Happy World Coding Day: Why Biologists Must Learn Coding and AI to Read the Hidden Language of Genomes

Oct 29, 2026 · @Hiba

SEO field	Value
SEO title	World Coding Day: Why Biologists Need Coding and AI
Meta description	Discover how coding, bioinformatics and AI help biologists analyze genomes, identify hidden genes and understand complex microbial data.
URL	/research-insights/world-coding-day-ai-genomics
Primary keyword	coding in biology
Secondary keywords	AI in genomics, coding for biologists, microbial genomics, bioinformatics, AI in microbiology, genome analysis, Python for biology, World Coding Day, computational biology

Happy World Coding Day!

Today, 29 October, the world marks World Coding Day. UNESCO Member States proclaimed the day in November 2023, and it was first celebrated in 2024. This year is its third edition.

The day reminds us that coding is no longer only for computer scientists. It is becoming an essential scientific language for microbiologists, geneticists, biotechnologists, bioinformaticians, microbiome researchers, medical researchers and agricultural scientists.

> A microscope helps us see microorganisms. Sequencing helps us read their DNA. Coding helps us understand what that DNA is telling us.

Modern biology has entered a data-rich era. Sequencing machines produce DNA faster and more cheaply than ever. Yet raw DNA is only a long string of A, T, G and C until computational tools help us organize, compare and interpret it.

The central idea of this article is simple: **one line of code can help us discover biological information that the human eye cannot see.**

1. Why biology now needs coding

Traditional microbiology was built on culture, microscopy, biochemical tests and phenotypic characterization. These methods remain essential. They tell us how a living organism actually behaves.

Sequencing, however, has changed the scale of the work. A single study may now involve:

- millions of sequencing reads
- thousands of genes and protein sequences
- hundreds of microbial genomes
- complex microbiome datasets
- antimicrobial resistance (AMR) genes, mutations and mobile genetic elements
- metabolic pathways and biosynthetic gene clusters

Genomics was already projected in 2015 to rival astronomy and online video as one of the largest generators of data on Earth. No researcher can analyze data at this scale by hand.

2. What is hidden inside a microbial genome?

A bacterial genome is a compact instruction manual. Depending on the species, it may carry genes linked to carbohydrate metabolism, acid and bile tolerance, stress response, adhesion, vitamin biosynthesis, transport systems and enzymes.

It may also carry genes of concern: antimicrobial resistance, virulence factors and mobile genetic elements that can move between bacteria. Some strains carry gene clusters for bacteriocins, the antimicrobial peptides that bacteria use against competitors.

None of these genes are visible under a microscope. They are hidden inside DNA. Coding and bioinformatics allow researchers to search genomes for them systematically, using curated resources such as NCBI, EMBL-EBI and specialist databases.

3. The human eye versus computational analysis

Each tool reveals a different layer of biology:

1. The human eye can see a colony.
2. A microscope can see a bacterial cell.
3. A sequencer can read its DNA.
4. Code can search the genome.
5. AI can help identify complex patterns.

A typical bacterial genome contains several million DNA letters. A scientist cannot read millions of letters, then compare thousands of genomes one by one. Computational methods make this possible in hours rather than lifetimes.

> We cannot examine millions of genomic signals with the human eye alone. We need computational tools to convert biological data into biological understanding.

4. What coding allows microbiologists to do

With basic coding skills, researchers can:

- clean raw sequencing data and assemble genomes
- identify and annotate genes
- compare genomes and calculate genomic similarity
- construct phylogenetic trees and analyze pangenomes
- detect AMR genes, virulence-associated genes and bacteriocin clusters
- study microbiomes and visualize biological data
- automate repetitive analysis and link genomic with phenotypic data

Three languages cover most of this work:

Language	What biologists use it for
Python	Processing genomic data, machine learning, automation and plotting. Libraries such as Biopython read and manipulate sequence files.
R	Statistics, biological data analysis, microbiome analysis and publication-quality graphics.

Language	What biologists use it for
Bash/Linux	Managing large sequencing files and running bioinformatics pipelines on servers. Most genomics tools run here.

Many widely used tools, such as Prokka for genome annotation and QIIME 2 for microbiome analysis, are driven from the command line. Even a few commands open the door to them.

5. Coding in probiotic genomics

Probiotic research shows clearly why coding matters. Before a candidate strain moves toward food or health use, researchers follow a logical path:

Strain identity → functional genes → bacteriocins → AMR → safety → comparative genomics → experimental validation

Coding and bioinformatics help answer questions such as:

- Is the strain correctly identified at species level?
- Does it carry genes associated with acid or bile tolerance?
- Are bacteriocin-related gene clusters present? Tools such as BAGEL4 and antiSMASH search for them.
- Are antimicrobial resistance genes present? Databases such as CARD support this search.
- Are any resistance genes located near mobile genetic elements, where they could transfer to other bacteria?
- How does the strain compare with related, well-studied probiotic strains?

Regulators already rely on this approach. The European Food Safety Authority (EFSA) expects whole-genome sequence analysis for microorganisms intentionally used in the food chain, including checks for acquired AMR genes.

