

微生物群落构建过程的空间可视化方法

Spatial Visualization of Microbial Community Assembly Processes

闫慧贞¹, 王凯^{1,*}, 张德民¹

¹海洋学院, 宁波大学, 宁波市, 浙江省

*通讯作者邮箱: wangkai@nbu.edu.cn

引用格式: 闫慧贞, 王凯, 张德民. (2021). 微生物群落构建过程的空间可视化方法. *Bio-101* e2003392.

Doi: 10.21769/BioProtoc.2003392.

How to cite: Yan, H. Z., Wang, K. and Zhang, D. M. (2021). Spatial Visualization of Microbial Community Assembly Processes. *Bio-101* e2003392. Doi: 10.21769/BioProtoc.2003392. (in Chinese)

摘要: 基于零模型的微生物群落构建过程判别方法 (Stegen *et al.*, 2013) 已被广泛运用于多种生境中微生物群落构建机制的研究。在微生物地理学研究中, 该方法的直接结果能反映各种生态过程在研究区域中的相对比例。但是, 如不进一步对群落构建过程的地理空间分布格局进行表征, 则会使得对微生物群落构建机制的认识趋于笼统化。本教程详述了应用 ArcGIS 等软件将对样点间微生物群落构建过程映射到地图的分析流程, 不仅能展示各种生态过程的相对重要性, 同时能清晰表征其在整个研究区域的地理空间分布格局, 从而加深研究者对微生物群落构建过程空间异质性的认识。

关键词: 微生物群落, 群落构建, 空间分布, 可视化

仪器设备

个人电脑 (安装 Windows 10 操作系统)

软件

1. ArcGIS (v10.4): <https://desktop.arcgis.com/>
2. Microsoft Office Excel (v2016)

实验步骤

Stegen 零模型方法的最终计算结果为 β NTI (β -Nearest Taxon Index, 最近种间指数) 和 RC_{Bray} (基于 Bray-Curtis 非相似性的 Raup-Crick 指数) 数值矩阵 (Stegen *et al.*, 2013), 能反映主导成对样本间微生物群落周转的各种生态过程, 即群落构建过程。将成对采样站位之间的微生物群落构建过程映射在地图上, 即可反映出研究区域中各种生态过程的地理空间分布格局。本教程用成对站位间不同颜色的连线来表示站位间的各类生态过程, 操作步骤如下:

1. 示例数据来源

本教程描述的微生物群落构建过程空间可视化方法由本课题组首次提出并应用 (Wang *et al.*, 2019), 因此以该论文中的部分数据作为示例数据。示例数据可从 Github 获取: <https://github.com/huizhen-yan/Huizhen-BioProtocol>。

2. 将采样站位经纬度表在 Excel 中转换为成对站位表, 随后在 ArcGIS 中通过 XY To Line 工具 (Esri, Redlands, CA) 生成采样站位成对连线图层。

2.1 组织采样站位经纬度表 (使用 Excel)

将站位经纬度坐标转换为十进制。将站位编号、纬度和经度三列合并一列并用逗号分隔 (图 1 中 D 列), 复制 D 列至 E 列。对 D 和 E 列进行成对排列组合, 生成 F 列 (图 1)。D 和 E 列组合操作如下: 鼠标右击图 1 采样站位经纬度表所在工作簿→查看代码→弹出 Microsoft Visual Basic for Applications (VBA) 窗口, 运用如下宏代码 (图 2), 运行结束后即出现 F 列。宏代码可从 Github 获取: <https://github.com/huizhen-yan/Huizhen-BioProtocol>。

