Journal of Immunological Methods 465 (2019) 67–71

Contents lists available at ScienceDirect

Journal of Immunological Methods

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

Technical Note

A primer set for comprehensive amplification of V-genes from rhesus macaque origin based on repertoire sequencing

Ronit Rosenfeld[a,1], Anat Zvi[a,*,1], Eitan Winter[b], Ronen Hope[b], Ofir Israeli[a], Ohad Mazor[a], Gur Yaari[b]

[a] Department of Biochemistry and Molecular Genetics, Israel Institute for Biological Research, Ness-Ziona, Israel
[b] Bioengineering Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel

ARTICLE INFO

Keywords:
Antibody
V-gene diversity
Germline/repertoire
High throughput sequencing
Rhesus macaque

ABSTRACT

Recombinant antibodies serve as therapeutic molecules for a broad range of applications. High affinity antibodies are typically isolated following an active and effective immunization. Human-like antibodies may be obtained from immunized nonhuman primates (NHP), such as rhesus macaque, when immunized human origin is not available. For the isolation of such antibodies, strategies like phage and yeast display, are employed. These strategies are primarily based on the amplification of the rearranged variable (V) regions coded by mRNA, obtained from lymphatic source of immunized animals. To amplify these genomic sequences, designated set of primers are required, ideally covering the immune animal V-gene repertoire. Such primer sets are commonly designed based on the germline repertoire of specific animals according to immunoglobulin databases. However, In case of rhesus macaque, however, the known immunoglobulin germline V-gene database is still limited. The emergence and continuous improvements in high-throughput sequencing (HTS) technologies now enable the profiling of an immune repertoire for both basic and applicative studies, among which is the identification and expression of novel alleles. We report here on the profiling of non-immunized rhesus macaque (*Macaca mulatta*) expressed antibody repertoire, using HTS and advanced tailored bioinformatics tools. This analysis resulted in 32,480 and 73,354 complete heavy and light variable gene (V_H and V_L) sequences, respectively. Further analysis of these sequences, using the IgDiscover tool, resulted in the identification of 102, 214 and 48 inferred V_H, V_κ and V_λ germline sequences, respectively, of which over 50% are novel alleles. This dataset, together with other recently published datasets, enabled the design of a comprehensive primer set (v2018), which demonstrated the broadest coverage of rhesus macaque germline genes identified up to date. The newly designed primer set was confirmed for its extent of coverage of the V-genes in various datasets of rhesus macaque germlines as well as the expressed repertoire mapped in this study. Among other things, an improvement of 28% and 50% in the coverage of the VH and VL expressed repertoire was demonstrated in comparison to a primer set we have previously designed. This primer set can be further used for various applications that require the complete coverage of the NHP V-gene repertoire.

1. Introduction

Immune gene libraries are commonly constructed for the isolation of therapeutic antibodies. An active and controlled immunization is a prerequisite for isolation of well-tolerated high-affinity antibodies. Owing to immunogenicity issues, nonhuman primates (NHP)-derived antibodies represent a preferred source for such therapeutic molecules, in cases where human immunization is unachievable. Particularly, macaque-derived antibodies demonstrated similarity to their human counterparts and thus, were suggested to exhibit good therapeutic tolerance (Avril et al., 2014).

Well established and powerful methodologies (of which phage display is the most frequently used) are implemented for the selection of antibodies with desirable specificities from an antibody display library. The construction of such libraries is primarily based on the amplification of the rearranged variable (V) regions coded by mRNA, usually

Abbreviations: V, variable; V_H, heavy chain variable; V_L, light chain variable; V_κ, kappa chain variable; V_λ, lambda chain variable; NHP, nonhuman primates; HTS, High throughput sequencing; IG, Immunoglobulin; RACE, Rapid Amplification of cDNA Ends

