

The Rise of the AI Scientist

The Story of OASIS Helping Plant Research



Attorney Kim Kyungjin

KIMKJ.COM

Prologue: The Rise of the AI Scientist Finding Data and Proposing Hypotheses

In life science research, papers to read and data to analyze keep piling up. Life science is the study of how living things live and respond to their environment. Figuring out how plants survive in salty soil is also part of this field. Researchers cannot simply stop at reading papers; they must connect different experimental results to discover the reasons behind them.

There are also multiple types of data that look inside the bodies of living organisms. A gene is a segment of DNA that contains the information needed to build and maintain an organism's body. The complete set of genetic information an organism possesses is called the genome. Studies that broadly investigate how actively genes function and what proteins and substances are produced are collectively called omics. Looking at such data together allows us to discover relationships that are hard to find with only one type of data.

However, having a large amount of data does not mean answers appear on their own. Experimental conditions differ from paper to paper, and the storage locations and display formats of data also vary. Confirming whether a gene found in one study is the same as a gene in another study also takes time. It is becoming increasingly difficult for researchers to read and connect all the necessary information.

To reduce this burden, AI has entered research. AI helps search the literature, summarize long texts, and analyze data. On top of this, systems are emerging that add functions to formulate research plans, use necessary tools, and review results. Systems that assist scientific research across multiple stages like this are called AI scientists.

A hypothesis proposed by an AI scientist is an explanation that has not yet been confirmed. In science, a hypothesis must explain observations while being testable through experiments or analysis. Finding a paper stating that a certain gene is related to salt damage in plants is different from connecting multiple data sources to propose new research candidates. The latter

requires a process of examining what relationship each piece of data supports.

OASIS is an AI scientist created to study the processes by which plants respond to difficult environments. It was developed by teams including the research team of Lianguang Shang (商连光) at the Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences. Yazhouwan National Laboratory and the National Technology Innovation Center for Comprehensive Utilization of Saline-Alkali Land also participated in joint development. Saline-alkali land refers to soil where crops have difficulty growing because it is high in salt or strongly alkaline.

The environments that OASIS deals with are salt damage, drought, high temperature, and low temperature. Salt damage is the harm plants suffer due to high salinity. The burden plants experience in such environments is called abiotic stress. This term distinguishes it from damage caused by living organisms such as insect pests or pathogens. OASIS is designed to answer

research questions by connecting literature and biological data on 6 plant species.

Inside this system, there are 6 AI agents with different assigned tasks. An agent is an AI component that performs necessary tasks to meet a given goal. In OASIS, they divide up roles: making plans, searching literature, screening evidence, querying data, reviewing progress, and synthesizing results. It is a structure focused on carrying out multiple steps of work rather than providing a single one-off answer.

It also features the capability to reuse past task experience. Self-Evolving Experience Learning, or SEEL, derives reusable procedural rules from task history and error-correction experience. However, just because it possesses the ability to accumulate experience does not mean that all new answers become correct. Verifying whether rules are helpful must be distinguished from testing whether a research hypothesis is correct.

The capabilities of OASIS were examined through benchmark evaluation questions and a rice experiment case study. The fact that it solves evaluation questions well serves as evidence for judging its ability to retrieve and connect data. Experiments conducted on actual plants serve as evidence to confirm whether a proposed candidate is related to biological responses. The two results answer different questions.

In the rice case study, the researchers guided the analysis by asking questions multiple times. OASIS connected relevant data to prioritize 5 candidate genes. The researchers selected OsPP2a among them and conducted actual experiments. The experimental results provided evidence that this gene is involved in rice salt stress responses.

The rise of the AI scientist makes researchers rethink their own roles. Even if relationships between data can be found quickly, researchers must still decide what to study and judge which evidence to trust. There also remains the process of refining plausible explanations into testable hypotheses and investigating the reasons when actual results differ from

expectations. The value of an answer provided by AI depends on verifiable evidence rather than the confidence of its phrasing.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]
<https://doi.org/10.1016/j.molp.2026.09.005>

김영길, 이용주, 임수중, 송화전, 정문영, & 권오욱. (2026). AI Scientist 기술 개발 동향. *전자통신동향분석*, 41(5), 72–84.
<https://doi.org/10.22648/ETRI.2026.J.410507>

한국과학기술한림원. (2026-07-14). AI Scientist 시대, 인간 과학자의 자리. YouTube. [동영상] https://www.youtube.com/watch?v=WxsUB_-eCYo

刘祯. (2026-08-24). 【新华社】从“工具”到“搭档” AI应怎样助力科研. 新华社 (中国科学院转载) .
https://www.cas.cn/cm/202608/t20260824_5118852.shtml

Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Sirkovic, P., Myaskovsky, A., Glowaty, G., Weissenberger, F., Orlandi, A., Popovici, D., Palepu, A., Rong, K., Tanno, R., Saab, K., Zhang, F., Blum, J., Carroll, A., Kulkarni, K., Tomashev, N., ... Natarajan, V. (2026). Accelerating scientific discovery with Co-Scientist. *Nature*, 655(8122), 487–496.
<https://doi.org/10.1038/s41586-026-10644-y>

Lu, C., Lu, C., Lange, R. T., Yamada, Y., Hu, S., Foerster, J., Ha, D., & Clune, J. (2026). Towards end-to-end automation of AI research. *Nature*, 651(8107), 914–919. <https://doi.org/10.1038/s41586-026-10265-5>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21). Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. <https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

Table of Contents

Prologue: The Rise of the AI Scientist Finding Data and Proposing Hypotheses

Chapter 1: Finding Answers in Data and Formulating Hypotheses

Chapter 2: Why Answers Are Hard to Find Despite Plentiful Papers and Biological Data

Chapter 3: Six AI Agents Sharing the Work of OASIS

Chapter 4: SEEL Turning Failures into Rules for the Next Research

Chapter 5: The Role OASIS Takes Among Various AI Scientists

Chapter 6: What the 84 Points on PlantStressQA Tell Us and What They Do Not

Chapter 7: Connecting Clues from Multiple Plants to Select OsPP2a

Chapter 8: Confirming the AI-Generated Hypothesis with Actual Rice

Chapter 9: Humans Decide What to Ask and What Conclusions to Draw

Epilogue: Better Questions in the Face of More Answers

Chapter 1: Finding Answers in Data and Formulating Hypotheses

A researcher finding answers in papers and formulating new hypotheses are connected. To find out whether a certain gene is related to salt damage in plants, one must start by reading existing research. However, just because a related paper is found does not mean one can state that the gene does the same thing in other plants. In between, a process of checking additional data and testing explanations is required.

A large language model is an AI trained on vast amounts of text to generate sentences tailored to questions. It is commonly referred to by the acronym LLM. Researchers can use such models to summarize long papers or get explanations for unfamiliar terms. However, natural-sounding sentences do not automatically mean accurate answers. The phenomenon in which a model convincingly generates information that differs from the facts is called hallucination.

One method to reduce this problem is Retrieval-Augmented Generation, or RAG. RAG is a method that retrieves external data relevant to a question and provides it to the AI. The AI creates answers by referring to that data. This helps incorporate information not encountered during training into answers or present the documents that served as the basis for the answer.

Understanding RAG merely as a literature summarization technique misjudges its scope. One can also compare multiple pieces of evidence and construct explanations using retrieved data. RAG can also be used within systems equipped with research planning or data analysis capabilities. However, the function of searching data to generate answers does not automatically include procedures for selecting analysis tools and testing hypotheses.

The difference lies in what tasks are required to answer a question. A question asking about the already known function of a gene can be answered through literature search. On the other hand, selecting a gene whose function has not yet been confirmed as a research candidate requires multiple types of

evidence. One must jointly examine functions in other plants, similarities in genetic information, and changes in experimental data.

Distinguishing technical terms properly is also important. In plant research, ABA refers to a plant hormone called abscisic acid. Hormones are signaling substances that regulate responses within an organism's body. Abscisic acid is involved in how plants respond to arid environments and other conditions. Because the same acronym is used with different meanings in other fields, it must be interpreted according to context.

