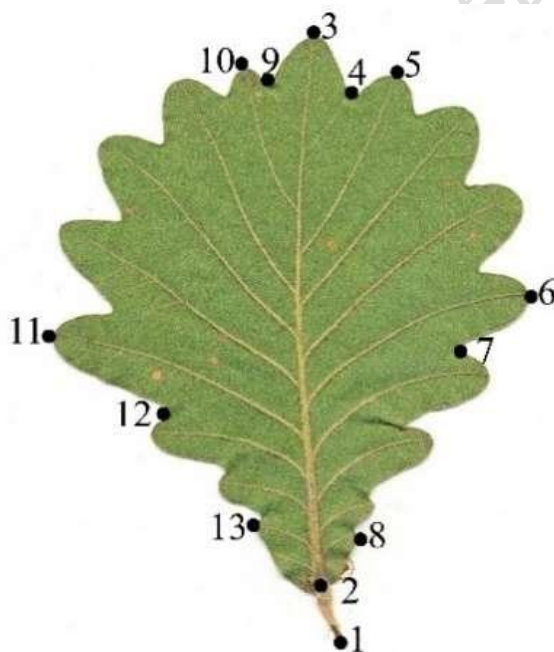




北京林业大学

杜芳课题组叶片形态分析流程

Leaf morphology analysis process in
the Du Fang's lab



杜芳，李月娟，祁敏，苏蔚，张艺

本资料仅限分子生态实验室内交流，有问题

请邮件至杜芳：dufang325@bjfu.edu.cn

最后更新: 4/24/2025

The analysis software obtained from open sources, please cite the relevant websites if you need to use them (relevant references are at the end of the documents).

software download address:

Image J: <https://imagej.net/ij/download.html>

Morpho J:

https://morphometrics.uk/MorphoJ_page.html

R: <https://www.r-project.org/>

VueScan: <https://www.hamrick.com/>

目录

一、叶型扫描(CanoScan 5600 F 扫描仪).....	1
1.安装 VueScan 软件	1
2.连接扫描仪	3
3.放置待扫描叶片	3
4.扫描叶片	3
5.保存图片	3
二、叶型标志点(ImageJ 软件).....	4
1.功能.....	4
2.下载地址.....	4
3.操作步骤.....	4
3.1 导入图像(Import image).....	4
3.2 标志点采集(Landmarks picking).....	5
3.3 标志点坐标数据处理 (Data processing for landmark coordinates).....	6
三、叶型分析(Morpho J 软件).....	7
1.功能.....	7
2.下载地址.....	7
3.操作步骤.....	7
3.1 导入文件格式(Input file formats).....	8
3.2 导入文件(Import file)	8
3.3 前期数据处理(Preliminary data processing).....	9

3.3.1 普式拟合(Procrustes fit)	9
3.3.2 查找异常值 (Find Outliers)	11
3.3.3 分离对称和非对称性 (Separate symmetry and asymmetry)	11
3.3.4 提取新分类层(Extract New Classifier)	14
3.3.5 生成新的子样本集(Generate new subsamples)....	16
3.3.6 构建线框(Create wireframes)	18
3.4 多变量分析(Multivariate analysis)	19
3.4.1 主成分分析(Principal component analyses)	19
3.4.2 典型变量分析(Canonical variate analysis)	29
3.4.3 判别分析(Discriminant analysis)	32
3.4.4 偏最小二乘法 (Two-Block Partial Least-Square analysis, 2B-PLS)	34
3.4.5 普式方差分析(Procrustes ANOVA)	37
3.5 导出数据(Export data)	39
四、检测叶形差异是否显著	41
五、地理、遗传和气候对叶形的影响	42
1.获取叶形、地理、遗传及气候因子	42
2.叶形与地理、遗传和气候的冗余分析	42
六、导入协变量(Import Covariates)	46
1.导入文件格式(Input file formats)	46
2.导入协变量	47

3.双标图(Biplot).....	49
七、映射到系统发育树(Map Onto Phylogeny) .	51
1.导入系统发育树	51
2.独立变量在系统发育上映射	52
八、假定模块化评估 (Modularity: Evaluate Hypothesis)	56
1.选择标志点子集	56
2.修改邻接图	57
3.比较分区(Comparing partitions).....	59
九、深度学习(Deep Learning)	60
1.选择体系结构	61
2.选择框架.....	61
3.选择模型.....	62
4.获取训练集	62
5.具体操作步骤:	62
5.1 准备软件	62
5.2 数据准备及处理	63
5.2.1 准备用于分类的图片	63
5.2.2 修改运行命令	64
5.3 数据上传及运行	66
5.3.1 数据上传	66
5.3.2 数据运行	66

十、测量叶片传统性状指标	67
1. 基础指标	67
1.1 批量导入图像	67
1.2 测量叶长	68
1.3 像素转换	69
1.4 其余叶性状指标	70
2. 叶脉密度(Leaf Vein Density).....	71
2.1 一级叶脉密度(1° vein density)	72
2.2 二级叶脉密度(2° vein density)	72
2.3 小脉密度(Minor vein density)	73
3. 比叶面积(Specific Leaf Area).....	78
3.1 导入图像(Import image).....	79
3.2 参数设置(Parameter Setting).....	79
3.3 叶片干重(Dry Leaf Weight).....	81
3.4 比叶面积(Specific Leaf Area).....	81
3.5 批量计算图片中叶片大小（叶面积、叶最大直径）	82
十一、气候因子与叶性状的线性关系	82
1. 获取叶性状数据	82
2. 获取气候数据	82
3. 叶性状与气候因子的相关分析	83
十二、R 语言常用的作图方法	85

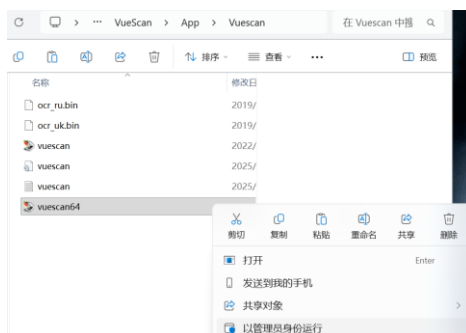
1. 直方图.....	85
2. 箱型图.....	85
3. 散点图加折线图	86
参考文献.....	87

北林分子生态实验室

一、叶型扫描(CanoScan 5600 F 扫描仪)

1.安装 VueScan 软件

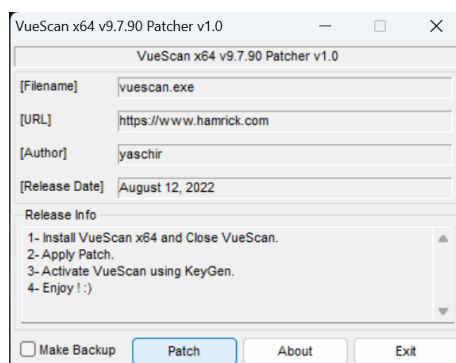
在 VueScan 官方网站 <https://www.hamrick.com/> 下载 VueScan 安装包，解压缩，右键单击 vuescan64，选择以管理员身份运行软件，



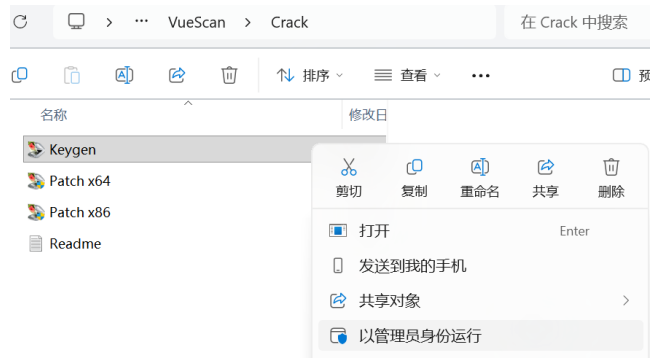
出现下图未注册界面，直接关闭该界面



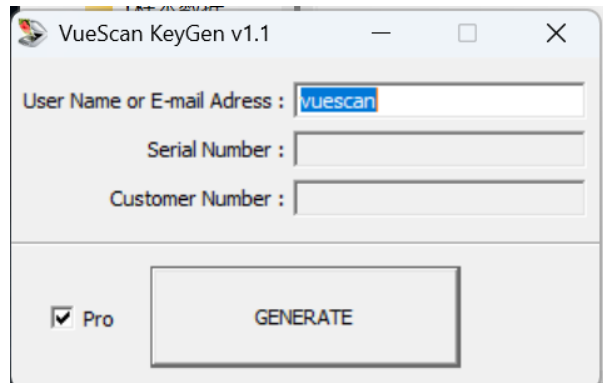
右键单击 Patch x64，选择以管理员身份运行软件，出现下图界面，点击 Patch



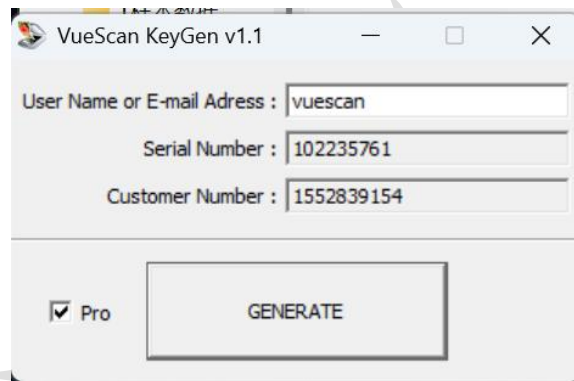
右键单击 Keygen，选择以管理员身份运行软件，



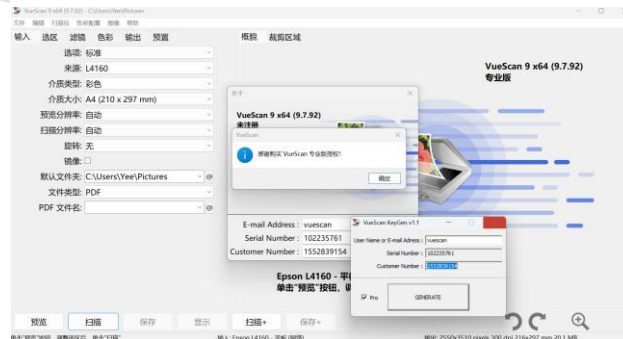
出现下图界面



点击 GENERATE，生成注册序列号，



打开软件进行复制注册，即可使用



2.连接扫描仪

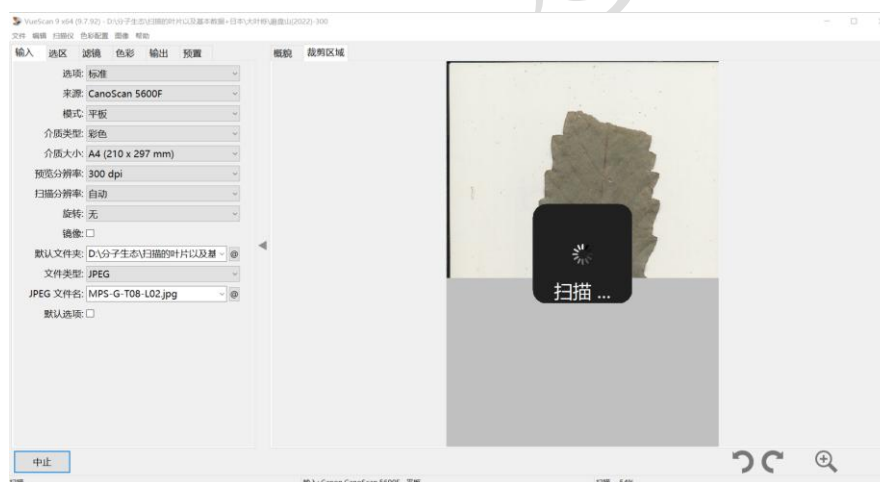
将扫描仪放置在一个平稳的桌面上，连接扫描仪的电源线并接通电源，使用扫描仪附带的 USB 数据线，一端连接扫描仪，另一端连接电脑的 USB 端口。

3.放置待扫描叶片

掀起扫描仪的盖板，将压制干燥后的叶片叶背朝下贴在扫描仪的玻璃板上，轻轻盖回盖板

4.扫描叶片

打开 vuescan 软件，进行叶片扫描，参数设置参考下图，分辨率选择 600dpi 或 300dpi



5.保存图片

扫描完成后图片会自动保存，但应去设置的保存文件地址查看图片是否保存成功

二、叶型标志点(ImageJ 软件)

1.功能:通过对样本进行标志点标记来测量面积、平均值、标准偏差、最大值、最小值、角度以及长度。

2.下载地址: <https://imagej.net/ij/download.html>

3.操作步骤:

3.1 导入图像(Import image)

选择“File”→“Import”→“Image Sequence”选定并打开一张叶片图像(图 1, 2)。

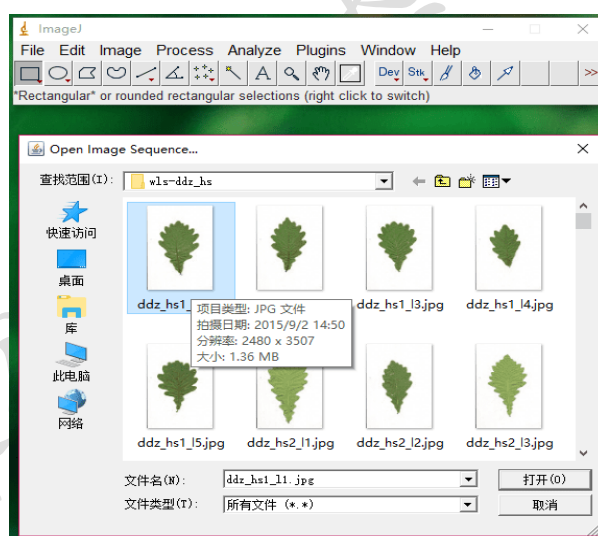


图 1 选择准备导入的叶片图像

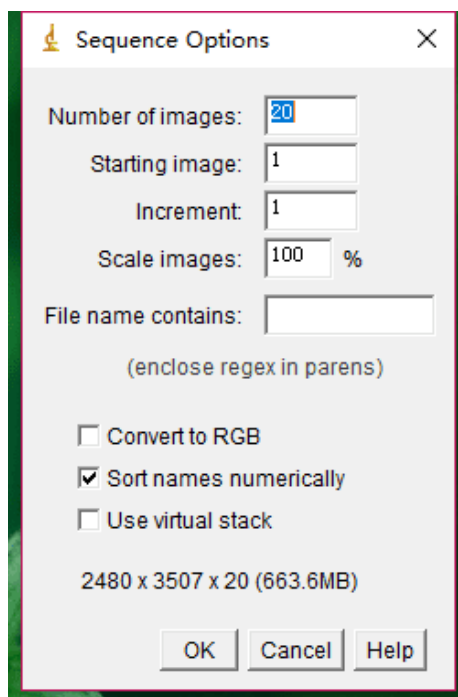


图 2 编辑导入图像数

3.2 标志点采集(Landmarks picking)

打开“Point Picker”：选择“Plugins”→“PointPicker”，在出现的工具条上选择带“+”钢笔头形状的工具在叶片的右键标点(图 3)(可用放大镜工具调整图像大小)。注意：给每一片叶子标点的时候务必按照统一的顺序,如果想调整顺序可以选择带“▼”的钢笔头形状的工具(图 3)。当标好所有的点之后，点击输出图标(纸张形状)，然后点击“show”选项，选中所有数据(图 4)后，复制并粘贴到 Excel 中，空出最上一行和最下一行。

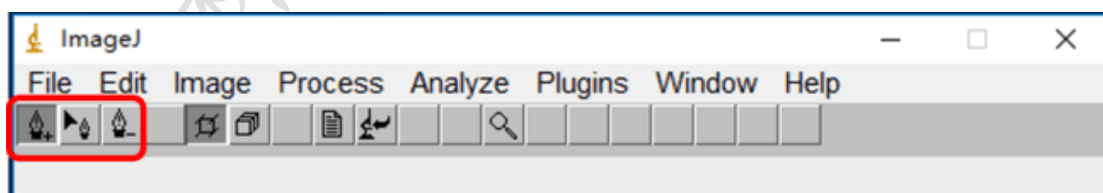


图 3 钢笔工具



Log

File	Edit	Font			
0	1197	628	1	11	12
1	1321	728	1	12	13
2	1501	872	1	13	14
3	1549	1136	1	14	15
4	1621	1685	1	15	16
5	1553	2033	1	16	17
6	1313	2414	1	17	18
7	1088	2321	1	18	19
8	916	1893	1	19	20
9	728	1385	1	20	21
10	824	1040	1	21	22
11	1004	852	1	22	23
12	1173	724	1	23	24

图 4 标志点标记结果

3.3 标志点坐标数据处理(Data processing for landmark coordinates)

在电子表格中,标志点的 x 坐标值和 y 坐标值分别为第 2 列和第 3 列,在 x 坐标值数列的上方一行输入文本“LM=”,其后为标志点的数目(比如:LM=13)并在 x 坐标值数列的下方一行输入文本“ID=”,后跟分类群的名字(图 5,比如:ID=LZD_T_T01_L01, LZD 代表种群, T 代表物种或分类组(当研究同一个物种时,按不同组进行分类), T01 是树的棵数, L01 叶片数)。

	A	B
1	LM=13	
2	1260	2878
3	1276	2847
4	1232	2289
5	1283	2297
6	1346	2297
7	1511	2546
8	1487	2685
9	1459	2743
10	1207	2293
11	1188	2295
12	1052	2546
13	1140	2753
14	1176	2794
15	ID=LZD_T_T01_L01	

图 5 标志点坐标数据整理

关闭数据窗口，点击带“-”钢笔头形状的工具在叶片的右键标点清除所有标点(图 3)。向下滑动鼠标翻到下张图片，重复以上步骤直到标完所有图像。

三、叶型分析(Morpho J 软件)

1.功能：基于普式叠印法去除二维或三维标志点坐标的非形态变异后，可以进行许多重要类型的几何形态分析 (geometric morphometric analyses, GMMs) ，包括识别离群值，生成协方差矩阵，进行主成分分析(Principal component analyses (PCA))、典型变量分析(Canonical Variate Analysis (CVA))、偏最小二乘法(Two-Block Partial Least-Square (2B-PLS))等分析操作，同时可以对形状变化的统计结果形成可视化的图形 (Klingenberg 2011)。

2.下载地址：https://morphometrics.uk/MorphoJ_page.html

3.操作步骤：

3.1 导入文件格式(Input file formats)

用于后续叶形分析软件 Morpho J 的数据输入格式为 txt.格式，所以在进行分析数据之前需要将标志点的坐标数据文件转化成 Morpho J 支持的文件类型(图 5)。

3.2 导入文件(Import file)

安装并打开 Morpho J 后，在“File”中选择“Create New Dataset”，出现图 6 对话框，选中“TPS”格式，找到标志点数据保存位置，最后点击 Create Dataset 导入数据。

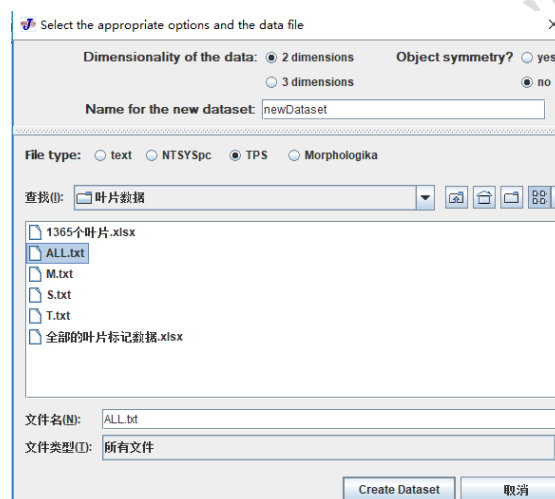


图 6 文件导入

如果数据成功导入出现图 7 对话框：

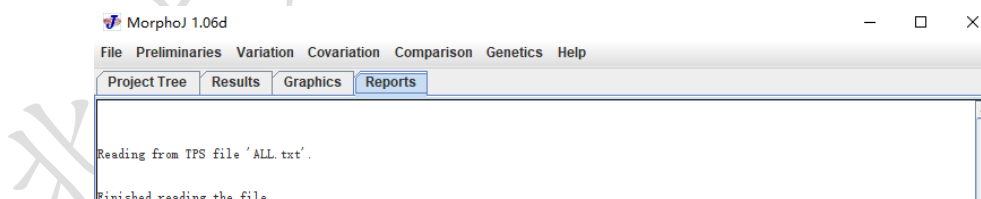


图 7 数据导入成功提示

如果数据导入失败，会出现图 8 对话框，具体信息如下：

Problems with reading the file.

The number of dimensions in some entry/entries in record 1471 appears to be inconsistent with the number of dimensions.

Errors occurred in the reading process.

The new dataset could not be created.

查找数据集出现错误的方法是用给出的行数 $1471 \times 15 = 22065$,

(乘 15：一片叶子 13 个标志点加 LM 和 ID 共 15 行)，然后用 EXCEL 打开原始数据，找到 22065 行，即可找到出现错误的数
据，改正后重新导入。



图 8 数据导入失败提示

点击 Project Tree 查看导入数据，单击右键，出现图 9 对话框，然后选 Rename 可以给导入数据命名，这里我们命名为：All data。

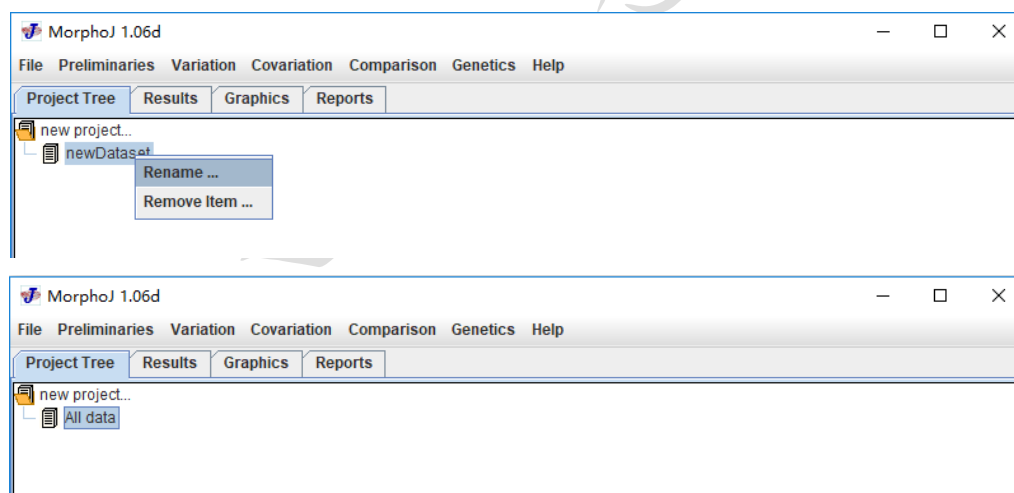


图 9 修改数据名称

3.3 前期数据处理(Preliminary data processing)

3.3.1 普式拟合(Procrustes fit)

叶片标本上标点的位置、取向和比例等非形态的变化会直接导致分析结果出现偏差，因此必须除去非形态变异对分析的干扰。本方法

使用普氏叠印法去除非形状变异 (Morpho J 手册, http://www.flywings.org.uk/MorphoJ_guide/frameset.htm?comparison/cv a.htm)。

步骤为：注意一定要首先在“Project Tree”中选定要处理的数据，然后选择“Preliminaries”→“New Procrustes Fit”，在弹框图 10 中选择“Align by principal axes”后点击“Perform Procrustes Fit”。

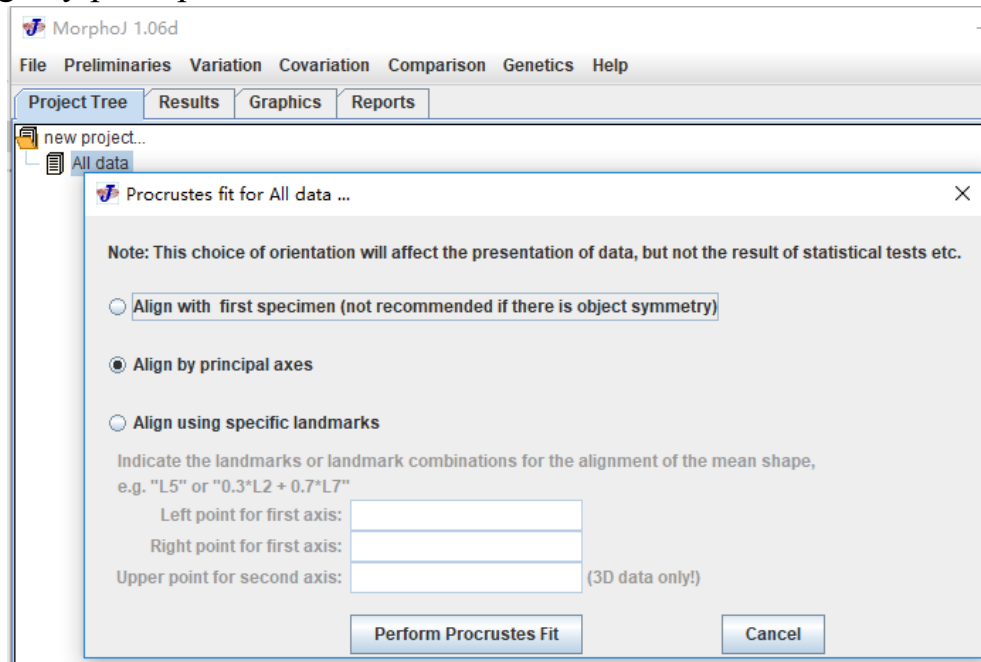


图 10 普式叠印选项

之后出现以下对话框：

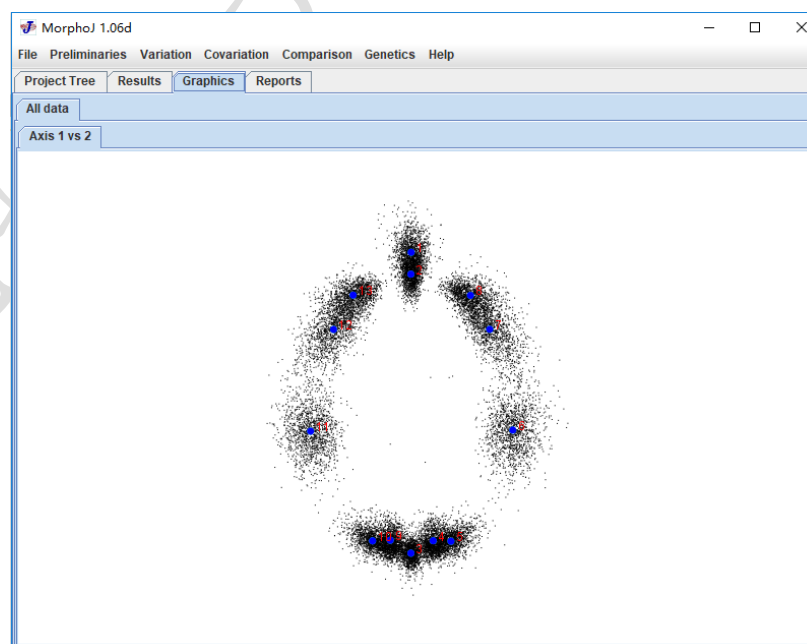


图 11 普式叠印结果图展示

图 11 显示所有输入数据经过平移、旋转、缩放最终最大限度的

重合在一起。

3.3.2 查找异常值 (Find Outliers)

可在“Preliminaries”菜单中选择“Find Outliers”在图 12 弹框的左下方“Included”中显示每个个体偏离平均值的数值，此时可根据偏差大小找到离群值并选择排除或保留。

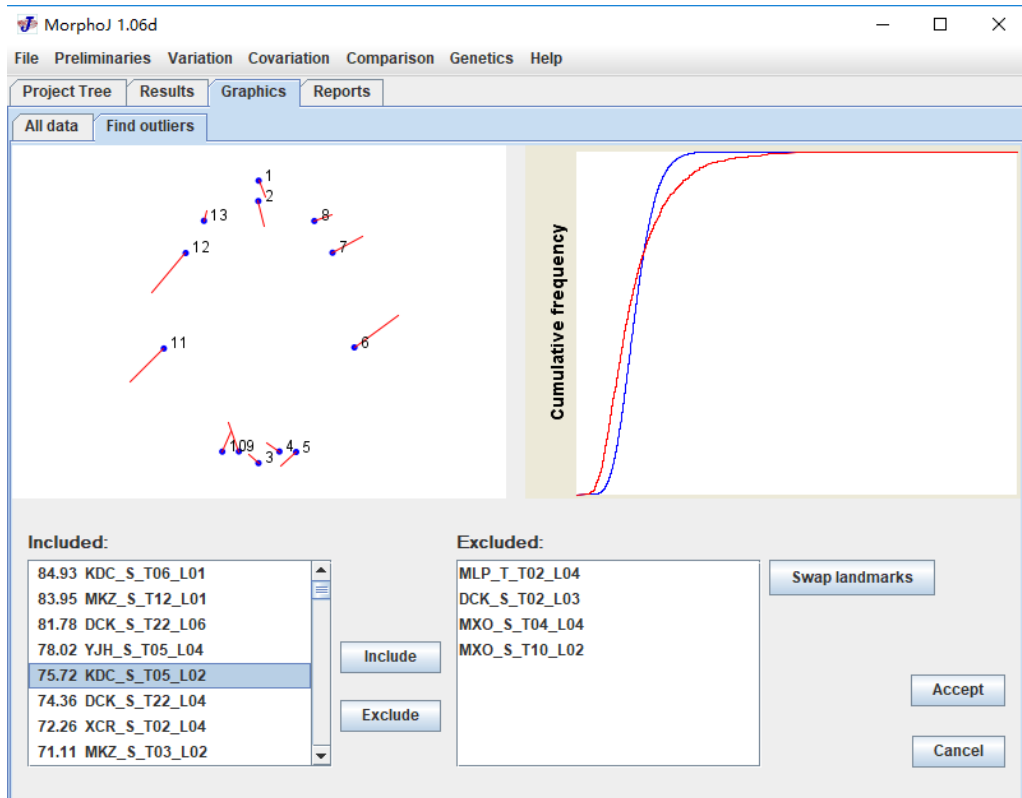


图 12 离群值筛选

3.3.3 分离对称和非对称性(Separate symmetry and asymmetry)

