如何挑选推断系统进化的形态性状

——形态性状的定义及挑选的四个基本规范

How to select morphological characters for inferring systematic phylogeny

– The definition of morphological character and the four basic precepts of selection

刘思培*

动物进化与系统学院重点实验室，中国科学院动物研究所，北京

*通讯作者邮箱: spliu@ioz.ac.cn

引用格式: 刘思培. (2021). 如何挑选推断系统进化的形态性状—形态性状的定义及挑选的四个基本规范. *Bio-101* e1010657. Doi: 10.21769/BioProtoc.1010657.

How to cite: Liu, S. P. (2021). **How to select morphological characters for inferring systematic phylogeny**– The definition of morphological character and the four basic precepts of selection. *Bio-101* e1010657. Doi: 10.21769/BioProtoc.1010657. (in Chinese)

摘要: 在生物学领域，根据不同的目的如推断系统进化、建立分类检索表等挑选的形态性状，既有部分相同又有属性区别。性状挑选是研究系统进化的第一步，但得到的关注远远小于建立系统进化树。性状挑选和由性状载体组成的系统进化树末端单元相互印证。用于推断系统进化的性状的定义是具有系统同源性属性。挑选性状需要遵循 4 个规范：(1) 挑选独立性强的性状；(2) 变化的性状状态是互斥的；(3) 性状状态要编写成离散形式；(4) 性状能够反映进化史信息。建立系统进化树的目标是尽可能多找到符合规范的性状作为独立证据，在特定理论模型计算下筛选出支持系统进化的最优证据。

关键词: 性状，形态学，系统进化，分类，末端单元

1. 性状是什么

在自然科学领域，只要涉及通过定义和描述判定具体事物，都需要挑选性状 (character)。哲学上，传统的观念认为定义事物要通过性状反映事物的基本属性 (essential properties) (Harris and Mishler, 2009)。但一些哲学家也指出通过指物定义即可定义一个事物，不需要关联事物的本质属性 (Kripke, 1980)。在生物学领域，

根据不同的目的如推断系统进化、建立分类检索表、区分生态群落等挑选性状，往往容易引起混乱。关键在于同样的性状被应用于不同的研究目的，其基本属性是有区别的 (Harris and Mishler, 2009)。

2. 用于推断系统进化的性状与用于物种分类的性状的属性差别

用于推断系统进化的性状和用于物种分类的性状最容易被混淆。上世纪中叶，系统学理论被建立之后，界定同一分类阶元从选择整体相似的性状变革为选择具有反映自然进化过程的共近裔性状 (apomorphy)，以保证分类阶元的单系性 (Mishler, 2009)。由于分类学和系统进化形成统一的体系，一部分用于分类的性状和一部分用于推断系统进化历史的性状集重叠。需要注意到，用于分类的性状侧重于描述和区分生物类群 (Hull, 1978)，用于推断系统进化的性状侧重于揭示自然进化历史。因此，根据研究的不同目的，同一性状也会存在属性差异。比如：双翅目具有平衡棒这个性状，在研究昆虫纲系统进化时可作为双翅目的共近裔性状，而对昆虫纲进行分类检索时可以作为双翅目的鉴别特征 (diagnosis)，这二者基本属性不同。

3. 末端单元

系统进化分析包含两个部分：构建性状矩阵和建立系统进化树 (Mishler, 2009)。性状挑选和性状编码连接着标本观察和最终推断结果，是系统进化研究的重要步骤，也可以说系统进化树是性状矩阵的另一种表达方式，然而得到的关注远远小于建立系统进化树 (Pleijel, 1995; Mishler, 2005 and 2009)。性状挑选和末端单元 (TUs: terminal units) 判定相互印证：一个好的性状能够标明末端单元之间的亲缘关系，同时确定一个末端单元意为在此阶元之下所有的性状状态 (character states) 相同 (Hennig, 1966)。系统进化树的基本末端单元不是类群或个体，而是性状载体 (semaphoront)。性状载体是指一个生物体在某个体发育时间点上具有的性状组 (Hennig, 1966)。单一的性状载体由于性状信息不完整不太可能单独成为一个末端单元，因此末端单元由一组具有同源的性状状态的性状载体组成。末端单元不是先验存在的，每次构建系统进化树都需要进行讨论和推断。在很多论文中，末端单元被称为操作分类单元 (OTUs: operational taxonomic units)。这种定义的逻辑缺陷在于默认为分类结果在建立系统进化树之前就已经被确定，而实际上物种分类界定要依据系统进化树的推断结果 (Mishler, 2005)。

