

学 位 論 文 題 名

Genomic study on integrations of endogenous rice tungro bacilliform virus-like sequences and episomal DNAs in the rice genomes

(イネゲノムに内在するウイルス様配列とエピソードDNAの挿入に関するゲノミクス研究)

学位論文内容の要旨

In the rice genomes, pararetrovirus-like sequences that are similar to rice tungro bacilliform virus (RTBV), called endogenous rice tungro bacilliform virus-like sequences (ERTBV), have been found, and none of the ERTBVs are functionally intact. A remarkable feature of the insertion sites of most ERTBVs is that they are flanked by AT di-nucleotide repeats (ATRs). Although AT-rich regions in chromosomal DNA have been described as preferential target sites for different external and internal DNAs, our current understanding is insufficient to fully describe the role of ATRs in terms of DNA integration. In this study, the integration event of ERTBV and episomal DNA segments into ATRs was investigated through comparative analyses between japonica and indica genomes. In addition, the origin and distribution of ERTBV were also analyzed in the rice-closely-related species.

(1) Integration of ERTBV into AT di-nucleotides repeats of rice genomes

The rice genomes contain a number of segments of ERTBV, many of which were present between ATRs. The possible processes involved in ERTBV integration into the rice genome were postulated as; 1) ERTBV integrated into the ATRs in the genome, and 2) ERTBV accompanied ATRs prior to integration into the genome. This study sought the process for the ERTBV integration into the rice genome. Comparison of genomic sequences between two closely related rice subspecies, japonica and indica, allowed me to verify the preferential insertion of ERTBV into ATRs. I conclude that ATRs were present before the ERTBV insertions occurred.

(2) AT di-nucleotide repeats as damping sites of episomal DNAs

I confirmed that ERTBVs are preferentially present between ATRs, which allowed me to speculate that AT-rich regions might be widely employed as a site to integrate external and internal

DNAs into chromosomes. Therefore, this study investigated the sequences present between the same SSRs by the comparative analysis between japonica and indica that was also used in the case of ERTBVs. In addition to ERTBVs, the comparative analyses showed that ATrs occasionally incorporate repeat sequences including transposable elements, and a wide range of other sequences. Besides the known genomic sequences, the insertion sequences also represented DNAs of unclear origins together with ERTBVs, suggesting that ATrs have integrated episomal DNAs that would have been suspended in the nucleus. Such insertion DNAs might be trapped by ATrs in the genome in a host-dependent manner. Conversely, other simple mono- and di-nucleotide sequence repeats (SSR) were less frequently involved in insertion events relative to ATrs. Therefore, ATrs could be regarded as hot spots of double-strand breaks that induce non-homologous end joining. The insertions within ATrs occasionally generated new gene-related sequences or involved structural modifications of existing genes. Likewise, in a comparison between *Arabidopsis thaliana* and *Arabidopsis lyrata*, the insertions preferred ATrs to other SSRs. ATrs in plant genomes could be considered as genomic dumping sites that have trapped various DNA molecules and may have exerted a powerful evolutionary force.

(3) The original distribution of ERTBV in the rice genomes

ERTBV originated from ancient viral infections that have been fixed in the germline of the rice genomes. A number of the ERTBV segments were also detected in the other *O. sativa* and *O. rufipogon* strains based on Southern blotting experiments. The evolution and origin of ERTBV in *Oryza* species remain to be clarified. Here, I explored original copies of ERTBV among the characterized copies in japonica genome. Among 162 ERTBV segments identified in the japonica and indica database, 22 are shared between the two genomes. With PCR analysis, nine of the 22 ERTBVs were detected in *japonica*, *indica* and *O. rufipogon*, and these were thought to be ancient copies that existed before japonica-indica differentiation. To further understand the origin of ERTBV, the phylogenetic tree was constructed for each of the seven functional regions in ERTBV sequence. One of the most clear outcome resulted from RT/RH gene tree that was distinctly divided into four major clusters. The nine old segments detected by PCR analysis were distributed in all the four clusters, implying the different phylogenies of the old segments. Therefore, these results strongly indicated that the multiple origins of ERTBV cognates were inferred. The different origins of ERTBV in the *Oryza* genomes suggest that the distribution of ERTBVs was geographically different.

学位論文審査の要旨

主査	教授	貴島 祐治
副査	特任教授	三上 哲夫
副査	准教授	小柳 香奈子 (情報科学研究科)
副査	講師	高牟禮 逸朗

学位論文題名

Genomic study on integrations of endogenous rice tungro bacilliform virus-like sequences and episomal DNAs in the rice genomes

(イネゲノムに内在するウイルス様配列とエピソードDNAの挿入に関するゲノミクス研究)

