

令和 6 年 6 月 16 日現在

機関番号：82401

研究種目：基盤研究(C) (一般)

研究期間：2020～2023

課題番号：20K06761

研究課題名(和文) The role of non-allelic gene conversion in the evolution of Transposable Elements

研究課題名(英文) The role of non-allelic gene conversion in the evolution of Transposable Elements

研究代表者

FAWCETT JEFFREY (FAWCETT, JEFFREY)

国立研究開発法人理化学研究所・数理創造プログラム・上級研究員

研究者番号：50727394

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：トランスポゾン(Transposable Elements)は生物の進化の主要な原動力と考えられている。我々は以前、トランスポゾンの進化における役割を理解するには、トランスポゾンそのものの進化、特にトランスポゾンの配列間で生じるnon-allelic gene conversionとトランスポゾン間での競争による自然選択を理解する必要があると提唱した。今回の主要な成果は、ソバ属植物のセントロメアの進化におけるトランスポゾンの役割の解明である。我々はある特定のLTR型レトロトランスポゾンの種特異的な増加が、新規セントロメアの獲得と関係していることを明らかにし、種文化と関係している可能性を提唱した。

研究成果の学術的意義や社会的意義

How Transposable Elements contribute to evolution is a central question in evolutionary biology. Here, we have provided a novel example showing that the rapid amplification and evolution of TEs is associated with the turnover of centromeres and most probably speciation.

研究成果の概要(英文)：Transposable Elements are thought to be key drivers of evolution and have most likely contributed to various evolutionary processes such as the rewiring of regulatory networks. We previously argued that understanding the evolution of TEs themselves, including the role of non-allelic gene conversion and selection amongst TEs is crucial for unravelling the role of TEs in evolution. One of the major achievement has been our results on the contribution of TEs to centromere evolution, a process in which rapid TE evolution via non-allelic gene conversion is thought to be involved, in buckwheat species. We found that the species specific activity of a particular LTR retrotransposon family of the CRM clade is associated with neocentromeres, suggesting its role in speciation via centromere turnover.

研究分野：Genome Evolution

キーワード：Transposable Elements

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様式 C - 19、F - 19 - 1 (共通)

1 . 研究開始当初の背景

Transposable Elements (TEs) are thought to be key drivers of evolution and have most likely contributed to various evolutionary processes such as the rewiring of regulatory networks. We previously argued that understanding the evolution of TEs themselves, including the role of non-allelic gene conversion and selection amongst TEs is crucial for unravelling the role of TEs in evolution. The drastic increase in whole-genome sequence data of various species in the recent years provides the opportunity to understand the evolution of TEs and their contribution to genome evolution in further detail that was not possible before.

2 . 研究の目的

Here, we aimed to understand the role of TEs in eukaryotic evolution by elucidating the sequence evolution of TE themselves. This is because TEs are highly diverse and undergo an evolutionary process similar to that of other organisms and the sequence diversity of TEs reflect these evolutionary processes. Thus, we aimed to understand the diversification process of TEs, including the role of selection within TEs and non-allelic gene conversion between TE sequences, and subsequently how the diversification processes is associated with various evolutionary processes of the species harboring those TEs.

3 . 研究の方法

We identified all LTR retrotransposons in the *Fagopyrum* genomes by a combined approach utilizing LTRharvest and RepeatMasker. We then performed sequence-based clustering with CD-hit using the reverse transcriptase sequences. We also performed phylogenetic analyses with the reverse transcriptase sequences of the chromovirus clade identified in the three *Fagopyrum* species and those identified in other plant species reported by a previous study.

4 . 研究成果

Our first candidate to use as a model to study the evolution of TE was the expansion of the B2 SINE family in mouse. This was because it was previously demonstrated that this family expanded in rodents and that many CTCF binding sites in mouse are derived from this family. We therefore aimed to investigate the relationship between the expansion of B2 SINE and the evolution of CTCF binding sites and obtained all the positions, sequences, and estimated insertion ages of the B2 SINE elements in mouse. Nevertheless, we found that the relationship between B2 SINE, CTCF binding motifs, and CTCF-mediated regulation

appears to be far more complex than we initially anticipated (e.g. Kaaij et al. 2019). We have therefore put this analysis on hold for the time being and have instead been focusing on another interesting topic that we recently came across through another collaboration involving TE expansion and centromere turnover and thus potentially speciation in *Fagopyrum* species.

