The Journal of Immunology

Ability To Develop Broadly Neutralizing HIV-1 Antibodies Is Not Restricted by the Germline Ig Gene Repertoire

Cathrine Scheepers, Ram K. Shrestha, Bronwen E. Lambson, Katherine J. L. Jackson, Imogen A. Wright, Dshanta Naicker, Mark Goosen, Leigh Berrie, Arshad Ismail, Nigel Garrett, Quarraisha Abdool Karim, Salim S. Abdool Karim, Penny L. Moore, Simon A. Travers and Lynn Morris

J Immunol 2015; 194:4371-4378; Prepublished online 30 March 2015;
doi: 10.4049/jimmunol.1500118
http://www.jimmunol.org/content/194/9/4371

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
9650 Rockville Pike, Bethesda, MD 20814-3994.

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Ability To Develop Broadly Neutralizing HIV-1 Antibodies Is Not Restricted by the Germline Ig Gene Repertoire

Cathrine Scheepers,[*,†] Ram K. Shrestha,[‡] Bronwen E. Lambson,[*,†] Katherine J. L. Jackson,[§,¶] Imogen A. Wright,[‡] Dshanta Naicker,[*] Mark Goosen,[*] Leigh Berrie,[*] Arshad Ismail,[*] Nigel Garrett,[‖,#] Quarraisha Abdool Karim,[‖,] Salim S. Abdool Karim,[‖,**] Penny L. Moore,[*,†,‖] Simon A. Travers,[‡] and Lynn Morris[*,†,‖]**

The human Ig repertoire is vast, producing billions of unique Abs from a limited number of germline Ig genes. The IgH V region (IGHV) is central to Ag binding and consists of 48 functional genes. In this study, we analyzed whether HIV-1–infected individuals who develop broadly neutralizing Abs show a distinctive germline IGHV profile. Using both 454 and Illumina technologies, we sequenced the IGHV repertoire of 28 HIV-infected South African women from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 and 004 cohorts, 13 of whom developed broadly neutralizing Abs. Of the 259 IGHV alleles identified in this study, approximately half were not found in the International Immunogenetics Database (IMGT). This included 85 entirely novel alleles and 38 alleles that matched rearranged sequences in non-IMGT databases. Analysis of the rearranged H chain V region genes of mAbs isolated from seven of these women, as well as previously isolated broadly neutralizing Abs from other donors, provided evidence that at least eight novel or non-IMGT alleles contributed to functional Abs. Importantly, we found that, despite a wide range in the number of IGHV alleles in each individual, including alleles used by known broadly neutralizing Abs, there were no significant differences in germline IGHV repertoires between individuals who do and do not develop broadly neutralizing Abs. This study reports novel IGHV repertoires and highlights the importance of a fully comprehensive Ig database for germline gene usage prediction. Furthermore, these data suggest a lack of genetic bias in broadly neutralizing Ab development in HIV-1 infection, with positive implications for HIV vaccine design. ***The Journal of Immunology*****, 2015, 194: 4371–4378.**

The induction of broadly neutralizing Abs (bNAbs) is likely to be crucial for an efficacious HIV vaccine. Although the majority of chronically HIV-infected individuals develop some level of cross-neutralizing activity (1), bNAbs are generally found in <20% of HIV-infected individuals (2). The mechanisms underlying bNAb emergence are largely unknown, but a better understanding of how these Abs arise in natural infection would provide a blueprint for a vaccine designed to elicit them. Over the last few years a large number of potent and broad bNAbs have been isolated from selected HIV-infected donors. These bNAbs target conserved epitopes on the HIV envelope, including the membrane proximal external region (MPER), V2 glycans, V3 glycans, CD4 binding site (CD4bs), and the gp120/gp41 interface (3, 4). However, most bNAbs have unusual genetic features, including high levels of somatic hypermutation (SHM), long CDRH3s (the complementary determining region 3 on the H chain) (3, 5), and, for some classes, biases in germline IgH V region (IGHV) gene usage (3, 5–9).

The VRC01 class of Abs to the CD4bs (including NIH45-46, 12A12, 3BNC117, VRC-PG04, and VRC-CH31 isolated from

*Centre for HIV and Sexually Transmitted Infections, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg 2131, South Africa; [†]Division of Virology and Communicable Disease Surveillance, School of Pathology, University of the Witwatersrand, Johannesburg 2050, South Africa; [‡]South African National Bioinformatics Institute, South African Medical Research Council Bioinformatics Unit, University of the Western Cape, Bellville 7535, South Africa; [§]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales 2052, Australia; [¶]Department of Pathology, School of Medicine, Stanford University, Stanford, CA 94305; [‖]Centre for the AIDS Programme of Research in South Africa, KwaZulu-Natal 4013, South Africa; [#]Department of Infectious Diseases, Nelson R. Mandela School of Medicine, University of KwaZulu-Natal, 4041 Durban, South Africa; and **Department of Epidemiology, Columbia University, New York, NY 10032

Received for publication January 16, 2015. Accepted for publication February 24, 2015.

