Journal of Virological Methods 266 (2019) 30–33

Contents lists available at ScienceDirect

Journal of Virological Methods

journal homepage: www.elsevier.com/locate/jviromet

Evaluation of two workflows for whole genome sequencing-based typing of influenza A viruses

Daniel Wüthrich[a,b,c], Daniela Lang[b], Nicola F. Müller[c,d], Richard A. Neher[c,e], Tanja Stadler[c,d], Adrian Egli[a,b,*]

[a] Division of Clinical Microbiology, University Hospital Basel, Basel, Switzerland
[b] Applied Microbiology Research, Department of Biomedicine, University of Basel, Basel, Switzerland
[c] Swiss Institute of Bioinformatics, Basel, Switzerland
[d] Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland
[e] Biozentrum, University of Basel, Basel, Switzerland

ARTICLE INFO

Keywords:
Whole genome sequencing
Influenza A
Virus
Typing
Method

ABSTRACT

We compared two sample preparation protocols for whole genome sequencing of influenza A viruses. Each protocol was assessed using cDNA quantity and quality and the resulting mean genome coverage after sequencing. Both protocols produced acceptable result for samples with high viral load, whereas one protocol performed slightly better with limited virus count.

Genome sequences are increasingly used to understand and monitor the transmission events of influenza viruses in various scenarios such as long term evolution in immunocompromised hosts (Xue et al., 2017), hospital epidemiology (Houghton et al., 2017; Pagani et al., 2015; Valley-Omar et al., 2015), transmissions within households (McCrone et al., 2018), and public health e.g. within larger communities (Ghedin et al., 2005; Virk et al., 2017). Most of these studies have been performed in retrospect using batch-wise sequencing of previously collected and stored samples. In the past, most studies have used Sanger sequencing of single viral RNA segments such as the haemagglutinin segment (Houghton et al., 2017; Pagani et al., 2015).

In recent years, whole genome sequencing (WGS) has become the method of choice for high-resolution typing of pathogens (Pallen et al., 2010), including influenza viruses. Multiple protocols have been described for WGS based typing of influenza A (Hoffmann et al., 2001; Meinel et al., 2018; Zhou and Wentworth, 2012) and influenza B (Oong et al., 2017; Vijaykrishna et al., 2015) viruses. The lack of standardization between WGS protocols and recommendation for influenza virus typing is a problem for routine diagnostic laboratories. RNA concentrations and stability is an additional important challenge in biobanked samples (Forster et al., 2008). The number of isolates that failed in the preparation of particular protocols is often not reported in publications. However, we expect, based on our own experience, that a substantial number of viruses from frozen swabs (between 10% and 20%) cannot be WGS-typed due to quality reasons.

In this study, we assess the performance of two previously published WGS based typing protocols for influenza A viruses (Method A (Meinel et al., 2018) and Method B (Zhou and Wentworth, 2012)). We chose two Qiagen-based methods as the related products are well established in our laboratory. We compared the two protocols in order to select the most suitable method for a large-scale sequencing of influenza A viruses. During the influenza season of 2016/2017, we collected nasopharyngeal swabs (Universal Transport Medium, UTM-RT, Copan) from patients with influenza like illness in the City of Basel and surrounding areas. All samples were confirmed to be influenza A positive by using the Xpert Xpress Flu/RSV (Cepheid, Sunnyvale, CA) assay. This assay provides a Ct-value that allows a semi quantitative estimation of the viral load of positive samples. Positive samples were aliquoted in 1 ml volume in 1.5 ml tubes (Sarstedt) and frozen at −80 °C until batch-wise sample processing was performed.

We used twelve isolates of the collection to directly compare the two sequencing protocols. We selected these samples to reflect a wide range of Ct values from influenza A specific PCR. Briefly, both protocols start with an RNA isolation step, followed by a RT-PCR and a PCR clean-up before WGS. The RT-PCR step transcribes all eight viral RNA segments into cDNA and amplifies the viral genome. A detailed comparison of the methods can be seen in Table 1. The resulting PCR products were analysed using Tape Station (Agilent Velocity, Lab901

* Corresponding author at: Division of Clinical Microbiology, University Hospital Basel, Petersgraben 4, 4031 Basel, Switzerland.
E-mail address: adrian.egli@usb.ch (A. Egli).

https://doi.org/10.1016/j.jviromet.2019.01.009
Received 12 October 2018; Received in revised form 4 January 2019; Accepted 17 January 2019
Available online 21 January 2019

Table 1
Comparison of the influenza whole genome amplification protocols.

