Evolution of the orthopoxvirus core genome

Cristian Molteni\textsuperscript{a, 1, *}, Diego Forni\textsuperscript{a, 1}, Rachele Cagliani\textsuperscript{a}, Alessandra Mozzi\textsuperscript{a}, Mario Clerici\textsuperscript{b, c}, Manuela Sironi\textsuperscript{a}

\textsuperscript{a} Scientific Institute IRCCS E. MEDEA, Bioinformatics, Bosisio Parini, Italy
\textsuperscript{b} University of Milan, Milan, Italy
\textsuperscript{c} Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

\begin{abstract}
Orthopoxviruses comprise several relevant pathogens, including the causative agent of smallpox and monkeypox virus. Analysis of orthopoxvirus genome evolution mainly focused on gene gains/losses. We instead analyzed core genes, which are conserved in all orthopoxviruses. We show that, despite their strong constraint, some genes involved in viral morphogenesis and transcription/replication were targets of pervasive positive selection, which was relatively uncommon in immunomodulatory genes. However at least three of the positively selected genes, E3L, A24R, and H3L, might have evolved in response to immune selection. Episodic positive selection was particularly common on the internal branches of the orthopox phylogeny and on the monkeypox virus lineage. The latter showed evidence of episodic positive selection at the D14L gene, which encodes a modulator of complement activation (MOPICE). Notably, two genes (B1R and A33R) targeted by episodic selection on more than one branch are involved in forms of intra-genomic conflict. Finally, we found that, in orthopoxvirus proteomes, intrinsically disordered regions (IDRs) tend to be less constrained and are common targets of positive selection. Extension of our analysis to all poxviruses showed no evidence that the IDR fraction differs with host range. Conversely, we found a strong effect of base composition, which was however not sufficient to explain IDR fraction. We thus suggest that, in poxviruses, the IDR fraction is maintained by modulating GC content to accommodate disorder-promoting codons. Overall, our data provide novel insight in orthopoxvirus evolution and provide a list of genes and sites that are expected to modulate viral phenotypes.
\end{abstract}

\section{Introduction}

The \textit{Poxviridae} family includes a number of diverse double-stranded (ds) DNA viruses with large genomes, ranging in length from ~135 to ~350 kbp. Poxviruses infect a wide spectrum of hosts, including insects, birds, reptiles, and mammals (Alonso et al., 2020; Gyuranecz et al., 2013; Lefkowitz et al., 2006; Moss, 2013; Sarker et al., 2019). Thus, the \textit{Poxviridae} family is divided into two subfamilies, \textit{Chordopoxvirinae} and \textit{Entomopoxvirinae}, for viruses that infect vertebrates and invertebrates, respectively. Chordopoxviruses are further classified into 18 genera (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/74/poxviridae). Among these, the \textit{Orthopoxvirus} genus comprises several viruses of great medical relevance, including variola virus (VARV), the causative agent of smallpox, and vaccinia virus (VACV), which was used in the smallpox eradication campaign (Fenner et al., 1988). Additional orthopoxviruses with zoonotic potential such as monkeypox virus (MPXV) and cowpox virus (CPXV), are increasingly reported as a cause of human disease, possibly because of waning population immunity caused by discontinuation of routine smallpox vaccination (Beer and Rao, 2019; Silva et al., 2020; Simpson et al., 2020). Indeed, a recent multi-country MPXV outbreak outside of Africa has generated international concerns of possible changes in the epidemiology of this virus (Bunge et al., 2022). In addition to viruses that have been known for decades, recent years have also witnessed the identification of novel orthopoxviruses in humans and other animals. Thus, Akhmeta (AKMV) and Alaska (AKPV) viruses have been sequenced in the last 15 years from people living in Georgia and Alaska (Gao et al., 2018; Gigante et al., 2019), whereas orthopoxvirus Abatino (OPVA) was first isolated in Italy during an outbreak in captive macaques in 2015 (Cardeti et al., 2017). To date, 12 orthopoxvirus species have been officially recognized by the ICTV (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses-2011/w/dsdna_viruses/74/poxviridae).

\textsuperscript{1} These authors contributed equally to this work.

\textsuperscript{*} Corresponding author.
E-mail address: cristian.molteni@lanostrafamiglia.it (C. Molteni).
Distinct orthopoxviruses largely differ in their host range. For instance, VARV and camelpox virus (CMLV) can only infect humans and camels, respectively, whereas CPXV and MPXV most likely have their natural reservoirs in rodents but can infect a wide range of mammals (Silva et al., 2020). Also, compared to variable coding sequences, core genes are more likely to become the targets of novel antipoxvirus preventive and therapeutic strategies. We therefore focused on genes that are conserved more likely to become the targets of novel antipoxvirus preventive and therapeutic strategies. 

Gene annotations and genome sequences were retrieved from the National Center for Biotechnology Information database (NCBI, http://www.ncbi.nlm.nih.gov/). Genome sequences and viral species were selected on the basis of the ICTV classification (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011-w/dsdna-viruses/74/poxviridae), with the addition of the Alaskapox virus (Gigante et al., 2019). In the case of CPXV, four sequences were included to be representative of major clades (Franke et al., 2017) (Fig. 1, Table S1). For VACV, the horsepox virus sequence, possibly representing a natural isolate, was selected (Duggan et al., 2020; Tulman et al., 2006). For VARV, an Indian isolate sampled in 1964 was preferred over the reference strain, as the latter has an unknown passage history and was sequenced by subcloning (Schelkunov et al., 1993). The CPXV reference strain was not included for similar reasons (Archard and Mackett, 1979). In all other cases, reference strains were used.

The whole genome alignment was obtained using Progressive Mauve (v.2.3.1) (Darling et al., 2004, 2010), a tool that aligns genome sequences by taking into account the presence of large deletions, rearrangements, and inversions. Mauve identifies and aligns regions of local collinearity; each block is a region of homologous sequences shared by two or more genomes. Using collinear block information and the Mauve output, we identified one-to-one orthologous genes that were present in all the 16 species analyzed. This generated a dataset of 123 orthologous genes (Supplementary Table S1).