	A	B	C	D	E	F
1	SampleID	LatitudeY	LongitudeX	XY	XYCopy	
2	S3	30.83	122.65	=A2&" "&B2&" "&C2	S3,30.83,122.65	S3,30.83,122.65,S7,30.81,122.5
3	S7	30.81	122.5	S7,30.81,122.5	S7,30.81,122.5	S3,30.83,122.65,S8,30.77,122.32
4	S8	30.77	122.32	S8,30.77,122.32	S8,30.77,122.32	S3,30.83,122.65,S10,30.69,121.74
5	S10	30.69	121.74	S10,30.69,121.74	S10,30.69,121.74	S3,30.83,122.65,S11,30.68,121.89
6	S11	30.68	121.89	S11,30.68,121.89	S11,30.68,121.89	S3,30.83,122.65,S12,30.68,121.56
7	S12	30.68	121.56	S12,30.68,121.56	S12,30.68,121.56	S3,30.83,122.65,S14,30.67,122.57
8	S14	30.67	122.57	S14,30.67,122.57	S14,30.67,122.57	S3,30.83,122.65,S18,30.67,122.8
9	S18	30.67	122.8	S18,30.67,122.8	S18,30.67,122.8	S3,30.83,122.65,S20,30.62,121.9
10	S20	30.62	121.9	S20,30.62,121.9	S20,30.62,121.9	S3,30.83,122.65,S21,30.6,121.6
11	S21	30.6	121.6	S21,30.6,121.6	S21,30.6,121.6	S3,30.83,122.65,S22,30.59,122.42
12	S22	30.59	122.42	S22,30.59,122.42	S22,30.59,122.42	S3,30.83,122.65,S23,30.57,122.25

图 1. 采样站位经纬度表

```

(通用) 两列组合
Sub 两列组合() '代码改编自Excel880.com
    Dim Dn, En, IDstart, IDend, i, j, k '声明变量
    Dn = [D1048576].End(xlUp).Row '将D列最后一个非空单元格的行号赋给Dn
    En = [E1048576].End(xlUp).Row
    IDstart = Range("D2:D" & Dn) '把D列非空区域赋给数组IDstart
    IDend = Range("E2:E" & En)
    ReDim pair(1 To Dn * En, 0) '声明动态数组pair, 该数组包含Dn*En个元素
    For i = 1 To UBound(IDstart) 'For循环, 步长默认值为1, i从1到数组IDstart的上界
        For j = i + 1 To UBound(IDend) 'j从2到数组IDend的上界
            Key = IDstart(i, 1) & "," & IDend(j, 1) '数组IDstart与IDend合并, 以逗号分隔, 并将该字符串赋给Key
            k = k + 1
            pair(k, 0) = Key '将Key赋给动态数组pair, 参与For循环
        Next
    Next
    [F2:F1048576].Clear
    [F2].Resize(k) = pair '从F2单元格开始, 扩展区域为k行, 写入数组pair
End Sub

```

图 2. 两列组合宏代码 (改编自 Excel 880.com)

2.2 整理成对站位经纬度表, 准备导入 ArcGIS

对图 1 中的 F 列进行分列后共包含 6 列, 分别是起始站位编号 (StationStart)、起始站位纬度 (StartY)、起始站位经度 (StartX)、终点站位编号 (StationEnd)、终点站位纬度 (EndY) 和终点站位经度 (EndX)。然后添加 SID 列 (图 3 中 A 列) 作为唯一值, 用于后续表格链接。

	A	B	C	D	E	F	G
1	SID	StationStart	StartY	StartX	StationEnd	EndY	EndX
2	ID1	S3	30.83	122.65	S7	30.81	122.5
3	ID2	S3	30.83	122.65	S8	30.77	122.32
4	ID3	S3	30.83	122.65	S10	30.69	121.74
5	ID4	S3	30.83	122.65	S11	30.68	121.89
6	ID5	S3	30.83	122.65	S12	30.68	121.56
7	ID6	S3	30.83	122.65	S14	30.67	122.57
8	ID7	S3	30.83	122.65	S18	30.67	122.8
9	ID8	S3	30.83	122.65	S20	30.62	121.9
10	ID9	S3	30.83	122.65	S21	30.6	121.6
11	ID10	S3	30.83	122.65	S22	30.59	122.42
12	ID11	S3	30.83	122.65	S23	30.57	122.25

图 3. 采样站位成对排列 (分列后)

2.3 在 ArcGIS 中使用 XY To Line 工具将成对采样站位连线转换为地图图层

在 ArcMap 10.4 中调用 ArcToolbox→Data Management Tools→Features→XY To Line, XY To Line 工具参数设置如图 4 所示。导入步骤 2.2 中得到的成对站位经纬度表, StartX 和 StartY 对应起始站位的经度和纬度, EndX 和 EndY 对应终点站位的经度和纬度, Line Type 选择默认推荐线形 GEODESIC, ID 是指在 XY To Line 中输入的图 3 所示表格中的 SID 列, Spatial Reference 是指空