One should not guess the meanings of FT and CO simply by looking at the letters, either. FT is the name of a gene involved in the flowering process. CO is the name of a protein that regulates how genes work. Proteins like CO that control the activity of specific genes are called transcription factors. The fact that the two names appear together and an explanation of how one regulates the other are two different pieces of information.

A knowledge graph is a way to organize such relationships. It is data that represents entities such as genes and proteins individually and connects the relationships between them. One can trace relationships to find which protein regulates the operation of which gene. Moving beyond finding documents with similar words, it helps investigate biological connections.

PlantScience.ai is a separate AI system that answers questions based on botanical knowledge. It constructed a knowledge graph through AutoSKG, which automatically organizes scientific knowledge. To this, it combined the capability to distinguish names of genes, proteins, and other entities in documents. When receiving a question, it performs both a search for data with similar meanings and a traversal following relationships in the knowledge graph. This approach is called graph-based RAG.

PlantScience.ai links the discovered relationships into knowledge chains to present the grounds for its answers. When evaluated on 1,770 questions curated by experts, it showed better or comparable results in 73.1~87.6% of the questions

depending on the general-purpose language model compared. This figure does not represent the correct answer rate for all questions. It is the proportion of questions where it was superior or equal compared to the baselines.

The ability to find relationships in literature well cannot resolve every research question on its own. Researchers also examine numerical values measuring gene activity levels and positional information where plant characteristics vary. Here, the transcriptome refers to the entirety of results where genetic information in cells has been transcribed into RNA. Comparing transcriptomes allows us to observe how the operation of specific genes changes depending on conditions.

A QTL is a genomic region associated with traits expressed as numerical values, such as height or the degree of tolerating salt damage. An eQTL is a genomic location associated with differences in gene activity levels. Just because a certain region is associated with a plant trait does not mean a specific gene within it can be immediately confirmed as the cause. One must

further connect explanations in literature with actual measurement data to narrow down research candidates.

A research system combining AI agents divides this work into multiple stages. It turns a research question into detailed subtasks and retrieves necessary literature and numerical data. Next, it examines whether the data supports the hypothesis and organizes the results. Depending on the system, it may also include functions to write or execute analysis code. Executing code is a process of performing computations, and it does not guarantee the biological significance of the results.

OASIS divided roles to connect literature and omics data in plant stress research. It simultaneously performs the task of finding mechanisms of action in literature and querying relevant evidence in data. The review role verifies whether the evidence covered the required scope and whether tasks have been completed. This cannot be viewed as a mechanism that filters out all hallucinations or logical errors.

In the rice research case study as well, discovery did not end with a single question. The research team guided the analysis by asking questions multiple times, and OASIS connected evidence for candidates during that process. The research team selected targets for actual experiments from among the proposed candidates. The difference shown by this case study lies not in the presence or absence of search, but in the scope of work connecting data to narrow down candidates to test.

There are also limitations to AI that creates hypotheses. A critical study by the research team of Bisht points out the problem of AI relying heavily on published writings. Know-how learned firsthand in the research field or undisclosed failure results are difficult to learn sufficiently through papers alone. Because of conditions not revealed in training data, proposals that sounded convincing in writing may not hold true in actual experiments.

Having multiple agents does not automatically make hypotheses diverse either. Models adjusted to answer according to human evaluations and preferences may converge on similar

explanations. The research team of Bisht discusses this as a risk of diminishing hypothesis diversity. Even if multiple answers agree, it is not the same as gathering independent experimental evidence.

The ability to find answers in data and the ability to formulate hypotheses are not in competition. Accurate retrieval becomes the starting point of a hypothesis, and a hypothesis indicates what data should be searched further. What research requires is continuing this process while verifying the evidence at each stage. If a new candidate is proposed, it must be clear why that candidate was chosen and how it will be tested.

References

Yu, H., Zhou, S., Huang, M., Ding, L., Chen, Y., Wang, Y., Ren, Y., Cheng, N., Wang, X., Liang, J., Zhang, H., Ding, Y., & Li, K. (2026). PlantScience.ai: An LLM-powered virtual scientist for plant science. *Molecular Plant*, 19(5), 1117–1123. <https://doi.org/10.1016/j.molp.2026.03.010>

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research.

Molecular Plant. [온라인 선게재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

Bisht, H., Kumar, V., Jablonka, K. M., Mausam, & Krishnan, N. M. A. (2026). Agentic AI Scientists Are Not Built For Autonomous Scientific Discovery. arXiv preprint (cs.AI). [프리프린트]

<https://arxiv.org/abs/2605.08956>

Lyu, Y., Zhang, X., Zhao, Y., Papakostas, K., Yi, X., Guo, S., Hu, W., Piotrowski, J., Kaliski, J., Urbani, J., Meng, Z., Zhou, L., Ren, Z., & Yan, X. (2026). EvoScientist: Towards Multi-Agent Evolving AI Scientists for End-to-End Scientific Discovery. arXiv preprint (cs.CL). [프리프린트]

<https://arxiv.org/abs/2603.08127>

Chapter 2: Why Answers Are Hard to Find Despite Plentiful Papers and Biological Data

While growing, crops must endure drought, high temperatures, and soil with high salinity. Such environments can hinder plant growth and reduce yields. As climate change increases the need to respond to extreme weather, researchers are seeking the mechanisms by which crops withstand environmental changes. Understanding those mechanisms is necessary to determine what traits crops should have when grown.

To understand plant responses, looking only at outward appearance is not enough. One must jointly examine how genetic information changes and which genes function more actively. Changes in proteins and internal substances also provide clues for understanding responses. Research that analyzes different kinds of data together is called multi-omics.

The data included in multi-omics each show different aspects. Genomic data deals with DNA information possessed by organisms. Transcriptomic data shows the state where genetic information has been transcribed into RNA. Proteomic data deals with the entire set of proteins, and metabolomic data deals with all substances produced or used in life activity processes. Epigenomic data examines markers and other factors that influence gene activity even without changes in the letter sequence of DNA.

With the advancement of technologies that read DNA information rapidly and equipment that measures plant characteristics on a large scale, data has also increased. Observable traits such as plant height and leaf shape are called phenotypes. Researchers look for relationships by comparing genetic information and phenotypes. However, the speed of producing data is not the same as the speed of uncovering its meaning.

One of the sticking points in research is that data is scattered across multiple places. A paper contains functional descriptions

of a specific gene, while a separate database holds measurement values. In yet another location, the position of that gene is recorded. A database is a repository created to collect and search such data in a consistent manner. Even after finding where the needed data is located, the task of matching them together remains.

The same gene is sometimes labeled with different names. Research uses a gene ID, which is an identification number to distinguish genes. However, ID systems can differ from database to database. In rice research, RAP-DB and MSU data are used, while there are sources such as TAIR for Arabidopsis and Spud DB for potato. Comparing names directly can lead to mistakenly counting the same target as different genes.

The reference standard indicating gene locations must also be verified. A reference genome is genomic information used as a standard when multiple studies compare locations. When this reference is updated, the coordinates of the same gene can change. If locations from different versions are overlaid directly,

targets that are actually close together might be judged as separated, or vice versa.

Data formats and descriptions also vary widely. Information explaining and annotating gene functions is called an annotation. While some data describes functions in detail, other data marks them as still unknown. Methods for writing measurement units and experimental conditions also differ. Merely gathering data in one place cannot resolve these differences.

Differences within the same crop must also be taken into account. Rice has lineages such as japonica and indica, as well as Tongil-type cultivars. Because the genetic background differs depending on lineages and cultivars, responses can vary even in the same environment. One should not assume that a salt stress-related signal found in one cultivar will appear identically in another cultivar.

The stage of a plant's growth and its surrounding environment also change the results. Responses can differ between when it is

a young seedling, when its stems branch out, or when its grain heads emerge. Temperature, humidity, and light conditions also have an effect. The relationship where environmental effects vary depending on genetic traits is called genotype-by-environment interaction. This means that looking only at genetic information or only at the environment makes it difficult to fully explain how a plant responds.

