Editorial Commentary | Omics Sciences

Towards haplotypes of blood group genes: the impact of long-read sequencing in molecular immunohematology

Lea Wörner, Peter Bugert

Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service Baden-Württemberg-Hessen, Mannheim, Germany

Correspondence to: Prof. Dr. rer. nat. Peter Bugert. Institute of Transfusion Medicine and Immunology, Medical Faculty Mannheim, Heidelberg University, German Red Cross Blood Service Baden-Württemberg-Hessen, Friedrich-Ebert-Str. 107, 68167 Mannheim, Germany. Email: peter.bugert@medma.uni-heidelberg.de.

Comment on: Song J, Kwon SS, Suh EJ, *et al.* Allelic Cataloging of RHD-RHCE Reference Sequences Using Targeted Long-Read Sequencing. Clin Chem 2025;71:1067-82.

Keywords: Blood group genes; RH blood group system; long-read sequencing; haplotyping

Submitted Feb 06, 2026. Accepted for publication Mar 19, 2026. Published online Apr 28, 2026.
doi: 10.21037/atm-2026-1-0022
View this article at: https://dx.doi.org/10.21037/atm-2026-1-0022

Among the currently recognized 48 human blood group systems with more than 370 blood group antigens, the Rh system, defined by the two homologous genes *RHD* and *RHCE*, is the most polymorphic. Both genes are closely located on chromosome 1 separated by only 32 kilo bases (kb) (*Figure 1*). By January 2026, the International Society of Blood Transfusion (ISBT) Blood Group Database listed 447 *RHD* alleles and 186 *RHCE* alleles including the reference alleles (1). Most alleles are defined by single nucleotide variations (SNVs). In addition, gene conversion events generated a variety of *RHD-RHCE* hybrid alleles due to the high sequence homology (over 93%) of the two genes. The Rh system comprised 56 antigens, including the five major antigens D, C, c, E and e. When the serologic phenotype is unclear (weak or partial antigen), the investigation of the molecular background is usually conducted by using specific polymerase chain reaction (PCR) methods for SNV genotyping or by short-read sequencing *RHD* and *RHCE*. The correct assignment of variants to one or the other allele can impact the interpretation of genotype-phenotype correlation. However, haplotype phasing of variants is often not possible by specific PCR methods or short-read sequencing.

With the introduction of long-read sequencing technologies such as single molecule real-time (SMRT) sequencing [Pacific Biosciences (PacBio)] and nanopore sequencing [Oxford Nanopore Technologies (ONT)] the characterization of gene haplotypes became more feasible. These technologies have a significant impact on blood group genetics and facilitate the setup of reference haplotype data for all blood group genes as several studies have already shown. For instance, long-read sequencing of the *ABO* gene, coding for the ABO blood group system, enabled the definition of haplotypes and a phylogenetic analysis (2). The *ACKR1* gene for the Duffy blood group system was analyzed by ONT sequencing of a single 5.7 kb PCR product and 25 alleles were defined by phasing of the identified SNVs in a different research work (3). The MNS blood group system is defined by the homologous *GYPA*, *GYPB* and *GYPE* genes. SMRT sequencing of large amplicons yielded standardized reference haplotype sequences, which can be used for MNS genotyping and identification of hybrid genes (4). As another example for the investigation of the *RH* locus, Zang *et al.* (5) used targeted DNA capture, SMRT long-read sequencing, and bioinformatics to fully assemble the *RH* region. Consensus sequences of 2.1 to 2.9 kb average length enabled phasing of distant SNVs to determine *RHD-RHCE* haplotypes, and identification of known and novel structural variations along with the breakpoints.

Recently, Song *et al.* (6) described the application of SMRT high-fidelity long-read sequencing to the entire

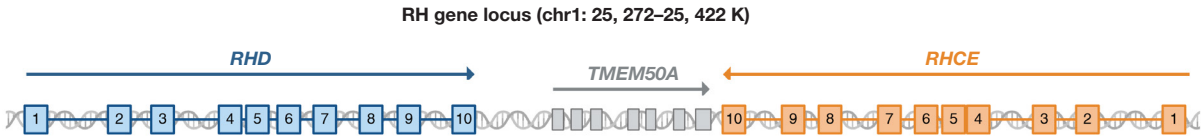

Figure 1 *RH* gene locus on chromosome 1 of the human genome GRCh38 spanning approximately 150 kb. The *RHD* and *RHCE* genes each have 10 exons and are oriented in the opposite direction. The non-homologous *TMEM50A* gene is located in between. Created in BioRender. Wörner L. [2026]. https://BioRender.com/limlf99.

RH gene locus by using overlapping amplicons. For the targeted so-called "meticulous amplicon tiling" protocol, 32 long-range polymerase chain reaction (LR-PCR) amplicons with a length of 25 kb each are combined. This allowed the generation of 170 kb long full-length *RHD-RHCE* haplotype sequences, spanning *RHD*, the intergenic region and *RHCE*. With the analysis of 63 samples, the setup of a full-length *RHD-RHCE* reference haplotype collection of the Eurasian population was aimed. For all samples, serological phenotypes matched the identified haplotypes. Complete *RHD-RHCE* phasing was achieved for 48 samples, from which 96 allelic full-length reference haplotype sequences were derived. Partially phased haplotype sequences (either complete *RHD* or *RHCE* phasing) were used to create shorter reference haplotype sequences. The 96 allelic full-length reference haplotype sequences were phylogenetically analyzed. The authors also focused on the detection of structural variants, such as *RHD* zygosity, rhesus box cross-overs, or gene conversion events in either *RHD* or *RHCE*, and were able to identify a variety of variants with their respective breakpoints. In summary, the introduction of the long-read sequencing-based approach leveraged the potential of this technology to analyze the highly homologous *RH* gene locus, thereby providing new perspectives for the molecular basis of *RH* haplotypes and their evolutionary history.

