

烟草种质资源图像数据库的维护与更新

张兴伟, 戴培刚, 佟英, 刘国祥 (中国农业科学院烟草研究所, 农业部烟草生物学与加工重点实验室, 山东青岛 266101)

摘要 阐述了建设烟草种质资源图像数据库对于烟草种质资源的保护、研究、共享利用的重要性, 并制定了拍摄、处理、存储的一套维护更新的规范标准, 形成了图像数据库维护更新作业流程, 并给出了图像维护更新过程中一些问题的处理方法。

关键词 烟草; 种质资源; 图像; 数据库; 维护; 更新

中图分类号 S-058 **文献标识码** A **文章编号** 0517-6611(2017)27-0235-04

Maintenance and Updating of Image Database for Tobacco Germplasm Resources

ZHANG Xing-wei, DAI Pei-gang, TONG Ying et al (Key Laboratory of Tobacco Biology and Processing of Ministry of Agriculture, Tobacco Research Institute, Chinese Academy of Agricultural Sciences, Qingdao, Shandong 266101)

Abstract The importance of constructing image database for tobacco germplasm resources for the protection, study and sharing of tobacco germplasm resources was expounded. A set of maintenance and updating standards for shooting, handling and storing tobacco images were developed. An operation flow on maintaining and updating image database for tobacco germplasm resources was formed. Some methods were given to solve problems occurred in updating image database for tobacco germplasm resources.

Key words Tobacco; Germplasm resources; Image; Database; Maintenance; Updating

随着人们对信息直观性要求的提高, 数据库中单纯的文字信息已不能满足人们的多层次需要。据统计, 人类接受的外部信息, 70%以上来自于视觉。图像作为一种内容丰富、表现直观的多媒体信息被大量广泛使用^[1-8], 如何有效检索与管理图像信息已成为迫切需要解决的问题, 以便让用户更加直观形象地了解所关心的对象。如果图像数据库与其他信息系统有机结合, 不仅可以加快相关科学基础知识的普及, 而且可以有效提高资源的利用效率。种质资源是人类选育新品种和发展农业生产的物质基础, 也是进行生物学研究的重要材料, 是极其宝贵的自然财富^[9]。烟草传入我国已经有 400 多年的历史, 同时由于生态条件、栽培措施和人文环境的复杂性和多样性, 形成了我国丰富多样的烟草种质资源^[10-11]。建国以来, 我国高度重视种质资源研究, 早在 20 世纪 50 年代就开展了全国性的烟草种质资源收集保存工作, 1979—1983 年又进行了全国范围的烟草种质资源补充征集工作。此后, 种质资源的收集工作逐步转向以重点地区考察收集为主, 同时日常性的收集与保存工作持续进行^[12]。从 2015 年开始, 开展了“第三次全国农作物种质资源考察与收集行动”, 收集了不少优异的地方烟草种质资源。截至 2016 年底, 国家烟草种质资源中期库中已编目的种质数量多达 5 607 份。目前, 我国已成为世界上烟草种质资源保存数量最多的国家。资源类型包括晾晒烟、烤烟、黄花烟、白肋烟、香料烟、雪茄烟及野生烟等七大类型^[13]。

随着社会、经济的快速发展, 资源整合与共享成为时代发展的必然选择。为此, 国家烟草专卖局启动了“中国烟草种质资源平台建设”专项^[14], 其核心内容就是要建立有效的资源整合和共享机制, 实现烟草种质资源实物与信息

的全面共享, 为行业可持续发展奠定牢固的基础。上述发展和变化对传统的烟草种质实物共享体系与手段提出了新的要求与挑战。因此, 建立更加便捷、高效的种质资源实物共享和管理系统十分必要。前期中国农业科学院烟草研究所已经构建了烟草种质资源信息网^[15-16], 并开发了在线共享系统^[17]和地理信息系统^[18], 近年来得到上级领导和科研人员的高度肯定。烟草种质资源图像数据库属于中国烟草种质资源信息网的重要模块之一, 已实现与烟草种质资源具体信息的链接和共享, 自 2008 年 9 月建成以来, 迄今运行已经 9 年整。为了更好地发挥其功用, 笔者对图像数据库进行了多次维护更新, 制定了一套关于拍摄、处理、存储图像的规范标准, 形成了图像数据库维护更新作业流程, 并给出了图像维护更新过程中一些问题的处理方法, 为快速、便捷地检索图像提供服务, 以期更好地为用户提供种质资源的直观信息, 促进烟草种质资源整合与共享的深入开展。

