



除虫菊類のゲノムや有用物質生合成酵素の生物情報 基盤構築に関する研究

山城, 敬範

(Degree)

博士 (学術)

(Date of Degree)

2023-03-25

(Date of Publication)

2024-03-01

(Resource Type)

doctoral thesis

(Report Number)

甲第8652号

(URL)

<https://hdl.handle.net/20.500.14094/0100482400>

※ 当コンテンツは神戸大学の学術成果です。無断複製・不正使用等を禁じます。著作権法で認められている範囲内で、適切にご利用ください。



博 士 論 文

除虫菊類のゲノムや有用物質生合成酵素の
生物情報基盤構築に関する研究

2023年1月

神戸大学大学院工学研究科

山城 敬範

目次

第 1 章		
緒言	2
第 2 章	<i>T. cineraiifolium</i> ゲノムの解読	
1 節	7
2 節	11
3 節	16
4 節	45
引用文献	51
第 3 章	<i>T. coccineum</i> ゲノムの解読	
1 節	60
2 節	61
3 節	69
引用文献	100
第 4 章	TciGLIP のトランスフェラーゼ活性に重要な部位の特定	
1 節	105
2 節	107
3 節	111
引用文献	124
第 5 章		
総括	127
謝辞	132

第1章 緒言

蚊は感染症を媒介することから、世界で最も人間を殺す生物として知られ、死者は年間 83 万人にも及ぶ。特にマラリアは、2020 年の発表で年間 2 億人以上が感染、60 万人以上の死者数を出し、対策費用だけでも約 30 億米ドルが投じられており、現在でも根絶が困難な感染症であることから、その社会的経済的損失は甚大なものとなっている。多くの昆虫媒介性感染症は、ワクチンなどの予防手段が乏しく、虫に刺されないことが主な対策法となっており、殺虫剤や忌避剤の使用が有効な手段となっている。近年の気候変動の影響もあり、感染症を媒介する昆虫の生息分布域が拡大していることから、殺虫剤のニーズは世界中で伸長している。1902 年に日本で開発された蚊取り線香を皮切りに、今日まで様々な用途、対象に合わせた剤形の殺虫剤が開発されているが、特に、最もシンプルな剤形である蚊取り線香は安価で火があれば使える簡便さから、電力供給の不安定な途上国をはじめとする国々で現在も主要に使用されている。

かつて世界中で使用されていた殺虫成分は、DDT（ジクロロジフェニルトリクロロエタン）や BHC（ベンゼンヘキサクロライド）といった有機塩素系化合物であったが、現在は天然成分のピレトリン類やその合成類縁体であるピレスロイドと呼ばれる化合物が使用されており、特に安全性が重要視される家庭用の殺虫剤においては、そのほとんどでピレトリン類もしくはピレスロイドが使用されている。ピレトリン類は除虫菊（和名：シロバナムシヨケギク、学名：*Tanacetum cinerariifolium*）という白いダイジー様の多年生キク科植物（図 1-1 左写真）の主に子房から抽出される種特異的な二次代謝産物である。同属植物のアカバナムシヨケギク（*Tanacetum coccineum*）（図 1-1 右写真）もピレトリン類を生産するが、除虫菊の方がその含有量のはるかに多いことから、現在では *T. coccineum* は観賞用途のみで栽培されている。除虫菊を乾燥させた粉は中世の時代からその殺虫効果が知られており、民間薬として使用されていたが、その殺虫成分の同定は 20 世紀に入ってからである。ピレトリン類は、1. 広汎な昆虫への高い即効性、2. 温血動物への毒性が低い、3. 光や酸素

で分解されやすく環境負荷が少ない、という特長を有しており、特に選択毒性の高さは有機塩素系化合物や有機リン系化合物などの他の殺虫成分になく、安全性という観点から他に類をみない殺虫成分として重宝されている。1958年にピレトリン類の化学構造が完全に解明されて以来、何十種類もの合成類縁体のピレスロイドの開発が進み、その物性や殺虫スペクトルに適合した殺虫剤の開発も同時に進めてられてきた。ピレスロイドは合成が容易で安価という点から、天然成分のピレトリン類に代わって主要な殺虫成分として使われてきたが、その乱用によるピレスロイド抵抗性害虫の出現が問題となっている。一方で、ピレトリン類を含む除虫菊抽出物はピレスロイド抵抗性害虫への効果が確認されており、また、市場的な天然由来成分への回帰傾向もあることから、近年除虫菊自体の利用価値が見直されている。さらに、ピレトリン類は構造が不安定で化学合成が難しく、現代でも除虫菊の植物体を刈り取り、溶媒で抽出するという原始的な方法がとられている。ピレトリン類は開花に伴ってその生合成量を増加させ、その大部分は除虫菊の子房に蓄積される。最大の収量を得るためには、播種から開花まで1年以上もの期間を要するため効率が悪く、現在は中国やオーストラリアといった国の広大な敷地を使用した大規模栽培でその需要をカバーしている。しかし、今後さらに高まる需要を満たすためには物量的な手法だけでなく、代謝工学的な手法を利用した生産性の改善を行い、さらなる効率化が求められる。

ピレトリン類はピレトリン I、ピレトリン II、ジャスモリン I、ジャスモリン II、シネリン I およびシネリン II という 6 種のエステル化合物の混合物であり、これらは菊酸とピレトリン酸という 2 種類の酸、ピレスロロン、ジャスモロンおよびシネロロンという 3 種のアルコールのそれぞれの組み合わせから成る (図 1-2)。ピレトリン類の生合成機構は完全に解明されておらず、前駆物質の異種発現系はいくつか報告されているが、完全な生合成経路の構築は未だなされていない。そのため、ピレトリン類の生合成機構の完全解明は現在、各国の研究グループが取り組むテーマとなっている。さらに、除虫菊が傷害を受けた際に放出する揮発性有機化合物 (volatile organic compounds、VOCs) を近隣の無傷の除虫菊が

感知することで、ピレトリン類生合成を増大させる現象が知られており、ピレトリン類の生産量を調節する技術としての応用が期待されているが、この VOCs を介したピレトリン類生産量制御の分子機構については明らかになっていない。以上の背景から、本研究では、ピレトリン類の生合成機構の解明をはじめとする問題の解決の端緒として、除虫菊のゲノム解読を主軸とした解析を行い、将来的なピレトリン類の増産や除虫菊という植物の有効活用を視野に入れた遺伝的情報基盤を構築することを主目的とした。

本論文は 5 章からなり、第 2、3 章においては、除虫菊 *T. cinerariifolium* とアカバナムシヨケギク *T. coccineum* のゲノム解析を行い、第 4 章ではピレトリン類の生合成酵素の一つである TeiGLIP (*T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ) の活性と分子の進化・多様性について *in silico* 解析を行った。第 2 章において、*T. cinerariifolium* のドラフトゲノム配列を作成し、ゲノム内の構成、遺伝子の解析を行うことで、*T. cinerariifolium* の遺伝的特徴を解析し、また、ピレトリン類の生合成機構の解明の手掛かりとなる遺伝子を見出すことができた。第 3 章では *T. cinerariifolium* と同様に、ピレトリン類生産能を有するがその量が少ない *T. coccineum* のドラフトゲノムを作成し、*T. cinerariifolium* と比較することで、これら 2 種の植物のピレトリン類生合成量の違いを解明するための手掛かりや、生体防御機構の違いについて解明した。第 4 章では、アミノ酸配列アラインメントの two entropy 解析と、タンパク質のモデル予測およびドッキングシミュレーションツールを使用した *in silico* 解析によって、TeiGLIP がトランスフェラーゼ活性を獲得した鍵となる部位を特定した。

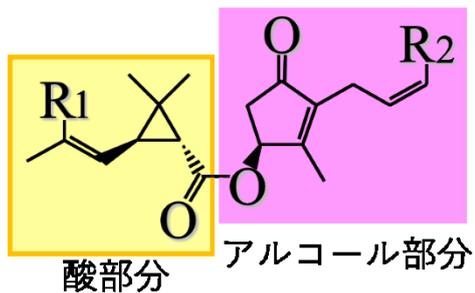


Tanacetum cinerariifolium



Tanacetum coccineum

図 1-1 除虫菊（和名：シロバナムシヨケギク、学名：*Tanacetum cinerariifolium*）および
アカバナムシヨケギク（*Tanacetum coccineum*）の花の写真



	R1	R2
ピレトリン I	CH ₃	CH=CH ₂
ピレトリン II	COOCH ₃	CH=CH ₂
ジャスモリン I	CH ₃	CH ₂ CH ₃
ジャスモリン II	COOCH ₃	CH ₂ CH ₃
シネリン I	CH ₃	CH ₃
シネリン II	COOCH ₃	CH ₃

【酸部分】
R₁: CH₃ (菊酸)
COOCH₃ (ピレトリン酸)

【アルコール部分】
R₂: CH=CH₂ (ピレスロロン)
CH₂CH₃ (ジャスモロン)
CH₃ (シネロロン)

図 1-2 ピレトリン類を構成する 6 種のエステル化合物の構造式

第2章 *T. cinerariifolium* ゲノムの解読

第1節 序論

ピレトリン類の生合成に関連する酵素は、今日までにいくつか同定されている(図 2-1)。ピレトリン類のアルコール部分はオキシリピン経路によって生成しており、リノレン酸が *T. cinerariifolium* 13-リポキシゲナーゼ (TciLOX1) によって 13-ヒドロペルオキシリノレン酸に変換され (Ramirez, Yang, et al. 2013)、その後、アレンオキシドシクラーゼ (AOS)、アレンオキシド合成酵素 (AOC)、3-オキソ-2-(2-ペンテニル)-シクロペンタン-1-オクタン酸レダクターゼ 3 (OPR3)、および別の未知の酵素によってジャスモンに変換されると考えられる。ジャスモンは、*T. cinerariifolium* ジャスモンヒドロキシラーゼ (TciJMH) によってジャスモロンに変換される (Li, Zhou, and Pichersky 2018)。ピレスロロンはピレスロロン合成酵素 (TciPYS) によりジャスモロンから誘導されていることが知られている (Li et al. 2019)が、シネロロンへの変換を担う機序および酵素は未だ解明されていない。ピレトリン類の酸部分の合成に関して、非メバロン酸経路を経て生成したジメチルアリルニリン酸 2 分子が *T. cinerariifolium* クリサンテミルニリン酸合成酵素 (TciCDS) によってクリサンテミルニリン酸に変換される (Rivera et al. 2001)。その後、クリサンテミルニリン酸は *T. cinerariifolium* アルコールデヒドロゲナーゼ 2 (TciADH2) および *T. cinerariifolium* アルデヒドデヒドロゲナーゼ 1 (TciALDH1) によって chrysanthemic acid (菊酸) に酸化される (Xu et al. 2018)。もう一方の酸部分である pyrethric acid (ピレトリン酸) は、*T. cinerariifolium* クリサンテモール 10-ヒドロキシラーゼ (TciCHH) および *T. cinerariifolium* 10-カルボキシクリサンテム酸 10-メチルトランスフェラーゼ (TciCCMT) によって生成される (Xu et al. 2019)。最終的に、chrysanthemoyl-CoA (CoA 活性型菊酸) およびピレスロロンが、*T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ (TciGLIP) によってピレトリン I にエステル化される。TciGLIP は、ピレトリン I のエステル化を触媒する酵素として単離・同定されている (Kikuta et al. 2012)が、ピレトリン II、

ジャスモリン I、ジャスモリン II、シネリン I、およびシネリン II の合成を触媒する酵素は明確に同定されていない。このような背景から、*T. cinerariifolium* のゲノム配列の解明により、ピレトリン類の全生合成経路および進化過程を解明するための決定的な手掛かりが得られることが期待できる。ゲノム配列はさらに、遺伝子組換え *T. cinerariifolium* を創出するための分子的ツールや、化学的に合成することが困難なピレトリン類を効率的かつ安定的に産生するための基盤データとして利用することも期待できる。

過去数年間でキクタニギク (*Chrysanthemum seticuspe*) (Hirakawa et al. 2019)、クソニンジン (*Artemisia annua*) (Shen et al. 2018) およびヒマワリ (*Helianthus annuus*) (Badouin et al. 2017) など、キク科の種々の植物のゲノムが報告されているが、ピレトリン類を産生するキク科植物のゲノム配列は決定されていない (Katsuda 2012)。*T. cinerariifolium* のゲノムサイズはフローサイトメトリー法によって約 7.1 Gb であると推定され (Siljak-Yakovlev et al. 2010)、他のキク科のゲノム (上記例の *C. seticuspe* : 2.72 Gb、*A. annua* : 1.74 Gb、*H. annuus* : 3.6 Gb) の 2 倍以上であることが示唆された。このような大きなゲノムのアセンブリは高度の反復配列、パラログ遺伝子の重複、およびその生物自体のヘテロ接合性によって依然として難しいままである。本章において、HiSeq 4000 および HiSeq X を用いた超並列シーケンシングから得られた *T. cinerariifolium* のドラフトゲノム配列を報告する。さらに、アセンブルしたゲノム配列に含まれる反復配列の含有率およびピレトリン類生合成関連遺伝子の重複についても解析した。

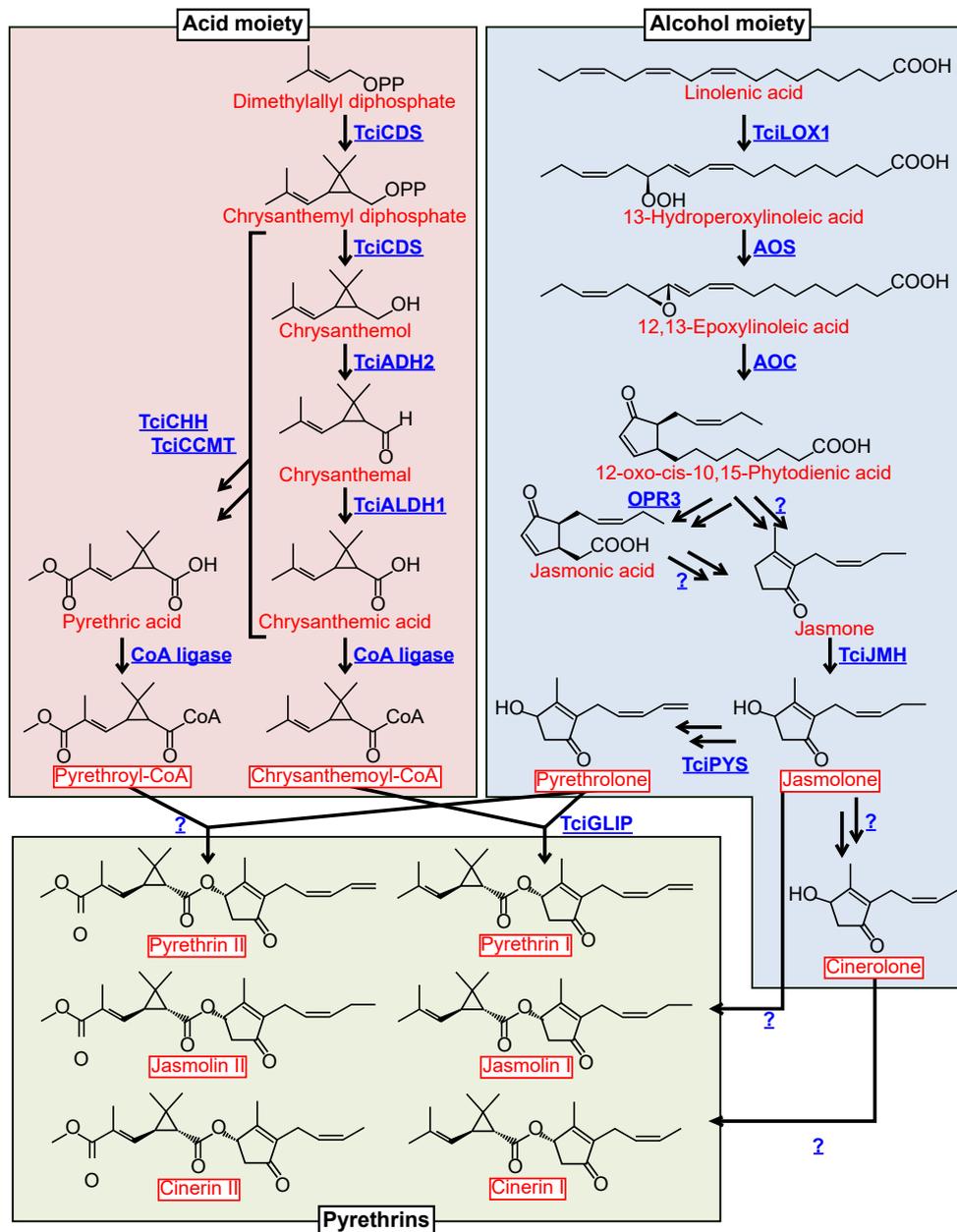


図 2-1 ピレトリン類の生合成経路

酸部分である chrysanthemic acid (菊酸) および pyrethric acid (ピレトリン酸) は、ジメチルアリルニリン酸から TciCDS、TciADH2、TciALDH1、TciCHH および TciCCMT によって合成される。アルコール部分であるピレスロロン、ジャスモロン、およびシネロロンは、リノレン酸から TciJMH、TciPYS および未知の酵素を含め、オキシリピン経路に関連する酵素によって合成される。最終的に、chrysanthemoyl-CoA (CoA 活性型菊酸) およびピレスロロンが TciGLIP によってピレトリン I にエステル化される。化合物および酵素はそれ

ぞれ赤色および青色で記載した。未同定の酵素はクエスチョンマークで示した。AOC: アレンオキシド合成酵素、AOS: アレンオキシドシクラーゼ、OPR3: 3-オキソ-2-(2-ペンテニル)-シクロペンタン-1-オクタン酸レダクターゼ 3、TciADH2: *T. cinerariifolium* アルコールデヒドロゲナーゼ 2、TciALDH1: *T. cinerariifolium* アルデヒドデヒドロゲナーゼ 1、TciCCMT: *T. cinerariifolium* 10-カルボキシクリサンテム酸 10-メチルトランスフェラーゼ、TciCDS: *T. cinerariifolium* クリサンテミルニリン酸合成酵素、TciCHH: *T. cinerariifolium* クリサンテモール 10-ヒドロキシラーゼ、TciGLIP: *T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ、TciJMH: *T. cinerariifolium* ジャスモンヒドロキシラーゼ、TciLOX1: *T. cinerariifolium* 13-リポキシゲナーゼ、TciPYS: *T. cinerariifolium* ピレスロロン合成酵素

第2節 実験の部

1. 植物材料およびゲノムシーケンシング

尾道市立美術館（広島県尾道市西土堂町）の敷地内において、野生型 *T. cinerariifolium* を野生条件下で成長させ、収穫し、植物サンプルとした。DNeasy Plant Mini Kit (Qiagen) を使用し、付属のプロトコルに従って、葉と茎からゲノム DNA を抽出した。抽出した DNA から、TruSeq DNA PCR-Free キット (Illumina) を用いてショートインサートライブラリ（以下、PE）、Nextera Mate-Pair Sample Prep Kit (Illumina) を用いて 3 つの異なるインサートサイズ（3 Kb、5 Kb、および 8 Kb）のメイトペアライブラリ（それぞれ MP-3kb、MP-5Kb および MP-8Kb）をそれぞれ構築した。その後、HiSeq X および HiSeq 4000 (Illumina) 機器をそれぞれ用いて、PE および MP ライブラリを 150 bp × 2 サイクルおよび 100 bp × 2 サイクルのペアエンドシーケンシングにかけた。

2. ゲノム配列の *de novo* アセンブリ

最初に、得られたリード配列のクリーニングとして、Trimmomatic version 0.36 (Langdon 2015)を用いて、Truseq 又は Nextera mate-pair Sample Prep Kit によって付加されたアダプター配列、クオリティの低いリード、および長さ 36 bp 未満の短いリードをトリミングした。その後、4 段階の過程（「プレアセンブリ」、「SOAPdenovo によるアセンブリ」、「クリーンアップ」、および「アセンブリ結果の併合」）によって、PE リードからコンティグ配列を作成した。第 1 段階では、PE リードについて、オーバーラップを含め、エラー補正とクリーニングが行われる PANDAseq (Masella et al. 2012)によってプレアセンブルした。第 2 段階では、第 1 段階でプレアセンブルしたリードおよび残りのリードを 80~127 の複数の k-mer を用いる SOAPdenovo v.2.04-r240 (Luo et al. 2012)にかけ、コンティグ配列を作成した。第 3 段階で、他のコンティグ配列と 95%以上同じ配列相同性を示すコンティグをトリムすることによって、第 2 段階で作成したコンティグ配列のクリーニングを行った。第 4 段階では、Platanus version 1.2.4 (Kajitani et al. 2014)を用いてデフォル

トパラメータで PE リードを処理して生じたコンティグを、SSPACE-longread v1-1 (Boetzer and Pirovano 2014)を用いて、最小オーバーラップ長を 36 bp に、最小リンク数を 1 に、最大リンク比を 0.5 に設定し、SOAPdenovo によって構築したコンティグにマーキングした。次に、3 段階の過程（「リード選択」、「スキヤフオールディング」、および「ギャップフィリング」）によってスキヤフオールド配列を作成した。第 1 段階で、--local オプションをつけた bowtie2 version 2.3.4.3 (Langdon 2015)を用いて、PE および MP リードをコンティグ配列にマップし、コンティグ間を跨ぐようにマップされたリードペアを選択した。第 2 段階で、SSPACE-STANDARD version 3.0 (Boetzer et al. 2011) (BaseClear) を用いて、最小リンク数を 3 に設定して、選択したリードおよびコンティグをスキヤフオールディングした。第 3 段階で、GapFiller v1-10 (Nadalin, Vezzi, and Policriti 2012) (BaseClear) によって、最小オーバーラップ数を 30、切り取るリード数を 50 に設定して、スキヤフオールドにおける未知の塩基を埋め、ドラフトゲノム配列を完成した。以上のシーケンスリードおよび完成したドラフトゲノム配列は DDBJ の bioproject accession code PRJDB8358 下にアップロードした。さらに、embryophyta_odb9 タンパク質セットとともに BUSCO-v3 (Waterhouse et al. 2018)を用いて、ドラフトゲノム配列の完全性を評価した。

3. 遺伝子予測およびアノテーション

NCBI Sequence Read Archive (accession No. SRR2062279、SRR206418024、および SRR5985187-598519425) において入手できる *T. cinerariifolium* の RNA-Seq データを、アセンブルしたゲノムに HISAT2 version 2.2.1 (Kim, Langmead, and Salzberg 2015)を用いてマッピングし、遺伝子予測のためのイントロン情報を得た。その後、アセンブルしたゲノムおよび得られたイントロン情報を、GenBank から入手できる既知の *T. cinerariifolium* の遺伝子情報を読み込ませた AUGUSTUS 3.3.1 (Stanke et al. 2008)にかけた。E-value のカットオフ値 1.0 を用いて、HMMER 2.3.1 (Eddy 1998)における hmmpfam によって、GyDB 2.0 (Llorens et al. 2011)に対比して、予測された遺伝子における転移因子

(Transposable elements、以下、TE) を識別した。信頼度の高い遺伝子を識別するために、InterProScan 5.6-48.0 (Jones et al. 2014)を用いて、既知タンパク質と類似の特徴をもつ非TE 遺伝子を検出し、Blast2GO (Conesa et al. 2005)を用いてアノテーションを行った。以上のアノテーション情報は DDBJ の bioproject accession code PRJDB8358 下にアップロードした。

4. TE 含有量と他の植物との属間比較解析

既にゲノムが解明されている 6 つの植物、シロイヌナズナ (*A. thaliana*) (TAIR10 (Lamesch et al. 2012))、タバコ (*N. tabacum*) (Ntab-TN90 (Sierra et al. 2014))、イネ (*O. sativa*) (assembly Build 4.0 (Rice Annotation et al. 2008))、ヒマワリ (*H. annuus*) (HA412HO_v1.1 (Badouin et al. 2017))、クソニンジン (*A. annua*) (ASM311234v1 (Shen et al. 2018))、およびキクタニギク (*C. seticuspe*) (CSE_r1.0 (Hirakawa et al. 2019)) についても *T. cinerariifolium* と同様に TE 含有量を推定した。*Arabidopsis* 属モデルセットおよびデフォルトパラメータとともに AUGUSTUS 3.3.1 を用いて、これら植物ゲノムの遺伝子コーディング領域を推定し、hmmpfam を用いて GyDB と対比させ、これらの予測された遺伝子における TE を検出した。GyDB の分類に従って、ゲノム領域において TE の各クレードが占める割合を蓄積スコアとして算出した。ORTHOSCOPE 法 (Inoue and Satoh 2019)を用いて、sire クレード TE および oryco クレード TE の分子系統樹も推定した。*T. cinerariifolium*、*H. annuus*、*A. annua*、および *C. seticuspe* において、hmmpfam で抽出された TE 内の逆転写ドメインのアミノ酸配列を CLUSTAL W-mpi 0.13 (Li 2003) によって整列させ、Fast Tree 2.1.10 (JTT モデル、CAT 近似) (Price, Dehal, and Arkin 2010)を用いて、マトリクススペースの JTT モデル (Jones, Taylor, and Thornton 1992)に基づきブートストラップ反復回数 100 回で最尤系統樹を作成した。

5. タンパク質スーパーファミリー含有量と他の植物との属間比較解析

T. cinerariifolium と同様に、本章第 2 節の 3 にて述べた既にゲノムが解明されている 6

つの植物におけるタンパク質群の特性も InterProScan を用いて解析した。各スーパーファミリーの特徴を有する遺伝子の数を測定し、InterProScan によって検出されたスーパーファミリーの特徴 (Sig) と植物の属 (Genus) の各組み合わせに対して、重複オッズスコアを次のように算出した：

$$\text{Multiplication odds score (Genus, Sig)} = \log_2 \frac{N(\text{Genus, Sig}) + PS}{N(\text{Sig}) + PS}$$

式中、N (Genus, Sig) はある植物の属 (Genus) における、InterProScan によって検出されたスーパーファミリーの特徴 (Sig) を有する遺伝子数を表し、PS は疑似カウント定数を表し、これを 0.5 に設定した。機能性タンパク質に関する更なる解析について、各々のカテゴリーで最も重複オッズスコアが高かったスーパーファミリーのタンパク質群の各々に対して BLASTP 解析 (ver 2.7.1) (Altschul et al. 1997) を行った。 *Sambucus nigra* アグルチニン I (SNA-I, accession No. O22415.1)、 *A. thaliana* のエチレンレセプター1 (ETR1, accession No. AAA70047.1)、 *A. thaliana* のフェリチン-1 (AtFer-1, accession No. CAA63932.1) をリボソーム不活性化タンパク質、シグナル伝達ヒスチジンキナーゼ、フェリチン様タンパク質のそれぞれのクエリとして用いた。検出されたタンパク質ペア (SNA-I と Tci_399175、ETR1 と Tci_144982、AtFer-1 と Tci_154278) は CLUSTAL W-MPI 0.13 によって配列アラインメントを行った。 *T. cinerariifolium* のピレトリン生合成関連酵素に関して、ORTHOSCOPE 法で、分子系統樹をさらに評価した。 *A. thaliana* (TAIR10)、 *N. tabacum* (Ntab-TN90)、 *O. sativa* (assembly Build 4.0)、 *H. annuus* (HA412HO_v1.114)、 *A. annua* (ASM311234v1)、 *C. seticuspe* (CSE_r1.0)、ブラックコットンウッド (*Populus trichocarpa*) (Pop_tri_v3 (Tuskan et al. 2006))、タルウマゴヤシ (*Medicago truncatula*) (MedtrA17_4.0 (Tang et al. 2014))、ダイズ (*Glycine max*) (Glycine_max_v2.1 (Schmutz et al. 2010))、トウモロコシ (*Zea mays*) (B73_RefGen_v4 (Jiao et al. 2017))、ワタ *Gossypium raimondii* (Graimondii2_0 (Paterson et al. 2012))、トマト (*Solanaum lycopersicum*) (SL3.0 (Shearer et al. 2014))、およびヨーロッパブドウ (*Vitis vinifera*)

(12X (Canaguier et al. 2017)) の各植物の予測タンパク質群に対して、BLASTP 2.7.1 を用いて、TciLOX1、TciJMH、TciCDS、および TciGLIP に関連する遺伝子を検出した。E-value のカットオフ値を 10^{-3} として BLAST 検索においてヒットした配列をスクリーニングし、上位 10 番目までのヒットした配列を後の系統解析に用いた。これらの配列を CLUSTAL W-mpi 0.13 によってアライメントし、MEGA ソフトウェア (Kumar et al. 2018) を用いて、マトリクスベースの JTT モデルに基づきブートストラップ反復回数 500 回で最尤系統樹を作成した。代謝遺伝子クラスタ (MGCs) は PhytoClust (Topfer, Fuchs, and Aharoni 2017) を用いて調査した。先行研究で使用されている値に従い (Topfer, Fuchs, and Aharoni 2017)、変数を 20 Kbp に設定した。ピレトリン類生合成関連酵素 (TciLOX1、TciJMH、TciCDS、および TciGLIP) をコードする遺伝子が存在するスキュフォールドに含まれる遺伝子は、Genomjack ソフトウェアプログラム (三菱スペース・ソフトウェア、東京、日本) を用いて可視化した。

第3節 結果

1. *T. cinerariifolium* ゲノムの配列アセンブリおよびアノテーション

T. cinerariifolium ゲノムのペアエンド (PE) ライブラリおよびメイトペア (MP) ライブラリ (それぞれ、インサートサイズ 3Kb、5Kb および 8Kb : MP-3Kb、MP-5Kb、および MP-8Kb) を作成し、その後、Hiseq X および Hiseq 4000 解析機器を用いて配列を決定した。PE、MP-3Kb、MP-5Kb、および MP-8Kb のリードの総塩基数はそれぞれ、1,497 Gb、135 Gb、166 Gb、および 95 Gb であった (表 2-1)。その後、出力されたリードに対してコンティグアセンブリおよびスキファールディングを行い、ゲノム配列を推定した (図 2-2)。SOAPdenovo および Platanus を用いて PE リードをアセンブルし、SSPACE-Longread を用いてコンティグを統合することによって、総長 7.7 Gb、3,892,368 のコンティグ (表 2-2) を構築した。その後、リードのペア情報からコンティグ配列を結合する SSPACE-STANDARD と、コンティグ間の未知の塩基を埋める Gapfiller を用いて、アセンブルしたコンティグを PE および MP リードによってスキファールディングした (図 2-2)。生じたスキファールドの総長は 7.1 Gb であり、これはフローサイトメトリーによって推定した *T. cinerariifolium* のゲノムサイズに相当した。スキファールドの N50 値は 14 Kb であり、スキファールドの最大長は 1.2 Mb であった (表 2-2)。その後、過去の *T. cinerariifolium* の RNA-seq 情報 (Xu et al. 2018; Khan et al. 2017) から作成された「イントロン情報」を用いて、ドラフトゲノム配列を AUGUSTUS によって解析し、その結果、935,992 個の推定遺伝子が予測された。ゲノムの完全性を評価するために、ドラフトゲノム配列を BUSCO による解析にかけ、ゲノム配列における完全な (C)、断片化した (F)、および欠失した (M) 保存遺伝子をカウントした。1440 個の保存されたコア植物遺伝子を用いた解析から、*T. cinerariifolium* のゲノムアセンブリにおいて、91.8% (82.3%が完全、9.5%が断片化) の保存遺伝子が存在することが認められた (表 2-3)。他の植物のゲノムで検出された保存遺伝子の割合は、アメリカネナシカズラ (*Cuscuta campestris*) で 83.8% (Vogel

et al. 2018)、*Ruellia speciosa* で 77.5% (Zhuang and Tripp 2017)、エゾムカシヨモギ (*Erigeron breviscapus*) で 86.9% (Yang et al. 2017)、ライムギ (*Secale cereal*) で 89.0% (Bauer et al. 2017)であるため、今回決定した *T. cinerariifolium* のドラフトゲノム配列はこれらと同等以上の高品質であることが示されたため、以後の解析にはこの *T. cinerariifolium* ドラフトゲノム配列を用いた。一般的に、植物ゲノム中には他の真核生物ゲノムと比較して転移因子 (Transposable Element: TE) が多く含まれる傾向があり、その種類や数 (重複度) が植物種によって異なるため、各植物のゲノムを特徴づけるうえで重要な要素となる。そのため、通常、植物ゲノムの特徴を調べるに際して、TE プロファイルの解析を伴うことが多い。そこで、最初に Gypsy データベース (GyDB) に対する hmmpfam を用いた *T. cinerariifolium* ゲノム中の TE のアノテーションを行ったところ、全ゲノムのうち 2,397 Mbp (33.84%) を占める 525,098 個の TE の存在が明らかになった。TE をさらに分類したところ、Ty1/Copia、Ty3/Gypsy、Retroviridae、Caulimoviridae、Bel/Pao、および他の TE がそれぞれ 1,135 Mbp (16.02%)、907 Mbp (12.81%)、222 Mbp (3.13%)、88 Mbp (1.24%)、18 Mbp (0.25%)、および 28 Mbp (0.39%) を占めることが確認された。Ty1/Copia および Ty3/Gypsy の間で、sire (Ty1/Copia、11.16%)、athila (Ty3/Gypsy、5.76%)、del (Ty3/Gypsy、4.05%)、および oryco (Ty1/Copia、2.14%) クレードが豊富であった (表 2-3)。残る 410,894 個の非 TE 遺伝子を InterProScan にかけたところ、既知のタンパク質の特徴を示す 60,080 個の推定遺伝子が明らかになった。以上より、高品質の 7.1 Gb の *T. cinerariifolium* ドラフトゲノム配列がアセンブルされ、1,497 Gb の PE リードおよび 396 Gb の MP リードに基づき、計 525,098 個の TE と 60,080 個の推定遺伝子を含むことが確認できた。以上のアセンブルに使用したシーケンスリード、完成したドラフトゲノム配列およびアノテーション情報は DDBJ の bioproject accession code PRJDB8358 下にアップロードした。

表 2-1 シーケンスリードの統計

Library	Insert size (bp)	Read length (bases)	Number of reads	Total read length (bases)	Expected depth (=Total read length / 7.1 Gb)
PE	350	151	4,956,379,975	1,496,826,752,450	210.82
MP-3Kb	3,000	101	665,902,933	134,512,392,466	18.95
MP-5Kb	5,000	101	822,858,639	166,217,445,078	23.41
MP-8Kb	8,000	101	471,281,672	95,198,897,744	13.41

