DOI :10.19800/j.cnki.aps.1996.03.008

第35卷第3期 古生物学报 Vol.35,No.3
1996年5月 ACTA PALAEONTOLOGICA SINICA May, 1996

距离计算法在化石群对比中的运用

邹欣庆[1)] 郝诒纯[2)] 葛晨东[1)] 许叶华[1)]

1) 南京大学海岸与海岛开发国家试点实验室,南京 210093

2) 中国地质大学,北京 100083

内 容 提 要

分析前人的几种化石群对比方法,指出其中存在的问题,提出了经过改进后的距离计算法。这一方法的特点是:1) 采用半定量取值法,准确地反映出化石群对比思想;2) 在进行世界范围对比时,采用对地方性分子逐步加权的方法,以降低世界性分子在对比过程中所产生的干扰。另外,还选取了全球12个赛诺曼期的钙质超微化石群,运用距离计算法进行对比,进而推测了当时全球表层洋流的格局。

关键词 化石群对比 距离计算法 赛诺曼期

1 前言

自从化石层序律(law of faunal succession)被揭示以后,古生物学研究领域中的化石对比就变成了古生物工作者的一个基本工作手段。这种化石的对比,是为了解决地层的新、老关系。

在古生物地理学的研究中,对比被赋予了新的含义,其目的不仅仅是确定地层新、老关系,更重要的是为了恢复生物地理史,为揭示地球的演变规律提供直接的依据。在过去的几十年中,古生物地理学研究中的化石对比,确实为一些新理论的提出和验证(如大陆漂移说、板块学说等),提供了不可取代的证据。

随着研究程度的深入以及人们对古生物地理分区的需要,化石的对比,已经从一个或几个特征种的对比,上升到整个化石群面貌的对比,并且逐步完成了由定性向定量的转变。

事实上,对化石群进行定量的对比研究早在本世纪初就已开始,当时的对比思想可用公式:

$$AI = \frac{C}{N_1 + N_2 - C} \quad (1)$$

来表示,它运用于两个化石群之间的比较,其中 N_1 为两个化石群中分异度较小的化石群中的分类群数,N_2 为另一化石群中的分类群数,而 C 是表示两个化石群中共有的分类群数。Simpson(1947)发现,两个化石群中共有的分类群数,主要取决于小样本的分类群数,即 N_1,因此他把公式(1)改进为 C/N_1,如果表示成百分比,则是:

$$AI = \frac{100C}{N_1} \quad (2)$$

这一公式一直流行了许多年。

60 年代末、70 年代初,有些古生物学家感觉到,世界性属种在化石群对比中有时会产生出很大的干扰作用,这些干扰往往会掩盖掉应该在对比中起重要作用的土著分类群的意义。因此,Johnson(1971)在对比腕足动物群的时候,用了公式:

$$PI(=AI)=\frac{C}{2E_1} \tag{3}$$

公式中的 C 是两个化石群中的相同属数,E_1 是小样本中的土著分类群数,这个公式与公式(1)、(2)相比,把世界性分子的干扰因素从分母中消除掉了。

对于公式(3),又有人提出,如果化石采集得不够多,对化石群的揭示不完全,那么在这个不完全的化石群中,土著分类群的含量会比实际情况少(土著分类群在化石群中所占的类型百分比,随着化石采集量的增大而增大,一直到趋近其实际百分比)。因此,用公式(3)来计算化石群的相似性时,有时就会显得超出实际相似的程度。

鉴于这种情况,大冢系数(otsuka similarity index)又成了古生物学家爱用的一项指标。大冢系数的数学表达式为:

$$(AI=)\frac{C}{\sqrt{N_1N_2}} \tag{4}$$

但与 Simpson 的公式(2)一样,由于没有考虑到世界性属种的干扰问题,公式(4)同样受到一些学者的指责。

Savage 等(1979)在研究早泥盆世的腕足动物分布时,用了公式:

$$AI=\frac{C-C^{cosm}}{N_1-N_1^{cosm}}\times 100\% \tag{5}$$

式中 C 是两个样本中的共有属,C^{cosm}为共有属中的世界性属,N_1 为小样本中的化石分类单位数,N_1^{cosm} 为小样本中的世界性属,这个公式的特点就是彻底地从分子、分母中把世界性属种的干扰因素全部消除掉,因此,它很好地避免了由于小样本中化石群面貌的不完全而产生的对比过程中的异常相似性,这一点很受古生物地理研究人员的欢迎。

