



藻類ゲノムに伝播したウイルス様遺伝子とその進化的役割

本郷 悠貴^{1*}・外丸 裕司²¹ 国立研究開発法人 水産研究・教育機構 水産資源研究所 (〒 236-8648 神奈川県横浜市金沢区福浦 2-12-4)² 国立研究開発法人 水産研究・教育機構 水産技術研究所 (〒 739-0452 広島県廿日市市丸石 2-17-5)Yuki Hongo^{1*} and Yuji Tomaru²: Virus-derived genes in algal genomes and their evolutionary roles. Jpn. J. Phycol. (Sôru) 69: 160–164, November 10, 2021

The viral genome sequences are often integrated into their host genomes during their infection processes. Moniruzzaman *et al.* (2020, Nature 588: 141–145) investigated 24 green algal genomes and detected giant endogenous viral elements (GEVEs) that had been derived from nucleocytoplasmic large DNA viruses. Comparison among the GEVEs revealed that the sequence segments of the GEVEs have been rearranged in the algal genomes due to their duplication and gene loss. Moreover, the genes in the GEVEs possibly affect the host metabolic state, and it was suggested that the GEVEs are one of the driving forces of the host evolution. In addition to reviewing Moniruzzaman *et al.* (2020), we also introduce our latest study in which the viral elements in the diatom genome, *Chaetoceros tenuissimus*, were discovered.

Key Index Words: endogenous viral elements, genome evolution, viral genome integration, virus

¹Fisheries Resources Institute, Japan Fisheries Research and Education Agency, 2-12-4 Fukuura, Kanazawa, Yokohama, Kanagawa 236-8648, Japan

²Fisheries Technology Institute, Japan Fisheries Research and Education Agency, 2-17-5 Maruishi, Hatsukaichi, Hiroshima 739-0452, Japan

* Author for correspondence: hongoy@affrc.go.jp

真核生物のゲノムにはレトロウイルスに由来する配列が多く存在し、ヒトゲノムには約8%の領域がこれに該当する (Griffiths 2001)。次世代シーケンサーの普及によって様々な種のゲノムが解読されるにつれ、そのゲノム中にレトロウイルスだけでなく DNA ウイルスや RNA ウイルスに由来する配列 (遺伝子やその断片) が多く挿入されていることが近年分かってきた。こうしたウイルス由来の配列は“内在性ウイルス様配列 (endogenous viral elements: EVEs)” と呼ばれ (Katzourakis & Gifford 2010)、宿主ゲノムの進化に大きく貢献している (Kazazian 2004, Ribet *et al.* 2008)。ウイルス配列の挿入は、ウイルス感染時にウイルスゲノムが宿主細胞に存在するので、ウイルスと宿主間で核酸配列の移動 (水平伝播) が容易であることが想像できる。このウイルスと宿主間の水平伝播について藻類に着目すれば、実際に、*Ectocarpus siliculosus* (Dillwyn) Lyngbye (シオミドロ) のゲノムに、感染性の二本鎖 (ds) DNA ウイルスである EsV-1 の遺伝子や断片が EVEs として高い相同性を保持して挿入されている (Delaroque & Boland 2008)。また逆に、*Emiliania huxleyi* (Lohmann) W.W. Hay & H.P. Mohler (円石藻) に感染する EhV (dsDNA ウイルス) は、宿主が持つスフィンゴ脂質の合成に関する 8 つの遺伝子のうち 7 つを保持しており、系統解析の結果はこれらの遺伝子が宿主からウイルスへ水平伝播してきたことを支持している (Monier *et al.* 2009)。上記で

