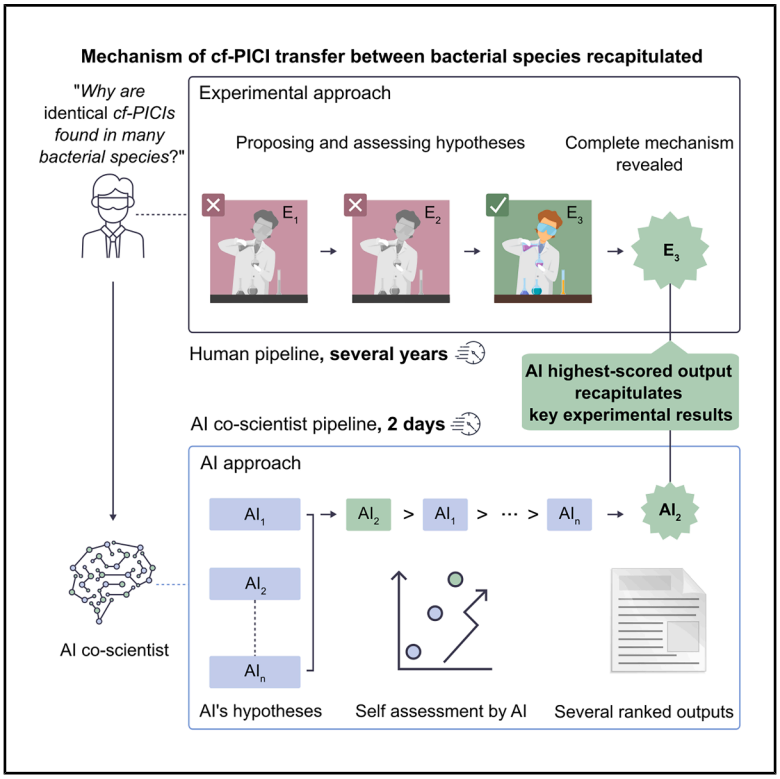
Cell

AI mirrors experimental science to uncover a mechanism of gene transfer crucial to bacterial evolution

Graphical abstract



Highlights

- AI co-scientist predicted a complex gene transfer mechanism before its publication
- Top AI-generated hypotheses opened new research directions
- AI bypassed human bias to propose overlooked biological possibilities
- Benchmarking showed AI co-scientist outperformed other LLMs on this task

Authors

José R. Penadés, Juraj Gottweis, Lingchen He, ..., Vivek Natarajan, Alan Karthikesalingam, Tiago R.D. Costa

Correspondence

j.penades@imperial.ac.uk (J.R.P.), natviv@google.com (V.N.), alankarthi@google.com (A.K.), t.costa@imperial.ac.uk (T.R.D.C.)

In brief

By solving a previously unsolved biological question, the AI co-scientist predicted a complex mechanism of gene transfer and generated hypotheses that opened new research directions, illustrating AI's potential as a creative engine in discovery.

Penadés et al., 2025, Cell *188*, 6654–6665
November 13, 2025 © 2025 The Authors. Published by Elsevier Inc.
https://doi.org/10.1016/j.cell.2025.08.018

CellPress

CellPress
OPEN ACCESS

Cell

Theory

AI mirrors experimental science to uncover a mechanism of gene transfer crucial to bacterial evolution

José R. Penadés,[1,2,3,4,7,8,*] Juraj Gottweis,[5,7] Lingchen He,[1,2] Jonasz B. Patkowski,[1,2,6] Alexander Daryin,[5] Wei-Hung Weng,[5] Tao Tu,[5] Anil Palepu,[5] Artiom Myaskovsky,[5] Annalisa Pawlosky,[5] Vivek Natarajan,[5,*] Alan Karthikesalingam,[5,*] and Tiago R.D. Costa[2,3,6,*]

[1]Department of Infectious Disease, Imperial College London, London SW7 2AZ, UK
[2]Centre for Bacterial Resistance Biology, Imperial College London, London SW7 2AZ, UK
[3]Fleming Initiative, Imperial College London, London W2 1NY, UK
[4]School of Health Sciences, Universidad CEU Cardenal Herrera, CEU Universities, Alfara del Patriarca, Valencia 46115, Spain
[5]Google Research, Mountain View, CA 94043, USA
[6]Department of Life Sciences, Imperial College London, London SW7 2AZ, UK
[7]These authors contributed equally
[8]Lead contact
*Correspondence: j.penades@imperial.ac.uk (J.R.P.), natviv@google.com (V.N.), alankarthi@google.com (A.K.), t.costa@imperial.ac.uk (T.R.D.C.)
https://doi.org/10.1016/j.cell.2025.08.018

SUMMARY

Artificial intelligence (AI) models have been proposed for hypothesis generation, but testing their ability to drive high-impact research is challenging since an AI-generated hypothesis can take decades to validate. Here, we challenge the ability of a recently developed large language model (LLM)-based platform, AI co-scientist, to generate high-level hypotheses by posing a question that took years to resolve experimentally but remained unpublished: how could capsid-forming phage-inducible chromosomal islands (cf-PICIs) spread across bacterial species? Remarkably, the AI co-scientist's top-ranked hypothesis matched our experimentally confirmed mechanism: cf-PICIs hijack diverse phage tails to expand their host range. We critically assess its five highest-ranked hypotheses, showing that some opened new research avenues in our laboratories. We benchmark its performance against other LLMs and outline best practices for integrating AI into scientific discovery. Our findings suggest that AI can act not just as a tool but as a creative engine, accelerating discovery and reshaping how we generate and test scientific hypotheses.

INTRODUCTION

Novel conversational artificial intelligence (AI) systems have tremendous potential to augment and accelerate biomedical discovery.[1–4] However, it remains uncertain whether AI systems can propose creative, novel, and impactful hypotheses that rival those of scientists and meet the rigorous standards for publication in reputed journals.

To explore this potential, we recently tested a novel AI system, named AI co-scientist,[5] on a series of unsolved questions in biology and biomedicine. While the AI-generated hypotheses were impressive, verifying them experimentally requires significant time and effort, as they represent new scientific areas needing multiple "wet lab" experiments. To test the system more efficiently, we challenged it with a specific unsolved question that had intrigued our groups for over a decade and whose answer was recently uncovered through extensive experimental work, yet not publicly disclosed. At the time of testing the AI co-scientist (last year; Figure 1), the experimental work addressing this question had just been submitted to *Cell* and was not publicly accessible, ensuring the AI could not draw on prior knowledge when tested. This allowed us to directly assess the AI's ability to generate plausible hypotheses by comparing its outputs to a newly known, unpublished, experimentally validated solution.

Our work focused on a unique family of phage satellites known as capsid-forming phage-inducible chromosomal islands (cf-PICIs).[6] Phage satellites are mobile genetic elements that rely on helper phages for their lifecycle, including the formation of infective particles. Several families are well characterized,[7] including P4-like elements (Gram-negative bacteria),[8] PICIs (in Gram-positive and Gram-negative bacteria),[9,10] PICI-like elements (PLEs; *Vibrio cholerae*),[11] and phage-inducible chromosomal minimalist islands (PICMIs; also in *Vibrio*).[12] These satellites depend entirely on helper phages to produce capsids and tails critical for packaging and dissemination.

 © 2025 The Authors. Published by Elsevier Inc.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

