

分类号

密级

UDC



北京林业大学

学术型硕士毕业论文

基于几何形态和深度学习的栎类叶片形态研究

Leaf Morphology Analysis Based on Geometric Morphometric Analyses (GMMs)
and Deep Learning in Oaks

苏蔚

指导教师 杜芳

学院 生态与自然保护学院

学科专业 生态学

研究方向 恢复生态与生物多样性保护

二〇二一年 六月 十六日

独创性声明

本人声明所呈交的论文是本人在导师指导下独立进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得北京林业大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

签 名：_____ 日 期：_____

关于学位论文使用授权的声明

本人完全了解北京林业大学有关保留、使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属北京林业大学；学校有权保留并向国家有关部门或机构送交论文的纸质版和电子版，允许学位论文被查阅、借阅和复印；学校可以将学位论文的全部或部分内容公开或编入有关数据库进行检索，可以允许采用影印、缩印或其它复制手段保存、汇编学位论文。

签 名：_____ 导师签名：_____ 日 期：_____

本论文研究工作受国家自然科学基金项目（项目编号：
42071060；项目名称：栎类生态适应的分子机理研究）的资助，
在此表示感谢！

摘要

栎类叶型变异巨大，物种鉴定困难，曾引起包括达尔文在内的进化生物学家关注。本论文首先利用分子标记结果作为先验，综合分析比较了几何形态学分析和深度学习对壳斗科 (Fagaceae) 栎属 (*Quercus* L.) 白栎组 (sect. *Quercus*) 两近缘物种槲栎 (*Quercus aliena* Bl.) 槲树 (*Quercus dentata* Thunb.) 的鉴别能力。并且基于几何形态学，探讨了叶片异速生长模式、评估假设的模块以及由地理和气候因子对叶形态特征的影响。主要结论如下：

(1) 利用 12 对 SSR (Simple Sequence Repeat) 分子标记对槲栎和槲树 54 个种群中的 710 个个体进行遗传分组。以混合系数 $Q = 0.9$ 为标准将槲栎槲树个体分为纯种和杂交个体， $0.1 < Q < 0.9$ 的个体视为杂交个体， $Q \geq 0.9$ 和 $Q \leq 0.1$ 的个体为纯种，结果得到槲栎个体 149 个，槲树个体 322 个，杂交个体 239 个。此遗传标记的分组结果作为先验。

(2) 选取落叶栎常用的 13 个标志点对 710 棵树上采集到 3657 片完整成熟的叶片进行标记。通过 Procrustes 叠印分析将叶片形态分为对称组分 (symmetric component) 和非对称组分 (asymmetric component)，并通过主成分分析 (Principal Component Analysis, PCA)、典型变量分析 (Canonical Variate Analysis, CVA) 和判别分析 (Discriminant Analysis, DA) 研究了叶形各组分的差异。研究发现，槲栎和槲树的叶形 (包括对称组分和非对称组分) 存在显著差异，并且叶片的非对称组分区分程度不高。槲栎槲树沿 CV1 几乎全部分离，而杂交个体散布于两个种的分布范围之间。在物种间有 90% 的槲栎个体和 95% 的槲树个体可以被正确判别。基于典型变量分析和判别分析研究了两近缘种同域异域对叶形的影响，结果表明异域的槲栎叶形有向槲树接近的趋势。

(3) 以槲栎槲树及杂交个体叶片扫描图像 2221 张为研究对象，基于卷积神经网络 (Convolutional Neural Network, CNN) 的 Xception 深度学习分类模型对栎类分类时，槲栎和槲树物种间的分类的准确率达到 95.8%；槲栎和槲树和杂交个体的分类准确率分别为 71.8% 和 67.9%；当对槲栎、槲树以及杂交个体进行三分类时准确率为 44.5%。深度学习分类结果表明，槲栎和槲树基于叶片扫描图的效果较好，杂交个体基于叶片特征无法与两个种进行区分。

(4) 通过双偏最小二乘法 (Two Blocks Partial Least Squares, 2B-PLS) 分析了叶片形状 (对称和非对称组分) 与叶片大小的共变关系。结果表明, 对称组分的槲栎槲树存在显著的异速生长模式, 而叶片非对称组分中为等距生长。

(5) 通过基于距离的冗余分析 (distance-based Redundancy Analysis, dbRDA) 和双偏最小二乘 (Two Blocks Partial Least Squares, 2B-PLS) 的双图 (biplots) 研究了叶片形态 (物种水平及对称性水平) 与地理和气候因素之间的关系。结果表明, 槲栎叶形与纬度变化显著相关; 八月份降水 (prec08) 和平均日较差 (bio15) 对叶形有影响。

(6) 基于 13 个叶形标志点对槲栎槲树 471 个体进行模块化 (Modularity) 分区, 提出模块化假设: 槲栎 (对称组分) 及槲树 (对称、非对称) 分为叶尖部、叶中下部及叶柄三个模块, 槲栎的非对称组分分为叶片和叶柄两个模块。

本论文通过比较分析深度学习和叶片几何形态对栎类叶片鉴定的能力表明以上两种方法都具有较高的分类精度, 可为其他叶型复杂的近缘物种鉴定提供参考。地理和气候因子对叶形的影响体现了叶片对环境的适应性, 同时探究叶片模块化也检测了其形态结构内的强相关性, 这些都有利于我们从多角度更深入的理解物种的生态与进化。

关键词: 栎属, 几何形态学, 深度学习, 叶片形态特征, 模块化

Leaf Morphology analysis based on Geometric Morphometric Analyses
(GMMs) and deep learning in oaks

Master Candidate: WEI SU

(Ecology)

Supervised by Prof. Dr. Fang DU

Abstract

The great variation of oak leaf shape and the difficulty of species identification have attracted the attention of evolutionary biologists since Charles Darwin. Based on a comprehensive comparison by leaf Geometric Morphometric Analyses (GMMs) and deep learning methods using microsatellite markers as *a priori*, this thesis aims to identify leaf shape difference between two closely related oaks of *Quercus* of Fagaceae, *Quercus aliena* Bl. and *Quercus dentata* Thunb.. Based on GMMs, allometric growth patterns, evaluated hypothesis modularity, and the effects of geographical and climatic factors on leaf morphological characteristics were discussed. The main conclusions are as follows:

(1) 710 individuals from 54 populations of *Q. aliena* and *Q. dentata* were assigned by 12 pairs of simple sequence repeat (SSR) molecular markers. 149 individuals were determined as *Q. aliena*, 322 individuals were determined as *Q. dentata*, and 239 individuals were determined as hybrids. Using *Q* value as the criteria for assignment. Individuals of *Q* value between 0.1 and 0.9 were considered to be hybrids, *Q* value more than 0.9 or *Q* value less than 0.1 were considered to be pure species. The result of genetic assignment used as *a priori* in the following analysis.

(2) Geometric morphometric analysis was conducted basing on 3657 leaves of 710 individuals with 13 commonly used landmarks. The leaf shape was divided into symmetrical

and asymmetric components by generalized procrustes analysis (GPA). The differences of leaf shape components were studied by principal component analysis (PCA), canonical variable analysis (CVA) and discriminant analysis (DA). Leaf shape of *Q. aliena* and *Q. dentata* (including symmetrical components and asymmetrical components) were significantly different, and it was not obvious in asymmetrical components of leaf discrimination. Almost all individuals were separated along CV1, and hybridized individuals were scattered between the two species. Among species, 90% of *Q. aliena* and 95% of *Q. dentata* can be correctly identified. DA and CVA results showed the influence of *Q. aliena* and *Q. dentata* on leaf shape. The leaf shape of allopatric *Q. aliena* was close to *Q. dentata*.

(3) 2,221 images of *Q. aliena*, *Q. dentata* and hybrid individual leaves were selected based on the convolutional neural network (CNN) Xception classification model. The classification accuracy of *Q. aliena* and *Q. dentata* reached 95.8%; The classification accuracy of *Q. aliena* with hybrid individuals, *Q. dentata* with hybrid individuals were 71.8% and 67.9% respectively; The accuracy of three classification of *Q. aliena*, *Q. dentata* and hybrid individuals was 44.5%. The results showed that *Q. aliena* and *Q. dentata* based on leaf scanned images had better classification effect, and hybrid individuals could not be distinguished from *Q. aliena* and *Q. dentata* based on leaf features.

(4) The covariant relationship between leaf shape (symmetric and asymmetric components) and leaf size were estimated by Two Blocks Partial Least Squares (2B-PLS). There was a significant allometric pattern for symmetrical components of *Q. aliena* and *Q. dentata*, but isometric in the asymmetric components.

(5) The relationship between leaf morphology (species level and symmetry level) and geographical and climatic factors was analysed by distance-based redundancy analysis (dbRDA) and 2B-PLS. The results showed that the leaf shape of *Q. aliena* and latitude change significantly correlated; August precipitation (Prec08) and mean diurnal range (Bio15) were interpreted the leaf shape.

(6) In order to test the existence of modularization and comparison of differences, we discussed the modularity hypothesis for leaf shape of *Q. aliena* and *Q. dentata* leaves. Based

on 13 leaf-shaped landmarks of the 471 individuals modular partition, we evaluate modular hypothesis: the leaf shape of *Q. aliena* (symmetrical components) and *Q. dentata* (symmetric and asymmetric components) were divided into leaf tip, lower part of the leaves and petiole three modules, but the asymmetric components of *Q. aliena* are divided into leaf and petiole two modules.

The comparative analysis of geometric morphometric analysis and deep learning for the identification of oak leaves showed that the above two methods have high classification accuracy. The results provide references for the identification of other complex leaf types of related species. The influence of geographical and climatic factors on leaf shape shows leaf shape variation adaptability to the environment. Exploring leaf modularity detects strong correlations within its morphological structure. This study will provide a deeper understanding of the ecology and evolution of oak species.

Key Words: *Quercus*, geometric morphometric analysis, deep learning, leaf morphological characteristics, modularity

目录

1 引言.....	1
1.1 物种分类及方法.....	1
1.1.1 物种分类的意义.....	1
1.1.2 基于叶片形状的分类研究.....	1
1.2 几何形态学.....	1
1.2.1 几何形态学的发展.....	1
1.2.2 几何形态学常用方法与分析.....	2
1.2.3 几何形态学的应用.....	5
1.3 深度学习研究现状.....	6
1.3.1 深度学习的发展.....	6
1.3.2 深度学习在植物分类中的应用.....	7
1.4 槲栎和槲树的研究现状.....	7
1.4.1 栎属简介.....	7
1.4.2 槲栎和槲树简介.....	8
1.4.3 槲栎和槲树叶形研究进展.....	9
1.5 本研究目的及意义.....	9
1.5.1 研究内容.....	9
1.5.2 研究目的和意义.....	9
1.5.3 叶片形态研究流程.....	10
2 材料与方法.....	11
2.1 样本采集与数据获取.....	11
2.1.1 样本采集与处理.....	11
2.2 遗传先验.....	14
2.2.1 引物筛选及 PCR 扩增.....	14
2.2.2 基于微卫星分子标记的先验分析.....	15
2.3 几何形态学分析.....	16
2.3.1 叶形数据获取.....	16
2.3.2 环境数据获取.....	17
2.3.3 广义普氏分析.....	17
2.3.4 薄版样条法.....	18
2.3.5 主成分分析.....	18

2.3.6 典型变量分析.....	18
2.3.7 判别分析.....	19
2.3.8 双偏最小二乘分析.....	19
2.3.9 冗余分析.....	19
2.3.10 标志点配置中的模块化.....	20
2.4 深度学习分类.....	21
2.4.1 数据集划分.....	21
2.4.2 实验步骤.....	21
2.4.3 评价指标.....	22
2.5 三种鉴别方法的综合比较.....	22
3 结果与分析.....	25
3.1 榲栌和榲树的遗传结构划分.....	25
3.2 榲栌和榲树的叶片形态特征.....	25
3.2.1 叶片形状差异分析.....	25
3.2.2 叶片的判别分析.....	27
3.2.3 叶片异速生长分析.....	29
3.2.4 同域和异域对叶片的影响.....	30
3.2.5 地理和气候对叶片的影响.....	32
3.2.6 叶片标志点配置中的模块化假设.....	34
3.3 榲栌榲树的深度学习分类.....	36
3.4 三种鉴别方法的综合比较结果.....	36
4 讨论.....	38
4.1 榲栌和榲树叶片的分类结果.....	38
4.2 榲栌与榲树的叶形态特征.....	39
4.3 榲栌与榲树的异速生长模式.....	40
4.4 榲栌与榲树的模块化分区.....	40
5 结论与展望.....	42
5.1 结论.....	42
5.2 展望.....	42
参考文献.....	44
个人简介.....	53
导师简介.....	54
成果目录清单.....	56
致谢.....	58

1 引言

1.1 物种分类及方法

1.1.1 物种分类的意义

“正确区分物种，正确归合物种”是物种分类的主要目的（陈世骧，1979）。生物界中没有完全相同的两个物种，每个物种都有属于自己的特征，而捕捉到这些主要特征并据此分类，是进行生态、进化等科学研究的基础（陈世骧，1983）。栎属植物形态变异多样，其分类标准曾引起包括达尔文在内的多位科学家的关注，达尔文就在其《物种起源》中写到：科学家没有使英国的栎类更清晰，反而使其更模糊（Darwin, 1872）。多年来，科学家也尝试使用包括形态学、孢粉和分子学等证据尝试理清栎类之间的物种界限，但是争议仍然存在（彭焱松，2007; Hipp *et al.*, 2020）。本论文从我国广泛分布的两种落叶栎类的叶片形态出发，通过叶片几何形态研究的方法解析其叶片形态标志点间的变异；同时以遗传分组结果作为先验，比较几何形态学方法和新近出现的深度学习方法对两种栎类鉴定的优缺点。同时，我们也探讨了气候、地理因子对两种近缘种叶片形态变异的影响，并尝试解析这些变异的异速生长和整体模块化模式。

1.1.2 基于叶片形状的分类研究

分类通常作为评价种质资源多样性的基础，由于表型性状非常直观有效，其常被应用于不同种植物的分类鉴定和核心种质的构建（李银霞，2006）。目前野外物种的识别仍然主要依赖于表型性状的判断，比如观察花、果、叶、枝干等形态，结合解剖特征及地理分布位置来进行一些基本的物种判断，并进行分类与命名（方伟，2012）。

叶子形状是一种宝贵的分类资源，叶形态研究也是植物分类学和系统学的核心研究内容。基于叶片的表型性状分类也在一些叶片编译特别复杂的物种类型判别中得到了很好的应用，比如栎类，例如 Kremer 等（2002）人通过 5 个叶形特征（叶柄长度、窦宽度，叶背有无毛，叶脉数和比率）将无梗花栎（*Quercus petraea* (Matt.) Liebl.）和夏栎（*Quercus robur* L.）进行了很好的判别（Kremer *et al.*, 2002）；Ponton 等（2004）基于叶片的绒毛密度和叶基部的耳廓角度两个性状对夏栎（*Q. robur*）和无梗花栎（*Q. petraea*）两物种进行判别，可以提供至少 99.5% 的正确分类结果。

1.2 几何形态学

1.2.1 几何形态学的发展

形态测量学 (Morphometric) 是描述生物体形态差异的分析方法, 主要分为传统形态学 (Traditional morphometric) 和几何形态学 (Geometric morphometric)。传统形态学 (多变量形态学), 是基于统计学分析得到形态上的变化规律 (形态差异), 研究者通常在样本上直接测量变量 (如长度、宽度、高度或角度等) 从而计算所得的间接变量 (比率及面积等)。几何形态学是一门基于图论的应用数学分支, 是一种利用标志点映射到数学坐标中采用多变量数据分析的方法获得其差异性 (Lele and Richtsmeier, 1991; Bookstein, 1996)。通过对其旋转和缩放等使其最大限度地重合, 在除去非形态因素 (位置和尺寸等) 的影响下, 基于线性代数矩阵运算获取网格图和线框图, 利用其相对扭曲程度来分析样本之间的形状差异。经过了 20 余年的发展, 几何形态学逐渐步入一个更加成熟的阶段 (Ashburner and Friston, 2000)。

1.2.2 几何形态学常用方法与分析

研究方法 几何形态学主要有标志点法 (landmarks) (Lele and Richtsmeier, 1991)、轮廓线法 (outline) (Rohlf and Bookstein, 1992)、半标志点法 (semi-landmarks) (Bookstein, 1991) 和有限元分析 (finite element analysis) (Cheverud and Richtsmeier, 1986) 等方法。标志点是生物体结构上具有一致性的位点, 且出现在所有被研究的样本形态上 (Lele and Richtsmeier, 1991), 用于描述观察对象的形态特征差异。当有些生物体结构无法找到能够重复标记的坐标点, 或者选择的标志点不能充分地反映被研究对象结构的总体形态变化, 例如贝壳和叶缘光滑的叶片等, 这就需要使用轮廓线法进行分析 (Rohlf and Bookstein, 1992)。在所测样本的结构边缘等距选取样点来表示样本的形态, 并将所得到的样点轮廓通过一定的数学函数将其数字化。半标志点法通常是被选作标志点的位置并不是均匀地分布在一个样本中, 为统计平滑的曲线和表面 (Bookstein, 1991) 并被应用于二维轮廓 (Bookstein, 1997) 和三维曲面 (Gunz, 2001) 的形态变化研究中。有限元分析过映射转换将一个有机体 (由标志点连接成的封闭的小区域) 映射到另一个有机体上, 并对相应区域进行比较 (Cheverud and Richtsmeier, 1986), 主要用于分析生物体功能结构的系统进化意义等 (Tseng and Stynder, 2011)。

标志点法 几何形态学中标志点法最为常用, 首先是对样本的形态进行数字化处理, 然后从形态数据中提取出大小和形状两个因素, 并比较计算不同样本间的大小和形状差异, 用统计学的方法分析类群内、类群间或不同形态单元间的变化规律或形态差异 (MacLeod, 2002; 葛德燕等, 2012)。标志点的选择原则要可以直观准确地表达信息, 并可以进行重复标记 (Zelditch *et al.*, 2004)。由于标本上非形态变异 (标志点的位置、取向和比例的变化) 会直接引起相关性的变化, 所以分析时必须去除非形态变异对分析的干扰, 目前常用的方法是叠印法 (superimposition)。叠印法包括 2 点配准、广义普氏

分析 (Generalized Procrustes Analysis, GPA) 和广义耐受适应 (Generalized Resistant Fit, GRF) 等。2 点配准是 Bookstein 提出的一种简单的叠印方法, 该方法基于最优化原则来叠加标点, 实现消除非形态变异的影响 (Bookstein, 1991)。GPA 是使用最小正方形估计平移和旋转参数将标志点进行叠印 (Rohlf and Slice, 1990)。GRF 是将叠印参数 (如旋转角度和比例) 作为标志点子集的中值进行叠印, 当多形态变量仅因为很少的标点而受到限制时, 可以用 GRF (Zelditch *et al.*, 2004)。GPA 主要针对样本的整体形态进行了对比, 基于坐标点找出各个样本之间的最小差距, 通过叠印之后每一个构象都对应形态空间中的一个点 (Rohlf and Bookstein, 1990; Rohlf and Slice, 1990)。相比之下, GPA 可以反复的允许样本叠印, 而 GRF 不能进行进一步的统计分析 (白明等, 2007)。GPA 在对研究植物个体间叶片差异中得到了很好的应用, 比如 Viscosi and Cardini (2011) 基于几何形态学对两个 *Q. petraea* 种群的叶片形状进行分析, 发现 GPA 在测量叶片的层次结构差异和揭示非常小的变化方面非常有效。