1 烟草种质资源图像数据库的技术架构

烟草种质资源图像数据库的网络拓扑结构及系统运行的软硬件环境参照文献[16]。

2 烟草种质资源图像数据库建设的规范标准

2.1 烟草种质资源图像数据的拍摄

2.1.1 拍摄时期。一般应选择盛花期, 特殊种质可选择种质特征充分表现的最佳时期。

2.1.2 拍摄时间。尽可能选择多云日上午 09:00—10:00 拍摄。

2.1.3 拍摄份数。一般至少拍摄 2 份, 这样可供图像存储时择优选择。

2.1.4 相机要求。单镜头反光数码相机, 且具有近摄和变焦功能。

2.1.5 像素设定。花序、花冠、蒴果应设定在 500 万像素以上, 株型、叶形要求设定 800 万像素以上。

2.1.6 图像格式。必须是 JPG 格式。文件名统一要求如

基金项目 国家烟草专卖局专项(国烟办综[2005]501号); 中国烟草总公司海南省公司项目(20164600020007); 中国农业科学院科技创新工程项目(ASTIP-TRIC01)。

作者简介 张兴伟(1978—), 男, 吉林德惠人, 副研究员, 博士, 从事烟草种质资源研究。

收稿日期 2017-07-21

下:①繁种试验。文件名采用8位统一编号+“-对象序号”。示例如下:株型00002290-1,叶形00002290-2,花序00002290-3,花冠00002290-4,蒴果00002290-5。②编目试验。文件名采用8位临时序号+“-对象序号”。示例如下:株型AHB00005-1,叶形AHB00005-2,花序AHB00005-3,花冠AHB00005-4,蒴果AHB00005-5。未指定序号的可用品种名称代替。

2.1.7 背景选择。一般应选择浅蓝色背景,野生种或特殊种质也可选择黑色背景,背景质地要求无皱、无反光布质或其他相当材质。

2.1.8 图像质量。要求无阴影、无外露(背景应超出拍摄实体30 cm以上)、无凋萎。图像中不得出现任何标签,最好放置标尺。

2.1.9 株型照相。一般要求整株挖出于试验地外空旷区域,采用长焦平视、于烟株中部略偏上方侧顺光拍摄,防止株型变形。若遇晴朗天气气温高,为避免烟株凋萎,也可直接在原地拍摄,但应在不影响试验数据采集的前提下,拔去周边烟株、杂草、已经萎蔫的烟花及已经结成的蒴果,以免影响视角。

2.1.10 叶形照相。取3片完整叶,自左向右“下、中、上”居中排列,叶片间距1~2 cm。要求叶片无病斑、无破损、无明显杂色,叶耳完整(可用医用手术刀片从烟株上轻轻剥离)。采用标准焦距拍摄。

2.1.11 花序照相。结合株型照相,拍摄完整花序,采用标准焦距拍摄。

2.1.12 花冠照相。取代表性完整开放花(含花柄)9朵,上4、下5(花冠向上),两排交错居中排列,行距、花距均为1 cm左右。采用近摄焦距拍摄。

2.1.13 蒴果照相。取代表性完整成熟蒴果9枚,上4、下5(果尖向上)两排交错居中排列,行距、果距均为1 cm左右。采用近摄焦距拍摄。

2.2 烟草种质资源图像数据的存储 照片拍摄后应及时从相机取出,按照上述命名原则命名并及时存储在图像数据库中。若图像数据库中已有此种质图像,则在相应名称后+“a”,示例如下:株型00002290-1a,叶形00002290-2a,花序00002290-3a,花冠00002290-4a,蒴果00002290-5a。若图像数据库中仍有此种质图像,则在相应名称后+“b”,以此类推。此外,为了安全起见,要做好烟草种质资源图像数据的备份存储。