PE: ショートインサートライブラリ、MP: メイトペアライブラリ

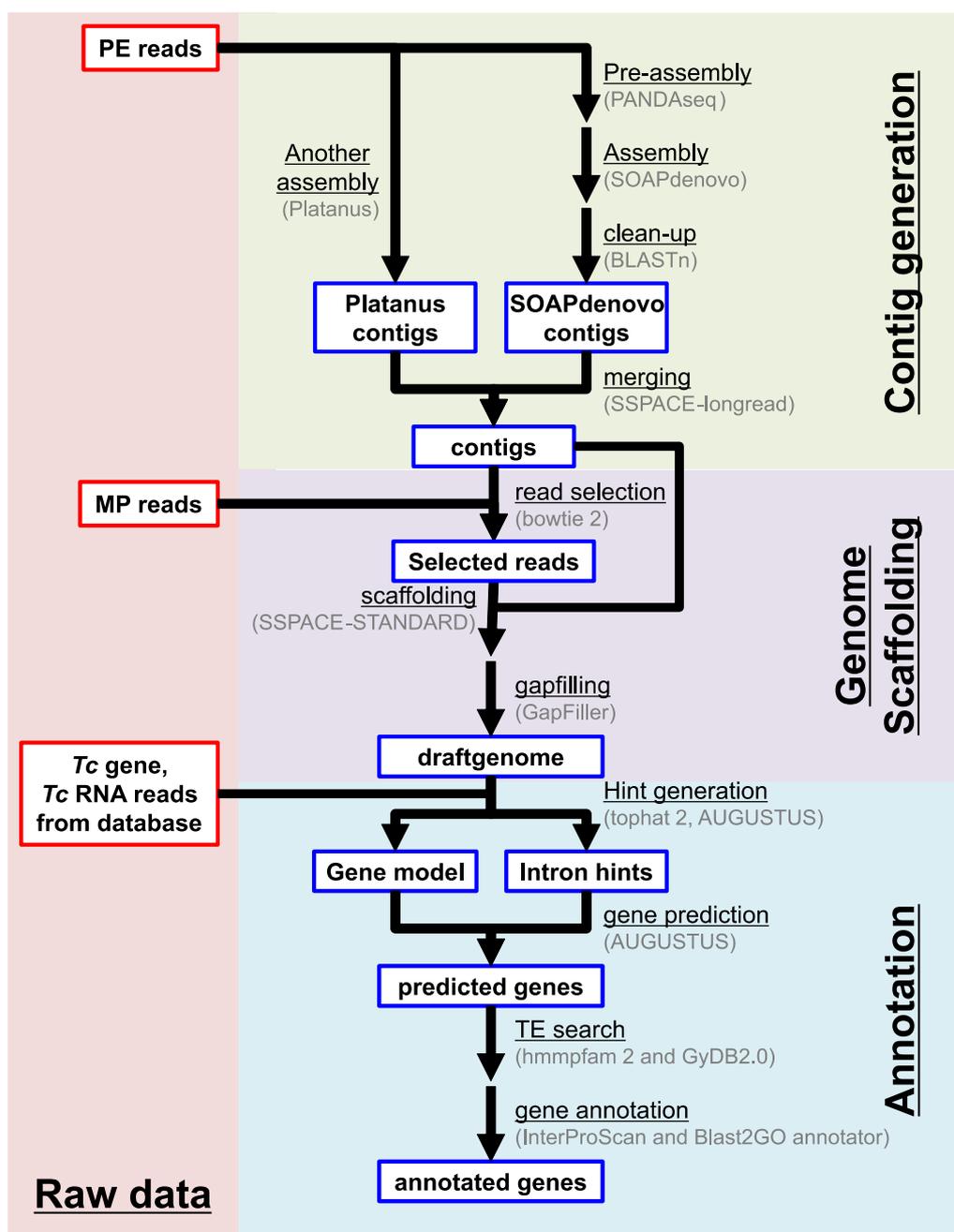


図 2-2 ゲノムアセンブリ、遺伝子予測のフローチャート

PANDAseq を用いる「プレアセンブリ」、SOAPdenovo および Platanus を用いる「コンテ ィグアセンブリ」、BLASTN を用いる「クリーンアップ」、および SSPACE-longread を用 いる「アセンブリ結果の併合」を含む 4 段階の過程によって、PE リードをコンテ ィグアセンブリにかけた。その後、bowtie2 を用いる「リード選択」、SSPACE-STANDARD を用い る「スキヤフォールディング」、および GapFiller を用いる「ギャップフィリング」を含む 3 段階の過程によって MP リードをスキヤフォールディングにかけた。これらの操作によ

り、ドラフトゲノム配列が得られた。その後、AUGUSTUS を用いる「遺伝子予測」、hmmpfam を用いる「TE 検出」、および InterProScan を用いる「タンパク質のスーパーファミリー予測」を含む 3 段階の過程によって、ドラフトゲノム配列におけるコーディング配列をアノテーションした。

表 2-2 ゲノムアセンブリの統計

	Contigs	Scaffolds
Total number of sequence fragments	3,892,368	2,016,451
Total length (bp)	7,710,571,398	7,084,225,540
N50 (bp)	8,078	13,813
Length of longest contig (bp)	1,189,490	1,185,691
GC content (%)	35.1	35.1

表 2-3 ドラフトゲノムのアノテーション統計

Number of predicted genes	935,992	
BUSCO v3	C: 82.3% (Single: 72.8%, Duplicated: 9.5%) F: 9.5% M: 8.2%	
Number of predicted TEs	525,098	
Region of TEs	Total	2,397 Mbp (33.84%)
	Ty1/Copia	1,135 Mbp (16.02%)
	Ty3/Gypsy	907 Mbp (12.81%)
	Retroviridae	222 Mbp (3.13%)
	Caulimoviridae	88 Mbp (1.24%)
	Bel/Pao others	18 Mbp (0.25%) 28 Mbp (0.39%)
Number of predicted genes encoding products with known protein signatures	60,080	

BUSCO の表記で C: 完全長の保存遺伝子の割合、F: 断片化している保存遺伝子の割合、M: 検出できなかった保存遺伝子の割合、TE: 転移因子

2. TE の属間比較解析

シロイヌナズナ (*A. thaliana*)、タバコ (*N. tabacum*)、およびイネ (*O. sativa*) のゲノムにおける TE を予測し、*T. cinerariifolium* の TE クレードと他の種のそれらとを比較したところ、*T. cinerariifolium* において他の植物ゲノムより 2 倍以上多い 2 つの TE クレードの存在が明らかになった (表 2-4)。*T. cinerariifolium* において、sire クレード TE は、*A. thaliana*、*N. tabacum* および *O. sativa* よりそれぞれ、14.3 倍、8.5 倍、および 8.1 倍高い割合で存在することが認められた (表 2-4)。同様に、oryco クレード TE はそれぞれ、2.2 倍、4.6 倍、および 2.6 倍高い割合で存在することが認められた。これらの結果から、sire クレード TE および oryco クレード TE は、*T. cinerariifolium* ゲノムにおいて、他の植物より大幅に高い割合を占めることが示された。さらに、*T. cinerariifolium* ゲノムにおけるこれらの TE の割合を、他のキク科植物 (*C. seticuspe*、*A. annua* および *H. annuus*) におけるこれらの TE の割合と比較したところ (表 2-5)、sire クレード TE および oryco クレード TE はこれらのキク科の種のいずれにも顕著に豊富に含まれていることが証明された。キク科の共通祖先において、又は個々の種に独立して、sire クレード TE および oryco クレード TE が重複したかどうかを評価するために、sire および oryco クレード TE 内でコードされた逆転写酵素ドメインの分子系統樹を解析し、種特異的な重複現象が起きた数を反映する、単一種のクラスタ内における共クラスタリング遺伝子を計数した (図 2-3)。これらの TE が共通祖先において重複していたのであれば、大半の TE が 50% より高いブートストラップ値でオーソログクラスタに位置すると予測され、そうでなければ、単一種で構成されるクラスタに位置すると予測される。実際、sire クレード TE の 98%、92%、93%、および 98% がそれぞれ *T. cinerariifolium* (図 2-3A)、*C. seticuspe* (図 2-3B)、*A. annua* (図 2-3C)、および *H. annuus* (図 2-3D) において重複し、このことから、sire クレード TE の大半が個々の種において重複したことが示される。同様に、oryco クレード TE の 85%、69%、78%、および 89% がそれぞれの生物 (図 2-3E~2-3H) において重複し、このことか

ら、oryco クレード TE も大半が個々の種において重複したことが示されている。さらに、sire クレード TE の 90% (図 2-3A) および oryco クレード TE の 49% (図 2-3E) が、8 倍又はそれ以上重複した遺伝子とクラスタを形成した。これらの結果から、sire クレード TE および oryco クレード TE は個々のキク科の種において重複したことが明らかになった。ゲノムスキャフォールドにおける TE の分布を、それぞれ 20 Kb 領域でコードされた sire および oryco クレード TE の遺伝子の数をカウントすることによって解析した。図 2-4 に示すように、20 Kb 毎に存在する sire (図 2-4A) および oryco (図 2-5B) クレード TE 遺伝子の頻度プロットを用いると、sire および oryco クレード TE 遺伝子が均一に分布していることが観察された。さらに、いかなる TE クレードについても頻度プロットからは分布の局在化は認められなかった (図 2-4C)。次に、当該 TE クレードの他のメンバーから 20 Kb の領域以内にどれだけの TE 数が存在するかを調べた。図 2-4D に示すように、sire および oryco クレード TE 遺伝子のそれぞれ 82.35%と 92.65%が、当該クレードの他の TE から 20 Kb 以上離れて分布していた。これらのデータから、*T. cinerariifolium* において重複している TE はゲノム全体にわたり散在していることが明らかになった。

表 2-4 *T. cinerariifolium* と非キク科植物間におけるゲノム中の TE 含有割合の比較

Family	Clade	Tc	Nt	Os	At
Ty1/ Copia	<i>sire</i>	11.16 %	1.31 %	1.38 %	0.78 %
		(790,364,218)	(47,853,545)	(5,299,804)	(936,463)
Ty3/ Gypsy	<i>athila</i>	5.76 %	4.30 %	1.26 %	3.21 %
		(408,032,509)	(156,681,791)	(4,813,373)	(3,843,212)
Ty3/ Gypsy	<i>del</i>	4.05 %	17.59 %	3.08 %	0.96 %
		(286,557,581)	(640,748,439)	(11,774,484)	(1,146,571)
Ty1/ Copia	<i>oryco</i>	2.14 %	0.47 %	0.83 %	0.96 %
		(151,866,453)	(17,163,885)	(3,191,553)	(1,150,943)
Retro- viridae	<i>lentiviridae</i>	1.67 %	0.53 %	0.84 %	1.02 %
		(118,004,479)	(19,437,617)	(3,226,018)	(1,219,271)

括弧内の数値は各々のクレードに振り分けられた領域の長さ (bp) を示す。Tc: *Tanacetum cinerariifolium*, Nt: *Nicotiana tabacum*, Os: *Oryza sativa*, At: *Arabidopsis thaliana*, TE: 転移因子

表 2-5 *T. cinerariifolium* とキク科植物におけるゲノム中の TE 含有割合の比較

Family	Clade	Tc	Cs	Aa	Ha
Ty1/ Copia	<i>sire</i>	11.16 %	9.87 %	6.34 %	2.30 %
		(790,364,218)	(268,554,348)	(113,663,806)	(83,863,222)
Ty3/ Gypsy	<i>athila</i>	5.76 %	3.36 %	5.71 %	1.21 %
		(408,032,509)	(91,399,675)	(102,303,891)	(44,010,125)
Ty3/ Gypsy	<i>del</i>	4.05 %	1.55 %	1.91 %	8.80 %
		(286,557,581)	(42,223,327)	(34,214,927)	(320,646,159)
Ty1/ Copia	<i>oryco</i>	2.14 %	1.58 %	1.33 %	1.04 %
		(151,866,453)	(42,884,278)	(23,928,063)	(37,770,638)
Retro- viridae	<i>lentiviridae</i>	1.67 %	1.56 %	1.04 %	2.04 %
		(118,004,479)	(42,471,725)	(18,691,180)	(74,263,619)

括弧内の数値は各々のクレードに振り分けられた領域の長さ (bp) を示す。Tc: *Tanacetum cinerariifolium*、Cs: *Chrysanthemum seticuspe*、Aa: *Artemisia annua*、Ha: *Helianthus annuus*、TE: 転移因子

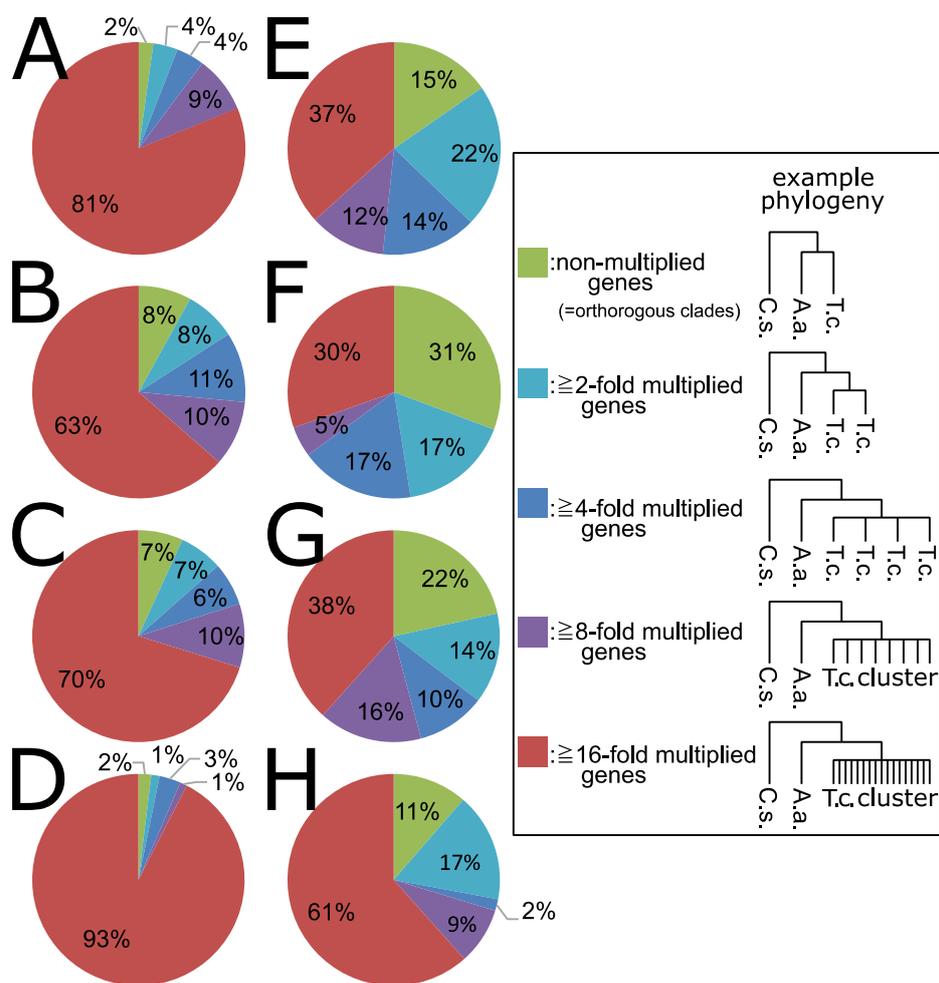


図 2-3 sire クレード TE (A~D)および oryco クレード TE (E~F)の重複解析

sire クレード TE および oryco クレード TE の逆転写酵素ドメインを用いた分子系統樹に基づき、各属に対して、単一の属クラスタにおいてクラスタ化している遺伝子の数をカウントした。この数は種特異的な重複現象の回数を反映する。*T. cinerariifolium* (A、E)、*C. seticuspe*(B、F)、*A. annua* (C、G)、および *H. annuus*(D、H)に対して、クラスタ化した遺伝子の数を円グラフに示す。

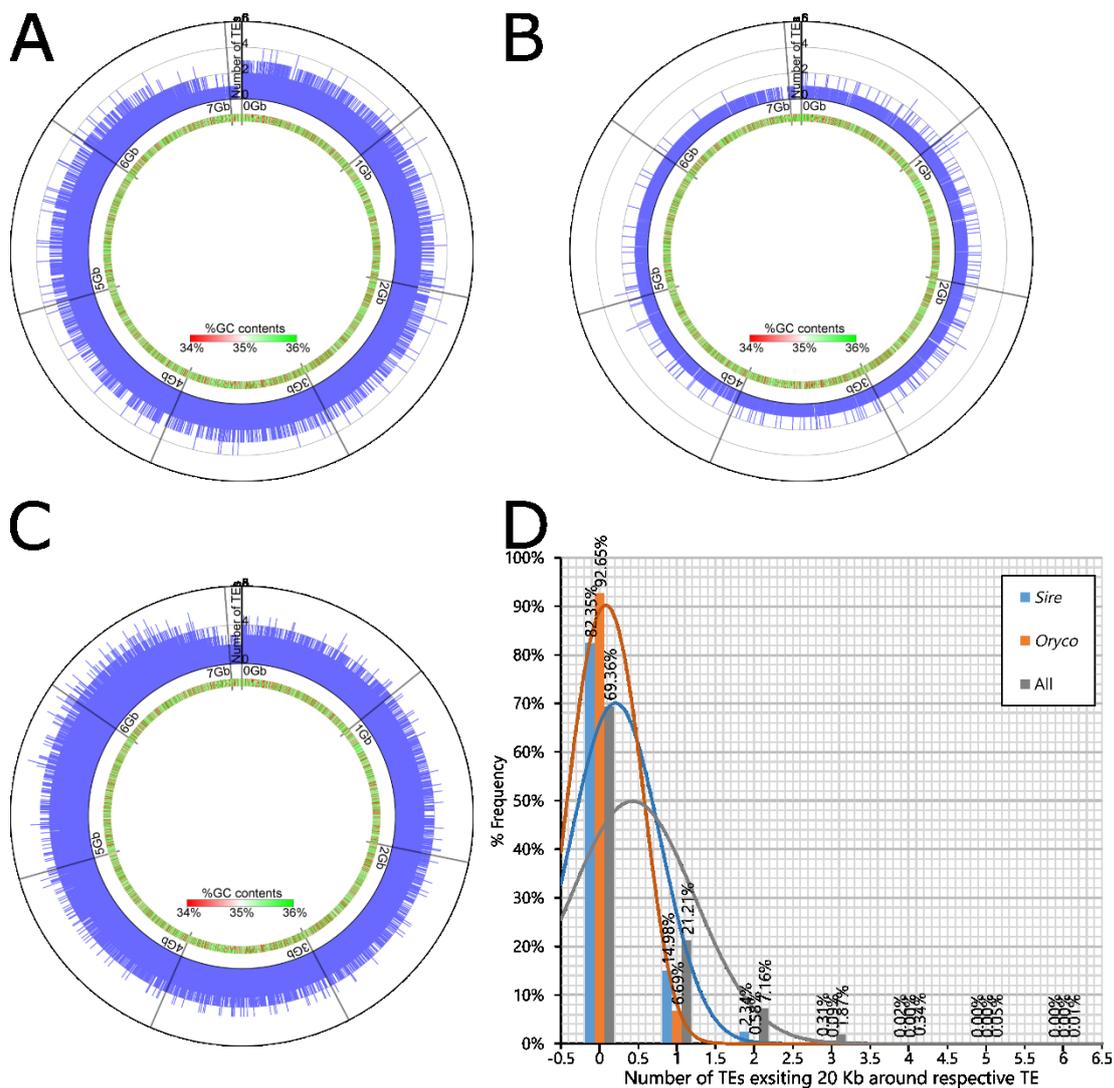


図 2-4 *T. cinerariifolium* のゲノム中における転移因子の分布

sire クレード TE (A)、oryco クレード TE (B) およびその他のクレードの TE (C) の分布を、円を用いて示した。20 Kb の領域毎で各々のスキファールド内にコードされている TE の数を外円にプロットし、20 Kb 毎の GC 含有量を内円にヒートマップで示した。(D) sire、oryco および全ての TE について、各々の TE の周囲 20 Kb に存在する TE の数をヒストグラムで示した。TE: 転移因子

3. *T. cinerariifolium* 遺伝子の機能的アノテーションと属間比較解析

次に、*T. cinerariifolium* におけるタンパク質スーパーファミリーの重複比について他の種との比較を行った。具体的には、InterProScan を用いて、*C. seticuspe*、*A. annua*、*H. annuus*、*N. tabacum*、*O. sativa*、および *A. thaliana* のタンパク質データセットを解析し、スーパーファミリーと種の各組み合わせに対して重複オッズスコアを算出した。重複オッズスコアは、特定のスーパーファミリーの遺伝子について、ゲノム内にどの程度存在、重複化しているかを異種生物間で比較し、数値化したものである。重複オッズスコアの正の値は、特定のスーパーファミリーに対する遺伝子について、ある種が他の種より多く重複していることを示唆している。生体防御、シグナリング、および代謝に関連するスーパーファミリーに対して、*T. cinerariifolium* について高い重複オッズスコアを有するものを表 2-6 に示した。

生体防御に関連するスーパーファミリーに関して、「リボソーム不活性化タンパク質」(IPR036041) および「リシン B 様レクチン」(IPR035992) は、いずれも II 型リボソーム不活性化タンパク質 (以下、RIP) として分類され、*T. cinerariifolium* においては、それぞれ 2.05 および 1.57 の重複オッズスコアを示した (表 2-6)。リシンを含む RIP は広範囲の生物種に対して高い毒性を示し、産生植物にとって生体防御分子となる。例えば、*N. tabacum* の葉から分離・精製されたタバコ RIP は、青枯病菌 (*Pseudomonas solanacearum*)、火傷病菌 (*Erwinia amylovora*)、および *Shigella asonei* に対して強力な抗菌活性を有することが報告されている (Sharma et al. 2004)。ヨウシュヤマゴボウ (*Phytolacca americana*) の葉の抽出物から精製した抗ウイルスタンパク質 (Obrig, Irvin, and Hardesty 1973) は、タバコモザイクウイルス感染に対する全身抵抗性を増強することが報告されている (Zhu et al. 2016)。セイヨウニワトコ (*Sambucus nigra*) のアグルチニン I (SNA-I) は、カメムシ目に対して殺虫効力を示すことが確認されている (Shahidi-Noghabi, Van Damme, and Smaghe 2008)。実際に、*T. cinerariifolium* の RIP に対する SNA-I をクエリとする

BLASTP 検索では、 2.99×10^{-112} の E-value 値、37.89%の配列相同性を有する Tci_399175 が検出された。配列アラインメントと BLASTP の結果から、Tci_399175 は、RIP ドメインと Q-X-W モチーフを含んだ二つの RICIN ドメインを有することが分かった (図 2-5A)。先行研究から、RIP ドメインは rRNA グリコシド結合の開裂に関与しており、一方で、RICIN ドメインは標的細胞のグリカンへの結合を介した侵入に関与していることが示されている (Walsh, Dodd, and Hautbergue 2013)。この配列解析から、*T. cinerariifolium* で重複している RIP には、SNA-I と類似の殺虫成分を含むことを示唆している。これらの所見から、*T. cinerariifolium* は、セスキテルペンラクトン類 (Ramirez, Saillard, et al. 2013) やピレトリン類 (Zito, Zieg, and Staba 1983) (DeMicco et al. 2010) のような低分子殺虫成分だけでなく、昆虫に対する天然の毒性タンパク質をコードしている遺伝子を、他の植物と比べて多数有することが示唆される。

シグナリングに関連するスーパーファミリーに関して、「シグナル伝達ヒスチジンキナーゼ、二量体化/ホスホアクセプタードメイン」(IPR036097)、「Rho GDP・解離抑制ドメイン」(IPR024792)、および「HECT、E3 リガーゼ触媒ドメイン」(IPR035983) は、*T. cinerariifolium* において特異的な重複を示し、重複オッズスコアはそれぞれ 1.39、1.32、および 1.14 であった (表 2-6)。いくつかの植物種はヒスチジンキナーゼ (エチレンレセプターのように機能するタンパク質) と非エチレンレセプタータイプのヒスチジンキナーゼ (例：サイトカイニンレセプター) (Sudre et al. 2013) を有することが知られている。*A. thaliana* のエチレン活性型キナーゼであるエチレンレセプター1 (ETR1) (Chang et al. 1993) をクエリとして用いた BLASTP 検索により、G-X-G モチーフを含む HATPase_c ドメインを有し、0 の E-value 値、69.18%の配列相同性を有する予測タンパク質 Tci_144982 が検出された (図 2-5B)。この配列解析から、*T. cinerariifolium* で重複しているヒスチジンキナーゼには、ETR1 と類似のエチレンレセプタータイプのヒスチジンキナーゼが含まれることを示唆している。これらのシグナリング分子が *T. cinerariifolium* において特異的に

重複していることから、*T. cinerariifolium* はシグナル伝達および調節タンパク質を種特異的に重複させていることが示唆される。ピレトリン類の濃度および組成を調節する分子機序は未だ明確にされていないが、これらの豊富なシグナリング関連タンパク質は、ピレトリン合成を調節する明らかな候補となりうる。

代謝に関連するスーパーファミリーで高い遺伝子重複を示した「エノラーゼ様、C 末端ドメイン」(IPR036849)、「シトクロム c オキシダーゼ様ドメイン」(IPR036909)、「リポキシゲナーゼ、C 末端ドメイン」(IPR036226)、「アシル CoA デヒドロゲナーゼ/オキシダーゼ N 末端ドメイン」(IPR037069)、および「金属依存性ヒドロラーゼ、複合ドメイン」(IPR011059) の重複オッズスコアは、それぞれ、1.28、1.22、1.16、1.13、および 1.11 であった (表 2-6)。リポキシゲナーゼ (例えば、TciLOX1) およびアシル CoA デヒドロゲナーゼ/オキシダーゼ (12-オキソ-cis-10,15-フィトジエン酸の酸化に関連する反応を触媒する酵素) は、ピレトリン類のアルコール部分のオキシリピン経路に関与することが知られているため (図 2-1)、これらの重複オッズスコアは *T. cinerariifolium* におけるピレトリン類およびジャスモン酸誘導体の特異的生合成に合致する。さらに、重複オッズスコア 0.37 の「シトクロム P450」(IPR036396) が検出されたが、これは *T. cinerariifolium* において計 745 個もの遺伝子が検出され、最も豊富な種類の遺伝子であった。実際、シトクロム P450 として分類されているタンパク質、TciJMH や TciPYS はジャスモロンやピレスロロンの生合成に関与することが知られている (図 2-1)。このため、シトクロム P450 の重複は、*T. cinerariifolium* 特異的なジャスモン酸誘導体の生合成においてこの種類のタンパク質が果たす役割と適合する。鉄輸送を担うフェリチン様スーパーファミリーの重複オッズスコアは 1.42 であることが確認された。*A. thaliana* において、フェリチン-1, 3 および 4 は、シトクロム P450 酵素をコードする多数の遺伝子の発現を制御することが知られている (Briat et al. 2010)。*A. thaliana* のフェリチン-1 (AtFer-1) をクエリとする、*T. cinerariifolium* のフェリチン様タンパク質に対する BLASTP 検索では、 5.64×10^{-115} の E-

value 値、69%の配列相同性を有する Tci_154278 が検出され、このタンパク質は保存された真核生物フェリチンドメインを含む鉄イオンチャネル、フェロオキシダーゼ二鉄中心、フェリハイドライト核形成中心を有していた (図 2-5C)。この結果は、金属依存性ヒドロラーゼ、シトクロム c、アコニターゼ、リポキシゲナーゼ、および シトクロム P450 のような金属利用酵素をコードしている遺伝子の重複と一致し (表 2-6)、*T. cinerariifolium* では、種々の酵素反応とその調節に対して使用するために他の植物と比較して大量の金属を移入する必要のあることが示唆される。

低い重複オッズスコアを示したスーパーファミリーは、「エクспанシン、セルロース結合様ドメイン」(IPR036749)、「ミトコンドリア糖タンパク質」(IPR036561)、および「ArfGAP ドメイン」(IPR038508)が認められ、重複オッズスコアはそれぞれ-0.53、-0.53、および 0.49 であった (表 2-7)。*O. sativa*、*C. seticuspe*、*A. annua*、および *H. annuus* におけるこれらのスーパーファミリーの重複オッズスコアが *T. cinerariifolium* における重複オッズスコアと同等に低かった (表 2-7) ことから、これらスーパーファミリーのタンパク質の数が少ないことは、キク科植物間で *T. cinerariifolium* に特異的でないことが示唆された。それに対して、真菌病原体に対する防御において重要な役割を果たすスーパーファミリー、「エンドキチナーゼ様タンパク質」(Kasprzewska 2003) は *T. cinerariifolium* 特異的な低重複状態を示した。RIP の蓄積 (表 2-6) およびエンドキチナーゼの数が少ないことから、これらに付随する生体防御戦略は *T. cinerariifolium* において特異的に進化していることが示唆された。

表 2-6 *T. cinerariifolium* において重複オッズスコアの高かったスーパーファミリー

Category	InterPro ID	Superfamily name	Tc	Cs	Aa	Ha	Nt	Os	At
Biodefense	IPR036041	Ribosome-inactivating protein	2.05 (98)	-1.05 (7)	-0.24 (16)	-2.31 (0)	-2.31 (0)	-0.17 (17)	-2.31 (0)
Biodefense	IPR035992	Ricin B-like lectins	1.57 (69)	-0.18 (17)	0.11 (22)	-1.06 (7)	-0.64 (11)	-0.94 (8)	-1.32 (5)
Signaling	IPR036097	Signal transduction histidine kinase, dimerization/phosphoacceptor domain	1.39 (101)	-0.64 (21)	-0.39 (26)	-0.3 (28)	0.33 (46)	-1.76 (7)	-0.76 (19)
Signaling	IPR024792	Rho GDP-dissociation inhibitor domain	1.32 (34)	-0.06 (10)	-0.51 (6)	-0.27 (8)	-0.06 (10)	-0.97 (3)	-0.97 (3)
Signaling	IPR035983	HECT, E3 ligase catalytic domain	1.14 (72)	0.4 (41)	-0.22 (25)	-0.6 (18)	0.09 (32)	-1.54 (7)	-0.95 (13)
Metabolism	IPR012347	Ferritin-like	1.42 (35)	0.1 (11)	-0.58 (5)	-0.44 (6)	-0.73 (4)	-1.09 (2)	-0.44 (6)
Metabolism	IPR036849	Enolase-like, C-terminal domain	1.28 (50)	-0.25 (14)	-0.18 (15)	-0.41 (12)	0.08 (19)	-1.04 (6)	-0.92 (7)
Metabolism	IPR036909	Cytochrome c-like domain	1.22 (39)	-0.43 (9)	-0.78 (6)	-0.15 (12)	0.22 (17)	-0.66 (7)	-0.66 (7)
Metabolism	IPR036226	Lipoxygenase, C-terminal domain	1.16 (86)	0.45 (51)	-0.14 (32)	-0.45 (25)	0.01 (36)	-1.18 (13)	-1.77 (7)
Metabolism	IPR037069	Acyl-CoA dehydrogenase/oxidase, N-terminal domain	1.13 (36)	-0.53 (8)	-0.23 (11)	-0.14 (12)	0.29 (18)	-0.91 (5)	-0.77 (6)
Metabolism	IPR032466	Metal-dependent hydrolase	1.11 (94)	-0.71 (23)	-0.47 (28)	0.1 (44)	0.37 (54)	-1.19 (15)	-0.56 (26)
Metabolism	IPR036396	Cytochrome P450	0.37 (745)	0.35 (732)	0.26 (688)	-0.02 (568)	0.06 (600)	-0.86 (314)	-0.66 (361)

括弧内の数値は各々のスーパーファミリーに振り分けられた遺伝子の数を示す。Tc: *Tanacetum cinerariifolium*, Cs: *Chrysanthemum seticuspe*, Aa: *Artemisia annua*, Ha: *Helianthus annuus*, Nt: *Nicotiana tabacum*, Os: *Oryza sativa*, At: *Arabidopsis thaliana*

表 2-7 *T. cinerariifolium* において重複オッズスコアの低かったスーパーファミリー

Category	InterPro ID	Superfamily name	Tc	Cs	Aa	Ha	Nt	Os	At
Biodefense	IPR036861	Endochitinase-like	-1.15 (1)	-0.41 (5)	-0.28 (6)	0.51 (14)	0.26 (11)	0.07 (9)	0.35 (12)
Non-specific	IPR036749	Expansin, cellulose-binding-like domain	-0.53 (36)	-0.53 (36)	-0.08 (51)	0.02 (55)	0.61 (85)	0.18 (62)	-0.05 (52)
Non-specific	IPR036561	Mitochondrial glycoprotein	-0.53 (6)	-0.29 (8)	-0.29 (8)	-0.29 (8)	0.53 (18)	-0.09 (10)	0.53 (18)
Non-specific	IPR038508	ArfGAP domain	-0.49 (24)	-0.65 (21)	-0.3 (28)	0.56 (55)	0.67 (60)	-0.96 (16)	0.3 (45)

括弧内の数値は各々のスーパーファミリーに振り分けられた遺伝子の数を示す。Tc: *Tanacetum cinerariifolium*、Cs: *Chrysanthemum seticuspe*、Aa: *Artemisia annua*、Ha: *Helianthus annuus*、Nt: *Nicotiana tabacum*、Os: *Oryza sativa*、At: *Arabidopsis thaliana*

A

Tci_399175	MRLFATLLLLVVAATY-LHLGGVQADNKYIRKKEFSTNNGQMNFKAYIKLLEELRDRLASG-TDVGILVTRPARNIPLQ	78
SNA-I (O22415.1)	MRVVTKLLYLWVLAICGLGIHGAIHFRVTPPPVYPSVSENLIGADTYGPFLLALQEKVILGNHTAFDLPVLPNPSQVSDS	80
Tci_399175	ERFIQVELQNSGYQVITVIIDTVNVYVVGVLVSSPLAPILHYLDGASEALRDALPNTTYIHRPLNFDGRYRSLP----	153
SNA-I (O22415.1)	NRFVLPVLTNPSGDTVTLAIDVNVLYVVAFFSSNG-----RSYFFSGSTAVQRDNLFEVDT-TQEELNFTGNYSILSERQVGF	154
Tci_399175	DRDQTPLGHGALNDAIR-----NLYYGQSQ-RSALLVIQMVAEAVRIRHIEHLILRNMYDERNPFIPTDKAINLENS	226
SNA-I (O22415.1)	GRVYIPLGPKSLAQATSSLRTYTLGAGDTKPLARGLLVVIQMVSEPAARFVYIELRIKTSITDAS--EFTPDLLMLSMENN	232
Tci_399175	WDALSTQIQSSCFESGVFLREVRVQIAPDPSSQLRIRNVESSMALAALALMLYKTKPTAIR-----HPVVPVPPVAV	297
SNA-I (O22415.1)	WSSMSSEIQQAQPGGIFP-GVQQLRDERNPLEVINERRLLELPIYAVILYGCAPVTSNSTNNAIDQILIKMPPVFRGG	310
Tci_399175	GADGQCPYGEPTTIIGRDGQMDVKENQYGNENPIILFPCGNAQRNQLWTFKSDGTIRSNKCLITTSNGYIMIFDCDLA	377
SNA-I (O22415.1)	GYEKVCSVVEVTRRLSGWDGLCVDRDGHYIDGNIVQLGPCGN-ECNQLWTFRTDGTIRWLKCLITTS-SSVMYDCNTV	388
Tci_399175	F-ETTKWILHNAGITMNPRLRLVIAAESSTPRVLTAAVDSNSSRQAWSPQNTYQPTLTYISGFLEMCLQANGENARVWL	456
SNA-I (O22415.1)	PPPATKVVVSTDGIITNPRSGLVLTAPQAAGETALSLENNIHAARQGWVVG-DVEPLVTVVGVYKQMCCLTFENGENNFWL	467
Tci_399175	ANCVIETEPQQWAIYGDRTIRLYSDRTLCTVSDGHEVSDSILFKCQGSERWTFMADATILNPAQLVMDVGRGSDVS	536
SNA-I (O22415.1)	EDCVLNR-VEQENALYGDGTIRVNSNRLCVTSDHEPSDLIVILKCEGSAQVWVFNVTNGTISNPNAKLVMDVAQSNVS	546
Tci_399175	LQETILYPTGNPNQKWLAF-----556	
SNA-I (O22415.1)	LRKILYPTGNPNQWITITQPA570	

