JOURNAL OF CLINICAL MICROBIOLOGY, Mar. 2009, p. 674–679 Vol. 47, No. 3
0095-1137/09/$08.00+0 doi:10.1128/JCM.01028-08

Simple, Sensitive, and Swift Sequencing of Complete H5N1 Avian Influenza Virus Genomes$^{\triangledown}$†

Dirk Höper, Bernd Hoffmann, and Martin Beer*

Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, Greifswald-Insel Riems 17493, Germany

Received 29 May 2008/Returned for modification 2 September 2008/Accepted 14 December 2008

The spread of highly pathogenic avian influenza A virus (HPAIV) of subtype H5N1 demands fast and reliable methods for in-depth, full-length sequence analysis. For this purpose, we designed a simple and sensitive method for the preparation of sequencing libraries from H5N1 HPAIV diagnostic RNA samples for sequencing with the Genome Sequencer FLX instrument. The method presented seamlessly integrates high-throughput pyrosequencing with the Roche/454 instrument into diagnostics without the need for additional equipment or molecular biological techniques besides standard PCR and the Genome Sequencer FLX sample preparation and sequencing pipeline.

In order to decide which action is required after the diagnosis of an infection with highly pathogenic avian influenza A virus (HPAIV) of subtype H5N1, detailed knowledge of the properties of the individual isolates at the genomic level is necessary. This information is needed immediately and needs to be as detailed as possible to take action against further spread of the virus, to detect important sequence changes modulating pathogenicity, or to permit molecular epidemiological analysis in an outbreak situation. To this end, reliable complete genomic sequences of isolates provide the most solid background as the basis for decisions on the actions to be taken after the diagnosis of infection with HPAIV subtype H5N1. Moreover, full genome sequences enable determination of the new isolate's properties, e.g., identification of features allowing enhanced replication in a mammalian host, as has been shown, for example, by Conenello and colleagues (2). However, the classical sequencing of complete avian influenza A virus (AIV) genomes by the method of Sanger et al. (7) needs several steps of sample preparation and therefore is laborious and time-consuming (times of up to weeks for sequencing and assembly of whole genomes, including repetitions for sequence validation).

In contrast, after proper sample preparation, which can be done within 3.5 h, followed by overnight clonal amplification of the library fragments, the new high-throughput pyrosequencing technique (5) available by use of the Genome Sequencer (GS) FLX instrument (Roche, Mannheim, Germany) generates large amounts of sequence information in a single run, thereby providing several repetitions of sequencing in a single experiment. The key features of this new technology are a sample preparation technique that does not rely on cloning for the generation of shotgun sequencing libraries; bead-bound clonal amplification of the library fragments in a single emulsion PCR (emPCR) run, which allows parallel amplification of the whole library; and massively parallel sequencing of the clonally amplified bead-bound shotgun library fragments in the wells of a PicoTiterPlate (PTP) that yields one sequencing read per bead. During preparation of the single-stranded template DNA (sstDNA) library, the target DNA is randomly fragmented and special adapters are ligated to the DNA fragments. The adapters serve as priming sites for PCR amplification of the library, thereby eliminating the need for sequence knowledge prior to sequencing. Via one of these adapters, the library fragments are bound to beads and are then clonally amplified in an emPCR. In the emPCR, the droplets of the water in an oil emulsion are microreactors containing the PCR reagents and one bead-bound DNA fragment, which thereby allows parallel clonal amplification of the complete sstDNA library in a single PCR. Subsequently, the beads carrying the amplified library are recovered from the emulsion, prepared for sequencing, and loaded into a PTP. This PTP has approximately 1.6 million wells, each of which holds only one bead. Thus, during loading, the beads carrying the amplified library are physically separated, which allows the sequencing of every bead-bound amplicon in a single sequencing read. Thus, up to 100 Mb of raw sequence data can be obtained in a single instrument run by sequencing hundreds of thousands of bead-bound DNA fragments in parallel. This enables sequencing of complete AIV genomes with a great deal of reliability within 3 days after sample receipt.

MATERIALS AND METHODS

Viruses. The virus strains used in this study, A/swan/Germany/R65/2006 (H5N1) (11), A/stone marten/Germany/R747/2006 (H5N1) (9), A/Cygnus olor/Germany/R1372/2007 (H5N1), and A/Beijing duck/Germany/R1959/2007 (H5N1), were obtained from the virus collection of the National Reference Laboratory for Avian Influenza at the Friedrich-Loeffler-Institut, Greifswald, Insel Riems, Germany. The two H5N1 HPAIV isolates, isolates R65/2006 and R747/2006, which were previously sequenced at our institute by the classical sequencing method of Sanger et al. (7), were used as references for validation of the protocol.