对称性 形态学中的对称包括匹配对称 (matching symmetry) 和对象对称 (object symmetry): 匹配对称是对称的成对结构, 分别位于身体的左右两侧, 各自独立, 且互为镜像 (Klingenberg *et al.*, 2002), 例如昆虫的翅膀, 人的两只手 (图 1.1A); 对象对称是对象本身的对称, 这意味着一些标志点可能位于中轴或中轴平面上, 而另一些则相应成对地存在于左右两侧 (Klingenberg *et al.*, 2002), 例如动物头骨和植物叶片等 (图 1.1B)。因此对象对称的样本在 GPA 过程中, 分为对称组分 (symmetric components) 和非对称组分 (asymmetric components), 对称组分被视为全部标志点沿中轴线和其镜像坐标形成的平均形态, 而非对称组分则由原始标志点的叠印构成 (Mardia *et al.*, 2005)。

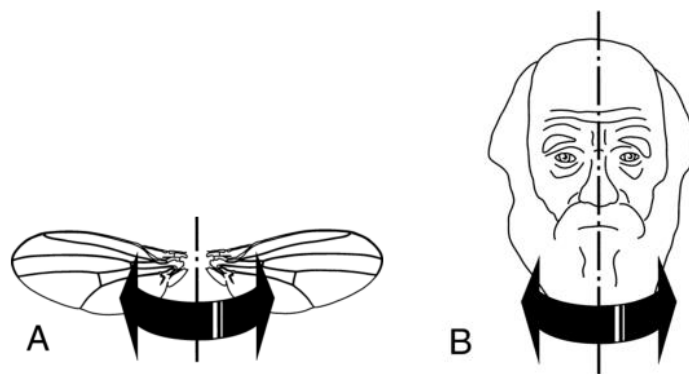


图 1.1 匹配对称 (A) 和对象对称 (B) 示意图 (引自 Klingenberg *et al.*, 2002)

Figure 1.1 Matching symmetry (A) and object symmetry (B) schematics. (cited from Klingenberg *et al.*, 2002)

异速生长 几何形态学同时也考虑了形态空间的结构的影响, 其中包括异速生长 (allometric) 和等距生长 (isometric)。异速生长在形态学中代表被观察对象大小和形

状的相关性 (Huxley 1924; 1932), 反应了植物有机体对异质环境的适应性 (Klingenberg, 1998); 等距生长代表尺寸和形状的变化互不影响 (Darroch and Mosimann, 1985)。在几何形态学中, 使用 Procrustes 形式空间或构象空间 (后者也称为尺寸与形状空间) 来实现, 虽然这些空间在整体结构上有很大的不同, 但它们在局部几何上有一定的相似性 (Klingenberg, 2016)。

模块化 大多数生物由可识别的部分组成, 根据它们的发育起源、结构和功能很容易被识别为独立的单位, 然而这种独立性并非完全, 因为生物体的各个部分相互协调, 并最终集合到整个生物体中 (Klingenberg, 2008)。集成 (integration) 和模块化 (modularity) 就体现了协调和相对独立的这种关系: 集成是指产生所研究的表型结构中生物相互作用而产生的性状间的凝聚力 (Olson and Miller, 1958; Pigliucci and Preston, 2004); 模块化是指系统中的相对连接程度, 模块是一个在内部紧密集成, 但相对独立于其他此类模块的单元 (Klingenberg, 2008)。Wagner 等人 (2007) 认为所有的生物都被认为是高度集成且具有凝聚力的系统, 并与环境相互作用。图 1.2 显示了两个模块 (虚线), 它们的各个部分通过许多交互 (箭头) 在内部集成, 但它们彼此相对独立, 因为模块之间只有很少的交互 (Klingenberg, 2009): 图 1.2A 表示粗体线与模块之间的边界重合; 图 1.2B 表示粗体线将整体结构划分为的两个子集跨越了两个模块。

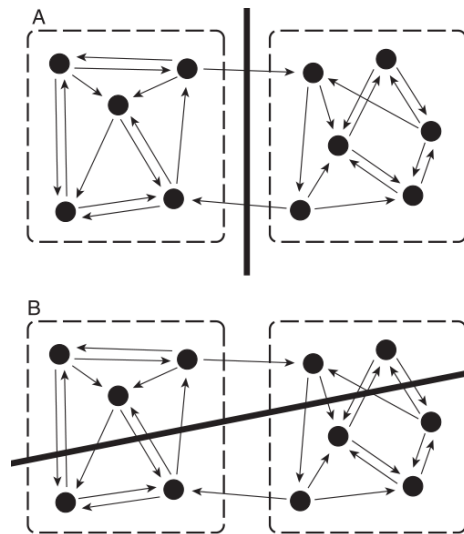


图 1.2 比较结构的不同分区来划分模块

A: 细分与模块边界重合; B: 细分与模块边界不一致, 因此跨越两个模块 (引自 Klingenberg, 2009)

Figure 1.2 Delimiting modules by comparing different partitions of the structure.

A: The subdivision coincides with the boundary between modules; B: The subdivision does not coincide with the modular boundary and therefore goes across both modules. (cited from Klingenberg, 2009)

生物学模块化的研究也包括基因调控网络中的分子相互作用（molecular interactions in networks of gene regulation）比如代谢网络（metabolic networks）、生态相互作用网络（networks of ecological interactions）和形态集成和模块化（morphometric integration and modularity）等（e.g. Dassow and Meir, 2000; Ravasz et al., 2002; Krause et al., 2003; Olesen et al., 2007; Klingenberg, 2009）。因此，根据特定研究的背景，建立模块的相互作用依赖于不同的机制（Wolsan *et al.*, 2019）。在交互网络信息很容易获得的情况下，模块化分析可以直接使用各部分之间的连接进行识别（例如，Girvan and Newman, 2002; Guimerà and Nunes Amaral, 2005; Rosvall and Bergstrom, 2007）。但在相互作用不容易观察到时，则需要从被测变量之间的共变模式推断出来，例如基因调控模块可能需从不同实验环境中共享的基因表达模式中推断出来（Somogyi *et al.*, 2004）。同样，对于形态学数据，发育或进化的相互作用必须从实验或比较分析形态学性状的共变中推断出来，需要提出假设并验证，即从性状间的共变模式来推断交互作用（Klingenberg, 2008; Klingenberg, 2009）。

标志点结构中的模块假设 根据形态学的研究背景，形态模块的划分原则上空间相邻，并且要求成对标志点应该包含在一个子集对中，即左右两边的对称标志点都属于一个子集或都排除在外。通过比较假设的标志点子集之间的协变强度，将假设模块之间的共变程度与整个结构的可选分区进行比较，如果划分对应于模块之间的真实边界，则子集之间的协变应该是弱的，并且与假设相对应的标志点子集之间的协变应该低于标志点不同细分的协变（Klingenberg, 2009）。*RV* 系数（*RV* coefficient）描述了两个定量变量矩阵之间的相关性程度，是皮尔森平方相关系数的正则化，*RV* 的值在 0 到 1 之间，接近于 1 说明两个矩阵越相似，接近于 0 说明两个矩阵差异大（Escoufier, 1973; Robert and Escoufier, 1976）。本研究中以 *RV* 系数作为两组标志点子集变量之间的关联强度，通过比较子集之间的 *RV* 系数来评估模块间的边界（Klingenberg, 2009）。

1.2.3 几何形态学的应用

几何形态学被广泛用于分析各种研究背景下的生物形状的变异，系统学、进化生物学、人类学、生物医学，发育生物学和遗传学等学科研究中均涉及到该方法（Klingenberg, 2010; Zelditch *et al.*, 2012）。在研究花朵形状（对称和非对称）的应用中，Savriama 等（2012）使用几何形态学方法分析了十字花科丹参（*Erysimum mediohispanicum*）野生种群的花形变化，将形状变化划分为对称变化和非对称变化，结果表明，对称组分占总方差的绝大部分，并提出了形态测量方法有望在进化和发育的背景下进一步分析花的对称性和非对称性。Klingenberg（2018）利用几何形态学方法对同质园中种植的蝴蝶花的花形状进行了分析，结果表明花不同部位的形状，尤其是非对称组分在不同方向上有显著

差异。几何形态学方法在栎属植物的叶片的研究中也得到了很好的应用，例如 Viscosi 等（2012）利用几何形态学方法强调了叶片大小和形状共变在白栎树进化过程中的作用，对意大利中部混交林三种同域栎树匈牙利栎（*Quercus frainetto*）、无梗花栎（*Q. petraea*）和毛白栎（*Quercus pubescens*）的叶形和叶大小进行了分析，区分鉴定叶形性状，并分析比较了杂交个体的表型与亲本种的异速生长，结果表明在纯栎树中，叶片异速生长是叶片变异与环境变异相关的重要组成部分，而在杂交发生时，叶片异速生长受叶片形状和大小性状的遗传力的影响。Liu（2018）基于 SSR 分子标记先验，利用几何形态学分析对 6 个同域分布的槲栎和槲树种群的叶片进行判别，结果表明两近缘种叶形态存在差异。李月娟（2020）基于 SSR 分子标记先验，并通过几何形态学的方法对川滇高山栎和刺叶高山栎的叶片进行形态学研究，同时也探索了叶片形态与地理、遗传和气候因素之间的关系。结果表明叶片的对称组分和叶片大小主要与遗传因素有关，而非对称组分只受地理因素的影响。

几何形态学为形态集成的研究开辟了实现评估模块化假设的方法（Bookstein, 1996; Dryden and Mardia, 1998; Klingenberg *et al.*, 2009）。模块的这些属性非常普遍，识别模块化交互作用的使用也非常多样。例如，模块化应用于工程和管理（Garud *et al.*, 2002）。在昆虫翅膀形态的研究系统中，集成化和模块化已经得到了广泛的研究（Klingenberg and Zaklan, 2000; Klingenberg *et al.*, 2001），形状的共变表明果蝇的翅膀为一个完整的发育模块，而蜜蜂的前翅和后翅是两个独立的发育模块（Klingenberg and Zaklan 2000; Klingenberg *et al.*, 2001）。同时啮齿动物的下颚（Atchley and Hall, 1991; Klingenberg *et al.*, 2003, 2004; Zelditch *et al.*, 2008），包括人类在内的各种哺乳动物的头骨（Cheverud, 1982; Goswami, 2006a, 2006b）以及人面部组织（Quinto *et al.*, 2018）都证实了模块化的重要性。几何形态学可以非常详细地表征形态变化的模式，并保持所研究结构的解剖学背景，例如在拟南芥中，花和叶的性状往往形成独立的遗传模块（Juenger *et al.*, 2005）。一个生物体是否受到进化的限制，或有适应的潜力，是基于这些性状的表型整合或模块化，Klápště（2020）在叶形表型性状模块化研究中，表明叶片性状集合中存在多个模块，表现出了最大的物种区分能力，并且可能导致了生长在不同小气候下的两种同域栎树无梗花栎（*Q. petraea*）和夏栎（*Q. robur*）之间的生态偏好。

1.3 深度学习研究现状

1.3.1 深度学习的发展

深度学习（Deep Learning）指基于样本数据通过一定的训练方法得到包含多个层级的深度网络结构的机器学习过程（Hinton, 2006）。深度学习具有多层非线性映射的深层

结构，自提出以来，广泛应用于视觉、音频和自然语言处理等领域，因其高效便捷已经成为互联网大数据和人工智能的一个热潮（余凯等，2013）。卷积神经网络（convolutional neural network, CNN）是第一个真正意义上具有多层结构的深度学习模型，提高了数据量过于庞大时的处理效率，实现了特征自动提取的需求（LeCun *et al.* 2014）。CNN 基本结构为：输入层，卷积层，池化层，全连接层和输出层。其中卷积层就是特征提取层，所包含的卷积核则是神经网络训练的核心。并且不同的卷积核能提取不同的输入数据相应的特征，提取到的特征随卷积核的数量增多而增多。在 ILSVRC（imagenet large scale visual recognition challenge）比赛上，研究者提出了一些新的 CNN 模型，其 2014 年的冠军 GoogLeNet 采用了 Inception 结构（在同一层中使用不同大小的卷积核），用以提高分类效果。并且在 Inception 之后又增加了一些改进，Xception 则为 Inception 的极限版本，在 ImageNet 数据库上测试的分类 top-5 正确率高达 94.5%，是分类模型中比较出色的选择。

1.3.2 深度学习在植物分类中的应用

深度学习在图像分类上取得了重要进展并在不断更新改进，包括特征提取算法和图像分类算法，叶在植物分类领域得到了广泛的运用，例如：Tan 等人（2018）建立了植物物种自动识别系统 D-Leaf，选取了 43 种热带树木，每种 30 个样品，共收集了 1,290 张叶片图像，与 AlexNet（93.26%）和微调 AlexNet（95.54%）模型相比，D-Leaf 模型实现了 94.88% 的可比测试精度，帮助植物学家和外行人快速识别植物。CNN 模型的表现优于传统的形态测量（66.55%），从神经网络中提取的特征被发现与神经网络分类器可以进行很好地拟合。Liu 等人（2019）将深度学习应用于大花菊花的识别，提出了一种大型菊花品种识别的新方法，通过上传一幅菊花图像到系统中，利用系统中相应的品种图像，可以快速获得预测出前 5 位的品种信息。Dunker 等人（2020）结合深度学习的多光谱成像流式细胞术，分析 35 种植物的花粉，共计 426,876 幅图像，实现了快速准确地进行物种鉴定。左羽等（2020）以植物图像为研究对象，将经典卷积神经网络 VGG16 与全卷积网络（fully convolutional network, FCN）相结合构造一个新的植物图像分类模型，以 21,500 张（43 类）植物图像作为模型的训练数据集，分类结果的准确率达到 97.23%。

1.4 榲栌和榲树的 research 现状

1.4.1 榲属简介

壳斗科植物是我国亚热带与温带常绿阔叶林和落叶阔叶林的重要树种（邱丽旻等，2017），分布西起喜马拉雅山脉东至台湾岛，北达中国东北地区，南到海南，在东亚木

本植被中占有突出地位 (Yang *et al.*, 2020)。栎属物种不仅在生态学中扮演重要的角色, 它们一直也被作为研究近缘种分类的模式物种, 其分类从 19 世纪开始至今一直是一个挑战 (Hipp *et al.*, 2020)。

在更新的分类系统中栎属 (genus *Quercus*) 被分为栎亚属 (subgenera *Quercus*) 和麻栎亚属 (subgenera *Cerris*), 分别对应于新大陆分支和旧大陆分支, 其中栎亚属包括以下 5 个组: 白栎组 (sect. *Quercus*)、红栎组 (sect. *Lobata*)、中间栎组 (sect. *Protobalanus*)、活栎组 (sect. *Virentes*) 和本都栎组 (sect. *Ponticae*) (Hipp *et al.*, 2020)。据《中国植物志》记录, 栎亚属仅有一个组 (白栎组) 分布在我国, 约有 15 个种。《Flora of China》将其合并为 7 个种, 分别为槲栎 (*Quercus aliena* Bl.)、槲树 (*Quercus dentata* Thunb.)、白栎 (*Quercus fabri* Hance.)、大叶栎 (*Quercus griffithii* Hook.)、蒙古栎 (*Quercus mongolica* Fisch.)、枹栎 (*Quercus serrata* Thunb.) 和云南波罗栎 (*Quercus yunnanensis* Franch.)。

栎属植物种间基因交流频繁, 近缘种的叶片表型特征极为相似, 在组间和种间关系的处理上一直存在很大的分歧 (Viscosi, 2012; Liu *et al.*, 2018; 李月娟, 2020)。栎属近缘种间的形态特征表现多有重复, 种间界限并不清晰, 进而导致在物种鉴定上及种间的亲缘关系研究中存在较大的困难 (彭焱松, 2007), 对栎属近缘种的表型性状深入研究, 具有重要参考价值。

1.4.2 槲栎和槲树简介

槲栎 (*Q. aliena*) 和槲树 (*Q. dentata*) 隶属于白栎组 (sect. *Quercus*), 据中国植物志记载: 槲栎 (*Q. aliena*), 落叶乔木, 高达 30m, 树皮暗灰色, 深纵裂, 小枝粗灰褐色, 近无毛, 具圆形淡褐色皮孔; 芽卵形, 芽鳞具缘毛, 叶片长椭圆状倒卵形至倒卵形, 长 10-20 (-30) cm, 宽 5-14 (-16) cm, 先端短钝尖或短渐尖, 基部宽楔形或近圆形, 叶缘具波状钝齿, 叶背被灰棕色细绒毛或近无毛, 侧脉每边 10-15 条, 叶面中脉侧脉不凹陷; 叶柄长 1-1.3cm, 无毛。

槲树 (*Q. dentata*), 落叶乔木, 高达 25m, 树皮暗灰褐色, 深纵裂, 小枝粗壮, 有沟槽, 密被灰黄色星状绒毛。芽宽卵形, 密被黄褐色绒毛, 叶片倒卵形或长倒卵形, 长 10-30cm, 宽 6-20cm, 顶端短钝尖, 叶面深绿色, 基部耳形, 叶缘波状裂片或粗锯齿, 幼时被毛, 后渐脱落, 叶背面密被灰褐色星状绒毛, 侧脉每边 4-10 条; 叶柄长 2-5mm, 密被棕色绒毛。

槲栎花期 3-5 月, 槲树花期 4-5 月, 果期都为 9-10 月。槲栎槲树在我国主要分布于陕西、安徽、浙江、山东、江苏、河南、湖北、湖南、四川、贵州、云南等, 在朝鲜、日本也有分布; 除以上外槲栎还分布于江西、广东、广西; 槲树还分布于黑龙江、吉林、辽宁、河北、山西、甘肃、台湾。槲栎和槲树均生于海拔 80-2,000m 的向阳山坡, 常与

其他树种组成混交林或成小片纯林。

1.4.3 槲栎和槲树叶形研究进展

白栎组植物叶片倒卵形至狭椭圆形，基部楔形到圆形，叶片边缘波形，由于组间形态接近，种间界限不明确，一些叶片中间型个体或者白栎组中的杂交个体的叶片在形态学上也与两种相似。槲栎和槲树分布范围广，在生态领域尤其是分类学中扮演着极其重要的角色，目前对两物种的研究主要集中在落叶阔叶林的物种多样性(谢晋阳, 1994)，种群结构与更新的研究(刘峰等, 2000)，盐胁迫条件对其生长的影响(韩亚琦等, 2007)，并且基于核基因构建的系统发育树结果表明二者亲缘关系较近 (Hubert *et al.*, 2014)。但其相关的叶片形态学的研究还比较少。Liu 等 (2018) 对 6 个同域分布的槲栎和槲树种群以分子标记先验的结果进行几何形态学分析，结果表明槲栎和槲树的叶片形态差异在同域显著，杂交个体分散且更接近槲栎个体的叶片形态。Lyu 等 (2018) 研究了小尺度内同域分布的槲栎 (*Q. aliena*)，槲树 (*Q. dentata*) 和栓皮栎 (*Q. variabilis*) 的基因流情况，基于微卫星分子标记和叶绿体 DNA 片段，发现三个种之间存在明显的种间界限。

