

RESEARCH

Open Access



ProbeTools: designing hybridization probes for targeted genomic sequencing of diverse and hypervariable viral taxa

Kevin S. Kuchinski^{1,2*}, Jun Duan¹, Chelsea Himsworth^{3,4}, William Hsiao^{1,5} and Natalie A. Prystajeky^{1,6}

Abstract

Background: Sequencing viruses in many specimens is hindered by excessive background material from hosts, microbiota, and environmental organisms. Consequently, enrichment of target genomic material is necessary for practical high-throughput viral genome sequencing. Hybridization probes are widely used for enrichment in many fields, but their application to viral sequencing faces a major obstacle: it is difficult to design panels of probe oligo sequences that broadly target many viral taxa due to their rapid evolution, extensive diversity, and genetic hypervariability. To address this challenge, we created ProbeTools, a package of bioinformatic tools for generating effective viral capture panels, and for assessing coverage of target sequences by probe panel designs *in silico*. In this study, we validated ProbeTools by designing a panel of 3600 probes for subtyping the hypervariable haemagglutinin (HA) and neuraminidase (NA) genome segments of avian-origin influenza A viruses (AIVs). Using *in silico* assessment of AIV reference sequences and *in vitro* capture on egg-cultured viral isolates, we demonstrated effective performance by our custom AIV panel and ProbeTools' suitability for challenging viral probe design applications.

Results: Based on ProbeTools' *in silico* analysis, our panel provided broadly inclusive coverage of 14,772 HA and 11,967 NA reference sequences. For each reference sequence, we calculated the percentage of nucleotide positions covered by our panel *in silico*; 90% of HA and NA reference sequences had at least 90.8 and 95.1% of their nucleotide positions covered respectively. We also observed effective *in vitro* capture on a representative collection of 23 egg-cultured AIVs that included isolates from wild birds, poultry, and humans and representatives from all HA and NA subtypes. Forty-two of forty-six HA and NA segments had over 98.3% of their nucleotide positions significantly enriched by our custom panel. These *in vitro* results were further used to validate ProbeTools' *in silico* coverage assessment algorithm; 89.2% of *in silico* predictions were concordant with *in vitro* results.

Conclusions: ProbeTools generated an effective panel for subtyping AIVs that can be deployed for genomic surveillance, outbreak prevention, and pandemic preparedness. Effective probe design against hypervariable AIV targets also validated ProbeTools' design and coverage assessment algorithms, demonstrating their suitability for other challenging viral capture applications.

Keywords: Influenza A viruses, Avian influenza viruses, Viral genomics, Hybridization probe capture, Targeted genomic sequencing, Viral surveillance

Background

Most viral specimens are characterized by low amounts of viral genomic material and excessive background from viral hosts and environmental organisms. Consequently, practical viral genome sequencing requires

*Correspondence: kevin.kuchinski@bccdc.ca

²Vancouver, Canada

Full list of author information is available at the end of the article



targeted enrichment for confident detection and accurate genotyping, especially in high-throughput surveillance and clinical applications [1–3]. Hybridization probe capture methods have been used for viral target enrichment [4–7], but designing probe oligo sequences for many viruses can be a major obstacle due to extensive genomic diversity and hypervariability within and between viral taxa [8–13].

Probe panels are typically designed by enumerating probe-length sub-sequences (k-mers) from reference sequences. The simplest approach to designing probes for hypervariable taxa is to enumerate k-mers from an exhaustive collection of reference sequences, thereby including as much genomic divergence in the design space as possible [7, 8]. This approach becomes problematic, however, when redundant probe sequences are enumerated from repeated instances of conserved genomic loci.

A few strategies have been used to address this redundancy problem. One common strategy is to cluster similar k-mers after they have been enumerated [6, 7]. Another strategy is to align candidate probe sequences against select reference genomes to identify and retain only those probes targeting divergent genotypes [8]. Redundancy has also been addressed by constraining the design space to a limited number of representative reference genomes, selected either by manual curation or clustering [9–12]. Some of these strategies have been supplemented with multiple sequence alignments over hypervariable loci or entire genomes so that probes are designed from consensus and degenerate sequences [9, 10].

Spacing between probe sequences is another complicated design consideration. Regular spacing (tiling) is the most common approach because it is easy to implement, but it does not ensure optimal positioning of probes. Reducing the spacing increases the likelihood that some enumerated probes are optimally positioned, but it also increases the number of probe candidates and any associated computation to collapse redundancy among them. Creating the smallest possible panel of probes that optimally covers the entire target space quickly becomes an intractable computational problem, one that had led to increasingly complicated approaches including sophisticated minimization of loss functions [13].

Efforts to address viral hypervariability have resulted in several elaborate probe design algorithms. Unfortunately, these have largely been implemented on a study-by-study basis and have not resulted in general-purpose software tools that can be easily used by others. Meanwhile, among the handful of published probe design packages, there is only one option that specifically addresses viral hypervariability [13]. The rest

are intended for comparatively conserved eukaryotic genomes and are inadequate for many viral applications [14–17]. This leaves virologists with limited options for designing their own hybridization probes, especially if they have minimal capacity for custom programming, sophisticated mathematics, and experimental bioinformatics.

Here, we present ProbeTools, a general-purpose software package for designing compact probe panels against diverse viral taxa and other hypervariable genomic targets. It can also be used to assess how well existing panels cover user-provided target sequences. ProbeTools implements the established K-mer clustering method, but it adds a novel incremental design heuristic to minimize the generation of redundant probes. It also provides a simple command line user interface for ease of use and automation.

In this study, we demonstrate ProbeTools' effectiveness by designing capture panels for avian-origin influenza A viruses (AIVs). These viruses are subtyped by two hypervariable viral surface proteins called haemagglutinin (HA) and neuraminidase (NA), making them an appropriately challenging case study for ProbeTools. The genome segments encoding these proteins have diversified into 16 avian-origin HA subtypes and 9 avian-origin NA subtypes, giving rise to 144 possible combinations and the HxNx nomenclature used in both animal and human contexts (e.g. H1N1, H3N2, H5N1, H7N9). Furthermore, each of these subtypes has diverged into numerous clades, many of which have been only partially characterized [12, 18, 19].

AIV lineages have varying potential for spillover from wild birds into poultry and humans [20–25], posing a perennial threat to agriculture and public health. Some lineages cause costly outbreaks of severe disease in poultry flocks which, in turn, expose humans to potentially dangerous zoonotic influenza infections. This threatens economic disruption, future pandemic crises, and new types of seasonal influenza, which remains an important global health burden and among the ten leading causes of death worldwide [12, 21–31]. Consequently, surveillance of AIVs in wild birds is a cornerstone of outbreak prevention and pandemic preparedness [12, 20, 32, 33]. An effective panel of AIV-specific probes would be instrumental for these genomics-based surveillance efforts.

In this study, we designed and validated a compact panel of 3600 probes for detecting and subtyping AIVs. Our results showed broad inclusivity against all avian-origin HA and NA subtypes based on *in silico* predictions against of tens-of-thousands of AIV reference sequences. We also demonstrated successful captures *in vitro* on a representative collection of 23 egg-cultured AIVs. These results validated the core ProbeTools algorithms and