挙げた dsDNA ウイルスは、核細胞質性大型 DNA ウイルス (Nucleocytoplasmic large DNA viruses: NCLDV) と呼ばれるゲノムサイズと粒径が比較的大きいウイルスに分類され、その中の Phycodnaviridae 科に属している。現在、NCLDV は国際ウイルス分類委員会 (ICTV) の分類基準で 7 つの科 (Phycodnaviridae, Mimiviridae, Ascoviridae, Iridoviridae, Marseilleviridae, Asfarviridae, Poxviridae) にまとめられている。この様に NCLDV は、藻類に感染することによってウイルスと宿主双方のゲノム情報を伝播しあっていることが垣間見られるが、藻類の多くにこうした現象が存在するのかどうかについては明らかになっていなかった。こうした問題に対して Moniruzzaman らのグループは、以前から NCLDV と関係が知られている緑藻植物門に注目して数多くの論文を出版している。その中でも、公開されているゲノム情報とバイオインフォマティクスの技術を用いて NCLDV 由来の giant EVEs (GEVEs) を発見した論文 (Moniruzzaman *et al.* 2020) を本報では紹介したい。現在、我々の研究グループでは海産珪藻 *Chaetoceros tenuissimus* Meunier とその感染性ウイルスの研究を実施しているが、本紹介論文と類似する内容があり興味深い。そのため、本稿の最後に我々の研究成果 (Hongo *et al.* 2021) も併せて紹介したい。

NCLDV 由来 GEVE の探索

Moniruzzaman *et al.* (2020) は、これまでに公開されている緑藻門内の 6 科・65 種のゲノムを、バクテリアやアーキアの遺伝子予測するツール、Prodigal (Hyatt *et al.* 2010) を用いて再び遺伝子予測を行い、open reading frames (ORFs) を取得した。得られた ORFs はウイルス由来のデータを除く Pfam データベースと Viral Orthologous Groups (VOG) データベース (<https://vogdb.org>) に対して、隠れマルコフモデルを実装した HMMER3 の hmmsearch ツール (Hyatt *et al.* 2010) を用いて細胞生物とウイルス由来の配列を検索し、ViralRecall (Aylward & Moniruzzaman 2021) ツールで上記の結果を統合してウイルス由来の配列をスコア化した。同定された配列 (コンティグ) における GEVE の存在有無を確かめるため、Moniruzzaman *et al.* (2020) は以下に示す様々な解析アプローチで検証した。

- (1) ViralRecall ツールで同定された GEVE を含むコンティグに NCLDV を特徴付けるコア遺伝子が存在するか、NCLDV のコア遺伝子だけを集めたデータベースを作成し、先程と同様に、hmmsearch ツールを使って検索した。ここで述べたコア遺伝子というのは、Iyer *et al.* (2006) によって提唱された NCLDV を構成する全ての科に共有される 9 つの遺伝子、そして少なくとも 3 つの科で共有される 22 の遺伝子を指す。また、コア遺伝子は、NCLDV が細胞生物と共通して保有する遺伝子 (複製や転写、DNA 修復など) で、NCLDV の単系統性が支持されているもの、また、他のウイルスには存在しない遺伝子で構成されている。
 - (2) 分子系統解析によって、NCLDV のコア遺伝子の系統学的位置を確認した。
 - (3) コンティグ配列のテトラヌクレオチド頻度を算出し、その相関距離によるクラスタリングでコンティグを分類した。さらに、コア遺伝子が確認されたコンティグとクラスターを組むコンティグ群を、ウイルス由来の配列を含むコンティグ候補として選択した。
 - (4) これまでに公開されている 127 の NCLDV ゲノム情報と Prodigal で予測されたタンパク質情報を Proteinortho (Lechner *et al.* 2011) ツールでオルソロググループを同定し、そのグループ内のタンパク質情報をデータベース化した。そのデータベースを参照して、ViralRecall ツールで同定されたコンティグを hmmsearch ツールで検索した。
 - (5) ViralRecall ツールで同定されたコンティグ内の遺伝子を NCBI RefSeq を対象に相同検索し、相同性を評価した。
 - (6) 真核生物に水平伝播した原核生物の遺伝子は、イントロンの密度が低いことが知られている (Marcet-Houben & Gabaldón 2010, Rossoni *et al.* 2019)。イントロンの密度パターンをホストの遺伝子座と GEVE で比較し、GEVE で密度が低くなっていることを確認した。
- 以上の検証項目を実施し、緑藻ゲノム内の GEVE を同定した。