但是,Savage 的公式也有它自身的弱点,概括起来说,有下面这几个方面:

1) 对于世界性属的定义,并没有一个统一的标准,各个研究者所把握的尺度显然是不一样的。

2) 就生物群面貌本身而言,世界性属与土著分类群一样重要,它们对生物群面貌的贡献,同样应当受到重视。

3) 在进行两个化石群相似性比较的时候,如果就两个化石群而言,绝没有世界性属和土著分类群之分。

4) 在上面所列举的所有公式中,说它们是定量对比,只是就计算结果而言,其实,数据的采集只是停留在定性的水平上,化石分类单位在所有的计算中,只是一种(0,1)分布形式,即把化石群中的化石存在形式分为出现与不出现两种,分别用 1 和 0 来表示。这种不考虑化石具体出现频率的数据采集方式,必定会造成大量信息的浪费,如果仅仅是浪费信息量,还可以通过别的工作弥补,可有时会造成一些原则上的失误。例如 3 个化石点中某一化石单元的产量分别为:0、0.5%、90%,在上面一系列计算方法的取值中,这 3 个数分别变成了 0、1、1,

这样的取值就与实际情况大相径庭。

因此，在进行白垩纪钙质超微化石群对比时，笔者试图弥补这些不足之处。基本的工作原理是多元统计的欧氏距离计算以及 Q 型聚类，当然这一工作不是简单的数学公式的复制，在整个计算过程中，包括数据的采集，都自始至终地贯穿着古生物学中的化石群对比思想。具体工作方法介绍如下。

2 采集数据

在所研究的时间面上，尽量多地收集世界各地的钙质超微化石研究资料，取保存比较完整，研究得比较好的化石群为样本，把所有同时间面的样本集中在一起，构成一个统计单元，把一个统计单元的所有钙质超微化石(以种为单位) 定义为每一个样本的变量，这样，在一个统计单元中共有 n 个样本，每个样本有 m 个变量。

变量的取值是一个关键问题。前人的 0、1 取值法虽然方便，但我们已经发现它有很大弊端，因此不可取。剩下的就是化石的绝对百分含量和相对含量。绝对百分含量很精确。但是，如果某一化石种(变量) 在 3 个化石群(样本) 中的百分含量分别为 0、20%、50%，从化石群对比的角度考虑，就这一化石种(变量) 而言，后两个化石群(样本) 的相关性更大，但用精确的百分含量进行数学计算，得到的结果恰恰相关，后两者的距离较远，为 30%，前两个样本的距离较近，只有 20%。因此，用绝对百分含量进行化石群对比的统计计算，与古生物学中的化石群对比思想相悖。

与绝对百分含量相比，化石的相对含量就有明显的优越性。化石含量通常被划分为 4 类，由多至少分别为丰富、常见、偶见、缺失。笔者对各个变量的赋值，就是以这四级划分为基础的，4 个等级分别用数字 3、2、1、0 来赋值。这样，不管是 20%的丰富，还是 50%的丰富，在进行化石群对比时，意义是一样的，笔者对它们的赋值都是 3，而与 0 .5%偶见(赋值为 1) ，就显示了距离。笔者认为这样的取值方法，基本上符合化石群对比的思想。

3 计算过程

将一个统计单元的所有样本(n 个) 集中起来，并对每一个样本中的所有变量(m 个) 赋值以后，就得到一个 m × n 的矩阵，下一步就必须进行计算。

计算的第一步就是对所有的变量值进行标准化变换，其目的就是统一量纲，以避免含量较少的化石种在对比过程中所起的作用被高含量的化石种所掩盖。原始数据为矩阵：

$$X=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1m} \\ X_{21} & X_{22} & \cdots & X_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ X_{n1} & X_{n2} & \cdots & X_{nm} \end{bmatrix}$$