This work was supported by the Poliomyelitis Research Foundation, the University of the Witwatersrand Health Sciences Faculty Research Council, and the National Research Foundation. The Centre for the AIDS Programme of Research in South Africa is funded by the National Institute of Allergy and Infectious Diseases/National Institutes of Health and the U.S. Department of Health and Human Services (Grant AI51794). The South African National Bioinformatics Institute received funding from the South African Department of Science and Technology, South African National Research Foundation Study Bursary, Atlantic Philanthropies, and the South African Medical Research Council. C.S. was supported by the Columbia University–Southern African Fogarty AIDS International Training and Research Program through the Fogarty International Center, National Institutes of Health (Grant 5 D43 TW000231). P.L.M. is a Wellcome Trust Intermediate Fellow in Public Health and Tropical Medicine (Grant 089933/Z/09/Z).

Address correspondence and reprint requests to Dr. Lynn Morris, National Institute for Communicable Diseases, a Division of the National Health Laboratory Service, Centre for HIV and Sexually Transmitted Infections: HIV Virology Section (Morris Laboratory), 1 Modderfontein Road, Sandringham 2131, South Africa. E-mail address: lynnm@nicd.ac.za

The online version of this article contains supplemental material.

Abbreviations used in this article: BCN, broadly cross-neutralizing; BLAST, basic local alignment search tool; bNAb, broadly neutralizing Ab; CAPRISA, Centre for the AIDS Programme of Research in South Africa; CD4bs, CD4 binding site; FR, framework; IGHV, IgH V region; IMGT, International Immunogenetics Database; MPER, membrane proximal external region; NCBI, National Center for Biotechnology Information; NGS, next-generation sequencing; non-BCN, non–broadly cross-neutralizing; SHM, somatic hypermutation; SNP, single nucleotide polymorphism; UCA, unmutated common ancestor; VH, H chain V region.

www.jimmunol.org/cgi/doi/10.4049/jimmunol.1500118

multiple donors) was shown to preferentially use either IGHV1-2*02 or IGHV1-46*02 germline alleles (3, 10, 11). This preference is thought to be due to the electrostatic and hydrophobic contacts afforded by conserved residues in framework (FR)1, CDRH2, and FR3 of this allele (some of these features are conserved in IGHV1-46*02 and IGHV1-3*01, which are also used by CD4bs Abs) (9, 10). Anti-HIV Abs frequently use IGHV1 genes compared with non-HIV Abs (5, 6). In particular, IGHV1-69 is used by mAbs that target V2, the CD4 induced site, and gp41 in HIV infection, as well as other viral infections, such as influenza (8, 12). The preference for this particular gene can be attributed to the interaction of hydrophobic residues in the CDRH2 with helical elements or hydrophobic β-sheets like those found on the hemagglutinin of influenza and gp41 and gp120 of the HIV envelope (8, 13). The use of IGHV5-51*01 and IGHV5-51*03, which are underrepresented in mature Abs, also was reported to be favored by anti-V3 HIV Abs (6, 7).

The human germline IGHV repertoire consists of seven IGHV subgroups, which are described in the International Immunogenetics Database (IMGT; www.imgt.org). These IGHV1–7 subgroups include functional genes, open-reading frames, and pseudogenes with only functional genes being involved in Ab production (14). IGHV3 is the largest of the IGHV subgroups, with 21 functional genes, followed by IGHV1 and IGHV4 (both with 10 functional genes), and IGHV2, IGHV5, IGHV6, and IGHV7 (with three or fewer genes) (14). Most of these genes have multiple alleles, including functional and nonfunctional alleles that differ by either a single nucleotide polymorphism (SNP) or by multiple SNPs, which can be either synonymous or nonsynonymous, or frameshift mutations caused by indels that contribute diversity to the Ig gene repertoire. Furthermore, whole IGHV genes were reported to have been duplicated or deleted from the germline repertoire of some individuals, resulting in varied gene copy numbers (15, 16). IGHV genes make up the majority of the H chain V region (VH) of mature Abs and are central to Ag binding. Differences in germline IGHV repertoires between different populations were highlighted recently in an extensive study of the human Ig gene locus (16), in which African individuals were found to be particularly diverse.

