

Comparative analysis of deep mutational scanning datasets in enteroviruses A and B identifies functional divergence and therapeutic targets

Received: 28 August 2025

Accepted: 7 January 2026

Published online: 23 February 2026

 Check for updates

Beatriz Álvarez-Rodríguez^{1,4}, William Bakhache^{2,4}, Lauren McCormick^{2,3}, Ron Geller¹✉ & Patrick T. Dolan²✉

Deep mutational scanning (DMS) can define functional constraints acting on viral proteomes by quantifying the effects of mutations on viral fitness. However, DMS analyses do not discern type-specific from species-level constraints, limiting their utility in understanding how selective pressures change as viral families diversify. Here we show that comparison of DMS datasets from related viruses can overcome these limitations. By contrasting two proteome-wide DMS datasets from prototypical members of the enterovirus A and B species, we identify evolutionary constraints at the species level to occur across core enzymatic machinery and capsid assembly interfaces. In contrast, type-level constraints are observed across host-interaction sites in both structural and non-structural proteins. Furthermore, we find DMS data to reflect both type- and species-level evolutionary signatures in nature yet diverge at conserved hotspots subjected to selection pressures that are lacking in vitro. Finally, we highlight the utility of comparative DMS studies for drug discovery by identifying a mutationally constrained pocket in the 2C helicase that is conserved across all major human enterovirus species. Our findings provide a framework for dissecting evolutionary pressures acting at different evolutionary scales and for guiding the rational design of broad-spectrum therapeutics with high barriers to resistance.

Enteroviruses constitute a diverse genus of RNA viruses in the *Picornaviridae* family and include several species that pose important and recurring public health challenges globally. Within the genus *Enterovirus* are several species that infect humans. These include the human enterovirus (HEV) A–D species, the rhinoviruses A–C and several parechoviruses. HEV species also exhibit regional patterns of circulation, with Asian countries experiencing outbreaks of *Enterovirus A*, the USA affected by *Enterovirus D* and Europe commonly affected by *Enterovirus B*¹.

Initially classified by immune cross-reactivity between their capsids, and now classified genetically by capsid sequence, the HEV species are genetically diverse, with similarity in their polyprotein coding regions ranging from 50% to 60%. This diversity extends to their pathogenicity and clinical outcomes. HEV infections, which commonly affect infants, young children and the elderly², have broad tissue tropism, infecting the skin, respiratory and gastrointestinal tissues, as well as the cardiovascular and central nervous systems. *Enterovirus A*, *Enterovirus*

¹Institute for Integrative Systems Biology (I2SysBio), Universitat de Valencia-CSIC, Valencia, Spain. ²Quantitative Virology and Evolution Unit, Laboratory of Viral Diseases, NIH–NIAID Division of Intramural Research, Bethesda, MD, USA. ³Department of Biology, University of Oxford, Oxford, UK. ⁴These authors contributed equally: Beatriz Álvarez-Rodríguez, William Bakhache. ✉e-mail: ron.geller@csic.es; Patrick.Dolan@nih.gov

C and *Enterovirus D* (for example, EVA71, poliovirus and EVD68) cause severe central nervous system complications such as acute flaccid myelitis and brain-stem encephalitis, which can result in paralysis and death³. *Enterovirus B* (for example, coxsackievirus B3 (CVB3)) infection is linked to myocarditis, which can lead to cardiomyopathy and heart failure^{4–6}, and pancreatic cancer development⁷.

Enteroviruses possess a single-stranded RNA genome of positive polarity ~7.5 kb long that is enclosed in a ~30-nm icosahedral particle. The viral genome is translated as a single polyprotein of ~2,150 amino acids (AAs). The polyprotein is cleaved by the viral 2A protease into a structural precursor protein (P1) that encodes the capsid proteins and a non-structural precursor (P2 and P3) that encodes the replication machinery. The P2–P3 precursor is processed further by the 3C viral protease into seven proteins (2A–2C and 3A–3D) that perform essential enzymatic and membrane remodelling activities required for viral replication^{8–10} and antagonize the innate immune system^{11–13}. The P1 precursor is cleaved by the 3CD protease to liberate three capsid proteins, VP0, VP1 and VP3 (ref. 8). These undergo sequential assembly to form the virus particle, starting as a heterotrimer (the ‘protomer’), five of which assemble to form the pentamer, and 12 pentamers form the viral capsid¹⁴. During maturation, VP0 undergoes RNA-dependent autocatalytic cleavage into VP2 and VP4 to produce infectious virions. The capsid proteins of enteroviruses are highly divergent in both sequence and function, engaging with diverse host attachment and entry factors¹⁵. Evolutionary pressure placed on them by adaptive immunity has also driven their diversification, primarily through substitutions in surface-exposed loops¹⁶. Overall, mutations affecting structural and non-structural proteins both contribute to the diversity of phenotypic traits observed in enteroviruses, with single non-synonymous changes having profound effects on pathogenesis^{17–20}. Understanding how mutations alter viral fitness and host interaction is thus of key importance for understanding the biology and pathogenesis of these viruses.

Except for poliovirus and a regionally approved vaccine for EVA71 in Asia, no broadly licensed vaccines exist for enteroviruses, and vaccine development is hindered by the high type diversity for each enterovirus species²¹. Similarly, while potent antivirals have been developed (for example, Rupintrivir^{22,23}), the high mutation rates²⁴ of RNA viruses make antiviral resistance a major obstacle to effective therapy. The severe clinical outcomes of enterovirus infections and the lack of effective interventions underscore the urgent need for diverse therapeutic strategies. In this regard, a deeper understanding of the evolutionary mechanisms governing enterovirus fitness is critical to inform rational vaccine design and the development of antivirals with high resistance barriers.

Deep mutational scanning (DMS) is a high-throughput experimental and computational framework that enables the measurement of mutational fitness effects (MFEs) for thousands of mutations in parallel. In the context of virology, DMS has been applied across different viral families to uncover functional constraints that affect viral fitness, interaction with host factors and evasion of antibody neutralization^{25–28}. As the majority of DMS studies have performed DMS on individual proteins, a direct comparison of MFEs across proteins within a viral proteome is not possible. Furthermore, the lack of DMS data for related viruses has also prevented dissecting virus-specific constraints from evolutionarily conserved pressures on a given viral family.

Here we leverage two full-proteome DMS studies of enteroviruses EVA71 and CVB3 (refs. 29,30) to develop a unified comparative framework that distinguishes between type-specific evolutionary pressures and the fundamental constraints shared by enteroviruses. Our approach identifies regions of convergent and divergent selection by directly comparing the MFEs of the viral proteomes. We then integrate these MFEs with structural analysis, showing that convergent constraints localize to core functional regions, including sites of capsid assembly and enzymatic activity, while divergent constraints map to distinct host-factor interactions. To contextualize our findings within

a real-world evolutionary framework, we compare DMS data to natural sequence variation and uncover distinct selective pressures operating in natural infections that are not captured in *in vitro* experimental systems. Finally, we integrate our results with computational prediction of druggable pockets to identify a pan-enterovirus-conserved, mutationally constrained pocket in the viral 2C helicase, highlighting it as a high-potential target for broad-spectrum antiviral development. As more DMS studies become available, the described analysis framework can be applied to additional viral families to uncover general functional and evolutionary constraints, define sites of interaction with host proteins and guide the rational design of broad-spectrum therapeutics.

Results

Comparative analysis of DMS data from prototypical members of enterovirus A and B species

To compare mutational tolerance across enterovirus species, we assessed global sequence similarity and phylogenetic relationships across the proteomes of the two prototype viruses used in this study: EVA71 (an *Enterovirus A* species) and CVB3 (an *Enterovirus B* species) (Fig. 1a). These viruses exhibit substantial divergence at the AA level, with an overall identity of ~55% across the full proteome (Fig. 1b). The capsid region (P1) is the most divergent, showing only ~43% identity between EVA71 and CVB3, whereas the non-structural region (P2–P3) shares higher similarity (~63% identity), with particularly high conservation in the 2A protease and the 3D RNA-dependent RNA polymerase.

This pattern of sequence divergence is recapitulated in phylogenetic analyses. Due to the known pervasiveness of recombination throughout the history of enteroviruses, we constructed trees using both capsid and 3D polymerase sequences to capture any differences in evolutionary relationships between these regions of the genome. In the tree constructed from capsid sequences ($n = 15$), EVA71 and CVB3 form distant branches, consistent with strong antigenic and functional diversification (Fig. 1c). In contrast, CVB3 and EVA71 cluster more closely in the tree based on 3D polymerase sequences ($n = 15$), reflecting stronger evolutionary conservation of replicative functions (Fig. 1d). This pattern extends beyond EVA71 and CVB3. Across different enterovirus species, capsid proteins demonstrate greater sequence divergence than non-structural proteins (Supplementary Fig. 2a), highlighting the distinct pressures at play on different functional regions of the proteome across various evolutionary timescales.

To dissect the shared and species-specific constraints shaping HEV evolution, we compared proteome-wide, site-specific mutational tolerance from the two published enterovirus DMS datasets, which encompass >80,000 mutations across the 11 proteins of each virus^{29,30}, containing nearly all possible single AA substitutions in the EVA71 and CVB3 genomes. Single codon deletions were also included across the full proteome of both viruses, except in the structural proteins of CVB3. These libraries were used to generate viral populations, which were subsequently subjected to two rounds of infection at low multiplicities of infection to minimize complementation and enable fitness-based selection. Mutation frequencies were then quantified by next-generation sequencing in both the plasmid libraries and the passed viral populations, as described in the original studies^{29,30}.

To enable direct comparison, we standardized both datasets using a common analysis workflow (Methods and Supplementary Fig. 1). Specifically, we quantified the relative enrichment of each mutation at each site by dividing its frequency by that of the corresponding wild-type (WT) codon. MFEs were then calculated as the ratio of relative enrichment after selection (that is, in viral populations) to that before selection (that is, in plasmid libraries). The resulting values were \log_2 -transformed and normalized such that the mean MFE of synonymous mutations was set to neutral (Methods).