1.5 本研究目的及意义

1.5.1 研究内容

本研究以槲栎和槲树为研究对象，参考中国植物志等资料在全国大范围进行采样，以基于 SSR 的分子标记为先验，通过几何形态学和深度学习方法进行物种鉴别分类，并综合对比三种方法的投入产出比。并基于几何形态学探究叶片形状是否随着叶片大小的共变关系，以及与环境因子之间的相关性，并探索叶片形态（对称组分、非对称组分）与地理和气候因素之间的关系，同时对两物种叶片模块化进行假设及验证。

1.5.2 研究目的和意义

本研究的目的是：

- (1) 槲栎和槲树基于不同方法下的分类结果，并比较异同；
- (2) 验证槲栎和槲树叶片是否具有模块化，基于标志点进行模块分区的假设；
- (3) 挖掘环境因子对栎类叶片形态特征的影响。

该研究不仅对同属槲栎和槲树种间叶片形态特征进行了相关研究，而且分析了这两个物种的叶片几何形态学特征，也为近缘种植物的分类提供参考；同时为研究形态特征的适应性和叶片模块化分区奠定基础，具有重要的理论意义。

1.5.3 叶片形态研究流程

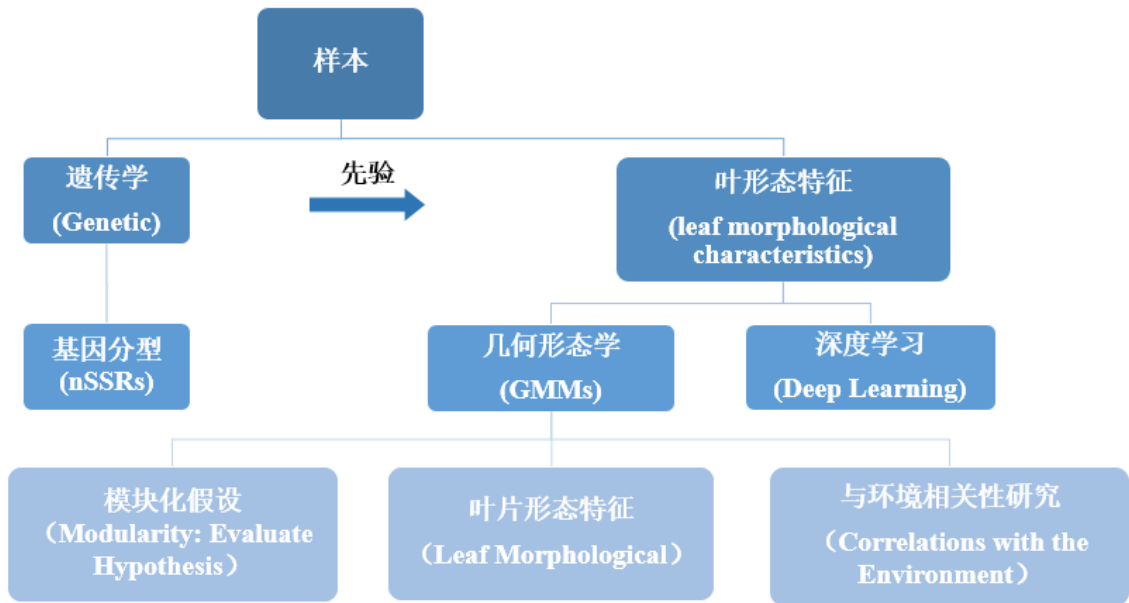


图 1.3 栎类叶片形态研究流程

Figure 1.3 The chart of oak leaf morphological analyses.

备注：

基因分型（nSSRs）：利用微卫星序列重复，并以混合系数 Q 以 0.9 为界，将个体的划分结果作为先验。

几何形态学（GMMs）：选取落叶栎常用的 13 个标志点，数字化叶形态特征，分离对称和非对称组分进行多变量分析。

深度学习（DL）：基于叶片扫描图像，以 Xception 模型对遗传先验的个体分类。

2 材料与方法

2.1 样本采集与数据获取

2.1.1 样本采集与处理

经过查阅中国植物志及榲栌榲树相关的文献，同时在中国数字标本馆 (<http://www.cvh.ac.cn>)、中国植物图像库 (<http://ppbc.iplant.cn>) 查找近二十年的标本信息完整的采样点，确定两近缘物种在中国各省的分布情况。于 2018~2019 年的 6~8 月份在中国 15 个省份采集了榲栌和榲树共计 54 个种群，其中 8 个采样点为两物种同域分布。每个种群通过 GPS 记录仪 (GRAMIN 621sc) 详细地记录采样点的经纬度信息，采集过程中每个种群至少相隔 30km，种群内的个体至少相隔 10m。由于不同采样地点的光照条件不同以及小环境的地形差异，可能会导致部分叶片有趋光性，为避免光照对叶形的影响，我们在每个个体上采集 4~7 片不同方向的具有完整叶形的叶片和一片干净的嫩叶或小枝用于提取 DNA。对每个个体按照种群及物种名称对其进行编号。最终采集 710 个个体，3,657 张叶片。采集叶片以硅胶干燥，平整地储存于信封中，并间隔一定时间更换硅胶，避免叶片潮湿发霉。用于叶片形态分析的样本需铺平并压置于标本夹中保存，用于后续叶形数据获取。采集点的分布和种群名称缩写见图 2.1，每个种群详见表 2.1。

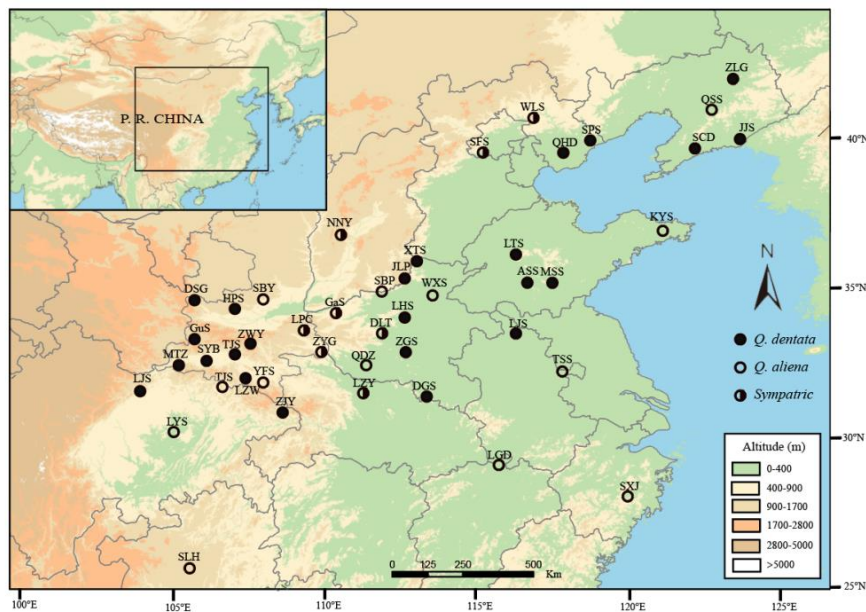


图 2.1 榲栌和榲树采样种群分布图

Figure 2.1 Sampling populations distribution area of *Q. aliena* and *Q. dentata*

基于几何形态和深度学习的栎类叶片形态研究

表 2.1 槲栎和槲树种群采样点信息表

Table 2.1 Sample information of *Q. aliena* and *Q. dentata*

采样点 Situation	降水季节性变异 系数 bio15	平均日较 差 bio02	最湿季平均 温度 bio08	8 月份降水 prec08	个体数 Numbers
DLT_A	73.86	11.38	24.35	117.00	9
NNY_A	100.44	11.03	16.42	120.00	14
GaS_A	76.89	9.27	18.67	107.00	10
LPC_A	73.01	10.38	19.60	110.00	10
SFS_A	131.61	10.97	22.53	149.00	22
WLS_A	124.84	12.18	21.38	140.00	20
ZYG_A	70.90	9.95	23.52	108.00	14
LZY_A	62.48	9.18	25.58	130.00	14
HPS_A	74.74	9.51	17.78	110.00	10
KYS_A	95.21	8.88	22.92	163.00	13
LGD_A	53.67	8.27	20.95	131.00	8
LYS_A	78.49	6.81	25.30	177.00	13
QDZ_A	68.37	10.46	23.77	114.00	14
QSS_A	108.39	10.94	22.23	181.00	10
SBP_A	89.44	10.20	22.08	122.00	9
SBY_A	77.53	11.32	20.85	107.00	10
SLH_A	70.26	7.38	22.08	131.00	9
SXJ_A	48.38	8.27	19.32	180.00	10
TJS_A	73.93	10.02	13.47	120.00	9
TSS_A	75.35	9.23	26.15	123.00	12
WXS_A	99.40	9.51	20.15	141.00	11
YFS_A	66.25	9.19	17.13	145.00	12
DLT_D	73.86	11.38	24.35	117.00	15
NNY_D	100.44	11.03	16.42	120.00	9
GaS_D	76.89	9.27	18.67	107.00	14
LPC_D	73.01	10.38	19.60	110.00	11
SFS_D	131.61	10.97	22.53	149.00	15
WLS_D	124.84	12.18	21.38	140.00	17
ZYG_D	70.90	9.95	23.52	108.00	10
LZY_D	62.48	9.18	25.58	130.00	11

材料与amp;方法

JLP_D	93.36	12.45	18.61	137.00	20
KYS_D	95.22	8.85	22.87	164.00	21
LHS_D	81.33	10.33	22.45	121.00	10
LJS_D	89.32	10.34	26.37	133.00	10
LJJ_D	91.55	8.09	22.23	159.00	10
LTS_D	113.31	10.74	25.44	159.00	14
LZW_D	70.02	9.52	18.79	111.00	5
MSS_D	108.45	10.13	23.92	163.00	18
MTZ_D	77.38	9.45	18.21	119.00	20
QDZ_D	68.57	10.48	23.77	114.00	21
QHD_D	126.04	9.68	21.88	174.00	11
SCD_D	101.80	7.22	22.00	176.00	18
SPS_D	125.60	10.30	22.75	172.00	20
SYB_D	75.59	9.34	17.40	118.00	20
XTS_D	104.54	10.52	21.87	141.00	10
ZGS_D	70.93	10.58	24.74	135.00	10
ZJY_D	62.85	8.40	20.02	160.00	11
ZLG_D	103.69	12.09	21.73	171.00	10
ZWY_D	71.29	9.53	17.04	112.00	19

2.2 遗传先验

2.2.1 引物筛选及 PCR 扩增

本论文中的遗传先验部分工作来自实验室的未公开发表数据。我们基于 DNA 试剂盒 (Tiangen 植物基因组) 对所有采样的个体提取基因组 DNA, 后在浓度为 1% 的琼脂糖凝胶中进行电泳检测, 以胶图的结果来检测 DNA 质量, 其成相清晰不拖尾。初步检测后, 共选用 12 对引物 (表 2.2) 对提取的 DNA 样本进行 PCR 扩增实验。其中, 8 对引物 (GOT021, FIR026, POR017, FIR015, WAG068, PIE271, GOT011, WAG066) 选自 Durand 等 (2010), 剩余的 4 对引物是选自 Ueno 等 (2008): QmC00716, QmC00932, DN950446, QmC02052。选用 12 对引物对榲栌榲树采集的所有样品进行基因分型, 并用于后续的几何形态学分析和深度学习分类的先验结果。

选取的 PCR (15 μ l) 的扩增体系包括: 个体 DNA 1 μ l、Mix 7.5 μ l、反向引物 -R 0.24 μ l、5' 末端带有 M13 的正向引物-F 0.06 μ l、带有荧光标记的 M13 引物 0.24 μ l 以及 ddH₂O 5.96 μ l。PCR 扩增后通过电泳检测其扩增条带是否清晰, 若扩增条带呈相清晰不托尾, 则进行毛细管电泳检测。然后数据导入 GeneMarker 软件读取峰图, 从而获得 SSR 数据 (Hulce *et al.*, 2011)。

表 2.2 12 对 SSR 引物详细信息

Table 2.2 Information on primers in 12 nuclear simple sequence repeats (nSSRs) loci

引物 SSR	温度 T _m (°C)	基序 Motif	引物序列 (5'-3') Primer Sequence (5'-3')	等位基 因长度 Allele size rang (bp)	多态性 Polym orphis m	参考文 献 Reference
GOT021	56	AT	AGAAAGTTCCAGGGAAAGCA CTTCGTCCCCAGTTGAATGT	111-128	Yes	Durand <i>et al.</i> 2010
FIR026	56	TC	CTTCATGCACCAATTCCTCA GGCCATGTATGTGTGCAAAA	208-217	Yes	Durand <i>et al.</i> 2010
QmC007 16	56	TC	AAGAGAACCCATTCCATCCCTGA GTTTCCCGAACAGTGGTTTCTTGA	261-287	Yes	Ueno <i>et al.</i> 2010
POR017	54	CT	CCCATATCCCTCTACGAAAGAA CTGGAGATGACATAGTGTCTCAAA	140-169	Yes	Durand <i>et al.</i> 2010
FIR015	56	AC	ACCCTAAAACCCCAATCACC	128-138	Yes	Durand

			CGGATCTTCGGCTATTCTTG				<i>et al.</i> 2010
QmC009	54	TC	AGGCTCAAAAACAAAACCAAACCG GTTTCCCCTTTCCCATAATCAAAC CCT	247-260	Yes		Ueno <i>et al.</i> 2010
DN95044	56	AG	TCTCTTTCTCCGTCCATTATCGC GTTTCTCCACAGACCCCATTTC TCTGCAACAAAACCAAACAC	155-185	Yes		Ueno <i>et al.</i> 2010 Durand
WAG068	56	AG	CGGAGGAGAGAGTCAGCAAC	165-195	Yes		<i>et al.</i> 2010 Durand
PIE271	56	TC	CACACTCACCAACCCTACCC GTGCGGTTGTAGACGGAGAT	197-247	Yes		<i>et al.</i> 2010
QmC020	56	AG	CACACCCAGATCCACAAAACCTCC GTTTGCCTCTACGGTCTCCCTCTT CCCCACCGTCTACTCTCAA	250-300	Yes		Ueno <i>et al.</i> 2008 Durand
GOT011	56	TC	GCGTTCACCACGTCCATAAT	197-255	Yes		<i>et al.</i> 2010 Durand
WAG066	56	AG	AACCTGTTTGGCTTCGTGTG AACAAAAGATTGGGAGGTGC	128-244	Yes		<i>et al.</i> 2010

2.2.2 基于微卫星分子标记的先验分析

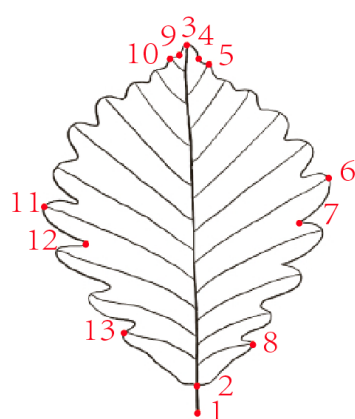
将 GeneMarker 读取的 SSR 数据首先通过 Flexbin (Excel 的宏) 进行归一化处理去除等位基因在读取过程中的片段长度的误差 (Amos *et al.*, 2007)。用英文符号“?”表示归一化数据中的缺失项, 另外在数据的第一行需标明种群数 (“npops =”) 和位点数 (“nloci =”), 以 “.txt” 格式保存。最后通过 CONVERT 软件 (Glaubitz, 2004) 将归一化的数据转化成后续分析所需要的格式。

使用 STRUCTURE V2.3 软件 (Pritchard *et al.*, 2000) 基于贝叶斯聚类分析对所有的榲栌榲树样品进行遗传结构分析。根据榲栌榲树不同个体间等位基因频率的差异, 检测其遗传结构和每个个体基于 nSSRs 被分配到同组的概率。在其运行过程中, 设置 K 值为 1-10, 并且每个 K 重复运行 20 次, Markov Chain Monte Carlo cycles (MCMC) 设置为 200000, burn-in cycles 设定为 100000。然后通过计算 ΔK 和 $\text{LnP}(K)$ 来确定最优的 K 值, 从而估计最可能的集群数 (Evanno *et al.*, 2005; Janes *et al.*, 2017)。最后通过在线软件 Structure Harvester (Earl and Vonholdt, 2012) 将所有结果下载压缩为压缩包, 并通过 Distruct 1.1 软件 (Rosenberg, 2004) 得到其按 K 分类柱状图。另外我们可以通过运行 STRUCTURE 结果得到混合系数 (Q) 值, 来统计样本中基因型纯合或者定义为杂交个体的数量, Q 值接近 0 或 1 则视为某一类群的纯合个体, Q 值接近我们假

定的界限值则被视为是假定的杂交个体。本研究选择 $Q = 0.9$ 作为分配榲栌榲树到不同物种和假定杂交个体的标准，其中 $Q \geq 0.9$ 或 $Q \leq 0.1$ 为纯合个体， $0.1 < Q < 0.9$ 为假定杂交个体（Vähä and Primmer, 2006）。

2.3 几何形态学分析

2.3.1 叶形数据获取



标志点	描述
Landmark	Description
LM1	叶柄起始点
LM2	叶片与叶柄的连接点
LM3	叶片的顶端
LM4	叶片右侧从顶端起第一个凹陷点
LM5	叶片右侧从顶端起第一片裂叶端点
LM6	叶片右侧离中脉最宽裂叶的端点
LM7	紧接标志点 6 下边的凹陷点
LM8	叶片右侧从叶柄起第一片裂叶端点
LM9	叶片左侧从顶端起第一个凹陷点
LM10	叶片左侧从顶端起第一片裂叶端点
LM11	叶片左侧离中脉最宽裂叶的端点
LM12	紧接标志点 11 下边的凹陷点
LM13	叶片左侧从叶柄起第一片裂叶端点

图 2.2 标志点位置

Figure 2.2 Landmark configuration and descriptions of the landmarks

将采集的叶片用高分辨率扫描仪（CanoScan 5600 F）对其脉络清晰的一面进行扫描，并保存为水平和垂直分辨率为 600dpi 的“.jpg”格式。通过 Image J 软件（Abràmoff *et al.*, 2004）对叶片扫描图像按照图 2.2 的标志点（landmarks）顺序进行标定，标志点构成了榲栌榲树叶形的几何轮廓。标志点的选取参考了 Viscosi（2015）在落叶栎 *Q. frainetto*, *Q. petraea* 和 *Q. pubescens* 中实行的标准，并且本课题组已经在部分同域分布的榲栌榲树种群中进行了应用，效果良好（Liu *et al.*, 2018）。这些标志点由每片叶子的 13 对笛卡尔坐标 (x, y) 转化为数字格式，获取所有叶片的标志点原始坐标数据后，将其整理为“.txt”文本格式，保存为每张叶片的 13（LM1-LM13）个点坐标值。其中 LM1-3 这三个点没有对称点，其余 LM4-8 与 LM9-13 互为对称点，其左右分别一一对应。并对其按照格式命名，每片叶子具有相应的 ID，其中包括采样点名称、物种名、个体序号及叶片序号，也可增加更多的分类器信息，比如是否同域生长或者是否为变种等，输入数据格式如下：

```

“LM=13
X0 Y0
X1 Y1

```

... ..