Researchers use statistical methods to reduce these variations. BLUP is a method that predicts trait values by taking into account multiple observations and differences in conditions. Meta-analysis, which analyzes results from multiple studies together, is also used. However, adjusting differences statistically does not make all conditions identical. There still remains a gap between finding an associated genomic region and uncovering the action of the actual causal gene.

When drought and high temperature strike together, the situation becomes even more complex. Stomata on plant leaves are tiny pores through which water and gases pass in and out. Reducing water loss through stomata helps endure drought.

Meanwhile, the cooling process as water evaporates is involved in responding to high temperatures. When both environments overlap, plants must respond to conflicting demands at the same time.

Therefore, even if one knows the changes during drought and the changes during high temperature separately, it is difficult to predict the outcome of combined conditions simply by adding them together. This is because gene activity and changes in internal substances can display distinct patterns. To apply data from studying a single stress to combined stress, one must re-examine under what conditions the results were obtained.

Even after finding many genes as candidates, interpretation remains. GWAS is an analysis that looks at what associations exist between genome-wide differences and organismal traits. QTL analysis finds genomic regions associated with traits expressed as numerical values. While such analyses can narrow down associated loci, they cannot immediately reveal which gene at that locus does what.

A similar problem occurs in transcriptomic analysis. Suppose the activity level of a certain gene changes after saltwater treatment. That change could be a response contributing to reducing damage, or it could be a consequence resulting from the plant already being damaged. One cannot determine cause and effect solely from the fact that measured values changed. If it is a gene whose function is unknown, literature needed for interpretation may also be lacking.

Literature search helps interpret such data. This is because one can find what similar genes do in other plants. However, one must check whether functions confirmed in the literature and currently measured numerical values deal with the same conditions. One must also verify whether gene names match, whether data coordinate reference standards are identical, and whether experimental conditions are comparable.

RAG, which is Retrieval-Augmented Generation, can also be used to search data and present evidence. However, a system equipped only with the ability to read literature and answer requires separate procedures to align and analyze numerical

data. Comparing measured values in tables with genomic coordinates demands capabilities to handle data formats and semantics in addition to strong reading comprehension.

To narrow this gap, OASIS connects literature and omics data into a single workflow. In the rice research, it narrowed down candidates by jointly examining functional studies in other plants and rice genetic data. In that it connected different pieces of evidence in response to questions from the research team, it demonstrates how data integration is used in hypothesis exploration.

The difficulty in life sciences does not arise solely from a shortage of needed numbers. One must confirm whether the same name refers to the same entity and whether the same number indicates the same condition. One must also distinguish whether an association implies a mechanism of action. To turn data into evidence usable for research, one must go through this process of interpretation and connection.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

Gao, J., Kim, T.-H., Baek, D.-H., Lee, C.-J., Park, W.-G., Yun, S. D., & Kim, S.-M. (2026). Integrative genomic and transcriptomic dissection of salt tolerance for Japonica rice improvement. *Frontiers in Plant Science*, 16, Article 1751273. <https://doi.org/10.3389/fpls.2025.1751273>

Xu, C., Chen, R., Li, W., Song, L., Huang, S., Zhang, F., Liu, Y., Huang, X., Li, N., Du, L., & Shen, S. (2026). Genomics, Multi-Omics, and Emerging Artificial Intelligence for Combined Drought–Heat Stress Resilience in Plants: A Structured Narrative Review. *Plants (Basel)*, 15(16), Article 2516. <https://doi.org/10.3390/plants15162516>

Dong, Y., Li, J., Li, F., Luo, J., Jia, Y., Li, D., Wang, L., Su, X., Hu, J., Shang, Y., Huang, S., Zhu, Y., & Jia, Y. (2026). Potato Agent: AI-Driven Data and Knowledge Exploration on an Agent-Ready Potato Multi-Omics Platform. *bioRxiv preprint (openRxiv)*. [프리프린트]

<https://doi.org/10.64898/2026.08.12.744101>

Wang, Z., Liang, C., Zhang, X., Wei, W., & Li, K. (2026). BOLE: A Knowledge-Enhanced Multi-Agent Framework for Intelligent Genomic Breeding. *bioRxiv preprint (openRxiv)*. [프리프린트]

<https://doi.org/10.64898/2026.02.04.703673>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21).
Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物
耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.
<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

Chapter 3: Six AI Agents

Sharing the Work of OASIS

Even answering a single research question requires many different types of tasks. One must search the literature, read experimental conditions, and compare numerical data. One must also examine whether the gathered evidence sufficiently answers the question. To connect these tasks of different natures, one must determine what each step handles and what results it passes on.

OASIS divides that work among 6 specialized agents. An agent is an AI component that performs tasks according to an assigned goal. A structure combining multiple agents is called a multi-agent system. In OASIS, six roles cooperate: planning, hybrid search, evidence extraction, data querying, review, and synthesis.

Dividing roles alone does not make the answer accurate. A gene found in the literature and a gene retrieved from data must be the same. One must also verify whether the searched content

relates to the research question. One must be able to trace which data each agent handled in sequence, so that if a problem arises in the results, the reason can be investigated.

At the foundation of this work is plant stress-related data. OASIS covers 6 plant species: Arabidopsis, rice, maize, wheat, soybean, and upland cotton. Upland cotton is a type of cotton. The data covers four types of stress: salt damage, drought, high temperature, and low temperature. While connecting results obtained from different plants, differences in species and environmental conditions must be considered together.

Literature data includes more than 150,000 paper abstracts and about 15,000 full texts. An abstract is a short text summarizing the purpose, methods, and results of a study. A full text means the entire paper. Such a collection of data built so that necessary knowledge can be gathered and searched is called a knowledge base. Even with a large volume of literature, finding evidence that fits the question remains a separate necessary task.

In addition to written literature, OASIS connects omics and genomic data. This includes data where gene activity levels changed depending on conditions, genetic differences among plant populations, and genomic locations related to traits. While literature explains the working mechanism of a gene, data allows researchers to check whether measurement results align with that explanation.

The first role is planning. The planning agent divides the researcher's question into small, actionable tasks. It organizes what needs to be known to answer the question and which data to look for first. If the result of one task is needed to perform another, it also takes that sequence into account.

Planning bridges the broad research goal and actual queries. Simply requesting an investigation of a plant's salt stress response leaves a broad scope of data to search. The required evidence differs depending on whether the task is finding related genes or comparing known functions. Dividing the goal into detailed tasks also makes it easier to check whether search results have strayed from the original question.

The second role is hybrid search. Hybrid means combining different methods. It uses both a method that searches for materials with similar sentence meanings and a method that searches for materials with matching gene names or technical terms. Searching by meaning helps find relevant materials even if the phrasing differs. Searching by matching names is necessary to distinguish specific genes or terms.

The two search methods focus on different targets. The same biological process can be described with different words across papers. Conversely, genes with similar names may have different functions. Therefore, both semantic similarity in expression and the exact names of targets must be checked together. The search step gathers materials to review rather than finalizing the answer.

The third role is evidence extraction. Searched texts also contain content that is not directly related to the question. The evidence extraction agent extracts grounds for working mechanisms relevant to the research goal. It preserves the content needed to

examine which gene connects to which response and what experimental conditions support that explanation.

In biology, the principle behind how something works is called a mechanism. A mechanism is not explained simply by the fact that two things are mentioned together. For example, what response followed when a gene's activity changed could be relevant to the research question. Picking out this kind of evidence from the literature clarifies what should be connected to actual data later.

The fourth role is data querying. This agent searches and examines information needed for the question within connected omics and genomic data. Unlike reading explanations written in literature, it handles data such as measurement values and genomic locations. It checks changes in gene activity under specific conditions or related signals at specific locations to connect them with literature evidence.

Genomic location information helps narrow down research candidates. A QTL is a genomic region associated with a trait

expressed numerically, and an eQTL is a location associated with differences in gene activity levels. When two signals appear at the same location or in overlapping regions, it is called colocalization. Colocalization can serve as a ground for examining candidates, but it alone cannot determine which gene is the causal one.