:RIP domain (pfam00161)
 :RICIN domain (smart00458)
 :Q-X-W conserved motif

B

Tci_144982	MMSDCCEETQWPADELLVRYQYISDFEIAEAYFSIPELELYFVQKSAFEPYRHWLMDQGAFTVLOGATHFINLMTFSH	80
ETR1 (AAA70047.1)	-MEVCNCEETQWPADELLMRYQYISDFEIAEAYFSIPELELYFVKKSAFEPYRHWLMDQGAFTVLOGATHFINLMTFTH	79
Tci_144982	SKTVALVMTAKVLTAVVSCFPAIMLVHITPDLISVKTRELEIKKKAELDREGLIIRIQEETGHWMLTHEIRSTLDR	160
ETR1 (AAA70047.1)	SRVALVMTAKVLTAVVSCFPAIMLVHITPDLISVKTRELEIKKKAELDREGLIIRIQEETGHWMLTHEIRSTLDR	159
Tci_144982	HTILKPTLIELGRTLDLEECVLWMPSRKGMILQLSHSLHNLIPFGSTVPIINLPIITEVFNSEAIRIFHNCPLARIRTPV	240
ETR1 (AAA70047.1)	HTILKPTLVELGRTLALEECALWMPTRTGLELQLSYTLRHQHPVEYTVPIQLPVINQVFGTSRAVKISPNSEVARLEFVS	239
Tci_144982	GSYIPPEVVAVRVPVLLHLSNFEIDNWPINDSAKSYAVMILLFMNGVRKWRDHEMELVTVADQVAVALSHAAILEESMRA	320
ETR1 (AAA70047.1)	GKIMLGEVVAVRVPVLLHLSNFQINDWPELSTKRYALMVLMLPSDSARQHWVHELELVEVVADQVAVALSHAAILEESMRA	319
Tci_144982	RDQLMDQNEALNARQFAEMATHARNDLAVNHEMRTMHAITLASSILLETETLTPDQAMDETILSSNILLALMNDV	400
ETR1 (AAA70047.1)	RDLMDQNEALNARQFAEMATHARNDLAVNHEMRTMHAITLASSILLETETLTPDQAMDETILSSNILLALMNDV	399
Tci_144982	LDSLRLDGSILESEVENLHGLEKREVVPLINPTASVKNISMAIACDILPARGIGDEKRIHQDILNVAQVAVKPTKGG	480
ETR1 (AAA70047.1)	LDSLRLDGSILQLELSTNHLTLERVIANLTKPLAVVKKLPTLNLAPDLPEFVQDEKRIHQDILNVAQVAVKFSQGS	479
Tci_144982	VSIQASILSPEYIQEWQIPFCPTHTDGLFYLLVQVKDGGSGIKQDIEHIFKRESEPRASNNNGGAGIGLAISKRFV	560
ETR1 (AAA70047.1)	ISVIALVTK-----SDTRAADFFVVPFGSHFYLRVVKDKGAGINPQILPKIPTKRAQIQSLATRSSGGSGIGLAISKRFV	555
Tci_144982	NLMGGHDIWIESGGLGKGTVAELVKLGLCNYPNDPIMQQLVPT-VPKTRPHQSGGELIQ-----	619
ETR1 (AAA70047.1)	NLMGGHDIWIESDGLGKCTAIPDVKLGLSERSNE-SKQSGILPKVPAIPRESNPTGLKVLWMDENGVSMVTRGLLWHIGC	634
Tci_144982	-----HRQLDRGGYQASSEPLLYNRTH-----	642
ETR1 (AAA70047.1)	EVTVSSNEECLRVVSHKRVVEMDVMCPGVENYQIATRIRHEKFTQRHQREPLVWALSGNTDKSKKCKMSRGLDGLK	714
Tci_144982	-----	
ETR1 (AAA70047.1)	PVSLDNRDVLSDILLEPRVLYEGM738	

:HATPase c domain (pfam02518)

C

Tci_154278	MSLN-ISTSSFNLLLN-KIDNDIG-YVPRKAIGFVKVNGRCRASVGEENAVVS-----SGVLPQPEEVEVKKDAFVVEIS	72
AtFer1(CAA63932.1)	MASNALSSFTAANPALSPKPLLPHGASAPSVSLGFSRKVGGGRAVVAARVVDNNHPIHGVVQPEEVEVKKADIAIPIIT	80
Tci_154278	PQMSLAKQNYFDSEAAINEQINVEYNVSYVYHALYAYFDRDNVAKGFAKFKVSSDEEREHAEKIMKYQNMRRGVRVTL	152
AtFer1(CAA63932.1)	SHASLARQRFADASEAVINEQINVEYNVSYVYHSMAYFDRDNVAKGLAKFKESSEERECHAEMKYQNRGRVRL	160
Tci_154278	HTIVTPPSEFEHVERGDALYAMELALSLEKLVNEKLLALHAVADRNDPQMADFIESEFLAEQVEAIKKISDYVSQLRV	232
AtFer1(CAA63932.1)	HPIVSPSEFEHAEGDALYAMELALSLEKLVNEKLLNKHVASENNDPQLADFVESEFLGEQIEAIKKISDYITQLRMI	240
Tci_154278	GRGHGVWHFDQMLLEGGVAA252	
AtFer1(CAA63932.1)	GRGHGVWHFDQMLLN-----255	

:eukaryotic ferritin domain (cd01056)
 :Iron ion channel
 :Ferroxidase diiron center

図 2-5 RIP、ヒスチジンキナーゼおよびフェリチンのアミノ酸配列アラインメント

(A) Tci_399175 と *Sambucus nigra* アグルチニン I (SNA-I, accession No. O22415.1)

(B) Tci_144982 と *Arabidopsis thaliana* エチレンレセプター1 (ETR1, accession No. AAA70047.1)

(C) Tci_154278 と *A. thaliana* フェリチン-1 (AtFer-1, accession No. CAA63932.1)
保存されているタンパク質ドメインは着色したハイライトもしくは下線で示す。

4. ピレトリン類生合成関連酵素の系統解析

T. cinerariifolium の最も特徴的な特徴の 1 つはピレトリン類生合成経路の存在であり、その合成には TciLOX1 (*T. cinerariifolium* 13-リポキシゲナーゼ)、TciJMH (*T. cinerariifolium* ジャスモンヒドロキシラーゼ)、TciCDS (*T. cinerariifolium* クリサンテミルニリン酸合成酵素)、および TciGLIP (*T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ) が関わる (図 2-1)。TciLOX1 および TciJMH は、InterProScan 解析によって *T. cinerariifolium* において豊富に存在することが認められたリポキシゲナーゼおよび P450 スーパーファミリーと関連する。他方、TciCDS は一般的なテルペン合成酵素とは異なり、イソプレノイド単位の非 head-to-tail シクラーゼ活性を有し、TciGLIP は一般的な GDSL リパーゼとは異なり、アシルトランスフェラーゼ活性を有するため、機能の観点からこれらの酵素を分類できる相応な InterProScan スーパーファミリーはない。したがって、公的データベースから入手できる 13 の異なる植物ゲノムから予測されたタンパク質群に対して、TciLOX1、TciJMH、TciCDS、および TciGLIP をコードしている遺伝子の重複を、BLAST 検索を用いてさらに解析し、最尤系統樹を作成して分子系統解析を行った。

TciLOX1 は広く存在する植物酵素群の一つである。TciLOX1 はリノレン酸の 13-ヒドロペルオキシリノレン酸への変換を介してピレトリン類のアルコール部分の産生を担うことが確認されている (図 2-1)。TciLOX1 配列をクエリとして用いた BLASTP 検索から、134 個のタンパク質が検出され、このうち 9 個が *T. cinerariifolium* タンパク質、29 個が他のキク科植物のタンパク質、96 個が非キク科植物のタンパク質であり、これらのタンパク質と TciLOX1 との配列相同性はそれぞれ 42.03~57.58%、42.31~95.26%、および 35.09~75.38%であった。これらのタンパク質の分子系統解析から、TciLOX1 は *C. seticuspe*、*A. annua*、および *H. annuus* タンパク質とオーソログスクレードを形成したことが明らかになった (図 2-6A、クレード I)。このクレードはキク科植物のタンパク質のみより成る別のクレードの近くに位置し (図 2-6A、クレード II)、これらのリポキシゲナーゼファミリーの

タンパク質はキク科植物において重複したことが示唆された。さらに、これらのクレード (図 2-6A、クレード I および II) は大きなリポキシゲナーゼクレード (図 2-6A、クレード III) に属し、これには TciLOX1 と類似の 13S-リポキシゲナーゼ活性を有する *A. thaliana* の LOX3 および 4 が含まれた。このリポキシゲナーゼクレードは、キク科植物がリノレン酸リポキシゲナーゼ活性を共有するとの報告と一致した (Ramirez, Yang, et al. 2013)。

TciJMH は、ピレトリン類を産生する植物属においてのみ報告されている、ジャスモンからジャスモロンへの変換を触媒する (図 2-1)。クエリとして TciJMH 配列を用いた BLASTP 検索により、136 個のタンパク質が検出され、このうち 9 個が *T. cinerariifolium* タンパク質、30 個が他のキク科植物のタンパク質、97 個が非キク科植物のタンパク質であり、これらのタンパク質と TciJMH との配列相同性はそれぞれ 34.22~81.53%、37.67~85.01%、および 36.44~58.01%であった。これらのタンパク質の分子系統解析から、TciJMH は、2 つの *T. cinerariifolium* タンパク質と 18 個のキク科植物のタンパク質より成る大きなクレード (図 2-6B、クレード II) 内に、他の 1 つの *T. cinerariifolium* タンパク質、2 つの *A. annua* タンパク質、および 1 つの *C. seticuspe* タンパク質とサブクレードを形成したことが明らかになった (図 2-6B、クレード I)。この大きなクレードへの包含は、ジャスモンのジャスモロンへの変換が *Tanacetum* 属によってのみ報告されているという事実とは反する (Li, Zhou, and Pichersky 2018)。これらの結果から、これら非 *Tanacetum* 属植物の TciJMH 様タンパク質は、非機能的であるか、又は別の植物性化学物質の生合成経路に関与することが示唆された。

TciCDS は、ジメチルアリルニリン酸のクリサンテミルニリン酸への変換によってテルペノイドの酸部分の形成に関与し (図 2-1)、これも種特異的なピレトリン類合成反応であることが報告されている。クエリとして TciCDS 配列を用いた BLASTP 検索により、122 個のタンパク質が検出され、このうち 9 個が *T. cinerariifolium* タンパク質、28 個が他のキク科植物のタンパク質、85 個が非キク科植物のタンパク質であり、これらのタンパク質

と TciCDS との配列相同性はそれぞれ 61.27~88.54%、36.36~93.52%、および 20.21~66.57%であった。これらのタンパク質の分子系統解析から、TciCDS は、7 つの *T. cinerariifolium* タンパク質と 17 個のキク科植物のタンパク質より成る大きなクラスタ (図 2-6C、クレード II) 内に、3 つの *A. annua* タンパク質および 3 つの他の *T. cinerariifolium* タンパク質とサブクレードを形成したことが明らかになった (図 2-6C、クレード I)。TciCDS 様タンパク質の系統樹には明白なオーソログスクレードは含まれなかった。この結果は、クリサンテミルニリン酸を含む種特異的モノテルペンの産生機能が各々のキク科の属において分岐するという報告と一致した (Liu et al. 2012)。これまでに、*A. annua* はそのゲノム内に多数のテルペンに関連する遺伝子座を有することが認められ、これらの遺伝子の産物が種々の種特異的テルペンの生合成を担うことが報告されている (Shen et al. 2018)。これらの所見から、*T. cinerariifolium* 内で多重化したテルペン生合成遺伝子も、*A. annua* で報告されたように、*T. cinerariifolium* において重複し、分岐しているという見解が裏付けられる。

TciGLIP は、ピレトリン類の生合成経路において、菊酸部分およびピレスロロン部分のエステル化によってピレトリン I を合成する (図 2-1)。クエリとして TciGLIP 配列を用いた BLASTP 検索により、135 個の配列が検出され、このうち 9 個が *T. cinerariifolium* タンパク質、30 個が他のキク科植物のタンパク質、96 個が非キク科植物のタンパク質であり、これらのタンパク質と TciGLIP との配列相同性はそれぞれ 48.73~81.78%、31.87~79.35%、および 30.38~53.35%であった。これらのタンパク質の分子系統解析から、TciGLIP は、9 つの他の *T. cinerariifolium* タンパク質および 12 個のキク科植物のタンパク質より成る大きなクラスタ (図 2-6D、クレード II) 内に、1 つの *A. annua* タンパク質、2 つの *C. seticuspe* タンパク質、および 5 つの TciGLIP 以外の *T. cinerariifolium* タンパク質とサブクレードを形成したことが明らかになった (図 2-6D、クレード I)。この系統樹から、*T.*

cinerariifolium は TciGLIP 関連タンパク質を大量に重複させたことが証明された。

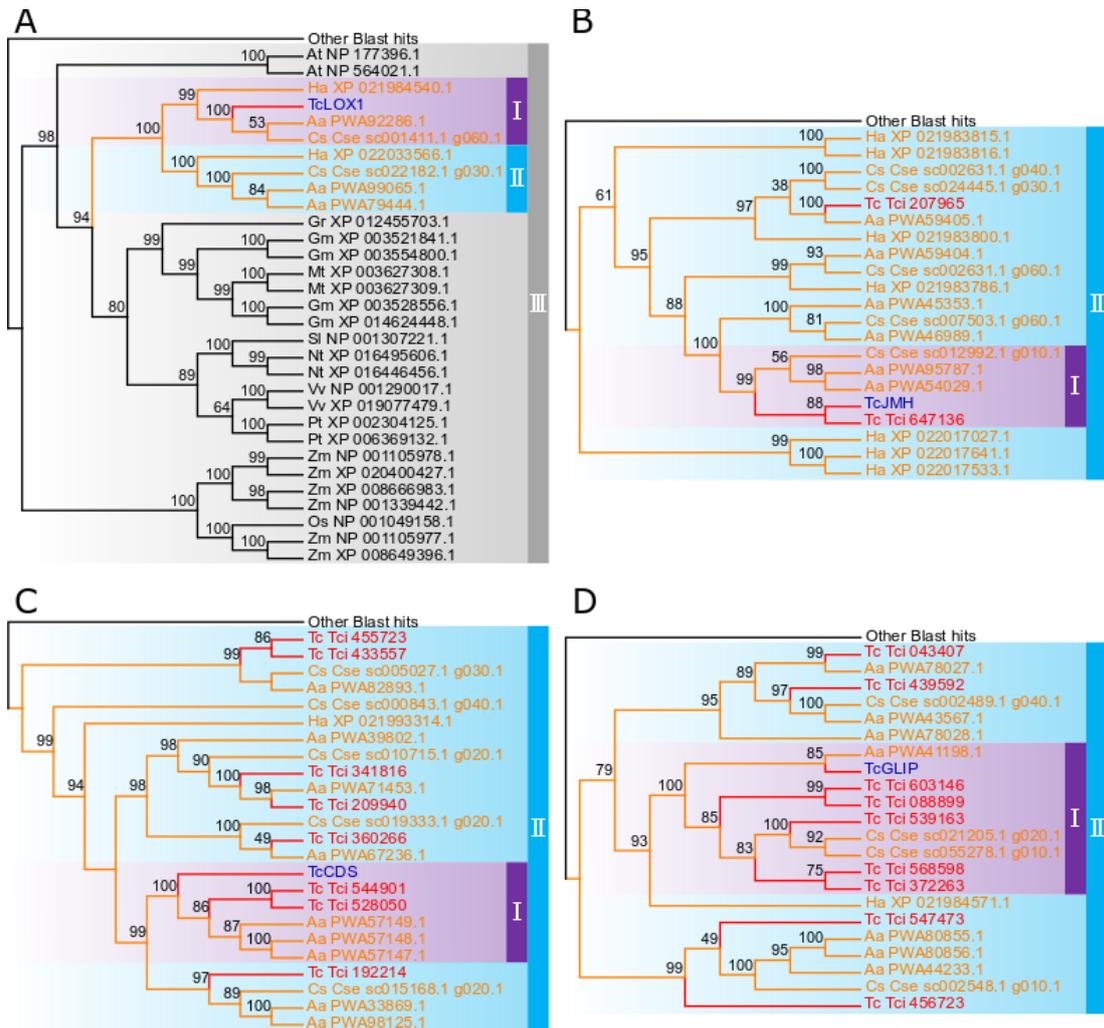


図 2-6 ピレトリン類生合成関連酵素の分子系統樹

TcLOX1 (A)、TcJMH (B)、TcCDS (C)、およびTcGLIP (D)と配列相同性を有するタンパク質の分子系統樹を示す。ピレトリン類生合成関連酵素は青色で示す。 *T. cinerariifolium* のノードは赤色、キク科植物のタンパク質およびノードは橙色で示す。TcLOX1 (TcLOX1): *T. cinerariifolium* 13-リポキシゲナーゼ、TcJMH (TcJMH): *T. cinerariifolium* ジャスモンヒドロキシラーゼ、TcCDS (TcCDS): *T. cinerariifolium* クリサンテミルニリン酸合成酵素、TcGLIP (TcGLIP): *T. cinerariifolium* GDSL (Gly-Asp-Ser-Leuモチーフ) リパーゼ

5. ピレトリン類生合成関連酵素の分布解析

近年の研究において、特有の植物代謝産物に関与している固有の代謝経路の構成要素をコードする遺伝子は、ゲノム的に代謝遺伝子クラスタ (MGCs) として共局在、共制御されていることが知られている (Field and Osbourn 2008) (Topfer, Fuchs, and Aharoni 2017)。ピレトリン類の MGCs で共局在している遺伝子を見出すために、ゲノム配列を PhytoClust (Topfer, Fuchs, and Aharoni 2017) を用いてクラスタ範囲と近隣領域をデフォルト設定で解析した。しかし、*T. cinerariifolium* ゲノム内において、ピレトリン類生合成関連酵素をコードする遺伝子座に MGCs は検出できなかった。更なる解析を可能にするために、TciLOX1、TciJMH、TciCDS、TciGLIP をコードする遺伝子座を含むスキュフォールド内の遺伝子分布を、Genomjack ソフトウェアプログラムを用いて解析した。TciLOX1 をコードする遺伝子はスキュフォールド sc00017724 に位置しており、1 つの推定タンパク質 (Tci_094327) をコードする遺伝子、4 つの TE 遺伝子が共局在していた (図 2-7A)。クエリとして Tci_094327 配列を用いた BLASTP 検索により、*A. annua* の SQUAMOSA プロモーター結合 (SPB) 転写因子 (PWA57483.1) が E-value 値が 2×10^{-35} 、52.35% の配列相同性でヒットした。*A. annua* の SPB 転写因子はアルテミシニン B の生合成に関与するタンパク質の発現を制御することから (Lv et al. 2019)、Tci_094327 はオキシリピン経路の制御因子の候補となる。TciJMH をコードする遺伝子はスキュフォールド sc00012411 に位置し、2 つの輸送タンパク質 (Tci_072426、Tci_072429) をコードする遺伝子、3 つの TE 遺伝子が共局在していた (図 2-7B)。TciCDS をコードする遺伝子はスキュフォールド sc00057709 に位置し、Tci_214196 (BLASTP で明確なホモログを欠いたタンパク質) をコードする遺伝子、1 つの TE 遺伝子が共局在していた (図 2-7C)。これらの結果から、他のピレトリン類生合成関連酵素をコードする遺伝子は、TciJMH や TciCDS の位置する遺伝子座の付近には存在していないことを示している。対照的に、TciGLIP をコードする遺伝子はスキュフォールド sc00006304 に位置し、2 つの GDSL ファミリーリパーゼ

(Tci_043407、Tci_043410)、1つのグルタチオン-S-トランスフェラーゼ (Tci_043405)、1つの推定タンパク質 (Tci_043408) をコードする遺伝子、2つの TE 遺伝子が共局在していた (図 2-7D)。Tci_043407 と Tci_043410 は、TciGLIP と各々 50.53%、41.10%の配列相同性を有しており、このことはこれらの遺伝子が遺伝子重複によって生じ、TciGLIP と同様の活性を有した酵素をコードしている可能性を示唆している。以上より、SPB 転写因子をコードする遺伝子が TciLOX1 をコードする遺伝子座の近傍に、2つの GDSSL リパーゼをコードする遺伝子が TciGLIP をコードする遺伝子座の近傍に存在していることを示しており、このことは、これら新規で発見された遺伝子が既知のピレトリン類生合成酵素をコードする遺伝子と共制御している可能性があることを示唆している。

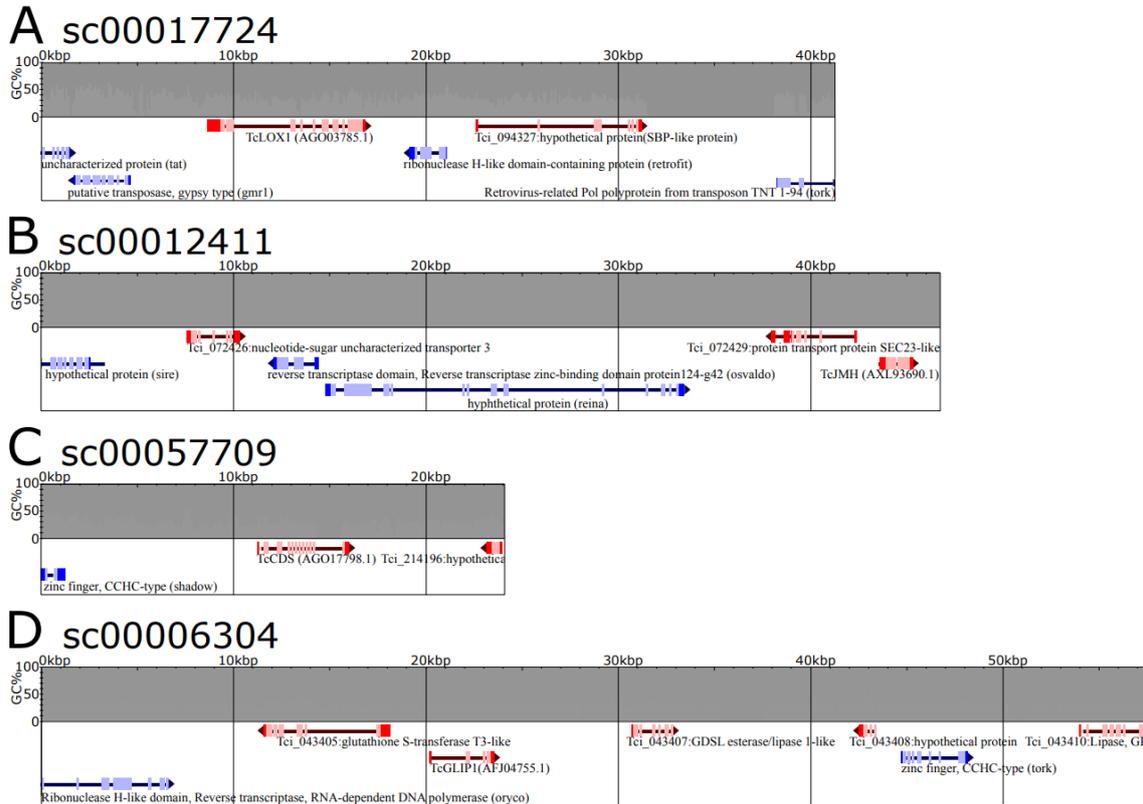


図 2-7 ピレトリン類合成関連酵素の遺伝子座を含むスキファールド

(A) TciLOX1、(B) TciJMH、(C) TciCDS、(D) TciGLIP をコードする遺伝子を含むスキファールド上に存在する予測遺伝子を、GenomeJack ソフトウェアを用いて可視化。GC 含有割合、非 TE コーディング領域、TE コーディング領域をそれぞれグレーの棒グラフ、赤色矢印、青色矢印で示す。非 TE 遺伝子上の赤色とピンク色の箱はそれぞれエキソンとタンパク質コーディング領域を示す。TE 遺伝子上の青色と水色の箱はそれぞれエキソンとタンパク質コーディング領域を示す。TcLOX1 (TciLOX1): *T. cinerariifolium* 13-リポキシゲナーゼ、TcJMH (TciJMH): *T. cinerariifolium* ジャスモンヒドロキシラーゼ、TcCDS (TciCDS): *T. cinerariifolium* クリサンテミルニリン酸合成酵素、TcGLIP (TciGLIP): *T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ、TE: 転移因子

第4節 考察

T. cinerariifolium は、他のキク科の植物 (*C. seticuspe* で 2.72 Gb、*A. annua* で 1.74 Gb、*H. annuus* で 3.6 Gb) と比べて著しく大きなゲノムサイズ (7.1 Gb) であることから、*T. cinerariifolium* 特異的な進化プロセスが生じていることが示唆される。これらの問題に取り組むために、本章において、*T. cinerariifolium* のドラフトゲノム配列の解析を行った。一般的に、*T. cinerariifolium* のような大きなサイズのゲノムのアセンブリは、不適切なシーケンシング法および反復配列の存在に起因する多数のシーケンスエラーにより困難となる (Michael and VanBuren 2015)。そのため、ゲノムの配列を高カバレッジにシーケンスし、エラーを修正するために PE リードによる 211 倍のカバレッジおよび反復配列のアセンブリを許容するために MP リードによる 56 倍のカバレッジを設けた (表 2-1)。さらに、大きなゲノムのアセンブリには PANDAseq (図 2-2) も有効であることが認められている。というのも、このプログラムには、Illumina シーケンサーに最適化されたエラー修正と、オーバーラップするペアエンド情報を用いたロングリードの作成工程が含まれているためである (Masella et al. 2012)。本研究において用いた高カバレッジのリードによるアセンブリ戦略から、7.1 Gb のゲノム配列が明らかになり、N50 は 14 Kb、BUSCO 解析によって完全性は 91.8%であると判断された (表 2-2 および 2-3)。生じたゲノム配列から、*T. cinerariifolium* ゲノムには TE および多数のスーパーファミリータンパク質をコードする遺伝子が重複していることが明らかになった。

sire クレード TE および oryco クレード TE は、*T. cinerariifolium* が他のキク科植物から分岐した後に蓄積され (図 2-3)、これらのクレードの TE はそれぞれゲノム配列の 11.16% および 2.14% を占めることが認められた (表 2-4)。sire クレード TE および oryco クレード TE の蓄積は *T. cinerariifolium* のゲノム肥大化の要因となりうる。さらに、TE はトウガラシ属 (*Capsicum*) においては植物耐病性関連遺伝子の重複に関与することが示唆されている (Kim et al. 2017)。このような背景から、*T. cinerariifolium* における

sire クレード TE および oryco クレード TE の蓄積は、InterProScan 解析によって検出されたシグナリング、生体防御、および二次代謝産物生合成に関連する酵素群の *T. cinerariifolium* 特異的な重複にも寄与していると思われる。

T. cinerariifolium で特に重複しているシグナリング関連スーパーファミリー遺伝子としてヒスチジンキナーゼが検出された (表 2-6)。このファミリーには、*A. thaliana* の ETR1 と配列相同性を有するエチレンレセプタータイプのヒスチジンキナーゼが含まれていることが示された (図 2-5B)。*T. cinerariifolium* は、植物体が傷つけられたときに特定の組み合わせの揮発性有機化合物 (VOCs) を発することによってピレトリン類生合成をアップレギュレートすることが知られている (Kikuta et al. 2011) (Ueda, Kikuta, and Matsuda 2012) ことを考慮すると、これらの結果は、*T. cinerariifolium* で重複しているヒスチジンキナーゼには、空气中に揮散される化合物への応答に関与する同キナーゼが含まれていることを示唆している。以上の結果と知見から、傷害誘導性 VOCs によるピレトリン類生合成経路の構成要素のアップレギュレーションは、ヒスチジンキナーゼ介在性シグナル伝達を介して生じることが示唆される。さらに、VOCs の生合成に関与することが知られているリポキシゲナーゼ (Scala et al. 2013) も、*T. cinerariifolium* において種特異的に重複していることが認められた (表 2-6)。これらの所見は、ピレトリン類生合成の VOCs 依存的調節機構を解明するための決定的な手掛かりとなる。

T. cinerariifolium で重複している生体防御に関連するスーパーファミリーの遺伝子には、植食性動物に対して毒性を示すリボソーム不活性化タンパク質 (RIP) が含まれていた (表 2-6)。さらに、BLATP および配列アラインメントを用いた解析から、*T. cinerariifolium* で重複している RIP には、SNA-I と配列相同性を有している殺虫性のタンパク質が含まれていることが示唆された (図 2-5A)。このことから、*T. cinerariifolium* は、ピレトリン類のような毒性低分子の産生に加えて、毒性タンパク質を生体防御分子として産生する可能性があることが示唆される。それに対して、*T. cinerariifolium* ゲノム

は、真菌類病原体に対する防御に関与することが知られているタンパク質の一種であるエンドキチナーゼをコードする遺伝子を持つのみである（表 2-7）。これらの結果は、*Stagonosporopsis tanacetii* および *Didymella tanacetii* のような真菌類によって引き起こされる深刻な葉病害に対して *T. cinerariifolium* の抵抗性が低いことと一致し（Scott et al. 2017）、*T. cinerariifolium* は栽培において、真菌類病原体感染の予防に配慮する必要があることを強調する。本ゲノム解析から、*T. cinerariifolium* は、その原産地、すなわち、現在の栽培地域より真菌類が少ないバルカン半島の乾燥した環境に適応しているという見解が裏付けられた。結果として、この植物は、進化の過程でピレトリン類の獲得に加え、植食性動物に対する効果的な防御物質として多くの RIP を獲得した種類が生き残った可能性がある。

T. cinerariifolium は、多くのリポキシゲナーゼおよびシトクロム P450 スーパーファミリーをコードしている遺伝子を重複させていることも認められ（表 2-6）、これらはピレトリン類の生合成経路に関与する。ピレトリン類の生合成に関連する特有の遺伝子の存在は、*T. cinerariifolium* ゲノムの最も顕著な特徴の 1 つである。ピレトリン類の生合成に対する酵素の分子系統解析から、ピレトリン I を合成する TciGLIP のパラログ（図 2-1）は、*T. cinerariifolium* に特異的なクレードにおいてクラスタを形成する（図 2-6D、クレード II）が、ジャスモンおよびクリサンテミルニリン酸をそれぞれ合成する TciJMH および TciCDS のパラログ（図 2-1）は、キク科植物に特異的なクレードにおいてクラスタを形成する（図 2-6B および 2-6C、クレード II）ことが証明された。それに対して、13-ヒドロペルオキシリノレン酸の合成を担い、広く存在する植物酵素の TciLOX1 関連酵素（図 2-1）は、非キク科植物のタンパク質を含むクレードにおいてクラスタを形成する（図 2-6A、クレード III）。特に、テルペノイド合成酵素として分類される TciCDS 関連酵素は、この分類の酵素が種々の種特異的テルペンを生合成することが知られている *A. annua* と同様に、*T. cinerariifolium* においても重複し変化していた（図 2-6C）。したが

って、これらの *T. cinerariifolium* の重複したテルペノイド生合成酵素は、*T. cinerariifolium* 特異的テルペノイド合成能と一致する。それに対して、ピレトリン類のアルコール部分の合成に対して重要な酵素である TciJMH は (図 2-1)、他の植物において高度に保存されていることが認められた (図 2-6B) が、ジャスモロン、ピレスロロン、およびシネロロンを含むピレトリン類のアルコール部分は、*Tanacetum* 属のメンバーにおいてのみ報告されている (Li, Zhou, and Pichersky 2018)。これらの結果から、これらのアルコール部分が、*Tanacetum* 属以外の植物では未だ確認されていないか、あるいは TciJMH 様タンパク質が他の基質に対して作用するのかどうかという疑問が浮上する。

InterProScan 解析により、*T. cinerariifolium* ゲノムでは、鉄輸送を担うタンパク質、フェリチン様スーパーファミリーのタンパク質をコードしている遺伝子が重複していることが明らかになった。さらに、種々の鉄利用酵素も重複することが認められた (表 2-6) が、これにはピレトリン類生合成経路の酵素 (TciLOX1、TciJMH、および TciPYS) を包含する分類であるリポキシゲナーゼおよびシトクロム P450 スーパーファミリーのタンパク質が含まれる。実際に、AtFer-1 は *A. thaliana* においてシトクロム P450 の発現を制御することが報告されており (Briat et al. 2010)、*T. cinerariifolium* で重複しているフェリチンをコードする遺伝子は、AtFer-1 に類似するフェリチンをコードする遺伝子座を含む (図 2-5C、Tci_154278)。これらの結果から、フェリチン様スーパーファミリーおよびピレトリン類を生合成するリポキシゲナーゼおよびシトクロム P450 が、鉄の取り込みおよび生化学的寄与において相補的役割を果たすことを考えると、これらのタンパク質は *T. cinerariifolium* において共進化してきた可能性があるとの見解が裏付けられる。したがって、これらの重複したフェリチン様スーパーファミリータンパク質と、重複した種特異的なピレトリン類生合成酵素との機能的相関は興味深い。

ピレトリン I の生合成における最終エステル化反応において決定的役割を果たす TciGLIP (図 2-1) の分子系統解析から、*T. cinerariifolium* のゲノムにおいて TciGLIP 関

連タンパク質をコードしている遺伝子が分岐したことが明らかになった (図 2-6D)。これらから、TciGLIP (および他の可能性のあるタンパク質) は、この植物属の特定の系統においてピレトリン類のエステル化における役割を獲得したという結果が示された。以前、ピレトリン酸とピレスロロン (ピレトリン II) に対する組換え TciGLIP の触媒活性は、ピレトリン酸とジャスモロン (ジャスモリン II)、ピレトリン酸とシネロロン (シネリン II)、菊酸とピレスロロン (ピレトリン I)、菊酸とジャスモロン (ジャスモリン I)、および菊酸とシネロロン (シネリン I) に対する活性より低いことが認められた (Kikuta et al. 2012)。これらの所見を合わせると、本研究から、新たに検出された TciGLIP ファミリータンパク質が、未同定のエステル化反応を触媒する可能性のあることが示唆される (図 2-1)。

特定の代謝経路の構成要素をコードする遺伝子は、特に特有の植物代謝物に対して、しばしば代謝遺伝子クラスター (MGCs) を形成することを考慮すると、代謝活性に関与する遺伝子座のゲノム中における共局在の情報も重要である。ピレトリン類の生合成に関与する既知の酵素で MGCs は検出されなかったが、SQUAMOSA プロモーター結合 (SPB) 転写因子様タンパク質をコードする遺伝子が、TciLOX1 をコードする遺伝子と共局在していることが分かった (図 2-7A)。 *A. annua* の SPB 転写因子は *A. annua* 特異のテルペノイドであるアルテミシニン B の生合成に関与するタンパク質をコードする遺伝子を制御する。TciLOX1 はピレトリン類のテルペノイド部位の生合成に関与しているのではなく、アルコール部位の生合成に関与していることから (図 2-1)、この共局在している SPB 転写因子をコードする遺伝子が、ピレトリン類の生合成に関与しているかどうかというのは興味深い所である。同様に、2つの GDSL リパーゼをコードする遺伝子が、TciGLIP をコードする遺伝子座に共局在することも分かった (図 2-7D)。これらの新たに見つかった GDSL リパーゼは、未解明のシネリン I、II、ジャスモリン I、II、ピレトリン II のエステル化反応の候補になる (図 2-1)。

本研究のゲノムおよび分子系統解析は、未同定のピレトリン類の生合成に至る経路および付随する調節機構の検証に寄与することが期待される。さらに、本研究において得られた分子基盤により、*T. cinerariifolium* の遺伝子組換え体やゲノム編集体の作成が容易になり、ピレトリン類を含め、有用な植物二次代謝物やタンパクの更なる効率的産生のための代謝工学への道が開かれる。

引用文献

- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Res*, 25: 3389-402.
- Badouin, H., J. Gouzy, C. J. Grassa, F. Murat, S. E. Staton, L. Cottret, C. Lelandais-Briere, G. L. Owens, S. Carrere, B. Mayjonade, L. Legrand, N. Gill, N. C. Kane, J. E. Bowers, S. Hubner, A. Bellec, A. Berard, H. Berges, N. Blanchet, M. C. Boniface, D. Brunel, O. Catrice, N. Chaidir, C. Claudel, C. Donnadiou, T. Faraut, G. Fievet, N. Helmstetter, M. King, S. J. Knapp, Z. Lai, M. C. Le Paslier, Y. Lippi, L. Lorenzon, J. R. Mandel, G. Marage, G. Marchand, E. Marquand, E. Bret-Mestries, E. Morien, S. Nambeesan, T. Nguyen, P. Pegot-Espagnet, N. Pouilly, F. Raftis, E. Sallet, T. Schiex, J. Thomas, C. Vandecasteele, D. Vares, F. Vear, S. Vautrin, M. Crespi, B. Mangin, J. M. Burke, J. Salse, S. Munos, P. Vincourt, L. H. Rieseberg, and N. B. Langlade. 2017. 'The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution', *Nature*, 546: 148-52.
- Bauer, E., T. Schmutzer, I. Barilar, M. Mascher, H. Gundlach, M. M. Martis, S. O. Twardziok, B. Hackauf, A. Gordillo, P. Wilde, M. Schmidt, V. Korzun, K. F. Mayer, K. Schmid, C. C. Schon, and U. Scholz. 2017. 'Towards a whole-genome sequence for rye (*Secale cereale* L.)', *Plant J*, 89: 853-69.
- Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano. 2011. 'Scaffolding pre-assembled contigs using SSPACE', *Bioinformatics*, 27: 578-9.
- Boetzer, M., and W. Pirovano. 2014. 'SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information', *BMC Bioinformatics*, 15: 211.
- Briat, J. F., K. Ravet, N. Arnaud, C. Duc, J. Boucherez, B. Touraine, F. Cellier, and F. Gaymard. 2010. 'New insights into ferritin synthesis and function highlight a link between iron homeostasis and oxidative stress in plants', *Ann Bot*, 105: 811-22.
- Canaguier, A., J. Grimplet, G. Di Gaspero, S. Scalabrin, E. Duchene, N. Choisne, N. Mohellibi, C. Guichard, S. Rombauts, I. Le Clainche, A. Berard, A. Chauveau, R. Bounon, C. Rustenholz, M. Morgante, M. C. Le Paslier, D. Brunel, and A. F. Adam-Blondon. 2017. 'A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3)', *Genom Data*, 14: 56-62.