X12 Y12

ID = DLT_D_T_T01_L01”

其中, LM = 13 代表每一组数据识别 13 个标志点, 共 13 个标志点 (landmark, LM); X0 Y0—X12 Y12 表示 13 个点的坐标数据; 示例中 ID = DLT_D_T_T01_L01 为每片叶子的 ID 号, 前 3 个字符 DLT 为所采种群样地名称 (大乐坛), 第 5 字符 D 表示所属物种 (本实验中的物种是通过 SSR 分子标记的研究结果来进行判定, $Q \geq 0.9$ 或 $Q \leq 0.1$ 为纯合个体, $0.1 < Q < 0.9$ 为假定杂交个体), 第 7 字符 T 表示为同域种群, 第 9-11 字符 T01 表示采集点编号 01 的一棵树, 第 13-15 字符 L01 表示为这棵树的 01 号叶片, 即该叶片为 DLT 种群中 D 物种的 01 号树的 01 号叶片, 并且该采样点的榭栎榭树生长在同一生境下。

2.3.2 环境数据获取

我们对榭栎榭树的叶形与地理和气候因子进行相关性研究。在 WorldClim 数据库 (<http://www.worldclim.org>) 中获取了近 30 年 (1970-2000) 的气候数据, 包括 19 个生物气候变量和全年 12 个月的降水、太阳辐射、平均温度、最高温度、最低温度、风速和水蒸汽压力等气候变量共计 103 个气候因子。使用 ArcMAP (Environmental Systems Research Institute, 2014) 将榭栎榭树采样点的 GPS 坐标 (经度和纬度) 映射到 30 arc-second 的 ESRI® 气候数据网格中, 提取出采样点的气候数据。通过 R 语言的“usdm”函数包 (Uncertainty Analysis for Species Distribution Models) 中的方差膨胀因子分析 (Variance Inflation Factor, VIF), 去除部分相关性较强的因子, 最终保留了 4 个气候因子 (VIF < 0.7): 降水季节性变异系数 (bio15)、平均日较差 (bio02)、最湿季平均温度 (bio08) 和 8 月份降水 (prec08) 用于后续的偏最小二乘分析和冗余分析。种群的主要气候数据记录在表 2.1。

2.3.3 广义普氏分析

为了满足叶形的坐标数据进行后续多变量分析, 需要对坐标数据预处理。将整理好的叶形数据的“.txt”格式导入 Morpho J 软件对获得的数据进行广义普氏分析 (Generalized Procrustes Analysis, GPA), 通过对坐标数据进行傅里叶转换, 将所有坐标数据平移、旋转和缩放, 使其最大限度的重合。广义普氏分析方法减少了叶片大小对叶片形态的影响, 以转化后的普氏坐标来代表叶片形状。其次, 根据数据的维数和数据集中样本数量之间的关系, 使用 Procrustes 距离可以检查哪些叶片与平均形状严重偏离, 此类异常的叶片被称为“离群值 (outlier)”。剔除离群值后生成新数据集用于后续分析。

将普氏坐标中剔除离群值后的叶形数据生成对称组分和非对称组分数据集。

以 ID 中不同位置的分类器进行分层,包括对称组分,非对称组分以及叶片大小,然后可以对不同分类层次的数据做叶片的平均构型,建立各个层次的协方差矩阵 (covariance matrix),用于后续的多元统计分析。

2.3.4 薄版样条法

利用分类器对叶形分层后,创建可编辑的变换网格 (transformation grid) 和线框图 (wireframe),可以直接观察到叶形态的轮廓,称为薄版样条法 (Thin-Plate Spline, TPS)。利用薄板样条法在后续分析中输出坐标轴上重要值 (最大值或最小值) 的变换网格和对应线框图,将变换网格与线框图重叠进而重建叶片形状,分析叶形收缩或扩张的变化趋势,并将某一物种叶形的变异范围可视化 (Viscosi and Cardini, 2011)。

2.3.5 主成分分析

主成分分析 (Principal Component Analysis, PCA) 是将原始的叶形数据转化,尽可能多的提取原始变量中解释变异的成分,来体现样本间存在的差异。它可以用来显示数据集中形状变化的主要特征,也可以作为一种排序方法来发现观测之间的关系模式。Procrustes 坐标的 PCA 不考虑弯曲能量的权重,也不考虑形状变化的均匀分量 (Rohlf, 2003)。在 MorphoJ 中,主成分分析从协方差矩阵调用,协方差矩阵可用于检验样本中形状变化的主要特征。每个主成分 (PC) 明确地属于一个子空间,每个子空间的变化量可以通过对具有各自对称或非对称类型的 PCs 相关特征值的累加得到。

我们将 13 个标志点基于个体水平的协方差矩阵 (对称组分及非对称组分) 进行线性组合,从中提取出对叶片形态变异贡献较大的组合变量。以 90% 的置信区间将所有叶片形态进行划分,根据 PCA 散点图的结果来分析榲栌榲树及杂交个体的叶片形状是否存在差异,并按照前两个 PCs 生成对应最大值和最小值的变换网格和线框图。

2.3.6 典型变量分析

典型变量分析 (Canonical Variate Analysis, CVA) 是一种通过马氏距离 (Mahalanobis distance) 和普氏距离 (Procrustes distance) 的排序分析和置换测试 (10,000 permutations per test),最大限度地分离我们预先设定的组 (基于 ID 中的分类器) 的常用方法。可检测至少三组间的组间差异并在同一坐标内以置信椭圆进行区分,设置置信区间为 90%, CVA 的结果用显著典型变量 (CV) 评分的散点图来表示。对于少量组进行区分时,默认值为每次比较 10,000 次迭代,这对于大多数研究应用程序来说是合理的。根据生成的变换网格和线框图界面,可以显示与规范变量 (CVs) 相关联的形状更改。这种变化通过形状回归得到的各自 CV 分数相同 (Rohlf *et al.*, 2000)。针对 CV 轴的不同位置生成相应位置的叶

片形状，CV 形状变化的尺度因子以马氏距离为单位（与 CV 分数的散点图坐标轴的单位相同）。根据 CVA 散点图的结果来分析榭栎榭树及杂交个体的叶片形状差异是否显著，并生成对应 CV 轴上物种边界处的变换网格和线框图。并可以在同一图中，分别将同域和异域的榭栎榭树叶形差异清晰地展现出来。

2.3.7 判别分析

判别分析 (Discriminant analysis, DA) 是通过交叉验证函数 (cross-validated scores) 和判别函数 (discriminant scores) 两种方法，为组间比较提供了可靠的信息。与 CVA 的不同之处在于，DA 只考虑了组间的两两判别。在两个以上的组比较时，对每一对进行单独的分析，或者只对选定的一对进行分析。MorphoJ 中的实现使用 Fisher 的分类规则，该规则将分界点设置为 0。分析自动包括了组间均值差异的参数 T^2 检验。DA 的结果用频率分布直方图的重叠程度来表示。本研究中用于榭栎、榭树以及其杂交个体三者间的两两判别，也可以生成同域异域的榭栎榭树四者之间的两两判别，观察其形状变化程度。

2.3.8 双偏最小二乘分析

双偏最小二乘 (Two Blocks Partial Least Squares, 2B-PLS) 方法用于检验两组或多组变量之间的共变模式 (Bookstein *et al.*, 2003)，广泛应用于几何形态学的研究中 (e.g. Rohlf and Corti, 2000)，例如研究叶片大小与形态的共变模式。本研究中利用 2B-PLS 分析榭栎榭树叶片大小与形状的共变关系，即将叶片形状性状映射到叶片大小的性状上，并生成变换网格和线框图相应极值时的形状，以表示出叶形随叶片大小的变化而发生的变化。PLS 也可用于将形状数据与其他类型的数据 (生态信息、实验条件等) 或其他形状变量相关联。如果块 (Blocks) 中的变量不是形状变量 (例如质心大小或协变量)，则使用双图 (biplots)。从 MorphoJ 中分别以榭栎榭树的叶片对称组分和非对称组分数据，作为叶形数据，将之前获取的 4 个环境因子及 2 个地理因子作为协变量数据导入，做二者的双偏最小二乘分析。

2.3.9 冗余分析

为了分析地理和气候因子是否解释了一些叶形变异 (对称组分和非对称组分)，我们使用 R 语言的函数包 “usdm” (Oksanen *et al.*, 2017) 对叶形和环境因子进行基于距离的冗余分析 (distance-based Redundancy Analyses, dbRDAs) 进行了边际测试 (marginal tests) 和条件测试 (conditional tests) (Legendre and Anderson, 1999)。dbRDA 可以直接执行在选择的距离或不同的响应矩阵，是多元多重回归的一种形式。

我们首先从 MorphoJ 中分别导出了榭栎榭树的叶片普氏坐标数据，作为叶

形数据，使用方差划分运行了包括解释变量的边际模型（地理+气候）。另外考虑到地理分布和局部气候因素在对叶片形态变异方面的相对影响，我们分别运行了三个不同的条件的 dbRDAs 模型，分别通过控制另外一个因素，测试了单一地理效应和单一气候效应：（1）控制气候效应时地理变量的条件模型（地理+Condition（气候））；（2）控制地理时气候变量的条件模型（气候+Condition（地理））。最后，通过方差分析（anova function）评估每个独立因素对叶形变异的影响，并统计 dbRDA 的 F 值、解释变异的百分比和 P 值（Oksanen *et al.*, 2017）。

2.3.10 标志点配置中的模块化

为了评估标志点配置中的模块化假设，我们将榲栌榲树标志点划分为两个或多个对应于假设模块的子集。通过 MorphoJ 比较了这种假定的细分与所有可能的替代分区之间的共变强度（Klingenberg, 2009），将假定分区的 RV 系数与备选分区的 RV 系数分布进行比较，假定分区的 RV 系数小于或等于备选分区的 RV 系数时的分区比例，可以解释为 P 值的模拟。RV 系数由以下公式获得：

当两组变量包含在随机向量 x_1 和 x_2 中，由 p 和 q 变量组成，可以写成一个组合的随机向量 $x = (x_1, x_2)$ ，长度是 $p+q$ ，变量的组合向量定义了一个协方差矩阵，其公式如下

$$S = \begin{bmatrix} S_1 & S_{12} \\ S_{21} & S_2 \end{bmatrix} \quad (2-1)$$

对角块 S_1 和 S_2 各自独立地对应于两组变量的协方差矩阵，而非对角块 S_{12} 是两组变量之间的协方差矩阵（矩阵 S_{21} 是 S_{12} 的转置）。RV 系数计算如下：

$$RV = \frac{\text{trace}(S_{12}S_{21})}{\sqrt{\text{trace}(S_1S_1) \text{trace}(S_2S_2)}} \quad (2-2)$$

如果有两组以上的假定模块，RV 系数可以用来评估每一对集合之间的关联强度，但它不能同时提供所有子集之间的关联的一个整体度量。Klingenberg 定义了一种新的多组变量间关联测度，即 RV_M 系数（Klingenberg, 2009），作为集合之间所有成对 RV_M 系数的平均值：

$$RV_M = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^k RV(i,j) \quad (2-3)$$

RV_M 系数可以通过一种置换方法来检验所有变量集之间独立的零假设，这种置换方法将上述方法扩展到两组变量。这可以看作是对子集之间集成的总体测试。

我们对榲栌榲树叶片形状首先进行如图 2.3 三个假设：A：分为叶柄（LM1-LM2）和叶片（LM3-LM13）两个模块（图 2.3A）；B：划分为叶基部（LM1、LM2、LM8、LM13）、叶中部（LM6、LM7、LM11、LM12）和叶尖部（LM3-LM5、LM9、LM10）三个模块（图 2.3B）；C：划分为叶柄（LM1-LM2）、叶尖（LM3-LM5、LM9、LM10）以及叶片下半部分（LM6-LM8、LM11-LM13）三

个模块（图 2.3C）。

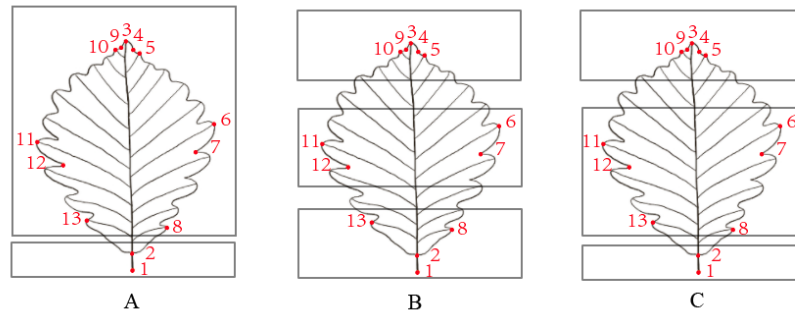


图 2.3 槲栎和槲树叶片标志点配置中的模块化假设

Figure 2.3 Modular hypothesis in the configuration of *Q. dentata*, *Q. aliena* leaf landmarks.

(A) 叶柄和叶片部分 (B) 叶基部、叶中部和叶尖部 (C) 叶柄、叶尖以及叶片下半部分

(A) petiole and leaf part, (B) base, middle and tip of leaf, and (C) petiole, tip and lower part of leaf

2.4 深度学习分类

2.4.1 数据集划分

Xception 模型使用了深度可分离的卷积增加了网络宽度，这样不仅可以提高分类的正确率，也增强了网络对细微特征的学习能力。将 Xception 应用于白栎组近缘种的分类上，基于扫描的叶片图像，按分类单元划分不同的数据集：**Data A**：槲栎和槲树的二分类；**Data B**：槲栎和杂交个体的二分类；**Data C**：槲树和杂交个体的二分类；**Data D**：槲栎、槲树以及杂交个体的三分类，如表 2.3 所示。

表 2.3 数据集包含的分类单元

Table 2.3 Taxon of the dataset

数据集	槲栎图片 (张)	槲树图片 (张)	杂交图片 (张)	共计 (张)
Dataset	<i>Q. aliena</i>	<i>Q. dentata</i>	Hybrid	Total
Data A	539	1202	0	1741
Data B	539	0	480	1019
Data C	0	1202	480	1682
Data D	539	1202	480	2221

2.4.2 实验步骤

利用 nSSRs 技术确定了本文使用的数据集组中的物种，使用包含 539 张 *Q. aliena* 图像、1,202 张 *Q. dentata* 图像和 480 张 *Q. aliena* 和 *Q. dentata* 杂交个体的

2,221 张扫描图像进行训练、验证和测试。图像采集设备是 CanoScan 5600 F 扫描仪 (Canon Inc.)，扫描精度为 600 dpi，图像尺寸为 4,960 像素×7,015 像素。

将数据分成三个不同的子集进行训练 (70%)、验证 (15%) 和测试 (15%)，然后将需要二分法的物种对应的数据集放在一起，随机排列，用于 Xception 模型训练和预测。为了获得一个清晰的分类，图像数据集被分为以下方面：Data A 包括 539 张 *Q. aliena* 图像和 1,202 张 *Q. dentata* 图像；Data B 包括 539 张 *Q. aliena* 图像和 480 张杂交个体的图像；Data C 包括 862 张 *Q. dentata* 图像和 480 张杂交个体的图像；Data D 包括 539 张 *Q. aliena* 图像，862 张 *Q. dentata* 图像和 480 张杂交个体的图像 (表 2.3)。对每个训练集中的物种图像进行相应编号，例如，格式为 BJFU_000001.jpg。另外，在三个数据集的 GroundTruth 文件夹中，对应序列号的物种都用 CSV 格式的数据清晰地标注出来。

2.4.3 评价指标

为了评价分类器的结果，确定了不同的参数。一个物种的分类率或准确度即预测正确类所占比重，由下列关系给出：

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2-4)$$

T 和 F 代表 True 和 False, 代表预测是否正确; P 和 N 代表 Positive 和 Negative, 代表预测结果。其中 TP (True Positive, 真阳性率) 表示将正类预测为正类的识别数; TN (True Negative, 真阴性率) 表示将负类预测为负类的识别数; FN (假阴性率) 表示将正类预测为负类的识别数; FP (假阳性率) 表示将负类预测为正类的识别数。重要的是要考虑更详细的参数:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (2-5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (2-6)$$

为了将两个参数同等地考虑，我们利用 F-score 表示 Precision 和 Recall 的加权调和平均，计算公式如下：

$$\text{F-score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (2-7)$$

2.5 三种鉴别方法的综合比较

为了能够比较鉴别栎类三种方法的投入，我们分别量化其成本，将三种方法所需的实验耗材、仪器以及人力转化为成本。以本论文中的实验为参考，预估鉴别不同个体所需成本：

分子标记方法:本方法所需的实验试剂包括:植物基因组 DNA 提取试剂盒, PCR Mix, 5' 末端带有 M13 的正向引物, 反向引物和 5' 末端带有荧光标记的 M13 引物(四种荧光标记为: FAM, HEX, TAMRA, ROX)。每个个体扩增一个 SSR 引物需要模板 DNA、正向引物、反向引物、荧光标记的 M13 引物以及 Mix (15 μ L 体系)。

植物基因组 DNA 提取试剂盒在天根生化科技有限公司订购, 每个样品约 5 元。引物和 Mix 均在生工生物工程有限公司订购, 每对引物的价格按碱基数量收费, 5' 末端带有 M13 的正向引物和反向引物的碱基单价是 0.5 元, 5' 末端带有荧光标记的 M13 引物的碱基单价是 1.2 元(4 种荧光标记修饰的价格不同)。根据本文所使用的引物(表 2.2), 每 OD(约 500 μ L) 正向引物的价格是: 256 碱基 \times 0.5 = 128 元, 每 OD(约 500 μ L) 反向引物的价格是: 262 碱基 \times 0.5 = 131 元, 每 OD(约 500 μ L) 带有荧光标记的 M13 引物的价格是: 1,287 元, 1,000mL Mix 的价格是: 10 元。

对提取的 DNA 进行 PCR 扩增后, 将扩增产物送到北京擎科生物有限公司进行毛细管电泳检测。为提高检测效率和降低成本, 本文使用 96 孔板进行混样检测, 如图 2.4 所示:



图 2.4 混样检测示意图

Figure 2.4 Schematic diagram of mixed sample detection.

