Arch Microbiol (2001) 176 : 386–390
DOI 10.1007/s002030100345

COMMENTARY

Milind G. Watve · Rashmi Tickoo · Maithili M. Jog
Bhalachandra D. Bhole

How many antibiotics are produced by the genus *Streptomyces*?

Received: 13 March 2001 / Revised: 30 July 2001 / Accepted: 14 August 2001 / Published online: 25 September 2001

Abstract *Streptomyces* is the largest antibiotic-producing genus in the microbial world discovered so far. The number of antimicrobial compounds reported from the species of this genus per year increased almost exponentially for about two decades, followed by a steady rise to reach a peak in the 1970s, and with a substantial decline in the late 1980s and 1990s. The cumulative number shows a sigmoid curve that is much flatter than what a logistic equation would predict. We attempted to fit a mathematical model to this curve in order to estimate the number of undiscovered antimicrobials from this genus as well as to predict the trends in the near future. A model assuming that the screening efforts are encouraged by a previous year's success and that the probability of finding a new antibiotic is a function of the fraction of antibiotics undiscovered so far offered a good fit after optimizing parameters. The model estimated the total number of antimicrobial compounds that this genus is capable of producing to be of the order of a 100,000 – a tiny fraction of which has been unearthed so far. The decline in the slope appeared to be due to a decline in screening efforts rather than an exhaustion of compounds. Left to itself, the slope will become zero in the next one or two decades, but if the screening efforts are maintained constant, the rate of discovery of new compounds will not decline for several decades to come.

Keywords Actinomycetes · Antibiotic diversity · Drug discovery · Drug screening · Logistic model · Mathematical model · Secondary metabolites · *Streptomyces*

M.G. Watve (✉) · R. Tickoo · M.M. Jog · B.D. Bhole
Department of Microbiology, Abasaheb Garware College, Pune 411 004, India
e-mail: watve@vsnl.com,
Tel.: +91-20-5440311, Fax: +91-20-4338009

Introduction

Natural products have been the largest contributors to drugs in the history of medicine. Of the 520 new drugs approved between 1983 and 1994, 39% were natural products or derived from natural products and 60–80% of antibacterial and anticancer drugs were derived from natural products (Harvey 2000). In recent years, however, screening of natural products, particularly microbial products has fallen out of favor. Modern techniques including *in silico* screening, combinatorial biosynthesis (Cane et al. 1998; Hutchinson 1998) and combinatorial biocatalysis (Michels et al. 1998) are increasingly occupying the central stage in drug discovery. However, it is becoming increasingly apparent that 99% of the diverse bacterial species is unexplored (Davies 1999; Ward et al. 1990; Watve et al. 2000). As attempts are being made to explore the uncultured or so-called unculturable bacteria using culture-independent molecular methods (Handelsman et al. 1998; Hugenholtz 1998; Seow et al. 1997), it is also becoming increasingly clear that the majority of these bacteria may in fact be culturable (Watve et al. 2000) and thus accessible to classical methods of screening. These novel organisms are a likely source of a large number of bioactive compounds.

Diminishing returns from *Streptomyces* screening

But are the taxa popularly screened for such compounds explored enough? *Streptomyces*, for example, is known to be the largest antibiotic-producing genus. The history of antibiotics derived from the genus *Streptomyces* begins with the discovery of the first antibiotic, streptothricin, in 1942. It was streptomycin, however, which came 2 years later, that triggered systematic screening of antibiotics from this genus (Berdy 1974; Vandamme 1984; Waksman 1963). Since then there have been continued efforts towards screening of novel antimicrobial compounds from the genus *Streptomyces*. Between 1955 and 1962, about

