A review of the literature organized into a new database: RHeference

Aline Floch , Stéphane Téletchéa , Christophe Tournamille , Alexandre G. de Brevern , France Pirenne

PII: S0887-7963(21)00010-9
DOI: https://doi.org/10.1016/j.tmrv.2021.04.002
Reference: YTMRV 50651

To appear in: *Transfusion Medicine Reviews*

Please cite this article as: Aline Floch , Stéphane Téletchéa , Christophe Tournamille , Alexandre G. de Brevern , France Pirenne , A review of the literature organized into a new database: RHeference, *Transfusion Medicine Reviews* (2021), doi: https://doi.org/10.1016/j.tmrv.2021.04.002

Highlights

- RHeference contains an extensive review of RHD with 805 entries and 467 sources
- RHeference is a powerful, modern database allowing simple and complex queries
- It focuses on practical observations in immunohematology and transfusion medicine

A review of the literature organized into a new database: RHeference

Aline Floch, [1,2,3] Stéphane Téletchéa,[4] Christophe Tournamille,[2,3] Alexandre G. de Brevern, [3,5,6], France Pirenne,[1,2,3]

1. *Univ Paris Est Creteil, INSERM, IMRB, F-94010 Creteil, France*
2. *EFS Ile-de-France Créteil, France*
3. *Laboratory of Excellence GR-Ex, F75739 Paris, France*
4. *Nantes Université, CNRS UMR 6286, UFIP, Nantes, France*
5. *Université de Paris, INSERM UMR_S 1134, BIGR, DSIMB, Univ de la Réunion, Univ des Antilles, Paris, France*
6. *Institut National de la Transfusion Sanguine, Paris, France*

Corresponding author:

Professor F. Pirenne

Etablissement Français du Sang, 51 avenue du Maréchal de Lattre de Tassigny, 94000 Créteil France

E-mail: France.pirenne@efs.sante.fr

Abstract

Hundreds of articles containing heterogeneous data describe D variants or add to the knowledge of known alleles. Data can be difficult to find despite existing online blood group resources and genetic and literature databases.

We have developed a modern, elaborate database for D variants, thanks to an extensive literature search with meticulous curation of 387 peer-reviewed articles and 80 abstracts from major conferences and other sources.

RHeference contains entries for 710 *RHD* alleles, 11 *RHCE* alleles, 30 phenotype descriptions (preventing data loss from historical sources), 35 partly characterized alleles, 3 haplotypes and 16 miscellaneous entries. The entries include molecular, phenotypic, serological, alloimmunization, haplotype, geographical and other data, detailed for each source. The main characteristics are summarized for each entry. The sources for all information are included and easily accessible through doi and PMID links. Overall, the database contains more than 10.000 individual pieces of data.

We have set up the database architecture based on our previous expertise on database setup and biocuration for other topics, using modern technologies such as the Django framework, BioPython, Bootstrap and Jquery. This architecture allows an easy access to data and enables simple and complex queries: combining multiple mutations, keywords, or any of the characteristics included in the database.

RHeference provides a complement to existing resources and will continue to grow as our knowledge expands and new articles are published.

The database url is http://www.rheference.org/

Keywords:

RH blood group; RHD gene; Immunogenetics; Database.

Abbreviations

ISBT	International Society of Blood Transfusion
RCIBGT	Red Cell Immunogenetics and Blood Group Terminology

INTRODUCTION

Since its discovery in the 1940s, the RH blood group system has been an important focus in transfusion medicine. It is the second most important blood group system due to its high clinical significance, and also the most diverse, with 55 antigens to-date.[1–3] Alloantibody formation to RH antigens may occur during non-identical red blood cell (RBC) transfusion or pregnancy, sometimes leading to hemolytic transfusion reactions and hemolytic disease of the newborn (HDN). The D (RH1) antigen is the most immunogenic RH antigen, and the most severe hemolytic reactions are caused by anti-D.

The existence of altered D antigens was recognized early on, because of weak D phenotypes and allo-anti-D formation in D positive individuals, some of which led to fatal HDN.[4,5] By cross-matching RBCs and anti-D, several phenotypes were described, dubbed DI to DVII. Then, some categories were subdivided, others became obsolete, and additional phenotypes were described.[6–11]

The molecular bases of the RH system were a source of questioning for decades, until the *RHD* and *RHCE* genes (NM_016124 and NM_020485 are the reference sequences in Genbank, respectively) were identified in the 1990s as two paralogous genes, with a high sequence identity.[12] The genes are separated by only one gene, *TMEM50A* (SMP-1), orientated head-to-tail and inherited together as a haplotype. Gene conversion events are responsible for a high frequency of hybrid *RHD-RHCE* alleles.[13]

The discovery of its molecular bases progressively led to a new understanding of the RH system. Serological classifications were defined more precisely or revisited in the light of molecular discoveries: studies made the link between phenotypes and genotypes, while others revealed that several heterogeneous genotypes were associated with what had been considered to be a single phenotype.[14–20] Extensive cross-matching with rare anti-D and the

characterization of endless D phenotype categories became increasingly difficult as new alleles were discovered.

More than 400 articles regarding D variants have now been published. This vast body of literature is composed of exceedingly heterogeneous, dispersed data. Some describe new molecular bases, others add to the knowledge of previously described alleles: phenotypes, anti-D formation, splicing, detection of alleles in various populations, etc. Conventional databases dedicated to human genetic variations and phenotypes, such as ClinVar,[21] mention Overhydrated Spherocytosis due to Rh_{null} syndrome,[22] but do not cover other clinical data relevant to immunohematologists: blood group phenotypes, risk of antibody formation, or clinical consequences of alloimmunization.

Resources dedicated to immunohematology exist. The International Society of Blood Transfusion (ISBT) manages the reference list and nomenclature for blood group genes and alleles, available online as pdf files, updated every year or two (see web resources). For the RH system, the ISBT Red Cell Immunogenetics and Blood Group Terminology (RCIBGT) Working Party provides 4 distinct tables, as of Jan, 2021: "*RHD* negative blood group alleles v4.0 180208", "*RHD* weak D and Del alleles v5.0 180207", "*RHD* Partial D blood group alleles v5.0 180207" and "RH Blood Group Alleles: *RHCE*". The tables contain alleles submitted to or collected by the dedicated Working Party, listing molecular descriptions of each allele (nucleotide change, amino acid(s), exons), allele names (common and numerical, e.g. *RHD*DIIIa* and *RHD*03.01*), some phenotypes and comments. The classification of the *RHD* alleles into the ISBT tables is also informative regarding D phenotypes, but occasional mistakes and inconsistencies in these tables may occur. For instance, the *RHD*525T* allele has been included despite incompatible events c. 525C>T and p.F175L being reported.[23] Several alleles have also been assigned multiple allele numbers (e.g. *RHD*13.02* and *RHD*16.03* for the *RHD* allele with c.667T>G, c.676G>C and c.697G>C substitutions).