* Corresponding author.
E-mail address: anatz@iibr.gov.il (A. Zvi).
[1] These authors contributed equally to this work

https://doi.org/10.1016/j.jim.2018.11.011
Received 5 September 2018; Received in revised form 19 November 2018; Accepted 20 November 2018
Available online 22 November 2018

obtained from lymphatic source of immunized animals, using a designated set of amplification primers. For any such downstream application, an optimal design of primer sets prone to fully cover the animal germline V-gene repertoire, is essential to achieve high coverage, yield and specificity on one hand, and to avoid biases and/or missing sequences on the other hand. Consequently, attempts to coverage optimization are therefore vital (Lim et al., 2010). Primer set design is mostly based on immunoglobulin databases (e.g. Kabat (Kabat et al., 1991), IMGT (www.imgt.org), VBASE (http://www.vbase2.org)). In the case of NHP, a complete germline V-gene database is not available yet and therefore V-gene amplification frequently relies on primers conferring restricted coverage, or otherwise on primers designed from human V-genes (Palat et al., 2009; Glamann et al., 1998). We previously demonstrated the successful isolation from NHP of extremely high affinity antibodies directed to the ricin toxin (Noy-Porat et al., 2016), using a specific primer set (v2012), designed according to the limited NHP V-gene databases available at the time.

HTS technologies currently enable sampling of the immune repertoire in unprecedented details and, consequently, large scale characterization of the expressed antibody repertoire (recently reviewed in (Miho et al., 2018; Yaari and Kleinstein, 2015) and references therein). Implementation of efficient methods and computational tools tailored specifically for antibody repertoire analysis, enabled us to profile the entire expressed antibody repertoire of non-immunized rhesus macaque peripheral blood lymphocytes, and to infer germline sequences. Based on the newly identified germline dataset, and on two other datasets that were recently documented (Corcoran et al., 2016; Ramesh et al., 2017), an updated set of primers was designed, which we coined v2018, meant to optimally cover the up-to-date known germline gene alleles. A comprehensive coverage of the three germline datasets by the v2018 primer set is demonstrated, as well as the extent of coverage of the expressed antibody repertoire identified in this study.

2. Materials & methods

2.1.1. Preparation of HTS libraries and repertoire sequencing

Total RNA was purified using RNAeasy Micro Kit (Qiagen, cat. no. 74004) from PBMCs of three non-immunized rhesus macaques, according to the manufacturer instructions. Immunoglobulin (IG) library preparation containing a mixture of isotypes, was performed essentially as described earlier (Turchaninova et al., 2016), based on RACE (Rapid Amplification of cDNA Ends) approach and template-switching.

Paired-end sequencing of the rhesus macaque generated library was performed using Illumina MiSeq, standard Illumina sequencing primers and MiSeq Reagent Kit v3 (600 cycles) at the presence of 25% PhiX library.

2.1.2. Pre-processing and generation of antibody sequences

The V_H and V_L repertoires were identified from ~2.4 M reads; data analysis was based on the RACE protocol described in Turchaninova et al. (Turchaninova et al., 2016) The pre-processing of the data was conducted using the pRESTO software (version 0.5.8) pipeline (Vander Heiden et al., 2014). A Phred quality score of 10 was applied as quality filtering. V-D-J assignment was conducted using the IgBLAST tool (Ye et al., 2013).

2.1.3. Identification of germline V-genes

The expressed repertoire was subjected to analysis towards de novo discovery of germline V-genes using the IgDiscover computational approach (Corcoran et al., 2016). The starting database used for the discovery process consisted of known rhesus macaque V_H, V_κ and V_λ sequences (19, 110 and 35, respectively) downloaded from the IMGT (October 2016), together with rhesus V_H, V_κ and V_λ sequences (400, 239 and 37, respectively) identified in the study of Corcoran et al. (Corcoran et al., 2016). The IgDiscover pipeline was subjected to three iterations, with default parameters in case of V_H sequences. Due to the relatively high degree of V_L CDR3 diversity, more stringent filtering criteria for CDR3 clustering were set for the light chain in order to minimize false positives.

2.1.4. Construction of the V2018 primer set and specificity evaluation

The V2018 primer set was constructed by manually updating a previously established primer set (V2012), according to this study newly identified germline genes as well as the currently available rhesus macaque V-gene datasets (IMGT (www.imgt.org), Corcoran and colleagues (Corcoran et al., 2016), Ramesh and colleagues (Ramesh et al., 2017)). The process of the primer set design included, among others, the following considerations: broad coverage, minimum number of non-silent degenerative bases (in the 5′ region of the primers in particular), and Tm as high as possible to assure high specificity (actual Tm values above 55 °C in average, and not < 52 °C).

The specificity of each primer was assessed using the command-line version of the MFEprimer-2.0 program (Qu and Chenggang, 2015). The custom database against which the primer specificities were evaluated included the expressed antibody sequences generated in this study, as well as germline sequences originating from the Corcoran et al. study (Corcoran et al., 2016), the Ramesh et al. study (Ramesh et al., 2017), and the germline sequences identified here (V_H and V_L, in accordance with the relevant primer sets). An *in house* pipeline was employed for processing of the MFEprimer-2.0 results, counting mismatches in the generated amplicons vs. the custom database and extracting hits which contained mismatches while retaining a predicted Tm cutoff of > 52 °C.

