Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism

Paul A. Hoskisson,[a] Ryan F. Seipke[b]

[a]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom
[b]Faculty of Biological Sciences, Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, United Kingdom

ABSTRACT Microbial natural products, particularly those produced by filamentous *Actinobacteria*, underpin the majority of clinically used antibiotics. Unfortunately, only a few new antibiotic classes have been discovered since the 1970s, which has exacerbated fears of a postapocalyptic world in which antibiotics have lost their utility. Excitingly, the genome sequencing revolution painted an entirely new picture, one in which an average strain of filamentous *Actinobacteria* harbors 20 to 50 natural product biosynthetic pathways but expresses very few of these under laboratory conditions. Development of methodology to access this "hidden" biochemical diversity has the potential to usher in a second Golden Era of antibiotic discovery. The proliferation of genomic data has led to inconsistent use of "cryptic" and "silent" when referring to biosynthetic gene clusters identified by bioinformatic analysis. In this Perspective, we discuss this issue and propose to formalize the use of this terminology.

KEYWORDS *Actinobacteria*, natural products, secondary metabolism, specialized metabolism, *Streptomyces*

Almost a century of research and industrial activity has earned *Streptomyces* species and other filamentous *Actinobacteria* a well-deserved reputation for their ability to produce a vast array of natural products, many of which have found utility in the clinic. Consequently, these microbes have been an important resource for the pharmaceutical industry. Indeed, some compounds that were developed into drugs (e.g., clavulanic acid, chloramphenicol, spectinomycin, ivermectin, amphotericin, and nystatin) are now considered essential medicines by the World Health Organization (1). The Golden Age of antibiotic discovery came to an end in the 1970s, because existing pipelines for new molecules ran dry and rediscovery of known compounds was common (the so-called "dereplication problem"). Generally, natural product discovery programs were replaced with combinatorial chemistry platforms that produced large chemical libraries for high-throughput screening, which largely failed to produce new lead compounds. In the longer term, industrial antibiotic discovery programs were no longer financially viable and Big Pharma divested from discovery programs (2). This reduction in antibiotic discovery capacity has been against a backdrop of increasing antimicrobial-resistant infections within the clinic and community to the point that there is now an urgent need to discover and develop novel antimicrobials for the clinic.

It is not all doom and gloom, however; we are living in the "Genomics Age" of antibiotic discovery, which began in 2002 with the sequencing of the first *Streptomyces* genome, *Streptomyces coelicolor* A3(2) (3). Analysis of this genome sequence identified a greater number of natural product biosynthetic gene clusters (BGCs) than the organism was known to produce. A further 18 BGCs were identified, but at the time, only four compounds had been observed in the laboratory. This exciting revelation gave rise to the notion that on the whole, secondary metabolism in *S. coelicolor* A3(2)

is relatively inactive in a laboratory setting, with most molecules detected in laboratory culture having been previously identified and linked to biosynthetic gene clusters (BGCs) (Known Knowns).

This was established as a widespread phenomenon, rather than a quirk of *S. coelicolor*, when the genomes of *Streptomyces avermitilis*, *Streptomyces griseus*, and *Streptomyces scabies* were sequenced and analyzed (4–6). Since these genomes were "mined" for secondary metabolites, it has become apparent that the genomes of *Actinobacteria* represent a vast repository of novel natural product chemistry. For example, a study of 830 actinobacterial genomes identified >11,000 natural product BGCs that represent >4,000 chemical families (with chemical families being compounds with similar chemical backbones that may possess similar physical and chemical properties [7]) which tells us that there is vast chemical space out there in nature. While there are rare cases of similar or identical molecules being produced by evolutionarily distinct BGCs (8, 9), the prevalence of this scenario will remain unknown until high-throughput chemical analysis catches up with genomics. The newfound abundance of genomic data has resulted in an explosion of methods to computationally mine actinobacterial genomes for natural product BGCs such as antiSMASH, PRISM, EvoMining, and DeepBGC (10–13), which can be linked to natural product databases (MIBiG, antiSMASH database, and Natural Product Atlas) (14–16). The prioritization of which BGCs to study remains a challenge in the field, but the wealth of data available indicates that there are many new compounds out there to be discovered (17).

