Figures and figure supplements

Long read sequencing reveals poxvirus evolution through rapid homogenization of gene arrays

Thomas A Sasani et al.
Figure 1. A single nucleotide variant accumulates following increases in K3L copy number. (A) Following 20 serial infections of the \( \Delta E3L \) strain (MOI 0.1 for 48 hr) in HeLa cells (see Materials and methods for further details), replication was measured in triplicate in HeLa cells for every fifth passage, and compared to wild-type (VC-2) or parent (\( \Delta E3L \)) virus. (B, C) Digested viral DNA from every 5th passage (B) and four plaque-purified clones (C) were probed with a K3L-specific probe by Southern blot analysis. Number of K3L copies (left) and size in kbp (right) are shown. K3L\textsubscript{His47Arg} allele frequency for each population (shown below) was estimated by PCR and Sanger sequencing of viral DNA. (D) Replication of plaque purified clones from (C) was measured in HeLa (D) or BHK (Figure 1—figure supplement 1) cells in triplicate. Statistical analysis was performed to compare the means of populations b or c relative to a, or between the means of populations b or c relative to d by one-way ANOVA followed by Dunnett’s multiple comparison test. *p<0.05, **p<0.01, ***p<0.005. K3L\textsubscript{His47Arg} and E9L\textsubscript{Glu495Gly} population-level allele frequencies estimated from Illumina MiSeq reads are shown in Figure 1—figure supplement 2. Replication of clone a compared to \( \Delta E3L \) is shown in Figure 1—figure supplement 3. All titers were measured multiple times in BHK cells by plaque assay, shown with median and 95% confidence intervals.

DOI: https://doi.org/10.7554/eLife.35453.002
Figure 1—figure supplement 1. K3L^His47Arg and K3L CNV are non-adaptive in the permissive BHK cell line. Replication was measured for plaque purified clones (as in Figure 1D) in BHK cells by 48 hr infection (MOI 0.1) in triplicate. All titers were measured multiple times in BHK cells by plaque assay, shown with median and 95% confidence intervals. Statistical analysis was performed to compare the means of populations b or c relative to a, or between the means of populations b or c relative to d by one-way ANOVA followed by Dunnett’s multiple comparison test. No significant differences were observed between any of the compared populations.

DOI: https://doi.org/10.7554/eLife.35453.003
Figure 1—figure supplement 2. Allele frequencies of the two high-frequency SNVs identified in vaccinia populations. Population-level K3L<sub>His47Arg</sub> and E9L<sub>Glu495Gly</sub> allele frequencies were estimated using freebayes on Illumina MiSeq reads from different passages.
DOI: https://doi.org/10.7554/eLife.35453.004
Figure 1—figure supplement 3. The E9L<sup>Glu495Gly</sup> variant does not contribute to virus replication. A virus clone containing the E9L<sup>Glu495Gly</sup> variant as the only genetic change relative to ∆E3L was isolated following four rounds of plaque purification in BHK cells (clone a in Figure 1C). Replication was measured in comparison to ∆E3L virus after 48 hr of infection (MOI 0.1) of HeLa cells in triplicate. Titers were measured multiple times in BHK cells by plaque assay, shown with median and 95% confidence intervals. ns, not significant by unpaired 2-tailed t test with Welch’s correction.

DOI: https://doi.org/10.7554/eLife.35453.005
**Figure 2.** ONT reads capture SNVs and copy number expansions in individual virus genomes. (A) Representative structure of the K3L locus in the VC-2 reference genome is shown on top, with representative Illumina MiSeq and ONT MinION reads shown to scale below. The K3L<sup>His47Arg</sup> variant within reads is indicated by an asterisk. ONT reads that split and re-align to the K3L duplicon are indicative of individual multicopy arrays (shown below). Tandem duplication breakpoints flanking the duplicon are indicated by arrowheads. (B) Population-level K3L<sup>His47Arg</sup> allele frequency was estimated using Illumina or ONT reads from different passages. E9L<sup>Glu495Gly</sup> allele frequencies are shown in Figure 2—figure supplement 1. Error rate calculations for different flow cell chemistries are shown in Figure 2—figure supplement 2. (C) For each sequenced passage, K3L copy number was assessed within each ONT read that aligned at least once to the K3L duplicon (see Materials and methods for further details). Detailed plot of reads Figure 2 continued on next page.
containing 6 + K3L copies is shown in Figure 2—figure supplement 3. (D) Representative reads from the specific long read preparation are depicted relative to the VC-2 reference genome. The locations of relevant genes are indicated by colored boxes (gene name above or below), and the locations of high frequency variants in K3L and E9L are indicated by arrowheads.