间坐标系统，应当与已导入 ArcMap 中的地图图层保持一致，随后生成采样站位成对连线图 (图 5)。

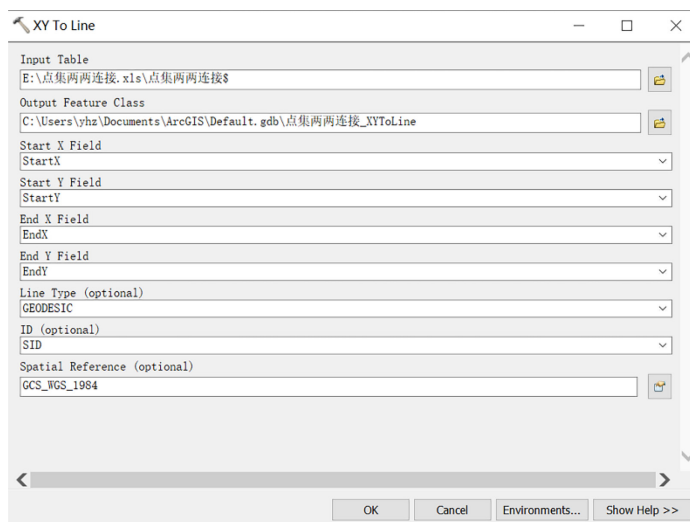


图 4. ArcGIS XY To Line 工具参数设置

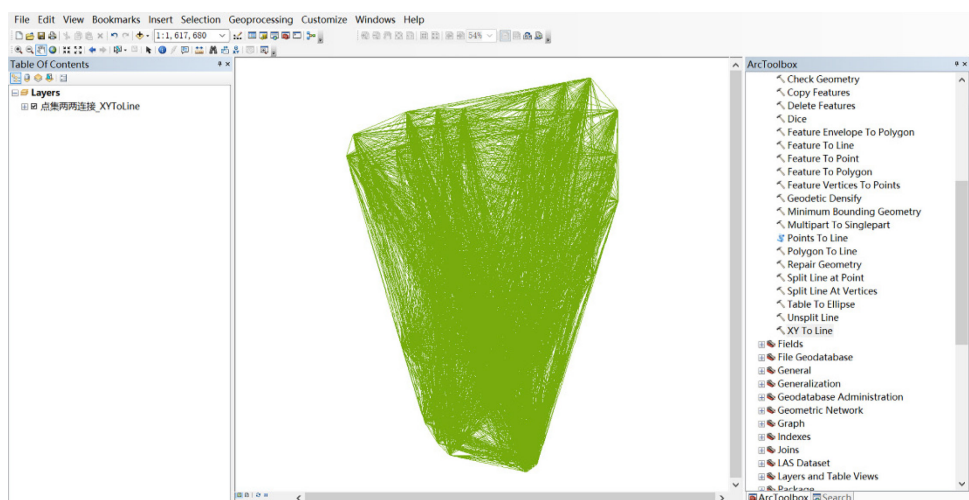


图 5. 成对站位连线图

3. 将 β NTI 和 RC_{Bray} 数值矩阵在 Excel 中转换为成对站位表，并整理群落构建过程属性表

利用 Excel 中 Power Query 编辑器 (点击数据→从表格; 不同 Office 版本调用方式略有不同), 选中首列, 随后点击转换→逆透视列→逆透视其他列, 将 β NTI 和 RC_{Bray} 数值矩阵转换为成对站位表 (图 6)。随后, 使用 Excel “筛选” 功能设置 β NTI 和