The increase in genome sequencing and genome mining studies has, however, led to inconsistency in the terminology used to refer to the "hidden treasures" encoded by the genomes of natural product-producing organisms. In the literature, the use of "cryptic" and "silent" has become pervasive, and these terms are often used interchangeably to describe natural product BGCs for which the genes encoding the enzymes for biosynthesis have been identified but for which no product has been observed during laboratory culture. Inconsistent use of these terms is confusing, especially for newcomers to the field. In this perspective, we disambiguate "silent" and "cryptic" and propose rules for their use when applied to natural product BGCs.

"Cryptic" is derived from the Greek *kruptós* meaning "hidden" and is used widely in biology to describe well-camouflaged species (https://en.wiktionary.org/wiki/cryptic) or where two or more species have been misclassified as a single taxon due to being morphologically indistinguishable but are genetically distinct (cryptic species) (18). In microbiology the use of "cryptic" to describe plasmids that do not encode any known beneficial traits (19) is found regularly, but this perhaps reflects our poor understanding of the role these plasmids play in the ecology of many organisms, rather than them not being beneficial. "Silent" is a straightforward term that most are familiar with. For instance, "gene silencing" is well understood in RNA interference (RNAi) scenarios and "silent" is also used to describe gene conversion events in antigenic variation—a gene is present, but is not expressed, but can be expressed following a mutational event. These systems often reflect complicated and/or poorly understood gene regulatory systems. Thus, we propose that the term "cryptic" should be used for describing BGCs and/or natural products that are hidden or unknown and the term "silent" should be used only for describing BGCs that are not expressed.

For instance, consider the following scenario and what the description of the BGC and product would be. One has bioinformatically analyzed a *Streptomyces* genome sequence, and a BGC has been identified that encodes a putative polyketide biosynthetic (PKS) system. Expression of the BGC has not been analyzed, so the use of "silent" is inappropriate. Future experimentation may establish that the PKS BGC is in fact unexpressed, which means the BGC should at this point be described as "silent." By definition this means that the associated polyketide product is, of course, "cryptic," because it cannot be produced if its biosynthetic genes are not expressed. We outline various other scenarios for the usage of cryptic and silent in Fig. 1.

The majority of the confusion has crept into literature when discussing the Known Unknowns and the Unknown Knowns. This is where it is important for the term "cryptic"
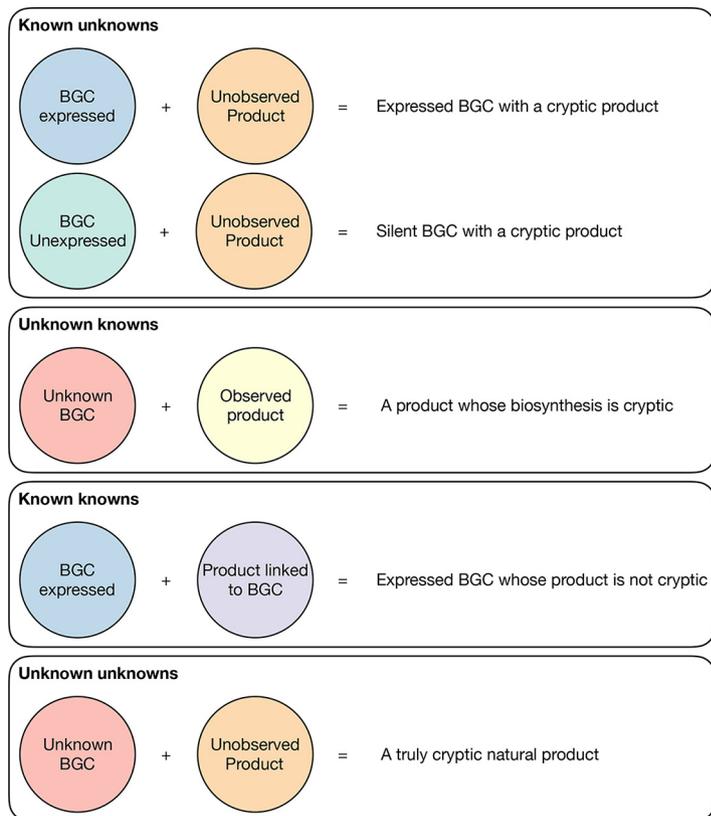
FIG 1 A summary of the description of various permutations of the cryptic or silent nature of secondary metabolism.