DOI: https://doi.org/10.7554/eLife.35453.014
Figure 2—figure supplement 1. E9L$^{\text{Glu}495\text{Gly}}$ variant dynamics. Population-level E9L$^{\text{Glu}495\text{Gly}}$ allele frequencies were estimated using freebayes and nanopolish on Illumina or ONT reads, respectively, from different passages as in Figure 2B.
DOI: https://doi.org/10.7554/eLife.35453.015
Figure 2—figure supplement 2. Error rate profiles in ONT reads. The proportions of non-reference bases aligned to the 5-mers containing the K3L\textsuperscript{WT} (A), K3L\textsuperscript{His47Arg} (B), E9L\textsuperscript{WT} (C), or E9L\textsuperscript{Glu495Gly} (D) sequences were calculated from alignments of ONT reads from the P15 population generated with Figure 2—figure supplement 2 continued on next page.
three distinct flowcell chemistries (listed above each plot). Kernel density plots represent the distributions of error proportions for each non-reference base across all 5-mers in the vaccinia reference genome.

DOI: https://doi.org/10.7554/eLife.35453.016
Passage: 10 15 20
Max K3L copies: 13 16 15

Figure 2—figure supplement 3. ONT reads capture high K3L copy number in vaccinia genomes. K3L copy number was assessed in ONT reads from P10, P15, and P20 as described in Figure 2C. Stacked bar plots indicate overall proportions of sequencing reads that contain between 6 and 16 copies of K3L, in ascending order (darker bars represent increasing copy number). The maximum copy number observed in any single ONT read is indicated below each passage.

DOI: https://doi.org/10.7554/eLife.35453.017
Figure 3. The K3L<sup>His47Arg</sup> variant homogenizes within multicopy arrays throughout experimental evolution. Stacked bar plots representing the proportions of mixed and homogeneous K3L arrays were generated from ONT reads for the indicated virus populations (passages are listed above each plot). The proportions of reads containing homogeneous K3L<sup>WT</sup>, homogeneous K3L<sup>His47Arg</sup>, or any combination of mixed alleles are shown for reads containing 1–5 K3L copies. A simulation of SNV accumulation under a binomial distribution is shown in Figure 3—figure supplement 1, and results from sequencing with different flow cell chemistries is shown in Figure 3—figure supplement 2. Simulations of the effects of ONT sequencing error rates on the identification of mixed and homogeneous arrays are shown in Figure 3—figure supplement 3, and the proportions of each combination of K3L alleles in 3, 4, and 5-copy arrays are shown in Figure 3—figure supplement 4.

DOI: https://doi.org/10.7554/eLife.35453.024
Figure 3—figure supplement 1. Simulated accumulation of the K3L\textsuperscript{His47Arg} SNV. The K3L\textsuperscript{His47Arg} allele was uniformly distributed in simulated vaccinia populations with copy number distributions identical to passages P10, P15, and P20 (see Materials and methods for further detail). Stacked bar plots representing the proportions of mixed and homogeneous K3L arrays were generated as in Figure 3.

