

nf-core/funcscan: Identifying natural products and functional gene clusters in ancient microbial DNA

**Jasmin Frangenberg¹, James A. Fellows Yates^{1,2},
Anan Ibrahim¹, Louisa Perelo³, Moritz E. Beber⁴
and Pierre Stallforth¹**

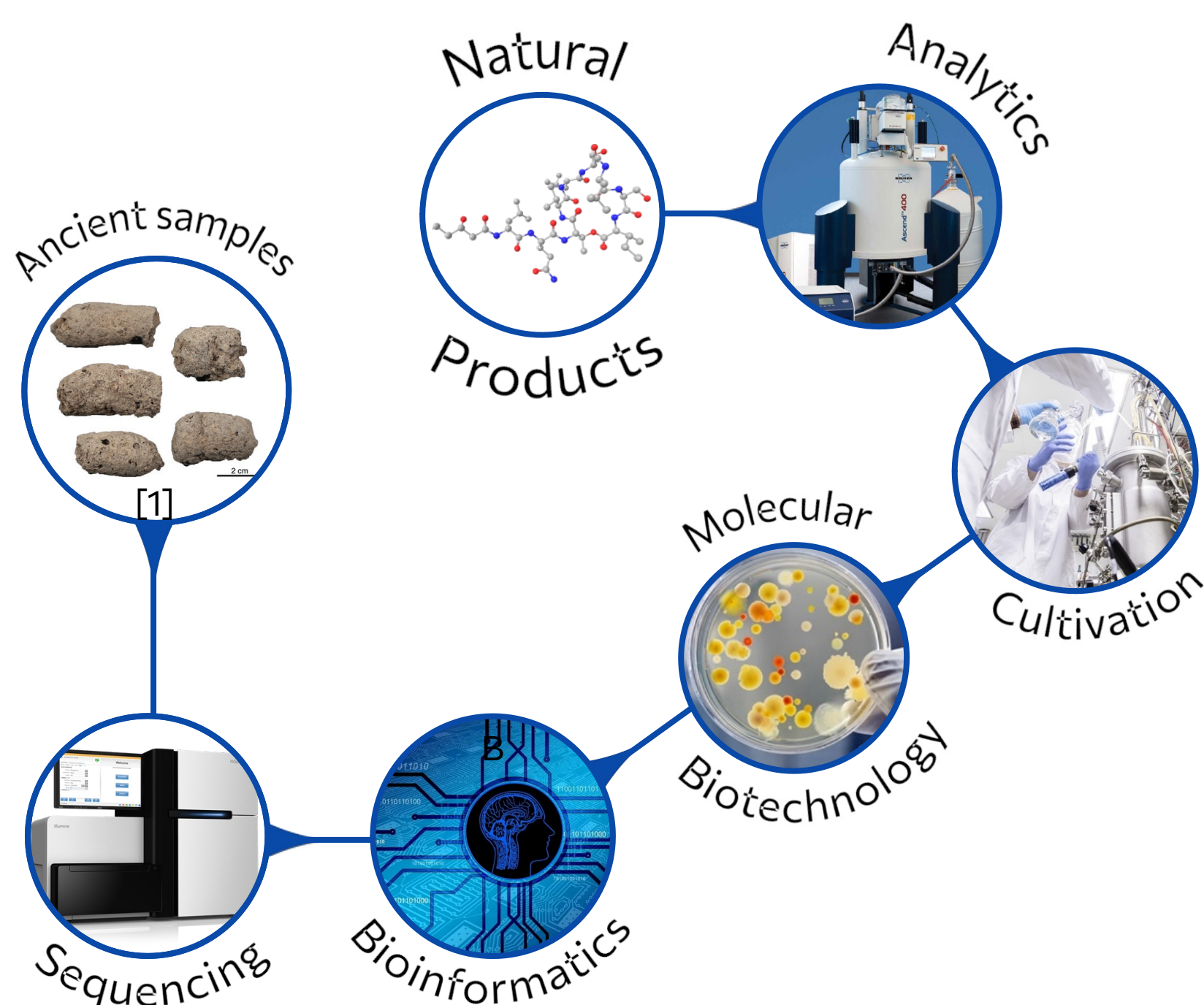
¹ Department of Paleobiotechnology, Leibniz Institute for Natural Product Research and Infection Biology – Hans Knöll Institute, Adolf-Reichwein-Straße 23, 07745 Jena, Germany