Gene alignments were generated using the GUIDANCE2 suite (Sela et al., 2015), setting sequence type as codons and using MAFFT (Katoh and Standley, 2013) as an aligner. GUIDANCE2 also allows to filter unreliably aligned positions; we removed codons with a score lower than 0.90 (Privman et al., 2012). Finally, the resulting alignments were manually checked.

A consensus tree of the 16 viruses was constructed from a concatenated alignment of 49 conserved genes in poxviruses. Nodes with bootstrap support higher than 0.8 are labeled with a dot.
2.2. Positive selection analyses

All gene alignments were analyzed for the presence of recombination signals using the RDP5 tool (Martin et al., 2020). This tool allows the implementation of several recombination methods. Specifically, we applied four methods (RDP, GENECONV, MaxChi, and Chimera) (Martin and Bybicki, 2006; Martin et al., 2017; Posada and Crandall, 2001; Sawyer, 1989; Smith, 1992) that showed good power (Bay and Biehn, 2000; Martin et al., 2017; Posada and Crandall, 2001; Martin et al., 2020). We applied four methods (RDP, GENECONV, MaxChi, and Chimera) (Martin and Bybicki, 2006; Martin et al., 2017; Posada and Crandall, 2001; Sawyer, 1989; Smith, 1992) that showed good power (Bay and Biehn, 2000; Martin et al., 2017; Posada and Crandall, 2001; Martin et al., 2020).

In order to identify specific branches with a proportion of sites evolving with dN/dS > 1, the adaptive Branch-Site Random Effects Likelihood method (aBS-REL) was used. This method applies sequential likelihood ratio tests to identify branches under positive selection without a priori knowledge about which lineages are of interest (Smith et al., 2015).

2.3. Protein-protein interaction analysis and epitope prediction

Data regarding viral protein-protein interactions were retrieved from a previous work (Mirzakhanyan and Gershon, 2019). Data regarding host-virus protein interactions were retrieved from the literature (Zhang et al., 2009) and from the Virus MINT database (https://maayanlab.cloud/Harmonizome/resource/Virus+MINT). A comprehensive interaction network was built using Cytoscape v3.9.1 (Shannon et al., 2003).

Linear B-cell epitopes were predicted using the B Cell Epitope Prediction Tools from the IEDB webserver (http://www.iedb.org/), using the amino acid sequence of VACV H3 protein as input and with the default threshold (0.5).

Prediction of discontinuous B-cell epitopes was carried out using the Discotope 2.0 webserver (Kringelum et al., 2012), using the 3D-structure of VACV H3 protein as input (PDB id:5ej0) and with a cutoff = 3.7.

2.4. Analysis of intrinsically disordered regions

The fraction of disordered residues for the core proteomes of the 16 orthopoxviruses were calculated using the IUPred3 (Dosztanyi, 2018; Mezuras et al., 2009, 2019) and the Espritz (Walsh et al., 2012) tools.

Predictions were run on individual proteins.

For IUPred3, the short disorder prediction type was selected and we defined as disordered all residues in a protein showing a IUPred3 score > 0.5. Espritz was run using X-ray method as prediction type and best Sw as decision threshold (Walsh et al., 2012).

The IUPred3 tool was also applied to a set of 22 proteomes (both core and accessory) that are representative of different genera in the Poxviridae family (Supplementary Table S3).

For all viruses analyzed herein, gene sequences were also retrieved without any prior knowledge about which lineages are of interest (Smith et al., 2015).

2.5. Statistical analyses

Differences in dN/dS or in fraction of disordered residues among gene groups were evaluated using Kruskal-Wallis or Wilcoxon rank sum test, when appropriate. Nemenyi tests with γ2 distribution to account for ties were used as post-hocs using the PMCMRplus R package (http://www.iedb.org/).
The binomial test was performed by counting, for the positively selected genes, the number of positively selected disordered residues, the total number of positively selected sites, and the fraction of disorder residues compared to the total length of the residues in the positively selected proteins.

All calculations were performed in the R environment.

**Fig. 2.** Relative evolutionary rates of orthopoxvirus genes. (A) Whole-genome alignment obtained with progressive Mauve. Each genome is laid out in a horizontal track, with annotated coding regions denoted as boxes. A similarity plot generated by progressive Mauve is shown above each genome, with colors indicating regions that align to part of any another genome and thus represent locally collinear blocks. A similarity profile is also plotted within blocks, with height proportional to the level of conservation in that region. For clarity, only few representative viral species are shown. The 123 genes we analyzed in this study are colored in the prototype VACV genome. Colors are according to conservation: grey, genes conserved in all poxviruses; red, genes conserved in chordopoxviruses; blue, genes conserved in orthopoxviruses. (B) Boxplots of dN/dS for the 123 analyzed genes based on their conservation level. Colors are as in panel A. Statistical significance calculated by the Nemenyi post-hoc test is also reported. *, p value < 0.05; **, p value < 0.01; ***, p value < 0.001. (C) Boxplots of dN/dS values for the analyzed genes based on their different functional classes. Statistical significance calculated by the Nemenyi post-hoc test is also reported. *, p value < 0.05; **, p value < 0.01; ***, p value < 0.001.
3. Results

3.1. Relative evolutionary rates of orthopoxvirus core genes

With the aim to explore the selective patterns acting on orthopoxvirus genes, we analyzed the genomes of 16 viruses (Table 1, Fig. 1). Eleven of them correspond to recognized species: AKMV, CMLV, Ectromelia virus (ECTV), MPXV, OPVA, Racoonpox virus (RPXV), Skunkpox virus (SPXV), Taterapox virus (GBLV, also known as TATV), VACV, VARV, and Volepox virus (VXPV). Four additional sequences were selected to represent the major clades within the CPXV genetic diversity (Franke et al., 2017) and one genome represents the recently sequenced AKPV (Gigante et al., 2019) (https://talk.ictvonline.org/ictv-reports/ictv_9th_report/dsdna-viruses-2011/w/dsdna_viruses/74/poxviridae) (Table 1, Fig. 1).