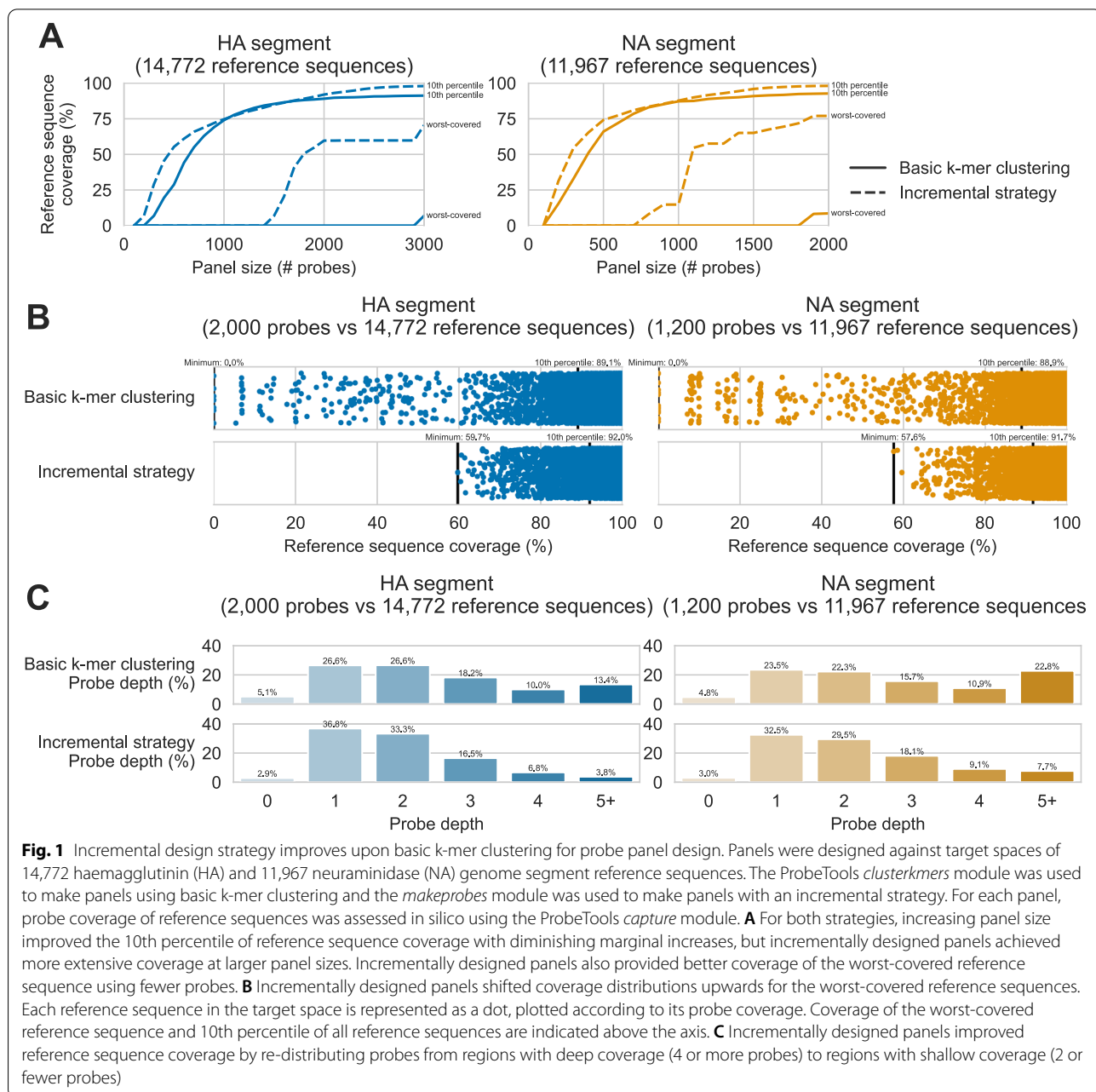
demonstrated its suitability for other challenging probe design applications with hypervariable viral targets.

Results

Assessing basic k-mer clustering and marginal improvements to target coverage with additional probes

We began by assessing probe design against hypervariable targets with a basic k-mer clustering algorithm, wherein all 120-mers were enumerated from a target space of AIV reference sequences then clustered based on 90% nucleotide sequence identity. We used

this strategy, implemented in the ProbeTools *clusterkmers* module, to generate probe panels of increasing size against 14,772 HA segment reference sequences and 11,967 NA segment reference sequences. We then used the ProbeTools *capture* module, which aligns probe sequences against target sequences, to assess target space coverage, i.e. the percentage of nucleotide positions in each target sequence covered by at least one probe in the panel (Fig. 1A, solid lines). As expected, panels with more probe sequences provided better target space coverage, however we observed diminishing marginal



improvements for both HA and NA genome segments. We also noted that reference sequences with no probe coverage remained in the target space past the point of diminishing marginal returns. These results highlighted two limitations of the basic k-mer clustering approach: some HA and NA segments remained undetected despite designing additional probes, and additional probes provided only modest and diminishing improvements to the distribution of target coverage.

Improving target coverage with incremental panel design focused on poorly covered targets

To address the limitations we observed with basic k-mer clustering, we devised an incremental design strategy to improve marginal coverage increases, especially for poorly covered targets. In this strategy, basic k-mer clustering was used to design panels in smaller batches of 100 probes. After adding each batch to the growing panel, target space regions without probe coverage were identified using the *capture* module. These low coverage regions were then extracted with another ProbeTools module called *getlowcov* and used as a new target space for designing the next batch. In this way, each subsequent batch of probes was focused on regions not already covered by the panel.

We compared target space coverage for panels designed with this incremental strategy against panels designed above using basic k-mer clustering (Fig. 1). The incremental strategy provided higher 10th percentiles of coverage, especially for HA panels larger than 2000 probes and NA panels larger than 1200 probes (Fig. 1A). Furthermore, the incremental strategy provided better coverage for the worst-covered reference sequences (Fig. 1AB). We also compared depth of probe coverage, i.e. the number of probes covering each nucleotide position in target sequences (Fig. 1C). This comparison indicated that the incremental strategy improved target coverage by redistributing probes from positions with deep coverage to shallow coverage. We speculate that the incremental approach, by removing already-covered regions from the target space after each batch, limited the enumeration of adjacent, partially overlapping k-mers that provided redundant coverage. Based on the observed performance improvements of the incremental strategy, it was implemented as an additional self-contained ProbeTools module called *makeprobes* (Fig. 2).

Predicted coverage of HA and NA subtypes by AIV_v1 panel

Using the incremental strategy implemented in the ProbeTools *makeprobes* module, we generated an AIV-targeting probe panel called AIV_v1. It was composed of 1935 HA-specific probes and 1435 NA-specific probes.

We also included 184 probes targeting the highly conserved matrix segment (M) which is the standard AIV diagnostic target [24, 34]. We then used the ProbeTools *capture* module to predict probe coverage using the AIV_v1 panel for all 36,313 AIV reference sequences in the target space. The minimum, maximum, and 10th percentile of reference sequence coverage was calculated for each HA and NA subtype and the M segment (Fig. 3A).

We observed that M segments had the best coverage followed by NA subtypes then HA subtypes, reflecting the comparative levels of genomic diversity within these genome segments. No reference sequence had less than 59.6% coverage, which is sufficient for segment and subtype identification. HA subtypes H5, H7, and H9 are considered high priority for AIV surveillance because they frequently cause agricultural outbreaks and novel influenza infections in humans [23–26, 34]; 90% of H5, H7, and H9 reference sequences had at least 94.4, 88.5, and 92.4% probe coverage respectively. We also noted a significant positive monotonic association between a subtype's target coverage distribution and number of reference sequences from that subtype in the target space (Fig. 3B). This indicated that over-representing subtypes in the target space resulted in preferential design and better probe coverage for these targets, e.g. the high priority subtypes H5, H7, and H9.

In vitro capture of diverse egg-cultured influenza isolates

After assessing the AIV_v1 panel in silico, we had it synthesized and used it to perform in vitro captures on a collection of diverse egg-cultured AIV isolates (Table 1). We ensured that each avian-origin HA and NA subtype was represented in the collection, and we included isolates from wild birds, poultry, and humans. The collection contained 22 egg cultures, including one mixed infection, providing 23 viruses and 69 distinct HA, NA, and M segments for in vitro capture.

Sequencing libraries were prepared from each isolate then pooled. AIV library pools were diluted 1:100 (ng/ng) in libraries of background material made from mock-infected egg cultures, then captured three times independently using the AIV_v1 panel. Pre- and post-capture pools were sequenced to calculate mean fold-enrichment at each nucleotide position in these 69 HA, NA, and M segments. Half of all nucleotide positions had a mean fold-enrichment greater than 351.2-fold, and 90% of nucleotide positions had a mean fold-enrichment greater than 195.0-fold (Fig. 4A). We also calculated the percentage of the capture pools composed of background material from the mock-infected egg cultures, then compared these percentages pre- and post-capture (Fig. 4B). Before capture, the mean background percentage was 99.17%, but this was reduced