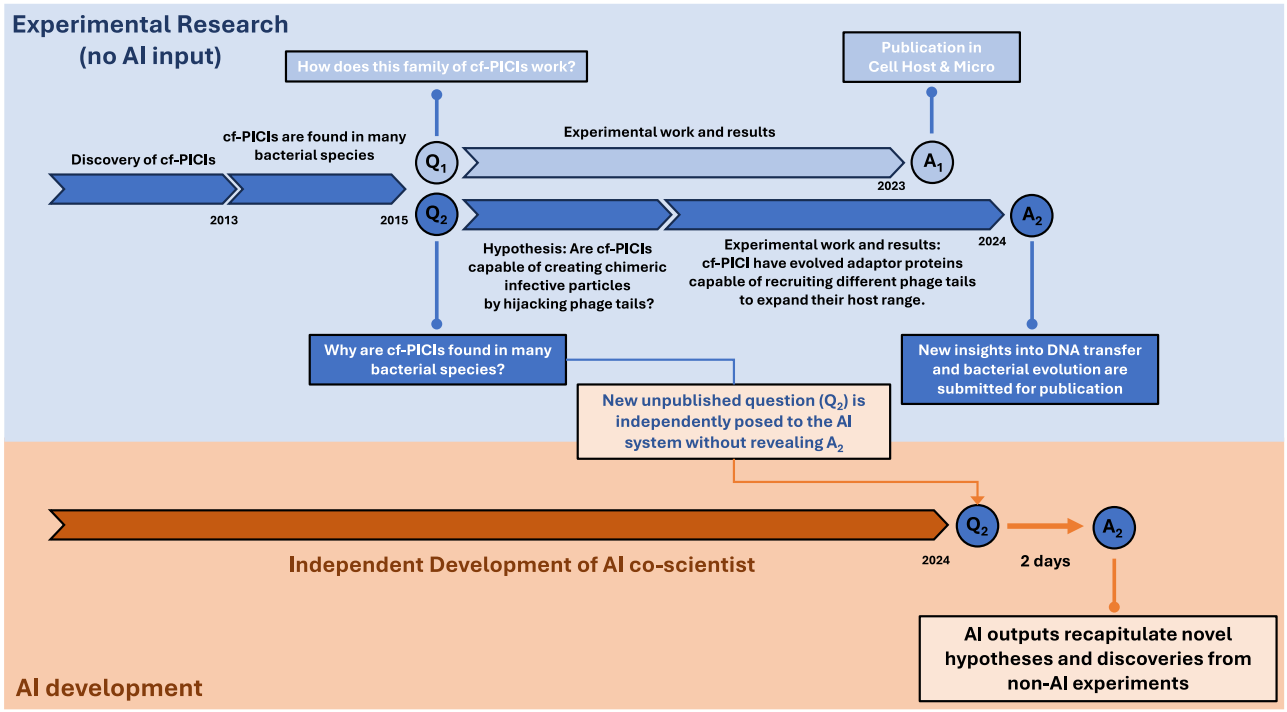


Figure 1. AI recapitulates the experimental discovery of a novel gene transfer mechanism

The blue section of the figure presents a flowchart outlining the timeline of the experimental research pipeline that led to the discovery of how cf-PICIs are mobilized between bacterial species. The orange section highlights the potential of AI to accelerate research by rapidly recapitulating, with no prior knowledge, the key experimental findings shown in blue.

Over a decade ago, we discovered a new family of satellites, cf-PICIs (Figure 1). Unlike classical PICIs and other satellites, cf-PICIs encode their own capsid-forming proteins, allowing them to produce small capsids independently of helper phages.[6] While these proteins are of phage origin, they have evolved to interact exclusively with cf-PICI components, exhibiting specificity that prevents interaction with phage-encoded proteins. This mechanism was published in 2023 after the concept of satellites as distinct genetic entities was established (Figure 1).[6]

Following the initial discovery of cf-PICIs, we observed an intriguing phenomenon: unlike other typically species-specific satellites, identical cf-PICIs were frequently detected across different bacterial species. This was further validated by methods developed to identify satellites in bacterial genomes.[13] Given the narrow host range of phages and other satellites, we hypothesized that cf-PICIs use an unprecedented mechanism of horizontal gene transfer to spread efficiently in nature. Over several years, we sought to uncover this mechanism and recently pieced together the solution.

With this knowledge, and confident the scientific community was unaware of this breakthrough, we challenged the AI co-scientist to hypothesize how identical cf-PICIs could exist in different bacterial species. Only one "input document" was provided (Data S1): a simple document comprising only previously published, openly available information as the system's research goal. We used this question to test the system's ability to answer scientific problems. Since the system was not specifically trained in phage or satellite biology, or bacterial evolution, we hypothesized this would provide an unbiased test of its ability to generate strong hypotheses. The results revealed an unexpected outcome. The AI co-scientist generated five ranked hypotheses (see results), with the top one recapitulating the key hypothesis and main experimental finding of our original manuscript submitted to *Cell* (Figure 1; Table 1). That manuscript has now been accepted for publication alongside this study,[14] providing essential context. This outcome demonstrates a fascinating and unsettling capability of AI systems: functioning at a level comparable to an experienced investigator, proposing hypotheses worthy of publication.

RESULTS

AI-driven research directions for cf-PICI host range expansion

Based on the background information and existing hypotheses, the AI co-scientist proposed five primary research directions to explain the broad host range of cf-PICIs, alongside a final point outlining potential areas of future interest. These were listed in order of preference and likelihood. To facilitate comparison, the system ranked the hypotheses using an Elo-score-based tournament framework,[15,16] similar to methods used in chess, where hypotheses compete iteratively and are scored based on their relative performance (see STAR Methods). Note that all hypotheses received a similar ranking, with only minor differences between them. A summary of the proposed hypotheses

Table 1. Summary of the AI's provided hypotheses

Elo Score[a]	Primary research directions[b]	Specific ideas[c]	Relevant to…[d]		
			Original question	Satellite biology	New areas of research[e]
1,777	capsid-tail interactions	conserved regions on capsids and tails	yes	yes	yes
		characterization of adaptor proteins	yes	yes	yes
1,701	integration mechanisms	–	no	yes	no
1,777	entry mechanisms	direct interaction with conserved bacterial receptors	yes	yes	yes
		role of bacterial OMVs	no	yes	no
		capsid interaction with bacterial membranes	no	yes	yes
1,701	helper phage and environmental factors	helper phage specificity	no	yes	no
		generalized transduction	no	yes	no
		influence of prophages	no	yes	no
		environmental stress and SOS response	no	yes	no
1,721	alternative transfer and stabilization mechanisms	transfer via conjugation	yes	yes	yes
1,545	unexpected areas of research and why to research them	cf-PICI interactions with other MGEs	no	yes	no
		role of quorum sensing in cf-PICI transfer	no	yes	no
		impact of biofilms on cf-PICI transfer	no	yes	no
		potential for cf-PICI transfer to eukaryotic cells	no	yes	yes
		evolutionary origins of cf-PICIs	no	yes	no

[a]The different hypotheses are ranked by an Elo score similar to that used to rank chess players.
[b]Main areas of research proposed by the AI co-scientist.
[c]For the areas highlighted in footnote b, the AI co-scientist provided specific research topics of interest.
[d]Some of the hypotheses directly address the original idea of the manuscript (why identical cf-PICIs can be present in different bacterial species; relevant to the original question), while others do not directly address that question but may provide hypotheses relevant to the biology of the satellites.
[e]The novelty of the hypotheses is evaluated by determining whether they are novel (yes) or whether these ideas have already been proposed in the field (no).
See also Data S1 and S2.

and their Elo scores is presented in Table 1, with more detailed information provided in Data S2. These correspond to:

(1) Capsid-tail interactions: investigate the interactions between cf-PICI capsids and a broad range of helper phage tails (ideas related to broad tail interacting, tail adaptor proteins, tail-binding sites, capsid-mediated interactions, etc).
(2) Integration mechanisms: examine the mechanisms by which cf-PICIs integrate into the genomes of diverse bacterial species (ideas related to integration, transposition, recombination, etc).
(3) Entry mechanisms: explore alternative cf-PICI entry mechanisms beyond traditional phage receptor recognition (ideas related to membrane vesicles, surface structures, membrane perturbation, etc).
(4) Helper phage and environmental factors: investigate the role of helper phages and broader ecological factors in cf-PICI transfer (ideas related to generalized transduction, prophages, community interactions, stress responses, etc).
(5) Alternative transfer and stabilization mechanisms: explore other potential transfer mechanisms, such as conjugation, extracellular vesicles (EVs), and unique stabilization strategies, that might contribute to cf-PICI's broad host range (ideas related to conjugative plasmids, EVs, membrane mimicry, unique stabilization strategies, etc).
(6) Unexpected areas of research and why to research them: beyond the core research directions outlined above, several unexpected areas might provide valuable insights into the broad host range of cf-PICIs.

Below, we will evaluate the accuracy of the different proposed hypotheses. However, our initial review of the AI co-scientist outcomes left a strong impression—not only because one of the hypotheses (the most likely) accurately summarizes the major finding of the experimental manuscript,[14] but also because the other hypotheses were plausible and, in some cases so compelling that they represent ideas that we plan to explore in the future.

Capsid-tail interactions

This hypothesis was ranked as the most promising, and to understand the magnitude of the outcome proposed by the AI system, it indeed summarized the experimental results presented in the accompanying *Cell* paper.[14] In that study, we showed that cf-PICIs form a new biological entity consisting of cf-PICI DNA

packaged into capsids encoded by cf-PICI genes. These capsids lack tails and cannot spread to new bacterial hosts. We found that tail-less capsids are produced even without helper phages, although their production increases in their presence. When cells carrying these particles lyse—through any mechanism disrupting cellular integrity—the tail-less cf-PICIs are released as inactive structures into the microbial community. There, they interact with tails from various phages infecting different species, forming functional and infective chimeric cf-PICI particles. Since the host range of phages and satellites is determined by their tails, which bind to bacterial surface structures,[17] this mechanism explains why identical cf-PICIs appear in different species. Depending on the tail used, cf-PICI DNA is delivered to different hosts.

Our experiments further demonstrated that some tail-less cf-PICI capsids bind to multiple tails from phages infecting different species. This ability is mediated by two cf-PICI-encoded proteins: the adaptor and connector.[14] By swapping these proteins between two cf-PICIs, we demonstrated that their expression determines tail-binding specificity and, consequently, host range.[14] Encouraged by these findings, we have begun characterizing these interactions using cryo-electron microscopy (cryo-EM).