The fifth role is review. The review agent verifies whether the collected evidence covers the scope required for the task. It also checks whether planned tasks were completed. Its role is to check the gap between the fact that materials were found and the judgment that the question was sufficiently answered. Even after going through this, the accuracy of scientific conclusions is not automatically guaranteed.

Review is necessary because completing a task is different from having sufficient evidence. Even if multiple papers are found, they may not cover the required plant or conditions. Even if data query results exist, they may address only a part of the question. Examining the scope of evidence and the task

completion status makes it possible to distinguish what has been verified from what remains.

The sixth role is synthesis. The synthesis agent connects the evidence extracted from literature with data query results, organizing them into an answer that researchers can read. Depending on the research question, it can propose hypotheses or the priority of candidate genes. At this time, presenting results along with their evidence allows researchers to examine the reasoning process.

In synthesized results, observed facts must be distinguished from inferences drawn from them. Finding a related signal at a specific location is a fact confirmed in the data. The explanation that a gene at that location might be involved in salt stress response is a hypothesis to be tested. Distinguishing between the two allows one to understand the reason for recommending a candidate and the parts requiring further verification.

The six roles operate in connection with the researcher's questions. In rice research, the research team guided the

direction of the analysis by asking questions multiple times. OASIS connected literature and data evidence accordingly and prioritized candidates. Therefore, dividing roles among agents does not mean the entire study was completed without human questions or choices.

In this structure, the important connection lies between the explanations in the literature and the signals in the data. Planning determines the necessary tasks, while search and evidence extraction provide literature grounds. Data querying examines numerical and location data, while review and synthesis organize the task scope and results. When researchers can see what evidence was added at each stage, they can genuinely evaluate the proposed hypotheses.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]
<https://doi.org/10.1016/j.molp.2026.09.005>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21).
Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物
耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.
<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

韩漾. (2026-10-20). AI智能体“绿洲”发布 助力植物抗逆研究. 中国农业科学院
（来源：农业基因组研究所，作者韩漾）.
<https://caas.cn/xwzx/kyhd/28164b3f0e7f4996a5749892288670bc.htm>

김영길, 이용주, 임수중, 송화전, 정문영, & 권오욱. (2026). AI Scientist 기술 개발
동향. 전자통신동향분석, 41(5), 72–84.
<https://doi.org/10.22648/ETRI.2026.J.410507>

Bu, D., Sun, J., Li, K., He, Z., Huang, W., Hu, J., Zhang, S., Lei, S., Huo, P.,
Wang, Z., et al. (2026). Empowering AI data scientists using a multi-agent
LLM framework with self-evolving capabilities for autonomous,
tool-aware biomedical data analyses. Nature Biomedical Engineering.
[온라인 선거재; 권·호·쪽 미확인]
<https://doi.org/10.1038/s41551-026-01634-6>

Chapter 4: SEEL Turning Failures into Rules for the Next Research

The process of AI searching for and analyzing research data does not end with a single answer. Errors can occur while reading multiple sources and using tools sequentially. Even for the same gene, if names or identification numbers differ across sources, the connection gets blocked. A gene is a part of DNA that contains information necessary for an organism's body to form and function. If an identification number assigned to a gene is incorrectly connected, data for a different gene might be retrieved.

Work can also halt if numerical notations in data differ or analysis programs do not match. If the order of using tools is mixed up, the required result may not be obtained. Even if such a problem is fixed once, repeating the same trial and error occurs if that experience cannot be used in the next task. An AI

that assists research needs the ability to reuse lessons learned from past work, along with the ability to generate answers.

OASIS includes a self-evolving experience learning feature for this purpose. Its English name is Self-Evolving Experience Learning, abbreviated as SEEL. Here, experience refers to the work records left behind while the AI executes tasks. Self-evolving means transforming lessons gained from those records so they can be used in subsequent tasks.

SEEL does not stop at looking only at results completed without any issues from start to finish. It also examines processes that initially failed but were completed after correcting errors. It compares what steps were taken during failure and what was changed before finishing the task. In other words, it looks at the correction process that might be missed if only successful answers were saved.

The core lies not in merely accumulating past records as they are. SEEL extracts work rules from those records that can be reused in other tasks. These are called procedural rules. While

knowledge gained as results corresponds to what was learned, procedural rules correspond to the order and method in which work should be done.

Comparing failed paths with corrected paths allows one to examine which parts of the work need to be changed. Experience obtained this way becomes easier to reuse than a long log of a single task. However, a method that helped once does not necessarily suit all tasks. If an incorrect lesson is preserved as a rule, there is also a risk of spreading the same error across multiple tasks.

Therefore, OASIS places verified experiences into a skill library. A skill library is a place where methods to be retrieved and reused in future work are stored. Extracting rules from experience is distinct from verifying whether those rules can be trusted and reused. The mere fact that there are many failure records does not mean the system has improved.

A specific case handled by SEEL is the execution history of 86 tasks from PlantStressQA. PlantStressQA is a collection of

benchmark problems that evaluates how well AI handles issues where plants respond to difficult environments such as salinity or drought. SEEL gained experience for future work from the execution records of these tasks. Here, 86 is the number of historical tasks used for experience learning.

What this experience learning targets is the repetition of the research process. Even in a new task, some steps of connecting data or using tools may resemble past tasks. If verified rules can be reused, the need to solve already resolved problems from scratch decreases. However, the process of checking whether conditions for applying the rules are met is still necessary.

Attempts to preserve and reuse experience continue in other AI research systems as well. EvoScientist proposes a structure that separately manages memories of research ideas and experimental execution. It can be compared with SEEL in that it connects records during research to subsequent work. However, different systems cannot be assumed to use the same memory structure or the same improvement methods.

Rules for executing tasks well must be distinguished from biological facts. Thoroughly finding data and reducing analysis errors does not mean a new hypothesis immediately becomes fact. What SEEL assists with is connecting past correction experiences to the next task. Whether hypotheses emerging through that process are correct in actual plants must still be confirmed through separate experiments.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21). *Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究*. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.

<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

Lyu, Y., Zhang, X., Zhao, Y., Papakostas, K., Yi, X., Guo, S., Hu, W., Piotrowski, J., Kaliski, J., Urbani, J., Meng, Z., Zhou, L., Ren, Z., & Yan, X. (2026). *EvoScientist: Towards Multi-Agent Evolving AI Scientists for*

End-to-End Scientific Discovery. arXiv preprint (cs.CL). [프리프린트]
<https://arxiv.org/abs/2603.08127>

Ghareeb, A. E., Chang, B., Mitchener, L., Yiu, A., Szostkiewicz, C. J., Shved, D., Gyimesi, G. J., Laurent, J. M., Wright, S. M., Razzak, M. T., White, A. D., Finnemann, S. C., Hinks, M. M., & Rodriques, S. G. (2026). A multi-agent system for automating scientific discovery. *Nature*, 655(8122), 497–505. <https://doi.org/10.1038/s41586-026-10652-y>

Srikanth, D., Zhao, B., Xu, D., Wu, Y., & Jiang, Z. (2026). Recursive self-improvement of AI research agents. arXiv preprint. [프리프린트]
<https://arxiv.org/abs/2609.26457>

Chapter 5: The Role OASIS Takes Among Various AI Scientists

Using the name AI scientist does not mean that all systems perform the same work. Some systems conduct experiments inside a computer, while others propose hypotheses regarding disease treatment. There are also systems focused on finding plant data. To compare them, one must first look at what problems they were designed to solve.

Here, an agent refers to an AI program that carries out work by finding materials or using tools according to a goal. Systems that use multiple agents together sometimes divide roles. For instance, the role of finding materials and the role of reviewing results are kept separate. Such an organization is a way to connect multiple stages of research, but the scope of work actually handled varies by system.

Sakana AI's The AI Scientist mainly deals with the automation of AI and machine learning research. Machine learning is a technology that enables computers to learn rules from data. This system connects the process of generating research ideas, writing and executing experimental code, and organizing results. Writing papers and automated review are also included in its scope of research. Its main stage is research conducted by computer.