Step	Method A (Meinel et al., 2018)	Method B Modified from (Zhou and Wentworth, 2012)
	RNA extraction	
Sample input	100 µl	100 µl
RNA extraction	RNeasy Plus Mini Kit [a] (Purification of total RNA from animal cells (protocol): gDNA elimination.)	RNeasy Mini Kit with QIAcube (Purification of total RNA containing small RNAs from cells.)
Elution volume	50 µl	30 µl
	RT-PCR	
RNA input	5 µl	2.5 µl
Polymerase used	SuperScript III One-Step RT-PCR System with Platinum *Taq* DNA Polymerase	SuperScript III One-Step RT-PCR System with Platinum *Taq* High Fidelity DNA Polymerase
Reaction volume	50 µl RT-PCR volume	25 µl RT-PCR volume
Forward primer(s)	5'-ACGCGTGATCAGCAAAAGCAGG-3'	5'-GGGGGGAGCAAAAGCAGG-3' 5'-GGGGGGAGCGAAAGCAGG-3'
Reverse primer	5'-ACGCGTGATCAGTAGAAACAAGG-3'	5'-ACGGGTTATTAGTAGAAACAAGG-3'
Thermocycler	55 °C for 40 min; 94 °C for 2 min [94 °C 15 s, 52 °C 30 s, 68 °C 3 min] × 35 [b] 68 °C for 5 min, 4 °C hold	42 °C for 60 min; 94 °C for 2 min [94 °C 30 s, 44 °C 30 s, 68 °C 3 min 30 s] × 5 [94 °C 30 s, 52 °C 30 s, 68 °C 3 min 30 s] × 30 68 °C for 10 min, 4 °C hold
PCR clean-up	AMPure XP protocol	AMPure XP protocol

[a] As we had no Hamilton Microlab Star device, we used a similar manual protocol.
[b] Number of cycles was reduced to from 40 to 35 in comparison to the publication.