For *RHD*, the other major resource is The Human RhesusBase[24] which currently has the most comprehensive list of *RHD* alleles. The allele list is regularly updated (the latest update was done in March, 2020), amongst other means by monitoring novel *RHD* alleles deposited in the Genbank database. The allele list is extensive, with some data regarding serology or descriptions for many entries. It is possible to access the data within the Human RhesusBase by several dedicated, pre-established lists (phenotypes, names, mutation mechanism, haplotypes...). More advanced queries (e.g. which *RHD* alleles include two specific mutations) require browsing the entries manually.

Other interesting online resources for the RH blood group system include the table for *RHCE* alleles, maintained by the New York Blood Center (see web resources); Erythrogene,[25] which analyzed the variation of blood groups in the 1000 Genomes project; and Bloodantigens.com,[26] for RBC antigen typing from Next-Generation Sequencing data. Finally, the Blood Group Antigen Mutation Database[27] of the National Center for Biotechnology Information was a useful tool until it was discontinued a few years ago.

Overall, the existing resources contain limited data compared to the extensive literature, propose few links to published material, and have a limited number of features. *RHD* genotyping is becoming increasingly widespread, and this scattering of data is becoming more and more challenging for patient blood management. There also remain many gray areas in our understanding of the RH system, in part due to the dispersion of the information. For all these reasons, it appears necessary to propose a modern, complete, user-friendly database for the *RHD* gene and D variants, containing all published information and links to each source, using state-of-the-art bioinformatics solutions that enable performing complex queries.

MATERIALS AND METHODS

Sources and literature search

The initial allele list was assembled in July 2016 from the *RHD* allele tables of the RCIBGT Working Party of the ISBT[1,2] and completed by searching Genbank for *RHD* entries.[28] Additional alleles were found through an extensive literature search, including the *RHCE* alleles responsible for the expression of some D epitopes.[29–31] Between July 2016 and December 2020, relevant literature was accessed with different strategies. In a first step, a literature search in PubMed with keywords ("RHD", "D variant", "RHD allele", "weak or partial D antigen", etc.) was performed. In a second step, other relevant articles were found by checking the reference list of the articles collected in the first step. In a third step, since many alleles were still not associated with a study published in a peer-reviewed journal, the abstract books of major international and national meetings were browsed for alleles that had not been described elsewhere and cases of anti-D formation. (see web resources) If no other source could be found for an allele, the comments of the Genbank entry were considered. Data from full-length articles was favored over other sources. Like in many other fields, abstracts often precede full-length publications, and the latter undergo a thorough review process.

Data processing and organization

Data was collected manually from the sources, curated, and organized into a fixed series of subsections for each entry. To enable efficient browsing of the entries at a glance, the data was also synthetized in curated summarized annotations, displayed for each subsection.

The first subsection of an entry page contains the ISBT nomenclature and approved names, the ISBT allele table in which the allele can be found, a list of other names by which the allele has been referred to previously in the literature. The entry type is stated.

The second subsection, Molecular Data, first shows a scheme of the allele (with *RHD* exons colored red, *RHCE* colored blue, and deleted exons devoid of color). When hovering over an exon, a tooltip box appears with the exon numbering and nucleic and proteic limits. When hovering over a mutation, a box appears with the genetic variations at the nucleotide and

protein level. The Molecular Data section also contains: the nucleotide and amino acid substitution(s), whether the allele is a *D-CE* hybrid, the position of the amino acid substitution(s) relative to the lipid bilayer,[32] splicing data, etc.

The third subsection, Phenotype, contains: all D phenotypes reported, phenotypes for other RH antigens, monoclonal anti-D testing and epitope patterns, D antigen density per RBC. The detailed reagents, methods and reactivities have not been collected but a link to the upstream source is provided. The D phenotypes have been organized into 6 classes: D positive (for apparently normal or undetailed positive D phenotype), discrepant (for negative, weak and/or positive reactions depending on reagents and techniques within the same report), weak D, very weak D, DEL, D negative, and undetailed ambiguous D phenotype. Whether adsorption-elution was performed is specified for each report. For the curated summarized annotation of the main D phenotype deduced from this data, priority was given to the phenotypes most informative for users: DEL and variable reactions (i.e. positive or negative reactions in different studies and/or with different anti-D reagents), over negative D and weak D, respectively.

The fourth section, Haplotype, contains phenotypic and molecular data regarding the association of *RHD* and *RHCE* alleles. For each *RHD* allele and entry, all the published RHCE phenotypes and genotypes reported were taken into consideration. The annotations of the main RHCE phenotype and *RHCE* allele associations were deduced from these reports.

The fifth section, Alloimmunization contains: anti-D formation (i) in carriers of each variant (allo- and auto-anti-D, with as much serological data as available; anti-D presented in conference abstracts are included in this section) and (ii) in D negative recipients (when the carrier of the variant is a blood donor). A curated summarized annotation of the risk of anti-D formation in carriers and in D negative recipients was deduced from these reports. Where applicable, the allele phenotype was also considered for this curated summarized annotation.

Indeed, anti-D formation may not have been reported in carriers of alleles expressing a D negative phenotype, but is expected to be possible. Similarly, anti-D formation in D negative recipients may not have been reported for alleles expressing D positive phenotypes, but is expected to be possible. When relevant, other RH antibody specificities were also included, like anti-C (anti-RH2) for *RHD*DIIIa-CEVS(4-7)-D (RHD*03N.01)*.

The sixth subsection, Reports, contains: a list of samples, situations, ethnicities and populations in which each allele has been described to date. The curated summarized annotation was deduced from the data and combines a summary of the population(s) and/or ethnicitie(s) in which the allele has been described, with a rough estimation of the number of samples reported (because some studies may overlap). This data is not prevalence data, since most studies rely on an initial phenotyping step, skewing the estimates, but an overview of the reports for the alleles in the immunohematology and transfusion contexts.

