Nucleic Acids Research, 2026, gkag416
https://doi.org/10.1093/nar/gkag416
Web Server issue

COMMBAT: a web platform for exploring expression control of biosynthetic gene clusters

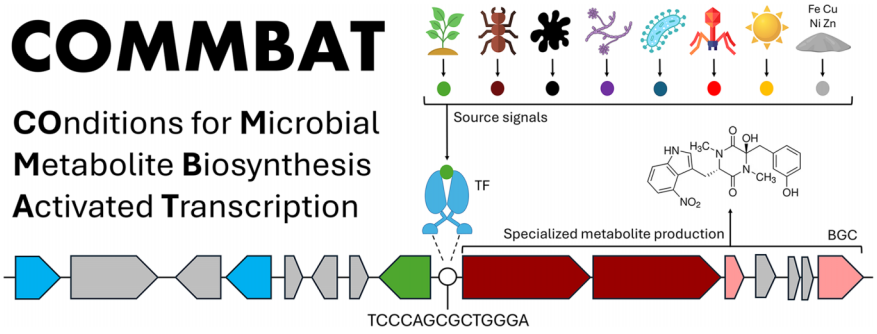
Silvia Ribeiro Monteiro [1], Augustin Rigolet [2], Clément Jeunehomme [1], Julianne Gathot [1], Yasmine Kerdel [1], Matthias Henry [1], Hannah E. Augustijn [2,3,4], Marnix H. Medema [3], Gilles P. van Wezel [2], Sébastien Rigali [1,*]

[1]InBioS—Center for Protein Engineering, University of Liège, Institut de Chimie, Liège B-4000, Belgium
[2]Molecular Biotechnology, Institute of Biology, Leiden University, Sylviusweg 72, 2333 BE Leiden, The Netherlands
[3]Bioinformatics Group, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands
[4]Present Address: Department of Chemistry & Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, United States
*To whom correspondence should be addressed. Email: srigali@uliege.be

Abstract

Bacterial genomes contain thousands of biosynthetic gene clusters (BGCs) responsible for the production of structurally diverse natural products with applications in medicine, agriculture, and biotechnology. Expression of these BGCs is tightly regulated by transcription factors (TFs) responding to environmental cues, yet predicting which TFs regulate specific BGCs remains challenging. In particular, TF binding sites (TFBSs) within BGCs often diverge from canonical motifs, limiting the effectiveness of standard motif-scanning approaches and hindering systematic exploration of BGC regulation. Here, we present COMMBAT (COnditions for Microbial Metabolite Biosynthesis Activated Transcription), a framework for large-scale prediction of TF–BGC regulatory interactions across bacterial genomes. COMMBAT integrates motif matching with genomic context and gene function information to predict functional TFBSs. The COMMBAT web platform (https://www.commbat.uliege.be) enables users to (i) identify BGCs potentially regulated by a given TF, and (ii) predict candidate TFs that control a specific BGC. With over 4000 TF position weight matrices from four public repositories and more than 400 000 BGCs from MIBiG and antiSMASH DB, COMMBAT provides a scalable resource to predict regulatory inputs and guide/prioritize culture conditions and genetic engineering strategies for natural product discovery.

Graphical abstract



Introduction

Bacteria produce structurally diverse specialized metabolites that play central roles in ecology, medicine, and biotechnology. The genes required for their production are typically organized in biosynthetic gene clusters (BGCs), which encode core and tailoring biosynthetic enzymes, transporters, resistance determinants, and regulatory elements [1]. Genome sequencing has revealed that only a small fraction of predicted BGCs has been experimentally characterized [2], leaving a vast reservoir of cryptic BGCs with immense pharmaceutical and biotechnological potential unexplored. Unlocking this hidden diversity requires not only advances in genome mining and metabolomics but also a deeper understanding of the regulatory mechanisms that govern BGC expression [3–5].

A major bottleneck in exploiting the biosynthetic potential encoded in microbial genomes is the limited understanding of transcription factor regulatory networks (TFRNs) controlling BGC expression. While genome mining has uncovered millions of BGCs, predicting the environmental signals that activate them remains challenging. TFs act as sensors of

Received: March 11, 2026. **Revised:** April 7, 2026. **Accepted:** April 17, 2026

environmental or intracellular signals (e.g. nutrients, stress, or signaling molecules) and modulate BGC expression accordingly. Thus, ecological interactions and environmental cues strongly influence natural product biosynthesis, underscoring the importance of regulatory networks for rational activation of cryptic BGCs [3, 6]. High-throughput approaches that map transcription factor (TF) binding are expected to greatly enhance our ability to predict BGC regulation and guide elicitor-based discovery strategies [3, 7, 8]. Regulatory signatures can also provide functional clues; for example, binding sites for the iron homeostasis regulator DmdR1 were harnessed to uncover novel BGCs involved in iron homeostasis [9]. Systematic exploitation of TF binding information therefore provides a promising route to connect environmental signals with specialized metabolism, and, occasionally, even infer the function (biological activity) and the structure (building blocks and precursors) of specialized metabolites [10].