The study of Song *et al.* (6) provides an important insight into current technological developments to expand the toolbox of molecular immunohematology. The straightforward approach demonstrated the potential of long-read sequencing to overcome limitations of current methods, particularly for the analysis of homologous blood group genes. Sequencing of LR-PCR amplicons enables accurate phasing of variants covering long distances (7). The analysis of haplotype sequences improves the interpretation of the blood group phenotype, especially in the context of compound heterozygosity and novel variants (7). Long-read sequencing of blood group genes can also help to resolve phenotype-genotype discrepancies (7,8). In addition, the technique offers great potential for sequencing genetic regions with high sequence homology, such as the *RH* gene locus, but also *GYPA* and *GYPB* of the MNS blood group system. While homologous regions are difficult to amplify specifically in targeted sequencing approaches and short-read sequences often show mapping errors, long-read sequencing can overcome these limitations, as demonstrated by Song *et al.* (6). Prospectively, the "meticulous amplification tiling" protocol applied to the PacBio sequencing technology could also be transferred to other homologous (blood group) genes.

The study highlights another important aspect of long-read sequencing: the need for the systematic generation of high-quality haplotype reference sequences of genes and also of the entire human genome (9,10). Since the majority of the currently available blood group allele data are limited to single variants in exons, further research work is needed to characterize high-quality reference haplotype sequences. The setup of reference haplotype sequence data is essential to improve the reliability of bioinformatic-based variant calling and genotype-phenotype correlation. It is also highly important to make such haplotype reference sequences publicly available so that this data can be accessed regardless of available resources. However, the high costs of comprehensive and systematic research approaches have often hampered to generate haplotype sequences collections. High costs are also an important issue for a possible implementation in a in a clinical diagnostic setting. Besides the expectable reduction in costs and time due to further technical improvement, one could also think of long-read nanopore sequencing as a probably more cost-effective alternative technology.

Song *et al.* (6) presented a comprehensive long-read sequencing approach for the generation of haplotype sequences as a more comprehensive basis for sequence interpretation. Current challenges of sequence analysis and interpretation for homologous blood group genes were addressed as well. As admitted by the authors, the identified haplotype sequences do not represent the whole diversity

and complexity of the *RHD-RHCE* region due to the restricted sample set. In addition, there are still challenges regarding the implementation in a routine diagnostic setting. However, the presented long-read sequencing approach can be used to improve reference data quality as an important step towards comprehensive reference sequence data collections for all blood group systems.

Acknowledgments

None.

Footnote

Provenance and Peer Review: This article was commissioned by the editorial office, *Annals of Translational Medicine*. The article has undergone external peer review.

Peer Review File: Available at https://atm.amegroups.com/article/view/10.21037/atm-2026-1-0022/prf

Funding: None.

Conflicts of Interest: Both authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-2026-1-0022/coif). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. International Society of Blood Transfusion (ISBT) Blood Group Database. Available online: https://blooddatabase.isbtweb.org/ [Accessed 2026-01-09].
2. Gueuning M, Thun GA, Wittig M, et al. Haplotype sequence collection of ABO blood group alleles by long-read sequencing reveals putative A1-diagnostic variants. Blood Adv 2023;7:878-92.
3. Srivastava K, Khil PP, Sippert E, et al. ACKR1 Alleles at 5.6 kb in a Well-Characterized Renewable US Food and Drug Administration (FDA) Reference Panel for Standardization of Blood Group Genotyping. J Mol Diagn 2020;22:1272-9.
4. Xu H, Su X, Zuo Q, et al. The Research of a Large-Scale Analysis Platform for MNS Blood Group Identification Based on Long-Read Sequencing. Transfus Med Rev 2024;38:150836.
5. Zhang Z, An HH, Vege S, et al. Accurate long-read sequencing allows assembly of the duplicated RHD and RHCE genes harboring variants relevant to blood transfusion. Am J Hum Genet 2022;109:180-91.
6. Song J, Kwon SS, Suh EJ, et al. Allelic Cataloging of RHD-RHCE Reference Sequences Using Targeted Long-Read Sequencing. Clin Chem 2025;71:1067-82.
7. Thun GA, Gueuning M, Mattle-Greminger M. Long-read sequencing in blood group genetics. Transfus Med Hemother 2023;50:184-97.
8. Wang Z, Chu Y, Xiao Y, et al. Detecting serologically difficult ABO blood groups using single-molecule real-time sequencing technology. Vox Sang 2024;119:1096-105.
9. Rhie A, McCarthy SA, Fedrigo O, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature 2021;592:737-46.
10. Warburton PE, Sebra RP. Long-Read DNA Sequencing: Recent Advances and Remaining Challenges. Annu Rev Genomics Hum Genet 2023;24:109-32.

Cite this article as: Wörner L, Bugert P. Towards haplotypes of blood group genes: the impact of long-read sequencing in molecular immunohematology. Ann Transl Med 2026;14(2):25. doi: 10.21037/atm-2026-1-0022