Given the propensity of some HIV mAbs to use restricted IGHV genes, we examined whether HIV-infected individuals who develop bNAbs have unique IGHV repertoires compared with individuals who do not, to determine whether the ability to develop these types of Abs is genetically restricted.

Materials and Methods

Samples and ethics statement

This study involved 28 adult (>18 y) women of African ancestry (mostly Zulu speaking) with HIV-1 subtype C infection from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 (17) and 004 (18) cohorts who were being followed at urban and rural clinics in KwaZulu-Natal, South Africa. Of the 28 women, 13 developed broadly neutralizing anti-HIV Abs, 13 did not develop broadly neutralizing Abs despite chronic HIV infection, and 2 were intermediate neutralizers (Supplemental Table I). Throughout this article, "BCN" is used to refer to the HIV-infected women who develop broadly neutralizing Abs, and "non-BCN" refers to women who did not develop broadly neutralizing Abs. BCN individuals were defined as those whose sera from 2 y postinfection were able to neutralize 33–94% (median 56%) of a panel of 18 viruses, made up of 6 subtype A, 6 subtype B, and 6 subtype C viruses, of which 2 viruses were isolated from the women in the cohort. Non-BCN individuals neutralized 0–11% of the 18 viruses and had the same viral loads as the BCN individuals at 6 mo postinfection [to remove viral load biases associated with the development of bNAbs (19)].

This study was given ethics clearance from the Human Research Ethics Committee for Medical Research in Johannesburg, South Africa (clearance number M111104 for this particular study and M080470 for the CAPRISA parent study).

DNA extraction

Genomic DNA was extracted from PBMCs from each individual. Prior to extraction, PBMCs were thawed (at 37˚C) and washed with 10 ml RPMI 1640 with 10% FBS. Genomic DNA was extracted from the pellet from all 28 individuals using a Promega Wizard Genomic purification kit.

Primer design and amplicon library construction

Alignments were created for each IGHV subgroup using sequences obtained from ENSEMBL (http://www.ensembl.org/index.html) and IMGT (http://www.imgt.org/). Forward primers (with the exception of IGHV1-F) were mapped to the intron in leader sequence. The reverse primers were mapped to the intron after the CDR3 region. IGHV1, IGHV3, and IGHV4 primers were designed based on previously published primer sets (20). New primers for IGHV2, IGHV5, IGHV6, and IGHV7 were designed based on related sequences for each subgroup. A total of eight primer sets was designed to amplify all seven subgroups. The binding properties of the primer sets were determined using University of California Santa Cruz's BLAT database (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start) and the National Center for Biotechnology Information (NCBI)'s basic local alignment search tool (BLAST). Additional next-generation sequencing (NGS)-specific sequences were added to the gene-specific primers to allow sequencing of the Roche 454 and Illumina MiSeq. The 454 primers included the 454 adapter sequence, key sequence, and 10-bp MID sequences (four unique sequences in total; primers are listed in Supplemental Table II). The Illumina MiSeq primers included a MiSeq index-binding tag and the read 1 or read 2 tags (used for paired-end reads; primers are listed in Supplemental Table III). Nextera XT Indexing tags, which allow pooling of samples on the MiSeq, were added during library construction rather than included into the primer.

Each IGHV subgroup was amplified three times for each individual to ensure adequate coverage of the subgroup and minimize PCR bias. The PCR conditions for all eight amplicons and both 454 and Illumina primers were the same, with the exception of the annealing temperatures: 56˚C for IGHV1, IGHV3a, and IGHV5; 59˚C for IGHV2, IGHV6, and IGHV7; and 55˚C for IGHV3b and IGHV4. The PCR conditions were as follows: initial denaturation at 94˚C for 3 min, 35 cycles of denaturation at 94˚C for 15 s, annealing for 45 s, extension for 1 min at 72˚C, final extension at 72˚C for 8 min, and held at 4˚C. Each PCR contained 17.5 μl dH_2O, 2.5 μl Roche FastStart High Fidelity 10× buffer with 18 mM $MgCl_2$, 0.5 μl deoxyribonucleotide triphosphates mix (10 mM each), 1 μl each primer (10 μM), 0.25 μl Roche FastStart High Fidelity Enzyme (5 U/μl), and 1 μl 10 ng/μl DNA. The PCR amplicon lengths ranged from ~390 to 440 bp. All replicates of the eight amplicons from each individual were pooled to create a full IGHV repertoire for each individual.

