

用 MaxEnt 模拟湖北海棠 (*Malus hupehensis*) 地理分布

王雷宏¹, 杨俊仙^{2*}, 黄成林¹

(1. 安徽农业大学林学与园林学院, 合肥 230036; 2. 安徽农业大学理学院, 合肥 230036)

摘要: 采用湖北海棠 551 个分布数据以及当前生物气候数据, 用 MaxEnt 软件模拟了它的地理分布。结果显示: ①潜在分布地区有巴基斯坦北部、印度北部、尼泊尔、不丹、缅甸东北部、越南西北部、广西北部、西藏南部及东南部、日本岛东南侧、朝鲜半岛中南部及沿海地区; 河北东北部、山东半岛北部和辽宁南部, 其中分布值在 0.6~1 的地区有不丹、韩国济州岛及附近岛屿和日本及附近岛屿。②分布记录非常少, 但分布值已达 0.6~1 的地区有山西东南部、江苏省南通、四川雅安市汉源和山东崂山。③以分布值在 0.6~1 的地区来表示现实地理分布格局, 自西向东分布为喜马拉雅南部, 东南部-横断山-秦岭-巫山-长江中下游山地-武夷山-朝鲜半岛-日本岛。自北向南, 北界在太行山南部, 南界武夷山西南部。④Jackknife test 检测表明干旱季节的平均温度对湖北海棠的分布增益贡献最大, 本种自然分布于干旱季节平均气温在-10℃~15℃的地区。

关键词: 湖北海棠; MaxEnt; 地理分布

中图分类号: S718.545

文献标识码: A

文章编号: 1672-352X (2013)03-0383-04

Modelling geographic distribution of *Malus hupehensis* with MaxEnt

WANG Lei-hong¹, YANG Jun-xian², HUANG Cheng-lin¹

(1. School of Forestry & Landscape of Architecture, Anhui Agricultural University, Hefei 230036;

2. School of Science, Anhui Agricultural University, Hefei 230036)

Abstract: Geographic distribution of *Malus hupehensis* was modelled with MaxEnt by 551 presence recorded data and current bioclimatic data. The results showed that: ①*Malus hupehensis* would be more likely to occur in those areas (potential distribution), such as northern Pakistan, northern India, Nepal, Bhutan, northeast Myanmar, northwest of Vietnam, northern of Guangxi, the southern and southeastern of Xizang, southeast of Japan, south, center and coast of Korea, northeast of Hebei, north of Shandong, southern of Liaoning. Of which, the areas with 0.6-1 distribution value (logistic) are Bhutan, Cheju Island and coastal islands of Korea, Japan islands and the its nearby islands. ② The areas with distribution values 0.6-1 were in the southeast of Shanxi, Nantong of Jiangsu, Hanyuan of Sichuan, and Laoshan of Shandong, although, there were a few present records in those areas. ③ Geographical distribution patterns of *M. hupehensis* could be expressed using those areas with distribution value 0.6-1, from west to east, the southern and southeastern of the Himalayas, Hengduan Mountains, Qinling Mountains, and Wu Mountains, the mountainous regions of middle and lower reaches of Yangtze River, Wuyi Mountains, Korean Peninsula, the Japan islands. From north to south, the most northern boundary was in the south of the Taihang Mountains, and the southern boundary was in the southwest of the Wuyi Mountains. ④ It was showed that the average temperature in the dry season was the greatest contribution to the distribution gain of *M. hupehensis* by Jackknife Test, and it was distributed naturally over those areas with average temperature of -10℃-15℃ in the dry season.