The AI co-scientist's suggestions for studying capsid-tail interactions were particularly relevant and insightful (see Table 1 and "what to research in this area" in Data S2). Among its plausible, previously unexplored ideas, two stood out:

(1) Conserved regions on capsids and tails: use cryo-EM to identify conserved regions on cf-PICI capsids and phage tails that mediate their interactions. The AI system also recommended comparing these structures to pinpoint conserved contact points and using mutagenesis to test their role in binding and transfer.
(2) Characterization of adaptor proteins: investigate the role of cf-PICI-encoded adaptor proteins in mediating interactions with diverse phage tails. It is also proposed to investigate the diversity of the genes encoding these proteins in the cf-PICI genomes.

Notably, our experimental study also introduced concepts not identified by the AI that were needed to complete the model. For example, while helper phages typically induce satellites and provide packaging components, some phages induce cf-PICIs but fail to supply compatible tails, resulting in tail-less particles. In such cases, two helper phages were needed—one to induce, the other to complete packaging or tail provision. We also found that tail-less cf-PICI capsids can occasionally form without phage induction. Still, the central finding, that cf-PICIs exploit tails from different phages to broaden their host range, mediated by cf-PICI adaptor and connector proteins, was accurately proposed by the AI. Had we received this idea earlier, it might have significantly accelerated our research. Our progress was likely slowed by bias from existing satellite models (see discussion).

Integration mechanisms

Even though this hypothesis does not directly address the original question, it must still be considered due to its importance once cf-PICIs are present in different species.

This hypothesis can be evaluated from two perspectives. On the one hand, while integration is undoubtedly essential for the stable maintenance of transferred mobile genetic elements in the recipient cell, it does not address the fundamental question of how identical cf-PICIs are found in multiple bacterial species. For integration to occur, cf-PICIs must first be mobilized across species, and the mechanism behind this inter-species transfer is not addressed here.

On the other hand, if the first hypothesis (capsid-tail interaction) provides a plausible means of inter-species transfer, then integration becomes critical for persistence. Transfer without integration would likely result in loss of the element. For example, a family of PICIs and cf-PICIs was recently shown to integrate into late genes of resident prophages.[18] For such integration to occur, the recipient cell must harbor the appropriate prophage. If that prophage cannot spread across species, it would limit establishment (but probably not the transfer) of cf-PICIs in other species.

In our own genomic analysis (related to the cf-PICIs described in the accompanying *Cell* paper[14]), we found that the attachment (*attB*) sites targeted by cf-PICI integrases are highly conserved across diverse bacterial species. This likely facilitates integration following transfer. Notably, this information was not included in our original challenge to the AI system, making its hypothesis particularly interesting given its alignment with genomic evidence.

Entry mechanisms

This hypothesis was particularly intriguing and, in fact, ranked equally to the primary one. It explores alternative pathways that could explain cf-PICI mobilization between species. Crucially, all these ideas assume that the main hypothesis—tail-mediated dissemination—is incorrect. The AI system explicitly recommended testing whether tails are not involved in cf-PICI transfer (see topic 4 in the "entry mechanisms" section of Data S2). Based on that premise, it proposed alternative entry mechanisms, some of which are genuinely exciting.

Importantly, although our experiments suggest these are not the primary routes for cf-PICI spread, we cannot rule out the possibility that such mechanisms may occasionally occur. Below is a summary of the proposed ideas:

(1) Direct interaction with conserved bacterial receptors: the AI's first hypothesis, assuming tails were not involved, suggested testing whether tail-less cf-PICIs could directly interact with specific and conserved bacterial receptors. While there is no direct evidence supporting this idea, the presence of cf-PICI-encoded proteins with unknown functions made it worth considering. In fact, we investigated it in our experimental manuscript,[14] and included strains that were either able or unable to produce tails compatible with tail-less cf-PICI capsids. These controls ultimately confirmed that tails are essential for transfer, effectively ruling out this mechanism.
(2) Role of bacterial outer membrane vesicles (OMVs): the second idea was equally compelling, but less novel, since it has been proposed for the transfer of other satellites. Specifically, the Tycheposons, a satellite family found in *Prochlorococcus*,[19] have an unknown transfer

mechanism. It has been hypothesized that some of these elements are mobilized between microbial communities via vesicles, while others are transferred in viral particles.[19] This dual mode of transfer might explain their remarkable genetic diversity and the unusual range of sizes for these satellites (4–200 kb). Although we found no direct evidence for this mechanism in cf-PICIs, the hypothesis remains an exciting avenue for future research.

(3) Capsid interaction with bacterial membranes: the final hypothesis suggested by the AI co-scientist proposed that cf-PICI capsids could directly interact with and penetrate bacterial membranes, bypassing receptor-mediated entry mechanisms. This might involve membrane fusion or pore formation for entry. To investigate this, the AI recommended experiments using liposome model systems to study interactions between purified cf-PICI capsids and artificial membranes with varying lipid compositions. Techniques such as cryo-EM and fluorescence microscopy were suggested to visualize these interactions and assess for evidence of membrane fusion or pore formation (see topic 3 in the section "entry mechanisms" of Data S2).

In summary, taken out of context, these ideas might seem implausible. However, in light of the possibility that the primary hypothesis could be incorrect, some of these alternative hypotheses are highly creative and, more importantly, straightforward to test. This is crucial. The AI system generates insights that could significantly enhance our understanding of how satellites function. Furthermore, the initial testing of these ideas is relatively simple and grounded in a clear rationale. In our experience, the necessity of including controls to confirm the role of tails in cf-PICI transfer demonstrates the importance of these alternative mechanisms. Even though our findings did not support these hypotheses as primary routes, they remain valuable for exploring the broader context of satellite dissemination and inter-species gene transfer.

Helper phage and environmental factors

In this section, the AI co-scientist proposed several interesting ideas. However, these ideas were not entirely novel and, in some cases, have already been suggested for the transfer of other satellites. Moreover, many of these ideas are not specific to cf-PICIs and could be applied to other satellites. As a result, while they provide valuable insights into satellite transfer mechanisms, they do not directly address the unique question of why identical cf-PICIs, but not other satellites, are frequently found in different bacterial species. Nevertheless, the hypotheses are thought-provoking and offer avenues for further exploration.

(1) Helper phage specificity: the first hypothesis suggested investigating the range of helper phages capable of supporting cf-PICI transfer and whether helper phages with broad host ranges could explain the inter-species dissemination of these elements (Table 1; Data S2). This idea has already been validated for classical PICIs in *Staphylococcus aureus (SaPIs)*. Research, including our own, has shown that helper phages for classical PICIs can inject their DNA into a range of *Staphylococcus* species, such as *S. epidermidis*, *S. xylosus*, and *S. chromogenes*, and even into unrelated species like *Listeria monocytogenes*.[20–22] This broad host range could explain the presence of SaPI-like elements in different species. However, contrary to cf-PICIs, which are found as identical elements in multiple species, identical SaPI elements have not been observed across species. This suggests that while the mechanism may facilitate transfer, as occurs in the laboratory, these elements are not maintained in recipient cells, likely because they do not confer additional benefits. Integration stability and other factors may also play a role here, as previously hypothesized.

(2) Generalized transduction: the second idea relates to generalized transduction, where phages inadvertently package host DNA instead of their own.[23] This could theoretically enable cf-PICI transfer without requiring specific interactions between cf-PICI capsids and phage tails. This mechanism has already been demonstrated for *S. aureus* PICI SaPIbov2.[20] However, as with helper phage specificity, this hypothesis does not explain why cf-PICIs are uniquely found in multiple species while other satellites are not. Since cf-PICIs and classical PICIs share similar sizes and genomic loci, it remains unclear why other satellites would not also benefit from these mechanisms.

(3) Influence of prophages: another hypothesis suggests that prophages may play an important role after cf-PICI transfer by providing *attB* sites for integration or by acting as helper phages in the new species (see topic 3 in Data S2). While this idea is plausible and aligns with existing knowledge,[18] it does not address the critical question of how cf-PICIs initially transfer between species. Once transfer occurs, it is reasonable to assume that cf-PICIs might interact with new helper phages to promote further dissemination in the recipient species, and these interactions must be investigated, as we did in our experimental *Cell*-submitted paper. However, the mechanism enabling the initial transfer remains unexplained.

(4) Environmental stress and SOS response: the final idea, outlined in topic 4 (Data S2), proposed that environmental stressors, including those triggering the SOS response, could induce resident prophages and thereby facilitate satellite transfer. While this concept is broadly applicable to satellite mobilization, it does not specifically address the unique dissemination of cf-PICIs. This hypothesis is more general and represents a common mechanism of phage and satellite activation under stress conditions.

In summary, as these hypotheses become broader and less specific, they lose focus on the primary question that initiated this study: why identical cf-PICIs, but not other satellites, are found across different bacterial species. Despite this limitation, these general ideas remain valuable and merit consideration. They align with previously published concepts and provide a broader framework for understanding satellite transfer. Although they do not fully resolve the unique case of cf-PICI dissemination, they offer a foundation for further exploration and may guide future studies into the complex interplay between satellites, phages, and their environments.