- Chang, C., S. F. Kwok, A. B. Bleecker, and E. M. Meyerowitz. 1993. 'Arabidopsis ethylene-response gene ETR1: similarity of product to two-component regulators', *Science*, 262: 539-44.
- Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon, and M. Robles. 2005. 'Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research', *Bioinformatics*, 21: 3674-6.
- DeMicco, A., K. R. Cooper, J. R. Richardson, and L. A. White. 2010. 'Developmental neurotoxicity of pyrethroid insecticides in zebrafish embryos', *Toxicol Sci*, 113: 177-86.
- Eddy, S. R. 1998. 'Profile hidden Markov models', *Bioinformatics*, 14: 755-63.
- Field, B., and A. E. Osbourn. 2008. 'Metabolic diversification--independent assembly of operon-like gene clusters in different plants', *Science*, 320: 543-7.
- Hirakawa, H., K. Sumitomo, T. Hisamatsu, S. Nagano, K. Shirasawa, Y. Higuchi, M. Kusaba, M. Koshioka, Y. Nakano, M. Yagi, H. Yamaguchi, K. Taniguchi, M. Nakano, and S. N. Isobe. 2019. 'De novo whole-genome assembly in *Chrysanthemum seticuspe*, a model species of *Chrysanthemums*, and its application to genetic and gene discovery analysis', *DNA Res*, 26: 195-203.
- Inoue, J., and N. Satoh. 2019. 'ORTHOSCOPE: An Automatic Web Tool for Phylogenetically Inferring Bilaterian Orthogroups with User-Selected Taxa', *Mol Biol Evol*, 36: 621-31.
- Jiao, Y., P. Peluso, J. Shi, T. Liang, M. C. Stitzer, B. Wang, M. S. Campbell, J. C. Stein, X. Wei, C. S. Chin, K. Guill, M. Regulski, S. Kumari, A. Olson, J. Gent, K. L. Schneider, T. K. Wolfgruber, M. R. May, N. M. Springer, E. Antoniou, W. R. McCombie, G. G. Presting, M. McMullen, J. Ross-Ibarra, R. K. Dawe, A. Hastie, D. R. Rank, and D. Ware. 2017. 'Improved maize reference genome with single-molecule technologies', *Nature*, 546: 524-27.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. 'The rapid generation of mutation data matrices from protein sequences', *Comput Appl Biosci*, 8: 275-82.
- Jones, P., D. Binns, H. Y. Chang, M. Fraser, W. Li, C. McAnulla, H. McWilliam, J. Maslen, A. Mitchell, G. Nuka, S. Pesseat, A. F. Quinn, A. Sangrador-Vegas, M. Scheremetjew, S. Y. Yong, R. Lopez, and S. Hunter. 2014. 'InterProScan 5: genome-scale protein function classification', *Bioinformatics*, 30: 1236-40.
- Kajitani, R., K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, M. Yabana, M.

- Harada, E. Nagayasu, H. Maruyama, Y. Kohara, A. Fujiyama, T. Hayashi, and T. Itoh. 2014. 'Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads', *Genome Res*, 24: 1384-95.
- Kasprzewska, A. 2003. 'Plant chitinases--regulation and function', *Cell Mol Biol Lett*, 8: 809-24.
- Katsuda, Y. 2012. 'Progress and future of pyrethroids', *Top Curr Chem*, 314: 1-30.
- Khan, S., S. Upadhyay, F. Khan, S. Tandon, R. K. Shukla, S. Ghosh, V. Gupta, S. Banerjee, and L. Ur Rahman. 2017. 'Comparative transcriptome analysis reveals candidate genes for the biosynthesis of natural insecticide in *Tanacetum cinerariifolium*', *BMC Genomics*, 18: 54.
- Kikuta, Y., H. Ueda, K. Nakayama, Y. Katsuda, R. Ozawa, J. Takabayashi, A. Hatanaka, and K. Matsuda. 2011. 'Specific regulation of pyrethrin biosynthesis in *Chrysanthemum cinerariaefolium* by a blend of volatiles emitted from artificially damaged conspecific plants', *Plant Cell Physiol*, 52: 588-96.
- Kikuta, Y., H. Ueda, M. Takahashi, T. Mitsumori, G. Yamada, K. Sakamori, K. Takeda, S. Furutani, K. Nakayama, Y. Katsuda, A. Hatanaka, and K. Matsuda. 2012. 'Identification and characterization of a GDSL lipase-like protein that catalyzes the ester-forming reaction for pyrethrin biosynthesis in *Tanacetum cinerariifolium*- a new target for plant protection', *Plant J*, 71: 183-93.
- Kim, D., B. Langmead, and S. L. Salzberg. 2015. 'HISAT: a fast spliced aligner with low memory requirements', *Nat Methods*, 12: 357-60.
- Kim, S., J. Park, S. I. Yeom, Y. M. Kim, E. Seo, K. T. Kim, M. S. Kim, J. M. Lee, K. Cheong, H. S. Shin, S. B. Kim, K. Han, J. Lee, M. Park, H. A. Lee, H. Y. Lee, Y. Lee, S. Oh, J. H. Lee, E. Choi, E. Choi, S. E. Lee, J. Jeon, H. Kim, G. Choi, H. Song, J. Lee, S. C. Lee, J. K. Kwon, H. Y. Lee, N. Koo, Y. Hong, R. W. Kim, W. H. Kang, J. H. Huh, B. C. Kang, T. J. Yang, Y. H. Lee, J. L. Bennetzen, and D. Choi. 2017. 'New reference genome sequences of hot pepper reveal the massive evolution of plant disease-resistance genes by retroduplication', *Genome Biol*, 18: 210.
- Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Mol Biol Evol*, 35: 1547-49.
- Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, A. S. Karthikeyan, C. H. Lee, W. D. Nelson, L.

- Ploetz, S. Singh, A. Wensel, and E. Huala. 2012. 'The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools', *Nucleic Acids Res*, 40: D1202-10.
- Langdon, W. B. 2015. 'Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks', *BioData Min*, 8: 1.
- Li, K. B. 2003. 'ClustalW-MPI: ClustalW analysis using distributed and parallel computing', *Bioinformatics*, 19: 1585-6.
- Li, W., D. B. Lybrand, F. Zhou, R. L. Last, and E. Pichersky. 2019. 'Pyrethrin Biosynthesis: The Cytochrome P450 Oxidoreductase CYP82Q3 Converts Jasmolone To Pyrethrolone', *Plant Physiol*, 181: 934-44.
- Li, W., F. Zhou, and E. Pichersky. 2018. 'Jasmone Hydroxylase, a Key Enzyme in the Synthesis of the Alcohol Moiety of Pyrethrin Insecticides', *Plant Physiol*, 177: 1498-509.
- Liu, P. L., J. N. Wan, Y. P. Guo, S. Ge, and G. Y. Rao. 2012. 'Adaptive evolution of the chrysanthemyl diphosphate synthase gene involved in irregular monoterpene metabolism', *BMC Evol Biol*, 12: 214.
- Llorens, C., R. Futami, L. Covelli, L. Dominguez-Escriba, J. M. Viu, D. Tamarit, J. Aguilar-Rodriguez, M. Vicente-Ripolles, G. Fuster, G. P. Bernet, F. Maumus, A. Munoz-Pomer, J. M. Sempere, A. Latorre, and A. Moya. 2011. 'The Gypsy Database (GyDB) of mobile genetic elements: release 2.0', *Nucleic Acids Res*, 39: D70-4.
- Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S. M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T. W. Lam, and J. Wang. 2012. 'SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler', *Gigascience*, 1: 18.
- Lv, Z., Y. Wang, Y. Liu, B. Peng, L. Zhang, K. Tang, and W. Chen. 2019. 'The SPB-Box Transcription Factor AaSPL2 Positively Regulates Artemisinin Biosynthesis in *Artemisia annua* L', *Front Plant Sci*, 10: 409.
- Masella, A. P., A. K. Bartram, J. M. Truszkowski, D. G. Brown, and J. D. Neufeld. 2012. 'PANDAsseq: paired-end assembler for illumina sequences', *BMC Bioinformatics*, 13: 31.
- Michael, T. P., and R. VanBuren. 2015. 'Progress, challenges and the future of crop genomes', *Curr Opin Plant Biol*, 24: 71-81.
- Nadalin, F., F. Vezzi, and A. Policriti. 2012. 'GapFiller: a de novo assembly approach to fill the

- gap within paired reads', *BMC Bioinformatics*, 13 Suppl 14: S8.
- Obrig, T. G., J. D. Irvin, and B. Hardesty. 1973. 'The effect of an antiviral peptide on the ribosomal reactions of the peptide elongation enzymes, EF-I and EF-II', *Arch Biochem Biophys*, 155: 278-89.
- Paterson, A. H., J. F. Wendel, H. Gundlach, H. Guo, J. Jenkins, D. Jin, D. Llewellyn, K. C. Showmaker, S. Shu, J. Udall, M. J. Yoo, R. Byers, W. Chen, A. Doron-Faigenboim, M. V. Duke, L. Gong, J. Grimwood, C. Grover, K. Grupp, G. Hu, T. H. Lee, J. Li, L. Lin, T. Liu, B. S. Marler, J. T. Page, A. W. Roberts, E. Romanel, W. S. Sanders, E. Szadkowski, X. Tan, H. Tang, C. Xu, J. Wang, Z. Wang, D. Zhang, L. Zhang, H. Ashrafi, F. Bedon, J. E. Bowers, C. L. Brubaker, P. W. Chee, S. Das, A. R. Gingle, C. H. Haigler, D. Harker, L. V. Hoffmann, R. Hovav, D. C. Jones, C. Lemke, S. Mansoor, M. ur Rahman, L. N. Rainville, A. Rambani, U. K. Reddy, J. K. Rong, Y. Saranga, B. E. Scheffler, J. A. Scheffler, D. M. Stelly, B. A. Triplett, A. Van Deynze, M. F. Vaslin, V. N. Waghmare, S. A. Walford, R. J. Wright, E. A. Zaki, T. Zhang, E. S. Dennis, K. F. Mayer, D. G. Peterson, D. S. Rokhsar, X. Wang, and J. Schmutz. 2012. 'Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres', *Nature*, 492: 423-7.
- Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. 'FastTree 2--approximately maximum-likelihood trees for large alignments', *PLoS One*, 5: e9490.
- Ramirez, A. M., N. Saillard, T. Yang, M. C. Franssen, H. J. Bouwmeester, and M. A. Jongsma. 2013. 'Biosynthesis of sesquiterpene lactones in pyrethrum (*Tanacetum cinerariifolium*)', *PLoS One*, 8: e65030.
- Ramirez, A. M., T. Yang, H. J. Bouwmeester, and M. A. Jongsma. 2013. 'A trichome-specific linoleate lipoxygenase expressed during pyrethrin biosynthesis in pyrethrum', *Lipids*, 48: 1005-15.
- Rice Annotation, Project, T. Tanaka, B. A. Antonio, S. Kikuchi, T. Matsumoto, Y. Nagamura, H. Numa, H. Sakai, J. Wu, T. Itoh, T. Sasaki, R. Aono, Y. Fujii, T. Habara, E. Harada, M. Kanno, Y. Kawahara, H. Kawashima, H. Kubooka, A. Matsuya, H. Nakaoka, N. Saichi, R. Sanbonmatsu, Y. Sato, Y. Shinso, M. Suzuki, J. Takeda, M. Tanino, F. Todokoro, K. Yamaguchi, N. Yamamoto, C. Yamasaki, T. Imanishi, T. Okido, M. Tada, K. Ikeo, Y. Tateno, T. Gojobori, Y. C. Lin, F. J. Wei, Y. I. Hsing, Q. Zhao, B. Han, M. R. Kramer, R. W. McCombie, D. Lonsdale, C. C. O'Donovan, E. J. Whitfield, R. Apweiler, K. O.

- Koyanagi, J. P. Khurana, S. Raghuvanshi, N. K. Singh, A. K. Tyagi, G. Haberer, M. Fujisawa, S. Hosokawa, Y. Ito, H. Ikawa, M. Shibata, M. Yamamoto, R. M. Bruskiwich, D. R. Hoen, T. E. Bureau, N. Namiki, H. Ohyanagi, Y. Sakai, S. Nobushima, K. Sakata, R. A. Barrero, Y. Sato, A. Souvorov, B. Smith-White, T. Tatusova, S. An, G. An, O. Ota S, G. Fuks, G. Fuks, J. Messing, K. R. Christie, D. Lieberherr, H. Kim, A. Zuccolo, R. A. Wing, K. Nobuta, P. J. Green, C. Lu, B. C. Meyers, C. Chaparro, B. Piegu, O. Panaud, and M. Echeverria. 2008. 'The Rice Annotation Project Database (RAP-DB): 2008 update', *Nucleic Acids Res*, 36: D1028-33.
- Rivera, S. B., B. D. Swedlund, G. J. King, R. N. Bell, C. E. Hussey, Jr., D. M. Shattuck-Eidens, W. M. Wrobel, G. D. Peiser, and C. D. Poulter. 2001. 'Chrysanthemyl diphosphate synthase: isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*', *Proc Natl Acad Sci US A*, 98: 4373-8.
- Scala, A., S. Allmann, R. Mirabella, M. A. Haring, and R. C. Schuurink. 2013. 'Green leaf volatiles: a plant's multifunctional weapon against herbivores and pathogens', *Int J Mol Sci*, 14: 17781-811.
- Schmutz, J., S. B. Cannon, J. Schlueter, J. Ma, T. Mitros, W. Nelson, D. L. Hyten, Q. Song, J. J. Thelen, J. Cheng, D. Xu, U. Hellsten, G. D. May, Y. Yu, T. Sakurai, T. Umezawa, M. K. Bhattacharyya, D. Sandhu, B. Valliyodan, E. Lindquist, M. Peto, D. Grant, S. Shu, D. Goodstein, K. Barry, M. Futrell-Griggs, B. Abernathy, J. Du, Z. Tian, L. Zhu, N. Gill, T. Joshi, M. Libault, A. Sethuraman, X. C. Zhang, K. Shinozaki, H. T. Nguyen, R. A. Wing, P. Cregan, J. Specht, J. Grimwood, D. Rokhsar, G. Stacey, R. C. Shoemaker, and S. A. Jackson. 2010. 'Genome sequence of the palaeopolyploid soybean', *Nature*, 463: 178-83.
- Scott, J. B., D. H. Gent, T. L. Pearce, S. J. Pethybridge, S. J. Pilkington, and F. S. Hay. 2017. 'Mycoflora Associated With Pyrethrum Seed and the Integration of Seed Steam Treatment Into Foliar Disease Management Strategies', *Plant Dis*, 101: 1874-84.
- Shahidi-Noghabi, S., E. J. Van Damme, and G. Smaghe. 2008. 'Carbohydrate-binding activity of the type-2 ribosome-inactivating protein SNA-I from elderberry (*Sambucus nigra*) is a determining factor for its insecticidal activity', *Phytochemistry*, 69: 2972-8.
- Sharma, N., S. W. Park, R. Vepachedu, L. Barbieri, M. Ciani, F. Stirpe, B. J. Savary, and J. M. Vivanco. 2004. 'Isolation and characterization of an RIP (ribosome-inactivating

- protein)-like protein from tobacco with dual enzymatic activity', *Plant Physiol*, 134: 171-81.
- Shearer, L. A., L. K. Anderson, H. de Jong, S. Smit, J. L. Goicoechea, B. A. Roe, A. Hua, J. J. Giovannoni, and S. M. Stack. 2014. 'Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome', *G3 (Bethesda)*, 4: 1395-405.
- Shen, Q., L. Zhang, Z. Liao, S. Wang, T. Yan, P. Shi, M. Liu, X. Fu, Q. Pan, Y. Wang, Z. Lv, X. Lu, F. Zhang, W. Jiang, Y. Ma, M. Chen, X. Hao, L. Li, Y. Tang, G. Lv, Y. Zhou, X. Sun, P. E. Brodelius, J. K. C. Rose, and K. Tang. 2018. 'The Genome of *Artemisia annua* Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis', *Mol Plant*, 11: 776-88.
- Sierro, N., J. N. Battey, S. Ouadi, N. Bakaher, L. Bovet, A. Willig, S. Goepfert, M. C. Peitsch, and N. V. Ivanov. 2014. 'The tobacco genome sequence and its comparison with those of tomato and potato', *Nat Commun*, 5: 3833.
- Siljak-Yakovlev, S., F. Pustahija, E. M. Šolić, F. Bogunić, E. Muratović, N. Bašić, O. Catrice, and S. C. Brown. 2010. 'Towards a Genome Size and Chromosome Number Database of Balkan Flora: C-Values in 343 Taxa with Novel Values for 242', *Advanced Science Letters*, 3: 24.
- Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. 'Using native and syntenically mapped cDNA alignments to improve de novo gene finding', *Bioinformatics*, 24: 637-44.
- Sudre, D., E. Gutierrez-Carbonell, G. Lattanzio, R. Rellan-Alvarez, F. Gaymard, G. Wohlgemuth, O. Fiehn, A. Alvarez-Fernandez, A. M. Zamarreno, E. Bacaicoa, D. Duy, J. M. Garcia-Mina, J. Abadia, K. Philippar, A. F. Lopez-Millan, and J. F. Briat. 2013. 'Iron-dependent modifications of the flower transcriptome, proteome, metabolome, and hormonal content in an *Arabidopsis* ferritin mutant', *J Exp Bot*, 64: 2665-88.
- Tang, H., V. Krishnakumar, S. Bidwell, B. Rosen, A. Chan, S. Zhou, L. Gentsbittel, K. L. Childs, M. Yandell, H. Gundlach, K. F. Mayer, D. C. Schwartz, and C. D. Town. 2014. 'An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*', *BMC Genomics*, 15: 312.
- Topfer, N., L. M. Fuchs, and A. Aharoni. 2017. 'The PhytoClust tool for metabolic gene clusters discovery in plant genomes', *Nucleic Acids Res*, 45: 7049-63.

- Tuskan, G. A., S. Difazio, S. Jansson, J. Bohlmann, I. Grigoriev, U. Hellsten, N. Putnam, S. Ralph, S. Rombauts, A. Salamov, J. Schein, L. Sterck, A. Aerts, R. R. Bhalerao, R. P. Bhalerao, D. Blaudez, W. Boerjan, A. Brun, A. Brunner, V. Busov, M. Campbell, J. Carlson, M. Chalot, J. Chapman, G. L. Chen, D. Cooper, P. M. Coutinho, J. Couturier, S. Covert, Q. Cronk, R. Cunningham, J. Davis, S. Degroeve, A. Dejardin, C. Depamphilis, J. Detter, B. Dirks, I. Dubchak, S. Duplessis, J. Ehlting, B. Ellis, K. Gendler, D. Goodstein, M. Gribskov, J. Grimwood, A. Groover, L. Gunter, B. Hamberger, B. Heinze, Y. Helariutta, B. Henrissat, D. Holligan, R. Holt, W. Huang, N. Islam-Faridi, S. Jones, M. Jones-Rhoades, R. Jorgensen, C. Joshi, J. Kangasjarvi, J. Karlsson, C. Kelleher, R. Kirkpatrick, M. Kirst, A. Kohler, U. Kalluri, F. Larimer, J. Leebens-Mack, J. C. Leple, P. Locascio, Y. Lou, S. Lucas, F. Martin, B. Montanini, C. Napoli, D. R. Nelson, C. Nelson, K. Nieminen, O. Nilsson, V. Pereda, G. Peter, R. Philippe, G. Pilate, A. Poliakov, J. Razumovskaya, P. Richardson, C. Rinaldi, K. Ritland, P. Rouze, D. Ryaboy, J. Schmutz, J. Schrader, B. Segerman, H. Shin, A. Siddiqui, F. Sterky, A. Terry, C. J. Tsai, E. Uberbacher, P. Unneberg, J. Vahala, K. Wall, S. Wessler, G. Yang, T. Yin, C. Douglas, M. Marra, G. Sandberg, Y. Van de Peer, and D. Rokhsar. 2006. 'The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)', *Science*, 313: 1596-604.
- Ueda, H., Y. Kikuta, and K. Matsuda. 2012. 'Plant communication: mediated by individual or blended VOCs?', *Plant Signal Behav*, 7: 222-6.
- Vogel, A., R. Schwacke, A. K. Denton, B. Usadel, J. Hollmann, K. Fischer, A. Bolger, M. H. Schmidt, M. E. Bolger, H. Gundlach, K. F. X. Mayer, H. Weiss-Schneeweiss, E. M. Temsch, and K. Krause. 2018. 'Footprints of parasitism in the genome of the parasitic flowering plant *Cuscuta campestris*', *Nat Commun*, 9: 2515.
- Walsh, M. J., J. E. Dodd, and G. M. Hautbergue. 2013. 'Ribosome-inactivating proteins: potent poisons and molecular tools', *Virulence*, 4: 774-84.
- Waterhouse, R. M., M. Seppey, F. A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva, and E. M. Zdobnov. 2018. 'BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics', *Mol Biol Evol*, 35: 543-48.
- Xu, H., W. Li, A. L. Schillmiller, H. van Eekelen, R. C. H. de Vos, M. A. Jongasma, and E. Pichersky. 2019. 'Pyrethric acid of natural pyrethrin insecticide: complete pathway

- elucidation and reconstitution in *Nicotiana benthamiana*', *New Phytol*, 223: 751-65.
- Xu, H., G. D. Moghe, K. Wiegert-Rininger, A. L. Schillmiller, C. S. Barry, R. L. Last, and E. Pichersky. 2018. 'Coexpression Analysis Identifies Two Oxidoreductases Involved in the Biosynthesis of the Monoterpene Acid Moiety of Natural Pyrethrin Insecticides in *Tanacetum cinerariifolium*', *Plant Physiol*, 176: 524-37.
- Yang, J., G. Zhang, J. Zhang, H. Liu, W. Chen, X. Wang, Y. Li, Y. Dong, and S. Yang. 2017. 'Hybrid de novo genome assembly of the Chinese herbal fleabane *Erigeron breviscapus*', *Gigascience*, 6: 1-7.
- Zhu, Feng, Shu Yuan, Zhong-Wei Zhang, Kun Qian, Jian-Guo Feng, and Yi-Zhong Yang. 2016. 'Pokeweed antiviral protein (PAP) increases plant systemic resistance to Tobacco mosaic virus infection in *Nicotiana benthamiana*', *European Journal of Plant Pathology*, 146: 9.
- Zhuang, Y., and E. A. Tripp. 2017. 'The draft genome of *Ruellia speciosa* (Beautiful Wild Petunia: Acanthaceae)', *DNA Res*, 24: 179-92.
- Zito, S. W., R. G. Zieg, and E. J. Staba. 1983. 'Distribution of pyrethrins in oil glands and leaf tissue of *Chrysanthemum cinerariaefolium*', *Planta Med*, 47: 205-7.

第3章 *T. coccineum* ゲノムの解読

第1節 序論

第2章ではピレトリン類高産生種の *Tanacetum cinerariifolium* のドラフトゲノムを明らかにした。*T. cinerariifolium* は、植物体が傷を受けた時に特定の揮発性有機化合物 (VOCs) を放出し、ピレトリン類の生合成が上昇することが知られている。興味深いことに、*T. cinerariifolium* において VOC 誘導性ヒスチジンキナーゼをはじめとする多くのヒスチジンキナーゼがコードされていた。このことは、ヒスチジンキナーゼとピレトリン類生合成に機能的な相関があることを示唆している。さらに、*T. cinerariifolium* は、リボソーム不活性化タンパク質 (RIP) コード遺伝子を他種より多く有している一方で、病原性真菌に対する防御に關与するエンドキチナーゼをコードする遺伝子の数は他種より少なくなっている。防御関連遺伝子に関する上記の性質から、*T. cinerariifolium* は、その原産地域であるダルマチア地方の乾燥した環境に存在する捕食者から防御できるよう、特有の進化を遂げたことが示唆された。

同属の種である *Tanacetum coccineum* は、*T. cinerariifolium* と系統発生的な関連が強いにも関わらず、ピレトリン類の産生量がかなり少ない (Katsuda 2012)。また、*T. coccineum* は、温度や湿度が高いペルシャからコーカサス地方に至る地域を原産としている一方で、*T. cinerariifolium* は比較的乾燥したダルマチア地方を原産としている。このように異なる気候の地域に分布することを考慮すると、上記2種の植物が環境によって異なる外敵に対して、それぞれに特異的な防御戦略を獲得した可能性を示唆している。そこで、これら2種のゲノムを比較することで、防御戦略の進化過程と多様化について明らかにすることができ、また、より効率的なピレトリン類生産や他の生体防御物質を同定することにつながると期待される。

本章において、*T. coccineum* のドラフトゲノムを明らかにし、*T. cinerariifolium* とのゲノムの比較解析を実施した。

第2節 実験の部

1. 本研究で引用した植物の系統解析

シロバナムシヨケギク (*T. cinerariifolium*)、アカバナムシヨケギク (*T. coccineum*)、クソニンジン (*Artemisia annua*)、キクタニギク (*Chrysanthemum seticuspe*)、ヒマワリ (*Helianthus annuus*)、タバコ (*Nicotiana tabacum*)、シロイヌナズナ (*Arabidopsis thaliana*) およびイネ (*Oryza sativa*) の内部転写スペーサー領域 (ITS) 1、ITS2、リブロース 1,5-ビスリン酸カルボキシラーゼ/オキシゲナーゼ大サブユニット (*rbcL*)、マチュラーゼ K (*matK*) の配列を NCBI データベース又は BLASTN から取得した。使用した塩基配列の accession No. は表 3-1 に示した。*C. seticuspe* の *ITS1*、*ITS2*、*rbcL*、*matK* の塩基配列は、*C. seticuspe* ゲノム配列 (CSE_r1.0, (Hirakawa et al. 2019)) に対して BLASTN (バージョン 2.7.1) (Altschul et al. 1997) を使用し、シロイヌナズナのものをクエリとして検出した。他植物との配列相同性の低かったイネを除き、塩基配列のアラインメントについて CLUSTAL W-mpi 0.13 (Li 2003) を用い、マトリクススペースの JTT 行列モデル (Jones, Taylor, and Thornton 1992) に基づいた最尤系統樹を、IQ-TREE 2.0.3 (Minh et al. 2020) および FigTree v1.4.4 ソフトウェア (<http://tree.bio.ed.ac.uk/software/figtree/>) を用いてブートストラップ反復回数 1000 回で作成した (図 3-1)。

表 3-1 本研究で引用した植物の系統解析に使用した塩基配列の accession No.

<i>species</i>	<i>ITS1</i>	<i>ITS2</i>	<i>rbcL</i>	<i>matK</i>
<i>T. cinerariifolium</i>	AB359720.1	AB359806.1	MT104464.1	MT104464.1
<i>T. coccineum</i>	AB359721.1	AB359807.1	MT104463.1	MT104463.1
<i>A. annua</i>	KC493085.1	KC493085.1	MF623173.1	MF623173.1
<i>H. annuus</i>	KF767534.1	KF767534.1	L13929.1	AY215805.1
<i>N. tabacum</i>	AJ300215.1	AJ300215.1	AP019625.1	MZ707522.1
<i>A. thaliana</i>	X52320.1	X52320.1	NC_000932.1	MK380721.1
<i>O. sativa</i>	KM036282.1	KM036282.1	D00207.1	KM103369.1

ITS1: 内部転写スペーサー領域 1、ITS2: 内部転写スペーサー領域 2、*rbcL*: リブロース 1,5-ビスリン酸カルボキシラーゼ/オキシゲナーゼ大サブユニット、*matK*: マチュラーゼ K

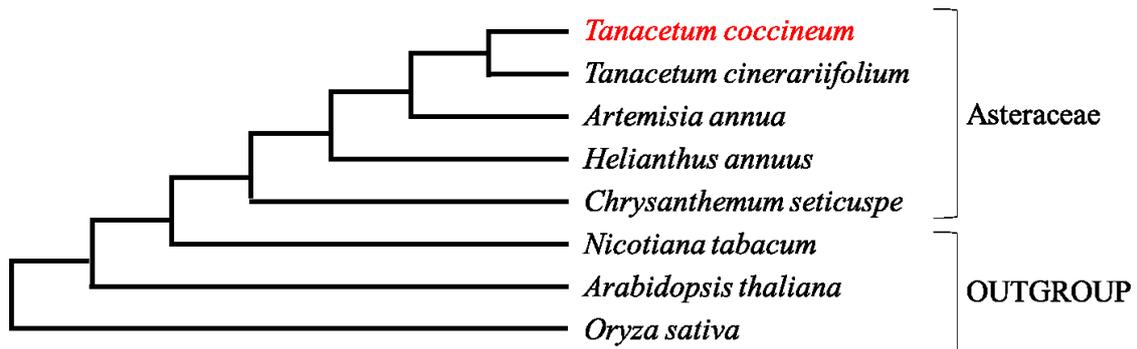


図 3-1 本研究で引用した植物の分子系統樹

系統樹の作成にあたって、各植物の内部転写スペーサー領域 (ITS) 1、ITS2、リブロース 1,5-ピスリン酸カルボキシラーゼ/オキシゲナーゼ大サブユニット (*rbcL*)、マチュラーゼ K (*matK*) の塩基配列を、NCBI データベースもしくは BLASTN を用いて取得した。*ITS1*、*ITS2*、*rbcL*、*matK* の塩基配列は CLUSTAL W-mpi 0.13 でアラインメントを行い (配列相同性の低い *O. sativa* のものを除く)、JTT 行列を用いたモデル (1000 回のブートストラップを実施) に基づく最尤系統樹を、IQ-TREE 2.0.3 および FigTree v1.4.4 ソフトウェアを用いて作成した。

2. 植物材料およびゲノムシーケンシング

T. coccineum (栽培品種: ロビンソンミックス) の種子を株式会社サカタのタネから購入した (カタログ番号: 906435)。DNeasy Plant Mini Kit (Qiagen) を、付属のプロトコルに従い用いて種子からゲノム DNA を抽出した。抽出した DNA から、TruSeq DNA PCR-Free キット (Illumina) を用いてショートインサートライブラリ (以下、PE)、Nextera Mate-Pair Sample Prep Kit (Illumina) を用いて 3 つの異なるインサートサイズ (3 Kb、5 Kb、および 8 Kb) のメイトペアライブラリ (それぞれ MP-3kb、MP-5Kb および MP-8Kb) をそれぞれ構築した。その後、PE および各 MP ライブラリに対し、NovaSeq 6000 Illumina 装置を用いて 151 bp × 2 サイクルのペアエンドシーケンスを行った。また、抽出した DNA より Miseq 用ライブラリ (以下、MS) を TrueSeq DNA PCR-Free キットを用いて構築した。MS ライブラリに対し、Illumina Miseq システムを用いて 301 bp × 2 サイクルのペアエンドシーケンスを行った。さらに、抽出した DNA より PacBio 用ライブラリ (以下、PB) を SMRTbell Express Template Prep Kit (PacBio 社) を用いて構築した。PB ライブラリに対し、PacBio Sequel II システムを用いてシーケンスを行った。

3. フローサイトメトリーを用いたゲノムサイズ推定

T. coccineum を野外 (大阪府豊中市) で種子より 6 カ月育成し、得られた植物体から葉を取得した。カミソリを用いて葉より 5mm 平方の切片を切り出し、Quantum Stain UV and PI for DNA (Quantum Analysis 社) を用い、付属のプロトコルに従い処理を行った。*T. coccineum* のゲノムサイズをサイトメーターCyFlow SL (Sysmex Partec 社) を用いて推定した。*T. cinerariifolium* (第 2 章第 2 節の 1 で使用した野生型 7.1 Gb) および *C. seticuspe* (3 Gb、栽培品種: Gojo-0) (Hirakawa et al. 2019) の葉の断片に対してもアッセイを行い、ゲノムサイズの対照とした。本研究で使用した *C. seticuspe* Gojo-0 とその関連情報については、ナショナルバイオリソースプロジェクトより得られる (<https://shigen.nig.ac.jp/chrysanthemum/top.jsp>)。

4. k-mer 深度解析を用いたゲノムサイズ推定

さらに、ゲノムサイズを k-mer 深度解析 (Alm Rosenblad et al. 2021)を用いて推定した。各 k-mer の出現数を Jellyfish ソフトウェア (Marcais and Kingsford 2011)で計測し、ホモ接合配列の k-mer 深度ピーク (C) を k-mer 深度のヒストグラムを用いて求めた。その後、ゲノムサイズを以下の式で求めた。

$$\text{ゲノムサイズ} = \frac{n * (L - k + 1)}{C}$$

n、L、k はそれぞれリード合計数、平均リード長、k-mer サイズを示す。

5. 遺伝子予測およびアノテーション

第 2 章と同様に、取得したリード配列のクリーニングを行った。Truseq 又は Nexter Mate Pair Sample Prep Kit 由来のアダプター、クオリティが低いリード、36 bp 未満の長さの短いリードを、Trimmomatic version 0.36 (Langdon 2015)を用いてトリミングした。PE リードおよび MS リードからコンティグ配列を 3 段階のプロセス (「プレアセンブリ」、「コンティグアセンブリ」、「クリーンアップ& 併合」) により作成した。第 1 段階ではオーバーラップを含むクリーンな PE リードについて PANDAsq (Masella et al. 2012)を用いてプレアセンブリした。第 2 段階では、第 1 段階でプレアセンブルしたリードと残ったリードで、80~127 の複数 k-mer に対し SOAPdenovo v2.04-r240 (Luo et al. 2012)を用いることでコンティグ配列を作成した。その後、SOAPdenovo で作成したコンティグ配列について、他のコンティグ配列との配列相同性 95%以上であるコンティグ配列を除去および併合することでクリーンアップし、コンティグ配列とした。

第 2 章とは異なり、PB リードを使用した「ハイブリッドアセンブル」のプロセスを取り入れてプレスキャフールドを作成した。このとき、SSPACE-longread v1-1 (Boetzer and Pirovano 2014)を用い、設定は最小オーバーラップ長を 20 bp、最小リンク数を 3、最大リンク割合を 0.3 とした。

スキャフールド配列は、3 段階 (「スキャフールドイング」、「ギャップフィリング」、

「ポリッシング」)のプロセスで作成した。まず第 1 段階では、PE および MP リードを bowtie バージョン 2.3.4.3 (Langdon 2015)を用い、`--local` オプションでマップし、コンテ ィグ間を跨ぐようにマップされたリードペアを選択した。SSPACE-STANDARD version 3.0(Boetzer et al. 2011) (BaseClear) を用いて、最小リンク数を 3 に設定して、選択した リードおよびコンテ ィグをスキヤフォールディングした。第 2 段階で、GapFiller v1-10 (Nadalin, Vezzi, and Policriti 2012) (BaseClear) を用い、最小オーバーラップ数を 30、 トリミングリード数を 10 に設定して埋め、さらに PB リードを使用してパラメーターをデ フォルト値とした TGS-Gapcloser (Xu et al. 2020)を用いて埋めた。第 3 段階では、スキヤ フォールドを POLCA プログラム (Zimin and Salzberg 2020)を用いてポリッシングし、ド ラフトゲノム配列を完成させた。以上のシーケンスリードおよび完成したドラフトゲノム 配列は DDBJ の bioproject accession code PRJDB12537 下にアップロードした。次に、こ のドラフトゲノム配列の完全性を、embryophyta_odb9 タンパク質セットを用い、BUSCO- v5 (Manni et al. 2021)で評価した。