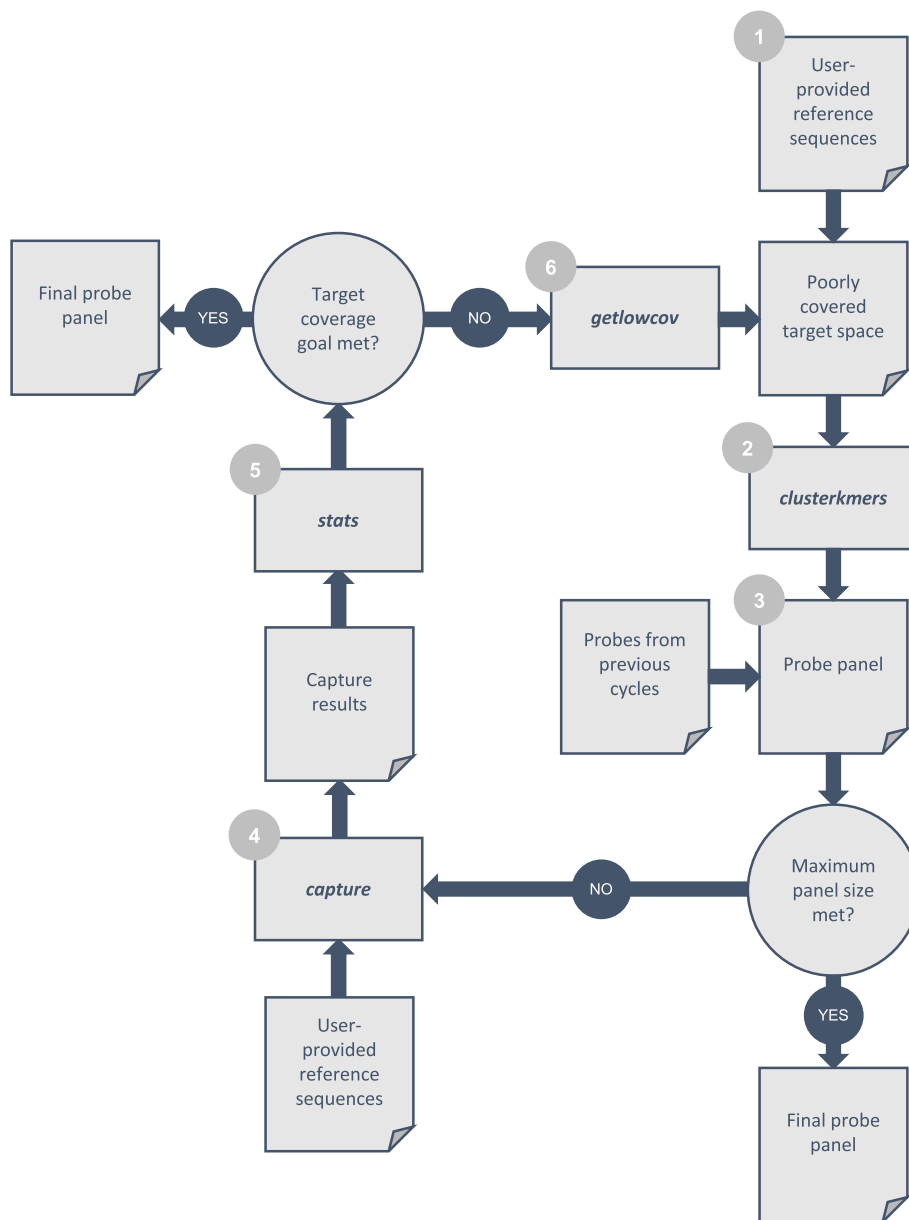


Fig. 2 ProbeTools *makeprobes* module implements a generalized incremental design algorithm. 1) The user provides a FASTA formatted file containing target sequences, which forms the total target space and becomes the poorly covered target space for the first cycle of the design loop. 2) The ProbeTools *clusterkmers* module generates a batch of probe sequences from the poorly covered target space using its k-mer clustering algorithm. 3) The latest batch of probes is combined with probes from previous batches to generate the current probe panel. If the size of the current probe panel meets the maximum panel size set by the user, the design loop ends and the current probe panel becomes the final panel, otherwise... 4) The ProbeTools *capture* module determines which nucleotide positions in the total target space are covered by the current probe panel. 5) The ProbeTools *stats* module calculates the 10th percentile of target coverage from the *capture* module results. If the target coverage goal set by the user is met, the current probe panel becomes the final probe panel, otherwise... 6) The *getlowcov* module extracts low coverage regions of the target space from the *capture* module results. These become the new poorly covered target space, and the design loop repeats

to 0.03% following capture. Together, these data demonstrate effective enrichment of AIV material and removal of background by probe capture with the AIV_v1 panel.

We also used these in vitro results to assess breadth of enrichment, i.e. the percentage of nucleotide positions in each HA, NA, and M segment that had been significantly enriched by probe capture (Fig. 4C, Table S1). Breadth

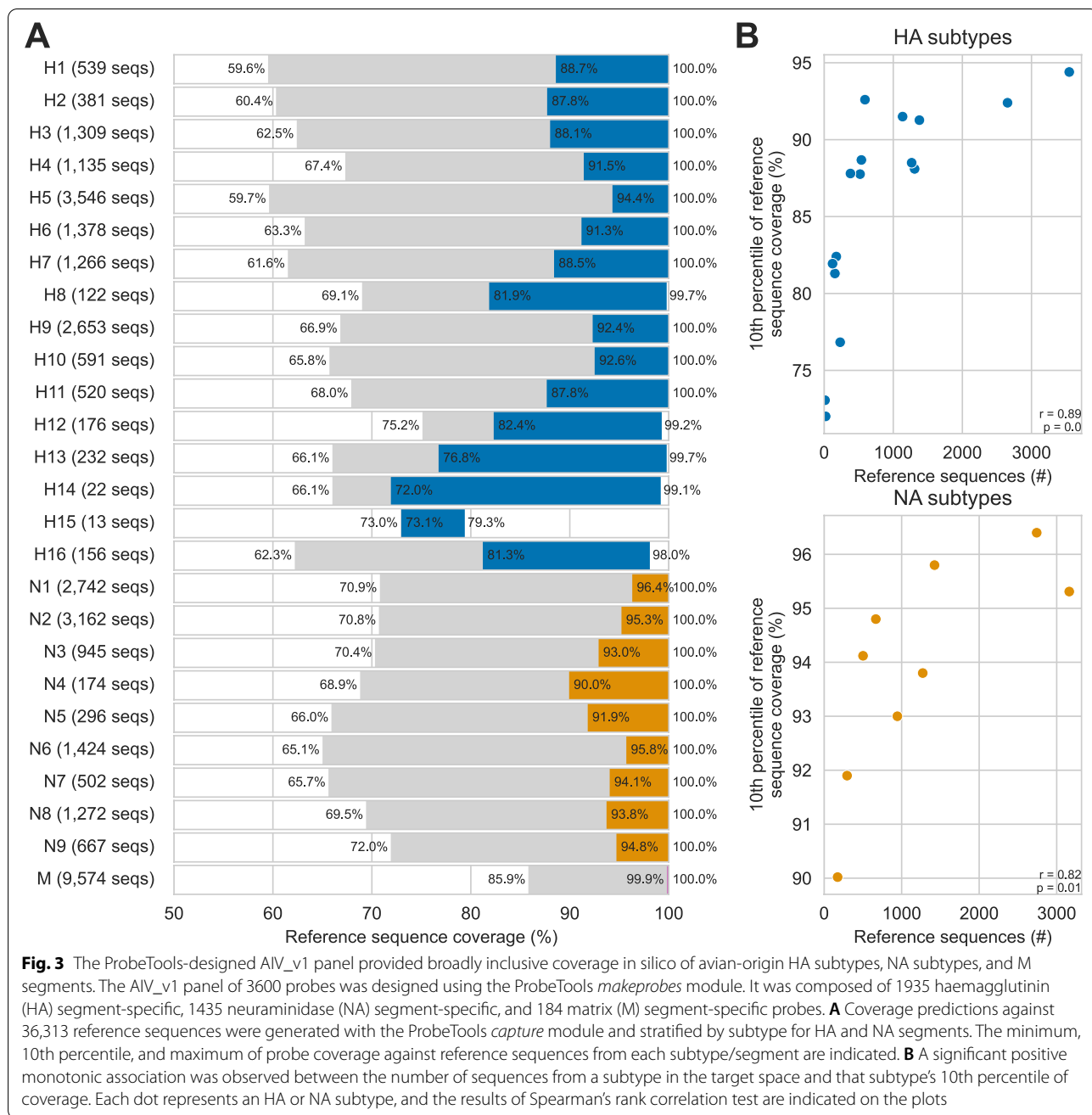


Fig. 3 The ProbeTools-designed AIV_v1 panel provided broadly inclusive coverage in silico of avian-origin HA subtypes, NA subtypes, and M segments. The AIV_v1 panel of 3600 probes was designed using the ProbeTools *makeprobes* module. It was composed of 1935 haemagglutinin (HA) segment-specific, 1435 neuraminidase (NA) segment-specific, and 184 matrix (M) segment-specific probes. **A** Coverage predictions against 36,313 reference sequences were generated with the ProbeTools *capture* module and stratified by subtype for HA and NA segments. The minimum, 10th percentile, and maximum of probe coverage against reference sequences from each subtype/segment are indicated. **B** A significant positive monotonic association was observed between the number of sequences from a subtype in the target space and that subtype's 10th percentile of coverage. Each dot represents an HA or NA subtype, and the results of Spearman's rank correlation test are indicated on the plots

of enrichment was greater than 96.3% for 64 of 69 segments in the collection, and it was not less than 46.5% for any segment, which is sufficient for segment and subtype identification (Table S3). Nine isolates contained high priority H5, H7, and H9 segments, all of which had greater than 98.7% breadth of enrichment. This included two isolates from zoonotic human infections (H5N1 and H7N9), which were extensively enriched despite the absence of reference sequences from human infections in the target space used for probe design.