In line with previous observations (Hatcher et al., 2014; Hendrickson et al., 2016; Senkevich et al., 2021; Upton et al., 2003), a whole-genome alignment with Mauve (a method for constructing multiple genome alignments in the presence of large-scale evolutionary events such as rearrangement and deletions/duplications (Darling et al., 2004, 2010)) revealed a large central collinear block, which encompasses the majority of highly conserved genes. Due to gene gains and losses, the terminal genome regions are known to be highly dynamic (Fig. 2A and Supplementary Table S1). In the central region, Mauve identified 123 one-to-one orthologs present in all the genomes. These include 49 genes shared by all poxviruses, 38 conserved in chordopoxviruses, and 36 that were common to all orthopox viruses (Fig. 2A) (Upton et al., 2003).

All one-to-one orthologs in the sixteen orthopoxvirus genomes were aligned and rigorously filtered to purge regions or codons with poor alignment quality. After accounting for recombination, we used the codeml M0 model to calculate the average non-synonymous substitution or synonymous substitution rate (dn/ds, also referred to as ω). Comparison among gene groups indicated that dn/ds is significantly different depending on the degree of conservation: genes that are present across all poxviruses have the lowest values, whereas those that are conserved in orthopoxviruses but not necessarily in the other genera or families have the highest (Kruskal-Wallis rank sum test, p value = 3.55 × 10⁻¹¹) (Fig. 2B).

We next used UniProt and bibliographic sources (Upton et al., 2003) to group viral genes into functional categories (Supplementary Table S1). Comparison of dn/ds among the seven functional classes revealed significant differences (Kruskal-Wallis rank sum test, p value = 9.02 × 10⁻⁸). The major contributors to the different evolutionary patterns were genes involved in immunomodulation, cell entry/attachment or cell-to-cell spread. These genes displayed higher dn/ds than those encoding proteins that participate to viral transcription, replication or morphogenesis (Fig. 2C).

3.2. Genome-wide scan for positive selection

The calculation of average dn/ds for viral genes resulted in values below 1, indicating that purifying selection is the major force acting on orthopoxvirus core genomes (Supplementary Table S2). This is consistent with observations on many other viral and non-viral genomes (Lequime et al., 2016; Li et al., 2020; Mozzii et al., 2020a; Sironi et al., 2015; Smith et al., 2013; Wertheim and Kosakovsky Pond, 2011). Nonetheless, positive selection often acts on specific sites in proteins that are otherwise selectively constrained. To test for this possibility, we analyzed the 123 ortholog alignments by applying the “site models” implemented in the codeml tool from the PAML suite (see methods). These models can be used to test whether a gene has experienced positive selection and which sites were the targets thereof. Specifically, we used likelihood ratio tests (LRTs) that compare models of gene evolution that allow (model M8) or disallow (model M7) a class of codons to evolve with dn/ds > 1. After false discovery rate (FDR) correction for multiple tests, 23 genes showed a significant LRT (Supplementary Table S2). These genes were thus tested with a more conservative M8/M8a test. Fifteen of them were significant and were considered targets of positive selection (Table 2, Supplementary Table S2). The specific sites subject to selection were identified using the BEB analysis from MB (Table 2).

Somehow surprisingly, only two of the positively selected genes are primarily involved in host range control and in immunomodulation. However, one of them (E3L) is a pleiotropic effector with the ability to counteract multiple pathways of the innate immune system (Szczersba et al., 2022). E3L is considered a host range gene in poxviruses and it modulates virulence in mouse models of VACV infection (Brandt et al., 2005; Bratke et al., 2013). The positively selected site we identified (V83) is located in the N-terminal region, which functions as a Z-nucleic-acid-binding domain and is essential for VACV pathogenesis in mice (Szczersba et al., 2022).

Three of the positively selected genes we identified encode proteins with a role in cell-to-cell spread or cell entry/attachment (Table 2). Notably, two such proteins, K2 and A56 interact. The K2/A56 complex prevents cell to cell fusion to avoid syncytia formation (Turner and Moyer, 2008). In turn, this process has a central role in superinfection exclusion, a mechanism that blocks infection of an already infected cell (Turner and Moyer, 2008). The positively selected H3L gene instead encodes a virion envelope protein which binds heparan sulfate on the cell surface. Thus, the protein mediates viral adhesion to the host cell, although it is likely to play a role in virion assembly, as well (da Fonseca et al., 2000; Lin et al., 2000). The protein product is also a major target of neutralizing antibodies (Davies et al., 2005), suggesting that the selective pressure might be exerted by the host humoral immune system. To evaluate this possibility, we predicted linear and conformational B-cell epitopes. Two of the three positively selected sites (P17 and N251) were located in predicted epitopes (Fig. 3).

The majority of positively selected genes is however involved in essential viral functions such as transcription, replication and morphogenesis (Table 2). Nevertheless, mutations in the A24R gene, which encodes the catalytic subunit of the viral RNA polymerase, have been associated with modulation of the RNAse L/PKR pathway, as well as to resistance to the antipoxviral drug isatin-beta-thiosemiacbazone (IBT) (Brennan et al., 2015; Cone et al., 2017; Cressawn et al., 2007). Specifically, mutations (T1121M and D1076G) in the C-terminal portion of A24, where one of the positively selected sites also maps, were associated with PKR evasion and IBT resistance (Brennan et al., 2015; Cressawn et al., 2007). Analysis of the crystal structure of the A24 protein indicated that the other positively selected site (P605) is spatially close to the L18F mutation, which confers fitness trade-offs depending on the cell type or host species (Cone et al., 2017). Three additional genes (E4L, E11L, and A8R) we identified as positive selection targets encode subunits of the RNA polymerase or transcription factors (Table 2). However, no evasion or IBT resistance mutations have to date been reported in the respective proteins. Most of these genes, as well as A20R, encoding a viral DNA polymerase processivity factor, are conserved across chordopoxviruses (Ishii and Moss, 2001). An additional positively selected gene, B1R, encodes a serine-threonine kinase essential for viral DNA synthesis (Lin et al., 1992). The kinase also interacts with a number of cellular proteins, possibly to promote cell survival (Nichols et al., 2006; Santos et al., 2004, 2006).