2.3 烟草种质资源图像数据的处理

2.3.1 烟草种质资源图像的预处理。将已经命名好的烟草种质资源图像存放于硬盘的原始图像相应目录下。对原始图像进行预处理加工,去掉照片中杂草、其他烟株等,对照片进行美化处理,包括调整饱和度、亮度及对比度等。同时,为加快网络浏览速度,将上述文件夹复制1份,使用ACDSee 9软件批处理调整图像方向以及统一大小,使图像最小边像素为440。将调整后图像存放于服务器硬盘的相应图像目录下,注意是覆盖原来图像。

2.3.2 烟草种质资源图像后处理。目前,只在中国烟草种质资源平台单位中提供相关服务。设置多级别的用户管理,不同的用户分配不同的权限。需要下载大分辨率图片的用户,须先填写下载申请,待走完手续后,由管理员给予用户图像。后台用户日志统计会详细记录用户申请理由、下载图像的时间及管理者的备注,及时追踪图像的去向。

3 烟草种质资源图像数据库更新的作业流程

3.1 功能实现 数据库构建、角色和权限设置参照文献[16]。

3.2 作业流程 烟草种质资源图像数据库更新维护的作业流程如图1所示。首先,对种质资源拍照人员依据拍照标准进行技能培训,以期获得高质量典型的烟草种质资源图像。再对该图像进行核对,要与原数据库中该种质的图像进行对比,按照上述命名标准重新命名。对重命名后的图像进行美图处理,以期能够更好地满足人们的审美要求。然后,一方面要对该图像进行安全存储,另一方面要从该种质众多图像中择优选出最能代表该种质的典型图像,进而更新到图像数据库中,达到充分共享及展示该种质资源的目的。

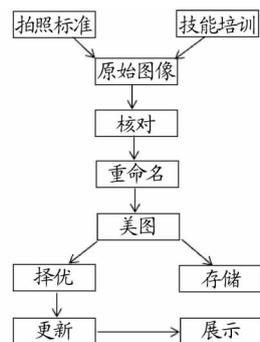


图1 烟草种质资源图像数据库更新维护的作业流程

Fig. 1 The operation flow of updating and maintaining image database for tobacco germplasm resources

3.3 应用简介 用户首先在浏览器中输入网址 <http://www.yesjk.com.cn>,打开“中国烟草品种资源网”的首页,在“烟草种质资源”栏中点击“种质图像查询”,或者在“烟草种质资源”栏中点击“共性数据查询”,按提示操作,前2种方式适用于“开放用户”;或者在“烟草种质资源”栏中点击“特性数据查询”,按提示操作,这种方式适用于“登录用户”。“种质图像查询”只接受种质名称和全国统一编号。如果用户还没有明确具体所需的种质,可以通过“中国烟草种质资源信息系统”的查询功能查找符合要求的种质资源。烟草种质资源的全国统一编号为一串8位数字,不足8位时用“0”补齐。在该系统中可只输入非零部分,系统会自动补齐,例如“00001273”可只输入“1273”。输完一个种质的编号后,请打回车键输入下一个,每行内只允许有一个种质的全国统一编号。种质名称接受模糊查询^[16],可允许一定程度的输入错误,这大大提高了检索效率。信息填写完毕提交后,即可浏览感兴趣的烟草种质资源的图像信息。图2为烟草种质资源图像数据库的操作流程,图3为一份烟草种质资源(林烟草,普通烟草的二倍体祖先种之一)图像信息的具体查询结果。