緑藻ゲノムの GEVEs

65 種の緑藻のうち 24 種で NCLDV に由来する遺伝的痕跡が確認され、うち 12 種から 78–1,925 kbp のサイズに分布する 18 の GEVEs が発見された。また、それら緑藻 12 種のうち、6 種は 2 つの異なる GEVEs を保有していることが確認された (図 1)。GEVEs には 76–1,782 の候補遺伝子が含まれている。特に *Tetrabaena socialis* (Dujardin) Nozaki & Mot. Ito には 2 つの GEVEs に合計 2,846 遺伝子が含まれていた。驚くことに、これは宿主が保有する遺伝子全体の 10% を占めるという。その他の種の GEVEs に含まれているコア遺伝子の分子系統解析から、4 つの GEVEs が *Phycodnaviridae* に、10 の GEVEs が *Mimiviridae* 科に由来することが推定されている (図 1)。残り 4 つの GEVEs はコア遺伝子が欠落していたため科レベルでの分類はされていないが、GEVEs に含まれている他の遺伝子と NCLDV が保有する遺伝子の比較によって、NCLDV 由来の GEVEs であることが推定されている。また、NCLDV に由来する遺伝的痕跡が確認された 24 種の緑藻のうち上記以外の 12 種については、NCLDV のコア遺伝子が数個同定されているが、GEVEs の確認には至っていない。これらは、NCLDV ゲノムの挿入から現在に至るまでに断片化が進んだ痕跡と考えられている。

T. socialis, *Yamagishiella* Nozaki, *Chlamydomonas eustigma* H. Ettl, *Coccomyxa* Schmidle, *Tetrademus obliquus* (Turpin) M.J. Wynne の GEVEs 中には複数のセグメントの重複が見られ、挿入後に配列の再編成が起こったことが示されている (図 1)。特に、自由生活型の NCLDV はコア遺伝子が一般的に 1 コピーであるのに対し、GEVEs は複数のコピーを持つこと、さらに GEVEs の重複領域を NCLDV ゲノムと比較すると有意に重複していることからゲノム再編成が裏付けられている。また、GEVEs 中に含まれていた 10 個の NCLDV コア遺伝子のうち 1 遺伝子 (D5-helicase/primase) は全 GEVEs で保持されているが、残りの 9 遺伝子については複数もしくは全て喪失している場合もあった (図 1)。これは、GEVE 中の遺伝子が時間の経過とともに喪失したことを示していると考えられている。以上のことをまとめると、GEVE はホストのゲノム中で様々な程度の重複と再編成、および遺伝子の喪失が起こっていたことを示唆している。NCLDV ゲノムには転移因子であるトランスポゾンが豊富であり (Filée 2015, Sun *et al.* 2015)、GEVE 中にも多く同定されるため、これらの転移因子が GEVE 内もしくは GEVE と宿主ゲノム間の相同組替えに寄与し、GEVE の可塑性を高めている可能性があること Moniruzzaman *et al.* (2020) は考えている。

NCLDV 内在化の進化歴

緑藻ゲノム内の GEVEs は、NCLDV 感染時に挿入されたものと予想できるが、異なるウイルスが独立して各々のゲノムに挿入した結果なのか、あるいは過去の内在化が藻類の種分化を通して継承された結果なのかは明確になっていない。今回発見された GEVEs は、緑藻門に限れば *Mimiviridae* と