RNA extraction. Total RNA was isolated from allantoic fluid (140 μl) with a QIAamp viral RNA minikit (Qiagen, Hilden, Germany), according to the manufacturer's instructions.

PCR, library preparation, sequencing, and sequence assembly. Figure 1 provides an overview of the complete procedure from RNA extraction to sequencing and an approximate timeline. DNA was generated from genomic RNA diluted 10-fold in RNA-safe buffer (50 ng/μl carrier RNA, 0.05% Tween 20, 0.05% sodium azide in RNase-free water) (4) by reverse transcription-PCR (RT-PCR)

* Corresponding author. Mailing address: Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, Greifswald-Insel Riems 17493, Germany. Phone: 49 38351 7200. Fax: 49 38351 7151. E-mail: Martin.Beer@fli.bund.de.

† This paper is dedicated to the memory of Daniela Hase.

$^{\triangledown}$ Published ahead of print on 24 December 2008.

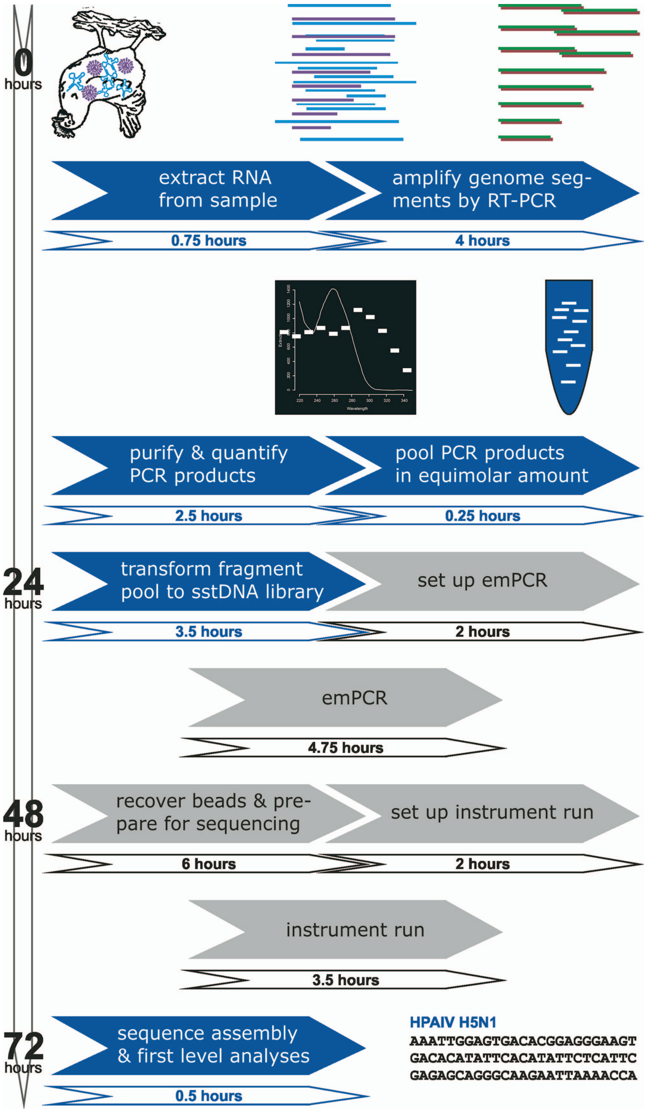
0
hours
extract RNA from sample
amplify genome segments by RT-PCR
0.75 hours
4 hours
purify & quantify PCR products
pool PCR products in equimolar amount
2.5 hours
0.25 hours
24
hours
transform fragment pool to sstDNA library
set up emPCR
3.5 hours
2 hours
emPCR
4.75 hours
48
hours
recover beads & prepare for sequencing
set up instrument run
6 hours
2 hours
instrument run
3.5 hours
72
hours
sequence assembly & first level analyses
0.5 hours
HPAIV H5N1
AAATTGGAGTGACACGGAGGGAAGT
GACACATATTCACATATTCTCATTC
GAGAGCAGGGCAAGAATTAAAACCA