选择“Preliminaries”→“Select Landmarks”，如图 13 所示：

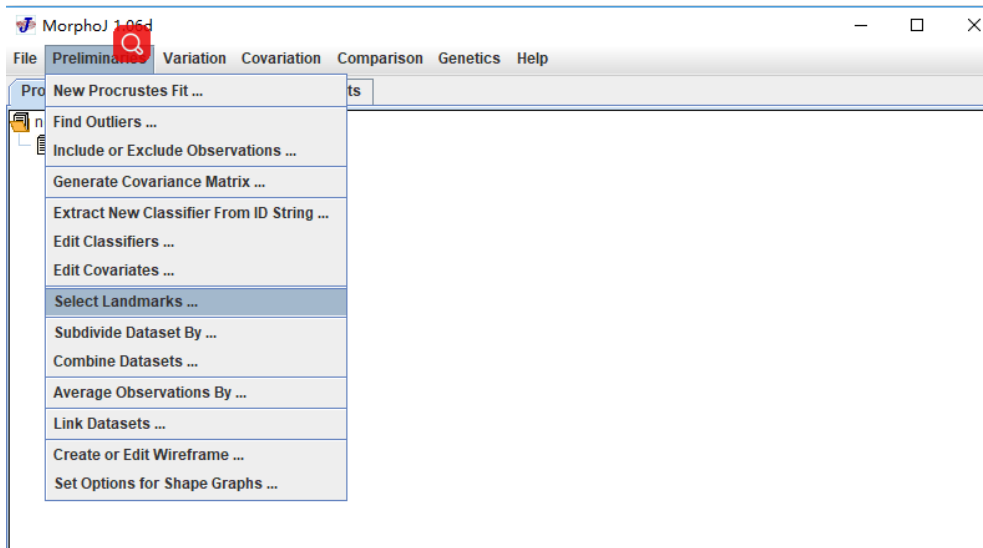


图 13 分离对称和非对称结构
出现以下对话框：

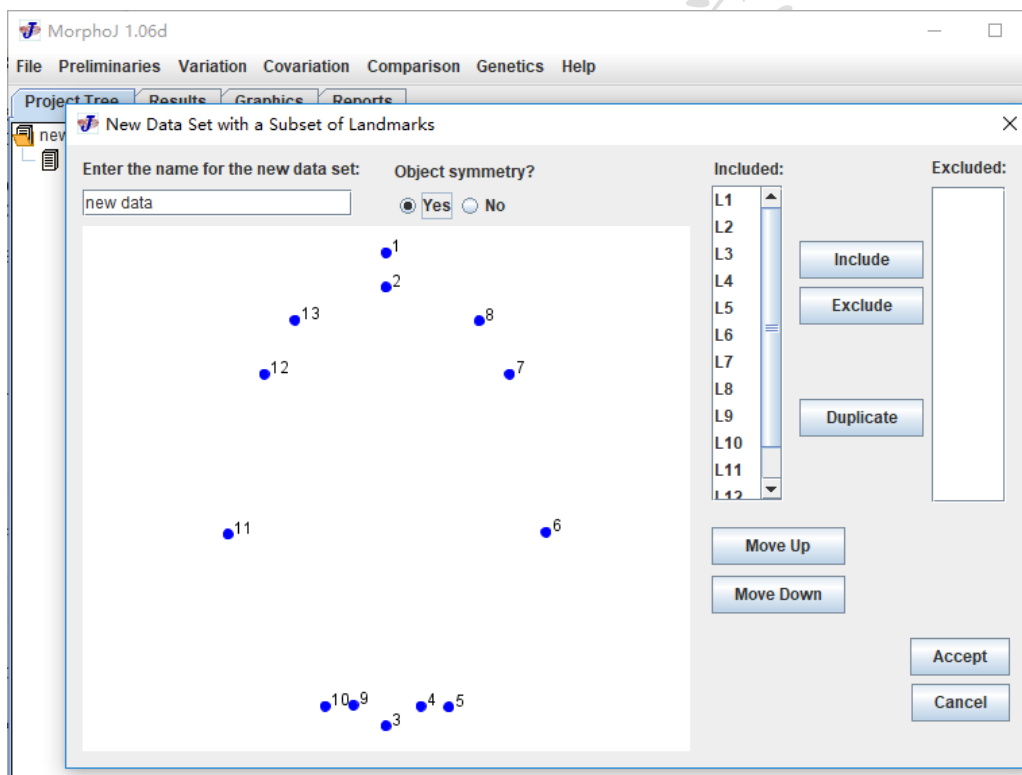


图 14 标志点筛选以及新数据命名

给新数据集命名，在图 14 中，选择 Yes(观察对称非对称结构)，标志点筛选结束后选 Move Down，最后点击 Accept，出现图 15 对话框：

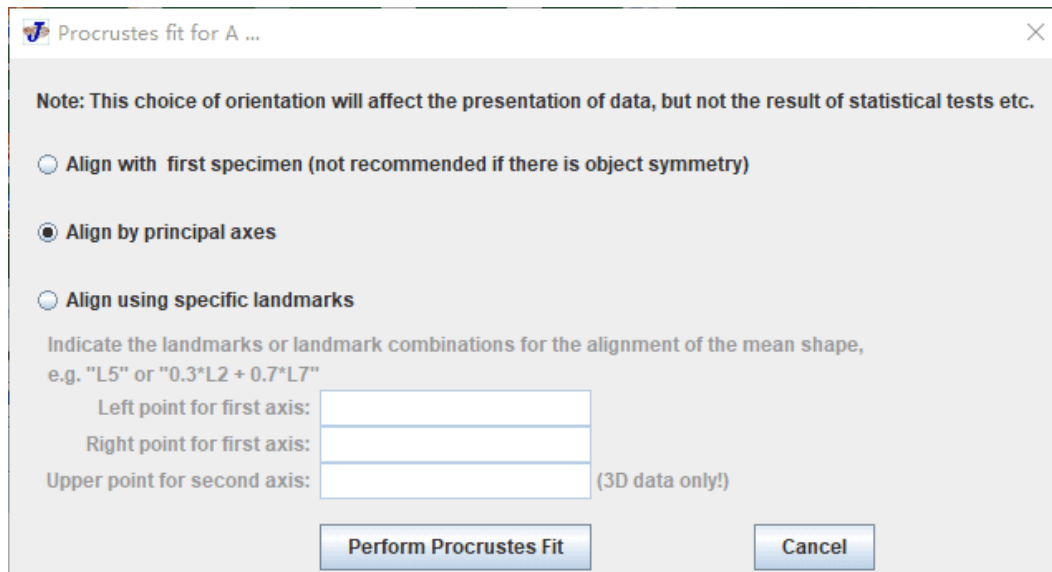


图 15 区分对称和非对称结构的普式叠印
点击 Perform Procrustes Fit，弹出图 16 对话框：

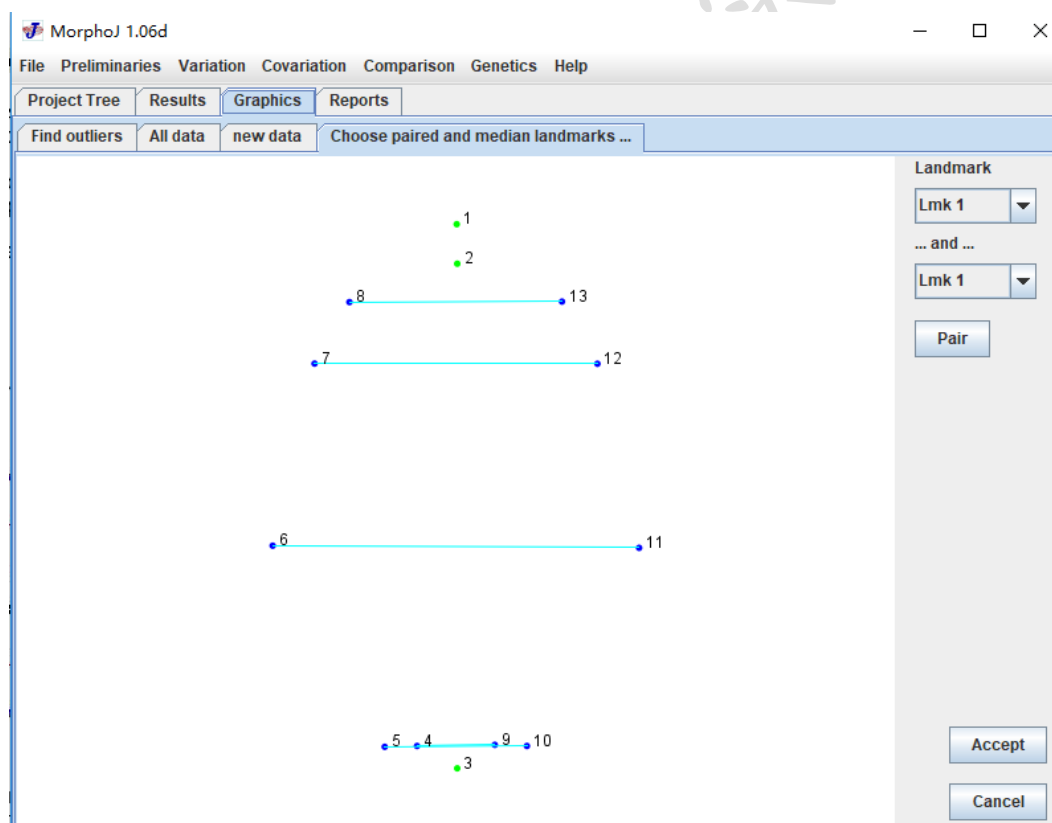


图 16 两边成对的标志点和中轴上的标志点选择
可以看出浅蓝色线条相连的为对称的标志点，该软件可以自动识别对称结构，点击 Accept，出现以下对话框，对话框中可以点击 Symm1vs2 和 Asym1vs2，查看对称(图 17)和非对称(图 18)结构的普氏叠印图：

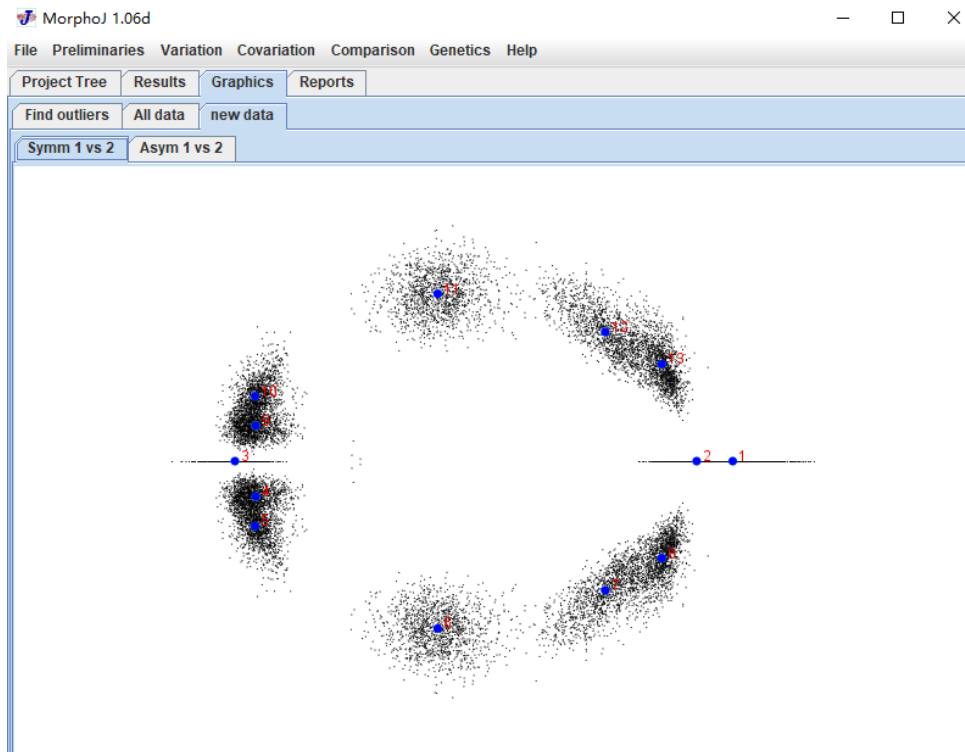


图 17 对称结构的普式叠印图

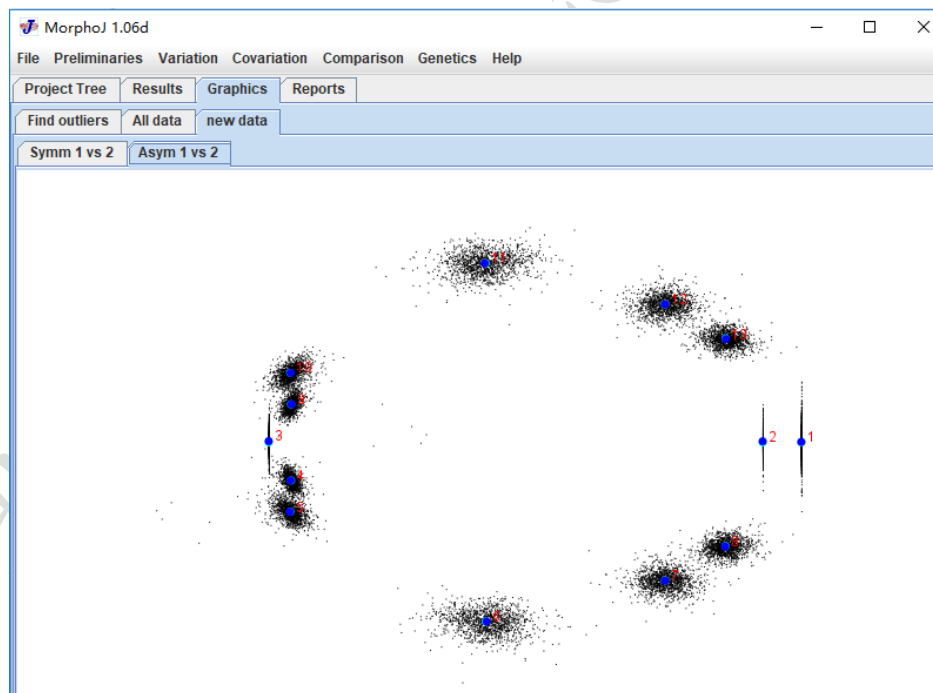


图 18 非对称结构的普式叠印图

3.3.4 提取新分类层(Extract New Classifier)

如果需要单独分析每个种群，就要对数据库中的信息进行提取并

重新分类。选择“Preliminaries”→“ Extract New Classifier from ID String”。弹出图 19 对话框，在“Name for new classifier”框中给新的分类命名，这里命名“leaf”：根据“Identifier strings”中的字符所显示叶片水平有 13 个字符长，从第 1 个第 13 个字符，因此“First character”为 1，“Last character”为 13，点击“Execute”。同样的方法可以命名为“tree”：根据“Identifier strings”中的字符所显示树水平有 9 个字符长，从第 1 个字符到第 9 个字符，因此“First character”为 1，“Last character”为 9。

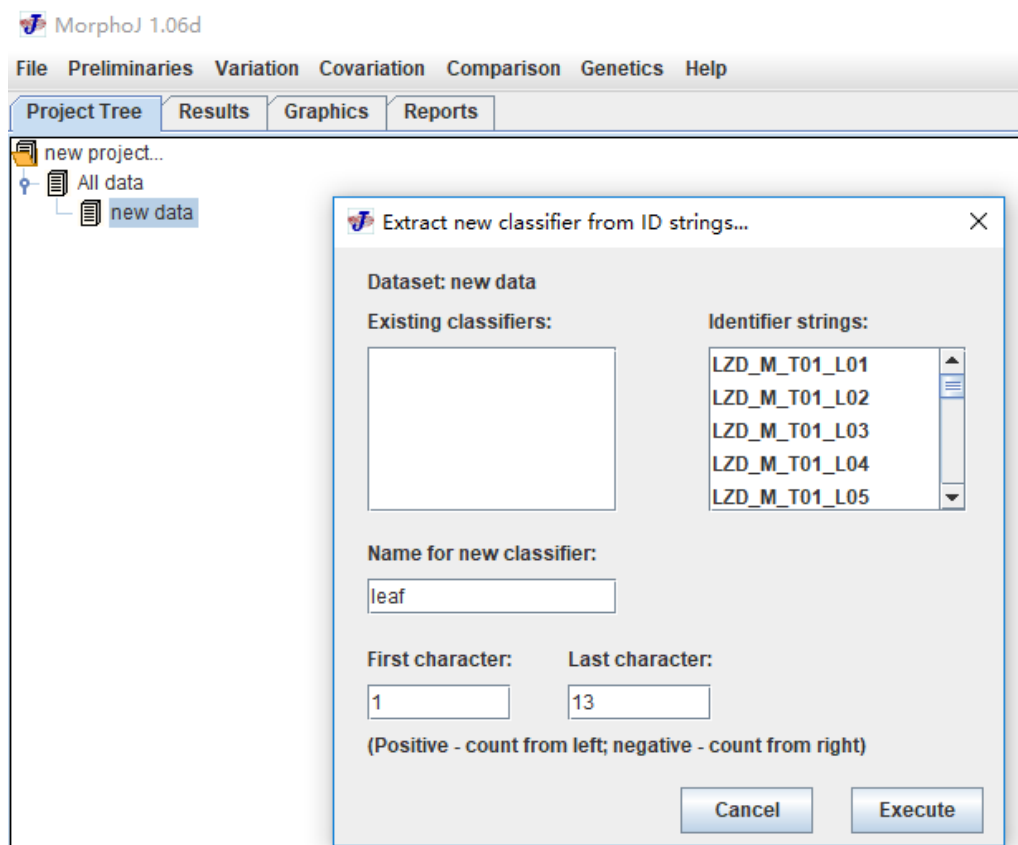


图 19 提取新的分类器

当提取新的分类后可以在“Preliminaries”下选择“Edit Classifier”检查并编辑结果(图 20)。

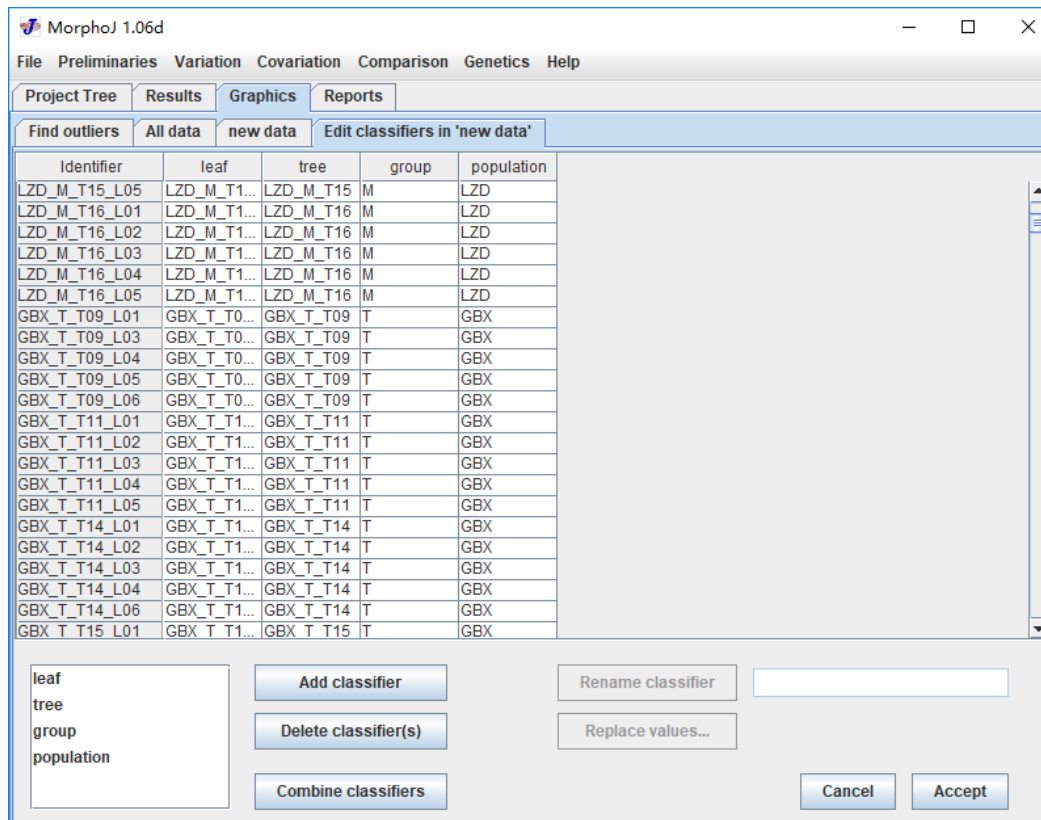


图 20 编辑分类器

3.3.5 生成新的子样本集(Generate new subsamples)

根据分组情况将叶片标志点的数据做平均,首先选择“Preliminaries → Average Observations By...”(图 21),然后弹出图 22 对话框,选择“Average by...”确定要数据平均的水平,最后点击 Execute.

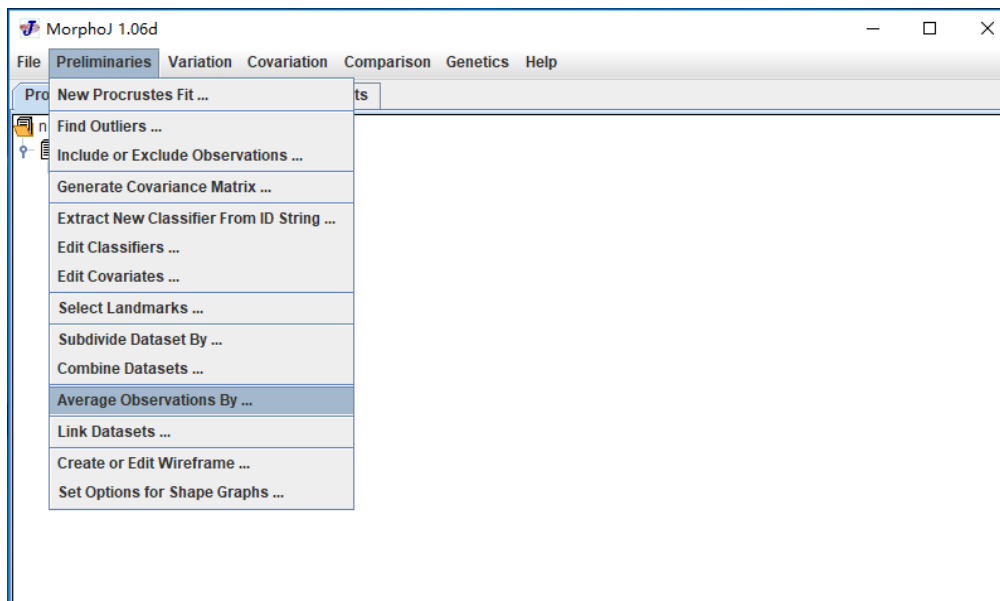


图 21 平均数据到不同的分组水平

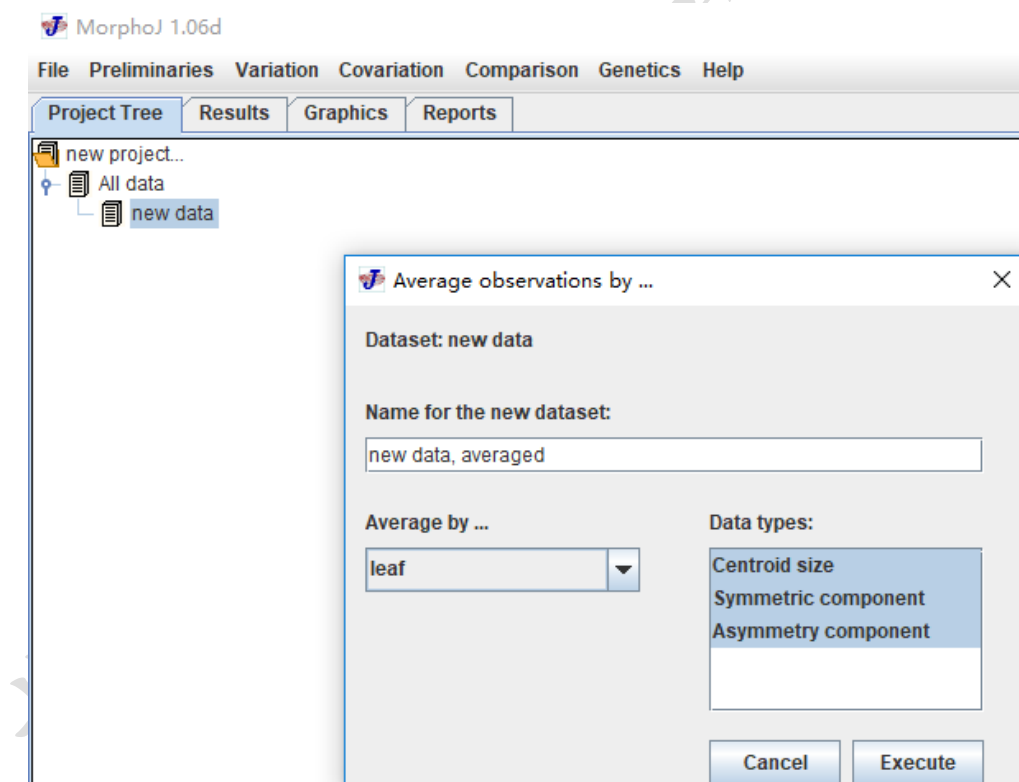


图 22 选择平均数据的分组水平

返回“Project Tree”中将新的数据集改名，如图 23 所示：

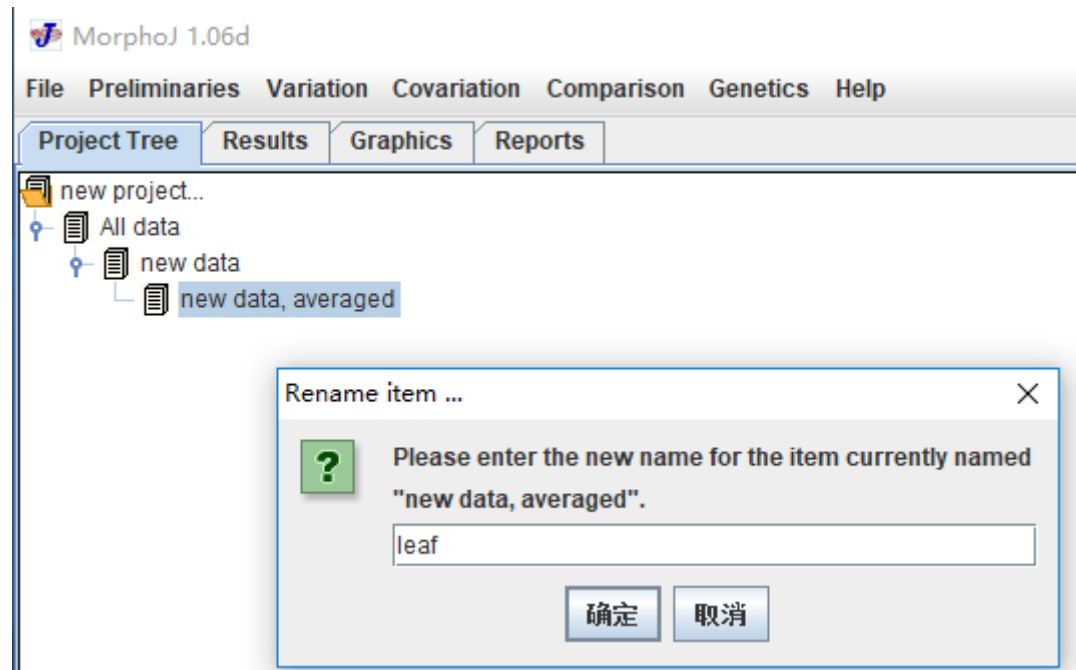


图 23 修改数据名称

3.3.6 构建线框(Create wireframes)

线框是标志点之间的连线的集合，可帮助研究者观测标志点的结构配置。使用 Morpho J 软件创建线框图的步骤如下：选择“Preliminaries”→“create or edit wireframe”，弹出图 24 对话框，在出现的图形背景中右击鼠标可以调整图形的方向。编辑线框，即在标志点之间进行划线，有两种方法：一是可以点击图中的一个标志点，然后拖拽到另一个标志点上；二是在右边下滑菜单中选择要连线的两个点然后点击“Link landmarks”。点击“Delete link”可删除线段。

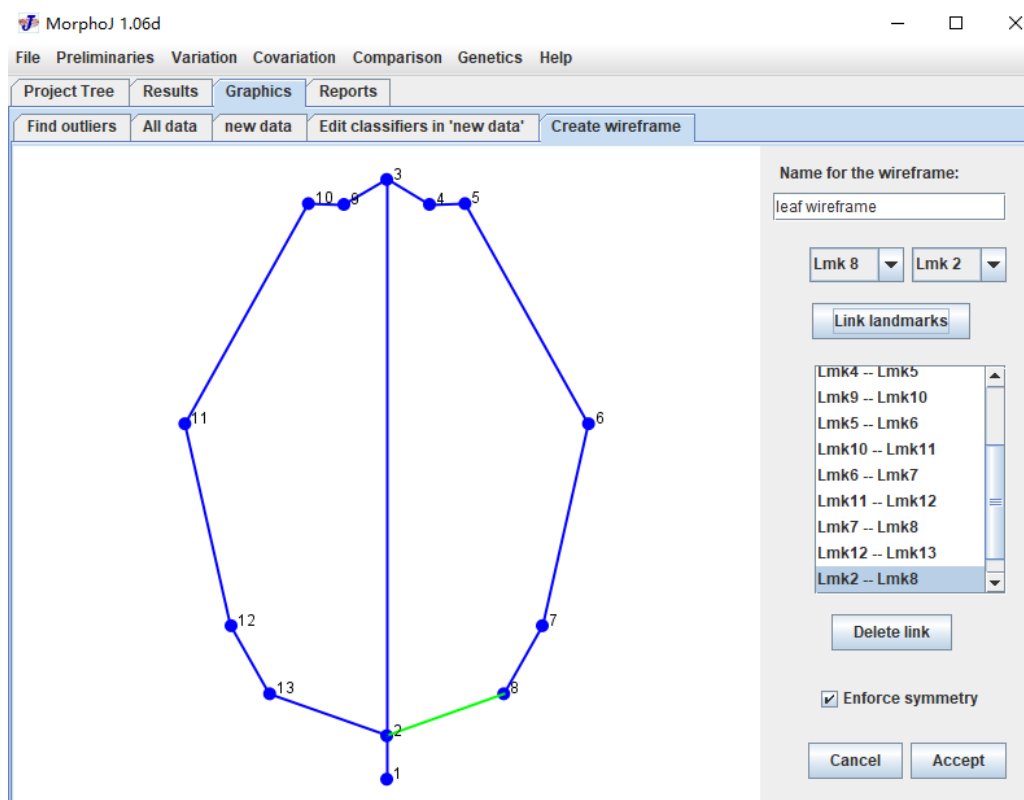


图 24 创建线框轮廓

3.4 多变量分析(Multivariate analysis)

3.4.1 主成分分析(Principal component analyses)

PCA 分析是将多个指标化为少数几个综合指标的一种多元统计分析方法。由于这种方法既可消除各指标不同量纲的影响，也可消除由各指标间相关性所带来的信息重叠，从而起到降维的作用，简化了指标的结构，使分析问题简单、直观、有效(杨维权等，1989)。主成分分析作为多变量分析中使用最广泛的方法之一，可用于检测数据集中形状变化的主要特征并可视化分析组之间的形状差异(Vujić *et al.*, 2016)

步骤：进行 PCA 之前首先要创建叠印后的叶形态坐标的协方差矩阵(covariance matrix)。点击“Preliminaries” → “Generate Covariance Matrix”；在“data types”框中选择“Symmetric component”或者“Asymmetry component”后点击“Execute”按钮。出现图 25 对话框：

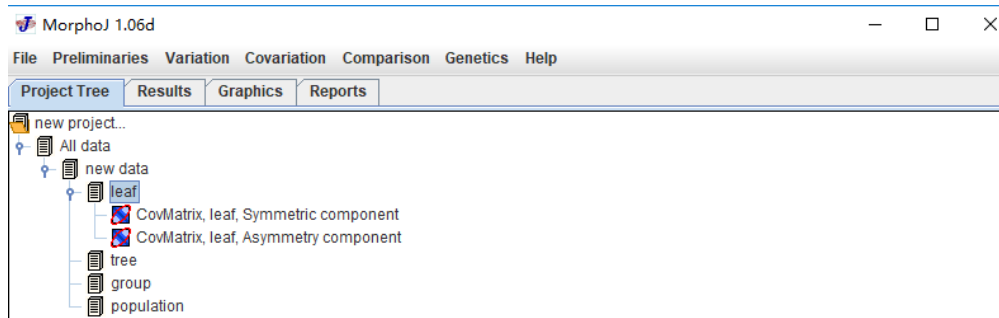


图 25 创建协方差矩阵

然后选中分析数据的协方差矩阵进行 PCA 分析以显示不同的叶片在几何空间上的异同。点击“Variation” → “Principal Components Analysis”，出现图 26：