6. 遺伝子の予測およびアノテーション

既にゲノムが解明されている 6 つの植物、*A. thaliana* (TAIR10 (Lamesch et al. 2012))、 *N. tabacum* (Ntab-TN90 (Sierro et al. 2014))、*O. sativa* (assembly Build 4.0 (Rice Annotation et al. 2008))、*H. annuus* (HA412HO_v1.1 (Badouin et al. 2017))、*A. annua* (ASM311234v1 (Shen et al. 2018))、*C. seticuspe* (CSE_r1.0 (Hirakawa et al. 2019)) のゲノムについて転移因子 (Transposable elements, TE) の含有量を推定し、*T. coccineum* との比較を行った。シロイヌナズナモデルセットおよびパラメーターをデフォルト値とし た AUGUSTUS 3.3.1 (Stanke et al. 2008)でこれら植物の遺伝子コーディング領域を推定 し、HMMER 2.3.1 (Eddy 1998)における hmmpfam を用いて GyDB (Llorens et al. 2011) と対比させ、予測された遺伝子における TE を検出した。GyDB の分類に従って、ゲノム領 域における TE の各クレードが占める割合を蓄積スコアとして算出し、第 2 章の *T*

cinerariifolium ゲノムの結果もあわせて比較を行った。

del および tork クレード TE の分子系統樹も、ORTHOSCOPE 法 (Inoue and Satoh 2019)を用いて第2章と同様に推定した。hmmpfam で抽出した *T. cinerariifolium*、*A. annua*、*C. seticuspe*、*H. annuus* の TE 内にコードされている逆転写酵素ドメインのアミノ酸配列とともに、CLUSTAL W-mpi 0.13 を用いてアライメントし、JTT 行列を用いたモデルに基づく最尤系統樹を、Fast Tree 2.1.10 (JTT モデル、CAT 近似) を用いてブートストラップ反復回数 100 回で作成した。

7. ピレトリン類生合成関連酵素をコードする遺伝子の相同性検索およびシンテニー解析

T. coccineum で予測されるタンパク質において、*T. cinerariifolium* の既知のピレトリン類生合成関連酵素と配列相同性が高いものを BLASTP 2.7.1 を用いて検出した。*T. cinerariifolium* のピレトリン類生合成関連酵素 TciADH2 (accession No. AUQ44118.1)、TciALDH1 (accession No. AUQ44119.1)、TciCCH (accession No. AGO03787.1)、TciCCMT (accession No. QCP80351.1)、TciCDS (accession No. ADO17798.1)、TciGLIP (accession No. AFJ04755.1)、TciJMH (accession No. AXL93690.1)、TciLOX1 (accession No. AGO03785.1) および TciPYS (accession No. AXL93709.1) との全体的な配列相同性が 85% 超であるトップヒット配列を、*T. coccineum* のピレトリン類生合成関連酵素とした。検出された各タンパク質に対して、CLUSTAL W-mpi 0.13 を用いた配列アライメント、InterProScan 5.33-72.0 を用いたドメイン検索および Genomjack ソフトウェアプログラム (三菱スペース・ソフトウェア、東京、日本) を用いたシンテニー解析を行った。

8. タンパク質スーパーファミリー含有量と他の植物におけるそれとの比較解析

本章第2節6で述べた、既にゲノムが解明されている植物6種および *T. cinerariifolium* に対し、InterProScan 解析を用い、タンパク質群の特徴を解析した。各スーパーファミリーの特徴を有する遺伝子の数を求めてから、InterProScan で検出したスーパーファミリーの特徴 (Sig) および植物の属 (Genus) の組み合わせに対する重複オッズスコアを第2章

と同様に、以下のように算出した。

$$\text{Multiplication odds score (Genus, Sig)} = \log_2 \frac{N(\text{Genus, Sig}) + PS}{N(\text{Sig}) + PS}$$

式中、N (Genus, Sig) はある植物の属 (Genus) における、InterProScan によって検出されたスーパーファミリーの特徴 (Sig) を有する遺伝子数を表し、PS は疑似カウント定数を表し、これを 0.5 に設定した。

機能性タンパク質を更に解析すべく、RIP 関連スーパーファミリーの予測タンパク質に対し、BLASTP 2.7.1 を行った。RIP のクエリには SNA-I (セイヨウニワトコ *Sambucus nigra* アグルチニン I、accession No. O22415.1) を用いた。SNA-I と *T. coccineum* にて検出したタンパク質 Tco_1336120 および *T. cinerariifolium* の Tci_399175 (第 2 章参照、accession No. GEY27201.1) に対して、CLUSTAL W-mpi 0.13 を用いてアミノ酸配列をアライメントした。

T. coccineum および *T. cinerariifolium* のヒスチジinkinナーゼ関連スーパーファミリーの予測タンパク質に対して、Conserved Domain Database (CDD v.3.19) (Lu et al. 2020) を用いたドメイン検索を行い、HATPアーゼ (ヒスチジinkinナーゼ様 ATPase) および REC (phosphoacceptor receiver) ドメインを検出した。さらに、*A. thaliana* エチレンレセプター1 (AtETR1、accession No. AAA70047.1) 並びに *T. coccineum* および *T. cinerariifolium* のヒスチジinkinナーゼスーパーファミリータンパク質に対し、CLUSTAL W-mpi 0.13 でアミノ酸配列をアライメントし、JTT 行列を用いたモデルに基づく最尤系統樹を、MEGA ソフトウェア (Kumar et al. 2018) を用いてブートストラップ反復回数 500 回で作成して分子系統解析を行った。

第3節 結果と考察

1. *T. coccineum* ゲノムのシーケンス

T. coccineum ゲノムの PE ライブラリと 3 種の MP ライブラリ (MP-3Kb、MP-5Kb、MP-8Kb) を作成して、Illumina NovaSeq 6000 装置を用いてシーケンスした。Miseq 用ライブラリ (MS) を作成して、Illumina MiSeq システムでシーケンスした。長いリードについては、PacBio 用ライブラリ (PB) を作成して、PacBio Sequel II システムを用いてシーケンスした。配列リードの合計塩基数は、PE、MP-3Kb、MP-5Kb、MP-8Kb、MS、PB で、それぞれ 854 Gb、99 Gb、108 Gb、109 Gb、27 Gb、93 Gb であった (表 3-2)。

表 3-2 シーケンスリードの統計

Library	Insert Size (bp)	Read Length (Bases)	Number of Reads	Total Read Length (Bases)
PE	350	151	5,732,398,372	854,270,829,961
MP-3Kb	3000	151	698,859,570	99,096,491,543
MP-5Kb	5000	151	750,513,382	107,931,325,410
MP-8Kb	8000	151	762,709,822	109,359,355,641
MS	550	301	97,731,712	26,503,089,921
PB		Ave. 10,738	8,670,092	93,100,193,428

PE、MP、MS ライブラリのリードの長さは、シーケンシングの最大サイズを示す。PB ライブラリは平均サイズを示す。PE: ショートインサートライブラリ、MP: メイトペアライブラリ、MS: Miseq 用ライブラリ、PB: PacBio 用ライブラリ

2. *T. coccineum* ゲノムのサイズ推定

DNA アセンブリに先立ち、*T. coccineum* の 1C DNA 量について、フローサイトメトリーを用い、*C. seticuspe* ゲノム（栽培品種：Gojo-0、3 pg/1C）（Hirakawa et al. 2019）を内部標準として推定した。推定した *T. coccineum* DNA 量は 9.4 pg/1C であり（図 3-2A）、これはゲノムサイズ約 9.4 Gb に相当する。このサイズは、Feulgen 染色の光学密度分析によるゲノム推定で得られた過去の測定値 5.8 pg/1C の約 1.6 倍である（Bennett, Smith, and Heslop-Harrison 1982）。この 9.4 Gb というゲノムサイズの妥当性をさらに検証するため、Jellyfish を用いて、*T. coccineum* ゲノムの PE リードの k-mer スペクトログラム解析を実施したところ、k-mer スペクトログラムには 2 つの主要な分布がみられた（図 3-2B）。一つはカバレッジ値 1 を最大値にとるものであり、もう一つはカバレッジ値 44 を最大値とする多峰性の分布であった。カバレッジ値 1 を最大値とする分布はシーケンスエラーによるものと考えられるため、両分布間で最小値をとっているカバレッジ値 11 以上のデータを正しいゲノム由来の k-mer として扱った。過去の研究に従うと（Alm Rosenblad et al. 2021）、カバレッジ値 11 以上からなりカバレッジ値 44 で最大値をとる k-mer 分布より推定したゲノムサイズは 9.8 Gb であった。以上の解析により、今回の研究で用いた *T. coccineum* のゲノムサイズは約 9 Gb であるという結論に達した。また、ピークがカバレッジ値 44 の他にカバレッジ値 133 にも別のピークが見られた（図 3-2B）。TE や単純反復配列などの繰り返し配列がゲノム配列中に存在する場合、k-mer スペクトルが多峰性の分布となることが知られている（Williams et al. 2013）。*T. cinerariifolium* が *T. coccineum* の近縁にあり、多くの TE をそのゲノム配列中に持つことから、*T. coccineum* についてもこの多峰性の分布はゲノム中の TE の割合を示すと考えられる。

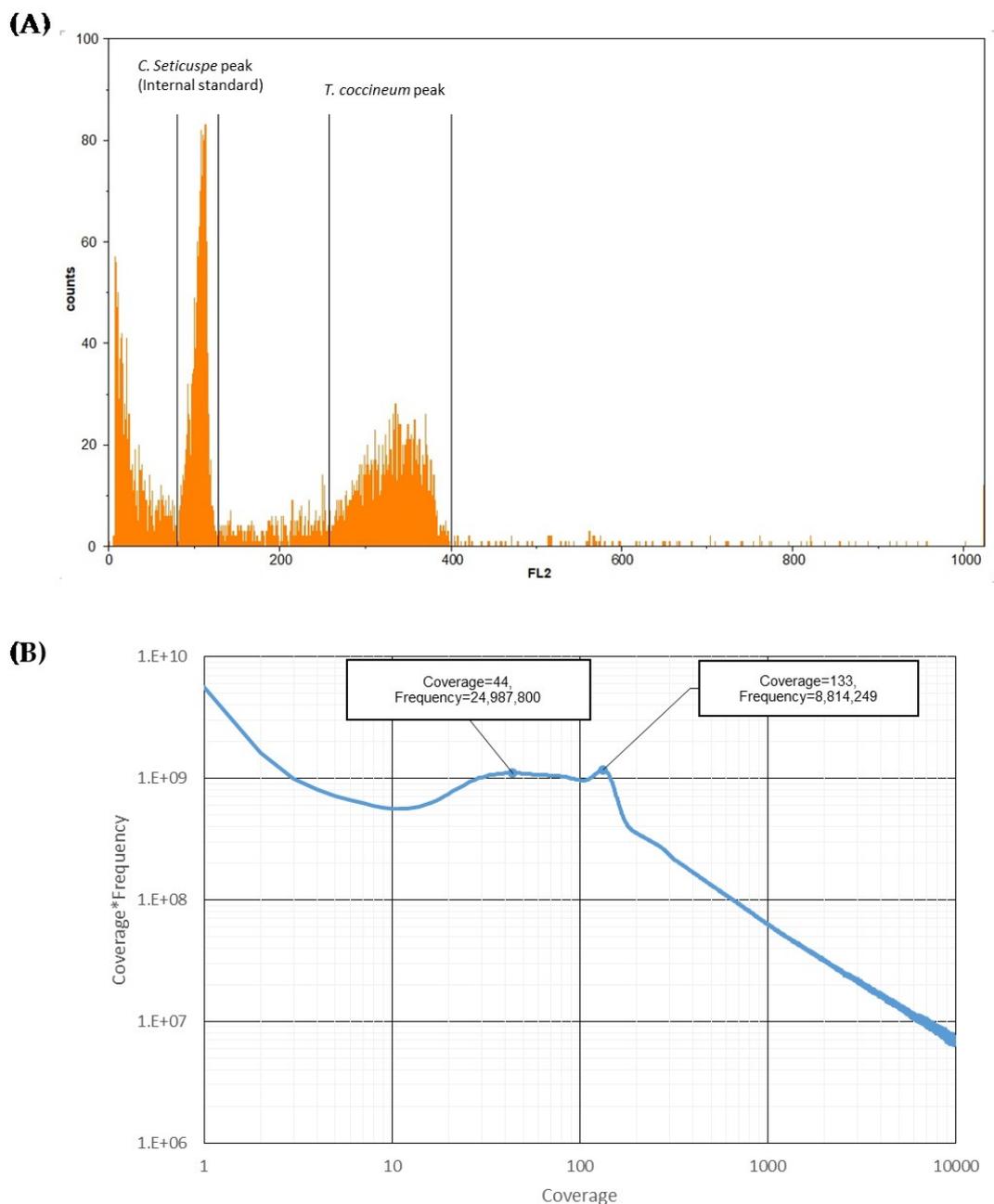


図 3-2 (A) フローサイトメトリー解析と(B) k-mer スペクトログラム解析による *T. coccineum* ゲノムのサイズ推定

(A) ヒストグラムは、フローサイトメトリー解析による *C. seticuspe* と *T. coccineum* の相対 DNA 量を示す。内部標準である *C. seticuspe* gojo-0 (3 pg/1C) のピークを 100 付近 (106.10) にセットしたところ、*T. coccineum* のピークが 330.75 となったことから、*T. coccineum* の DNA 量は約 9.4 pg/1C と推定された。count: 核の個数、FL2: 相対 DNA 量
 (B) *T. coccineum* の PE リードの k-mer スペクトログラムを示す。

3. *T. coccineum* ゲノムの配列アセンブリおよびアノテーション

各種シーケンサーより得られたリードに、コンティグアセンブリおよびスキファオールディングを行った (図 3-3)。SOAPdenovo を用いて PE リードおよび MS リードをアセンブリすることで、6,500,576 個のコンティグ配列を作成し、合計長 8.57 Gb となった (表 3-3)。アセンブリ後のコンティグは、第 2 章と同様、PB リードおよび MP リードに対して、コンティグ配列を連結する SSPACE、スキファオールド中のコンティグ間にある不明な塩基を埋める Gapfiller やさらに TGS-Gapcloser プログラムを用いてスキファオールドにした (図 3-3)。ギャップを埋めた配列の正確性は、PE リードよりも配列の正確性が落ちる PB リードに依存するため、スキファオールド配列を POLCA でポリッシングした。このとき得られたスキファオールドの合計長は 9.46 Gb であり、これは、フローサイトメトリーで推定した *T. coccineum* のゲノムサイズとよく一致していた (図 3-2A)。このスキファオールドの N50 値は 27.8 Kb、スキファオールドにおける最大コンティグ長は 331 Kb であった (表 3-3)。次にこのドラフトゲノムに対して、AUGUSTUS による解析を行ったところ、推定遺伝子数は 1,582,136 であった。これらの予測遺伝子には TE などの非機能的遺伝子が含まれている。*T. cinerariifolium* (第 2 章参照) と比較すると、*T. coccineum* ゲノムの予測遺伝子数は 167,245/Gb と予測されるため、*T. cinerariifolium* ゲノム遺伝子数の観察値 131,830/Gb と比べて大きくなっていた。

ゲノム配列中の保存遺伝子について、完全 (C)、断片化 (F)、欠損 (M) としてそれぞれ計数する BUSCO を用いて、ドラフトゲノム配列の完全性を評価した。コア植物遺伝子として登録されている 1614 個の配列解析では、保存遺伝子の 97.8%が *T. coccineum* ゲノムアセンブリに存在していること (92.7%が完全、5.1%が断片化) が確認できた (表 3-4)。このスコアは第 2 章で得られた *T. cinerariifolium* のドラフトゲノムと同様にクオリティが高いことを示している。よって、この *T. coccineum* のドラフトゲノムを以降の解析に用いた。

T. cinerariifolium を始めとする植物のゲノムでは機能性遺伝子よりも多くの TE が検出されることから、アセンブリ後の *T. coccineum* ゲノムの TE について解析した。第 2 章と同様に、hmmpfam を GyDB に対して用いることで TE の検出およびアノテーションを行い、772,794 個の TE の存在を確認した。この結果、*T. coccineum* ゲノムの TE 数は 82,212/Gb と予測され、*T. cinerariifolium* ゲノムでの TE 数 73,957/Gb (第 2 章の結果より算出) と比べて若干大きくなっていった。さらに、信頼度の高いアノテーションを得るために、予測遺伝子に対し InterProScan を実施したところ、既知タンパク質の特徴を示す産物をコードする、103,680 個の推定遺伝子の存在が明らかになった。以上から、PE リード 854 Gb、MP リード 316 Gb、MS リード 26.5 Gb、PB リード 93.1 Gb より、高品質の 9.4 Gb の *T. coccineum* のドラフトゲノムがアセンブルされ、また、772,794 個の TE と 103,680 個の信頼性の高い遺伝子を含んでいることが示された。以上のアセンブルに使用したシーケンスリード、完成したドラフトゲノム配列およびアノテーション情報は DDBJ の bioproject accession code PRJDB12537 下にアップロードした。

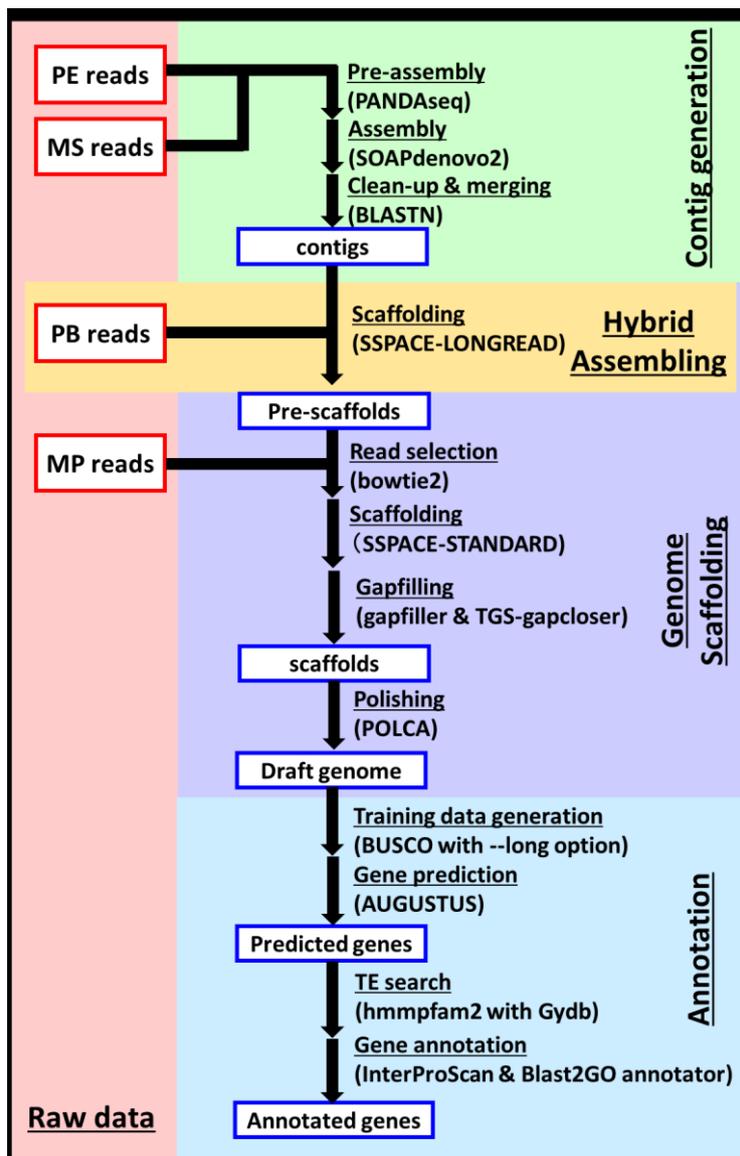


図 3-3 ゲノムアセンブリおよび遺伝子予測のフローチャート

PE リードと MS リードは、コンティグアセンブリに供し、PANDA-seq を使用した「プレアセンブリ」、SOAP denovo を使用した「コンティグアセンブリ」、BLASTN を使用した「クリーンアップ & 併合」の 3 行程を実施した。PB リードは SSPACE-longread を使用したハイブリッドアセンブリに供し、プレスキャフォールドを作成した。MP リードはスキャフォールド作成に供し、bowtie2 を使用したリード選出を行い、SSPACE-STANDARD を使用した「スキャフォールディング」、Gapfiller と TGS-Gapcloser を使用した「ギャップフィリング」、POLCA を使用した「ポリッシング」の 3 工程を実施した。これらの工程を経て作成されたドラフトゲノムに対し、コーディング領域は、BUSCO を使用した「訓練データの作成」、AUGUSTUS を使用した「遺伝子予測」、hmmpfam を使用した「TE 検出」、

InterProScan および Blast2GO を使用した「遺伝子アノテーション」の 4 工程を実施した。

PE: ショートインサートライブラリ、MP: メイトペアライブラリ、MS: Miseq 用ライブラリ、PB: PacBio 用ライブラリ、TE: 転移因子

表 3-3 ゲノムアセンブリの統計

	Contigs	Scaffolds (before gapfilling)	Draft Genome
Total number of sequence fragments	6,500,576	3,061,809	2,836,647
Total length (bp)	8,565,698,618	9,395,951,224	9,463,677,832
N50 (bp)	8,465	25,397	27,784
Length of longest contig (bp)	149,916	329,693	331,286
Gaps (bp)	0	777,041,487	724,210,424
GC content (%)	34.9	35.1	35.1

表 3-4 ドラフトゲノムのアノテーション統計

Number of Predicted Genes	1,582,136
BUSCO v5	C: 92.7% (Single: 70.8%, Duplicated: 21.9%) F: 5.1% M: 2.2%
Number of predicted TEs	772,794
Number of predicted genes encoding products with known protein signatures	103,680

BUSCO の表記で C: 完全長の保存遺伝子の割合、F: 断片化している保存遺伝子の割合、M: 検出できなかった保存遺伝子の割合、TE: 転移因子

4. TE の属間比較解析

T. coccineum, *T. cinerariifolium*, *A. annua*, *C. seticuspe*, *H. annuus*, *N. tabacum*, *O. sativa* および *A. thaliana* のゲノムにおいてアノテーションされた TE を、GyDB 分類に基づくクレードごとに分類し、各 TE クレードの全 TE 領域に対する割合を解析した。キク科に属する *A. annua*, *C. seticuspe* および *H. annuus*、並びにモデル生物の *N. tabacum*, *O. sativa* および *A. thaliana* を第 2 章の *T. cinerariifolium* と同様に用いた。各植物の上位 5 つのクレードを表 3-5 に示す。*T. coccineum* では TE クレードとして sire クレードの TE が最も多くなっており、これは他キク科植物 3 種 (*T. cinerariifolium*, *A. annua* および *C. seticuspe*) でも同様であった。*T. cinerariifolium* ゲノムでは、athila クレードが 2 番目に多く、以下 del、oryco、lentiviridae の順であった。また *T. coccineum* ゲノムでは、del クレードが 2 番目に多く、以下 athila、oryco、tork の順であった。*T. coccineum* に特異的な del および tork クレード TE の重複は、*T. cinerariifolium* と *T. coccineum* が共通の祖先から進化上分岐してから、これらの TE が重複したことを示唆している。

del および tork クレード TE の重複が共通のキク科の祖先のものなのか、あるいは個別の種で独立したものなのかを検証するため、第 2 章と同様の手法を用い、del および tork クレード TE 内にコードされている逆転写酵素ドメインの分子系統樹を解析し、単一種で共クラスタリングしている遺伝子数を計数した。この系統解析によると、*T. coccineum*, *T. cinerariifolium*, *A. annua*, *C. seticuspe*, *H. annuus* の del クレード TE のそれぞれ 67%、62%、73%、68%、86% が重複クラスタを構成していることが明らかとなった (図 3-4A)。同様に、tork クレード TE ではそれぞれ 57%、37%、54%、38%、71% が重複クラスタを構成していることが示された (図 3-4B)。上記の結果によると、各種において del クレード TE の半分超と tork クレード TE の 3 分の 1 超が重複していたが、これ以外の TE では、キク科の共通祖先の TE が保存されていることが示された。これらの結果から、それぞれキク科に属する各々の種で、del および tork クレード TE がおおむね重複しており、このことが *T.*

coccineum で TE が多くなっている理由と考えられた。

表 3-5 *T. coccineum* と他植物間における全 TE 領域中の TE 含有割合の比較

rank	Tco	Tci	Cs	Aa	Ha	Nt	Os	At
1	sire (25.7)	sire (33.0)	sire (32.0)	sire (21.8)	del (37.7)	del (40.4)	tat (11.4)	athila (9.54)
2	del (15.3)	athila (17.0)	athila (10.9)	athila (19.6)	sire (9.85)	tat (20.5)	retro- viridae (8.97)	retro- viridae (4.89)
3	athila (12.5)	del (12.0)	oryco (5.11)	del (6.57)	lenti- viridae (8.72)	athila (9.87)	del (8.39)	caulimo virus (4.15)
4	oryco (7.25)	oryco (6.34)	lenti- viridae (5.06)	oryco (4.59)	tat (6.76)	sire (3.02)	tork (4.73)	badnav irus (4.05)
5	tork (4.70)	lenti- viridae (4.92)	del (5.03)	tork (4.01)	athila (5.17)	tork (2.80)	alpha- retrovir idae (4.66)	tork (3.08)

括弧内の数値は、全 TE 領域に対する各クレード TE 領域の占める割合 (%) を示す。

Tco: *Tanacetum coccineum*、Tci: *Tanacetum cinerariifolium*、Cs: *Chrysanthemum seticuspe*、Aa: *Artemisia annua*、Ha: *Helianthus annuus*、Nt: *Nicotiana tabacum*、Os: *Oryza sativa*、At: *Arabidopsis thaliana*

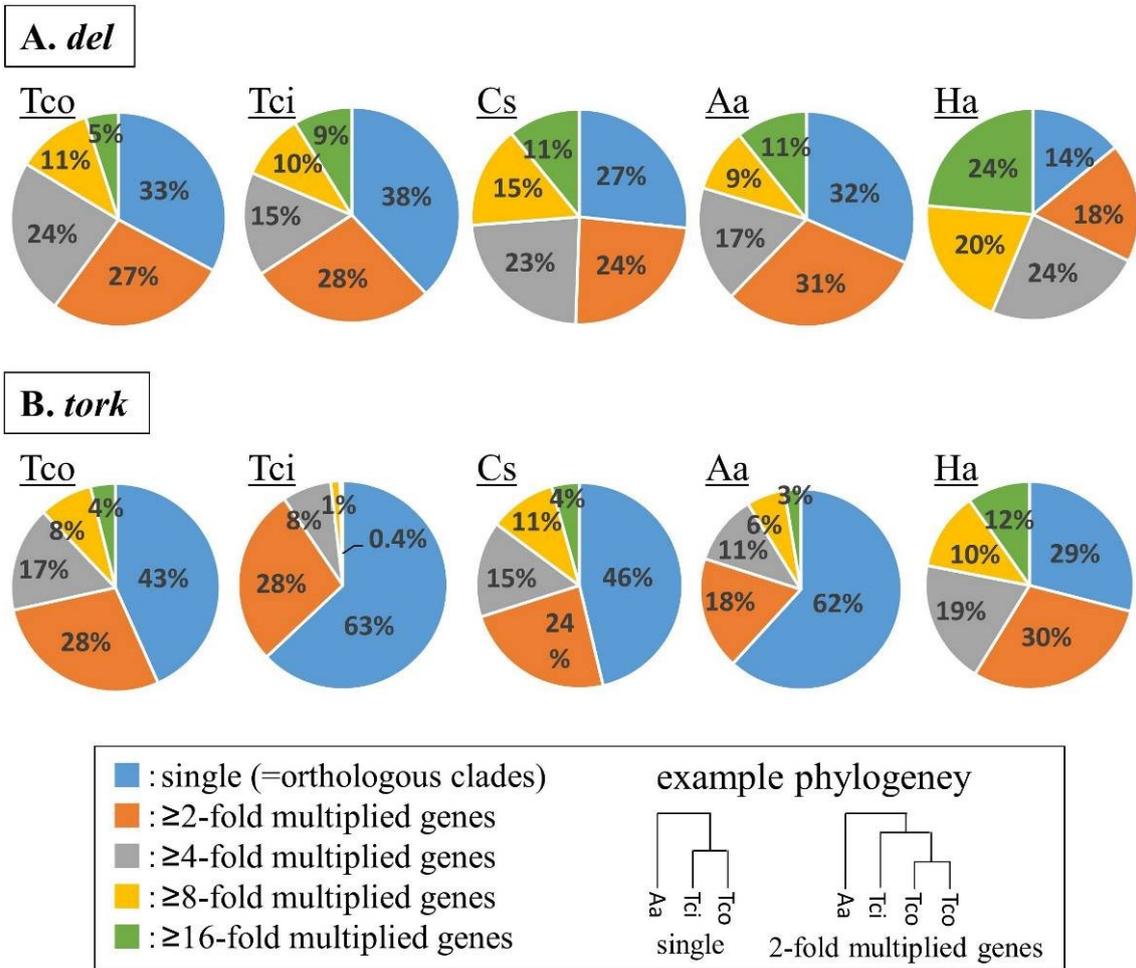


図 3-4 (A) del クレードと (B) tork クレードの TE の重複解析

del クレード TE と tork クレード TE の分子系統樹において、種特異的な重複現象の回数を反映する、単一植物種内でクラスタを形成する遺伝子の数をカウントした。Tco: *Tanacetum coccineum*、Tci: *Tanacetum cinerariifolium*、Cs: *Chrysanthemum seticuspe*、Aa: *Artemisia annua*、Ha: *Helianthus annuus*

5. *T. coccineum* ゲノム内にコードされるピレトリン類生合成関連酵素

T. coccineum で予測されるタンパク質の中で、*T. cinerariifolium* のピレトリン類生合成に関与すると知られている 9 種のタンパク質 (TciADH2 (Xu et al. 2018)、TciALDH1 (Xu et al. 2018) 、TciCCH (Xu et al. 2019)、TciCCMT (Xu et al. 2019)、TciCDS (Rivera et al. 2001)、TciGLIP (Kikuta et al. 2012)、TciJMH (Li, Zhou, and Pichersky 2018)、TciLOX1 (Ramirez et al. 2013)および TciPYS (Li et al. 2019)) について、それぞれのアミノ酸の配列相同性が高いものが BLASTP で検出された (表 3-6)。これは、既知のピレトリン類生合成関連酵素が完全なセットとして *T. coccineum* ゲノム内に保存されていることを示している。また、InterProScan を用いて、各タンパク質の保存ドメインが検出され、これらのタンパク質が機能性であり、信頼性が高いものであることが確認できた。

表 3-6 *T. coccineum* ゲノム中にコードされている、既知のピレトリン類生合関連タンパク質と一致するタンパク質

Known Pyrethrin-Related Enzymes	Corresponding Proteins of <i>T. coccineum</i>	Protein Sequence Similarity
TciADH2	Tco_0487905	Identities = 340/378 (90%), Positives = 359/378 (95%), Gaps = 2/378 (1%)
TciALDH1	Tco_0682217	Identities = 448/499 (90%), Positives = 471/499 (94%), Gaps = 1/499 (0%)
TciCCH	Tco_0360514	Identities = 470/498 (94%), Positives = 484/498 (97%), Gaps = 1/498 (0%)
TciCCMT	Tco_1190813	Identities = 358/374 (96%), Positives = 361/374 (97%), Gaps = 5/374 (1%)
TciCDS	Tco_1315810	Identities = 358/395 (91%), Positives = 374/395 (95%), Gaps = 0/395 (0%)
TciGLIP	Tco_1108878	Identities = 337/365 (92%), Positives = 348/365 (95%), Gaps = 0/365 (0%)
TciJMH	Tco_0572988	Identities = 450/512 (88%), Positives = 479/512 (94%), Gaps = 2/512 (0%)
TciLOX1	Tco_0863779	Identities = 847/861 (98%), Positives = 853/861 (99%), Gaps = 0/861 (0%)
TciPYS	Tco_1240348	Identities = 465/488 (95%), Positives = 475/488 (97%), Gaps = 0/488 (0%)

Tci: *Tanacetum cinerariifolium*, Tco: *Tanacetum coccineum*, ADH2: アルコールデヒドロゲナーゼ 2、ALDH1: アルデヒドデヒドロゲナーゼ 1、CCMT: 10-カルボキシクリサンテム酸 10-メチルトランスフェラーゼ、CDS: クリサンテミルニリン酸合成酵素、CHH: クリサンテモール 10-ヒドロキシラーゼ、GLIP: GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ、JMH: ジャスモンヒドロキシラーゼ、LOX1: 13-リポキシゲナーゼ、PYS: ピレスロロン合成酵素

6. ピレトリン類生合成関連酵素コード遺伝子のシンテニー解析

GenomeJack ソフトウェアプログラムを用いて、TciADH2、TciALDH1、TciCCH、TciCCMT、TciCDS、TciGLIP、TciJMH、TciLOX1 および TciPYS に相当するタンパク質をコードする遺伝子座を含むスキヤフォールド中の遺伝子分布を解析した。全ての遺伝子が別々のスキヤフォールドに位置しており、また、TcoCCMT および TcoGLIP コード遺伝子を除き、いずれのピレトリン類生合成関連酵素コード遺伝子もその近傍領域に TE が存在していた。Jacalin 様レクチンドメインを含む Tco_1190812 タンパク質コード遺伝子が、TcoCCMT タンパク質コード遺伝子の上流に位置していた (図 3-5A)。Tco_1190812 のアミノ酸配列をクエリとして用いた BLASTP 検索では、予測タンパク質として *A. annua* のマンノース結合レクチン (accession No. PWA73033.1) と配列相同性を有するものが E-value 値が 3×10^{-93} 、相同性が 92.72% で得られた。*T. cinerariifolium* ゲノムでも、対応する Jacalin 様レクチンコード遺伝子 (accession No. GEW32189.1) が TciCCMT コード遺伝子座の上流にみられた。以上の結果から、当該 Jacalin 様レクチンコード遺伝子と Tci(o)CCMT コード遺伝子のシンテニーが保存されていることが示された。

TciGLIP は、ピレトリン類生合成の最終エステル化において主要な酵素である (Kikuta et al. 2012) (図 2-1)。シンテニー解析によると、TciGLIP および TcoGLIP コード遺伝子双方の下流に、GLIP (Gly-Asp-Ser-Leu モチーフリパーゼ) をコードしていると推定されるオープンリーディングフレームが存在していることが示された。しかし、TciGLIP の上流に位置する *T. cinerariifolium* グルタチオン S-トランスフェラーゼコード遺伝子 (accession No. GEU71427.1) および *T. cinerariifolium* のタンデム GLIP コード遺伝子の下流に位置する推定タンパク質コード遺伝子 (accession No. GEU71430.1) は、*T. coccineum* ゲノムでは TE に置き換えられていた (図 3-5B)。これは、*T. coccineum* と *T. cinerariifolium* が分岐した後、当該タンデム GLIP コード遺伝子座が転座していることを示している。TciGLIP の転写制御機序はまだ明らかになっていないが、これら GLIP コード遺伝子の近

傍領域の違いが、TciGLIP および TcoGLIP をコードする遺伝子で発現が異なる機序を検討する手がかりとなりうる。今後、両者のゲノム構成がそれぞれの GLIP 遺伝子の転写にどのように影響しているかを調べる予定である。