BGC transcription is typically controlled by both cluster-situated (pathway-specific) and global TFs [11, 12]. However, TF binding sites (TFBSs) of global TFs within BGCs often deviate from the canonical motifs recognized in a TF's core regulon, making BGC-associated TFBSs challenging to detect using standard motif-scanning tools [12]. Consequently, studies of BGC regulation have largely focused on individual TF–BGC pairs [12, 13], an approach that does not scale to the rapidly growing number of sequenced genomes and BGC repositories such as MIBiG [14] and antiSMASH DB [15]. As a result, systematic prediction of TF/elicitor pairs remains an underexploited strategy, despite its promise for activating silent BGCs and uncovering novel natural products [3].

To enable systematic exploration of regulatory interactions in specialized metabolism, we developed COMMBAT (COnditions for Microbial Metabolite Biosynthesis Activated Transcription). COMMBAT integrates TF position weight matrix (PWM)-based motif detection with genomic and functional context to improve TFBS prediction within BGCs [13]. The associated web platform provides a user-friendly interface for high-throughput analyses, allowing users to (i) identify BGCs potentially regulated by a given TF or (ii) predict TFs likely to control a selected BGC. By facilitating large-scale mapping of TF–BGC interactions, COMMBAT broadens our ability to connect environmental cues and transcriptional control with specialized metabolism expression, thereby supporting the rational activation of cryptic BGCs and the discovery of novel natural products.

Materials and methods

Packages and frameworks

COMMBAT is coded in Bash, Perl and R, and is freely available at https://gitlab.uliege.be/Silvia.RibeiroMonteiro/commbat.git. The scripts use a series of Perl and R dependencies, which are listed in the "README" and "LICENSE-3RD-PARTY" files. The software requires as input a set of BGCs and TF position weight matrices (PWMs). The BGCs are scanned by the PREDetector software [16] (available at https://gitlab.uliege.be/ptocquin/RPREDetector.git), which uses the BGC's sequence and description files (in FASTA and GFF3 format, respectively) and the TF PWMs (in FASTA format) generated according to the expression described by Hertz and Stormo [17]. Only TFBSs with a PWM score equal or superior to zero are retained for subsequent COMMBAT scoring.

COMMBAT score calculation

The COMMBAT score estimates the likelihood that a TF regulates the expression of a given BGC. The scoring formula was developed and described by Ribeiro Monteiro *et al.* [13], and is based on the INTERACTION SCORE (I) and the TARGET SCORE (T), following the expression:

$$\text{COMMBATscore} = \frac{1}{2}[I+T] = \frac{1}{2}\left[\frac{I_{\text{TFBS}}}{I_{\max}} + \frac{R+\max\ (F)}{2}\right],$$

where:

- I evaluates the binding affinity between a TF and its predicted binding site based on the normalized PWM score: ratio of a predicted TFBS (I_{TFBS}) to the maximum PWM score ($I_{\max}$) [13].
- T integrates (i) the Region Score (R), which reflects the genomic location of the predicted TFBS, and (ii) the Function Score (F), based on the functional categories of genes within BGC that are predicted to be regulated by the TF [13]. The Function Score (F) is derived from the genomic context of the predicted TFBS by identifying co-transcribed genes and their functional categories (regulatory, core biosynthetic, additional biosynthetic, transport/resistance, or other genes) using antiSMASH annotations). Co-transcribed genes are predicted based on experimentally validated data from a meta-analysis of 71 polycistronic BGC transcription units, whereby genes located within −90 to +170 nt relative to the upstream gene stop codon are considered part of the same transcription unit [13]. The selected F score, max(F), corresponds to the score of the gene within the transcription unit whose functional category is predicted to most significantly impact the expression of the BGC [13]. Category weights are based on our previous meta-analysis of over 300 experimentally validated TFBSs, which revealed a hierarchy in TF targeting preferences [12]. Based on these observations, weights were assigned as follows: regulatory genes (1.0), core biosynthetic genes (0.8), additional biosynthetic genes (0.5), transport and self-resistance genes (0.2), and other/unknown function genes (0.1).

The performance of the COMMBAT scoring approach compared to conventional PWM-based methods has been assessed and demonstrates its superior ability to identify TFBSs involved in BGC expression control [13].