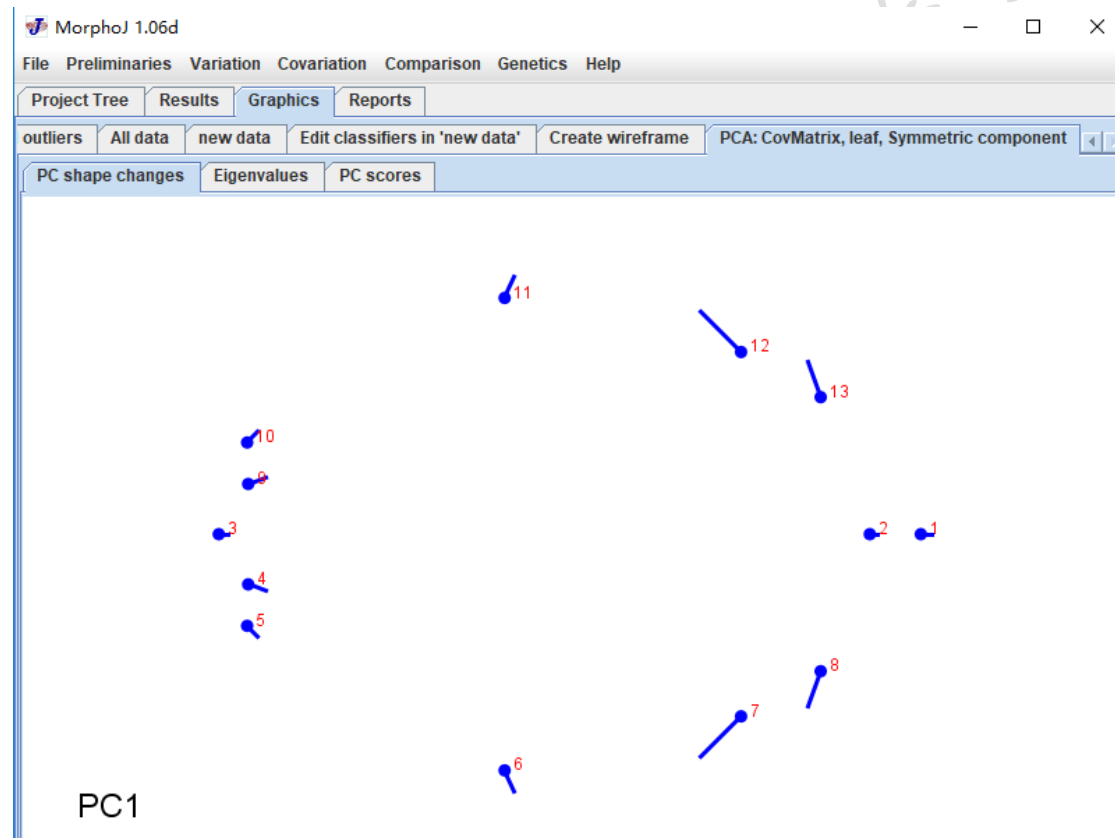


图 26 PC1 轴上的叶形变化

在出现的“graphics”窗口上点击“PC scores”，然后出现散点图(图 27)：

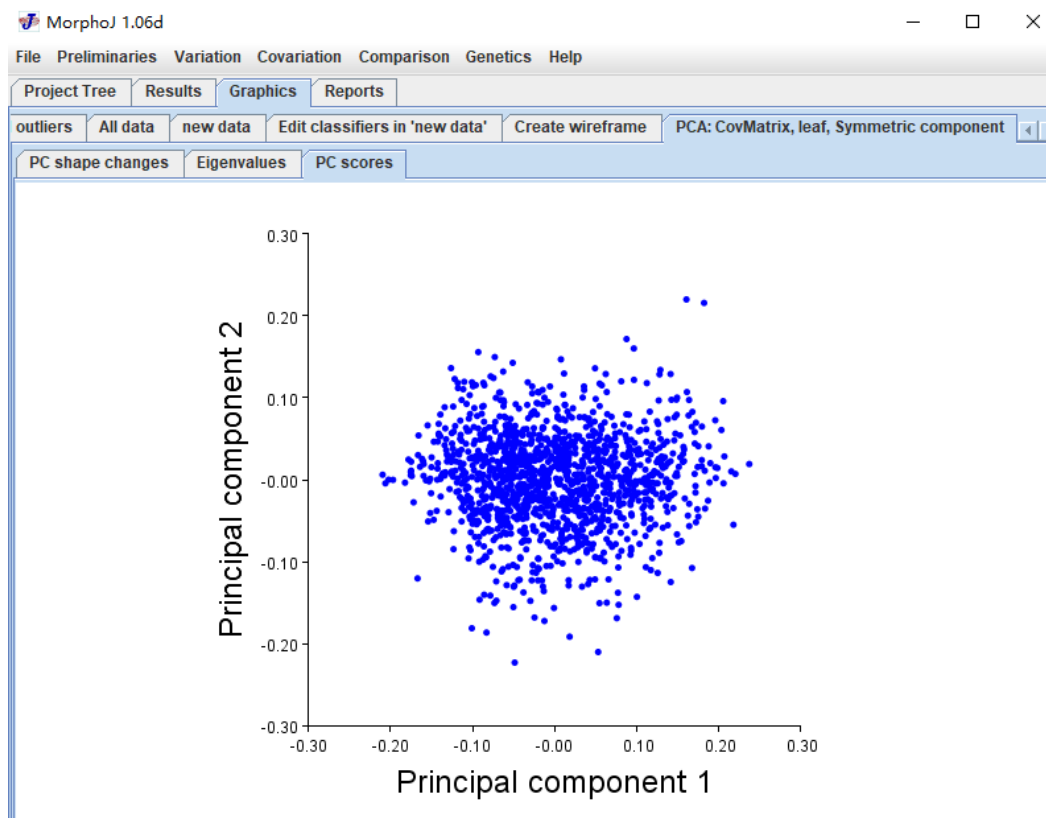


图 27 PCA 散点图

右击鼠标可编辑图像，出现图 28 对话框：

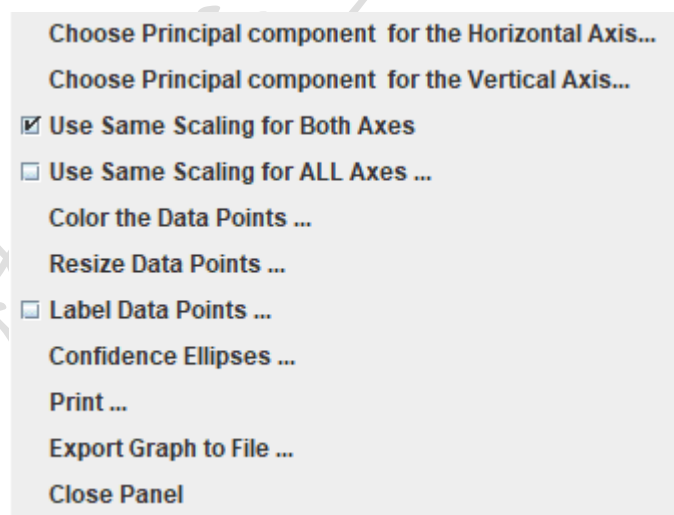


图 28 散点图的设置选项

“Label Data Points”，可对所有标志点名字显示出来；“Color the Data Points”选中“Use a classifier variable to determine the color for each category”，然后弹出图 29，选择物种 spe 作为分类单元，

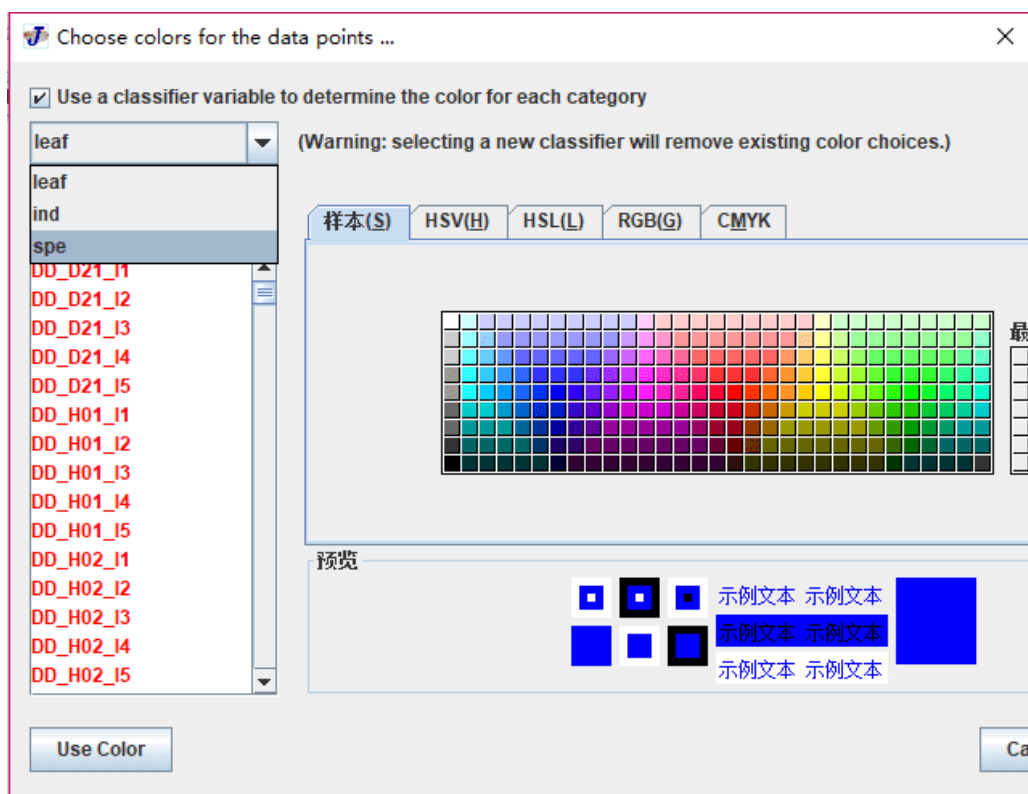


图 29 设置 PCA 散点图的分组水平
选中颜色，点击物种名可以人为更换颜色，如图 30 所示：

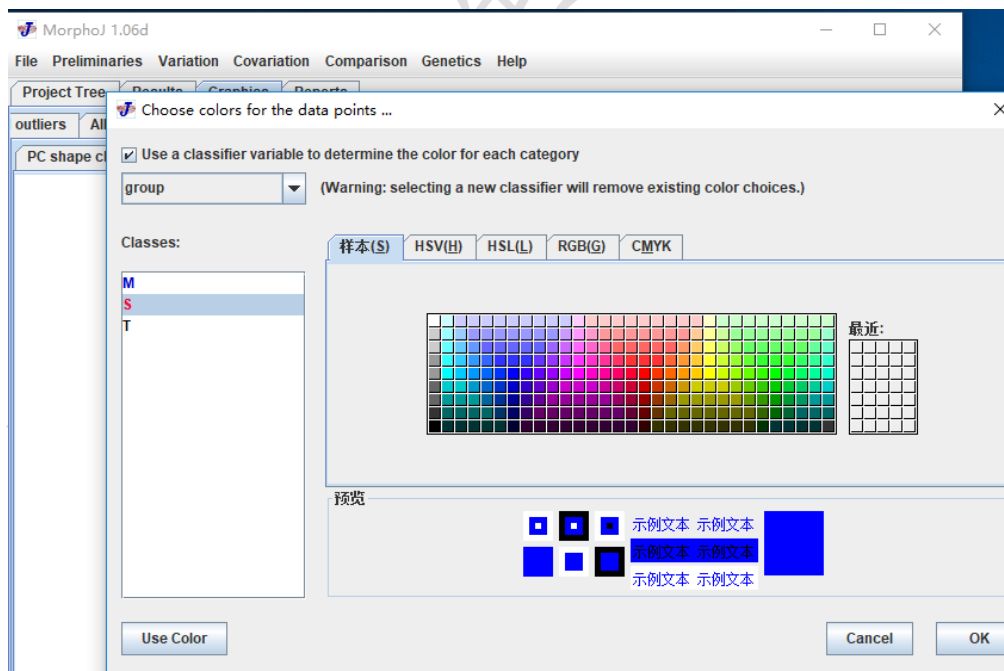


图 30 设置各组散点颜色
点击“Use Color”最后点“OK”，即可出现不同颜色的散点图，如图 31 所示：

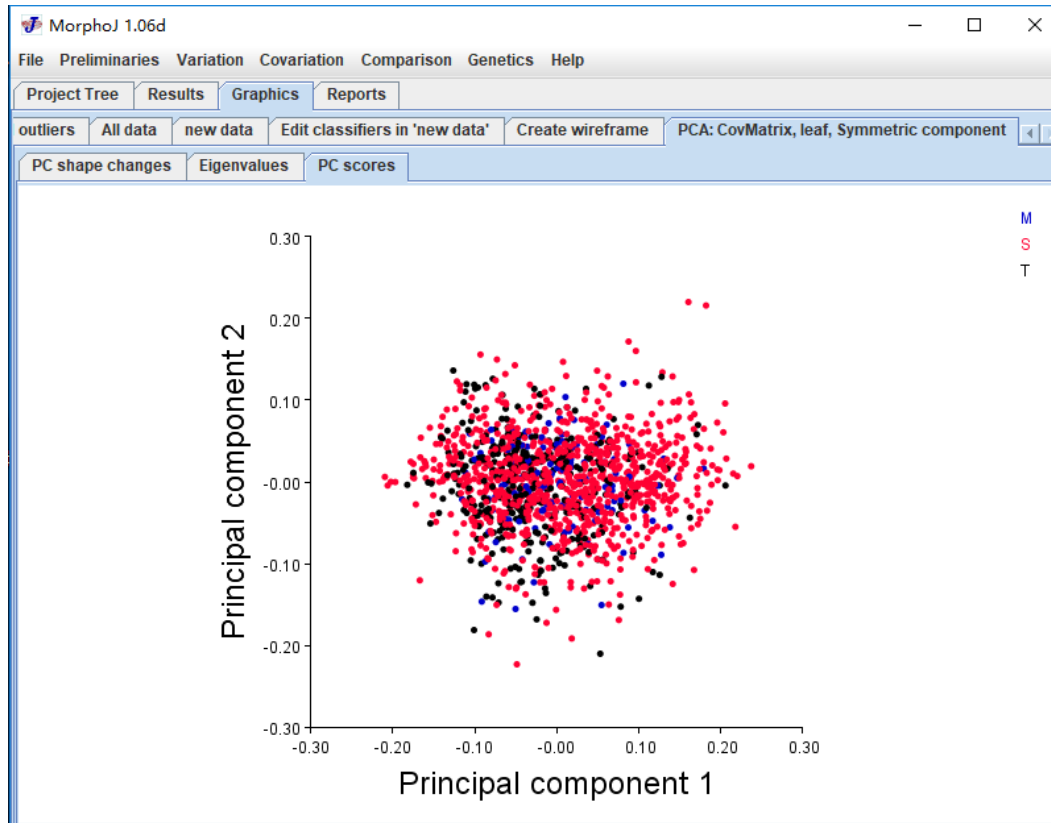


图 31 分组完成的 PCA 散点图

此外，单击右键出现图 32 对话框中阴影部分：

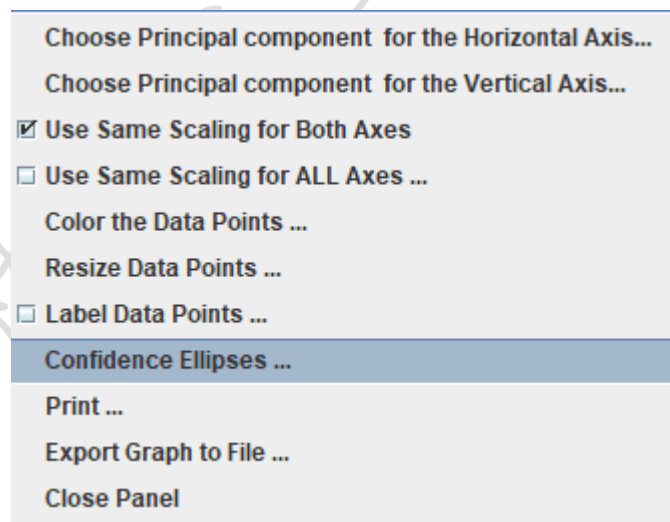


图 32 PCA 散点图的设置选项

弹出图 33，选中“Draw ellipse(s)”，点击 OK，图 34 是设置完成的 PCA 散点图

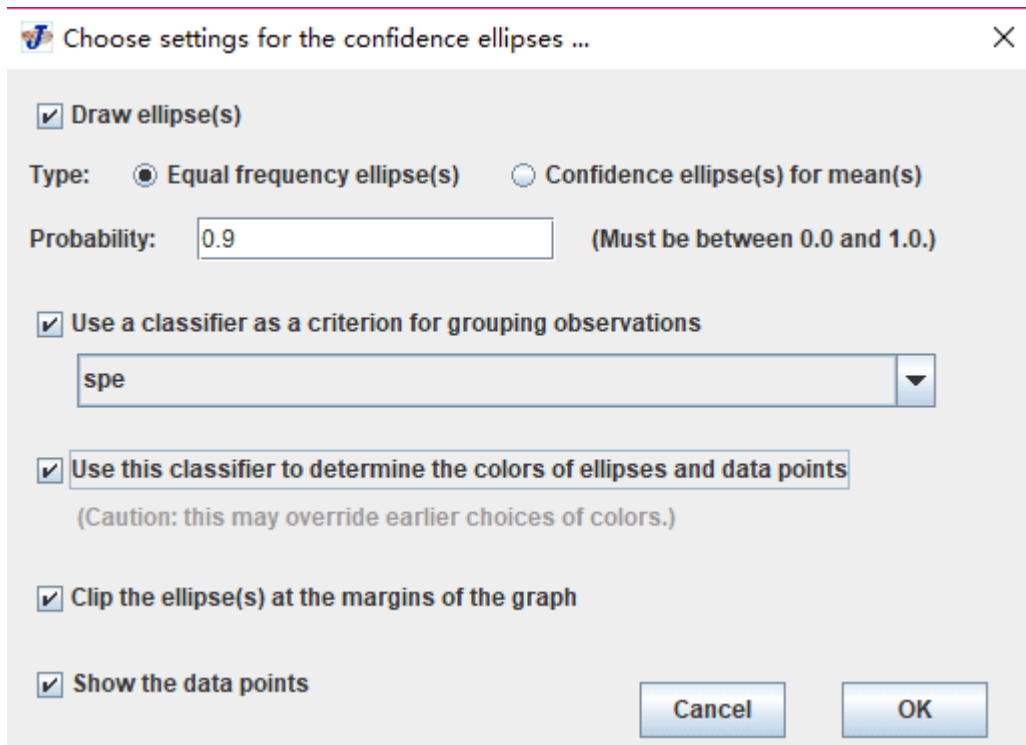


图 33 PCA 散点图的置信区间设置

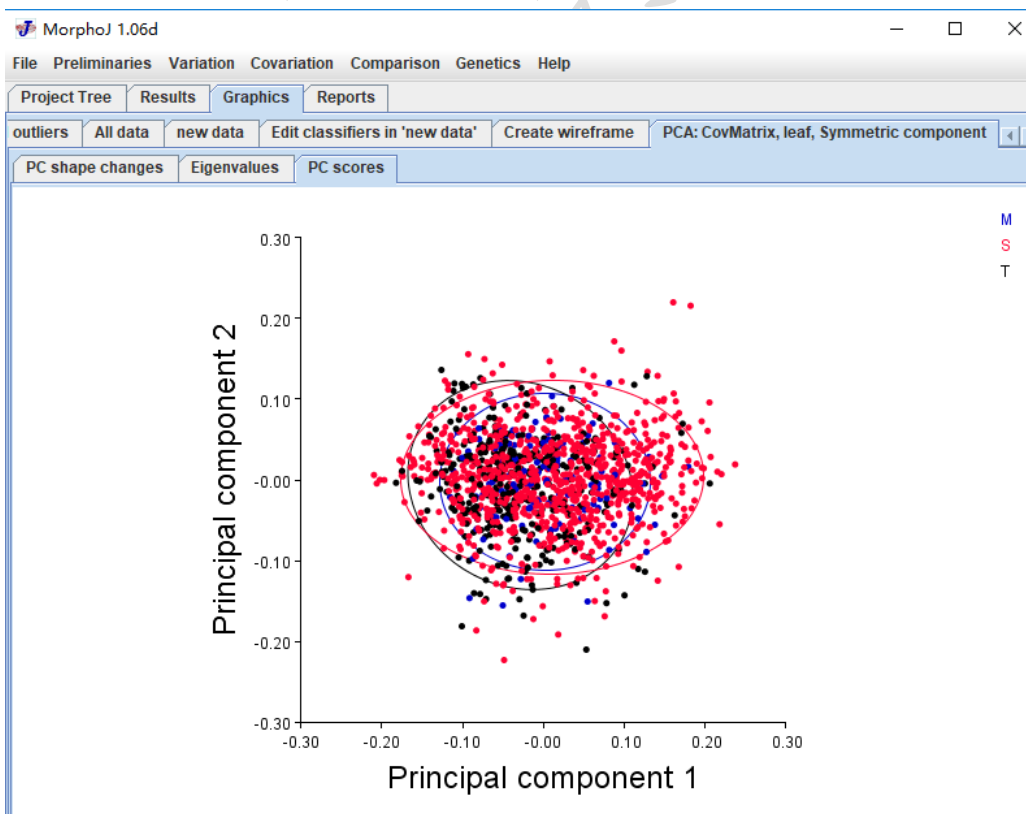


图 34 设置完成的 PCA 散点图

增加散点的大小，单击右键出现图 35 对话框中阴影部分：

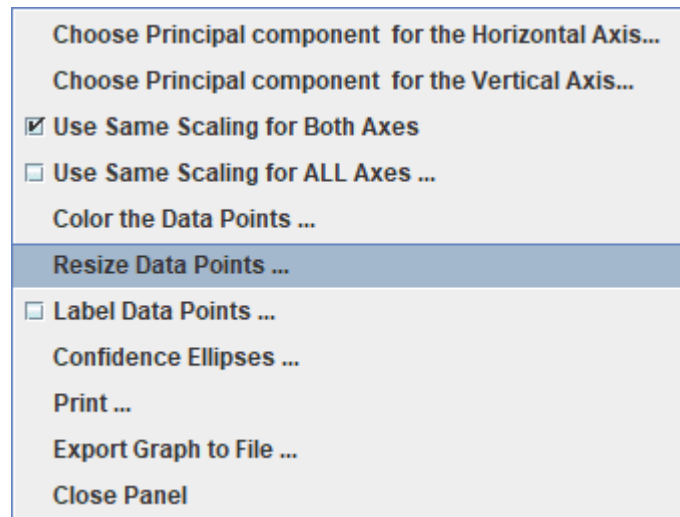


图 35 PCA 散点图的设置选项

在图 36 中，输入 1 到 10 可以改变散点大小，



图 36 修改散点图中点的大小

然后选择图 37 中的“Export Graph to File”，将设置完成的散点图导出

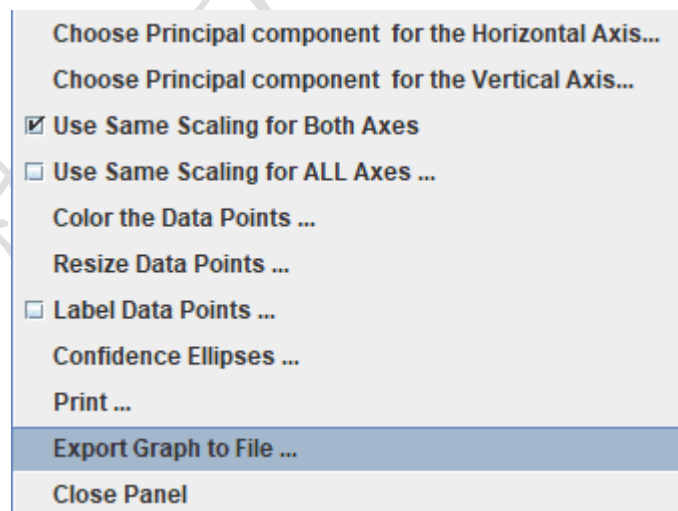


图 37 导出散点图选项

其中，图 38 展示了所有可导出的文件格式：

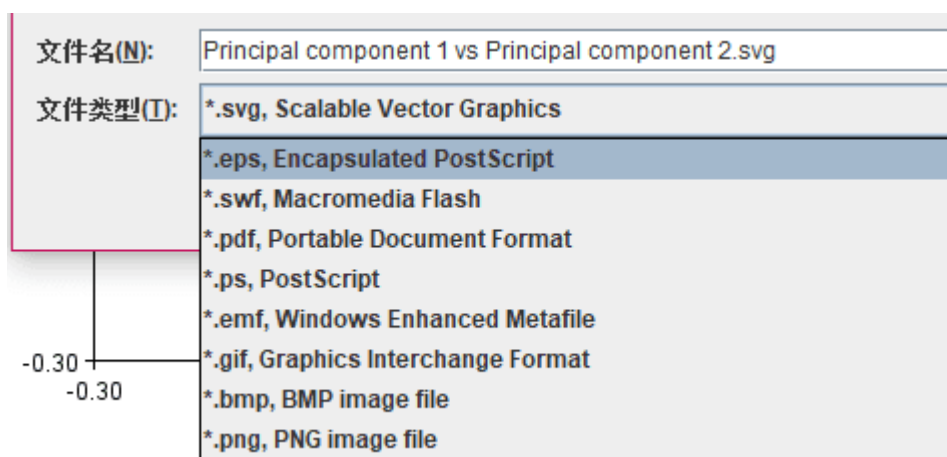


图 38 导出图的文件类型

另外还可以通过点击 PC shape changes，查看不同 PC 轴上叶片形态的变化，图 39 是 PC1 轴上叶形变化的棒棒糖图：

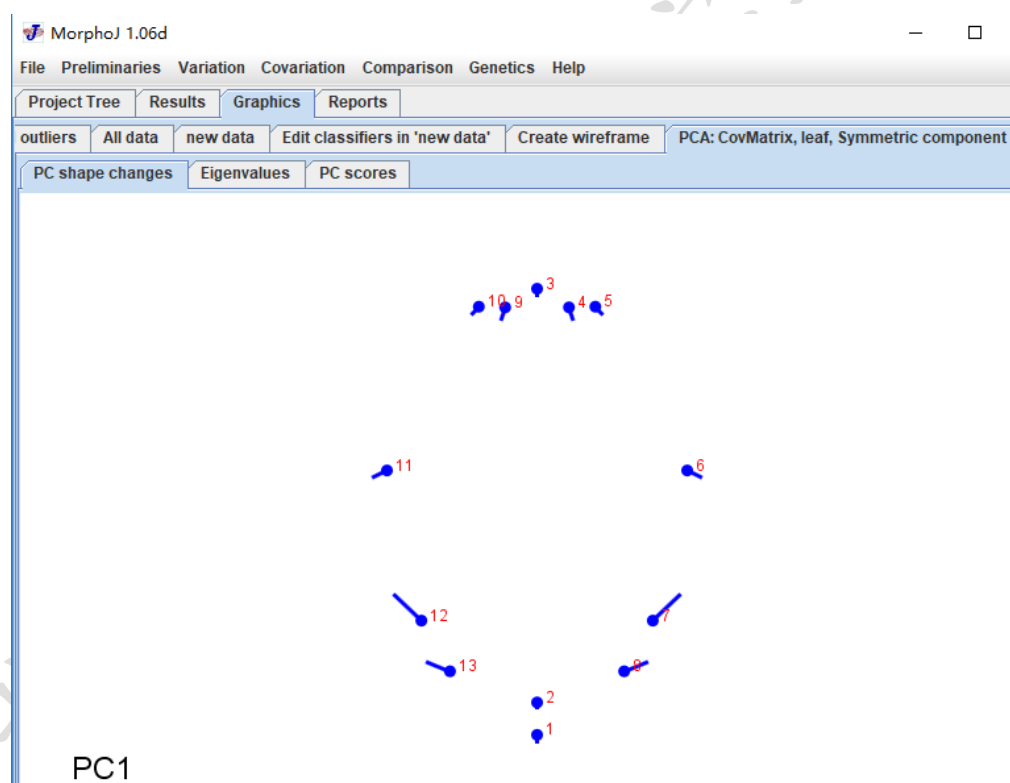


图 39 PC1 轴上叶形变化的棒棒糖图

选中图像，右击鼠标，改变 PC 轴，设置 PC 值的大小，改变图形类型等，以修改图形类型为例：

选择“Change the type of Graph”→“Wireframe Graph”(图 40),弹出



图，然后点击“确定”，图形类型转换完成，出现线

框图(图 41), 点击下图为 **Warped Outline Drawing** 可转换成网格(42)

