様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　４ 年　 ６ 月 １６ 日現在

機関番号：８２４０１
研究種目：若手研究
研究期間：2019～2021
課題番号：１９Ｋ１６３９１

研究課題名（和文）Systems biology approach to unravel biosynthesis and evolution of camptothecin, a potent anti-cancer natural product

研究課題名（英文）Systems biology approach to unravel biosynthesis and evolution of camptothecin, a potent anti-cancer natural product

研究代表者
Ｒａｉ　Ａｍｉｔ（Rai, Amit）

国立研究開発法人理化学研究所・環境資源科学研究センター・研究員

研究者番号：６０７６０５３５

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：ここでは、Ophiorrhiza pumila, Ophiorrhiza japonica, およびNothapodytesfoetidaの高品質なゲノムアセンブリを報告します。 比較ゲノム解析は、カンプトテシン（CPT）生合成の出現を導いた重要な酵素としてストリクトシジンシンターゼ（STR）を示しました。 進化に向けた代謝遺伝子クラスターの重要性と、特殊な代謝物を合成する能力を維持することの重要性を示しました。 CPT生合成から推定代謝中間体を同定しました。 最後に、生合成経路に関連する候補遺伝子を特定するために、STRノックアウトラインを確立しました。

研究成果の学術的意義や社会的意義
This study have produced three high quality genome assemblies of important medicinal plants to explore evolutionary basis of camptothecin biosynthesis, one of the most potent anti-cancer drug in the world. Results and resources will be useful to establish sustainable source of camptothecin.

研究成果の概要（英文）：During this study, we finalized high-quality genome assemblies of Ophiorrhiza pumila (genome size 440Mb), Ophiorrhiza japonica (genome size 800Mb), and Nothapodytes foetida (genome size 9.79Mb). We also used complete carbon labeling and nitrogen labeling of the Ophiorrhiza pumila metabolome to establish a high-quality metabolome resource to discover unknown intermediates of camptothecin biosynthesis.
Comparative genome analysis showed strictosidine synthase (STR) as the key enzyme that guided the emergence of camptothecin biosynthesis, loss of which resulted in losing the ability of its biosynthesis in coffee. We showed the importance of metabolic gene clusters towards the evolution and retaining the ability to synthesize specialized metabolites in plants. We also identified putative metabolite intermediates from camptothecin biosynthesis. Finally, we established STR knockout lines, which block camptothecin biosynthesis to identify candidate genes associated with biosynthesis pathways.

研究分野：環境および天然医薬資源学関連

キーワード：Camptothecin　Medicinal plants　MIAs　Whole genome sequencing　Phytogenomics　Ophiorrhiza pumila　Comparative genomics　Systems biology

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

The monoterpene indole alkaloids (MIA) are a large group of plant-derived specialized metabolites, many of which have valuable pharmaceutical or biological activity. The diverse chemical structures in this metabolite class originate from strictosidine, which undergoes a series of oxidation, reduction, and ring rearrangement reactions leading to synthesizing metabolites such as vincosamide, catharanthine, vinblastine, tabersonine, vindoline, ajmalicine and camptothecin (CPT) among others. CPT, one of the most potent drug precursors for cancer treatment, is in huge demand with a limited supply of the plant's extracts. Two of the CPT-analogs, topotecan, and irinotecan, are used extensively for the treatment of metastatic colorectal cancer, ovarian cancer, cervical cancer, and small cell lung cancer, with combined sales reaching 20 billion US dollars in the year 2017. Against a world demand of 3000 kg, only 600 kg of CPT was made available in 2015. There is an urgent need to establish the required resources to undertake studies on identifying candidate genes involved in the CPT biosynthesis and to establish a sustainable resource for CPT production through means including synthetic biology. Prior to this study, the available genomics and metabolomics resources were limited to exploring the biosynthesis of specialized metabolites in plants focused on MIAs biosynthesis in general and CPT biosynthesis in specific. Further, little was known about the mechanism that resulted in the emergence of MIAs biosynthesis.

２．研究の目的

The purpose of this study is to identify involved genes and metabolite-intermediates associated with CPT biosynthesis. The limited number of CPT-producing plants, low plant tissue yields per gram, and associated environmental concerns are bottlenecks vis-à-vis increased demand. With an estimated 20 million new cancer cases globally by 2025 and an economic burden estimated at $1.16 trillion in 2010, meeting the increasing demands for CPT and other anticancer MIAs has become a daunting challenge and requires immediate attention. It is imperative and urgent to address this demand and supply gap, which is also related to the national security of countries such as Japan, which has seen a rise in cancer patients with the aging population. Through this study, we also aimed to address the fundamental question as; "Why do plants make diverse chemical ingredients with bioactive properties?". Understanding MIAs' evolution and biosynthesis is essential for building sustainable alternate production platforms to facilitate access to these lifesaving compounds.