Key words: *Malus hupehensis* (Pamp.) Rehd.; maximum entropy method; geographic distribution

收稿日期: 2012-03-28

基金项目: 安徽省优秀青年自然科学基金 (10040606Q18) 和安徽农业大学人才引进基金 (2008010) 共同资助。

作者简介: 王雷宏, 男, 博士, 讲师。E-mail: leihong_wang@yahoo.cn

* 通信作者: 杨俊仙, 女, 讲师。E-mail: yangjunxian@ahau.edu.cn

湖北海棠 (*Malus hupehensis* (Pamp.) Rehd.) 是一种重要的苹果砧木和观赏海棠育种材料, 与山荆子 (*Malus baccata*), 毛山荆子 (*Malus mandshurica*) 近缘, 它们之间形态区分较难, 分布区范围仍有不明确处, 近年经大量的变异式样调查研究, 已明确其形态特征^[1-5], 依此对中国主要标本馆的腊叶标本重新进行了鉴定, 结果粗略显示秦岭淮河以北的地区也有分布, 如山西的中南部也有分布, 由于各地区采集的标本数量和质量差别较大, 很难知晓较详细的地理分布格局以及潜在分布地区, 而物种地理分布模型模拟可以详细反映物种的分布格局及潜在分布区, 此类软件较多, 常见的有 BIOCLIM, SVM, 广义模型及回归树等, 近年对大量的不同模型模拟比较研究证实, MaxEnt 模拟分布较为客观、细腻、准确, 可以逼近自然的分布格局, 并能反映出潜在分布区以及分布与生物气候因子的关系^[6-7]。本研究的目的在于用最大熵模型 (MaxEnt) 模拟湖北海棠的地理分布, 明确这个种的地理分布格局及潜在分布区, 分布多度, 分析其与生物气候因子的关系, 明确其基本的生态特性, 发现采集遗漏点, 为下一步开展分子谱系地理学研究居群采样做好准备。

1 材料与方法

1.1 数据来源

依据湖北海棠和山荆子变异式样研究得出的形态区分特征, 对中国科学院北京植物所标本馆(PE), 江苏省中国科学院植物研究所植物标本馆(NAS), 南京大学植物标本馆(N), 南京林业大学树木标本馆(NF)等苹果属标本进行了重新鉴定, 共获得分布记录数据 551 个, 用地名数据库(<http://v2.cvh.org.cn/>)和 Google Earth (<http://www.google.com/earth/index.html>) 转换出采集点的经纬度坐标, 输入 Excel 表格, 保存成 *.csv 格式。气候数据来自 WorldClim (<http://www.worldclim.org/>) 的 1950~2000 年生物气候数据 (Bioclim), 分辨率 2.5 arc-minutes^[8], 包括 19 个生物气候变量, bio1: 年均温; bio2: 温度较差; bio3: 年均温占年温较差百分比($\text{bio1}/\text{bio7} \times 100$); bio4: 温度季节变化方差; bio5: 最热月极高温; bio6: 最冷月极低温度; bio7: 年温较差; bio8: 雨季均温; bio9: 干季均温; bio10: 暖季均温; bio11: 冷季均温; bio12: 年降雨量; bio13: 最湿月降雨量; bio14: 最干月降雨量; bio15: 季降雨量变异系数; bio16: 雨季降水量; bio17: 干季降水量; bio18: 暖季降雨量; bio19: 冷季降雨量。下载后, 用

DIVA-GIS^[9-10]把数据转换成 ASC II 格式, 分析底图用 1:400 万的亚洲(<http://www.diva-gis.org>)和中国地图(<http://nfgis.nsd.gov.cn>)。

1.2 研究方法

将以上分布数据和环境数据导入 MaxEnt (www.cs.princeton.edu/~schapire/maxent)^[6], 随机选取 25% 的分布点作为测试集(test data), 剩余的作为训练集(training data), 其中再从训练集中随机取 10% 的分布数据来计算分布阈值, 最大迭代次数为 500 次, 收敛阈值为 10^{-5} , 输出数值逻辑斯蒂值(logistic), 表示分布概率(简称分布值), 输出文件类型 ASCII, 其余参数仍为软件默认值, 用 DIVA-GIS 软件可以直接打开 MaxEnt 生成的 ASCII 文件, 显示矢量图结果。MaxEnt 用受试者工作特征曲线(Roc Curve)检验模拟分布的适合度。用刀切法(Jackknife Test)检测生物气候变量与物种分布值之间的关系。