式中 X_{ij} 为第 i 个化石群中的相对含量。标准化变换的公式为：

$$X_{ij}^{'}=\frac{X_{ij}-X_j}{S(j)} \qquad (6)$$

式中：

$$X_j = \frac{1}{n}\sum_{i=1}^{n} X_{ij},$$

$$S(j) = \sqrt{\frac{\sum_{i=1}^{n}(X_{ij} - X_j)^2}{n}}$$

变换后,就得到一个新的矩阵 $\{X'_{ij}\}$,下一步就可以对这一新矩阵进行距离计算。多元统计中欧氏距离的数学表达式为:

$$d_{ik} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(X'_{ij} - X'_{kj})^2} \qquad (7)$$

式中 d_{ik} 为样本 i 和样本 k 的距离,m 为变量数。

这一公式并不能完全反映古生物学中的化石群对比思想,化石群作为样本,化石作为样本中的变量,与其它多元统计的情况有一个很大的不同,就是由于变量是同一时间面上的世界各地的化石,因此在一个化石群样本中,不可能出现所有的变量,有相当一部分的变量为零,也就是化石不出现,即使在两个化石群相比时,也有许多变量,在两个样本中都为零(也就是两个化石群中同时不出现的化石),这些没有出现的化石(变量),在(7)式中被同时赋值为零,这就大大提高了它们在化石群对比中的重要性,而使得两个化石群异乎寻常地相似(d_{ik}值小)。

为解决这一问题,笔者对(7)式进行了修改。目的是对两个化石群中同时不出现的化石在化石群对比中所引起的异常相似性进行校正。具体做法是:从(7)式分母 m 中减去同时不出现的化石种数,以达到增加 d_{ik} 的目的,这样就使得计算结果更接近两个化石群的真实距离。那么,对于这样一个由化石含量组成的特殊矩阵,应该能找到一个新的 $\{m_{ik}\}$ 矩阵,m_{ik} 为第 i 行和第 k 行上不同时为零的变量数,用 m_{ik} 代替(7)式中的 m,就得到一个用于化石群对比的距离计算公式:

$$d_{ik} = \sqrt{\frac{1}{m_{ik}}\sum_{j=1}^{m}(X'_{ij} - X'_{kj})^2} \qquad (8)$$

这就是进行化石群两两对比的数学方法。

上面讨论过关于世界性属的干扰问题,笔者认为在化石群两两相比时,不存在世界性与地方性之分,但要进行全球化石群面貌对比时,不考虑世界性分子的干扰则是不科学的,不过像公式(5)那样,把世界性属一刀切的武断做法,笔者认为也不可取,从某种意义上讲,公式(5)所做的对比,不是化石群总体面貌的对比,而是少数几个地方性分子的对比,这跟化石群总体面貌对比的宗旨相悖。

在进行化石群两两对比以后,就得到了一个距离矩阵 $\{d_{ik}\}$,这是第一次运算的结果。要进行全球化石群面貌的对比,必须做 Q 型聚类运算。既然是全球对比,世界性属的干扰因素就必须被考虑进去。笔者对世界性属干扰因素的处理办法是:对土著分类群加权,使其对化石群面貌的贡献不被世界性属所掩盖,进而削弱世界性属的干扰。因为土著分类群和世界性属不可能截然地区分开,因此,对土著分类群的"加权"处理也不能一视同仁,而是要根据它们出现频率的不一样,进行不同次数的加权,出现频率越高,加权次数越少(所有样本中都出

现的世界性属，则享受不到加权的机会)，反之，出现频率越低，就被视为越具地方性，其享受加权的机会就越多。

具体的计算过程是：从第一次计算的结果矩阵 $\{d_{ik}\}$ 中，取最小值，这个最小距离所反映的两个化石群 a、b 将被视作一类而遭到合并，合并以后它们将与其它样本进行重新比较。