Novelty score calculation

To provide a rough estimation on whether a BGC is known or cryptic (the genetic material is not yet associated with a metabolite) COMMBAT computes a Novelty score based on antiSMASH's "KnownClusterBlast" similarity values [15]. The Novelty score ranges from 0 (high similarity to a characterized BGC) to 1 (no detectable similarity), allowing prioritization of cryptic BGCs with potentially novel chemistry.

Databases

BGC datasets can be selected from: MIBiG repository (v. 4.0) [18], antiSMASH DB (version 5) [15], and user's antiSMASH predictions (version 8) [19]. The PWMs from four databases are available: COMMBAT DB, LogoMotif [20], Prodoric [21], and RegPrecise [22].

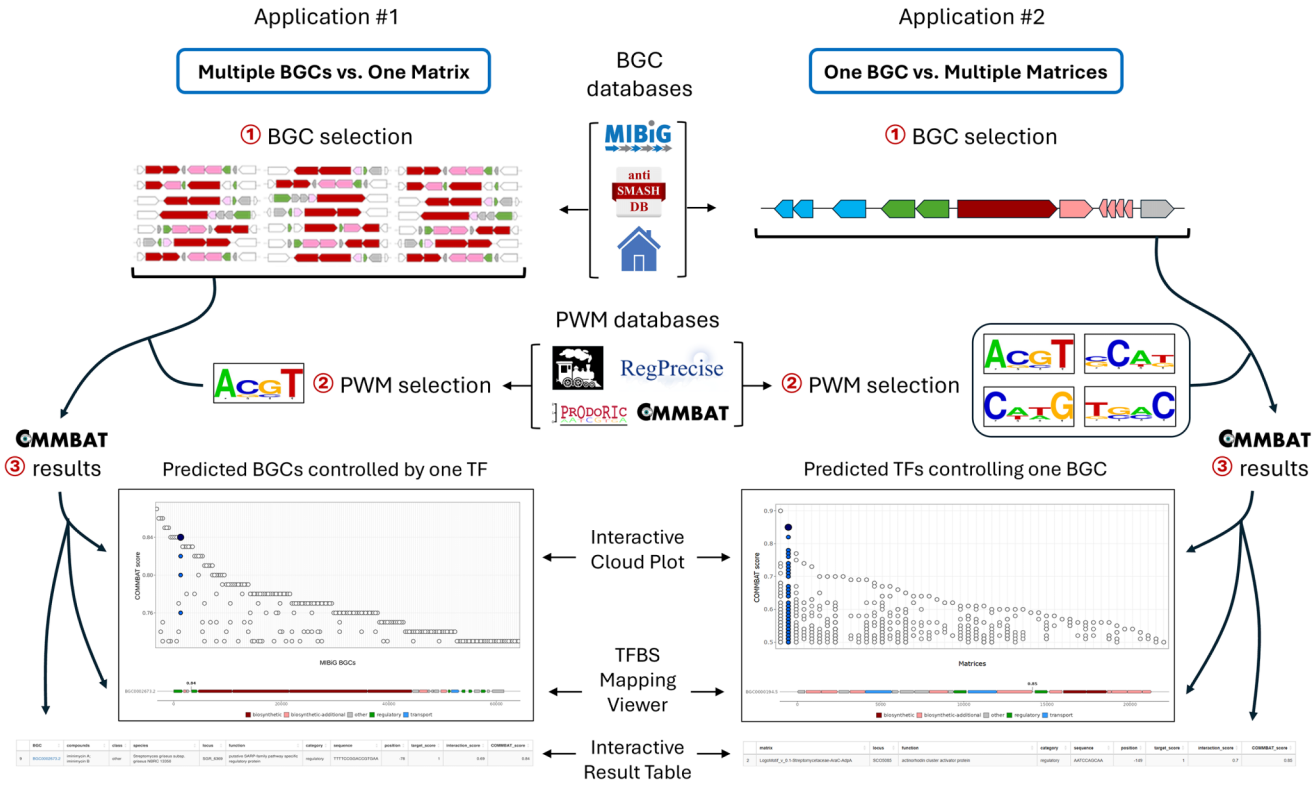


Figure 1. COMMBAT workflow. COMMBAT supports two types of analysis: Application #1, prediction of BGCs potentially controlled by one TF; Application #2, prediction of candidate TFs controlling the expression of a specific BGC. The result output page includes an Interactive Cloud Plot, a TFBS Mapping Viewer, and an Interactive Result Table.