The remaining positively selected genes are poorly characterized (Table 2). Three of them are involved in different stages of morphogenesis. In particular, the H7 protein contributes to the formation of crescent membranes, a very early event in virion formation (Meng et al., 2013). Conversely, F12 is involved in later stages, as it is necessary for the egress of virions from the host cell (Carpentier et al., 2017), and for actin tail formation (Zhang et al., 2000). Thus, deletion mutants lacking the F12L gene display a small plaque phenotype and are attenuated in a mouse model of infection (Zhang et al., 2000). Finally, E6R is necessary for viroplasm package into viral membranes (Resch et al., 2009).
evidence of selection on multiple branches (see below). The other three genes on this branch, one is the ortholog of VACV A33R, which showed episodic positive selection (Fig. 4A and Table 3). Among the four positively selected genes, two are involved in morphogenesis and transcription, VACV C3L and its MPXV ortholog, DI4L (also referred to as MOPICE), function as modulators of the complement system and represent virulence factors (Chen et al., 2005). Surprisingly, two of the four genes showing evidence of episodic positive selection on multiple branches (Fig. 4B), B1R and A33R, are involved in peculiar poxviral mechanisms that control viral fitness and spread. Thus, B1R encodes a protein kinase that antagonizes both a cellular antiviral factor (BAF) and the poxvirus-encoded pseudokinase D14L (also referred to as MOPICE), function as modulators of the complement system and represent virulence factors (Chen et al., 2005).

### 3.3. Episodic selection across the orthopox virus phylogeny

The LRTs we applied to search for positively selected genes are devised to detect pervasive positive selection. However, selection is often episodic, acting on a few branches in a phylogenetic tree. To explore the frequency of episodic positive selection and to assess whether any branch(es) was particularly targeted by it, we applied the adaptive Branch-Site Random Effects Likelihood (aBSREL) model. In particular, we run aBSREL to test whether, on each branch of the phylogeny, a proportion of sites have evolved under positive selection (note that aBSREL does not test for selection at specific sites). This method allowed the identification of 22 genes with at least one branch showing evidence of episodic positive selection (Fig. 4A). Again, these genes belonged to different functional classes, but those involved in morphogenesis were the most abundant (Table 3). Analysis of selection signals indicated that the longer internal branches of the orthopoxvirus phylogeny were frequent targets of selection. Among terminal branches, the lineage showing the strongest evidence of episodic positive selection was MPXV (Fig. 4A and Table 3). Among the four positively selected genes on this branch, one is the ortholog of VACV A33R, which showed evidence of selection on multiple branches (see below). The other three are orthologs of J1R, J6R, and C3L. Whereas J1R and J6R play roles in

| Gene namea | Functional class | Protein description | Function | Positively selected sitesb | Fraction of disordered residues (%) | Refs. |
|------------|------------------|---------------------|----------|--------------------------|--------------------------------------|-------|
| A8R        | Transcription    | 32 kDa small subunit of transcription factor VITF-3 | Initiates transcription from intermediate gene promoters | Q145, H169 | 4.04 | Sanz and Moss (1999) |
| A20R       | Viral replication/repair | DNA polymerase processivity factor component A20 | Plays an essential role in viral DNA replication | V231, K191, V416 | 1.88 | Ishii and Moss (2001) |
| A24R       | Transcription    | DNA-dependent RNA polymerase subunit rpo132 | Catalytic subunit of the viral RNA polymerase | P605, V1164* | 1.03 | Broyles (2003) |
| A56R       | Cell-to-cell fusion or spread | Hemagglutinin | Inhibits cell-cell fusion and reduces superinfection; interacts with K2 | R96 | 28.14 | Turner and Moyer (2008) |
| B1R        | Viral replication/repair | Serine/threonin kinase | Essential for viral DNA replication; interacts with several cellular proteins | K76, V94, Q296 | 1.95 | Lin et al. (1992), Nichols et al. (2006), Santos et al. (2004, 2006) |
| C12L       | Immunomodulation | Serine protease inhibitor SPI-1 | Plays a role in mediating viral host range; may act to inhibit apoptosis | S125 | 4.61 | Liu et al. (2019), Shi et al. (1999) |
| E3L        | Immunomodulation | Nucleic acid-binding protein | Counters multiple pathways of the innate immune system | V83* | 11.97 | Szczesna et al. (2022) |
| E4L        | Transcription    | RNA polymerase subunit | Transcription of early, intermediate, and late genes | A36, A264*, A194, S428, H564* | 1.16 | Resch et al. (2009) |
| E6R        | Morphogenesis    | Hypothetical protein | Late protein which plays an essential role in virion assembly | 18A | 4.90 | Grimm et al. (2019) |
| E11L       | Transcription    | Putative virion core protein | Part of the viral RNA polymerase complex | V203, V211, V560, S628 | 1.25 | Carpentier et al. (2017), Zhang et al. (2000) |
| F12L       | Morphogenesis    | EEV maturation protein | Virion egress and actin tail formation | V203, V211, V560, S628 | 1.25 | Carpentier et al. (2017), Zhang et al. (2000) |
| F8L        | Unknown          | Hypothetical protein | Protein with iActA-like proline repeats not required for actin tail formation | P20* | 61.45 | Higley and Way (1997) |
| H3L        | Cell entry/attachment | IMV heparin binding surface protein | Might provide virion attachment to target cell | A63, N251, P17 | 7.02 | Lin et al. (2000) |
| H7R        | Morphogenesis    | Hypothetical protein | Formation of crescent membranes | T67 | 5.96 | Meng et al. (2013) |
| K2L        | Cell-to-cell fusion or spread | Serine proteinase inhibitor SPI-3 | Inhibits cell-cell fusion and reduces superinfection; interacts with A56 | A3, R183 | 1.13 | Turner and Moyer (2008) |

*As in VACV Copenhagen nomenclature.