Comparison of the standardized datasets revealed striking similarities between the two DMS datasets despite the divergence of EVA and EVB (Fig. 1e and Supplementary Fig. 2b,c), as evidenced by the

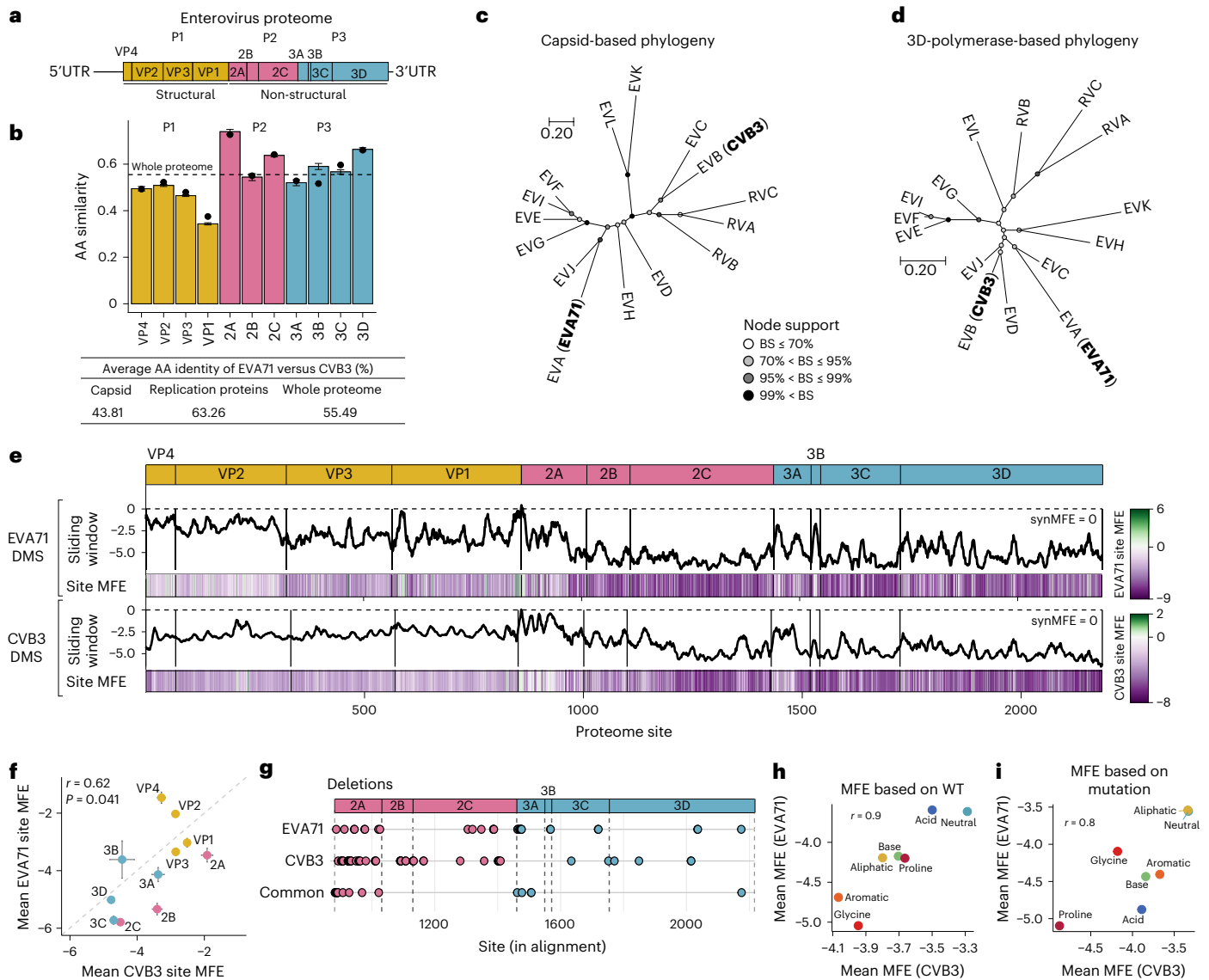


Fig. 1 | Overview of viral systems and DMS datasets. **a**, Schematic of the enterovirus genome. UTR, untranslated region. **b**, Per-protein similarity at the AA level between EVA71 and CVB3 ($n = 325$ and $n = 49$, respectively). **c, d**, Phylogenetic trees of all known enterovirus species, illustrating sequence divergence across structural coding regions (**c**) and the 3D polymerase coding region. Nodes are shaded by bootstrap support levels (BS) shown in legend (**d**). The tip labels indicate enterovirus species. **e**, Heat maps of site MFEs across EVA71 and CVB3 proteins with line plots showing a five-residue sliding window average. **f**, Correlation of

mean site MFE between EVA71 and CVB3 across proteins. The error bars indicate the standard error of the mean across all positions of each protein. The dashed line represents a 1:1 relationship between both axes. **g**, Genomic distribution of tolerated deletions unique to EVA71 or CVB3 or shared between the two viruses (common). **h**, Correlation of mean MFE based on the nature of the WT residue across AA physicochemical categories between the EVA71 and CVB3 DMS datasets. **i**, Correlation of mean MFE based on the nature of the mutation across AA physicochemical categories between the EVA71 and CVB3 DMS datasets.

strong correlation of the average MFE per site (site MFEs) between the two viruses (Pearson's $r = 0.66$, $P < 2.2 \times 10^{-16}$; see Fig. 1f for per-protein MFE comparison). Specifically, capsid proteins exhibit significantly greater mutational tolerance than non-structural proteins across both viruses. Among the non-structural proteins, 2A and 3A are the most permissive. We also analysed tolerance to single codon deletions, albeit only for the non-structural region, as deletion data are unavailable for the CVB3 capsid²⁹ (Supplementary Fig. 1). Overlapping deletion-tolerant sites between the two viruses in the 2A and 3A proteins were observed, further supporting the mutational tolerance of these two proteins (Fig. 1g).

The extensive MFE datasets enabled the examination of trends in structural and sequence composition that govern global mutational tolerance. Across both viruses, secondary structure context strongly influences mutational constraint, with α -helices being the

most constrained structural elements, loops exhibiting the greatest mutational tolerance and β -strands displaying intermediate tolerance levels (Supplementary Fig. 2d,e). We also observed strong similarities in mutation tolerance based on the nature of the WT or mutant AA between the datasets when comparing the average MFEs per WT AA class (aromatic, acidic and so on) (Pearson's $r = 0.9$, $P = 0.0056$; Fig. 1h). Glycine and aromatic residues are least tolerant to substitutions, reflecting their unique structural roles, whereas neutral AAs demonstrate the greatest mutational flexibility. Similarly, MFEs based on the nature of the mutant AA are strongly correlated between both viruses (Pearson's $r = 0.8$, $P = 0.034$), with substitutions introducing aliphatic and neutral residues being well tolerated, while mutations introducing proline or charged residues are predominantly deleterious (Fig. 1i). Mutations that maintain physicochemical similarity are consistently the most tolerated (Supplementary Fig. 2f,g). Interestingly, the impact

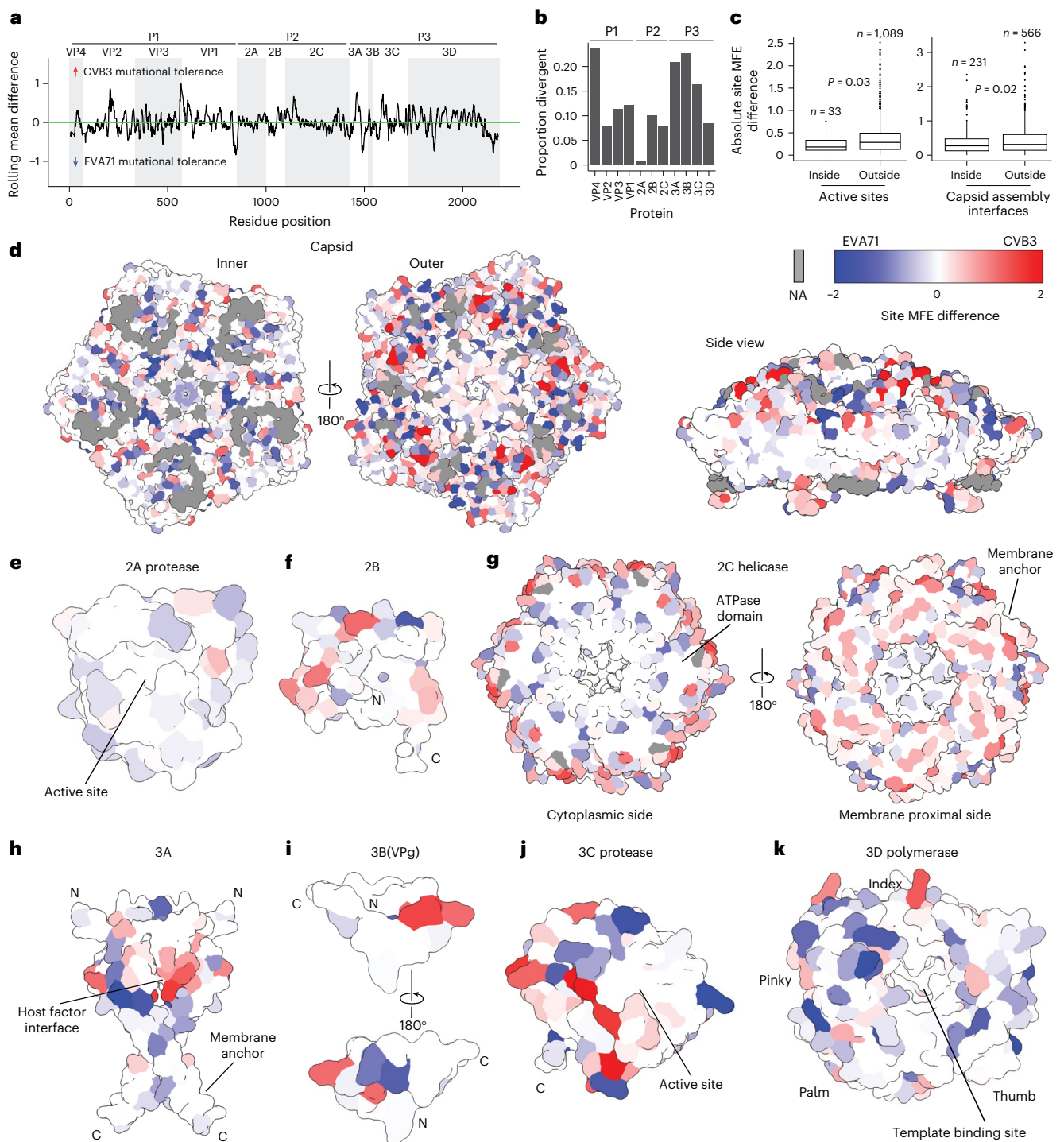


Fig. 2 | Structural mapping of enterovirus A and B differential mutational tolerance. **a**, Site difference in MFE scores (MFE in CVB3 minus MFE in EVA71) across the aligned viral proteomes, averaged using a ten-residue sliding window. Positive values indicate higher mutational tolerance in CVB3, while negative values indicate higher tolerance in EVA71. A horizontal green line at $y = 0$ denotes similar tolerance across both viruses. **b**, Proportion of divergent sites for each viral protein. Divergent sites are defined as those with an absolute site MFE difference greater than the median plus two median absolute deviations (threshold, 0.86). Proportions are calculated relative to the total number of aligned residues per protein. **c**, Distribution of absolute site MFE difference for sites located inside versus outside active sites and capsid assembly interfaces. In each box plot, the lower and upper hinges correspond to the 25th and 75th percentiles, respectively, with the centre line indicating the median. The whiskers extend to the most extreme data points within $1.5 \times$ the interquartile

range (IQR) from the hinges. Observations falling outside this range are treated as outliers and are plotted individually. Statistical significance was calculated using a two-sided Wilcoxon test. **d–k**, Structural mapping of site MFE difference on the structure of the capsid pentamer (Protein Data Bank (PDB): [8E2X](#)) (**d**), 2A protease (**e**), 2B protein (**f**), 2C helicase (hexameric form) (**g**), 3A protein (dimeric form) (**h**), 3B protein (VPg) (**i**), 3C protease (**j**) and 3D polymerase (**k**). All models are based on the EVA71 proteins (from experimental structures or AlphaFold3 predictions), which serve as a representative scaffold. Surfaces are coloured by the site MFE difference according to the key: red indicates higher tolerance in CVB3, blue indicates higher tolerance in EVA71 and white represents sites under convergent constraint (similar MFE scores). Key functional regions are labelled, and grey denotes positions with missing data due to structural alignment gaps (NA, not available).

of AA substitutions depends not only on the nature of the change but also on its directionality, reflecting fundamental constraints shaped by biochemical compatibility across proteins (Supplementary Fig. 2f,g).