Alternative transfer and stabilization mechanisms

A key advantage of AI systems is their ability to propose research directions that differ from those of human scientists. A compelling example is the first hypothesis in this section: topic 1 suggests exploring conjugation as a potential mechanism for satellite transfer, an exciting idea in the context of phage satellites.

Just as some satellites integrate into prophages, it is plausible they could also integrate into conjugative elements such as conjugative plasmids, integrative-conjugative elements (ICEs), or integrative mobilizable elements (IMEs), enabling mobilization via conjugation. Since conjugation is generally more promiscuous than phage-mediated transfer, this hypothesis is currently under investigation in our lab.

The AI system further proposed that cf-PICIs (and possibly other satellites) could hitchhike on conjugative elements. In addition to integrating, satellites could carry an *oriT* sequence, recognized by conjugative machinery to initiate transfer.[24] If so, they could hijack this machinery to move between species.

Supporting this idea, some SaPIs carry *cos* sequences, allowing them to hijack the packaging machinery of *cos* phages.[21,25,26] An identical strategy is used by the P4 satellite.[27] As *cos* sequences initiate phage-mediated packaging, *oriT* could similarly enable satellites to hijack conjugation. This innovative hypothesis, inspired by the AI system and not previously proposed by human researchers, is now being actively explored in our lab. In collaboration with Prof. Eduardo Rocha's group at the Pasteur Institute, we have identified potential satellite candidates with *oriT* sequences and are investigating their role.

The remaining hypotheses in this section were less impactful. Topic 2 revisited extracellular vesicles as a transfer mechanism, a concept previously proposed here and in other contexts. Topic 3, focusing on stabilization mechanisms, explored the role of the immune system in facilitating satellite entry into new cells but did not provide a strong link to inter-species transfer. Topic 4 discussed alternative replication strategies, which are less relevant given that cf-PICIs replicate like classical PICIs.[6,9] While stabilization (topic 3) and replication (topic 4) may influence post-transfer persistence, their roles in initial dissemination remain unclear.

In conclusion, the first hypothesis—conjugation as a mechanism for cf-PICI transfer—stands out as innovative and promising. The other ideas were less relevant and often repetitive. This case highlights the strength of AI systems in generating fresh, unbiased hypotheses and delivering a novel avenue we are now actively pursuing. It underscores the potential of AI not only to enhance our understanding of cf-PICI transfer but to reveal broader principles of satellite biology.

Unexpected areas of research and why to research them

The topics proposed in this section are relevant to satellite biology, including cf-PICIs, but do not directly address the original question. It appears the AI system aimed to broaden the research scope by highlighting areas beyond our specific focus.

The first proposed area relates to *cf-PICI interactions with other Mobile genetic elements (MGEs)*. This is a highly relevant and active area of research in satellite biology. We and others have investigated how satellites affect helper and non-helper phages,[28–31] as well as other MGEs like plasmids and satellites. For example, we showed that some PICIs are induced by other PICIs,[32,33] and that PICIs and phages influence the size of non-conjugative plasmids relying on transduction.[32,33] This area holds enormous potential and is currently being pursued by members of our team.

The second topic suggests examining the *role of quorum sensing in cf-PICI transfer*. Recent studies highlight quorum sensing's role in phage biology,[34] and extending this to satellites is logical, given their dependence on phages for induction and transfer. This hypothesis could open new avenues for understanding the regulatory mechanisms of satellite mobilization.

The third idea proposes analyzing the *impact of biofilms on cf-PICI transfer*. This area remains largely unexplored but has significant potential. Biofilms promote gene transfer[35] and often contain diverse species, offering a potential niche for inter-species satellite dissemination. Notably, SaPIbov2 encodes Bap,[36] a protein involved in biofilm formation,[37] supporting this concept's relevance.

The fourth suggestion, examining the *potential for cf-PICI transfer to eukaryotic cells*, seems less biologically relevant. No satellite sequences have been found in eukaryotes, making this idea speculative and less impactful for this study.

The final topic addresses the *evolutionary origins of cf-PICIs*, a classic question often raised in discussions about satellites. While this is undoubtedly an important and exciting area of research, it diverges from the original scope of this work.

Overall, the AI-generated ideas highlight diverse and exciting directions that could deepen our understanding of satellite biology. While not directly addressing cf-PICI inter-species transfer, they underscore the broader relevance of satellites and their interactions with other biological systems.

Analysis of other LLM systems

Having been impressed by the outcomes generated by the AI co-scientist system, we were intrigued to see what kind of hypotheses would be provided by other recently available large language and reasoning models as well as "deep research" agentic systems, which were challenged using the same document that was used to challenge the AI co-scientist (Data S1). These include OpenAI o1 (https://openai.com/index/openai-o1-system-card/),[38] Gemini 2.0 Pro Experimental (https://blog.google/feed/gemini-exp-1206/),[39] Gemini 2.0 Flash Thinking Experimental 12–19 (https://deepmind.google/technologies/gemini/flash-thinking/),[39] OpenAI deep research (https://openai.com/index/introducing-deep-research/), OpenAI o3-mini-high (https://openai.com/index/openai-o3-mini/),[40] Claude Sonnet 3.7 (https://www.anthropic.com/), and DeepSeek-R1 (https://huggingface.co/deepseek-ai/DeepSeek-R1).[41] Since a very specific question was posed in these analyses, we did not perform this competition to rank the different models but rather to observe the type of thinking they developed. It is worth noting that these systems were reasoning and using compute that spanned in the order of seconds and minutes and we took only one sample from the system. The co-scientist on the other hand, performs iterative self-improvement and reasoning that can span days to come up with its solution. Additionally, the AI co-scientist does use Gemini 2.0 models as specialized agents in the system. The AI co-scientist is a flexible system and can easily incorporate the other AI systems compared here as agents or tools, so they are not necessarily competing with each other. However, we thought it was still helpful to highlight the differences.

Table 2. Comparison of the different AI systems

Primary research directions[a,b]	Specific ideas[c]	o1	Gemini	Gemini 2.0	DR	o3	CS	D R1
Capsid-tail interactions	conserved regions on capsids and tails	no	no	no[d]	no	no	no[d]	no
	characterization of adaptor proteins	no	no	no	no	no	no	no
Integration mechanisms	–	yes	yes	yes	yes	no	yes	yes
Entry mechanisms	direct interaction with conserved bacterial receptors	no	yes	yes	no	no	no	yes
	role of bacterial OMVs	no	yes	yes	no	no	no	no
	capsid interaction with bacterial membranes	no	yes	yes	no	no	no	no
Helper phage and environmental factors	helper phage specificity	no	no	no	yes	yes	no	no
	generalized transduction	no	no	no	yes	no	no	yes
	influence of prophages	no	no	no	no	no	no	no
	environmental stress and SOS response	no	no	no	no	no	no	no
Alternative transfer and stabilization mechanisms	transfer via conjugation	no	no	no	no[e]	no	no	no
Unexpected areas of research and why to research them	cf-PICI interactions with other MGEs	no	yes	no	no	no	no	no
	role of quorum sensing in cf-PICI transfer	no	no	no	no	no	no	no
	impact of biofilms on cf-PICI transfer	no	no	no	no	no	no	no
	potential for cf-PICI transfer to eukaryotic cells	no	no	no	no	no	no	no
	evolutionary origins of cf-PICIs	no	no	no	no	no	no	no

[a]The table indicates whether the primary research directions highlighted by the AI co-scientist system were also identified by the other LLM reasoning and agentic deep research AI systems, including OpenAI o1 (o1), Gemini 2.0 Pro Experimental (Gemini), Gemini 2.0 Flash Thinking Experimental 12–19 (Gemini 2.0), OpenAI deep research (DR), OpenAI o3-mini-high (o3), Claude Sonnet 3.7 (CS), and Deepseek-R1 (D R1).
[b]Main areas of research proposed by the AI co-scientist system.
[c]For the areas highlighted in footnote b, the AI co-scientist system provided specific research topics of interest.
[d]The idea of cf-PICIs binding to different tails is mentioned, although without additional details.
[e]Transformation is tangentially mentioned.
See also Data S3 and S4.

The results, summarized in Table 2, indicate that, contrary to the AI co-scientist, none of the other AI systems were able to recapitulate the findings of the experimental manuscript. However, they proposed some ideas, many of which were also suggested by the AI co-scientist. Notably, the inability of other systems to reproduce the findings presented in the accompanying experimental manuscript serves as an internal control, confirming that the information related to that paper was not publicly available. Data S3 contains the outputs generated by the different AI systems, which are summarized below.