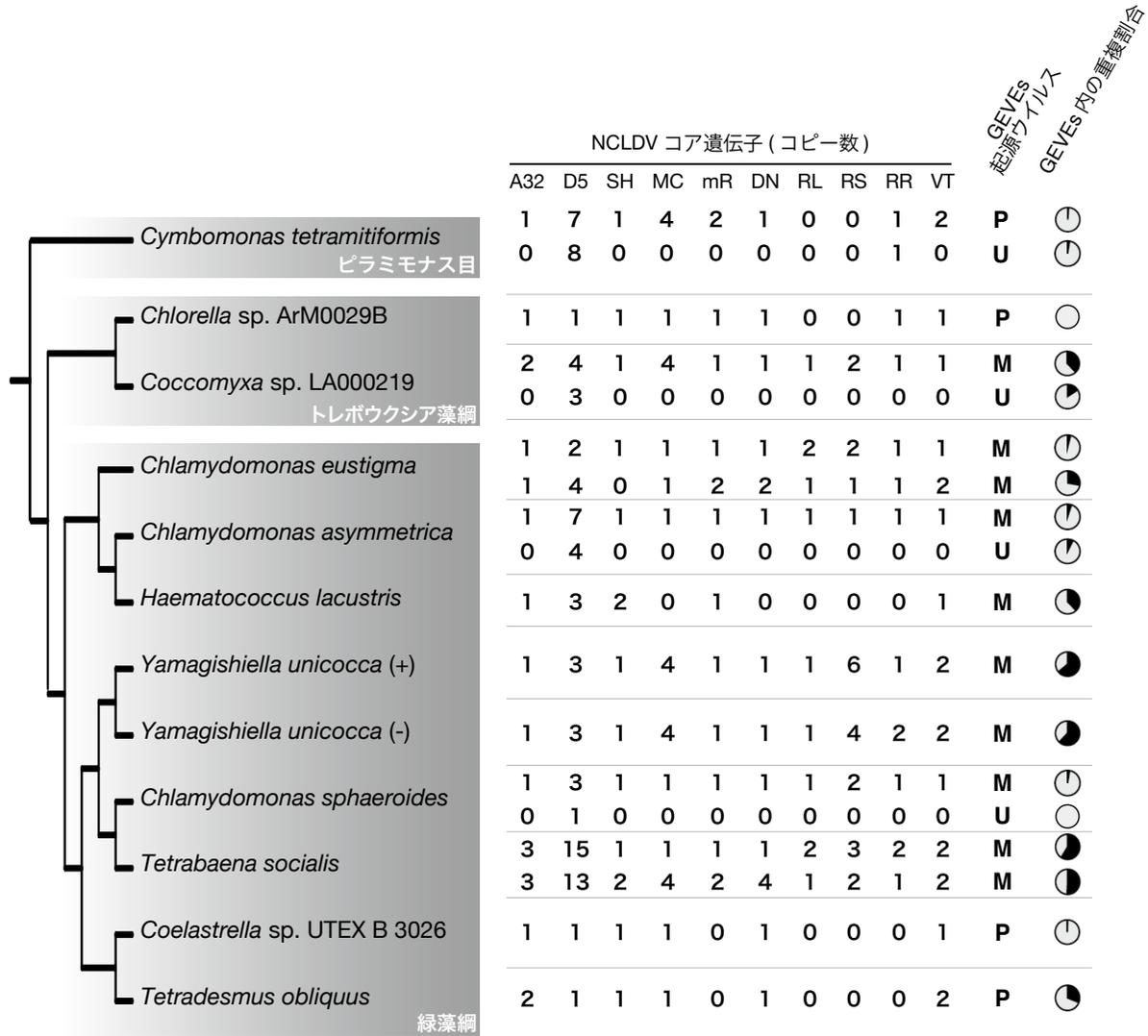


図1. 緑藻の系統関係と GEVEs の特徴. NCLDV のコア遺伝子のコピー数とその遺伝子配列情報から GEVEs の由来となるウイルスの推定. GEVEs 内の重複割合を円グラフで示す. A32, A32-like virion packaging ATPase; D5, D5 helicase/primase; DN, DNA polymerase; GC, guanine + cytosine content; MC, major capsid protein; mR, mRNA capping enzyme; RL, RNA polymerase large subunit; RR, ribonucleotide reductase; RS, RNA polymerase small subunit; SH, superfamily II helicase; VT, VLT3-like transcription factor. P, *Phycodnaviridae*; M, *Mimiviridae*; U, Undetermined. (Moniruzzaman *et al.* 2020, Fig. 1 を元に作成)