A structural map of evolutionary constraints in EVA71 and CVB3

To place the evolutionary constraints between EVA71 and CVB3 in a functional context, we calculated the site-wise differences in mean MFE and defined divergent sites as those exceeding the median plus two median absolute deviations (Methods and Fig. 2a). This reveals that VP4, 3A and 3B have the greatest proportion of sites with divergent MFEs between the two viruses. In contrast, the 2A protease shows the lowest proportion of divergent sites, indicating similar mutational flexibility in both viruses (Fig. 2b).

To understand the structural basis for these differences, we mapped site-wise divergence values onto the viral protein structures. As expected from the high structural and functional conservation across the two proteomes, sites in enzymatic active sites or those participating in capsid assembly are significantly more conserved than sites outside of these regions (Fig. 2c and Supplementary Fig. 3a,b). This observation provides an important internal validation that independent experimental approaches produced consistent and comparable MFE measurements. Regions of convergent constraint correspond to common essential functions, including the interfaces between capsid pentamers important for virion assembly (Fig. 2d, side view), the catalytic sites of the 2A and 3C proteases (Fig. 2e,j), the ATPase domain of the 2C helicase (Fig. 2g, cytoplasmic view) and the template-binding channel of the 3D polymerase (Fig. 2k). In contrast, regions mediating host-factor interactions are dominated by divergent mutational tolerance. This is most evident on the exterior surface of the viral capsid, where large patches corresponding to surface-exposed loops containing virus-specific determinants for receptor binding^{20,31} are substantially more tolerant to mutation in one virus than in the other (Fig. 2d, outer view). Moreover, the mutational flexibility of surface loops may also be shaped by type-specific host immune pressures, as previously described for CVB3 (ref. 16), although this possibility remains to be validated in EVA71. A similar pattern of divergence occurs in the amino-terminal region of the 3A protein, which harbours a known host-factor binding interface^{32,33} (Fig. 2h).

Divergent mutational tolerance at the receptor-binding footprints of enterovirus capsids

Given that EVA71 and CVB3 use different host entry factors, Scavenger receptor class B member 2 (SCARB2)^{31,34} for EVA71 and Coxsackie and Adenovirus Receptor (CAR) and Decay-Accelerating Factor (DAF; CD55) for CVB3 (ref. 20), we hypothesized that structural divergence on the capsid surface would map to these receptor-binding footprints. Indeed, our analysis shows that site-wise MFE divergence is significantly enriched within these footprints compared with other capsid regions (Fig. 3a and Supplementary Fig. 3c). Notably, ~60% of the most divergent sites (top 1%) are located within a receptor-binding region, making a highly divergent site nearly 20 times more likely to be found in this region (Fig. 3b). This pattern reflects a functional constraint in which interaction interfaces with a cognate receptor limit the mutation tolerance of an otherwise variable region in viruses employing other entry factors (Fig. 3c,d).

To visualize these results, we mapped the site MFE differences onto a structure of the EVA71 capsid in complex with its receptor, SCARB2 (PDB: 6I2K). The structure shows SCARB2 binding to the southern rim of the viral canyon through contacts in the VP1 GH and VP2 EF loops (Fig. 3e). These specific receptor-binding sites are under strong mutational constraint in EVA71 but not in CVB3 (Fig. 3c,d). This divergence is driven by the residues Gly137–Gly138, Glu142 and Phe158 in the VP2 EF loop (Fig. 3f,g). In contrast, the capsid assembly interfaces are under strong convergent constraint (white regions) (Fig. 3f,g).

In the reciprocal analysis, we mapped the site MFE differences onto the structure of CVB3 in complex with its principal entry receptor, CAR (PDB: 7VYK). CAR binds within the CVB3 canyon at an interface involving the VP1, VP2 and VP3 capsid proteins (Fig. 3h). The CVB3 VP2 CAR-binding region involves an inserted sequence relative to EVA71 that accommodates CAR binding. Consequently, comparing constraints at all sites in the footprint is not possible. Nonetheless, the canyon region generally exhibits strong mutational constraint, and we observe clear divergence at the CAR contact residues Val150 in VP1 and Asp182 in VP3, which are under stronger mutational constraint in CVB3 than in EVA71 (Fig. 3i,j).

Counterintuitively, a more complex pattern emerged when we analysed the interaction sites for the CVB3 attachment factor, DAF (PDB: 7VY5). Some residues interacting with DAF show higher mutational tolerance in CVB3 than in EVA71 (Fig. 3k–m). This unexpected tolerance is particularly notable at residue Gln234 in VP3. Interestingly, this residue has been identified as a key molecular switch that modulates the ability of the virus to shift from DAF-dependent to DAF-independent entry pathways^{20,29}, providing a potential explanation. Alternatively, since CAR expression is sufficient to mediate viral entry into HeLa-H1 cells³⁵, DAF dependence may be reduced, alleviating constraints pertaining to the use of this coreceptor.

Membrane and host-factor interactions drive enterovirus 3A mutational divergence

The divergent mutational tolerance we observed between the EVA71 and CVB3 3A proteins suggests that they engage host factors through distinct interfaces (Fig. 4a). The 3A proteins of both viruses bind host membranes to facilitate the formation of viral replication organelles¹⁰. Additionally, CVB3 3A blocks protein trafficking in the cell via interaction with GBF1 (ref. 36), while EVA71 does not, despite also interacting with it³⁷. Hence, we hypothesized that distinct interactions with both the lipid membrane and the host protein GBF1 could explain these differences.

We used DeepTMHMM³⁸ to predict the topology of the 3A proteins to investigate the 3A membrane-spanning region. The software predicts a single α -helical transmembrane helix for both proteins, but with different lengths of the transmembrane region, suggesting that CVB3 3A inserts deeper into the membrane than EVA71 3A. This prediction aligns with our DMS data, which show that the membrane-embedded region in CVB3 is mutationally constrained, whereas the corresponding region in EVA71 3A is predicted to lie outside the membrane and is tolerant to substitutions (Fig. 4b,c and Supplementary Fig. 4a,b). This suggests that distinct lipid interactions impose unique pressures on 3A in each virus. However, further work is needed to interpret the observed differences in selection.

Next, to examine potential differences in the interaction with GBF1, we used AlphaFold3 to model the 3A dimer from both viruses in complex with the complete form of this host factor (Fig. 4d,e and Supplementary Fig. 4c). Again, our models predict a difference in this virus–host interaction interface. Although these predictions exhibit low interface predicted template modeling (ipTM) scores (–0.2), the predicted interfaces correspond to regions of differential selection in EVA71 and CVB3. Recent work has shown that virus–host protein interactions are often associated with similarly low ipTM scores³⁹. In EVA71, GBF1 makes contact with both the α 1 and α 2 helices of the 3A dimer. In contrast, the corresponding residues on the CVB3 3A dimer are oriented away from GBF1 and not involved in the binding interface. This region also exhibits greater mutational tolerance in CVB3, consistent with the distinct predicted binding modes in CVB3 and EVA71 3A (Fig. 4f,g). Moreover, these findings are also congruent with published biochemical data and computational work suggesting the importance of the N-terminal region in mediating GBF1 interaction and COP-I transport^{32,33,40}. While our predictions explain MFE differences between EVA71 and CVB3 3A, additional structural studies are needed to map the distinct interaction surfaces between enterovirus 3A proteins and GBF1.

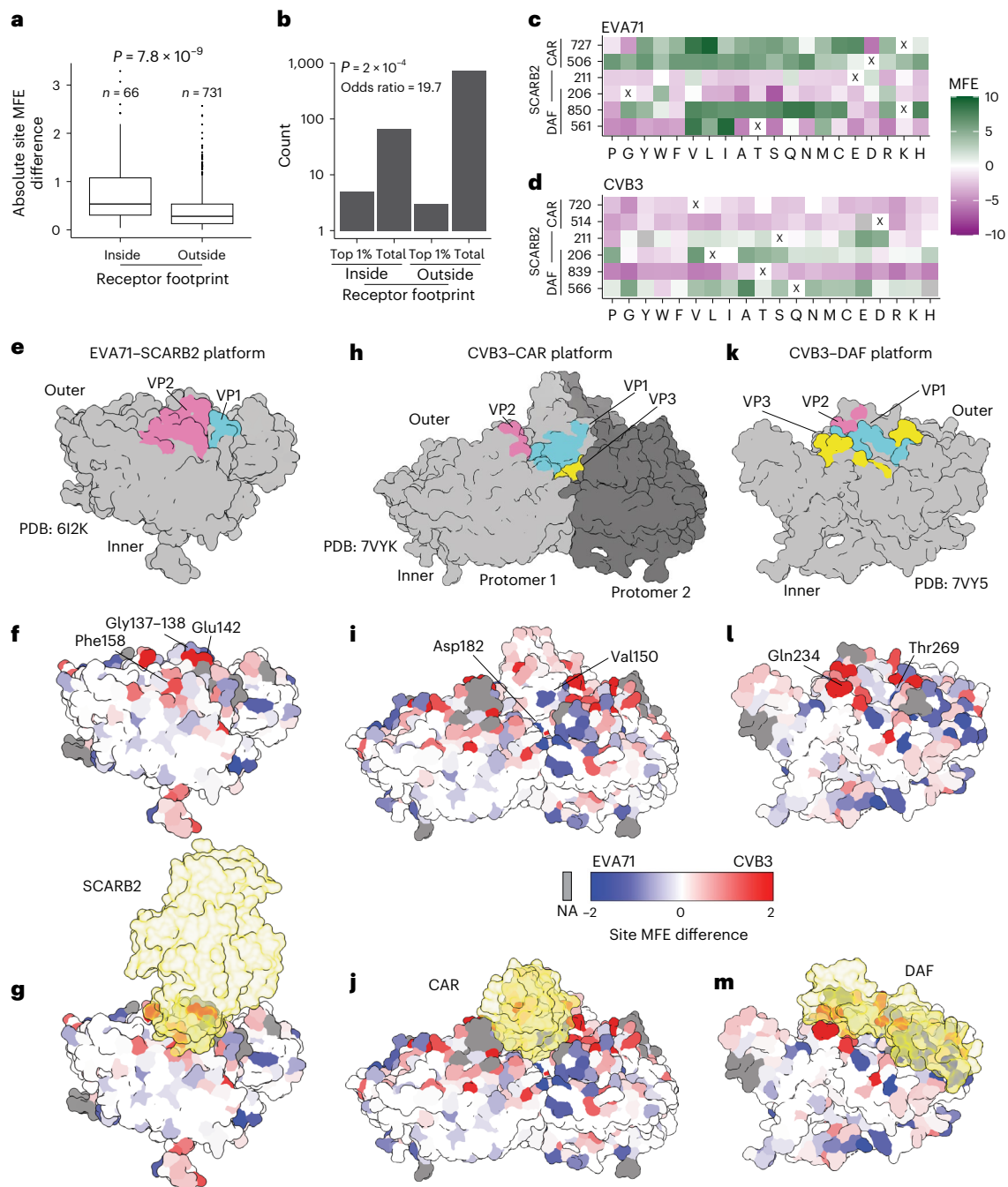


Fig. 3 | Divergence in mutational tolerance at the EVA71 and CVB3 receptor-binding footprints. **a**, Distribution of absolute site MFE difference for sites located inside versus outside receptor footprints. In each box plot, the lower and upper hinges correspond to the 25th and 75th percentiles, respectively, with the centre line indicating the median. The whiskers extend to the most extreme data points within $1.5 \times$ the IQR from the hinges. Observations falling outside this range are treated as outliers and are plotted individually. Statistical significance was calculated using a two-sided Wilcoxon test. **b**, The total number of analysed sites alongside the count of the most divergent sites (top 1% by MFE difference), categorized by their location inside or outside the receptor-binding interface. Statistical significance was calculated using a two-sided Fisher's exact test.