Centromere evolution is one aspect in which TE expansion possibly involving selection amongst TEs and non-allelic gene conversion between TE sequences is likely to play an important role. We recently sequenced and the genomes of *F. esculentum* (common buckwheat) and its close wild relative *F. homotropicum* and constructed chromosome-scale assemblies as part of a large collaboration between different researchers involved in various aspects of this species. We then annotated TEs in these two species and another sister group species *F. tataricum* whose genome had already been sequenced. We found that the large genome size difference between *F. tataricum* (~490 Mb) and the other two species *F. esculentum* (~1270 Mb) and *F. homotropicum* (~1250 Mb) can be largely ascribed to the differences in Gypsy-type LTR retrotransposons, which comprised of 826 Mb, 798 Mb, and 148 Mb of *F. esculentum*, *F. homotropicum*, and *F. tataricum*, respectively.

Gypsy-type LTR retrotransposons can be divided into two major clades, chromoviruses and non-chromodomain retroviruses. Sequence-based clustering analyses and evolutionary analyses showed that the lineage-specific expansion of a few subfamilies belonging to the Athila family of the non-chromodomain retrovirus in *F. esculentum* and *F. homotropicum* is largely responsible for the genome size difference between *F. tataricum* and the other two species. Furthermore, we found that some lineages of the chromoviruses cluster at specific chromosomal regions of *F. esculentum* and *F. homotropicum* but not *F. tataricum* (Figs 1&2). We confirmed that these regions correspond to primary constrictions by FISH analysis with probes corresponding to these LTR retrotransposon sequences, indicating that these lineages are associated with centromeres in these two species.

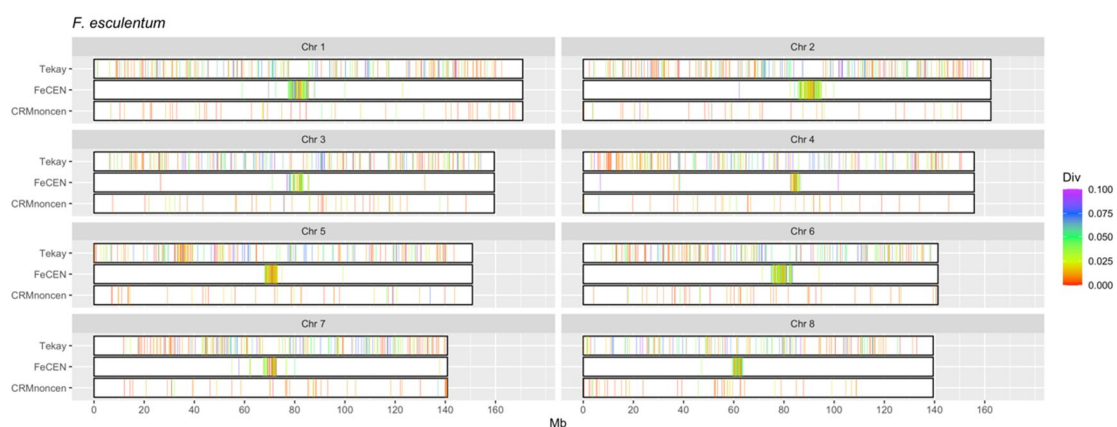


Fig 1. Distribution of the Tekay family and two different CRM subfamilies across each chromosome of *F. esculentum*.