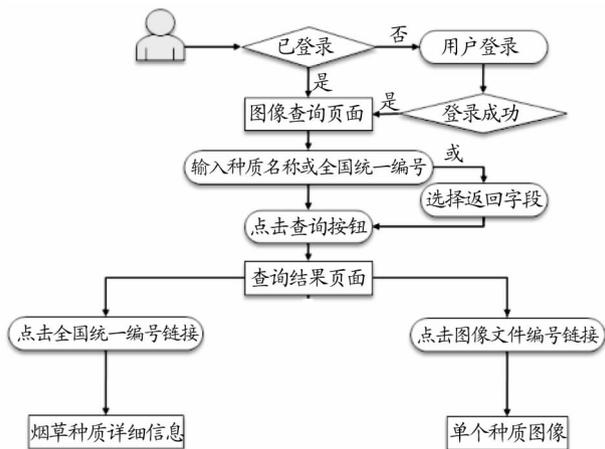


图2 烟草种质资源图像数据库的操作流程

Fig. 2 The workflow of image database for tobacco germplasm resources

4 常见问题的处理

4.1 错误数据处理 在烟草种质资源图像数据库的建设过程中,由于各种原因会造成少量错误,如个别烟草种质资源平台建设单位提供的原始照片有误、不育系种质有蒴果照片、普通烟草种质却拥有黄花烟草的花序照片等。建库初期由于审核机制不健全和图像数据库操作人员专业知识不足等原因,导致这些照片被录入图像数据库中。在图像数据库的使用过程中,随着自查机制的推出及图像数据库使用人员的问题提出,及时核对原始数据库,仔细甄别相关信息,将错误数据及时更正。

4.2 非典型数据处理 在烟草种质资源图像数据库的建设过程中,有些种质拥有不止一套照片(株型、叶形、花序、花冠和蒴果)。建库初期,为了达到最快的展示效果,通常是选择一套整体质量突出的照片录入到图像数据库中。但是,不能保证该种质的5张照片都是典型照片。在此后的图像数据



图3 种质图像查询具体结果

Fig. 3 The querying result of germplasm images

库更新维护中,应仔细核对拥有多套照片的种质,择优选择,动态更新,做到每张照片都是典型照片。

5 讨论

图像数据库建设的最终目的是为更多的用户提供更大

的便利,以便加深对种质资源的直观认识。如果没有用户需求,建库也就失去意义。只有从用户需求出发,图像数据库的建设才能实现其价值,并不断提高资源利用率。烟草种质资源图像本身就是各种符号的集合体,而对于符号的掌握程

度与科研人员知识储备密切相关。研究者首先被图像吸引,再被文字丰富,用文字来理解图像,用图像来印证文字,在图像与文字的积极互动中,研究者不知不觉加深了对烟草种质资源的认知程度。

5.1 图像数据库的标准化建设 标准化是建设高质量数据库的重要支撑和坚强保障,对于保持数据库的系统性和兼容性,实现网络化和资源共享均具有极其重要的意义。因此,建设图像数据库时要严格遵循标准化原则,在图像数据采集、加工、存储及检索的过程中,一定要采用通用的标准。该标准是在国家烟草专卖局“中国烟草种质资源平台建设专项”的实施过程中逐渐形成的,并经过了实践的检验。在烟草种质资源图像数据库的维护更新中,应严格遵循上述标准。

5.2 图像数据库的共建共享 图像数据库建设是一项基础性工作,更是一项长期的系统工程,需要投入大量的人力、物力和财力,需要国家烟草专卖局给予稳定的财政支持,需要“中国烟草种质资源平台建设”的平台单位合力共建。目前,保存的5607份烟草种质资源中仍有一些种质缺乏照片,有不少种质照片不全,不少种质图片质量还有待提高。烟草种质资源的典型图片拍摄时间很短,错过之后就只能等待下一个生长周期。任何一家单位独立完成全部烟草种质资源图片的拍摄基本是不可能的。综上所述,这些客观因素要求高层要给予稳定的财政支持,一定要多家平台单位通力合作,并且要长期坚持。每个平台单位获得的资源照片要及时传到图像数据总库中,只有在此基础上才能真正实现共享。

5.3 图像数据库的维护更新 为避免浪费,在确定入库图像时,应特别注意该图像是否有可靠的来源,避免不符合标准甚至错误的图像入库。图像数据录入的完成并不意味着数据库建设的完成,还要经常进行维护和更新。资源网应安排专门人员收集图像数据库使用过程中的反馈信息,定期对图像数据进行清理、修正和补充更新。

5.4 烟草突变体图像的整合共享 由中国农业科学院烟草研究所牵头主持的“烟草突变体创制、筛选与鉴定”项目,历时5年,经过多家合作单位共同努力,采用EMS诱变和T-DNA激活标签插入等方法,创制了27万余份烟草突变体材料,经鉴定获得了700余份性状稳定遗传的烟草突变体^[19]。