３．研究の方法

In this study, we used a multi-omics and systems biology approach together with phylogenomic to identify candidate genes and intermediates associated with CPT biosynthesis. Next-generation sequencing technology and hybrid genome assembly strategy were used to establish high-quality genome assemblies of *Ophiorrhiza pumila* and *Nothapodytes foetida*, two of the CPT-producing plants. We also established whole-genome assembly of non-CPT producing plant, *Ophiorrhiza japonica*, and comparative genome analysis was used to identify genes essential for CPT biosynthesis. We next established metabolome resources for *O. pumila* and used isotope labeling experimental setup to identify metabolite intermediates of CPT biosynthesis pathways. We used integrative omics analysis, phylogenomic, comparative genome analysis, and evolutionary analysis to identify candidate genes and potential metabolite intermediates of the CPT biosynthesis pathway. Our approach also allowed us to identify events that resulted in the evolution of MIAs. We also performed gene cluster analysis and combined synteny with associated gene clusters across four MIA-producing plants as a strategy was used to further assign confidence scores on genes to select them for further characterization.

４．研究成果

One of the major achievements of this study is the high-quality public resources that were generated through this study including valuable datasets that are undergoing analysis at the time of preparing this report. Specially, whole genome assembly of *N. foetida* and *O. japonica* was assembled and finalized towards the end of the financial year of 2022, and thereafter, gene prediction and annotation are currently on-going at the time of preparing this report. Major results, and conclusions emerging from this study is summarized below-

(1) **Comparative genome analysis revealed emergence of MIAs biosynthesis centered around strictosidine synthase**- In this study, we completed and characterized the whole-genome assembly of *O. pumila* and performed comparative genome analysis with three other MIA-producing plants, including *Catharanthus roseus*, *Camptotheca acuminata* (which also produces CPT), and *Gelsemium sempervirens*. Compared to all medicinal plant genomes published to date, *O. pumila* genome assembly is the best and most contiguous plant genome reported[1]. We also showed the importance of experimental validation and suggested its inclusion as the key to accurately representing the plant genomes. Comparative genome analysis showed the emergence of two copies of strictosidine synthase (STR), an enzyme that determines the formation of strictosidine and the first step toward MIAs biosynthesis from secologanins. The STR enzyme remains conserved across wider lineages of plant species, suggesting its essential role in plants. At the same time, we observed the emergence of STR orthologous genes, which were retained only in MIA-producing plants while lost across other plant lineages. Gene cluster analysis showed STR and tryptophan decarboxylase exists within a metabolic gene cluster and are conserved across all MIA-producing plants except *C. acuminata*. We reported the parallel biosynthesis pathway towards strictosidine biosynthesis and thus explained the absence of the gene cluster in major CPT-producing plants. When we compared the *O. pumila* genome with *Coffee canephora*, another plant from the Rubiaceae family, we observed the conserved gene cluster but missing the key STR enzyme. Using phylogenomic analysis, we confirmed and established that the emergence of MIAs was centered around STR (Fig1). The evolution of key enzymes associated with MIAs and CPT biosynthesis emerged post evolution of STR. Further, we showed the importance of combining phylogenomics with gene cluster analysis to discover candidate genes associated with specialized metabolites biosynthesis. <u>The entire dataset, including *O. pumila* genome assembly, phylogenomic analysis, MIAs associated gene clusters, potential candidate genes associated with MIA biosynthesis, and candidate transcription factors associated with CPT biosynthesis, have been made publicly available to expand the research on the functional characterization of genes involved in the biosynthesis pathways</u>.

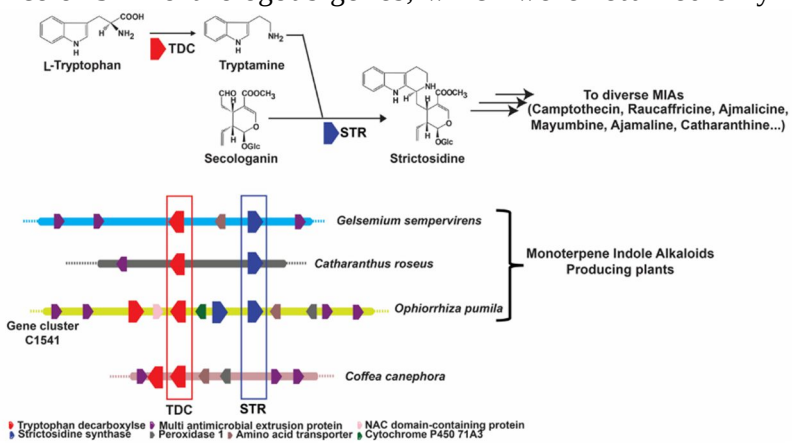

Fig 1. Loss of strictosidine synthase enzyme from the gene-cluster lost the evolution of MIAs in Coffee.