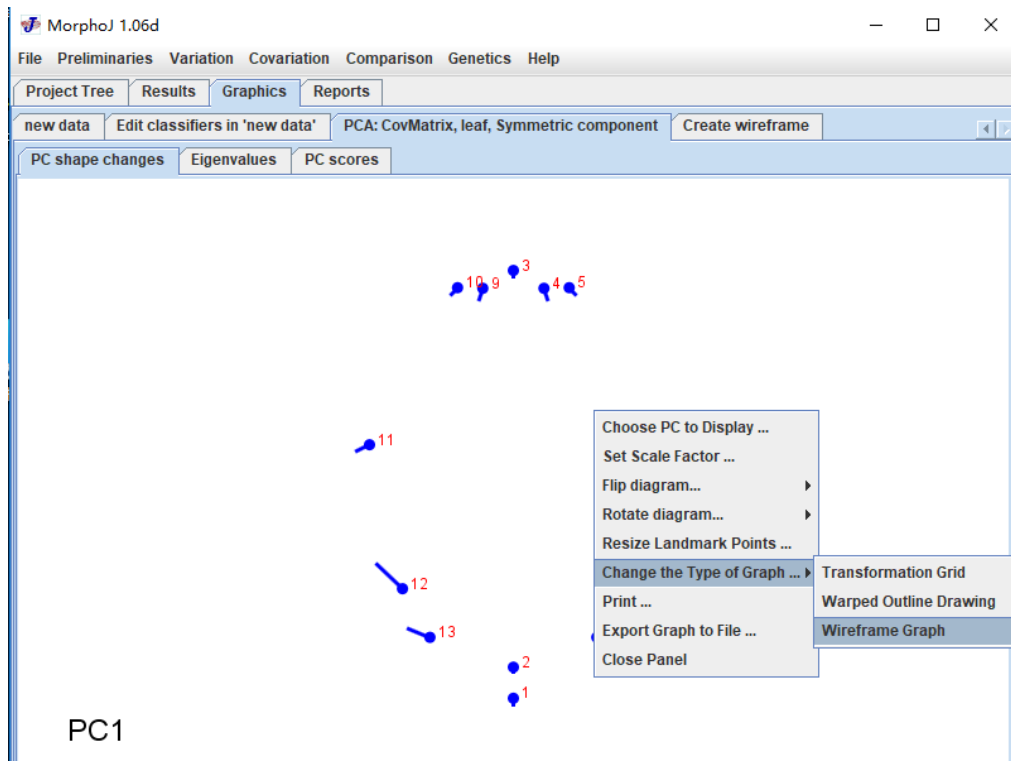


图 40 图形转换的设置选项

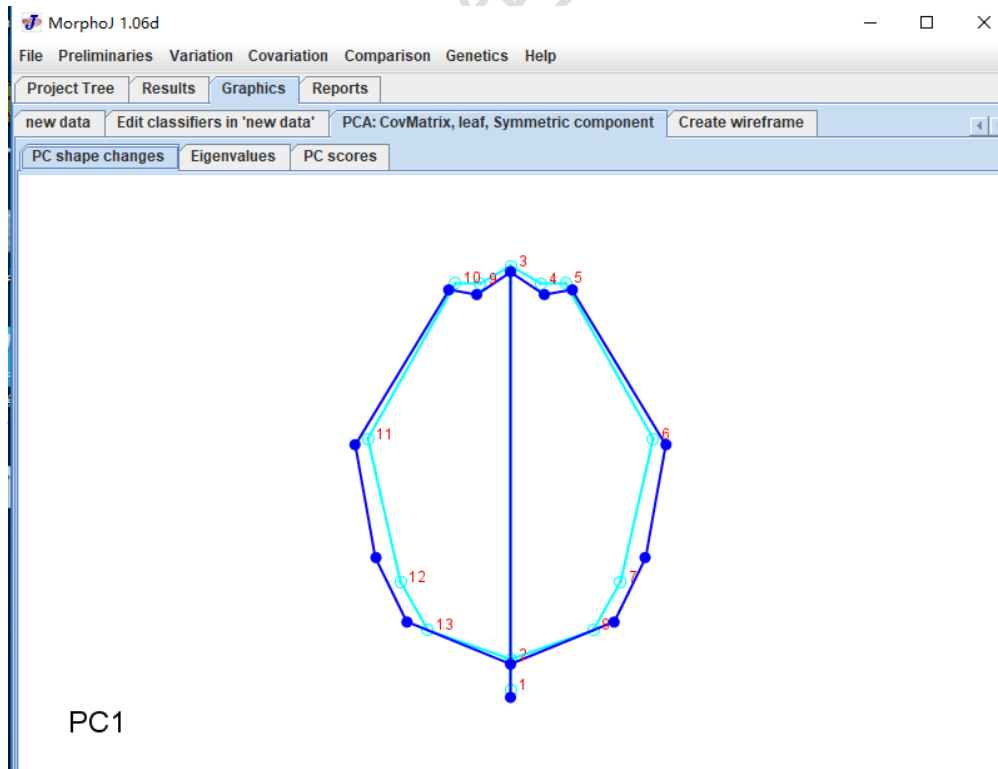


图 41 PC1 轴上叶形变化的线框图

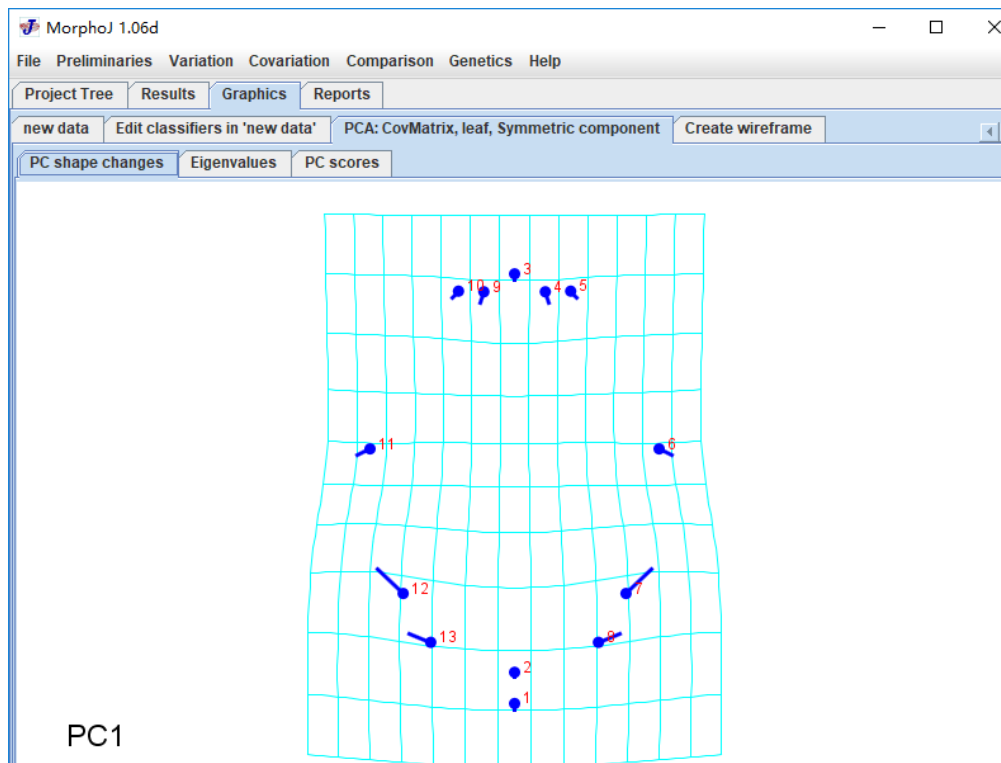
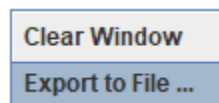


图 42 PC1 轴上叶形变化的网格图

我们可以通过右击鼠标点击 Export to File 导出线框图与网格图：



, 然后在作图软件中叠加起来(修图软件 Ai), 例如图 43, 更形象的表示叶片形态的变化。

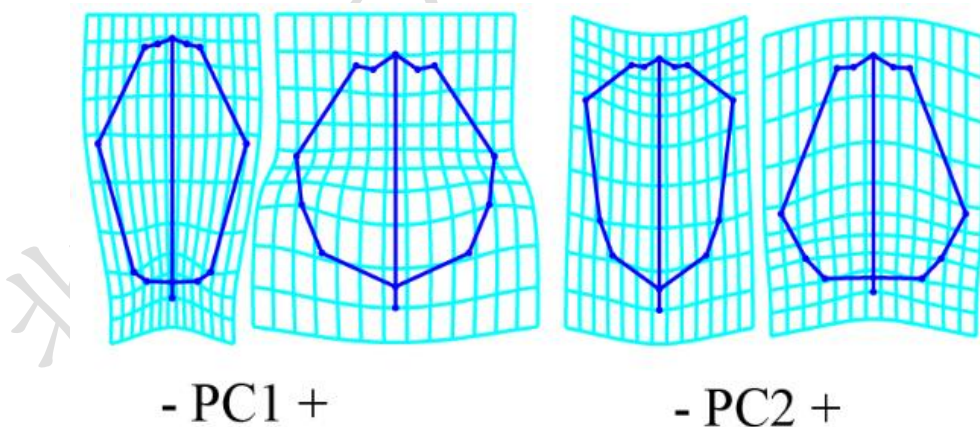


图 43 最终合成的重组图

除了散点图和叶型轮廓图，在“result”中还显示了分析结果(图 44)，右击可选择输出结果，显示各个 PC 占总变异的比例：

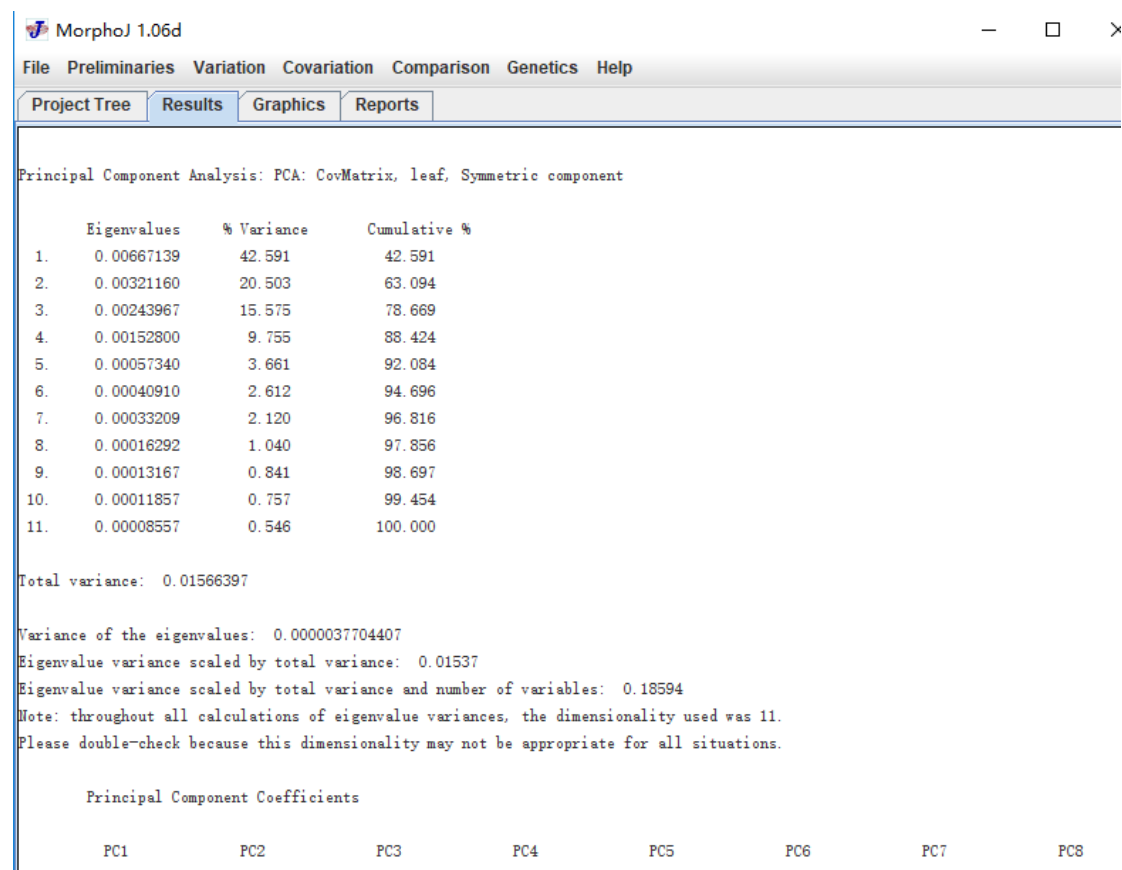


图 44 PCA 结果显示

3.4.2 典型变量分析(Canonical variate analysis)

CVA 分析是用于检测具相同或相似特征的某一类样品总体之间的差异的一种多元统计方法。不仅可以消除数据变量之间的相关性，而且可以解决变量自身相关性 (胡坤, 1992) 。CVA 产生的是散点图，输出的散点图可以通过 Morpho J 添加相同频率椭圆(Equal frequency ellipses)或者椭圆置信区间(Confidence ellipses for means)。

操作步骤：选择“Comparison”→“Canonical Variate Analysis”，在出现的对话框中的“Classifier variables”选择之前提取的新的分类单元 group(组)用于分类，“Classifier variable(s) to use for grouping”选择“group”作为分类单元，并且勾选执行 10000 次(图 45)。

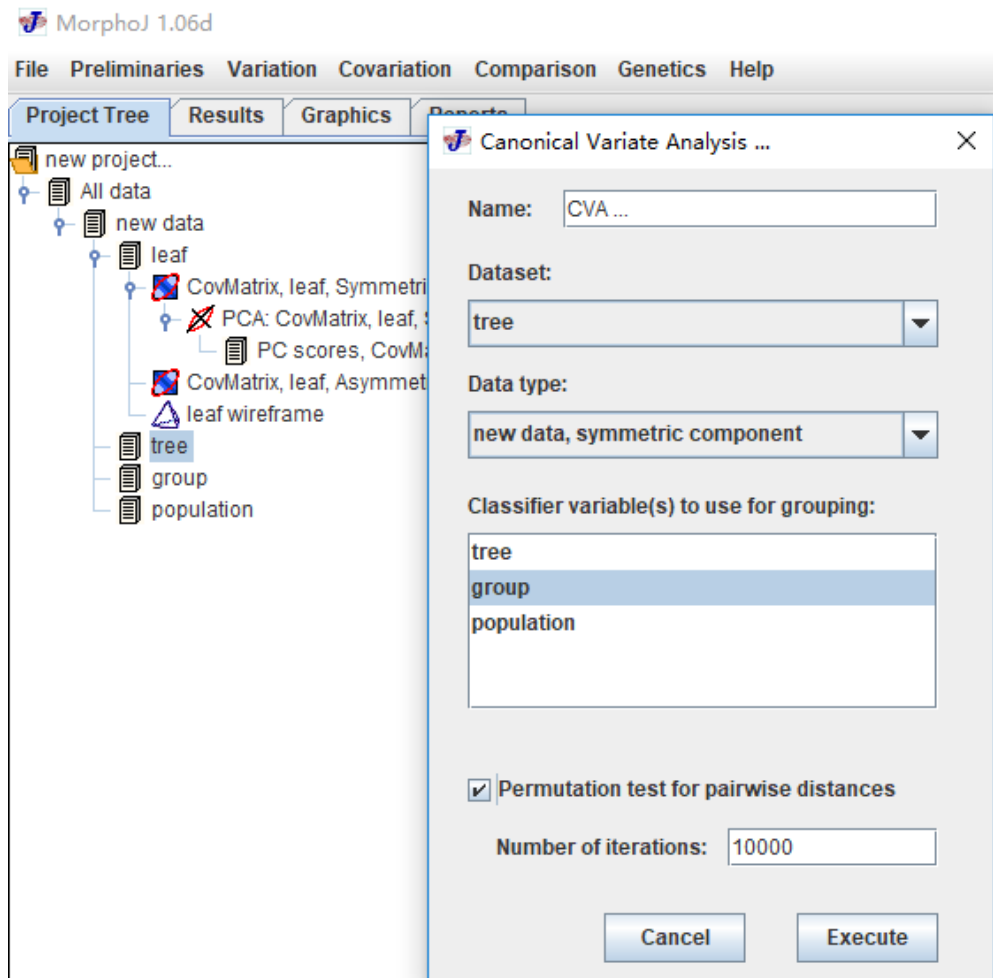


图 45 CVA 数据设置
执行后会出现叶型轮廓图(图 46):

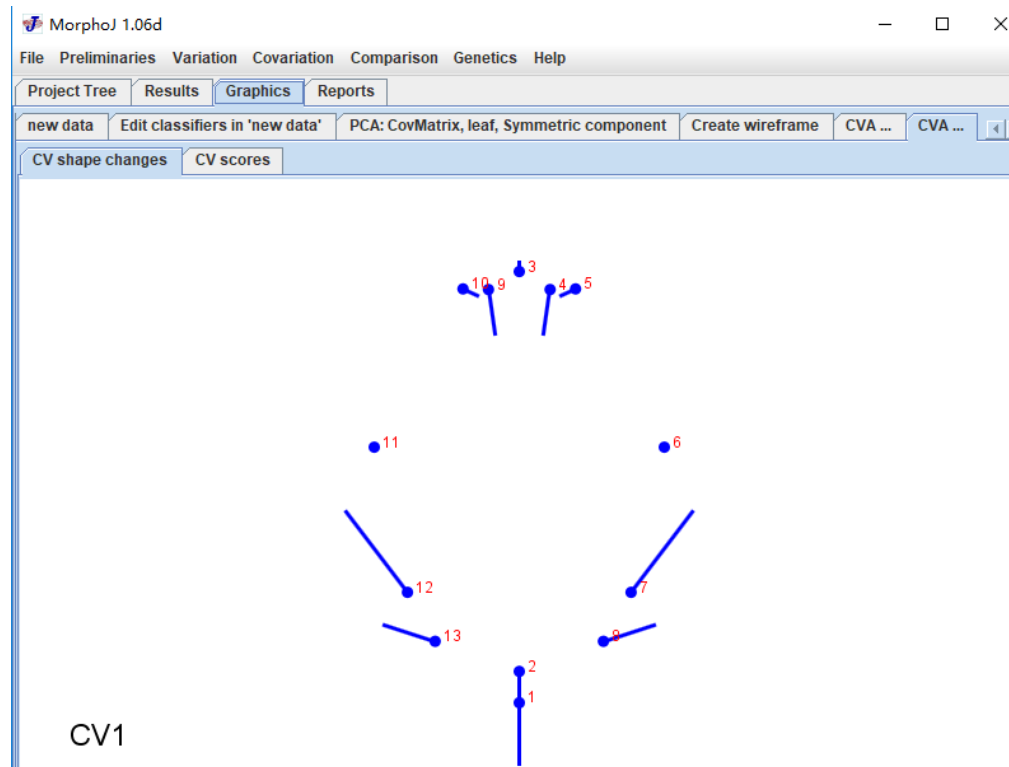


图 46 CV1 叶形变化

点击“CV scores”可以看到 CVA 的散点图(图 47):

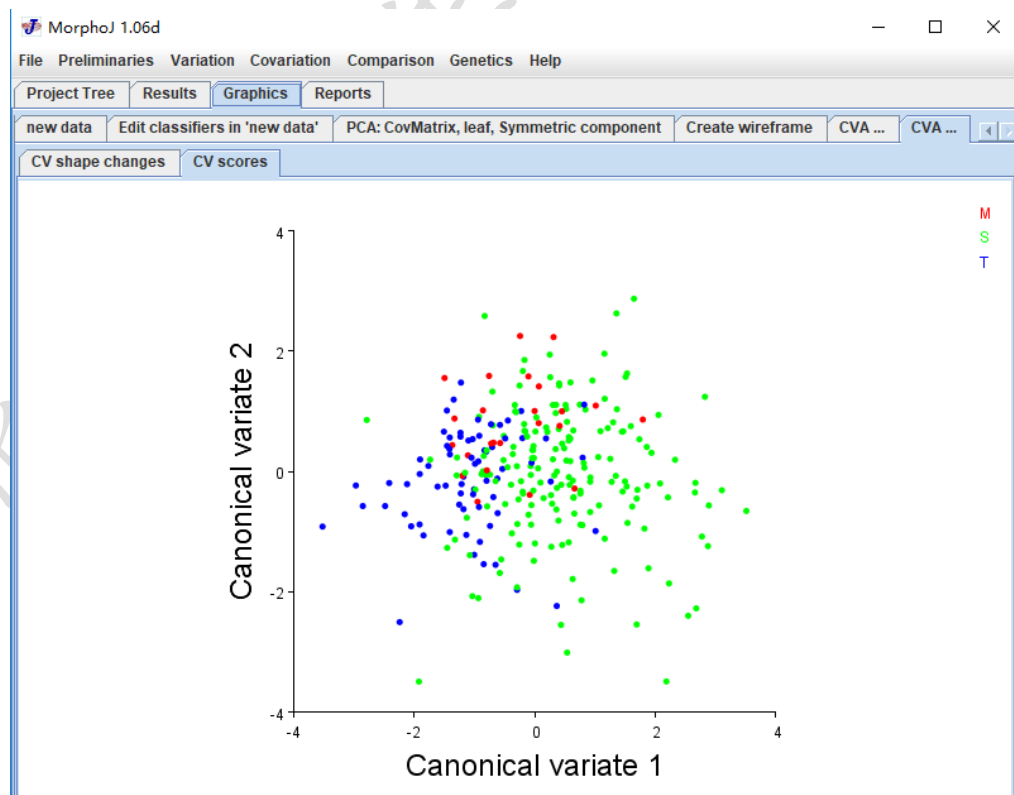


图 47 CVA 散点图

之后和 PCA 一样单击右键对图片进行颜色、散点大小、椭圆频

率的制作及修改，同时也要通过 CV shape changes 将 CV 轴上叶片形态展现出来。

最后在“Results”菜单中可以显示 CVA 分析结果。

3.4.3 判别分析(Discriminant analysis)

DA 分析通过建立一个或多个判别函数，用多元数据确定判别函数中的待定系数，并计算判别指标(陈德钊, 1998)。据此即可确定某一新样品属于哪一类。在形态学研究中一般用于分析两个类群(Viscosi and Cardini, 2011; Vieira *et al.*, 2014)间的变异。DA 分析通过交叉交叉函数(Cross-validated scores)和判别函数(Discriminant scores)两种方法分别进行分析(T^2 P -value for tests with 1000 permutations < 0.0001)，提供可靠的种群分类信息，输出的结果为条形图。

步骤：DA 分析不区分对称和非对称组分，首先在“Project Tree”选中原始数据(All data)重做“select landmarks”(图 14)，选择 No(不观察对称和非对称结构)，得到新的数据集(命名为“no data”)，新数据同样需要进行分类及做平均，然后进行 DA 分析，在菜单栏点击“Comparison”选择“Discriminant Function”(图 48)。

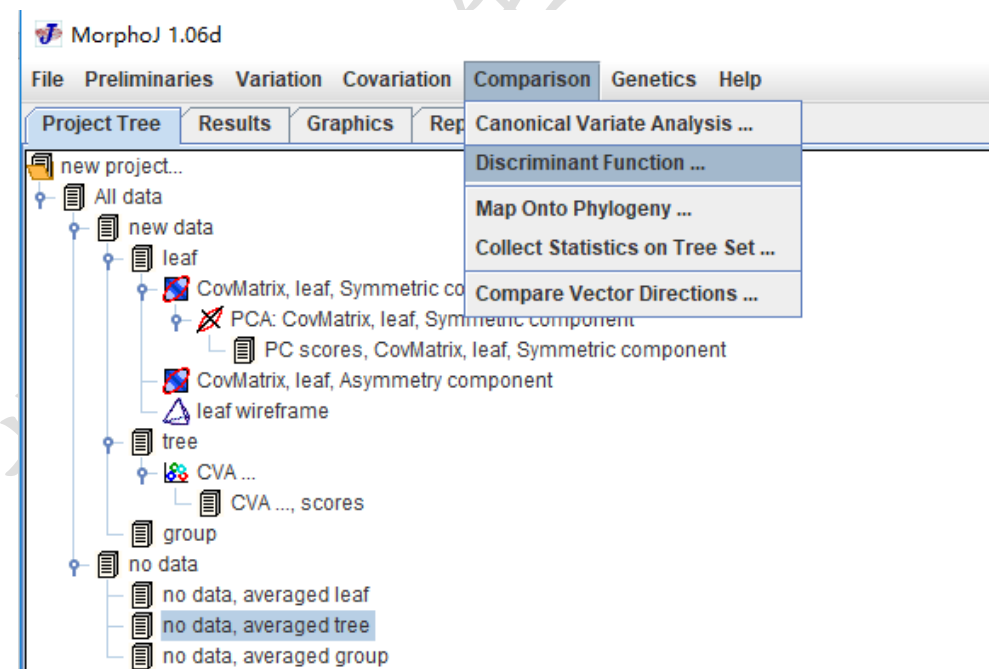


图 48 DA 选项

选择好数据集，数据类型，数据分类水平以及成对比较的组，重复执行 1000 次，点击“Execute”(图 49)。

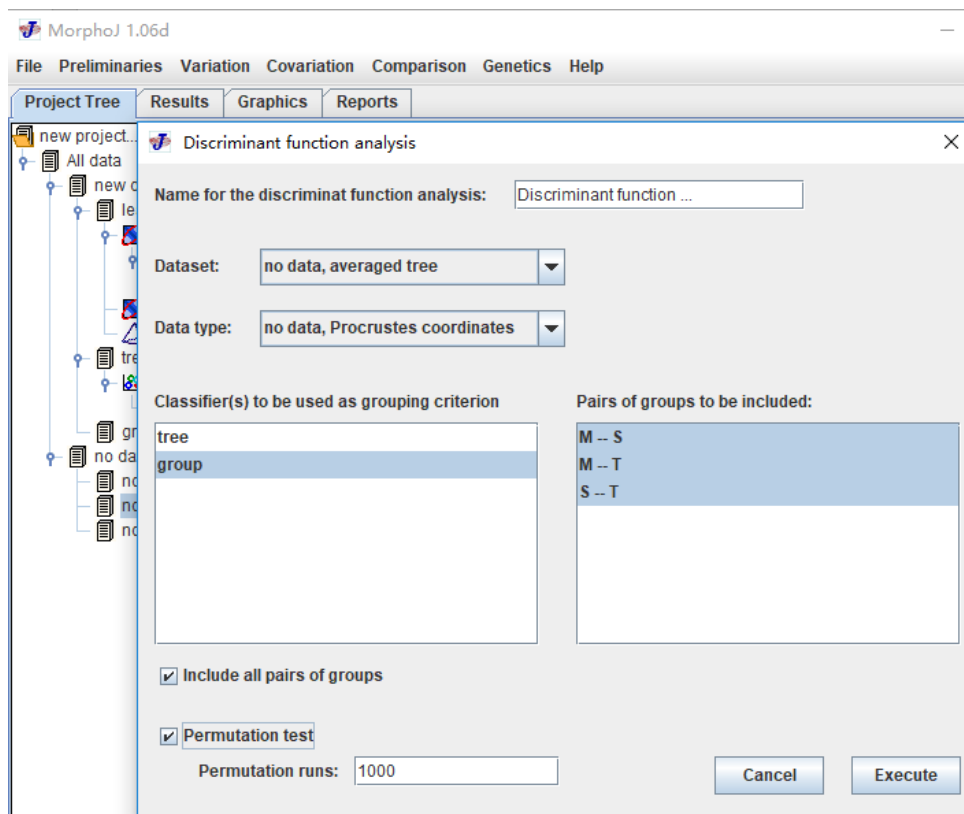


图 49 DA 数据设置

结果显示如图 50，可以根据 DA 轴的值改变叶片形态。

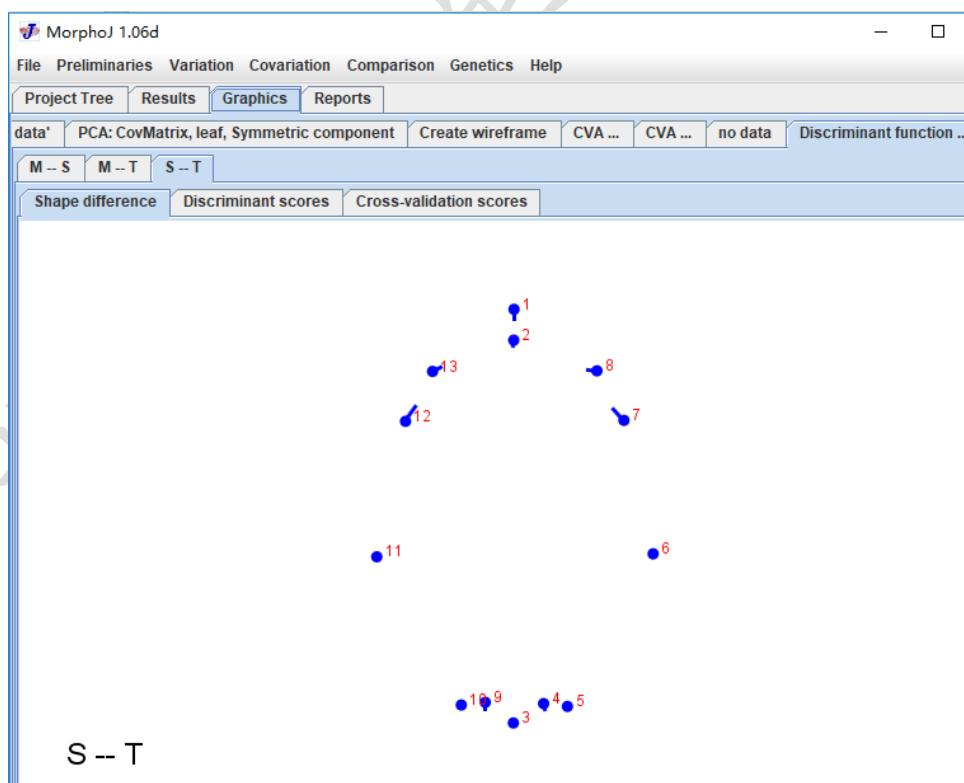


图 50 DA 的棒棒糖图

点击“Cross-validation scores” 结果如图 51 所示，再单击右键，改变条

形图颜色，菜单栏第四行，M-S, M-T, S-T 分别为不同组间两两比较的结果：

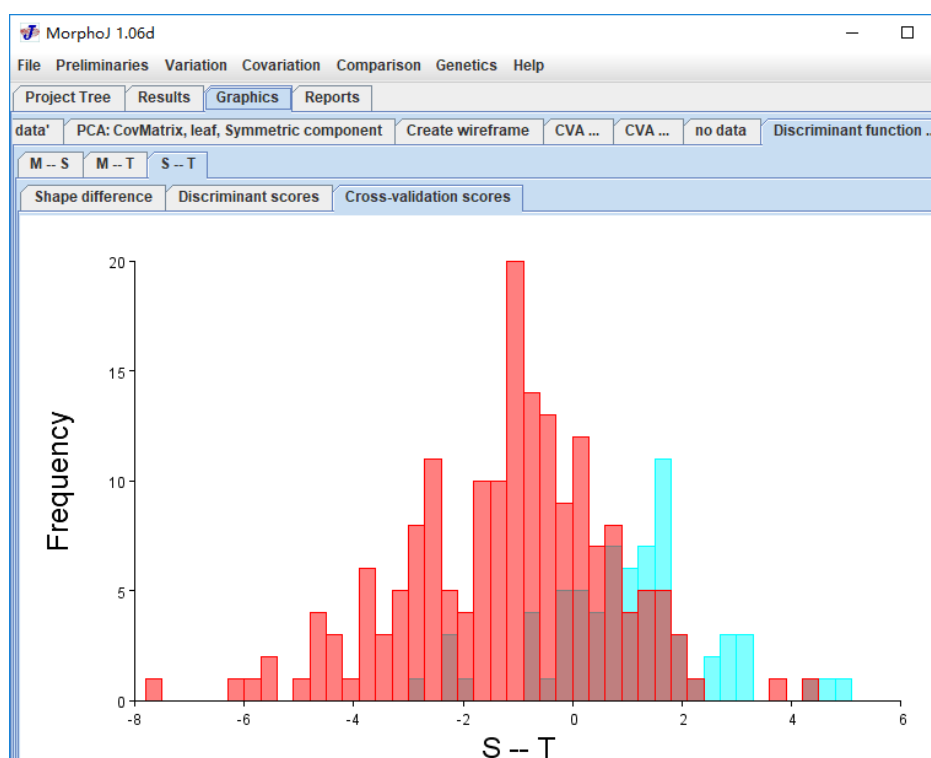


图 51 DA 的条形图

3.4.4 偏最小二乘法(Two-Block Partial Least-Square analysis, 2B-PLS)

2B-PLS 分析是几何形态测量学中用于测试两个定量数据源之间的协方差的多变量测量工具(Marugán-Lobón, 2010)，该方法可用来分析植物的异速生长。异速生长指的是因为大小的变化从而影响形态发生的变化(Klingenberg and McIntyre, 1998; Viscosi 2015)。叶片的大小和形状是可以相互作用相互影响的，从统计学上可以通过测量叶片大小比例的变化衡量异速生长。每一个物种叶片形状的对称和非对称结构均可通过测量实验证明其在统计学上是否显著的形态学变异。通过回归分析 2B-PLS 将形态性状映射到大小的性状上来确认不同的形态变异是否由叶片大小的变化引起，计算结果将给出 *P*-value 值以及 RV 值来确定结果是否显著。

步骤：该分析同样可以通过 Morpho J 软件计算，在菜单栏选中“Covariation”然后选择“Partial Least Squares”再选择“Two Separate Blocks”(图 52)。

