

氏名 西村 瑠佳

学位(専攻分野) 博士(理学)

学位記番号 総研大甲第 2503 号

学位授与の日付 2024 年 3 月 22 日

学位授与の要件 生命科学研究科 遺伝学専攻
学位規則第6条第1項該当

学位論文題目 Ancient Viral Discovery and Characterization Using Ancient
DNA Sequencing Data

論文審査委員 主査 中村 保一
遺伝学コース 教授
黒川 顕
遺伝学コース 教授
北野 潤
遺伝学コース 教授
川本 祥子
遺伝学コース 准教授
太田 博樹
東京大学 大学院理学系研究科 教授

博士論文の要旨

氏 名：西村瑠佳

論文題目：Ancient Viral Discovery and Characterization Using Ancient DNA Sequencing Data

古代 DNA シーケンシングデータを用いた古代ウイルスの探索と特徴の解明

Ancient DNA can be extracted from archaeological specimens such as bones, teeth, and mummified tissues. The advent of genome sequencing techniques has allowed for extensive analysis of these ancient genetic materials since 1984 when the first ancient DNA was sequenced from a museum specimen. For instance, ancient DNAs extracted from human samples have provided insights into past human migration patterns and the characteristics of ancient human populations.

These ancient DNAs, derived from human samples, not only contain information about human genomes but also harbor genomic information of bacteria, viruses, and food remnants that coexisted within ancient individuals' bodies. These ancient genomic information offers valuable insights into various aspects, including host human health conditions, the lifestyle of ancient humans, past pandemics, and viral evolution. Previous research has discovered pathogenic bacterial and viral genomes, such as *Yersinia pestis* and the Hepatitis B virus, from ancient human samples, shedding light on infection dynamics in ancient human populations thousands of years ago. The time information encoded in these ancient viral sequences contributes significantly to our understanding of long-term viral evolution. The identification of these ancient non-human genomes can be achieved through bioinformatic analyses like homology search or comparison with known genomes. Furthermore, various methods, including machine-learning approaches, can also be employed for viral detection due to the substantial divergence present in viral genomes.

Despite these advancements, the number of detected ancient viruses has been limited, with the majority identified being human pathogenic viruses. Non-human pathogenic viruses, which constitute vital elements of the human virome with significant implications for human health, have been largely overlooked. In this study, I aimed to comprehensively detect ancient viral genomes and characterize ancient viromes, employing various bioinformatic techniques, including *de novo* assembly, homology search, metagenomic profiling, and non-homologous viral detection.

Firstly, I utilized whole genomic sequencing (WGS) data from dental pulps of five Jomon individuals who lived in the Japanese archipelago for thousands of years. Conducting a homology search with known viral genomes, I discovered eleven potential ancient viral genomes. From this set, I successfully reconstructed the likely complete

genome of the Siphovirus contig89 (CT89) virus, uncovering 59 open reading frames. These included phage proteins under strong purifying selection. The host of the CT89 virus was estimated to be *Schaalia meyeri*, based on clustered regularly interspaced short palindromic repeats (CRISPR) estimation. Phylogenetic analysis using both ancient and modern CT89 genomes revealed the most recent common ancestor of CT89 existed around 7900 years ago, offering insights into viral evolutionary processes over millennia. These findings indicate that metagenomic data extracted from the dental pulps of the Jomon people is crucial for recovering ancient viral genomes, allowing for the exploration of their evolutionary histories.

Secondly, I involved the analysis of coprolites or ancient feces from four Jomon individuals dating back 7000 to 5500 years. Coprolites WGS data analysis revealed a significant number of reads showed homology with known gut microbes, viruses, and food-derived genomes typically found in modern human feces. Detection of reads from various gut phages and their host bacteria simultaneously suggested the coexistence of viruses and hosts. The food genomes provide biological evidence for the dietary behavior of the Jomon people, consistent with previous archaeological findings. These results indicate that ancient genomic analysis of coprolites is useful for understanding the gut environment and lifestyle of ancient peoples.