We further examined the five segments with less than 96.3% breadth of enrichment to understand why they were apparently not captured in full. First, we used the ProbeTools *capture* module to assess if the AIV_v1 panel lacked probes targeting their particular genome segment sequences. We observed that most positions without significant enriched were nonetheless extensively covered by the probe panel (Fig. 5A). This indicated that insufficient design by ProbeTools was not a major explanation for the lack of significant capture of these segments.

Table 1 Representative collection of egg-cultured avian influenza virus isolates. Isolates were selected to provide representation of each avian-origin haemagglutinin (HA) and neuraminidase (NA) subtype as well as infections from poultry, wild bird, and human hosts. Each specimen was given a name based on an abbreviation of its host type and a sequential number (P for poultry, WB for wild bird, and H for human). Poultry and wild bird isolates were obtained from the Canadian Food Inspection Agency's National Centre for Foreign Animal Disease (CFIA NCFAD), and human isolates were obtained from the Public Health Agency of Canada's National Microbiology Laboratory (PHAC NML). Isolate subtypes were confirmed in-house by genome sequencing

Specimen name	Host type	Strain name	HA subtype	NA subtype	Source laboratory
P1	Poultry	A/Turkey/Ontario/844-2/2006	H6	N1	CFIA NCFAD
P2	Poultry	A/Chicken/Germany/N/1949	H10	N7	
P3	Poultry	A/Turkey/Ontario/18-2/2000	H7	N1	
P4	Poultry	A/Emu/Texas/39924/1993	H5	N2	
P5	Poultry	A/Turkey/Ontario/6118/1967	H8	N4	
P6	Poultry	A/Chicken/Quebec/IM-109/2010	H6	N1	
WB1	Wild bird	A/Duck/British Columbia/26-2/2005	H5	N2	
WB2	Wild bird	A/Swan/Alberta/OTH33-8/2009	H1	N1	
WB3	Wild bird	A/Teal/Germany/Wv632/2005	H5	N1	
WB4	Wild bird	A/Duck/Alberta/C-16/2007	H7	N7	
WB5	Wild bird	A/Duck/Australia/341/1983	H15	N8	
WB6	Wild bird	A/Duck/Alberta/60/1976	H12	N5	
WB7	Wild bird	A/Gull/Maryland/4/1977	H13/H7	N6/N3	
WB8	Wild bird	A/Pheasant/Washington/37349/1985	H9	N9	
WB9	Wild bird	A/Mallard/Gurjev/263/1982	H14	N5	
WB10	Wild bird	A/Duck/British Columbia/14/1999	H4	N6	
WB11	Wild bird	A/Duck/Prince Edward Island/274.1/2006	H16	N3	
WB12	Wild bird	A/Duck/Alberta/431/2006	H3	N8	
WB13	Wild bird	A/Pintail/Alberta/293/1977	H2	N9	
WB14	Wild bird	A/Mallard/Manitoba/OTH27-1186/2017	H11	N9	
H1	Human	A/Alberta/01/2014	H5	N1	PHAC NML
H2	Human	A/Anhui/1/2013	H7	N9	

Next, we assessed whether experimental factors were responsible for nucleotide positions in these segments failing to achieve statistically significant enrichment. Fold-enrichment values between positions with and without significant enrichment were comparable, but variation between capture replicates were significantly different, with higher variation for positions that were not significantly enriched (Fig. 5BC). We attribute this to sub-optimal cDNA synthesis for the affected positions, causing under-representation of these positions in the material that was captured, lower depths of coverage, and higher stochasticity (Fig. 5D). Despite this source of experimental variation, and the limited number of replicates that was practical for us to perform, only 3.1% of nucleotide positions across all HA, NA, and M segments were impacted, and most of these positions only barely failed the enrichment significance test (half achieved a p -value < 0.07) (Fig. 5E). Overall, our in vitro capture results demonstrated that the ProbeTools-designed AIV_v1 panel performed well on real viral isolates, effectively removing background material and providing high

breadths of enrichment across HA, NA, and M segment targets.

Comparison of in silico probe coverage prediction and in vitro probe capture enrichment

ProbeTools relies on in silico coverage assessment by the *capture* module, both for final panel evaluation and for identifying poorly covered sequences during incremental design. To validate ProbeTools' coverage assessment algorithm, we examined how closely its in silico predictions agreed with in vitro capture results on egg-cultured AIV isolates.

Using the ProbeTools *capture* module, we determined which nucleotide positions in the egg-cultured AIVs were predicted to be covered by the AIV_v1 probe panel. We then compared these predictions to our in vitro capture results to see if significant enrichment had actually occurred at these nucleotide positions (Fig. 6 and Fig. S1). Predicted probe coverage and significant enrichment results were concordant for 89.2% of nucleotide positions. Only 2.3% of nucleotide