Web application implementation and availability

The website interface was developed in R (v. 4.5.0) [23] with the R package RShiny (v. 1.10.0) [24] and deployed with the open-source Shiny Server v1.5.20.1002 (https://github.com/rstudio/shiny-server). Web visualizations are generated using R packages "ggplot2" (v. 3.5.2) [25] and "gggenes" (v. 0.5.1) [26], and interactive plots are created with "plotly" (v. 4.10.4) [27]. The web application is freely available at "https://www.commbat.uliege.be'' and uses the developed COMMBAT software code, available at https://gitlab.uliege.be/Silvia.RibeiroMonteiro/commbat.git. Both the COMMBAT web-interface and code will be regularly maintained.

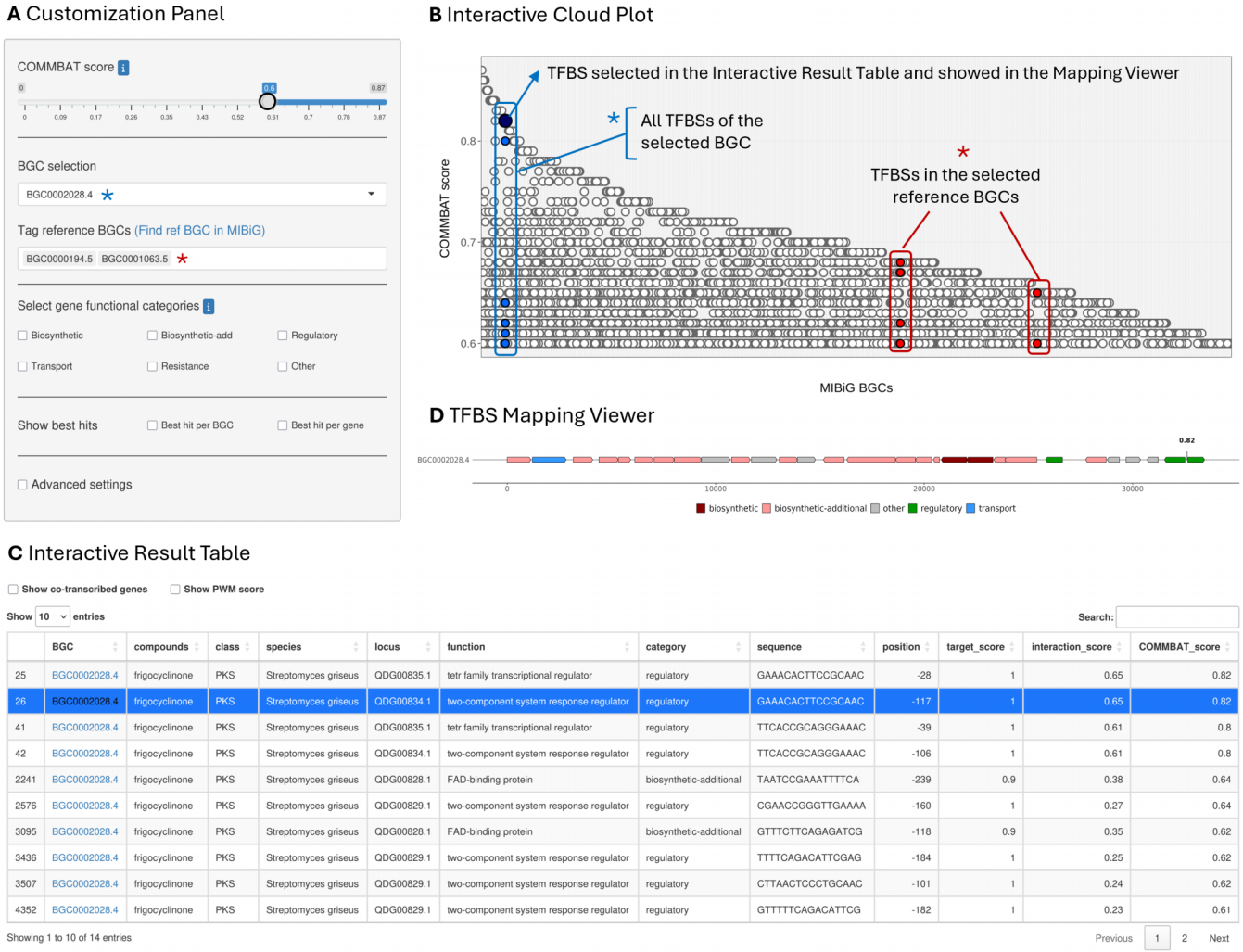
Results

Web server overview

The COMMBAT web tool enables large-scale predictions of the transcriptional controls governing specialized metabolism. It provides two applications (Fig. 1): "*Multiple BGCs versus One Matrix*" which identifies BGCs potentially regulated by a single TF, and "*One BGC versus Multiple Matrices*" which predicts candidate TFs controlling a specific BGC.ced The current implementation integrates more than 4000 TF position weight matrices (PWMs) and enables exploration of over 400 000 BGCs from public databases. The COMMBAT score represents the likelihood that a TF regulates a given BGC by integrating predicted TF binding affinity with the genomic position of the binding site and the functional relevance of the associated target gene [13]. The selected BGCs can be retrieved from the MIBiG repository (only known BGCs), or from antiSMASH database or predictions (both including known and cryptic BGCs). For each application, users select BGCs (step 1, BGC selection) and TF PWMs (step 2, PWM selection). The results (step 3) are displayed through an interactive cloud plot, a result table, and a TFBS mapping viewer. In this visualization, BGCs associated with higher COMMBAT scores are predicted to have more biologically relevant regulatory interactions with the queried TF.