Next-generation sequencing

Fifteen individuals were sequenced on the Roche 454 GS Junior, with four individuals pooled per sequencing run. All clean-up procedures and sequencing on the Roche 454 GS Junior (Titanium) was done as per the manufacturer's recommendations.

To confirm rare alleles that were only found in single individuals, 11 of the individuals sequenced on the GS Junior were resequenced on the Illumina MiSeq along with 13 additional individuals. A custom amplicon-sequencing approach was used for the MiSeq sequencing. Nextera XT Indexing tags were added to the pooled MiSeq amplicon libraries for each individual using 5 μl each Nextera XT index (two per sample), 1 μl cleaned PCR product (amplicon library prep), 0.5 μl Epicentre FailSafe enzyme, 25 μl Epicentre FailSafe PCR PreMix, and 13.5 μl dH_2O. Thermocycler conditions were 72˚C for 3 min, 95˚C for 30 s, 12 cycles of 95˚C for 10 s, 55˚C for 30 s, and 72˚C for 30 s, followed by a final extension at 72˚C for 5 min. All products were checked on an Agilent Bioanalyzer and cleaned up using 0.75× AMPure Beads, according to the manufacturer's protocol. Each sample was quantified on a Qubit and diluted to 8 nM. A single 8-nM pooled library was created by pooling 4 μl each diluted sample, 5 μl of which was denatured using 5 μl 0.2 N NaOH, according to the MiSeq protocol. A final concentration of 12 pM denatured DNA library with 15% PhiX control was run onto the Illumina MiSeq, using the MiSeq reagent kit (version 2) with 2 × 250 paired-end reads.

Sequence data analysis

FASTQ files were extracted from the raw GS Junior output (.sff files) or automatically generated on the MiSeq. The resulting sequences were quality trimmed using QTrim (21). Sequences with a minimum read length of 260 bp and a mean quality score ≥ 23 (≥ 99% confidence) were considered high quality and used for downstream analyses. Paired-end MiSeq reads were merged using PEAR (22), after quality trimming. All identical sequences

were collapsed to single unique reads. Unique reads compiled from six or more identical sequences were used for further analysis, whereas all unique reads compiled from less than six sequences were discarded from further analysis. The unique sequences were compared with a database of IGHV alleles downloaded from IMGT using a custom BLAST approach with a match reward score of 2, a mismatch penalty of 5, a gap penalty for insertions of 16, and a gap extension for deletions with a penalty of 4. All related sequences for each IGHV gene were aligned using RAMICS (23).

Sequences with 100% matches to functional IMGT alleles were assigned as that particular IMGT allele. Sequences matching open-reading frames or pseudogenes were removed from further analysis. Sequences with non-synonymous and/or synonymous SNPs, insertions, or deletions compared with their top matched functional IMGT sequence were given the name of the top-scoring IMGT match with the suffix "m" to denote a single mismatch (nonsynonymous and synonymous SNPs) (e.g., IGHV1-2*5m) and "mm" for multiple SNPs, insertions, or deletions (e.g., IGHV1-2*5mm). If more than one sequence with a single mismatch or multiple mismatches to the same top-scoring IMGT functional sequence was observed, an additional numerical identifier was given to the name (e.g., IGHV1-2*5m2 and IGHV1-2*5mm2). These sequences were compared with those listed in the Ig Polymorphism Database, GenBank, NCBI's dbSNP, NCBI's IgBLAST, and other published IGHV data (24–28). If 100% matches were found in any of these databases or publications, these sequences are described as non-IMGT alleles. If no match was found, and sequences were observed in multiple individuals or in both sequencing platforms in a single individual, we report these as novel alleles.

bNAb sequences were obtained from the HIV Sequence Database - CATNAP (http://www.hiv.lanl.gov/components/sequence/HIV/neutralization/main.comp) or GenBank. The germline IGHV gene usage for these sequences also was obtained from CATNAP or the relevant publication. The bNAb sequences were compared with our IMGT, non-IMGT, and novel allele sequences, using blastn from BLAST 2.2.29+ with a match reward of 1, a mismatch penalty of −1, gap deletions of 5, a gap extension of 2, a word size of 7, and an E value threshold of $1e^{-10}$. mAbs were isolated previously from seven of the CAPRISA participants examined in this study using single-cell sorting and PCR amplification of the H and L chains (N. Mkhize, E. Gray, and L. Morris, unpublished observations). The majority of these mAbs were not HIV specific. The sequences obtained from these mAbs were used to assign germline IGHV gene usage in each participant.

Statistical analysis

The two-tailed Fishers exact test with 95% confidence intervals was used to assess the difference in germline gene repertoires between the BCN and non-BCN sample groups.