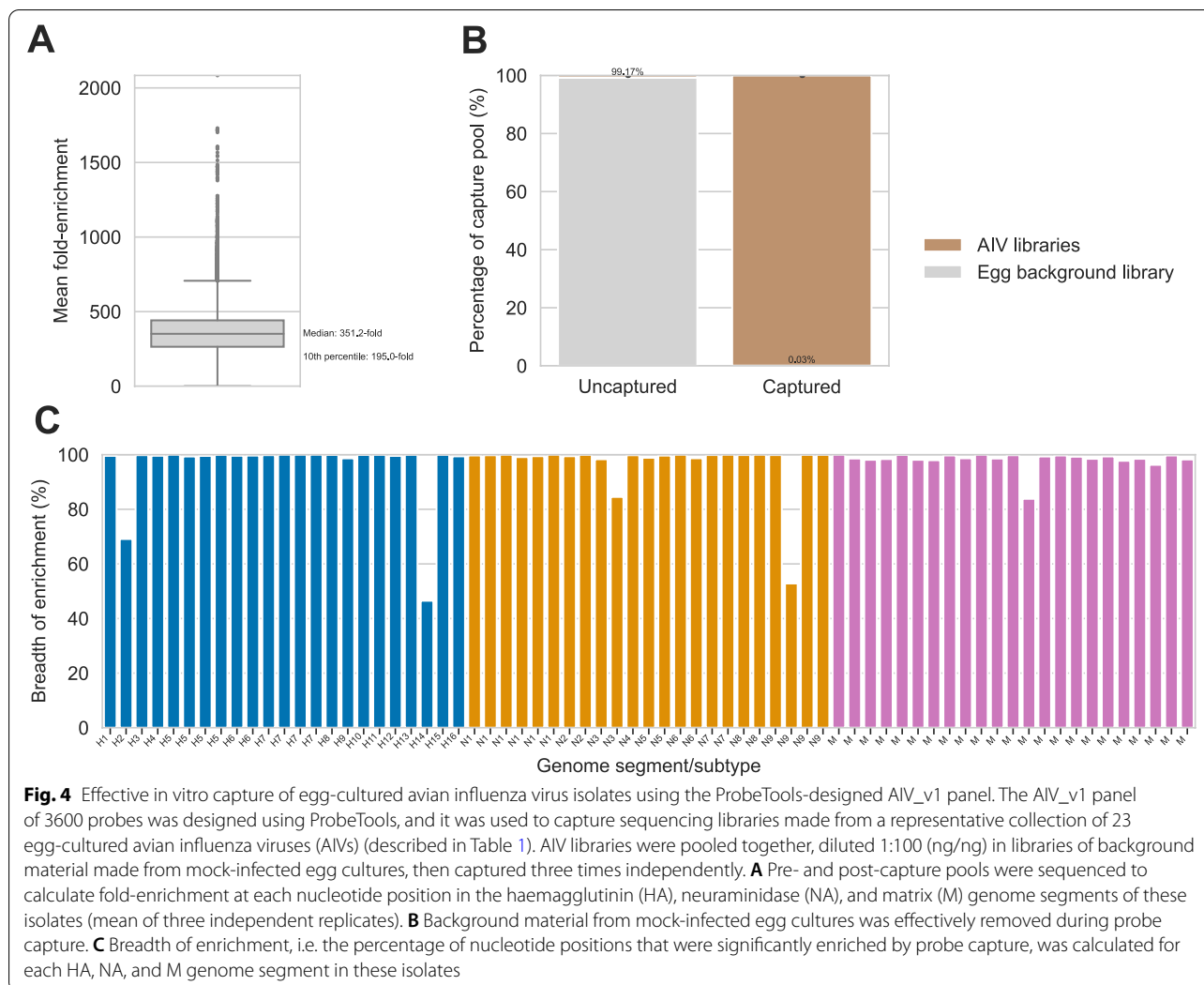


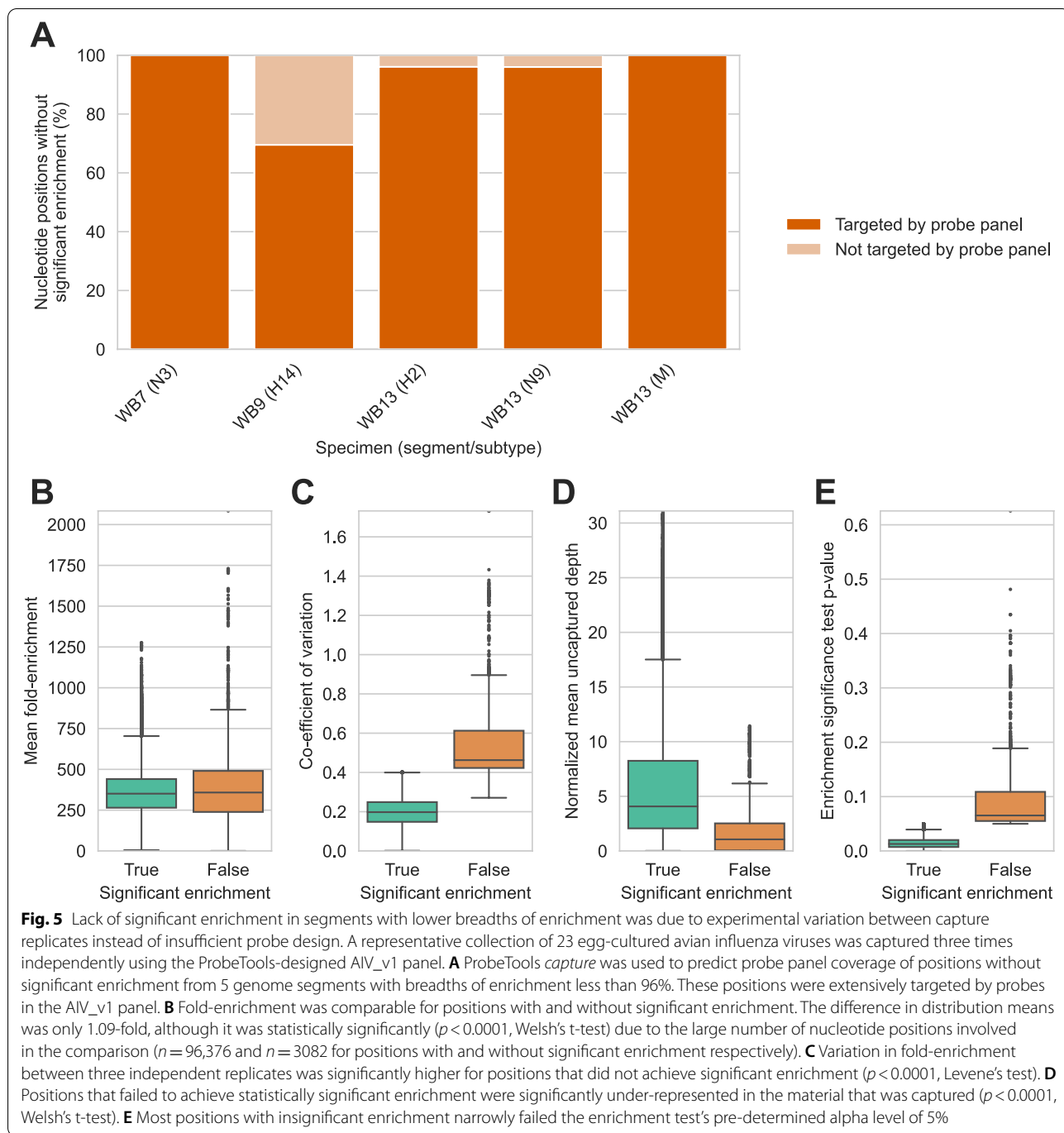
Fig. 4 Effective in vitro capture of egg-cultured avian influenza virus isolates using the ProbeTools-designed AIV_v1 panel. The AIV_v1 panel of 3600 probes was designed using ProbeTools, and it was used to capture sequencing libraries made from a representative collection of 23 egg-cultured avian influenza viruses (AIVs) (described in Table 1). AIV libraries were pooled together, diluted 1:100 (ng/ng) in libraries of background material made from mock-infected egg cultures, then captured three times independently. **A** Pre- and post-capture pools were sequenced to calculate fold-enrichment at each nucleotide position in the haemagglutinin (HA), neuraminidase (NA), and matrix (M) genome segments of these isolates (mean of three independent replicates). **B** Background material from mock-infected egg cultures was effectively removed during probe capture. **C** Breadth of enrichment, i.e. the percentage of nucleotide positions that were significantly enriched by probe capture, was calculated for each HA, NA, and M genome segment in these isolates

positions targeted by the AIV_v1 panel were not significantly enriched. These were concentrated in the five segments discussed above that were impacted by variability between replicates (Fig. S2). We also noted that 7.7% of nucleotide positions were significantly enriched despite not being targeted by the AIV_v1 panel, a phenomenon that was observed in most segments across all isolates (Fig. 6 and Fig. S1). We attribute this to the capture of larger fragments containing untargeted sequences adjacent to the location annealed by the probe. It might also indicate that local alignment parameters used by ProbeTools capture are more conservative than actual annealing thermodynamics. Either way, these results showed that ProbeTools predictions generally reflected actual capture of target genomic material, and in silico predictions more often underestimated panel performance when predictions were incorrect.

Discussion

This study highlighted some important considerations when designing panels using ProbeTools. Foremost among these was the effect of target space composition on panel inclusivity. In this AIV case study, we noted a significant positive monotonic association between panel coverage and the number of reference sequences representing a particular subtype in the target space. Based on how the ProbeTools algorithm ranks probe candidates by the number of k-mers in the cluster they represent, it stands to reason that over-representing similar taxa (which would contain many similar k-mers) would bias the resulting panel towards these taxa.

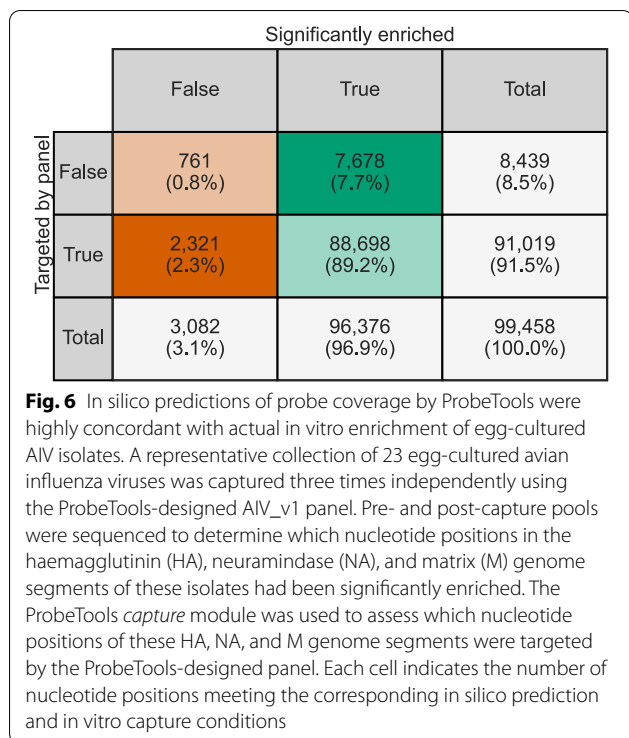
Consequently, ProbeTools users should have a thorough knowledge of the contents of their target space and the possible sources of sampling bias in the databases from which they obtain their reference sequences. In the case of AIVs, the agricultural impacts and public health threats of



certain HA subtypes have led to more frequent sequencing of these subtypes and accessioning of their genome sequences in popular databases. For our panel, this contributed to bias towards subtypes like H5, H7 and H9. Whether this is a benefit or limitation will depend on the intended application. In the context of outbreak prevention and pandemic preparedness, a panel biased towards taxa that are known for their agricultural impact and

zoonotic potential is beneficial. If the objective is to characterize viral diversity and ecology in wildlife, however, this could be a limitation.

To obtain the best results, ProbeTools users should purposefully curate their target space to serve their probe capture objectives. Users may want to identify taxa whose detection is a priority and over-represent them in the target space. Conversely, users may want



to ‘flatten’ their target space to ensure no particular taxa, clades, subtypes, *etc* dominate. This could be done manually, by selecting specific sequences to represent relevant groups, or it could be attempted bioinformatically by pre-clustering target sequences, providing the number and length of target sequences do not make this computationally prohibitive.