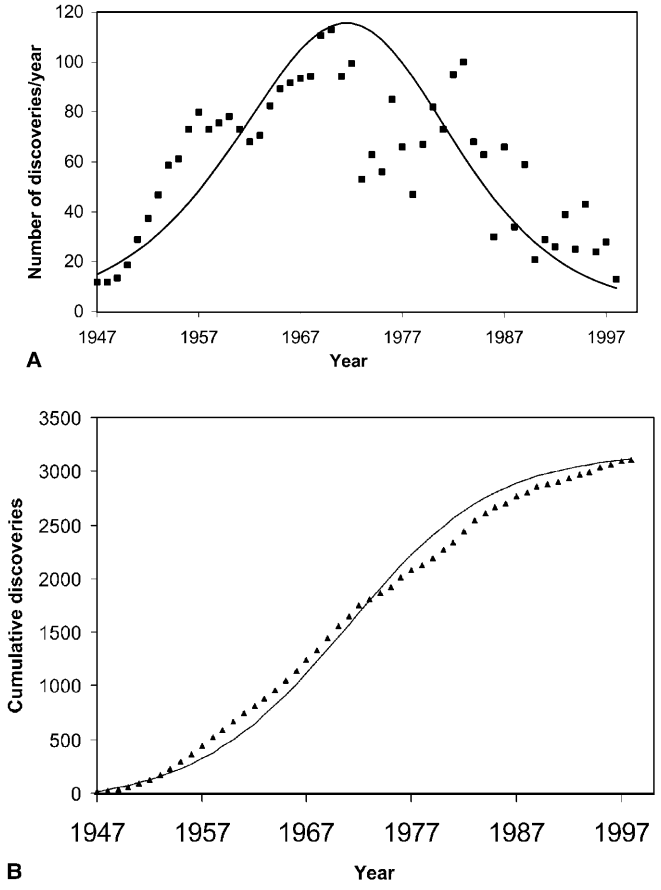

Fig. 1 The best-fit logistic model for antibiotic discovery. The *markers* denote data and the *line* the fitted logistic curve. The trend of accumulation of antimicrobial compounds is considerably flatter than that predicted by the logistic model. **A** Discoveries per year, **B** cumulative trend

80% of the antibiotics originated from actinomycetes, among which the major contributor was *Streptomyces* (Berdy 1974). Here, we attempt to analyze the trend in the discovery of antimicrobial compounds from this genus, a process that has been going on for over 50 years, and answer the question how many more compounds may be left to discover in this genus?

Asking such questions is relevant since infectious diseases continue to be a major health problem (Kerr and Lacey 1995; Zaehner and Fiedler 1995). This is due to a number of reasons, including the changing spectrum of pathogens, antibiotic resistance in pathogens (Cassell 1997; Culotta 1994; Travis 1994), immunocompromising treatments, and immunodeficiency diseases (Zaehner and Fiedler 1995). To meet the continued demand for new antibiotics against the resistant pathogens (Chopra et al. 1997), screening efforts need to be continued. Contrary to this need, antibiotic screening, particularly from actinomycetes, has stagnated. It is becoming increasingly difficult to obtain novel compounds, and screening more often yields the same compounds again and again (Goodfellow and O'Donell 1989; Zaehner and Fiedler 1995). The reduced probability of finding novel compounds in screening, reduced funding for screening efforts, exhaustion of compounds produced by this genus, and/or a change in the emphasis from drug screening to combinatorial chemistry and drug designing may explain the cause of the decline in the rate at which novel antimicrobial compounds are reported from this genus.

In order to plot the trend in the discovery of antimicrobials, we compiled a list of all the antibiotics coming from this genus that have been reported in *Chemical Abstracts* and/or the *Journal of Antibiotics*. After eliminating any obvious double reporting, we plotted the trend from 1947 to 1999 (Fig. 1A). The number of compounds discovered in a calendar year increased almost exponentially for about 17 years and then continued to rise at a lesser linear rate, reaching a peak in the 1970s. In the 1970s and early 1980s, the number of new compounds fluctuated between 70 and 100 per year, but in the late 1980s and 1990s declined rapidly. The cumulative trend is a sigmoid curve with an increasing slope in the initial years and a decreasing slope in the later phase (Fig. 1B).

Fitting a model to the trend

A sigmoid curve in population biology is typically described by a logistic equation (Brown and Rothery 1993; Maynard Smith 1968). A discrete form of logistic equation is: $X(t)–X(t–1)=X(t–1)\cdot r\cdot(1–X(t–1)/K)$, where, X is the population, r the growth rate of the population, and K the carrying capacity of the environment. The logistic equation has been used for a number of other applications (Brown and Rothery 1993; Gore and Lavraj 1992). In this case, we considered X to be the cumulative number of antibiotics discovered and K the total number of antibiotics produced by this genus. We attempted to fit a logistic equation to the curve with r and K optimized iteratively to minimize the sum of squared deviations.