3. Results and discussion

In the current study, we determined the expressed IG repertoire of non-immunized rhesus macaque and employed it to determine germline V-genes and to construct a primer set directed towards an optimal coverage of up-to-date known rhesus sequences. A flowchart summarizing the entire process is presented in Fig. 1.

3.1.1. Non-immunized rhesus macaque V-gene identification

The gap of knowledge documented in recent studies with regard to the existing repertoire of monkey germline genes (Corcoran et al., 2016), necessitates the discovery of novel alleles. A total of ~2.4 M paired-end reads generated by HTS were processed (see *Materials and Methods*), resulting in the identification of 32,480 and 73,354 assembled heavy and light chain sequences, respectively; this list of sequences constituted the IIBR expressed database. Based on this repertoire, the IgDiscover computational approach, developed by Corcoran et al. (Corcoran et al., 2016), was subsequently implemented to discover germline sequences. The total of ~2.4 M paired-end reads were provided as input data, and the starting database used for primary assignment included the germline sequences (52 IGHV and 279 IGLV) retrieved from the IMGT repository (www.imgt.org) and the germline sequences identified in the Corcoran study (Corcoran et al., 2016) (400 IGHV and 273 IGLV). The input data was subjected to three sequential iterations of discovery process, each iteration generating an improved database, which then replaced the starting database for the next discovery process. After completion of three iterations, final databases were produced, comprising of 102, 31, and 4 unique sequences for the V, D and J sequences of the heavy chain, respectively, and 214, 48, 6 and 4 for the light chain V_κ, V_λ, J_κ, J_λ, respectively. A substantial fraction of the V alleles detected were found to be novel, namely 51 for the V_H (50%) and 189 for the V_L (72%), not present in any of the germline databases of rhesus macaque previously published. This proportion of novel alleles identified in this study, further emphasizes the gap of knowledge in the repertoire of rhesus macaque germline

