEGSGGPGTIKKITFG
      10      20      30      40      50      60

      70      80      90     100     110     120
Mor-PR EGSQFKYVKQRIDVVDHANFTYGYSVIEGDALSQTLEKISYEMKLVASPDGGCILKSTSK
      .....
ACJ238 EGSQFNYYKHRIDEIDNANFTYACTLIEGDAISLTLEKIAYEIKLVASPDGGGILKSTSK
      70      80      90     100     110     120
  
```

d



注: d 图中 IS 代表输入的询问序列, R 代表匹配到的序列, 绿色代表完全匹配的氨基酸, 黄色代表相似的氨基酸, 红色代表错配的氨基酸序列。

图 4 COMPARE 数据库在线序列比对(a)及其结果(b~d)

Fig.4 Online sequence alignment (a) and the results interfaces (b~d) in COMPARE database

(3)数据库下载

此外, 用户还可以根据需求, 从数据库主页的文档选项卡中下载过敏原相关文章、PDF 或 FASTA 格式的过敏

原序列文件、COMPARE 数据库中所有收录的过敏原分子相关信息的 EXCEL 文件以及数据库年度更新中所收录的公开文档。

5 结束语

生物信息学和蛋白组学技术的发展为过敏原的鉴定及食物潜在致敏性的评价提供了更加便捷、快速的渠道,而过敏原数据库的建立更是加快了这一进程。COMPARE 作为近年来新建立的过敏原综合数据库,由于其公开、透明的监督运转体系,严格的过敏原审核标准以及持续的更新完善途径已经被越来越多相关研究及产业人员以及食品安全监管机构认可,并在过敏原基础临床研究以及含有转基因动植物成分的食品安全性评价中得到广泛应用^[18-20]。此外,应用 COMPARE 数据库完成的有关蛋白质过敏性的生物信息学评价报告,已经为各国农业、食品、药品监管部门对有关上市产品的安全性评估提供了参考和依据。但该数据库暂时还未开放主动提交过敏原的功能,所有收录序列均由数据库工作人员来自公共数据库或已发表的相关文献中检索得到,因此可能造成部分过敏原数据难以及时收录或者滞后,此外该数据库还缺少过敏原结构和表位等相关信息,用户无法在同一数据库中获得过敏原的完整信息,而随着不同类型的过敏原数据库的增多,建立有统一标准的综合过敏原数据库,加强不同类型的数据库之间的连接将大大提高数据库在过敏原分子研究中的应用效率。

本文详细介绍了 COMPARE 数据库的建立更新过程,数据收录标准以及使用方式,以期提高该数据库在我国的应用价值,进而促进过敏原分子鉴定研究。

参考文献

- [1] TANG MLK, MULLINS RJ. Food allergy: Is prevalence increasing? [J]. *Int Med J*, 2017, 47(3): 256–261.
- [2] 席令仪, 李国梁, 张现龙, 等. 食品过敏原生物传感检测技术研究进展 [J]. *食品安全质量检测学报*, 2022, 13(6): 1782–1790.
- XI LY, LI GL, ZHANG XL, *et al.* Research progress on biosensing technology for detection of food allergens [J]. *J Food Saf Qual*, 2022, 13(6): 1782–1790.
- [3] 于闯, 雍凌, 李振兴, 等. 从过敏原危害评估食物过敏风险 [J]. *中国食品卫生杂志*, 2021, 33(3): 383–391.
- YU C, YONG L, LI ZX, *et al.* Assess food allergy risk from allergen hazards [J]. *Chin J Food Hyg*, 2021, 33(3): 383–391.
- [4] ANVARI S, MILLER J, YEH CY, *et al.* IgE-mediated food allergy [J]. *Clin Rev Allergy Immunol*, 2019, 57(2): 244–260.