c, d, Heat maps displaying the EVA71 (**c**) and CVB3 (**d**) MFE scores for divergent residues in receptor contact sites. The 'X' symbols indicate WT AA positions, and grey squares represent missing data. **e–m**, Surface representations of EVA71 in complex with its receptor SCARB2 (PDB: 6I2K) (**e–g**) and CVB3 with its receptors CAR (PDB: 7VYK) (**h–j**) and DAF (PDB: 7VY5) (**k–m**). In **e, h** and **k**, contact loops are highlighted: VP1 (cyan), VP2 (pink) and VP3 (yellow). In **f, g, i, j, l** and **m**, the surface is coloured by the site MFE difference between EVA71 and CVB3 according to the key: red indicates higher mutational tolerance in CVB3, while blue indicates higher tolerance in EVA71. White denotes sites of convergent constraint. Grey denotes positions with missing data due to structural alignment gaps. In **g, j** and **m**, the receptors are shown as transparent yellow surface representations.

Comparison of laboratory and natural selection pressures

We next assessed how the observed trends in mutational tolerance in EVA71 and CVB3 relate to natural sequence variation. We used phydms⁴¹ to incorporate experimentally measured MFEs into phylogenetic models derived from natural sequence alignments across two evolutionary

scales, type (EVA71 and CVB3) and species (*Enterovirus A* and *Enterovirus B*). In all cases, phylogenetic models that incorporate MFE information outperform standard models (YNGKP/Goldman–Yang) (Supplementary Table 1), suggesting DMS-derived MFEs reflect selection processes operating in nature across both evolutionary timescales.

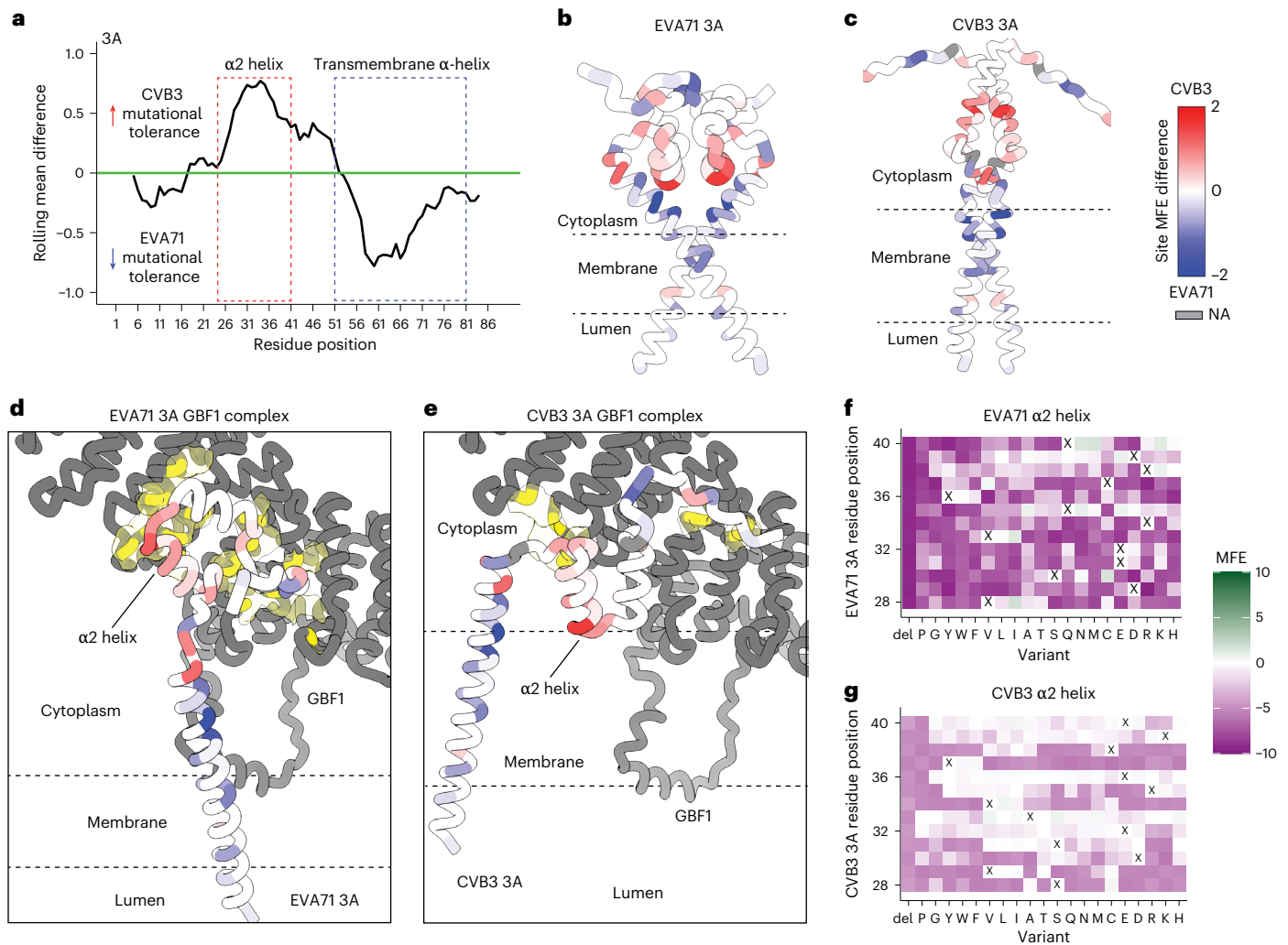


Fig. 4 | Mapping of the divergent structural and functional regions in enterovirus 3A. **a**, Site difference in MFE scores (MFE in CVB3 minus MFE in EVA71) across the aligned 3A proteins, averaged using a ten-residue sliding window. **b, c**, AlphaFold3-predicted structures of the EVA71 (**b**) and CVB3 (**c**) 3A homodimers. Residues are coloured by site MFE difference according to the colour key. **d, e**, AlphaFold3-predicted structural models of the EVA71 (**d**) and CVB3 (**e**) 3A dimers in complex with the host factor GBF1. 3A residues are coloured by site MFE difference according to the colour key, and GBF1 is shown as a grey surface, with its 3A contact residues highlighted in yellow. EVA71 3A–GBF1 prediction: pTM = 0.60, ipTM = 0.20, ranking score = 0.45; CVB3 3A–GBF1 prediction:

pTM = 0.59, ipTM = 0.19, ranking score = 0.44. **f, g**, Heat maps displaying MFE scores for the specific set of residues that are mutationally divergent and that constitute the EVA71 3A–GBF1 binding interface. Data are shown for EVA71 (**f**) and CVB3 (**g**). Each ‘X’ indicates the WT AA position. For all relevant panels, the predicted protein topology (cytoplasm, membrane, lumen), as determined by DeepTMHMM, is indicated. For all structural views, a common colour key is used for the MFE difference: red indicates higher tolerance in CVB3, blue indicates higher tolerance in EVA71 and white denotes sites of convergent constraint. Grey denotes positions with missing data due to structural alignment gaps.

We next analysed the correlation between MFEs and variation in natural sequence alignments across both evolutionary scales, calculated as Shannon entropy (*H*; Fig. 5a). This revealed that DMS-derived mutational tolerance correlates more strongly with site entropy at the species level (Pearson’s *r* = 0.52, *P* = 6.84 × 10⁻¹⁵⁰; *r* = 0.46, *P* = 3.99 × 10⁻¹¹⁴; for EVA and EVB, respectively) than at the type level (Pearson’s *r* = 0.24, *P* = 7.49 × 10⁻³¹; *r* = 0.3, *P* = 2.34 × 10⁻⁴⁶; for EVA71 and CVB3, respectively; Fig. 5b). This correlation is primarily driven by differences in the variability of the structural proteins, which show strong conservation at the type level but high variability at the species level (Supplementary Fig. 5a,b). Moreover, DMS-derived MFEs correlate more strongly with site entropy from other enterovirus species than with variation within their own types (Supplementary Fig. 5e,f), suggesting DMS captures broader patterns of conserved capsid constraints operating at longer evolutionary timescales.