所谓合并，就是建立一个新的样本 1，这个新样本的每个变量 X_{lj} 必须满足：

$$X_{lj}=\frac{X_{aj}+X_{bj}}{2}$$

在合并的同时，就要考虑给土著分类群加权的问题。

土著分类群是指那些分布局限的分类单元，那么在两个样本合并时，我们定义仅在一个样本中出现的分子为土著分类群，则在新合并的样本中给这些变量(土著分类群) 进行第一次加权。土著分类群加权以后，就完成了一个新样本的建立，接着就可以用这个新样本代替合并之前的两个老样本，进行第二次距离计算以及最相似化石群的寻找，这样循环往返，一直到最后两个样本的比较完成以后，整个聚类工作也就结束。

从上面的算法中可以看出，对不同的土著分类群加权的次数是不一样的，这是因为土著分类群的"分布局限性"并没有，也不可能有一个统一的标准。例如有 15 个化石群，在 1 个、2 个，甚至 5 个、8 个化石群中出现的化石，都能被称为土著分类群，但事实上，这些"分布局限性"的大小是不一样的。因此，对它们的加权也不能一视同仁，对局限性大的分子进行多次加权，而分布局限性小的分子则加权次数较少。

上面所介绍的这套化石群整体面貌对比的方法，与前人的工作相比，有两个方面的特点：

1) 克服了定性取值法的弊端，采用了半定量的赋值方式，较正确地反映了化石群对比的思想。

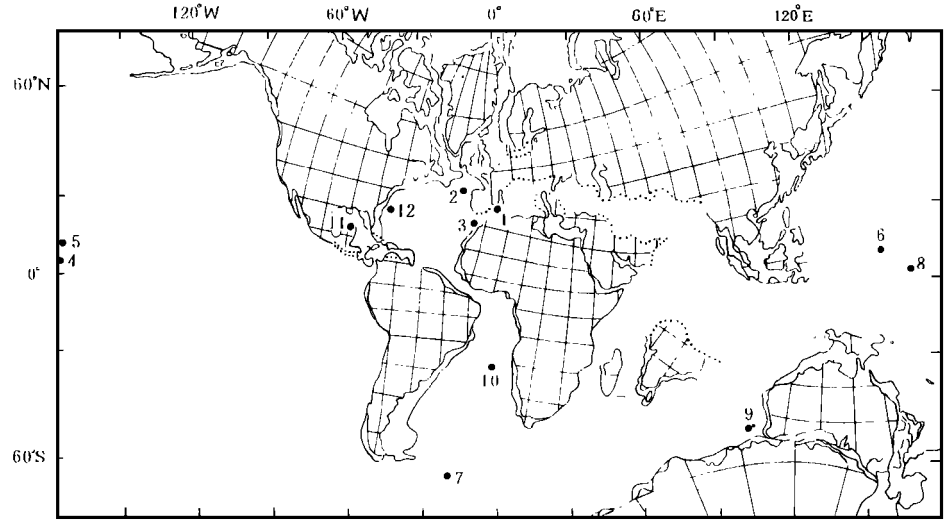


插图 1 赛诺曼期 12 个化石点位置示意图
The location of 12 sample points in Cenomanian

2) 对世界性属产生干扰的处理,不是主观地一刀切,而是采用对土著分类群逐步加权的方式,过滤掉世界性分子所产生的干扰,但保留其对化石群整体面貌的贡献。

4　赛诺曼期全球钙质超微化石群面貌的对比

笔者曾用上述距离计算法,对早阿尔比期－坎潘期(Early Albian－Campanian)共选取 7 个时间面,作了化石群面貌的对比,取得满意的结果,现以赛诺曼期(Cenomanian)为例,作一简单介绍。

