

为生物学家造些树

Joseph Malkevitch / 文 巫斌 / 译

在本文中，我将介绍近期数学在遗传学和基因组学上的贡献，以及我们在使用数学工具给遗传学和基因组学问题建立数学模型的过程中是怎样推动数学的学科发展的。

简介

在物理、化学、工程等学科中，数学从来都是非常有价值的研究工具。相比之下，数学直到近期才在生物学上显示出同样重要的作用。数学和生物学相互促进的例子很多，包括发展新的研究方法，分析两个研究对象（比如 DNA 序列和祖先树）之间的关联性，或是研究神经元或细胞的工作机制。不过，数学在生物学的一个小分支上的影响力特别深入，应用特别成功——这就是遗传

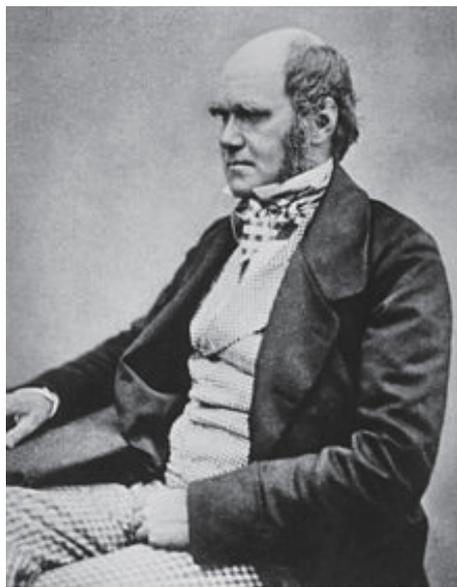


图1 达尔文的照片（来自维基百科）

学领域。虽然在较早的时期基因组学和数学在计算生物学上的应用更为成功，但“经典的”数学遗传学的各种方法和思想直到现在仍然不断地取得新的应用成果。其中一些经典结论有时会被归入人口遗传学的范畴。

在达尔文（Charles Darwin, 1809-1882）以前，就有人提出了物种进化的思想，但达尔文在1859年出版的《物种起源》无疑是思想史上的一个里程碑。达尔文是大自然的一个伟大的观察者，他根据他在英国和（搭乘“贝格尔号”航海探险的）游历过程中对自然界的观察进行理论推演。不过，达尔文在他的研究工作中很少用到数学工具。他写道：“我感到很遗憾，因为我连那些最主要的数学原理都没办法理解得很好。而懂数学的人则常常能比我思考得更深远。”时至今日，很多数学知识都已经广泛应用在遗传学和基因的研究中，包括提高农作物的产量、增加家畜的存活率、研发新药，治疗遗传病以及理解地球上的生命发展进程。

进化论的基本思想是，我们现在看到的物种都是由过去曾经存在的一些物种进化而来，并且任意物种都一定有它的起源。为了理解生物的复杂性，学者们建立了一个分类体系，按照从特殊到一般的顺序对各种生物进行分类。经过很多学者的不懈努力，我们现在终于构建起下面的分类系统：

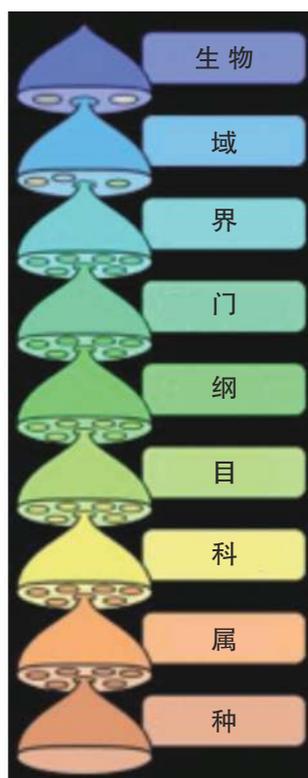


图2 生命形式的层次结构（来自维基百科）

所以，在同一个属内有很多个物种。在美国的教科书中，所有生物被划分为6个界，而英国的教科书只划分5个界。它们所使用的界的名称有：动物界、植物界、真菌界、原生生物界、原生细菌界、细菌界、

无核原生物界。分类系统的每个层次都会引发争议。在很长一段时间内，人们都认为只有一种长颈鹿，但最近有人提出长颈鹿应该被分成四个物种。物种的基本概念是不同的物种不能繁衍共同的后代。然而现实很复杂，这在不同学科中都有所体现。人们早就发现，不同的“物种”是可以交配的，而且已经有很多这样的实例。这种现象被称为“杂交”。杂交形成的杂种有时会分离出来形成新的物种，有时会通过杂种和母种之间的交配导致物种的“融合”。现在的两个典型的杂种例子是狮虎（狮子父亲和老虎母亲）和虎狮（老虎父亲和狮子母亲）。



图3 狮虎（左图）和虎狮（右图）（来自维基百科）

通常情况下，杂交的雄性后代不能继续繁衍，而雌性可以。但有时候也有例外。最近，通过基因组的研究发现，两个原人“物种”——智人和尼安德特人——杂交后形成了现代智人，相关的DNA证据引发了很多关注。

在继续讨论生物分类的问题以前，我们先简单说明一下自达尔文以来发展至今的分类学是如何帮助我们认识生物的。

早期的博物学者在分辨不同的鸟是否属于同一个种类时，通常会使用形体的大小、喙的大小和形状等等标准进行分类。然而，随着现代生物学发现了DNA在遗传过程中所扮演的角色，人们开始使用新的方法来判定哪些物种更接近，哪些距离较远。

我们有很多种方法来表示一个物种的遗传物质的构成。你可以把物种X的基因组想象成一段很长的DNA序列，或是一组染色体，或是位于染色体的不同位置上的一组基因。当我们比较非洲黑猩猩和人类的基因组，会发现很多复杂的情况。人类有46个（23对）染色体，而黑猩猩有48个（24对）染色体。我们也很清楚这些染色体是怎么配对的，并且知道黑猩猩的其中两对染色体和人类的一对染色体是“相对应”的。对于大多数的染色体对，父亲的一条染色体和母亲的一条染色体将组成后代的一对染色体。

通常情况下，遗传物质都出现在细胞核内，即染色体，但线粒体内也包含了遗传物质。线粒体是存在于很多细胞中的结构（细胞器），可以提供细胞所需的能量。人类的线粒体中的遗传物质只能来自母亲。某种程度上，追踪线粒体中的遗传物质简化了遗传特征的决定“机制”，但这也打破了个体的完整遗