Phycodnaviridae の2つに由来するもので、さらに、GEVEs は単離されたウイルスとクラスターを組むことから、過去の内在化を藻類の種分化を通して継承してきたのではなく、個々のウイルスが複数内在化したと解釈するのが妥当かもしれない。しかしながら、GEVEs に由来する NCLDV は、感染先とする宿主藻類を反映している場合もある。例えば、*Yamagishiella unicocca* (Rayburn & R.C. Starr Nozaki, *T. socialis*, *Chlamydomonas sphaeroides* Gerloff は *Mimiviridae* GEVE を保持し、*T. obliquus* と *Coelastrella* Chodat は *Phycodnaviridae* GEVE を保持していた。つまり、共通の祖先に起こった内在化イベントを起源として継承されているようにも見えるが、

種間 GEVEs のアミノ酸の相同性は低く、遺伝子の並び方も異なっており、さらにイントロンの配置も共有されていないことから、過去の内在化を継承してきたという強い証拠とは言えない。

NCLDV は、様々なウイルスや細胞生物から水平伝播に由来するモザイク状の遺伝子を保有することが知られており (Filée 2015, Van Etten *et al.* 2017), 細胞生物でしか見られない代謝経路の遺伝子を保持している (Van Etten *et al.* 2017, Wilson *et al.* 2005)。つまり、NCLDV には、バクテリアオフアージの「補助代謝遺伝子」と同様、宿主の代謝状態を変化させる可能性のある遺伝子が含まれている (Schulz *et*

al. 2020)。実際に, Moniruzzaman *et al.* (2020) が発見した GEVEs 中にも, 糖代謝やクロマチンリモデリング, シグナル伝達, エネルギー生成, 翻訳などで働く遺伝子の多くが含まれており, こうしたウイルス遺伝子のレパートリーが, 宿主ゲノムの進化促進に寄与していると推察されている。これまでバクテリオファージの宿主ゲノムへの内在化は, 進化の主要な推進力として長い間認識されていたが, 真核生物ではあまり一般的でないと考えられてきた。しかし, 今回, Moniruzzaman *et al.* (2020) による緑藻ゲノムにおける広範な GEVEs の発見が, この考え方に一石を投じる形となった。つまり, NCLDV の緑藻植物への内在化の発見は, 多くの真核生物の系統が NCLDV を介して多種多様な遺伝子にアクセスできることを示唆しており, 真核生物全体のゲノム進化過程は過小評価されていたかもしれない。

海産性珪藻 *Chaetoceros tenuissimus* と感染性 dsDNA ウイルス

著者らは珪藻 *Chaetoceros tenuissimus* とその感染性ウイルスとの進化的関係や感染の攻防を明らかにすることを目的として, ホストのゲノムを解読した。本稿で紹介した緑藻ゲノム内にウイルスが内在化している知見を受けて, 珪藻ゲノム内にウイルスの配列が存在するか調べたところ, ウイルスの replication associated protein に類似する配列が断片的に挿入されていることを発見した。緑藻の例のように GEVEs ではないものの, 単離されて性状が明らかになっている感染性ウイルスの遺伝子断片がゲノム中に挿入されているのは, 珪藻内では初めての知見となった。この挿入断片を詳細に調べると, 遺伝子の 5' 末端が欠損しており, さらに遺伝子内で塩基の挿入によりフレームシフトや途中終止コドンが入っていることから不完全な遺伝子であることがわかった。また, RT-PCR により本領域の転写は確認されたが, 今のところその役割は明らかになっていない。