to be used accurately. The use of "cryptic BGC" is appropriate when a natural product has been observed, but where its cognate BGC has not been experimentally identified, i.e., the compound is known, but its biosynthesis is cryptic (Unknown Knowns). Equally, the term "cryptic" can be appropriately used when expression of a BGC has been experimentally validated, but a product predicted from that BGC cannot be observed under laboratory culture conditions (Known Unknowns), i.e., the product of the BGC is cryptic.

Interestingly, it is the last category, the Unknown Unknowns, where the real hidden treasures likely lie. Currently, around one-third of all protein-encoding genes in bacterial genomes lack functional annotation (20). It is within these hypothetical protein-encoding genes where truly "cryptic BGCs" (Unknown Unknowns) may be located. These are completely novel BGCs, where the lack of functional annotation precludes them from being identified by the multitude of bioinformatic tools and natural product databases. The resulting secondary metabolites are also more likely to be missed when analyzing culture supernatants for bioactive molecules—we often look for the kinds of molecules that we are already familiar with. Again, screening natural product libraries for bioactivity does not investigate all "bioactivity space," since it is usually confined to activity against a panel of well-known surrogate indicator organisms/cell lines (21).

Progress has been made in identifying these kinds of truly cryptic gene clusters using evolutionary mining methods to identify the expansion and repurposing of enzyme families from central metabolism. This is exemplified by the discovery of a BGC for an arsenopolyketide (22) and an unusual thiotemplated nonribosomal peptide synthetase (NRPS)-independent BGC encoding closthioamide biosynthesis (23, 24). The same evolutionary rationale applies to the duplication of some housekeeping genes, which confer resistance to the antibiotic, so-called target duplication (25). Broadly speaking, these evolutionarily guided approaches can expand the predictions of com-

monly used BGC prediction algorithms by up to 26% (12) and can potentially identify BGCs for natural products that have novel modes of action (25). With the exciting expansion of genomic, metagenomic, and metabolomic resources and their integration with both small- and large-scale functional studies (26), we owe it to ourselves to be precise in our terminology.

REFERENCES

1. World Health Organization. 2019. WHO model list of essential medicines, 21st list (amended March 2020).

2. Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. Nat Rev Drug Discov 6:29–40. https://doi.org/10.1038/nrd2201.

3. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang C-H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabbinowitsch E, Rajandream M-A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature 417:141–147. https://doi.org/10.1038/417141a.

4. Ikeda H, Ishikawa J, Hanamoto A, Shinose M, Kikuchi H, Shiba T, Sakaki Y, Hattori M, Omura S. 2003. Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. Nat Biotechnol 21:526–531. https://doi.org/10.1038/nbt820.

5. Ohnishi Y, Ishikawa J, Hara H, Suzuki H, Ikenoya M, Ikeda H, Yamashita A, Hattori M, Horinouchi S. 2008. Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. J Bacteriol 190:4050–4060. https://doi.org/10.1128/JB.00204-08.

6. Bignell DRD, Seipke RF, Huguet-Tapia JC, Chambers AH, Parry RJ, Loria R. 2010. *Streptomyces scabies* 87-22 contains a coronafacic acid-like biosynthetic cluster that contributes to plant–microbe interactions. Mol Plant Microbe Interact 23:161–175. https://doi.org/10.1094/MPMI-23-2 -0161.