**Position refer to the VACV Copenhagen sequence.

*Residue localized into IDRs (intrinsically disordered regions).
A10L) represent major hubs in this network, whereas only B13 and E3L show abundant interactions with host proteins.

3.4. Protein evolution and intrinsically disordered regions

In viral and non-viral proteomes, protein regions that are intrinsically disordered (intrinsically disordered regions, IDRs)—that is, domains that do not adopt compact three-dimensional structures—are known to be fast evolving (Afanasyeva et al., 2018; Brown et al., 2010; Dyson and Wright, 2005; Hagai et al., 2014; Mozzi et al., 2020b; Nilsson et al., 2011). Previous studies also indicated that, compared with cellular organisms, viral proteomes have a wider variation in IDR fraction, with poxviruses showing a low representation of IDRs (Kumar et al., 2011). IDRs are frequently involved in PPIs, and data on other dsDNA viruses showed that viral proteins that interact with host factors have a higher fraction of disordered regions (Mozzi et al., 2020b). We thus assessed whether the same tendency is observed for orthopoxvirus proteins. Comparisons indicated that orthopox virus proteins that engage in V-H PPIs have a similar fraction of disordered regions as those that do not (Supplementary Fig. 2C). Also, no differences in IDR fraction was evident for V-V PPIs (Supplementary Fig. 2D).

Finally, we investigated whether differences exist in terms of IDR representation across poxvirus genera. We thus retrieved whole proteome information for 22 viruses that are representative of different genera in the Poxviridae family (for orthopoxviruses we used the VACV proteome (Supplementary Table S3). Calculation of the fraction of IDRs revealed significant differences (Kruskall-Wallis p value < 2.2 × 10^-16) with the highest levels observed in molluscipoxviruses and parapoxviruses, the lowest in alphavirus, betavirus, and deltaviruses (Fig. 6A).

Because previous studies showed that the level of protein disorder in viral proteomes is influenced by genome size and base composition (G+C content) (Pushker et al., 2013), we finally correlated the G+C content and length of poxvirus genomes to the average IDR fraction in whole proteomes. Whereas no correlation was observed with genome size, a very strong and positive correlation was detected with G+C content (Spearman’s rank correlation rho = 0.94, p value = 4.11 × 10^-16) (Fig. 6B). This strong relationship may suggest that IDR fraction in poxviruses is simply a secondary effect of genome base composition.

To assess this hypothesis we compared the G+C content in IDRs and in the structured regions of the same poxvirus proteins where IDRs are located. Results indicated that, among low-G+C content genomes IDRs tend to have a higher G+C than structured regions, whereas the opposite situation was observed in some high-G+C content genomes (Fig. 6C). Thus, the fraction of IDRs in poxvirus proteomes is not merely the result of base composition, but most likely represents an actively maintained feature.

4. Discussion

Gene-wise estimates of dN/dS indicated that purifying selection is a major force acting on orthopoxivirus genomes and that the strength of selective constraint is stronger for genes involved in core processes such as viral morphogenesis, transcription, and DNA replication or repair. Despite their stronger constraint, though, some of these genes were also...
targets of pervasive positive selection and most of these were conserved in Orthopoxviruses have greatly impacted the course of human history. Smallpox represented one of the most devastating afflictions of human societies, but also became the first (and so far the only) disease to be eradicated through a worldwide vaccination campaign. The vaccinating agent was another orthopoxvirus, VACV. In the modern world, orthopoxviruses still account for a substantial health burden and fears of accidental or deliberate VARV release prompt ongoing research on the virus. Moreover, orthopoxviruses are actively investigated to develop recombinant vaccines and as oncolytic viruses, in addition to representing common laboratory models to study virus evolution and host-pathogen interactions. However, analysis of the evolution of orthopoxvirus genomes has mainly focused on gene gains/losses (Hatcher et al., 2014; Hendrickson et al., 2010; Senkevich et al., 2021; Upton et al., 2003). Here, we focused on core genes - i.e., those that are conserved in 16 recognized or tentative orthopoxvirus species - and we

Fig. 4. Episodic positive selection in orthopoxviruses. (A) Consensus phylogenetic tree of the 49 genes conserved in all poxviruses. As in Fig. 1, nodes with bootstrap support higher than 0.8 are labeled with a dot. Colors indicate, for all branches, the number of genes that showed evidence of episodic positive selection using aBSREL analysis (see the legend for details). (B) Phylogenetic trees of the four genes for which multiple branches (colored in red) have been found under episodic positive selection.
measured selection using dN/dS, one of the most commonly used metrics. Importantly, the interpretation of dN/dS and its maximum likelihood estimate are based on the assumption that the sequences being compared are sampled from separate populations (Kryazhimskiy and Plotkin, 2008). This is because synonymous and non-synonymous differences among distantly related sequences represent substitutions that have gone to fixation, whereas differences among sequences sampled from a single population represent segregating polymorphisms (Kryazhimskiy and Plotkin, 2008). Thus, although a number of complete genomic sequences are available for different orthopoxviruses, we only analyzed 16 genomes that represent divergent species in the whole Poxviridae family. This situation is reminiscent of the positive selection patterns observed in cytomegaloviruses, whereby core genes, despite being more constrained, also showed signatures of positive selection (Mozi et al., 2020a). That finding was interpreted in terms of concerted evolution of genes that concur to the same process, an explanation that might also hold for orthopoxviruses, as some positively selected genes interact physically or functionally, as demonstrated by their PPIs. It should however be added that at least three of the positively selected genes, A24R, E3L, and H3L, might have evolved in response to selective pressures exerted by the host immune system. As mentioned above, mutations in the A24 protein have been associated with evasion of the PKR system, although the underlying mechanisms are unknown. Brennan and coworkers proposed that mutations that emerged in A24R during in vitro replication might represent an adaptation to evade dsRNA sensors; the identification of A24R as a target of positive selection suggests that such variants confer a fitness advantage not only during experimental evolution in cell culture but also in natural settings. Notably, E3L also functions as an antagonist of PKR, as well as of other innate immune pathways. The positively selected site is located in the N-terminal domain, which contributes to PKR antagonism, but also binds Z-nucleic acids. Major functions of the E3L N-terminus include the inhibition of necroptosis, a caspase-independent form of programmed cell death, and the counteracting the type I IFN system (Szczepan et al., 2022).