Results

Novel germline IGHV alleles in South African women

The germline IGHV gene repertoire of 28 HIV-infected women in the CAPRISA 002 and 004 cohorts, based in KwaZulu-Natal, South Africa, were sequenced using both the 454 and Illumina NGS platforms. Sequences with perfect matches to alleles listed in IMGT (29) were given the IMGT allele name and are referred to as

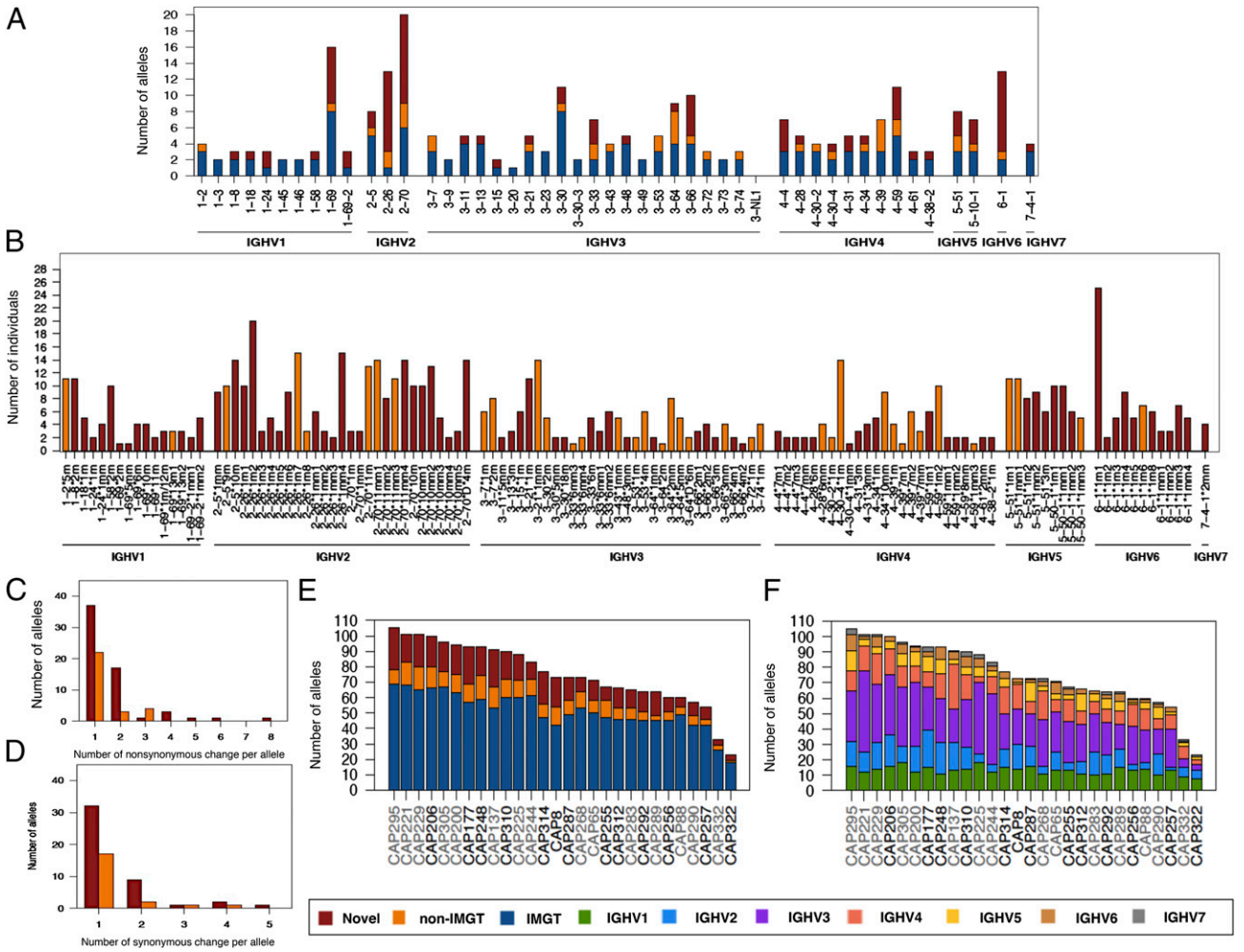


FIGURE 1. Novel germline IGHV alleles in 28 South African individuals. (**A**) Number of alleles observed for each IGHV gene. Shown are the numbers of novel (red), non-IMGT (orange), and IMGT (blue) alleles. (**B**) Prevalence of novel and non-IMGT germline IGHV alleles. (**C**) Number of non-synonymous changes in each allele compared with top-matched IMGT allele. (**D**) Number of synonymous changes in each allele compared with top-matched IMGT allele. (**E**) Total number of IMGT, novel, and non-IMGT germline IGHV alleles. BCN individuals are highlighted in black, and non-BCN individuals are highlighted in gray. (**F**) Total number of germline alleles from each IGHV subgroup. Shown are the total numbers of alleles colored according to IGHV subgroup.