7. Doroghazi JR, Albright JC, Goering AW, Ju K-S, Haines RR, Tchalukov KA, Labeda DP, Kelleher NL, Metcalf WW. 2014. A roadmap for natural product discovery based on large-scale genomics and metabolomics. Nat Chem Biol 10:963–968. https://doi.org/10.1038/nchembio.1659.

8. Kim SY, Ju K-S, Metcalf WW, Evans BS, Kuzuyama T, van der Donk WA. 2012. Different biosynthetic pathways to fosfomycin in *Pseudomonas syringae* and *Streptomyces* species. Antimicrob Agents Chemother 56: 4175–4183. https://doi.org/10.1128/AAC.06478-11.

9. Sit CS, Ruzzini AC, Van Arnam EB, Ramadhar TR, Currie CR, Clardy J. 2015. Variable genetic architectures produce virtually identical molecules in bacterial symbionts of fungus-growing ants. Proc Natl Acad Sci U S A 112:13150–13154. https://doi.org/10.1073/pnas.1515348112.

10. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi .org/10.1093/nar/gkz310.

11. Skinnider MA, Merwin NJ, Johnston CW, Magarvey NA. 2017. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. Nucleic Acids Res 45:W49–W54. https://doi.org/10.1093/ nar/gkx320.

12. Sélem-Mojica N, Aguilar C, Gutiérrez-García K, Martínez-Guerrero CE, Barona-Gómez F. 2019. EvoMining reveals the origin and fate of natural product biosynthetic enzymes. Microb Genom 5:e000260. https://doi .org/10.1099/mgen.0.000260.

13. Hannigan GD, Prihoda D, Palicka A, Soukup J, Klempir O, Rampula L, Durcak J, Wurst M, Kotowski J, Chang D, Wang R, Piizzi G, Temesi G, Hazuda DJ, Woelk CH, Bitton DA. 2019. A deep learning genome-mining

strategy for biosynthetic gene cluster prediction. Nucleic Acids Res 47:e110. https://doi.org/10.1093/nar/gkz654.

14. Medema MH, Kottmann R, Yilmaz P, Cummings M, Biggins JB, Blin K, de Bruijn I, Chooi YH, Claesen J, Coates RC, Cruz-Morales P, Duddela S, Düsterhus S, Edwards DJ, Fewer DP, Garg N, Geiger C, Gomez-Escribano JP, Greule A, Hadjithomas M, Haines AS, Helfrich EJN, Hillwig ML, Ishida K, Jones AC, Jones CS, Jungmann K, Kegler C, Kim HU, Kötter P, Krug D, Masschelein J, Melnik AV, Mantovani SM, Monroe EA, Moore M, Moss N, Nützmann H-W, Pan G, Pati A, Petras D, Reen FJ, Rosconi F, Rui Z, Tian Z, Tobias NJ, Tsunematsu Y, Wiemann P, Wyckoff E, Yan X, et al. 2015. Minimum information about a biosynthetic gene cluster. Nat Chem Biol 11:625–631. https://doi.org/10.1038/nchembio.1890.

15. Blin K, Pascal Andreu V, de los Santos ELC, Del Carratore F, Lee SY, Medema MH, Weber T. 2019. The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res 47:D625–D630. https://doi.org/10.1093/nar/ gky1060.