本实验所需的人员为受过系统的分子生物学知识的科研人员, 估计劳务报酬为 100 元/h。提取一个 DNA 所需 4h, 一次可提取 24 个 DNA, 本实验所需提取 DNA 的时间大约是 167h。扩增 1 个 PCR 所需 2h, 一次可扩增 96 个 DNA, 每个 DNA 需要使用 12 对引物进行扩增, 所需总时间为 264h。

几何形态学方法:本方法所需的实验步骤包括: 扫描叶片、标记叶片标志点以及数据分析。每个个体扫描五张叶片, 所需时间为 10min; 每张叶片标记 13 个

坐标点并导出数据，每个个体所需时间为 8min；扫描标记步骤无需科研知识，按一般工人劳务报酬 50 元/h 计算成本。后续的多元分析所需人员应具有 MorphoJ 软件操作基础及多元统计分析基础，需花费 1h，劳务报酬为科研人员报酬约 100 元/h。

深度学习方法：本方法所需的实验步骤包括：图像预处理及模型的修改调试，将所有图片随机排列并重新命名，对每一张图片写清分类标签，预计每张图片所需 1min，本方法所需的人员需具有人工智能相关背景知识，按科研人员劳务平均标准 100 元/h 计算成本。

3 结果与分析

3.1 榲栌和榲树按遗传结构划分

通过 STRUCTURE 的分析结果表明榲栌（红色）和榲树（绿色）聚到一起，明显分为两个组（图 3.1）。我们通过 STRUCTURE 分析结果进行统计，根据 Q 值等于 0.9 为标准，将所研究的这 54 个榲栌榲树种群中的 710 个个体进行划分，发现有 149 个个体被分到了榲栌，322 个个体被分到了榲树，另外有 239 个个体在遗传上被归类为假定的杂交个体。

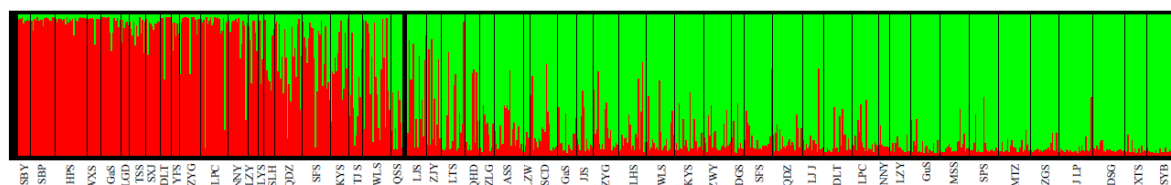


图 3.1 榲栌和榲树遗传结构图

Figure 3.1 STRUCTURE analysis between *Q. aliena* and *Q. dentata*

3.2 榲栌和榲树的叶片形态特征

3.2.1 叶片形状差异分析

在叶片坐标普式叠印的过程中，去除叶片的位置摆放和大小的影响，散点集中地分布在对应叶片 13 个标志点周围（图 3.2A）。对称组分（图 3.2B）和非对称组分（图 3.2C）的散点也表明 GPA 很好地排除了叶片位置和大小对形态研究的干扰。通过 GPA 对两个种的叶形进行初步可视化操作，剔除离群值，最终从 3,657 片叶片中挑选出 3,612 片叶片用于后续分析。

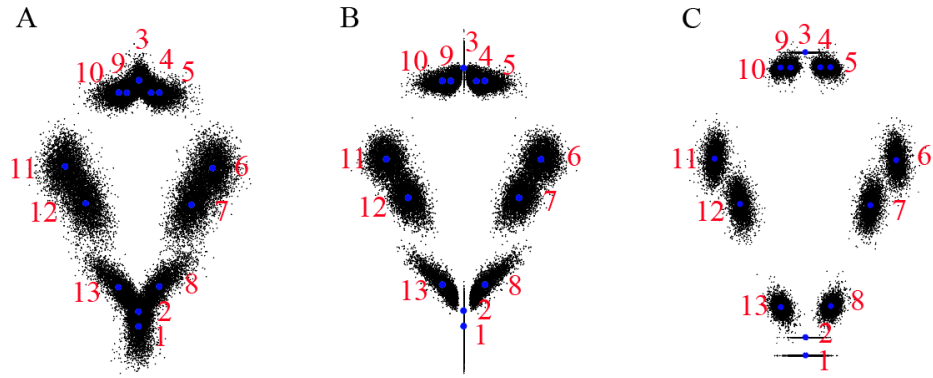


图 3.2 叶片 procrustes 叠印分析散点图基于普氏坐标数据 (A)、对称组分 (B) 和非对称组分 (C)

Figure 3.2 Generalized procrustes analysis of the leaf shape based on Procrustes coordinates (A), symmetric (B) and asymmetric components (C)

我们基于个体水平的 PCA，分别对对称组分和非对称组分的槲栎槲树叶片形状变异进行了研究（图 3.3）。主成分分析基于 13 个坐标点数据，由所有转换和重新标记的标志点配置副本组成，对对称组分和非对称组分分别产生 11 个具有非零特征值的 pc （即解释一些变化的 pc ）。基于对称组分，我们发现槲栎和槲树之间存在明显的种间差异，其杂交个体的置信椭圆包含了绝大部分的纯种（图 3.3A）。其中，PC1 和 PC2 的累积贡献率占 65%，尤其是沿着 PC1 轴，两物种出现了明显分离。因此我们分别生成了槲栎和槲树在 PC1、PC2 轴正负方向上极值处的变换网格和线框图。从变换网格图和线框图的叠加可以看出其沿某一 PC 轴的叶片形状变化趋势，叶形沿 PC1 轴（-0.1~0.15）的变化为：叶柄的相对长度由短变长，叶尖从短钝尖变长渐尖，叶上半部分由宽到窄，叶片最宽的叶窦位置相对下移，叶基部由窄变宽，叶片总体呈上半部分收缩，下半部分扩张的趋势。叶形沿 PC2 轴（-0.1~0.1）的变化为：叶片整体变窄，呈收缩趋势，叶尖从短钝尖变长渐尖，叶上半部分由宽到窄，叶基部呈收缩趋势。

非对称组分中，前两个 PCs 分别占总累计频率的 61%和 17%。PCA 的散点图显示，大部分散点重叠分布，两物种的样本存在差异较小（图 3.3B）。在 PC1、PC2 轴正负方向上极值处的变换网格和线框图也无规律性变化。

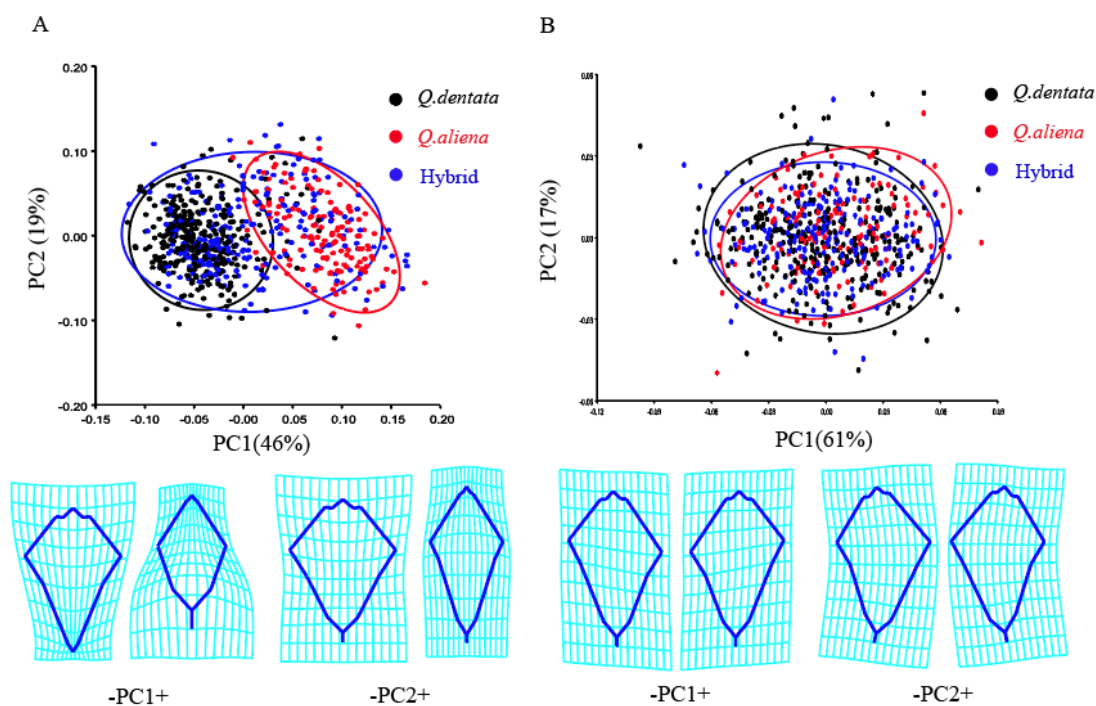


图 3.3 槲栎和槲栎及杂交个体的 PCA 分析

Figure 3.3 Results of PCA of *Q. dentata*, *Q. aliena* and hybrids

3.2.2 叶片的判别分析

我们将每棵树采集的全部叶片做平均，以叶片平均轮廓代表个体来减小误差，在个体水平对槲栎槲栎叶片进行了判别。CVA 和 DA 的结果显示槲栎槲栎的叶形存在显著差异，遗传假定的杂交个体处于二者中间形态。

从 CVA 的散点图可以看出（图 3.4），沿着 CV1 轴（累积贡献率为 96%），槲栎与槲栎几乎全部分离，而杂交个体散布于两个种的分布范围之间。沿着 CV2 的置信椭圆基本上都重叠（累积贡献率为 4%）。分别以 CV1 轴上的 -4、0 和 3 三个位置生成相应的转换网格和线框图，分别代表槲栎、杂交个体和槲栎的叶形轮廓。由变换网格和线框图可以看出，槲栎和槲栎的区别在于槲栎叶形呈上窄下宽，叶柄相对长度较长，顶端尖基部收缩慢。槲栎叶形上宽下窄，叶柄相对长度较短，顶端钝且基部收缩快。而假定杂交个体的叶片形状介于这两种形态之间，但更接近槲栎的叶片形态（沿 CV1 方向）。

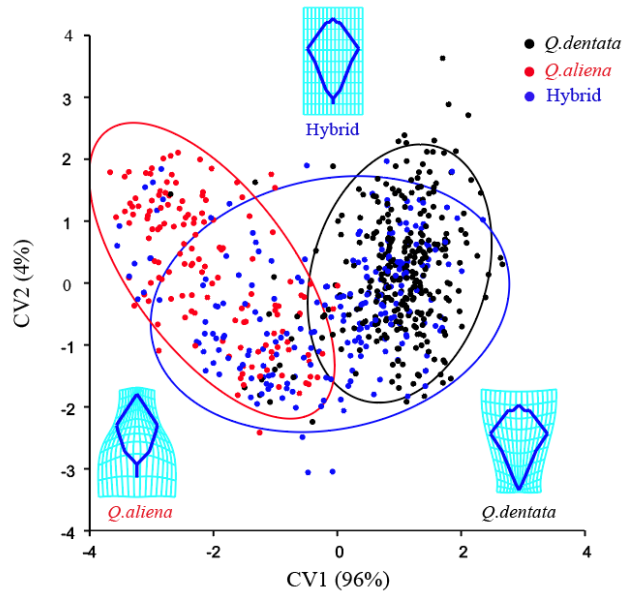


图 3.4 槲栎和槲树及杂交个体 CVA 分析

Figure 3.4 Results of CVA of *Q. dentata*, *Q. aliena* and hybrid

DA 中交叉验证的结果（图 3.5）表明，槲栎和槲树之间叶形差异显著（ $P < 0.0001$ ）。DA 判别结果的统计发现，在槲树槲栎的判别中（图 3.5a），90%的槲栎可以被正确判别，槲树有 95%的可以被正确判别。在槲栎与杂交个体的判别中（图 3.5b），79%的槲栎可以被正确判别，杂交个体有 73%的可以被正确判别。在槲树与杂交个体的判别中（图 3.5c），84%的槲树可以被正确判别，杂交个体有 54%的可以被正确判别。两两之间的马氏距离分别为 3.3886（槲栎和槲树），1.3229（槲栎和杂交个体），1.0443（槲树和杂交个体）；普氏距离分别为 0.1268（槲栎和槲树），0.0727（槲栎和杂交个体），0.0554（槲树和杂交个体）。从马氏和普氏距离都可以看出，假定的杂交个体与槲树的距离更小，表明杂交个体的叶形与槲树的叶形更相似。

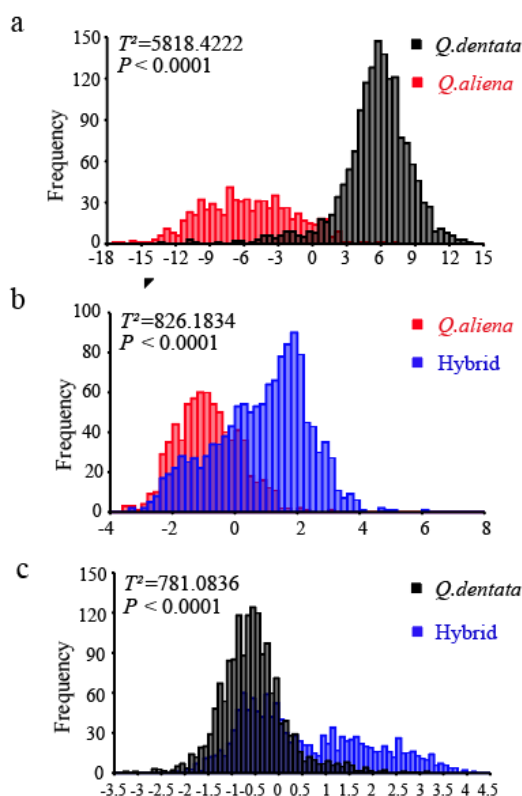


图 3.5 槲栎和槲树及杂交个体基于叶片水平 DA 分析

Figure 3.5 Results of DA of *Q. dentata*, *Q. aliena* and hybrid based on leaf level

3.2.3 叶片异速生长分析

我们还分别检测了槲栎槲树的对称组分和非对称组分的叶片形状与叶片大小的共变关系，即异速生长（图 3.6）。基于对称组分的结果表明，槲栎槲树的叶片形状和叶片大小存在显著的异速生长模式（图 3.6A）（ $P < 0.0001$ ），非对称组分的数据不显著（图 3.6B）（ $P = 0.5980$ ）。从变换网格与线框图叠加图可以看出，随着叶片由小到大，Log centroid size 从负值到正值，叶片形状由近椭圆形趋向近四边形变化，叶片尖端和叶片基部都呈现收缩趋势，叶柄的相对长度变短（图 3.6A）。另外非对称组分的叶形和叶片大小不存在异速生长关系（图 3.6B）。

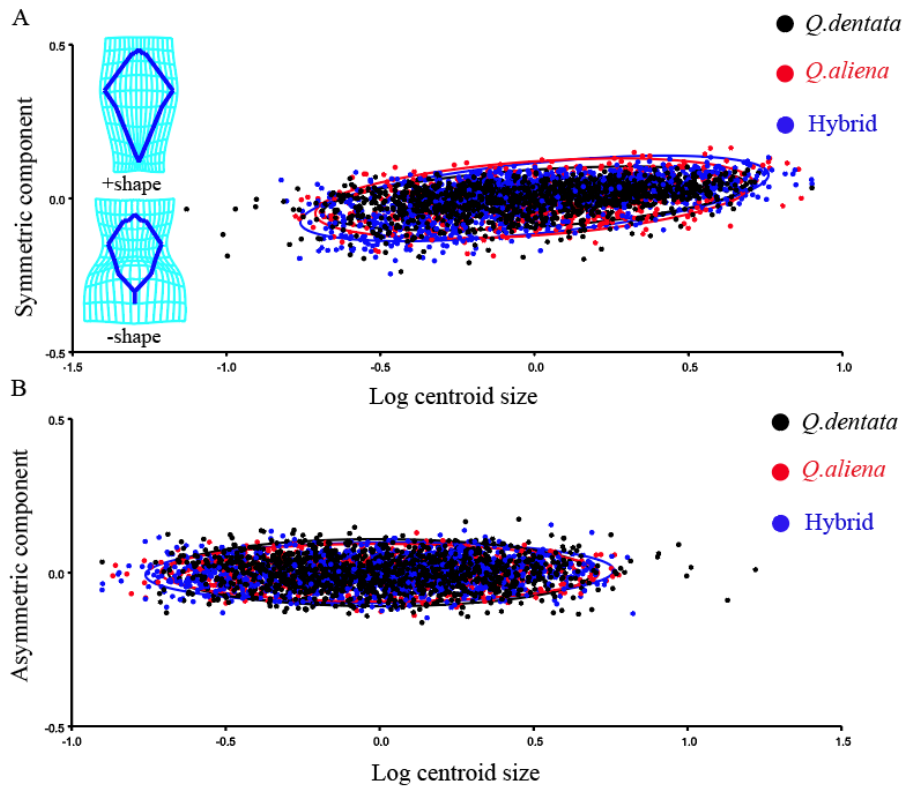


图 3.6 槲栎和槲树及杂交个体基于对称组分 (A) 和非对称组分 (B) 的 2B-PLS 分析
 Figure 3.6 Results of 2B-PLS analysis of *Q. dentata*, *Q. aliena* and hybrid based on symmetric components (A) and asymmetric components (B)

3.2.4 同域和异域对叶片的影响

从 CVA 的散点图可以看出 (图 3.7), 沿着 CV1 轴 (累积贡献率为 94.8%), 槲栎槲树整体上呈分离, 同域槲树的置信椭圆有向槲栎方向相对较小的偏移, 异域的槲栎有向槲树方向相对较大的偏移。沿着 CV2 轴, 各类群置信椭圆基本上都重叠 (占总贡献率的 4.1%), 槲栎在 CV2 上的跨度较大。同域的槲栎和异域的槲栎间马氏距离为 1.7567 ($P < 0.0001$), 同域的槲树和异域的槲树间马氏距离为 0.7102 ($P = 0.0219$)。

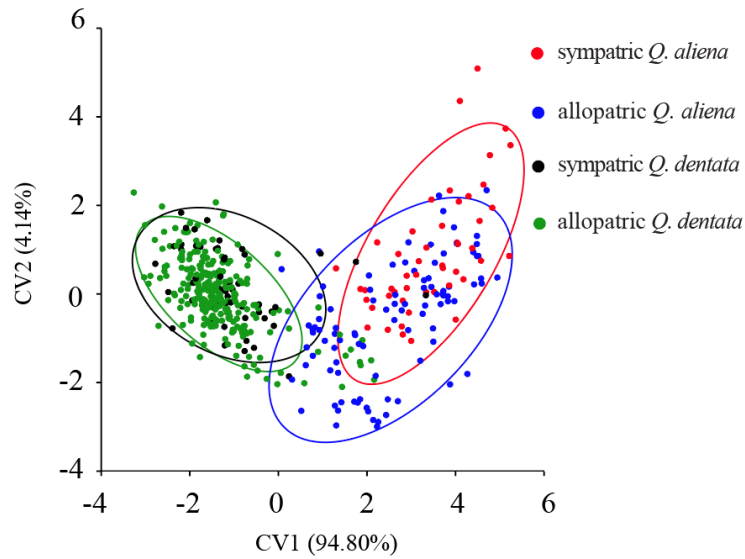


图 3.7 榲栌和榲树同域异域的 CVA 分析

Figure 3.7 Results of CVA of *Q. dentata* and *Q. aliena* of sympatric and allopatric

DA 中交叉验证的结果 (图 3.8) 表明, 叶片在榲栌的同域和异域的判别中 (图 3.8A), 同域有 68.12% 的叶片被正确区分, 异域有 60.32% 的叶片被正确区分 ($T^2=217.0449, P<0.0001$)。叶片在榲树的同域和异域的判别中 (图 3.8C), 同域有 68.12% 的叶片被正确区分, 异域有 60.32% 的叶片被正确区分 ($T^2=76.8588, P<0.0001$)。同域的榲栌和榲树之间存在的叶形差异要比异域之间存在的叶形差异更为显著。

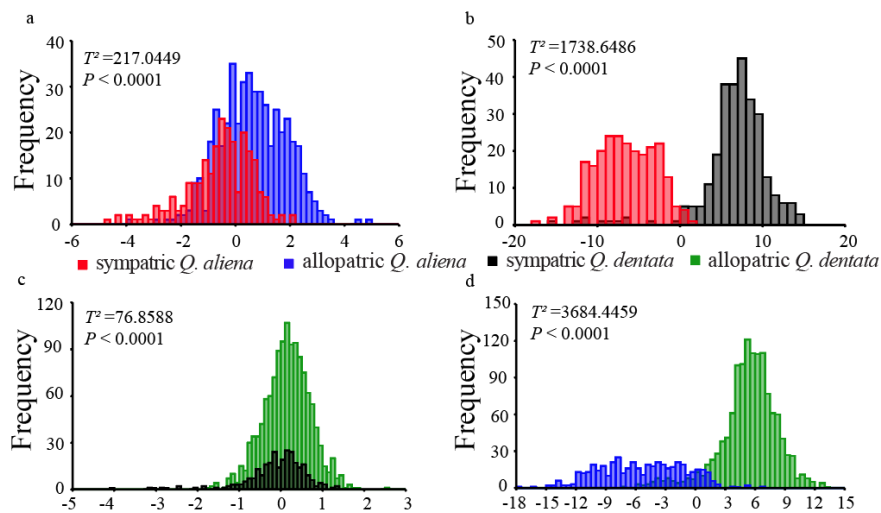


图 3.8 榲栌和榲树同域异域的 DA 分析

Figure 3.8 DA analysis of *Q. dentata* and *Q. aliena* of sympatric and allopatric