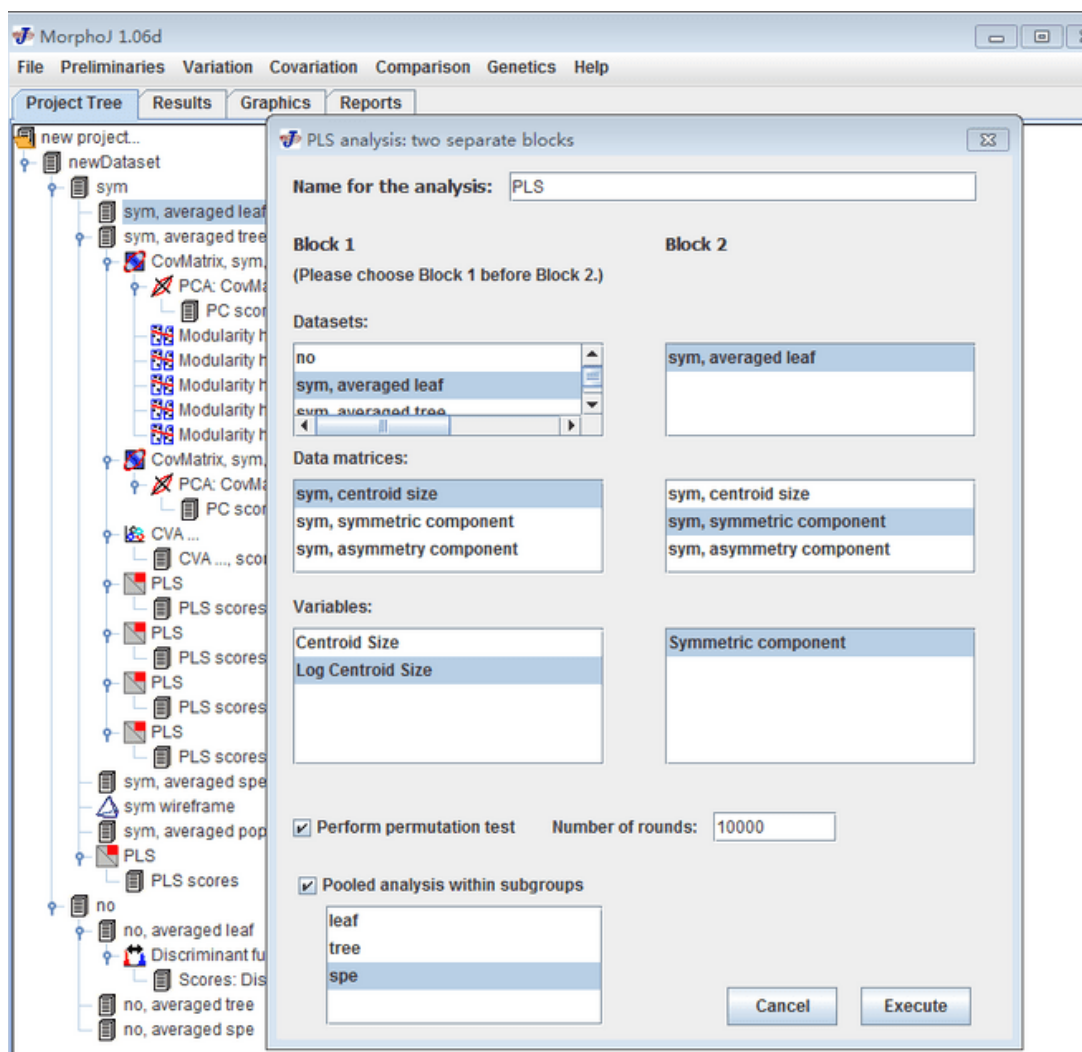


图 52 2B-PLS 数据设置

Block1 和 Block2 的数据类型、矩阵、变量根据自己的问题来设定，本示例中我们以个体水平为单位，选择叶片的对称组分作为 Block1，选择叶片的大小作为 Block2，进行回归分析，以物种来分组，通过 10 000 次迭代来检验其统计学意义。

结果如图 53 所示，可以单击右键，

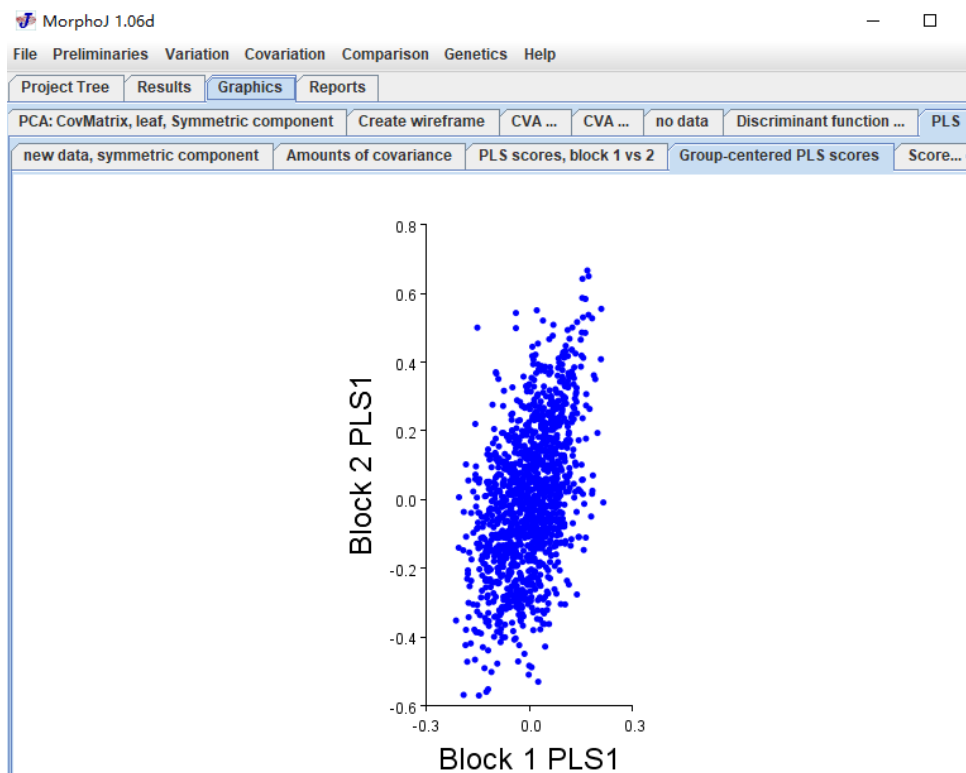


图 53 2B-PLS 散点图

不勾选 ☒ Use Same Scaling for Both Axes，来调节图片横轴和纵轴的长度，使图片看起来更协调，之后再改变点的颜色：

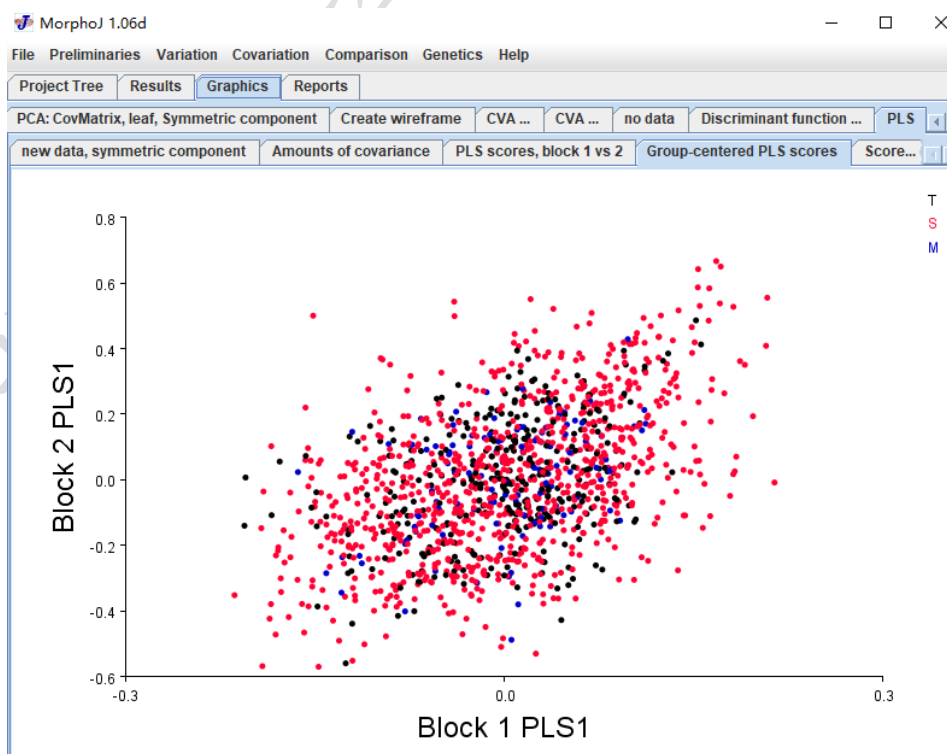


图 54 2B-PLS 分组后的散点图

图 54 是对称结构与叶片大小的回归分析结果，之后再做非对称结构与叶片大小的回归分析，步骤与以上相似，只有选择数据矩阵时选择非对称结构。

3.4.5 普式方差分析(Procrustes ANOVA)

Procrustes ANOVA 是一种量化不同水平的相对变异量的方法，在叶片几何分析中用于误差测量和测试不对称变异的相对量 (Klingenberg and McIntyre, 1998; Klingenberg *et al.*, 2002; Klingenberg, 2003; Viscosi and Cardini, 2011)。它主要使用基于 Procrustes 距离的方法来评估影响的相对大小，并使用 MANOVA 方法进行统计推断 (Morpho J 手册)。

ANOVA 分析需要导入新的数据(命名为“ANOVA data”), 每个叶片需要重复标记两次, 数据格式如右图, 每个叶片命名(ID=种群名_物种_树_叶_重复), 例 ID=MXO_S_T01_L01_1,表示种群 MXO 中 S 物种的第一棵树的第 一片叶子第一次标记的叶形数据(标志点坐标)。

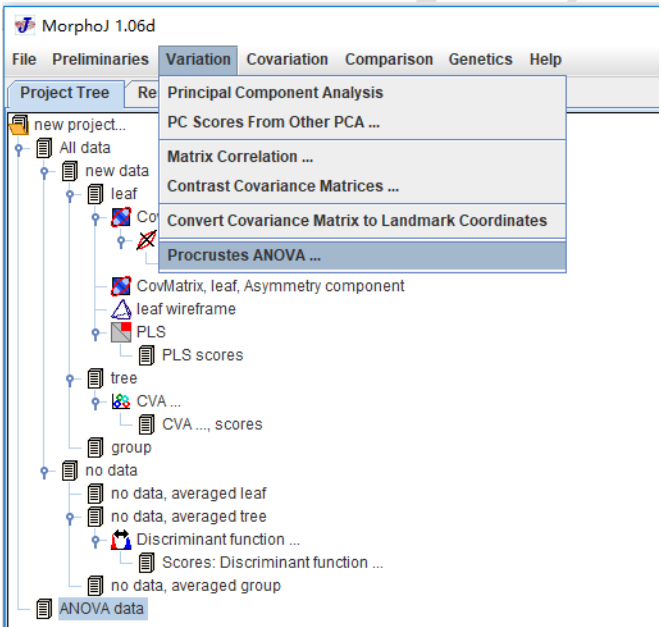


图 55 ANOVA 选项

先将该数据导入(“File→Create New Dataset”), 然后对数据进行预处理，即普氏叠印 (Preliminaries”→“New Procrustes Fit);

LM=13	
1219	2134
1229	2105
1208	1578
1244	1585
1265	1573
1415	1808
1423	1904
1371	1998
1156	1600
1120	1588
1037	1814
1065	1929
1081	2024
ID=MXO_S_T01_L01_1	
LM=13	
1221	2135
1232	2102
1210	1576
1246	1587
1267	1571
1415	1807
1423	1903
1374	1996
1155	1600
1121	1589
1035	1815
1067	1929
1081	2024
ID=MXO_S_T01_L01_2	

检验离群值(“Preliminaries” → “Find Outliers”);
 分离对称和不对称(“Preliminaries” → “Select Landmarks ”);
 数据分类(“Preliminaries” → “Extract New Classifier from ID String”);
 将数据做平均(“Preliminaries” → “Average observations by”)。
 开始 ANOVA 分析, 点击(“Variation” → “Procrustes ANOVA”)(图 55),
 出现图 56 对话框, Individual(一般是指你的样本)选叶片水平, Side 是
 系统指定的, Error 有两个, 一个是检验图像扫描误差, 另一个是检验
 标记点标记的误差, 由于我们每片叶子只扫描了一次, 重复标记了两次,
 所以这只能检验标记误差, Additional main effects 就选剩下的几个
 分类水平(物种, 种群, 树), 然后点击 Execute。

Procrustes ANOVA

Name: Procrustes ANOVA ...

Dataset: ANOVA

Classifiers for the effects:

Individual: LEF

Side: -- none --

Error 1: REP

Error 2: -- none --

Additional main effects:

- none --
- CLADE
- POP
- IND

Cancel Execute

图 56 ANOVA 数据设置

ANOVA 结果如图 57, 叶片大小和形状变异被划分到物种, 种群, 树, 叶和误差几个水平。另外形状变异还被分解到定向不对称(Side)和波动不对称(Ind * Side), 分别表示叶片左右两侧之间的系统差异和微小的随机偏差。最后各个水平对叶大小和形态的影响大小是根据 F 比率和每种效应解释的方差百分比(将效果的平方和(SS)除以总 SS 并将该比率乘以 100)来衡量的。

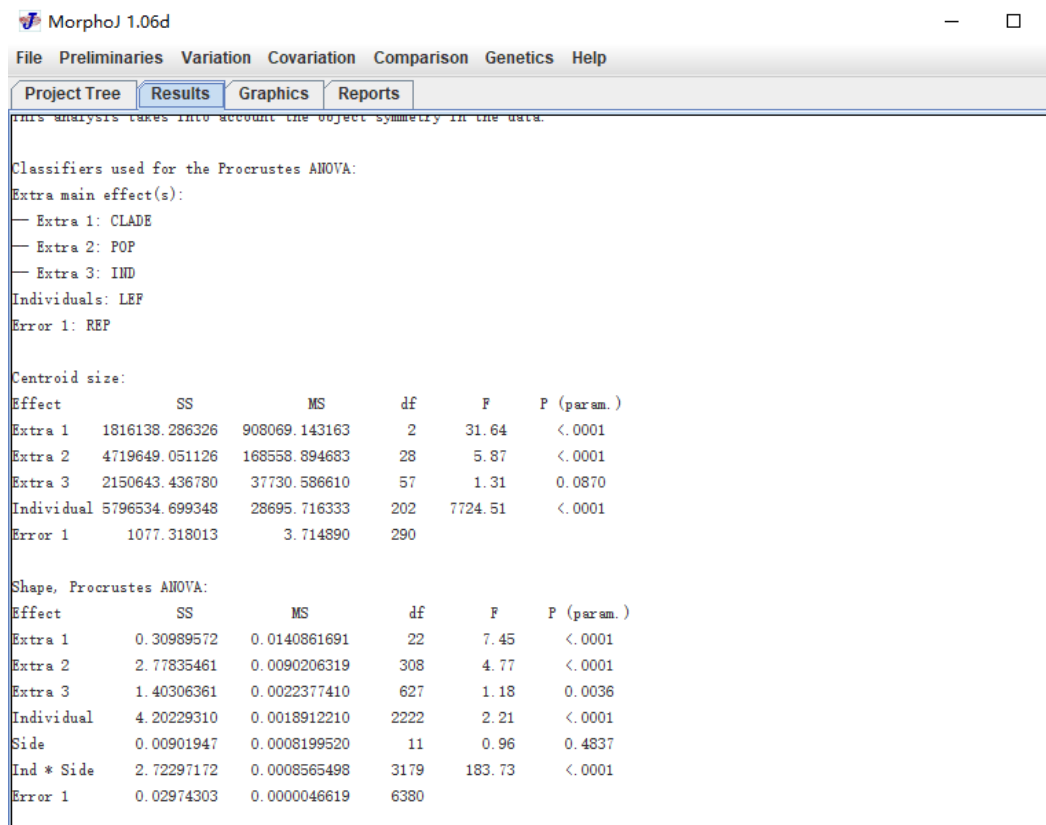


图 57 ANOVA 结果呈现

3.5 导出数据(Export data)

所有的分析都展示在 Project Tree 上(图 58)，叶片形态数据可以与环境数据或者遗传数据做相关性分析，所以我们需要获得叶片形态数据，Morpho J 软件可以将 PCA 分析之后的叶型数据导出。这部分相关分析例如 RDA 未包括进来

步骤：PCA 分析后，如图选中 PCA 数据

CovMatrix, leaf, Symmetric component
 PCA: CovMatrix, leaf, Symmetric component
 PC scores, CovMatrix, leaf, Symmetric component，选择“File”后“Export Dataset”(图 59)。

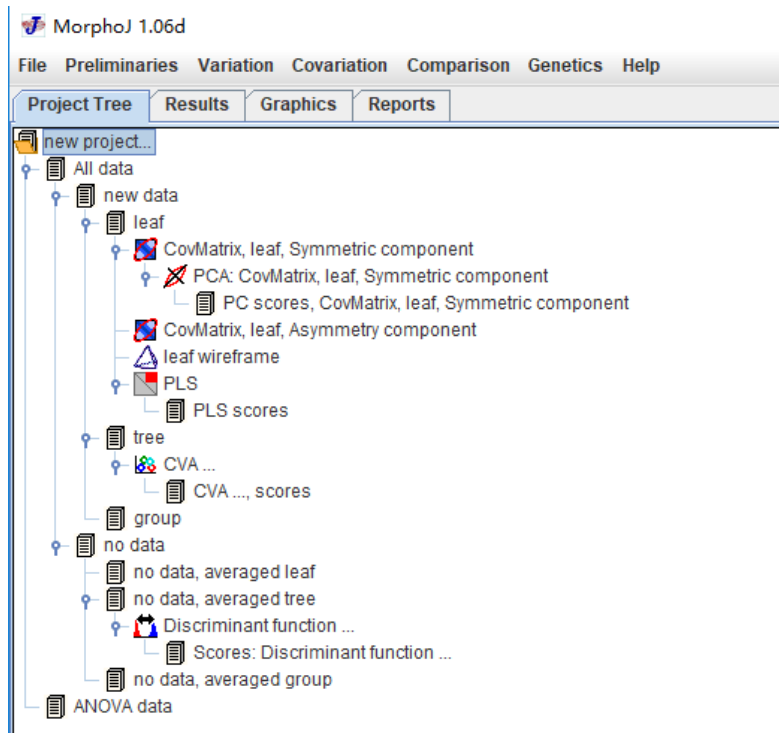


图 58 系统树展示所有的分析

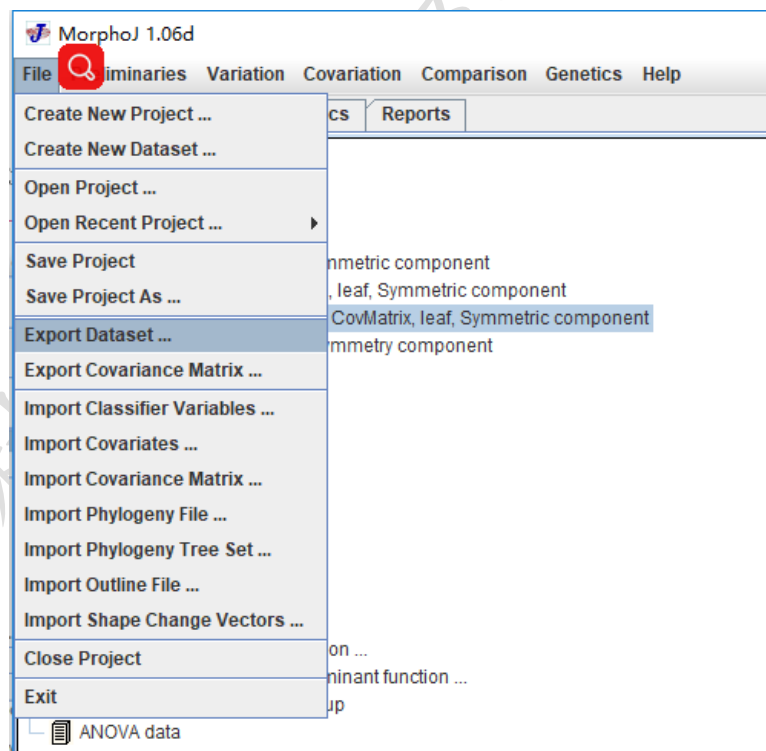


图 59 导出数据

注：导出数据时，若准备导出的数据之前进行了分层，则可以导出不同的分层数据(图 60)：

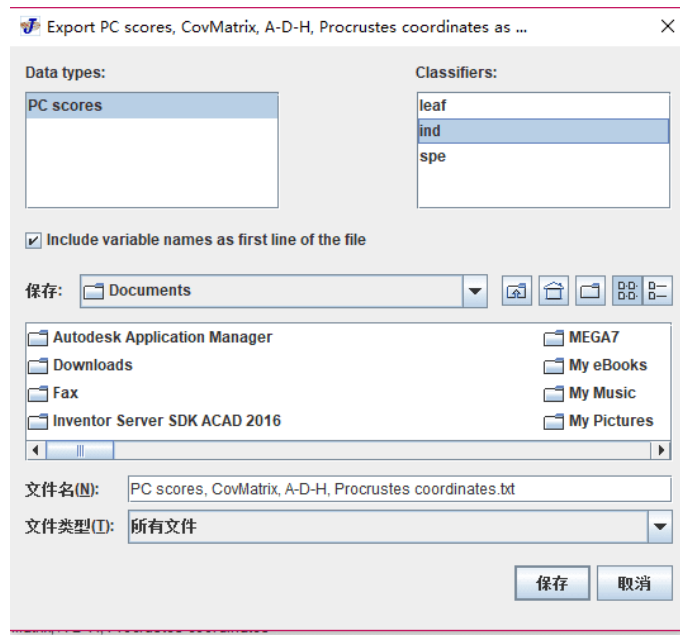


图 60 数据类型选择

四、检测叶形差异是否显著

我们从 Morpho J 中导出了对称组分和非对称组分的 PCA 数据以及叶片大小数据，并利用 R 语言的 agricolae 包进行了多重比较分析，来检测不同物种间叶形对称组分，非对称组分以及叶片大小变异是否显著。运行代码如下：

```
>library(agricolae)
>data=read.table("clipboard",header=T,sep="\t") #导入数据
>model = aov(syPC1 ~ species, data)
>summary(model)
>out =scheffe.test (model,"species")
>out #查看结果
```

五、地理、遗传和气候对叶形的影响

1. 获取叶形、地理、遗传及气候因子

我们首先从 Morpho J 中导出了对称组分、非对称组分和叶片大小的数据，作为叶形数据。微卫星分子 (Simple Sequence Repeats, nSSRs) 标记的 PCA 分析的前三个主成分(PCs)作为中性遗传变量。为了获取气候变量，我们在 WorldClim 数据库 (<http://www.worldclim.org>) 中获取了 1970-2000 近 30 年的气候数据，共 103 个气候因子。并使用 ArcMAP (Environmental Systems Research Institute, 2014) 将采样点的 GPS 坐标(经度和纬度)映射到 30 arc-second 的 ESRI®气候数据网格中，获得 19 个采样点的气候因子。接着，我们通过 R 语言的“usdm”函数包(Naimi *et al.*, 2014)中的方差膨胀因子分析(Variance Inflation Factor, VIF)，从这 19 个气候变量中去除部分相关性较强的因子，最终保留 3-5 个气候因子。保留的气候因子、SSR 分子标记的前三个 PCs 和经纬度及海拔数据作为叶形的预测因子，被用于后续的冗余分析。

2. 叶形与地理、遗传和气候的冗余分析

为了调查地理、遗传和气候因子是否解释了一些叶形变异(对称组分、非对称组分和叶片大小)，我们使用 R 语言的函数包“usdm”(Oksanen *et al.*, 2017)进行一系列基于距离的冗余分析(distance-based redundancy analyses, dbRDAs)，包括边际测试(Marginal tests)和条件测试(Conditional tests) (Legendre and Anderson, 1999)。我们使用方差划分可以运行一个包括地理、遗传和气候解释变量的边际模型(地理+遗传+气候)。另外也可以通过控制另外两种因素运行三个不同的条件 dbRDAs 模型，分别用来测试纯的地理效应、纯的遗传效应和纯的气候效应：(1)控制遗传和气候效应的有关地理变量的条件模型(地理+Condition(遗传+气候))；(2)控制地理和气候效应的有关遗传变量的条件模型(遗传+Condition(地理+气候))；(3)控制地理和遗传效应的有关气候变量的条件模型(气候+Condition(地理+遗传))。最后，通过方差分析函数(anova function)计算 dbRDA 的 F 值、解释变异的百分比和 p 值来评估每个独立因素对叶形变异的影响(Oksanen *et al.*,

2017)。

运行代码如下：

1. 气候，地理和遗传对叶形的影响

```
###导入数据
##获取遗传数据
>library(vegan)
>data=read.table("clipboard", header=F) #导入 SSR 归一化数据
>p=prcomp(data, retx = TRUE, center = TRUE, scale. = T, tol = NULL)
>summary(p) #PCA 结果作为后续分析的遗传数据
>predict(p) #计算标准化的主成分得分
>gene = read.table("clipboard", header=T) #导入遗传数据
##导入叶形数据
>sy= read.table("clipboard", header=T) #导入从 Morpho J 中导出的
对称组分 PCA 数据
>asy= read.table("clipboard", header=T) #导入从 Morpho J 中导出
的非对称组分 PCA 数据
>size= read.table("clipboard", header=T) #导入从 Morpho J 中导出
的叶片大小数据
##筛选气候数据
>library(usdm)
>data=read.table("clipboard", header=F) #导入气候因子数据
>x=vifcor(data,th=0.5) #通过设置参数(th=0.*)来调节保留的因子个数
>x #查看保留了哪些气候因子
##导气候和地理数据
>envo = read.table("clipboard", header=T)
###数据标准化
>sy1= decostand(sy, "max")
>asy1= decostand(asy, "max")
>size1= decostand(size, "max")
>envo1= decostand(envo, "max")
>gene1= decostand(gene, "max")
##建立模型
>mod1=varpart(asy1,~ bio02 + bio04 + bio08 + bio14+ bio15, ~ lon + lat,
gene1, data = envo1)
>mm1 = model.matrix(~ bio02 + bio04 + bio08 + bio14+ bio15, envo1)
[, -1]
```

```

> mm2 = model.matrix (~ lon+ lat, env1) [, -1]
> mod2 = varpart (asy1, mm1, mm2, gene1)
###运行模型
> sy_dbrda = capscale (asy1 ~ mm1 + mm2 + as.matrix (gene1), dist =
"euclidean")
> summary (sy_dbrda) #查看结果
> sy_anova = anova (sy_dbrda, by = "term",permutations = 999)  #方差
分析
> sy_anova  #查看方差分析结果
> sy_dbrda  #得 total 结果
###控制地理和遗传的影响
> sy_dbrda1 = capscale (sy1 ~ mm1 + Condition(mm2) + Condition
(as.matrix (gene1)), dist = "euclidean")
> sy_anova1 = anova (sy_dbrda1, by = "term",permutations = 999)
> sy_anova1
> sy_dbrda1  #得 total 结果
###控制气候和遗传的影响
> sy_dbrda2 = capscale (sy1 ~ mm2 + Condition(mm1) + Condition
(as.matrix (gene1)), dist = "euclidean")
> sy_anova2 = anova (sy_dbrda2, by = "term",permutations = 999)
> sy_anova2
> sy_dbrda2  #得 total 结果
###控制气候和地理的影响
> sy_dbrda3 = capscale (sy1 ~ Condition (mm1) + Condition(mm2) +
as.matrix (gene1), dist = "euclidean")
> sy_anova3 = anova (sy_dbrda3, by = "term",permutations = 999)
> sy_anova3
> sy_dbrda3  #得 total 结果
####注： 以上是基于对称组分的代码，基于非对称组分和叶片大小的
代码，只需循环执行建立和运行模型，修改标黄的部分即可。标黄
的部分为叶形数据，这里是以 sy1：从 Morpho J 中导出的对称组分 PCA
数据为例。
2. 具体的气候，地理和遗传因子对叶形的影响
###导入数据
> sy= read.table ("clipboard", header=T)  #导入从 Morpho J 中导出的
对称组分 PCA 数据

```

```

> asy= read.table ("clipboard", header=T) #导入从 Morpho J 中导出的
的非对称组分 PCA 数据
> size= read.table ("clipboard", header=T) #导入从 Morpho J 中导出的
的叶片大小数据
> factors = read.table ("clipboard", header=T) #导入解释因子，包括气
候，地理和 SSR 的前 3 个 PC 值
###运行模型
> sy_dbrda = capscale (sy1 ~ bio02 + bio04 + bio08 + bio14+ bio15 + lon
+ lat + PC1 +PC2 +PC3, factors, dist = "euclidean")
> summary (sy_dbrda) #查看结果
> sy_anova = anova (sy_dbrda, by = "term",permutations = 999) #方差
分析
> sy_anova #查看方差分析结果
> sy_dbrda #得 total 结果
##控制地理和遗传的影响
> sy_dbrda1 = capscale (sy1 ~ bio02 + bio04 + bio08 + bio14+ bio15+
Condition (lon + lat) + Condition (PC1 +PC2 +PC3), factors, dist =
"euclidean")
> sy_anova1 = anova (sy_dbrda1, by = "term",permutations = 999)
> sy_anova1
> sy_dbrda1 #得 total 结果
##控制气候和遗传的影响
> sy_dbrda2 = capscale (sy1 ~ lon + lat + Condition (bio02 + bio04 + bio08
+ bio14+ bio15) + Condition (PC1 +PC2 +PC3), factors, dist = "euclidean")
> sy_anova2 = anova (sy_dbrda2, by = "term", permutations = 999)
> sy_anova2
> sy_dbrda2 #得 total 结果
##控制气候和地理的影响
> sy_dbrda3 = capscale (sy1 ~ PC1 +PC2 +PC3 + Condition (bio02 +
bio04 + bio08 + bio14+ bio15) + Condition (lon + lat), factors, dist =
"euclidean")
> sy_anova3 = anova (sy_dbrda3, by = "term", permutations = 999)
> sy_anova3
> sy_dbrda3 #得 total 结果
###注： 以上是基于对称组分的代码，基于非对称组分和叶片大小的
代码，只需循环执行建立和运行模型，修改标黄的部分即可。标黄的

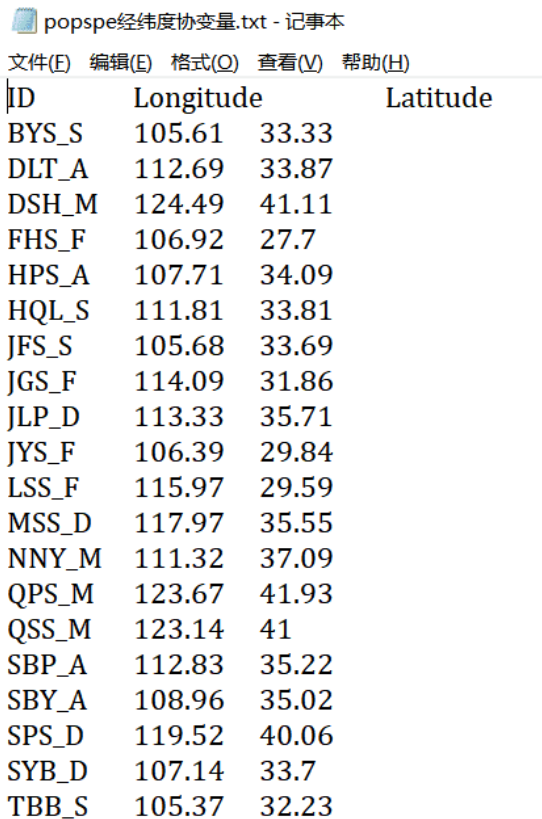
```

部分为需要进行分析的叶形数据，这里是以 syl：从 Morpho J 中导出的对称组分 PCA 数据为例。

六、导入协变量(Import Covariates)

1.导入文件格式(Input file formats)

用于叶形分析软件 Morpho J 的数据输入格式为 .txt 格式，所以在导入协变量之前需要将标志点的坐标数据文件转化成 Morpho J 支持的文件类型。该文件可以在 Excel 中准备，并保存为以制表符或逗号分隔的文本，文件格式如图 61:



ID	Longitude	Latitude
BYS_S	105.61	33.33
DLT_A	112.69	33.87
DSH_M	124.49	41.11
FHS_F	106.92	27.7
HPS_A	107.71	34.09
HQL_S	111.81	33.81
JFS_S	105.68	33.69
JGS_F	114.09	31.86
JLP_D	113.33	35.71
JYS_F	106.39	29.84
LSS_F	115.97	29.59
MSS_D	117.97	35.55
NNY_M	111.32	37.09
QPS_M	123.67	41.93
QSS_M	123.14	41
SBP_A	112.83	35.22
SBY_A	108.96	35.02
SPS_D	119.52	40.06
SYB_D	107.14	33.7
TBB_S	105.37	32.23

图 61 协变量数据的整理

第一行包含协变量的名称。第一个条目(上面的例子中的“ID”)将被忽略，但是剩下的条目将被用作协变量的名称。每个协变量都应该有一个不同的名字，必须匹配精确，包括拼写和大小写字母，数字可以是整数也可以有小数(例如:“34”，“30.5”)，也可以用科学记数

法(例如: “1.753 e-09”)。

2. 导入协变量

打开 Morpho J, 在对话框的顶部, 有一个标签指示已选择的数据集。协变量将只导入到这个数据集。在 “Project Tree” 选项卡中选择数据集, 然后在 “File” 中点击 “Import Covariates”, 导入数据, 如图 62:

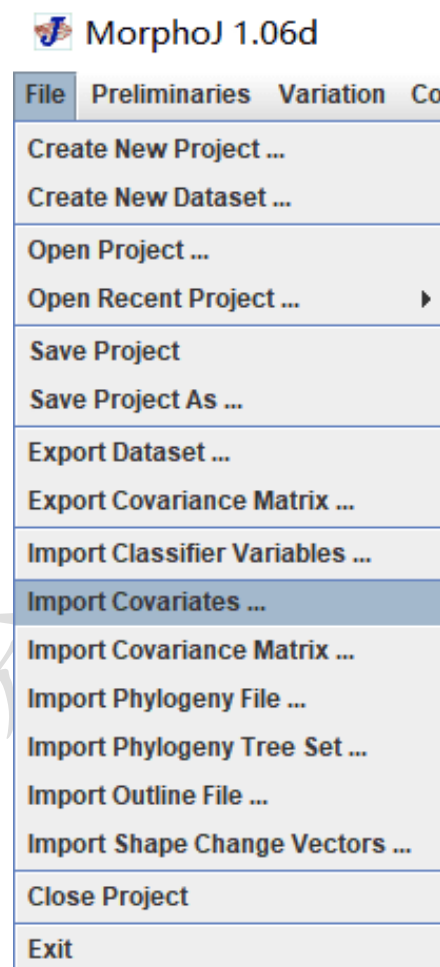


图 62 协变量导入

将调用以下对话框:

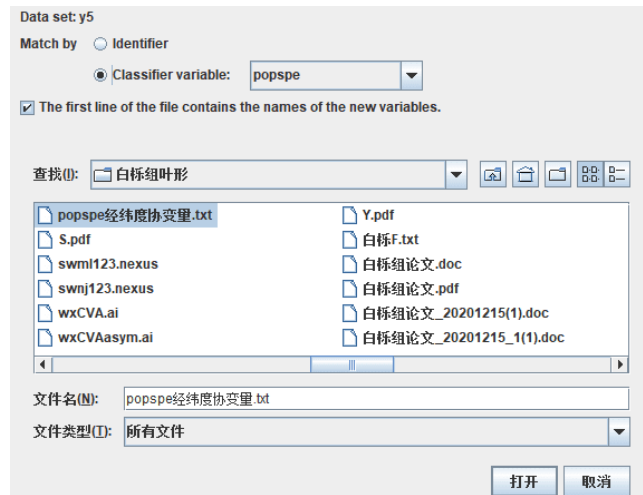


图 63 导入协变量至已经编辑好的分类器中

其中导入有两个分类选择,分别是 Identifier 和 Classifier variable,用于选择协变量如何与数据集中的观察值匹配。如果选择了 Identifier,则 Identifier 变量的值将用于匹配。如果选择了 Classifier variable,则可以从下拉菜单中选择我们已经定义过的分类器,并使用其值进行匹配。导入协变量后,通过在初步菜单中使用编辑协变量来再次检查结果。导入成功后选择该数据集在 Preliminaries 中点击“Edit Covariates”查看编辑协变量:

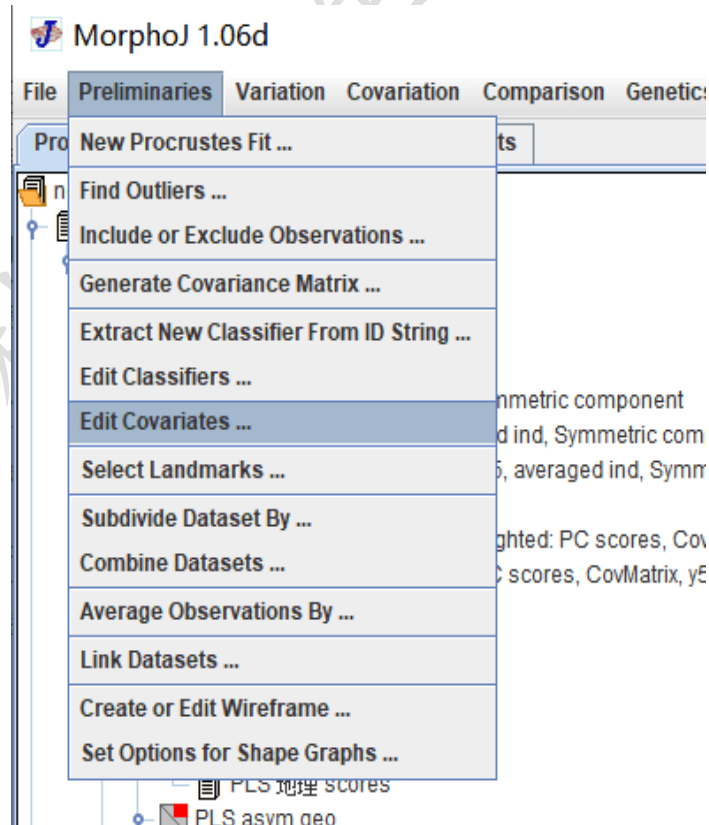


图 64 编辑协变量

导入成功如下图所示，在此界面可进行复制粘贴，删除等操作，并单击 Accept:

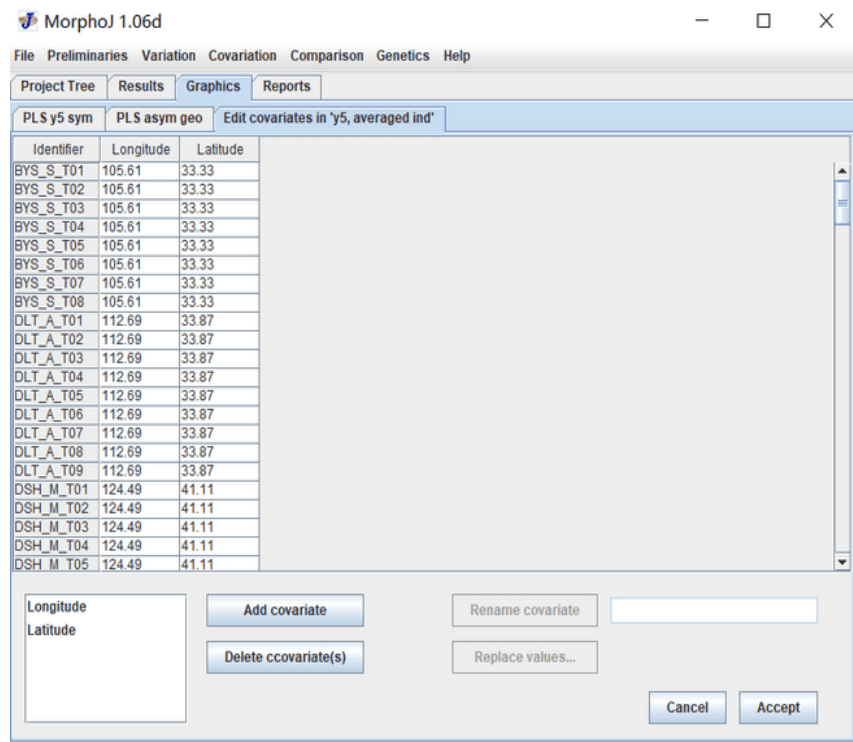


图 65 查看接受协变量

3.双标图(Biplot)

在 Morpho J 中，双块偏最小二乘(2B-PLS)的实现使用了两块变量间协方差矩阵的奇异值分解(e.g. Rohlf and Corti, 2000)。Biplot（双标图）是一种图形表示形式，最初开发它是为了在同一图中结合观测变量和观测结果的信息，但在 Morpho J 中仅用于说明不能以形状变化的形式来可视化的变量。

步骤：点击“Covariation”→“Partial Least Squares”→“Two separate blocks”，出现图 66 对话框：先选 Block1 为协变量，再选 Block2 为对称组分或非对称组分，点击执行

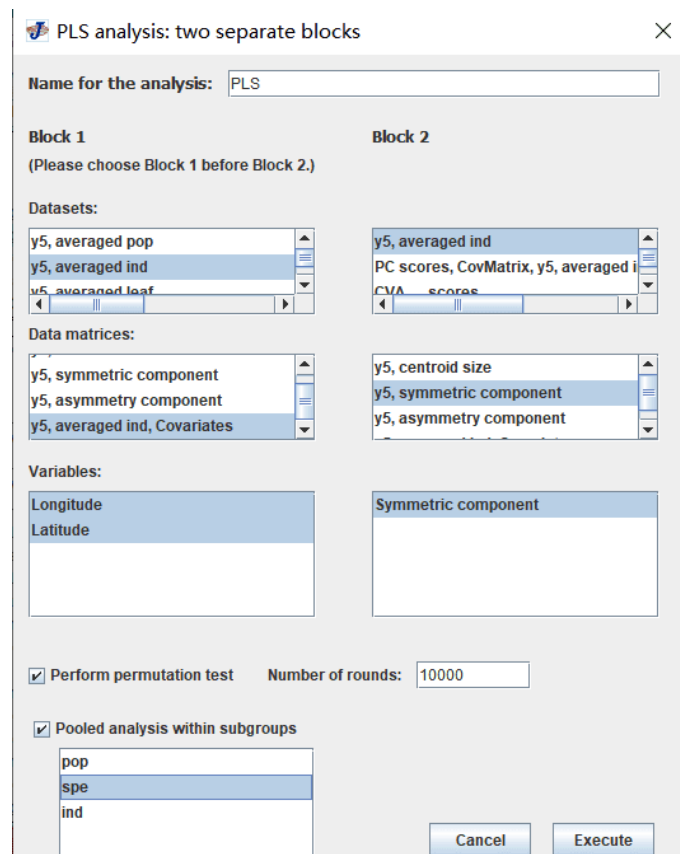


图 66 双标图(Biplot)参数选择

结果如图 67:

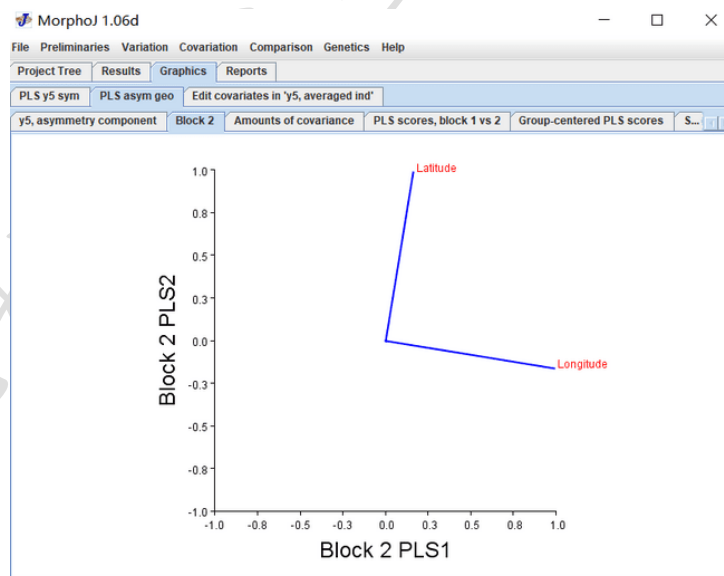


图 67 双标图(Biplot)结果

该图显示了两条 PLS 轴(水平方向的 PLS1 和垂直方向的 PLS2)。可以看出, PLS1 主要是变量“经度”, PLS2 轴主要是“纬度”。

七、映射到系统发育树(Map Onto Phylogeny)

1. 导入系统发育树

Morpho J 可以使用平方变化简约法(squared-change parsimony)将形状数据映射到系统发育树上 (Maddison, 1991)。系统发育文件以 NEXUS 文件格式导入 Morpho J, 再导入数据之前确保 PCA 分析已经完成。

NEXUS 文件是由 Mesquite、PAUP*、MacClade 等程序生成的, 格式如下图:

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

#NEXUS

BEGIN TREES;

Tree tree = ((F:0.00029610,M:0.00014854)1.0000:0.00026224,S:0.00014932,
(A:0.00053028,D:0.00018217)0.9910:0.00008139);

END;

图 68 导入数据格式

在“File”中选择“Import Phylogeny File”导入文件:

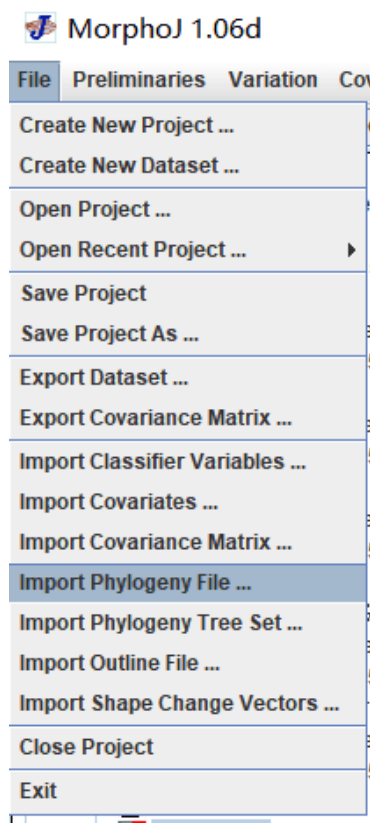


图 69 导入上一步 NEXUS 文件

2. 独立变量在系统发育上映射

将形态测量数据映射到系统发育树上，点击“Comparison”中的“Map Onto Phylogeny”。将出现以下对话框：

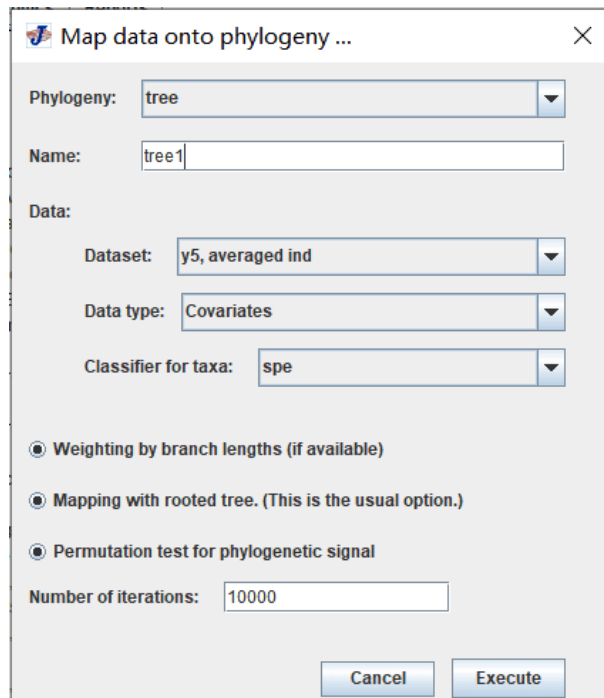


图 70 将独立变量映射到系统发育树上
点击“Excute”得到如下结果：

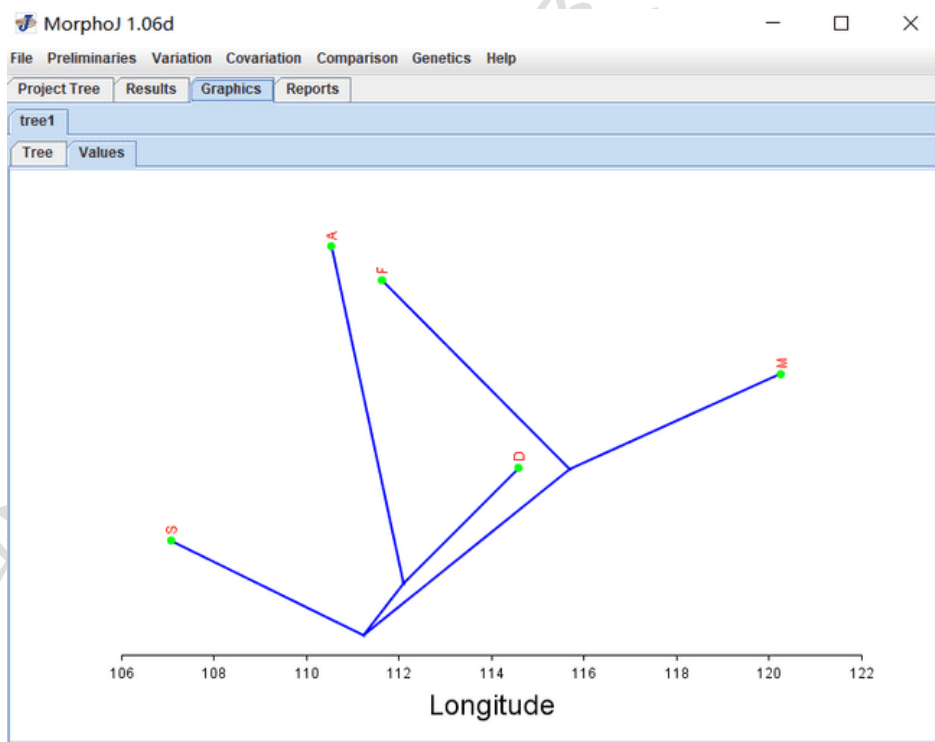


图 71 系统发育树结果

横轴提供变量的值(这里是质心大小)，而垂直方向对应于从树的根开始的累积分支长度(如果映射是通过加权的平方变化简化完成的)。如果分支长度可以被解释为进化时间，那么这个图就代表了进化谱系沿着变量轴的运动。

在该界面任意地方左击鼠标，就会出现一个弹出式菜单。该弹出菜单可用于改变图形的方向，比例因子，PC 轴显示，导出图形文件并保存。

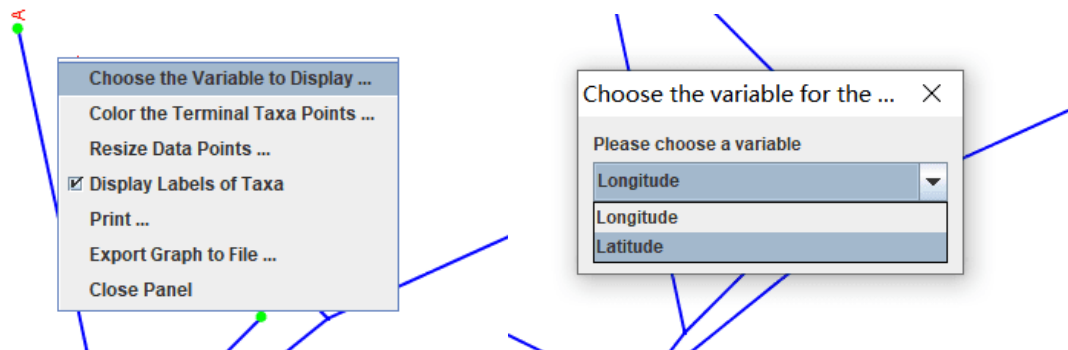


图 72 更换变量

如果数据集包含多变量分析的得分，Morpho J 将显示这些得分的分类单元平均值的散点图：

首先选择图 73 PC scores 文件：

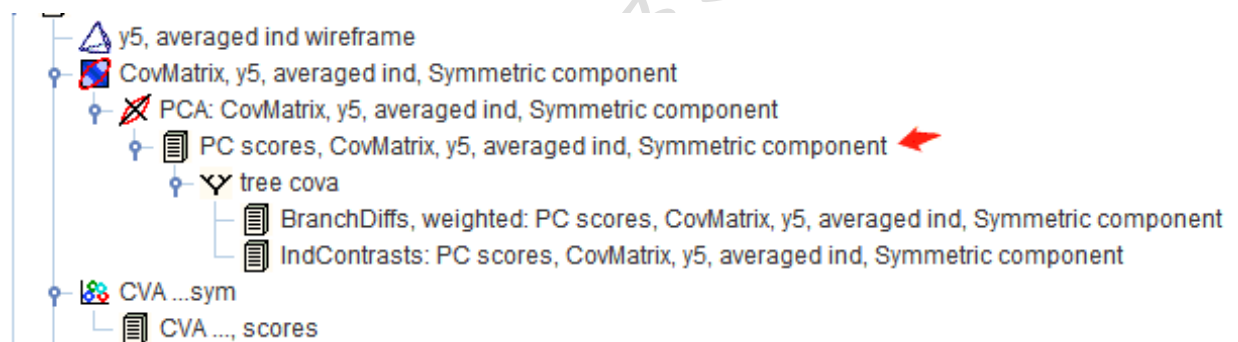


图 73 选择 PC scores 文件（红色箭头）

再进行映射，依次点击“Comparison”→“Map Onto Phylogeny”，做如下图操作：

Map data onto phylogeny ...

Phylogeny: tree

Name: tree

Data:

Dataset: PC scores, CovMatrix, y5, averaged ind, S...

Data type: PC scores

Classifier for taxa: spe

☒ Weighting by branch lengths (if available)

☒ Mapping with rooted tree. (This is the usual option.)

☒ Permutation test for phylogenetic signal

Number of iterations: 10000

Cancel Execute

图 74 PC scores 文件映射到系统发育树

得出结果:

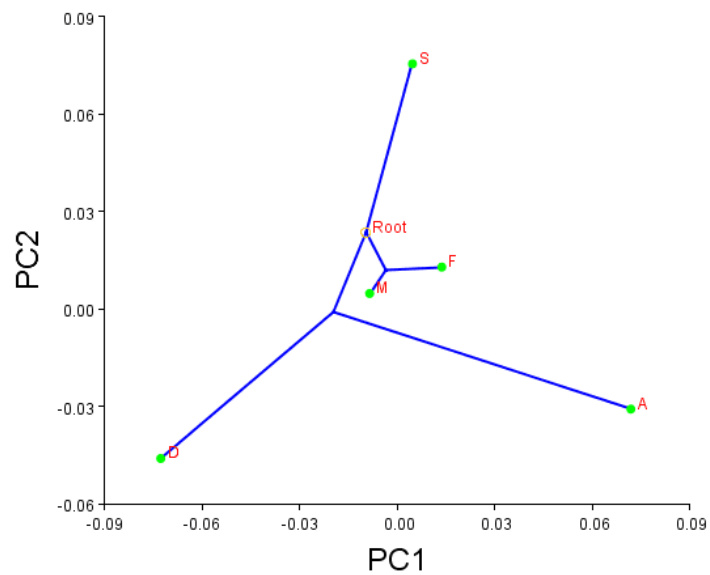


图 75 PC scores 文件映射系统发育树结果

八、假定模块化评估(Modularity: Evaluate Hypothesis)

模块是由许多交互作用形成的高度整合的单元,但又相对独立于其他这样的单元(Klingenberg, 2008)。在形态测量分析中,形态模块化是指不同模块之间的弱共变和模块内部的强共变。注意:模块化分析适合于单一物种。

为了评估模块化假设,可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。Klingenberg 等(2009)提出 RV 系数, RV 系数是平方相关的多元类比(Escoufier, 1973), 它的取值范围从 0 (完全不相关) 到 1 (完全相关)。通过比较与可选分区子集之间的 RV 系数来衡量两个或多个标志点子集之间相关性。如果一个假设划分与模块之间的真正边界一致,那么子集之间的相关性应该是最小的。

1.选择标志点子集

在 Project Tree 选项卡中选择一个协方差矩阵(通过这个图标  在项目树中可以识别)。在“Covariation”中点击“Modularity: Evaluate Hypothesis”, 如图 76:

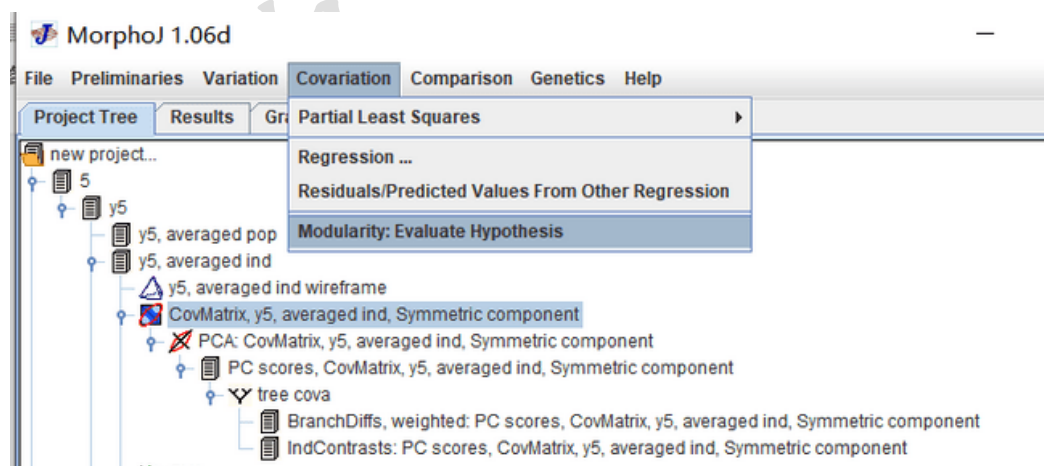


图 76 选择 Covariation 菜单下的 Modularity: Evaluate Hypothesis 然后出现下图对话框:

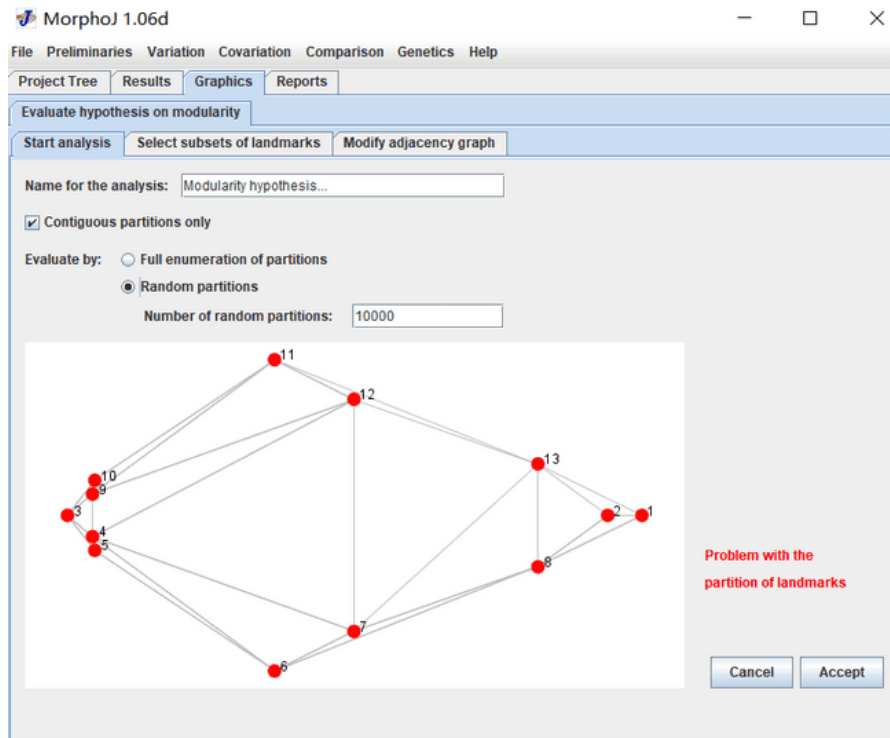


图 77 Modularity: Evaluate Hypothesis

第一个元素是一个用于指定分析名称的文本字段 (Name for the analysis)，可自行定义名称。在它下面有一个复选框，用于指定在比较中只包含标志点的空间连续分区还是包含所有分区。是否选择这个选项取决于研究的生物学背景(Klingenberg, 2009)。下一个选择是分析应该使用所有可能分区的完整枚举还是大量随机分区；如果选择了 Full enumeration of partitions 选项，则限制为空间连续的分区，如果选择 Random partitions 选项，将激活用于输入随机分区数量的文本字段。Klingenberg 建议 20 个标志点以内可以实现完整枚举，除此之外推荐 10,000 数量级的随机分区进行比较，提供 RV 系数分布的合理表征。

2.修改邻接图

点击“Accept”之后，我们在第三个选项 Modify adjacency graphs 中，将对称组分中的邻接点的线补充完整（可以用鼠标点击标志点直接进行连接，也可以在右侧操作界面点击标志点名称增加连接线段）并点击“Accept”：

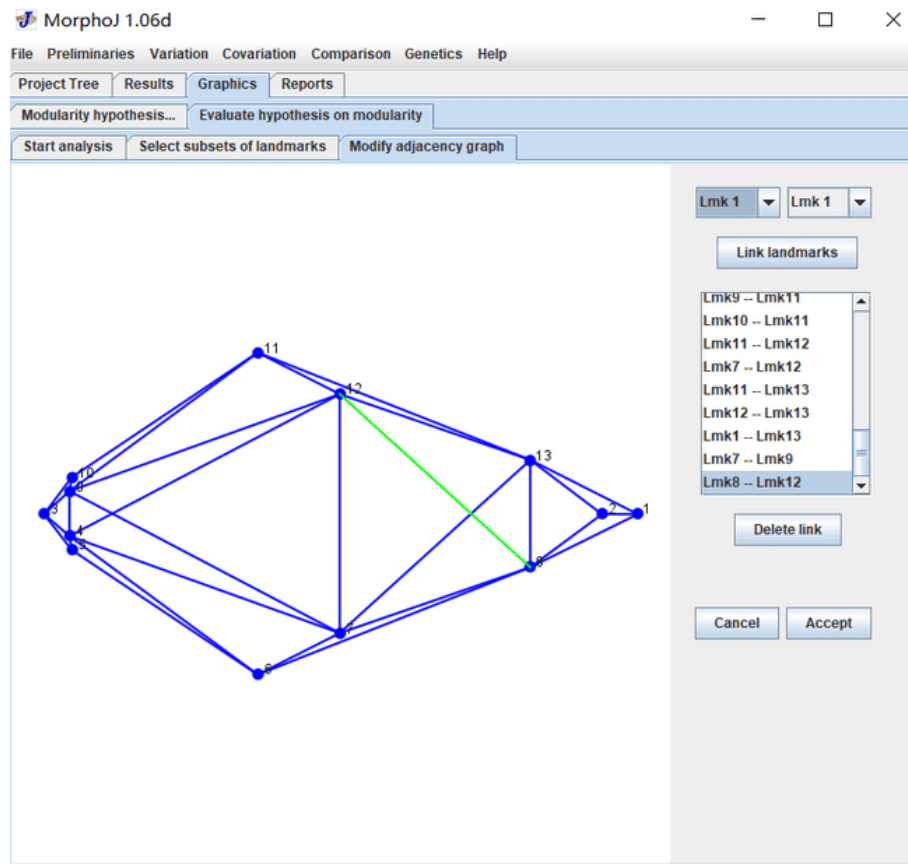


图 78 修改邻接图

点击“Accept”后我们在跳转页面中右上角（Number of subsets）选择分区数量，每个分区至少包含 2 对标志点，左侧图是一个标志点配置的图，显示了作为模块化假设指定的标志点子集(彩色点)和用于建立空间连续分区的邻接图(仅在选择该选择时使用)