16. van Santen JA, Jacob G, Singh AL, Aniebok V, Balunas MJ, Bunsko D, Neto FC, Castaño-Espriu L, Chang C, Clark TN, Cleary Little JL, Delgadillo DA, Dorrestein PC, Duncan KR, Egan JM, Galey MM, Haeckl FPJ, Hua A, Hughes AH, Iskakova D, Khadilkar A, Lee J-H, Lee S, LeGrow N, Liu DY, Macho JM, McCaughey CS, Medema MH, Neupane RP, O'Donnell TJ, Paula JS, Sanchez LM, Shaikh AF, Soldatou S, Terlouw BR, Tran TA, Valentine M, van der Hooft JJJ, Vo DA, Wang M, Wilson D, Zink KE, Linington RG. 2019. The Natural Products Atlas: an open access knowledge base for microbial natural products discovery. ACS Cent Sci 5:1824–1833. https://doi.org/10.1021/acscentsci.9b00806.

17. Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-Olvera A, Ramos-Aboites HE, Hoskisson PA, Barona-Gómez F. 2020. Evolutionary dynamics of natural product biosynthesis in bacteria. Nat Prod Rep 37:566–599. https://doi.org/10.1039/C9NP00048H.

18. Darlington CD. 1941. Taxonomic systems and genetic systems, p 137–160. *In* Huxley J (ed), The new systematics. Clarendon Press, Oxford, United Kingdom.

19. Wein T, Hülter NF, Mizrahi I, Dagan T. 2019. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. Nat Commun 10:2595. https://doi.org/10.1038/s41467-019-10600-7.

20. Antczak M, Michaelis M, Wass MN. 2019. Environmental conditions shape the nature of a minimal bacterial genome. Nat Commun 10:3100. https://doi.org/10.1038/s41467-019-10837-2.

21. Newman D. 2017. Screening and identification of novel biologically active natural compounds. F1000Res 6:783. https://doi.org/10.12688/ f1000research.11221.1.

22. Cruz-Morales P, Kopp JF, Martínez-Guerrero C, Yáñez-Guerra LA, Selem-Mojica N, Ramos-Aboites H, Feldmann J, Barona-Gómez F. 2016. Phylogenomic analysis of natural products biosynthetic gene clusters allows discovery of arseno-organic metabolites in model streptomycetes. Genome Biol Evol 8:1906–1916. https://doi.org/10.1093/gbe/evw125.

23. Dunbar KL, Büttner H, Molloy EM, Dell M, Kumpfmüller J, Hertweck C. 2018. Genome editing reveals novel thiotemplated assembly of polythioamide antibiotics in anaerobic bacteria. Angew Chem Int Ed Engl 57:14080–14084. https://doi.org/10.1002/anie.201807970.

24. Dunbar KL, Dell M, Gude F, Hertweck C. 2020. Reconstitution of polythioamide antibiotic backbone formation reveals unusual thiotemplated

assembly strategy. Proc Natl Acad Sci U S A 117:8850–8858. https://doi
.org/10.1073/pnas.1918759117.

25. Mungan MD, Alanjary M, Blin K, Weber T, Medema MH, Ziemert N. 2020.
ARTS 2.0: feature updates and expansion of the antibiotic resistant
target seeker for comparative genome mining. Nucleic Acids Res 48:
W546–W552. https://doi.org/10.1093/nar/gkaa374.

26. Aron AT, Gentry EC, McPhail KL, Nothias L-F, Nothias-Esposito M, Bous-
limani A, Petras D, Gauglitz JM, Sikora N, Vargas F, van der Hooft JJJ,
Ernst M, Kang KB, Aceves CM, Caraballo-Rodríguez AM, Koester I, Weldon
KC, Bertrand S, Roullier C, Sun K, Tehan RM, Boya P CA, Christian MH,
Gutiérrez M, Ulloa AM, Tejeda Mora JA, Mojica-Flores R, Lakey-Beitia J,
Vásquez-Chaves V, Zhang Y, Calderón AI, Tayler N, Keyzers RA, Tu-
gizimana F, Ndlovu N, Aksenov AA, Jarmusch AK, Schmid R, Truman AW,
Bandeira N, Wang M, Dorrestein PC. 2020. Reproducible molecular net-
working of untargeted mass spectrometry data using GNPS. Nat Protoc
15:1954–1991. https://doi.org/10.1038/s41596-020-0317-5.