Step I: Input parameter, BGC selection

COMMBAT supports BGC selection from three sources: (i) MIBiG database [18], (ii) antiSMASH database [15], and (iii) user-uploaded collections derived from antiSMASH predictions [19]. From MIBiG repository (version 4.0; 2636 curated BGCs), users may select one, multiple, or all BGCs using reference identifiers, organism names or taxonomic queries. antiSMASH DB entries (version 5; over 400 000 bacterial BGCs from 54 800 bacterial species) can be also filtered by BGC reference ID, genome accession number, or organism name. To maintain computational efficiency, analyses are currently limited to 20 bacterial species per session. Users may also analyze their own antiSMASH results by providing the antiSMASH job ID.



	BGC	compounds	class	species	locus	function	category	sequence	position	target_score	interaction_score	COMMBAT_score
25	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00835.1	tetr family transcriptional regulator	regulatory	GAAACACTTCCGCAAC	-28	1	0.65	0.82
26	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00834.1	two-component system response regulator	regulatory	GAAACACTTCCGCAAC	-117	1	0.65	0.82
41	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00835.1	tetr family transcriptional regulator	regulatory	TTCACCGCAGGGAAAC	-39	1	0.61	0.8
42	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00834.1	two-component system response regulator	regulatory	TTCACCGCAGGGAAAC	-106	1	0.61	0.8
2241	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00828.1	FAD-binding protein	biosynthetic-additional	TAATCCGAAATTTTCA	-239	0.9	0.38	0.64
2576	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00829.1	two-component system response regulator	regulatory	CGAACCGGGTTGAAAA	-160	1	0.27	0.64
3095	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00828.1	FAD-binding protein	biosynthetic-additional	GTTTCTTCAGAGATCG	-118	0.9	0.35	0.62
3436	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00829.1	two-component system response regulator	regulatory	TTTTCAGACATTCGAG	-184	1	0.25	0.62
3507	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00829.1	two-component system response regulator	regulatory	CTTAACTCCCTGCAAC	-101	1	0.24	0.62
4352	BGC0002028.4	frigocyclinone	PKS	Streptomyces griseus	QDG00829.1	two-component system response regulator	regulatory	GTTTTTCAGACATTCG	-182	1	0.23	0.61

Figure 2. Overview of the four sections of the COMMBAT output results page using option "Multiple BGCs versus One Matrix" on the MIBiG database. (**A**) Customization Panel for adjusting score thresholds and filtering options, with real-time updates of displayed results and optional highlighting of reference BGCs. Users can also highlight specific experimentally validated BGCs (red circles in the Cloud Plot) to gauge the relevance of the COMMBAT score. (**B**) Interactive Cloud Plot showing predicted TFBSs ranked by highest COMMBAT score per BGC. (**C**) Interactive Result Table listing predicted TFBSs with sortable and searchable metadata. (**D**) TFBS Mapping Viewer displaying the genetic organization of a selected BGC and the positions of predicted TFBSs.

Step II: Input parameter, position weight matrix selection

Depending on the selected application, users provide one PWM to scan multiple BGCs or multiple PWMs to interrogate a single BGC. COMMBAT includes a collection of 2278 PWMs sourced from four TF databases: COMMBAT DB (this work), RegPrecise [22], Logomotif [20], and Prodoric [21]. Users can select the PWM by TF name, TF family, bacterial family, or database source. Additionally, users may also upload or manually define custom PWMs from FASTA formatted sequences.

The integrated PWM databases differ in scope, curation strategy, and data origin. Prodoric is a manually curated resource focused on experimentally validated TFBSs, primarily from model bacteria such as *Bacillus* spp., *Escherichia coli*, and model pathogens. RegPrecise combines curated and computationally inferred PWMs, providing broad taxonomic coverage. LogoMotif contains experimentally validated TFBSs from *Streptomyces* species, which are particularly relevant for COMMBAT given their high BGC content. To complement these resources and avoid delays in database updates, we developed COMMBAT_DB, an up-to-date collection of PWMs derived from manually curated, experimentally validated TFBSs as well as high-confidence predicted motifs. Particular care was taken to construct sets of TFBSs with diverse sequences to limit redundancy, thereby reducing bias from overrepresented motifs and improving prediction sensitivity.