² Department of Archaeogenetics, Max Planck Institute for Evolutionary Anthropology, Deutscher Pl. 6, 4103 Leipzig, Germany

³ Quantitative Biology Center (QBiC), University of Tübingen, Auf der Morgenstelle 10, 72076 Tübingen, Germany

⁴ Unseen Bio ApS, Fruerbjergvej 3, 2100 Copenhagen, Denmark

Paleobiotechnology



Genome-mining of ancient DNA (aDNA) has the potential to be a powerful strategy to discover new bioactive natural products, along with their evolutionary and structural diversity. **Paleofeces**^[1] is rich in microbial aDNA, which allows to gain insight into the composition of the ancient human microbiome and unravels an **untapped diversity of "new" functional genes**. However, mining the sequences for different types of functional genes is rendered ineffective by the diversity in computational platforms, accessibility, scalability, and heterogeneous reporting of results of bioinformatic analysis tools.

The nf-core project^[2] is an initiative to establish **standardised, high-quality Nextflow pipelines** with active support from a wide community. nf-core pipelines are simple to install, executable across multiple platforms from personal laptops to HPCs, highly reproducible via the use of software containers, and offer both **user-friendly** graphical and command-line interfaces. Since 2018, an ever-growing number of contributors and collaborators ensures ongoing maintenance of all workflows and their modules.

Pipeline purpose

- nf-core/funccscan^[3] is capable of **identifying AMPs, ARGs and BGCs** in microbial genome sequences.
- Results of all tools of the 3 functional classes are aggregated and standardized.
→ Easy cross-comparison of results
- **Effective mining strategies** (e.g. deep learning or rule-based approaches)
→ identify functional genes of even low quality DNA (e.g. aDNA of paleofeces)

➡ **nf-core/funscan** elevates the level of genome mining scalability in all research fields. Here specifically, it supports the HKI PBT projects in potentially **resurrecting natural products with medical and economical importance.**