OpenAI o1

The first idea provided by this system, entitled "The Minimal Dependence on Helper Phage Tails Enables Broad 'Helper,'" seemed to identify the mechanism explaining the broader distribution of cf-PICIs between different species. However, the wording was a bit confusing, and after carefully analyzing it, the hypothesis proved to be incorrect. OpenAI o1 hypothesized that tail genes are more conserved than capsid genes, and therefore, the same tails can be used for many capsids, even for unrelated capsids. In fact, as summarized at the end of the report, OpenA o1 hypotheses that "because cf-PICIs provide their own head morphogenesis and genome-packaging modules yet borrow a tail module widely conserved among many phage 'helpers,' they can efficiently transfer between taxonomically distant bacteria. This unique autonomy of capsid and packaging—combined with broad phage-tail compatibility—drives their cross-species spread."

However, even though the idea of combining tails and capsids may seem correct, it is not in this case. Tails define the tropism, and if several phages use the same tails, they will have the same tropism, so their DNAs will be present in the same species. Tail genes are not more conserved than capsid genes, and each phage has a combination of these, which determines the compatibility of their capsids and tails. The correct answer is not the use of the same tail with different elements, but rather the opposite: the use of different tails, with different tropisms, by the same cf-PICI.

OpenAI o1 also mentioned mechanisms related to integration, which, as previously discussed, are necessary for transfer once the elements have arrived in the new species. It also suggested that the fact that cf-PICIs package their DNA independently of their helper phages may be important to overcome the species barrier, but the rationale behind this seems incorrect. Note that some satellites (cf-PICIs) encode their own packaging mechanisms,[14] which allow them to preferentially package their own DNA instead of the helper phage DNA.

Gemini 2.0 Pro Experimental

This model proposed two main ideas that were also suggested by the AI co-scientist system. The first relates to the potential ability of tail-less cf-PICIs to interact, via different mechanisms,

with bacterial receptors in a way that does not require tails. The second idea suggested exploring mechanisms of cf-PICI integration once the transfer had occurred. As previously mentioned, so far it seems clear that tails are required for cf-PICI transfer, while the mechanisms involved in the integration of the elements in the recipient cells, though not directly addressing the main question, are definitely relevant to the biology of the phage satellites.

Gemini 2.0 Flash Thinking Experimental 12–19

As with Gemini 2.0 Pro Experimental, the main hypothesis suggested by this Gemini Thinking model related to the idea that tail-less cf-PICIs can directly interact with different bacterial structures present in different species, explaining why these elements can be mobilized between species (hypothesis 1: relaxed host specificity of cf-PICI capsid receptor binding). Compared to the AI co-scientist, and in addition to outer membrane proteins, this system even included additional potential receptors, such as the lipopolysaccharide (LPS) core regions or peptidoglycan components.

This Gemini Thinking model also mentioned integration as a key factor to be analyzed in answering the question of inter-species cf-PICI transfer (hypothesis 3: post-entry survival and integration in diverse hosts). Related to this, the system also proposed ideas not previously suggested, such as analyzing the impact that the immune systems present in the recipient species may have on controlling cf-PICI transfer.

Gemini 2.0 Flash Thinking Experimental 12–19 also proposed an additional line of investigation (hypothesis 2: "universal" packaging and tail compatibility), which included two main ideas. One relates to the fact that cf-PICIs may be less dependent on host factors to package their DNA into their assembled capsids. There is no evidence of this. Importantly, in this aim, it was also suggested that "the interface between the cf-PICI capsid and the helper phage tail might be less stringent than in typical phage systems. This could allow cf-PICIs to utilize tails from a broader range of helper phages, even those from distantly related bacteria. This could be due to a simpler, more generic interaction mechanism at the capsid-tail junction." This definitely fits with the mechanism that explains how cf-PICIs spread inter-species. However, this result was less spectacular than that reported by the AI co-scientist because, overall, the report proposed that the mechanism of inter-species transfer depends on the capsid, not the tail. In fact, in the section *Novelty and Specificity of the Mechanism*, it is indicated that "this mechanism is novel because it highlights the capsid itself as the primary driver of broad host range for cf-PICIs, contrasting with the typical view where phage host range is primarily determined by tail fiber specificity and host factors." This suggests that the system assumes inter-species transfer occurs via capsid interaction and probably, intra-species transfer occurs via interaction with different tails. In any case, it would have been nice to know how we would have interpreted this in a scenario where we did not know the correct answer. It is tempting to speculate that even indirectly, maybe these ideas would have served to initiate a discussion that would eventually lead to the identification of the right answer.

OpenAI deep research

This system provided an extensive overview of the system, including general concepts relevant to the biology of the cf-PICIs and other satellites. The complete report can be found in Data S4. The system also provided a very clear hypothesis about how cf-PICIs move between species. This detailed pathway has been incorporated into Data S3.

In this report, the system proposed that the inter-species transfer of cf-PICIs depends on phages that can infect multiple bacterial species. The report states that "cf-PICIs across species likely exploit phages that serve as 'bridges' between different bacteria. Indeed, broad host range or generalized transducing phages could carry a cf-PICI from one species to another in a single transduction event. Even phages with a narrower range might facilitate stepwise transfer—for instance, a PICI could move from species A to closely related species B via one phage, and from species B to species C via another phage that infects B and C, and so on, eventually appearing in distantly related hosts."

This idea was also proposed by the AI co-scientist, and as mentioned before, it works in the laboratory with the classical PICIs from *S. aureus* (SaPIs) but does not seem to be relevant *in vivo*, since SaPIs are exclusive to *S. aureus*, with identical elements not being present in other species. OpenAI deep research does not recapitulate the main mechanism that explains inter-species cf-PICI transfer, although it does emphasize the relevance of the integration mechanism in the success of the transfer. Tangentially, the system also mentions the unlikely role that other mechanisms of gene transfer, such as transformation, may play in this process. There is no evidence supporting this idea.

OpenAI o3-mini-high

Contrary to other AI systems, OpenAI o3-mini-high provided only a hypothesis with a promising title: "autonomous capsid assembly coupled with promiscuous tail exploitation" (Data S3). However, upon closer examination, the system proposes that "many temperate phages have tail proteins that are more conserved and functionally promiscuous across diverse bacterial species." This idea was also suggested by other systems and, as previously mentioned, is incorrect. The answer is not that the same tail can be used to inject cf-PICI DNA into different species, but rather that the tail-less packaged cf-PICI DNA can interact with different tails, each one specifically infecting a single species.

Claude Sonnet 3.7 and Deepseek-R1

The analysis of the Claude Sonnet 3.7 and Deepseek-R1 systems was conducted during the review process, when the preprints related to this work[42] and the associated experimental data[14] were publicly available. As it was not possible to disable web search in these models, their outputs may have been influenced by the availability of our preprints. Regardless, neither system was able to recapitulate the key findings of our experimental study.

Claude Sonnet 3.7

To explain the widespread distribution of cf-PICIs, this system introduced a hypothesis that is difficult to interpret. Specifically, it stated that cf-PICIs "utilize tails from diverse phage families present across *Enterobacteriaceae*" (Data S3). The term "families" is ambiguous, potentially referring either to phages with different morphologies (e.g., Siphovirus, Myovirus, or Podovirus) or to Siphoviruses infecting different bacterial species. There is

currently no evidence that cf-PICIs can utilize tails from phages other than Siphoviruses, making the first interpretation unlikely. Regarding the second, the system proposed a set of experiments to test its hypotheses. However, these experiments focus on the idea that dissemination depends on tails recognizing different receptors, rather than the correct hypothesis that cf-PICIs hijack tails from phages infecting different species. The system, however, did suggest a potentially interesting idea: that defective prophages could serve as tail donors for cf-PICI elements. It also proposed that cf-PICIs might use conserved *attB* sites present across species, as proposed by other systems and confirmed in our experimental work.

In presenting its hypothesis, the system also introduced several biological inaccuracies regarding PICIs and cf-PICIs. For example, it claimed that PICIs depend on phage-encoded terminases for packaging. While this is true for some PICIs, others encode their own terminase[43] or redirect the phage's terminase specificity using dedicated proteins.[44] The system also incorrectly suggested that PICIs and cf-PICIs differ in their dependency on host factors for replication and packaging.[6] Furthermore, it stated that cf-PICIs possess a unique induction mechanism, whereas it has been shown that both PICI and cf-PICI elements are induced through a conserved mechanism.[6]

Deepseek-R1

This system proposed that cf-PICIs utilize broad host-range tails from different phage families that recognize widely conserved bacterial receptors. As an alternative hypothesis, which is less plausible, it suggested that cf-PICIs might encode adaptor proteins that modify phage tails to recognize different receptors. Both hypotheses are incorrect.