With respect to H3L, the gene encodes an immunodominant protein and a major target of neutralizing antibodies. Indeed, antibodies raised against the VACV H3 protein contribute to cross-protection against other orthopoxviruses (Gilchuk et al., 2016). Two of the positively selected sites were located in predicted B-cell epitopes, suggesting that they evolved to elude humoral immune responses in orthopoxvirus hosts.

The relatively low abundance of pervasive positive selection signals and the fact that these mainly involve genes that are not directly involved in immune modulation, is most likely the consequence of different effects. First, the limited number of sequences affords low power to detect positive selection and the power of the dN/dS statistic is further reduced when the majority of sites in the protein evolve under purifying selection (Kryazhimskiy and Plotkin, 2008). Second, ample evidence indicates that variation in gene copy number in the peripheral regions of poxvirus genomes provides an efficient and versatile mechanism to adapt to selective pressures exerted by the host immune response. Thus, the core genome might experience relatively limited immune-mediated pressures, which instead impinge on accessory genes.

Compared to pervasive positive selection, we found more genes to be targeted by episodic positive selection. The long internal branches of the orthopox phylogeny were preferential targets of episodic selection, as was the MPXV lineage. In analogy to CPXV, MPXV is a zoonotic virus with a broad host range. It is very difficult to speculate on the reason(s) why it experienced selection more frequently than other orthopoxviruses. It is however worth noting that the genetic diversity of MPXV in the endemic region is structured into two major clades: one is mainly transmitted in the Congo Basin (clade1), whereas the other, which is composed of two subclades (clades 2a and 2b), is mainly found in West Africa. A higher virulence of clade 1 viruses compared to those in clade 2 was consistently reported (Bunge et al., 2022). One of the genes showing evidence of episodic positive selection (D14L or MOPICE) functions as a modulator of complement activation and it is only encoded by clade 1 viruses (among which the reference genome is classified), as it is deleted in clade 2 genomes (Chen et al., 2005). Experiments with recombinant MPXV revealed that MOPICE is a virulence factor in clade 1 viruses, at

| Gene name | Functional classes | N of branches under selection | Branch under episodic selection | asSRE corrected p value |
|-----------|--------------------|-------------------------------|-------------------------------|------------------------|
| A3L       | morphogenesis      | 1                             | Node 2                        | 0.01393                |
| A4L       | morphogenesis      | 3                             | Node 5                        | 0.00164                |
| A56R      | cell entry/attachment | 1                          | Node 7                        | 0.04439                |
| A60R      | cell-to-cell fusion or spread | 2                        | Node 23                       | 0.00137                |
| B1R       | viral replication/repair | 2                         | ECTV                          | 0.01261                |
| B5R       | cell entry/attachment | 1                         | Node 13                       | 0.01671                |
| B13R      | immunomodulation | 1                             | Node 8                        | 0.00305                |
| D10R      | morphogenesis | 1                             | MPXV                          | 0.00012                |
| D110R     | transcription | 1                             | AKMV                          | 0.01797                |
| D15R      | morphogenesis | 1                             | VPXV                          | 0.02118                |
| F11L      | cell-to-cell fusion or spread | 1                      | Node 5                        | 0.00279                |
| F12L      | morphogenesis | 1                             | AKPV                          | 0.003753               |
| F13L      | morphogenesis | 1                             | Node 2                        | 0.00096                |
| F14L      | transcription | 1                             | MPXV                          | 0.00012                |
| F17R      | morphogenesis | 1                             | VARV                          | 0.00011                |
| F21L      | morphogenesis | 1                             | SPXV                          | 0.00468                |
| F22L      | morphogenesis | 1                             | MPXV                          | 0.003081               |
| F23L      | viral replication/repair | 1                       | Node 7                        | 0.00335                |
| F26L      | transcription | 1                             | MPXV                          | 0.00100                |

* As in VACV Copenhagen nomenclature.
least in prairie dogs, although it is not the only determinant of the lower pathogenicity of clade 2 MPXV (Hudson et al., 2012). Because the mammalian complement system is also fast evolving, due to host-pathogen conflicts (Cagliani et al., 2016), MOPICE might have evolved to adapt to complement components encoded by MPXV natural reservoir(s). Loss of MOPICE in clade 2/3 viruses might be framed within the common phenomenon of orthopoxvirus reductive evolution or adaptive gene loss.

Another intriguing observation about the signals we detected with aBSREL is that two genes showing evidence of selection on more than one branch are involved in some form of intra-genomic conflict. Thus, the B1 kinase antagonizes another poxvirus-encoded protein (B12), which in turn functions as a repressor of viral replication. The A33 protein is instead involved in superinfection exclusion, a phenomenon that allows faster viral spread, but also excludes from the infected cell additional viral genomes, eventually reducing the opportunities for cheating (Leeks et al., 2021). Notably, two genes (K2L and A56R) involved in a distinct mechanism of superinfection exclusion were also found to be targeted by pervasive positive selection, supporting the view that incoming viruses, even belonging to the same species, act as an important selective pressure. Intra-genomic conflicts have been described in a number of organisms (Gardner and Úbeda, 2017). It will be interesting to assess whether they also play a role in the evolution of viral genes, at least in viruses with complex genomes as orthopoxviruses.