To further identify differences between selection pressures in the lab and those in nature, we examined whether the same AAs were

preferred across each site between lab-derived MFEs and natural sequences at both the type and species level. For this, we used phydms to calculate differential preference scores for each site, with larger values indicating greater discordance between AAs preferred in DMS and those preferred in nature (Fig. 5a). In contrast to the observation that site-wise mutational tolerance aligns more closely with natural diversity at the species level (Fig. 5b), greater concordance with natural variation at the type level was observed for AA preferences for both viruses (*P* < 2.2 × 10⁻¹⁶ for type versus species according to Wilcoxon tests for both EVA71 and CVB3; Fig. 5c). This pattern is again primarily driven by the capsid proteins (Fig. 5c and Supplementary Fig. 5g). Notably, although overall similar AAs were preferred across the capsid in both the DMS results and sequence alignments at the type level, sites showing larger differences in preferred AAs were specifically enriched in surface-exposed residues in both viruses (*P* = 4.1 × 10⁻⁵ and 5.6 × 10⁻⁷ for EVA71 and CVB3, respectively, according to Wilcoxon tests; Fig. 5d). This probably reflects host interactions not captured in our in vitro DMS

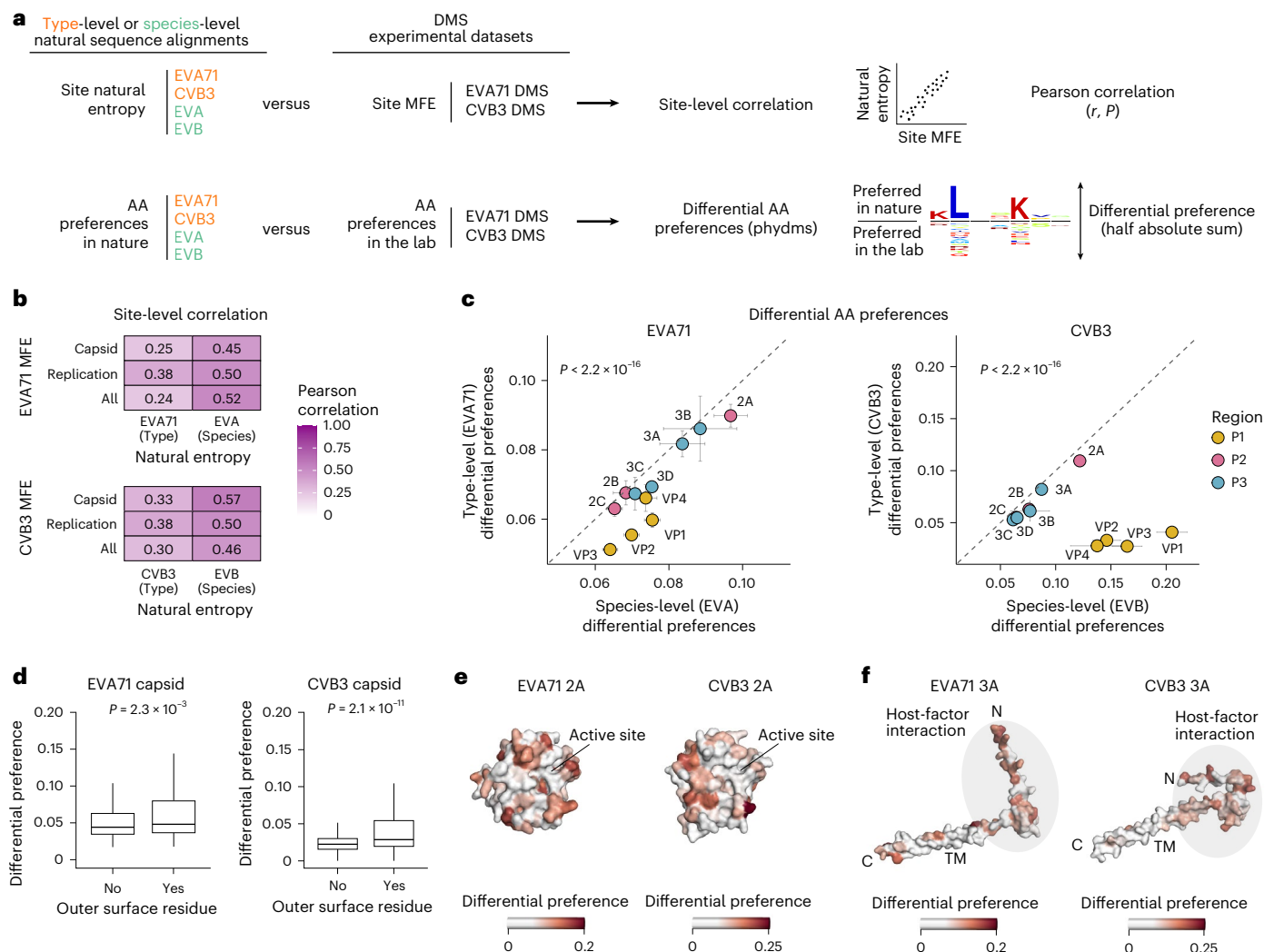


Fig. 5 | Comparison of laboratory and natural selection pressures. **a**, Workflow for the comparison of natural diversity and DMS-derived MFES. Sequence alignments were used to obtain a site-level diversity measure (Shannon entropy) and AA preferences. These metrics were compared with site MFES and AA preferences derived from the DMS datasets. Site-level analyses were performed using Pearson correlation tests, and differential AA preferences between nature and laboratory were obtained using phydms. **b**, Correlation between the experimental DMS datasets and natural sequence diversity at the species and type levels, calculated separately for different functional regions (capsid, replication or whole proteome). All Pearson correlations shown are two-sided and statistically significant ($P < 0.05$). **c**, Differential AA preferences per protein between the experimental DMS datasets and natural sequence diversity for CVB3 and EVA71 at the species and type levels. The error bars indicate the standard

error of the mean across all positions in each protein. Statistical significance was calculated using a two-sided Wilcoxon test across all positions in each viral proteome (species versus type, $P < 2.2 \times 10^{-16}$ for EVA71 and CVB3 proteins). The dashed line represents a 1:1 relationship between both axes. **d**, Differential AA preferences between the DMS datasets and natural sequences at the type level in surface-exposed versus buried capsid residues for EVA71 ($n = 618$ versus $n = 1,106$) and CVB3 ($n = 608$ versus $n = 1,094$). In each box plot, the horizontal line indicates the median, the box shows the IQR and the whiskers extend to values within 1.5 \times the IQR. **e, f**, Structural mapping of differential preferences between the DMS datasets and natural sequences at the type level onto the 2A (**e**) and 3A (**f**) proteins for EVA71 and CVB3. TM, transmembrane. Host-factor interaction regions in 3A include residues involved in ACBD3 and GBF1 binding.

experiments. Indeed, for CVB3, where we have previously comprehensively defined neutralizing antibody binding sites in human sera⁴², we found differences in AA preferences to be significantly enriched in sites targeted by neutralizing antibodies compared with other surface residues ($P = 4.7 \times 10^{-4}$ according to a Wilcoxon test; Supplementary Fig. 5i).

In contrast to the structural proteins, AA preferences in the non-structural proteins are consistent across the different evolutionary timescales, with the proteins 2A and 3A exhibiting the greatest divergence from natural AA preferences (Fig. 5e,f). Differences in 2A predominantly map to surface-exposed residues ($P = 0.3$ and $P = 0.0012$ for EVA71 and CVB3, respectively, according to Wilcoxon tests; Supplementary Fig. 5j), which are probably involved in interactions with host factors³⁵. Similarly, for 3A, differential selection

clusters in the N-terminal region for both viruses ($P = 0.0098$ and $P = 4.4 \times 10^{-6}$ for EVA71 and CVB3, respectively, according to Wilcoxon tests for the N-terminal domain versus the membrane-binding domain; Supplementary Fig. 5k), which mediates interactions with host factors to promote membrane remodelling and disrupt intracellular trafficking (Fig. 5f)⁴³.

Druggable pocket screen identifies a mutationally constrained target in the enterovirus 2C helicase

Building on our structural and functional analyses, we next explored whether our cross-species mutational tolerance data can be used to identify potential antiviral targets with high barriers to resistance (Fig. 6a). Specifically, we aimed to identify druggable pockets that

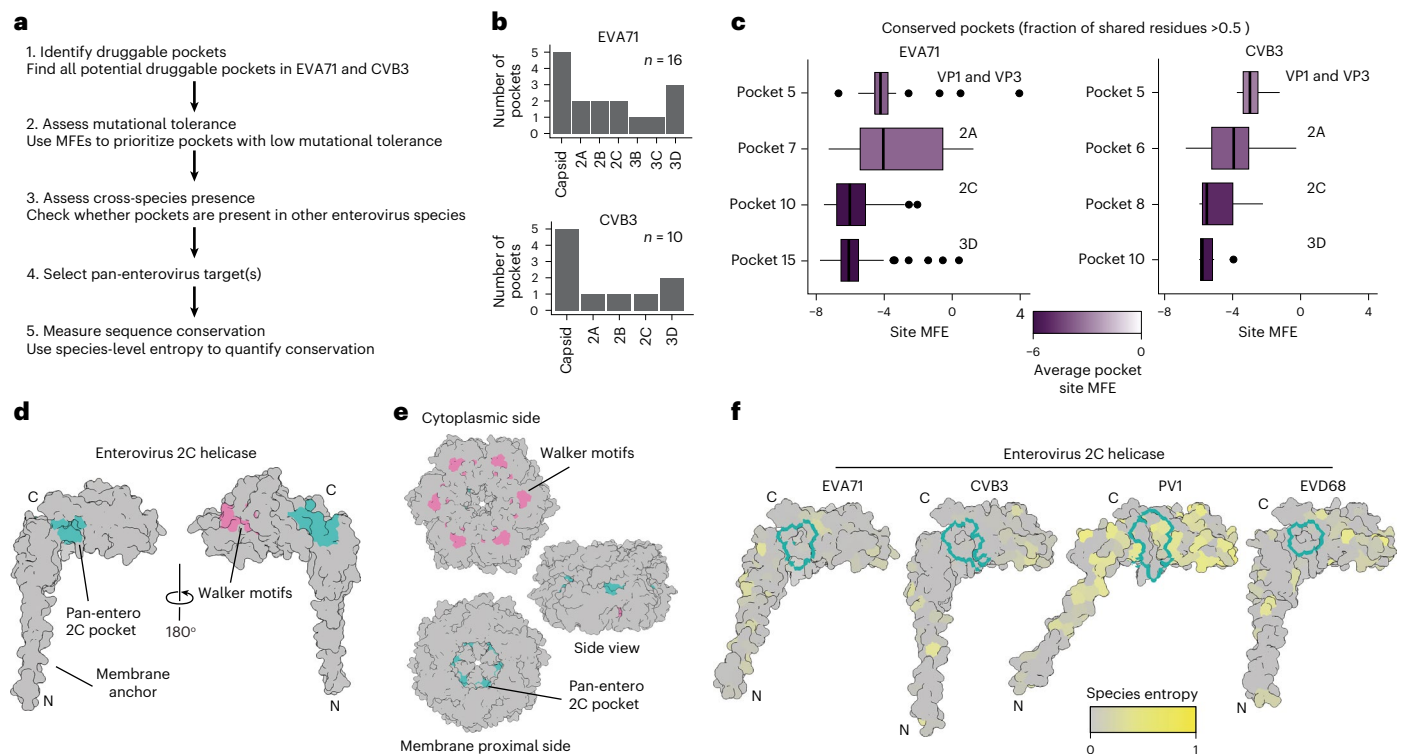


Fig. 6 | Identification of a conserved druggable pocket in the enterovirus 2C helicase. a, Workflow for identifying and prioritizing druggable pockets, combining computational druggable pocket screening with experimental determination of mutation tolerance from DMS studies. **b**, The total number of druggable pockets detected in the EVA71 and CVB3 proteomes using SiteMap. **c**, Box plots comparing the average MFEs for residues within conserved druggable pockets (fraction of shared residues >0.5, $n \geq 12$ residues). In each box plot, the horizontal line indicates the median, the box shows the IQR, the whiskers extend to values within $1.5 \times$ the IQR and outliers are plotted individually. Lower MFE scores indicate higher mutational constraint. **d**, Surface representation of the AlphaFold3-predicted EVA71 2C helicase monomer. The identified pan-

are intolerant to mutations such that mutations required to bypass drug binding would come at a high fitness cost for the virus and thus reduce its ability to spread in the host. For this, we first performed a computational screen for druggable pockets in the EVA71 and CVB3 proteomes using SiteMap⁴⁴, which identifies and scores protein pockets on the basis of size, geometry and physicochemical properties that are favourable for drug binding. This screen identified 16 and 10 candidate pockets in EVA71 and CVB3, respectively (Fig. 6b). Of these, four pockets were selected due to their conservation between EVA71 and CVB3 (fraction of shared residues relative to the smallest pocket >0.5; Fig. 6c and Supplementary Fig. 6a). We then applied our DMS data as a functional filter to prioritize pockets located in regions of high mutational constraint—that is, where potential escape mutants are likely to be lethal or have markedly reduced fitness. This analysis revealed that pockets in structural proteins are generally mutationally tolerant, whereas those in non-structural proteins are highly constrained, particularly those in the 2C and 3D proteins, making them ideal candidates for antiviral targeting (Fig. 6c).