3.2.5 地理和气候对叶片的影响

我们以地理和气候数据作为协变量,与叶片形状做偏最小二乘分析(图 3.9),发现在叶片对称组分中,PLS1 占累积贡献率的 89%,PLS2 占累积贡献率的 10%,解释度最大的是 bio15 和 prec08 (图 3.9A)。在叶片非对称组分中,PLS1 占累积贡献率的 96%,PLS2 占累积贡献率的 3%,解释度最大的也是 bio15 和 prec08。由此我们得出,地理和气候因子中对叶形(对称/非对称组分)影响最突出的因子是 bio15 和 prec08 (图 3.9B)。

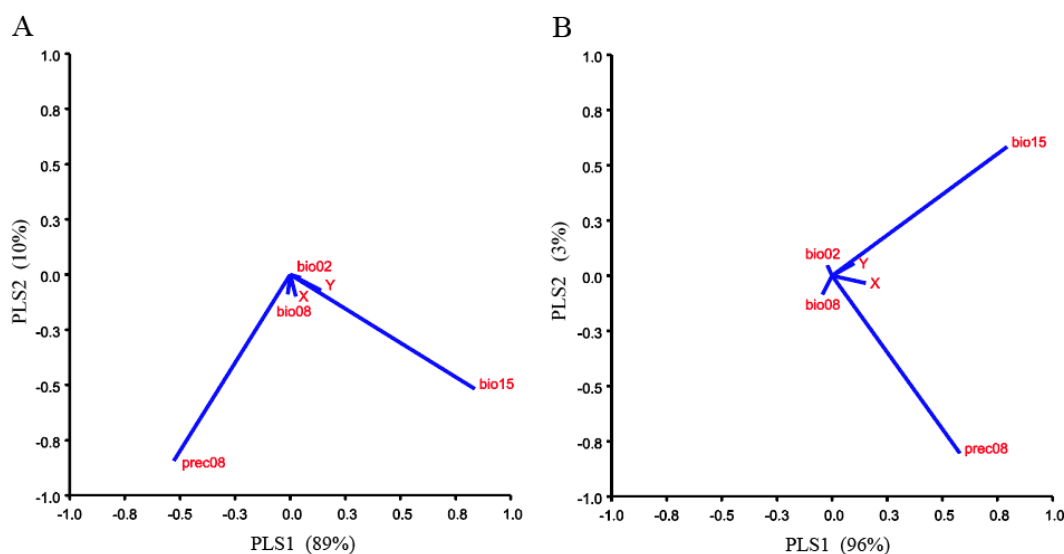


图 3.9 槲栎和槲树及杂交个体与环境协变量基于对称组分 (A) 和非对称组分 (B) 的 2B-PLS 分析

Figure 3.9 2B-PLS analysis of environmental covariates of *Q. dentata*, *Q. aliena* and hybrid based on symmetric components (A) and asymmetric components (B)

基于边际和条件的 dbRDA 分析了地理和气候因子对叶形的影响,并控制其中一种预测因子使其单独分析相关性, F 值代表整个拟合方程的显著性, F 值越大,拟合程度越好, P 值代表不拒绝原假设的程度, P 值越小,则更能证明该假设是正确的,方差代表解释变异的百分比。结果表明:槲栎的叶片形状在边际测试中受经度影响显著,槲树的叶形受地理和气候影响不显著。但在控制地理或气候因子后,各指标对槲栎槲树的叶形影响都不显著(表 3.1)。

表 3.1 利用基于距离的冗余分析测试叶片形状与地理、气候变量之间的关系

Table 3.1 Tests of the association of leaf shape on the total sample with geographical and climatic of predictor variables using distance-based Redundancy Analysis (dbRDA)

测试 test	变量预测因子 Variable predictors	榭栎 <i>Q.aliena</i>			榭树 <i>Q.dentata</i>		
		<i>F</i>	<i>P</i>	方差 variance (%)	<i>F</i>	<i>P</i>	方差 variance (%)
边际测试 (全部变量) Marginal (all variables)	bio15	0.5422	0.683	2.53	0.794	0.569	2.47
	bio02	0.5348	0.728	2.50	1.165	0.336	3.62
	bio08	0.9075	0.46	4.24	1.902	0.069	5.91
	prec08	1.1559	0.333	5.40	1.317	0.266	4.09
	Latitude	0.3428	0.871	1.60	0.429	0.897	1.33
	Longitude	3.6024	0.010 *	16.82	1.23	0.283	3.82
	Altitude	0.3365	0.869	1.57	1.359	0.229	4.22
条件测试 (地理) Conditional (Geography)	bio15	0.5422	0.674	2.53	0.794	0.601	2.47
	bio02	0.5348	0.706	2.50	1.165	0.314	3.62
	bio08	0.9075	0.445	4.24	1.902	0.071	5.91
	prec08	1.1559	0.289	5.40	1.317	0.251	4.09
条件测试 (气候) Conditional (Climate)	Latitude	0.3759	0.834	1.75	1.239	0.274	3.85
	Longitude	0.6022	0.631	2.81	0.788	0.567	2.45
	Altitude	0.3754	0.835	1.75	1.718	0.116	5.34

注：边际测试包括所有变量，而条件测试则只考虑所选变量的变化。 * $P < 0.05$

The marginal test includes all variables, while the conditional tests account for variation in the selected variables.

3.2.6 叶片标志点配置中的模块化假设

榲栌榲树（对称组分和非对称组分）的平方轨迹相关直方图中，可以看出当榲栌（图 3.10A）和榲树（图 3.11A）按照图 2.3A 划分的结果中，榲栌非对称组分 RV 小于或等于先验假设的分区数为 11，与真实模块的边界非常接近。而在按照图 2.3B 分成三部分时，榲栌对称组分和非对称组分（图 3.10B）RV 小于或等于先验假设的分区数为 2,079 和 1,210，榲树对称组分和非对称组分（图 3.11B）RV 小于或等于先验假设的分区数为 372 和 374；而在按照图 2.3C 分成三部分时，榲栌对称组分和非对称组分（图 3.10C）RV 小于或等于先验假设的分区数为 870 和 390，榲树对称组分和非对称组分（图 3.11C）RV 小于或等于先验假设的分区数为 28 和 43（表 3.2）。因此，榲栌对称组分和榲树的对称组分和非对称组分按照图 2.3C 分成三部分，RV 系数明显接近系数分布的下限（图 3.10 和 3.11 箭头），这一结果说明 C 假设与模块分区是一致的。

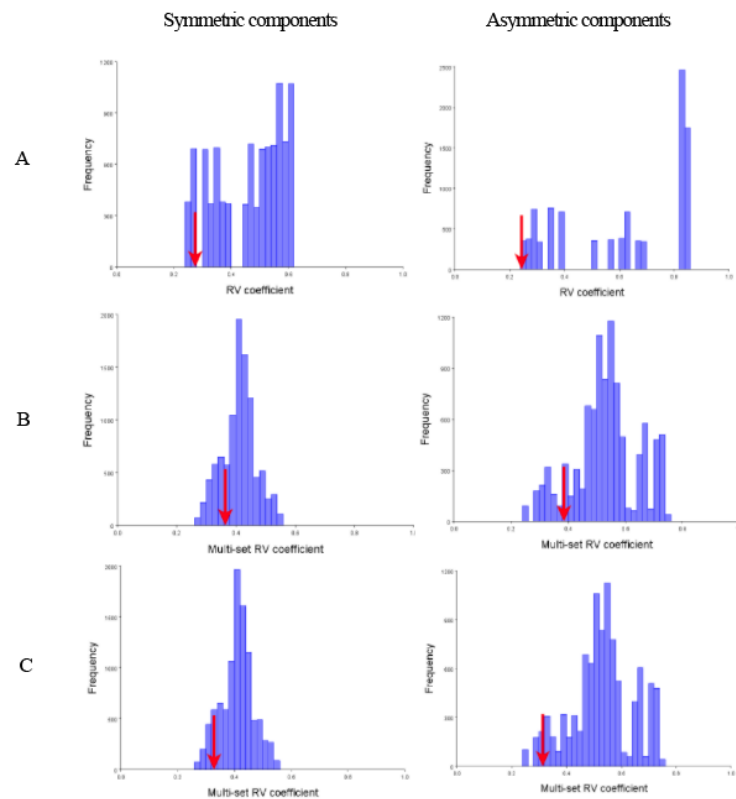


图 3.10 榲栌模块化假设的所有可能分区的平方轨迹相关直方图。箭头是全部分区中观察到的最低值。

Figure 3.10 Histograms of the squared trace correlations for all possible partitions of Q . *aliena*. The arrow is the lowest value observed for all partitions of the configuration.

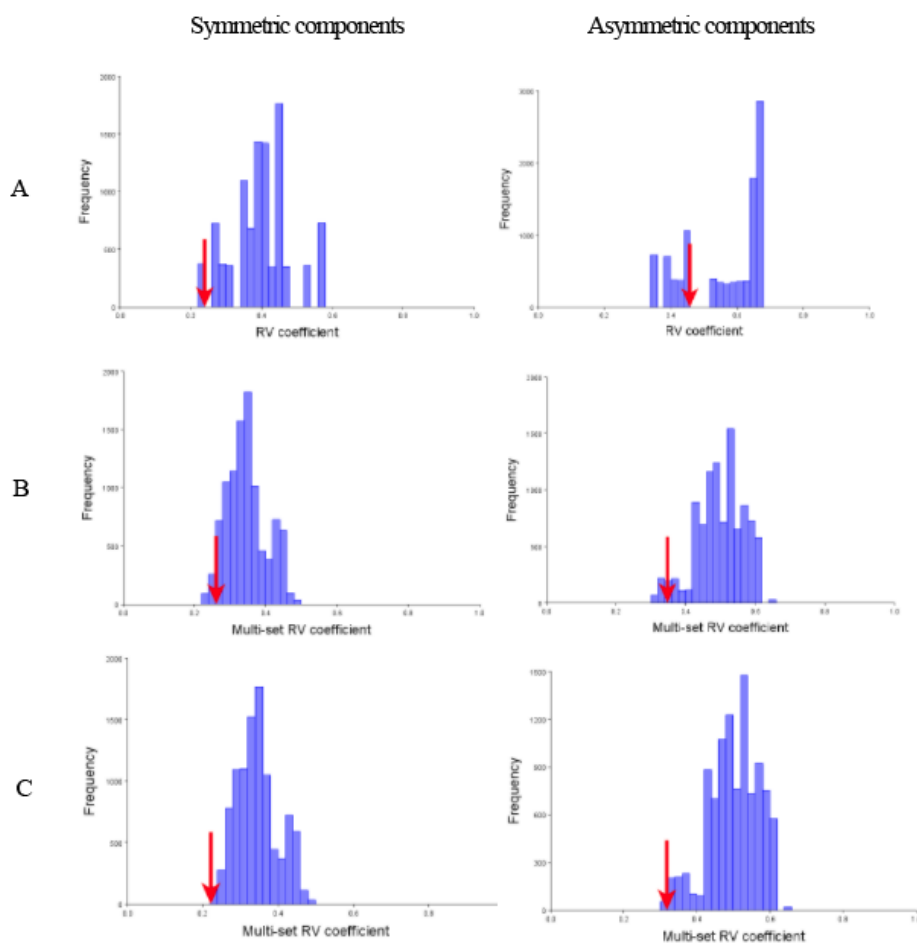


图 3.11 柵树模块化假设的所有可能分区的平方迹相关直方图。箭头是全部分区中观察到的最低值。

Figure 3.11 Histograms of the squared trace correlations for all possible partitions of *Q. dentata*. The arrow is the lowest value observed for all partitions of the configuration

表 3.2 柵栎柵树及杂交个体对应排列试验的 RV_M 系数和分区比例

Table 3.2 Multi-set RV coefficients and Proportion from the corresponding permutation tests for different subdivisions of *Q. dentata*, *Q. aliena* and its hybrids

	柵栎 <i>Q. aliena</i>				柵树 <i>Q. dentata</i>			
	对称组分 symmetric components		非对称组分 asymmetric components		对称组分 symmetric components		非对称组分 asymmetric components	
	$RV/$ RV_M	Proport ion	$RV/$ RV_M	Proport ion	$RV/$ RV_M	Proport ion	$RV/$ RV_M	Proport ion
A	0.275	0.0947	0.243	0.0011	0.239	0.0266	0.458	0.3118
B	0.364	0.2079	0.385	0.1210	0.261	0.0372	0.348	0.0374
C	0.328	0.0870	0.313	0.0390	0.222	0.0028	0.318	0.0043

3.3 槲栎槲树的深度学习分类

从表 3.3 可以看出，Xception 模型在槲栎和槲树叶片扫描图像分类上的分类效果整体较好，准确率可达 95.8%。但在槲栎槲树与杂交个体的三分类中的准确率极低，只有 44.5%，说明杂交个体从形态学上无法与两个种进行区分，与槲树区别的准确率（67.9%）要比与槲栎区别的准确率（71.8%）低，说明杂交个体在形态上与槲树更难区分，即叶形与槲树更为接近。

表 3.3 槲栎槲树及杂交个体基于 Xception 模型的深度学习结果

Table 3.3 Deep learning results of *Q. dentata*, *Q. aliena* and its hybrids with Xception models

	精确率 Precision	召回率 Recall	F 值 F-score	准确率 Accuracy
Data A	0.980	0.935	0.961	95.8%
Data B	0.815	0.662	0.731	71.8%
Data C	0.873	0.554	0.678	67.9%
Data D	0.335	0.329	0.332	44.5%

3.4 三种鉴别方法的综合比较结果

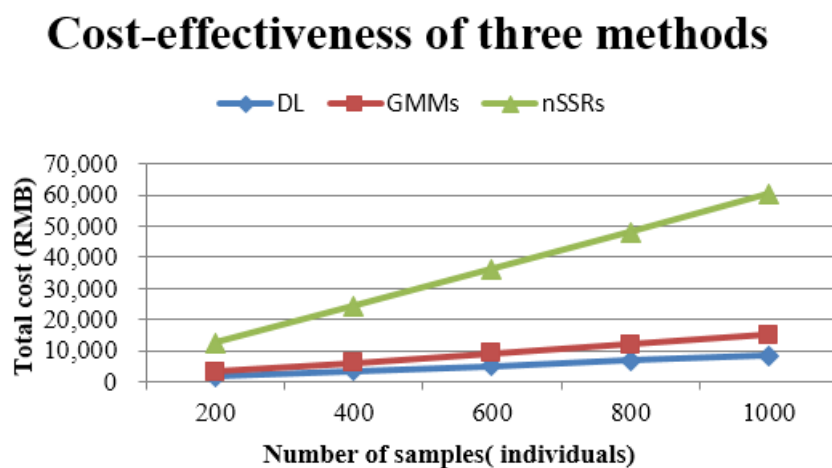


图 3.13 综合比较几何形态学和深度学习两种方法判别栎类的投入产出比

DL: 深度学习; GMMs:几何形态学; nSSRs:分子标记方法

Figure 3.13 The input-output ratios of GMMs and DL methods were compared comprehensively of oaks identification.

DL: Deep Learning; GMMs: Geometric Morphometric Analyses; nSSRs: nuclear Simple Sequence Repeats

我们综合比较了以 nSSRs, GMMs 和 DL 三种方法对栎类鉴别成本。如图 3.13 所示, 随着鉴定个体数量的增加时, nSSRs 需要进行大量的分子实验, 并对测序后的数据进行整理分析, 进行物种鉴定。即 nSSRs 的成本在于人工实验、试剂、引物及耗材等; GMMs 是对叶片进行扫描并人工选取 13 个标志点, 基于标志点的形状变化进行统计分析, 相当于人工提取分类特征, 工作量主要集中在叶片扫描及标记标志点; 在 DL 中, 通过机器学习的自动特征提取, 对扫描的叶片图像进行整合, 随机排列并重命名, 工作量主要集中在图像的预处理和模型的调试。我们将花费的人力成本换算成支出费用, 结合本研究 710 个个体的实际花费情况, 预估不同个体数量的判别所需总支出费用, 以 1,000 个个体为例做如下计算:

nSSRs: 植物基因组 DNA 提取试剂盒: $1,000 \times 5 = 5,000$ 元; Mix: $7,500\mu\text{L} \times 10 / 1,000 = 75$ 元; 正向引物: $60\mu\text{L} \times 128 / 500 = 16$ 元; 反向引物: $240\mu\text{L} \times 131 / 500 = 63$ 元; M13 荧光标记引物: $240\mu\text{L} \times 1,287 / 500 = 618$ 元; 提取 1,000 个个体的 DNA 所需的试剂的总价格是: $5,000 + 75 + 16 + 63 + 618 = 5772$ 元。

将 1,000 个 DNA 个体分别装入 96 孔板中, 共 11 板 DNA 原液, 最后混样检测的样品为 $11 \times 3 = 33$ 包。对于每包混样检测的样品, 擎科生物有限公司的收费标准是 350 元/包, 因此, 毛细管电泳检测的总费用为: $33 \times 350 = 11,550$ 元; 所需的人力总费用为: $(167 + 264) \times 100 = 43,100$ 元。

最终使用分子标记方法对 1,000 个个体先验所需总费用约为: $5,772 + 11,550 + 43,100 = 60,422$ 元。

GMMs: 叶片扫描标记人力费用为: $(10 + 8) / 60 \times 1,000 \times 50 = 15,000$ 元; 叶形数据分析 $1 \times 100 = 100$ 元, 最终几何形态学分析鉴别 1,000 个个体所需总费用为 $15,000 + 100 = 15,100$ 元。

DL: 图片与处理人力费用为: $1,000 \times 1 / 60 \times 5 \times 100 = 8,333$ 元; 模型调试费用 $1 \times 100 = 100$ 元, 最终深度学习鉴别 1,000 个个体所需 $8,333 + 100 = 8,433$ 元。

4 讨论

4.1 槲栎和槲树叶片的分类结果

本研究首先通过 12 对 SSR 引物对 54 个槲栎和槲树种群的进行个体划分，结果表明当 $K = 2$ 时是最佳分类，基于分子标记的 STRUCTURE 聚类结果可以将槲栎和槲树的每个个体准确的划分归类。以 $Q = 0.9$ 为纯合与杂交界限，种群内个体分配给纯合和杂交个体。我们以遗传先验的结果将采样个体划分为槲栎、槲树以及杂交个体，分析了基于叶片形状的分类研究。基于几何形态学分析的槲栎槲树以及杂交个体的物种判别中，CVA 的结果证实了槲栎槲树以及杂交个体的叶形存在一定的变异，但杂交个体主要介于它们之间（图 3.4）。从变换网格图可以清晰地看出它们的叶形差异主要集中在叶尖和基部相关区域。而假定的杂交个体受两亲本物种影响，叶片形状为二者的平均叶型，但与槲树的置信椭圆重叠更大，叶形更相似。DA 的结果也表明，90%的槲栎和 95%的槲树可以被正确判别（图 3.5a）；79%的槲栎与 73%的杂交个体可以被正确判别（图 3.5b）；84%的槲树与 54%的杂交个体可以被正确判别（图 3.5c）。有研究表明，遗传界定的杂交个体的叶片形状具中间态（相对于其亲本物种）（Stephan *et al.*, 2018; Viscosi *et al.*, 2009）。这说明杂交个体的叶片形态可能是亲本物种形态特征的嵌合体。