The seventh section, Links, contains links to other databases and websites: The Human RhesusBase,[24] Genbank,[28] and Erythrogene.[25] The final subsection of each entry, References, is the complete list of references and sources for the entry (see features below). The date when a given annotation or a full entry revision was performed is indicated where appropriate.

The data are included in RHeference as published. The allele assignments done by authors were not reinterpreted in the light of subsequent allele descriptions or by considering the limits of the methodologies used, e.g. alleles are often listed as *RHD*weak D type 1 (RHD*01W.1)* in the literature based on the presence of c.809T>G, but additional mutations are not excluded. Any discordant data between different sources was included with the respective references, e.g. *RHD*01EL.01 (RHD*1227A)* has been associated with weak D, DEL and D negative phenotypes.[33–35]

Three types of entries were created to accommodate some data that could not be associated with an exact allele: “phenotypic characterization”, “partly characterized allele”, “haplotype”. This was done to prevent the loss of (i) historical data: the phenotypes of some samples were studied in early works with or without partial molecular typing in subsequent studies (e.g. PCR patterns) and (ii) modern population studies: alleles are incompletely typed by recent genotyping strategies, particularly when a more precise typing will not modify patient blood management (e.g. the silent mutations of the subtypes of *RHD*DAR (RHD*09.01)* are rarely investigated).

Database design, architecture and querying.

RHeference is based on Django framework v1.11 LTS, an open Python framework, combined with BioPython,[36] Boostrap 3 and Jquery 1.12 to provide specific developments and biocuration for bioinformatics databases, as previously published.[37–39] RHeference is an open database freely accessible online at http://www.rheference.org/. A complete schema of the database’s architecture, with the specifications of the data and their relationships is available as Supplementary Figure S1.

The website is organized into 7 pages and menus: RHeference (the homepage), Allele, References, Documentation, Statistics, Search and Contact. These pages, detailed in Results, present RHeference, its content, and illustrate how to use the database.

RESULTS

Sources and entries

All the data incorporated in RHeference was carefully collected, checked manually, verified between sources for coherence, and compared with existing resources (see Methods). To date, RHeference v.1.00 incorporates 467 unique source references (Figure 1), published between

1955 and 2020, including 387 peer-reviewed articles (82.9%) and 80 conference abstracts or other sources (online resources, thesis, etc.).

RHeference v.1.00 contains 805 entries: 710 *RHD* alleles (including *RHD*01*), 11 *RHCE* alleles, 35 entries with partial molecular characterization, 30 with phenotypic characterization only, 3 haplotypes, and 16 miscellaneous entries. The latter include: erroneous sequences corrected in subsequent studies, intronic variants in remote zones of the *RHD* gene, and a few other, unclassifiable cases. There are 67 hybrid alleles.

No article or abstract could be found for 82 (10.1%) entries, 77 of which come from data deposited in Genbank. RHeference links 718 Genbank accession numbers; 548 entries have at least one accession number.

Overview of *RHD* alleles

Many positions within the *RHD* coding sequence have at least one known single nucleotide variation (SNV) (Figure 2). As expected, the most prevalent SNV in *RHD* alleles are those that match the *RHCE* sequence. The substitution c.667T>G is the most prevalent, found in 104 *RHD* alleles (14.6%), counting the hybrid alleles. The most common non-*RHCE* substitutions are c.186G>T, c.410C>T, c.819G>A and c.1136C>T, present in 18 (2.5%), 23 (3.2%), 20 (2.8%) and 334 (4.6%) alleles respectively. Table 1 presents the *RHD* alleles most frequently reported in the literature. Table 2 gives an overview of the characteristics and of the completion of the database for the *RHD* alleles.

Entry layout

Figure 3 shows the default, compact view for a RHeference entry, organized following the subsections detailed in Methods. The compact display gives access to the curated summarized annotation of the main characteristics derived from the data collected. Clicking each arrow on the right opens the expanded view for each section to reveal the detailed data linked to each source (e.g. in the Phenotype subsection: RH phenotypes other than for the D antigen, antigen

density per Red Blood Cell, etc.) (Supplementary Figure S2). Many sections can be expanded further by clicking the plus signs.

Website layout

The 7 pages and menus of the website mentioned in Methods are accessible from the header (visible at the top of Figure 3). The homepage provides a brief description of the database. The Allele menu allows rapid browsing with a direct access to the list of all the entries. The References menu leads to the complete list of sources linked in the database. For each published article, the [Citation] tag leads to the corresponding PubMed entry or, whenever possible, to the Digital Object Identifier (doi). The [RHeference] tag leads to a page within the database dedicated to the source's details: title, authors, publication year, PMID, etc., with links to all the entries discussed within the source and a list of the annotations.

The Documentation page (http://www.rheference.org/documentation) provides a variety of information for the user, additional references (e.g. for each high- and low-frequency RH antigen), allowing for a better understanding of the significance and context of the data provided in the database. This page contains 6 tabs: (i) an Overview of the database with current database usage including querying, (ii) a commented Example of an entry, (iii) a schema of the database architecture, (iv) Funding and licensing information, (v) a presentation of the Collaborators on this project, and (vi) answers to Frequently Asked Questions (FAQ). The Statistics page is updated automatically when new data are incorporated to the database (http://www.rheference.org/statistics), shows the distribution of variants along the sequence as in Figure 2, and a number of statistics and counts including those listed in Tables 1 and 2. The Search menu enables several types of queries, from simple to more complex possibilities, detailed in the next paragraph. The Contact page provides information about the database designers.

Queries

As one of the main added values and specificities of RHeference, there are several query possibilities. First, within any entry, each characteristic can be clicked and opens a list of entries with similar data: a nucleotide or amino acid substitution, categorical data like phenotypes or haplotypes, etc. Second, the Quick search box in the website's header is used to search by keyword or position. Words (e.g. "weak" or "DAU") are searched in allele names, and numbers are searched in nucleotide and amino acid substitutions (e.g. "1025" or "342"). The user must click on an entry name from the drop-down list to be redirected to the entry page. A search by keyword within allele names is also possible from the page By name and a search by position from the page By mutation, both accessed from the Search menu. Third, from the Search menu In exons, any number and combination of exonic mutations can be selected with an option to output alleles with all or any of the mutations. For instance, clicking on exons 5 and 8 to select c.697G>C and c.1136C>T, with the option "All positions selected present" lists 8 alleles, while a search with the option "Any positions selected present" lists 95 alleles. Finally, a Complex search form is available from the menu for advanced queries. For instance, selecting the DEL phenotype associated with the main phenotype 'ce' outputs a list of 15 entries. For all query types except the Quick search, the output is a list of entries organized in a table with sorting options.