FIG. 1. Work flow and timeline for the generation of sstDNA libraries from samples containing H5N1 AIV and for sequencing with the GS FLX instrument. The image displays the main steps for the sequencing of AIV genomes as separate filled arrows labeled with a short descriptor. Blue arrows, steps of the protocol established or modified in the present study; gray arrows, standard procedures of the sequencing protocol with the GS FLX instrument. Every step is accompanied by an open arrow labeled with the approximate duration of the respective experimental procedure. The vertical open arrow displays the overall timeline from sample receipt to retrieval of the final sequence and first-level data analyses. For details on the experimental procedures, refer to the text (blue steps) and the manufacturer's documentation (gray steps).

with a SuperScript III one-step RT-PCR system with Platinum *Taq* high-fidelity polymerase (Invitrogen, Carlsbad, CA). For every H5N1 HPAIV genome, 11 PCRs were set up in duplicate with the primers listed in Table 1. Large segments 1, 2, and 3 were amplified as two fragments each. To allow for proper sequence assembly, the PCR products for each of these genome segments overlap approximately 320, 130, and 180 bp, respectively. Five microliters of the dilute template and 1 μl of each primer (10 μM; final concentration in the PCR mixture, 0.4 μM) were mixed and denatured for 2 min at 95°C. Immediately after denaturation, the samples were frozen in liquid nitrogen, and subsequently, 18 μl of the PCR master mixture (12.5 μl 2× reaction mixture, 1 μl Superscript III reverse transcriptase-Platinum *Taq* high-fidelity polymerase enzyme mixture, 4.5 μl RNase- and DNase-free distilled water) was added. PCR was performed in a model 2720 thermal cycler (Applied Biosystems, Foster City, CA). The cycling parameters were 30 min at 55°C (reverse transcription); 2 min at 94°C (inactivation of reverse transcriptase and activation *Taq* polymerase); 40 cycles of 15 s at 94°C (denaturation), 30 s at 55°C (annealing), and 3 min at 68°C (elongation); and 1 cycle of a hold for 5 min at 68°C (final synthesis). The PCR products were gel purified with a QIAquick gel extraction kit (Qiagen). The purified DNA was quantified with a model 1000 spectrophotometer (NanoDrop Technologies, Inc., Wilmington, DE) in the double-stranded DNA mode and was then pooled in equimolar amounts. This DNA pool (3 to 5 μg) was transformed to sstDNA libraries by using a GS DNA library preparation kit (Roche), according to the manufacturer's instructions, but the Ampure bead purification step was omitted. Instead, the fragmented DNA was cleaned with one column from a MinElute PCR purification kit (Qiagen). The purified fragmented DNA was filtered through a PCR Kleen spin column (Bio-Rad Laboratories, Munich, Germany) and was finally concentrated with a MinElute PCR purification kit. The sstDNA libraries were subjected to duplicate emPCRs with the GS emPCR kit I (Roche), according to the manufacturer's instructions, with 0.5 and 1 copy per bead. After bead recovery and enrichment, all beads available from the duplicate emPCRs (up to a maximum of 45,000 beads) were loaded and sequenced in one small lane (totaling two lanes per library with two emPCRs per lane) of the PTP by using a GS SR70 sequencing kit (Roche) and the appropriate instrument run protocol. The resulting sequencing reads were sorted according to the genome segments to which they related and were subsequently assembled into one contig (i.e., a set of overlapping sequencing reads) per segment by using the GS FLX sequence assembly software newbler (version 1.1.03.24; Roche). During the assembly, the primer sequences were subtracted from the raw data.

Real-time RT-PCR. For real-time RT-PCR estimation of the approximate genome content of the RNA samples, a diagnostic RT-PCR assay directed against the M segment (8) in combination with a QuantiTect Probe RT-PCR kit (Qiagen) and a 7500 real-time PCR system (Applied Biosystems) was used.

Nucleotide sequence accession numbers. The nucleotide sequences obtained in this study are available from the EMBL/GenBank/DDBJ database under accession numbers FM165519 to FM165526 (isolate R1372/2007) and FM165527 to FM165534 (isolate R1959/2007).