Deepseek-R1 also proposed generalized transduction as a possible mechanism for cf-PICI transfer between species. While theoretically possible, this mechanism is not specific to cf-PICIs and could apply to many types of phage satellites. Finally, Deepseek-R1 introduced biological inaccuracies, such as the claim that cf-PICIs require different elements for replication or integration compared to PICIs. As previously mentioned, the mechanisms by which PICIs and cf-PICIs are induced, replicate, or integrate are conserved.[6]

DISCUSSION

AI has demonstrated groundbreaking potential for enhancing biological research across the scientific pipeline, and our study offers a concrete example of how such systems can contribute to hypothesis generation and experimental planning. Specifically, we investigated how cf-PICIs disseminate and found that the AI co-scientist system not only proposed plausible mechanisms but independently recapitulated the very hypothesis we had experimentally validated: that cf-PICIs hijack phage tails from different species to achieve inter-species transfer. This alignment was not superficial. The system deduced that if tails determine tropism,[17] and cf-PICIs exist in multiple species, then cf-PICIs must interact with different tails. This insight was reached without prior access to our results or any specific biological fine-tuning, highlighting the power of large language models (LLMs) to generate testable, mechanistically sound hypotheses rooted in logical reasoning.

The co-scientist's hypothesis mirrored the path we had taken, but reached its conclusion without the delays imposed by human bias. For years, we had observed efficient induction of cf-PICIs without successful transfer and assumed that induction implied full helper phage dependence. We assumed that, whether following phage infection, or induction of a prophage or satellite, functional infectious particles were generated. We overlooked the possibility that cf-PICIs might be released as tail-less particles, a notion that, in hindsight, seems obvious. Our fixation on canonical paradigms delayed progress. In contrast, the AI system, unburdened by such assumptions, quickly suggested that tail exchange could underlie inter-species transfer. This contrast illustrates how LLMs may excel in problem spaces where established dogma limits human creativity.

Interestingly, this was not an isolated success. When we benchmarked the AI co-scientist against other state-of-the-art reasoning systems, we found that while several systems generated creative and novel hypotheses, only co-scientist recapitulated our experimental findings in detail. The AI co-scientist also opened a possibility that is currently under investigation in our labs, and is conjugation as a means of satellite mobility, an idea traditionally overlooked due to the assumption that satellites rely solely on phages. Some ideas appeared across multiple systems yet did not directly address the specific questions we posed. This underscores an important lesson: recurrence of a hypothesis across AI systems does not equate to scientific robustness. Rather, the merit of a hypothesis must still be evaluated through domain knowledge and experimentation.

Our experience also aligns with recent trends in the development of autonomous AI scientists. Prior systems such as Co--scientist,[45] designed for autonomous execution of chemical experiments, and more recent platforms like Virtual Lab,[3] PaperQA2,[46] HypoGeniC,[47] and the AI Scientist,[2] exemplify the growing ambition to integrate AI across the full research workflow, from literature analysis and hypothesis generation to experimental execution and manuscript writing. While these systems represent major progress, they often remain constrained to specific domains, retrospective analyses, or lack end-to-end experimental validation. In contrast, our application of the AI co-scientist demonstrated flexibility in tackling an open-ended biological problem and contributed novel, testable hypotheses that aligned with our empirical findings. Its scientist-in-the-loop design prioritizes cognitive collaboration over full automation, suggesting broader generalizability and a valuable role for AI as a true partner in scientific reasoning.

Nevertheless, this transformation comes with challenges. One immediate concern is how to prioritize and evaluate the growing volume of plausible hypotheses generated by AI systems. Traditional experimental pipelines are neither fast nor inexpensive enough to test every promising idea. Moreover, intellectual property frameworks may need to be reconsidered, especially when ideas emerge from generative systems rather than individual researchers. The attribution of credit and authorship also becomes murkier in collaborative environments where AI plays a substantial creative role.

Just as importantly, the growing power of AI systems raises the risk of over-reliance. While LLMs can generate novel and logical ideas, they lack the depth of contextual judgment that

comes from years of domain expertise. Without critical oversight, there is a risk of pursuing hypotheses that, although appealing, do not meaningfully advance knowledge. In our study, the Elo rankings of proposed hypotheses were similar. In this scenario, human expertise will be essential to recognize which ones merited further exploration and how to design the right experiments. As new data emerge during validation, these must be reintegrated into the AI's reasoning pipeline, fostering a continuous feedback loop between machine inference and human insight.

Ultimately, our results show that AI systems can do more than assist: they can catalyze paradigm shifts in research by identifying overlooked possibilities and accelerating hypothesis-driven discovery. The ability to generate mechanistic, testable, and biologically relevant hypotheses at scale could transform research in fields ranging from microbiology to drug discovery and personalized medicine. But this revolution will only be as valuable as the frameworks we build to support it. These include strategies for hypothesis triage, transparency in AI reasoning, co-authorship norms, and a rethinking of peer review and funding mechanisms in an AI-augmented research ecosystem. The future of science will depend not only on the power of AI but also on our ability to integrate it responsibly, critically, and creatively into the scientific process.

Limitations of the study

While our results highlight the potential of the AI co-scientist to contribute meaningfully to hypothesis generation and scientific reasoning, several limitations remain. First, our evaluation is based on a specific biological system, and while the system is complex and open-ended, further testing is needed to establish generalizability across other scientific domains. Second, while we report instances where AI-derived hypotheses aligned with or inspired novel experimental findings, the evaluation remains qualitative and limited in scale. A more systematic benchmarking framework is needed to assess the creativity, utility, and reproducibility of AI-generated scientific ideas. Finally, the system's performance depends heavily on the quality of underlying models and training data, which may embed biases or omit relevant domain-specific knowledge. Addressing these limitations will be critical to realizing the full potential of AI as a collaborative scientific partner.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact (j.penades@imperial.ac.uk).

Materials availability

This study did not generate new unique reagents.

Data and code availability

This study did not generate new unique code.

ACKNOWLEDGMENTS

We would like to acknowledge the contribution of Professor Darzi, Executive Chair of the Fleming Initiative, which convened this unique partnership. This work was supported by grants MR/X020223/1, MR/M003876/1, MR/V000772/1, and MR/S00940X/1 from the Medical Research Council (UK); BB/V002376/1 and BB/V009583/1 from the Biotechnology and Biological Sciences Research Council (BBSRC, UK); EP/X026671/1 from the Engineering and Physical Sciences Research Council (EPSRC, UK); and ERC-2023-SyG project 101118890—TalkingPhages to J.R.P. and 215164/Z/18/Z/WT to T.R.D.C.

AUTHOR CONTRIBUTIONS

J.R.P., J.G., V.N., A.K., and T.R.D.C. conceived the study; J.R.P., J.G., and T.R.D.C. conducted the experiments; J.R.P., J.G., L.H., J.B.P., A.D., W.-H. W., T.T., A. Palepu, A.M., A. Pawlosky, V.N., A.K., and T.R.D.C. analyzed the data; J.R.P. wrote the manuscript with inputs from all the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - AI System utilized for this challenge
  - Framing the challenge as a prompt to the AI system: minimal input for maximum insight
  - Comparison of the AI co-scientist with other AI models
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2025.08.018.

Received: February 23, 2025
Revised: June 17, 2025
Accepted: August 13, 2025
Published: September 9, 2025

REFERENCES

1. Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., Chandak, P., Liu, S., Van Katwyk, P.V., Deac, A., et al. (2023). Scientific discovery in the age of artificial intelligence. Nature *620*, 47–60. https://doi.org/10.1038/s41586-023-06221-2.
2. Lu, C., Lu, C., Lange, R.T., Foerster, J., Clune, J., and Ha, D. (2024). The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery. Preprint at arXiv. https://doi.org/10.48550/arxiv.2408.06292.
3. Swanson, K., Wu, W., Bulaong, N.L., Pak, J.E., and Zou, J. (2024). The Virtual Lab: AI Agents Design New SARS-CoV-2 Nanobodies with Experimental Validation. Preprint at bioRxiv. https://doi.org/10.1101/2024.11.11.623004.
4. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
5. Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Palepu, A., Sirkovic, P., Myaskovsky, A., Weissenberger, F., Rong, K., Tanno, R., et al. (2025). Towards an AI co-scientist. Preprint at arXiv. https://doi.org/10.48550/arXiv.2502.18864.
6. Alqurainy, N., Miguel-Romero, L., de Sousa, J.M., Chen, J., Rocha, E.P.C., Fillol-Salom, A., and Penadés, J.R. (2023). A widespread family of phage-inducible chromosomal islands only steals bacteriophage tails to spread in nature. Cell Host Microbe *31*, 69–82.e5. https://doi.org/10.1016/j.chom.2022.12.001.

7. Penadés, J.R., Seed, K.D., Chen, J., Bikard, D., and Rocha, E.P.C. (2025). Genetics, ecology and evolution of phage satellites. Nat. Rev. Microbiol. *23*, 410–422. https://doi.org/10.1038/s41579-025-01156-z.

8. Six, E.W., and Klug, C.A.C. (1973). Bacteriophage P4: a satellite virus depending on a helper such as prophage P2. Virology *51*, 327–344. https://doi.org/10.1016/0042-6822(73)90432-7.

9. Fillol-Salom, A., Martínez-Rubio, R., Abdulrahman, R.F., Chen, J., Davies, R., and Penadés, J.R. (2018). Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. ISME J. *12*, 2114–2128. https://doi.org/10.1038/s41396-018-0156-3.

