

基于 Maxent 与 GIS 的我国建兰地理分布预测及关键生物气候因子分析

梁 春, 罗 清, 陆 祖 正, 谢 振 兴, 覃 茜, 黄 欣 怡

(广西壮族自治区亚热带作物研究所, 广西 南宁 530001)

摘 要:该研究基于生态位模型理论对我国建兰地理分布及气候影响因素进行研究。以建兰在国内的现有地理分布数据为基础,通过运用地理信息系统(GIS),建立了基于最大熵模型(Maxent)理论的建兰国内适生区预测模型,并对影响建兰地理分布的关键生物气候影响因子做出分析。结果表明:经 ROC 验证,模型的训练 AUC 为 0.977,测试 AUC 为 0.944,数据检验显示所建立模型的预测精度良好;刀切法检验结果表明,最冷季均温、最早季均温、年均气温、最早月降水量为影响建兰潜在分布的最关键生物气候因子;中国建兰潜在分布区主要位于广东北部、广西东北、江西及湖南南部的南岭山脉地区,福建北部武夷山南段地区以及广东莲花山以南等地。通过对模型数据与预测结果进行分析,为未来国内建兰种质资源保护、引种驯化与规模化室外种植研究提供了科学建议与理论基础。

关键词:建兰;最大熵模型;地理信息系统;地理分布预测;生物气候因子

中图分类号:S 682.31 **文献标识码:**A **文章编号:**1001-0009(2017)09-0199-06

建兰(*Cymbidium ensifolium*)属兰科(Orchidaceae)兰属(*Cymbidium*)植物,别名四季兰,是主要栽培观赏品种之一,具有悠久的栽培历史,其野生与人工栽培品种资源极为丰富^[1]。近年来有大量学者开展了对包括建兰在内的我国兰花遗传多样性分析、种质资源调查等研究工作。唐源江等^[2]基于 SRAP 标记对包括建兰在内的国兰栽培及野生种质资源开展遗传多样性分析,对国兰种质资源在分子水平上进行鉴定与分类;黄暖爱^[3]利用地理信息系统对我国重点兰科植物地理分布格局进行分析,并基于研究数据对保护现状给予评价。但由于多数兰科植物的地理分布影响因素尚不确定,相当一部分稀有野生兰花资源仍缺失野外相关研究。与此同时,国内兰花产业也针对建兰开展了试验性的引种

驯化及规模化种植,但由于建兰对种植地区气候等因素的要求较高,盲目引种往往造成极大的损失。

生态位建模(Ecological niche modelling)又被称为栖息地分布预测建模(Predictive habitat distribution modelling),即通过对目标物种在现实环境空间的分布(即实际生态位)进行分析,运用特定算法建立数学模型来预测其在目标地理空间分布的过程^[4]。生态位模型作为生物多样性信息学的组成部分,近年来在濒危物种保护、入侵物种潜在分布评估以及气候对物种分布影响预测等研究领域发挥着重要作用。其中,最大熵模型法得到广泛关注。最大熵模型基于最大熵理论,由 JAYNES^[5]于 1957 年提出,即对于精确信息集或特定概率分布函数的可测值,在评估该信息系列的所有概率分布情况下,选择具有最大信息熵的分布,是一种对该随机量的有效处理方法与准则。ELITH 等^[6]最早将最大熵模型引入物种栖息地建模,即根据目标物种现实分布点和现实分布地区的环境数据,通过学习样本与环境变量间的关系构建预测模型,对物种在目标地区的分布概率进行测算,寻找具有最大熵的分布预测。对模型的评估通常通过分析接受者操作特性(Receiver operating characteristic,ROC)曲线进行,即

第一作者简介:梁春(1987-),男,本科,研究实习员,现主要从事亚热带植物等研究工作。E-mail:8772080@qq.com.

责任作者:罗清(1986-),女,硕士,助理研究员,现主要从事植物栽培育种等研究工作。E-mail:lqlq522@163.com.