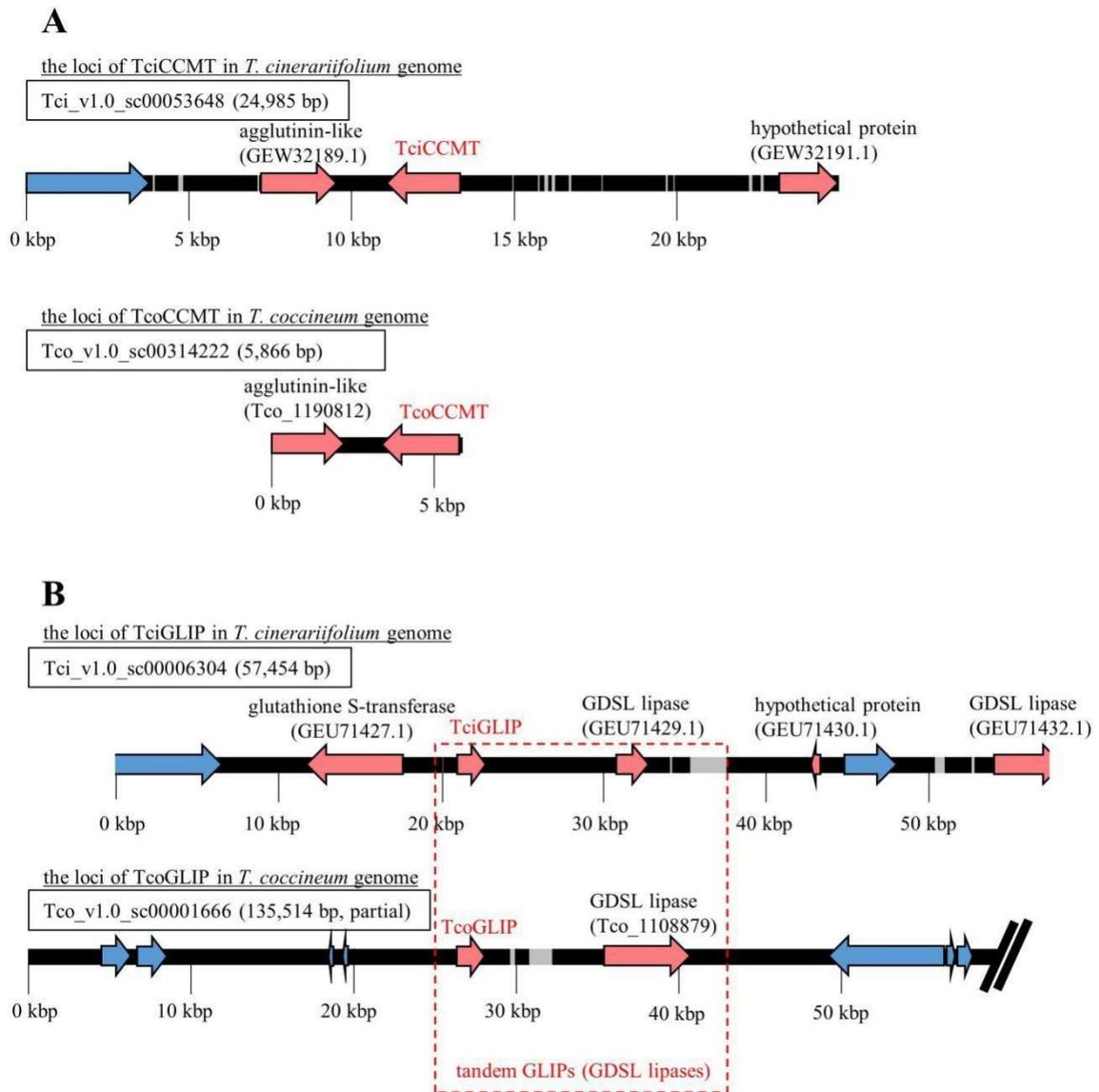


図 3-5 ピレトリン類生合成関連酵素(A) Tci(o) CCMT と(B) Tci(o)GLIP をコードする遺伝子を含むスキファールド

灰色の領域: ギャップ、赤色の矢印: タンパク質コード遺伝子、青色の矢印: 転移因子 (TE)

7. *T. coccineum* 遺伝子の機能アノテーションと属間の比較分析

次に、タンパク質スーパーファミリーの重複比について、*T. coccineum* と別種とで比較を行った。具体的には、*T. coccineum*、*T. cinerariifolium*、*C. seticuspe*、*A. annua*、*H. annuus*、*N. tabacum*、*O. sativa* および *A. thaliana* の予測タンパク質データセットに対し InterProScan を用いた解析を行い、各スーパーファミリーにおける重複オッズスコアを算出した。生体防御、シグナル伝達、代謝に関連するスーパーファミリーで、*T. coccineum* で最大値を示したものと最小値を示したものに対し、*T. cinerariifolium* など他植物に対する比較を行った（それぞれ表 3-7、3-8）。*T. coccineum* と *T. cinerariifolium* を更に詳しく比較するため、第 2 章で確認したスーパーファミリーと最大、最小値を示すスーパーファミリー表（それぞれ表 3-7、3-8）に記載されていないスーパーファミリーの重複オッズスコアを表 3-9 に示す。

生体防御に関するスーパーファミリーのうち、「リボソーム不活性化タンパク質 (RIP)」(IPR036041) ドメインを含むタンパク質をコードする遺伝子で *T. coccineum* ゲノムにおける重複が認められた（重複オッズスコアが 1.96、表 3-7）。第 2 章での *T. cinerariifolium* でもこのスーパーファミリーのタンパク質には重複が認められていた。「リボソーム不活性化タンパク質 (RIP)」(IPR036041) の重複オッズスコアは、*T. coccineum* ゲノムは *T. cinerariifolium* ゲノムの 1.5 倍（表 3-7）であったが、「リシン B 様レクチン」(IPR035992) の重複オッズスコアは *T. coccineum* ゲノムで、*T. cinerariifolium* と比べて低くなっていた（表 3-9）。リシンを含め RIP は、昆虫や細菌、ウイルスなど様々な生物種に対し高い毒性を示しており、これを産生する植物の生体防御分子として働く (Bolognesi et al. 2016)。RIP は、タイプ I とタイプ II に分けられ、リシン B レクチンドメインを有しない方が前者、有するほうが後者である (Zhu et al. 2018)。リシン B レクチンドメインは、標的細胞のグリカンに対する結合を通じ細胞内部へ移行するため、タイプ II RIP のほうがタイプ I RIP と比べて毒性が高くなる。これらの結果より、毒性が高いタイプ II の RIP をコードしている遺

伝子が、*T. coccineum* ゲノムと比べて *T. cinerariifolium* ゲノムでより多く重複しており、自然環境では *T. coccineum* のほうが、*T. cinerariifolium* と比べて外敵から影響を受けやすい、あるいは感受性が高いと考えられる。第 2 章において、*Sambucus nigra* の殺虫性 RIP の SNA-I (*S. nigra* アグルチニン-I、accession No. O22415.1) (Shahidi-Noghabi, Van Damme, and Smaghe 2008) と類似のアミノ酸配列を持ち、殺虫性を有するタイプ II RIP と推定される Tci_399175 (accession No. GEY27201.1) をコードする遺伝子が、*T. cinerariifolium* ゲノム中に確認されている。*T. coccineum* ゲノムの予測タンパク質群に対する SNA-I アミノ酸配列をクエリとする BLASTP 検索では、Tco_1336120 が検出された。標的細胞の認識に重要な RICIN ドメインのアライメントでは、当該殺虫性 RIP が *T. coccineum* ゲノムでもコードされるとの推定が確認できた (図 3-6)。この結果は、SNA-I 様の殺虫性 RIP が *Tanacetum* 属の両種で保存されていることを示すものであった。さらに、今回の RIP 遺伝子比較解析によって、*T. coccineum* ゲノムでタイプ I RIP が多く、*T. cinerariifolium* ゲノムでタイプ II RIP が多いことが確認された。以上の配列解析から、上記二種類の植物で RIP を用いた防御戦略が大きく異なることが示唆された。

病原性真菌に対する防御に重要な役割「エンドキチナーゼ様スーパーファミリー」(IPR036861) が、*T. coccineum* ゲノムにも他の植物種と同等のレベルで存在しており、またこれは *T. cinerariifolium* よりも多かった (表 3-9)。これらの結果からは、*T. cinerariifolium* が乾燥地域を原産としているのに対し、*T. coccineum* は湿度の高い地域を原産としていることを合わせると、これらの植物は病原性真菌に対しても防御機構を多様化させていると推定される。

代謝関連スーパーファミリーでは、「ウレアーゼ α サブユニット」(IPR005848)、「リポキシゲナーゼ、C 末端ドメイン」(IPR036226)、「RuBisCo」(IPR033966)、「金属依存性加水分解酵素」(IPR032466)、「エノラーゼ様、C 末端ドメイン」(IPR036849)、「シトクロム P450」(IPR036396) コード遺伝子は *T. coccineum* において最大の遺伝子重複が起きて

おり、重複オッズスコアはそれぞれ 2.36、1.86、1.60、1.48、1.38、0.90 であった (表 3-7)。リポキシゲナーゼ、金属依存性加水分解酵素、シトクロム P450 といった金属タンパク質をコードする遺伝子の重複が特徴的であった。なお、同様の重複は第 2 章において *T. cinerariifolium* ゲノムでもみられている。一部のピレトリン類生合成関連タンパク質はシトクロム P450 又はリポキシゲナーゼのスーパーファミリーに属していることから、*T. coccineum* および *T. cinerariifolium* 両種でピレトリン類を合成できることを考えると、これらの遺伝子は、共通の祖先植物において重複したことが示唆される。*T. coccineum* ゲノムにおけるシトクロム P450 スーパーファミリーコード遺伝子数は、*T. cinerariifolium* ゲノムのものと比べて 1.6 倍であった。分子系統解析によると、*T. coccineum* シトクロム P450 の 57% はオーソログ遺伝子クラスタに含まれていなかったが、独立したクラスタを構成していた (図 3-7)。この分子系統樹は、一部のオーソログなシトクロム P450 の重複が *T. coccineum* のみの系統で生じたことを示唆している。以上より、*T. cinerariifolium* および *T. coccineum* のシトクロム P450 がそれぞれの進化過程において重複することで、ピレトリン類を始めとする種特異的な植物特有の代謝物の産生能を獲得したとする考えを支持するものである。また、この結果は、種特異的な二次的代謝物が *T. coccineum* で *T. cinerariifolium* より豊富である可能性を示唆している。*T. coccineum* の二次代謝物はほとんど調べられていないため、同植物の二次代謝物についてさらなる検証が必要である。

また、*T. cinerariifolium* と同様に、「HECT E3 ユビキチンリガーゼ触媒ドメイン」(IPR035983) を有するタンパク質の高度な重複が *T. coccineum* でも見られた (表 3-7)。一方、RING フィンガー型 E3 ユビキチンリガーゼに含まれるドメインである「RCHY zinc-ribbon」(IPR039512) を含むタンパク質の重複度は低かった (表 3-8)。この結果は、*T. coccineum* ゲノムにおいて HECT 型 E3 ユビキチンリガーゼをコードする遺伝子は重複されているが、RING フィンガー型 E3 ユビキチンリガーゼは重複されていないことを示唆している。このような E3 ユビキチンリガーゼコード遺伝子の重複に明らかな不均衡がみられ

ることの生物学的な意義については、さらなる研究が必要である。同様に、「シグナル伝達ヒスチジンキナーゼの二量体化／リン酸受容体ドメイン」(IPR036097)を含むタンパク質をコードする遺伝子が、*T. cinerariifolium* ゲノムでは重複していたが、*T. coccineum* ゲノムでは重複していなかった(表 3-9)。植物では、日光、植物ホルモン、エチレンなど環境刺激に対する反応にヒスチジンキナーゼが関与している。シロイヌナズナのエチレン受容体 1 (AtETR1) は、典型的な VOC 誘導性ヒスチジンキナーゼであり、HATP アーゼ(ヒスチジンキナーゼ様 ATP アーゼ)ドメインおよび REC (phosphoacceptor receiver) ドメインを有する(Shakeel et al. 2013)。*T. coccineum* および *T. cinerariifolium* のヒスチジンキナーゼコード遺伝子に対し、HATP アーゼドメインと REC ドメインの存在を調べた。これらドメインを含む遺伝子数のデータは、図 3-8A にベン図で示している。HATP アーゼドメインおよび REC ドメインの両方がコードされている遺伝子の数は、*T. coccineum* では 13、*T. cinerariifolium* では 38 であった。ヒスチジンキナーゼと予測されるタンパク質の分子系統樹を図 3-8B に示す。*T. cinerariifolium* のパラログ 5 つと *T. coccineum* のパラログ 4 つを含むクラスタの中に AtETR1 が位置している。比較解析によると、オーソログのクラスタだけではなく、*T. cinerariifolium* 特異的なクラスタ重複も確認された(図 3-8B、緑)。第 2 章において、VOCs を介した *T. cinerariifolium* のピレトリン類産生の調節にかかわる VOC 誘導性ヒスチジンキナーゼの存在が示唆された。本研究でも、*T. coccineum* および *T. cinerariifolium* におけるピレトリン類産生量とヒスチジンキナーゼタンパク質の数との相関性が示唆された。今回の結果は、*T. coccineum* では、*T. cinerariifolium* と異なり種特異的なヒスチジンキナーゼコード遺伝子の重複によるガス(VOCs)誘導型のピレトリン類産生機構の獲得がなされていないことを示唆している。ヒスチジンキナーゼと VOCs 誘導型ピレトリン類産生との機能的な関係性については研究を続行中である。

本研究では、*T. coccineum* のドラフトゲノムを明らかにした。*T. coccineum* と近縁の *T. cinerariifolium* の間でゲノムの比較分析を行ったところ、*T. coccineum* 遺伝子の特徴的な

性質が明らかになり、これが *T. coccineum* と *T. cinerariifolium* の間でピレトリン類の産生量が異なる原因と考えられた。

表 3-7 *T. coccineum* の中で重複オッズスコアの高いスーパーファミリー

Category	IPR ID	Superfamily Name	Tco	Tci	Cs	Aa	Ha	Nt	Os	At
Biodefense	IPR036041	Ribosome-inactivating protein	1.96 (159)	1.29 (98)	-1.81 (7)	-1.00 (16)	-3.07 (0)	-3.07 (0)	-0.94 (17)	-3.07 (0)
Metabolism	IPR005848	Urease, alpha subunit	2.36 (108)	-0.14 (15)	-1.87 (1)	-1.46 (3)	-1.87 (1)	-0.87 (7)	-2.14 (0)	-1.87 (1)
Metabolism	IPR036226	Lipoxygenase, C-terminal domain	1.86 (232)	0.48 (86)	-0.22 (51)	-0.82 (32)	-1.12 (25)	-0.67 (36)	-1.86 (13)	-2.44 (7)
Metabolism	IPR033966	RuBisCO	1.60 (42)	-0.25 (8)	0.05 (11)	-0.15 (9)	-1.37 (1)	-0.95 (3)	-0.25 (8)	-1.15 (2)
Metabolism	IPR032466	Metal- dependent hydrolase	1.48 (166)	0.69 (94)	-1.13 (23)	-0.89 (28)	-0.32 (44)	-0.05 (54)	-1.61 (15)	-0.98 (26)
Metabolism	IPR036849	Enolase-like, C-terminal domain	1.38 (71)	0.91 (50)	-0.62 (14)	-0.55 (15)	-0.78 (12)	-0.29 (19)	-1.41 (6)	-1.29 (7)
Metabolism	IPR036396	Cytochrome P450	0.90 (1220)	0.19 (745)	0.16 (732)	0.07 (688)	-0.20 (568)	-0.12 (600)	-1.05 (314)	-0.85 (361)
Signaling	IPR035983	HECT, E3 ligase	1.22 (95)	0.84 (72)	0.10 (41)	-0.52 (25)	-0.90 (18)	-0.21 (32)	-1.84 (7)	-1.25 (13)

括弧内の数値は各々のスーパーファミリーに振り分けられた遺伝子の数を示す。Tco: *Tanacetum coccineum*、Tci: *Tanacetum cinerariifolium*、Cs: *Chrysanthemum seticuspe*、Aa: *Artemisia annua*、Ha: *Helianthus annuus*、Nt: *Nicotiana tabacum*、Os: *Oryza sativa*、At: *Arabidopsis thaliana*

表 3-8 *T. coccineum* の中で重複オッズスコアの低いスーパーファミリー

Category	IPR ID	Superfamily Name	Tco	Tci	Cs	Aa	Ha	Nt	Os	At
Signaling	IPR039512	RCHY1, zinc-ribbon	1.18 (3)	-0.48 (8)	-0.09 (12)	-0.09 (12)	-0.01 (13)	1.28 (39)	-0.72 (6)	0.09 (12)

括弧内の数値は各々のスーパーファミリーに振り分けられた遺伝子の数を示す。Tco: *Tanacetum coccineum*、Tci: *Tanacetum cinerariifolium*、Cs: *Chrysanthemum seticuspe*、Aa: *Artemisia annua*、Ha: *Helianthus annuus*、Nt: *Nicotiana tabacum*、Os: *Oryza sativa*、At: *Arabidopsis thaliana*

表 3-9 *T. cinerariifolium* (第 2 章参照) で重複オッズスコアが特徴的なスーパーファミリー

Category	IPR ID	Superfamily Name	Tco	Tci	Cs	Aa	Ha	Nt	Os	At
Biodefense	IPR035992	Ricin B-like lectins	0.81 (44)	1.41 (69)	-0.34 (17)	-0.05 (22)	-1.22 (7)	-0.80 (11)	-1.10 (8)	-1.48 (5)
Biodefense	IPR036861	Endochitinase-like	-0.13 (7)	-1.13 (1)	-0.39 (5)	-0.25 (6)	0.53 (14)	0.29 (11)	0.09 (9)	0.37 (12)
Signaling	IPR036097	Signal transduction histidine kinase, dimerization/phosphoacceptor domain	-0.11 (32)	-1.41 (101)	-0.62 (21)	-0.37 (26)	-0.28 (28)	0.35 (46)	-1.74 (7)	-0.74 (19)
Signaling	IPR024792	Rho GDP-dissociation inhibitor domain	0.48 (18)	1.24 (34)	-0.14 (10)	-0.58 (6)	-0.34 (8)	-0.14 (10)	-1.04 (3)	-1.04 (3)
Metabolism	IPR012347	Ferritin-like	0.72 (22)	1.29 (35)	-0.03 (11)	-0.71 (5)	-0.57 (6)	-0.86 (4)	-1.23 (2)	-0.57 (6)
Metabolism	IPR036909	Cytochrome c-like domain	0.40 (21)	1.16 (39)	-0.50 (9)	-0.84 (6)	-0.22 (12)	-0.16 (17)	-0.82 (7)	-0.82 (7)
Metabolism	IPR037069	Acyl-CoA dehydrogenase/oxidase, N-terminal domain	0.45 (22)	1.05 (36)	-0.60 (8)	-0.30 (11)	-0.22 (12)	0.22 (18)	-0.98 (5)	-0.84 (6)

括弧内の数値は各々のスーパーファミリーに振り分けられた遺伝子の数を示す。Tco: *Tanacetum coccineum*, Tci: *Tanacetum cinerariifolium*, Cs: *Chrysanthemum seticuspe*, Aa: *Artemisia annua*, Ha: *Helianthus annuus*, Nt: *Nicotiana tabacum*, Os: *Oryza sativa*, At: *Arabidopsis thaliana*

```

SNA-I (O22415.1)          SYTNNAIDAQIIKMPVFR---GGGYEKVCSVVEVTRRISGWDLGCVDVRDGHYIDGNTVQLGPCGN-ECNQQLWTF
Tci_399175 (GEY27201.1) -----KPTAIRMPVPPVPPVAVGADEQCYPYGEPTTNIIGRDGQCMDVKENQYGNNGNPIILFPCGNAQRNQLWTF
Tco_1336120              -----NPRAIRMPVPPV---TVGANEQCPNGEPTTNIIGRDGQCVHVKDNQYNNGNSIILWACGNALRNQLWTF
                          ..  *---          *  ::  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *  *
SNA-I (O22415.1)          RTDGTIRWLKCLTT-----SSVMIYDCNTVPPEATKWVVSTDGTITNPRSGLVLTAPQAEGTALSLENNIHAA
Tci_399175 (GEY27201.1) KSDGTIRSNGKCLTT---SGNYIMIFDCD-LAPETTKWILHNAGTIMNPRIRLVIAAESSTPRTVLTAAVDSNSS
Tco_1336120              KSDGTIRSNGKCLTTPGYASGNYIMIFDCDTAVPEATKWFLYNAGTIMNPRLVIAAETSTQGTVLTVAKDNNSS
                          :*****  *****          ..  ***:***:  **:-***:  ..  ***  ***  ***:  ..  **:  *:  ..  :
SNA-I (O22415.1)          RQGWTVG-DVEPLVTFIVGYKQKCLFENGENFVWLEDCVLNR-VEQEWALYGDGTIRVNSNRSICVTSEDHEPSD
Tci_399175 (GEY27201.1) RQAWSAGNYTQPTITYISGFLEMCLQANGENARVWLANCVIDTEPRQQWALYGDRTIRLYSDRTLCVTSDGHESVD
Tco_1336120              RQAWSAGNYTQPTINYISGYREMCLQANGANARVWLANCVIGTEPRQQWVLYGDSTIRLYSDRTLCVTSDGHESSD
                          **_*_*  _*  _*_*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*  _*
SNA-I (O22415.1)          LIVILKCESGNQRWVENTNGTISNPNAKLVMDAQSNVSLRKIILYPPTGNPNQRWITTTQPA
Tci_399175 (GEY27201.1) SIILFKCQSEAQRWTFMADATILNPYAQLVMDVRGSDVSLQEIILYPPTGNPNQRWLAF----
Tco_1336120              SIILLKCQWGDQRWTFMADGTILNPARLVMDVRNSDVSLQEIILYPPTGNPNQRWLAF----
                          *:-**:*  ***_*  ..**  **  *-*****  *-***:-*****  *****-:-*

```

図 3-6 RIP の RICIN ドメインのアミノ酸配列アラインメント

セイヨウニワトコ *Sambucus nigra* agglutinin I (SNA-I, accession No. O22415.1)、*T. cinerariifolium* Tci_399175 (accession No. GEY27201.1) および *T. coccineum* Tco_1336120 の RICIN ドメインのアミノ酸配列アラインメントを示す。本アラインメントは CLUSTAL W-mpi を用いて作成した。RICIN ドメインは緑色で強調した領域、RICIN ドメイン中に保存されている Q-X-W モチーフは赤字で示す。

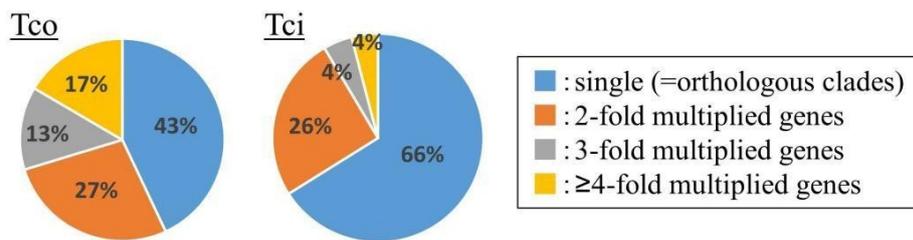
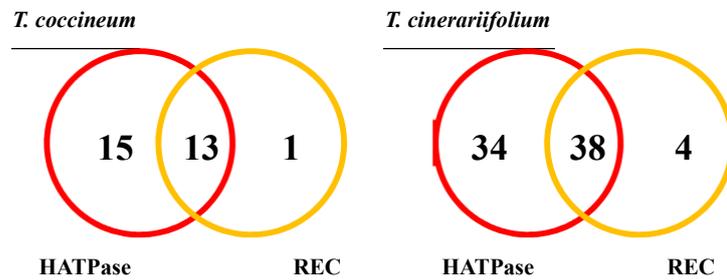


図 3-7 *T. coccineum* と *T. cinerariifolium* ゲノムにおける、“シトクロム P450” (IPR036396) スーパーファミリドメインを有するタンパク質をコードする遺伝子の重複解析

Tco: *Tanacetum coccineum*, Tci: *Tanacetum cinerariifolium*

A



B

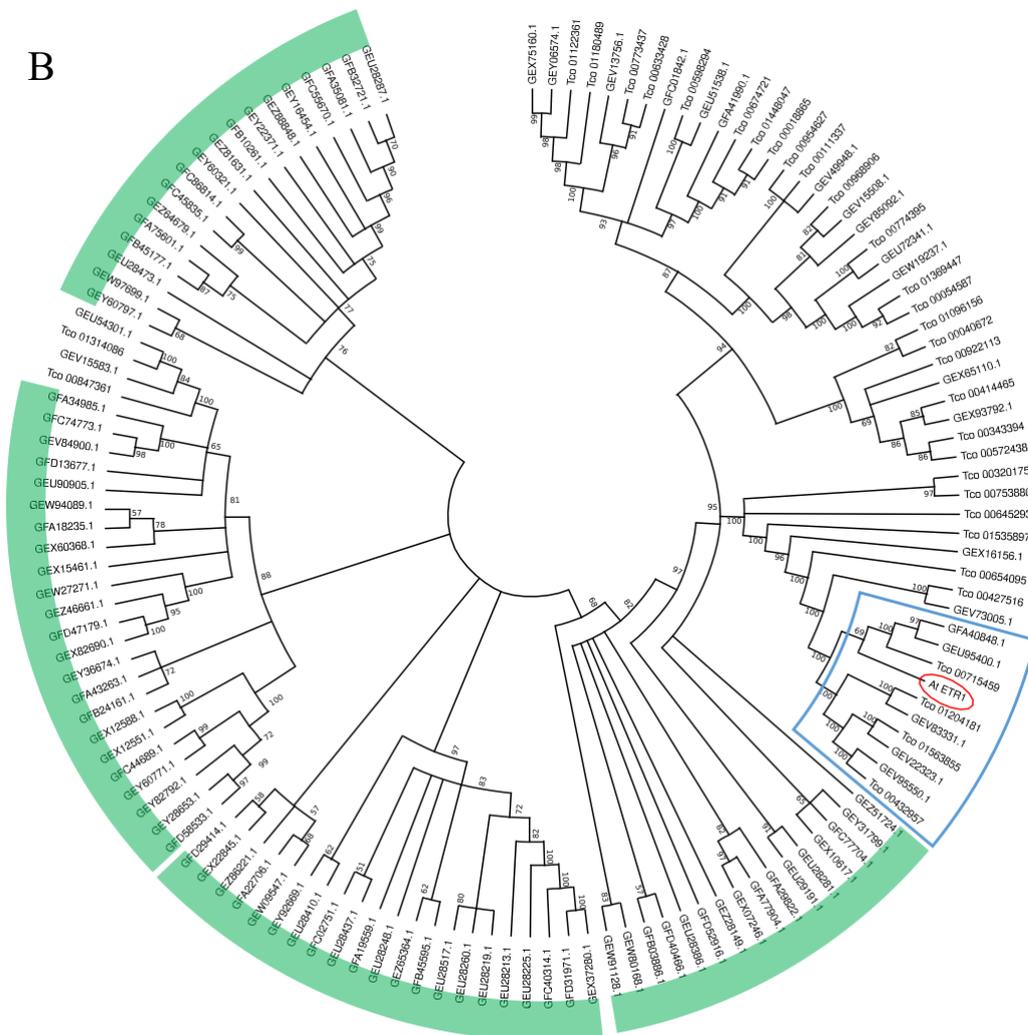


図 3-8 (A)ヒスチジンキナーゼドメインを含むタンパク質をコードする遺伝子の数、(B)ヒスチジンキナーゼスーパーファミリー遺伝子の系統解析

(A) *T. coccineum* と *T. cinerariifolium* の、ヒスチジンキナーゼ様 ATPase (HATPase) もしくは phosphoacceptor receiver (REC) ドメインを有するタンパク質をコードするヒスチジンキナーゼスーパーファミリー遺伝子の数を示したベン図。

(B) *T. coccineum* と *T. cinerariifolium* のシグナル伝達ヒスチジンキナーゼスーパーファミリーに属するタンパク質のアミノ酸配列を用いた系統樹。 *T. coccineum* と *T. cinerariifolium* の遺伝子の名前は、それぞれ、Tco の接頭語がついたものと accession No. で示している。 *A. thaliana* のエチレンレセプター1 (AtETR1) は赤色の円で囲っている。 *T. cinerariifolium* の 5 つのパラログ *T. coccineum* の 4 つのパラログを含む AtETR1 のクラスターは青色の線で囲っている。 *T. cinerariifolium* 特異的なクラスターは緑色の円弧で示している。

引用文献

- Alm Rosenblad, M., A. Abramova, U. Lind, P. Olason, S. Giacomello, B. Nystedt, and A. Blomberg. 2021. 'Genomic Characterization of the Barnacle *Balanus improvisus* Reveals Extreme Nucleotide Diversity in Coding Regions', *Mar Biotechnol (NY)*, 23: 402-16.
- Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. 'Gapped BLAST and PSI-BLAST: a new generation of protein database search programs', *Nucleic Acids Res*, 25: 3389-402.
- Badouin, H., J. Gouzy, C. J. Grassa, F. Murat, S. E. Staton, L. Cottret, C. Lelandais-Briere, G. L. Owens, S. Carrere, B. Mayjonade, L. Legrand, N. Gill, N. C. Kane, J. E. Bowers, S. Hubner, A. Bellec, A. Berard, H. Berges, N. Blanchet, M. C. Boniface, D. Brunel, O. Catrice, N. Chaidir, C. Claudel, C. Donnadiou, T. Faraut, G. Fievet, N. Helmstetter, M. King, S. J. Knapp, Z. Lai, M. C. Le Paslier, Y. Lippi, L. Lorenzon, J. R. Mandel, G. Marage, G. Marchand, E. Marquand, E. Bret-Mestries, E. Morien, S. Nambeesan, T. Nguyen, P. Pegot-Espagnet, N. Pouilly, F. Raftis, E. Sallet, T. Schiex, J. Thomas, C. Vandecasteele, D. Vares, F. Vear, S. Vautrin, M. Crespi, B. Mangin, J. M. Burke, J. Salse, S. Munos, P. Vincourt, L. H. Rieseberg, and N. B. Langlade. 2017. 'The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution', *Nature*, 546: 148-52.
- Bennett, M. D., J. B. Smith, and J. S. Heslop-Harrison. 1982. 'Nuclear DNA amounts in angiosperms', *Proc R Soc Lond B*, 216.
- Boetzer, M., C. V. Henkel, H. J. Jansen, D. Butler, and W. Pirovano. 2011. 'Scaffolding pre-assembled contigs using SSPACE', *Bioinformatics*, 27: 578-9.
- Boetzer, M., and W. Pirovano. 2014. 'SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information', *BMC Bioinformatics*, 15: 211.
- Bolognesi, A., M. Bortolotti, S. Maiello, M. G. Battelli, and L. Polito. 2016. 'Ribosome-Inactivating Proteins from Plants: A Historical Overview', *Molecules*, 21.
- Eddy, S. R. 1998. 'Profile hidden Markov models', *Bioinformatics*, 14: 755-63.
- Hirakawa, H., K. Sumitomo, T. Hisamatsu, S. Nagano, K. Shirasawa, Y. Higuchi, M. Kusaba, M. Koshioka, Y. Nakano, M. Yagi, H. Yamaguchi, K. Taniguchi, M. Nakano, and S. N. Isobe. 2019. 'De novo whole-genome assembly in *Chrysanthemum seticuspe*, a model species of *Chrysanthemums*, and its application to genetic and gene discovery analysis',

- DNA Res*, 26: 195-203.
- Inoue, J., and N. Satoh. 2019. 'ORTHOSCOPE: An Automatic Web Tool for Phylogenetically Inferring Bilaterian Orthogroups with User-Selected Taxa', *Mol Biol Evol*, 36: 621-31.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. 'The rapid generation of mutation data matrices from protein sequences', *Comput Appl Biosci*, 8: 275-82.
- Katsuda, Y. 2012. 'Progress and future of pyrethroids', *Top Curr Chem*, 314: 1-30.
- Kikuta, Y., H. Ueda, M. Takahashi, T. Mitsumori, G. Yamada, K. Sakamori, K. Takeda, S. Furutani, K. Nakayama, Y. Katsuda, A. Hatanaka, and K. Matsuda. 2012. 'Identification and characterization of a GDSL lipase-like protein that catalyzes the ester-forming reaction for pyrethrin biosynthesis in *Tanacetum cinerariifolium*- a new target for plant protection', *Plant J*, 71: 183-93.
- Kumar, S., G. Stecher, M. Li, C. Knyaz, and K. Tamura. 2018. 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Mol Biol Evol*, 35: 1547-49.
- Lamesch, P., T. Z. Berardini, D. Li, D. Swarbreck, C. Wilks, R. Sasidharan, R. Muller, K. Dreher, D. L. Alexander, M. Garcia-Hernandez, A. S. Karthikeyan, C. H. Lee, W. D. Nelson, L. Ploetz, S. Singh, A. Wensel, and E. Huala. 2012. 'The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools', *Nucleic Acids Res*, 40: D1202-10.
- Langdon, W. B. 2015. 'Performance of genetic programming optimised Bowtie2 on genome comparison and analytic testing (GCAT) benchmarks', *BioData Min*, 8: 1.
- Li, K. B. 2003. 'ClustalW-MPI: ClustalW analysis using distributed and parallel computing', *Bioinformatics*, 19: 1585-6.
- Li, W., D. B. Lybrand, F. Zhou, R. L. Last, and E. Pichersky. 2019. 'Pyrethrin Biosynthesis: The Cytochrome P450 Oxidoreductase CYP82Q3 Converts Jasmolone To Pyrethrolone', *Plant Physiol*, 181: 934-44.
- Li, W., F. Zhou, and E. Pichersky. 2018. 'Jasmone Hydroxylase, a Key Enzyme in the Synthesis of the Alcohol Moiety of Pyrethrin Insecticides', *Plant Physiol*, 177: 1498-509.
- Llorens, C., R. Futami, L. Covelli, L. Dominguez-Escriba, J. M. Viu, D. Tamarit, J. Aguilar-Rodriguez, M. Vicente-Ripolles, G. Fuster, G. P. Bernet, F. Maumus, A. Munoz-Pomer, J. M. Sempere, A. Latorre, and A. Moya. 2011. 'The Gypsy Database (GyDB) of mobile genetic elements: release 2.0', *Nucleic Acids Res*, 39: D70-4.
- Lu, S., J. Wang, F. Chitsaz, M. K. Derbyshire, R. C. Geer, N. R. Gonzales, M. Gwadz, D. I.

- Hurwitz, G. H. Marchler, J. S. Song, N. Thanki, R. A. Yamashita, M. Yang, D. Zhang, C. Zheng, C. J. Lanczycki, and A. Marchler-Bauer. 2020. 'CDD/SPARCLE: the conserved domain database in 2020', *Nucleic Acids Res*, 48: D265-D68.
- Luo, R., B. Liu, Y. Xie, Z. Li, W. Huang, J. Yuan, G. He, Y. Chen, Q. Pan, Y. Liu, J. Tang, G. Wu, H. Zhang, Y. Shi, Y. Liu, C. Yu, B. Wang, Y. Lu, C. Han, D. W. Cheung, S. M. Yiu, S. Peng, Z. Xiaoqian, G. Liu, X. Liao, Y. Li, H. Yang, J. Wang, T. W. Lam, and J. Wang. 2012. 'SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler', *Gigascience*, 1: 18.
- Manni, M., M. R. Berkeley, M. Seppey, F. A. Simao, and E. M. Zdobnov. 2021. 'BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes', *Mol Biol Evol*, 38: 4647-54.
- Marcais, G., and C. Kingsford. 2011. 'A fast, lock-free approach for efficient parallel counting of occurrences of k-mers', *Bioinformatics*, 27: 764-70.
- Masella, A. P., A. K. Bartram, J. M. Truszkowski, D. G. Brown, and J. D. Neufeld. 2012. 'PANDAseq: paired-end assembler for illumina sequences', *BMC Bioinformatics*, 13: 31.
- Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, and R. Lanfear. 2020. 'IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era', *Mol Biol Evol*, 37: 1530-34.
- Nadalín, F., F. Vezzi, and A. Policriti. 2012. 'GapFiller: a de novo assembly approach to fill the gap within paired reads', *BMC Bioinformatics*, 13 Suppl 14: S8.
- Ramirez, A. M., T. Yang, H. J. Bouwmeester, and M. A. Jongsma. 2013. 'A trichome-specific linoleate lipoxygenase expressed during pyrethrin biosynthesis in pyrethrum', *Lipids*, 48: 1005-15.
- Rice Annotation, Project, T. Tanaka, B. A. Antonio, S. Kikuchi, T. Matsumoto, Y. Nagamura, H. Numa, H. Sakai, J. Wu, T. Itoh, T. Sasaki, R. Aono, Y. Fujii, T. Habara, E. Harada, M. Kanno, Y. Kawahara, H. Kawashima, H. Kubooka, A. Matsuya, H. Nakaoka, N. Saichi, R. Sanbonmatsu, Y. Sato, Y. Shinso, M. Suzuki, J. Takeda, M. Tanino, F. Todokoro, K. Yamaguchi, N. Yamamoto, C. Yamasaki, T. Imanishi, T. Okido, M. Tada, K. Ikeo, Y. Tateno, T. Gojobori, Y. C. Lin, F. J. Wei, Y. I. Hsing, Q. Zhao, B. Han, M. R. Kramer, R. W. McCombie, D. Lonsdale, C. C. O'Donovan, E. J. Whitfield, R. Apweiler, K. O.