Applications

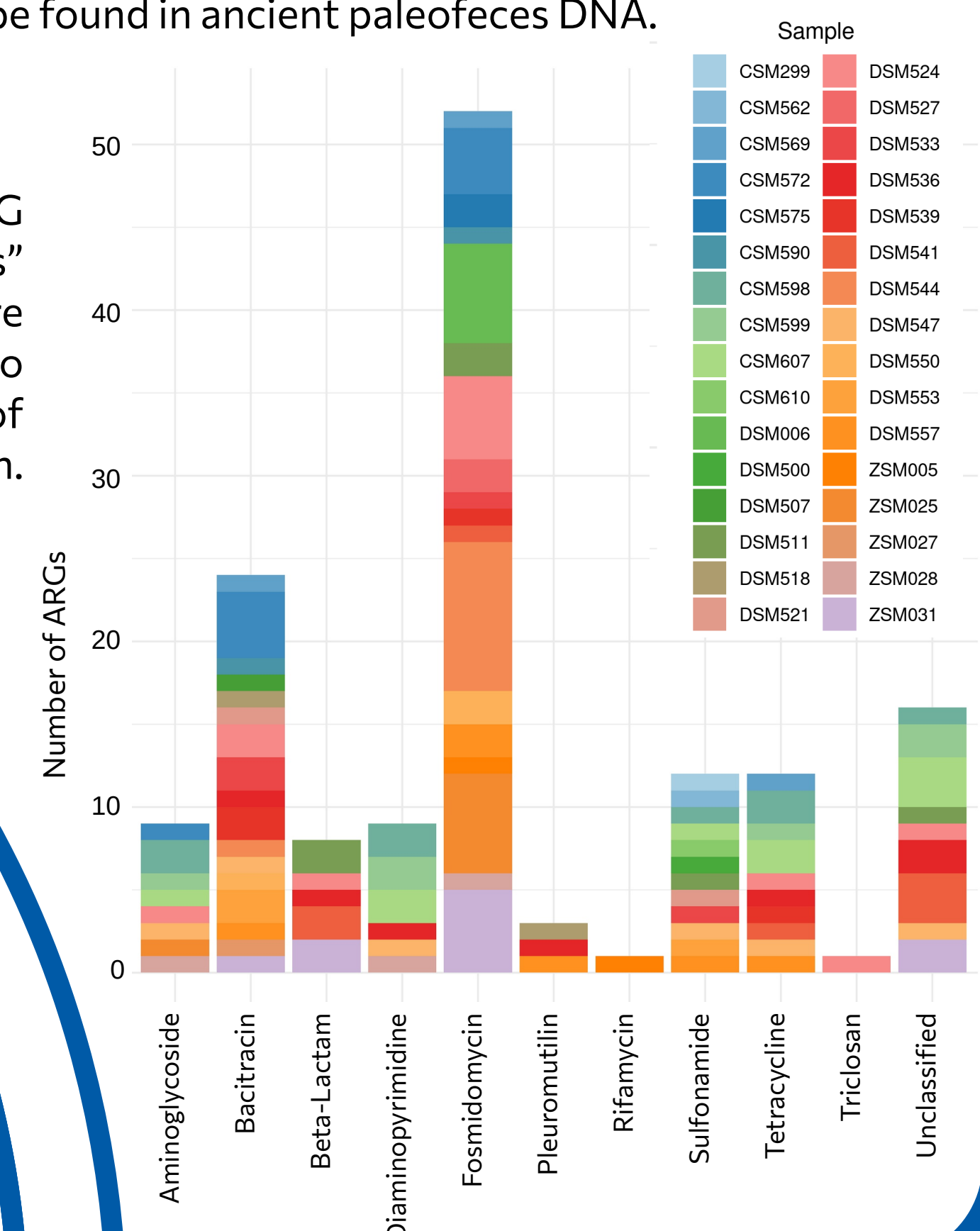
nf-core/funcscan is...

- **Easy to install:** only 2 dependencies
→ nextflow + docker/conda/singularity
- **Easy to use:**

