Development of an Influenza A Sequencing Workflow on Ion PGM™ Sequencer for Improved Surveillance

LIFE TECHNOLOGIES™
ANIMAL HEALTH

Angela M. Burrell, Craig Cumming, Wayne Ge, and Cate O'Connell, Life Technologies 2130 Woodward Street, Austin, Texas 78744

ABSTRACT

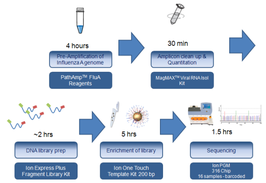
Complete genome sequencing is crucial for ongoing identification, surveillance and characterization of Influenza A. When sequencing viral genomes, background host nucleic acids may be co-processed during library preparation resulting in a sequencing reaction in which a majority of reads are taken up by the host genome. One solution to this problem is to run a pre-amplification RT-PCR on the extracted nucleic acid with primers designed to amplify only the viral genome prior to library preparation.

We have developed the PathAmp™ FluA Reagents, a set of highly-specific primers and high-fidelity master mix for the amplification of all eight influenza genomic segments to obtain cDNA suitable for direct input into library preparation.

Performance was tested on a diverse panel of avian and swine influenza viral strains (n=16) with the PathAmp™ reagents and a collaborator's pre-amplification workflow on the Ion PGM™ Sequencer. The genomic fraction covered by reads significantly improved when the PathAmp™ reagents were used. Mean coverage depth was approximately 700-fold higher and host genome contamination dropped from >26% to <0.25% with the PathAmp™ reagents.

Sensitivity of the PathAmp™ reagents was tested by serially diluting swine influenza virus in porcine nasal swabs and tonsil tissue. Lineage calls for both sample matrices were correct down to 200 viral copies demonstrating the sensitivity of this technique. In conclusion, the PathAmp™ FluA Reagents provide a rapid, accurate, and sensitive solution for Influenza sequencing on the Ion PGM™ allowing for faster responses to emerging strains.

Figure 1. PathAmp™ Influenza A Whole Genome Sequencing Workflow



INTRODUCTION

Influenza A is a negative-sense RNA virus and a major source of economic loss in the animal health field. Complete genome sequencing is crucial for ongoing surveillance of influenza. It provides detailed information about virus origin and evolution, which is particularly important in identifying emerging strains. Sequencing allows for detection of small mutations across the genome as well as monitoring for larger genetic re-assortments. The information can also be used to characterize the virus for vaccine development and to provide information about antiviral resistance.

When sequencing viral nucleic acid, high amounts of background host nucleic acids may be co-processed during library preparation, resulting in a sequencing reaction in which a majority of reads are taken up by the host genome. A solution to this problem is to run a pre-amplification RT-PCR on the extracted nucleic acid with primers specifically designed to amplify only the viral nucleic acid prior to library preparation.

We have developed the PathAmp™ FluA Reagents, a set of highly specific, universal primers along with high-fidelity enzymes and buffers for the amplification of all eight Influenza A genomic segments, which range in size from 900 bp to 2.4 kbp. cDNA obtained from the PathAmp™ workflow is suitable for direct input into the Ion Xpress™ Plus library preparation kit.

MATERIALS AND METHODS

Reagent performance was verified on a diverse panel of 16 influenza A strains (Figure 2) from both avian and swine sources. The PathAmp™ reagent workflow was compared to another published pre-amplification workflow on the same panel of samples. For both workflows, samples were barcoded allowing for all 16 to be run on a single sequencing reaction saving time and money. Average coverage was greater than 95% with the PathAmp™ reagent workflow compared to just 89% for the same samples processed with the collaborator workflow (Figure 3). Mean coverage depth was approximately 700-fold higher for the PathAmp™ reagents as compared to the published workflow. (Figure 4) Host contamination also dropped from 26% down to less than 0.25% when the PathAmp™ reagents were used (Figure 5).

Sensitivity of the PathAmp™ reagents was tested by serially diluting a swine influenza virus sample in porcine nasal swab and tonsil tissue nucleic acid. Lineage calls for both sample matrices were correct down to 200 viral copies. Greater than 99% of each segment was covered by reads when virus was present at 1000 copies or 200 copies when diluted in tonsil or nasal swab nucleic acid, respectively (Figure 6). In conclusion, the PathAmp™ FluA Reagents provide a rapid, accurate, and sensitive solution for Influenza A viral sequencing on the Ion PGM™ Sequencer.