2 结果与分析

2.1 标本记录分布现状

从标本记录来看, 湖北海棠分布在北纬 22.40° ~ 37.87° 之间, 东经 98.50° ~ 127.78° 之间, 湖北海棠的分布北界已达山西晋中地区, 西北界可达甘肃兰州西部地区, 西南界达云南腾冲县, 南界可达香港, 东界可达台湾。从标本的密集度来看, 主要集中分布于重庆、湖北、江西、浙江等省, 然而这种分布与标本采集的强度、频度和质量是密切相关的, 仍然不能反映可能的潜在分布区和地理分布格局, 具体见图 1。图中实心圆点为标本采集记录点。

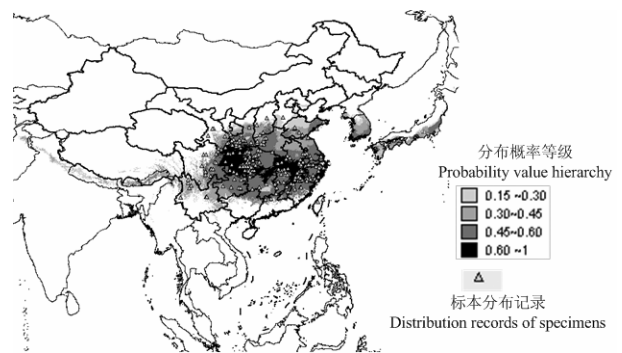


图 1 湖北海棠标本记录和最大熵模拟分布

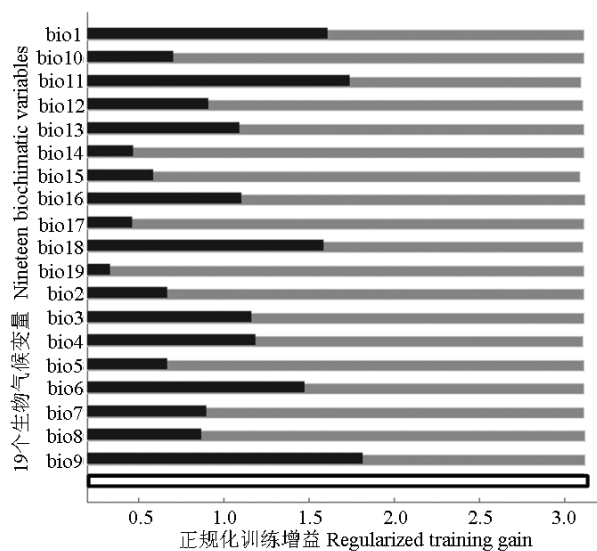
Figure 1 Distributions of specimen records and modeling with MaxEnt in *Malus hupehensis*

2.2 MaxEnt 模拟分布

模拟检验显示训练集、验证集的遗漏率与预测遗漏率一致性较高, 受试者特征曲线显示训练集

的 AUC 值为 0.986, 验证数集的 AUC 值为 0.984, 远大于随机预测的 AUC 值 0.5, 本次模拟效果良好。训练集 10% 的分布点产生分布阈值 0.1496, 意义在于分布值小于 0.15 地区可以认为不存在湖北海棠。在 DIVA-GIS 软件中, 取原分布值 2 位小数, 把分布值设为 4 个等级, 其中 0.15~0.30 为浅灰色, 0.3~0.45 为灰色, 0.45~0.6 为灰黑色, 0.60~1 为黑色。三角形表示已有采集标本的分布点。具体见图 1 中图例。分布值在 0.15~1 之间, 没有采集记录的地区都可定性为潜在分布区域, 有巴基斯坦北部、印度北部、尼泊尔、不丹、缅甸东北部、越南西北部、广西北部、西藏东南部、日本岛东南侧、朝鲜半岛中南部及沿海地区、河北东北部、山东半岛北部和辽宁省南部, 其中分布值在 0.6~1 间的不丹、韩国济州岛及附近岛屿、日本岛及附近岛屿属于采集遗漏点的可能性较大。标本记录非常少, 分布值已达 0.6~1 的地区有山西南南部、江苏南通、四川汉源和山东崂山。