基金项目:广西科技计划资助项目(桂科 AB16380060);广西壮族自治区亚热带作物研究所基本科研业务费专项资金资助项目(桂热研 201610)。

收稿日期:2017-02-09

以命中概率为纵轴,虚报概率为横轴,通过计算训练集的 ROC 曲线下面积(Area under the ROC curve, AUC)并与随机分布的 AUC 值比较来评估预测地理分布与环境变量的相关性,从而判断模型预测效果。张路^[7]分析了最大熵模型在预测物种潜在分布范围方面的应用,指出该模型在解决深层次的理论问题时具有重要的应用价值;崔晋亮等^[8]通过最大熵模型模拟蓝莓当前在全球范围的潜在地理分布,并基于气候因子模拟了蓝莓在末次冰盛期和全新世中期以及 21 世纪 50 年代的潜在地理分布,取得了良好的效果。

地理信息系统(Geographic information system, GIS),是一种旨在获取、存储、处理、分析、管理和展现各类空间或地理数据的综合性系统,最早由 TOMLINSON^[9]于 1968 年提出。近年来随着计算机技术的发展,GIS 不仅广泛应用于国民经济建设的所有领域,而且也在地质、大气、环保等多个科研领域发挥着重要作用。在生物地理学方面,地理信息系统正以与计算机科学结合的方式逐步取代传统的人工数据整理与图表绘制,推动着物种栖息地分析、自然保护区规划、外来物种分布等学术领域的迅猛发展。

该研究通过对建兰在国内的地理分布以及相关气候因素进行分析,对建兰在国内的潜在地理分布区进行预测,同时获取影响建兰野生分布的关键气候影响因子,从而确定建兰在国内的地理分布,以期为在国内开展野生品种资源收集调查以及进行建兰引种提供参考依据。

1 材料与方法

1.1 数据处理工具

该研究使用 Maxent 3.3.3k 软件进行最大熵建模。Maxent 基于最大熵模型理论,通过对一系列数据层或环境变量(海拔、降水量等)以及地理分布数据做出分析,进行物种栖息地建模。Maxent 是目前最大熵模型法在物种分布建模中应用最广的软件。

对于 GIS 数据,使用 DIVA-GIS 7.5 软件及 Google Earth 7.1.5 进行分析处理。DIVA-GIS 是一款用于地理绘图及 GIS 数据分析的软件,针对当前 GIS 数据系统大多较为冗杂且软件许可费用昂贵的问题,DIVA-GIS 旨在为科研人员提供一款免费、简便且多用途的 GIS 软件,其开发得到了国际农业研究磋商组织(CGIAR)下属多个机构及其他多家科研院所的支持。Google Earth 由谷歌公司开发,是一款基于卫星及航拍数据、GIS、GPS 以及地理三维模型

的软件,是目前全球范围内使用量最大的虚拟地球仪系统。

1.2 建兰地理分布数据的采集

该研究通过检索标本确定建兰已知分布点。通过中国数字植物标本馆(Chinese virtual herbarium, CVH)取得中国科学院华南植物园标本馆(IBSC)标本信息 77 份,中科院植物研究所标本馆(PE)68 份,中科院昆明植物研究所标本(KUN)30 份,中山大学生物系植物标本室(SYS)20 份,广西植物研究所标本馆(IBK)22 份,其它馆藏 30 份^[10];通过全球生物多样性信息网络(Global biodiversity information facility, GBIF)获得建兰在国内的分布地数据 38 份^[11],共计 285 份地理分布信息。通过 Google Earth 获得各地理分布点经纬度数据。

1.3 气候数据的选择与获取

气候数据选择 World Clim 提供的 19 项生物气候变量(Bioclimatic variables)数据记录作为环境变量,空间分辨率 2.5 arc minutes^[12]。生物气候变量以历年各月降水量与气温数据为基础,将相关信息转化为更具有生物学意义的统计变量,各变量信息见表 1。