In recent years, evidence has accumulated that IDRs represent a dynamic fraction of viral and non-viral proteomes. IDRs tend to be less constrained and are common targets of positive selection (Afanasyeva et al., 2018; Brown et al., 2010; Dyson and Wright, 2005; Hagai et al., 2014; Mozzi et al., 2020b; Nilsson et al., 2011). Our data confirm a recent analysis of viral proteomes by showing that poxviruses display, on average, a low fraction of IDRs (Kumar et al., 2021). In their study, Kumar and co-workers showed that dsDNA viruses that replicated in the cytoplasm (as is the case of poxviruses) have lower IDR fractions than those that replicated in the nucleus. The authors also showed that, in animal viruses, the IDR fraction tends to negatively correlate with proteome size. These observations might partially explain the low IDR representation in poxviruses. It is also worth noting that, in contrast to herpesviruses, with large dsDNA genomes and a higher fraction of IDRs, poxviruses typically cause acute and transient infections. Whether these differences account for different IDR content remains to be clarified.

Fig. 5. Network of protein-protein interactions. Virus-virus and virus-host protein-protein interactions are shown in a network generated with Cytoscape. Human proteins are shown in green and viral proteins in cyan. Viral proteins under pervasive or episodic positive selection are denoted in orange and violet, respectively. Protein-protein interactions have been retrieved from previous works (see methods for details). Proteins identified as positive selection targets are shown only if at least an interactor is known (therefore they are not all present in the figure).
Fig. 6. Level of protein disorder in poxvirus genera. (A) Phylogenetic tree of representative members of 22 genera in the Poxviridae family. The tree was constructed with IQ-TREE v.1.6.12 (Trifinopoulos et al., 2016) using a protein alignment of VACV H4 (RNA polymerase-associated transcription-specificity factor) orthologs. Nodes with bootstrap support higher than 0.8 are labeled with a dot. The fraction of disordered residues (IUPred3 score > 0.5) was calculated for each protein of the representative viral species in 22 poxvirus genera and plotted as boxplots (see Supplementary Table S3 for virus abbreviation). Viral genera are colored accordingly to the legend. Whole proteomes were analyzed. (B) Correlation plot between the mean fraction of disordered residues per gene and the mean G+C percentage for the 22 representative viral species. Colors are as in panel A. (C) Percentage of G+C content in disordered (IDR) and structured (SR) residues. The G+C content (%) in genes having both disordered and structured residues was calculated and displayed as boxplots. Red asterisks denote a statistical difference (*, p value < 0.05; **, p value < 0.01; ***, p value < 0.001) among IDR and SR for each viral proteome. Statistical significance was calculated using Wilcoxon rank sum tests and corrected for multiple testing using the false discover rate (FDR).
Kumar et al., 2021). In any case, our data suggest that, as in other viral and non-viral proteomes, IDRs show dynamic evolution in poxviruses. It should however be mentioned that, although we found that positively selected signals are more frequently detected in IDRs than expected by chance, this result is based on a small number of sites. Likewise, the gene-wise correlation between dN/dS and IDR fraction revealed a significant but weak correlation. Additional analyses, possibly over deeper evolutionary distances, might contribute to shed light into the relationship between evolutionary rates and disorder in poxviruses. Also, in contrast to observations in some herpesviruses, orthopoxvirus proteins that interact with host components are not more disordered than those without known host interactors.

It was previously suggested that disordered regions in viral proteins afford evolutionary plasticity and host adaptation while preserving protein function (Hagai et al., 2014; Mozzi et al., 2020b). In fact, in herpesviruses, a higher fraction of IDRs seems to be associated with a wider host range (Mozzi et al., 2020b). We did not find differences among orthopoxviruses in IDR fraction, despite a large diversity in host wider host range (Mozzi et al., 2020b). We did not find differences in molluscipoxviruses, which are very well protein function (Hagai et al., 2014; Mozzi et al., 2020b). In fact, in contrast to observations in some herpesviruses, orthopoxvirus proteins that interact with host molecules are not significantly more disordered than those that do not, suggest that IDRs do not play a relevant role in host adaptation in poxviruses. Conversely, we found a very strong effect of base composition on IDR fraction. This is partially expected because residues enriched in disordered regions are mainly encoded by G+C rich codons (Basile et al., 2019; Pushker et al., 2013). Clearly, this observation opens the question as to whether the fraction of IDRs is simply a consequence of base composition or if it is an adaptive feature driven by specific viral characteristcs. To indirectly address this question, we thus compared, in the proteins showing evidence of intrinsic disorder, the G+C content of IDRs and of structured regions. Differences were observed in both G+C-poor and in G+C-rich genomes. Although not conclusive, these data suggest that, in poxviruses, the IDR fraction is actively maintained by increasing or decreasing the G+C content to accommodate disorder-promoting codons. Clearly, though, these data have important limitations. The most relevant one is that IDRs were derived from predicted genes and sites we identified should be prioritized in functional analyses.

CRediT authorship contribution statement

Cristian Molteni: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft. Diego Forni: Conceptualization, Formal analysis, Investigation, Visualization, Writing – original draft. Rachele Cagliani: Formal analysis, Writing – review & editing. Alessandra Mozzi: Formal analysis, Visualization. Mario Clerici: Supervision, Writing – review & editing. Manuela Sironi: Conceptualization, Investigation, Supervision, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no competing interests.

Data Availability

NCBI accession IDs are provided in Supplementary Tables 1 and 3.

Acknowledgment

This work was supported by the Italian Ministry of Health (‘Ricerca Corrente 2022’ to MS).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.virusres.2022.198975.

References

Afanas′ayeva, A., Bockwoldt, M., Cooney, C.R., Heiland, I., Gossmann, T.J., 2018. Human long intrinsically disordered protein regions are frequent targets of positive selection. Genome Res. 28, 975–982.