模拟软件默认分布值 0.5 的地区为典型分布区, 结合实际存在的记录数据, 可用分布值在 0.6~1 的地区来表示湖北海棠现实地理分布格局, 从西向东, 依次为喜马拉雅南部, 东南部-横断山-秦岭-巫山-长江中下游山区-武夷山-朝鲜半岛-日本岛。南北向, 北界在太行山南部, 南界在江西省南部。



黑边条形: 所有变量; 黑色条形: 单一变量; 灰色条形: 无变量

Bar with black border: With all variables; Black bar: With only variable; Gray bar: Without variable

图 2 Jackknife test 检测生物气候变量对分布的重要性

Figure 2 Using jackknife to measure Bioclim variables for gain of distribution

2.3 地理分布与生物气候变量的关系

用 Jackknife test 检测生物气候变量对于分布增益的贡献, 创建了 3 种模型, 具体结果见图 2。第 1 种依次排除一个变量, 用其余变量建模; 第 2 种用单独每个变量建模; 第 3 种是用所有变量进行建模, 用深蓝色条代表仅一个变量时对分布的增益值, 浅灰色条代表排除本变量, 其余变量对分布的增益值, 红色代表所有变量对分布的增益值。干旱季节的平均温度 (bio9) 对分布的增益最大, bio19 (冷季降雨量) 对分布几乎没有增益。具体结果见图 3。其中干旱季节的平均温度与分布值之间的响应曲线显示, 从 -10°C 到 5°C 这个温度区间, 分布值随温度升高单调增大, 高过 5~15°C 时, 分布值随温度升高单调减少, 在干旱季节时, 低于 -10°C 或高于 15°C 的地区都不适合湖北海棠分布。显然在干旱季节, 温度过高或过低都会造成进一步的干旱, 湖北海棠喜温暖湿润的亚热带气候。

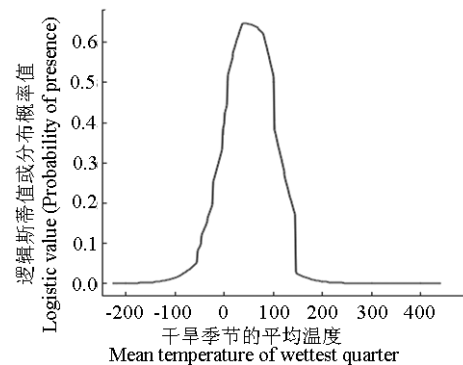


图 3 预测分布值与变量 bio9 的响应曲线

Figure 3 Response curves between logistic prediction value and bio9

3 讨论

MaxEnt 是一个优秀的生物地理分布模拟软件, 近来应用较多, 在多种模拟软件对比研究中, 已证实其优良的可靠性^[7]。本次直接用 MaxEnt 对湖北海棠的地理分布进行模拟, 采用了重新鉴定的标本共 551 个分布记录, 生物气候数据空间分辨率是 arc 2.5 min, 未考虑土壤和植被的影响。分布阈值 0.1496 是以一般情况下 MaxEnt 默认的 10% 训练值计算出, 4 个分布值等级也由软件自然划分, 为了方便, 笔者只取分布值两位小数。软件默认分布值 0.5 地区是典型的分布区, 只是表明这些分布区与标本采集记录吻合程度较高, 并不代表 0.5 以下就绝对没有分布记录。以 10% 训练集计算出的分布阈值 0.1496, 代表的意义在于, 只允许训练集中 10% 的标本记录可能是标本鉴定失误、分布逸出或者为栽培分布。