To assess the potential of these two pockets as pan-enterovirus targets, we expanded our drug pocket analysis in 2C and 3D to representative species from *Enterovirus C* (poliovirus 1; PV1) and *Enterovirus D* (enterovirus D68; EVD68). To define conserved drug pockets, we calculated the fraction of shared residues in each pocket across the viruses, using the smallest drug pocket as a reference (Supplementary Fig. 6b). The 3D pocket shows significant variation between viruses, sharing few or no residues across all four viruses (Supplementary Fig. 6b).

enterovirus druggable pocket (green) is shown in relation to the ATPase domain (pink), which includes the Walker A, B and C motifs. **e**, Surface representation of the AlphaFold3-predicted EVA71 2C helicase in its hexameric form, shown from the cytoplasmic, membrane and side views. **f**, Multi-species structural comparison of the 2C pocket. The surface of the AlphaFold3-predicted 2C protein monomer from four representative enterovirus species is shown: EVA71, CVB3, PV1 and EVD68. Surfaces are coloured by the natural sequence entropy at each position within their respective species, with grey indicating high conservation (low entropy) and yellow indicating high variability. The prioritized pocket region is outlined in green.

In contrast, the 2C pocket shows strong conservation across all four viruses, with at least one pocket in the 2C protein of each virus sharing >0.8 of its residues with that of CVB3 (Supplementary Fig. 6b). These pockets in 2C share nine conserved residues, five of which are identical across all four viruses (Supplementary Fig. 6c,d). Structurally, the pocket is situated away from the well-characterized ATPase Walker motifs (Fig. 6d,e), and the architecture of the pocket is preserved across all four enterovirus species, with root mean square deviation values of the shared residues relative to the CVB3 2C pocket ranging from 0.165 to 0.204 Å (Supplementary Fig. 6f). Finally, to validate the conservation of these nine conserved drug pocket residues in nature, we examined their variability compared to the remaining sites in the proteome across large alignments of natural sequences from each virus species ($n = 1,081, 762, 292$ and 71 for enteroviruses A, B, C and D, respectively). Overall, these residues showed lower variability than the remainder of the proteome across the four enterovirus species ($P < 0.05$ for enteroviruses B–D and $P = 0.2$ for enterovirus A according to Wilcoxon tests; Fig. 6f and Supplementary Fig. 6g). The high structural conservation, low natural variability and strong fitness costs in cell culture of the residues comprising this core drug pocket support its potential as a promising target for drug development.

Discussion

Understanding the functional consequences of mutations is a fundamental goal of evolutionary virology, holding important implications in molecular epidemiology and the design of successful therapeutics.

The systematic derivation of MFEs via DMS has proved to be a powerful tool for this purpose, allowing for the comprehensive mapping of MFEs across viral proteins and the definition of sites affecting interaction with host factors, including antiviral proteins, receptors and neutralizing antibodies^{16,26,27,45}. DMS studies on related viral species allow us to assess how evolutionary constraints on viral proteomes change over longer evolutionary timescales, as viral families diversify phenotypically. Capitalizing on the recent publication of the first two proteome-wide DMS studies, we compared genome-wide MFEs for two species of enteroviruses, EVA71 and CVB3. Although these viruses share only ~50% sequence identity, we found that trends in MFEs are largely consistent across their proteomes (Figs. 1 and 2), highlighting biophysical and functional constraints shared across species. These shared constraints map to catalytic sites (that is, proteases, polymerase and helicase) and capsid assembly interfaces, probably representing the conserved functional core for the major HEV group.

Beyond this conserved core, we identified several sites exhibiting divergent mutation tolerance across the two viruses. In the capsid proteins, distinct constraints in the two viruses were observed across sites that interact with the cognate receptor of each virus, SCARB2 for EVA71 and CAR for CVB3 (Figs. 2 and 3), providing a clear, type-specific virus–host interaction interface. Interestingly, we also observed divergence in constraints underlying functions of the non-structural proteins (Figs. 2 and 4). For example, divergence in mutational tolerance was observed in 3A at residues implicated in binding the host GBF1 protein, suggesting that EVA71 and CVB3 have tuned their interaction with this critical host factor to their specific replication requirements. This phenomenon is reminiscent of the distinct membrane interactions and binding modes observed for the host factor ACBD3 by 3A proteins from distinct genera of picornaviruses: enterovirus and kobuvirus⁴⁶. Overall, our results underscore the utility of comparative DMS analyses in uncovering type-specific host–interaction interfaces, which can help uncover the mechanisms underlying the distinct tissue tropism and pathogenic outcomes of infections with related viruses.

By integrating DMS data with natural sequence variation across viral types and species, we provide a comprehensive analysis linking experimentally derived MFEs to evolutionary patterns across multiple evolutionary scales (Fig. 5). Site-wise MFEs correlate more strongly with broader species-level mutational flexibility than with within-type variation, highlighting the ability of DMS to capture broad structural and functional constraints. In contrast, the preference for specific AAs at each site better correlates with type-specific evolutionary constraints, consistent with findings in influenza haemagglutinin, where DMS profiles correspond better to natural variation within a strain than across strains⁴⁷.

Although DMS provides a high-resolution map of functional constraints under controlled, *in vitro* conditions, natural evolution unfolds in more complex environments that integrate diverse and dynamic host-specific pressures, including immune responses, tissue-specific replication constraints and other selective forces absent from cell culture. Type-level differences in capsid proteins were concentrated at surface-exposed immune-evasion sites, revealing host-driven selection that is incompletely captured by DMS. In contrast, non-structural proteins show better agreement between experimental and natural constraints across evolutionary scales, reflecting their conserved roles in viral replication and intracellular processes. The non-structural proteins 2A and 3A are exceptions, exhibiting a high mutational tolerance in the DMS datasets and large discrepancies from natural sequences in AA preferences. The elevated tolerance of 2A *in vitro* aligns with previous findings that this protein is dispensable for replication in cell culture, despite its essential role in antagonizing the host interferon response^{48,49}. For 3A, differential AA preferences are concentrated in the N-terminal region of the protein across both viruses, which is known to interact with multiple host factors. Although the precise drivers of this divergence remain unclear, mutations in these regions have been

shown to impair the ability of the virus to block host immune responses (antiviral RNA interference responses in EVA71 (ref. 42) and antigen presentation in CVB3 (ref. 50)). These observations support the idea that key immune-related functions of 2A and 3A are not fully captured in our experimental system. Although EVA71 and CVB3 may encounter distinct host factors and immune pressures in their respective natural contexts, the regions where *in vitro* and *in vivo* constraints are the strongest converge on the same proteins (the capsid, 2A and 3A), implicating these as key hotspots of immune-driven selection across enteroviruses.

Drug resistance remains a major obstacle to efficacious antiviral therapy. DMS data can guide the prioritization of drug targets where mutations conferring reduced drug binding come at a high cost to viral replication^{29,51}. Here we integrated MFE data of both enteroviruses with computational druggable pocket prediction to identify such targets (Fig. 6). In general, druggable pockets in structural regions showed higher mutational tolerance, in agreement with prior experimental findings from different capsid-binding antivirals (for example, WIN compounds⁵²). In the non-structural proteins, we identified a drug pocket in 2C that is both mutationally intolerant and structurally conserved in both viruses, and appears to be conserved across major HEV species (enterovirus A–D), suggesting that it is a promising candidate for future development of pan-enterovirus antivirals.

Overall, these findings highlight the power of integrating DMS data from related viruses with structural and evolutionary analyses to uncover both general and specific constraints acting on viral proteomes. This analysis framework can be readily extended to additional viral families and similarly applied to discovering hotspots of type-specific host interactions, conserved druggable pockets, and both convergent and divergent selection pressures.

Methods

Cell lines and viruses

Rhabdomyosarcoma (RD, CCL-136), HeLa-H1 (CRL-1958; RRID: [CVCL_3334](#)) and RPE (CRL-4000) cells were obtained from ATCC. All cell lines were periodically validated to be free of mycoplasma. Viral libraries were constructed from plasmid molecular clones of EVA71 (4643/TW/1998) and CVB3 (Nancy strain, taxon identifier 103903). Detailed information on the production of DMS plasmid libraries and viral populations can be found in their corresponding publications^{29,30}. Briefly, mutagenized plasmid libraries encompassing nearly all possible single AA alterations across the EVA71 and CVB3 genomes, as well as all single codon deletions across all regions except the CVB3 capsid, were used to generate the corresponding high-diversity viral populations. These populations were subjected to two sequential passages at low multiplicity of infection and mutation frequencies in both the input plasmid libraries, and the passaged viral populations were quantified by next-generation sequencing.

Calculation of MFEs

For libraries generated using the synthetic biology approach (P2 and P3 for the CVB3 and EVA71 datasets), codon tables were filtered to keep only codons where engineered mutations were originally introduced. For the P1 libraries of CVB3, generated via PCR-based mutagenesis, all single mutations within codons were excluded from the analysis to increase the signal-to-noise ratio⁵³. MFEs were calculated following the general procedure described in `dms_tools2` for the ratio method⁵⁴ using custom scripts (https://github.com/QVEU/EV_DMS_Comparison). Briefly, for each site, the sum of all codons giving rise to a particular AA substitution was divided by that of the WT AA to obtain the relative enrichment of each mutation at each site. The relative enrichment for a given mutation in the viral populations was divided by its relative enrichment in the plasmid libraries to obtain MFEs. To avoid zeros in the numerator, a coverage-scaled pseudocount of 1 was added to the count of each mutation. Additionally, all mutations that were not observed at least five times in the mutagenized libraries were omitted from analysis.

