Genome Sequences of Parvovirus B19 Reference Strains

Jan-Hendrik Trösemeier,ᵃ Annika Branting,ᵇ Vladimir V. Lukashov,ᶜ Johannes Blümel,ᵃ Sally A. Baylisᵃ

Paul-Ehrlich-Institut, Langen, Germanyᵃ; Thermo Fisher Scientific, Carlsbad, California, USAᵇ; Center for Infection and Immunity, Amsterdam, the Netherlandsᶜ

We report here the sequences of two reference strains of parvovirus B19 (B19V) used for quantitation of B19V DNA. One reference strain has been established by the World Health Organization (WHO) and the other by the European Pharmacopeia (Ph. Eur.) and belong to B19V genotype 1a1 and 1a2, respectively.

Address correspondence to Sally A. Baylis, sally.baylis@pei.de.

Parvovirus B19 (B19V) is a single-stranded DNA virus belonging to the genus *Erythroparvovirus*. Diseases associated with B19V include erythema infectiosum, aplastic crisis, hydrops fetalis, and arthritis. Even in asymptomatic individuals, viremia may exceed 12 $\log_{10}$ IU/ml of B19V DNA, and consequently plasma used in the manufacture of medicinal products is screened by nucleic acid testing (NAT) to exclude units containing high titers of B19V (1). The World Health Organization (WHO) has established an international standard (IS) and an international reference panel (IRP) comprising three different genotypes of B19V which have been used in the standardization of NAT assays for both blood/plasma safety as well as for clinical diagnostic purposes (2). The 1st and 2nd ISs correspond to the genotype 1 strain included in the 1st WHO IRP, and this represents the most predominant strain circulating worldwide. Although the ISs, together with a calibrated biological reference preparation (BRP) from the European Pharmacopeia (Ph. Eur.), have been used to define the maximum threshold for B19V DNA levels (10 IU/μl) in plasma pools (3, 4), the genomic DNA sequences of these strains have not been reported.

The B19V genome sequences, excluding the inverted terminal repeats (ITRs) of ~350 nucleotides, were amplified by PCR. Amplicons were sequenced using the Ion Torrent PGM (Life Technologies GmbH, Darmstadt, Germany). Sequencing reads were preprocessed by PCR-adapter trimming using FASTA-Tools (5) and quality filtering using PRINSEQ (6). *de novo* assembly was performed with the MIRA 4 assembler (7). On the basis of BLAST searching against the NCBI database, the sequence with the highest identity and *E* score was chosen for reference-assisted scaffolding using Mira and BWA (8). Consensus sequences, based on reference-assisted scaffolding, were constructed using samtools (9).

A comparison of the two B19V reference strains revealed that they share 98.63% nucleotide sequence identity; the A+T content of both strains is 58.1%. Homopolymer tracts of 5 consecutive Cs, 7 As, 7 Ts, and 8 Gs were resolved without difficulties. In the case of the WHO strain, 9 nucleotides upstream of the nonstructural protein (NS1) start codon, a polymorphism was identified with a T present in 87% of the reads and a C in the rest. Analysis of B19V sequences in the NCBI database demonstrated that most genotype 1 B19V strains contained a C at this position (including the BRP), although a small number contain a T, which is more usually found in genotype 2 and 3 B19V strains.

Phylogenetic analysis of the two strains, using the complete NS1-VP1/2 fragment of the genome, was performed as previously described (10) and demonstrated that the WHO strain grouped with B19V genotype 1a1 and that the BRP strain grouped with genotype 1a2.

These results indicate that the 1st and 2nd WHO IS and the Ph. Eur. reference B19V strains are highly conserved genotype 1 viruses representing two separate genogroups which are both widespread throughout the world. Knowledge of the DNA sequences of these viruses is important for oligonucleotide design to ensure traceability and proper calibration of assays, particularly where virus load determination is mandated.

Nucleotide sequence accession numbers. The sequence of the WHO B19V strain has been deposited in GenBank under the name WHO IRP, and the Ph. Eur. strain was deposited under the name EDQM BRP; the respective accession numbers are KM065414 and KM065415.

REFERENCES

1. **Baylis SA.** 2008. Standardization of nucleic acid amplification technique (NAT)-based assays for different genotypes of parvovirus B19: a meeting summary. Vox Sang. **94:**74–80. http://dx.doi.org/10.1111/j.1423-0410.2007.00992.x.
2. **Baylis SA, Ma L, Padley DJ, Heath AB, Yu MW, Collaborative Study Group.** 2012. Collaborative study to establish a World Health Organization International genotype panel for parvovirus B19 DNA nucleic acid amplification technology (NAT)-based assays. Vox Sang. **102:**204–211. http://dx.doi.org/10.1111/j.1423-0410.2011.01541.x.
3. **Saldanha J, Lelie N, Yu MW, Heath A, B19 Collaborative Study Group.** 2002. Establishment of the first World Health Organization international standard for human parvovirus B19 DNA nucleic acid amplification techniques. Vox Sang. **82:**24–31. http://dx.doi.org/10.1046/j.1423-0410.2002.00132.x.
4. **Nübling CM, Daas A, Buchheit KH.** 2004. Collaborative study for establishment of a European Pharmacopoeia biological reference preparation

(BRP) for B19 virus DNA testing of plasma pools by nucleic acid amplification technique. Pharmeuropa Bio. **2003:**27–34.

5. **Pearson WR, Wood T, Zhang Z, Miller W.** 1997. Comparison of DNA sequences with protein sequences. Genomics **46:**24–36. http://dx.doi.org/10.1006/geno.1997.4995.

6. **Schmieder R, Edwards R.** 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics **27:**863–864. http://dx.doi.org/10.1093/bioinformatics/btr026.

7. **Chevreux B, Wetter T, Suhai S.** 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology: Proceedings of the German Conference on Bioinformatics GCB 1999. GBF-Braunschweig, Hannover, Germany.

8. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:**589–595. http://dx.doi.org/10.1093/bioinformatics/btp698.

9. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. http://dx.doi.org/10.1093/bioinformatics/btp352.

10. **Molenaar-de Backer MW, Lukashov VV, van Binnendijk RS, Boot HJ, Zaaijer HL.** 2012. Global co-existence of two evolutionary lineages of parvovirus B19 1a, different in genome-wide synonymous positions. PLoS One **7:**e43206. http://dx.doi.org/10.1371/journal.pone.0043206.