- Koyanagi, J. P. Khurana, S. Raghuvanshi, N. K. Singh, A. K. Tyagi, G. Haberer, M. Fujisawa, S. Hosokawa, Y. Ito, H. Ikawa, M. Shibata, M. Yamamoto, R. M. Bruskiwich, D. R. Hoen, T. E. Bureau, N. Namiki, H. Ohyanagi, Y. Sakai, S. Nobushima, K. Sakata, R. A. Barrero, Y. Sato, A. Souvorov, B. Smith-White, T. Tatusova, S. An, G. An, O. Ota S, G. Fuks, G. Fuks, J. Messing, K. R. Christie, D. Lieberherr, H. Kim, A. Zuccolo, R. A. Wing, K. Nobuta, P. J. Green, C. Lu, B. C. Meyers, C. Chaparro, B. Piegu, O. Panaud, and M. Echeverria. 2008. 'The Rice Annotation Project Database (RAP-DB): 2008 update', *Nucleic Acids Res*, 36: D1028-33.
- Rivera, S. B., B. D. Swedlund, G. J. King, R. N. Bell, C. E. Hussey, Jr., D. M. Shattuck-Eidens, W. M. Wrobel, G. D. Peiser, and C. D. Poulter. 2001. 'Chrysanthemyl diphosphate synthase: isolation of the gene and characterization of the recombinant non-head-to-tail monoterpene synthase from *Chrysanthemum cinerariaefolium*', *Proc Natl Acad Sci US A*, 98: 4373-8.
- Shahidi-Noghabi, S., E. J. Van Damme, and G. Smagghe. 2008. 'Carbohydrate-binding activity of the type-2 ribosome-inactivating protein SNA-I from elderberry (*Sambucus nigra*) is a determining factor for its insecticidal activity', *Phytochemistry*, 69: 2972-8.
- Shakeel, S. N., X. Wang, B. M. Binder, and G. E. Schaller. 2013. 'Mechanisms of signal transduction by ethylene: overlapping and non-overlapping signalling roles in a receptor family', *AoB Plants*, 5: plt010.
- Shen, Q., L. Zhang, Z. Liao, S. Wang, T. Yan, P. Shi, M. Liu, X. Fu, Q. Pan, Y. Wang, Z. Lv, X. Lu, F. Zhang, W. Jiang, Y. Ma, M. Chen, X. Hao, L. Li, Y. Tang, G. Lv, Y. Zhou, X. Sun, P. E. Brodelius, J. K. C. Rose, and K. Tang. 2018. 'The Genome of *Artemisia annua* Provides Insight into the Evolution of Asteraceae Family and Artemisinin Biosynthesis', *Mol Plant*, 11: 776-88.
- Sierro, N., J. N. Battey, S. Ouadi, N. Bakaher, L. Bovet, A. Willig, S. Goepfert, M. C. Peitsch, and N. V. Ivanov. 2014. 'The tobacco genome sequence and its comparison with those of tomato and potato', *Nat Commun*, 5: 3833.
- Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler. 2008. 'Using native and syntenically mapped cDNA alignments to improve de novo gene finding', *Bioinformatics*, 24: 637-44.
- Williams, D., W. L. Trimble, M. Shilts, F. Meyer, and H. Ochman. 2013. 'Rapid quantification of sequence repeats to resolve the size, structure and contents of bacterial genomes', *BMC*

- Genomics*, 14: 537.
- Xu, H., W. Li, A. L. Schillmiller, H. van Eekelen, R. C. H. de Vos, M. A. Jongsma, and E. Pichersky. 2019. 'Pyrethric acid of natural pyrethrin insecticide: complete pathway elucidation and reconstitution in *Nicotiana benthamiana*', *New Phytol*, 223: 751-65.
- Xu, H., G. D. Moghe, K. Wiegert-Rininger, A. L. Schillmiller, C. S. Barry, R. L. Last, and E. Pichersky. 2018. 'Coexpression Analysis Identifies Two Oxidoreductases Involved in the Biosynthesis of the Monoterpene Acid Moiety of Natural Pyrethrin Insecticides in *Tanacetum cinerariifolium*', *Plant Physiol*, 176: 524-37.
- Xu, M., L. Guo, S. Gu, O. Wang, R. Zhang, B. A. Peters, G. Fan, X. Liu, X. Xu, L. Deng, and Y. Zhang. 2020. 'TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads', *Gigascience*, 9.
- Zhu, F., Y. K. Zhou, Z. L. Ji, and X. R. Chen. 2018. 'The Plant Ribosome-Inactivating Proteins Play Important Roles in Defense against Pathogens and Insect Pest Attacks', *Front Plant Sci*, 9: 146.
- Zimin, A. V., and S. L. Salzberg. 2020. 'The genome polishing tool POLCA makes fast and accurate corrections in genome assemblies', *PLoS Comput Biol*, 16: e1007981.

第4章 TciGLIP のトランスフェラーゼ活性に重要な部位の特定

第1節 序論

第2章序論で述べた通り、*Tanacetum cinerariifolium* のピレトリン類生合成において、*T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu モチーフ) リパーゼ (TciGLIP) は、ピレトリン類のうちの一つであるピレトリン I の最終エステル化を担う重要な酵素である (Kikuta et al. 2012)。TciGLIP は GDSL エステラーゼ/リパーゼファミリータンパク質 (GELPs) の一員であり、GELPs は GDSX (Gly-Asp-Ser-X) コンセンサス配列に特徴づけられるリパーゼの一種である。GELPs は植物において、発芽 (Clauss et al. 2008)、受粉における花粉の相互作用 (Takahashi et al. 2010)、脂質代謝 (Gao et al. 2017)、二次代謝 (Ruppert et al. 2005) といった様々な生物学的機能に関与している (Shen et al. 2022)。標準的な GELPs はエステラーゼ活性 (加水分解活性) を有するが、数種の GELPs はトランスフェラーゼ活性を有する。TciGLIP はピレトリン I のエステル化、つまりトランスフェラーゼ活性の他に、ピレトリン I に対するエステラーゼ活性も有するが、その活性はトランスフェラーゼ活性よりも低い (Kikuta et al. 2012)。このトランスフェラーゼ活性を有する GELPs (tr-GELPs) は TciGLIP の他に、*Tanacetum coccineum* における TciGLIP のオーソログである、*T. coccineum* GDSL リパーゼ (TcoGLIP) がある (本論文第3章および (Zeng et al. 2021))。また、別の植物属では、コムギ (*Triticum aestivum*) の *T. aestivum* キサントフィルアシルトランスフェラーゼ (TaXAT) は、キサントフィルとトリアシルグリセリドをキサントフィルエステルに変換する tr-GELP であり、TaXAT も TciGLIP と同様に、キサントフィルエステルに対するエステラーゼ活性よりもトランスフェラーゼ活性の方が高い (Watkins et al. 2019)。さらに、トマト (*Solanum lycopersicum*) の *S. lycopersicum* クロロゲネート: グルカレート カフェオイルトランスフェラーゼ (SICGT) は、トランスフェラーゼ活性を有する一方で、エステラーゼ活性を失った tr-GELP として最初に発見された (Teutschbein et al. 2010)。このような数種の GELPs におけるトランスフェラーゼ活

性は、タンパク質構造に起因する基質親和性によるものであり、これらの特徴は tr-GELPs 間で共有されていると考えられる。GELP に関する多くの研究において、エステラーゼ活性の一般的な活性中心は、Gly-Asp-His の 3 残基からなる catalytic triad と呼ばれる部位であることが示されている (Akoh et al. 2004)。TciGLIP でも、catalytic triad の点変異 (S40A、D318A および H321A) によって、トランスフェラーゼ活性が失われるが (Kikuta et al. 2012) (Kikuta et al. 2013)、一方で、SlCGT における同様の点変異 (H331A) はトランスフェラーゼ活性に影響を与えないことが報告されており (Teutschbein et al. 2010)、トランスフェラーゼ活性に重要な残基は、エステラーゼ活性のそれと完全に一致していないことを示している。これらの知見は、tr-GELP のトランスフェラーゼ活性に重要な残基が catalytic triad 以外にも存在することを示唆する。本章では、tr-GELPs 間で共通のトランスフェラーゼ活性に重要な残基を two entropy 解析を用いて特定し、TciGLIP におけるこれら残基の変異による影響を構造モデル予測および結合シミュレーションによって検証した。

第2節 実験の部

1. シーケンスアラインメントおよび分子系統解析

基質既知の GELPs について、4 つの tr-GELP (TciGLIP、TcoGLIP、TaXAT および SICGT)、およびエステラーゼ活性を有する GELPs (est-GELPs) として AtCDEF1 (*Arabidopsis thaliana* cuticle destructing factor 1) (Takahashi et al. 2010)、BnSCE3 (*Brassica napus* sinapine esterase) (Clauss et al. 2008)、CpEST (*Carica papaya* esterase) (Abdelkafi et al. 2009)、FvGELP1 (*Fragaria vesca* GDSL esterase/lipase) (Lucia, Fong-Chin, and Wilfried 2022)、OsGLIP1 (*Oryza sativa* GDSL lipase) (Gao et al. 2017) および RsAAE (*Rauvolfia serpentine* acetylajmalan acetylerase) (Ruppert et al. 2005) の 6 つのアミノ酸配列を NCBI データベースから取得した。それぞれのアミノ酸配列の accession No. を表 4-1 に示す。また、上記 10 個の各 GELPs のアミノ酸配列をクエリとして NR データベースに対して BLASTP を行い、各クエリに対してヒットした上位 30 個のアミノ酸配列を E-value 値 10^{-3} を閾値としてスクリーニングし、重複している配列を除去して推定 GELPs として用いた。得られた全 GELPs のアミノ酸配列について、CLUSTAL W-mpi 0.13 (Li 2003) を用いてアラインメントし、JTT 行列を用いたモデル (Jones, Taylor, and Thornton 1992) に基づく最尤系統樹を、Fast Tree 2.1.10 (JTT モデル、CAT 近似) (Price, Dehal, and Arkin 2010) を用いてブートストラップ反復回数 100 回で作成した。

表 4-1 本研究で引用した基質既知の GELPs の accession No.

Group	Protein name		Accession No.
tr-GELPs	TciGLIP	<i>Tanacetum cinerariifolium</i> GDSL lipase	AFJ04755.1
	TcoGLIP	<i>Tanacetum coccineum</i> GDSL liapse	GJR32646.1
	TaXAT	<i>Triticum aestivum</i> xanthophyll acyltransferase	QEM23753.1
	SICGT	<i>Solanum lycopersicum</i> chlorogenate: glucarate caffeoyltransferase	CBV37053.1
est-GELPs	AtCDEF1	<i>Arabidopsis thaliana</i> cuticle destructing factor 1	NP_194743.1
	BnSCE3	<i>Brassica napus</i> sinapine esterase	Q3ZFI4.1
	CpEST	<i>Carica papaya</i> esterase	P86276.1
	FvGELP1	<i>Fragaria vesca</i> GDSL esterase/lipase	XP_004304671.2
	OsGELP1	<i>Oryza sativa</i> GDSL lipase	APX55003.1
	RsAAE	<i>Rauvolfia serpentine</i> acetylajmalan acetyesterase	AAW88320.1

GELPs: GDSL (Gly-Asp-Ser-Leu モチーフ) エステラーゼ/リパーゼファミリータンパク質

tr-GELPs: トランスフェラーゼ活性を示す GELPs

est-GELPs: エステラーゼ活性を示す GELPs

2. Two entropy 解析

GELPs のアミノ酸配列アラインメント中で、トランスフェラーゼ活性と相関のあるアミノ酸位置の特定のため、過去の研究に則り (Ye et al. 2006)、Shannon's entropy の計算を行った。要約すると、tr-GELPs とその他の GELPs (est-GELPs および推定 GELPs) について、アラインメント中の位置 p における entropy 値 (E_p) を下記計算式で求めた。

$$E_p = - \sum_{a=1}^{20} \frac{N_{a,p}}{N_{all}} \log_{10} \frac{N_{a,p}}{N_{all}}$$

$N_{a,p}$ はアラインメント中の位置 p におけるアミノ酸 a を有する配列の個数を、BLOSUM62-based pseudo-count strategy (Zea et al. 2017) を用いて補正した値である。 N_{all} はアラインメント内の全ての配列の数を示している。Pseudo-count は 2.00 と設定した。

3. TciGLIP のタンパク質構造モデリングおよび基質結合シミュレーション

天然型の TciGLIP および 4 種の点変異体 (S339A、G64A、D336A および R153A) の各々のアミノ酸配列について、ColabFold (AlphaFold2 with MMseqs2) (Mirdita et al. 2022) のデフォルト設定にてタンパク質構造モデルの予測を行い、各タンパク質のエネルギー下位 5 つのモデルを以降の解析に使用した。Pyrethrin I の基質である chrysanthemoyl-CoA (CHEBI: 143950) および pyrethrolone (CHEBI: 39111) の分子構造ファイルは、ChEBI データベース (<https://www.ebi.ac.uk/chebi/init.do>) からダウンロードし、それぞれの基質分子の CHARMM (Chemistry at Harvard Macromolecular Mechanics) 力場は、Spartan'18 v1.4.5 (Wavefunction, Inc.) を用いて最小化処理を行った。天然型の TciGLIP および変異体の各々の構造モデルに対し、chrysanthemoyl-CoA を基質として AutoDock Vina 1.1.2 (Trott and Olson 2010) を使用して結合モデルを作成した。条件は、グリッドポイントを x: 36、y: 42、z: 44 とし、0.375 Å で分割、*exhaustiveness* を 100、*num_modes* を最大値の 20、その他のパラメーターをデフォルトに設定した。Chrysanthemoyl-CoA のチオールエステル部分の硫黄原子と、TciGLIP の catalytic triad を構成する His321 の C-

2 原子との距離を基準に、chrysanthemoyl-CoA 結合 TciGLIP モデルを選出した。さらに同様に AutoDock Vina を用いて、chrysanthemoyl-CoA 結合 TciGLIP モデルに対する pyrethrolone の結合モデルの作成を行った。グリッドポイントを x: 18、y: 24、z: 22 とし、0.375 Å で分割、*exhaustiveness* を 100、*num_modes* を最大値の 20、その他のパラメーターをデフォルトに設定した。各 TciGLIP について、chrysanthemoyl-CoA と pyrethrolone の両基質が catalytic triad の近傍に存在するモデルを計数し、天然型 TciGLIP を対象とする Dunnett's test を用いて有意水準 0.05 にて有意差を調べた。タンパク質モデルの視覚化は UCSF Chimera 1.16 (Pettersen et al. 2004)を用いて行った。

第3節 結果と考察

1. GELPs のシーケンスアラインメントおよび two entropy 解析

基質既知の GELPs について、TciGLIP、TcoGLIP、TaXAT および SICGT の 4 つの tr-GELP および AtCDEF1、BnSCE3、CpEST、FvGELP1、OsGLIP1 および RsAAE の 6 つの est-GELP のアミノ酸配列を NCBI データベースから取得した。上記 GELPs に加え、BLASTP を用いて上記 GELPs をクエリとした推定 GELPs を取得した。本論文では、est-GELPs と推定 GELPs をまとめて「その他の GELPs」と呼称する。取得した全 GELPs (図 4-1) のアミノ酸配列について、CLUSTAL W-mpi を用いてアラインメントを作成した。

Tr-GELPs で選択的に保存されている残基を検出するために、アラインメント内のそれぞれの位置におけるアミノ酸の two entropy 解析を、過去の研究の方法 (Ye et al. 2006) (Zea et al. 2017) に倣って実施した。Two-entropy 解析は、アデノシン受容体のリガンドの特定 (van Westen et al. 2012) や、カンナビノイド受容体のリガンド認識機構の解明 (Gonzalez et al. 2008) といった研究で使用されている。あるタンパク質の集団間のアミノ酸配列アラインメントにおいて、アミノ酸タイプとタンパク質の機能に相関性がない位置の entropy 値の差は 0 となる。アミノ酸タイプとタンパク質の機能に高い相関性がある位置の entropy 値の差は大きくなるため、tr-GELPs の entropy 値が低く、その他の GELPs の entropy 値が高くなる位置の残基は、tr-GELPs で高度に保存されており、トランスフェラーゼ活性に重要な残基であることを意味する。図 4-2A と B は、tr-GELPs とその他の GELPs との entropy 値の差を、アラインメント中の位置および TciGLIP の catalytic triad からの距離でそれぞれ表した散布図である。N 末端シグナル配列部分において、y 値が負の値をとりやすい (tr-GELPs の entropy 値が比較的高くなる) ことを除いて、領域特異的な entropy 値の差異は見られなかった。Tr-GELPs の entropy 値がその他の GELPs の entropy 値より低く、その差が大きいものからトップ 10 の位置残基のうち、アラインメント中 208 番目の Asn もしくは Arg ([N/R]208) および 484 番目の Asp (D484) は、6 つの est-GELPs

では見られなかった (図 4-3)。対照的に、est-GELPs で選択的に保存されている残基は検出されなかった。これらの結果から、[N/R]208 および D484 は tr-GELP のトランスフェラーゼ活性の獲得に重要な残基であることが示された。

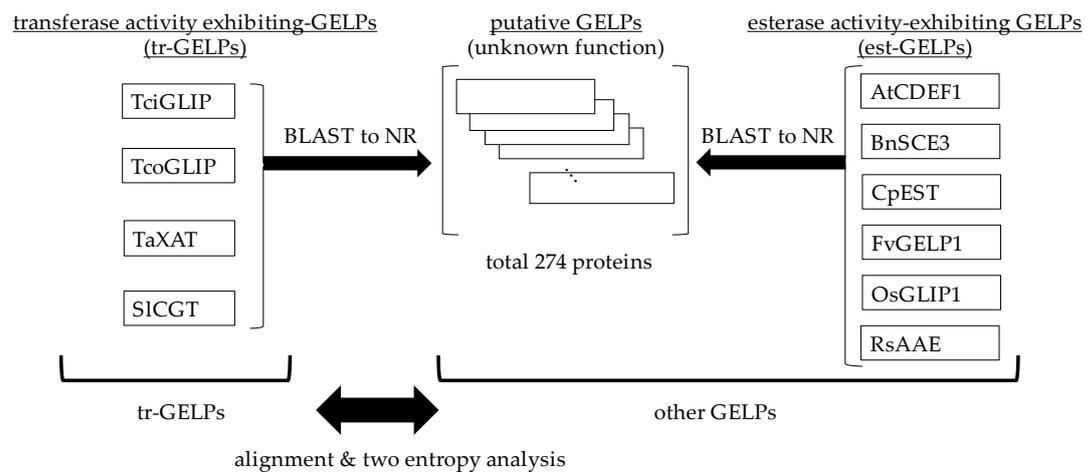


図 4-1 本研究で使用した GELPs

TciGLIP、TcoGLIP、TaXAT および SICGT は tr-GELPs。AtCDEF1、BnSCE3、CpEST、FvGELP1、OsGLIP1 および RsAAE は est-GELPs。推定 GELPs は、tr-GELPs と est-GELPs のアミノ酸配列をクエリとして BLASTP で取得した機能未知の GELP である。est-GELPs と推定 GELPs をまとめて「その他の GELPs」と呼称する。

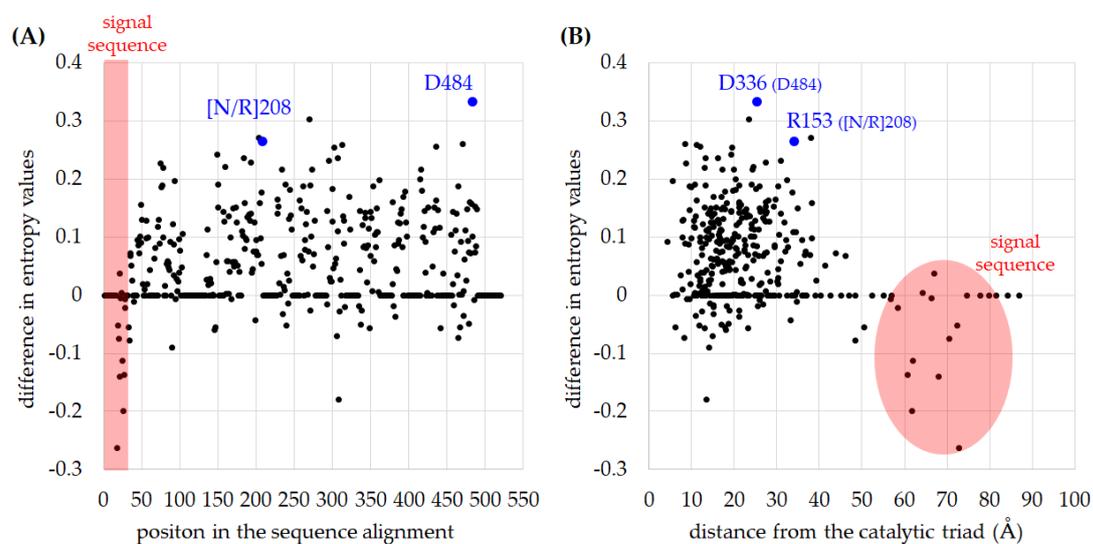


図 4-2 tr-GELPs とその他の GELPs の entropy 値の差を示した散布図

それぞれの散布図の y 軸の値は、その他の GELPs の entropy 値から tr-GELPs の entropy 値を引いた差を示す。この値が大きい残基ほど、その他の GELPs よりも tr-GELPs において選択的に保存されていることを示す。

(A) GELPs のアラインメント中の位置を x 軸とした図。

(B) TciGLIP の catalytic triad からの距離 (Å) を x 軸とした図。TciGLIP における R153 と D336 は、GELPs アラインメント中の [N/R]208 と D484 にそれぞれ対応する。

Tcin_TciGLIP -----MAVASRKLGAIVLVAVLCLSLPTGCLSSQ----QAAALFVFGDSVFDPGNNNHINTH---VNFKANFWPY 63
Tcooc_TcoGLIP -----MAVASRKLGAIVLVAVLCLSLPTGCLSSQ----QAAALFVFGDSVFDPGNNNHINTH---VNFKANFWPY 63
Taes_TaXAT -----MSSRCLLLLAIFLS-TQPTGSGEEGPAPALMVFGDSLVVDVGN---NNYIFTIAKANFPY 58
Slyc_SlCGT MALGMRVVVLLHLISLILLQLIKGDDM---VIKLEPKLKKCGIDRI FQFGDSLSDTGN---CLRESYCGAQTGTGLPY 75
Atha_AtCDEF1 -----MVEGESKALWILATVFAV-AAVAVAVHGQTPCYFVFGDSVFDVNGN---NNALNTKAKVNVLPY 61
Bnap_BnSCE3 -----MASSIKKLITSFLLFFFYT-IIVASSEPCRRYKSIISFGDSLADTGN---YLHLSVDNVHPPQAAFLPY 65
Cpap_CpEST -----MEKPSGQFLGLSLLLLPLLLISCNAQQLFI FGDSDLYDNGNKPFLATD----VPSTFWPY 56
Fves_FvGELP1 -----MLGMEIWKQLVLLPCGLFLTLNVLGVSGATLP---SCSFPALYNFGDSNSDTGG-----ISAAFEPIRTPY 63
Osat_OsGLIP1 ---MAQQAHAAAVVGVLLYCCLCFLFVGVVAGEHGGGGD IKRQYKAMFSGDSLTDGTGICVNMMSAVNRTELTAQPPY 77
Rser_RsAAE -----MGFARLLHLVFSLLVFAGITNGLICPFDSDIYQLGDSLSDTGN---LIRLPDPDPTTAAHFY 60

Block I

Tcin_TciGLIP GQSYFSS-PTGRFSDGRIIPDFI-----AEYASLP--IIPAYLEPNND---FTHG 107
Tcooc_TcoGLIP GQSYFSS-PTGRFSDGRIIPDFI-----AEYASLP--IIPAYLEPNND---FTHG 107
Taes_TaXAT GRDFDKHVATGRFCNGKLLIDFI-----AEKVGFN-GSPLAYLSPEASQGNLLG 107
Slyc_SlCGT GMNFYQN-ATGRCSDGFTILDYI-----AMECGLP--LLNPSLEE---NADFSHG 119
Atha_AtCDEF1 GIDYF-QGPTGRFNGRNIPDVI-----AELAGFN-NPIPPFAG--ASQAQANIG 107
Bnap_BnSCE3 GETFFSV-PTGRSDGRLIIDFI-----AEFLGLP--YVPPYFQSQ--NVSFEQG 110
Cpap_CpEST GLSIDF--PNGRWSDGRIVPDFI-----AEFLGIPFP--PPVLDRSAN---FSSG 99
Fves_FvGELP1 GEGFFRK-PAGRSDGRLIIDFI-----AEGKLP--YLSAYLNS--LGTNYKHG 108
Osat_OsGLIP1 GITFFGH-PTCRSDGRLVVDLFI-----AEGGLP--LLPESKVI---GGDFRRG 121
Rser_RsAAE GETFFGT-PTGRSDGRLIIDFI-----ATAINLP--LLNPYLQ--NVSFRHG 104

Block II

Tcin_TciGLIP ANFASAGAGALIASH--AGLAVGLQTQLRYFGDLVDHYRQNLGDIKSRQL-----LSDAVYLFSCGGND 169
Tcooc_TcoGLIP ANFASAGAGALIASH--AGLAVGLQTQLRYFGDLVDHYRQNLGDIKSRQL-----LSDAVYLLSCGGND 169
Taes_TaXAT ANFASAAAGYNDHGTLLK--AIVSVQQLKFKDYQAKLAVVAGSSHARSII-----ISGSLYIICAGSD 169
Slyc_SlCGT VNFVSGATALSAYELISR-DIAMSTNSSLVQMRMSSYFKSVCSN-----DCAK---YLENSLFLICEIGD 186
Atha_AtCDEF1 LNYASGAGGIREETSEN-----MGERISLRQVNNHNSAIIITAAVPLSR-----LRQCLYTTINIGSD 165
Bnap_BnSCE3 VNFVYAGATLDRAFFIEK-GIVSDFTNVLSVQLNFTFKQILPTLCASS-----RDCREM---LGDLSILMGESGND 180
Cpap_CpEST VTFATADATILGTFP--QTLTLG-----DQVKAFAQIKST-WTDAQ-----RQKGIYMFYIGAND 151
Fves_FvGELP1 ANFATGGSSTIRRPN---ETIFENGISPFSLDMQTAQFLQKSRATDLFRQAKNPFYERSRPNQDFAKALYTFDIGND 184
Osat_OsGLIP1 ANMAIVGTALDFEFESIGVGFPPWYVGSNNVQLRWRDLPSCATA-----APQSIAY---LAESEFLFGSLGGND 192
Rser_RsAAE VNFVAVAGATLDRSFLAAR-GVQVSDIHSLSLAQLNWFRTYLGSI-CSTP-----KECSN---KLKNALFILGNI GND 173

Block III

Tcin_TciGLIP YQSPYY-----PYTQEYVDIVIGNMNTVIK-----GIYEGGRKFGVVNVPLIGCWPGMR--- 220
Tcooc_TcoGLIP YQSPYY-----PYTQEYVDIVIGNMNTVIK-----GIYEGGRKFGIVTVPHIGCWPGMR--- 220
Taes_TaXAT FVYNYIINP---FLDTNQTAEQFSDRLVGMFNNSVT-----QLYEMGARRIGVSLPPFGCFPMAITLYG 231
Slyc_SlCGT VTYGFKQKGP-----IEEVRR--IVPDIVKNIHSVR-----TVIGFGATRILVPGNFPFGCFPIILLFLYM 245
Atha_AtCDEF1 YLNNYFLSPP-TLARRLFNPDQYARSLISLRYIYLT-----QLYVLGARNVALFGIGIGCTPRIVATLQ 229
Bnap_BnSCE3 YNYPFFEDKS-----INEIKE-LTFLIIKAISDAIV-----DLIDLGGKTLVVPGSFPVGCSSAAYLTLFG 239
Cpap_CpEST YLNYTNANLN-----ATAQQEAFVSVQVI AKLKDQLL-----AIYGLGGRKFAFQNLAPLGLCLPIVK--- 208
Fves_FvGELP1 LSAG-FRKLK-----FDQLRA-QIPDIVNQLATAVR-----RIYEQGGRAFVLIHNTGPIGCLPINLFLYNL 242
Osat_OsGLIP1 YNAMVLFGFT-----IDQARN-YTPKIVDQIASGVE-----KLIAMGAVDII VPGVMPFGCFALYLTELK 251
Rser_RsAAE VNYAFPNT-RT-----IEEIRA-YVPFITEAVANATR-----EIIIRLGGSRVIVPGIIPGCVARNLNFNLN 231

Tcin_TciGLIP -AKQPGNT---CNTEVDELTRLHNQAFARLEHLEKQLEG-----FVYAKFDLSTAILNR--MKNPS- 276
Tcooc_TcoGLIP -AKLPGNT---CHTADELTRLHNQAFARLEHLEKQLEG-----FMYAKFDLSTAISDR--MKNPS- 276
Taes_TaXAT HGR-----SGCVSRLLNDAQYNNMKLKAADVSLSKKYHDLK-----IVVLDYAPLYNLATSPVPSQ--- 287
Slyc_SlCGT NDSSTVYDEY-HCAEEWNNFTI SYNLLQQSIIHELNEEYFNIS-----IYGDYNNAYYWLRLNAVAL--- 307
Atha_AtCDEF1 GG-----TGCAEEVNOAVIIFNTKLKALVTD FNNK-PGAM-----FTYVDLFS---GNEDFAAL- 280
Bnap_BnSCE3 TAKEKDYDPLTGLCPWLNDFGKHDEQLKTEIRRLRKLYPHVN-----IMYADYNSLYRLYQKPTKY--- 302
Cpap_CpEST -QDFKTG--NFCLPLASNLAAQHNQLLSETLENLSETLD-----GFNYIYDYFNSLRR--MARPN- 265
Fves_FvGELP1 NPAPGYVDEH-GCVKQNDMAIEFNRQKRVKIKLRAELPQAA-----ITYVDAYAAKYGLISNAKNEG-- 305
Osat_OsGLIP1 SSNKSDYDDY-GCLKPLNELATHHNSLLQTSLAAVQARHRRSPSSPSPPAAAVRIMYADYYAVVAEMMQAPARL--- 327
Rser_RsAAE FFPDGDKDDL-GCLSSLNLSIYFNSLFRALASLSEFPQAV-----IYADYNAWRFLFRNGPAL--- 293

Tcin_TciGLIP -----KYGFKEGESACCG----SGPFGGNYDCG---R---IKEFGLCDNATEYFFDFPHFNPVNASRQFAE 332
Tcooc_TcoGLIP -----KYGFKEGETACCG----SGPFGGIYNCG---R---TKEFKLCDNVTEYFFDFPHFNPVNASRQFAE 332
Taes_TaXAT -----GFTEAKRACCG---TGTVEAS--ILCNLLPG-----TCPSARTVFWVDVWHPSEAANKVVVD 340
Slyc_SlCGT -----GFNKKTLQISCCG--IGGEYNYTESRR--CGKPGAEEK-----ACADPSYSLSWDSGSLTQKAYGWI TK 366
Atha_AtCDEF1 -----GITVGRSCTT-----VNPGEELCAANGP-----VCPDRNKFI FWDNVHTTEVINTVVAN 330
Bnap_BnSCE3 -----GFKNR-PLAACCG-VGGQYNFTIGEE--CYEGEVG-----YCNPSEYINWDGYHI TEAAHQRM AH 359
Cpap_CpEST -----NYGYFTNLACCG---TGSHD-AFGCGFKNV-----HSNLCYQRYGMFDFGRHNAEKTNEVAH 321
Fves_FvGELP1 -----FVDEPMKVC CG---YHVRDHYVWCNKAIIVNGREVYGA.SCGNASSAISWGVHYTQAANQVAVN 365
Osat_OsGLIP1 -----GFRS---GTAACCGAGGGEYNWYVAR--CGMRGAA-----ACANPSSAVCWDGAGHTTEAANRVIAG 384
Rser_RsAAE -----GSNSTSLKCCCGIGG-PYNYDPDRE--CGSRGVP-----VCPNPTQYIQWDTHTFTQAAVYRVAE 351

Block V

Tcin_TciGLIP MFWDGDSMVTQPYNLKALFEGKPSKYLPNDEL----- 405
Tcooc_TcoGLIP MFWDGDSMVIQPYNLKALFEGKPSKFLPNDEL----- 405
Taes_TaXAT SILVDEINNLVA----- 391
Slyc_SlCGT WLIDIDLPQ---LNCRV----- 420
Atha_AtCDEF1 AAFNGPIASPFNISQLVN----- 388
Bnap_BnSCE3 GIILNGPYATPAFNWSCIDA-ASVDNESSFGS----- 429
Cpap_CpEST LIFSADPSVVFPNMLREL FVHP----- 383
Fves_FvGELP1 HILNGALS-NPPIPIITQACHRS----- 426
Osat_OsGLIP1 GWLRGPHYCHP---PILL----- 438
Rser_RsAAE YVIPGIKA---LKCSYNSIQPFLREGEGRQALRLNERE 427

図 4-3 基質既知の GELPs のアミノ酸配列アラインメント

1 列あたり 80 文字のアミノ酸もしくはギャップ (-) で記載。赤色の文字および青色の文字は、それぞれ catalytic triad および本研究で検出したトランスフェラーゼ活性に重要な残基候補 ([N/R]208 および D484) を示す。Block I、II、III および V は GELPs で一般的に保存されている領域を示す。

2. GELPs の分子系統解析

Tr-GELPs、est-GELPs および推定 GELPs のアミノ酸配列から作成した分子系統樹を 図 4-4 に示す。キク科植物 (*T. cinerariifolium* と *T. coccineum* を含む) のタンパク質クラスタ (図 4-4、クレード A) において、[N/R]208 と D484 を両方有する GELP は TciGLIP と TcoGLIP のみであり、[N/R]208 もしくは D484 のどちらか片方を有する GELP がいくつか存在していた。ナス科植物 (*S. lycopersicum* を含む) のタンパク質において、D484 を有する GELPs がクラスタを形成しており (図 4-4、クレード B)、さらに、[N/R]208 と D484 を両方有する GELP が SICGT の他に 2 つ存在していた。 *Triticum* 属および *Hordeum* 属植物のタンパク質において、TaXAT の他に [N/R]208 と D484 を両方有する 6 つの GELPs および D484 を有する 3 つの GELPs がクラスタを形成していた (図 4-4、クレード C)。この [N/R]208 と D484 を両方有する 6 つの GELPs は、TaXAT と同様のトランスフェラーゼ活性を示すことが考えられる。TciGLIP、TcoGLIP、TAXAT および SICGT の 4 つの tr-GELPs の他に、[N/R]208 と D484 を両方有する 9 つのアブラナ科植物由来の GELPs がクラスタを形成していた (図 4-4、クレード D)。興味深いことに、このクラスタは、est-GELP の BnSCE3 を含む別のアブラナ科植物タンパク質クラスタ (図 4-4、クレード D-II) から離れた位置に存在していた。この結果は、これらの GELPs が同じ植物ファミリー内で独立して重複化してきたことを示唆している。

