

基于生态位模型的石蒜适生区预测^{*1}

毛志远¹ 张兆金² 周坚^{1*}

(1. 南京林业大学生物与环境学院, 南京 210037;

2. 江苏省东海县现代农业园区管理委员会)

【摘要】:运用预设预测规则的遗传算法(GARP)和最大熵(MaxEnt)两种生态位预测模型,以及受试者工作曲线(ROC)分析方法,预测石蒜属石蒜潜在适生区。结果表明,GARP和MaxEnt模型ROC曲线下方的面积AUC(area under the ROC curve)均值分别为0.910和0.988,MaxEnt模型的AUC值更大,预测结果更准确,运行速度更快,更适合用于石蒜的适生区预测。对环境变量进行刀切法表明,在所有环境变量中,最冷季度平均温度对于石蒜分布的影响(贡献)最大,其次是年均温、最冷月的最低温度和最暖季度降水量,而海拔、降水量变化方差对石蒜分布的影响比较小。预测结果显示,石蒜在世界范围内主要分布在亚洲东部以及亚洲中部一小部分,另外北美洲的东部,欧洲南部一小部分地区也适合其生长。在中国范围内主要分布在云南、贵州、福建、江苏、浙江、安徽、江西、重庆、湖北、湖南等省,以及山东、河南、陕西、甘肃等省南部;四川东部和广东、广西(除了南部沿海地区)均预测为适生区,海南、台湾、西藏部分地区也是适生区。

【关键词】:预测模型;石蒜;GARP;MaxEnt;ROC曲线

石蒜(*Lycoris radiata*)为石蒜属多年生草本花卉植物,该物种在包括江苏、广东、四川和陕西等15个省区有分布^[1],是石蒜属植物中分布较广的物种之一。

生态位模型是基于生态位理论,根据物种的生态需求来预测物种的实际分布和潜在分布^[2]的模型。目前,许多研究利用了基于生态位理论的计算机模型软件来开展物种适生区的预测^[3-10]。GARP和Max-Ent是两种最常用的预测物种潜在分布区的生态位模型。本研究采用这两种生态位模型软件开展了对石蒜在全世界范围的物种分布预测,模拟了其潜在适生区,并讨论了石蒜在国内的潜在分布区,为今后石蒜的保护、利用和推广提供参考依据。

1 材料与方法

1.1 数据来源

1.1.1 石蒜分布数据

¹ 收稿日期:2014-07-04 修回日期:2014-09-12

基金项目:国家林业公益性行业科研专项(201004056)。

作者简介:毛志远(1986—),男,硕士生,主要研究方向为植物发育学。

通讯作者:周坚,男,教授。E-mail:zhiwu@njfu.edu.cn

石蒜数据来自世界生物多样性信息机构(<http://www.gbif.org/>)和中国数字植物标本馆(<http://www.cvh.org.cn/cms/>)网站,共获得石蒜在国外的分布点129个和在中国的分布点114个(地理单位尽量统一到县级单位),最后整理为243个全球已知石蒜的分布点资料,见图1。



图1 石蒜在全球的分布点

1. 1. 2 环境数据来源

参考张海娟等^[10]方法,全球气候环境数据为1950—2000年间气象资料,含19个生物气候变量和1个海拔变量(来自WORLDCLIM)。

1. 1. 3 分析底图

全球矢量地图下载地址为<http://www.diva-gis.org/data/misc>,文件名为word_adm0.zip。中国省级行政区划图从国家基础地理信息系统下载(<http://nfgis.nsd.gov.cn/>,1:400万)。

1. 2 预测模型及相关软件

主要使用4种软件,即基于GARP算法模型的Desktopgarp(版本1.1.6),基于最大熵模型 MaxEnt(版本3.3.3k),DIVA-GIS(版本7.5.0)和Esri公司的Arc-GIS^[10]。

1. 2. 1 GARP模型的预测

物种分布数据的处理:物种已知分布点数据转换成Desktopgarp模型软件支持的格式,主要包含3列内容,即物种名,物种分布经度和纬度,再通过模型软件中Upload Data Points功能加载。选择70%的分布数据(约172个点)作为训练数据,其余数据作为检验数据^[10]。

环境数据的处理:通过ArcGIS软件中的Raster toascii功能可以直接将ESRI栅格下载文件转换成.asc格式文件。Asc文件再通过Desktopgarp中的DatasetManager处理转换成Desktopgarp可识别的格式.raw。环境数据是以数据集的形式加载到软件中的^[3]。