(2) **Integrative omics to identify association between putative intermediates and enzymes from CPT biosynthesis pathways**- Expression analysis showed that genes associated with secoiridoids biosynthesis were co-expressed and showed the highest expression in the root and hairy root of *O. pumila*. These tissues also showed the highest levels of CPT and its potential intermediates. Using integrative omics analysis, we identified potential candidate genes that were highly correlated with metabolites accumulation across multiple tissues[1]. Several of the candidate genes are being pursued at this moment for functional characteristics.

(3) **High-quality metabolome resources for *O. pumila***- We used complete carbon and complete nitrogen isotope labeling approach to identify putative metabolite

intermediates of the CPT biosynthesis pathway. Previously, we published a metabolome database for 12 plant species, including *O. pumila* using a computational metabolomics approach2. We further aided our metabolome database by performing metabolome analysis for *O. pumila* hairy roots, the metabolome of which was completely N15 isotope-labeled using. Combining our datasets from metabolome analysis of total carbon and nitrogen labeling allowed us to identify 273 nitrogen-containing metabolites, the majority of which include two nitrogen atoms, thus including potential intermediates of CPT biosynthesis pathways1. The updated metabolome database is now publicly available and the only resource to explore CPT metabolite intermediates. We also performed a time-series experiment of feeding both nitrogen atoms in tryptophan as isotope-labeled and fed to *O. pumila* hairy root. We observed complete isotope labeling of both nitrogen atoms of CPT five days post-feeding the isotope-labeled substrate. The dataset is currently under investigation, and we hope to identify and confirm intermediates of CPT and complete the biosynthesis pathways.

(4) **Time-series multi-omics analysis of *O. pumila* hairy roots treated with four phytohormones**- Phytohormones are known to alter the original state of the metabolome and often result in the induction of specialized metabolites biosynthesis in plants. Jasmonic acid, ethylene, salicylic acid, and ABA have previously been reported to include increased accumulation of CPT. We treated three weeks old *O. pumila* hairy roots with jasmonic acid, ethylene, salicylic acid, and ABA, and samples were collected at 0hr, 2hr, 12hr, 24hr, 48hr, 72hr, 96hr, and 120hr. RNA-seq analysis was performed for these four treatments, 8-time points, two conditions (treated and not treated), and three biological replicates. Our results showed that genes associated with secoiridoids biosynthesis pathways were highly correlated. STR enzymes showed a very high increase in the expression level within 2hours of all four phytohormones treatments. In contrast, putative enzymes assigned based on comparative genome analysis and gene cluster analysis showed increased expression after 48 hours of phytohormone treatment. We performed WGCNA analysis and grouped enzymes in specific groups based on the expression pattern. Metabolome analysis for the same set of samples is ongoing. Once the metabolome results are acquired, integrative omics analysis will be performed to establish metabolome levels-gene expression association. This is an unpublished result, and analysis is still ongoing.

(5) **Whole genome assembly of *Nothapodytes foetida* and *Ophiorrhiza japonica*-** Among major plant species used for the commercial extraction of CPT includes *C. acuminata* and *N. foetida*. We firstly estimated genome assembly of *N. foetida*, and flow cytometer-based analysis estimated genome size as 9.79Gb. Compared to *O. pumila* and *C. acuminata*, *N. foetida* estimated genome size was thus over 20x, suggesting that the genome must have undergone massive expansion through whole-genome duplications and transposable elements. We next performed karyotyping for N. foetida, and we observed 56 chromosomes with a possibility of a diploid genome of 28x2 or a tetraploid genome of 14x4. We are now investigating the *N. foetida* ploidy level. Further, we performed whole-genome sequencing using PacBio HiFi sequencing technology and acquired 35x genome coverage for N. foetida. We also completed HiC library preparation and acquired 90x genome coverage worth sequencing using Illumina Novo-seq paired-end sequencing. We are now finalizing genome assemblies and annotation of *N. foetida*. Our preliminary scaffolding showed 28 chromosomes with a very high level of heterozygosity. We also finalized the genome assembly of *O. japonica*. *O. japonica* is a close relative of *O. pumila*, which does produce several CPT intermediates but not CPT. Experiments have shown that *O. japonica* cannot survive the CPT toxicity. Therefore, the genome assembly of *O. japonica* is an excellent resource for studying the evolution of CPT biosynthesis. We established chromosome-scale genome assembly of *O. japonica*, including 11 pseudomolecules with contig N50 as 4.8Mb. Gene prediction and genome characterization are ongoing. We hope to make the genome assemblies for both these plant species public shortly.