DISCUSSION

RHeference is a modern, powerful database for D variants. From the expertise gained on previous biocuration and databases on other topics,[37–39] we set up the database with modern technologies allowing simple and complex queries depending on the user's interest,

many intradatabase links, a high flexibility for future developments and database updates and interconnection with other online databases and resources.

RHeference incorporates 467 sources, most of which are peer-reviewed articles. In RHeference v1.00, we focused on including one or more sources for each entry and for anti-D formation (Figure 1). We are (and will be) continuously including new sources, with yearly updates, or more frequently if a group of curators can be formed. Using the contact form (http://www.rheference.org/contact), users can report any relevant article or abstract, as well as any bugs or typographical errors. The most up-to-date statistics will be regularly updated and available at http://www.rheference.org/statistics

We have chosen to include some data presented in conference abstracts, particularly for alleles with no peer-reviewed reference and regarding anti-D formation. It is sometimes argued that the existence of anti-D which have not made it into peer-reviewed journals may be doubted.[40] However, publishing a full-length article for anti-D observed with limited serological data is challenging when the allele has been reported previously. Until some transfusion journals created "New allele report" sections,[41] it was also nearly impossible to publish a new allele separately. Nowadays, it may still be quite challenging for teams focusing on patient management to conduct the full serological and molecular analysis necessary for a submission. Many teams, including ours, retain several (or many!) new, partly analyzed alleles awaiting more data or new samples and analyses before considering full-length publication. We consider reports in abstract form to be important data, if they are recognized as such, and have included some in RHeference. Each page dedicated to a source is clearly titled as "Article", "Abstract", or "Online Resource" (for links to The Human RhesusBase,[24] Genbank,[28] etc.).

For RHeference v1.00, we have chosen not to solicit data proceeding from personal communications or observations, which has not been peer reviewed. However, such data

could be considered for future evolutions of RHeference, as it could easily be included and flagged as unpublished observations. Sharing such data resulting from the experience and expertise of immunohematology reference laboratories and transfusion specialists would be of great value to the community. E.g. underlining which D variants are particularly prone to anti-D formation would be informative for blood banks with limited D negative pRBC resources.

As a point of comparison with the Human RhesusBase,[24] the best online resource to date, RHeference contains 721 alleles versus 591 for the Human RhesusBase, and 805 entries overall versus 620. Because all Genbank accession numbers have been included (i.e. when all 10 *RHD* exons were deposited separately for a single allele and explicitly linked in the publication), RHeference lists 718 Genbank accession numbers, while The Human RhesusBase has chosen to focus on the one or two most relevant accession number(s) for each allele and lists 552 Genbank accession numbers. The main way in which the data in RHeference complements The Human RhesusBase is that RHeference includes many more sources (467 versus 139 unique sources) and that all fields have been filled in RHeference when the data could be located. The number of alleles/entries with at least one link to an article or abstract is 89.9% in RHeference and 58.3% in the Human RhesusBase. The data within The Human RhesusBase was counted manually in January 2021.

One of the limitations of RHeference is that the allele assignments done by authors were not reinterpreted in the light of the limits of the methodologies used or subsequent allele descriptions, unless the authors themselves noted the limits of their genotyping methodology. When the authors noted limits, the data was associated with entries flagged as "partly characterized", e.g. *RHD*weak D type 45.1 (RHD*01W.45.1)* or *RHD*weak D type 45.2 (RHD*01W.45.2)*.[42] This may skew the data for some alleles for which subtypes are rarely separated, but probably with limited consequences overall. For instance, the silent mutations allowing the distinction between *RHD*weak D type 4.2.0, .1, .2* and .3 (respectively

*RHD*09.01.00, .01, .02* and *.03*) are not included in recent genotyping platforms, but all these alleles code for the same protein sequence with p.T201R, p.F223V, and p.I342T.

Reports of *RHD* alleles in many populations and countries are listed in RHeference as observed by immunohematologists. Immunogenomics studies on the RH blood group system have been performed in a relatively small number of areas (Europe, North America, Eastern Asia), and this disequilibrium influences the alleles described to-date in an immunohematology setting, the number of reports in patients, and some of the early assumptions of the geographic origin of some alleles. *RHD* genotyping is spreading worldwide[43–45]. Making data easily accessible through RHeference will help teams pinpoint relevant alleles to look for within local minority communities, inform decision-making for countries starting to implement *RHD* genotyping and encourage underrepresented countries to perform studies or collaborate [46–56] to fill in the gaps that remain from an immunogenomics standpoint. Such data included in RHeference are not allele prevalence data but reflect practical observations in immunohematology, as a complement to the databases mapping human genetic variation, such as the Genome Aggregation Database (gnomAD),[57,58] The ISBT blood group database project for all blood groups [59] is likely to include links to these databases. In the meantime, RHeference provides links to the Erythrogene project website.[25] Interconnections between RHeference and other databases will be discussed with existing, established and future databases maintainers.

Experts currently agree on the management of a minority of *RHD* alleles[40] because of the rarity of most alleles, the heterogeneous observations or simply the insufficient data regarding anti-D antibody formation. Patient management depends on local policies for each variant: some countries focus on preventing potential anti-D formation[40,60] while others treat individuals with weak D phenotypes as “normal” D positive unless anti-D is detected.[57,61]

By linking to all the anti-D reports, RHeference will help highlight which alleles may benefit from a discussion between experts.

CONCLUSIONS

With 467 references and 805 entries to date, RHeference provides an extensive review and pinpoints the gaps in current knowledge regarding D variants. Thanks to its features and numerous query options, it provides easy access and opens new possibilities to articulate available data, as a complement to existing resources. It also provides many additional information and links to publications of different kinds, useful for specialists and non-specialists. The overview it provides and the capacity to focus on points of interest by flexible queries paves the way to a better understanding of the D antigen and the RH blood group system. The database will continue to grow as our knowledge expands and new articles are published.