Another strategy could be to use the various ProbeTools modules to extract low coverage sequences from specific groups whose target sequences have poor probe coverage after a core panel is designed. For instance, had H15 subtype AIVs been a surveillance priority in this study, supplemental H15-specific probes could have been designed by running the *capture*, *getlowcov*, and *makeprobes* modules on the H15 subset of target sequences after noting their comparatively low coverage by the main panel. In this way, the modular nature of ProbeTools and the relatively simple-to-understand algorithms within each module empower users to experiment and find creative solutions. This flexibility is instrumental for tailoring probe panels to the needs of the user and their specific viral capture application.

Conclusions

In this study, we used ProbeTools to create an effective and broadly inclusive panel of hybridization capture probes for subtyping AIVs. Our results show the utility of this panel as a tool for AIV surveillance, outbreak

prevention, and pandemic preparedness. They also demonstrate that ProbeTools can effectively design probes against hypervariable genomic targets like avian-origin HA and NA segments. This validation of ProbeTools’ core design and coverage assessment algorithms shows that they are suitable for other challenging design applications, e.g. other viruses with hypervariable genes and pan-viral capture panels targeting multiple diverse taxa.

An increasing frequency of zoonotic outbreaks, epidemics, and pandemic crises has renewed interest in characterizing viral diversity at the interface of wildlife, livestock, game, and humans [35–38]. Genomic sequencing is becoming central to these One Health ventures. Viral capture panels will need designing and updating as our knowledge of viral threats continues to expand [39, 40].

The on-going COVID-19 pandemic has also demonstrated the value of viral genomics to public health [41–44], resulting in unprecedented investments in sequencing capacity at public health laboratories. This will expand routine genomics for numerous common pathogens, requiring the development of new target enrichment protocols. The COVID-19 pandemic has popularized the use of tiled multiplex PCR for viral genome enrichment in clinical and public health applications [45, 46], but on-going genomic drift is likely to cause amplicon dropouts and require frequent primer scheme redesigns for many pathogens, as has already been observed for SARS-CoV-2 [47]. Due to their longer length and, thus, higher tolerance of nucleotide mismatches [6], hybridization probe panels would require less frequent assay upkeep. To illustrate this principle, we used ProbeTools to design a SARS-CoV-2 panel containing 322 probes based on 1899 reference sequences from the first 2 months of the pandemic (January and February 2020). We then assessed in silico how well this panel covered 36,038 sequences from the most recent 2 months of the pandemic (May and June 2022); the tenth percentile of target coverage was 99.41% and the minimum was 98.19%, demonstrating that hybridization probes, especially panels designed by ProbeTools, can withstand genetic drift.

Furthermore, targeted enrichment protocols could be easily parallelized for multiple pathogens with probe capture; specimens containing different pathogens could be prepared into libraries concurrently and even pooled for a single capture using a pan-pathogen panel [8, 9, 11]. Amplicon sequencing, on the other hand, would require separately performed multiplex PCR reactions for each different pathogen, decreasing laboratory throughput.

Genomic sequencing is maturing into a routine tool for viral discovery, OneHealth surveillance, and clinical microbiology. Hybridization probe capture offers an

enrichment method that is durable against genomic drift and conducive to high-throughput, parallelized workflows for numerous pathogens. ProbeTools facilitates probe design tasks for these endeavours.

Methods

ProbeTools modules

ProbeTools consists of five main modules written in Python (v3.7.3) that perform essential tasks in the probe design process. ProbeTools is freely available under the MIT License. It can be installed easily using the Anaconda/Miniconda package and environment manager. Alternatively, it can be installed via the Python Package Index, followed by separate installation of its VSEARCH and BLASTn dependencies. Installation instructions, source code, documentation, and usage examples are available at <https://github.com/KevinKuchinski/ProbeTools>.

The *clusterkmers* module enumerates and clusters probe-length k-mers from user-provided target sequences. 1) K-mers are enumerated using a sliding window that advances by a specified number of bases. The user may also specify the width of the window. 2) K-mers are clustered based on nucleotide sequence similarity using VSEARCH *cluster_fast* [48]. 3) Centroid sequences from each cluster are ranked by the size of the cluster they represent. Centroids from larger clusters are assumed to be better probe candidates by virtue of having similarity to more k-mers in the target space. By default, *clusterkmers* enumerates 120-mers, advancing the window one base at a time, and it clusters using a nucleotide sequence identity threshold of 90%. Previous studies have observed effective hybridization between targets and probes with this degree of sequence similarity [9, 11].

The *capture* module predicts how well user-provided probe sequences cover user-provided target sequences. 1) Each probe sequence is locally aligned against each target sequence using BLASTn [49]. 2) Alignments are filtered, retaining those with a minimum sequence identity over a minimum alignment length. 3) Subject alignment start and end coordinates are extracted from the BLASTn results to determine which nucleotide positions in the target sequences are covered by probes. By default, *capture* requires 90% sequence identity over at least 60 bases to assign probe coverage to the aligned positions.

The *getlowcov* module uses the output of *capture* to extract genomic regions with low coverage from the provided targets. This allows for additional probe design focused on poorly covered regions of the target space. This module returns all sub-sequences where a minimum number of consecutive bases were covered by fewer than a specified number of probes. By default, *getlowcov* returns all sub-sequences over 40 bases in length

where all bases in the sub-sequence were covered by zero probes.

The *stats* module uses the output of *capture* to calculate coverage statistics. For each provided target, it calculates the percentage of nucleotide positions covered by varying numbers of probes (“target coverage” and “probe depth”).

The *makeprobes* module chains the previous modules together to implement a generalized incremental design strategy (Fig. 2). In this strategy, probes are designed in batches, and regions of the target space with probe coverage are removed between batches so that additional probes are focused on poorly covered areas. This module can be used as a convenient departure point for custom designs. The user is only required to provide target sequences and select a batch size. They can optionally specify a maximum panel size and target space coverage goal. The *makeprobes* module iterates through its design loop, adding batches of probes to the panel until the maximum panel size is met, the target space coverage goal is achieved, or no further probes can be generated.

Preparation of AIV target space

All available full-length influenza A virus genome segment sequences from avian hosts were downloaded from the Influenza Research Database (www.fludb.org) on Dec 5, 2017 [50]. Sequences containing degenerate bases were removed to avoid low quality entries. Sequences were then clustered using VSEARCH *cluster_fast* (v1.0.7) [48] with a 100% sequence identity threshold to remove redundant entries. The remaining entries were used as our final AIV target space (described in Table 2).

AIV_v1 probe panel design and in silico coverage assessment

The AIV_v1 panel was designed against our final AIV target space using the ProbeTools *makeprobes* module as follows: 2000 probes were designed against HA targets in 20 batches of 100 probes; 1500 probes were designed against NA targets in 15 batches of 100 probes, and 200 probes were designed against M targets in 20 batches of 10 probes. All probes were 120 nucleotides in length, and designs were conducted using *makeprobes* with default parameters. Designs were conducted with ProbeTools v0.0.5, VSEARCH v1.0.7, and BLASTn v2.2.31.

The top-ranked 1935 HA probes, 1435 NA probes, and 184M probes were combined into the final panel. Additional probes were added to the panel for potential control and validation applications, including 36 probes targeting the common reference strain A/Puerto Rico/8/34 and 10 probes targeting synthetic spike-in DNA oligomers with randomly generated artificial sequences. This provided a final panel of 3600 probes

Table 2 Avian influenza virus reference sequences used as target space in this study. Full-length genome segment sequences from avian hosts were downloaded from the Influenza Research Database (www.fludb.org). Sequences containing degenerate bases were removed, then the remaining sequences were clustered using a 100% nucleotide sequence identity threshold to discard redundant entries. This provided a final target space of 36,313 reference sequences representing all avian-origin haemagglutinin (HA) subtypes, neuraminidase (NA) subtypes, and matrix (M) segments

Genome segment	Subtype	Reference sequences in target space (#)	Target space size (KB)
HA	H1	539	939.1
	H2	381	664.0
	H3	1309	2267.6
	H4	1135	1944.1
	H5	3546	6129.7
	H6	1378	2361.3
	H7	1266	2148.5
	H8	122	209.9
	H9	2653	4498.9
	H10	591	1005.4
	H11	520	897.5
	H12	176	301.5
	H13	232	405.4
	H14	22	38.3
	H15	13	22.7
	H16	156	271.7
		HA untyped	733
	HA total	14,772	25,360.4
NA	N1	2742	3804.9
	N2	3162	4498.5
	N3	945	1347.3
	N4	174	249.8
	N5	296	427.4
	N6	1424	2037.0
	N7	502	718.4
	N8	1272	1822.9
	N9	667	948.5
		NA untyped	783
	NA total	11,967	16,971.0
M	none	9574	9582.4

(a breakpoint in the manufacturer's pricing structure), which was synthesized as a custom panel by Twist Bioscience (San Francisco, CA, USA). Sequences for probes in the AIV_v1 panel are provided in Supplemental Material 1. In silico coverage assessment of the AIV_v1 panel, both against the reference sequence target space and the consensus sequences of the egg-cultured isolate collection,

were conducted using the *capture* and *stats* modules with default parameters.