RESULTS AND DISCUSSION

Validation of the protocol. For validation of the protocol, the two defined H5N1 HPAIV isolates A/swan/Germany/R65/2006 (H5N1) (11) and A/stone marten/Germany/R747/2006 (H5N1) (9) were analyzed. To this end, RNA was isolated from the allantoic fluid of inoculated embryonated chicken eggs. By application of the protocol described above, the sequencing of isolate R65/2006 generated approximately 1.45 Mb of raw sequence from 15,342 reads, with an average read length of 94.6 bases. The assembled unrevised sequence of 13,105 bases covered 98.94% of the previously reported coding sequence and had 99.91% identity with the reference sequence. The assembled sequence had a median depth (the sequence depth is the number of times that every base was sequenced) of 67, and 99.68% of the bases had a quality score of ≥40 (i.e., a probability of ≥99.99% that the individual bases were correctly identified). Sequencing of isolate R747/2006 yielded 7,673 reads, with an average read length of 98.8 bases, totaling 0.76 Mb of raw sequence. The assembly resulted in 13,154 bases with a median depth of 45, and 99.57% of the bases had a quality score of ≥40. These assembled unrevised sequences covered 98.99% of the previously reported coding sequences and had 99.94% identity with the reference sequence. Table 2 summarizes all data regarding sequence quality. Table 3 shows details of the comparisons of the coding portions of the new and the previously published sequences. No differences between the coding sequences for both the M and the NS genes were detected for either isolate R65/2006 or isolate

TABLE 1. Primers used for RT-PCR amplification of H5N1 HPAIV genomic segments

Amplification of segment/fragment	Primer designation	Primer sequence	Position (bp) in reference sequence	Reference GenBank accession no.
1/A	H5N1-pb2_1Fv2 H5N1-pb2_1421R	CGAAAGCAGGTCAAATATATTCAATA GGYAATATTCCGATCATHCCCAT	3–28 1421–1399	AM744956
1/B	H5N1-pb2_1102F H5N1-pb2_1Rv2	GGGTATGARGAATTCACAATGGT GGTCGTTTTTAAACAATTCGACAC	1102–1124 2330–2307	AM744956
2/A	H5N1-pb1_2Fv3 H5N1-pb1_2R-1230	AGCAGGCAAACCATTTGAATGG ATGTTGAACATGCCCATCATCAT	7–28 1265–1243	AM744955
2/B	H5N1-pb1_2F-1100 H5N1-pb1_2R-2300	RACACAAATACCAGCAGAAATGCT AYGAAGGACAAGCTAAATTCACTA	1119–1142 2319–2296	AM744955
3/A	H5N1-pa_3Fv2 H5N1-pa_1312R	AGCGAAAGCAGGTACTGATTCAAAA CHCCTATYTCATCAAGTTCTATCCA	1–25 1312–1288	AM744954
3/B	H5N1-pa_1131F H5N1-pa_3Rv2	ACTYGGTGAGAAYATGGCACC AAGGTACTTTTTTGGACAGTATGG	1131–1151 2224–2201	AM744954
4	H5N1-ha_4Fv2 H5N1-ha_4Rv2	AGCAGGGGTTCAATCTGTCAAAA AGTAGAAACAAGGGTGTTTTTAACTA	7–29 1779–1754	AM492165
5	H5N1-np_5Fv2 H5N1-np_5Rv2	TCACCGAGTGACATCAACATCA AAGGGTATTTTTCTTTAATTGTCATAC	25–46 1556–1530	AM744952
6	H5N1-na_6Fv2 H5N1-na_6Rv2	AGCAAAAGCAGGAGTTTAAAATGA TAGAAACAAGGAGTTTTTTGAACAAAC	1–24 1396–1370	AM492166
7	H5N1-m_7Fv2 H5N1-m_7Rv2	CAAAAGCAGGTAGATGTTGAAAGA AACAAGGTAGTTTTTTACTCCAATTC	3–26 1021–996	AM744957
8	H5N1-ns_8Fv2 H5N1-ns_8Rv2	AAAAGCAGGGTGACAAAGACATAA AGTAGAAACAAGGGTGTTTTTTATCA	4–27 875–850	AM744953

R747/2006. Moreover, no deletions in any of the raw sequences were found. All insertions traced were part of homopolymers that were overcalled, i.e., when the assembler software identified more bases than the number actually present in the sequence. In most cases, these inserted bases had a base quality score of <20, i.e., a probability of <99% that the individual bases were correctly identified. On the contrary, the majority of all bases had a quality score of ≥40, which corresponds to a probability of ≥99.99% that the individual bases were correctly identified (Table 2). Therefore, it must be assumed that the insertions are not true insertions but sequencing artifacts. In contrast to the insertions, the base quality score for the substitutions, with two exceptions, was 64 (a probability of 99.99996% that the individual bases were correctly identified, which is the highest possible score). Only half of the base substitutions detected caused amino acid substitutions; the remaining half were silent substitutions. One possible cause for these de-