Thirdly, I analyzed 384 WGS data, including 27 samples from various ages in the Japanese archipelago, 9 samples from Indonesia, and 348 WGS data of ancient individuals obtained from the NCBI database. Employing multiple viral detection methods, including homology search, machine-learning based methods, and a CRISPR based method, I identified 50,288 candidate ancient viral genomes. Most of these were classified as phages, particularly *Caudoviricetes* viruses, with some showing no resemblance to modern viral sequences, potentially indicating highly diverged or extinct ancient viral genomes. Furthermore, I identified host bacterial genomes and analyzed the relationships between viruses and bacterial through host CRISPR loci. The red complex bacteria, associated with severe periodontal diseases, targeted several ancient viral sequences. These findings indicated that a thorough approach to ancient viral detection can elucidate ancient viromes by examining both viral compositions and their relationships with the host.

In summary, the findings demonstrate that ancient WGS data can facilitate the detection of genomes from ancient viruses, microbes, and food remnants. Analyzing these genomes allows for the characterization of ancient genomes, ancient viromes, and insights into long-term viral evolution.

博士論文審査結果

Name in Full
氏名 西村 瑠佳

Title
論文題目 Ancient Viral Discovery and Characterization Using Ancient DNA Sequencing Data

近年の次世代シーケンシング技術の進歩により、骨や歯、ミイラ化した組織などの考古学的サンプルから古代の DNA が抽出され、大規模な配列決定が行われるようになってきている。こうした古代の DNA の解析により、歴史的にみた人類の移動経路や、古代の人類集団の特徴が明らかにされてきている。古代の DNA サンプルにはヒトゲノムだけでなく古代のウイルスゲノムに由来する情報も含まれるが、これまで、古代ウイルスの同定は限られておりその多くはヒト病原性ウイルスであった。しかしながら非ヒト病原性ウイルスもヒトの健康に多大な影響を及ぼすヒトウイルス叢の重要な構成要素である。西村さんは、多くが未解明の領域であった古代のウイルスゲノムを網羅的に検出するために生命情報解析による研究を展開した。

最初に、西村さんは 3,800 年前の遺跡から出土した縄文人 5 個体由来の歯髄を試料とした塩基配列データを解析し、歯髄中に残されていた古代ウイルスゲノムを検出した。解読された塩基配列からヒト由来のリードを除去した後に *de novo* アセンブルを行い Siphovirus contig89 (CT89) ウイルスのほぼ完全な塩基配列を再構築した。この配列を現代の CT89 ゲノムと比較解析することにより CT89 の共通祖先は約 7,900 年前に存在したことを示した。次に西村さんは 7,000 年から 5,500 年前の縄文人の糞石由来の古代 DNA を分析することで、現代人の糞便にも典型的に見られる既知の腸内微生物、ウイルス、食物由来 DNA と類似のあることを示した。最後に西村さんは、より古代のウイルスを包括的に発見するために、世界各地に由来する 384 セットの古代の WGS データを公共データベースから取得し解析した。多様なウイルス検出法を組み合わせた解析により、ゲノムとしてある程度の品質をもつものから完全なものまで 483 の古代ウイルスゲノムを得た。中には現代のウイルス配列と一致しないものもあり、高度に分岐した、あるいは絶滅した古代のウイルスゲノムが存在する可能性を示唆している。これらの研究結果により、西村さんは古代ウイルスゲノムの特徴を明らかにするとともに、長期的なウイルス進化の解明の可能性を示すことで、古代ウイルス解析研究の有用性を強調した。

この博士論文は遺跡由来の試料にもとづく塩基配列からウイルスゲノムを再構築すること、またそのゲノム塩基配列からの分子進化研究を実施することが可能であることを明瞭に示し、この分野における新規の多くの知見を記載しており優れた研究成果と認められる。したがって審査委員会は本論文が学位の授与に値すると判断した。