The antibiotic curve was much flatter than what a best-fit logistic equation predicted (Fig. 1A, B). A logistic model therefore was found to be unsuitable to describe the trend. There are a number of intuitive reasons why a logistic model could fail. The assumption of a logistic model, that the increment in X is directly proportional to X when X is small, may not be applicable to the rate at which discoveries are made. There is no intuitive reason why discoveries per unit time should depend upon the numbers discovered so far. Instead they might depend upon the rate at which discoveries were made in the recent past. A good standing rate of discovery would encourage more research, attract funding as well as ensure ready availability of expertise. We therefore modified the equation as: $X'(t)=X'(t–1)\cdot r\cdot(1–X(t–1)/K)$, where, $X'(t)$ is the number of antibiotics discovered in the year t. This equation was not radically different than the logistic equation in its shape and therefore this modification alone did not appreciably improve the fit.

The term $(1–X/K)$ in a logistic equation indicates the fraction of the carrying capacity unoccupied and functions as a feedback term that regulates the growth rate of a pop-

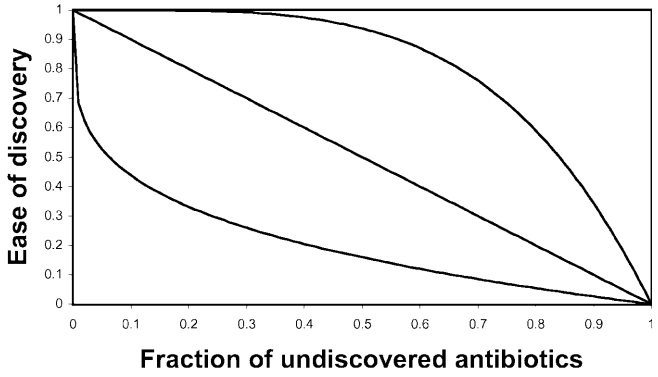

Fig. 2 The decline in the probability of finding new compounds. If all compounds were equally abundant, the relationship would have been linear. Due to the commonness of some compounds and the rarity of others, a concave relationship is expected, which is obtained at fractional values of Y in the equation $X'(t)=X'(t–1)\cdot r\cdot(1–(X(t–1)/K)^Y)$ (see text). The *convex line* shown represents Y=4 and the concave one Y=0.25

ulation as the carrying capacity is saturated. This term is intuitively logical for the rate of discoveries, since the rate will depend upon the fraction of undiscovered antibiotics. The rate should be high when no or few antibiotics have been discovered, since any compound detected has a high probability of being novel. The rate would drop with more discoveries accumulating and tend to zero as most compounds have been discovered. The simplistic assumption that the rate depends linearly on the number of undiscovered compounds, however, need not be correct. The relationship could be linear if all the compounds are more or less equally abundant. The distribution of relative abundances in biology is well known to be highly skewed, with a large number being rare or moderately abundant and a smaller number being very abundant (Krebs 1989; May 1975). Abundance distribution from diverse communities is typically described by log-normal distribution (Preston 1948, 1962). If sampled randomly, the abundant types have a high probability of being discovered. They are therefore the ones to be discovered first. With increasing efforts, more and more of the left hand side of the distribution is uncovered. If the abundance distribution of antibiotics is similar, the probability of discovering new compounds would not decline linearly with the number of discoveries. There would be a rapid decline initially as the commoner ones are discovered resulting in a concave relationship. We modified this term as $(1–(X/K)^Y)$. At Y=1 the function was linear. At Y>1 it was convex and at Y<1 it was concave (Fig. 2), the degree of concavity being decided by the value of Y. The predictive model now took the form: $X'(t)=X'(t–1)\cdot r\cdot(1–(X(t–1)/K)^Y)$, where, $X'(t)$ is the number of antibiotics discovered in the year t, $X(t–1)$ is the cumulative number of discoveries up to the year t–1, r is a rate constant and K is an estimator of the total number of antimicrobials produced by the genus.

In this form the equation modeled the number of discoveries per unit time as a product of screening efforts in that unit time denoted by the term $X'(t–1)\cdot r$ and the probability of a detected compound being novel as $(1–(X(t–1)/K)^Y)$. Unlike the logistic equation this equation had a positive and a negative feedback. The rate could decline with a decreasing number of undiscovered compounds. But perhaps much before this could happen another positive feedback cycle could dominate. A decrease in X' would cause a decrease in the screening efforts further decreasing X'. If such a feedback cycle was triggered, the trend could saturate much before the compounds were exhausted.