References-

1. **A. Rai**, H. Hirakawa, R. Nakabayashi, S. Kikuchi, K. Hayashi, M. Rai, H. Tsugawa, T. Nakaya, T. Mori, H. Nagasaki, R. Fukushi, Y. Kusuya, H. Takahashi, H. Uchiyama, A. Toyoda, S. Hikosaka, E. Goto, K. Saito and M. Yamazaki, Nat. Commun., 2021, 12, 405

2. H. Tsugawa, R. Nakabayashi, T. Mori, Y. Yamada, M. Takahashi, **A. Rai**, R. Sugiyama, H. Yamamoto, T. Nakaya, M. Yamazaki, R. Kooke, J. A. Bac-Molenaar, N. Oztolan-Erol, J. J. B. Keurentjes, M. Arita and K. Saito, Nat. Methods, 2019, 16, 295–298

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　2件／うちオープンアクセス　2件）

１．著者名 Rai Amit、Hirakawa Hideki、Nakabayashi Ryo、Kikuchi Shinji、Hayashi Koki、Rai Megha、Tsugawa Hiroshi、Nakaya Taiki、Mori Tetsuya、Nagasaki Hideki、Fukushi Runa、Kusuya Yoko、Takahashi Hiroki、Uchiyama Hiroshi、Toyoda Atsushi、Hikosaka Shoko、Goto Eiji、Saito Kazuki、Yamazaki Mami	４．巻 12
２．論文標題 Chromosome-level genome assembly of Ophiorrhiza pumila reveals the evolution of camptothecin biosynthesis	５．発行年 2021年
３．雑誌名 Nature Communications	６．最初と最後の頁 1-19
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41467-020-20508-2	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

１．著者名 Rai Amit、Rai Megha、Kamochi Hidetaka、Mori Tetsuya、Nakabayashi Ryo、Nakamura Michimi、Suzuki Hideyuki、Saito Kazuki、Yamazaki Mami	４．巻 27
２．論文標題 Multiomics-based characterization of specialized metabolites biosynthesis in Cornus Officinalis	５．発行年 2020年
３．雑誌名 DNA Research	６．最初と最後の頁 1-15
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/dnares/dsaa009	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

〔学会発表〕　計5件（うち招待講演　2件／うち国際学会　5件）

１．発表者名 Amit Rai, Hideki Hirakawa, Ryo Nakabayashi, Shinji Kikuchi, Koki Hayashi, Megha Rai, Hiroshi Tsugawa Tetsuya Mori, Mami Yamazaki, and Kazuki Saito
２．発表標題 Genome-guided multi-omics approach to explore evolution of specialized metabolites in plants
３．学会等名 International plant systems biology, EMBO workshop, April 26-27（招待講演）（国際学会）
４．発表年 2020年～2021年

１．発表者名 Amit Rai
２．発表標題 Integrated Omics for Hypothesis Generation
３．学会等名 The 69th Annual Conference on Mass Spectrometry, Japan, May 19-21（招待講演）（国際学会）
４．発表年 2020年～2021年

１．発表者名
Koki Hayashi, Amit Rai, Kazuki Saito, Mami Yamazaki

２．発表標題
Predicting genes related to the biosynthesis of Camptothecin by integrating promoter similarity score with gene co-expression analysis

３．学会等名
The 61th Annual Meeting of the Japanese Society of Plant Physiologists, Osaka（国際学会）

４．発表年
2019年～2020年

１．発表者名
Koki Hayashi, Amit Rai, Kazuki Saito, Mami Yamazaki

２．発表標題
Integration of cis-element motif analysis with co-expression analysis to predict genes related to Camptothecin biosynthesis

３．学会等名
Genome Support Meeting, PAG 2019, Nagoya, Japan（国際学会）

４．発表年
2019年～2020年

１．発表者名
Koki Hayashi, Amit Rai, Kazuki Saito, Mami Yamazaki

２．発表標題
シス調節配列のモチーフ解析と共発現解析の統合による カンプトテシン生合成候補遺伝子の探索

３．学会等名
The 56th Plant Chemistry Symposium（国際学会）

４．発表年
2019年～2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕
Ophiorrhiza pumila Genome DataBase
http://pumila.kazusa.or.jp/

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関