ACKNOWLEDGEMENTS

The authors would like to thank all the people who showed enthusiastic support for this project, including at the 2019 Congress of the International Society of Blood Transfusion in Basel and on ResearchGate.net. We also thank Philip Berardi (The Ottawa Hospital/University of Ottawa, Canada), Gwen Clarke (University of Alberta, Canadian Blood Services, Canada), Yann Fichou (Université de Brest, Etablissement français du sang de Brest, France), Jelena Lukacevic Krstic (Split University Hospital Center, Croatia) and Maria José Rodrigues (Centro de Sangue e da Transplantação de Lisboa, Portugal) for their responses to our inquiries regarding their papers and/or abstracts.

WEB RESOURCES

- New York Blood Center's Table for *RHCE* alleles: www.bloodgroupgenomics.org/rhce/rhce-table/ (accessed 20/03/2021)
- Erythrogene: www.erythrogene.com/ (accessed 27/07/2020)
- Genbank: www.ncbi.nlm.nih.gov/genbank/ (accessed 14/01/2021)
- ISBT Working Party for Red Cell Immunogenetics and Blood Group Terminology allele tables: www.isbtweb.org/working-parties/red-cell-immunogenetics-and-blood-group-terminology/ (accessed 27/12/2020)
- PubMed: www.pubmed.ncbi.nlm.nih.gov/ (accessed 14/01/2021)
- RHeference: www.rheference.org/ (accessed 14/01/2021)
- The Human RhesusBase www.rhesusbase.info/ (accessed 14/01/2021)

FUNDING AND RESOURCES

This work was supported by grants (2016 and 2018) from Laboratory of Excellence GR-Ex, reference ANR-11-LABX-0051. The labex GR-Ex is funded by the program "Investissements d'avenir" of the French National Research Agency, reference ANR-11-IDEX-0005-02.

This work was performed using high performing computing resources from GENCI–CINES (Grants 2018-A0040710370, 2020- A0070710961 and 2020-A0080711465).

The authors thank the Nantes Université SCI for infrastructure support and web hosting of the database.

AdB acknowledges the Indo-French Centre for the Promotion of Advanced Research / CEFIPRA for collaborative grants (numbers 5302-2), and French National Research Agency with grant ANR-19-CE17-0021 (BASIN).

The funding bodies have no role in the design of the study and collection, analysis, interpretation of data nor in writing the manuscript.

AUTHORSHIP CONTRIBUTIONS

Aline Floch: Conceptualization, Methodology, Validation, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization, Funding acquisition; **Stéphane Téletchéa**: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Review & Editing, Visualization; **Christophe Tournamille**: Conceptualization, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Review & Editing, Supervision, Funding acquisition; **Alexandre G. de Brevern**: Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Data Curation, Writing - Original Draft, Visualization, Supervision, Project administration, Funding acquisition; **France Pirenne**: Conceptualization, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Review & Editing, Supervision, Project administration, Funding acquisition

DISCLOSURE OF CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest relevant to the manuscript submitted.

REFERENCES

[1] Storry JR, Castilho L, Chen Q, Daniels GL, Denomme GA, Flegel WA, et al. International society of blood transfusion working party on red cell immunogenetics and terminology: report of the Seoul and London meetings. ISBT Sci Ser 2016;11:118–22. https://doi.org/10.1111/voxs.12280.

[2] Storry JR, Clausen FB, Castilho L, Chen Q, Daniels GL, Denomme G, et al. International Society of Blood Transfusion Working Party on Red Cell Immunogenetics and Blood Group Terminology: Report of the Dubai, Copenhagen and Toronto meetings. Vox Sang 2019;114:95–102. https://doi.org/10.1111/vox.12717.

[3] Chou ST, Westhoff CM. The Rh and RhAG blood group systems. Immunohematology 2010;26:178–86.

[4] Lacey PA, Caskey CR, Werner DJ, Moulds JJ. Fatal hemolytic disease of a newborn due to anti-D in an Rh-positive Du variant mother. Transfusion 1983;23:91–4. https://doi.org/10.1046/j.1537-2995.1983.23283172867.x.

[5] Gibble JW, Shirey RS, Ness PM. The Du phenotype, serology, and clinical relevance. Crit Rev Clin Lab Sci 1984;21:229–53. https://doi.org/10.3109/10408368409165783.

[6] Tippett P, Sanger R. Observations on subdivisions of the Rh antigen D. Vox Sang 1962;7:9–13. https://doi.org/10.1111/j.1423-0410.1962.tb03223.x.

[7] Lomas C, McColl K, Tippett P. Further complexities of the Rh antigen D disclosed by testing category DII cells with monoclonal anti-D. Transfus Med 1993;3:67–9. https://doi.org/10.1111/j.1365-3148.1993.tb00106.x.

[8] Tippett P, Lomas-Francis C, Wallace M. The Rh antigen D: partial D antigens and associated low incidence antigens. Vox Sang 1996;70:123–31. https://doi.org/10.1111/j.1423-0410.1996.tb01309.x.

[9] Lomas C, Grässmann W, Ford D, Watt J, Gooch A, Jones J, et al. FPTT is a low-incidence Rh antigen associated with a "new" partial Rh D phenotype, DFR. Transfusion 1994;34:612–6. https://doi.org/10.1046/j.1537-2995.1994.34794330017.x.

[10] Wallace M, Lomas-Francis C, Beckers EA, Bruce M, Campbell G, Chatfield S, et al. DBT: a partial D phenotype associated with the low-incidence antigen Rh32. Transfus Med 1997;7:233–8. https://doi.org/10.1046/j.1365-3148.1997.d01-32.x.

[11] Noizat-Pirenne F, Le Pennec P-Y, Mouro I, Roussel M, Cartron J-P, Ansart-Pirenne H. DAL: a new partial RHD phenotype. Transfusion 2001;41:971–2. https://doi.org/10.1046/j.1537-2995.2001.41070971.x.

[12] Colin Y, Chérif-Zahar B, Le Van Kim C, Raynal V, Van Huffel V, Cartron J-P. Genetic basis of the RhD-positive and RhD-negative blood group polymorphism as determined by Southern analysis. Blood 1991;78:2747–52.

[13] Wagner FF, Flegel WA. RHCE represents the ancestral RH position, while RHD is the duplicated gene. - PubMed - NCBI. Blood 2002;99:2272–4. https://doi.org/10.1182/blood-2001-12-0153.

[14] Wagner FF, Gassner C, Muller TH, Schonitzer D, Schunter F, Flegel WA. Three molecular structures cause rhesus D category VI phenotypes with distinct immunohematologic features. Blood 1998;91:2157–68.

