capacity lost through the formation of isolated lithium during cycling can be recovered during calendar ageing. This is the opposite of what has typically been reported (irreversible lithium corrosion and loss of capacity) during calendar ageing of charged batteries[2–5].

Zhang and colleagues observed similar capacity recovery during calendar ageing in discharged coin cells when they tested a range of battery electrolytes and cycling conditions, both of which influence the efficiency of lithium-metal cycling. This demonstrated that the phenomenon is general. However, cycling measurements alone cannot prove that the capacity recovery was caused by the recovery of isolated lithium. The authors therefore used several other methods to investigate the capacity recovery, of which operando measurements — in which an optical microscope was used to observe lithium electrodeposition–electrodissolution cycles on a copper mesh electrode in a coin cell over time — provided most insight into the mechanism.

The optical time-lapse data showed that isolated lithium formed as early as the first cycle. However, when aged after discharge, the electrically insulating products of electrolyte decomposition surrounding the isolated lithium dissolved into the electrolyte (Fig. 1). This allowed the isolated lithium to become electrically reconnected to the electrode after subsequent charging.

The small coin cells used by Zhang *et al.* were helpful models for evaluating lithium-metal cycling, but they did not contain a battery cathode. When the authors added a cathode, they observed the same benefits of calendar ageing. It should be noted, however, that addition of a cathode creates a more complex system, the behaviour of which can be difficult to interpret.

The authors also studied capacity recovery in larger pouch cells, a type of battery characterized by soft packaging. These cells contained a cathode made of lithium iron phosphate ($LiFePO_4$, a frequently used lithium-ion cathode material) and a copper substrate. The cells lacked a lithium-metal anode, which both decreases the total mass of the active material in the cells compared with cells that contain lithium-metal anodes and increases the energy density.

During the first charging step of the pouch cells, lithium metal was electrodeposited onto the copper substrate using the cathode as the lithium source. Zhang *et al.* observed that, as in the smaller coin cells, the pouch cells showed signs of capacity recovery after ageing in the discharged state — although further characterization was not carried out to confirm that this was due to recovery of isolated lithium. Nevertheless, the finding suggests that isolated lithium could be recovered to mitigate capacity losses in the large-format, cathode-containing batteries that are likely to be developed for practical applications.

Although the study aids our understanding of calendar ageing and capacity loss in lithium-metal batteries, gaps in our knowledge still exist. For example, the authors implemented a carefully prescribed protocol that enabled recovery of isolated lithium, but it is too soon to say whether this can be translated to consumer applications in which the number of cycles and rest time will vary. The authors also investigated pouch cells that used a low-voltage cathode material, but high-voltage materials will be needed to make batteries that have high energy densities, and this might complicate or accelerate ageing mechanisms. Furthermore, this work did not study ageing in partly charged or partly discharged cells; this should be evaluated, given that ageing might occur in such states in practical applications. Nevertheless, Zhang *et al.* offer a crucial perspective: calendar ageing might not be detrimental to the performance of lithium-metal batteries, and could be used to improve it.

Laura C. Merrill is in the Nanoscale Sciences Department, Sandia National Laboratories, Albuquerque, New Mexico 87185, USA.
e-mail: lcmerri@sandia.gov

1. Zhang, W. *et al. Nature* **626**, 306–312 (2024).
2. Lin, D. *et al. Nature Chem.* **11**, 382–389 (2019).
3. Boyle, D. T. *et al. Nature Energy* **6**, 487–494 (2021).
4. Kolesnikov, A. *et al. Adv. Energy Mater.* **10**, 2000017 (2020).
5. Gao, Y. *et al. Nature Energy* **5**, 534–542 (2020).
6. Wood, S. M. *et al. Adv. Energy Mater.* **8**, 1801427 (2018).
7. Merrill, L. C., Rosenberg, S. G., Jungjohann, K. L. & Harrison, K. L. *ACS Appl. Energy Mater.* **4**, 7589–7598 (2021).
8. Harrison, K. L. *et al. ACS Nano* **11**, 11194–11205 (2017).

The author declares no competing interests.