4. 挑选性状的 4 个规范

重建不同尺度的进化史，所选择的末端单元也会不同，因此要挑选便于寻找同源性和进化速度适中的性状 (Mishler, 2005 and 2009)。人们一直尝试为系统进化性状明确定义，并将如何挑选用于建立系统进化树的性状形成一个可重复的过程 (Coless, 1985; Grant and Kluge, 2004; Harris and Mishler, 2009)。其中涉及的理论基础源于 Hennig (1966) 提出自然进化历史变化结果为追溯进化历史提供证据，因此自然进化过程中的后验条件在建立系统进化树时变成了先验条件，即可以使用通过变化后形成的共近裔性状去标记进化历史过程。Harris 和 Mishler (2009) 认为推断系统进化性状的定义是具有系统同源性 (phylogenetic homology) 属性，即物种间享有相同的性状状态以示它们来自共同祖先。这里不涉及存在于同一生物体内的旁系同源 (paralogy) 和序列同源 (serial homology)。Serenio (2007) 强调了挑选推断系统进化性状最重要的两个规范：(1) 独立性；(2) 互斥性。此外，他还规定了性状包含的基本元素，即：初级位符 (primary locator: L1)，二级位符 (secondary locator: L2)，变量 (variable: V)，变量修饰语 (variable qualifier: q)：变量状态 (v0, v1,)。如果将描述性状的语句视作哲学命题 (proposition)，性状 (包括位符、变量) 可被看做该句的主词 (subject)，性状状态 (包括变量修饰语) 则是该句的谓语 (predicate) (图 1)。

图 1. 性状包含的基本元素和命题结构

4.1 独立性 (Independent)

不同的性状之间具有从完全互赖连锁到完全彼此独立的连续型变化。建立系统进化树需要尽可能挑选彼此独立性强的性状。这些独立的性状可以分别作为支持一种进化历史过程的独立证据，并在最大简约法、最大似然法等特定的理论模型计算下被筛选 (Mishler, 2005)。Maddison (1993) 为解决状态缺失所导致的最大简约计算结果偏差，试图将尾巴颜色和尾巴存在状态合并为一个性状，但这

样就违背了性状相互独立原则 (Sereno, 2007) (图 2)。

1. Tail, color: red (0), blue (1).
2. Tail: present (0), absent (1).
3. Tail, color: red (0), blue (1), absent (2).

图 2. Maddison (1993)提出的修正方案。他写道: One possibility is to code tail condition as a single three-state character that incorporates both tail presence and color. 该方案将尾巴颜色 (性状 1) 和尾巴的存在状态 (性状 2) 合并为一个性状 (性状 3), 违背了性状相互独立原则。

4.2 互斥性 (Mutually exclusive)

性状状态具有从完全相等到完全不同的连续型变化, 需要根据具体的进化分支选择具有适度变化状态的性状, 因为只有适度变化的性状状态才能够作为重建系统进化历史的潜在标记 (Mishler, 2005)。因此, 性状状态必须是互斥的, 以展现性状状态变化 (Freudenstein, 2005), 例如一条翅脉的状态是存在, 就不能同时拥有不存在的状态。大多数非互斥的形态性状实际上属于两个混在一起的相互独立的性状 (Sereno, 2007), 例如尾巴的颜色状态之间是互斥的, 但颜色状态与存在状态并不互斥 (图 2)。

4.3 离散性 (Discretely describable)