[15] Wagner FF, Gassner C, Müller TH, Schönitzer D, Schunter F, Flegel WA. Molecular basis of weak D phenotypes. Blood 1999;93:385–93.

[16] Wagner FF, Frohmajer A, Ladewig B, Eicher NI, Lonicer CB, Müller TH, et al. Weak D alleles express distinct phenotypes. Blood 2000;95:2699–708.

[17] Westhoff CM, Vege S, Halter-Hipsky C, Whorley T, Hue-Roye K, Lomas-Francis C, et al. DIIIa and DIII Type 5 are encoded by the same allele and are associated with altered

RHCE*ce alleles: clinical implications. Transfusion 2010;50:1303–11. https://doi.org/10.1111/j.1537-2995.2009.02573.x.

[18] Reid ME, Hipsky CH, Velliquette RW, Lomas-Francis C, Larimore K, Olsen C. Molecular background of RH in Bastiaan, the RH:-31,-34 index case, and two novel RHD alleles. Immunohematology 2012;28:97–103.

[19] Lomas-Francis C, Halter Hipsky C, Velliquette RW, Reid ME. DIII Type 7 is likely the original serologically defined DIIIb. Transfusion 2012;52:39–42. https://doi.org/10.1111/j.1537-2995.2011.03233.x.

[20] von Zabern I, Wagner FF, Moulds JM, Moulds JJ, Flegel WA. D category IV: a group of clinically relevant and phylogenetically diverse partial D. Transfusion 2013;53:2960–73. https://doi.org/10.1111/trf.12145.

[21] Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res 2018;46:D1062–7. https://doi.org/10.1093/nar/gkx1153.

[22] Huang C-H, Cheng G, Liu Z, Chen Y, Reid ME, Halverson G, et al. Molecular basis for Rhnull syndrome: Identification of three new missense mutations in the Rh50 glycoprotein gene. Am J Hematol 1999;62:25–32. https://doi.org/10.1002/(SICI)1096-8652(199909)62:1<25::AID-AJH5>3.0.CO;2-K.

[23] Haer-Wigman L, Veldhuisen B, Jonkers R, Lodén M, Madgett TE, Avent ND, et al. RHD and RHCE variant and zygosity genotyping via multiplex ligation-dependent probe amplification. Transfusion 2013;53:1559–74. https://doi.org/10.1111/j.1537-2995.2012.03919.x.

[24] Wagner FF, Flegel WA. The Rhesus Site. Transfus Med Hemother 2014;41:357–63. https://doi.org/10.1159/000366176.

[25] Möller M, Jöud M, Storry JR, Olsson ML. Erythrogene: a database for in-depth analysis of the extensive variation in 36 blood group systems in the 1000 Genomes Project. Blood Adv 2016;1:240–9. https://doi.org/10.1182/bloodadvances.2016001867.

[26] Lane WJ, Westhoff CM, Gleadall NS, Aguad M, Smeland-Wagman R, Vege S, et al. Automated typing of red blood cell and platelet antigens: a whole-genome sequencing study. Lancet Haematol 2018;5:e241–51. https://doi.org/10.1016/S2352-3026(18)30053-X.

[27] Patnaik SK, Helmberg W, Blumenfeld OO. BGMUT: NCBI dbRBC database of allelic variations of genes encoding antigens of blood group systems. Nucleic Acids Res 2012;40:D1023-1029. https://doi.org/10.1093/nar/gkr958.

[28] Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, et al. GenBank. Nucleic Acids Res 2013;41:D36-42. https://doi.org/10.1093/nar/gks1195.

[29] Flegel WA, Wagner FF, Chen Q, Schlanser G, Frame T, Westhoff CM, et al. The RHCE allele ceCF: the molecular basis of Crawford (RH43). Transfusion 2006;46:1334–42. https://doi.org/10.1111/j.1537-2995.2006.00901.x.

[30] Hipsky CH, Lomas-Francis C, Fuchisawa A, Reid ME, Moulds M, Christensen J, et al. RHCE*ceCF encodes partial c and partial e but not CELO, an antigen antithetical to Crawford. Transfusion 2011;51:25–31. https://doi.org/10.1111/j.1537-2995.2010.02764.x.

[31] Chen Q, Hustinx H, Flegel WA. The RHCE allele ceSL: the second example for D antigen expression without D-specific amino acids. Transfusion 2006;46:766–72. https://doi.org/10.1111/j.1537-2995.2006.00795.x.

[32] Floch A, Pirenne F, Barrault A, Chami B, Toly-Ndour C, Tournamille C, et al. Insights into anti-D formation in carriers of RhD variants through studies of 3D intraprotein interactions. Transfusion. 2021 Feb 14. doi: 10.1111/trf.16301.

[33] Scott SA, Nagl L, Tilley L, Liew Y-W, Condon J, Flower R, et al. The RHD(1227G>A) DEL-associated allele is the most prevalent DEL allele in Australian D- blood donors with C+ and/or E+ phenotypes. Transfusion 2014;54:2931–40. https://doi.org/10.1111/trf.12701.

[34] Mota M, Dezan, Marcia Regina M, Valgueiro MC, Sakashita AM, Kutner JM, Castilho L. RHD allelic identification among D-Brazilian blood donors as a routine test using pools of DNA. J Clin Lab Anal 2012;26:104–8. https://doi.org/10.1002/jcla.21489.

[35] Yan L, Wu J, Zhu F, Hong X, Xu X. Molecular basis of D variants in Chinese persons. Transfusion 2007;47:471–7. https://doi.org/10.1111/j.1537-2995.2006.01138.x.

[36] Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 2009;25:1422–3. https://doi.org/10.1093/bioinformatics/btp163.

[37] Téletchéa S, Santuz, Hubert H, Léonard S, Etchebest C. Repository of Enriched Structures of Proteins Involved in the Red Blood Cell Environment (RESPIRE). PLoS One 2019;14:e0211043. https://doi.org/10.1371/journal.pone.0211043.

[38] Gheyouche E, Launay R, Lethiec J, Labeeuw A, Roze C, Amossé A, et al. DockNmine, a Web Portal to Assemble and Analyse Virtual and Experimental Interaction Data. Int J Mol Sci 2019;20. https://doi.org/10.3390/ijms20205062.

[39] Teletchea S, Teletchea F. STOREFISH 2.0: a database on the reproductive strategies of teleost fishes. Database (Oxford) 2020;2020. https://doi.org/10.1093/database/baaa095.