Google's Co-Scientist generates and refines scientific hypotheses according to goals set by researchers. Multiple agents produce proposals, review them, and rank them. Through the process of comparing and revising candidates, it narrows down hypotheses to explore further. Automating the entire process of computer experiments has a different purpose from improving the quality of hypotheses to be reviewed.

Application cases of Co-Scientist include research seeking candidate treatments for acute myeloid leukemia. This approaches the problem by exploring the possibility of using existing drugs for other treatments or examining drug

combinations. It also addressed research on liver fibrosis, where scar tissue accumulates in the liver, and antibiotic resistance. In such cases, hypotheses proposed by AI and results confirmed through experiments by research teams must be read separately.

Biomni is a general-purpose agent that performs various tasks in biomedical research. It carries out analysis by connecting a diverse array of tools, data, and program execution environments. Its scope includes narrowing down gene candidates related to diseases and assisting tasks in molecular biology research. It is an approach aimed at handling multiple types of research problems by linking a wide range of tools.

CellVoyager focuses on analyzing biological data obtained from individual cells. Among them, single-cell transcriptome data shows what gene information was read and how much in each cell. Even cells within the same tissue can have different activity states, making cell-by-cell analysis necessary. CellVoyager assists in analyzing such data and testing hypotheses.

Robin is another life sciences research system that must be distinguished from OASIS. It connects literature search with data analysis, and deals with the flow of reflecting experimental results into follow-up research. Its application cases include research searching for candidate treatments for dry macular degeneration, which impairs central vision in the eye. Simply sharing the common ground of dealing with life sciences does not allow one to transfer these achievements to those of OASIS.

Agents targeting plants also have different purposes. BotAI assists in finding and exploring information based on plant diversity data. PotatoAgent focuses on potato research. It supports tasks like gene exploration by connecting various biological data and genomic data of potatoes. While they share plants as their subject, the data they handle and the research questions they address are different.

OASIS focuses on the problem of plants responding to harsh environments. Burdens caused by non-living environmental factors such as salinity, drought, high temperatures, and low temperatures are called abiotic stress. Even under the same

stress, responses can differ depending on the plant species and genes. OASIS was designed to examine these differences by connecting literature and biological data.

The scope covered by OASIS includes 6 plant species: Arabidopsis, rice, maize, wheat, soybean, and upland cotton. Literature data includes more than 150,000 abstracts and about 15,000 full-text papers. An abstract is a short text summarizing the core of a paper, and a full text is the entire paper. To this, it connects omics data that broadly measures gene activities and genomic differences.

When plant species change, gene names, identifiers, and data formats can also change. To examine whether the known role of a gene in one plant continues in another plant, these differences must be reconciled. The specialization of OASIS lies in handling together the data needed for the research question of plant stress. This does not mean other systems cannot handle such data at all.

OASIS works by dividing the roles of planning, search, evidence organization, data querying, review, and synthesis. It also features SEEL, which converts execution history into reusable rules. This configuration is tailored to the purpose of connecting explanations in literature with actual measurement data. The value of research results cannot be judged solely by the fact that there are many tools or multiple roles.

When comparing AI scientists, what tasks they can handle is more important than which system has a broader name. Computer experiments, medical hypothesis generation, cellular analysis, and crop data exploration require different preparations. The potential shown by OASIS lies in connecting scattered evidence in plant stress research and narrowing down candidates to review. Only by making that scope clear can researchers choose the system that fits their questions.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research.

Molecular Plant. [온라인 선게재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Sirkovic, P., Myaskovsky, A., Glowaty, G., Weissenberger, F., Orlandi, A., Popovici, D., Palepu, A., Rong, K., Tanno, R., Saab, K., Zhang, F., Blum, J., Carroll, A., Kulkarni, K., Tomashev, N., ... Natarajan, V. (2026). Accelerating scientific discovery with Co-Scientist. *Nature*, 655(8122), 487–496.

<https://doi.org/10.1038/s41586-026-10644-y>

Lu, C., Lu, C., Lange, R. T., Yamada, Y., Hu, S., Foerster, J., Ha, D., & Clune, J. (2026). Towards end-to-end automation of AI research. *Nature*, 651(8107), 914–919. <https://doi.org/10.1038/s41586-026-10265-5>

Huang, K., Zhang, S., Wang, H., Qu, Y., Lu, Y., Li, R., Roohani, Y., Qiu, L., Cao, S., Li, G., Zhang, J., Yin, D., et al. (2026). Autonomous biomedical research with an artificial intelligence agent. *Science*, 393(6813), ead4351. <https://doi.org/10.1126/science.adz4351>

Dong, Y., Li, J., Li, F., Luo, J., Jia, Y., Li, D., Wang, L., Su, X., Hu, J., Shang, Y., Huang, S., Zhu, Y., & Jia, Y. (2026). Potato Agent: AI-Driven Data and Knowledge Exploration on an Agent-Ready Potato Multi-Omics Platform. *bioRxiv preprint (openRxiv)*. [프리프린트]
<https://doi.org/10.64898/2026.08.12.744101>

Ghareeb, A. E., Chang, B., Mitchener, L., Yiu, A., Szostkiewicz, C. J., Shved, D., Gyimesi, G. J., Laurent, J. M., Wright, S. M., Razzak, M. T., White, A. D., Finnemann, S. C., Hinks, M. M., & Rodrigues, S. G. (2026). A multi-agent system for automating scientific discovery. *Nature*, 655(8122), 497–505. <https://doi.org/10.1038/s41586-026-10652-y>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21).
Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物
耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.
<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

Alber, S., Chen, B., Sun, E., Isakova, A., Wilk, A. J., & Zou, J. (2026).
CellVoyager: AI CompBio agent generates new insights by autonomously
analyzing biological data. *Nature Methods*, 23(4), 749–759.
<https://doi.org/10.1038/s41592-026-03029-6>

中国科学院植物研究所. (2026-10-28). 植物所植物科学数据中心团队发布植
物多样性智能体BotAI. 中国科学院植物研究所 官网 科研进展. [기관 소식]
https://ib.cas.cn/2019gb/kyjz2019/202610/t20261028_8299763.html

Chapter 6: What the 84 Points on PlantStressQA Tell Us and What They Do Not

To know how well AI helps with research, we need to compare its abilities using the same standards. The tests or evaluation criteria used for this purpose are called benchmarks. However, the meaning of a score changes depending on what kind of test was taken. Just because an AI solves math problems well does not mean we can judge that it analyzes plant research data well.

In plant stress research, the ability to understand the meaning of text alone is not enough. One must find necessary evidence in research papers, read measurement data, and examine how different results connect to each other. Knowing the name of a gene does not mean one immediately knows whether that gene is involved in responding to salt. There is a need to separately evaluate the ability to find and connect related evidence.

PlantStressQA is a question set created to evaluate such research problems. It consists of 200 questions prepared by experts. Its purpose is to examine how well AI handles knowledge and data about plant stress. The evaluation scope includes 6 plant species and 4 types of stress: salinity, drought, heat, and cold.

To solve these questions, one must read explanations about genes together with actual measurement results. The function of a gene refers to what work it is involved in inside the body. Gene expression is the process in which information from that gene is read and used. A result showing that the expression level changed in a certain environment provides a clue to infer function, but that alone cannot confirm its role.

Data are also needed to find which locations in the genome are related to plant characteristics. At this time, a broad region of a chromosome may appear as a clue. Because multiple genes can be located within that region, finding the location is different from choosing the causal gene. This is why one must examine the literature together with other measurement results.

OASIS scored 84.0 points out of 100 on PlantStressQA. It was 33.9 points to 50.2 points higher than the language models used as comparison baselines. This figure is a difference in points. It must not be interpreted as meaning that performance increased by 33.9 percent to 50.2 percent.