表 1　赛诺曼期各化石点之间的距离

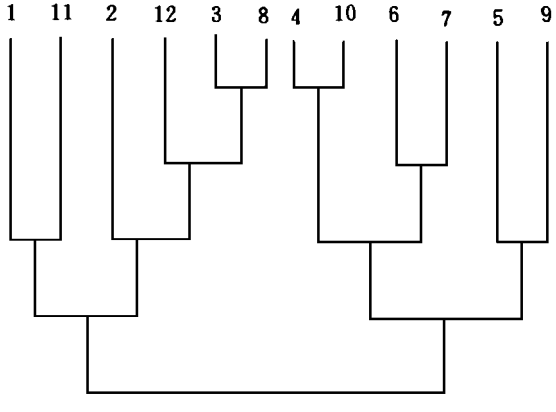
Distance among sample points in Cenomanian

	1	2	3	4	5	6	7	8	9	10	11	12
1	.00	2.32	2.46	2.70	2.60	2.49	2.84	2.73	2.53	2.66	2.38	2.43
2	2.32	.00	2.08	2.20	2.26	2.08	2.41	2.22	2.23	2.18	2.26	2.06
3	2.46	2.08	.00	1.73	1.99	1.61	1.64	1.36	1.85	1.49	2.11	1.85
4	2.70	2.20	1.73	.00	2.08	1.82	2.13	1.73	2.05	1.68	2.31	1.99
5	2.68	2.26	1.99	2.08	.00	2.09	2.39	2.21	2.10	2.08	2.39	2.11
6	2.49	2.08	1.61	1.82	2.09	.00	1.84	1.62	2.02	1.72	2.19	2.00
7	2.84	2.41	1.64	2.13	2.39	1.84	.00	1.53	2.12	1.72	2.47	2.13
8	2.73	2.22	1.36	1.73	2.21	1.62	1.53	.00	2.01	1.37	2.31	1.96
9	2.53	2.23	1.85	2.05	2.10	2.02	2.12	2.01	.00	1.98	2.30	1.93
10	2.66	2.18	1.49	1.68	2.08	1.72	1.72	1.37	1.98	.00	2.27	1.96
11	2.38	2.26	2.11	2.31	2.39	2.19	2.47	2.31	2.30	2.27	.00	2.23
12	2.43	2.06	1.85	1.99	2.11	2.00	2.13	1.96	1.93	1.96	2.23	.00

样本数量:12;种数:94

聚类结果:

样本		距离
3	8	1.36
4	10	1.68
6	7	1.84
3	12	1.91
5	9	2.10
2	3	2.21
1	11	2.38
6	4	2.08
6	5	2.46
1	2	2.58
1	6	2.65



笔者共收集了赛诺曼期 12 个化石点的统计结果(插图 1),资料主要来源于 Thierstein (1974)、Stradner 等(1984)、Watkins 等(1984)。在这 12 个化石点中,共有钙质超微化石 94 种,通过对这 94 个化石种在每个化石点的含量进行赋值,就得到了一个 12×94 的矩阵($n=12, m=94$),运用距离计算法对这一矩阵进行计算(计算过程在计算机上实现),得到的结果列于表Ⅰ。

从表Ⅰ中可以看出,赛诺曼期全球钙质超微生物可以分为两个大区,即北大西洋区和南大西洋-太平洋区,由于钙质超微生物营漂浮生活,大洋表层洋流对它的分布起着重要的作用,因此,从表Ⅰ的结果我们可以判断:赛诺曼期南、北大西洋水体之间并无畅通的表层洋流,相反,南大西洋与太平洋的表层水体有着广泛的交流。

参 考 文 献

Barron, E. J., Harrison, C. G. A., Sloan, J. L. and Hay, W. W., 1981: Paleogeography, 180 million years ago to the present. Ecologae Geol. Helv., **74**: 443–470.

Bergen, J. A., 1986: Nannofossil Biostratigraphy at Site 585, East Mariana Basin. in DSDP V. 89, ed. Orlofsky, S.

Cepek, P., Gartner, S. and Cool, T., 1980: Mesozoic Calcareous Nannofossils DSDP Site 415 and 416, Morocan Basin, DSDP, **50**.