10. Martínez-Rubio, R., Quiles-Puchalt, N., Martí, M., Humphrey, S., Ram, G., Smyth, D., Chen, J., Novick, R.P., and Penadés, J.R. (2017). Phage-inducible islands in the Gram-positive cocci. ISME J. *11*, 1029–1042. https://doi.org/10.1038/ismej.2016.163.

11. O'Hara, B.J., Barth, Z.K., McKitterick, A.C., and Seed, K.D. (2017). A highly specific phage defense system is a conserved feature of the Vibrio cholerae mobilome. PLoS Genet. *13*, e1006838. https://doi.org/10.1371/journal.pgen.1006838.

12. Barcia-Cruz, R., Goudenège, D., Moura de Sousa, J.A., Piel, D., Marbouty, M., Rocha, E.P.C., and Le Roux, F. (2024). Phage-inducible chromosomal minimalist islands (PICMIs), a novel family of small marine satellites of virulent phages. Nat. Commun. *15*, 664. https://doi.org/10.1038/s41467-024-44965-1.

13. de Sousa, J.A.M., Fillol-Salom, A., Penadés, J.R., and Rocha, E.P.C. (2023). Identification and characterization of thousands of bacteriophage satellites across bacteria. Nucleic Acids Res. *51*, 2759–2777. https://doi.org/10.1093/nar/gkad123.

14. He, L., Patkowski, J.B., Miguel-Romero, L., Aylett, C.H.S., Fillol-Salom, A., Costa, T.R.D., and Penadés, J.R. (2025). Chimeric infective particles expand species boundaries in phage inducible chromosomal island mobilization. Cell *188*, 6636–6653.e6. https://doi.org/10.1016/j.cell.2025.08.019.

15. Elo, A.E. (1986). The Rating of Chessplayers, Past and Present (ARCO PUBLISHING. INC.).

16. Coulom, R. (2007). Computing "ELO ratings" of move patterns in the game of go. ICGA J. *30*, 198–208. https://doi.org/10.3233/ICG-2007-30403.

17. Nobrega, F.L., Vlot, M., de de Jonge, P.A., Dreesens, L.L., Beaumont, H.J.E., Lavigne, R., Dutilh, B.E., and Brouns, S.J.J. (2018). Targeting mechanisms of tailed bacteriophages. Nat. Rev. Microbiol. *16*, 760–773. https://doi.org/10.1038/s41579-018-0070-8.

18. Tommasini, D., Mageeney, C.M., and Williams, K.P. (2023). Helper-embedded satellites from an integrase clade that repeatedly targets prophage late genes. Nar. Genom. Bioinform. *5*, lqad036. https://doi.org/10.1093/nargab/lqad036.

19. Hackl, T., Laurenceau, R., Ankenbrand, M.J., Bliem, C., Cariani, Z., Thomas, E., Dooley, K.D., Arellano, A.A., Hogle, S.L., Berube, P., et al. (2023). Novel integrative elements and genomic plasticity in ocean ecosystems. Cell *186*, 47–62.e16. https://doi.org/10.1016/j.cell.2022.12.006.

20. Maiques, E., Ubeda, C., Tormo, M.A., Ferrer, M.D., Lasa, I., Novick, R.P., and Penadés, J.R. (2007). Role of staphylococcal phage and SaPI integrase in intra- and interspecies SaPI transfer. J. Bacteriol. *189*, 5608–5616. https://doi.org/10.1128/JB.00619-07.

21. Chen, J., Carpena, N., Quiles-Puchalt, N., Ram, G., Novick, R.P., and Penadés, J.R. (2015). Intra- and inter-generic transfer of pathogenicity island-encoded virulence genes by cos phages. ISME J. *9*, 1260–1263. https://doi.org/10.1038/ismej.2014.187.

22. Chen, J., and Novick, R.P. (2009). Phage-Mediated Intergeneric Transfer of Toxin Genes. Science *323*, 139–141. https://doi.org/10.1126/science.1164783.

23. Zinder, N.D., and Lederberg, J. (1952). Genetic exchange in Salmonella. J. Bacteriol. *64*, 679–699. https://doi.org/10.1128/jb.64.5.679-699.1952.

24. Ares-Arroyo, M., Nucci, A., and Rocha, E.P.C. (2024). Expanding the diversity of origin of transfer-containing sequences in mobilizable plasmids. Nat. Microbiol. *9*, 3240–3253. https://doi.org/10.1038/s41564-024-01844-1.

25. Quiles-Puchalt, N., Carpena, N., Alonso, J.C., Novick, R.P., Marina, A., and Penadés, J.R. (2014). Staphylococcal pathogenicity island DNA packaging system involving cos-site packaging and phage-encoded HNH endonucleases. Proc. Natl. Acad. Sci. USA *111*, 6016–6021. https://doi.org/10.1073/pnas.1320538111.

26. Carpena, N., Manning, K.A., Dokland, T., Marina, A., and Penadés, J.R. (2016). Convergent evolution of pathogenicity islands in helper cos phage interference. Philos. Trans. R. Soc. Lond. B Biol. Sci. *371*, 20150505. https://doi.org/10.1098/rstb.2015.0505.

27. Ziermann, R., and Calendar, R. (1990). Characterization of the cos sites of bacteriophages P2 and P4. Gene *96*, 9–15. https://doi.org/10.1016/0378-1119(90)90334-n.

28. Rostøl, J.T., Quiles-Puchalt, N., Iturbe-Sanz, P., Lasa, Í., and Penadés, J.R. (2024). Bacteriophages avoid autoimmunity from cognate immune systems as an intrinsic part of their life cycles. Nat. Microbiol. *9*, 1312–1324. https://doi.org/10.1038/s41564-024-01661-6.

29. Chee, M.S.J., Serrano, E., Chiang, Y.N., Harling-Lee, J., Man, R., Bacigalupe, R., Fitzgerald, J.R., Penadés, J.R., and Chen, J. (2023). Dual pathogenicity island transfer by piggybacking lateral transduction. Cell *186*, 3414–3426.e16. https://doi.org/10.1016/j.cell.2023.07.001.

30. Fillol-Salom, A., Rostøl, J.T., Ojiogu, A.D., Chen, J., Douce, G., Humphrey, S., and Penadés, J.R. (2022). Bacteriophages benefit from mobilizing pathogenicity islands encoding immune systems against competitors. Cell *185*, 3248–3262.e20. https://doi.org/10.1016/j.cell.2022.07.014.

31. Frígols, B., Quiles-Puchalt, N., Mir-Sanchis, I., Donderis, J., Elena, S.F., Buckling, A., Novick, R.P., Marina, A., and Penadés, J.R. (2015). Virus Satellites Drive Viral Evolution and Ecology. PLoS Genet. *11*, e1005609. https://doi.org/10.1371/journal.pgen.1005609.

32. Haag, A.F., Podkowik, M., Ibarra-Chávez, R., del Gallego Del Sol, F.G., Ram, G., Chen, J., Marina, A., Novick, R.P., and Penadés, J.R. (2021). A regulatory cascade controls *Staphylococcus aureus* pathogenicity island activation. Nat. Microbiol. *6*, 1300–1308. https://doi.org/10.1038/s41564-021-00956-2.

33. Humphrey, S., San Millán, Á.S., Toll-Riera, M., Connolly, J., Flor-Duro, A., Chen, J., Ubeda, C., MacLean, R.C., and Penadés, J.R. (2021). Staphylococcal phages and pathogenicity islands drive plasmid evolution. Nat. Commun. *12*, 5845. https://doi.org/10.1038/s41467-021-26101-5.

34. Erez, Z., Steinberger-Levy, I., Shamir, M., Doron, S., Stokar-Avihail, A., Peleg, Y., Melamed, S., Leavitt, A., Savidor, A., Albeck, S., et al. (2017). Communication between viruses guides lysis-lysogeny decisions. Nature *541*, 488–493. https://doi.org/10.1038/nature21049.

35. Molin, S., and Tolker-Nielsen, T. (2003). Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure. Curr. Opin. Biotechnol. *14*, 255–261. https://doi.org/10.1016/s0958-1669(03)00036-3.

36. Ubeda, C., Tormo, M.A., Cucarella, C., Trotonda, P., Foster, T.J., Lasa, I., and Penadés, J.R. (2003). Sip, an integrase protein with excision, circularization and integration activities, defines a new family of mobile *Staphylococcus aureus* pathogenicity islands. Mol. Microbiol. *49*, 193–210. https://doi.org/10.1046/j.1365-2958.2003.03577.x.

37. Cucarella, C., Solano, C., Valle, J., Amorena, B., Lasa, I., and Penadés, J.R. (2001). Bap, a *Staphylococcus aureus* surface protein involved in biofilm formation. J. Bacteriol. *183*, 2888–2896. https://doi.org/10.1128/JB.183.9.2888-2896.2001.