除了上述两个规范, Mishler (2005) 指出性状态要编写成离散形式, 因为进化历史是离散的、由前状态变为后状态的过程。比如使用叶细胞宽度作为判断 6 种山墙藓系统进化的形态特征, 可分别测量每种山墙藓 10 个个体的叶细胞宽度, 并进行方差分析 (ANOVA) 和多重比较 (multiple range tests), 最后筛选出 4 个性状状态 (Mishler and De Luna, 1991) (图 3)。

图 3. 叶细胞宽度的 4 个性状状态。分别测量 6 种山墙藓 *Tortula papillosum* (pa)、*T. caninervis* (c)、*T. norvegica* (n)、*T. obtusissima* (o)、*T. princeps* (pr) 和 *T. ruralis* (r) 每种 10 个个体的叶细胞宽度，进行方差分析和多重范围测试后推断出叶细胞宽度的 4 个性状状态。

4.4 反应进化史信息 (Historically informative)

此外，性状要能够尽量真实地反应进化史信息。可以根据以往经验偏向选择具有可遗传且没有显著趋同进化 (homoplasy) 可能性的性状。由于分类学和系统进化形成统一体系，用于物种分类的性状也需要反映自然进化过程 (Mishler, 2009)。

5. 性状分析

根据上述 4 条规范挑选性状后，基于观察所判定的相同的性状状态称为初级同源 (primary homology) (de Pinna, 1991)。建立系统进化树，即性状分析过程，是使用最大简约法、最大似然法等理论模型再次推断系统进化同源性的过程，且最终推断的系统进化树需要得到最大数量初级同源的支持 (Harris and Mishler, 2009)。建立系统进化树的目标是尽可能多的找到符合规范的性状作为推断系统进化的证据，以确保结果最小程度失真 (Mishler, 2005)。在此过程中，如果相同的性状状态被证实具有同源性，即称为二级同源 (secondary homology) (de Pinna, 1991)，或者被证

实的同源 (corroborated homology); 如果相同的性状状态被拒绝具有同源性就称为趋同进化。需要注意到, 二级同源只能被大概率证实, 也就是具有可证伪性, 而不是像几何定理一样被严格证明。

致谢

感谢 Brent D. Mishler 教授 (美国加州大学伯克利分校) 提供的参考文献。感谢朱朝东研究员 (中国科学院动物研究所) 对本文初稿的修改。感谢匿名审稿人对文章提出的宝贵意见。

竞争性利益声明

本文不存在经济或非经济性竞争利益。

参考文献

1. de Pinna, M.C.C. (1991). [Concepts and tests of homology in the cladistic paradigm](#). *Cladistics* 7: 367-394.
2. Freudenstein, J.V. (2005). [Characters, States and Homology](#). *Syst Biol* 54(6): 965-973.
3. Harris, E.S.J. and Mishler, B.D. (2009). [The delimitation of phylogenetic characters](#). *Biological Theory* 4(3): 2003-234.
4. Hennig, W. (1966). [Phylogenetic systematics](#). Univeristy of Illinois Press. Urbana.
5. Hull, D.L. (1978). [A matter of individuality](#). *Philos Sci* 45: 335-360.
6. Kripke, S. (1980). *Naming and Necessity*. Harvard Univeristy Press. Cambridge, MA.
7. Maddison, W.P. (1993). [Missing data versus missing characters in phylogenetic analysis](#). *Syst Biol* 42(4): 576-581.
8. Mishler, B.D. (2005). The logic of the data matrix in phylogenetic analysis. In: Albert, V.A. (Ed.). In: *Parimony, phylogeny, and genomics*. Oxford University Press, 57-70.
9. Mishler, B.D. (2009). [Three centuries of paradigm changes in biological classication: Is the end in sight?](#) *Taxon* 58(1): 61-67.

10. Mishler, B.D. and de Luna, E. (1991). [The use of ontogenetic data in phylogenetic analyses of mosses](#). *Adv Bryol* 4: 121-167.
11. Pleijel, F. (1995). [On character coding for phylogeny reconstruction](#). *Cladistics* 11: 309-315.
12. Sereno, P.C. (2007). [Logical basis for morphological characters in phylogenetics](#). *Cladistics* 23: 565-587.