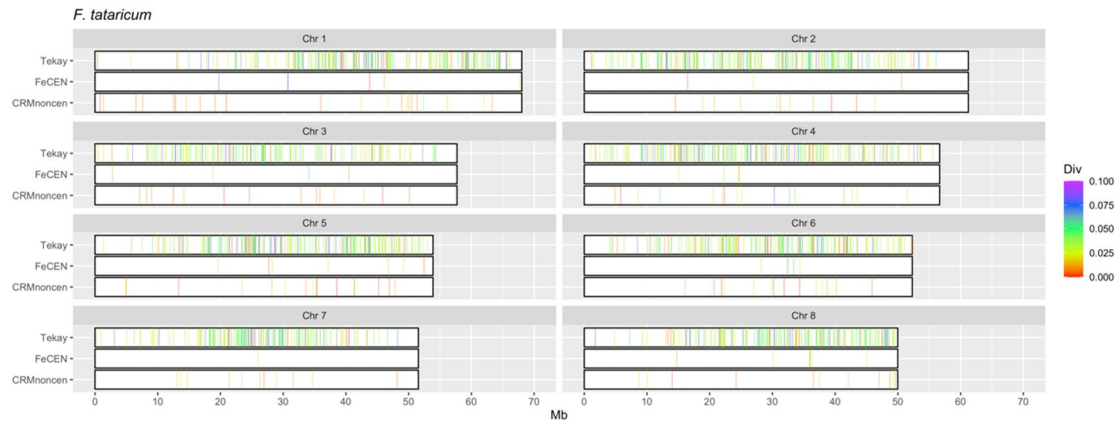


Fig 2. Distribution of the Tekay family and two different CRM subfamilies across each chromosome of *F. tataricum*.

To further investigate the evolution of these centromere-associated TEs, we performed phylogenetic analyses of the chromoviruses. We found that the CRM family in *Fagopyrum* could be divided into those associated with centromeres in *F. esculentum* and *F. homotropicum* (FeCEN) and those that are not associated with centromeres (CRMnoncen) (Fig 3). It has been demonstrated that some but not all lineages of the CRM family are associated with centromeres in various plant species (Neumann et al. 2011). FeCEN had amplified in the lineage leading to *F. esculentum* and *F. homotropicum* since its divergence from the lineage of *F. tataricum*, although its amplification rate has decreased recently in contrast to CRMnoncen (Fig 4). We could not find any TE lineages associated with centromeric regions in *F. tataricum*, including the few FeCEN elements.

We then examined whether the putative centromeric regions are collinear between each species. The centromeric regions of all 8 chromosomes were collinear between *F. esculentum* and *F. homotropicum*. Although the chromosomal location of *F. tataricum* have not been determined, we can use the distribution of the GC content and TE density across each chromosome as a proxy. Indeed, we found that the chromosomal locations of the centromeric regions are most probably different between the two species for at least three of the eight chromosomes. Frequent repositioning of centromeres, i.e. de novo centromere formation, has been reported in various animals, and also recently in other plants such as maize, *Oryza*, Arabidae, and cucurbit species (Han et al. 2009, Schneider et al. 2016, Liao et al. 2018, Mandakova et al. 2020). Our results suggest that similar repositioning has occurred for at least some of the chromosomes in *F. esculentum* and/or *F. tataricum*. Considering the

proposed role of centromere, especially of new repositioned centromeres, in generating reproductive barriers (Lu et al. PNAS2019), the evolution of centromeres may have helped reinforce reproductive isolation between *F. tataricum* and the common ancestor of *F. esculentum* and *F. homotropicum*.