A large number of antibiotics waiting to be discovered

The three parameters of the model, namely r, K and Y, were optimized iteratively to give a best-fit equation to existing data and minimizing the sum of squared deviations (Fig. 3A, B). This equation offered a considerably better fit than a logistic equation and validated our new set of assumptions leading to the equation. The equation gave a flatter curve at fractional values of Y, and the desired flatness that fitted the observed trend was obtained at Y=0.27. The equation that offered the best fit was: $X'(t)=1.306\cdot X'(t–1)\cdot(1–(X(t–1)/294300)^{0.27})$

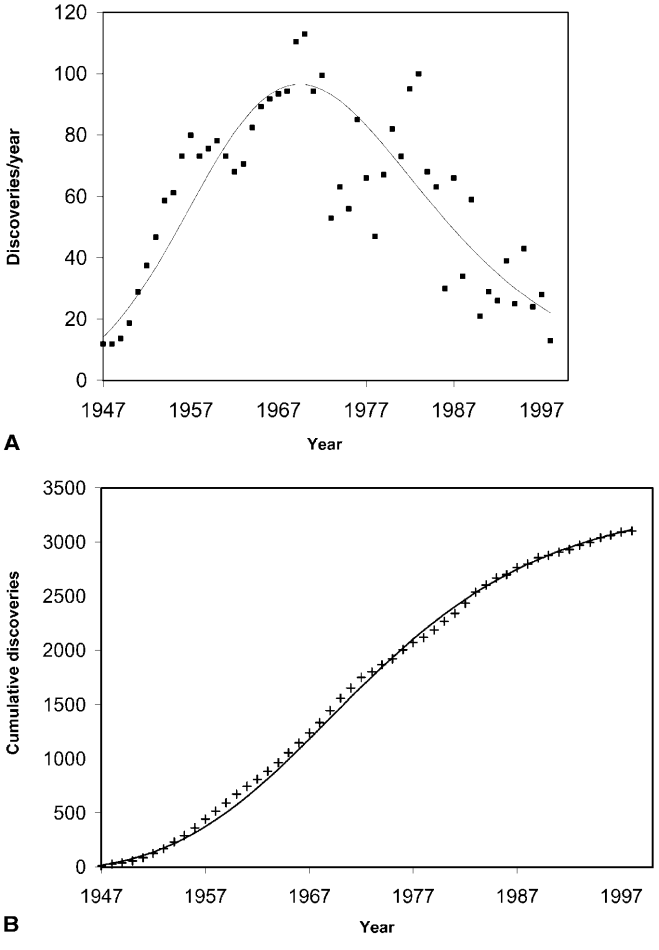

Fig. 3 The fractional power logistic equation.This equation offers a better fit to the data than a classical logistic equation. **A** Discoveries per year, **B** cumulative trend

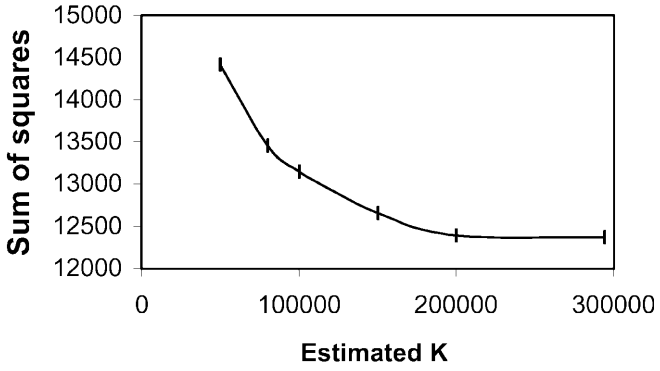

Fig. 4 The sum of squares at different values of *K*, the estimated total number of antibiotics, after optimizing *r* and *Y*. Although the best estimator of *K* is 294,000, the sum of squares is not substantially different for somewhat smaller values of *K*. A more conservative estimate therefore could be close to 150,000