A prototype bioinformatics workflow was developed to easily analyze next-generation sequencing results in order to obtain lineage calls. Reads were mapped to a set of 185 genome segment reference sequences, one representing each lineage of all 8 segments. The depth of coverage at each of the reference sequences was recoded. The segment lineage with the highest average coverage was identified resulting in an output report of top hits for each segment. A consensus sequence was generated for each of the best matching segments and the FASTA file was uploaded to flugenome.org for further analysis.

Figure 2. Influenza A Subtypes Tested

Subtypes tested by traditional sequencing	Subtypes tested by Ion PGM
Subtype	Subtype
Swine H1N1	Swine H1N2
Avian H5N2	Swine H3N2
Avian H7N1	Swine H1N1
Avian H7N4	Avian H7N9
Avian H5N3	Avian H9N2
Avian H7N7	Avian H5N2
Avian H7N3	Avian H5N9
Avian H4N6	Avian H3N6
Equine H3N8	Avian H8N4
Equine H2N3	Avian H10N7
Avian H6N5	
Avian H11N9	
Avian H10N7	

Performance was verified on 20 Influenza A subtypes with the PathAmp™ reagents. A panel 16 diverse samples from10 influenza subtypes (both avian and swine strains) were successfully tested on the Ion PGM™ Sequencer. All 16 cultured samples were barcoded and run in a single sequencing reaction, saving both time and reagents. Subtyping lineage calls were 100% concordant with real-time PCR subtypes and Sanger sequencing results.

RESULTS

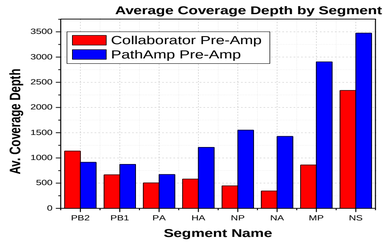
Figure 3. PathAmp™ and collaborator workflow segment coverage.

Collaborator Workflow															
	Swine isolates								Avian isolates						
Reported subtype	H1N?	H3N2	H3N2	H1N1	H3N2	H1N?	H1N1	H3N2	H7N9	H9N2	H5N2	H5N2	H5N9	H3N6	H8N4
PB2	1	1	1	1	1	1	1	1	1	1	0.69	1	0.99	1	0.96
PB1	1	1	1	1	1	1	1	1	1	1	0.79	1	1	1	0.95
PA	1	1	1	1	1	1	1	1	0.97	0.999	0.63	0.87	1	0.96	0.96
HA	1	1	1	1	1	1	1	1	1	1	1	1	0.92	1	1
NP	1	1	1	1	1	1	1	1	0.98	1	1	1	0.95	1	1
NA	1	1	1	1	1	1	1	1	1	1	0.99	0.98	1	1	1
MP	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NS	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Subtype by sequencing	H1N2	H3N2	H3N2	H1N1	H3N2	H1N2	H1N1	H3N2	H7N9	H9N2	H5N2	H5N2	H5N9	H3N6	H8N4

PathAmp FluA Workflow															
	Swine isolates								Avian isolates						
Reported subtype	H1N?	H3N2	H3N2	H1N1	H3N2	H1N?	H1N1	H3N2	H7N9	H9N2	H5N2	H5N2	H5N9	H3N6	H8N4
PB2	1	1	1	1	1	1	1	1	1	1	1	1	0.996	1	1
PB1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
PA	1	1	1	1	1	1	1	1	0.99	1	0.8789	0.872	1	1	0.962
HA	1	1	1	1	1	1	1	1	1	1	1	1	0.873	1	1
NP	1	1	1	1	1	1	1	1	1	1	1	1	0.988	1	1
NA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
MP	1	1	1	1	1	1	1	1	0.994	1	1	1	1	1	1
NS	1	1	1	1	1	1	1	1	0.999	1	1	1	1	1	1
Subtype by sequencing	H1N2	H3N2	H3N2	H1N1	H3N2	H1N2	H1N1	H3N2	H7N9	H9N2	H5N2	H5N2	H5N9	H3N6	H8N4

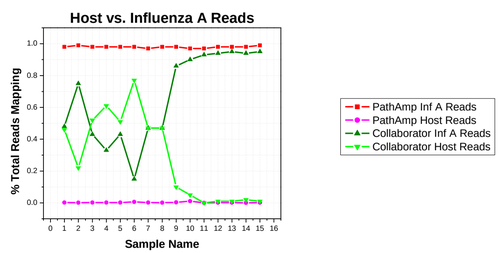
Figure 3 shows a comparison of the % segment coverage for each sample tested with the PathAmp™ and collaborator workflow. Improved coverage was observed with the PathAmp™ workflow when testing the AIV samples. Average coverage was greater than 95% with the PathAmp™ reagent workflow compared to just 89% for the same samples processed with the collaborator workflow

Figure 4. Average Coverage Depth by Segment



Comparison of coverage depth for all samples shows that for both the collaborator and PathAmp™ pre-amp protocols, coverage depth varies inversely with segment length (shorter segments tending to have more coverage). Mean coverage depth for all segments was ~1630 for the PathAmp™ kit vs. ~ 860 for the collaborator workflow.