TABLE 2. Summary of data characterizing the quantity and quality of the sequences obtained

Isolate	Median depth[c]	Coding sequence[a]			Overall sequence[b]		
		Length (no. of bases)[d]	% Coverage[e]	No. (%) of Q40+ bases[f]	Length (no. of bases)	% Coverage	No. (%) of Q40+ bases
R65/2006	67	12,561	98.94	12,527 (99.73)	13,105	96.65	13,063 (99.68)
R747/2006	45	12,568	98.99	12,537 (99.75)	13,154	97.01	13,097 (99.57)
R1959/2007	98	12,696	100.00	12,665 (99.76)	13,281	97.95	13,231 (99.62)
R1372/2007	32	12,358	97.34	12,251 (99.13)	12,795	94.37	12,677 (99.08)

[a] Only the coding part in the de novo assembly is considered.

[b] The whole-genome sequences in the assembly are considered.

[c] Median of the overall sequence depth (the number of times that each nucleotide was sequenced).

[d] Length of the assembly covering the portion of the genome considered.

[e] Percent coverage of the reference sequence (GenBank accession numbers for segments 1 to 8 for isolate R65/2006, DQ464357, DQ464361, DQ464360, DQ464354, DQ464359, DQ464355, DQ464356, and DQ464358, respectively; accession numbers for segments 1 to 8 for for isolates R747/2006, R1372/07, and R1959/07, AM744956, AM744955, AM744954, AM492165, AM744952, AM492166, AM744957, and AM744953, respectively).

[f] Number of bases with a quality score of ≥40 (Q40+), i.e., with a probability of ≥99.99% for the individual bases, and the percentage of bases with a quality score of ≥40 in the sequence portion considered.

TABLE 3. Details of the comparison of the coding portion of the new GS FLX instrument-generated sequences for isolates R65/2006 and R747/2006 and the previously reported sequences for these isolates as a reference[a]

Gene	R65/2006			R747/2006		
	Reference GenBank accession no.	Substitution(s)[b]	No. of insertions	Reference GenBank accession no.	Substitution(s)	No. of insertions
PB2	DQ464357	c1009t (L328F)	0	AM744956	None	0
PB1	DQ464361	a1603g + g1604t (S527V)[c]	0	AM744955	None	1
PA	DQ464360	t189c (silent) g1393a (E457K)	0	AM744954	c1058t (P345L), c1143t (silent)	4
HA	DQ464354	a316g (silent) g1534a (silent)	0	AM492165	None	1
NP	DQ464359	None	1	AM744952	None	0
NA	DQ464355	None	2	AM492166	None	0
M	DQ464356	None	0	AM744957	None	0
NS	DQ464358	None	0	AM744953	None	0

[a] The previously reported sequences were sequenced by the method of Sanger et al. (7).

[b] Nucleotide substitutions compared to the reference sequences indicated in the reference GenBank accession no. column; the impact on the coded amino acid is given in parentheses.

[c] Both base substitutions together caused only one amino acid substitution.

viations of the new sequence data from the previously reported sequence data might be the fact that the virus had been passaged between the sequencing experiments.

On the basis of the results of the validation, we concluded that our method is suitable for the very fast and reliable in-depth sequencing of whole H5N1 HPAIV genomes.

Determination of PCR sensitivity. In order to assess the sensitivity of our approach, template RNA from isolates R65/2006 and R747/2006 was serially diluted in RNA isolated from AIV-negative sample material. These template dilutions were used for both the generation of all PCR products for sequencing and quantitation of the approximate number of genome copies by real-time RT-PCR by use of a copy number standard. To this end, the dilute samples were examined by quantitative RT-PCR directed against the M segment (8). Table 4 summarizes the data from the dilution series experiment. Here, the PCRs for fragments 1B and 3A turned out to be the least sensitive. Most dilute samples which allowed the production of sufficient amounts of fragments 1B and 3A had a threshold cycle (C_T) value of approximately 26. This C_T value corresponds to roughly 5.6 × 10^3 copies/μl of RNA extract. Because during the extraction RNA from 140 μl of allantoic fluid was eluted in 50 μl elution buffer, our procedure enables the sequencing of complete AIV genomes from 6 μl of total RNA isolated from allantoic fluid, in which the concentration is about 2 × 10^3 genome copies per μl. This would allow the sequencing of complete AIV genomes from 6 μl of RNA extracted from swabs containing approximately 3.0 × 10^5 genome copies when the RNA is finally eluted in 50 μl of buffer.