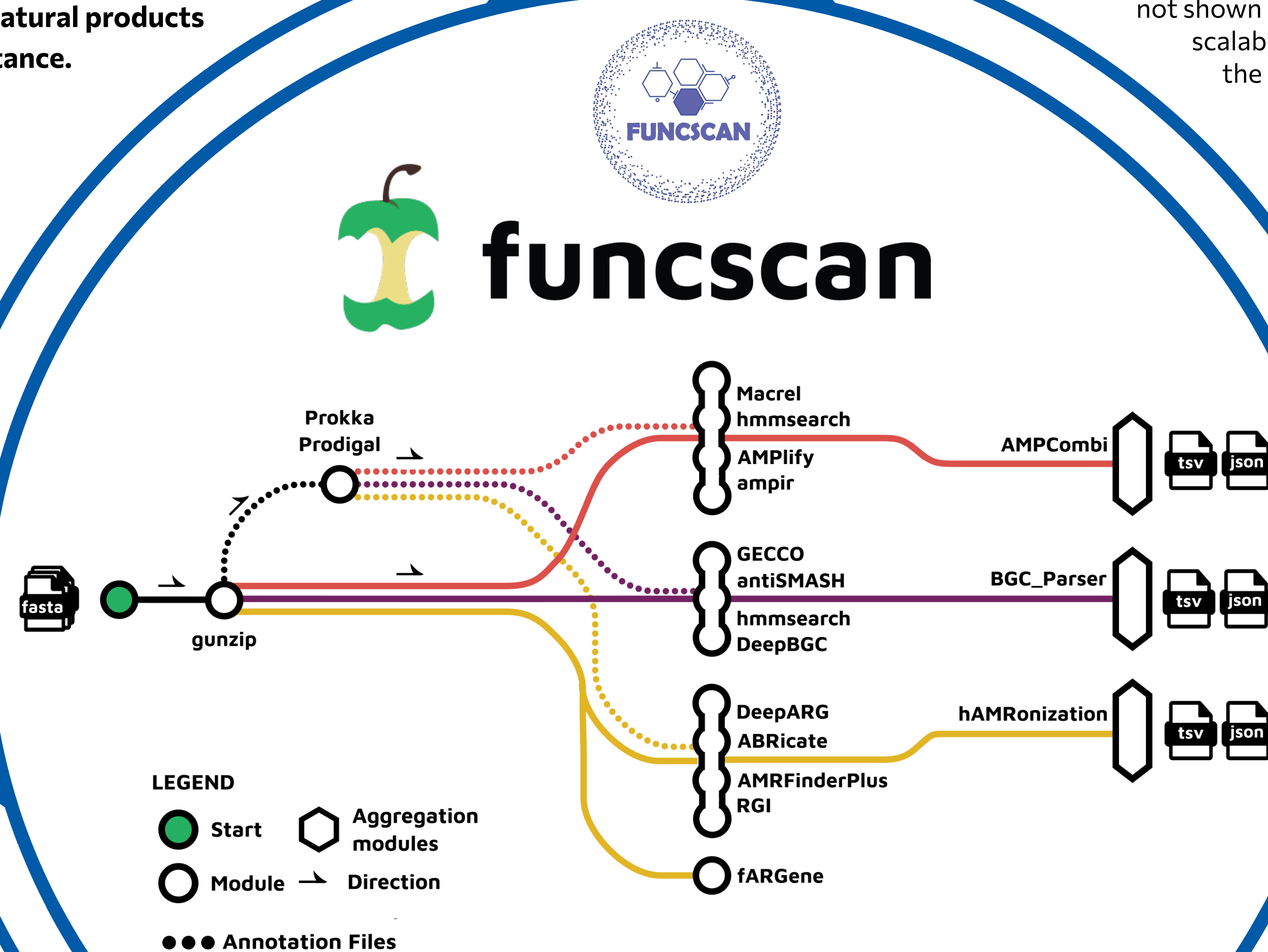
```
nextflow run nf-core/funcscan \
  -profile my_profile --outdir my_output
```
- Applicable to any microbial genome sequence
- Creating **sample-wise summaries**
- **Scalable** across different magnitudes of infrastructures
- Connectable to Nextflow Tower^[4] for elevated productivity and collaboration

Workflow:
Antibiotic
resistance genes
(ARGs)