生物气候变量通常用于物种分布及其它相关生态学建模。这些变量有效反映了各类生物统计学信息,如气候年度趋势(年均气温、年度降水量等)、季节性(年度气温及降水量范围等)、极限或限制性水热环境因子(最冷与最热月份气温、最早与最湿季度降水等)。

表 1 生物气候变量

Table 1 Biological climatic variables			
变量名 Variable	内容 Description	变量名 Variable	内容 Description
BIO1	年均气温	BIO11	最冷季均温
BIO2	月均昼夜温差	BIO12	年均降水量
BIO3	等温性	BIO13	最湿月降水量
BIO4	温度季节性变化方差	BIO14	最早月降水量
BIO5	最暖月最高温	BIO15	降水量方差
BIO6	最冷月最低温	BIO16	最湿季降水量
BIO7	年度气温范围	BIO17	最早季降水量
BIO8	最湿季均温	BIO18	最暖季降水量
BIO9	最早季均温	BIO19	最冷季降水量
BIO10	最暖季均温		

1.4 GIS 地理地图的获取

该研究使用的我国国界及行政区划 GIS 底图通过国家基础地理信息中心取得,空间比例 1:400 万。

1.5 数据处理

将通过 Google Earth 定位的建兰分布点经纬度数据保存于逗号分隔值(Comma-separated values,

CSV)文件,导入 Maxent 作为数据样本。将从 World Clim 获得的波段按行交叉格式(Band interleaved by line,BIL)当前全球生物气候变量数据导入 DIVA-GIS,根据我国国境地理位置裁切成相应经纬度区块数据,导入 Maxent 作为环境层变量。

在 Maxent 中设置随机选取 75%分布点作为训练集数据(Training data),25%地理分布点作为测试集数据(Test data)以验证模型。开启刀切法(Jackknife)测试检验各变量对分布增益的贡献,从而检验各气候因子对建兰地理分布的影响,同时绘制各气候因子的响应曲线。输出类型为逻辑值(Logistic),最大迭代次数 500,分布概率为[0,1]。

2 结果与分析

AUC 值可视为 ROC 曲线与坐标轴围成的面积值,作为模型预测准确性的衡量指标可划分为:模型预测准确性较差(0.5~0.6),预测精度一般(0.6~0.7),预测精度较准确(0.7~0.8),预测精度很准确(0.8~0.9),预测精度极准确(0.9~1.0)^[13]。软件依据该研究模型运算得出的 AUC 分析结果见图 1。根据软件运算结果,模型的训练 AUC 为 0.977,测试 AUC 为 0.944,表明研究模型的预测效果非常理想,模型可用于我国建兰的适生区地理分布预测以及气候影响因子分析。

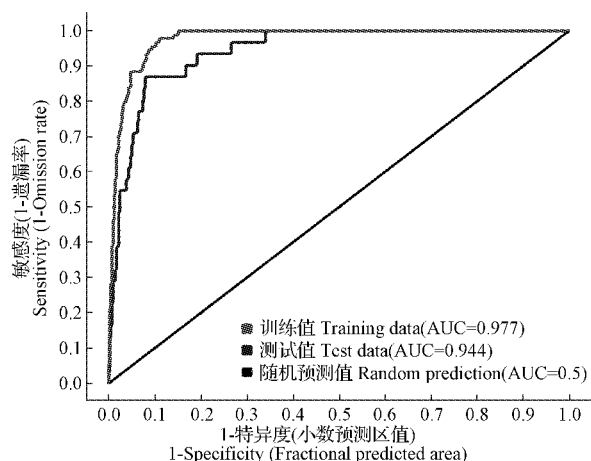


图1 模型接受者操作特性曲线 AUC 分析

Fig.1 AUC analysis of receiver operating characteristic curve for Maxent model

2.1 模型关键气候影响因子的判断

根据测试统计,对研究模型的相对贡献率最高的环境变量为年度气温范围(20.9%)、最暖季降水量(17.7%)、最早月降水量(16.6%)以及月均昼夜温差(12.8%)。4项变量的贡献率总和大于60%。然