Approval of applicability of the protocol. Furthermore, to test the functionality of our method, we used the protocol outlined above to sequence the genomes of two H5N1 HPAIV isolates (A/Cygnus olor/Germany/R1372/2007 and A/Beijing duck/Germany/R1959/2007) from recent outbreaks in southern Germany. For isolate R1959/2007, sequencing yielded 16,392 reads with an average read length of 95.5 bases and altogether approximately 1.57 Mb of raw sequence. The assembly of these raw data resulted in 13,281 nucleotides of assembled sequence. The sequence of the second newly sequenced isolate, isolate R1372/2007, was assembled from 0.57 Mb of raw sequence derived from 6,354 reads with an average length of 90.3 bases. Details on the sequence quantity and quality for isolates R1959/2007 and R1372/2007 are summarized in Tables 2 and 5. Figure 2 displays the depths of the sequences for the genomic segments 4 from both R1372/2007 and R1959/2007. The plot shows a considerably lower sequence depth for isolate R1372/2007. Nevertheless, even this suboptimal sequencing result obtained reliable, high-quality sequences (Table 5). First-level analyses of the sequences obtained by using the BLAST program (1) clearly identified both R1959/2007 and R1372/2007 as H5N1 HPAIV, thus confirming the previous diagnostic results and simultaneously considerably extending our knowledge of the isolates.

TABLE 4. Summary of results of dilution series experiment for the determination of PCR sensitivity

Dilution[a]	C_T[b]	No. of copies/μl[c]	PCR result for segment:										
			1A	1B	2A	2B	3A	3B	4	5	6	7	8
10^0	ND[d]	ND	+	+	+	+	+	+	+	+	+	+	+
10^{-1}	23.1 ± 0.4	5.66 × 10^4 ± 1.20 × 10^4	+	+	+	+	+	+	+	+	+	+	+
10^{-2}	26.6 ± 0.5	5.55 × 10^3 ± 1.00 × 10^3	+	+	+	+	+	+	+	+	+	+	+
10^{-3}	30.2 ± 0.6	5.07 × 10^2 ± 1.23 × 10^2	+	−	+	+	−	+	+	+	+	+	+
10^{-4}	33.8 ± 0.5	4.84 × 10^1 ± 7.45 × 10^0	+	−	+	−	−	+	+	−	−	+	+

[a] Dilution factor of the AIV-positive total RNA in AIV-negative total RNA.

[b] Threshold cycle of the diluted sample as determined by real-time RT-PCR directed against the M segment (8) (values are means ± standard deviations; $n = 6$).

[c] Approximate number of genome copies per μl of allantoic fluid, as determined by quantitative PCR with a copy number standard (values are means ± standard deviations; $n = 6$).

[d] ND, not determined.

TABLE 5. Data characterizing for each segment the quantity and quality of the sequences obtained for isolates R1959/2007 and R1372/2007

Isolate	Segment/gene	Median depth[c]	Coding sequence[a]			Overall sequence[b]		
			Length[d]	% Coverage[e]	No. (%) of Q40+ bases[f]	Length	% Coverage	No. (%) of Q40+ bases
R1959	1/PB2	166	2,280	100.00	2,273 (99.69)	2,332	99.62	2,325 (99.70)
	2/PB1	129	2,274	100.00	2,269 (99.78)	2,303	98.38	2,292 (99.52)
	3/PA	92	2,151	100.00	2,141 (99.54)	2,201	98.57	2,190 (99.50)
	4/HA	64	1,707	100.00	1,703 (99.77)	1,748	98.26	1,744 (99.77)
	5/NP	92	1,497	100.00	1,494 (99.80)	1,534	98.02	1,526 (99.48)
	6/NA	73	1,350	100.00	1,349 (99.93)	1,401	100.00	1,393 (99.43)
	7/M	129	759	100.00	759 (100.00)	892	86.85	892 (100.00)
	8/NS	60	678	100.00	677 (99.85)	870	99.43	869 (99.89)
R1372	1/PB2	62	2,250	98.68	2,245 (99.78)	2,274	97.14	2,269 (99.78)
	2/PB1	47	2,275	99.96	2,265 (99.56)	2,303	98.38	2,289 (99.39)
	3/PA	32	2,097	96.37	2,080 (99.19)	2,097	93.91	2,080 (99.19)
	4/HA	18	1,707	100.00	1,678 (98.30)	1,748	98.26	1,718 (98.28)
	5/NP	28	1,405	93.65	1,390 (98.93)	1,415	90.42	1,399 (98.87)
	6/NA	13	1,187	87.93	1,160 (97.73)	1,187	84.91	1,160 (97.73)
	7/M	25	759	100.00	756 (99.60)	918	89.39	912 (99.35)
	8/NS	28	678	100.00	677 (99.85)	853	97.49	850 (99.65)