In task-by-task comparisons, the areas presented with the largest differences were population variation and genetic mapping analysis. Population variation is a task that deals with differences appearing in the genomes of multiple populations of individuals. The gap in this area was 55.6 points. Genetic mapping analysis is a task of finding genomic locations related to specific traits, and the gap was 52.3 points. The score difference in the overall evaluation and the score differences by task must be read distinctively.

This result is evidence that OASIS solved problems better than the comparison targets in that evaluation. It shows that a system designed to handle literature and measurement data together can be helpful in specialized research problems. However, a single overall score alone cannot tell us which

capability made how much of a difference. To separate the contributions of retrieval, data lookup, review, and experiential learning, appropriate comparisons are needed.

The 84.0 points must not be read as an 84 percent success rate in actual experiments either. This score is an evaluation result for the PlantStressQA questions. It does not mean that if AI proposes 100 gene candidates, 84 of them will be correct in experiments. Nor is it a guarantee that it will deliver the same performance across all types of plant research.

There is a difference between solving evaluation questions and solving new research problems. In real research, necessary data may be missing, or results may not match one another. Living organisms' responses can also vary depending on experimental conditions. It is difficult to judge how well an AI will handle such situations based solely on a score obtained from a fixed set of questions.

The case of OsPP2a in rice also shows that test scores and experimental evidence are different kinds of grounds. The

research team guided the analysis of OASIS by asking multiple questions. From among the 5 candidates prioritized by OASIS, the researchers selected OsPP2a and conducted actual experiments. Those experiments provided genetic and molecular evidence that this gene is involved in rice salt response. This is not a result that automatically follows from the PlantStressQA score.

Benchmarks are evidence that can be used when selecting and improving research tools. They allow us to see which areas are relatively strong and which tasks show large differences. However, we must respect the scope of questions that a score answers. The 84.0 points on PlantStressQA evaluate the ability to handle defined plant stress questions, and it is not a number that decides the truth or falsity of a new hypothesis in place of experiments.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research.

Molecular Plant. [온라인 선거재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21).

Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.

<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

韩漾. (2026-10-20). AI智能体“绿洲”发布 助力植物抗逆研究. 中国农业科学院（来源：农业基因组研究所，作者韩漾）.

<https://caas.cn/xwzx/kyhd/28164b3f0e7f4996a5749892288670bc.htm>

Bisht, H., Kumar, V., Jablonka, K. M., Mausam, & Krishnan, N. M. A.

(2026). Agentic AI Scientists Are Not Built For Autonomous Scientific Discovery. arXiv preprint (cs.AI). [프리프린트]

<https://arxiv.org/abs/2605.08956>

Yu, H., Zhou, S., Huang, M., Ding, L., Chen, Y., Wang, Y., Ren, Y., Cheng, N., Wang, X., Liang, J., Zhang, H., Ding, Y., & Li, K. (2026). PlantScience.ai: An LLM-powered virtual scientist for plant science. Molecular Plant, 19(5), 1117–1123. <https://doi.org/10.1016/j.molp.2026.03.010>

Chapter 7: Connecting Clues from Multiple Plants to Select OsPP2a

When too much salt accumulates in the soil, rice has difficulty growing. The ability to endure such an environment is called salt tolerance. To find genes that can help improve salt tolerance, one must narrow down many candidates to those genuinely likely to be involved. Collecting a long list of gene names is different from choosing which targets to test in experiments.

Different research materials also provide different information. Some papers explain the roles that genes perform inside the plant body. Other materials show how gene activity changes when exposed to salt. Still other materials identify genomic locations related to salt tolerance. A process is needed to examine whether these clues point to the same candidate.

The rice research with OASIS is not an instance where AI alone decided the questions and carried out the entire process to the end. The research team guided the analysis by asking multiple rounds of questions. Following those questions, OASIS connected literature and biological data to narrow down candidates. The researchers' questions and judgments were embedded throughout the analysis process.

At the starting point of the analysis were 284 regulatory genes from Arabidopsis. Arabidopsis is a small plant widely used when studying life phenomena in plants. Regulatory genes are involved in regulating the activity of other genes or reactions inside the body. The research team and OASIS examined findings gained from Arabidopsis that could serve as clues for rice research.

The responses organized by OASIS in its initial literature search included ion balance, sodium and potassium movement, ABA signaling, and osmotic regulation. Osmotic regulation is the process of adjusting the balance related to water movement inside and outside cells. The balance of reactive oxygen species

(ROS), MAPK signaling, and calcium ion (Ca^{2+}) signaling were also examined. Reactive oxygen species are highly reactive oxygen-related substances, and MAPK signaling is a process in which proteins relay signals in sequence. Calcium ions are also used to relay signals inside cells. Organizing these responses provided the background for finding candidates, and it does not confirm that OsPP2a regulates all of these responses.

When connecting genes from different plants, researchers examine homology. Homology refers to a relationship where genes resemble each other because they originate from a common ancestor. Among these, the relationship of genes passed down into respective species as species diverge is called orthology. Such relationships serve as a starting point for exploring research findings from one plant in another plant. However, just because genes are similar does not allow one to conclude that they perform the same role under all conditions.

One of the clues used to examine candidates was TAP46 in Arabidopsis. TAP46 is related to signaling of the plant hormone abscisic acid, namely ABA. Hormones are signaling substances

that regulate reactions inside the body. By connecting such mechanisms found in the literature with similarities between genes, researchers can obtain grounds for deciding which genes to examine in rice.

To this, genomic data obtained from rice itself were connected. A genome is the entire set of genetic information possessed by an organism. GWAS is an analysis that compares genomic differences and traits across multiple individuals to find relevant locations. QTL refers to chromosome regions related to traits that show quantitative differences, such as salt tolerance. Such data allow researchers to see whether a candidate gene is located in a position associated with salt tolerance in rice.

A region analyzed by combining results from multiple QTL studies is called a meta-QTL. This is a method of examining positional information from one study by adding clues from other studies. However, the mere fact that a gene is located within that region cannot confirm it as the causal gene. This is because multiple candidates may be present together within a broad region.

An eQTL refers to a genomic location associated with differences in gene expression levels. The gene expression level is a value showing how much of the gene's information is read and used. When the location of a candidate gene overlaps with the location of an eQTL signal, it is called colocalization. The result that locations overlap also provides grounds for examining a candidate, but in itself it does not explain the entire mechanism of action.

OsPP2a is one of the candidates prioritized by examining these various clues together. Its identifier is LOC_Os12g04290. This gene was reviewed as a rice gene corresponding to TAP46 in Arabidopsis. There were also clues showing that it is located within a meta-QTL region related to rice salt tolerance and that its location overlaps with an eQTL signal.

Each clue answers a different question. The literature provides information on related biological reactions, and homology shows connections between plant species. Rice genomic data add reasons to examine that candidate within rice. Rather than

relying on any single clue alone, candidates were narrowed down by considering how these pieces of evidence connect.

Through this process, OASIS prioritized 5 candidates. This does not mean that the salt-tolerance regulatory functions of all 5 genes were confirmed. It was a stage of selecting candidates worthy of investigation through experiments. The researchers chose OsPP2a among them as the subject for verification. The fact that the involvement of OsPP2a in regulating rice salt tolerance had not been directly reported previously was another reason to investigate this candidate.

Such analysis goes further than merely finding and transferring already known information. It connects mechanisms known in other plants with data obtained from rice to generate new hypotheses to test. However, this does not mean that retrieval-augmented generation, or RAG, can only summarize. RAG is a method of retrieving external materials to generate answers, and it can also be used to assist reasoning based on those materials. The core of this case lies not in the name of a

specific technology, but in how evidence supporting the candidates was connected.

After selecting OsPP2a, a step was needed to confirm it in actual rice plants. Follow-up experiments by the researchers provided evidence that this gene is involved in rice salt stress responses. Hypotheses formed during the candidate selection stage must be distinguished from evidence obtained through experiments. A search that began with 284 Arabidopsis genes was narrowed down to a research question that could be asked directly in rice.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선거재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21). *Molecular Plant* | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.