To normalize the MFE scores across regions and between the two viruses, we normalized each dataset to the MFE score of synonymous mutations, assuming these would have a neutral fitness score ($\log_2(\text{syn MFE}) = 0$). For the CVB3 libraries, since synonymous mutations were present throughout the proteome, the MFEs of each replicate were standardized to the average of all synonymous MFEs observed. These normalized MFEs were averaged across replicates to calculate the average MFE for each mutation, as long as at least two MFE values were observed among the independent replicates. As synonymous mutations were not introduced throughout the original EVA71 libraries, a sublibrary containing a synonymous mutation (642 GAA → GAG) and a representative mutagenized region from each of the two original libraries (sites 577–610 and 863–910 of the capsid and replication libraries, respectively) was used for normalization. MFEs for this sublibrary were calculated and normalized to synonymous mutations as performed for the CVB3 libraries. A linear model was obtained between the synonymous-normalized MFEs of the sublibrary and the non-normalized MFEs of the corresponding subset of capsid and non-structural mutations. This model was then applied to the full dataset, enabling normalization across regions and standardization to synonymous mutation MFEs in both viruses. Site MFEs were calculated as the average mutation MFE per site for each replicate and averaged across replicates to calculate the average MFE for each site, as long as at least two site MFE values were observed among the independent replicates. High reproducibility was observed between experimental replicates (correlation of mutation-level MFE: Pearson's $r = 0.7$ – 0.95 for CVB3 and $r = 0.89$ – 0.96 for EVA71).

Protein structure analyses

The EVA71 capsid structure was obtained from the PDB: [8E2X](#). The CVB3 capsid structure was based on PDB structure [4GB3](#) but modified to match the Nancy sequence as previously reported⁵³. The CVB3 3D structure used was PDB: [3CDW](#). Structure prediction for all remaining CVB3 and EVA71 proteins was performed using the AlphaFold (version 3) server (<https://alphafoldserver.com/>)⁵⁵. The structures of 2C and 3D from PV1 and EVD68 were predicted using AlphaFold3, on the basis of the individual protein sequences extracted from the full polyproteins in UniProt entries [P03300](#) and [Q68T42](#), respectively. We used AlphaFold3 to model the interactions between the EVA71 and CVB3 3A proteins and human GBF1 (UniProt entry: [Q92538](#)). Secondary structure predictions were obtained from STRIDE⁵⁶. Surface residues of the capsid and 2A were identified using the findSurfaceRes script in PyMOL v3.1 with a cut-off of 2.5 \AA^2 . For the capsid, outer surface residues were defined as those located more than 130 \AA from the centre of the capsid. Receptor footprints were obtained using the interfaceResidues script in PyMOL, with a cut-off of 3 \AA to detect capsid residues in contact with the receptor in the PDB structures [6I2K](#) (EVA71–SCARB2), [7VYK](#) (CVB3–CAR) and [7VY5](#) (CVB3–DAF). The locations of the Walker motifs were extracted using a publication that describes the structure of the CVB3 2C protein⁵⁷. Structures were coloured by the site MFE difference attribute using the Render by Attribute tool in UCSF ChimeraX^{58,59}. Capsid assembly interfaces between adjacent protomers and pentamers were identified using the select contact function in UCSF ChimeraX. Interacting surfaces were defined by having a buried solvent-accessible surface area of 15 \AA^2 or greater. To determine the transmembrane topology of the 3A viral proteins from CVB3 and EVA71, we uploaded their AA sequences to the DeepTMHMM³⁸ (version 1.0.44) web server (<https://dtu.biolib.com/DeepTMHMM/>). Domains of 3A (N-terminal, membrane-binding and C-terminal domains) were obtained from UniProt annotations and published work (EVA71 (ref. [46](#)) and UniProt entry [P03313](#) for CVB3).

Structural alignments and per-protein comparison

EVA71 and CVB3 proteins were structurally aligned in PyMOL using the align command to perform pairwise alignments based on sequence

and structural similarities. Gaps in the alignments were excluded from subsequent analyses. To compare the two datasets, a normalization per protein was performed to account for differences in the two datasets, such as differences in sequencing coverage. For this, site MFEs were standardized using the formula below so that the average corrected MFE of each protein would be 0, and positive and negative scores would indicate sites with MFE above or below the average MFE of that particular protein, respectively.

For each site i of a protein p :

$$\text{corrected MFE}_i = \frac{\text{MFE}_i - \min(\text{MFE}_p)}{\text{mean}(\text{MFE}_p) - \min(\text{MFE}_p)}$$

Natural diversity analyses

To examine the variability of natural enteroviral sequences, all available sequences matching the criteria “Enterovirus A”, “Enterovirus B”, “Enterovirus C” or “Enterovirus D” were downloaded from NCBI Virus into four separate FASTA files on 26 September 2024. Sequences marked as being lab-passaged or vaccine strains were excluded. Full-length polyprotein open reading frames were then identified (EVA $n = 4,034$, EVB $n = 1,644$, EVC $n = 2,007$, EVD $n = 1,415$) and extracted using EMBOSS (version 6.6.0), before being clustered at 98% similarity with cd-hit (version 4.8.1)⁶⁰ (EVA $n = 1,081$, EVB $n = 762$, EVC $n = 292$, EVD $n = 71$). Vaccine-derived sequences were removed from the *Enterovirus C* sequence set. The *Enterovirus A* and *Enterovirus B* sequence sets were combined so they could be aligned against one another. Sequences were first codon aligned using RevTrans version 2.0 (ref. [61](#)) and then translated to AA sequences using the transeq function from EMBOSS⁶² and re-aligned using MAFFT⁶³ (version 7.505). The AA alignment was indexed to the EVA71 and CVB3 DMS WT strains using custom R code. For Shannon entropy and similarity calculations, *Enterovirus A*, *Enterovirus B*, EVA71 ($n = 325$) and CVB3 ($n = 49$) subalignments were extracted. For phydms⁴¹ analyses, the 98% clustered *Enterovirus A* and *Enterovirus B* sequence sets were subclustered at 95% AA similarity by cd-hit to accommodate the data-processing limits of the pipeline (EVA $n = 39$, EVB $n = 78$).

To calculate Shannon entropy, AA counts at each position were summed using the FreqMat function from the QSutils⁶⁴ package (version 1.20.0); then, the formula below was used, where p is the frequency of a given AA at a given site:

$$\text{Shannon Entropy} = - \sum p(x) \log(p(x))$$

To assess how similar EVA71 and CVB3 are to one another and to other *Enterovirus A* and *Enterovirus B* sequences, we used the DistanceMatrix function from the DECIPHER package⁶⁵ (version 2.30.0). Pairwise per cent dissimilarities were calculated between all sequences in the combined *Enterovirus A* and *Enterovirus B* alignment; then, the results were filtered to the desired comparison (that is, EVA71 versus CVB3 sequences). Per cent similarity is simply one minus per cent dissimilarity. Per cent similarities at the protein, capsid and replication protein levels were determined using the same procedure on the appropriate alignment fragment.

Phylogenetic trees were constructed with RAxML⁶⁶ (version 8.2.12) and visualized in TreeViewer⁶⁷ (version 2.2.0). The ICTV representative AA sequence for each species was used for all species except *Enterovirus A* and *Enterovirus B*, for which the EVA71 and CVB3 DMS WT sequences were used instead. Trees were constructed using the PROTGAMMAWAG model and bootstrapped 1,000 times.

Phydms analysis

Phydms⁴¹ was used to evaluate whether the incorporation of MFE into phylogenetic models could improve model fit compared with standard models and to calculate differential AA preferences between the DMS datasets and natural variation. For this, EVA71 and CVB3 sequences, or

EVA and EVB sequences, were aligned and processed as indicated above for entropy measurements and split into the capsid (P1) and replication (P2 and P3) regions. MFE values were transformed into site preferences by dividing each MFE value by the sum of all MFE values at each site. Phydms was then run using the default settings following preparation of the sequence alignment with the phydms_prealignment function.

Drug pocket prediction

Druggable pockets were identified using Schrödinger (version 13.8). Proteins were first subjected to the protein preparation workflow in Maestro using the default settings, and then SiteMap was used to define druggable pockets using the default settings, with sites having a SiteMap score >0.8 considered as druggable according to previously suggested criteria⁶⁸. Homologous pockets were defined as those sharing more than 50% of their residues, relative to the size of the smaller pocket.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All sequencing data are available in the NCBI Sequence Read Archive under project numbers [PRJNA1066851](#), [PRJNA643896](#), [PRJNA779606](#), [PRJNA1013170](#) and [PRJNA1033421](#).

Code availability

All code used to analyse and generate the analysis described here is available via GitHub at https://github.com/QVEU/EV_DMS_Comparison.