次に, このウイルス断片のゲノムへの挿入過程の解析を行ったところ, 挿入断片の前後に共通する配列の存在が明らかになった。この共通配列はレトロトランスポゾンである Long interspersed nuclear element (LINE) が挿入する際に見られる Target site duplication (TSD) と言われる特徴的な配列であった。ゲノム上に含まれている遺伝子の機能的なレパートリーを集計し, ゲノムが公開されている珪藻 (*Thalassiosira pseudonana* Hasle & Heimdal, *Phaeodactylum tricorutum* Bohlin, *Fistulifera solaris* Mayama, M. Matsumoto, K. Nemoto & Tuy. Tanaka) と比較すると, レトロトランスポゾンに由来する逆転写酵素が *C. tenuissimus* により多く含まれている事が判明した。また, LINE の存在比も他の珪藻より比較的多い事が明らかになった。つまり, 感染時にウイルスの replication associated protein の mRNA が, ホストの LINE によってゲノム中に挿入されたものと推察された。この挿入断片は, PCR 解析から種内の他株にも存在することが明らかになったため, おそらく *C. tenuissimus* の共

通祖先によって獲得されたものと考えられた。以上のことから, *C. tenuissimus* と感染性のウイルスは進化上, 密接に関係しているものと考えられた (Hongo *et al.* 2021)。今後, *C. tenuissimus* だけでなく, *Chaetoceros* 属はもちろん珪藻内でウイルス由来の配列断片が挿入されているか調査し, 珪藻とウイルスの進化的関係を明らかにしていきたい。

引用文献

- Aylward, F. O. & Moniruzzaman, M. 2021. ViralRecall—a flexible command-line tool for the detection of giant virus signatures in 'omic data. *Viruses* 13: 15–17. doi.org/10.3390/v13020150
- Delaroque, N. & Boland, W. 2008. The genome of the brown alga *Ectocarpus siliculosus* contains a series of viral DNA pieces, suggesting an ancient association with large dsDNA viruses. *BMC Evol. Biol.* 8: 1–12. doi.org/10.1186/1471-2148-8-110
- Filée, J. 2015. Genomic comparison of closely related Giant Viruses supports an accordion-like model of evolution. *Front. Microbiol.* 6: 593. doi.org/10.3389/fmicb.2015.00593
- Griffiths, D. J. 2001. Endogenous retroviruses in the human genome sequence. *Genome Biol.* 2: 1–5. doi.org/10.1186/gb-2001-2-6-reviews1017
- Hongo, Y., Kimura, K., Takaki, Y. *et al.* 2021. The genome of the diatom *Chaetoceros tenuissimus* carries an ancient integrated fragment of an extant virus. *Sci. Rep.* (in press)
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W. & Hauser, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11: 119. doi.org/10.1186/1471-2105-11-119
- Iyer, L. M., Balaji, S., Koonin, E. V. & Aravind, L. 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* 117: 156–184. doi.org/10.1016/j.virusres.2006.01.009
- Katzourakis, A. & Gifford, R. J. 2010. Endogenous viral elements in animal genomes. *PLoS Genet.* 6: e1001191. doi.org/10.1371/journal.pgen.1001191
- Kazazian, H. H. 2004. Mobile elements: drivers of genome evolution. *Science* 303: 1626–1632. doi.org/10.1126/science.1089670
- Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P. F. & Prohaska, S. J. 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 12: 124. doi.org/10.1186/1471-2105-12-124
- Marcet-Houben, M. & Gabaldón, T. 2010. Acquisition of prokaryotic genes by fungal genomes. *Trends Genet.* 26: 5–8. doi.org/10.1016/j.tig.2009.11.007
- Monier, A., Pagarete, A., De Vargas, C. *et al.* 2009. Horizontal gene transfer of an entire metabolic pathway between a eukaryotic alga and its DNA virus. *Genome Res.* 19: 1441–1449. doi.org/10.1101/gr.091686.109
- Moniruzzaman, M., Weinheimer, A. R., Martinez-Gutierrez, C. A. & Aylward, F. O. 2020. Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* 588: 141–145. doi.org/10.1038/s41586-020-2924-2
- Ribet, D., Harper, E., Dupressoir, A., Dewannieux, M., Pierron, G. & Heidmann, T. 2008. An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. *Genome Res.* 18: 597–609. doi.org/10.1101/gr.073486.107