本研究基于卷积神经网络的深度学习算法自动提取图像特征且无需人工干预，克服了传统植物叶片识别依靠人力来提取特征的弊端。Xception 模型分类准确率较出色，具有很好的泛化能力，使用 Xception 模型在槲栎槲树及杂交个体的叶片扫描图像分类上进行研究分析，按照将其视为三个分类单元进行三分类以及两两分类的二分类。本研究使用的用于训练、验证及测试的叶片图像均为高分辨率扫描图像，背景统一为白色，并非真是生境下的拍摄照片，极大的减少了机器学习算法的误差。由槲栎槲树的二分类准确率（95.8%）远高于其他三组数据，说明槲栎槲树间叶形差异虽然相近，但可以作为区分标准。分别由槲栎和槲树与杂交个体组成的二分类结果表示，杂交个体与槲树的分辨率更低一些，从而证明杂交个体的叶片形状更接近槲树，这一结果与几个形态学中的 CVA 和 DA 结果一致。刘媛（2018）对于槲栎槲树同域分布的 6 个种群的研究中表明，槲栎槲树的杂交个体更倾向于槲栎的形状，推测为槲栎槲树的叶形状特征受同异域影响存在差异。本研究中选取了包括同异域在内的共 54 个种群，在槲栎和槲树的同异域判别分析（图 3.8）中可以发现，槲树的叶形受同异域影响不显著，而异域的槲栎叶形有向槲树叶形靠近的趋势。从而证明，在采样点信息扩大时，尤其是增加异域采样点数据时，杂交个体的叶形会更接近槲树的叶形。

GMMs 可以为叶片形状变化的定量和图形分辨率提供可靠的方法（Viscosi *et*

al., 2012), 它不仅区分物种, 还可以显示叶片收缩和扩张的趋势。GMMs 与 DL 对榲栌和榲树的鉴别准确率都具有重要的参考价值, DL 在精度上优于 GMMs, 在成本和时间上具有显著优势。在杂交个体的鉴定上, GMMs 中 DA 的交叉验证结果略优于 DL, 因为 DL 关注的是整个叶形轮廓, 但杂交个体的整体轮廓与两种物种极为相似, 因此判别误差较大。甚至物种间仅用两个变量就可以进行判别 (Ponton *et al.*, 2004), 故传统形态学在某些情况下的判别准确率略高于 GMMs (Stephan *et al.*, 2018)。nSSRs 方法成本高、时间久、人工实验操作复杂, 需要大规模基因组材料和基础设施, 鉴别结果具有生物学依据, 并且 nSSRs 数据还可用于遗传学等其他研究中; GMMs 方法可以数字化表型特征, 比较形态差异, 但标志点在标记的过程中需由操作者独立完成, 以减少由主观判断而产生的误差; DL 方法省时省力, 分类的准确率高, 且基础数据量越大, 分类精度越高, 但操作者需具备人工智能相关背景知识。综上, 考虑到了时间及人力成本, 深度学习在物种的鉴别中是一种高效便捷的方法。

4.2 榲栌与榲树的叶形态特征

对称性 Morpho J 将形状变化分为对称变化和非对称变化, 它们提供了与不同生物问题相关的信息 (Klingenberg *et al.*, 2002)。对称组分主要用于研究个体之间的形状差异, 而非对称组分往往被用于相关性分析, 比如作为在杂交环境下发育不稳定性的衡量标准 (Mikula and Machola'n, 2008)。对于此类相关研究, MorphoJ 提供了个体非对称的措施 (Klingenberg and Monteiro, 2005), 用于与外部变量做相关性研究, 如地理、气候因子。本研究中将榲栌榲树叶片坐标普氏叠印 (图 3.2A) 后分为对称组分 (图 3.2B) 和非对称组分 (图 3.2C), 对称组分主要用于分类判别方向的研究, 非对称组分主要用于与变量的相关性研究。

叶形变异 PCA 可以用来检测样本中形状变化的主要特征, 我们通过榲栌榲树及杂交个体进行 PCA 分析 (图 3.3), 研究了它们叶片形状 (对称组分和非对称组分) 变异的整体趋势。对于对称组分的 PCA 分析 (图 3.3A), 我们发现两物种叶片形状变异主要体现在 PC1 轴上, 榲栌在 PC1 上范围大致为 0~0.15, 榲树在 PC1 上范围大致为 -0.10~0, 这说明榲栌具有更广的变异范围。而对于非对称组分, 两物种的种间差异不显著 (图 3.3B)。我们利用 PCA 分析的 PC1、PC2 极值所绘制的转换网格图 (图 3.3), 可以发现榲栌榲树的叶形变化, 主要体现在叶片基部轮廓和叶片尖端以及叶片最宽位置的变化。

同域异域对叶形影响 为细化榲栌榲树的叶形状特征受同异域影响, 我们将榲栌榲树的同域和异域作为分类器, 分别进行 CVA 和 DA。CVA 结果 (图 3.7) 可以看出榲树的叶形比较稳定, 榲栌的叶形存在一定变异, 并且异域生长的榲栌

叶形更加接近槲树，有可能是在异域地区杂交渐渗的概率大于同域区域，即增强（reinforcement）作用的原因。DA 结果表明槲栎叶片在同域和异域的判别中，同域有 68.12%的叶片被正确区分，异域有 60.32%的叶片被正确区分；槲树叶片在同域和异域的判别中，同域有 68.12%的叶片被正确区分，异域有 60.32%的叶片被正确区分。两者都证明在异域种群下两物种叶形态上比同域情况更相似，推测是由于同一生境下物种竞争促使物种形态差异增大，即生态特征替代（Ecological Character Displacement, ECD）增加（Adams & Rohlf, 2000）。

环境对叶形影响 我们对于槲栎槲树受环境因子影响的研究结果表明，叶片形态的变异在与地理和气候因素中，叶片不同组分对叶形变异存在差异（图 3.9）。在叶片对称和非对称组分中，对叶形的影响因子主要是八月份降水（prec08）和平均日较差（bio15）。同时基于边际和条件的 dbRDA 分析了环境因子对叶形的影响（表 3.1），结果表明槲栎的叶片形状在边际测试中受经度影响显著，但槲树的叶形受地理和气候影响均不显著。在控制地理或气候因子后，各指标对槲栎槲树的叶形影响都不显著。说明槲树叶片形态稳定，槲栎的叶形态变异受纬度的影响较大，并且同时受各地点的小环境影响产生相应的变异。

4.3 槲栎与槲树的异速生长模式

槲栎槲树的叶形对称组分中的异速生长关系显著（图 3.6A），叶片大小在由小到大的变化中，叶形由槲栎形态过渡到槲树形态；而在非对称组分中存在显著的等距关系（图 3.6B），这与其他落叶栎树的报道一致（Viscosi, 2015）。无论是槲栎或槲树都在对称组分中存在异速生长模式，在非对称组分中为等距生长，所以异速生长模式更主要为个体间差异提供参考。陆霞梅（2007）提出，植物的特定基因型在几何学上表型差异的表达，除了异速生长以外也存在一定的环境因素的影响。环境因素还是植株大小引起植物异速生长模式还有待进一步研究。

异速生长分析还可以区分包括时间功能（function of time）和大小功能（function of size）两种不同的生态学进程（Weiner, 2004）。有些生态学进程是时间上的功能，比如生命表（life table），但是很多生态进程更容易用大小（size）来理解。但是大小和时间不能独立地进行区分，即异速生长模式与异速生长分析。任何植物都存在相应程度的异速生长模式，比如竞争等作用（低出生率、高死亡率等）就可以达到影响种群数量动态的目的，从而影响个体在生长速度和形状变化上的表达，从而影响植物体的异速生长模式。

4.4 槲栎与槲树的模块化分区

在几何形态学研究中，假设一个形态模块内由标志点构成的子集之间的关联最小，因此，通过比较子集间的协变强度，可直接评估模块化（Klingenberg, 2009）。

一般来说几何形态学分析中邻近的标志点划分为同一子集；对自身对称的结构，如左和右对称及对称轴上的标志点，成对的标志点需要在同一子集内，使划分的模块内具有强协变且模块间相对独立（Klingenberg, 2009）。

本研究基于标志点配置分析了槲栎和槲树叶片形态的模块化，将叶片对称组分和非对称组分的标志点划分为不同的子集，以探讨其几何形态的模块化分区，验证整片叶子是否为单一的模块以及两种三等分划分模块。假设分区的结果表明，槲栎（图 3.10）的对称组分和槲树（图 3.11）的对称组分以及非对称组分，叶片分为叶尖部、中下部分以及叶柄（图 2.3C）都是比较合适的模块化分区。而槲栎的非对称组分倾向于整张叶片为一个模块，叶柄为一个模块。

本研究中考虑了所有的分区，这种组合包括许多细分，比如其中一个或两个标志点子集不是连续的，这些标志点在空间上是分开的。虽然考虑标志点配置的所有可能分区的集合不是生物学上的比较基础，这种空间上不相交的标志点可能不是模块的合理候选（Klingenberg, 2009）。比如限制空间的连续性显然会减缓分区数量的增长，因为限制了配置中的标志点数量和子集数量。但本研究中的标志点及划分的模块数量都可以满足列举出全部的分区数量，也是由于 Klingenberg 建议 20 个标志点以下可以全部列举出，所得结果在形态学上的具有一定的参考意义。此外，如果形态模块的内部集成来源于其前体内的组织结合发育相互作用，它们可能与空间定义的发育单元有关（Kornberg and Guha, 2007）。

在对象对称的生物体模块研究中，Mirsha 等（2018）探讨了几个拉丁美洲混血儿面部组织的模块分区，发现了基于功能需求和面部组织的模块化信号，并提出对称和非对称模块模式之间的差异可能与左右两侧发育过程中基因表达的差异有关。我们在进行植物的二维模块化分区时，若能参考遗传和环境对形态变异的影响可能对其模块化分区有更加深入、完整的解释。

5 结论与展望

5.1 结论

叶片形状长期以来一直是众多学科关注的焦点，但对叶片形状变异的复杂遗传和环境交互作用机制的研究尚未深入。我们的研究分别探讨了地理和气候因子以及种间关系，对叶片形态变异的影响。基于多位点基因型进行物种的先验判别，并以叶形为分类性状利用几何形态学和深度学习进行物种分类，得出以下结论：

(1) 几何形态学和深度学习基于叶片形态都可以区分榲栌和榲树，但遗传鉴定的杂交个体在叶片形态上，与两亲本形态密切相关且处于中间型。

(2) 榲树叶片变异较稳定，但榲栌叶片形状在异域种群中与榲树叶片相近，在同域种群中差别较大，有可能是增强作用和生态特征替代的结果。

(3) 榲栌和榲树叶片对称组分中存在异速生长模式，叶片大小与叶片形状变异存在强相关；榲栌和榲树叶片非对称组分中存在显著的等距关系，叶片大小与叶片形状变异不相关。

(4) 榲栌的叶形受经度影响显著。榲栌和榲树的叶形变异还与降水季节性变异系数 (bio15) 和 8 月份降水 (prec08) 有关，且在不同组分上存在差异。

(5) 两种栎类叶片中存在模块化分区：榲栌（对称组分）及榲树（对称、非对称）分为叶尖部、叶中下部及叶柄三个模块，榲栌的非对称组分分为叶片和叶柄两个模块。

我们的研究为榲栌榲树的判别及叶形态特征提供了必要的信息，并为近缘种的研究提供了新的参考。

5.2 展望

本研究基于榲栌榲树两白栎组近缘种为研究对象，对其以遗传先验，几何形态学和深度学习进行物种鉴别分类，并对叶形态变异及其影响因素进行深入研究，然而还有一些问题在今后的研究中值得拓展，列举如下：

(1) 本研究进一步发现了榲栌榲树叶片对称组分的异速生长关系，以及非对称组分的等距关系。其异速生长关系不受种间同域异域的影响，这种生长关系是否受其他生物因素或者环境因素影响还需要进一步证实，并测试其保守程度。未来可通过异速生长分析叶形的可塑性，全面地了解植物叶片表型可塑性及其适应机制。

(2) 本研究利用几何形态学在叶形态方面的分析中，只考虑了地理和气候因子对叶片的对称组分和非对称组分的解释度。未来可以考虑结合种群遗传学相关数据，从遗传角度解释叶形态变异，三种因子互相作用并在叶形态上的表达。波

动不对称性在生态学和进化研究中被广泛用于量化发育不稳定性,对于植物等固着生物来说,波动不对称性可能不仅仅来源于发育不稳定性,因为响应环境异质性的表型可塑性也可能产生波动不对称性。在之后的研究中可以补充波动不对称的数据并进行相关研究。另外,同域异域对榲栌榲树叶形的影响还需从物种分化和基因流方向等角度证明。

(3) 基于 Xception 模型的深度学习算法对栎类叶形态鉴别中,两近缘种榲栌和榲树的区分效果比较好,而对于我国广泛分布的栎类来说,与榲栌榲树同为白栎组的其他五个种在叶片形态上也相近,深度学习模型是否应用在白栎组的分类中还需进一步验证。考虑到其叶片形态的相似程度,可以结合其他器官进行物种识别,比如花型、枝干和种子以及花粉在显微镜下的成像。后续的研究中可扩大训练和测试的数据,并进行数据增广,如添加镜像数据和缩放数据等,使运行结果更加精准。同时可以引入其他分类模型比较分类精度。

参考文献

- 白明, 杨星科. 几何形态测量法在生物形态学研究中的应用[J]. 昆虫知识. 2007,44(01): 143-147.
- 陈君. 中国白栎组单拷贝核基因标记的开发及系统发育关系研究[D]. 中国林业科学研究院硕士学位论文. 2017.
- 陈世骧. 关于物种定义[J]. 动物分类学报. 1979, 4(04): 425-426.
- 陈世骧. 物种概念与分类原理[J]. 中国科学. 1983, (04): 315-320.
- 方伟, 刘恩德. 经典植物分类学的发展与 iFlora. 植物分类与资源学报. 2012, 34(6): 532-538.
- 葛德燕, 夏霖, 吕雪霏, 黄乘明, 杨奇森, 黄建华. 几何形态学方法及其在动物发育与系统进化研究中的应用[J]. 动物分类学报. 2012, 37(02): 296-304.
- 韩亚琦, 唐宇丹, 张少英, 李向应, 何志. 盐胁迫抑制榲栌 2 变种光合作用的机理研究[J]. 西北植物学报, 2007(03): 583-587.
- 李银霞, 高其洁, 李天红. 基于果实相关性状的桃品种初级核心种质取样策略研究. 果树学报. 2006, 23(3): 359-364.
- 李月娟. 通过几何形态学方法研究川滇高山栎和刺叶高山栎的叶片形态变异[D]. 北京林业大学硕士学位论文. 2020.
- 刘峰, 陈伟烈, 贺金生. 神农架地区锐齿榲栌种群结构与更新的研究[J]. 植物生态学报, 2000(04): 396-401.
- 刘媛. 同域分布栎属两近缘种形态变异与种群遗传研究[D]. 北京林业大学硕士学位论文. 2018.
- 陆霞梅, 周长芳, 安树青, 方超, 赵晖, 杨茜, 颜超. 植物的表型可塑性、异速生长及其入侵能力[J]. 生态学杂志. 2007, 26(09): 1438-1444.
- 彭焱松, 陈丽, 李建强. 中国栎属植物的数量分类研究[J]. 武汉植物学研究. 2007, 25(02): 149-157.
- 邱丽胤, 柳涛, 常虹, 吴晓媛. 中国壳斗科植物空间多样性格局研究. 植物科学学报. 2017, 35(2): 234-242.
- 谢晋阳, 陈灵芝. 暖温带落叶阔叶林的物种多样性特征[J]. 生态学报, 1994(04): 337-344.
- 余凯, 贾磊, 陈雨强, 徐伟. 深度学习的昨天、今天和明天[J]. 计算机研究与发

- 展,2013,50(09):1799-1804.
- 左羽,陶倩,吴恋,等. 基于卷积神经网络的植物图像分类方法研究[J]. 物联网技术, 2020, 010(003):72-75.
- Abràmoff M D, Magalhães P J, Ram S J. Image Processing with ImageJ[J]. Biophotonics International. 2004, 11(7): 36-43.
- Adams D C, Rohlf F J. Ecological Character Displacement in Plethodon: Biomechanical Differences Found from a Geometric Morphometric Study[J]. Proceedings of the National Academy of Sciences of the United States of America, 2000, 97(8): 4106-4111.
- Amos W, Hoffman J I, Frodsham A, Zhang L, Best S, Hill A. Automated binning of microsatellite alleles: problems and solutions[J]. Molecular Ecology Resources. 2007, 7(1): 10-14.
- Ann E K, Kenneth A F, Doran M M, Robert E U, William W T. Compartments revealed in food-web structure[J]. Nature: International weekly journal of science. 2003, 426(6964): 282-285.
- Ashburner J, Friston K J. Voxel-based morphometry—the methods[J]. Neuroimage. 2000, 11(6): 805-821.
- Atchley W R, Hall B K. A model for development and evolution of complex morphological structures[J]. Biological Reviews of the Cambridge Philosophical Society. 1991, 66(2): 101-157.
- Bookstein F L, Gunz P, Mitteroecker H, Prossinger H, Schaefer K, Seidler H. Cranial integration in Homo: singular warps analysis of the midsagittal plane in ontogeny and evolution[J]. Journal of Human Evolution, 2003, 44(2):167-187.
- Bookstein F L. Thin-plate splines and the Atlas problem for biomedical images[C]. Springer Berlin Heidelberg, 1991.
- Bookstein F L. Biometrics, biomathematics and the morphometric synthesis[J]. Bulletin of Mathematical Biology, 1996, 58(2): 313-365.
- Bookstein F L. Landmark methods for forms without landmarks: Morphometrics of group differences in outline shape[J]. Medical Image Analysis. 1997, 1(3): 225-243.
- Bookstein F L. Procrustes Distance, bending Energy, and the biometrical Modeling of Shape Phenomena[J]. International Journal of Shape Modeling. 1997, 3(1-2): 17-38.
- Cheverud J M, Richtsmeier J T. Finite-element scaling applied to sexual dimorphism in Rhesus Macaque (*Macaca Mulatta*) facial growth[J]. Systematic Zoology. 1986, 35(3): 381-399.
- Cheverud J M. Developmental Integration and the Evolution of Pleiotropy[J]. Integrative and

-
- Comparative Biology, 1996, 36(1): 44-50.
- Cheverud J M. Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution[J]. 1982, 36(3): 499-516.
- Claudia E M, Gonzalo C C, Verdú J R. Taxonomic Diversity as Complementary Information to Assess Plant Species Diversity in Secondary Vegetation and Primary Tropical Deciduous Forest[J]. Journal of Vegetation Science, 2009, 20(5): 935-943.
- Darroch J N, Mosimann J E. Canonical and principal components of shape[J]. Biometrika. 1985, 2(2): 241-252.
- Darwin C R. On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life[J]. Nature, 1872, 144(11), 971-972.
- Dryden I L, Mardia K V. 1998. Statistical Shape Analysis [J]. John Wiley & Sons, Ltd. 2010, 213(6): 663- 669.
- Dunker S, Motivans E, Rakosy D, Boho D, Knight TM. Pollen analysis using multispectral imaging flow cytometry and deep learning[J]. New Phytologist, 2020, 229(1).
- Durand J M, Bodénès C, Chancerel E, Daruvar A, Kremer A, Plomion C, et al. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study[J]. BioMed Central, 2010, 11(1).
- Earl D A, Vonholdt B M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method[J]. Conservation Genetics Resources. 2012, 4(2): 359-361.
- Escoufier Y. Le traitement des variables vectorielles[J]. Biometrics. 1973, 29(4): 751-760.
- Evanno G S, Regnaut S J, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study[J]. Molecular Ecology. 2005, 14(8): 2611-2620.
- Garud R, Kumaraswamy A, Langlois R. Managing in the Modular Age: Architectures, Networks, and Organizations[M]. Bristol: Blackwell, 2002.
- Gerhard Schlosser. Modularity and the units of evolution[J]. Theory in Biosciences. 2002, 121(1): 1-80.
- Girvan M, Newman M E. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences. 2002, 99(12): 7821-7826.
- Glaubitz J C. CONVERT: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages[J]. Molecular Ecology Resources. 2004, 4(2): 309-310.
- Goswami A. Cranial modularity shifts during mammalian evolution[J]. American Naturalist. 2006a,

168: 270-280.