IMGT alleles. Sequences that did not match IMGT alleles as a result of synonymous or nonsynonymous mutations or indels, but were matches to non-IMGT Ig gene databases (30–33) or Ig gene publications (24–28), were given the published name and are referred to as non-IMGT alleles. Sequences with no matches to either IMGT or non-IMGT alleles and were observed in multiple individuals or across both technologies in one individual are referred to as "novel" alleles.

A total of 47 functional IGHV genes representing all seven subgroups was found in our cohort (Fig. 1A). Only one published gene, IGHV3-NL1, first identified in Papua New Guinea (20), was not detected. There was a wide range in the number of alleles associated with each gene, with 20 alleles in the IGHV2-70 gene compared with only a single allele for the IGHV3-20 gene. Of the 259 alleles identified in this population, just over half (~52%, n = 136) had an exact match to those listed in IMGT, whereas ~15% (n = 38) were non-IMGT alleles, and ~33% (n = 85) were novel alleles. Eighty-one of the eighty-five novel alleles (95%) were observed in more than one individual (Fig. 1B). The four novel alleles observed in single individuals were confirmed with both 454 and Illumina sequences. The most commonly observed novel alleles were IGHV6-1*1m1 found in 25 individuals (~89%) and IGHV2-26*1m2 found in 20 individuals (~71%), both of which contained a single SNP compared with the known alleles. Overall, 48% (n = 60) of the novel and non-IMGT alleles identified in this study were found in at least four individuals, indicating that many are fairly common. Furthermore, this study significantly expanded the number of alleles for certain genes. For example, IGHV6-1, which only had 2 alleles reported in IMGT was found to have 10 novel alleles and 1 non-IMGT allele (Fig. 1A, 1B). Similarly, IGHV2-26 was found to have 13 alleles (including 10 novel alleles and 2 non-IMGT alleles) compared with 1 recorded in IMGT. Other significant expansions were in the IGHV1-69 and IGHV2-70 genes.

The majority (~97%, n = 119) of novel and non-IMGT alleles had single or multiple SNPs (synonymous and nonsynonymous),