参数设置:Desktopgarp采用由不同法则共同组成的模型方案,设置最大迭代次数1000次,收敛值0.01,运行次数设定为20^[4],并启用最优规则集合参数选项^[5]。输出结果为ARC/INFO Grids格式。用Arcmap将其中10个最佳模型叠加,最后形成石蒜在世界的潜在分布区。然后加入中国区域的分析底图,得到在中国区域的潜在分布区。

1. 2. 2 MaxEnt 模型的预测

最大熵 MaxEnt 模型原理是确保可能分布满足所有的已知信息和条件，即达到了最大熵^[6]。

物种数据的处理:将 Desktopgarp 的物种已知分布点数据文件格式在 Excel 中转为 . csv 格式文件导出，通过 Browse 加载到 MaxEnt 软件中。

环境数据的处理:下载的 20 个生物气候变量和海拔变量，其下载格式为 ESRI 栅格，通过 ArcGIS 软件中的 Raster toascii 功能可以直接转换成 MaxEnt 软件可识别的 . asc 格式文件。把格式 . asc 文件的环境数据通过 Browse 加载到 MaxEnt 软件中^[7]。

参数设置:设置 30%的分布数据为检验数据，其余为训练数据，最大迭代次数 1000 次，收敛值 0. 01^[7]。

选择自动特征规则，获得 cumulative 类型和 . asc 格式的结果输出文件。通过 ArcGIS 软件中的 asciito Raster 功能将输出文件转成 . grd 格式，最后在 Arcmap 中处理，最后形成石蒜在世界的潜在分布区。然后加入中国区域的分析底图，得到中国区域的潜在分布区^[8]。

1. 2. 3 ROC 曲线的绘制

ROC 曲线的应用非常广泛，是一个基本的评价技术。曲线绘制通过 DAVIDA-GIS 软件完成，将石蒜预测图转换成 . grd 栅格图，使用软件的 modeling 功能生成相应的 . roc 文件作为验证文件，最后利用 ShowROC/Kappa 选项中导入 . roc 文件，得到石蒜的 ROC 曲线和 AUC 值^[9]。ROC 曲线的绘制方法和判别标准参照文献^[10]。

2 结果与分析

2. 1 两种模型 AUC 值及其比较

通过 DAVIDA-GIS 软件计算 GARP 和 MaxEnt 的 AUC 值均值分别是 0. 910 和 0. 988，都大于 0. 9，说明两种软件计算的石蒜模型都具有较高的诊断价值。相对而言，因为 MaxEnt 的 AUC 值较大(0. 988)，所以它可以更准确地预测石蒜的适生区^[10]。本研究最后选用 MaxEnt 模型的分析结果来进行石蒜在中国的适生区预测。

2. 2MaxEnt 模型下的石蒜适生区预测

在 MaxEnt 模型下，生成石蒜的全球适生性指数图(图 2)以及中国适生性指数图(图 3)。根据重叠系数(OverlayValue, OV)的大小将石蒜预测图平均划分为 5 个等级，绿色区域为石蒜非适生区，浅绿色区域为石蒜低适生区，黄色区域为石蒜中适生区，橘黄色区域为石蒜高适生区和红色区域为石蒜最佳适生区^[11]。

在 MaxEnt 模型下，石蒜在全球适生区的分布范围，主要在亚洲东部以及亚洲中部一小部分，另外北美洲的东部，欧洲南部一小部分地区也适合其生长(图 2)。



图2 MaxEnt 模型石蒜全球适生区预测

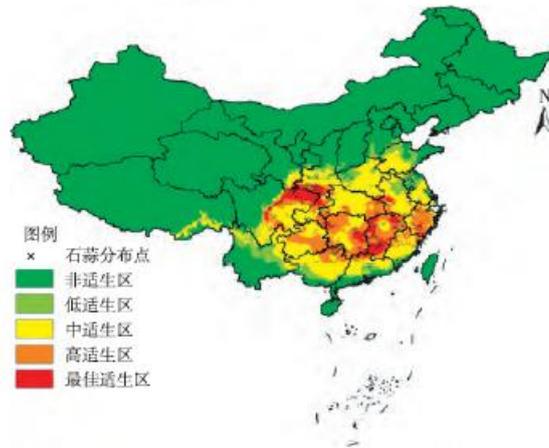


图3 MaxEnt 模型石蒜中国适生区预测

由图3也可以看出:已知在中国大陆生长石蒜的114个分布点都在适生区预测范围内,说明模型的预测结论具有一定的可靠性。石蒜在中国大陆适生区主要集中在秦岭—淮河以南及附近地区,包括了云南、贵州、福建、江苏、浙江、安徽、江西、重庆、湖北、湖南等省大部;山东、河南、陕西、甘肃等省南部;四川东部,广东和广西(除了南部沿海地区)也预测是适生区;另外海南、台湾沿海、西藏有一小部分地区也是适生区。