Goswami A. Morphological integration in the carnivoran skull[J]. *Evolution*. 2006b, 60: 169-183.

Gunz P. Using semilandmarks in three dimensions to model human neurocranial shape[D]. Vienna: University of Vienna, 2001.

Hinton G E, Osindero S, The Y W. A Fast Learning Algorithm for Deep Belief Nets[J]. *Neural Computation*, 2006,18(7): 1572-1554.

Hipp A L, Mano P S, Hahn M, Avishai M , Bodénès C, Cavende-Bares J, Crowl A A, Deng M, Denk T, Fitz-Gibbon S. Genomic landscape of the global oak phylogeny. *New Phytologist*, 2020, 226(4).

Hubert F, Grimm G W, Jousselin E, Berry V, Franc A, Kremer A. Multiple nuclear genes stabilize the phylogenetic backbone of the genus *Quercus*[J]. *Systematics & Biodiversity*, 2014, 12(4):405-423.

Hulce D, Li X, Snyderleiby T, Liu C S. GeneMarker® Genotyping Software: Tools to Increase the Statistical Power of DNA Fragment Analysis[J]. *Journal of Biomolecular Techniques*. 2011, 22(Suppl): S35.

Huxley J S. Constant differential growth-ratios and their significance[J]. *Nature*. 1924, 115(2883): 155-156.

Huxley J S. *Problems of relative growth*[M]. London and New York: Methuen and Co. 1932.

James M. Cheverud. Developmental Integration and the Evolution of Pleiotropy[J]. *American Zoologist*. 1996, 36(1): 44-50.

Janes J K, Miller J M, Dupuis J R, Malenfant R M, Gorrell J C, Cullingham C I, Andrew R L. The $K=2$ conundrum[J]. *Molecular Ecology*. 2017, 26(14): 3594-3602.

Klápště J, Kremer A, Burg K, Garnier G P, El-Dien O G, Ratcliffe B, El-Kassaby Y A, Porth I. *Quercus* species divergence is driven by natural selection on evolutionarily less integrated traits[J]. *Heredity*, 2020, 126(2).

Klingenberg C P, Badyaev A V, Sowry S M, Beckwith N J. Inferring developmental modularity from morphological integration: analysis of individual variation and asymmetry in bumblebee wings[J]. *American Naturalist*. 2001, 157(1): 11-23.

Klingenberg C P, Barluenga M, Meyer A. Shape Analysis of Symmetric Structures: Quantifying Variation among Individuals and Asymmetry[J]. *Evolution*. 2002, 56(10): 1909-1920.

Klingenberg C P, Duttke S, Whelan S, Kim M. Developmental plasticity, morphological variation

- and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves[J]. *Journal of Evolutionary Biology*. 2015, 25(1): 115-129.
- Klingenberg C P, Leamy L J, Cheverud J M. Integration and modularity of quantitative trait locus effects on geometric shape in the mouse mandible[J]. *Genetics*. 2004; 166(4): 1909-1921.
- Klingenberg C P, Leamy L J. Quantitative genetics of geometric shape in the mouse mandible[J]. *Evolution*. 2001, 55(11): 2342-2352.
- Klingenberg C P, Mebus K, Auffray J C. Developmental integration in a complex morphological structure: how distinct are the modules in the mouse mandible[J]? *Evolution and Development*. 2003, 5(5): 522-531.
- Klingenberg C P, Monteiro L R. Distances and directions in multidimensional shape spaces: implications for morphometric applications[J]. *Systematic Biology*. 2005, 54(4): 678–688.
- Klingenberg C P, Stefanie D, Zaklan K. Morphological integration between developmental compartments in the drosophila wing. *Evolution*. 2010, 54(4): 1273-1285.
- Klingenberg C P, Zaklan S D. Morphological integration between developmental compartments in the *Drosophila* wing[J]. *Evolution*. 2000, 54(4): 1273-1285.
- Klingenberg C P. Heterochrony and allometry: the analysis of evolutionary change in ontogeny[J]. *Biological Reviews of the Cambridge Philosophical Society*. 1998, 73(1): 79-123.
- Klingenberg C P. Morphological Integration and Developmental Modularity[J]. *Annual Review of Ecology, Evolution, and Systematics*. 2008, 39: 115-132.
- Klingenberg C P. Morphometric integration and modularity in configurations of landmarks: Tools for evaluating a-priori hypotheses[J]. *Evolution and Development*. 2009, 11(4): 405–421
- Klingenberg C P. Studying morphological integration and modularity at multiple levels: concepts and analysis[J]. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2014, 369(1649): 20130249.
- Klingenberg P C. Size, shape, and form: concepts of allometry in geometric morphometrics[J]. *Development Genes and Evolution*. 2016, 226(3): 113-137.
- Kornberg T B, Guha A. Understanding morphogen gradients: a problem of dispersion and containment[J]. *Current Opinion in Genetics & Development*. 2007,17(4): 264–271.
- Kremer A, Dupouey J L, Deans J D, Munro R, Steinhoff S, Badeau V. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak stands[J]. *Annals of Forest Science*. 2002, 59(7): 777-787.
- Lecun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L. Backpropagation

- Applied to Handwritten Zip Code Recognition[J]. *Neural Computation*, 2014, 1(4):541-551.
- Lecun Y, Bottou L. Gradient-based learning applied to document recognition[J]. *Proceedings of the IEEE*, 1998, 86(11):2278-2324.
- Legendre P, Anderson M J. Distance-based redundancy analysis: testing multi-species responses in multi-factorial ecological experiments[J]. *Ecological Monographs*. 1999, 69: 1-24.
- Lele S, Richtsmeier J T. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data[J]. *American Journal of Physical Anthropology*. 1991, 86(3): 415-427.
- Liu Y, Li Y, Song J, Zhang R, Yan Y, Wang Y, Du F K. Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae)[J]. *Annals of Forest Science*. 2018, 75(4): 90.
- Liu Z, Wang J, Tian Y, Dai S L. Deep learning for image-based large-flowered chrysanthemum cultivar recognition[J]. *Plant Methods*, 2019, 15: 146.
- Lyu J, Song J, Liu Y, Wang Y Y, Li J Q, Du F K. Species Boundaries Between Three Sympatric Oak Species: *Quercus aliena*, *Q. dentata*, and *Q. variabilis* at the Northern Edge of Their Distribution in China. [J]. *Frontiers in plant science*, 2018, 9: 414.
- MacLeod N. Geometric morphometrics and geological shape-classification systems[J]. *Earth-Science Reviews*. 2002, 59(1): 27-47.
- Mardia K V, Bookstein F L, Moreton I J. Statistical Assessment of Bilateral Symmetry of Shapes[J]. *Biometrika*. 2005, 92(1): 249-250.
- Mikula O, Machola 'n M. There is no heterotic effect upon developmental stability in the ventral side of the skull within the house mouse hybrid zone[J]. *Journal of Evolutionary Biology*. 2008, 21: 1055–1067
- Mirsha Q S, Francese M M, Jorge G V, Javier R, Andres R L, Ro G J, et al. Developmental pathways inferred from modularity, morphological integration and fluctuating asymmetry patterns in the human face[J]. *Scientific Reports*. 2018, 8(963): 1-15.
- Oksanen J, Blanchet F G, Kindt R, Legendre P, Minchin P, O'Hara R B, Simpson G, Solymos P, Stevens M, Wagner H. *vegan: Community ecology package*[J]. R package version. 2017.
- Olesen J M, Bascompte J, Dupont Y L, Jordano P. The Modularity of Pollination Networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*. 2007, 104(50): 19891-19896.
- Olson E C, Miller R L. Morphological Integration[J]. *Evolution*. 1958, 12(4): 558-559.
- Pigliucci M. Phenotypic integration: studying the ecology and evolution of complex phenotypes[J]. *Ecology Letters*. 2010, 6(3): 265-272.

- Ponton S, Dupouey J L, Dreyer E. Leaf morphology as species indicator in seedlings of *Quercus robur* L. and *Q. petraea* (Matt.) Liebl.: Modulation by irradiance and growth flush[J]. *Annals of Forest Science*. 2004, 61: 73-80
- Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data [J]. *Genetics*. 2000, 155(2): 945-959.
- Ravasz E, Somera L A, Mongru A D, Oltvai N Z, Barabási A L. Hierarchical Organization of Modularity in Metabolic Networks[J]. *Science*. 2002, 297(5586): 1551-1555.
- Robert P, Escoufier Y. A unifying tool for linear multivariate statistical analysis: the RV-coefficient[J]. *Annals of Applied Statistics*. 1976, 25(3): 257–265.
- Rohlf F J, Brookstein F L. Proceedings of the michigan morphometrics workshop[J]. *Systematic Biology*. 1992, 41(3): 392-395.
- Rohlf F J, Corti M. Use of Two-Block Partial Least-Squares to Study Covariation in Shape[J]. *Systematic Biology*. 2000, (4): 740-753.
- Rohlf F J. Bias and error in estimates of mean shape in geometric morphometrics[J]. *Journal of Human Evolution*. 2003, 44(6): 665-683.
- Rohlf F J. Statistical power comparisons among alternative morphometric methods[J]. *American Journal of Physical Anthropology*. 2000, 111(4): 463-478.
- Rosenberg N A. distruct: a program for the graphical display of population structure[J]. *Molecular Ecology Resources*. 2004, 4(1): 137-138.
- Salakhutdinov R, Hinton G. Deep Boltzmann machines. *Journal of machine learning research proceedinds track*. 2009, 9(1): 448-455.
- Savriama Y, Gómez J M, Perfectti F, Klingenberg C P. Geometric morphometrics of corolla shape: dissecting components of symmetric and asymmetric variation in *erisimum mediohispanicum* (brassicaceae) [J]. *New Phytologist*, 2012, 196:945-954.
- Stephan J M, Teeny P W, Vessella F, Schirone B. Oak morphological traits: Between taxa and environmental variability[J]. *Flora*,2018, 243: 32-44.
- Thomas J, Manuel P J, Salvador B, Luis M J. Quantitative trait loci mapping of floral and leaf morphology traits in *Arabidopsis thaliana*: evidence for modular genetic architecture[J]. *Evolution and development*. 2005, 7(3): 259-271.
- Tseng Z J, Stynder D. Mosaic functionality in a transitional ecomorphology: skull biomechanics in stem Hyaeninae compared to modern South African carnivorans[J]. *Biological Journal of the Linnean Society*. 2011, 102(3): 540-559.

- Tucić B, Budečević S, Jovanović S M, Vuleta A, Klingenberg C P. Phenotypic plasticity in response to environmental heterogeneity contributes to fluctuating asymmetry in plants: first empirical evidence[J]. *Journal of Evolutionary Biology*, 2018, 31(2).
- Ueno S, Taguchi Y, Tsumura Y. Microsatellite markers derived from *Quercus mongolica* var. *crispula* (Fagaceae) inner bark expressed sequence tags[J]. *Genes and Genetic Systems*. 2008, 83(2): 179-187.
- Vähä J P, Primmer C R. Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci[J]. *Molecular Ecology*. 2006, 15(1): 63-72.
- Viscosi V, Antonecchia G, Lepais O, Fortini P, Gerber S, Loy A. Leaf Shape and Size Differentiation in White Oaks: Assessment of Allometric Relationships among Three Sympatric Species and Their Hybrids[J]. *International Journal of Plant Sciences*. 2012, 173(8): 875-884.
- Viscosi V, Cardini A. Leaf Morphology, Taxonomy and Geometric Morphometrics: A Simplified Protocol for Beginners[J]. *Plos One*. 2011, 6(10).
- Viscosi V. Geometric morphometrics and leaf phenotypic plasticity: assessing fluctuating asymmetry and allometry in European white oaks (*Quercus*)[J]. *Botanical Journal of the Linnean Society*. 2015, 179(2): 335-348.
- von Dassow G, Meir E, Munro E M, Odell G M. The segment polarity network is a robust developmental module[J]. *Nature*. 2000, 406(6792): 188-192.
- Wagner G P, Pavlicev M, Cheverud J M. The road to modularity[J]. *Nature Reviews Genetics*. 2007, 8(12): 921-931.
- Wei T J, Chang S W, Sameem B, Jen Y H, Yong K T. Deep Learning for Plant Species Classification using Leaf Vein Morphometric[J]. *IEEE-ACM Transactions on Computational Biology and Bioinformatics*, 2018: 1545-5963.
- Weiner J. Allocation, plasticity and allometry in plants. *Perspect Plant Ecol* 6:207-215[J]. *Perspectives in Plant Ecology Evolution and Systematics*. 2004, 6(4): 207-215.
- Wolsan M, Suzuki S, Asahara M, Motokawa M. Dental integration and modularity in pinnipeds[J]. *Scientific Reports*, 2019, 9(1): 4184.
- Yang J, Guo Y F, Chen X D, Zhang X, Ju M M, Bai G Q, Liu Z L, Zhao G F. Framework phylogeny, evolution and complex diversification of Chinese oaks[J]. *Plants*. 2020, 9(8): 1024.
- Zelditch M L, Lundrigan B L, Garland T. Developmental regulation of skull morphology. I. Ontogenetic dynamics of variance. [J]. *Evolution & development*, 2004, 6(3): 194-206.
- Zelditch M L, Swiderski D L, Sheets H D. *Geometric morphometrics for biologists*[M]. Amsterdam:

Elsevier. 2012.

Zelditch M L, Wood A R, Bonett R M, Swiderski D L. Modularity of the rodent mandible: integrating bones, muscles, and teeth[J]. *Evolution and Development*. 2008, 10: 756-768.

个人简介

个人简介

苏蔚，女，1995年12月生，籍贯黑龙江省鸡西市。2013年9月至2017年7月就读于齐齐哈尔大学生命科学与农林学院园林专业，取得农学学士学位。2018年9月至2021年6月就读于北京林业大学生态与自然保护学院生态学专业，导师为杜芳教授，研究方向为恢复生态学。

硕士期间，主修了高级森林生态学、生态学研究方法、生态学数据分析—R语言、多元统计分析和硕士英语等课程，成绩良好。硕士期间主要研究内容为栎属近缘物种叶片形态特征等相关性研究，并以第二作者的发明专利1篇，并以第一作者发表CSCD学术论文一篇，以第四作者发表SCI学术论文一篇。

导师简介

杜芳，女，1981年11月生，甘肃甘南人。2010年6月获得兰州大学生态学博士学位，2010年12月获得法国波尔多第一大学生态学博士学位。现任北京林业大学生态与自然保护学院教授和云南大学生态与环境学院客座教授。

主要致力于森林树种遗传多样性及物种形成的机理研究。主持国家自然科学基金项目三项，参与国家自然科学基金重点项目和科技部重点研发计划。以第一作者或通讯作者发表SCI论文21篇，发明专利1项，论文引用700余次（谷歌学术统计）。2014年在北京林业大学创建了首个分子生态实验室，为从微观角度解决宏观森林生态学问题搭建了良好的平台。

2009年获法国农业科学学院三大科研进展奖励；2014年获得教育部高等学校科学研究优秀成果奖自然科学奖一等奖（排名第四）；2014、2018年获得梁希青年论文奖；2015年当选北京市科技新星；2018年起担任国际栎树协会专业委员会成员和北京生态修复协会理事；2019年当选中国植物学会女科学家协会会员。2021年开始担任SCI期刊*Ecology & Evolution*的Associate Editor。

成果目录清单

[1] Yang Xu, Hao Chen, Min Qi, **Wei Su**, Yue Zhang, Fang K. Du. The complete chloroplast genome of *Quercus fabri* (Fagaceae) from China [J]. *Mitochondrial DNA Part B*, 2019, 4(2): 2857-2858.

[2] **苏蔚**, 宋以刚, 祁敏, 杜芳. 基于几何形态分析的栎属白栎组叶片形态特征 [J/OL]. *应用生态学报*. <https://doi.org/10.13287/j.1001-9332.202107.001>. (已被应用生态学期刊录用, CSCD 源刊)

[3] 杜芳, **苏蔚**, 李迎超, 尹康权. 天女木兰的快速繁育方法 [P]. 北京市: CN110169359B, 2021-03-23.

致谢

学业生涯将圆满结束，在林业大学的校园里，我度过了三年的美好时光，结识了许多的良师益友，这都是我一生受用的精神财富。在此，我要对我的父母、师长、亲友、同窗表示深深的感谢。

首先，我要感谢我的父母，是他们的生育之情、养育之恩才有了现在的我，他们一直默默的支持与关怀，理解我包容我，每一次的关心问候都是我前行的动力。

其次，特别要感谢我的导师杜芳教授，这篇论文的完成离不开杜老师的心血和劳动，从论文的选题与结构到核心都无不透露着杜老师的悉心指导。为了使文章能够更加完善，老师经常给我们反复修改、推敲，才有现在的论文终稿的成型。

再者，需要感谢南卡罗来纳大学唐继军老师，天津大学计算机科学与技术学院郭菲老师及宋洁对我在深度学习上的帮助。感谢林科院李迎超老师、北京林业大学草业与草原学院尹康权老师对植物组织培养上的指导。感谢林科院黄继红老师、北京林业大学刘艳红、李景文、侯继华老师对论文提出了宝贵的指导意见，以及感谢百忙之中匿名对本文进行审阅、评议的各位老师。感谢祁敏帮助分子实验，感谢许阳、师兄王天瑞和师弟刘可可帮助采样，师姐李月娟指导分析数据。感谢杜芳老师课题组全体成员在论文完成过程中的帮助。感谢北林 10#1723 宿舍成员在生活上的陪伴支持与帮助。感谢身边的挚友们，你们的理解和关心，你们的陪伴，使我的生活丰富多彩。

最后，我要感谢北京林业大学，美丽的校园使我度过了美好的三年研究生生活。