Step III-1a: data output, application "*Multiple BGCs versus One Matrix*" using MIBiG.

Results are presented through four interactive modules (Fig. 2): a Customization Panel (Fig. 2A), an Interactive Cloud Plot (Fig. 2B), an Interactive Result Table (Fig. 2C), and a TFBS Mapping Viewer (Fig. 2D).

The Customization Panel (Fig. 2A) enables dynamic filtering of results by (i) the COMMBAT score, (ii) gene functional categories, and (iii) limiting results to either the best TFBS per

Figure 3. Overview of the four sections of the COMMBAT output results page using option "Multiple BGCs versus One Matrix" on antiSMASH results. (**A**) Customization Panel with the filter for adjusting the Novelty score threshold in addition to the filter for the COMMBAT score threshold. (**B**) Interactive Cloud Plot. Visual representation where each circle corresponds to a predicted TFBS within a BGC. Predicted TFBSs are plotted according to the COMMBAT score (*x* axis) and BGCs are ranked according to their Novelty score (*y* axis). (**C**) Interactive Result Table. (**D**) TFBS Mapping Viewer with localization of the two TFBSs selected in the Interactive Result Table.

BGC or the best TFBS per gene. Because TFs exhibit highly variable binding specificities, with differing tolerance to sequence variation and often incomplete binding site information, universal score thresholds are not biologically meaningful [28]. Therefore, COMMBAT enables flexible adjustment of threshold parameters, allowing users to tailor sensitivity and specificity according to the TF and biological context. For optimal interpretation of results, users are encouraged to consider the known binding characteristics of their TF of interest, particularly its tolerance to sequence variability. Additionally, users may highlight experimentally validated TF–BGC interactions (reference BGCs in Fig. 2A and B) to gauge the relevance of the COMMBAT score. All filters update results in real time. Optionally, users can tailor the output of their results by clicking on the "Advanced settings" case, allowing them to access additional filtering parameters including the "TFBS position" (the region where to identify TFBSs), the "Target score," the "Interaction score," and the "PWM score." All filters update results in real time.

The Interactive Cloud Plot (Fig. 2B) ranks BGCs according to their highest COMMBAT score, facilitating rapid identification of candidate regulatory interactions. Results can be displayed as best hits per BGC or per gene to display only the most reliable predicted TFBS per BGC or per gene.

The Interactive Result Table (Fig. 2C). A sortable and searchable table listing TFBSs from BGCs according to the COMMBAT score. The table provides detailed metadata for each entry, including the BGC reference ID, the associated compound name(s), the BGC class, the producing organism, the gene locus and name (when available) linked to the predicted TFBS, the known or predicted gene function, the gene BGC functional category, the TFBS sequence, the TFBS position relative to the gene's start codon, and scoring metrics (Target, Interaction, and COMMBAT scores). Optional columns include PWM scores and predicted co-transcribed genes, defined as genes located between − 90 and + 170 nt relative to the upstream stop codon as described by Ribeiro Monteiro *et al.* [12].

The TFBS Mapping Viewer (Fig. 2D) displays the genetic organization of the selected BGC and highlights predicted TFBS positions within the BGC.

Step III-1b: data output, application "*Multiple BGCs versus One Matrix*" using antiSMASH results

COMMBAT also supports predictions on BGCs derived from antiSMASH, either selected from the database or retrieved from user-uploaded collection predictions (Fig. 3). antiSMASH "KnownClusterBlast" similarity values to known MIBiG BGCs [18] are converted into a "Novelty score" defined as 1 minus the similarity percentage, ranging from 0 (fully characterized, known BGC) to 1 (no similarity, cryptic BGC). Results are displayed through the COMMBAT interface described above, with additional visualization of a Novelty score to highlight potentially cryptic BGCs. The cloud plot can display TFBSs according to COMMBAT score while

Figure 4. Overview of the four sections of the COMMBAT output results page using option "One BGC versus Multiple Matrices." (**A**) Customization Panel allowing tagging of experimentally validated TFs as benchmarks to assess other candidate regulators (see selected PWMs 1, 2, 3 in the "Tag reference matrices" window). (**B**) Interactive Cloud Plot. PWMs are ranked in decreasing order of their highest TFBS COMMBAT score and the selected reference TFs/PWMs (1, 2, 3) are highlighted in red. (**C**) Interactive Result Table. In the example, the table only shows the TFBSs associated with the PWM selected in the Customization Panel. (**D**) TFBS Mapping Viewer with localization of the TFBS selected in the Interactive Result Table.

ranking BGCs by Novelty, or by ranking BGCs solely by their highest COMMBAT score (Fig. 3B). The combined use of both metrics facilitates identification of cryptic BGCs (high Novelty score) under predicted transcriptional control (high COMMBAT score).