<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

韩漾. (2026-10-20). AI智能体“绿洲”发布 助力植物抗逆研究. 中国农业科学院
(来源：农业基因组研究所，作者韩漾) .

<https://caas.cn/xwzx/kyhd/28164b3f0e7f4996a5749892288670bc.htm>

Gao, J., Kim, T.-H., Baek, D.-H., Lee, C.-J., Park, W.-G., Yun, S. D., & Kim, S.-M. (2026). Integrative genomic and transcriptomic dissection of salt tolerance for Japonica rice improvement. *Frontiers in Plant Science*, 16, Article 1751273. <https://doi.org/10.3389/fpls.2025.1751273>

Chapter 8: Confirming the AI-Generated Hypothesis with Actual Rice

Just because a computer selects a gene as a candidate for research does not mean the role of that gene has been uncovered. A gene is a segment of DNA that contains information necessary for an organism to grow and live. Analyzing data allows us to guess the role of a gene. To know whether that guess holds true in a living organism, it must be verified through experiments.

In the rice research with OASIS, these two processes were also distinct. OASIS connected research papers and biological data to narrow down genes likely to be related to the ability to tolerate salt. Human researchers guided the direction of the analysis and selected experimental targets. It was also the human researchers who actually grew the rice and measured its responses to salt.

The researchers did not simply ask a single question and wait for results. They guided the analysis by asking multiple questions. OASIS took 284 regulatory genes from Arabidopsis, a plant frequently used in research, as its starting point. It then connected similar genes in rice, mechanisms described in papers, and genetic data from rice. Through this process, it prioritized 5 candidates for study.

The researchers selected OsPP2a among them as the subject for experiments. The identifier distinguishing this gene in datasets is LOC_Os12g04290. Choosing the candidate was a judgment that determined the starting point of the experiments. After reviewing the evidence presented by OASIS, they decided to verify its role in actual rice plants.

Experiments that directly handle living plants, cells, or reagents are called wet-lab experiments. Unlike analyses comparing data inside a computer, this is a process of observing how living organisms respond under given conditions. When studying OsPP2a, it was also necessary to confirm whether the

connections found in the data appeared as actual changes in rice.

The researchers investigated two rice lines in which the function of the OsPP2a gene was lost. The two lines were independent Ospp2a mutant lines. A mutant line refers to a group of plants with changes in their genetic information. The researchers examined these rice plants alongside wild-type rice, which served as the comparison baseline.

In the experiment, young rice plants were subjected to a 150 mM NaCl condition. NaCl is sodium chloride, the main component of salt, and mM is a unit representing the concentration of a solution. The burden rice experiences due to salt under this condition is called salt stress. The researchers compared how wild-type rice and mutant rice responded to the same salt treatment.

The two Ospp2a mutant lines appeared more sensitive to salt than wild-type rice. Their ability to tolerate salt did not improve; rather, they were more severely affected by salt. Differences also

appeared after completing the salt treatment and returning them to normal conditions. The mutant rice had a lower survival rate during the recovery process.

Visible responses and survival rates alone make it difficult to know what changes occurred inside the rice plants. The researchers also measured the amount of ions contained within the plant bodies. Ions are small electrically charged particles. What they examined here were sodium ions, Na^+ , and potassium ions, K^+ .

Under salt stress conditions, more sodium ions accumulated in the mutant rice. The balance between sodium and potassium also changed. The Na^+/K^+ ratio, which is the ratio of sodium relative to potassium, increased. This result shows that the salt sensitivity of the mutant rice was accompanied by ion changes inside the plant bodies.

The researchers also investigated how much genes were expressed. Gene expression is the process in which genetic information is read and used for functioning. Measuring the

amount of RNA produced in that process allows one to examine the degree of expression. The RT-qPCR used by the researchers is a method of comparing the amount of RNA derived from specific genes.

In this analysis, the OsSOS1 and OsHKT1;5 genes, which are related to sodium transport, were examined. In Ospp2a mutant rice, changes appeared in the expression levels of both genes. Along with measurements of survival rate and ion content, this result demonstrated that differences also existed in gene expression. However, this result alone cannot be said to explain all the reasons why the two genes changed or to uncover the entire mechanism of action inside the cell.

The pieces of evidence obtained from the experiments show different aspects. Survival rate after salt treatment shows the impact experienced by the rice. Ion measurements reveal the status of sodium and potassium. Gene expression analysis allows observation of changes at the molecular level. Viewing these results together provides evidence that OsPP2a is involved in rice salt stress responses.

At this point, it is important to respect the scope of what experimental results state. The researchers compared two independent mutant lines under specified salt treatment conditions. The results are genetic and molecular evidence supporting the role of OsPP2a. They do not imply that the same effect would appear across all cultivation environments, nor do they mean that altering this gene alone can create salt-tolerant rice.

In this research, the AI analysis contributed to narrowing down what to test. Human researchers reviewed those suggestions to decide experimental targets, set up comparison conditions, and measured results. All of these steps were necessary for hypotheses derived from data to be connected with evidence obtained from actual rice.

The same distinction applies when interpreting experimental results. Instead of reading observations to fit explanations offered by AI, one must evaluate how well observed results support those explanations. The significance of the OsPP2a research does not lie in accepting the AI suggestion as the

correct answer as is. It lies in testing a testable hypothesis on actual plants and confirming the answer through multiple kinds of measurements.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]
<https://doi.org/10.1016/j.molp.2026.09.005>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21). *Molecular Plant* | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. <https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

韩漾. (2026-10-20). AI智能体“绿洲”发布 助力植物抗逆研究. 中国农业科学院（来源：农业基因组研究所，作者韩漾）. <https://caas.cn/xwzx/kyhd/28164b3f0e7f4996a5749892288670bc.htm>

농촌진흥청. (2026-08-20). 농촌진흥청, 벼 '염 스트레스' 내성 조절 유전자 기능 규명. 대한민국 정책브리핑. <https://m.korea.kr/briefing/pressReleaseView.do?gubun=pressRelease&newsId=156774969&pageIndex=1&repCode=B00006>

Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Sirkovic, P., Myaskovsky, A., Glowaty, G., Weissenberger, F., Orlandi, A., Popovici, D., Palepu, A., Rong, K., Tanno, R., Saab, K., Zhang, F., Blum, J., Carroll, A., Kulkarni, K., Tomashev, N., ... Natarajan, V. (2026). Accelerating scientific discovery with Co-Scientist. *Nature*, 655(8122), 487–496.

<https://doi.org/10.1038/s41586-026-10644-y>

Chapter 9: Humans Decide

What to Ask and What Conclusions to Draw

What should humans do when AI suggests candidates for research? It is not just a matter of taking the results and running experiments. Humans must decide what to find out, weigh which evidence to trust, and judge how far to draw conclusions. As AI participates more deeply in research, the weight of these choices also grows.

The ability to analyze data quickly does not mean the same thing as the ability to accurately explain nature. Even if a computer finds connections among multiple pieces of data, what role that connection plays in a living organism must be verified separately. Well-organized explanations or confidently written sentences cannot replace experimental evidence.

A hypothesis is an idea formed to explain an observed phenomenon. A hypothesis is worth checking, but the

possibility of being wrong also remains. Just because AI suggests many hypotheses does not mean that just as many facts are uncovered. The judgment of choosing which hypothesis to test and the process of interpreting test results must follow.

This is where research can get blocked. This is because suggestions increase rapidly, but the time, cost, and equipment for experiments are limited. A state where candidates to verify pile up and make it difficult for research to move forward is called a verification bottleneck. What is needed at this time is not just making the list of candidates longer. It is comparing the evidence and deciding which targets to test first.

A good research question sets the direction of exploration. You must decide what you want to find out, which organism to target, and under what conditions to check. If the goal is unclear, it is hard to tell what evidence is important even if you gather a lot of data. Researchers must set these standards before assigning work to AI and refine questions even during analysis.

In OASIS's rice research, the human research team guided the analysis by asking questions multiple times. OASIS connected data to prioritize 5 gene candidates. Among them, it was the research team that chose OsPP2a as the actual experimental target. This is an example showing how suggestions during the analysis process and decisions to proceed with research were connected.