Preparation of sequencing libraries from egg-cultured influenza isolates

Detailed laboratory procedures for the following are provided in Supplemental Material 2. RNA extracts from egg-cultured AIV isolates and mock infected eggs were provided by the Canadian Food Inspection Agency's National Centre for Foreign Animal Disease (Winnipeg, Manitoba, Canada) and the Public Health Agency of Canada's National Microbiology Laboratory (Winnipeg, Manitoba, Canada). Eggs were not directly handled by the authors. cDNA was prepared from each RNA extract using a previously described method [51]. cDNA was fragmented by sonication, then prepared into sequencing libraries for Illumina platforms with unique dual index barcodes. Adapter-ligated cDNA was split into three separate barcoding reactions, providing three separately barcoded replicate libraries for each isolate.

Probe capture enrichment and genomic sequencing of libraries prepared from egg-cultured influenza isolates

Detailed laboratory and bioinformatic procedures for the following are provided in Supplemental Material 2. 1) Three pools were prepared, with each pool containing one replicate library from each AIV isolate. These pools were sequenced in-house on Illumina MiSeq to generate full HA, NA, and M segment sequences for each isolate and to confirm HA and NA subtypes. 2) Each pool was diluted in 1:100 (ng/ng) in one of three replicate libraries of background genomic material that had been prepared from a mock-infected chicken egg. Aliquots of each diluted pool were sequenced pre-capture at Canada's Michael Smith Genome Sciences Centre (Vancouver, BC) on one Illumina HiSeq X lane to establish baseline HA, NA, and M segment abundance. 3) Each diluted pool was independently captured using the AIV_v1 probe panel. Captured pools were then sequenced in-house on Illumina MiSeq to assess target enrichment of HA, NA, and M segments post-capture.

Analysis of significant probe capture enrichment for egg-cultured AIV isolates

1) Pre- and post-capture depths of coverage were determined by mapping each library's sequencing reads to the HA, NA, and M segment sequences of its corresponding AIV isolate. 2) Depths of coverage were normalized by dividing raw pre- and post-capture read depths by the total reads in the corresponding pre- and post-capture pools (Table S2). 3) For each library, fold-enrichment at each nucleotide position was calculated by dividing the normalized post-capture read depth by

the normalized pre-capture read depth. 4) For each AIV isolate, mean fold-enrichment was calculated at every nucleotide position from the fold-enrichment values of its three independently captured replicate libraries. 5) Mean fold-enrichment values and their standard deviations were used to determine if significant enrichment had occurred at all nucleotide positions using a one-sample T-test against the fixed value of one-fold enrichment with an alpha level of 5%.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-022-08790-4>.

Additional file 1.

Additional file 2.

Additional file 3.

Additional file 4.

Additional file 5.

Additional file 6.

Additional file 7.

Acknowledgements

We would like to acknowledge the efforts of all laboratories world-wide who have submitted genomic sequences to the Influenza Research Database. Dr. Yohannes Berhane and Matthew Suderman at the Canadian Food Inspection Agency's National Centre for Animal Disease were instrumental in providing diverse egg-cultured AIV validation material from wild birds and poultry. We also thank Dr. Agatha Jassem at the British Columbia Centre for Disease Control's Public Health Laboratory and Dr. Nathalie Bastien at the Public Health Agency of Canada's National Microbiology Laboratory for providing H5N1 and H7N9 validation material from human infections. Additionally, we thank Tracy Lee at the British Columbia Centre for Disease Control's Public Health Laboratory for providing primers used to generate cDNA from AIV egg-cultures.

Authors' contributions

KK designed and implemented the ProbeTools algorithms, wrote the ProbeTools source code, designed the AIV_v1 probe panel, prepared sequencing libraries, performed probe captures and in-house sequencing, analyzed the data, and wrote the manuscript. JD performed preliminary studies with k-mer clustering, assisted with the design and implementation of the ProbeTools algorithms, and provided guidance on bioinformatic data analysis. CH helped assemble the validation collection of egg-cultured AIV isolates, ensured relevant strains were included, and provided direction for AIV probe panel design to ensure its suitability for agricultural surveillance applications. WH provided guidance on implementing ProbeTools algorithms, best practices for constructing and distributing bioinformatics tools and packages, and bioinformatic data analysis. NP provided guidance on experiment design for in vitro captures, troubleshooting for library preparation, probe capture, and sequencing of egg-cultured AIV isolates, and provided direction for AIV probe panel design to ensure its suitability for public health surveillance applications. All authors reviewed and contributed comments on the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was funded through research grants from Genome British Columbia (UPP025), Investment Agriculture Foundation of British Columbia (A0822), and the CANARIE Research Software Program (RS3–073).

Availability of data and materials

ProbeTools v0.0.5 source code, which was used to design the final probe panel and assess its coverage of target sequences in silico for this manuscript, is available on GitHub at <https://github.com/Kevinkuchinski/ProbeTools>. FASTA

files of the HA, NA, and M genome segment reference sequences used as a target space for design and assessment in this manuscript (described in Table 2) are provided as part of the ProbeTools v0.0.5 release. The sequences of the AIV_v1 probe panel are also provided as part of the ProbeTools v0.0.5 release, and they are also included in this manuscript's supplemental information as Supplemental Material 1. Data from the in vitro captures are provided in BAM format with pre- and post-capture libraries mapped to the HA, NA, and M genome segment sequences of their corresponding egg-cultured AIV isolate. These can be accessed from the NCBI Short Read Archive as part of BioProject PRJNA796698. Total read counts used to normalize depths of coverage in these libraries are provided in the manuscript's supplemental material as Table S2.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada. ²Vancouver, Canada. ³Animal Health Centre, British Columbia Ministry of Agriculture, Food, and Fisheries, Abbotsford, British Columbia, Canada. ⁴School of Population and Public Health, University of British Columbia, Vancouver, British Columbia, Canada. ⁵Faculty of Health Sciences, Simon Fraser University, Burnaby, British Columbia, Canada. ⁶Public Health Laboratory, British Columbia Centre for Disease Control, Vancouver, British Columbia, Canada.

Received: 25 February 2022 Accepted: 18 July 2022

Published online: 12 August 2022

References

- Fitzpatrick AH, Rupnik A, O'Shea H, Crispie F, Keaveney S, Cotter P. High throughput sequencing for the detection and characterization of RNA viruses. *Front Microbiol.* 2021;12:621719.
- Xiao M, Liu X, Ji X, Li M, Li J, Yang L, et al. Multiple approaches for massively parallel sequencing of SARS-CoV-2 genomes directly from clinical samples. *Genome Med.* 2020;12(1):57.
- Houldcroft CJ, Beale MA, Breuer J. Clinical and biological insights from viral genome sequencing. *Nat Rev Microbiol.* 2017;15(3):183–92.
- Depledge DP, Palsler AL, Watson SJ, Lai IY, Gray ER, Grant P, et al. Specific capture and whole-genome sequencing of viruses from clinical samples. *PLoS One.* 2011;6(11):e27805.
- Paskey AC, Frey KG, Schroth G, Gross S, Hamilton T, Bishop-Lilly KA. Enrichment post-library preparation enhances the sensitivity of high-throughput sequencing-based detection and characterization of viruses from complex samples. *BMC Genomics.* 2019;20(1):155.
- Brown JR, Roy S, Ruis C, Yara Romero E, Shah D, Williams R, et al. Norovirus whole-genome sequencing by SureSelect target enrichment: a robust and sensitive method. *J Clin Microbiol.* 2016;54(10):2530–7.
- Wylezich C, Calvelage S, Schlottau K, Ziegler U, Pohlmann A, Höper D, et al. Next-generation diagnostics: virus capture facilitates a sensitive viral diagnosis for epizootic and zoonotic pathogens including SARS-CoV-2. *Microbiome.* 2021;9(1):51.
- Wylie TN, Wylie KM, Herter BN, Storch GA. Enhanced virome sequencing using targeted sequence capture. *Genome Res.* 2015;25(12):1910–20.
- O'Flaherty BM, Li Y, Tao Y, Paden CR, Queen K, Zhang J, et al. Comprehensive viral enrichment enables sensitive respiratory virus genomic identification and analysis by next generation sequencing. *Genome Res.* 2018;28(6):869–77.
- Bonsall D, Ansari MA, Ip C, Trebes A, Brown A, Klenerman P, et al. Ve-SEQ: robust, unbiased enrichment for streamlined detection and whole-genome