2.3 适生区分布与环境变量的关系

利用MaxEnt模型中的“刀切法”(Dojackknifetomeasurevariableimportance)分析了环境变量对于石蒜分布的贡献(图4)。分别建立了3种模型:用所有变量建模(红色条带)、单独用每一变量进行建模(深蓝色条带)和依次排除1个变量并用剩余的变量建模(绿色条带)。观察绿色条带与红色条带之间的变化,如果变化大,说明排除的那个变量对石蒜分布贡献大,反之,对石蒜分布贡献小^[9]。从图4可以看出,最冷季度平均温度(bio_11)对于石蒜分布的贡献最大,次之年均温(bio_1)、最冷月份最低温度(bio_6)和最暖季度降水量(bio_18)对石蒜分布的影响也比较大,海拔(alt)和降水量变化方差(bio_15)对石蒜分布的影响比较小。

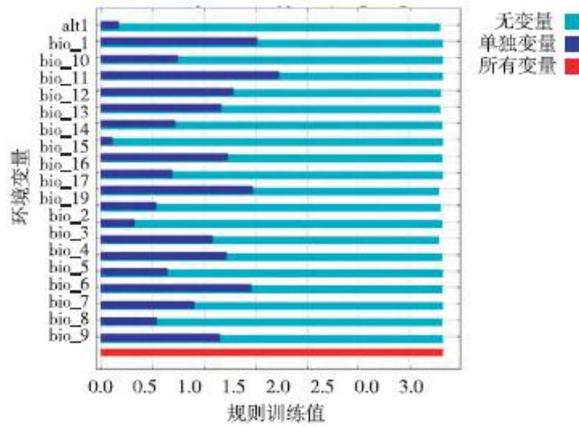


图4 刀切法检测生物气候变量对分布的重要性

石蒜分布值之间的响应曲线见图5。图5A是Maxent软件输出图，bio_11数据单位为摄氏温度，但数据合成时候被放大10倍。因此，图5B为修正后石蒜分布值之间的响应曲线。从图5B可见，在-2~0℃这个温度区间，分布值随温度升高单调增大，高过8℃时，分布值随温度升高单调减少。这表明：最冷季度平均温低于-2℃或高于8℃的地区都不适合石蒜分布。根据石蒜分布值之间的响应曲线分析，石蒜喜欢温暖湿润亚热带气候和暖温带气候区域的冬季季节。如果最冷季度温度过高或过低都会对石蒜叶片生长及光合作用产生不利影响。

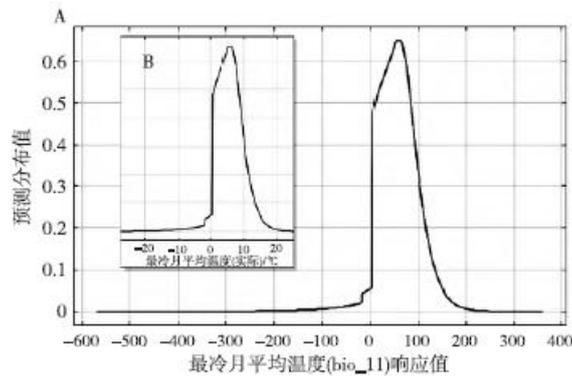


图5 预测分布值与最冷季度平均温度的响应曲线

3 结论与讨论

(1)模型的选择。在过去的20多年里，物种分布模型有很快的发展，提出了多个可用的方法模型^[12]。Wisz等^[13]提出MaxEnt在大、小样本数据下均能很好地预测物种分布，即MaxEnt模型的预测效果比GARP模型好。吴文浩等^[14]通过比较也发现，与GARP模型相比，MaxEnt模型不管是在物种现实生境模拟，还是在环境因子对物种生境影响的定量描述方面都有较高的性能。笔者在使用MaxEnt和GARP模型对石蒜适生区预测时，也是前者模型预测效果比较好，即采用MaxEnt软件可以更准确地预测石蒜的适生区。这也与Wisz和吴文浩等的实验结果相似^[13-14]。