[a] Only the coding part in the de novo assembly is considered.
[b] The complete sequences of the segments in the assembly are considered.
[c] Median of the overall sequence depth (the number of times that each nucleotide was sequenced).
[d] Length of the assembly covering the portion of the segment considered.
[e] Percent coverage of the reference sequence (the reference sequences are as described in Table 3 for isolate R747/2006).
[f] Number of bases with a quality score of ≥40 (Q40+), i.e., with a probability of ≥99.99% for the individual bases, and the percentage of bases with a quality score of ≥40 in the sequence portion considered.

To the best of our knowledge, this is the first published method enabling the sequencing of complete H5N1 AIV genomes directly from diagnostic samples by use of the GS FLX instrument. Although Pourmand and colleagues (6) also used pyrosequencing for AIV diagnostics, only some very short diagnostic sites of the HA segment were sequenced, and they used a completely different approach. However, there are several other sites in the influenza virus genome that determine the pathogenicity and host range, as described elsewhere (2, 3, 10). Therefore, by our approach, complete genomic sequences are generated with a different technology to yield information far beyond the properties of the HA segment. In order to achieve this goal, we used the Roche/454 GS FLX instrument, which allows massively parallel sequencing of shotgun libraries and which generates up to 100 Mb of raw sequence in a single instrument run. In the complete coding sequences that were finally generated, every nucleotide was sequenced in several reads. This is not comparable to conventional pyrosequencing and by far surpasses the single reads of diagnostic sites of the HA generated by Pourmand et al. (6). Taken together, our method allows the sequencing of complete H5N1 AIV genomes from routine samples and does not require either previous virus propagation in eggs or cell culture or the cloning and amplification of cDNA in vectors. Our approach permits sequencing of up to eight complete viral genomes within only 3 days and provides sequences of unprecedented depth and, consequently, with unprecedented reliability. The procedure described here can be seamlessly integrated into the normal diagnostic routine. The PCR that we applied for the generation of DNA requires sequence knowledge for priming. Therefore, with our technique we lose the advantage provided by the protocol with the GS FLX instrument that no sequence information is needed prior to sequencing. However, we thereby obtain the ability to sequence complete H5N1 HPAIV genomes even directly from samples from infected individuals and require information only about the virus subtype, which is obtained in the diagnostic procedure anyway. Moreover, our protocol enables for the first time the simple and swift in-depth sequencing of complete H5N1 HPAIV genomes without the introduction of any detectable sequence bias by DNA cloning or virus propagation in eggs. Furthermore, the procedure described here can easily be adapted to different AIV subtypes, therefore allowing sequencing of full-length genomes immediately after identification of the subtype by the use of routine diagnostics.

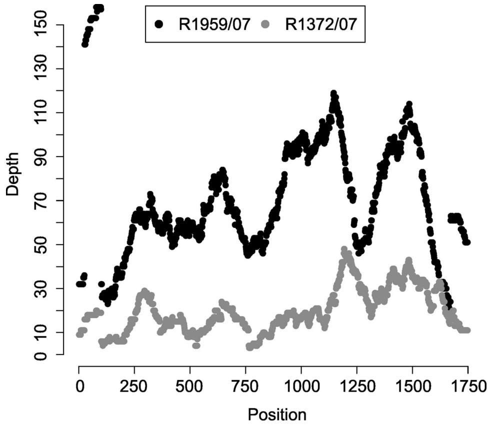

FIG. 2. Comparison of the sequence depth of segment 4 from both isolate R1372/2007 and isolate R1959/2007. The sequence depth (ordinate) plotted against the nucleotide position of the mRNA (abscissa) within the sequence of segment 4 is shown. The depth is the number of times that every nucleotide in the final sequence was sequenced, i.e., how many independent reads contributed information to the base called at the respective position.