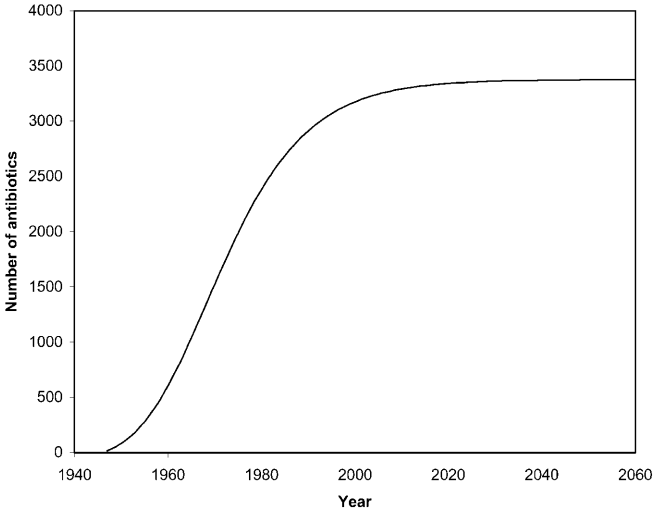

Fig. 5 Predicting antibiotic discovery with the current trend. Screening for antimicrobial compounds from *Streptomyces* will practically come to an end within a decade or so if the present trend continues

The best estimator of the number of antimicrobials present in the genus was 294,300. A more conservative estimate of 150,000 should be acceptable since there was a small difference in the sum of squares at these two values of *K* when the other two parameters were optimized. Below 150,000 the sum of squares increased rapidly (Fig. 4). In any case, the decline in the rate of discovering new compounds in the 1990s is unlikely to be due to near exhaustion of compounds, since even if we accept the more conservative estimate, only about 3% of the existing compounds have been reported so far.

According to the model, a decline in screening efforts was the major cause of the saturating trend seen in the 1990s. The dynamics of discoveries probably entered a vicious positive feedback cycle during this decade. Left to itself, the declining trend will bring *Streptomyces* screening almost to an end before the year 2020 (Fig. 5), leaving

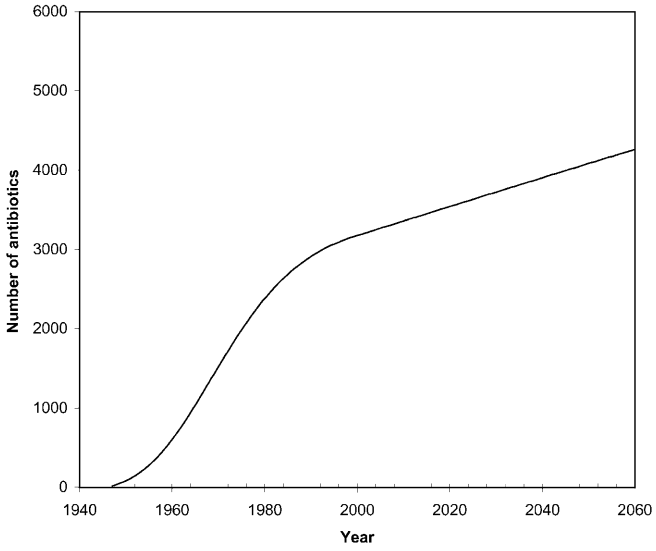

Fig. 6 Predicting with an assumption that the effort level remains constant. If the effort levels are kept constant in the model, by making it independent of $X'(t-1)$, it is predicted that novel compounds continue to accumulate for several decades

a vast majority of compounds still undiscovered. Since the end of screening is likely to be brought about by declining efforts, it can very well be prevented. Keeping the effort constant in the model at the 1995 level, the model predicted that a steady increase in the number of discovered compounds will continue for several decades (Fig. 6).

A fractional value of *Y* (0.27) indicated that the ease curve was considerably concave. An encouraging prediction can be made from the concave function. The initial steep decline of the curve, which represents the discovery of the more abundant compounds, would slowly transform into almost a straight line sloping gently. After having discovered a few thousand compounds, a typical screening yields more "old friends" than novel compounds (Goodfellow and O'Donell 1989; Zaehner and Fiedler 1995). However, since the steepest part of the curve is over by now, the difficulty will not increase as rapidly as it has in the previous years. Therefore the efforts needed to discover a novel compound will remain fairly stable for several decades to come. In any case, the realization that a large number of antibiotics from this genus are waiting to be discovered should rejuvenate efforts to screen for novel compounds and to develop more efficient methods for screening and identifying compounds.