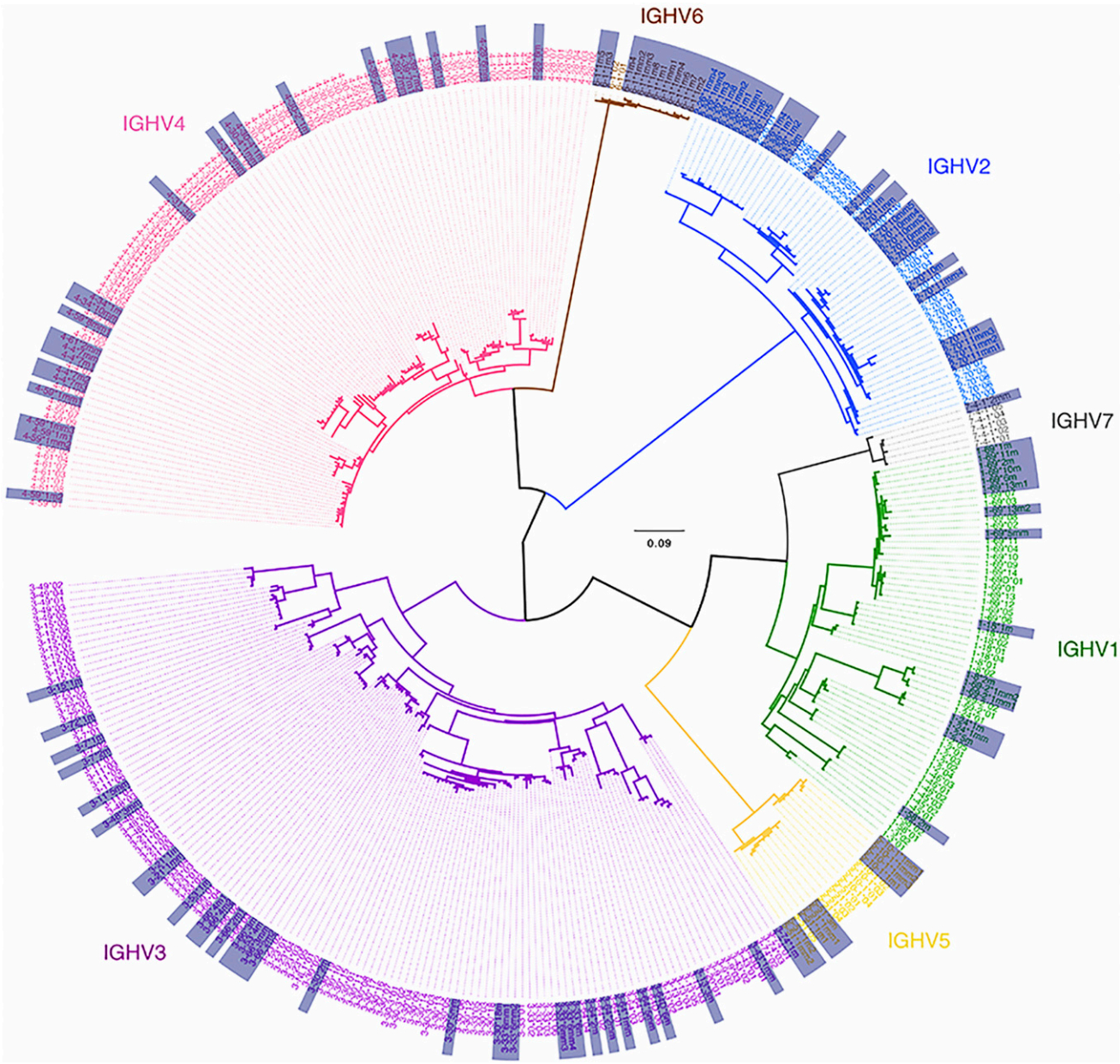


FIGURE 2. Phylogenetic tree of all germline IGHV alleles observed in 28 South African individuals. Shown are all IMGT, non-IMGT, and novel alleles observed in this study. Alleles are colored according to IGHV subgroup, and the non-IMGT and novel alleles are highlighted in the gray boxes.

with single nonsynonymous SNPs being the most common (Fig. 1C, 1D). One novel allele (IGHV2-26*1mm4, observed in 15 individuals) had eight nonsynonymous SNPs (Fig. 1B, 1C). Four novel alleles had indels relative to their most closely matched IMGT allele, and three of the four had additional SNPs. IGHV3-11*5mm, observed in two individuals, had a full codon insertion, whereas frameshifts as a result of single nucleotide deletions were found in IGHV4-59*8mm and IGHV4-61*2mm, each observed in two individuals, and IGHV6-1*1m8, which was found in six individuals. It is likely that these three alleles with the frameshift mutations are not functional, but rather are novel pseudogenes.

When the repertoires for each of the 28 individuals were analyzed separately, it was apparent that there was a wide range in the number of alleles/individual (Fig. 1E, 1F). CAP295 had the largest number of alleles (n = 105), whereas CAP322 had the fewest (n = 23). Each individual's repertoire included IMGT, non-IMGT, and novel alleles. There was variation in the number of alleles/IGHV family, because some individuals lacked alleles for IGHV6 or IGHV7, whereas others had smaller numbers of IGHV2, IGHV4, or IGHV5 alleles. A phylogenetic tree of all 259 alleles, including IMGT, non-IMGT, and novel alleles, sequenced in this study revealed separate clustering of each of the IGHV subgroups, with all of the novel and non-IMGT alleles clustering with their respective IMGT genes (Fig. 2).

Novel germline IGHV allele usage in bNAbs

To determine whether these novel and non-IMGT germline alleles are being used by the human immune system to generate functional Abs, we analyzed the germline gene usage of mAbs isolated from seven of the CAPRISA participants in this study. We found 15 mAbs that made use of seven novel or non-IMGT alleles identified from NGS of the 28 individuals (Fig. 3A). Of these, two clonally related mAbs were HIV-1 specific (CAP255-37C and CAP255-18F). An example of one of the non-IMGT alleles used by two mAbs isolated from two individuals (CAP177 and CAP248) is shown in Fig. 3B. Based on the IGHV repertoires from CAP177 and CAP248, the germline IGHV gene predicted to be used by these two mAbs is IGHV2-5*9m, which has a serine (S) at position 62 rather than a glycine (G) present in the closest related IMGT allele IGHV2-5*09. Compared with the germline allele, all other SNPs found in these two mAbs are likely the results of SHM. Of the remaining seven novel and non-IMGT alleles used by these mAbs, four (57%) had a better match at the amino acid level and three (43%) had a better match at the nucleotide level compared with the closest matching IMGT allele (Fig. 3A). We also analyzed the IGHV gene usage of 57 well-known HIV-1 bNAbs and found 14 instances in which the novel or non-IMGT alleles, identified in this study, provided the same or a better match than did their currently predicted IMGT allele (Fig. 3C). Two of the predicted alleles (IGHV4-39*7m2 and IGHV4-59*1m2) also were observed among the mAbs isolated from the CAPRISA participants. An example of alternate IGHV allele usage for the bNAbs is shown in Fig. 3D: PGT130 and PGT131 are more likely to be using IGHV4-39*7m1 or IGHV4-39*7mm, which like the mAbs have a valine (V) at the second amino acid, rather than a leucine (L), as seen in IGHV4-39*07.