烟草突变体是一类特殊的烟草种质资源,在进行基因功能研究和品种定向改良方面具有独特的利用价值。为进一步加强这些突变体资源的利用,有必要将突变体相关信息整合到烟草种质资源数据库(www.ycsjk.com.cn)中,以利用其成熟的分发利用体系,不断提高资源利用率。烟草种质资源研究团队与烟草突变体研究团队应加强合作。烟草突变体应主动采用烟草种质资源图像数据库的建设标准,以便于整合共享;烟草种质资源应主动吸纳稳定的具有特殊利用价值的突变体材料到烟草种质资源数据库中,以便更好地服务于生产实践和广大科研工作者。

参考文献

- [1] 张帆,方如明,蔡健荣.标准烟叶数据库的图像检索[J].农业机械学报,2001,32(1):66-68,71.
- [2] 吴健,俞天秀,张若识.敦煌艺术图像数据库的建设[J].敦煌研究,2008(6):68-71,116.
- [3] 王星,陈楠,李坤成,等.基于SQL server 2000数据库管理系统的正常人脑MRI图像数据库[J].中国医疗设备,2008,23(10):25-27.
- [4] 向红,刘嘉敏,谢海军,等.中国人人耳图像数据库的设计与建立[J].计算机工程,2009,35(18):275-277.
- [5] 李宁,程霞,杨伟,等.黄斑疾病OCT图像数据库的设立与建立[J].医疗卫生装备,2013,34(9):39-40,48.
- [6] 宁静,杜中赫,刘雅慧,等.植物种子信息图像数据库的建立[J].种子,2016,35(4):126-129.
- [7] 沈霞,王昌利,胡本祥,等.构建秦岭中草药标本馆图像数据库的探索[J].陕西中医学院学报,2015,38(3):98-100.
- [8] 许磊,黎智辉,王俊娟,等.高精度三维人脸图像数据库[J].刑事技术,2015,40(2):94-97.
- [9] 董玉琛.我国作物种质资源研究的现状与展望[J].中国农业科技导报,1999(2):36-40.
- [10] 于梅芳.我国烟草品种资源的研究[J].中国种业,1986(1):11-14.
- [11] 蒋子恩.我国烟草资源概况[J].中国烟草科学,1988(1):42-46.
- [12] 李毅军,牛佩兰,蒋子恩,等.我国烟草品种资源概况及研究战略[J].中国烟草,1995(1):11-14.
- [13] 张兴伟,王志德,牟建民,等.我国烟草种质资源现状与展望[J].中国烟草科学,2009,30(6):78-83.
- [14] 张兴伟,冯全福,杨爱国,等.中国烟草种质资源分发利用情况分析[J].植物遗传资源学报,2016,17(3):507-516.
- [15] 张兴伟,王志德,张久权,等.中国烟草种质资源信息网的开发与应用[J].中国烟草科学,2009,30(S1):32-36.
- [16] 任民,张兴伟,张久权,等.中国烟草种质资源网络信息系统的开发[J].农业工程学报,2010,26(3):209-215.
- [17] 任民,张兴伟,张久权,等.烟草种质资源在线共享系统的开发应用[J].中国烟草科学,2011,32(4):51-55.
- [18] 任民,张兴伟,王志德,等.基于Google Maps API的烟草种质资源WebGIS开发研究[J].植物遗传资源学报,2010,11(5):522-526.
- [19] 刘贯山,孙玉合.烟草突变体[M].上海:上海科学技术出版社,2016.
- [20] (上接第222页)
- [21] 董微,周增产,刘文玺,等.光伏温室室内外环境条件对比[J].农业工程,2015,5(5):44-48.
- [22] 董微,周增产,卓杰强,等.光伏低碳温室设计与应用[J].农业工程,2013,3(4):54-57,51.
- [23] 张良.夏季屋顶全开窗型温室的通风率与温度预测研究[D].镇江:江苏大学,2016.
- [24] AHMED H A, AL-FARAJ A A, ABDEL-GHANY A M. Shading greenhouses to improve the microclimate, energy and water saving in hot regions: A review[J]. Scientia horticulturae, 2016, 201: 36-45.
- [25] SONI P, SALOKHE V M, TANTAU H J. Effect of screen mesh size on vertical temperature distribution in naturally ventilated tropical greenhouses[J]. Biosystems engineering, 2005, 92(4): 469-482.