References

- Xie, Z., Khamrin, P., Maneekarn, N. & Kumthip, K. Epidemiology of enterovirus genotypes in association with human diseases. *Viruses* **16**, 1165 (2024).
- Lugo, D. & Krogstad, P. Enteroviruses in the early 21st century: new manifestations and challenges. *Curr. Opin. Pediatr.* **28**, 107–113 (2016).
- Vogt Matthew, R. et al. Enterovirus D68 in the anterior horn cells of a child with acute flaccid myelitis. *N. Engl. J. Med.* **386**, 2059–2060 (2022).
- Cihakova, D. & Rose, N. R. in *Advances in Immunology* Vol. 99 (ed. Alt, F. W.) 95–114 (Academic Press, 2008).
- Fairweather, D., Stafford, K. A. & Sung, Y. K. Update on coxsackievirus B3 myocarditis. *Curr. Opin. Rheumatol.* **24**, 401–407 (2012).
- Gaaloul, I. et al. Coxsackievirus B detection in cases of myocarditis, myopericarditis, pericarditis and dilated cardiomyopathy in hospitalized patients. *Mol. Med. Rep.* **10**, 2811–2818 (2014).
- Bastea, L. I. et al. Coxsackievirus and adenovirus receptor expression facilitates enteroviral infections to drive the development of pancreatic cancer. *Nat. Commun.* **15**, 10547 (2024).
- Ypma-Wong, M. F., Dewalt, P. G., Johnson, V. H., Lamb, J. G. & Semler, B. L. Protein 3CD is the major poliovirus proteinase responsible for cleavage of the P1 capsid precursor. *Virology* **166**, 265–270 (1988).
- Gamarnik, A. V. & Andino, R. Switch from translation to RNA replication in a positive-stranded RNA virus. *Genes Dev.* **12**, 2293–2304 (1998).
- Suhy, D. A., Giddings, T. H. Jr. & Kirkegaard, K. Remodeling the endoplasmic reticulum by poliovirus infection and by individual viral proteins: an autophagy-like origin for virus-induced vesicles. *J. Virol.* **74**, 8953–8965 (2000).
- Feng, Q. et al. Enterovirus 2A^{pro} targets MDA5 and MAVS in infected cells. *J. Virol.* **88**, 3369–3378 (2014).
- Wang, Y. et al. The strategy of Picornavirus evading host antiviral responses: non-structural proteins suppress the production of IFNs. *Front. Microbiol.* **9**, 2943 (2018).
- Zhang, X., Paget, M., Wang, C., Zhu, Z. & Zheng, H. Innate immune evasion by picornaviruses. *Eur. J. Immunol.* **50**, 1268–1282 (2020).
- Jiang, P., Liu, Y., Ma, H.-C., Paul, A. V. & Wimmer, E. Picornavirus morphogenesis. *Microbiol. Mol. Biol. Rev.* **78**, 418–437 (2014).
- Staring, J. et al. KREMEN1 is a host entry receptor for a major group of enteroviruses. *Cell Host Microbe* **23**, 636–643.e5 (2018).
- Álvarez-Rodríguez, B., Buceta, J. & Geller, R. Comprehensive profiling of neutralizing polyclonal sera targeting coxsackievirus B3. *Nat. Commun.* **14**, 6417 (2023).
- Ang, P. Y., Chong, C. W. H. & Alonso, S. Viral determinants that drive *Enterovirus-A71* fitness and virulence. *Emerg. Microbes Infect.* **10**, 713–724 (2021).
- Wang, Y. & Pfeiffer, J. K. Emergence of a large-plaque variant in mice infected with coxsackievirus B3. *mBio* **7**, e00119 (2016).
- Stadnick, E., Dan, M., Sadeghi, A. & Chantler, J. K. Attenuating mutations in coxsackievirus B3 map to a conformational epitope that comprises the puff region of VP2 and the knob of VP3. *J. Virol.* **78**, 13987–14002 (2004).
- Wang, Q. et al. Molecular basis of differential receptor usage for naturally occurring CD55-binding and -nonbinding coxsackievirus B3 strains. *Proc. Natl Acad. Sci. USA* **119**, e2118590119 (2022).
- Jartti, M., Flodström-Tullberg, M. & Hankaniemi, M. M. Enteroviruses: epidemic potential, challenges and opportunities with vaccines. *J. Biomed. Sci.* **31**, 73 (2024).
- Patick, A. K. et al. In vitro antiviral activity of AG7088, a potent inhibitor of human rhinovirus 3C protease. *Antimicrob. Agents Chemother.* **43**, 2444–2450 (1999).
- Smee, D. F., Evans, W. J., Nicolaou, K. C., Tarbet, E. B. & Day, C. W. Susceptibilities of enterovirus D68, enterovirus 71, and rhinovirus 87 strains to various antiviral compounds. *Antiviral Res.* **131**, 61–65 (2016).
- Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS Biol.* **16**, e3000003 (2018).
- Moratorio, G. mSphere of influence: experimental evolution of RNA viruses. *mSphere* **5**, e00179-20 (2020).
- Burton, T. D. & Eyre, N. S. Applications of deep mutational scanning in virology. *Viruses* **13**, 1020 (2021).
- Dadonaite, B. et al. A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2 spike. *Cell* **186**, 1263–1278.e20 (2023).
- Lei, R. et al. Mutational fitness landscape of human influenza H3N2 neuraminidase. *Cell Rep.* **42**, 111951 (2023).
- Álvarez-Rodríguez, B., Velandia-Álvarez, S., Toft, C. & Geller, R. Mapping mutational fitness effects across the coxsackievirus B3 proteome reveals distinct profiles of mutation tolerability. *PLoS Biol.* **22**, e3002709 (2024).
- Bakhache, W., Symonds-Orr, W., McCormick, L. & Dolan, P. T. Deep mutation, insertion and deletion scanning across the *Enterovirus A* proteome reveals constraints shaping viral evolution. *Nat. Microbiol.* **10**, 158–168 (2025).
- Zhou, D. et al. Unexpected mode of engagement between enterovirus 71 and its receptor SCARB2. *Nat. Microbiol.* **4**, 414–419 (2019).
- Wessels, E. et al. Molecular determinants of the interaction between coxsackievirus protein 3A and guanine nucleotide exchange factor GBF1. *J. Virol.* **81**, 5238–5245 (2007).
- Hirano, J. et al. Enterovirus 3A protein disrupts endoplasmic reticulum homeostasis through interaction with GBF1. *J. Virol.* **98**, e0081324 (2024).
- Yamayoshi, S. et al. Scavenger receptor B2 is a cellular receptor for enterovirus 71. *Nat. Med.* **15**, 798–801 (2009).

35. Chung, S.-K. et al. Internalization and trafficking mechanisms of coxsackievirus B3 in HeLa cells. *Virology* **333**, 31–40 (2005).
36. Wessels, E. et al. A viral protein that blocks Arf1-mediated COP-I assembly by inhibiting the guanine nucleotide exchange factor GBF1. *Dev. Cell* **11**, 191–201 (2006).
37. Choe, S. S., Dodd, D. A. & Kirkegaard, K. Inhibition of cellular protein secretion by picornaviral 3A proteins. *Virology* **337**, 18–29 (2005).
38. Hallgren, J. et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.04.08.487609> (2022).
39. Baptista, D. et al. AlphaFold models of host–pathogen interactions elucidate the prevalence and structural modes of molecular mimicry. Preprint at *bioRxiv* <https://doi.org/10.1101/2022.04.08.487609> (2025).
40. Dolan, P., Bakhache, W. & Symonds-Orr, W. Genotype-by-inhibitor interactions to dissect *Enterovirus* replication. Preprint at <https://doi.org/10.21203/rs.3.rs-7660613/v1> (2025).
41. Hilton, S. K., Doud, M. B. & Bloom, J. D. Phydms: software for phylogenetic analyses informed by deep mutational scanning. *PeerJ* **5**, e3657 (2017).
42. Qiu, Y. et al. Human virus-derived small RNAs can confer antiviral immunity in mammals. *Immunity* **46**, 992–1004.e5 (2017).
43. Jackson, T. & Belsham, G. J. Picornaviruses: a view from 3A. *Viruses* **13**, 456 (2021).
44. Halgren, T. New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **69**, 146–148 (2007).
45. Ashenberg, O., Padmakumar, J., Doud, M. B. & Bloom, J. D. Deep mutational scanning identifies sites in influenza nucleoprotein that affect viral inhibition by MxA. *PLoS Pathog.* **13**, e1006288 (2017).
46. Horova, V. et al. Convergent evolution in the mechanisms of ACBD3 recruitment to picornavirus replication sites. *PLoS Pathog.* **15**, e1007962 (2019).
47. Lee, J. M. et al. Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl Acad. Sci. USA* **115**, E8276–E8285 (2018).
48. Visser, L. J. et al. Essential role of *Enterovirus* 2A protease in counteracting stress granule formation and the induction of type I interferon. *J. Virol.* **93**, e00222-19 (2019).
49. Schipper, J. G. et al. The multifaceted role of the viral 2A protease in enterovirus replication and antagonism of host antiviral responses. *PLoS Pathog.* **21**, e1013443 (2025).
50. Cornell, C. T., Kiosses, W. B., Harkins, S. & Whitton, J. L. Coxsackievirus B3 proteins directionally complement each other to downregulate surface major histocompatibility complex class I. *J. Virol.* **81**, 6785–6797 (2007).
51. Ni, X. et al. Combined crystallographic fragment screening and deep mutational scanning enable discovery of Zika virus NS2B-NS3 protease inhibitors. *Nat. Commun.* **16**, 8930 (2025).
52. Lewis, J. K., Bothner, B., Smith, T. J. & Siuzdak, G. Antiviral agent blocks breathing of the common cold virus. *Proc. Natl Acad. Sci. USA* **95**, 6774–6778 (1998).
53. Mattenberger, F., Latorre, V., Tirosh, O., Stern, A. & Geller, R. Globally defining the effects of mutations in a picornavirus capsid. *eLife* **10** (2021).
54. Bloom, J. D. Software for the analysis and visualization of deep mutational scanning data. *BMC Bioinform.* **16**, 168 (2015).
55. Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
56. Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500–W502 (2004).
57. Hurdiss, D. L. et al. Fluoxetine targets an allosteric site in the enterovirus 2C AAA+ ATPase and stabilizes a ring-shaped hexameric complex. *Sci. Adv.* **8**, eabj7615 (2022).
58. Pettersen, E. F. et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
59. Meng, E. C. et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).
60. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
61. Wernersson, R. & Pedersen, A. G. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. *Nucleic Acids Res.* **31**, 3537–3539 (2003).
62. Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **16**, 276–277 (2000).
63. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
64. Guerrero-Murillo, M. & Font, J. G. QSutils: Quasispecies Diversity; Bioconductor Version: Release (3.11) <http://bioconductor.org/packages/release/bioc/html/QSutils.html> (2020).
65. Wright, E. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J.* <https://doi.org/10.32614/RJ-2016-025> (2016).
66. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
67. Bianchini, G. & Sánchez-Baracaldo, P. TreeViewer: flexible, modular software to visualise and manipulate phylogenetic trees. *Ecol. Evol.* **14**, e10873 (2024).
68. Halgren, T. A. Identifying and characterizing binding sites and assessing druggability. *J. Chem. Inf. Model.* **49**, 377–389 (2009).

Acknowledgements

The project was funded by grant nos PID2021-125063NB-I00, PID2024-157587NB-I00 and CNS2022-135100 to R.G. from the Spanish Ministerio de Ciencia, Innovación y Universidades, the Agencia Estatal de Investigación and the European Union (MICIU/AEI/10.13039/501100011033 and FEDER, UE or NextGenerationEU/PRTR). B.Á.-R. acknowledges postdoctoral funding (JDC2022-050122-I) from the Spanish Ministerio de Ciencia, Innovación y Universidades, the Agencia Estatal de Investigación and the European Union (MICIU/AEI/10.13039/501100011033 and the European Union NextGenerationEU/PRTR). This work was supported by funding to P.T.D. from the Division of Intramural Research, NIH–NIAID, under project number 1ZIAI001360. This research was supported in part by the Intramural Research Program of the National Institutes of Health (NIH). The contributions of the NIH authors were made as part of their official duties as NIH federal employees, are in compliance with agency policy requirements and are considered Works of the United States Government. However, the findings and conclusions presented in this paper are those of the authors and do not necessarily reflect the views of the NIH or the US Department of Health and Human Services.

Author contributions

Conceptualization: P.T.D., R.G., B.Á.-R. and W.B. Data curation: B.Á.-R., W.B. and L.M. Formal analysis: B.Á.-R., W.B. and L.M. Writing—original draft: B.Á.-R., W.B. and L.M. Writing—review and editing: R.G., P.T.D., B.Á.-R., W.B. and L.M.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41559-026-02993-8>.

Correspondence and requests for materials should be addressed to Ron Geller or Patrick T. Dolan.

Peer review information *Nature Ecology & Evolution* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2026

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All sequencing data has been uploaded to SRA (PRJNA1066851, PRJNA643896, PRJNA779606, PRJNA1013170, PRJNA1033421). The GitHub page can be found here: https://github.com/QVEU/EV_DMS_Comparison

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample sizes was decided based on previous deep mutational scanning studies.
Data exclusions	None
Replication	Data represent three biological replicates of each screen.
Randomization	No randomization was used in this design.
Blinding	No blinding was used or appropriate for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)	American Type Culture Collection (ATCC): RD (CCL-136), HeLa-H1 (CRL-1958), HEK293 (CRL-3216)
---------------------	--

Authentication	<i>Describe the authentication procedures for each cell line used OR declare that none of the cell lines used were authenticated.</i>
Mycoplasma contamination	Yes all cell lines were regularly tested for mycoplasma
Commonly misidentified lines (See ICLAC register)	<i>Name any commonly misidentified cell lines used in the study and provide a rationale for their use.</i>

Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>