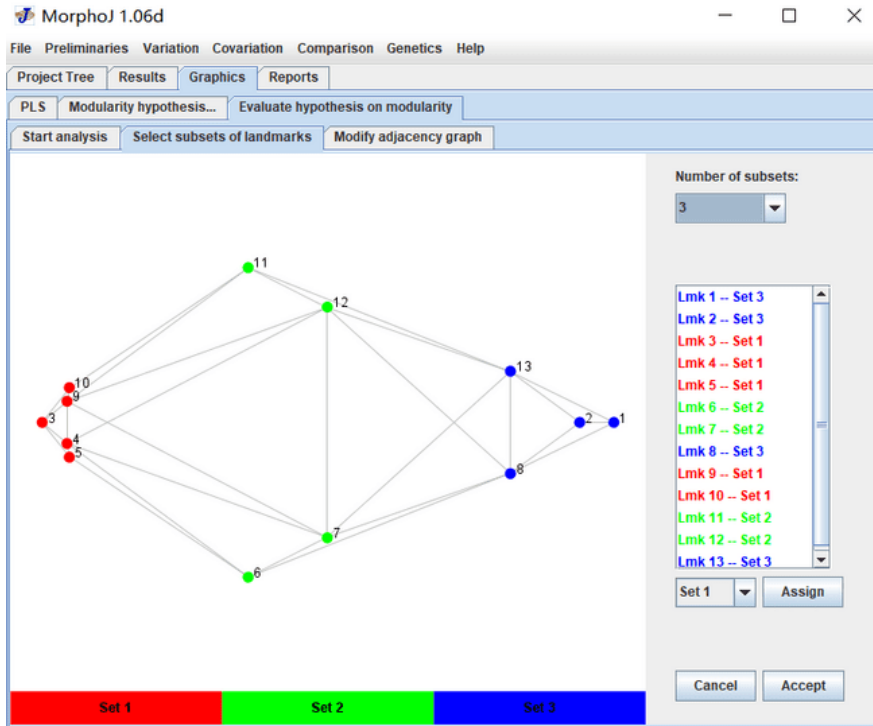


图 79 选择分区

点击“Accept”后跳转到如下界面：

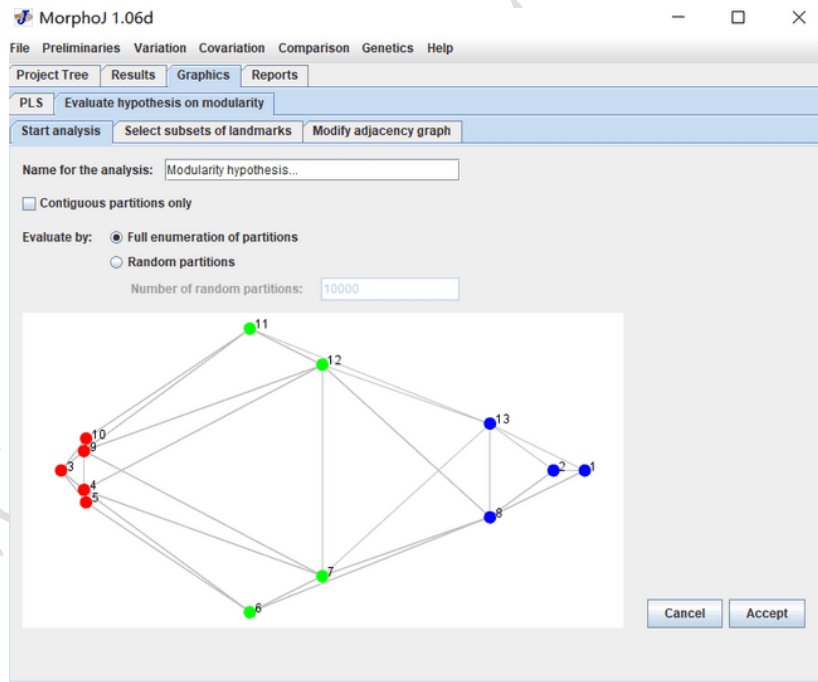


图 80 选择分区子集

3.比较分区(Comparing partitions)

点击“Accept”之后出现 RV 系数分布直方图：

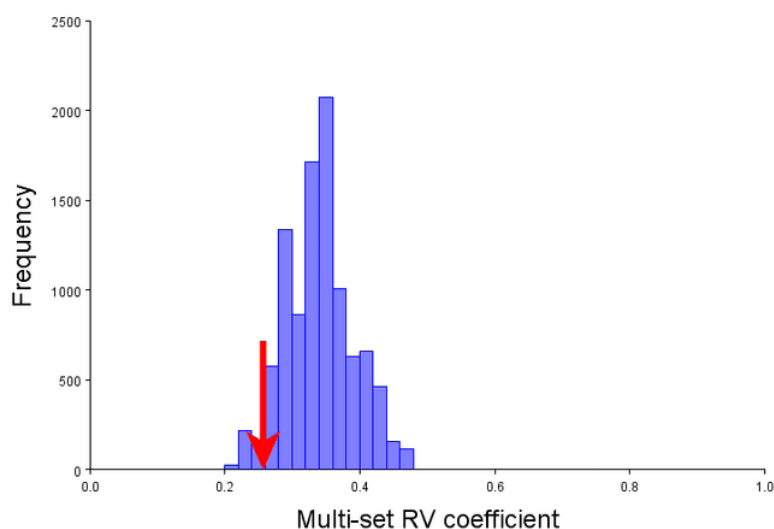


图 81 RV 系数的频数分布直方图

评估将标志点配置划分为两个模块的假设的最直接的方法是：计算所有可能的划分为适当大小子集的分区 RV 系数。如果模块化假设成立，则根据该假设分区的 RV 系数应该是最低值，或者至少应该接近所有分区的 RV 系数分布的下限。

当红色箭头在整个分区的接近左侧末端时，这一结果与划分的子集是独立模块的假设一致。当箭头在右侧时则与假定的划分相反。将设置不同分区的 RV 系数进行比较，可以提供比统计测试 P 值更多的信息。当 RV 系数小于或等于设置的分区的 RV 值的分区比例，可以解释为 P 值的模拟(The proportion of partitions for which the RV coefficient is less than or equal to the RV value for the partition of interest, which can be interpreted as the analog of such a P-value, is one piece of information that can be obtained. Klingenberg, 2009)

九、深度学习(Deep Learning)

机器学习指一类可以通过检测数据中的模式自动生成预测模型的算法(Olden *et al.*, 2008)。深度学习 (Deep Learning) 作为机器学习的一个分支，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)。使用一般的学习过程，深度学习算法能够自动地从数据中检测和提取特征。这种自动学习过程是通过将数据分解成多个层来实现的，每个层都有不同的抽象级别，允许算法学习表示数据的复杂特征(Christin *et al.*, 2019)。

1.选择体系结构

生态学中许多层面的研究通常提供高度维度的数据集,通过这些数据集,深度学习尤其准确和高效。在实践中,使用不同的深度学习架构可以实现这些结果。其中,使用最广泛的是卷积神经网络(convolutional neural network, CNN),由于其在图像分类中的性能,该架构有助于普及深度学习(Krizhevsky *et al.*, 2017)。CNN 是因其易于训练和良好的泛化能力而受到最多关注,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。

2.选择框架

随着深度学习的快速发展,创建了大量框架。所有框架(如下图)在易用性、可用资源、架构支持、可定制性或硬件支持方面提供了不同的方案(Christin *et al.*, 2019)。

Framework	Language	ONNX Support	Implementation of popular networks	Pre-trained models available	URL
Tensorflow	Python, C/C++, R, Java, Go, Julia	Needs conversion from external tool	Yes	Yes	https://www.tensorflow.org/
PyTorch	Python	Yes	Yes	Yes	https://pytorch.org/
Keras ^a	Python, R	Needs conversion from external tool	Yes	Yes	https://keras.io/
Microsoft Cognitive Toolkit (CNTK)	C#, C++, Python	Yes	Yes	Yes	https://docs.microsoft.com/en-us/cognitive-toolkit/
Deeplearning4J	Java, Scala	Basic support	Yes	Yes	https://deeplearning4j.org/
MATLAB + Deep Learning Toolbox	MATLAB	Yes	Yes	Yes	https://www.mathworks.com/products/deep-learning.html
Apache MXNET	C++, Python, Julia, Matlab, JavaScript, Go, R, Scala, Perl	Yes	Yes	Yes	http://mxnet.incubator.apache.org/
PlaidML	Python	Yes	Yes (via keras)	Yes (via keras)	https://github.com/plaidml/plaidml

图 82 不同框架对比图

3.选择模型

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。

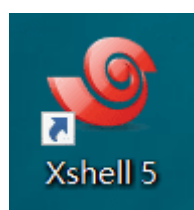
Xception 结构是一个由深度可分离的卷积层和残留连接组成的线性堆栈。这使得架构非常容易定义和修改, 使用像 Keras (<https://github.com/fchollet/keras>) 或 TensorFlow-Slim 这样的高级库, 它只需要 30 到 40 行代码, 这与 VGG-16 这样的架构没有什么不同, 而像 Inception V2 或 V3 这样的架构, 它们的定义要复杂得多。

4.获取训练集

训练数据集通常包含数以百万计的例子, 这取决于任务、检测项目的数量和期望的性能(Pondenkandath *et al.*, 2018)。总的来说, 训练数据集越大, 分类准确率就越高。然而, 重要的是要确保每个分类类别都有足够的测试, 以避免识别偏差(Wearn *et al.*, 2019)。还要注意, 训练数据集通常分为两个子集, 一个用于有效地训练模型, 另一个用于评估其性能。因此, 每个子集必须包含足够的例子代表所有分类类别, 以便有效的训练。

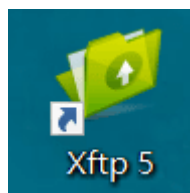
5.具体操作步骤:

5.1 准备软件



用于连接服务器, 输入命令等操作, 下载地址:

<https://xshell.en.softonic.com/>



用于上传查看文件，下载地址：
<https://www.pcsoft.com.cn/soft/13932.html>



用于编辑修改 python 文件，下载地址：
<https://www.jetbrains.com/pycharm/>

5.2 数据准备及处理

5.2.1 准备用于分类的图片

准备图片数据，按训练集：验证集：测试集分别是 70：15：15 的比例准备图片，将图片随机排列并重命名 (BJFU_001594)，使每一张图片有自己的编号，放进如下图所示的位置相应的 input 文件夹

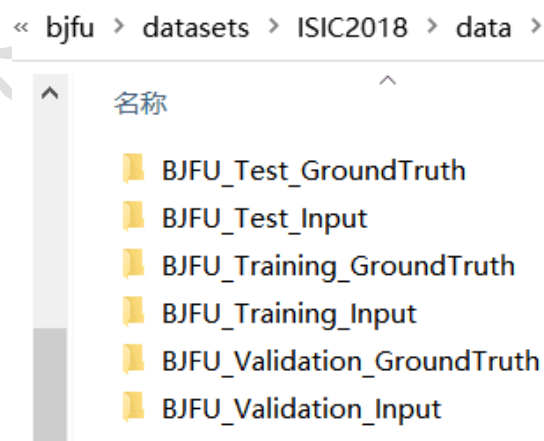


图 83 文件夹命名

并在 GroundTruth 中 (.csv 格式) 对每个编号的图片进行如下图

定义（以三分类为例 DEN、ALI、HYB），1 代表正确，0 代表否定，注意 input 的图片总数和 groundtruth 的总数要一一对应（某一编号可缺失，但不可对应不上）

	A	B	C	D
1	image	DEN	ALI	HYB
2	BJFU_001926	1	0	0
3	BJFU_001927	1	0	0
4	BJFU_001931	0	1	0
5	BJFU_001933	1	0	0
6	BJFU_001934	1	0	0
7	BJFU_001935	0	1	0
8	BJFU_001936	1	0	0
9	BJFU_001937	1	0	0
10	BJFU_001938	1	0	0
11	BJFU_001940	1	0	0
12	BJFU_001941	0	1	0
13	BJFU_001944	1	0	0
14	BJFU_001945	0	1	0
15	BJFU_001946	0	1	0
16	BJFU_001947	0	1	0
17	BJFU_001949	0	1	0
18	BJFU_001950	1	0	0
19	BJFU_001953	1	0	0
20	BJFU_001954	1	0	0
21	BJFU_001955	1	0	0
22	BJFU_001957	1	0	0

图 84 GroundTruth 示例

5.2.2 修改运行命令

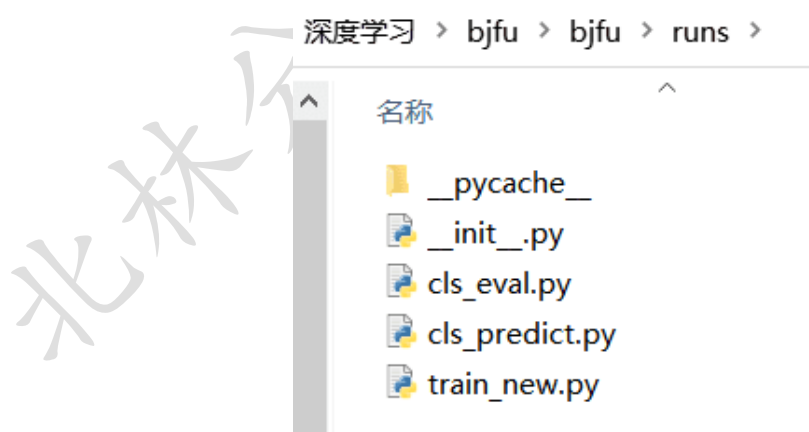


图 85 运行命令位置图

修改在上图位置的 python 文件（__init__、train_new），只需要改相应的路径和命名

在__init__.py 中如下更改：

```
# 图片文件夹
task3_img = 'BJFU_Training_Input'
task3_validation_img = 'BJFU_Validation_Input'
task3_test_img = 'BJFU_Test_Input'
```

图 86 更改位置示例

图片文件夹如果没有重命名无需改动

分类的名称需要以物种为基础进行更改，如下图所示（以三分类为例 DEN、ALI、HYB）

```
# 标签文件夹名字
task3_gt = 'BJFU_Training_GroundTruth'
task3_val_gt = 'BJFU_Validation_GroundTruth'
task3_test_gt = 'BJFU_Test_GroundTruth'

DEN = 0 # Melanoma
ALI = 1 # Melanocytic nevus
HYB = 2 # Basal cell carcinoma

classes = [DEN, ALI, HYB]
class_names = ['DEN', 'ALI', 'HYB']
```

图 87 更改位置示例

在 train_new 中如下更改：

```
3
4     import sys
5
6     sys.path.append("/home/wangtr/bjfu/bjfu")
7
```

图 88 更改位置示例

文件路径需按所存文件夹更改

5.3 数据上传及运行

5.3.1 数据上传

通过 Xftp 上传图片、groundtruth、_init_.py、train_new.py 到相应位置（一千张图片约 7 小时）

5.3.2 数据运行

在 Xshell 中进入到 runs 文件夹直接输入 train_new.py 进行图片的训练与验证（一千张图片约 10 小时），运行训练集输出结果（以三分类为例 DEN、ALI、HYB）如下：

```
50/50 [=====] - 22s 447ms/step - loss: 0.9094 - accuracy: 0.8706  
val_loss: 3.2083 - val_accuracy: 0.4355 - val_sensitivity_at_specificity: 0.9395
```

True\Pred	DEN	ALI	HYB	TOTAL
DEN	0.66	0.22	0.12	170
ALI	0.64	0.24	0.12	75
HYB	0.68	0.25	0.08	65
TOTAL	204	72	34	

	DEN	ALI	HYB	MEAN
precision	0.549	0.250	0.147	0.315
recall	0.659	0.240	0.077	0.325

图 89 运行训练集结果

训练后输入 cls_predict.py 文件进行预测，输出结果如下：

```
310/310 [=====] - 2s 6ms/step  
[5.7973313331604, 0.44516128301620483]
```

	DEN	ALI	HYB	MEAN
precision	0.545	0.224	0.238	0.335
recall	0.682	0.227	0.077	0.329

图 90 运行测试集结果

若重复操作将下面位置 cache 文件夹内容清空

```
« bjfu > bjfu > datasets > ISIC2018 > cache
```

图 91 清空数据位置

十、测量叶片传统性状指标

1. 基础指标

这里以叶片长为例，此方法适用于测量叶长(lamina length) (叶顶端到叶基部)，叶宽(lobe width) (叶片最宽处)，窦宽(sinuous width) (最宽叶片下边的凹陷点)，叶柄长(petiole length)，叶基部到叶最宽处的长度(wp: length of the lamina base to the widest part)等 (图 92)。

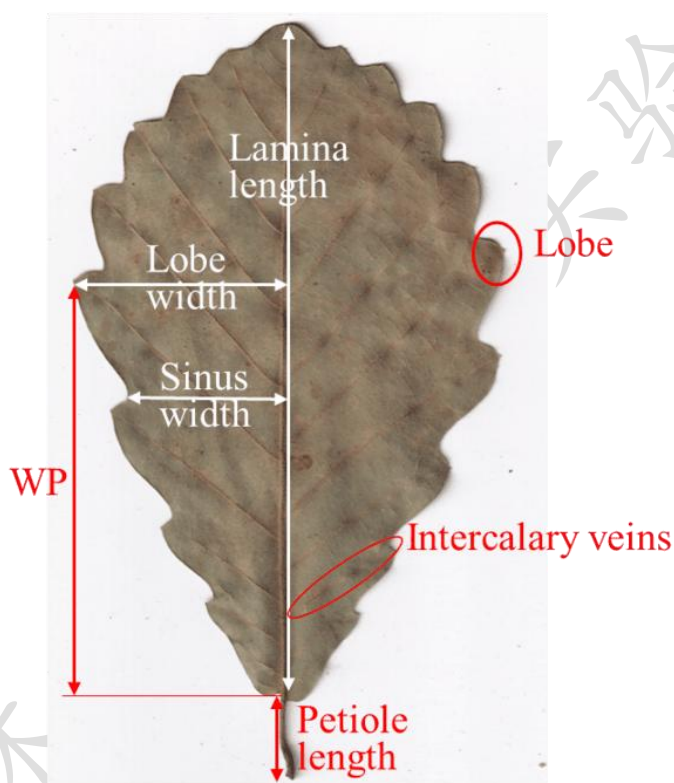


图 92 叶性状指标

具体步骤：

1.1 批量导入图像

选择“File” → “Import” → “Image Sequence”选定并打开叶片图像 (图 93, 94)。

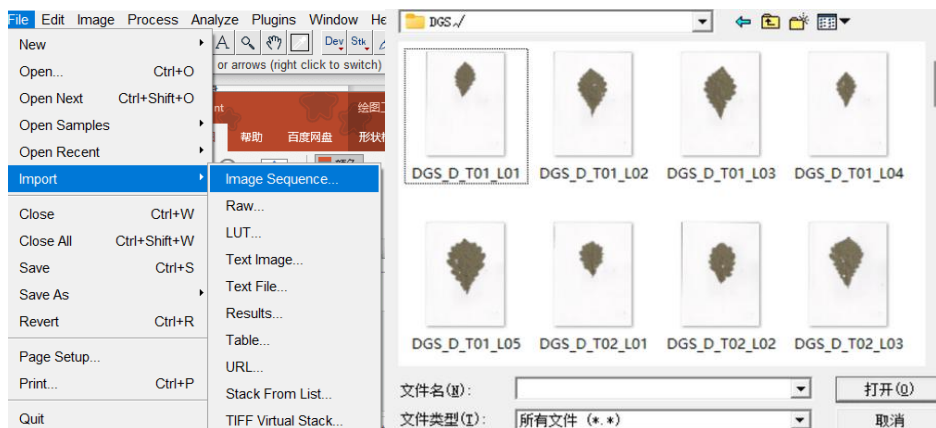


图 93 选择准备导入的叶片图像

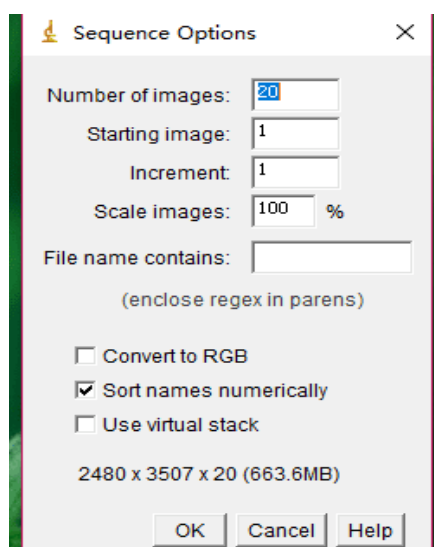


图 94 编辑导入图像的数量

1.2 测量叶长

将叶片放大到合适的位置，点击直线工具，选取叶顶端到叶基部的距离来测量叶片的长度像素点（图 95），选取完成后按 ctrl+M 完成长度的测量，单位为 dot。结果会自动记录在“Results”窗口里（图 96），测完一批后点击“Results”窗口里的“File”→“Save as”选择保存的位置并命名。



图 95 选取直线工具测量叶长

Results										
	Area	Perim.	Angle	Feret	FeretX	FeretY	FeretAngle	MinFeret	Length	
1	1599	1598	-90	1598	1644	1216	90	0	1598	

图 96 获得结果

1.3 像素转换

转换之前查看每张图片的分辨率，以图 97 为例，400 dpi (dot per inch)表示每英寸长度内有 400 个像素点数(dot)，换算一下，一个像素点(dot) = $(1/400)$ inch，其中 1 inch = 2.54 cm，则 1 dot = $(2.54/400)$ cm。
 实际叶长 = 测得的长度像素点数 $\times$ (2.54/分辨率)



图 97 图片属性

1.4 其余叶性状指标

其余指标如裂叶数量(lobe number), 夹层脉数量(number of intercalary veins) (an intercalary vein was a secondary vein irrigating a sinus and extending at least half way from the midrib to the base of the sinus), 叶基部形状(basal shape of the lamina)直接计数即可。其中叶基部形状记录为 1-9 (图 98):

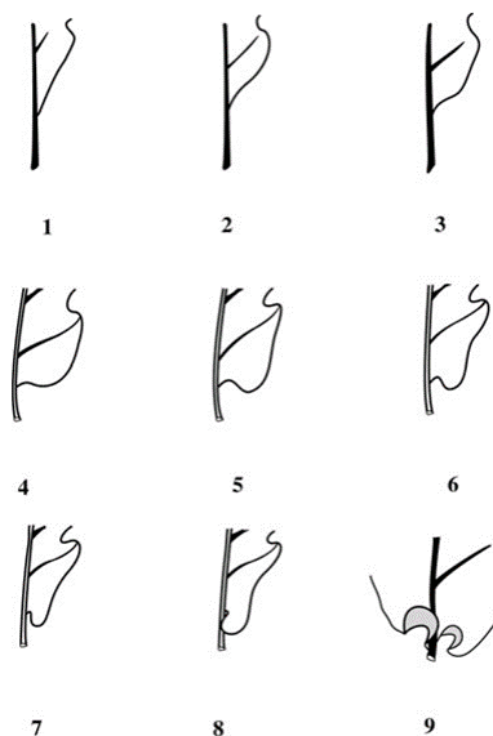


图 98 叶基部形状 (Kremer *et al.*, 2002)

2. 叶脉密度(Leaf Vein Density)

叶脉密度(mm mm^{-2})是指单位面积的叶脉长度。

每片叶子从叶片的底部到顶部都有一条主脉 (1° vein), 还有许多平行的二级脉 (2° vein) 分支。在目前的研究中, 一级和二级叶脉, 也称为主要叶脉(major vein), 从二级叶脉分支的叶脉被定义为三级叶脉 (3° vein)。在栎属的叶片中, 三级及以上的叶脉无法清楚地区分, 因此将三级及以上级别叶脉的叶脉密度统称为小脉密度(Minor vein density)。

测量叶脉密度, 包括一级叶脉密度(1° vein density)、二级叶脉密度(2° vein density)和小脉密度(Minor vein density)。

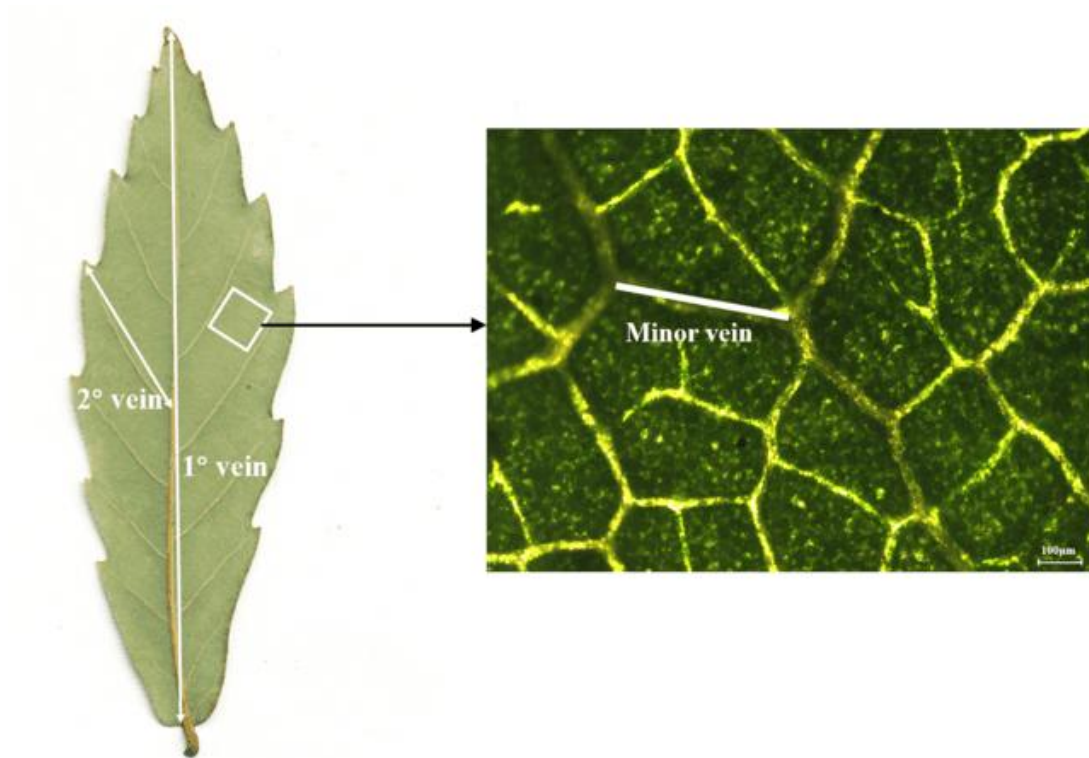


图 1 叶脉密度

2.1 一级叶脉密度(1° vein density)

用 ImageJ 软件的直线工具测量一级叶脉的长度，除以叶面积，即为该叶片的一级叶脉密度。



图 2 直线工具测量一级叶脉长度

2.2 二级叶脉密度(2° vein density)

用 ImageJ 软件的直线工具测量所有二级叶脉的长度总和，除以叶面积，即为该叶片的二级叶脉密度。

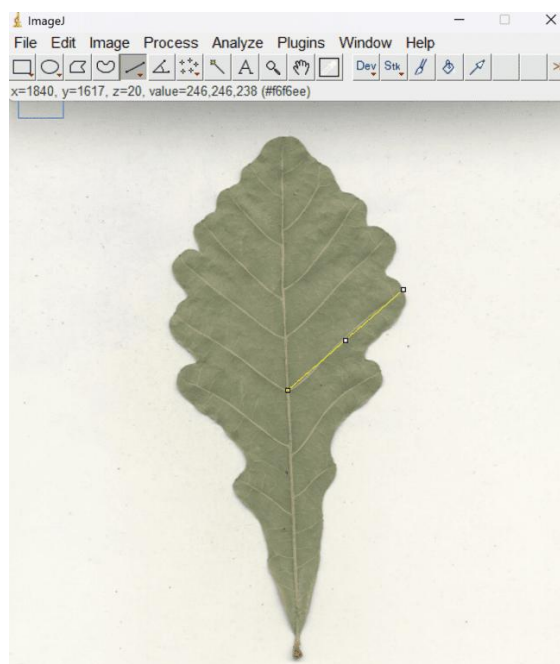


图 3 直线工具测量二级叶脉长度

2.3 小脉密度(Minor vein density)

为了得到小脉密度，我们需要使用带有图像软件（徕卡应用套件）的计算机连接的显微镜拍摄小脉图像，再结合 ImageJ 软件进行测量。

具体操作步骤：

2.3.1 叶片样本准备

选择生长良好、具有代表性的叶片，尽量保证叶片完整且无病虫害损伤。避开主脉剪下一小片叶片放到载玻片上（叶背朝上），在叶片上滴加适量的清水，盖上盖玻片。

2.3.2 显微镜观察与图像采集

(1) 打开仪器，顺序为①荧光→②显微镜→③电脑，关机顺序相反



图 4 仪器开关机顺序

(2) 将制备好的叶片切片放在显微镜载物台上，在电脑上打开 LAS



X 软件



(3) 点击“Live”开始拍照

(4) 选择明场“TL-BF”、“color mode”，通过调节显微镜的焦距和载物台的位置，在 10 倍物镜下调节 Exposure[ms]、Gain、Intensity、Aperture 等参数至合适亮度，使叶片小脉清晰地呈现在视野中



图 5 参数设置

(5)单击右键，选择“Export Image”，可保存为 JPEG、TIFF 等格式，点击“save”保存

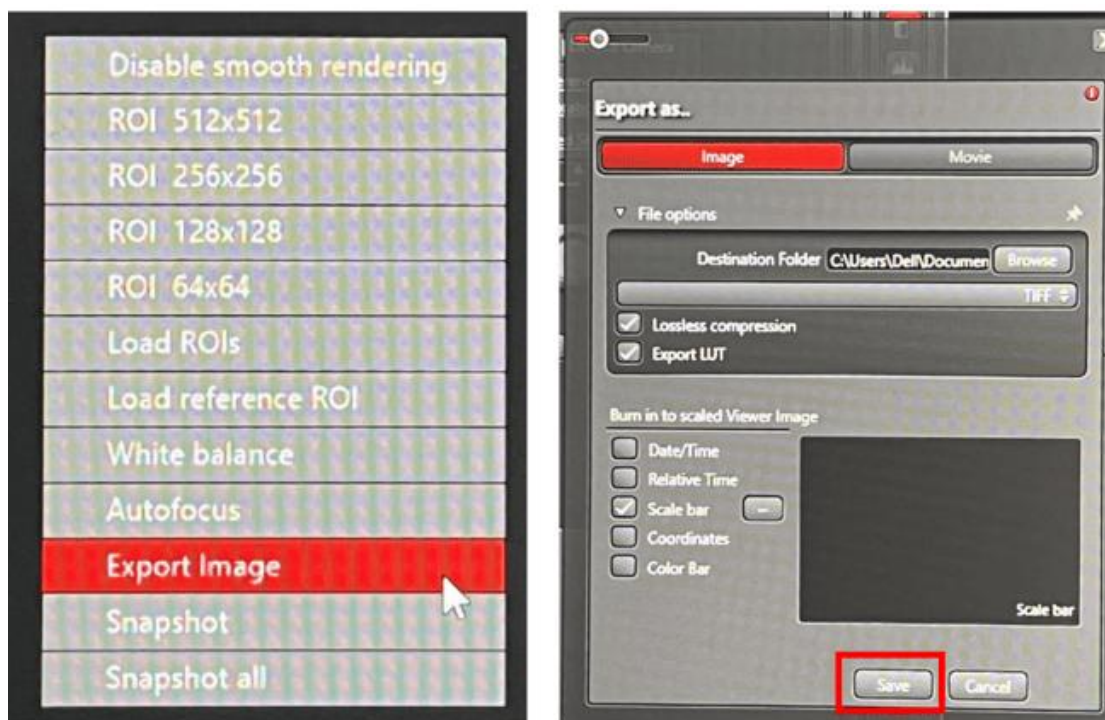


图 6 保存图片

2.3.3 ImageJ 软件测量

(1) 批量导入图像

选择“File”→“Import”→“Image Sequence”选定并打开叶片图像。

(2) 设置添加比例尺

① 用直线工具沿比例尺画线

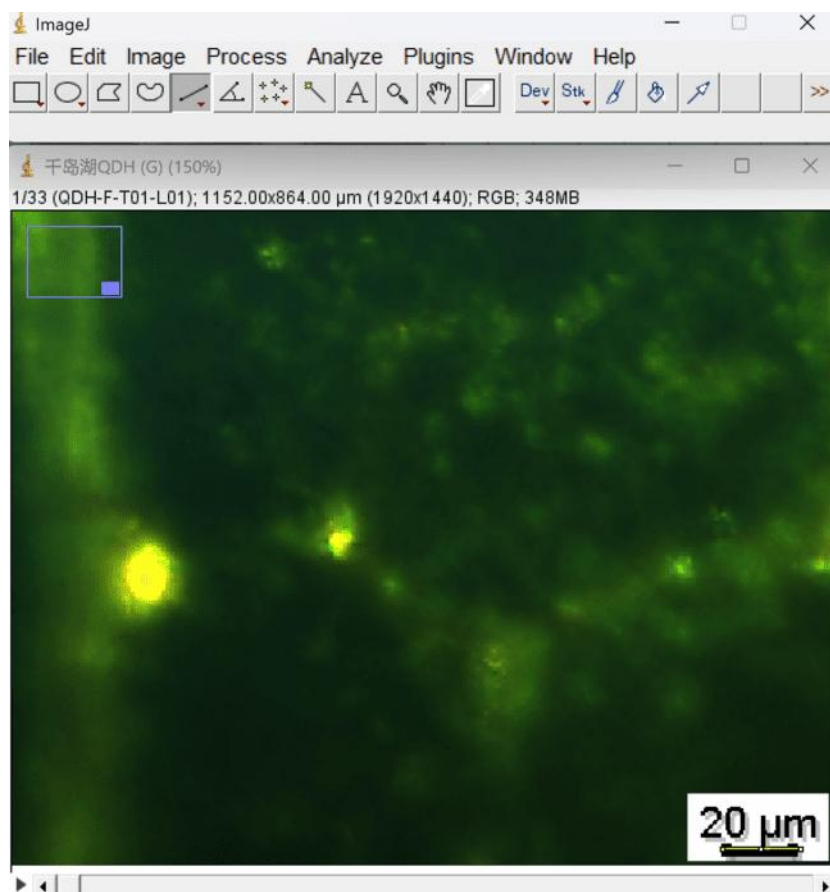


图 7 比例尺画线

②设置比例尺，选择“Analyze”→“Set Scale”，在“Known distance”和“Unit of length”设置该直线的已知长度和单位，勾选“Global”对所有照片应用这一比例。