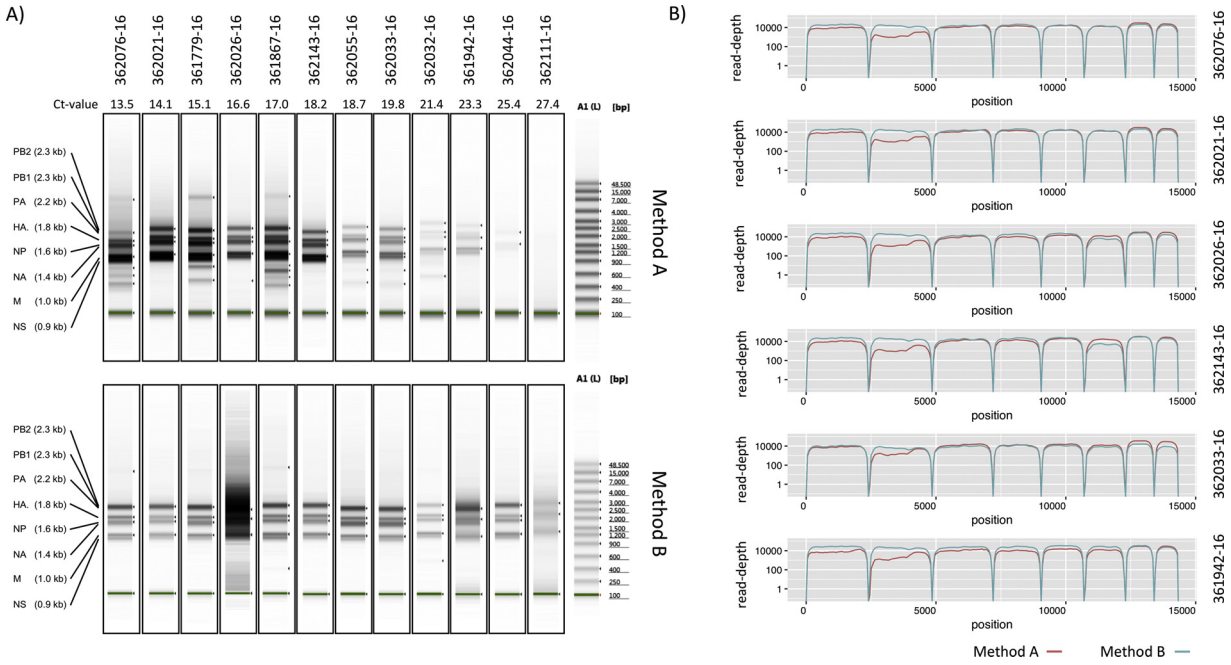

Fig. 1. Quality assessments of RT-PCR products. A) The RT-PCR products were analysed using Tape Station. Each sample is showed with respective Ct-values and a DNA size ladder. Bands show different amplified DNA segments. B) Read-depth of six isolates using both methods (method A in red, method B in blue) is shown. The y-axis shows the relative position on the concatenated genome. The x-axis shows the read-depth in log scale. Each row corresponds to a different isolate (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

Tape Station). This allows the evaluation of the PCR specificity and yield by inspecting the band patterns and their intensity (Fig. 1a). After the final PCR clean-up, the DNA concentrations from both protocols were fluorometrically quantified using a Qubit system (Thermofisher, dsDNA HS Assay Kit) (Table 2).

Out of these twelve Isolates, we selected six for whole genome sequencing that had a DNA concentration of more than 2 ng/µl in both methods. Additionally, these six samples represent a similar range of Ct-values as determined by the Xpert Xpress Flu/RSV assay compared to all twelve samples. We first fragmented the PCR products and processed them into a library by using the Nextera XT protocol (according to manufacturer's instructions). We next sequenced the fragmented and processed products using a MiSeq device (2 × 300 bp) at the NGS core facility at the Clinical Microbiology of the University Hospital of Basel (ISO-accredited (ISO/EC 17025)). The resulting reads were mapped against the influenza A (A/New York/392/2004(H3N2)) virus genome using BWA (Li and Durbin, 2009). Read-depth and genome consensus sequence was called using Pilon (Walker et al., 2014). A phylogenetic tree of all Isolates was constructed using FastTree (Price et al., 2010).

We found that the average DNA quantity was higher in samples processed with method B compared to samples processed with method A (25.44 vs. 13.55 ng/µl). In addition, method B provided a better DNA quality in comparison to method A, as the band pattern were more distinct and less unspecific bands appeared (Fig. 1a). Furthermore, the band intensities were more even across all samples. Although visual inspection did not allow a quantifiable scoring of the DNA quality, the

Table 2
Summary of RT-PCR and NGS analysis.

Sample ID	Ct-value	DNA concentration [ng/µl]		Average read-depth NGS		Undetermined Bases	
Method	Both	Method A	Method B	Method A	Method B	Method A	Method B
362076	13.5	33.4	29.2	10,103	13,243	26	0
362021	14.1	35	27.2	9054	13,524	35	13
361779	15.1	24.4	25.8	Not performed	Not performed	Not performed	Not performed
362026	16.6	12.3	29.6	10,395	15,089	54	5
361867	17	25.8	26	Not performed	Not performed	Not performed	Not performed
362143	18.2	21	33.6	10,236	7754	34	15
362055	18.7	2.66	30.2	Not performed	Not performed	Not performed	Not performed
362033	19.8	3.18	29	9606	20,543	40	1
362032	21.4	1.2	17	Not performed	Not performed	Not performed	Not performed
361942	23.3	2.88	27	9473	13,169	57	17
362044	25.4	0.624	22.4	Not performed	Not performed	Not performed	Not performed
362111	27.4	0.186	8.24	Not performed	Not performed	Not performed	Not performed
Average	19	14	25	9811	13,887	41	8.5