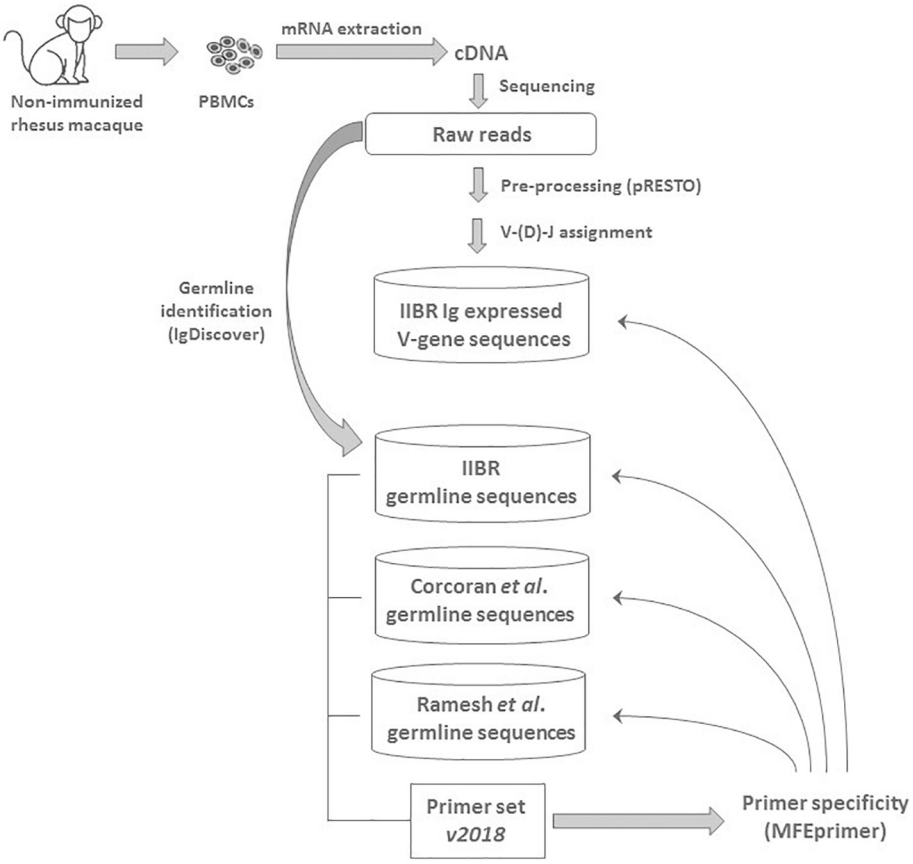

Fig. 1. Flowchart of the analysis process. A schematic representation of the workflow described herein: IIBR expressed antibody repertoire and germline sequences of non-immunized rhesus macaque were determined; a comprehensive primer set (v2018), based on the IIBR, Corcoran et al. (Corcoran et al., 2016) and Ramesh et al. (Ramesh et al., 2017) germline datasets was designed; the coverage achieved by the v2018 primer set against the various datasets was evaluated.

Table 1
Rhesus V-gene amplification v2018 primer set.

Primer name	5′-3′ sequence	Primer name	5′-3′ sequence	Primer name	5′-3′ sequence
Forward primers					
rhVH_For_1	CAGGAGCAGCTGGTGC	rhVκ_For_1	GA**Y**AT**H**GTGATGAC**M**CAGACTC	rhVλ_For_1	CAGGCTGTAGTGACTCAGG
rhVH_For_2	CAGGTCCAGCTGGTGC	rhVκ_For_2	GAT**R**TTG**Y**G**M**TGACTCAGTCTCC	rhVλ_For_2	CAGTCT**R**TGCTGAC**W**CA**R**CC
rhVH_For_3	CAGGT**R**CAGCT**R**CAGG	rhVκ_For_3	GACATTCAG**M**TG**W**CCCAGTC	rhVλ_For_3	CAGCCTG**Y**GCTGAC**Y**CAG
rhVH_For_4	CA**R**CTGCAGCTGCAGGAG	rhVκ_For_4	GAAAT**W**GT**R**ATGAC**K**CAGTC**Y**CC	rhVλ_For_4	CAGTCTGTGCTGACGCAG
rhVH_For_5	GAGGTGCA**R**CTGGTGC	rhVκ_For_5	GACAT**Y**CAGTTGACCCAGTC	rhVλ_For_5	CAG**K**CTGCCCTGAC**Y**CAG
rhVH_For_6	CAGGT**S**ACCTTGAAGGAGTC	rhVκ_For_6	GAAATTGTGATGACACAATCTCC	rhVλ_For_6	CAGTCTGCCCCGA**Y**TCAG
rhVH_For_7	GAGGTGCA**V****B**TGGTGG	rhVκ_For_7	CAAGTTATATTGAC**W**CAGTCTCCAG	rhVλ_For_7	CAGGCTGCCC**Y**GACTC
rhVH_For_8	GA**M**GTGCAGCTGGT**R**G	rhVκ_For_8	GACAT**Y****S**TGATGAC**Y**CAGTCTC	rhVλ_For_8	CAGGCAGGGCTGACTCAG