而由于生物气候变量各项变量不同程度相关(年均、季度及月份统计数据存在交叉、降水与气温潜在相关等),因此无法据此确认4项变量即为影响最显著的气候因子,需要进一步通过刀切法作出判断。

刀切法即在模型运行过程中,先将某个变量独立置入模型进行运算,随后排除并利用剩余的变量进行运算;逐项完成后,再将全部变量置入模型进行运算^[14]。根据以上3项的测试增益(Test gain)——在 Maxent 中定义为预测模型拟合样本的平均对数概率(Average log probability),来对该变量作出分析^[6]。如果该变量的独立测试增益较高,或者其缺失导致增益下降的程度较其它变量明显,则表明该变量对模型的预测结果影响显著。

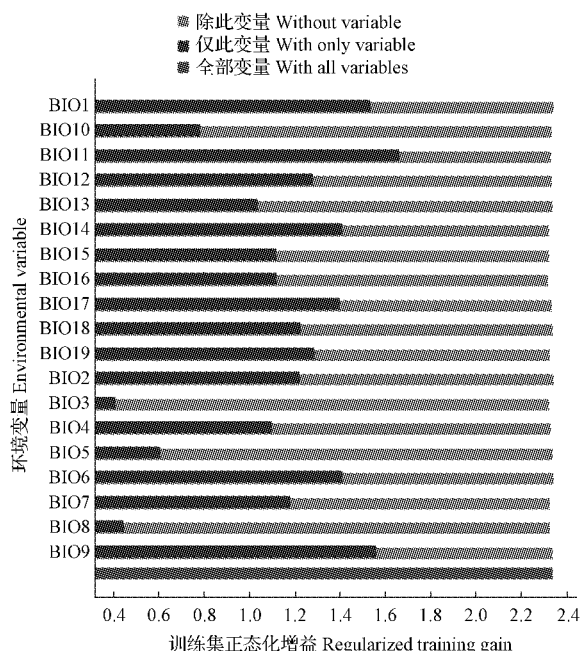


图2 模型各生物气候因子训练集正态化数据刀切法检验结果

Fig.2 Jackknife result of bioclimatic variables regularized training data

该模型对训练集数据进行刀切法检验结果见图2。其中,各变量深蓝色条目表示该变量独立测试增益,浅绿色条目表示剔除该变量的测试增益,底部红色条目表示包含全部变量的测试增益。结果表明各变量单独测试增益差异较明显,其中最冷季均温(BIO11),最早季均温(BIO9),年均气温(BIO1),最早月降水量(BIO14)最显著。由此可见,在诸项生物环境变量中,建兰受气温因素影响较大。根据 Maxent 绘制的变量响应曲线,得到各因子的响应区间见图3。以概率分布逻辑输出值0.5为界,得到各变量