[40] Sandler SG, Flegel WA, Westhoff CM, Denomme GA, Delaney M, Keller MA, et al. It's time to phase in RHD genotyping for patients with a serologic weak D phenotype. College of American Pathologists Transfusion Medicine Resource Committee Work Group. Transfusion 2015;55:680–9. https://doi.org/10.1111/trf.12941.

[41] Westhoff CM, Ness PM. Single-page reports of new alleles or antigens. Transfusion 2013;53:929–929. https://doi.org/10.1111/trf.12210.

[42] Fichou Y, Le Maréchal C, Jamet D, Bryckaert L, Ka C, Audrézet M-P, et al. Establishment of a medium-throughput approach for the genotyping of RHD variants and report of nine novel rare alleles. Transfusion 2013;53:1821–8. https://doi.org/10.1111/trf.12009.

[43] Ouchari M, Srivastava K, Romdhane H, Jemni Yacoub S, Flegel WA. Transfusion strategy for weak D Type 4.0 based on RHD alleles and RH haplotypes in Tunisia. Transfusion 2018;58:306–12. https://doi.org/10.1111/trf.14411.

[44] Brajovich MEL, Boggione CT, Biondi CS, Racca AL, Tarragó M, Nogués N, et al. Comprehensive analysis of RHD alleles in Argentineans with variant D phenotypes. Transfusion 2012;52:389–96. https://doi.org/10.1111/j.1537-2995.2011.03297.x.

[45] Lukacevic Krstic J, Dajak S, Bingulac-Popovic J, Dogic V, Mratinovic-Mikulandra J. Anti-D Antibodies in Pregnant D Variant Antigen Carriers Initially Typed as RhD. Transfus Med Hemother 2016;43:419–24. https://doi.org/10.1159/000446816.

[46] Koutsouri T, Chaikali A, Giannopoulos A, Douramani P, Flesiopoulou I, Panos P, et al. Frequency distribution of RHD alleles among Greek donors with weak D phenotypes demonstrates a distinct pattern in central European countries. Transfus Med 2019;29:468–70. https://doi.org/10.1111/tme.12623.

[47] Guzijan G, Jovanovic Srzentic S, Pavlovic Jankovic N, Djilas I, Lilić M. Implementation of Molecular RHD Typing at Two Blood Transfusion Institutes from Southeastern Europe. Transfus Med Hemother 2019;46:114–20. https://doi.org/10.1159/000496751.

[48] Safic Stanic H, Dogic V, Herceg I, Jagnjic S, Bingulac-Popovic J, Babic I, et al. D variants in the population of D-negative blood donors in the north-eastern region of Croatia. Transfus Med 2020. https://doi.org/10.1111/tme.12726.

[49] Tammi SM, Tounsi WA, Sainio S, Kiernan M, Avent ND, Madgett TE, et al. Next-generation sequencing of 35 RHD variants in 16 253 serologically D- pregnant women in the Finnish population. Blood Adv 2020;4:4994–5001. https://doi.org/10.1182/bloodadvances.2020001569.

[50] Vege S, Sprogøe U, Lomas-Francis C, Jakobsen MA, Antonsen B, Aeschlimann J, et al. Impact of RHD genotyping on transfusion practice in Denmark and the United States and identification of novel RHD alleles. Transfusion 2020. https://doi.org/10.1111/trf.16100.

[51] Granier T, Beley S, Chiaroni J, Bailly P, Silvy M. A comprehensive survey of both RHD and RHCE allele frequencies in sub-Saharan Africa. Transfusion 2013;53:3009–17. https://doi.org/10.1111/trf.12409.

[52] Xhetani M, Seferi I, Férec C, Zoraqi G, Fichou Y. Distribution of Rhesus blood group antigens and weak D alleles in the population of Albania. Blood Transfus 2014;12:565–9. https://doi.org/10.2450/2014.0240-13.

[53] Fichou Y, Parchure D, Gogri H, Gopalkrishnan V, Le Maréchal C, Chen J-M, et al. Molecular basis of weak D expression in the Indian population and report of a novel, predominant variant RHD allele. Transfusion 2018;58:1540–9. https://doi.org/10.1111/trf.14552.

[54] Schoeman EM, Roulis EV, Perry MA, Flower RL, Hyland CA. Comprehensive blood group antigen profile predictions for Western Desert Indigenous Australians from whole exome sequence data. Transfusion 2019;59:768–78. https://doi.org/10.1111/trf.15047.

[55] El Housse H, El Wafi M, Ouabdelmoumene Z, Zarati F, Alid R, Nourichafi N, et al. Comprehensive phenotypic and molecular investigation of RhD and RhCE variants in Moroccan blood donors. Blood Transfus 2019;17:151–6. https://doi.org/10.2450/2018.0153-18.

[56] Thongbut J, Raud L, Férec C, Promwong C, Nuchnoi P, Fichou Y. Comprehensive Molecular Analysis of Serologically D-Negative and Weak/Partial D Phenotype in Thai Blood Donors. Transfus Med Hemother 2020;47:54–60. https://doi.org/10.1159/000499087.

[57] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581(7809):434-443. doi: 10.1038/s41586-020-2308-7. Erratum in: Nature. 2021;590(7846):E53.

[58] Wagner FF. The RHD allele inventory: lessons from high-throughput genome databases. Vox Sang 2019;114(S1):185.

[59] Hyland CA, Gassner C, Daly J, Flower R. Blood group antigens in the genomic era. Vox Sang 2020;115(S1):15.

[60] Floch A, Tournamille C, Chami B, Pirenne F. Genotyping in Sickle Cell Disease Patients: The French Strategy. Transfus Med Hemother 2018;45:264–70. https://doi.org/10.1159/000490858.

[61] Boateng LA, Ngoma AM, Bates I, Schonewille H. Red Blood Cell Alloimmunization in Transfused Patients With Sickle Cell Disease in Sub-Saharan Africa; a Systematic Review and Meta-Analysis. Transfus Med Rev 2019;33:162–9. https://doi.org/10.1016/j.tmrv.2019.06.003.

TABLES

Table 1: *RHD* alleles most frequently reported in the literature to-date, by number of samples, excluding standard *RHD* (*RHD*01)* and *RHD* deletion *(RHD*01N.01*).