- sequencing of HCV and other highly diverse pathogens. *F1000Res*. 2015;4:1062.
11. Briese T, Kapoor A, Mishra N, Jain K, Kumar A, Jabado OJ, et al. Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio*. 2015;6(5):e01491–15.
 12. Xiao Y, Nolting JM, Sheng ZM, et al. Design and validation of a universal influenza virus enrichment probe set and its utility in deep sequence analysis of primary cloacal swab surveillance samples of wild birds. *Virology*. 2018;524:182–91.
 13. Metsky HC, Siddle KJ, Gladden-Young A, Qu J, Yang DK, Brehio P, et al. Capturing sequence diversity in metagenomes with comprehensive and scalable probe design. *Nat Biotechnol*. 2019;37(2):160–8.
 14. Chafin TK, Douglas MR, Douglas ME. MrBait: universal identification and design of targeted-enrichment capture probes. *Bioinformatics*. 2018;34(24):4293–6.
 15. Beliveau BJ, Kishi JY, Nir G, Sasaki HM, Saka SK, Nguyen SC, et al. OligoMiner provides a rapid, flexible environment for the design of genome-scale oligonucleotide in situ hybridization probes. *Proc Natl Acad Sci U S A*. 2018;115(10):E2183–92.
 16. Mayer C, Sann M, Donath A, Meixner M, Podsiadlowski L, Peters RS, et al. BaitFisher: a software package for multispecies target DNA enrichment probe design. *Mol Biol Evol*. 2016;33(7):1875–86.
 17. Kushwaha SK, Manoharan L, Meerupati T, Hedlund K, Ahrén D. MetCap: a bioinformatics probe design pipeline for large-scale targeted metagenomics. *BMC Bioinformatics*. 2015;16(1):65.
 18. Dugan VG, Chen R, Spiro DJ, et al. The evolutionary genetics and emergence of avian influenza viruses in wild birds. *PLoS Pathog*. 2008;4(5):e1000076 Published 2008 May 30.
 19. Wille M, Tolf C, Avrii A, Latorre-Margalef N, Wallerström S, Olsen B, et al. Frequency and patterns of reassortment in natural influenza A virus infection in a reservoir host. *Virology*. 2013;443(1):150–60. <https://doi.org/10.1016/j.virol.2013.05.004> Epub 2013 May 28.
 20. Verhagen JH, Fouchier RAM, Lewis N. Highly pathogenic avian influenza viruses at the wild-domestic bird Interface in Europe: future directions for research and surveillance. *Viruses*. 2021;13(2):212 Published 2021 Jan 30.
 21. Widdowson MA, Bresee JS, Jernigan DB. The global threat of animal influenza viruses of zoonotic concern: then and now. *J Infect Dis*. 2017;216(suppl_4):S493–8.
 22. Mostafa A, Abdelwhab EM, Mettenleiter TC, Pleschka S. Zoonotic potential of influenza A viruses: a comprehensive overview. *Viruses*. 2018;10(9):497 Published 2018 Sep 13.
 23. Sutton TC. The pandemic threat of emerging H5 and H7 avian influenza viruses. *Viruses*. 2018;10(9):461 Published 2018 Aug 28.
 24. Peiris JS, de Jong MD, Guan Y. Avian influenza virus (H5N1): a threat to human health. *Clin Microbiol Rev*. 2007;20(2):243–67.
 25. Watanabe T, Watanabe S, Maher EA, Neumann G, Kawaoka Y. Pandemic potential of avian influenza A (H7N9) viruses. *Trends Microbiol*. 2014;22(11):623–31.
 26. Nuñez IA, Ross TM. A review of H5Nx avian influenza viruses. *Ther Adv Vaccines Immunother*. 2019;7:2515135518821625 Published 2019 Feb 22.
 27. Macias AE, McElhaney JE, Chaves SS, Nealon J, Nunes MC, Samson SJ, et al. The disease burden of influenza beyond respiratory illness. *Vaccine*. 2021;39(Suppl 1):A6–A14.
 28. Lafond KE, Porter RM, Whaley MJ, Suizan Z, Ran Z, Aleem MA, et al. Global burden of influenza-associated lower respiratory tract infections and hospitalizations among adults: a systematic review and meta-analysis. *PLoS Med*. 2021;18(3):e1003550.
 29. Gordon A, Reingold A. The burden of influenza: a complex problem. *Curr Epidemiol Rep*. 2018;5(1):1–9.
 30. Sellers SA, Hagan RS, Hayden FG, Fischer WA 2nd. The hidden burden of influenza: a review of the extra-pulmonary complications of influenza infection. *Influenza Other Respir Viruses*. 2017;11(5):372–93.
 31. GBD 2017 Influenza Collaborators. Mortality, morbidity, and hospitalisations due to influenza lower respiratory tract infections, 2017: an analysis for the global burden of disease study 2017. *Lancet Respir Med*. 2019;7(1):69–89.
 32. Global Consortium for H5N8 and Related Influenza Viruses. Role for migratory wild birds in the global spread of avian influenza H5N8. *Science*. 2016;354(6309):213–7.
 33. Runstadler J, Hill N, Hussein IT, Puryear W, Keogh M. Connecting the study of wild influenza with the potential for pandemic disease. *Infect Genet Evol*. 2013;17:162–87.
 34. Spackman E, Senne DA, Myers TJ, et al. Development of a real-time reverse transcriptase PCR assay for type A influenza virus and the avian H5 and H7 hemagglutinin subtypes. *J Clin Microbiol*. 2002;40(9):3256–60. <https://doi.org/10.1128/JCM.40.9.3256-3260.2002>.
 35. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451(7181):990–3.
 36. Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C, et al. Global rise in human infectious disease outbreaks. *J R Soc Interface*. 2014;11(101):20140950. <https://doi.org/10.1098/rsif.2014.0950>.
 37. Carroll D, Daszak P, Wolfe ND, Gao GF, Morel CM, Morzaria S, et al. The global Virome project. *Science*. 2018;359(6378):872–4. <https://doi.org/10.1126/science.aap7463>.
 38. Lipkin WI, Firth C. Viral surveillance and discovery. *Curr Opin Virol*. 2013;3(2):199–204. <https://doi.org/10.1016/j.coviro.2013.03.010> Epub 2013 Apr 17.
 39. Gardy JL, Loman NJ. Towards a genomics-informed, real-time, global pathogen surveillance system. *Nat Rev Genet*. 2018;19(1):9–20. <https://doi.org/10.1038/nrg.2017.88> Epub 2017 Nov 13.
 40. Kress WJ, Mazet JAK, Hebert PDN. Opinion: intercepting pandemics through genomics. *Proc Natl Acad Sci U S A*. 2020;117(25):13852–5. <https://doi.org/10.1073/pnas.2009508117> Epub 2020 Jun 3.
 41. Khoury MJ, Holt KE. The impact of genomics on precision public health: beyond the pandemic. *Genome Med*. 2021;13(1):67. Published 2021 Apr 23. <https://doi.org/10.1186/s13073-021-00886-y>.
 42. Grad YH, Lipsitch M. Epidemiologic data and pathogen genome sequences: a powerful synergy for public health. *Genome Biol*. 2014;15(11):538. <https://doi.org/10.1186/s13059-014-0538-4>.
 43. Sintchenko V, Holmes EC. The role of pathogen genomics in assessing disease transmission. *BMJ*. 2015;350:h1314. <https://doi.org/10.1136/bmj.h1314>.
 44. Armstrong GL, MacCannell DR, Taylor J, Carleton HA, Neuhaus EB, Bradbury RS, et al. Pathogen genomics in public health. *N Engl J Med*. 2019;381(26):2569–80.
 45. Tyson JR, James P, Stoddard D, Sparks N, Wickenhagen A, Hall G, et al. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *bioRxiv*. 2020;2020:283077.
 46. Freed NE, Vlková M, Faisal MB, Silander OK. Rapid and inexpensive whole-genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford Nanopore rapid barcoding. *Biol Methods Protoc*. 2020;5(1):bpaa014.
 47. Kuchinski KS, Nguyen J, Lee TD, Hickman R, Jassem AN, Hoang LMN, et al. Mutations in emerging variant of concern lineages disrupt genomic sequencing of SARS-CoV-2 clinical specimens. *Int J Infect Dis*. 2022;114:51–4.
 48. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ*. 2016;4:e2584.
 49. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics*. 2009;10:421.
 50. Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, et al. Influenza research database: an integrated bioinformatics resource for influenza virus research. *Nucleic Acids Res*. 2017;45(D1):D466–74.
 51. Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, et al. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza A viruses. *J Virol*. 2009;83(19):10309–13.

Publisher's Note

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