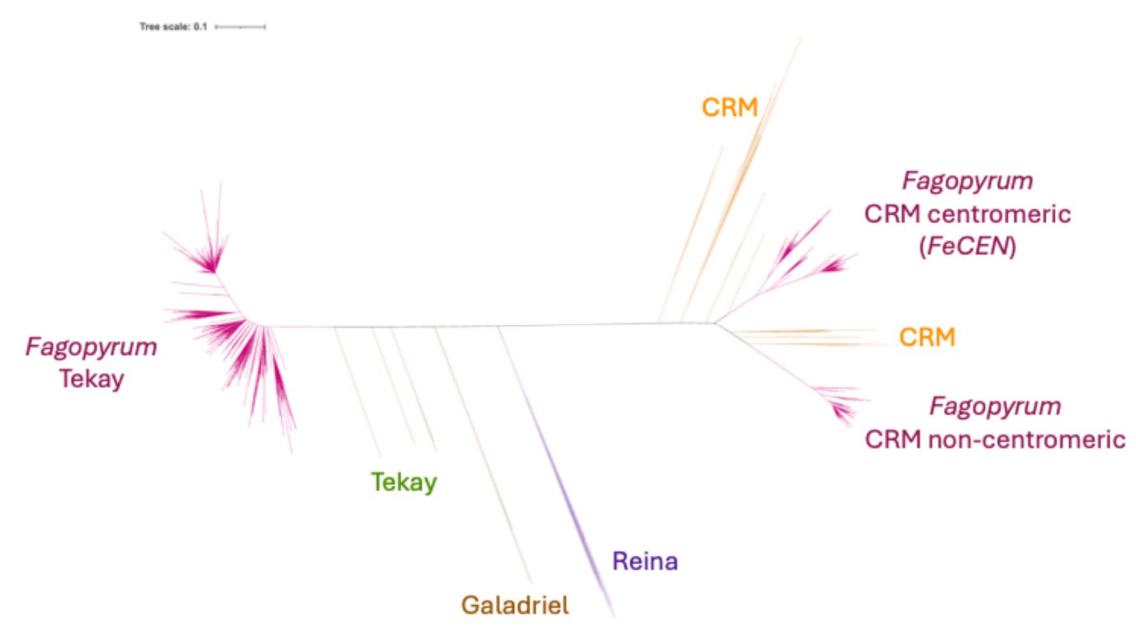


Fig 3. Phylogenetic analysis of the amino acid sequences of the reverse transcriptase domain of LTR retrotransposons of the chromovirus clade identified in *F. esculentum*, *F. tataricum*, and other plant species from Neumann et al. 2011.

Taken together, our results suggest that the lineage-specific expansion of a specific CRM family is associated with multiple newly-formed centromeres in *Fagopyrum* species, demonstrating the role of TE expansion in centromere evolution and possibly speciation. Our collaborators are currently performing experiments to determine the centromere positions in *F. tataricum*, which should help confirm our findings. Our results also suggest that FeCEN acquired mutations that enabled their expansion and centromeric association in *F. esculentum* and *F. homotropicum*. We are currently investigating the roles of selection and non-allelic gene conversion in the evolution of FeCEN.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 1件/うちオープンアクセス 0件）

1. 著者名 Fawcett JA, Takeshima R, Kikuchi S, Yazaki E, Katsube-Tanaka T, Dong Y, Li M, Hunt HV, Jones MK, Lister DL, Ohsako T, Ogiso-Tanaka E, Fujii K, Hara T, Matsui K, Mizuno N, Nishimura K, Nakazaki T, Saito H, Takeuchi N, Ueno M, Matsumoto D, Norizuki M, Shirasawa K, Li C, Hirakawa H, Ota T, Yasui Y.	4. 巻 9
2. 論文標題 Genome sequencing reveals the genetic architecture of heterostyly and domestication history of common buckwheat	5. 発行年 2023年
3. 雑誌名 Nature Plants	6. 最初と最後の頁 1236 ~ 1251
掲載論文のDOI（デジタルオブジェクト識別子） 10.1038/s41477-023-01474-1	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

〔学会発表〕 計2件（うち招待講演 2件/うち国際学会 0件）

1. 発表者名 Jeffrey Fawcett, Hideki Innan
2. 発表標題 The role of gene conversion between Transposable Elements in rewiring regulatory systems
3. 学会等名 Annual Meeting of the Molecular Biology Society of Japan（招待講演）
4. 発表年 2021年

1. 発表者名 Jeffrey Fawcett
2. 発表標題 The origin and evolution of cultivated common buckwheat
3. 学会等名 Annual Meeting of the Society of Evolutionary Studies Japan（招待講演）
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	印南 秀樹 (Innan Hideki) (90444140)	総合研究大学院大学・先導科学研究科・教授 (12702)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関