- Rossoni, A. W., Price, D. C., Seger, M. *et al.* 2019. The genomes of polyextremophilic cyanidiales contain 1% horizontally transferred genes with diverse adaptive functions. *eLife* 8: e45017. doi.org/10.7554/eLife.45017
- Schulz, F., Roux, S., Paez-Espino, D. *et al.* 2020. Giant virus diversity and host interactions through global metagenomics. *Nature* 578: 432–436. doi.org/10.1038/s41586-020-1957-x
- Sun, C., Feschotte, C., Wu, Z. & Mueller, R. L. 2015. DNA transposons have colonized the genome of the giant virus *Pandoravirus salinus*. *BMC Biol.* 13: 38. doi.org/10.1186/s12915-015-0145-1
- Van Etten, J. L., Agarkova, I., Dunigan, D. D., Tonetti, M., De Castro, C. & Duncan, G. A. 2017. Chloroviruses have a sweet tooth. *Viruses* 9: 88. doi.org/10.3390/v9040088
- Wilson, W. H., Schroeder, D. C., Allen, M. J. *et al.* 2005. Complete genome sequence and lytic phase transcription profile of a *Coccolithovirus*. *Science* 309: 1090–1092. doi.org/10.1126/science.1113109

(2021年8月10日受付, 2021年10月9日受理)
通信担当編集委員: 矢吹 彬憲



『万葉集』に隠れたアラメとカジメ?

仲田 崇志

この歌は『万葉集』の一首(740年前後作か*)で、「もうすでにこれ程までに噂されてさてわが君よこの先どうなるの」との意味である。『万葉集』にはいくつかの海藻が詠まれたが、アラメやカジメを詠んだ歌はない(中西1985.『万葉集事典 万葉集全訳注原文付 別巻』pp.304–333)。しかしこの歌にはアラメが隠れている。末尾の「あらめ」は「なるだろうか」の意だが、「荒海藻」と書かれた。身近な食材だったアラメを言葉遊びに使ったようだ(廣野1998.『食の万葉集』pp.146–147)。

冒頭の「あらかじめ」もカジメに掛かっているかと思いきや、これは考えすぎらしい。古代、カジメは「かちめ」、「豫」は「あらかしめ」と読まれ、発音が異なっていた(例えば『類聚名義抄』)。ちなみに当時「かちめ」は粉末状に加工したアラメを指していた(関根1969.『奈良朝食生活の研究』pp.96–97)。

安土桃山時代までに「かちめ」は一部で「かぢめ」に、「あらかしめ」は「あらかじめ」に濁った(例えば『日葡辞書』)。「ぢ」と「じ」の発音の区別も江戸時代にはなくなり(松本1995.佐藤編『概説 日本語の歴史』pp.98–114)、1946年告示の『現代かなづかい』からは表記も「かじめ」と「あらかじめ」になった。1200年の時を超え、カジメの言葉遊びまで成立してしまったわけだ。

*坂上郎女が娘婿に送った歌とされるため(岡田1990.日本文学研究26:1–11)、娘の誕生時期(720年代後半か)と『万葉集』1～16巻の成立時期(744年まで)から740年前後と推定した。

あらかじめひとごとしげしかくしあらば
豫 人事 繁 如是有者
しゑやわがせこおくもいかにあらめ
四惠也 吾背子 與裳 何如 荒海藻

おほとものさかのうへのいらつめ
大伴坂上郎女

『万葉集』4巻659番. 原文・読みは中西(1978)『万葉集全訳注原文付(一)』p.322に依る。