Albaranz, J.D., Ren, H., Torres, A.A., Shemeleva, E.V., Meo, C.A., Bannister, A.J., Bremher, M.P., Chung, B.Y., Smith, G.L., 2022. Molecular mimicry of NF-kB by vaccinia virus protein enables selective inhibition of antiviral responses. Nat. Microbiol. 7, 154–168.

Alonso, R.C., Moura, P.F., Caldeira, D.F., Mendes, M.H., Pinto, M.H., Carmelhutti, J.F., Flores, E.F., de Sant Ana, F.J., 2020. Poxviruses diagnosed in cattle from Distrírito Federal, Brazil (2015–2018). Transbound. Emerg. Dis. 67, 1563–1573.

Anisimova, M., Bielaszewski, J.P., Yang, Z., 2002. Accuracy and power of bayes prediction of amino acid sites under positive selection. Mol. Biol. Evol. 19, 950–958.

Archard, L.C., Mackett, M., 1981. Restriction endonuclease analysis of red cowpox virus and its white pock variant. J. Gen. Virol. 45, 51–63.

Basile, W., Salvatore, M., Basset, C., Elossohn, A., 2019. Why do eukaryotic proteins contain more intrinsically disordered regions? PLoS Comput. Biol. 15, e1007186.

Bay, R.A., Bielaszewski, J.P., 2007. Reconstructing ancestral protein and evolutionary scenarios relevant to functional divergence. J. Mol. Evol. 73, 273–286.

Beer, E.M., Rao, V.R., 2019. A systematic review of the epidemiology of human monkeypox outbreaks and implications for outbreak strategy. PLoS Negl. Trop. Dis. 13, e007971.

Brandt, T., Heck, M.C., Vijaysri, S., Jentzarla, G.M., Cameron, J.M., Jacobs, B.L., 2005. The N-terminal domain of the vaccinia virus E3L-protein is required for neurovirulence, but not induction of a protective immune response. Virology 333, 263–276.

Bratke, K.A., McLysaght, A., Rothenburg, S., 2013. A survey of host range genes in poxvirus genomes. Infect. Genet. Evol. 14, 406–425.

Brenneman, D.G., Kizerman, J.O., Shendure, J., Geballe, A.P., 2015. Experimental evolution identifies vaccinia virus mutations in A24R and A15R that antagonize the protein kinase R pathway and accompany collapse of an extragenic gene amplification. J. Virol. 89, 9986–9997.

Brown, C.J., Johnson, A.K., Daughdril, G.W., 2010. Comparing models of evolution for ordered and disordered proteins. Mol. Biol. Evol. 27, 609–621.

Broyles, S.S., 2003. Vaccinia virus transcription. J. Gen. Virol. 84, 2293–2300.

Bunge, E.M., Hoet, B., Chen, L., Lienert, F., Weidenhaller, H., Baer, L.H., Steffen, R., 2003. The changing epidemiology of human monkeypox: A potential threat? A systematic review. PLoS Negl. Trop. Dis. 6, e010141.

Burles, K., Irwin, C.R., Burton, R.L., Schriever, J., Evans, D.H., Buller, R.M., Barry, M., 2014. Initial characterization of vaccinia virus B4 suggests a role in virus spread. Virology 456-457, 108–120.

Caglioni, R., Forni, D., Filippi, G., Mozzi, A., De Gioia, L., Pontremoli, C., Pozzoli, U., Bresolin, N., Clerici, M., Sironi, M., 2016. The mammalian complement system as an innate immune system. Mol. Immunol. 68, 153–157.

Carpentier, D.C.J., Van Loggerenberg, A., Dieckmann, N.M.G., Smith, G.L., 2017. The changing epidemiology of human monkeypox-A potential threat? A systematic review. J. Gen. Virol. 98, 1500–1509.

Carpentier, D.C.J., Van Loggerenberg, A., Dieckmann, N.M.G., Smith, G.L., 2017. The changing epidemiology of human monkeypox-A potential threat? A systematic review. J. Gen. Virol. 98, 1500–1509.

Cresawn, S.G., Prins, C., Latner, D.R., Condit, R.C., 2007. Mapping and phenotypic analysis of spontaneous isatin-beta-thiosemicarbazone resistant mutants of vaccinia virus. Virology 363, 319–332.

da Fonseca, F.G., Wolffe, E.J., Weisberg, A., Moss, B., 2000. Effects of deletion or repression of the H3L envelope gene on vaccinia virus replication. J. Virol. 74, 1339–1348.

Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T., 2002. Mauve: multiple alignment of consensus genomic sequence with rearrangements. Genome Res. 14, 1394–1403.
Domingo E, Rodrigo AM, 2008. Evolution of RNA viruses. Cold Spring Harb Perspect Biol 4(2):a009055.

Domingo E, Butch WJ, 2018. The dynamics of RNA virus evolution. Annu Rev Virol 5:331–352.

Domingo E, Peeters M, 2011. Overview of virus evolution. Philos Trans R Soc Lond B Biol Sci 366(1573):1153–1167.

Domingo E, 2016. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 7(1):a026189.

Domingo E, 2019. Virus evolution. Annu Rev Virol 6:209–232.

Domingo E, 2020. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 11(2):a026189.

Domingo E, 2021. Virus evolution. Annu Rev Virol 8:211–233.

Domingo E, 2022. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 13(2):a026189.

Domingo E, 2023. Virus evolution. Annu Rev Virol 10:211–233.

Domingo E, 2024. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 15(2):a026189.

Domingo E, 2025. Virus evolution. Annu Rev Virol 12:211–233.

Domingo E, 2026. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 16(2):a026189.

Domingo E, 2027. Virus evolution. Annu Rev Virol 14:211–233.

Domingo E, 2028. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 17(2):a026189.

Domingo E, 2029. Virus evolution. Annu Rev Virol 16:211–233.

Domingo E, 2030. Antivirus and antiviral therapy. Cold Spring Harb Perspect Med 18(2):a026189.