- [5] RADAUER C, BUBLIN M, WAGNER S, *et al.* Allergens are distributed into few protein families and possess a restricted number of biochemical functions [J]. *J Allergy Clin Immunol*, 2008, 121(4): 847–852.
- [6] FABER MA, GASSE AL, DECUYPER II, *et al.* Cross-reactive aeroallergens: Which need to cross our mind in food allergy diagnosis? [J]. *J Allergy Clin Immunol Pract*, 2018, 6(6): 1813–1824.
- [7] ABDELMOTELEB M, ZHANG C, FUREY B, *et al.* Evaluating potential risks of food allergy of novel food sources based on comparison of proteins predicted from genomes and compared to www. AllergenOnline.

- org [J]. *Food Chem Toxicol*, 2021. DOI: 10.1016/j.fct.2020.111888
- [8] MARI A, RASI C, PALAZZO P, *et al.* Allergen databases: Current status and perspectives [J]. *Curr Allergy Asthma Rep*, 2009, 9(5): 376–383.
- [9] RADAUER C, BREITENEDER H. Allergen databases-A critical evaluation [J]. *Allergy*, 2019, 74(11): 2057–2060.
- [10] GOODMAN RE, BREITENEDER H. The WHO/IUIS allergen nomenclature [J]. *Allergy*, 2019, 74(3): 429–431.
- [11] GOODMAN RE, EBISAWA M, FERREIRA F, *et al.* Allergen Online: A peer-reviewed, curated allergen database to assess novel food proteins for potential cross-reactivity [J]. *Mol Nut Food Res*, 2016, 60(5): 1183–1198.
- [12] IVANCIUC O, SCHEIN CH, GARCIA T, *et al.* Structural analysis of linear and conformational epitopes of allergens [J]. *Regul Toxicol Pharm*, 2009, 54(3): S11–S19.
- [13] MARI A, SCALA E, PALAZZO P, *et al.* Bioinformatics applied to allergy: Allergen databases, from collecting sequence information to data integration. The Allergome platform as a model [J]. *Cell Immunol*, 2006, 244(2): 97–100.
- [14] VITA R, OVERTON JA, GREENBAUM JA, *et al.* The immune epitope database (IEDB) 3.0 [J]. *Nucleic Acids Res*, 2015, 43(D1): D405–D412.
- [15] HUANG Y, TAO AL. Biomedical and life sciences: Allergen database [M]. Dordrecht: Springer, 2015.
- [16] REE R, SAPITER BD, BERIN MC, *et al.* The COMPARE database: A public resource for allergen identification, adapted for continuous improvement [J]. *Front Allergy*, 2021, 2: 700533.
- [17] Health and Environmental Sciences Institute. COMPARE 2022 Update Documentation. [EB/OL]. [2022-01-26]. <http://db.comparedatabase.org/docs/COMPARE-2022-Documentation.pdf> [2022-10-02].
- [18] HERMAN R, SONG P. Comprehensive COMPARE database reduces allergenic risk of novel food proteins [J]. *GM Grops Food*, 2022, 13(1): 112–118.
- [19] CARLSON AB, MATHESIUS CA, BALLOU S, *et al.* Safety assessment of the insecticidal protein IPD079Ea from the fern, *Ophioglossum pendulum* [J]. *Food Chem Toxicol*, 2022, 166: 113187.
- [20] WANG C, BEAN GJ, CHEN CJ, *et al.* Safety assessment of Mpp75Aa1.1, a new ETX_MTX2 protein from *Brevibacillus laterosporus* that controls western corn rootworm [J]. *PLoS One*, 2022, 17(9): e0274204.

(责任编辑: 张晓寒 黄周梅)

作者简介

赵 岚, 博士, 主要研究方向为花粉和水果过敏原的鉴定和分子诊断。
E-mail: zhaoalan663@zju.edu.cn

高中山, 博士, 教授, 主要研究方向为过敏原分子基础和诊断应用。
E-mail: gaozhongshan@zju.edu.cn