Douglas, R. G. and Savin, S. M., 1973: Oxygen and carbon isotope analyses of Cretaceous and Tertiary foraminifera from the central North Pacific. DSDP, **17**: 591–605.

Fornaciari, E., Raffi, I., Rio, D., Villa, G., Baekman, J., Olafsson, G., 1990: Quantitative Distribution Patterns of Oligocene and Miocene Calcareous Nannofossils from the Western Equatorial Indian Ocean. Proceeding ODP, **115**.

Johnson, J. G., 1971: A quantitative approach to faunal province analysis. Am. Jour. Sci., **270**: 257–280.

Okada, H. and Thierstein, H. R., 1979: Calcareous Nannoplankton-leg 43, DSDP. DSDP, **43**.

Savage, N. M., Perry, D. G. and Boucot, A. J., 1979: A Quantitative Analysis of Lower Devonian Brachiopod Distribution; from Historical Biogeography, Plate Tectonics, and the Changing Environment. Oregon State University Press.

Simpson, G. G., 1947: Holarctic mammalian faunas and continental relationships during the Cenozoic: Geol. Soc. America Bull., **58**: 613–687.

Smith, A. G. and Briden, J. C., 1977: Mesozoic and Cenozoic paleocontinental maps. Cambridge Univ. Press, Cambridge.

Stradner, H. and Steinmetz, H. C. with a contribution by Svabenicka, L. 1984: Cretaceous Calcareous Nannofossils from the Angola Basin, DSDP Site 530. DSDP, **75**: (Ⅱ).

Thierstein, H. R., 1974: Calcareous Nannoplankton-Leg 26, DSDP. DSDP, **26**.

Watkins D. K. and Bowdler, J. L., 1984: Cretaceous Calcareous Nannofossils from DSDP Leg 77, Southeast Gulf of Mexico. DSDP, **77**.

Wiegand, G. E., 1984: Cretaceous Nannofossils from the Northwest African Margin, DSDP Leg 79. DSDP, **79**.

Wise, S. W. Jr. and Wind, F. H., 1977: Mesozoic and Cenozoic Calcareous Nannofossils Recovered by DSDP Leg 36 Drilling on the Falkland Plateau, Southwest Atlantic Sector of the Southern Ocean. DSDP, **36**.

[1994 年 9 月 10 日收到,1995 年 11 月修改]

APPLICATION OF DISTANCE CALCULATION TO FAUNAL COMPARISON

Zou Xin-qing[1), Hao Yi-chun[2), Ge Chen-dong[1) and Xu Ye-hua[1)

1) *State Pilot Laboratory of Coast and Island Exploitation, Nanjing University, Nanjing* 210093

2) *China University of Geosciences, Beijing* 100083

Key words faunal comparison, distance calculation, Cenomanian

Abstract

It is a common knowledge that faunal comparison plays an important role in paleobiogeography study. Many comparison methods have been proposed since the beginning of this century.

After summarizing some mathematical methods, which have been widely used for several decades, 4 shortages are pointed out:

1) There is no fixed standard to distinguish cosmopolitan and endemic taxa; 2) Since cosmopolitan taxa make the same important contribution to the whole fauna as endemic taxa, they can not be omitted; 3) There is no difference between cosmopolitan elements and endemic elements when comparison is made between only two faunas; 4) Only qualitative valuation is adopted in the prior calculation, which may cause serious mistakes in faunal comparison.

In view of these facts, distance-calculation is brought forward as a new method, which is superior to other methods at least in two aspects:

1) Instead of qualitative valuation, hemi-quantitative valuation is adopted during calculation; 2) In order to remove the interference produced by cosmopolitan taxa when world-wide comparison is made, endemic taxa are weighted progressively. In addition, 12 Cenomanian calcareous nannofloras from all over the world are gathered together to make a comparison by the new method.

Two interesting conclusions can be drawn through calculation:

1) Two biotic provinces are outlined based on the calculation results, which are: North Atlantic Province, and South Atlantic-Pacific Province; 2) Surface current connects South Atlantic with Pacific instead of North Atlantic in Cenomanian.