rhVH_For_9	GAGGTGCAGCTGGCG	rhVκ_For_9	GA**M**ATTGTGCTGAC**Y**CAGTCTC	rhVλ_For_9	AAGCCTATGCTGACTCAGC
rhVH_For_10	GAGGTGCAGCTGGTAGAG	rhVκ_For_10	GAGACTGTGGTGACCCAG	rhVλ_For_10	TCCTATGAGCTGACACAGC
rhVH_For_11	GAAGTGCAGTTGGTGGAGTC	rhVκ_For_11	**R**AAACAGTGGTGACGCAGTC	rhVλ_For_11	TCCTATGA**K****K**TGACTCAGCC
rhVH_For_12	GAGGTGCG**R**CTGGTGG	rhVκ_For_12	GATATCCAGATGACCCAGTC	rhVλ_For_12	TCCTA**Y****R**A**B**CTGACTCAG**Y**C
rhVH_For_13	CAGGT**R****M**AGCTGCAGCAG	rhVκ_For_13	GAAATTGTG**W**TGAC**V**CAGTCTCC	rhVλ_For_13	TCCTC**Y**G**R**GCTGACTCAG
rhVH_For_14	CAGGT**R**CAGCTG**R**AGGAG	rhVκ_For_14	GACAT**M**CAGATGACGCAGTC	rhVλ_For_14	TCCTATGAACTGAC**W**CAG**Y**C
rhVH_For_15	CAGGTGCAGTTGGTGCAG	rhVκ_For_15	GAC**R**T**Y**CA**D**ATGACCCAGTC	rhVλ_For_15	GA**R**GTTGTGTTCACTCAGCC
rhVH_For_16	GAGGT**W**CA**R**CTGGTGGAG	rhVκ_For_16	GACATCCAGATGACTCAGTCTC		
rhVH_For_17	CAGGTGAAGCTGCAGGAG	rhVκ_For_17	GAAACTGTGATGATGCAGTC		
rhVH_For_18	CAGGTGCAGCTGCAAGAG	rhVκ_For_18	GACATC**S**AGGTGACCCAGTC		
rhVH_For_19	CAGGTGCA**M**CT**R**CAGGAG	rhVκ_For_19	GCCATCCAGATGACCCAGTC		
rhVH_For_20	GAGGTCCAGCTGGT**R**CAG	rhVκ_For_20	GAT**R**TTGTGATGA**Y****Y**CAGACTC		
rhVH_For_21	GAGGTGCAGCTAGTGGAG				
rhVH_For_22	GTGGAGCAGCTGGTGG				
rhVH_For_23	GAGGCGCAGCTG**R**TG				
rhVH_For_24	CAGGT**R**CA**R**CTGGTGCAG				
rhVH_For_25	GAGGTGCAGTTGCTGGAG				
Reverse primers					
rhVH_Rev_1	CTGA**R**GAGACGGTGACC	rhVκ_Rev_1	GGGACCAAA**S**TGGA**Y**ATCAAA	rhVλ_Rev_1	CCCGGCTCACCGTCCTAG
rhVH_Rev_2	CTGAGGAGACGGTGACG	rhVκ_Rev_2	GGGACCAA**R**GTGGAGATCAAA	rhVλ_Rev_2	CCAA**R**TTGACCGTCCTCG
		rhVκ_Rev_3	GGGACCA**R**GGTGGA**M**ATCAAA		
		rhVκ_Rev_4	GGGACCAAGGTGGAGCTCAAA		