Forum: Microbiology

The journey to understand previously unknown genes

The analysis of DNA sequences sheds light on microbial biology, but it is difficult to assess the function of genes that have little or no similarity to characterized genes. Here, scientists discuss this challenge from genomic and microbial perspectives. **See p.377**

The topic in brief

- **Some aspects of microbiology remain mysterious because of a lack of information about the identity and role of many microbial genes and proteins.**
- **The ability to obtain and analyse microbial sequences at scale and across species, including those that cannot be grown under laboratory conditions, are providing insights and data to explore.**
- **Writing in *Nature*, Rodríguez del Río *et al.*[1] report their analysis of 149,842 bacterial genomes sampled from a variety of habitats in the wild.**
- **The data were used to select sequences to generate a catalogue of 404,085 previously unknown gene families that could be prioritized for further study.**
- **The investigation of these previously unknown genes could lead to new clinical tools or offer fresh perspectives about how microorganisms evolved to survive in their natural environments.**

Jakob Wirbel & Ami S. Bhatt

Bringing structure and context to gene mysteries

The function of most microbial genes is unknown. Some of this microbial 'dark matter' might encode previously unknown types of enzyme or classes of antibiotic. As ever more genes of unknown function are discovered through sequencing of DNA from mixtures of multiple genomes, termed metagenomic sequencing, the difficulty of experimentally characterizing these enigmatic genes has led to a focus on computationally predicting their function[2]. Two publications in *Nature*, one on page 377 by Rodríguez del Río *et al.*[1], and one by Pavlopoulos *et al.*[3] published last October, tackle this challenge by cleverly leveraging advances in clustering algorithms (computational tools that group genes on the basis of similarities in amino-acid sequence) and protein-structure prediction tools[4] such as AlphaFold.

Despite distinct technical approaches,

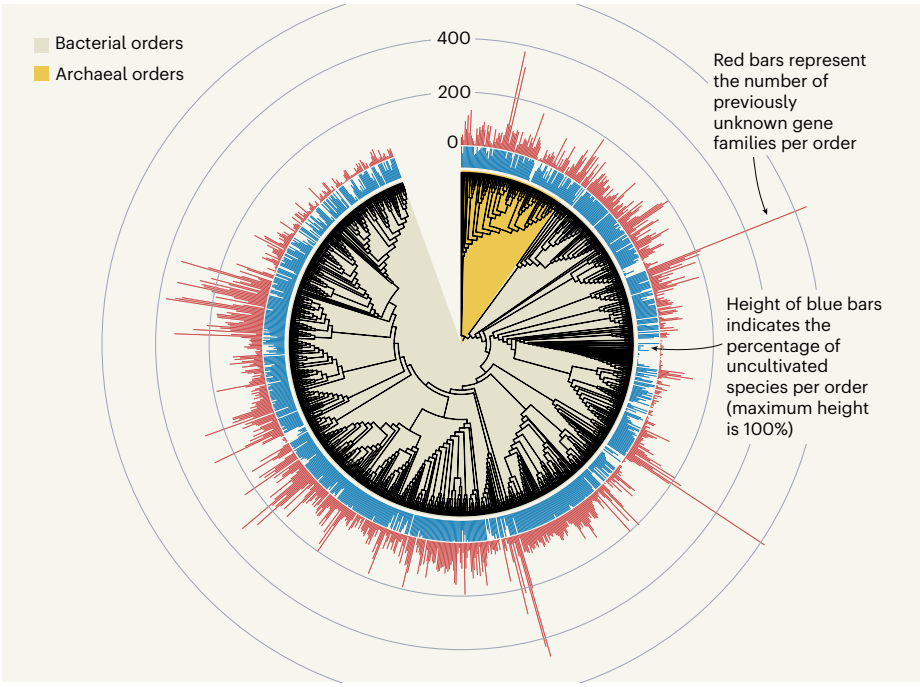

Figure 1 | Previously unknown microbial gene families. The large-scale analysis of DNA sequences captured from microbial samples as reported by Rodríguez del Río *et al.*[1] and by Pavlopoulos *et al.*[3] has revealed hundreds of thousands of previously unknown gene families. These data — which were gathered from microbes in the wild and across different habitats, and include species that have not been cultivated in the laboratory — provide a starting point for gaining insights into unexplored aspects of the biology of bacterial and archaeal microorganisms. Figure adapted from Fig. 3a of ref. 1.

the core strategy used by Pavlopoulos *et al.* and Rodríguez del Río *et al.* was similar. Both clustered hundreds of millions of protein sequences from metagenomic data sets into previously unknown protein families. Rodríguez del Río and colleagues filtered their data to examine genes only from prokaryotes (organisms whose cells lack a nucleus), whereas Pavlopoulos *et al.* used data that also included sequences from eukaryotes (organisms whose cells have a nucleus) and viruses.

With these catalogues of previously unknown families at hand, both teams set out to predict the function of their newly described families, capitalizing on genomic-context analysis, which involves examining adjacent genes for clues about function, as well as harnessing breakthroughs in methods to predict protein structures. In prokaryotic genomes, genes involved in the same pathway are often present close to one other. Genomic-context analysis, which proposes 'guilt by association', has been used effectively to predict previously unknown antiviral defence systems used by bacteria[5]. The second approach, comparing predicted protein structures to find similar (homologous) proteins, is more sensitive than simply comparing amino-acid sequences alone[6]. Both teams predicted structures for their protein families and compared them with databases of known structures, thereby generating informed predictions about the function of some of these enigmatic proteins.