Acknowledgements We thank Madhav Gadgil, A.P. Gore and S.A. Paranjape for comments on the model.

References

Berdy J (1974) Recent developments of antibiotic research and classification of antibiotics according to chemical structure. In: Perlman D (ed) Advances in applied microbiology. Academic, New York, pp 308–406

Brown D, Rothery P (1993) Models in biology: mathematics, statistics and computing. Wiley, London
Cane DE, Walsh CT, Khosla C (1998) Harnessing the biosynthetic code: combinations, permutations and mutations. Science 282 : 63–68
Cassell H (1997) Emergent antibiotic resistance: health risks and economic impact. FEMS Immunol Med Microbiol 18:271–274
Chopra I, Hodgson J, Metcalf B, Poste G (1997) The search for antimicrobial agents effective against bacteria resistant to multiple antibiotics. Antimicrob Agents Chemother 41:497–503
Culotta E (1994) Funding crunch hobbles antibiotic resistance research. Science 264:362–363
Davies J (1999) Millennium bugs. Trends Biochem Sci 24:M2–M5
Goodfellow M, O'Donell AG (1989) Search and discovery of industrially significant actinomycetes. In: Baumberg S, Hunter I, Rhodes M (eds) Microbial products: new approaches Cambridge University Press, Cambridge, pp 343–383
Gore AP, Lavraj UA (1992) Modeling innovation diffusion: some methodological issues. J Sci Industr Res (CSIR) 51:291–295
Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245–R249
Harvey A (2000) Strategies for discovering drugs from previously unexplored natural products. Drug Discovery Today 5:294–300
Hugenholtz P, Goebel BM, Pace NR. (1998) Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol 180:4765–4774
Hutchinson CR (1998) Combinatorial biosynthesis for new drug discovery. Curr Opin Microbiol 1:319–329
Kerr KG, Lacey RW (1995) Why do we still get epidemics. In: Hunter PA, Darby GK, Russell NJ (Ed) Fifty years of antimicrobials: past perspective and future trends. SGM symposium 53. Cambridge University Press, Cambridge, pp 179–203
Krebs CJ (1989) Ecological methodology. Harper and Row, New York
May RM (1975) Patterns of species abundance and diversity In: Cody ML, Diamond JM (eds) Ecology and evolution of communities. Belknap, Cambridge, Massachusetts, pp 81–120
Maynard Smith J (1968) Mathematical ideas in biology. Cambridge University Press, Cambridge
Michels PC, Khmelnitsky YL, Dordick JS, Clark DS (1998) Combinatorial biocatalysis: a natural approach to drug discovery. Trends Biotechnol 16:210–215
Preston FW (1948) The commonness and rarity of species. Ecology 29:254–283
Preston FW (1962) The canonical distribution of commonness and rarity. Ecology 43:410–432
Seow KT, Meurer G, Gerlitz M, Wendt-Pienkowski E, Hutchinson CR, Davies J A (1997) Study of iterative type II polyketide synthases, using bacterial genes cloned from soil DNA: a means to access and use genes from uncultured microorganisms. J Bacteriol 179:7360–8
Travis J (1994) Reviving the antibiotic miracle? Science 264:360–362
Vandamme EJ (1984) Antibiotic search and production: an overview. In: Vandamme EJ (ed) Biotechnology of industrial antibiotics. Marcel Dekker, New York, pp 3–32
Waksman SA (1963) The actinomycetes and their antibiotics. In: Umbreit W W (ed) Advances in applied microbiology, vol 5. Academic, New York, pp 235–316
Ward DM, Weller R, Bateson MM (1990) 16S rRNA sequence reveal numerous uncultured microorganisms in a natural community. Nature 345:63–65
Watve MG, Shejval V, Sonawane C, Rahalkar M, Matapurkar M, Shouche Y, Patole M, Phadnis N, Champhekar A, Damle K, Karandikar S, Kshirsagar V, Jog M (2000) The 'K' selected oligophilic bacteria: a key to uncultured diversity? Curr Sci 78:1535–1542
Zaehner H, Fiedler H (1995) The need for new antibiotics: possible ways forward. In: Hunter PA, Darby GK, Russell NJ (eds) Fifty years of antimicrobials: past perspective and future trends. SGM symposium 53 Cambridge University Press, Cambridge, pp 67–85