(2)环境因子对于模型预测的影响。石蒜的分布跟许多外界因素有关。本实验由于条件的限制，只使用了19个生物气候因子和地形因子中的海拔变量，缺少土壤因子、其他地形因子、人为干扰因子、全球变暖和生物相互作用因子等对预测结果可能有大的影响的因子，因此，不是预测区域中所有地点都适合石蒜的生长。

(3)环境因子与石蒜分布的关系。从以上模拟可以看出，温度、降水量、光照对于石蒜分布的影响显著，尤其是最冷季度平

均温度对它的影响最大。中国西北、东北的温带地区最冷季度温度偏低，降雨偏少，有大片的荒漠和盐碱地，光照过强，因此被预测为石蒜的非适生区；而南方大多数雨水充沛的地区，温度和光照比较适宜，是石蒜较易繁殖、较易适生和较易引种的地区。这些模拟有石蒜分布的区域是建立石蒜种植基地、发展石蒜产业优先考虑的选择地点。

(4)潜在分布区分析。本次模拟发现一些预测为有分布但没有标本采集记录的地区，如亚洲中部一小部分以及欧洲南部一小部分，查询世界生物多样性信息机构(GBIF)数据库也未发现相关记录。这种现象在多数情况下可以认为是原本就不是石蒜的自然分布区。而在中国境内也发现一些预测为有分布但没有标本采集记录的地区，如西藏的一小部分地区，云南北部一部分地区，河南省、山东省等。产生这一现象的原因是非常复杂的，有可能是调查遗漏，也可能是该物种本身在那些地区并无分布，这都会对模型模拟的准确性产生影响^[9]。潜在分布区分析对未来指导石蒜引种驯化工作有一定的理论参考意义。

参考文献:

- [1] 中国植物志编委会. 中国植物志: 16 卷 [M]. 北京: 科学出版社, 1981.
- [2] 朱耿平, 刘国卿, 卜文俊, 等. 生态位模型的基本原理及其在生物多样性保护中的应用 [J]. 生物多样性, 2013, 21(1) : 90 — 98.
- [3] Stockwell DRB. Genetic algorithms II. Machine learning methods for ecological applications [M]. Boston: Kluwer Academic Publishers, 1999: 123—144.
- [4] Stockwell DB, Peters DP. The GARP modeling system: problems and solutions to automated spatial prediction [J]. International Journal of Geographical Information System, 1999, 13(2) : 143—158.
- [5] Anderson R P, Lew D, Peterson A T. Evaluating predictive models of species' distributions: criteria for selecting optimal models [J]. Ecological Modelling, 2003, 162(3) : 211—232.
- [6] 郭水良, 高平磊, 娄玉霞. 应用 MaxEnt 模型预测检疫性杂草毒莠苣在中国的潜分布范围 [J]. 上海交通大学学报: 农业科学版, 2011, 29(5) : 15—19.
- [7] Phillips S J, Dudik M. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation [J]. Ecography, 2008, 31(2) : 161—175.
- [8] Phillips S J, Anderson R P, Schapire R E. Maximum entropy modeling of species geographic distributions [J]. Ecological Modelling, 2006, 190(3 /4) : 231—259.
- [9] 王雷宏, 杨俊仙, 黄成林. 用 MaxEnt 模拟湖北海棠 (Malus hupehensis) 地理分布 [J]. 安徽农业大学学报, 2013, 40(3) : 146—149.
- [10] 张海娟, 陈勇, 黄烈健, 等. 基于生态位模型的薇甘菊在中国适生区的预测 [J]. 农业工程学报, 2011, 27(S1) : 413—418, 420.
- [11] 余岩, 陈立立, 何兴金. 基于 GARP 的加拿大一枝黄花在中国的分布区预测 [J]. 云南植物研究, 2009, 31(1) : 57—62.

[12] Elith J, Phillips S J, Hastie T, et al. A statistical explanation of MaxEnt for ecologists [J]. *Diversity and Distributions*, 2011, 17(1) : 43–57.

[13] Wisz M S, Hijmans R J, Li J, et al. Effects of sample size on the performance of species distribution models [J]. *Diversity and Distributions*, 2008, 14(5) : 763–773.

[14] 吴文浩, 李明阳. 基于生态位模型的松材线虫潜在生境预测方法研究 [J]. *林业调查规划*, 2009, 34(5) : 33–38.