The sheer scale and computational investment involved in these efforts, which yielded hundreds of thousands of newly discovered protein families (Fig. 1), is impressive. Yet, the number of previously unknown genes that have a functional prediction still remains relatively small. In both publications, only around 15% of the previously unknown protein families could be annotated on the basis of structural similarity; genomic-context analysis enabled functions to be proposed for 7.4% of families in Pavlopoulos *et al.* and 13% in Rodríguez del Río and co-workers. In addition, some assigned functional categories (such as 'ribosome') lack detailed specificity and this might obscure the precise role of these genes. Ultimately, the reliability of these predictions will have to be determined experimentally. Indeed, Rodríguez del Río *et al.* took the first step towards this objective by experimentally verifying the annotation for two of their predicted families.

> **"These two studies unlock a wealth of previously hidden knowledge."**

By delving deeper into the microbial dark matter, these two studies unlock a wealth of previously hidden knowledge, paving the way for future discoveries in diverse fields from medicine to biotechnology. Follow-up experiments might include the study of protein families with completely new protein folds, possibly revealing unexplored biological functions. Similarly, synapomorphic genes — corresponding to protein families that are specific to a group of organisms sharing a common ancestor but absent in others — might hold clues to key evolutionary processes. With further refinement and validation, these computational approaches offer a powerful tool for unlocking the functional secrets of the unseen microbial world.

Jakob Wirbel and **Ami S. Bhatt** are in the Department of Medicine, Division of Hematology, Stanford University School of Medicine, Stanford, California 94305, USA. **A.S.B.** is also in the Department of Genetics, Stanford University School of Medicine.
e-mail: asbhatt@stanford.edu

Alexander J. Probst
Microbial sequences reveal ecology and evolution

Genes are the ultimate source of all biological information on Earth, from human eye colour to the cell shape of microorganisms. The proteins they encode can be grouped using bioinformatics into families, usually with shared functionality. The ensemble of all known proteins in databases is continuously expanding as genomes are sequenced and the functions of the encoded proteins are predicted. The greatest fraction of biological functional diversity on our planet is attributed to microbial proteins. With the advent of sequencing of mixed microbial genomes from the environment (an approach that explores multiple genomes and is called metagenomics[7]), the increase in the rate at which data are being added to genome and protein databases is striking. However, the functional capacity of most protein families is unknown and part of the microbial dark matter.

Rodríguez del Río and colleagues' work, as well as the study by Pavlopoulos *et al.*, analysed large-scale metagenomic data and explored the potential function and distribution of unknown protein families, which might have evolutionary and ecological importance. Rodríguez del Río analysed nearly 150,000 microbial genomes (Fig. 1), and Pavlopoulos and colleagues investigated nearly 27,000 metagenomic data sets retrieved from diverse ecosystems with various bioinformatics approaches — going well beyond the scale of public-database entries used in previous such studies[8]. Surprisingly, a method called rarefaction analysis used by Pavlopoulos and colleagues revealed no slowing down in the detection of previously unknown protein

families as new metagenomes were added to their analysis. Instead, the detection of protein families increased exponentially, warranting an array of follow-on studies.

The distribution of protein families across Earth's categories of ecosystem (biomes) presented by Pavlopoulos and colleagues corroborates the findings of previous investigations regarding the distribution of microbial genes[8]. Some biological entities, however, were particularly rich sources of newly discovered protein families, including viruses, as Pavlopoulos *et al.* report, and microbes called Asgardarchaeota, as presented by Rodríguez del Río and colleagues. The latter are a group of microorganisms called archaea that are closely related to the first ancestor of eukaryotes. As such, studying their proteins might reveal new insights into the evolution of the eukaryotic cell[9].

One major challenge in exploring the wealth of previously unknown protein families encoded in genomes of natural samples is the identification of eukaryotic genes in metagenomes. Although certain algorithms exist for the recovery of eukaryotic genomes from metagenomes, accurately predicting eukaryotic genes in mixed DNA sequences — equivalent to Pavlopoulos and colleagues' method of identifying microbial genes — is still not possible bioinformatically. Once this shortcoming is overcome with the development of new algorithms, scientists will substantially expand the protein 'sequence space' and will identify protein families of unknown function that drive the ecology and evolution of eukaryotes.

The greatest advance in painstakingly organizing the protein families of nearly 27,000 metagenomes and across the tree of life lies in the identification of ecosystem-specific protein clusters that differ in terms of their presence or absence, or relative abundance between varying conditions of a given ecosystem — for example, between the contexts of health or disease. Applying this strategy to examine microbial data for healthy people and those with colorectal cancer, Rodríguez del Río and colleagues found that specific unknown protein families were enriched in the gut bacteria of people with cancer. These protein families were associated with microbial motility, adhesion and invasion potentially of human tissue, as revealed through genomic-context analysis. Harnessing this approach in other fields of research should be extremely helpful for deciphering the different functions of sample sets, in the hope of identifying new targets for biochemical analyses to shed light on a tiny fraction of the microbial dark matter.