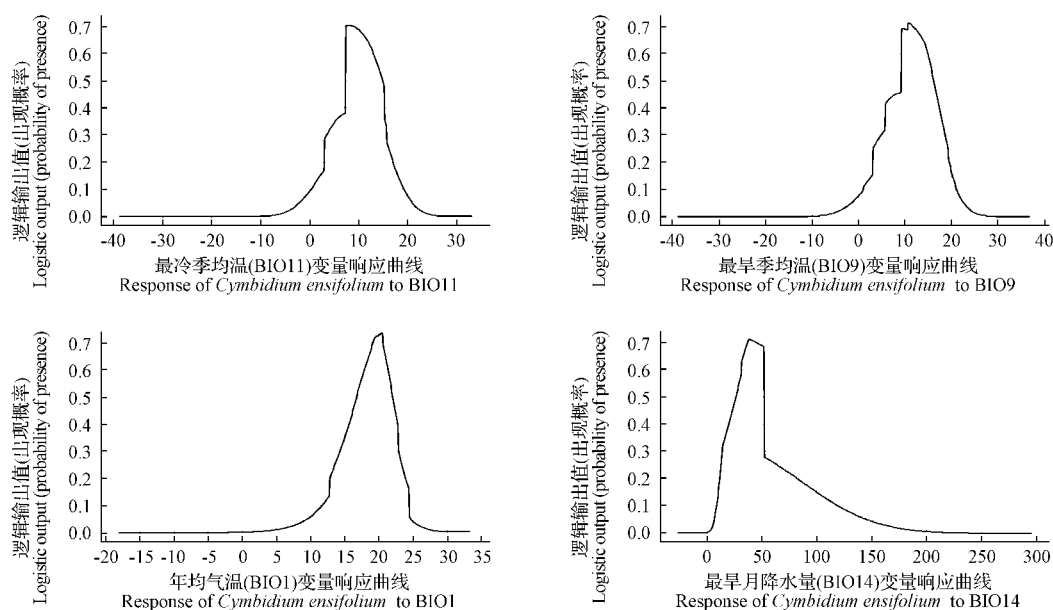


图3 模型最冷季均温(BIO11),最早季均温(BIO9),年均气温(BIO1),最早月降水量(BIO14)变量响应曲线

Fig. 3 Response curve of mean temperature of the coldest quarter (BIO11), mean temperature of the driest quarter (BIO9), annual mean temperature (BIO1), precipitation of driest month (BIO14)

的显著响应区间分别为:最冷季均温 6.7~15.0 °C, 最早季均温 9.5~16.4 °C, 年均气温 17.5~22.0 °C, 最早月降水量 26.9~57.7 mm。

2.2 基于模型的建兰适生区分布预测

将 Maxent 测试结果输出为 grd 格式栅格文件,

导入 DIVA-GIS。根据测试结果,将建兰分布概率值划分为不适宜分布区(0~0.25),较适宜分布区(0.25~0.50),适宜分布区(0.50~0.75),极适宜分布区(0.75~1.00)。同时叠加我国行政区划底图,获得建兰的适生区地理分布预测(图4)。

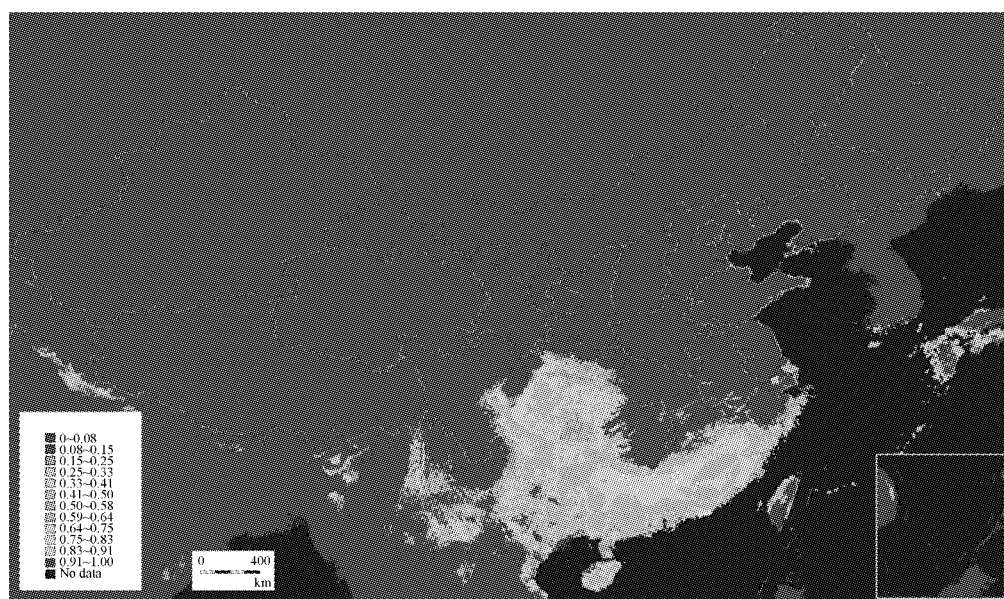


图4 建兰在我国及周边地区的适生区地理分布预测

Fig. 4 Geographical distribution prediction of *C. ensifolium* in China and surrounding area