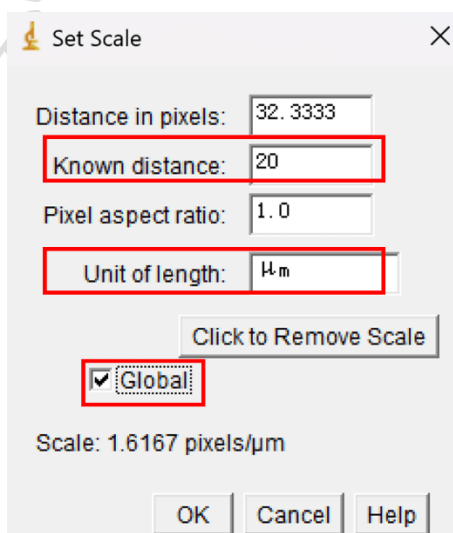


图 8 设置比例尺长度

(3)测量小脉长度

用直线工具测量所有小脉长度

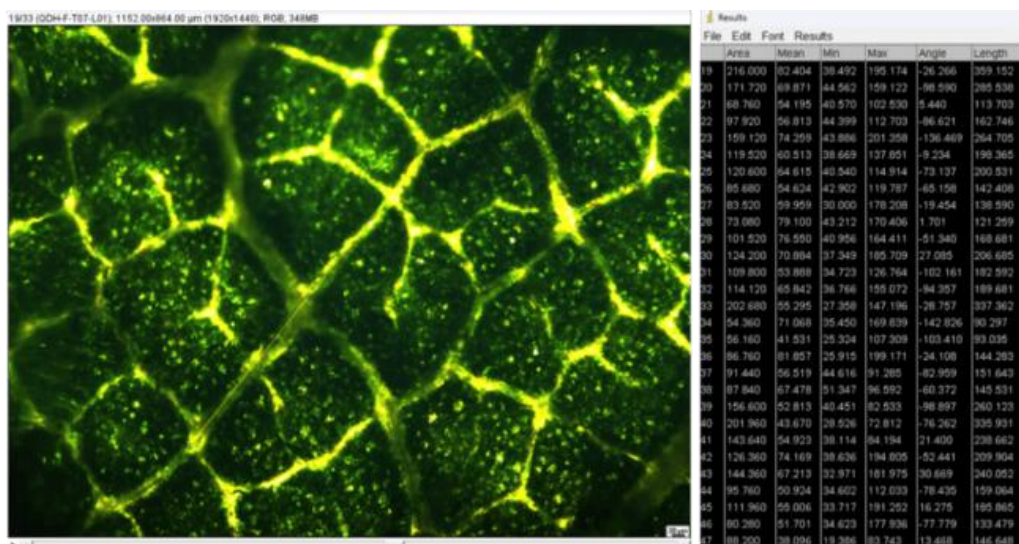


图 9 小脉长度测量

(4)计算小脉密度

将所有小脉长度求和，除以视野面积，即为小脉密度。（需将单位换算为微米， $1\mu\text{m} = 0.001\text{mm}$ ）

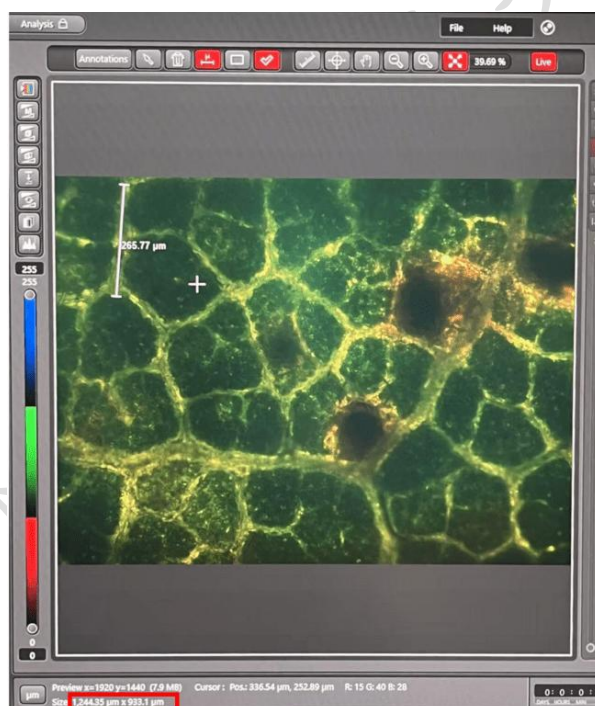


图 10 小脉视野面积

3. 比叶面积(Specific Leaf Area)

比叶面积是指单位叶面积与干重的比值(cm^2/g)，是叶功能性状指标，与叶片光合作用和产量相关。

为了获得叶面积，我们使用扫描仪以 300 dpi 的分辨率批量扫描

带叶柄的完全伸展的叶片，并使用 ImageJ 软件 (<https://imagej.nih.gov/ij/download.html>)估算叶面积。

具体操作步骤：

3.1 导入图像(Import image)

批量导入图像：选择“File”→“Import”→“Image Sequence”选定并打开叶片图像。

3.2 参数设置(Parameter Setting)

(1) 依次选择“Image”→“Type”→“8bit”(图 99)，分解 RGB 图像为灰度图，所得结果最为精确。

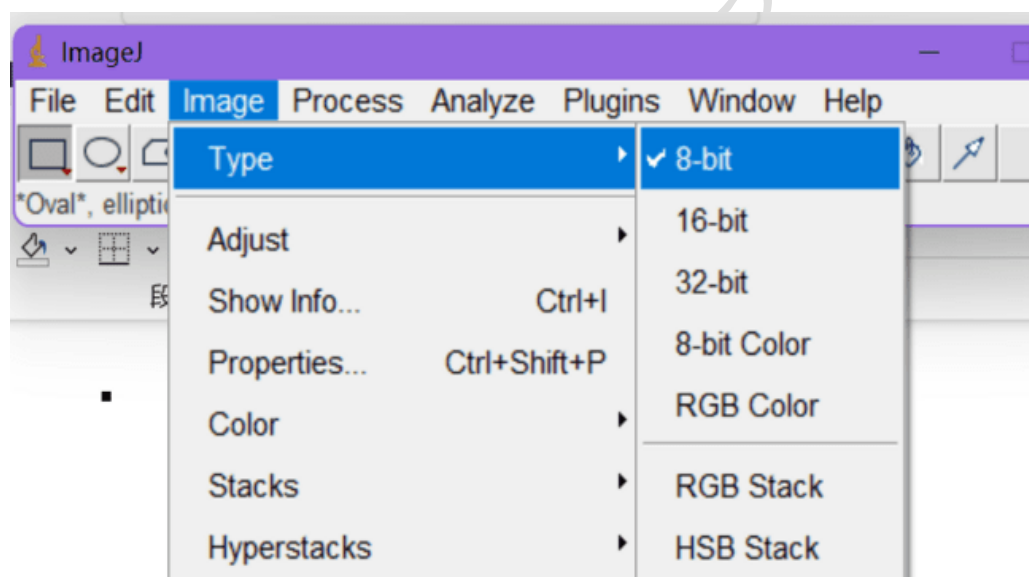


图 99 分解 RGB 图像为灰度图

(2) 选择“Image”→“Adjust”→“Threshold”顺序调整待测对象为 Red (图 100)，点击 Apply。

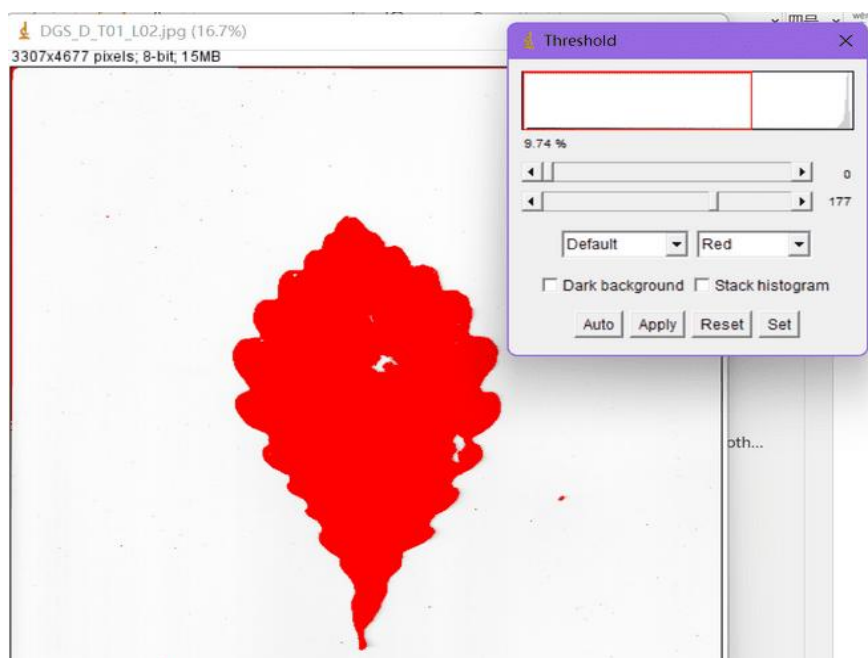


图 100 选取测定面积

(3) 再选择 “Analyze” → “Set Scale” 设置比例参数 (图 101)。

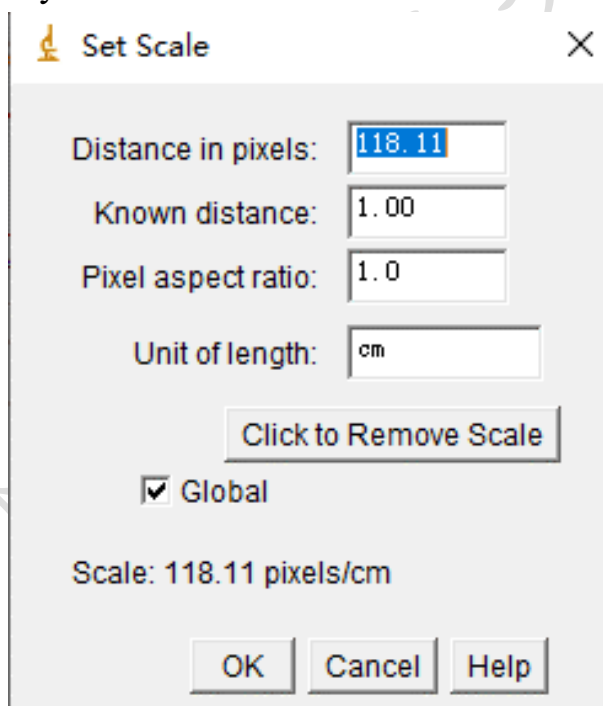


图 101 设置比例参数

Distance in pixels: 分辨率单位 dpi 指的是每英寸所打印的点数, 因为 1 英寸(in)=2.54 厘米(cm), 同时我们的分辨率 300dpi=300 pix per in, 则像素距离(Distance in pixels) = $300 \text{ pix} / 2.54 \text{ cm} = 118.11 \text{ pix/cm}$ 。这里有一点需要注意: 测叶面积之前要看一下图片的分辨率具体是多少 (点击图片右键→属性→详细信息)。

Known distance: 这个值不用设置, 自动生成 1

Unit of length: 设置单位为 cm

Global: 对于所有打开的照片应用该比例关系。如果有同批次同镜头视野的多张照片，可以勾选此选项，一次性将所有照片中像素所对应的实际距离进行设置。

有了这样的关系之后，我们利用 Image J 对位图以像素为单位进行测量或者计算的时候，就与实际长度联系起来了。

(4) 选择“Analyze”→“Analyze particles”，导出叶面积值（图 102）。

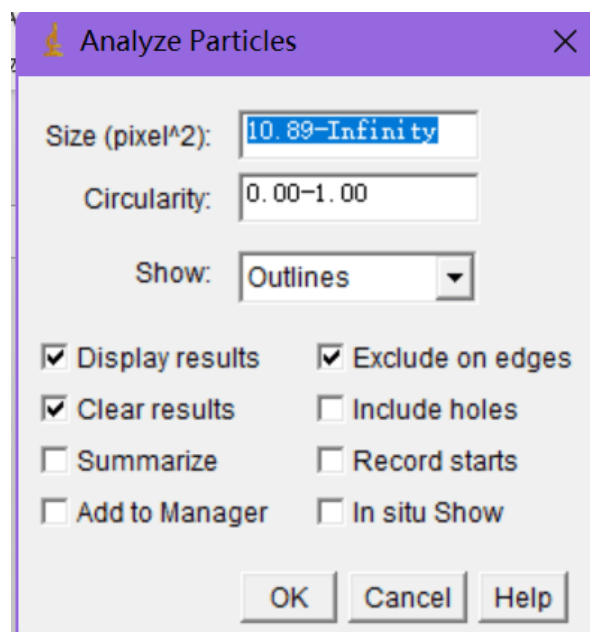


图 102 过滤非目标的小杂质

Size 可以设置 10 或者更大起始，以过滤非目标的小杂质；我们的叶子不是圆形，Circularity 的最大值可以设置的小一些；Exclude on edges 可以去除边界的非目标，但扫描时需要注意叶片不能接触边界；show 可以选择 Outlines 看一下计算是否正确。由于我们的扫描叶片数(count)是非常明确的，所以可以依次来检查匹配是否正确。Result 界面可以连续出现数值，最后将所有结果统一“Save as”导出成 Excel 表格。

3.3 叶片干重(Dry Leaf Weight)

新鲜叶片在 80℃通风烘箱中放置 3 d 后，用天平称重。

3.4 比叶面积(Specific Leaf Area)

每个个体的比叶面积：4-5 片叶片的总面积/4-5 片叶片的总干重。

3.5 批量计算图片中叶片大小（叶面积、叶最大直径）

(1) 下载 imgJ 加载程序包 RosetteTracker （图 103）；

<https://telin.ugent.be/~jdvylder/RosetteTracker/>



图 103 RosetteTracker 下载

- (2) 打开 imgJ, 点击 Plugins 加载 RosetteTracker, 点击 import images, 选择 good contrast（目标叶片与背景颜色差异较大的图片），导入图片；
- (3) 点击 settings, 选择 diameter/are（最大直径/叶面积），a single image 中 segmentation 填写植物数量和列数；
- (4) 图片标尺的换算：先在有标尺的图片上用直线工具描绘标尺，选择 Anysis 中的 set scale 换算 1plex=?mm，换算后的数据填入 set scale 中的 set manual；
- (5) 点击 Day&night 页面，取消 holes，选取需要识别的叶片颜色，clutter 选择 3 默认，点击 OK, 开始 run！

十一、气候因子与叶性状的线性关系

1. 获取叶性状数据

从 Morpho J 中导出叶片大小和叶片形状 PC 的数据（参见三、3.5），作为叶形输入数据。

2. 获取气候数据

采样点气候数据的获取参考第五节中从 WorldClim 数据库中获取气候因子的方法，第五节中使用方差膨胀因子分析(VIF)去除部分相关性较强的因子，使用共线性低的具体气候因子做相关分析。

本节使用 R 语言的函数包 “vegan” (Oksanen *et al.*, 2017) 将 19

个气候因子进行 PCA 降维分析，使用 PC1 轴的数据来代表每个采样点的气候（以下写为 cli PC1）。

PCA 分析代码：

```
#####加载函数包
>library(vegan)
#####导入数据
>envo=read.table("clipboard",header=T)
#####建立模型
>pca_princomp <- princomp(envo[,-19],cor=T)
#####19 个主成分的累积方差贡献率
>summary(pca_princomp)
#####前两个主成分的载荷值
>predict(pca_princomp,loadings=TRUE)
>pca_princomp$loadings[,1:2]
#####每个采样点 19 个主成分的得分
>x=predict(pca_princomp)
>write.table(x,"output.txt")
```

3. 叶性状与气候因子的相关分析

为了调查不同物种之间叶性状与气候因子的相关关系，我们使用 R 语言的函数包“ggplot2”，“ggpubr”，“lme4”，“car”和“lmerTest” (Fox *et al.*, 2013; Wickham, 2016)，进行线性分析和显著性分析。

运行代码如下：

```
#####加载函数包
>library(ggplot2)
#####设置 xy 轴的参数
>windowsFonts(RMN=windowsFont("Times New Roman"))
>theme_updatek=theme (
  >plot.title=element_text(family="RMN", size=18, face="bold",
    colour="black", hjust = 0.5),
  >axis.title.x=element_text(family="RMN", size=14, face="bold",
    colour="black"),
  >axis.title.y=element_text(family="RMN", size=14, face="bold",
    colour="black"),
  >axis.text.x=element_text(family="RMN", size=10, face="bold",
```

```

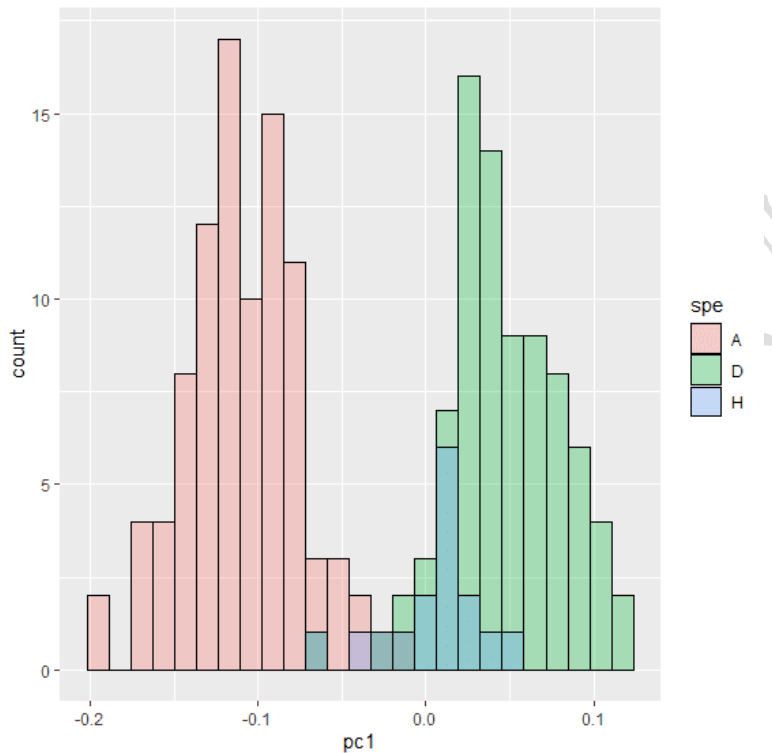
colour="black"),
  >axis.text.y=element_text(family="RMN", size=10, face="bold",
colour="black"),
  >axis.line = element_line(size=1,colour = "black"))
theme_legend=theme (
  >legend.text = element_text(colour = 'black', size = 10,face = 'bold'),
  >legend.title =element_text (colour = 'purple', size = 10,face =
'bold',angle = 0, hjust = 1, vjust =1),
  >legend.background = element_blank(),legend.key = element_blank()
)
#####导入数据(个体水平, 一系列 cli pc1 数据,一系列叶型 PC1 数据, 一系列
pop 数据 (用 1, 2, 3...来表示), 一系列物种(species)数据)
  >data=read.table("clipboard",header=T)
#####建立模型
  >p = ggplot (data, aes (x= cli pc1, y= leaf shape PC1, colour = species))
+ geom_point () + scale_x_continuous (limits = c (-10,10)) + stat_smooth
(method = lm,se = FALSE) + theme_bw () + theme_updatek +
theme_legend
#####得出散点图
  >p
#####每个物种气候因子与叶型的回归系数
#####加载函数包
  >library(ggpubr)
#####回归系数 R 值和 P 值
  >p+stat_cor ()
  >ggsave ("tmp4.pdf", width = 8, height = 8,units = "cm")
#####显著性:
  >library(lme4)
  >library(car)
  >library(lmerTest)
#####建立模型
  >m1<-lmer (leaf shape PC1~ cli PC1*species + (1|pop), data)
#####显著性分析
  >anova (m1, type="I")
####注: 以上是基于叶片形状的代码, 基于叶片大小或其他性状的代
码, 只需循环执行建立和运行模型。

```

十二、R 语言常用的作图方法

1. 直方图

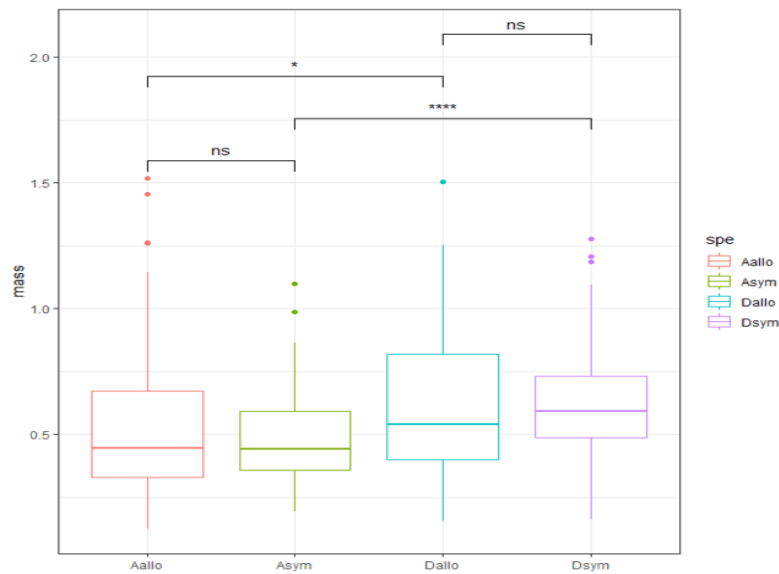
e.g., 三个组 A, D, H 的叶型 PC1 数据 (Du *et al.*, 2022)



```
#####加载包并导入数据
>library(ggplot2)
>data=read.table("clipboard",header=T)
#####作图 直方图重叠
>ggplot(data, aes(x = pc1, fill = spe)) + geom_histogram(position =
"identity", binwidth = 0.5, alpha = 0.4, colour = "black")
#####作图 直方图不重叠
>ggplot(data, aes(x = pc1, fill = spe)) + geom_histogram(position =
'dodge', binwidth = 0.1, alpha = 0.4, colour = "black")
```

2. 箱型图

e.g., 四个组 Aallo, Asym, Dallo, Dsym 的叶干重 (Du *et al.*, 2022)



####加载包并导入数据

```
>library(ggplot2)
```

```
>library(ggpubr)
```

```
>data=read.table("clipboard",header=T)
```

####作图

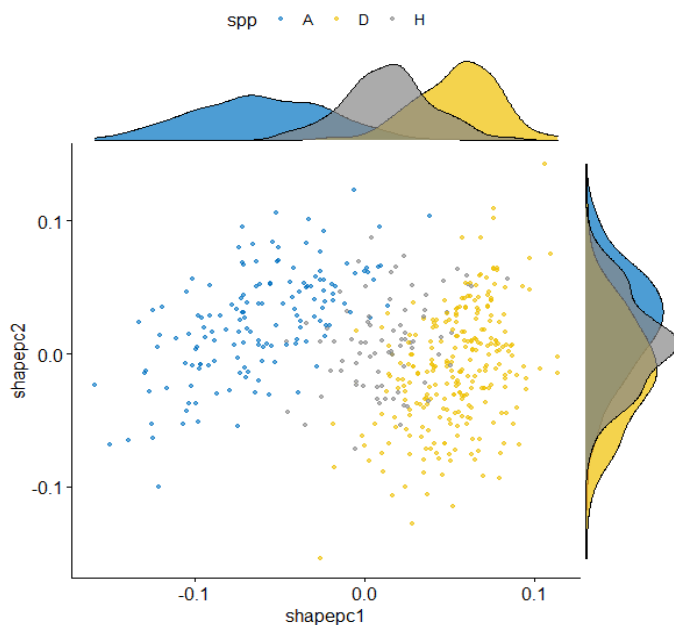
```
>ggplot (data, aes (x = spe, y = mass, color = spe)) + geom_boxplot () +  
theme _ bw ()
```

####添加两两之间的显著性检验

```
>stat _ compare _ means (aes (label = .. p. signif ..), comparisons = list (c  
( 'Asym', 'Aallo'), c('Asym', 'Dsym'), c('Aallo', 'Dallo'), c('Dsym',  
'Dallo')) )
```

3. 散点图加折线图

e.g., 三个组 A, D, H 的叶型 pc1 和 pc2 轴的数据 (Du *et al.*, 2022)



```
#####加载包并导入数据
>library(cowplot)
>data1=read.table("clipboard",header=T)
#####作图
>ggscatterhist (
  data1, x = "shapepc1", y = "shapepc2",
  color = "species", size = 1, alpha = 0.6,
  palette = "jco",
  margin. params = list(fill = " species ", color = "black", size =
0.2)
)
```

参考文献

- 陈德钊. 多元数据处理. 化学工业出版社, 1998.
- 胡坤. 1992. 典型变量分析在化探样品分类中的应用. *物探化探计算技术*, 14(3):196-199.
- *刘媛, 吕佳, 宋佳, 王玉焱, 王肖静, & 杜芳. 2017. 基于几何形态测量学的植物鉴定方法. *植物科学学报*, (6), 894-899.
- *宋佳, 侯盟, 鲁四海, 李俊清, & 杜芳. 2015. 一种基于标志点的叶片几何形态分析方法. *兰州大学学报(自科版)*, 51(5), 705-710.

-
- *苏蔚, 宋以刚, 祁敏, 杜芳. 2021. 基于几何形态分析的栎属白栎组叶片形态特征. *应用生态学报*, 32, 2309.
- 杨维权等. 1989. 多元统计分析. 高等教育出版社.
- Abràmoff MD, Magalhães PJ, Paulo J, Ram SJ. 2005. Image processing with IMAGEJ Part II. *Biophotonics International*, 11: 36-42.
- Christin S, Hervet É, Lecomte N. 2019. Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10): 1632-1644.
- *Du F, Qi M, Zhang YY, Petit RJ. 2022. Asymmetric character displacement in mixed oak stands. *New Phytologist*, 236, 1212-1224.
- Escoufier, Y. 1973. Le traitement des variables vectorielles. *Biometrics*, 29: 751-760.
- Hinton G E, Osindero S, The Y W. 2006. A Fast Learning Algorithm for Deep Belief Nets. *Neural Computation*, 18(7): 1572-1554.
- Klingenberg CP. 2003. A developmental perspective on developmental instability: theory, models and mechanisms. In: Polak M, ed. *Developmental instability: causes and consequences*. New York: Oxford University Press, 14-34.
- Klingenberg CP. 2011. MorphoJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, 11:353-357.
- Klingenberg CP and McIntyre GS. 1998. Geometric morphometrics of developmental instability: analyzing patterns of fluctuating asymmetry with Procrustes methods. *Evolution*, 52:1363-1375.
- Klingenberg CP, Barluenga M, Meyer A. 2002. Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution*, 56: 1909-1920.
- Klingenberg CP. 2008. Morphological integration and developmental modularity. *Annual Review of Ecology, Evolution and Systematics*, 39: 115-132.
- Klingenberg CP. 2009. Morphometric integration and modularity in configurations of landmarks: Tools for evaluating a-priori hypotheses. *Evolution & Development*, 11: 405-421.
- Krizhevsky, Alex, Sutskever, et al. 2017. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*.
- Kremer A, Dupouey J L, Deans J D, Cottrell J, Csaikl U, Finkeldey R, et al. 2002. Leaf morphological differentiation between *Quercus robur* and *Quercus petraea* is stable across western European mixed oak

-
- stands. *Annals of Forest Science*, 59(7): 777-787.
- Lecun Y, Boser B, Denker J, Henderson D, Howard R, Hubbard W, Jackel L. 2014. Backpropagation applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4): 541-551.
- Lele S, Richtsmeier JT. 1991. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86(3): 415-427.
- *Liu Y, Li Y, Song J, Zhang R, Yan Y, Wang Y, Du F. 2018. Geometric morphometric analyses of leaf shapes in two sympatric Chinese oaks: *Quercus dentata* Thunberg and *Quercus aliena* Blume (Fagaceae). *Annals of Forest Science*, 75: 90.
- *Li YJ, Zhang Y, Liao PC, Wang TR, Wang XY, Ueno S, Du F. 2021. Genetic, geographic, and climatic factors jointly shape leaf morphology of an alpine oak, *Quercus aquifolioides* Rehder & EH Wilson. *Annals of Forest Science*, 78, 1-18.
- Maddison WP. Squared-change parsimony reconstructions of ancestral states for continuous-valued characters on a phylogenetic tree. 1991. *Systematic Biology*, 40(3): 304-14.
- Maddison DR, Swofford DL, and Maddison WP. 1997. NEXUS: an extensible file format for systematic information. *Systematic Biology*, 46: 590–621.
- Marugán-Lobón J. 2010. Combining Shape Data and Traditional Measurements with the 2B-PLS: Testing the Covariation Between Avian Brain Size and Cranial Shape Variation as an Example. *Morphometrics for Nonmorphometricians*.
- Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara RB, Simpson G, Solymos P, Stevens M, Wagner H. 2017. vegan: Community ecology package. R package version.
- Olden, Jd, Lawler, et al. 2008. Machine learning methods without tears: A primer for ecologists. *Quarterly Review of Biology*, 83(2): 171-193.
- Pondenkandath V, Seuret M, Ingold R, et al. 2018. Exploiting State-of-the-Art Deep Learning Methods for Document Image Analysis.
- *Qi M, Du F, Guo F, Yin KQ, Tang JJ. 2024. Species identification through deep learning and geometrical morphology in oaks (*Quercus* spp.): Pros and cons. *Ecology and Evolution*, 14(2): e11032.

-
- Rohlf FJ, Corti M. 2000. The use of two-block partial least-squares to study covariation in shape. *Syst. Biol.*, 49: 740-753.
- Vieira M, Mayo SJ and Andrade IMD. 2014. Geometric morphometrics of leaves of *Anacardium microcarpum* Ducke and *A. occidentale* L. (Anacardiaceae) from the coastal region of Piauí, Brazil. *Brazilian Journal of Botany*, 37(3): 315-327.
- Viscosi V. 2015. Geometric morphometrics and leaf phenotypic plasticity: assessing fluctuating asymmetry and allometry in European white oaks (*Quercus*). *Bot J Linn Soc*, 179: 335–348.
- Vujić V, Rubinjoni L, Selaković S and Cvetković D. 2016. Small-scale variations in leaf shape under anthropogenic disturbance in dioecious forest forb *Mercurialis perennis*: a geometric morphometric examination. *Arch Biol Sci*, 68(4): 705-713.
- Wearn OR, Freeman R, Jacoby DMP. 2019. Responsible AI for conservation. *Nature Machine Intelligence*, 1(2): 72-73.

*表示本实验室发表论文。