Step III-2: data output, application "*One BGCs versus Multiple Matrices*"

In the "*One BGC versus Multiple Matrices*" application, COMMBAT predicts TFs potentially regulating a selected BGC. Results are visualized through the interactive modules described above, with TFBSs ranked according to their COMMBAT scores to highlight candidate regulators with the strongest predicted interactions (Fig. 4). Additionally, users may restrict outputs to the top hit per BGC or per gene and optionally highlight experimentally validated TF–BGC interactions (reference PWMS in Fig. 4A and B) to gauge the relevance of the COMMBAT score.

Future developments

Future development of COMMBAT will focus on expanding both the scope of its regulatory predictions and the usability of the web platform. PWM and BGC databases will be regularly updated, and support for gapped and variable-length motifs will be implemented to overcome current limitations to ungapped PWMs. Expanding motif diversity and TF coverage will improve predictions across broader bacterial taxa. Planned developments also include increased computational capacity for large-scale antiSMASH analyses and integration of rhizoSMASH, gutSMASH, and other upcoming BGC resources. We are also considering extending COMMBAT to archaeal, fungal and plant BGCs, However, this will require addressing key challenges, including differences in promoter architecture, more complex gene organization such as intron–exon structures, and the limited availability of well-characterized PWMs in nonbacterial systems.

Acknowledgements

Author contributions: Silvia Ribeiro Monteiro (Conceptualization [equal], Data curation [equal], Formal analysis [equal], Investigation [equal], Methodology [equal], Resources [equal], Software [equal], Validation [equal], Visualization [equal], Writing – original draft [equal], Writing – review & editing [equal]), Augustin Rigolet (Conceptualization [equal], Investigation [equal], Validation [equal], Writing – review & editing [equal]), Clément Jeunehomme (Investigation [equal], Validation [equal], Writing – review & editing [equal]), Julianne Gathot (Investigation [equal], Validation [equal], Writing – review & editing [equal]), Yasmine Kerdel (Investigation [equal], Validation [equal], Writing – review & editing [equal]), Matthias Henry (Investigation [equal], Validation [equal], Writing – review & editing [equal]), Hannah E Augustijn (Investigation [equal], Validation [equal], Writing – review & editing [equal]), Marnix H Medema (Investigation [equal], Validation [equal], Writing – review & editing [equal]), Gilles van Wezel (Investigation [equal], Validation [equal], Writing – review & editing [equal]), and Sébastien Rigali (Conceptualization [equal], Funding acquisition [equal], Project administration [equal], Supervision [equal], Writing – original draft [equal], Writing – review & editing [equal])

Conflict of interest

None declared.

Funding

The work was supported by le Fonds de la Recherche Scientifique by an FNRS aspirant Grant to S.R.M., an FNRS-PDR T.0195.23 Grant to S.R., FRIA grants to J.G. and C.J., an Advanced Grant 101055020-COMMUNITY from the European Research Council to G.P.v.W. and A.R., and by Starting Grant 948770-DECIPHER from the European Research Council to M.H.M.

Data availability

Source code for the COMMBAT webserver can be found at https://gitlab.uliege.be/Silvia.RibeiroMonteiro/commbat.git.