将适生区预测分布图转存为 kmz 格式数据,导入 Google Earth,可观察到建兰在我国的适宜及极适宜分布区主要涵盖北纬 18°~28°的地区,主要分布省份为广西、广东、福建、江西、湖南、贵州及海南等。极适宜分布区位于广西桂林、广东韶关及江西赣州等地市辖区。在 DIVA-GIS 中计算各省份覆盖面,极适宜区面积较大主要有广东(22 616.2 km²)、广西(17 561.9 km²)、江西(10 920.1 km²)等,适宜区面积较大的依次是广西(122 265.6 km²)、广东(105 987.4 km²)、福建(105 699.7 km²)。建兰各适宜分布区主要覆盖省份具体状况见表 2。

表 2 建兰极适宜分布区主要覆盖省份数据

Table 2 Major provincial coverage data of highly suitable distribution area prediction for *C. ensifolium* km²

省份 Province	极适宜区 Highly suitable distribution area	适宜区 Suitable distribution area	较适宜区 Good potential distribution area
广东	22 616.2	105 987.4	15 064.2
广西	17 561.9	122 265.6	58 795.1
江西	10 920.1	32 056.4	10 882.0
福建	6 710.5	105 699.7	2 676.4
湖南	2 401.5	20 469.9	60 437.6

3 结论

该研究运用最大熵模型与地理信息系统对建兰在我国的适生区进行预测,其预测结果经接受者操作特性曲线分析法验证,预测精度极准确。

3.1 适生区关键气候影响因素与地理特征预测

经刀切法对用于预测的生物气候变量分析后获得 4 个建兰关键影响因子,其中年均气温及最冷季均温阈值提示我国西北及秦岭—淮河以北大部分地区可能不存在建兰适生区,热带高温地区也可能不适宜建兰野外生长;合并分析最早季均温及最早月降水量阈值,可观察到建兰适生区内最早季与最冷季(或冬季)可能相对重叠,该时间段区域内气温较温和,同时依然保持一定降水量。根据以上气候特征可知,建兰在我国最佳适生区应分布在华南受亚热带季风气候影响的较高海拔山区,森林覆盖率较高的热带林区也可能有林下分布。亚热带季风气候区域冬季温和少雨,年均降雨量大,属湿润区,这正符合建兰喜半阴、好湿润和耐寒性差的特性^[15]。

3.2 适生区预测分析

Maxent 模型利用有限的发生数据可预测得到适生区,目前已在动植物方面进行较深入研究。通过 Maxent 对珍稀物种的分布进行评估^[16],还对植物檀香^[17]、南丹参^[18]、川贝母^[19]及动物原麝^[20]、松

材线虫^[21]等在中国的潜在适生区进行预测,对保护稀有物种、指导实际生产具有重要意义。因此课题组选用 Maxent 对我国传统名花—建兰的适生性分布进行评估并用于指导保护珍稀濒危兰花及引种。结果显示,根据地理分布区预测可得,建兰的高适生区主要分布于广东北部、广西东北、江西及湖南南部的南岭山脉地区,福建北部武夷山南段地区以及广东莲花山以南等地,与对生物气候因子的分析结果一致;四川、云南、海南、浙江及湖北等省份也有建兰适生区分布。丁长春等^[22]、关文灵等^[23]在云南成功引种建兰;刘翠华^[24]、白坚等^[25]、胡薇等^[26]分别对江西、四川、广东及福建的建兰进行研究;以上研究都说明了建兰的已知适生区与利用最大熵模型计算预测出的适生区符合。总体而言建兰在我国的主要适宜分布区与热带亚热带地区分布较为吻合,提示这些地区的气候适合建兰生长繁殖与室外人工种植。区域内涵盖了广东(南岭、丹霞山、罗坑、车八岭)、广西(七冲、猫儿山)、江西(九连山)、福建(武夷山、黄桑)等多省份国家级自然保护区,野生建兰种质资源也可能比较丰富。