38. OpenAI (2025). OpenAI o1 model: Reasoning and problem-solving advancements. Microsoft Azure OpenAI Documentation.

39. Team, G., Anil, R., Borgeaud, S., Alayrac, J.-B., Yu, J., Soricut, R., Schalkwyk, J., Dai, A.M., Hauth, A., Millican, K., et al. (2023). Gemini: A Family of Highly Capable Multimodal Models. Preprint at arXiv. https://doi.org/10.48550/arxiv.2312.11805.

40. OpenAI (2025). OpenAI o3 model wins gold at IOI: Advancements in AI reasoning. Geeky Gadgets.

41. DeepSeek, A.I., Guo, D., Yang, D., Zhang, H., Song, J., Zhang, R., Xu, R., Zhu, Q., Ma, S., Wang, P., et al. (2025). DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. Preprint at arXiv. https://doi.org/10.48550/arxiv.2501.12948.

42. Penadés, J.R., Gottweis, J., He, L., Patkowski, J.B., Shurick, A., Weng, W.-H., Tu, T., Palepu, A., Myaskovsky, A., Pawlosky, A., et al. (2025). AI mirrors experimental science to uncover a novel mechanism of gene transfer crucial to bacterial evolution. Preprint at bioRxiv. https://doi.org/10.1101/2025.02.19.639094.

43. Ubeda, C., Maiques, E., Tormo, M.A., Campoy, S., Lasa, I., Barbé, J., Novick, R.P., and Penadés, J.R. (2007). SaPI operon I is required for SaPI packaging and is controlled by LexA. Mol. Microbiol. *65*, 41–50. https://doi.org/10.1111/j.1365-2958.2007.05758.x.

44. Fillol-Salom, A., Bacarizo, J., Alqasmi, M., Ciges-Tomas, J.R., Martínez-Rubio, R., Roszak, A.W., Cogdell, R.J., Chen, J., Marina, A., and Penadés, J.R. (2019). Hijacking the Hijackers: Escherichia coli Pathogenicity Islands Redirect Helper Phage Packaging for Their Own Benefit. Mol. Cell *75*, 1020–1030.e4. https://doi.org/10.1016/j.molcel.2019.06.017.

45. Boiko, D.A., MacKnight, R., Kline, B., and Gomes, G. (2023). Autonomous chemical research with large language models. Nature *624*, 570–578. https://doi.org/10.1038/s41586-023-06792-0.

46. Skarlinski, M.D., Cox, S., Laurent, J.M., Braza, J.D., Hinks, M., Hammerling, M.J., Ponnapati, M., Rodriques, S.G., and White, A.D. (2024). Language agents achieve superhuman synthesis of scientific knowledge. Preprint at arXiv. https://doi.org/10.48550/arxiv.2409.13740.

47. Zhou, D., Schärli, N., Hou, L., Wei, J., Scales, N., Wang, X., Schuurmans, D., Cui, C., Bousquet, O., Le, Q., et al. (2022). Least-to-Most Prompting Enables Complex Reasoning in Large Language Models. Preprint at arXiv. https://doi.org/10.48550/arxiv.2205.10625.

48. DeepMind, G. (2024). Gemini AI: Multimodal artificial intelligence for text, images, audio, and video processing. https://deepmind.google/technologies/gemini.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
OpenAI o1	OpenAI [38]	https://openai.com/index/openai-o1-system-card/
Gemini 2.0 Pro Experimental	Team et al.[39]	https://blog.google/feed/gemini-exp-1206/
Gemini 2.0 Flash Thinking Experimental 12–19	Team et al.[39]	https://deepmind.google/technologies/gemini/flash-thinking
OpenAI deep research	N/A	https://openai.com/index/introducing-deep-research/
Claude Sonnet 3.7	N/A	https://www.anthropic.com/
OpenAI o3-mini-high	OpenAI[40]	https://openai.com/index/openai-o3-mini/
DeepSeek-R1	DeepSeek-AI et al.[41]	https://huggingface.co/deepseek-ai/DeepSeek-R1

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

The experimental model was computational: an AI co-scientist multi-agent system[5] and a set of comparator large language/reasoning models accessed via their public interfaces. No biological samples, cell lines or animal models were required or used.

METHOD DETAILS

AI System utilized for this challenge

As detailed elsewhere,[5] the AI system utilized for this challenge comprised a LLM-driven (large language model) novel hypothesis generation system, termed AI co-scientist. The AI co-scientist is a multi-agent system based on Gemini 2.0[39] with design echoing the reasoning processes underpinning the scientific method. The system is designed to act as an AI collaborator for scientists and promotes an expert-in-the-loop workflow. Scientists can specify research goals in natural language with additional preferences and constraints reflecting their topic of expertise or experimental setup. They can also provide their own best-guess solution to the research or other feedback to guide the system. The system parses the research goal into a research plan configuration and employs several Gemini 2.0-based agents that continuously generate, debate and improve hypothesis and research proposals. These agents are specialized for generating, reviewing, ranking and improving the ideas with access to tools such as web search to enable literature review and summarization. In particular, the system scales up test-time compute for scientific reasoning with an iterative, self-improvement loop, where the generated hypothesis competes in an Elo-score-based tournament.[15,16] The tournament accounts for multiple facets when scoring the proposed hypothesis and research directions including the criteria and constraints pre-specified in the initial research goal. The feedback and the win-loss patterns in the tournament are used to further refine and improve the ideas in the next iteration. At the end of the computation process, which can span for days instead of minutes with other frontier reasoning LLMs,[38,40,48] the system comes up with a ranked list of hypotheses and research proposals that attempt to satisfy the provided research goal with an additional research summary overview.

In this specific challenge, only one ‘‘input document’’ was provided by expert scientists, as detailed in Data S1, comprising only previously published and openly-available information as the research goal to the system. This document aimed to show the type of information that was used to challenge co-scientist. The amount of input delivered into the system was deliberately kept low. The document was also designed to be something any individual researcher could generate, tailored to their own interests, in order to interrogate co-scientist. This approach underscores the accessibility and general applicability of the system for hypothesis generation based on minimal, curated input.

No specialist tools or proprietary databases were utilized by the system for this experiment. As introduced earlier, the proposed AI-driven research directions are ranked by an Elo score similar to those used to rank chess players competing in tournaments.[15,16]

Framing the challenge as a prompt to the AI system: minimal input for maximum insight

We were extremely restrictive with the general information provided to the AI co-scientist system (Data S1). Essentially, we supplied only a single page containing a brief background on phage satellites, including some references, highlighting two key papers. One manuscript described the original discovery of cf-PICIs,[6] while the other detailed the development of a computational tool to identify phage satellites in bacterial genomes.[13] Using rudimentary methodology, we had observed many years ago that identical cf-PICIs

could be found in different bacterial species – a phenomenon we noted as the foundation of our unpublished *Cell* manuscript. The developed tool[13] confirmed this observation and provided multiple additional examples of identical cf-PICIs present in different bacterial species.

To illustrate this observation, we included an example from the unpublished *Cell* paper, in which we realized that one cf-PICI, EcCIGN02175 (renamed EcCI1), was present in five different bacterial genera and seven different bacterial species, including *Escherichia coli*, *Klebsiella pneumoniae*, *Shigella flexneri*, *Citrobacter freundii*, *Citrobacter amalonaticus*, *Enterobacter asburiae*, and *Enterobacter hormaechei* (Data S1). In the original input, we also included a cf-PICI (KpCIDSM30104; renamed KpCI1), which in our initial genomic analyses seemed to be present in both *K. pneumoniae* and *E. coli*. However, further analysis revealed that, while an identical island could not be found in other species (although similar islands are present), it was detected in multiple unrelated *K. pneumoniae* strains.

With this limited information, we posed the central question to the AI co-scientist system: why can cf-PICIs, but not other types of PICIs or satellites, be easily found across diverse bacterial species? What mechanism explains this phenomenon? (Data S1.)

Comparison of the AI co-scientist with other AI models

To enable a qualitative comparison with other AI systems, we also provided the same prompt as input to other state-of-the-art large language and reasoning models and a recently released "deep research" tool. This was done using their public-facing user interfaces. The specific models compared were OpenAI o1 (https://openai.com/index/openai-o1-system-card/), Gemini 2.0 Pro Experimental (https://blog.google/feed/gemini-exp-1206/), Gemini 2.0 Flash Thinking Experimental 12–19 (https://deepmind.google/technologies/gemini/flash-thinking/), OpenAI deep research (https://openai.com/index/introducing-deep-research/), OpenAI o3-mini-high (https://openai.com/index/openai-o3-mini/), Claude Sonnet 3.7 (https://www.anthropic.com/) and DeepSeek-R1 (https://huggingface.co/deepseek-ai/DeepSeek-R1).

QUANTIFICATION AND STATISTICAL ANALYSIS

No quantitative biological data or statistical hypothesis testing were performed in this study. The outputs of the AI co-scientist system were ranked using an Elo-style tournament scoring method,[15,16] analogous to ranking systems used in competitive games such as chess.