DOI: https://doi.org/10.7554/eLife.35453.025
Figure 3—figure supplement 2. ONT flowcell chemistries do not affect observed proportions of homogeneous and mixed K3L arrays. The P15 population was sequenced using R7.3, R9, and R9.4 ONT flowcell chemistries, and stacked bar plots representing the proportions of mixed and homogeneous K3L arrays were generated as in Figure 3.
DOI: https://doi.org/10.7554/eLife.35453.026
Figure 3—figure supplement 3. ONT sequencing error rates do not affect observed proportions of homogeneous and mixed K3L arrays. (A) Using reads from passage 15 populations sequenced with R7.3, R9, and R9.4 flowcell chemistries, all mixed arrays were converted into homogeneous arrays (see Materials and methods for further detail). K3L\textsuperscript{WT} and K3L\textsuperscript{His47Arg} alleles were then switched as a proxy for sequencing errors, and the proportions of mixed arrays recovered were calculated. Stacked bar plots representing the proportions of mixed and homogeneous arrays were then generated following one simulation. (B) Kernel density plots were also generated, representing the distributions of expected mixed array proportions recovered across 1000 simulations. The observed proportion of mixed arrays using the indicated flowcell chemistry is shown as a dotted red line.

DOI: https://doi.org/10.7554/eLife.35453.027
Figure 3—figure supplement 4. Multicopy K3L arrays contain diverse combinations of K3L^{WT} and K3L^{His47Arg} alleles. (A) The proportions of 3-copy K3L arrays containing each possible combination of K3L^{WT} and K3L^{His47Arg} alleles at P10, P15, and P20 were counted. Dotted red lines separate mixed and homogeneous arrays, and the total fraction of mixed arrays is indicated above each plot. The proportions of (B) 4-copy and (C) 5-copy arrays with each possible combination of K3L alleles are shown for passage 15, which was re-sequenced to generate a larger number of reads.

DOI: https://doi.org/10.7554/eLife.35453.028
Figure 4. The K3LHis47Arg variant homogenizes in K3L arrays regardless of copy number. (A) ONT reads from every 5th passage were grouped by K3L copy number, and each K3L copy was assessed for the presence or absence of the K3LHis47Arg SNV. Reads containing 1–5 K3L copies are shown. (B) Using reads from the P15 population, homogeneous K3L arrays were removed from the dataset, and K3LHis47Arg SNV frequency was plotted exclusively in mixed arrays. The number of reads of each copy number is indicated to the right of each row. Reads are oriented 5’ to 3’ relative to the VC-2 reference sequence, and the K3LHis47Arg allele frequency in each copy is indicated in blue.

DOI: https://doi.org/10.7554/eLife.35453.029
**Figure 5.** K3L^{His47Arg} homogenization within multicopy arrays is independent of intergenomic recombination rate. The P10 population was serially passaged in HeLa cells at different MOIs (listed above each plot), and each of the resulting P15 populations was sequenced with ONT. Stacked bar plots representing the proportions of mixed and homogeneous arrays were generated as in Figure 3.

DOI: https://doi.org/10.7554/eLife.35453.030
Figure 6. K3L^{His47Arg} variant homogenization is dependent on selection. The P10 population was serially passaged five times in BHK cells (MOI = 0.1, 48 hr; P15-BHK). P10 and P15 data are included from previous figures for comparison with P15-BHK. (A) K3L copy number was assessed for all passages.

Figure 6 continued on next page.
sequenced reads that unambiguously aligned to K3L at least once, as in Figure 2C. (B) K3L<sup>H47R</sup> and E9L<sup>G49G</sup> allele frequencies in each population were estimated using ONT reads, as in Figure 2B. Allele frequencies for all sequenced populations are included in Figure 6—source data 1. (C) Stacked bar plots representing the proportions of mixed and homogeneous arrays were generated from sequenced ONT reads, as in Figure 3. (D) ONT reads were assessed for the presence of each breakpoint (shown relative to the genome to the right) by aligning reads to a query sequence containing K3L using BLAST and extracting the starts and ends of individual alignments to the K3L duplicon. Due to sequencing errors, a proportion of reads do not match either breakpoint 1 or breakpoint 2.

DOI: https://doi.org/10.7554/eLife.35453.031
Figure 7. Model of K3L_{His47Arg} homogenization within K3L CNV via gene conversion. DOI: https://doi.org/10.7554/eLife.35453.034