3.3 存在问题与展望

该研究旨在利用已知分布点的生物气候数据作为建兰适生区的环境特征,通过最大熵模型在全国范围内模拟具有类似环境特征的地理信息点,从而构建建兰的适生区分布,以期对未来建兰种质资源保护、引种驯化与规模化室外种植研究提供评估指导和理论基础。需要指出的是,影响建兰分布的环境因素除了生物气候因子,还包括光照、土壤、植被、物种间互动以及人为影响等外界环境因素。尤其是多年的人为环境影响造成的生境片段化所导致的林内光照增强、温度升高和土壤含水量降低等一系列边缘效应,可能导致原栖息地中已不存在野生分布。一系列因素的影响需要在今后的研究中予以进一步探讨。由于建兰适生区地跨多个省份,需要建立统一或多地联动的野生兰种质资源保护体系,避免对建兰及其它兰科资源开发利用的盲目性造成种质资源的浪费和破坏。

参考文献

- [1] 朱根发,王碧青,吕复兵. 建兰与纹瓣兰种间杂种胚培养研究[J]. 热带亚热带植物学报, 2005, 13(5): 447-450.
- [2] 唐源江,曹雯静,吴坤林. 基于 SRAP 标记的国兰种质资源遗传多样性分析及分子身份证构建[J]. 中国农业科学, 2015(9): 1795-1806.
- [3] 黄暖爱. GIS 在重点兰科植物地理分布格局与保护现状评价中的应用[D]. 海口: 海南大学, 2008.