RC_{Bray} 值区间进行批量筛选，将 $\beta\text{NTI} > 2$ 的行填充为 HeterogeneousSelection，将 $\beta\text{NTI} < -2$ 的行填充为 HomogeneousSelection；当 $|\beta\text{NTI}| \leq 2$ 时，将 $\text{RC}_{\text{Bray}} > 0.95$ 的行填充为 DispersalLimitation，将 $\text{RC}_{\text{Bray}} < -0.95$ 的行填充为 HomogenizingDispersal，将 $|\text{RC}_{\text{Bray}}| \leq 0.95$ 的行填充为 Undominated。此步骤操作中应注意顺序，首先对 $|\beta\text{NTI}| > 2$ 的行进行筛选并填充过程属性，随后再根据 RC_{Bray} 值填充 $|\beta\text{NTI}| \leq 2$ 的行。注意添加 SID 列（图 6 中 A 列）作为唯一值，以便与采样站位经纬度表链接。

	A	B	C	D	E	F
1	SID	Process	StationStart	StationEnd	RC _{Bray}	βNTI
2	ID1	HomogenizingDispersal	S3	S7	-1	0.06
3	ID2	HomogenizingDispersal	S3	S8	-0.99	0.49
4	ID3	HeterogeneousSelection	S3	S10	0.09	10.12
5	ID4	HeterogeneousSelection	S3	S11	-0.02	8.73
6	ID5	HeterogeneousSelection	S3	S12	-0.1	3.78
7	ID6	HomogenizingDispersal	S3	S14	-0.99	-0.27
8	ID7	HomogenizingDispersal	S3	S18	-1	-0.75
9	ID8	Undominated	S3	S20	-0.48	1.86
10	ID9	HeterogeneousSelection	S3	S21	-0.49	3.09
11	ID10	HomogenizingDispersal	S3	S22	-1	-0.62
12	ID11	HomogenizingDispersal	S3	S23	-1	-0.41

图 6. 成对站位群落构建过程属性表

- 在 ArcGIS 中将群落构建过程属性表与站位经纬度表链接，并根据各类过程调整成对站位连线颜色

4.1 将群落构建过程属性表链接至站位经纬度表

右击步骤 2.3 中经 XY To Line 工具转换得到的成对站位连线图层“点集两两连接_XYToLine”图层→Open Attribute Table→Table Options→Joins and Relates→Join，进入 Join Data 对话框设置参数（图 7）。

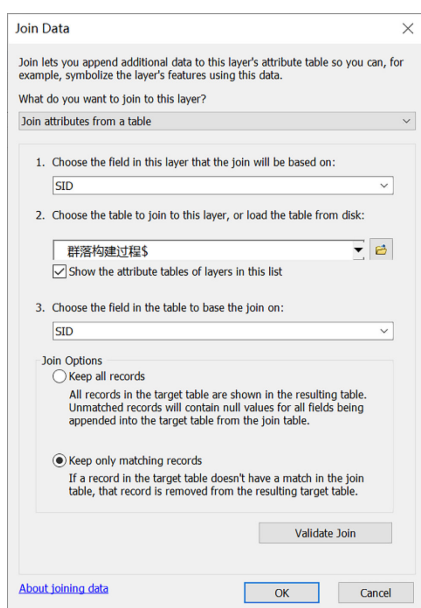


图 7. 表格链接参数设置

4.2 根据群落构建过程为成对站位连线赋色

右击成对站位连线图层，“点集两两连接_XYToLine”图层，**Properties**→**Symbology**→**Categories**→**Unique values**→**Value Field**，选择 **Process** 字段，确定后根据群落构建过程调整线条颜色、宽度、透明度等属性，随后将该图层与已制作好的地图图层（指研究区域地图，包括采样站位、行政区、河流、国界与省界四个图层，实际操作中可根据研究区域针对性选用地图图层）结合，即可生成微生物群落构建过程的空间分布图（图 8）。调整线条属性的操作如下：在图 8 所示界面中双击 **Table Of Contents** 窗口中的“点集两两连接_XYToLine”图层，打开 **Layer Properties** 窗口（图 9），点击 **Symbology**，**Value Field** 选择 **Process**（即图 6 所示 **Process**），然后点击 **Add All Values**，即可显示出所有 5 个生态过程（图 8）。此时，双击每个过程左侧的线条，打开 **Symbol Selector** 窗口，可调整线条颜色和宽度，在 **Layer Properties** 窗口点击 **Display** 则可调节线条的透明度。