When choosing candidates, looking only at the recommendation ranking is not enough. You must examine what data supports the candidate and whether there are inconsistencies among the data. If evidence obtained from different plants was connected, you must ask whether a known role in one plant would be the same in another plant. It is necessary to distinguish between possibilities pointed to by data and already confirmed facts.

Researchers must also check where the AI's answers came from. AI can present non-existent content as facts or misread the meaning of data. It can also miss important conditions in the process of connecting separate pieces of data. Rather than

trusting an answer because it is smooth, one must examine the supporting data and the scope that the data covers.

After that, an experiment suited to the hypothesis must be designed. Depending on what is used as a comparison standard, what conditions are given, and what is measured, the questions that can be answered will differ. In the OsPP2a research, mutant rice and wild-type rice were compared, and survival rates, ion contents, and gene expression were investigated. Researchers must interpret what answers these measurements provide to the initial questions.

Even if expected results appear, the scope of the conclusion must be weighed. Evidence obtained under set experimental conditions cannot be applied to all situations. The OsPP2a experiment provided evidence that this gene is involved in rice salt stress responses. Broadening that fact immediately into a conclusion that salt-tolerant rice can be created in all environments would go ahead of the evidence.

Even if AI has a feature that modifies work procedures based on experience, this principle does not change. SEEL, the self-evolving experience learning of OASIS, is a method that turns execution history and error-correction experiences into reusable procedural rules. Gathering verified experiences and checking whether a new hypothesis is biologically correct are two different things. A procedure that was helpful in the past does not guarantee this conclusion.

The responsibility of handling research results also remains. If data is cherry-picked only favorably, or results differing from expectations are hidden, trust in research collapses. If it is research that can affect humans and the environment, its risks must also be examined. The fact that AI participated in the analysis cannot be a reason for researchers to escape the responsibility for such judgment and explanation.

To judge whether an AI's suggestion is worth researching, deep questions are needed rather than quick answers. What do we learn if this hypothesis is confirmed? Can it be tested with the evidence we currently have? Is there something to learn even if

the result differs from expectations? With these questions, researchers decide where to use limited resources.

The role of humans does not need to stay only in searching for and organizing data. If that work can be reduced with the help of AI, room opens up to handle questions, experiments, and interpretations more carefully. However, such a change does not happen on its own. It is a change made possible when researchers understand the grounds for analysis and critically examine the results.

What must be left at the end of research is the boundary between what has been confirmed and what is not yet known. When that boundary is drawn clearly, other researchers can also review the results and formulate new questions. No matter how persuasive an answer AI provides, judging to what extent to accept that answer and explaining it remains up to the humans in charge of the research.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]

<https://doi.org/10.1016/j.molp.2026.09.005>

刘祯. (2026-08-24). 【新华社】从“工具”到“搭档” AI应怎样助力科研. 新华社 (中国科学院转载) .

https://www.cas.cn/cm/202608/t20260824_5118852.shtml

한국과학기술한림원. (2026-07-14). AI Scientist 시대, 인간 과학자의 자리.

YouTube. [동영상] https://www.youtube.com/watch?v=WxsUB_-eCYo

Bisht, H., Kumar, V., Jablonka, K. M., Mausam, & Krishnan, N. M. A. (2026). Agentic AI Scientists Are Not Built For Autonomous Scientific Discovery. *arXiv preprint (cs.AI)*. [프리프린트]

<https://arxiv.org/abs/2605.08956>

김영길, 이용주, 임수중, 송화전, 정문영, & 권오욱. (2026). AI Scientist 기술 개발 동향. *전자통신동향분석*, 41(5), 72–84.

<https://doi.org/10.22648/ETRI.2026.J.410507>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21).

Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.

<https://www.agis.org.cn/xwzx/kyjz/b0960d967c964346b50cb50c4886c209.htm>

Epilogue: Better Questions in the Face of More Answers

The questions faced by agriculture do not end inside the laboratory. Drought, heat, and soil salinity place burdens on crop growth. Finding out how crops endure in such environments is connected to efforts to stably produce food. Research asking about the role of small genes is also one part of that big question.

To find that answer, life sciences have accumulated a vast amount of data. They read the genetic information of living organisms, investigate which genes operate and how much, and record changes according to conditions. As data increases, the connections that can be found also increase. However, important connections do not automatically reveal themselves just because there is a lot of data.

AI helping scientific research is taking on new work in between. It examines separate data together and suggests explanations worth testing by humans. The significance of OASIS also lies in

not stopping at gathering data, but suggesting candidates that could lead to experiments. One of those candidates, through the research team's choice and experiments using actual rice, gained evidence that it is involved in salt responses.

The value of this evidence does not lie in having ended all questions. It lies in showing specifically what more can be asked. Results confirmed under set conditions become the basis for subsequent judgments. At the same time, conditions and explanations that have not yet been confirmed remain. Only when distinguishing between what is known and what is unknown can research move forward.

As AI produces more suggestions, choice becomes even more important. The time and cost available for experiments are limited. Rather than the number of suggestions, the ability to judge which questions to spend those resources on determines the direction of research. That judgment must take into account how solid the evidence is, the feasibility of testing it, and the value of the question itself.

Even in researching crops to respond to climate change, one must distinguish between possibilities and achievements. The result of understanding a gene's role does not immediately become a new crop variety or a solution for the agricultural field. Rather than saying AI bridged all those gaps, stating precisely which stage of exploration it helped reveals the value of the research much better.

Efforts to increase speed and an attitude of examining evidence meticulously can go together. While AI connects data, humans can ask whether that connection is valid. If experiments yield answers different from expectations, initial explanations can be revised. What is important is not defending the first answer given, but changing one's thoughts to match the evidence obtained.

Trust in research grows from verifiable evidence rather than grand promises. One must disclose what data was examined, under what conditions tests were conducted, and how far the results speak. Only then can others examine the results and

continue research. In science involving AI, this standard remains just as necessary.

The potential of science with AI is not yet a predetermined conclusion. Its use changes depending on what questions are asked and what evidence is demanded. Channeling the power to obtain more answers into better questions, and honestly revealing where answers reach and where they cannot, is the way to continue exploration.

References

Han, Y., Wei, H., Wang, X., Li, Y., Zhang, Z., Liu, B., Wen, S., Pan, N., He, H., Qian, Q., & Shang, L. (2026). OASIS, a self-evolving AI scientist that integrates omics data and literature knowledge for plant stress research. *Molecular Plant*. [온라인 선게재; 권·호·쪽 미확인]
<https://doi.org/10.1016/j.molp.2026.09.005>

한국과학기술한림원. (2026-07-14). AI Scientist 시대, 인간 과학자의 자리. YouTube. [동영상] https://www.youtube.com/watch?v=WxsUB_-eCYo

刘祯. (2026-08-24). 【新华社】从“工具”到“搭档” AI应怎样助力科研. 新华社 (中国科学院转载) .
https://www.cas.cn/cm/202608/t20260824_5118852.shtml

윤시아. (2026-10-12). 과학의 다음 질문, 이제 AI가 고른다. MIT 테크놀로지 리뷰 코리아 (MIT Technology Review Korea).

<https://www.technologyreview.kr/과학의-다음-질문-이제-ai가-고른다/>

김영길, 이용주, 임수중, 송화전, 정문영, & 권오욱. (2026). AI Scientist 기술 개발 동향. 전자통신동향분석, 41(5), 72-84.

<https://doi.org/10.22648/ETRI.2026.J.410507>

中国农业科学院深圳农业基因组研究所（大鹏湾实验室）. (2026-09-21).

Molecular Plant | 基因组所商连光团队开发自进化AI科学家“绿洲”助力作物耐盐碱研究. 中国农业科学院深圳农业基因组研究所（大鹏湾实验室）.

<https://www.agis.org.cn/xwzx/kjyz/b0960d967c964346b50cb50c4886c209.htm>

© 2026 Kim Kyungjin. All rights reserved.

KIMKJ.COM