Identifying differences in microbial communities (microbiomes) that might explain, for example, the disease state of a person, rely heavily on comparing which species are present and how abundant they are (the taxonomic composition), and examining genes that are associated with certain functions. Finding specific but differentially abundant protein families of unknown function, as demonstrated by Rodríguez del Río and co-workers, has the potential not only to replace current marker-gene-based approaches for differentiating microbiomes but also to advance microbiome research to a new and causality-driven level.

Alexander J. Probst is in the Research Center One Health Ruhr, University Alliance Ruhr, Department of Chemistry at University of Duisburg-Essen, Essen, 45141, Germany.
e-mail: alexander.probst@uni-due.de

1. Rodríguez del Río, A. *et al. Nature* **626**, 377–384 (2024).
2. Vanni, C. *et al. eLife* **11**, e67667 (2022).
3. Pavlopoulos, G. A. *et al. Nature* **622**, 594–602 (2023).
4. Jumper, J. *et al. Nature* **596**, 583–589 (2021).
5. Doron, S. *et al. Science* **359**, eaar4120 (2018).
6. Illergård, K., Ardell, D. H. & Elofsson, A. *Proteins* **77**, 499–508 (2009).
7. Tyson, G. W. *et al. Nature* **428**, 37–43 (2004).
8. Coelho, L. P. *et al. Nature* **601**, 252–256 (2022).
9. Eme, L. *et al. Nature* **618**, 992–999 (2023).

The authors declare no competing interests.
This article was published online on 30 January 2024.

Cancer

Natural inhibitor found for cell death by ferroptosis

Donna D. Zhang

The discovery that an evolutionarily conserved molecule used to make cholesterol also acts as a defence against a cell-death mechanism called ferroptosis might lead to new ways to treat cancer and other clinical conditions. **See p.401 & p.411**

Biology remains nothing short of astonishing, as researchers unveil the underpinnings of its myriad systems, especially those that are involved in protecting against cell death. On pages 401 and 411, respectively, Freitas *et al.*[1] and Li *et al.*[2] shed light on a regulated form of cell death called ferroptosis, which is driven by an iron-dependent modification of lipids in cellular membranes. The results bring into sharp focus an unexpected hero, the molecule 7-dehydrocholesterol (7-DHC).

The term ferroptosis was coined in 2012 (ref. 3). This cell death encompasses a variety of processes that include lipid oxidation by the action of reactive molecules called radicals (versions of molecules that have an unpaired electron) and the fragmentation of lipids at cellular membranes, culminating in membrane disruption, shrunken mitochondrial organelles and swelling of cells ('ballooning'). Ferroptosis occurs when there are problems in the regulation of normal iron levels (iron homeostasis) and in the oxidation of lipids. Preventing ferroptosis might be beneficial in alleviating neurodegenerative and kidney diseases, and activation of ferroptosis can kill cancer cells[4–7].

Freitas *et al.* and Li *et al.* report that 7-DHC, a molecule in the cholesterol-synthesis pathway (Fig. 1), acts to suppress ferroptosis. Both teams independently discovered the anti-ferroptotic role of the cholesterol-synthesis pathway. The authors reveal that several enzymes in this pathway function as potential suppressors of ferroptosis. However, one of the enzymes, DHCR7, which catalyses the reaction that converts 7-DHC to cholesterol, was found to promote ferroptosis. This indicates that 7-DHC, produced by the enzyme SC5D and used by DHCR7, operates as a key protection against ferroptosis.

Both teams then explored the mechanism of action of 7-DHC in more detail. They highlighted a key characteristic of its structure — a part that is described as a conjugated double bond in the sterol B-ring. This component of its structure enables 7-DHC to absorb radicals, thereby reducing lipid fragmentation driven by an oxidation process called peroxidation.

Both teams recognized that it is mainly lipid components called phospholipids — especially if these are fragmented into smaller pieces — that can initiate ferroptosis. These findings underscore the protective function of 7-DHC, especially in scenarios in which ferroptosis might otherwise occur if this molecule wasn't present.

Most remarkably, the two teams also found that the molecule ergosterol, which is found in yeast and fungi and has structural similarity to 7-DHC, also offers protection against ferroptosis. This finding suggests that the anti-ferroptotic effect of a molecule in the cholesterol-synthesis pathway might be evolutionarily conserved, serving to safeguard a variety of organisms from ferroptosis.