Bold letter indicates degenerative base; underlined letter indicates a non-silent degeneration.

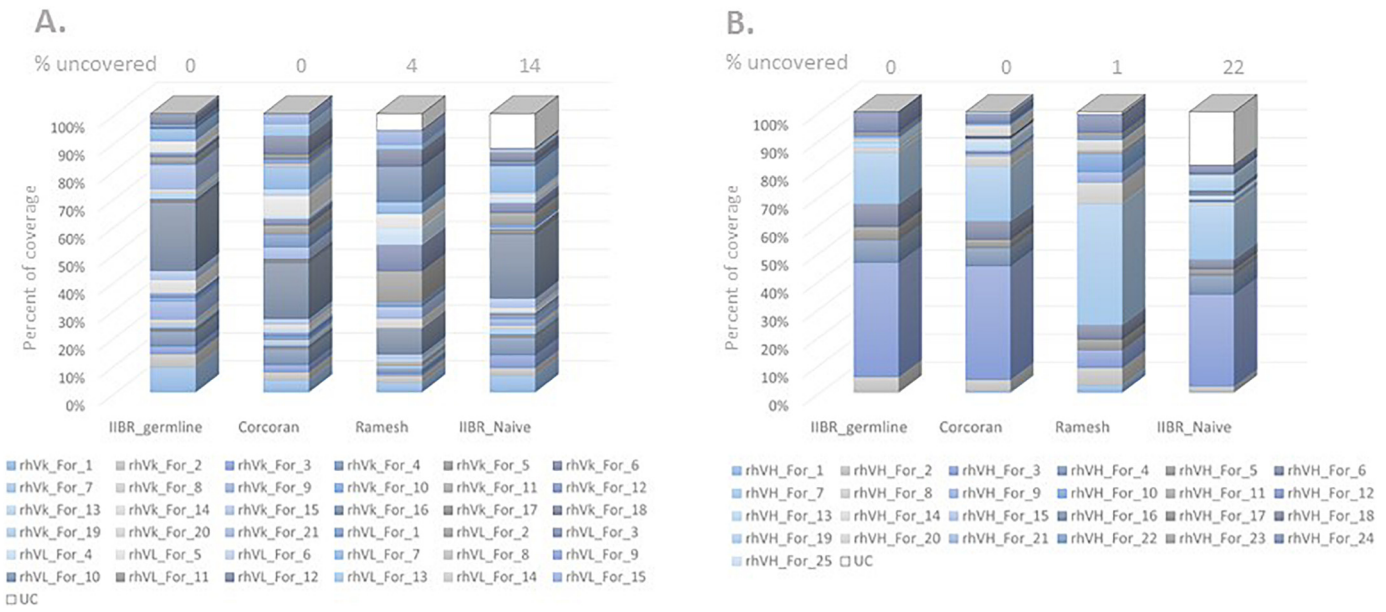

Fig. 2. Coverage evaluation of v2018 primers. Individual primer specificities of the V_H (A) and V_L (B) were evaluated against the various datasets, as indicated, using the MFEprimer tool (Qu and Zhang, 2015). The extent of coverage is provided by the percent of sequences covered by each of the particular primers (colored in variations of blues and grays). In case where the summation of the coverage of all primers in the set against a particular dataset is < 100%, the uncovered fraction is denoted by "UC" (white-colored in the column chart and actual value indicated on the top row).

sequences. When comparing the IIBR resulting database with currently available databases (that were used as starting set for germline discovery), only a limited overlap is perceived (Supplementary Fig. S1). This observation is not surprising in light of recent studies (Corcoran et al., 2016) which disclosed a relatively high genetic diversity between primate individuals. It might be therefore suggested that construction of individualized germline datasets may be required as the basis for subsequent analysis of the repertoire following an intervention such as vaccination.

3.1.2. Evaluation of the primer set coverage

In the past, we have constructed a primer set (denoted v2012), comprising of 16, 13 and 13 forward primers directed to the framework 1 (FR1) region of V_H,V_κ and V_λ, respectively, a set which was designed to cover the known rhesus macaque germline sequences available at the time in public databases (Noy-Porat et al., 2016). The availability of the IIBR germline sequences from non-immunized animals identified in this study as well as other recently published germline sequences, paved the way to refine and update the v2012 primer set, and establish a novel primer set designed to optimally cover currently known rhesus macaque germline sequences. The updated primer set, v2018, comprises of 25, 20 and 15 primers for V_H, V_κ and V_λ, respectively (Table 1). The specificity of each primer against the different datasets was assessed using the MFEprimer tool (Qu and Chenggang, 2015). The analysis was conducted initially with strict parameters (no mismatches allowed). A subsequent, less stringent examination of the results and of the predicted amplicons generated was employed, accepting primer matches with limited divergence (1–2 mismatches, and expected Tm not lower than 52 °C). The total coverage of the degenerate primers included in the novel primer set against each of the datasets examined is presented in Fig. 2 (V-gene sequences which included "N" within FR1, or missed a part of it, were excluded). It may be concluded that the v2018 primer set fully covers the germline repertoire of rhesus macaque determined in this study and the repertoire identified by Corcoran et al. (Corcoran et al., 2016), both for the V_H and V_L sequences, while as few as 1 V_H (out of 79) and 5 V_L germline gene sequences out of 112 documented by Ramesh et al. (*Ramesh* et al., *2017*), remained unrecognized by the primers.

For the purpose of efficient isolation of rhesus macaque antibodies, evaluation of the primer set coverage against the expressed repertoire is of high relevance. As one can expect, the coverage of the expressed repertoire, composed of V-genes subjected to SHM, was found to be somewhat lower. Yet, we show in this study that the v2018 primer set enabled to attain a coverage of over 78 and 86% for the V_H and V_L, respectively (Fig. 2). This extent of high coverage discloses a substantial improvement over previously published primer sets, either of primate or human origin. Furthermore, the v2018 outperforms the v2012 set, which resulted in ~50% and ~36% coverage for the V_H and V_L expressed repertoire, respectively.

Due to the high stringency applied in the theoretical specificity evaluation, it is most probable that experimentally the coverage attained by the primers will be even higher. To further confirm the specificity of the V2018 primers and the sizes of the amplicons generated, we evaluated experimentally the primers using cDNA from both non- and immunized rhesus samples as templates. This analysis resulted in efficient amplification (with comparable performance between the individual primers), yielding products with expected average sizes ranging between 370 and 386 and 293–385 for the VH and VL, respectively (Supplementary Fig. S2A). Furthermore, the calculated sizes of the V-gene expressed sequences identified in the current study are in agreement with the experimental results (Supplementary Fig. S2B).