图 8 整体展示了 5 类生态过程的分布情况，若仅展示某一过程，则保留该过程的线条颜色，将其他 4 类过程的线条宽度调为 0 即可。

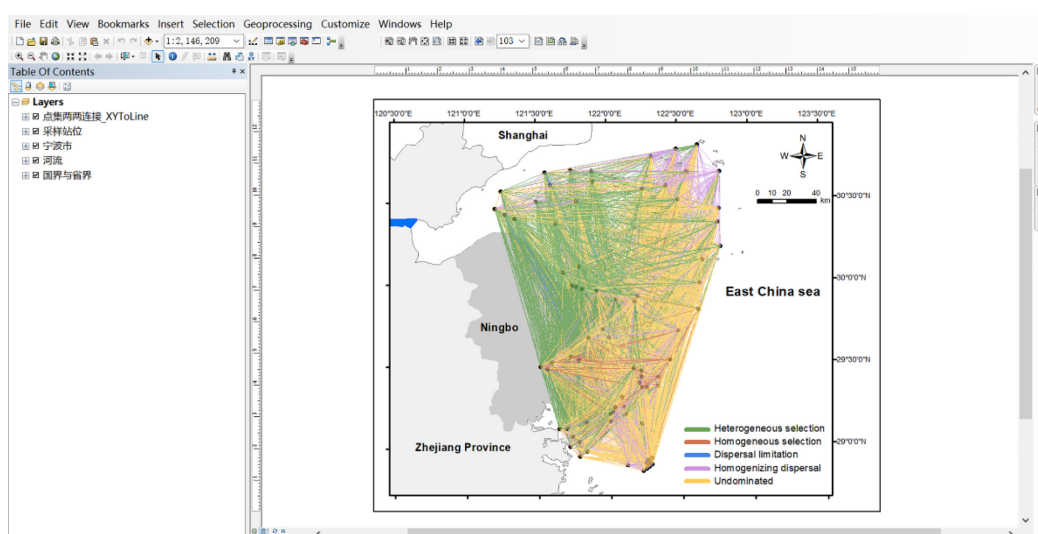


图 8. 微生物群落构建过程的地理空间分布图

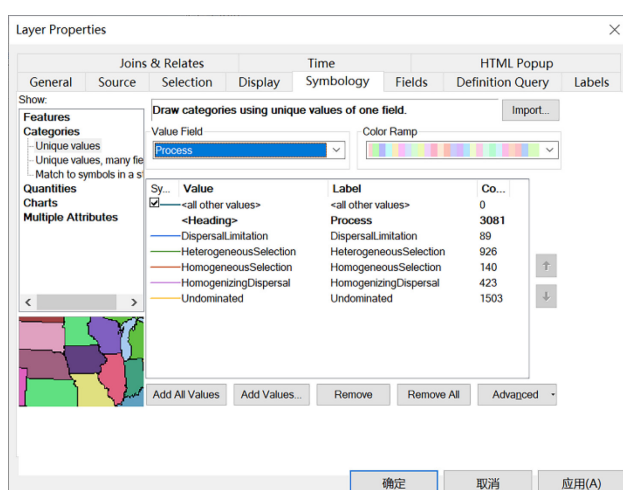


图 9. 线条属性设置窗口

致谢

本项工作得到了国家自然科学基金委员会 (项目编号: 41977192) 的资助。

参考文献

1. Esri, Redlands, CA. ArcGIS for Desktop. Retrieved from: <https://desktop.arcgis.com/>.
2. Stegen, J. C., Lin, X., Fredrickson, J. K., Chen, X., Kennedy, D. W., Murray, C. J., Rockhold, M. L. and Konopka, A. E. (2013). [Quantifying community assembly](#)

- [processes and identifying features that impose them.](#) *The ISME Journal* 7(11): 2069-2079.
3. Wang, K., Hu, H. J., Yan, H. Z., Hou, D. D., Wang, Y. T. Dong, P. S. and Zhang, D. M. (2019). [Archaeal biogeography and interactions with microbial community across complex subtropical coastal waters.](#) *Molecular Ecology* 28(12): 3101-3118.