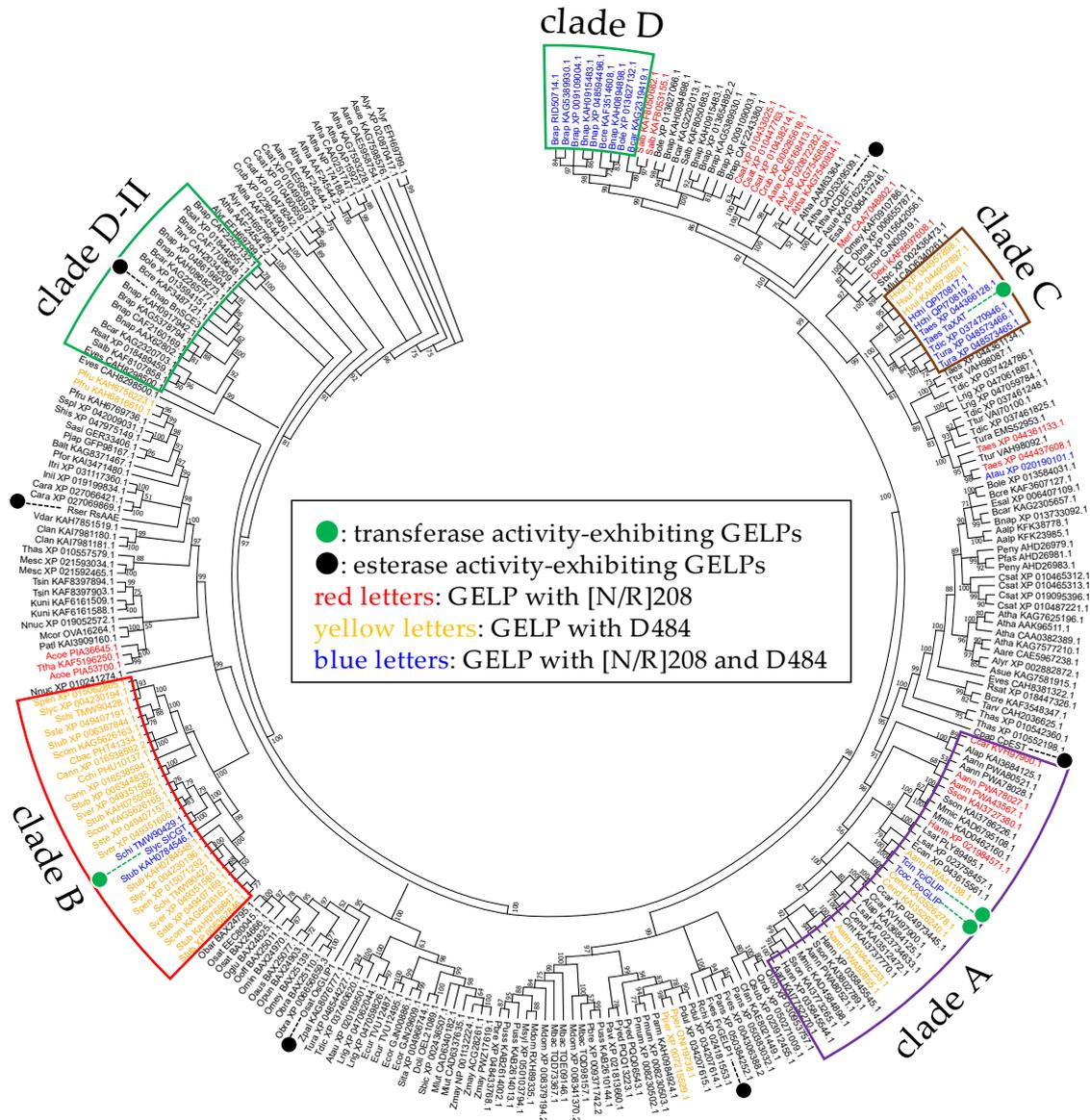


図 4-4 本研究で使用した GELPs の分子系統樹
 各々のタンパク質の名前は、植物の学名の省略名称 (表 4-2) と、慣用名もしくは accession No. で示す。緑色の丸と黒色の丸は、それぞれ tr-GELPs と est-GELPs を示す。赤色文字、黄色文字、青色文字は、それぞれ [N/R]208 を有する GELP、D484 を有する GELP、[N/R]208 と D484 を両方有する GELP をそれぞれ示す。キク科植物のタンパク質クラスタ (クレード A)、ナス科植物のタンパク質クラスタ (クレード B)、*Triticum* 属および *Hordeum* 属植物のタンパク質クラスタ (クレード C) およびアブラナ科植物のタンパク質クラスタ (クレード D、D-II) はそれぞれ紫色、赤色、茶色および緑色の線で囲った。

表 4-2 図 4-4 の分子系統樹中で使用した植物の学名略称対応表

Abbrev.	Species name	Abbrev.	Species name
Aalp	<i>Arabis alpina</i>	Obra	<i>Oryza brachyantha</i>
Aann	<i>Artemisia annua</i>	Oglu	<i>Oryza glumipatula</i>
Aare	<i>Arabidopsis arenosa</i>	Omey	<i>Oryza meyeriana</i>
Aart	<i>Ambrosia artemisiifolia</i>	Omin	<i>Oryza minuta</i>
Acoe	<i>Aquilegia coerulea</i>	Ooff	<i>Oryza officinalis</i>
Alap	<i>Arctium lappa</i>	Opun	<i>Oryza punctata</i>
Alyr	<i>Arabidopsis lyrata</i>	Osat	<i>Oryza sativa</i>
Asue	<i>Arabidopsis suecica</i>	Pans	<i>Potentilla anserina</i>
Atau	<i>Aegilops tauschii</i>	Parm	<i>Prunus armeniaca</i>
Atha	<i>Arabidopsis thaliana</i>	Patl	<i>Papaver atlanticum</i>
Balt	<i>Buddleja alternifolia</i>	Pavi	<i>Prunus avium</i>
Bcar	<i>Brassica carinata</i>	Pbre	<i>Pyrus x bretschneideri</i>
Bcre	<i>Brassica cretica</i>	Pdul	<i>Prunus dulcis</i>
Bnap	<i>Brassica napus</i>	Peny	<i>Pachycladon enysii</i>
Bole	<i>Brassica oleracea</i>	Pfas	<i>Pachycladon fastigiatum</i>
Brap	<i>Brassica rapa</i>	Pfor	<i>Paulownia fortunei</i>
Cann	<i>Capsicum annuum</i>	Pfru	<i>Perilla frutescens</i>
Cara	<i>Coffea arabica</i>	Pjap	<i>Phtheirospermum japonicum</i>
Cbac	<i>Capsicum baccatum</i>	Pmum	<i>Prunus mume</i>
Ccar	<i>Cynara cardunculus</i>	Pper	<i>Prunus persica</i>
Cchi	<i>Capsicum chinense</i>	Puss	<i>Pyrus ussuriensis</i>
Cend	<i>Cichorium endivia</i>	Pyed	<i>Prunus yedoensis</i>
Cfan	<i>Carpinus fangiana</i>	Qlob	<i>Quercus lobata</i>
Cint	<i>Cichorium intybus</i>	Qrob	<i>Quercus robur</i>
Clan	<i>Camellia lanceoleosa</i>	Qsub	<i>Quercus suber</i>
Cpap	<i>Carica papaya</i>	Rchi	<i>Rosa chinensis</i>
Crub	<i>Capsella rubella</i>	Rsat	<i>Raphanus sativus</i>
Csat	<i>Camelina sativa</i>	Rser	<i>Rauwolfia serpentina</i>
Dexi	<i>Digitaria exilis</i>	Salb	<i>Sinapis alba</i>
Doli	<i>Dichanthelium oligosanthes</i>	Sasi	<i>Striga asiatica</i>
Ecan	<i>Erigeron canadensis</i>	Sbic	<i>Sorghum bicolor</i>
Ecor	<i>Eleusine coracana</i>	Schi	<i>Solanum chilense</i>
Ecur	<i>Eragrostis curvula</i>	Scom	<i>Solanum commersonii</i>
Esal	<i>Eutrema salsugineum</i>	Shis	<i>Salvia hispanica</i>
Eves	<i>Eruca vesicaria</i>	Sita	<i>Setaria italica</i>
Fves	<i>Fragaria vesca</i>	Slyc	<i>Solanum lycopersicum</i>
Hann	<i>Helianthus annuus</i>	Spen	<i>Solanum pennellii</i>
Hchi	<i>Hordeum chilense</i>	Sson	<i>Smallanthus sonchifolius</i>
Hvul	<i>Hordeum vulgare</i>	Sspl	<i>Salvia splendens</i>
Inil	<i>Ipomoea nil</i>	Sste	<i>Solanum stenotomum</i>
Itri	<i>Ipomoea triloba</i>	Stub	<i>Solanum tuberosum</i>
Kuni	<i>Kingdonia uniflora</i>	Sver	<i>Solanum verrucosum</i>
Lrig	<i>Lolium rigidum</i>	Taes	<i>Triticum aestivum</i>
Lsat	<i>Lactuca sativa</i>	Tarv	<i>Thlaspi arvense</i>
Mbac	<i>Malus baccata</i>	Tcin	<i>Tanacetum cinerariifolium</i>
Mcor	<i>Macleaya cordata</i>	Tcoc	<i>Tanacetum coccineum</i>
Mdom	<i>Malus domestica</i>	Tdic	<i>Triticum dicoccoides</i>
Merr	<i>Microthlaspi erraticum</i>	Thas	<i>Tarenaya hassleriana</i>
Mesc	<i>Manihot esculenta</i>	Tsin	<i>Tetracentron sinense</i>
Mlut	<i>Miscanthus lutarioriparius</i>	Ttha	<i>Thalictrum thalictroides</i>
Mmic	<i>Mikania micrantha</i>	Ttur	<i>Triticum turgidum</i>
Msyl	<i>Malus sylvestris</i>	Tura	<i>Triticum urartu</i>
Nnuc	<i>Nelumbo nucifera</i>	Vdar	<i>Vaccinium darrowii</i>
Oaus	<i>Oryza australiensis</i>	Zmay	<i>Zea mays</i>
Obar	<i>Oryza barthii</i>	Zpal	<i>Zizania palustris</i>

3. TciGLIP のタンパク質構造予測および基質結合シミュレーション

酵素が活性を発揮する上で、活性中心への基質の接近は必要不可欠である。基質が活性中心へ適した接近をしている構造モデルは「合理的モデル」として定義され、この合理的モデルの数と基質親和性との関係 (Szollosi et al. 2016) (Rabal et al. 2016)は、酵素の構造と機能の相関を予測する際に用いられる。TciGLIP において、GELPs アラインメント中の [N/R]208 と D484 にそれぞれ対応する残基である R153 と D336 のトランスフェラーゼ活性への寄与を検証するために、TciGLIP の catalytic triad に対して、pyrethrin I の基質 (chrysanthemoyl-CoA と pyrethrolone) が接近できたモデルの数を調べた。TciGLIP は天然型と候補残基を Ala で置換した R153A と D336A 仮想変異体に加えて、S339A と G64A 変異体を使用した。これは、S339A 変異体はトランスフェラーゼ活性が失われないことに對し、G64A 変異体は失われるということが実験的に確かめられている (Kikuta et al. 2012) (Kikuta et al. 2013)ことから、本解析のネガティブコントロールならびにポジティブコントロールとして利用するためである。天然型の TciGLIP と 4 つの変異体 (S339A、G64A、R153A および D336A) のアミノ酸配列について、ColabFold を用いてタンパク質構造モデルを予測した。各モデルに対して、AutoDock Vina を用いて chrysanthemoyl-CoA と pyrethrolone を結合させ、合理的モデルの数を調べた (図 4-5)。

これらの解析により、天然型の TciGLIP は 45.3 ± 7.2 個の合理的モデルが生成された (図 4-6A)。同様に、S339A 変異体は 35.0 ± 4.5 個の合理的モデルが生成され (図 4-6A)、これら 2 つのトランスフェラーゼ活性ポジティブな TciGLIP 間で、有意差は検出されなかった (図 4-6A、 $P < 0.05$)。対照的に、G64A 変異体の合理的モデルの数は 25.0 ± 10.2 個であり、天然型のそれと比較して有意に少なかった (図 4-6A、 $P < 0.05$)。これらの結果は、S339A 変異体が天然型と同等のトランスフェラーゼ活性を有する一方で (Kikuta et al. 2012)、G64A 変異体がトランスフェラーゼ活性を失っていた (Kikuta et al. 2013) という実験的結果と十分に一致している。特に興味深いことに、D336A 変異体と R153A 変異体の合理的

モデルの数はそれぞれ 25.3 ± 6.3 個と 21.0 ± 5.4 個であり、これらの値は天然型のそれと比較して優位に少なく、G64A 変異体とほぼ同等であった (図 4-6A、 $P < 0.05$)。これらの結果から、R153 と D336 は TciGLIP のトランスフェラーゼ活性において、重要な残基であることが示唆された。とりわけ、R153 と D336 はどちらも、TciGLIP の catalytic triad から離れた場所に位置し (図 4-2B、図 4-6B)、GELPs で一般的に保存されている Block I、II、III および V 領域内には存在していなかった (図 4-3)。以上をまとめると、これらの構造解析から、R153 と D336 は、TciGLIP のトランスフェラーゼ活性の活性型立体配座を遠位で制御していることが示唆された。

本章において、two entropy 解析、タンパク質構造予測および結合シミュレーションの組み合わせによって、TciGLIP のトランスフェラーゼ活性に重要な残基が示された。本研究は、トランスフェラーゼ活性獲得の根底にある進化的分子メカニズム究明への道を開く。TciGLIP の R153 および D336 変異によるトランスフェラーゼ活性への影響の実験的検証は、現在進行中である。

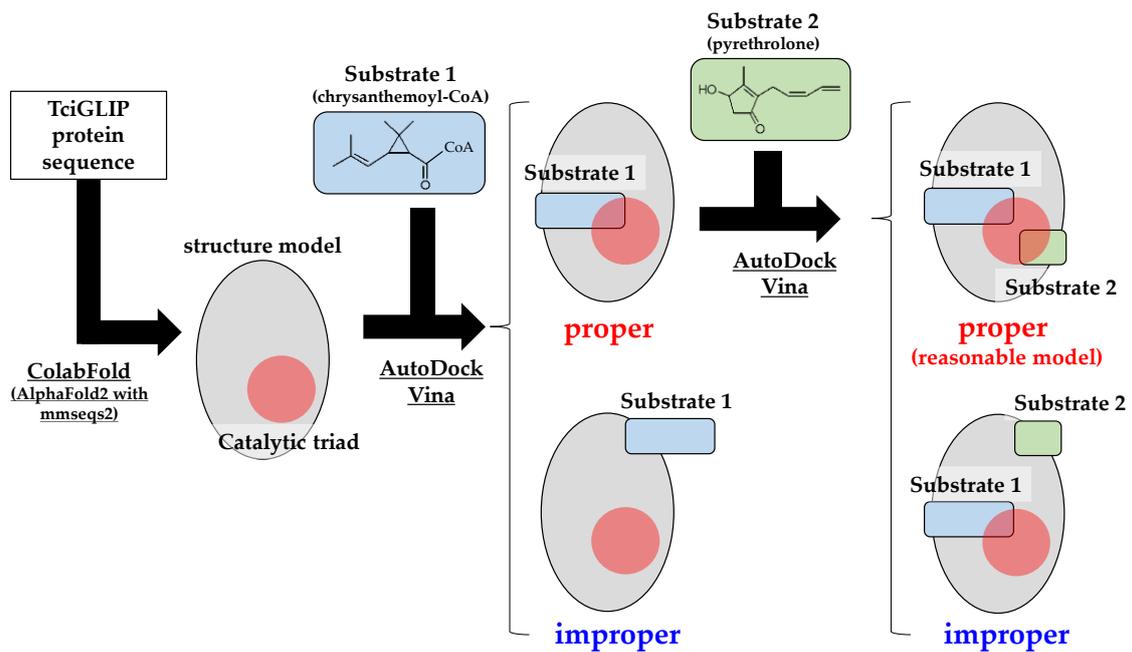


図 4-5 合理的モデルの選出スキーム

TciGLIP の天然型および変異体のアミノ酸配列を ColabFold に供し、タンパク質構造モデルを予測。AutoDock Vina を使用して、各々のモデルに対して chrysanthemoyl-CoA (substrate 1) を結合させる。substrate 1 が TciGLIP の catalytic triad に接近できたモデルに対し、さらに AutoDock Vina を用いて、pyrethrolone (substrate 2) を結合させる。Substrate 1 と 2 がどちらも TciGLIP の catalytic triad に接近できたモデルを「合理的モデル」とみなした。

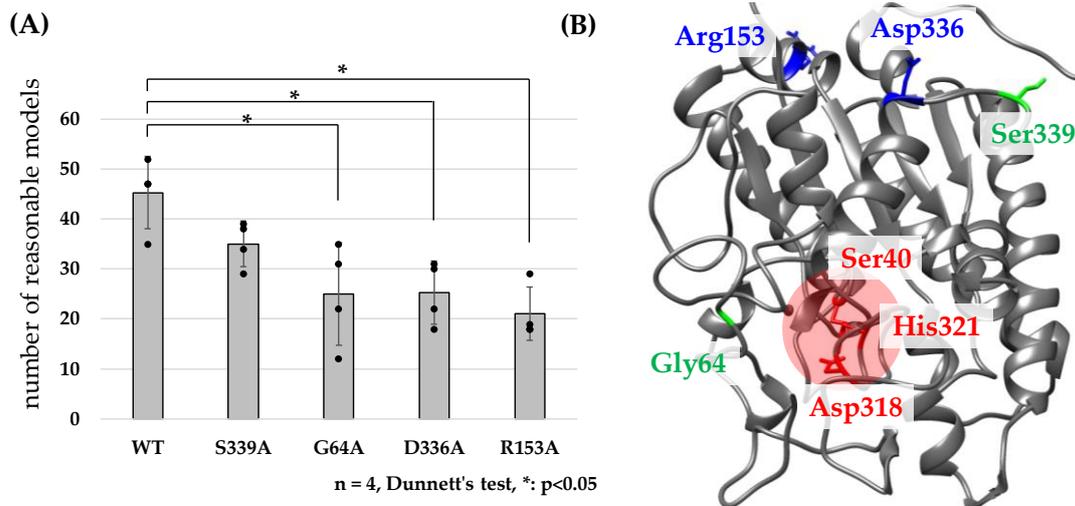


図 4-6 (A) 各 TciGLIP の合理的モデルの数 (B) 天然型 TciGLIP の予測モデル
 (B) 特徴的な残基を着色し、UCSF Chimera 1.16 を用いて視覚化した図。赤色の文字 (Ser40、Asp318 および His321) は TciGLIP の catalytic triad を示す。緑色の文字 (Gly64 および Ser339) は、以前の研究で、変異によるトランスフェラーゼ活性への影響が実験的に確かめられた残基を示す。青色の文字 (Arg153 および Asp336) は本研究で特定したトランスフェラーゼ活性に重要な残基候補を示す。

引用文献

- Abdelkafi, S., H. Ogata, N. Barouh, B. Fouquet, R. Lebrun, M. Pina, F. Scheirlinckx, P. Villeneuve, and F. Carriere. 2009. 'Identification and biochemical characterization of a GDSL-motif carboxylester hydrolase from *Carica papaya* latex', *Biochim Biophys Acta*, 1791: 1048-56.
- Akoh, C. C., G. C. Lee, Y. C. Liaw, T. H. Huang, and J. F. Shaw. 2004. 'GDSL family of serine esterases/lipases', *Prog Lipid Res*, 43: 534-52.
- Clauss, K., A. Baumert, M. Nimtz, C. Milkowski, and D. Strack. 2008. 'Role of a GDSL lipase-like protein as sinapine esterase in Brassicaceae', *Plant J*, 53: 802-13.
- Gao, M., X. Yin, W. Yang, S. M. Lam, X. Tong, J. Liu, X. Wang, Q. Li, G. Shui, and Z. He. 2017. 'GDSL lipases modulate immunity through lipid homeostasis in rice', *PLoS Pathog*, 13: e1006724.
- Gonzalez, A., L. S. Duran, R. Araya-Secchi, J. A. Garate, C. D. Pessoa-Mahana, C. F. Lagos, and T. Perez-Acle. 2008. 'Computational modeling study of functional microdomains in cannabinoid receptor type 1', *Bioorg Med Chem*, 16: 4378-89.
- Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. 'The rapid generation of mutation data matrices from protein sequences', *Comput Appl Biosci*, 8: 275-82.
- Kikuta, Y., H. Ueda, M. Takahashi, T. Mitsumori, G. Yamada, K. Sakamori, K. Takeda, S. Furutani, K. Nakayama, Y. Katsuda, A. Hatanaka, and K. Matsuda. 2012. 'Identification and characterization of a GDSL lipase-like protein that catalyzes the ester-forming reaction for pyrethrin biosynthesis in *Tanacetum cinerariifolium*- a new target for plant protection', *Plant J*, 71: 183-93.
- Kikuta, Y., G. Yamada, T. Mitsumori, T. Takeuchi, K. Nakayama, Y. Katsuda, A. Hatanaka, and K. Matsuda. 2013. 'Requirement of catalytic-triad and related amino acids for the acyltransferase activity of *Tanacetum cinerariifolium* GDSL lipase/esterase TcGLIP for ester-bond formation in pyrethrin biosynthesis', *Biosci Biotechnol Biochem*, 77: 1822-5.
- Li, K. B. 2003. 'ClustalW-MPI: ClustalW analysis using distributed and parallel computing', *Bioinformatics*, 19: 1585-6.
- Lucia, D. W., H. Fong-Chin, and S. Wilfried. 2022. 'Identification and characterization of strawberry FvGELP1 (*Fragaria vesca* GDSL esterase/lipase)', *Biodiversitas Journal of Biological Diversity*, 23: 9.
- Mirdita, M., K. Schutze, Y. Moriwaki, L. Heo, S. Ovchinnikov, and M. Steinegger. 2022. 'ColabFold: making protein folding accessible to all', *Nat Methods*, 19: 679-82.
- Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. 2004. 'UCSF Chimera--a visualization system for exploratory research and analysis', *J Comput Chem*, 25: 1605-12.

- Price, M. N., P. S. Dehal, and A. P. Arkin. 2010. 'FastTree 2--approximately maximum-likelihood trees for large alignments', *PLoS One*, 5: e9490.
- Rabal, O., F. Pastor, H. Villanueva, M. M. Soldevilla, S. Hervás-Stubbs, and J. Oyarzabal. 2016. 'In Silico Aptamer Docking Studies: From a Retrospective Validation to a Prospective Case Study-TIM3 Aptamers Binding', *Mol Ther Nucleic Acids*, 5: e376.
- Ruppert, M., J. Woll, A. Giritch, E. Genady, X. Ma, and J. Stockigt. 2005. 'Functional expression of an ajmaline pathway-specific esterase from Rauvolfia in a novel plant-virus expression system', *Planta*, 222: 888-98.
- Shen, G., W. Sun, Z. Chen, L. Shi, J. Hong, and J. Shi. 2022. 'Plant GDSL Esterases/Lipases: Evolutionary, Physiological and Molecular Functions in Plant Development', *Plants (Basel)*, 11.
- Szollosi, D., A. Erdei, G. Gyimesi, C. Magyar, and T. Hegedus. 2016. 'Access Path to the Ligand Binding Pocket May Play a Role in Xenobiotics Selection by AhR', *PLoS One*, 11: e0146066.
- Takahashi, K., T. Shimada, M. Kondo, A. Tamai, M. Mori, M. Nishimura, and I. Hara-Nishimura. 2010. 'Ectopic expression of an esterase, which is a candidate for the unidentified plant cutinase, causes cuticular defects in Arabidopsis thaliana', *Plant Cell Physiol*, 51: 123-31.
- Teutschbein, J., W. Gross, M. Nimtz, C. Milkowski, B. Hause, and D. Strack. 2010. 'Identification and localization of a lipase-like acyltransferase in phenylpropanoid metabolism of tomato (*Solanum lycopersicum*)', *J Biol Chem*, 285: 38374-81.
- Trott, O., and A. J. Olson. 2010. 'AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading', *J Comput Chem*, 31: 455-61.
- van Westen, G. J., O. O. van den Hoven, R. van der Pijl, T. Mulder-Krieger, H. de Vries, J. K. Wegner, A. P. Ijzerman, H. W. van Vlijmen, and A. Bender. 2012. 'Identifying novel adenosine receptor ligands by simultaneous proteochemometric modeling of rat and human bioactivity data', *J Med Chem*, 55: 7010-20.
- Watkins, J. L., M. Li, R. P. McQuinn, K. X. Chan, H. E. McFarlane, M. Ermakova, R. T. Furbank, D. Mares, C. Dong, K. J. Chalmers, P. Sharp, D. E. Mather, and B. J. Pogson. 2019. 'A GDSL Esterase/Lipase Catalyzes the Esterification of Lutein in Bread Wheat', *Plant Cell*, 31: 3092-112.
- Ye, K., E. W. Lameijer, M. W. Beukers, and A. P. Ijzerman. 2006. 'A two-entropies analysis to identify functional positions in the transmembrane region of class A G protein-coupled receptors', *Proteins*, 63: 1018-30.
- Zea, D. J., D. Anfossi, M. Nielsen, and C. Marino-Buslje. 2017. 'MIToS.jl: mutual information

tools for protein sequence analysis in the Julia language', *Bioinformatics*, 33: 564-65.

Zeng, T., J. W. Li, L. Zhou, Z. Z. Xu, J. J. Li, H. Hu, J. Luo, R. R. Zheng, Y. Y. Wang, and C. Y. Wang. 2021. 'Transcriptional Responses and GCMS Analysis for the Biosynthesis of Pyrethrins and Volatile Terpenes in *Tanacetum coccineum*', *Int J Mol Sci*, 22.

第5章 総括

本論文では、除虫菊 (*Tanacetum cinerariifolium*) の特性を明らかにし、ピレトリン類をはじめとする除虫菊の植物資源としての重要性、および、トランスジェニック体やゲノム編集体を用いる新たな植物代謝工学的な有用物質生産体として活用する道を拓くため、遺伝的情報基盤の構築を目的とした基礎研究について論述した。具体的に述べると、除虫菊および類縁植物であるアカバナムシヨケギク (*Tanacetum coccineum*) のゲノム解析およびゲノム比較を行い、種特異的二次代謝産物であるピレトリン類をはじめとする生体防御機構の違いを明らかにした。また、ピレトリン類生合成酵素の一つである TciGLIP (*T. cinerariifolium* GDSL (Gly-Asp-Ser-Leu motif) リパーゼ) について、その活性に重要なアミノ酸残基を *in silico* 解析により特定した。

ピレトリン類とその合成類縁体であるピレスロイドは世界的に広汎で使用されている殺虫成分であり、昆虫が媒介する感染症から人の生命を守る役割を果たしている。化学合成が困難なピレトリン類と異なり、ピレスロイドは安定的な生産が可能のため、頻繁に使用されているが、その乱用によりピレスロイド抵抗性害虫の発生が問題となっている。一方で、ピレトリン類を含む除虫菊抽出物はこのピレスロイド抵抗性害虫への効果が確認されており、将来的にピレトリン類の需要の拡大が予想される。ピレトリン類は 6 種のエステル化合物の混合物であり、その生合成には多段階の反応を必要とし、生合成経路は完全に解明されていない。そのため、現在でも除虫菊の大規模栽培による物量的生産方法でその需要を賄っているが、今後拡大する需要に対応するために、さらなる効率的な生産方法の開発が求められる。過去の研究から、除虫菊が傷害を受けた際に放出する揮発性有機化合物 (VOCs) は近隣の無傷の除虫菊個体のピレトリン類生合成を促進させる現象が確認されており、ピレトリン類の生合成制御技術に応用できることが考えられているが、この VOCs を介したピレトリン類生産量増大メカニズムについては解明されていない。さらに、上述した除虫菊抽出物のピレスロイド抵抗性害虫への効果は、ピレトリン類以外の成分も関与している可能性

が考えられており、その解明は除虫菊のさらなる有効活用につながる。このように、植物科学の面でも、有用物質生産の面でも高い潜在的な重要性を秘める除虫菊類について本質的な研究を進展させるためには、除虫菊やその近縁種のゲノム情報とピレトリンの鍵酵素である TciGLIP のタンパク質構造情報は不可欠である。

第 2 章において、除虫菊のドラフトゲノム解読を行い、全長約 7.1 Gb、N50 が 14 Kb、BUSCO 解析による完全性が 91.8%のドラフトゲノムを作成した。植物ゲノムは他の真核生物ゲノムと比較して、ゲノム上を移動する塩基配列である転移因子 (Transposable Element: TE) が重複しやすい傾向があることから、植物ゲノムの特徴を調べる際には TE の解析を伴うことが多い。そこで、除虫菊ゲノム内の TE を解析し、他の植物のそれらと比較を行ったところ、sire および oryco クレード TE が他のキク科植物から分岐した後に重複されたことが示された。除虫菊のゲノムサイズは他のキク科植物のゲノムと比較しても大きく、除虫菊ゲノムの肥大化には、この sire および oryco クレード TE の重複が関与していることが示唆された。遺伝子の属間比較解析では、シグナリング伝達関連タンパク質としてヒスチジinkinナーゼ、生体防御関連タンパク質としてリボソーム不活性化タンパク質 (RIP)、二次代謝関連タンパク質としてリボキシゲナーゼやシトクロム P450 が、除虫菊で特異的に重複しているタンパク質スーパーファミリーとして検出された。ヒスチジinkinナーゼは、VOC 誘導性のシグナル伝達機能を有する *Arabidopsis thaliana* のエチレンレセプター AtETR1 を含むことから、除虫菊におけるヒスチジinkinナーゼの重複は、ピレトリン類の VOCs 依存的調節機構を解明する上での決定的な手掛かりとなりうる。RIP は生体防御成分として産生される毒性タンパク質であり、RIP の重複は除虫菊の生体防御機構の特徴と考えられる。さらに、セイヨウニワトコの殺虫性 RIP である SNA-I (*Sambucus nigra* agglutinin I) と相同性を有する RIP が検出されたことから、除虫菊は対食植性昆虫防御成分としてピレトリン類だけでなく、RIP も利用していることが示唆された。一方で、対真菌防御成分とされるエンドキチナーゼ遺伝子の数が除虫菊では他の植物よりも少なかった。

これらの結果は、除虫菊が原産地であるバルカン半島の乾燥した気候を反映した防御機構を構築してきたことを示唆していた。ピレトリン類生合成関連酵素の系統解析では、特に TciGLIP において、同じタンパク質ファミリーで配列相同性の高い GDSL リパーゼが除虫菊ゲノム内で大量に重複されたことが示された。さらに、TciGLIP の分布解析では、TciGLIP をコードする遺伝子座の近傍に、他の機能未知な GDSL リパーゼをコードする遺伝子が共局在することが示された。これらの結果は、未解明のピレトリン類生合成酵素の解明への手掛かりとなる。

除虫菊の近縁種であるアカバナムシヨケギクは、除虫菊と同様にピレトリン類を生合成するが、その産生量はかなり少ない。そこで第 3 章において、アカバナムシヨケギクのドラフトゲノムを解読し、除虫菊のゲノムとの比較を行うことで、ピレトリン類生合成関連酵素をはじめとする遺伝的特徴の違いを網羅的に調べた。アカバナムシヨケギクにおいて、全長約 9.46 Gb、N50 が 27.8 Kb、BUSCO 解析による完全性が 97.8%のドラフトゲノムを作成した。除虫菊の既知のピレトリン類生合成関連酵素について、アカバナムシヨケギクゲノム内での存在を確認したところ、いずれの酵素についても高い配列相同性を有する対応した酵素が検出された。これら酵素をコードする遺伝子のシntenニー解析を行ったところ、除虫菊における TciGLIP の遺伝子座と、その同族体であるアカバナムシヨケギクの TcoGLIP (*T. coccineum* GDSL lipase) 遺伝子座の近傍領域が異なっていた。この違いは、Tci(o)GLIP をコードする遺伝子の発現調節、ひいてはピレトリン類産生量の違いを解明する手掛かりになることが考えられる。遺伝子の属間比較解析では、除虫菊と同様に、アカバナムシヨケギクゲノムにおいても RIP が重複していたが、毒性の弱い I 型 RIP はアカバナムシヨケギクで、毒性の強い II 型 RIP は除虫菊で多いことが示された。一方で、アカバナムシヨケギクのエンドキチナーゼの数は他の植物種と同程度であり、2 章で示された除虫菊におけるエンドキチナーゼ数の少なさをより強調する結果となった。これらの違いは、除虫菊とアカバナムシヨケギクの防御機構がそれぞれの原産地に合わせて進化し、多様化していることを示

唆している。また、除虫菊で重複していたヒスチジンキナーゼは、アカバナムシヨケギクにおいて同様の重複傾向がみられなかった。除虫菊とアカバナムシヨケギクにおけるヒスチジンキナーゼの数とピレトリン類産生量の多寡に相関性があり、2章で示唆された VOC 誘導性ヒスチジンキナーゼとピレトリン類の VOCs 依存的調節機構との関連性を裏付ける結果となった。以上、除虫菊とアカバナムシヨケギクのゲノム比較から防御機構をはじめとする遺伝的特徴の違いが明らかになり、さらにこれら 2 種の植物のピレトリン類産生量の違いを解明する上での手掛かりを得ることができた。

TciGLIP はピレトリン I の生合成において最終のエステル化反応を担う重要な酵素である。TciGLIP が属する GDSL (Gly-Asp-Ser-Leu モチーフ) エステラーゼ/リパーゼファミリータンパク質 (GELPs) は一般的に加水分解つまりエステラーゼ活性を有しているが、数種の GELPs は TciGLIP のようにトランスフェラーゼ活性を示す。第 4 章においては、除虫菊がピレトリン類生合成能を獲得するにあたり、TciGLIP がトランスフェラーゼ活性を獲得した鍵となる部位を *in silico* 解析を用いて特定した。TciGLIP および TcoGLIP を含む 4 種のトランスフェラーゼ活性が報告されている GELPs (tr-GELPs) およびエステラーゼ活性が報告されている 6 種の GELPs、さらに BLASTP で得られた 274 種の推定 GELPs のアミノ酸配列を用いてアラインメントを作成し、two entropy 解析を用いて tr-GELPs に特徴的なアミノ酸残基を調べたところ、トランスフェラーゼ活性に重要な候補残基として [N/R]208 および D484 が検出された。分子系統解析を行ったところ、tr-GELP は tr-GELPs 以外の他の GELPs と同じクラスターに属し、そこに属するほとんどの GELPs は [N/R]208 または D484 の少なくとも一つを有していた。この結果から、tr-GELP はそれぞれの植物種の GELP から遺伝子変異により誕生したことが明らかになった。また、タンパク質構造予測と基質結合シミュレーションを用いて、TciGLIP の活性中心に基質が接近できる合理的モデルの数を調べたところ、これら 2 つの候補残基の Ala 変異体の合理的モデルの数は、天然型の TciGLIP の合理的モデルの数と比較して優位に少なく、この結果

は、これら 2 つの残基が TciGLIP のトランスフェラーゼ活性に重要な残基であることを示唆していた。TciGLIP のこれら 2 つの候補残基の変異によるトランスフェラーゼ活性への影響の実験的検証は、現在進行中である。

本研究では、除虫菊および近縁種のアカバナムシヨケギクのドラフトゲノム解読と比較ゲノム解析を行ったことで、ピレトリン類の生合成をはじめとする防御機構の違いを明らかにした。また、ピレトリン類の生合成酵素の一つである TciGLIP について、その活性に重要な部位を *in silico* 解析で特定した。これらの研究成果は、ピレトリン類の生合成経路とその制御機構の完全解明への手がかりとなる。今後のピレトリン類の生合成に関する研究にあたっては、候補となる酵素の実験的活性検証による未解明のピレトリン類生合成酵素の同定や、除虫菊の部位・時期・条件による種々 RNA-seq データを取得し、関連遺伝子群の発現量とピレトリン類生合成量との相関を網羅的に解析することで、生合成機構の全容を解明していく方針である。さらに、本研究で得られたゲノムや生合成酵素の生物情報学的な基盤は、遺伝子改変個体の創出やそれらの栽培の最適化にも大きく貢献することが期待される。たとえば、除虫菊ゲノム情報を活用することで、ピレトリン類の収量に関与した遺伝子マーカーのような分子育種ツールの開発や遺伝子改変技術の確立といった応用研究へ展開していく予定である。そして、最終的には、ピレトリン類高生産除虫菊株の創出や、合成生物学的手法などを駆使した微生物や別種植物を宿主とする異種発現系の確立を行い、ピレトリン類の大量生産体制を構築することで、世界的な昆虫媒介性感染症の被害の根絶に貢献することを目指していきたい。

謝辞

本研究に際し、終始御懇篤なる御指導と御鞭撻を賜りました恩師サントリー生命科学財団生物有機科学研究所研究部長兼主幹研究員および神戸大学大学院工学研究科連携講座教授の佐竹炎先生に謹んで感謝の意を表します。

本研究に関して御指導と御討議いただき、貴重な意見を賜りました神戸大学大学院工学研究科教授荻野千秋先生と同教授山地秀樹先生に深く感謝いたします。

本研究に際し、御指導、御討議いただき、数多く御協力いただいたサントリー生命科学財団生物有機科学研究所研究員白石慧博士、ならびに小山知嗣博士をはじめとする、同研究所の皆様へ深く感謝いたします。

本研究に関して、数多く御協力いただいた大日本除蟲菊株式会社取締役兼同社中央研究所研究所長中山幸治博士をはじめとする、同研究所の皆様へ深く感謝いたします。

最後に、本研究に関わる研究費および本研究を行う機会を御提供くださった大日本除蟲菊株式会社に深く感謝いたします。

神戸大学博士論文

「除虫菊類のゲノムや有用物質生合成酵素の生物情報基盤構築に関する研究」全 132 頁

提出日 2023年1月17日

本博士論文が神戸大学機関リポジトリ **Kernel** にて掲載される場合、掲載登録日（公開日）はリポジトリの該当ページ上に掲載されます。

© 山城敬範

本論文の内容の一部あるいは全部を無断で複製・転載・翻訳することを禁じます。