ACKNOWLEDGMENTS

We thank Timm Harder for supplying H5N1 AIV samples and Daniela Hase for excellent technical assistance.

This project was funded by the Federal Ministry of Food, Agriculture and Consumer Protection (BMELV), Germany (project no. FSI 1-1.1.2.).

REFERENCES

1. **Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:**3389–3402.
2. **Conenello, G. M., D. Zamarin, L. A. Perrone, T. Tumpey, and P. Palese.** 2007. A single mutation in the PB1-F2 of H5N1 (HK/97) and 1918 influenza A viruses contributes to increased virulence. PLoS Pathog. **3:**1414–1421.
3. **Finkelstein, D. B., S. Mukatira, P. K. Mehta, J. C. Obenauer, X. Su, R. G. Webster, and C. W. Naeve.** 2007. Persistent host markers in pandemic and H5N1 influenza viruses, J. Virol. **81:**10292–10299.
4. **Hoffmann, B., K. Depner, H. Schirrmeier, and M. Beer.** 2006. A universal heterologous internal control system for duplex real-time RT-PCR assays used in a detection system for pestiviruses. J. Virol. Methods **136:**200–209.
5. **Margulies, M., M. Egholm, W. E. Altman, S. Attiya, J. S. Bader, L. A. Bemben, J. Berka, M. S. Braverman, Y.-J. Chen, Z. Chen, S. B. Dewell, L. Du, J. M. Fierro, X. V. Gomes, B. C. Godwin, W. He, S. Helgesen, C. He Ho, G. P. Irzyk, S. C. Jando, M. L. I. Alenquer, T. P. Jarvie, K. B. Jirage, J.-B. Kim, J. R. Knight, J. R. Lanza, J. H. Leamon, S. M. Lefkowitz, M. Lei, J. Li, K. L. Lohman, H. Lu, V. B. Makhijani, K. E. McDade, M. P. McKenna, E. W. Myers, E. Nickerson, J. R. Nobile, R. Plant, B. P. Puc, M. T. Ronan, G. T. Roth, G. J. Sarkis, J. F. Simons, J. W. Simpson, M. Srinivasan, K. R. Tartaro, A. Tomasz, K. A. Vogt, G. A. Volkmer, S. H. Wang, Y. Wang, M. P. Weiner, P. Yu, R. F. Begley, and J. M. Rothberg.** 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature **437:**376–380.
6. **Pourmand, N., L. Diamond, R. Garten, J. P. Erickson, J. Kumm, R. O. Donis, and R. W. Davis.** 2006. Rapid and highly informative diagnostic assay for H5N1 influenza viruses. PLoS One **1:**e95.
7. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
8. **Spackman, E., D. A. Senne, T. J. Myers, L. L. Bulaga, L. P. Garber, M. L. Perdue, K. Lohman, L. T. Daum, and D. L. Suarez.** 2002. 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. **40:**3256–3260.
9. **Starick, E., M. Beer, B. Hoffmann, C. Staubach, O. Werner, A. Globig, G. Strebelow, C. Grund, M. Durban, F. J. Conraths, T. Mettenleiter, and T. Harder.** 2008. Phylogenetic analyses of highly pathogenic avian influenza virus isolates from Germany in 2006 and 2007 suggest at least three separate introductions of H5N1 virus. Vet. Microbiol. **128:**243–252.
10. **Wasilenko, J. L., C. W. Lee, L. Sarmento, E. Spackman, D. R. Kapczynski, D. L. Suarez, and M. J. Pantin-Jackwood.** 2008. NP, PB1, and PB2 viral genes contribute to altered replication of H5N1 avian influenza viruses in chickens. J. Virol. **82:**4544–4553.
11. **Weber, S., T. Harder, E. Starick, M. Beer, O. Werner, B. Hoffmann, T. C. Mettenleiter, and E. Mundt.** 2007. Molecular analysis of highly pathogenic avian influenza virus of subtype H5N1 isolated from wild birds and mammals in northern Germany. J. Gen. Virol. **88:**554–558.