References

1. Ziemert N, Alanjary M, Weber T. The evolution of genome mining in microbes—a review. *Nat Prod Rep* 2016;33:988–1005. https://doi.org/10.1039/C6NP00025H
2. Gavriilidou A, Kautsar SA, Zaburannyi N *et al.* Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat Microbiol* 2022;7:726–35. https://doi.org/10.1038/s41564-022-01110-2
3. Rigali S, Anderssen S, Naômé A *et al.* Cracking the regulatory code of biosynthetic gene clusters as a strategy for natural product discovery. *Biochem Pharmacol* 2018;153:24–34. https://doi.org/10.1016/j.bcp.2018.01.007
4. Kolter R, van Wezel GP. Goodbye to brute force in antibiotic discovery? *Nat Microbiol* 2016;1:15020. https://doi.org/10.1038/nmicrobiol.2015.20
5. Augustijn HE, Roseboom AM, Medema MH *et al.* Harnessing regulatory networks in Actinobacteria for natural product discovery. *J Ind Microbiol Biotechnol* 2024;51:kuae011. https://doi.org/10.1093/jimb/kuae011
6. van Bergeijk DA, Terlouw BR, Medema MH *et al.* Ecology and genomics of Actinobacteria: new concepts for natural product discovery. *Nat Rev Micro* 2020;18:546–58. https://doi.org/10.1038/s41579-020-0379-y
7. Seyedsayamdost MR. High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. *Proc Natl Acad Sci USA* 2014;111:7266–71. https://doi.org/10.1073/pnas.1400019111
8. Medema MH, van Wezel GP. New solutions for antibiotic discovery: prioritizing microbial biosynthetic space using ecology and machine learning. *PLoS Biol* 2025;23:e3003058. https://doi.org/10.1371/journal.pbio.3003058
9. Augustijn HE, Reitz ZL, Zhang L *et al.* Genome mining based on transcriptional regulatory networks uncovers a novel locus involved in desferrioxamine biosynthesis. *PLoS Biol* 2025;23:e3003183. https://doi.org/10.1371/journal.pbio.3003183
10. Rigali S. When, where, and why specialised metabolites are produced: inferring function from expression control. *Essays Biochem* 2026;EBC20250024. https://doi.org/10.1042/EBC20250024
11. van der Heul HU, Bilyk BL, McDowall KJ *et al.* Regulation of antibiotic production in Actinobacteria: new perspectives from the post-genomic era. *Nat Prod Rep* 2018;35:575–604. https://doi.org/10.1039/C8NP00012C
12. Ribeiro Monteiro S, Kerdel Y, Gathot J *et al.* The transcriptional architecture of bacterial biosynthetic gene clusters. *J Nat Prod* 2025;88:1772–80. https://doi.org/10.1021/acs.jnatprod.5c00529
13. Ribeiro Monteiro S, Rigali S. Enhanced prediction of expression control in bacterial biosynthetic gene clusters via genomic and functional data integration. *Microb Genom* 2025;11:001512. https://doi.org/10.1099/mgen.0.001512
14. Medema MH, Kottmann R, Yilmaz P *et al.* Minimum information about a biosynthetic gene cluster. *Nat Chem Biol* 2015;11:625–31. https://doi.org/10.1038/nchembio.1890
15. Blin K, Shaw S, Medema MH *et al.* The antiSMASH database version 5. *Nucleic Acids Res* 2026;54:D522–6. https://doi.org/10.1093/nar/gkaf1210
16. Hiard S, Marée R, Colson S *et al.* PREDetector: a new tool to identify regulatory elements in bacterial genomes. *Biochem Biophys Res Commun* 2007;357:861–4. https://doi.org/10.1016/j.bbrc.2007.03.180
17. Hertz GZ, Stormo GD. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. *Bioinformatics* 1999;15:563–77. https://doi.org/10.1093/bioinformatics/15.7.563
18. Zdouc MM, Blin K, Louwen NLL *et al.* MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration. *Nucleic Acids Res* 2024;53:D678–90. https://doi.org/10.1093/nar/gkae1115
19. Blin K, Shaw S, Vader L *et al.* antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. *Nucleic Acids Res* 2025;53:W32–8. https://doi.org/10.1093/nar/gkaf334
20. Augustijn HE, Karapliafis D, Joosten KMM *et al.* LogoMotif: a comprehensive database of transcription factor binding site profiles in Actinobacteria. *J Mol Biol* 2024;436:168558. https://doi.org/10.1016/j.jmb.2024.168558
21. Dudek C-A, Jahn D. PRODORIC: state-of-the-art database of prokaryotic gene regulation. *Nucleic Acids Res* 2022;50:D295–302. https://doi.org/10.1093/nar/gkab1110
22. Novichkov PS, Kazakov AE, Ravcheev DA *et al.* RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in bacteria. *Bmc Genomics [Electronic Resource]* 2013;14:745. https://doi.org/10.1186/1471-2164-14-745

23. R Core Team, (2025). R: a language and environment for statistical computing. R Foundation for Statistical Computing. (7 April 2026, date last accessed).
24. Chang W, Cheng J, Allaire JJ *et al.* shiny: web application framework for R, *(7 April 2026, date last accessed)*.
25. Wickham H. *ggplot2*. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-24277-4
26. Wilkins D. gggenes: draw Gene Arrow Maps in 'ggplot2'. 2017. https://doi.org/10.32614/CRAN.package.gggenes
27. Sievert C. *Interactive Web-based Data Visualization with R, Plotly, and Shiny*. New York: Chapman and Hall/CRC, 2020. https://doi.org/10.1201/9780429447273
28. Rigali S, Nivelle R, Tocquin P. On the necessity and biological significance of threshold-free regulon prediction outputs. *Mol BioSyst* 2015;11:333–7. https://doi.org/10.1039/C4MB00485J

Received: March 11, 2026. **Revised:** April 7, 2026. **Accepted:** April 17, 2026