Number of samples	**Common name**	**ISBT**	**Articles in RHeference (N)**
50 to 150 samples	*RHD*DIIIa*	*RHD*03.01*	30
	*RHD*DVI type 1*	*RHD*06.01*	36
	*RHD*DVI type 2*	*RHD*06.02*	32
	*RHD*weak D type 4.2.0, RHD*DAR1*	*RHD*09.01.00*	8
	*RHD*weak D type 5*	*RHD*01W.5*	40
	*RHD*weak D type 42*	*RHD*01W.42*	10
	*RHD*93dupT*	*RHD*01N.50 and RHD*01EL.18*	17
	*RHD*ex10del type 1*	*Not listed*	9
	*RHD*960A*	*Not listed*	7
More than 150 samples	*RHD-RHCe(2-9)-RHD*	*RHD*01N.03*	44
	*RHD*D psi*	*RHD*08N.01*	62
	*RHD*DVI type 3*	*RHD*06.03.01*	25
	*RHD*weak D type 4.2.2, RHD*DAR1.2*	*RHD*09.01.02*	28
	*RHD*weak D type 4.0*	*RHD*09.03.01*	58
	*RHD*DAU0*	*RHD*10.00*	30
	*RHD*partial weak D type 11*	*RHD*11*	55
	*RHD*partial weak D type 15*	*RHD*15*	50
	*RHD*weak D type 1*	*RHD*01W.1*	80
	*RHD*weak D type 2*	*RHD*01W.2*	77
	*RHD*weak D type 3*	*RHD*01W.3*	60
	*RHD*weak D type 38*	*RHD*01W.38*	23
	*RHD*weak D type 150*	*RHD*01W.150*	5
	*RHD*1227A*	*RHD*01EL.01*	69
	*RHD*486+1A*	*RHD*01EL.08*	36

Table 2: Summary of RHeference's content and completion for 710 *RHD* alleles.

		N	%
D antigen phenotype	Alleles with at least one D phenotype reported	616	85.4
Annotated "main" D phenotype	D positive phenotype	36	5.1
	Weak or very weak D	225	31.7
	Variable or discrepant D phenotype (positive or negative depending on reagents and methods)	144	20.0
	DEL phenotype	60	8.5
	Negative D phenotype, DEL excluded	72	10.1
	Negative D phenotype, DEL not excluded*	46	6.5
Monoclonal anti-D	Alleles with data for testing monoclonal anti-D	239	33.7
Ag density/RBC	Alleles with at least one Ag density/RBC	150	21.1
Haplotype association	Alleles with at least one RHCE haplotype association	499	70.3
Annotated "main" RHCE haplotype association†	With ce	141	20.2
	With Ce	308	44.2
	With cE	83	11.9
	With CE	3	0.4
Anti-D formation in carriers	Alleles with any explicit data regarding anti-D formation in carriers	111	15.6
	Alleles with one or more allo-anti-D (including in abstract form)	77	10.9
	Alleles with no anti-D	501	70.6

* For 12 alleles, it was not completely clear whether DEL phenotype was excluded or not, either because the alleles have been classified as "D negative" by the ISBT but an article or abstract confirming that a DEL phenotype had been excluded could not be found, or because the source mentioned performing adsorption-elution but did not explicitly state the results for the allele. † Including probable associations e.g. if the known samples had C+E–c+e+ phenotype, the allele is most likely associated with Ce because of the likelihood of the *dce*

haplotype being in trans. The most up-to-date statistics will be regularly updated and available at http://www.rheference.org/statistics

FIGURE LEGENDS

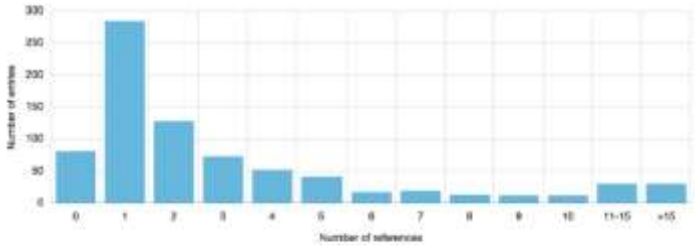

Figure 1: Data sources included in RHeference: number of entries with 0, 1, 2 etc. sources (peer-reviewed articles and conference abstracts).

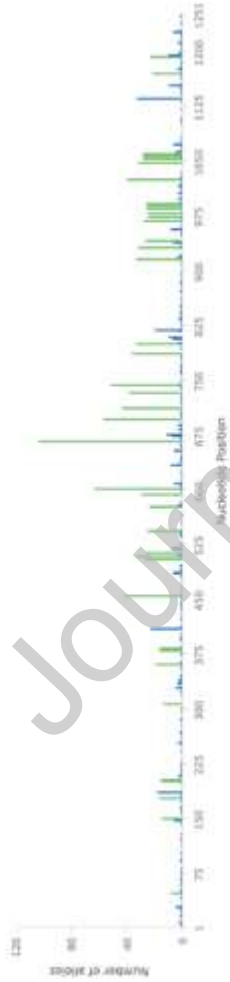

Figure 2: Genetic variants found in *RHD* alleles within the coding region.

Histogram of the number of alleles, including hybrid alleles, per nucleotide position (positions with mutations to *RHCE* are in green). The sequence stops at 1251 because the stop-codon is excluded.

Figure 3: Description of an entry available in RHeference.

The condensed display for *RHD*DNT (RHD*38)* serves as an example. Each subsection can be expanded by clicking on the corresponding arrow to the right. The expanded display is available as Supplementary Figure S2. The following information is visible in the condensed display. Subsection 1 serves as a header: ISBT nomenclature and table. Subsection 2: Molecular Data, with a schema of the allele. When hovering over an exon, a box appears with the exon limits. When hovering over a mutation, a box appears with the genetic variations at the nucleotide and protein level. Subsection 3: Phenotype, with the main phenotype for the D antigen.* Subsection 4: Haplotype, with the main RHCE phenotype associated (as an haplotype) with the allele.* Subsection 5: with a summary of alloimmunization data for anti-D antibody formation in carriers of the variant exposed to standard D.* Subsection 6: A summary of the situations and populations in which the allele has been reported.* Subsection

7: Links to other databases and websites (The Human RhesusBase,[24] Genbank,[28] Erythrogene[25]). Subsection 8: References to all sources relating to the entry. The last update for the whole entry appears at the bottom of the page.* The curated summarized annotations are interpreted from the data collected, with the last revision date for each summary shown next to it in parentheses.