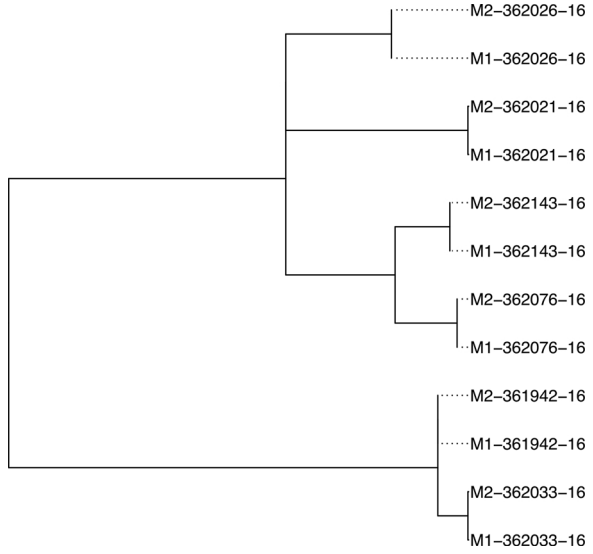

Fig. 2. WGS-based phylogenetic tree. The phylogenetic tree is based on the concatenated sequences of all eight viral segments and indicates the relationship between the different isolates. Sample names starting with M1 or M2 indicate if they were processed using Method A or Method B, respectively.

difference of the band patterns quality of the methods was striking.

The comparison of the Ct-values and DNA concertation showed that a higher Ct-value (smaller viral load) resulted in a lower DNA yield (Table 2). Furthermore, we also found that a high Ct-value resulted in weaker band pattern in the Tape Station analysis (Fig. 1a). These trends were far more pronounced in method A than method B.

After WGS and mapping of the reads, the read-depth distribution was analysed as a surrogate for the DNA quantity and quality. On average, the samples prepared with method A showed an average read-depth of 9811-fold, whereas samples with method B had an average read-depth of 13,887-fold (Table 2). Comparing the read-depth among the eight segments revealed that method A had a reduced read-depth of the NA Segment (PB1, NC_007372.1) (Fig. 1b). This trend was more pronounced in samples with low viral load (high Ct-values). Additionally, we also counted the sites that could not be used to generate a consensus sequence for the genome by Pilon (default parameter). This resulted on average in 41 and 8.5 not determined bases in the genome for the samples processed with method A and method B, respectively (Table 2). However, in the phylogenetic analysis the two methods result in the identical results (Fig. 2).

Poor sample quality may reflect a mixture of pre-analytical (collection, storage) and analytical (RNA-extraction, RT-PCR, library preparation) reasons: sample storage could result in disintegration and subsequently RNA degradation of particularly RNA viruses. In this study however, sample storage was the same for both methods. The protocols differed in the RNA extraction protocol and kit, primers used for the RT-PCR, the DNA Polymerase used in the PCRs, and the cycling conditions (Table 1). The bioinformatics analysis was the same for both methods.

In conclusion, both methods provided decent results that are comparable, however method B generated slightly more reliable results for WGS-based typing of influenza A viruses and more even read-depths across all segments. Improving WGS-based typing for influenza does not only include the analytical part investigated here, but also pre-analytical processes that potentially can be optimized and standardized. Ideally, the samples are sequenced shortly after collection and shared data for real-time surveillance. Therefore, we provide a detailed step-by-step protocol of the suggested method B in the supplementary material.

Conflict of interest

All authors declare no conflict of interest

Funding

The study has been funded by the Swiss National Science Foundation (interdisciplinary research grant CR32I3_166258/1). AE received a salary grant by the Swiss National Science Foundation (Ambizione PZ00P3_154709 /1).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jviromet.2019.01.009.

References

Forster, J.L., Harkin, V.B., Graham, D.A., McCullough, S.J., 2008. The effect of sample type, temperature and RNAlater™ on the stability of avian influenza virus RNA. J. Virol. Methods. https://doi.org/10.1016/j.jviromet.2007.12.020.

Ghedin, E., Sengamalay, N.A., Shumway, M., Zaborsky, J., Feldblyum, T., Subbu, V., Spiro, D.J., Sitz, J., Koo, H., Bolotov, P., Dernovoy, D., Tatusova, T., Bao, Y., St George, K., Taylor, J., Lipman, D.J., Fraser, C.M., Taubenberger, J.K., Salzberg, S.L., 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. Nature. https://doi.org/10.1038/nature04239.