In conclusion, in this study, we exploited rhesus expressed immunoglobulin sequences, to design a comprehensive primer set, which was shown to optimally cover the majority of up-to-date known rhesus V-genes. Such a primer set may contribute to an improved profiling of antibody repertoires as well as for correct and efficient isolation of antibodies with potential diagnostic and/or protective value.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jim.2018.11.011.

References

Avril, A., Froude, J.W., Mathieu, J., Pelat, T., Thullier, P., 2014. Isolation of antibodies from non-human primates for clinical use. Curr Drug Discov Technol 11, 20–27.
Corcoran, M.M., Phad, G.E., Vazquez Bernat, N., Stahl-Hennig, C., Sumida, N., Persson, M.A., Martin, M., Karlsson Hedestam, G.B., 2016. Production of individualized V gene databases reveals high levels of immunoglobulin genetic diversity. Nat. Commun. 7, 13642.

Glamann, J., Burton, R.D., Parren, W.H.I.P., Ditzel, J.H., Kent, A.K., Arnold, C., Montefiori, D., Hirsch, V., 1998. Simian immunodeficiency virus (SIV) envelop-specific Fabs with high-level homologous neutralizing activity: recovery from a long-term-nonprogressor SIV-infected macaque. J. Virol. 72, 585–592.

Kabat, E.A., Wu, T.T., Perry, H., Gottesman, K., Foeller, C., 1991. Sequences of Proteins of Immunological Interest. 5th edit. NIH Publication, Bethesda, MD, pp. 91–3242.

Lim, T.S., Mollova, S., Rubelt, F., Sievert, V., Dubel, S., Lehrach, H., Konthur, Z., 2010. V-gene amplification revisited - an optimised procedure for amplification of rearranged human antibody genes of different isotypes. New Biotechnol. 27, 108–117.

Miho, E., Yermanos, A., Weber, C.R., Berger, C.T., Reddy, S.T., Greiff, V., 2018. Computational strategies for Dissecting the High-Dimensional Complexity of Adaptive Immune Repertoires. Front. Immunol. 9, 224.

Noy-Porat, T., Rosenfeld, R., Ariel, N., Epstein, E., Alcalay, R., Zvi, A., Kronman, C., Ordentlich, A., Mazor, O., 2016. Isolation of Anti-Ricin protective Antibodies Exhibiting High Affinity from Immunized Non-Human Primates. Toxins 8, 1–15.

Palat, T., Hust, M., Thullier, P., 2009. Obtention and engineering of non-human primate (NHP) antibodies for therapeutics. Mini-Rev. Med. Chem. 9 (14), 1–5.

Qu, W., Chenggang, Z., 2015. Selecting specific PCR primers with MFEprimer. In: Basu, C. (Ed.), PCR Primer Design, Methods in Molecular Biology. Vol. 1275. Springer Science and business media, New York, pp. 201–213.

Qu, W., Zhang, C., 2015. Selecting specific PCR primers with MFEprimer. Methods Mol. Biol. 1275, 201–213.

Ramesh, A., Darko, S., Hua, A., Overman, G., Ransier, A., Francica, J.R., Trama, A., Tomaras, G.D., Haynes, B.F., Douek, D.C., Kepler, T.B., 2017. Structure and Diversity of the Rhesus Macaque Immunoglobulin Loci through Multiple De Novo Genome Assemblies. Front. Immunol. 8, 1407.

Turchaninova, M.A., Davydov, A., Britanova, O.V., Shugay, M., Bikos, V., Egorov, E.S., Kirgizova, V.I., Merzlyak, E.M., Staroverov, D.B., Bolotin, D.A., Mamedov, I.Z., Izraelson, M., Logacheva, M.D., Kladova, O., Plevova, K., Pospisilova, S., Chudakov, D.M., 2016. High-quality full-length immunoglobulin profiling with unique molecular barcoding. Nat. Protoc. 11, 1599–1616.

Vander Heiden, J.A., Yaari, G., Uduman, M., Stern, J.N., O'Connor, K.C., Hafler, D.A., Vigneault, F., Kleinstein, S.H., 2014. pRESTO: a toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics 30, 1930–1932.

Yaari, G., Kleinstein, S.H., 2015. Practical guidelines for B-cell receptor repertoire sequencing analysis. Genome Med 7, 121.

Ye, J., Ma, N., Madden, T.L., Ostell, J.M., 2013. IgBLAST: an immunoglobulin variable domain sequence analysis tool. Nucleic Acids Res. 41, W34–W40.