Yet one principle must stay firm: **computational prediction does not replace laboratory validation.**

6. Why AI matters

When datasets grow very large, machine learning becomes a useful partner to classical bioinformatics. AI can potentially help researchers:

- recognize patterns across thousands of genomic features
- classify microbial strains
- predict selected biological characteristics
- prioritize candidate probiotic strains for testing
- analyze antimicrobial resistance patterns
- integrate genome, phenotype and microbiome data

However, AI models are only as good as their training data. Small, biased or poorly labelled datasets produce unreliable predictions, and models can fail on organisms unlike those they learned from.

AI can generate predictions and prioritize biological questions, but experiments are still required to confirm function and safety.

7. From one genome to millions of genes

A typical bacterial genome contains several thousand genes. *Escherichia coli*, for example, has roughly 4,000 to 5,000.

The numbers change dramatically when researchers study whole communities. A single catalogue of the human gut microbiome, published in *Nature Biotechnology* in 2021, combined 204,938 genomes and about 170 million protein sequences. Metagenomic studies of soil, oceans and animal guts produce datasets of similar scale.

At this level, computational biology is not optional. It is indispensable.

> The challenge of modern biology is no longer only generating data — it is understanding it.

8. Coding is becoming a laboratory skill

Future biologists should not see coding as separate from laboratory science. A modern project often moves in a loop:

Petri dish → DNA extraction → sequencer → computer → bioinformatics → AI → laboratory validation

The results of the lab inform the next round of analysis, and the analysis tells the lab which experiment to run next. Wet-lab and computational scientists increasingly need a shared vocabulary. A microbiologist who can read a script, and a programmer who understands a growth curve, make a far stronger team.

9. Why students should learn coding

Biology students do not need to become software engineers. They need enough skill to handle their own data with confidence. A practical starting path:

1. Linux basics: navigating folders and handling files
2. Simple Python
3. Basic R
4. FASTA and FASTQ file formats
5. Genome databases such as NCBI and EMBL-EBI
6. Basic statistics
7. Data visualization
8. Simple bioinformatics pipelines
9. Responsible use of AI, including checking every output

Start small. Rename a hundred files with one loop. Count the genes in one genome. Each small success builds the confidence for the next.

10. Coding, AI and the future of microbiology

Coding gives scientists a new way to interrogate biological systems. Its impact is growing in microbial genomics, probiotic discovery, antimicrobial resistance surveillance, precision microbiome science, synthetic biology, food biotechnology, animal health and agriculture.

It also supports the One Health approach, which links human, animal and environmental health. Tracking a resistance gene as it moves between farms, food and hospitals is, at its heart, a data problem.

> **A microscope reveals the microbial cell. Sequencing reveals its DNA. Coding reveals the patterns hidden within that DNA.**

Code Finds Patterns — Biology Provides the Proof

> - Finding a gene does not prove that it is expressed.
> - Predicting a function does not prove biological activity.
> - An AI prediction does not prove safety.
> - Genomic evidence must be interpreted together with phenotypic and experimental data.

Closing: Happy World Coding Day!

Coding is not replacing biology. It is giving biologists a new microscope — one designed not for cells, but for data.

The future of biological discovery will increasingly depend on researchers who can connect:

Microbiology + Genomics + Bioinformatics + Coding + AI + Experimental Validation

One line of code may not change the world by itself — but it may help us discover the gene, pathway or microbial function that does.

References

1. UNESCO. *World Coding Day, 29 October.* unesco.org/en/days/coding
2. Stephens ZD, Lee SY, Faghri F, et al. (2015). Big Data: Astronomical or Genomical? *PLoS Biology* 13(7): e1002195. doi:10.1371/journal.pbio.1002195
3. Sayers EW, Beck J, Bolton EE, et al. (2024). Database resources of the National Center for Biotechnology Information in 2024. *Nucleic Acids Research* 52(D1): D33–D43. doi:10.1093/nar/gkad1044
4. Seemann T (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30(14): 2068–2069. doi:10.1093/bioinformatics/btu153
5. Cock PJA, Antao T, Chang JT, et al. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* 25(11): 1422–1423. doi:10.1093/bioinformatics/btp163
6. Bolyen E, Rideout JR, Dillon MR, et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology* 37: 852–857. doi:10.1038/s41587-019-0209-9

7. Almeida A, Nayfach S, Boland M, et al. (2021). A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nature Biotechnology* 39: 105–114. doi:10.1038/s41587-020-0603-3
8. Alcock BP, Huynh W, Chalil R, et al. (2023). CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Research* 51(D1): D690–D699. doi:10.1093/nar/gkac920
9. van Heel AJ, de Jong A, Song C, et al. (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic Acids Research* 46(W1): W278–W281. doi:10.1093/nar/gky383
10. Blin K, Shaw S, Augustijn HE, et al. (2023). antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Research* 51(W1): W46–W50. doi:10.1093/nar/gkad344
11. Greener JG, Kandathil SM, Moffat L, Jones DT (2022). A guide to machine learning for biologists. *Nature Reviews Molecular Cell Biology* 23: 40–55. doi:10.1038/s41580-021-00407-0
12. EFSA (European Food Safety Authority) (2021). EFSA statement on the requirements for whole genome sequence analysis of microorganisms intentionally used in the food chain. *EFSA Journal* 19(7): e06506. doi:10.2903/j.efsa.2021.6506