- [4] YAÑEZ-ARENAS C, GUEVARA R, MARTÍNEZ-MEYER E, et al. Predicting species' abundances from occurrence data: effects of sample size and bias[J]. *Ecological Modelling*, 2014, 294: 36-41.
- [5] JAYNES E T. Information theory and statistical mechanics[J]. *Physical Review*, 1957, 106(4): 620-630.
- [6] ELITH J, PHILLIPS S J, HASTIE T, et al. A statistical explanation of Maxent for ecologists[J]. *Diversity & Distributions*, 2011, 17(1): 43-57.
- [7] 张路. Maxent 最大熵模型在预测物种潜在分布范围方面的应用[J]. *生物学通报*, 2015, 50(11): 9-12.
- [8] 崔晋亮, 卫海燕, 桑满杰, 等. 基于最大熵模型的蓝莓潜在分布区预测[J]. *山东农业科学*, 2015(9): 36-41.
- [9] TOMLINSON R F. A geographic information system for regional planning[J]. *J Geogr*, 1969, 78(1): 45-48.
- [10] 中国数字植物标本馆. CVH[EB/OL]. <http://www.cvh.org.cn>, 2016.
- [11] 全球生物多样性信息网络. GBIF[DB/OL]. <http://www.gbif.org>, 2016.
- [12] Museum of Vertebrate Zoology. World Clim[EB/OL]. <http://www.worldclim.org>, 2016.
- [13] SWETS J. Measuring the accuracy of diagnostic systems[J]. *Science*, 1988, 240(4857): 1285-1293.
- [14] EFRON B, STEIN C. The jackknife estimate of variance[J]. *Annals of Statistics*, 1981, 9(3): 586-596.
- [15] 黄萍萍, 刘霖, 陈红玉. 建兰的栽培管理[J]. *现代园艺*, 2014(8): 37.
- [16] 李明阳, 席庆, 徐海根. 基于 WEB 数据库和 MaxEnt 模型的白头叶猴潜在生境评价研究[J]. *林业资源管理*, 2010(4): 32-36, 57.
- [17] 胡秀, 吴福川, 郭微, 等. 基于 MaxEnt 生态学模型的檀香在中国的潜在种植区预测[J]. *林业科学*, 2015, 50(5): 27-33.
- [18] 何淑婷, 白碧玉, 但佳惠, 等. 基于 MaxEnt 的南丹参在中国的潜在分布区预测及适生性分析[J]. *安徽农业科学*, 2014, 42(8): 2311-2314.
- [19] 王娟娟, 曹博, 白成科, 等. 基于 Maxent 和 ArcGIS 预测川贝母潜在分布及适宜性评价[J]. *植物研究*, 2014, 34(5): 642-649.
- [20] 弓晓敏, 王建军, 白建伟. 基于 Maxent 和 GIS 的历山国家级自然保护区原麝栖息地预测[J]. *山西林业科技*, 2015, 44(3): 7-9, 23.
- [21] 韩阳阳, 王焱, 项杨, 等. 基于 Maxent 生态位模型的松材线虫在中国的适生区预测分析[J]. *南京林业大学学报(自然科学版)*, 2015, 39(1): 6-10.
- [22] 丁长春, 刘伟, 罗清, 等. 兰属植物引种及栽培研究[J]. *文山学院学报*, 2016, 29(3): 5-9.
- [23] 关文灵, 李枝林, 余朝秀, 等. 云南野生兰花资源的多样性及其引种驯化[J]. *西南农业学报*, 2006, 19(4): 688-691.
- [24] 刘翠华. 江西省野生建兰的 ISSR 遗传多样性研究[D]. 南昌: 南昌大学, 2012.
- [25] 白坚, 胡旭, 周淑婷, 等. 47 个建兰品种的 SRAP 遗传多样性分析[J]. *植物遗传资源学报*, 2012, 13(3): 376-380.
- [26] 胡薇, 黄儒珠, 潘晓华, 等. 建兰 38 个品种的 RAPD 分析[J]. *园艺学报*, 2008, 35(2): 289-294.

Geographical Distribution Prediction and Key Biological Climatic Factors Analysis of *Cymbidium ensifolium* in China Based on Maximum Entropy Model and Geographic Information System

LIANG Chun, LUO Qing, LU Zuzheng, XIE Zhenxing, QIN Qian, HUANG Xinyi

(Guangxi Subtropical Crops Research Institute, Nanning, Guangxi 530001)

Abstract: This research studied the geographical distribution and climatic factors of *Cymbidium ensifolium* in China based on the theory of ecological niche model. The research established the suitable distribution area prediction model and analyzed the key climatic factors of *Cymbidium ensifolium* with geographic information system, based on maximum entropy model theory and existing domestic geographical distribution data of *Cymbidium ensifolium* in China. The results showed that ROC test, the training data AUC was 0.977, testing data AUC was 0.944, indicated the model had good accuracy in prediction, with high practical value. Jackknife test showed that the key biological climatic factors for potential distribution of *Cymbidium ensifolium* were mean temperature of coldest quarter, mean temperature of driest quarter, annual mean temperature, and precipitation of driest month. The major potential distribution areas of *Cymbidium ensifolium* in China were Nanling Mountains area located in northern Guangdong, northeastern Guangxi, and southern part of Jiangxi and Hunan, southern part of Wuyi Mountain located in northern Fujian and south of Lianhua Mountain in Guangdong. With the analysis of model data and predictions, the aim was to provide scientific advice and theoretical basis for the future research of germplasm resources protection, introduction and domestication and large-scale outdoor cultivation for *Cymbidium ensifolium* in China.

Keywords: *Cymbidium ensifolium*; maximum entropy model; geographic information system; geographical distribution prediction; biological climatic factor