Figure 5. % Host Contamination of PathAmp™ and Collaborator Pre-Amp Influenza A Workflows



When all reads were mapped to both Influenza A and the host genome, significantly fewer reads were taken up by the host genome when the PathAmp™ workflow was utilized, especially in swine influenza samples (samples 1-8).

Figure 6. SIV Detection Sensitivity Lineage Calls

Sample	Barcode	Approx Copy #	SIV CT	Lineage Call	Total # reads	% Reads matching Inf A	% Reads matching Host	% Fraction Covered	Av. Read Depth
SIV diluted in porcine tonsil tissue	1	~40,000,000	13.97	*	65,902	98.8%	0.3%	100%	810
	2	662126	20.64	*	155,996	97.6%	1.6%	100%	1803
	3	78440	23.99	*	180,118	92.0%	6.4%	100%	1921
	4	9191	27.37	*	82,189	60.5%	33.3%	100%	563
	5	1182	30.60	*	162,057	7.1%	78.3%	100%	145
	6	222	33.23	*	174,372	0.3%	81.6%	86%	3
	7	10	38.14	**1/8 incorrect**	97,842	0.2%	84.2%	35%	0
	8	0	Undetermined	**Multiple incorrect**	162,430	0.2%	84.0%	19%	0
SIV diluted in porcine nasal swabs	9	~40,000,000	13.97	*	160,450	99.1%	0.2%	100%	1964
	10	791955	20.35	*	223,193	98.7%	0.1%	100%	2536
	11	89426	23.79	*	172,326	98.6%	0.1%	100%	2030
	12	10083	27.22	*	181,147	95.2%	0.7%	100%	2011
	13	1486	30.24	*	184,680	76.4%	2.9%	100%	1702
	14	201	33.39	*	68,766	8.3%	17.1%	100%	63
	15	9	38.29	**1/8 incorrect**	159,746	1.8%	12.6%	77%	25
	16	0	Undetermined	**Multiple incorrect**	84,073	0.3%	14.5%	32%	3

* Lineage calls for all segments were correct.

Sensitivity of the PathAmp™ reagents was tested by serially diluting a swine influenza virus sample in porcine nasal swab and tonsil tissue. Lineage calls for both sample matrices were correct down to 200 viral copies. >99% of all SIV segments are covered by reads down to the 1:100,000X for the tonsil dilutions and 1:1,000,000X for the nasal swab dilutions This corresponds to a approximately 1,000 copies and 200 copies, respectively. We would like to incorporate a quality score or "no-call" into the analysis pipeline to enable the ability to differentiate between low positive samples and Influenza-negative samples. Average read depth for correct calls was >50X with >85% fraction covered.

CONCLUSIONS

The PathAmp™ FluA workflow accuracy was verified internally and externally on a diverse panel of influenza subtypes. Sensitivity of the sequencing workflow was demonstrated to provide accurate lineage calls down to approximately 200 viral copies in the presence of high amounts of contaminating host nucleic acid. The workflow was shown to significantly decrease the amount of contaminating host nucleic acid reads to improve influenza coverage during the sequencing reaction. The average read depth and coverage was significantly improved as compared to other collaborator workflows demonstrating that the PathAmp™ FluA sequencing workflow provides a robust solution for Influenza A virus sequencing.

Future improvements being considered include the incorporation of "no-call" to low-quality sequences and low titer samples as well as improvements to the analysis method to allow for detection of co-infections.

In conclusion, the PathAmp™ FluA Reagents provide a rapid, accurate, and sensitive solution for Influenza A viral sequencing on the Ion PGM™ Sequencer.

REFERENCES

1. E. Hoffmann et al,. (2001) Universal Primer Set for the Full-Length Amplification of all Influenza A Viruses. Arch Virol146: 2275–2289
2. Bin Zhou et al,. (2009) Single-Reaction Genomic Amplification Accelerates Sequencing and Vaccine Production for Classical and Swine Origin Human Influenza A Viruses. *J. Virol, 83(19):10309.*

ACKNOWLEDGEMENTS

We would like to acknowledge Dr. Amy Vincent and Dr. Mary Lea Killian from the USDA for providing samples for us to test.

TRADEMARKS/LICENSING