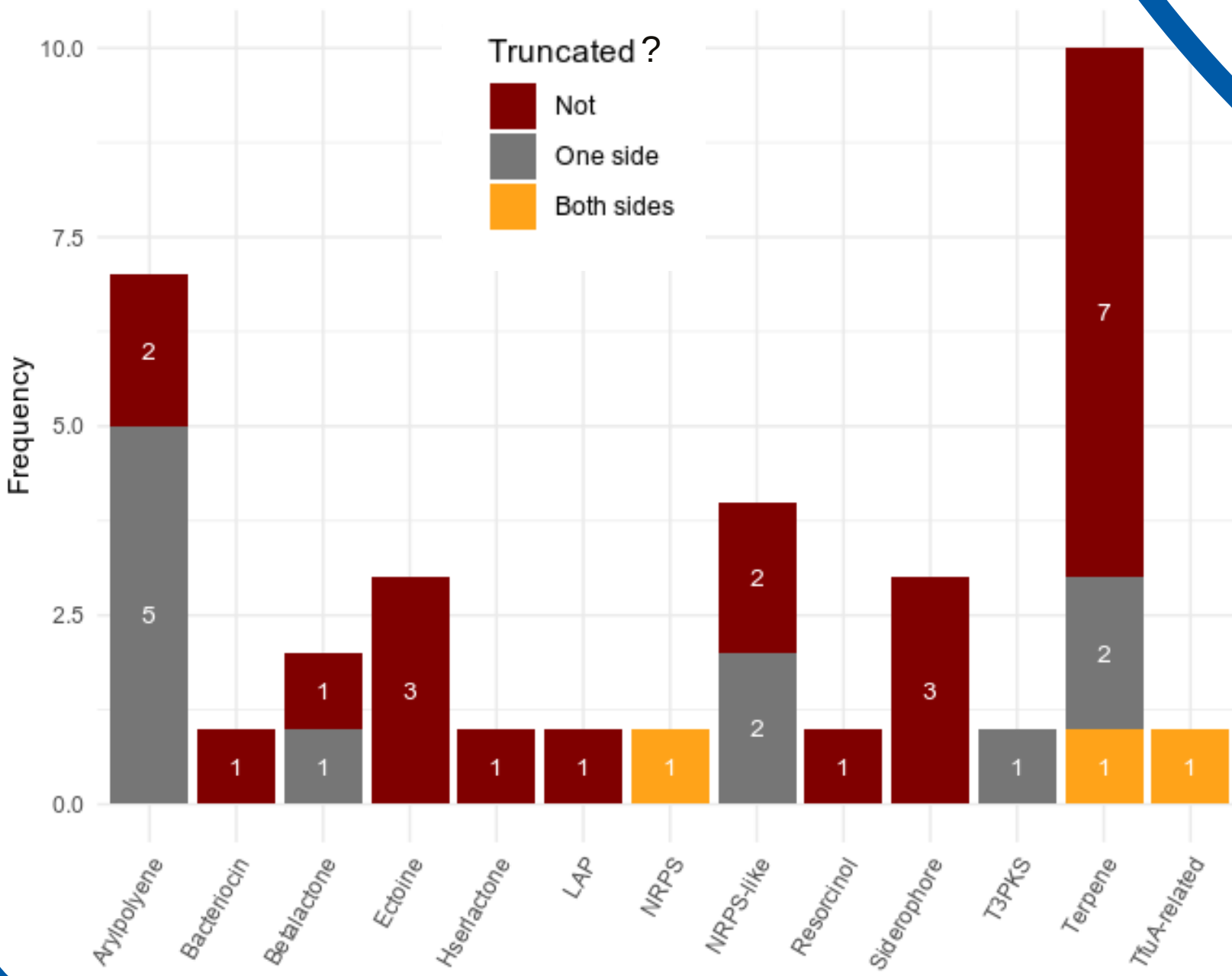
Many of the natural products of microorganisms show antimicrobial effects. Since the producing organisms need to protect themselves against these antimicrobials, they typically also harbor the corresponding ARGs. These small genes can abundantly be found in ancient paleofeces DNA.



The most abundant ARG classes are “Glycoproteins” and “Multidrug” which are not shown due to scalability of the graph.