考虑到湖北海棠分布之广, 这样的处理显得更为科学合理, 并不能一概以传统默认的分布值 0.5 为界进行人为强制划分。

本次模拟发现一些具高分布值的地区, 但没有采集记录, 如不丹、韩国、日本的一些地区, 查 Global Biodiversity Information Facility (GBIF) 数据库也未发现这些地区有分布记录。从湖北海棠的繁殖特性、生态习性、形态变异来看, 受地理隔离的可能性较少, 与大陆隔离的台湾岛有分布记录, 分布值也只有 0.15~0.3, 具高分布值没有采集记录的最大可能是调查的遗漏。在中国境内也发现一些具高分布值, 没有分布记录的地区, 如山西太行山东南部、山东青岛崂山沿海地区, 而这些地区却有苹果属山荆子 (*Malus baccata*) 的记录, 实际情况是往往把湖北海棠误鉴定为山荆子。山东泰山和蒙山分布虽有集中大片的群落^[2-3], 模拟结果显示这些地方分布值在 0.45~0.6 间。

与已有的山荆子的地理分布研究^[11]相比较, 显示湖北海棠、山荆子在地理分布上存在嵌套分布, 然而对分布影响的生物气候变量不同, 山荆子受干旱季节的降雨量影响较大, 湖北海棠受干旱季节的温度影响较大, 表明即使分布在同一地区也占有不同的生态位。已有证据表明湖北海棠是一个杂交起源种, 而与近缘种山荆子、三叶海棠等系统发育关系又不明确^[12], 那么这些近缘种之间在进化历史上可能存在一些特定的种群杂交历史事件, 当然近期的自然杂交也在不断进行, 这也正是交错分布区鉴定不明的重要原因。下一步计划开展湖北海棠的谱系地理学研究, 地理居群取样主要依据本次分布值在 0.6 以上的地区, 这些地区的周边, 分布值在 0.6 以下分布记录较多地区可作为辅助采样区, 对本研究模拟发现的可能遗漏点、鉴定可能有误的地区要进行重新野外调查研究。

参考文献:

- [1] 钱关泽. 苹果属(*Malus* Mill.)分类学研究[D]. 南京: 南京林业大学, 2005.
- [2] 钱关泽, 邵文豪, 刘莲芬, 等. 山东湖北海棠(*Malus hupehensis* (Pamp.) Rehd.)两新变种[J]. 植物研究, 2007, 119(05): 521-524.
- [3] 陈曦. 湖北海棠(*Malus hupehensis*)不同居群变异式样及遗传多样性的研究[D]. 南京: 南京林业大学, 2009.
- [4] 王雷宏, 汤庚国, 夏海武, 等. 山荆子叶脉序的研究[J]. 南京林业大学学报: 自然科学版, 2008, 32(2): 39-42.
- [5] 王雷宏. 山荆子变异式样研究[D]. 南京: 南京林业大学, 2008.
- [6] Phillips S J, Anderson R P, Schapire R E. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation[J]. Ecography, 2008, 31(2): 161-175.
- [7] 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.
- [8] Hijmans R J, Cameron S E, Parra J L, et al. Very high resolution interpolated climate surfaces for global land areas[J]. International Journal of Climatology, 2005, 25: 1965-1978.
- [9] Jarvis A, Ferguson M E, Williams D E, et al. Biogeography of wild Arachis: Assessing conservation status and setting future priorities[J]. Crop Science, 2003, 43(3): 1100-1108.
- [10] Jarvis A, Guarino L, Williams D, et al. The use of GIS in the spatial analysis of wild peanut distributions and the implications for plant genetic resources conservation[J]. Plant Genetic Resources Newsletter, 2002, 131: 29-35.
- [11] 王雷宏, 杨俊仙, 郑玉红, 等. 苹果属山荆子地理分布模拟[J]. 北京林业大学学报, 2011, 33(03): 70-74.
- [12] Robinson J P, Harris S A, Juniper B E. Taxonomy of the genus *Malus* Mill. (Rosaceae) with emphasis on the cultivated apple, *Malus domestica* Borkh[J]. Plant Systematics and Evolution, 2001, 226(1): 35-58.