Hoffmann, E., Stech, J., Guan, Y., Webster, R.G., Perez, D.R., 2001. Universal primer set for the full-length amplification of all influenza A viruses. Arch. Virol. https://doi.org/10.1007/s007050170002.

Houghton, R., Ellis, J., Galiano, M., Clark, T.W., Wyllie, S., 2017. Haemagglutinin and neuraminidase sequencing delineate nosocomial influenza outbreaks with accuracy equivalent to whole genome sequencing. J. Infect. https://doi.org/10.1016/j.jinf.2016.12.015.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

McCrone, J.T., Woods, R.J., Martin, E.T., Malosh, R.E., Monto, A.S., Lauring, A.S., 2018. Stochastic processes constrain the within and between host evolution of influenza virus. Elife. https://doi.org/10.7554/eLife.35962.

Meinel, D.M., Heinzinger, S., Eberle, U., Ackermann, N., Schönberger, K., Sing, A., 2018. Whole genome sequencing identifies influenza A H3N2 transmission and offers superior resolution to classical typing methods. Infection. https://doi.org/10.1007/s15010-017-1091-3.

Oong, X.Y., Ng, K.T., Tan, J.L., Chan, K.G., Kamarulzaman, A., Chan, Y.F., Sam, I.C., Tee, K.K., 2017. Whole-genome phylogenetic analysis of influenza B/Phuket/3073/2013-Like viruses and unique reassortants detected in Malaysia between 2012 and 2014. PLoS One. https://doi.org/10.1371/journal.pone.0170610.

Pagani, L., Thomas, Y., Huttner, B., Sauvan, V., Notaridis, G., Kaiser, L., Iten, A., Pittet, D., Harbarth, S., 2015. Transmission and effect of multiple clusters of seasonal influenza in a Swiss geriatric hospital. J. Am. Geriatr. Soc. https://doi.org/10.1111/jgs.13339.

Pallen, M.J., Loman, N.J., Penn, C.W., 2010. High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr. Opin. Microbiol. https://doi.org/10.1016/j.mib.2010.08.003.

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 - approximately maximum-likelihood trees for large alignments. PLoS One 5. https://doi.org/10.1371/journal.pone.0009490.

Valley-Omar, Z., Nindo, F., Mudau, M., Hsiao, M., Martin, D.P., 2015. Phylogenetic exploration of nosocomial transmission chains of 2009 influenza A/H1N1 among children admitted at red cross war memorial children's hospital, Cape Town, South Africa in 2011. PLoS One. https://doi.org/10.1371/journal.pone.0141744.

Vijaykrishna, D., Holmes, E.C., Joseph, U., Fourment, M., Su, Y.C.F., Halpin, R., Lee, R.T.C., Deng, Y.-M., Gunalan, V., Lin, X., Stockwell, T.B., Fedorova, N.B., Zhou, B., Spirason, N., Kühnert, D., Bošková, V., Stadler, T., Costa, A.-M., Dwyer, D.E., Huang, Q.S., Jennings, L.C., Rawlinson, W., Sullivan, S.G., Hurt, A.C., Maurer-Stroh, S., Wentworth, D.E., Smith, G.J.D., Barr, I.G., 2015. The contrasting phylodynamics of human influenza B viruses. Elife. https://doi.org/10.7554/eLife.05055.

Virk, R.K., Gunalan, V., Lee, H.K., Inoue, M., Chua, C., Tan, B.H., Tambyah, P.A., 2017. Molecular evidence of transmission of influenza A/H1N1 2009 on a university campus. PLoS One. https://doi.org/10.1371/journal.pone.0168596.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9. https://doi.org/10.1371/journal.pone.0112963.

Xue, K.S., Stevens-Ayers, T., Campbell, A.P., Englund, J.A., Pergam, S.A., Boeckh, M., Bloom, J.D., 2017. Parallel evolution of influenza across multiple spatiotemporal scales. Elife. https://doi.org/10.7554/eLife.26875.

Zhou, B., Wentworth, D.E., 2012. Influenza A virus molecular virology techniques. Methods Mol. Biol. 865, 175–192. https://doi.org/10.1007/978-1-61779-621-0_11.