本論文は図 14、表 10、附図 3、附表 6 を含み、5 章からなる総頁数 110 の論文であり、別に参考論文 1 編が添えられている。

イネには二本鎖 DNA ウイルス (パラレトロウイルス) である Rice tungro bacilliform virus (RTBV) が感染し、ツングロ病を起こすことが知られている。イネゲノムにはこの RTBV と類似の配列 Endogenous RTBV-like sequence (ERTBV) が多数存在することが以前に報告されている。ERTBV の挿入部位を調べると 8 割は AT の連続配列が両端に配置されており、ERTBV がどのようにゲノムに入ったのかを知る手掛かりを与えている。また、AT 連続配列には ERTBV に限らず他の DNA 配列も挿入している可能性がある。本研究では主に ERTBV のイネゲノムへの挿入様式をジャポニカとインディカのゲノムデータを解析し、明らかにした (1)。その過程で ERTBV 以外の配列についても AT 連続配列に特異的な挿入が生じていることを見出し、AT 連続配列の特性と挿入配列の構造を詳細に解析した (2)。さらに、ERTBV の挿入とイネの種分化は密接に関連している。この事実を基に、各 ERTBV 断片の挿入時期の推定と ERTBV の分布について調べた (3)。以上の結果をまとめ、ERTBV や核に浮遊するエピソード配列のイネゲノムへの挿入様式、ERTBV と RTBV の構造的な相違とイネゲノムへの挿入の有無、およびイネゲノムの進化と ERTBV 挿入の関係について考察した。

(1) ERTBV の AT 連続配列への挿入

Kunii ら (2004) は AT 連続配列が ERTBV の両端に高頻度に出現することを見出した。この事実から、私は 2 つの仮説を立てて、ERTBV の挿入様式について解析した。1 つは ERTBV が AT 連続配列を標的に挿入した可能性。もう 1 つは ERTBV が AT 連続配列を結合して、ゲノムに入った可能性。日本晴 (RAP build 5) と 93-11 (BGI) のデータベースからそれぞれ 88 個および 74 個の ERTBV が同定された。この内、ジャポニカでは 84%、インディカでは 77% の ERTBV の両端に AT 連続配列が検出された。2 つのゲノムどちらか一方で ERTBV の挿入を持つ 32 サイトでは、ERTBV のないゲノム側のいわゆる empty donor site で AT 連続配列が存在していた。また、ERTBV を持つサイトのパラログス配列においても AT 連続配列が見出された。これらの結果から、AT 連続配列は ERTBV がゲノムに挿入する前に存在していたこ

とが確認された。

(2) エピゾーム DNA の集積場としての AT 連続配列

全ての ERTBV は再編成され機能的な役割は持たず、単にウイルスの断片として存在しているため、積極的に AT 連続配列を標的に挿入しているとは考えづらい。ゲノム全体の反復配列を調べたところ、ERTBV だけではなく、他の配列も取り込まれることが判明した。挿入配列は既知のゲノム配列に加えて、起源が不明な DNA も含まれた。一方、AT 以外の連続配列、例えば AA/TT、GG/CC、AC/GT、AG/CT あるいは GC 連続配列では AT 連続配列のように ERTBV や反復配列などの挿入は認められなかった。従って、AT 連続配列は、広く細胞に浮遊している ERTBV などのエピゾーム様配列を取り込む性質を持っている可能性は高い。AT 連続配列はマトリックスアタッチメントサイト (MAR) にあたり、ゲノムの中でも、二本鎖切断されやすい配列であることが推定できる。AT 連続配列は切断後、修復過程で細胞核に浮遊する DNA 断片を取り込んできたと考えられる。AT 連続配列内の挿入はゲノムに存在しない未知の DNA 配列も含み、DNA の水平伝搬との関連も示唆できる。中には転写されている遺伝子配列を生じたり、既存の遺伝子の構造を変更してイネ独自の遺伝子構造を持つ配列も AT 連続配列内に見出された。*Arabidopsis thaliana* と *A. lyrata* の間の比較でも、イネと同様な結果を得た。従って、植物ゲノムの AT 連続配列は様々な DNA 断片を集積したダンピングサイト (ゴミ集積場) であると同時に、強力に進化を推進する機能を持つと考えられた。

(3) ERTBV の起源と分布

ツングロ病の抵抗性システムを探索するためには、その起源地での遺伝資源を調べることが有効と思われる。本章では日本晴と 93-11 のデータベースから得られた 164 個の ERTBV を基にその起源を調査した。日本晴と 93-11 で共通なサイトに挿入している 22 個の ERTBV に着目した。これら 22 個の ERTBV について PCR で起源の異なるジャポニカ、インディカ、ルフィポーゴン 10 系統のゲノムを調べ、9 つの ERTBV が全ての系統に存在していた。さらに、ウイルス配列を 7 つの部位に分け各断片ごとに系統樹を作成した。最も明瞭な系統樹は RT/RH 遺伝子で得られ、4 つのクラスターに分けられた。PCR 実験により検出された 9 つの断片は 4 つのクラスターに全て分布していた。このことから、ERTBV は 4 つの異なる起源を持つウイルスがイネに感染し、ゲノムに取り込まれ日本晴ゲノムを構成するに至ったと推定できる。同時に ERTBV の起源地の分布も地理的に異なっていたことを示唆している。

以上の成果は、イネに内在するウイルス様配列 (ERTBV) を手掛かりにして、植物ゲノムが外来 DNA をゲノムに取り込んできたことをはじめて明らかにした点で、学術上高く評価できる。また、今後の展望としてイネツングロ病抵抗性機構を解明し、イネ品種へ付与することについても基盤となる知見を与えた。審査員一同は、劉瑞芳氏が博士 (農学) の学位を受けるに十分な資格を有するものと認めた。