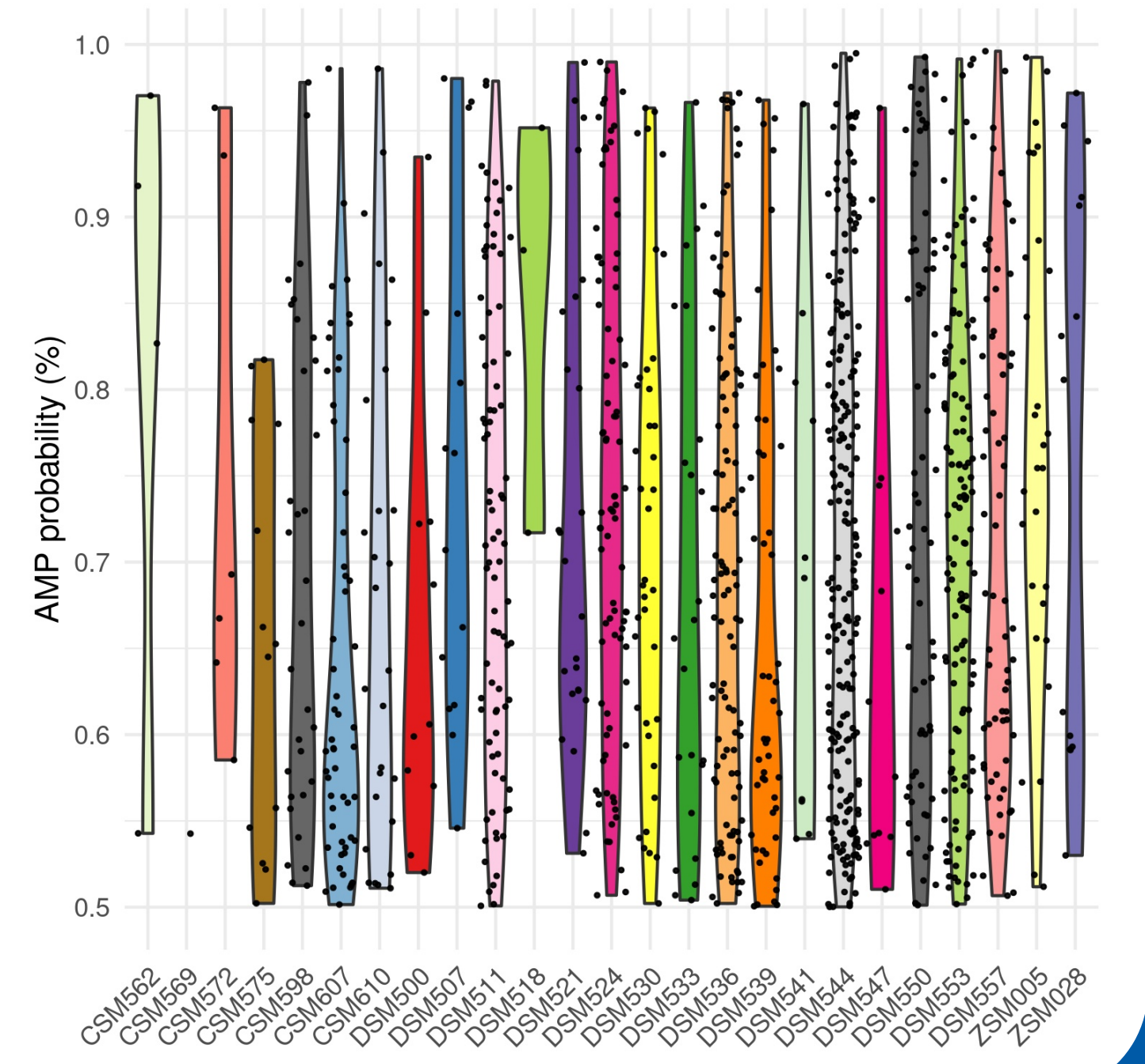


Workflow:
Biosynthetic
gene clusters
(BGCs)



BGCs are sets of physically close genes in a genome or on a plasmid which are involved in a single functional pathway. Since BGCs are typically very large, we often find truncated BGCs in the relatively short assembled contigs of paleofeces aDNA.

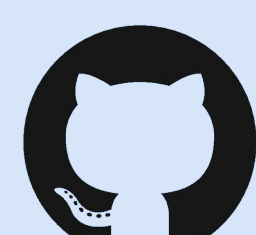
Workflow:
Antimicrobial
peptides
(AMPs)



AMPs are bioactive proteins, naturally produced by all living organisms, and representing the first line of defense against fungi, viruses and bacteria. Since these proteins are typically very small (15–100 amino acids), their genes can be found even in the fragmented aDNA from paleofecal samples.



jasmin.frangenberg@leibniz-hki.de



@jasmezz

Want to get involved?



nf-co.re/join

www.leibniz-hki.de

References

¹ Borry et al. (2020) *PeerJ*. doi: 10.7717/peerj.9001

² Philip Ewels et al. (2020) *Nat Biotechnol.* doi:10.1038/s41587-020-0439-x

³ <https://nf-co.re/funcscan>

⁴ <https://tower.nf>

WSS
WERNER SIEMENS-STIFTUNG

FRIEDRICH-SCHILLER-
UNIVERSITÄT
JENA