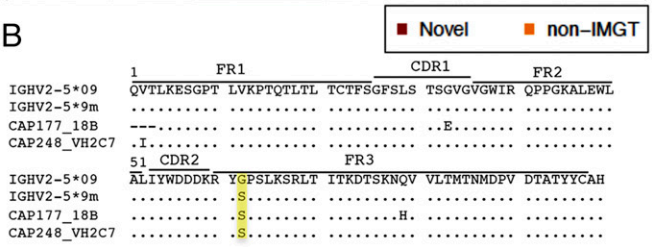
Comparison of germline IGHV repertoires in BCN and non-BCN individuals

Given that some HIV broadly neutralizing mAbs show biased variable H chain gene usage, we analyzed the germline IGHV repertoires of BCN and non-BCN individuals to determine whether the ability to develop bNAbs was associated with a particular genotypic profile. Among the 28 CAPRISA women, 13 developed bNAbs in their plasma after 2–4 y of HIV infection; 13 women did not develop bNAbs, despite matching viral loads; and 2 were intermediate neutralizers [Supplemental Table I (19) and N. Mkhize, E. Gray, and L. Morris, unpublished observations]. Comparison of the BCN and non-BCN individuals revealed that they had similar profiles of IGHV genes (Fig. 4A), approximately the same number of alleles for each of those genes (Fig. 4B), and the same number of potential alleles (Fig. 1E, 1F). The only gene that showed

A

Participant	mAb	Predicted Allele	Corresponding IMGT Allele	Nucleotide Change	Amino Acid Change
CAP255	75D	IGHV1-18*1m	IGHV1-18*01	A169G	N57D
CAP255	37C, 18F	IGHV1-69*6m	IGHV1-69*06	G148A	G50R
CAP177, CAP248	18B, VH2C7	IGHV2-5*9m	IGHV2-5*09	G184A	G62S
CAP255	53E	IGHV3-15*1m	IGHV3-15*01	G132A	none
CAP177	2D, 13F, 60B, 6C, 13G	IGHV3-30*2m	IGHV3-30*02	T180C	none
CAP177	23-3G, 1C, 5G	IGHV4-39*7m2	IGHV4-39*07	G298C	none
CAP255	71E, 91B	IGHV4-59*1m2	IGHV4-59*01	T94C	Y32H

B



C

bNab	IMGT Allele	Alternate Allele(s)	Nucleotide Change	Amino Acid Change
2G12	IGHV3-21*01	IGHV3-21*1m	G291A	none
CH103	IGHV4-59*01	IGHV4-59*1m2	T94C	Y32H
PGT121-123	IGHV4-59*01	IGHV4-59*1m2	T94C	Y32H
PGT125-128	IGHV4-39*07	IGHV4-39*7m2	G298C	none
PGT130-131	IGHV4-39*07	IGHV4-39*7m1 IGHV4-39*7mm	C4G C4G, G298C	L2V L2V
PGT135-137	IGHV4-39*07	IGHV4-39*7m1 IGHV4-39*7m2 IGHV4-39*7mm	C4G G298C C4G, G298C	L2V none L2V

D

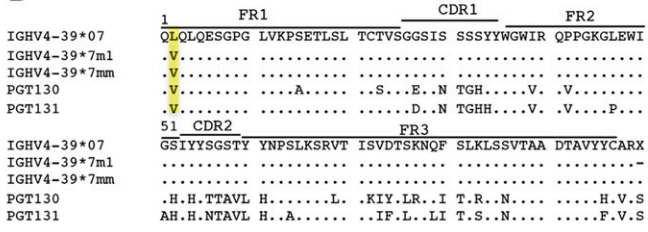


FIGURE 3. Novel IGHV alleles used by isolated mAbs and some bNAbs. (**A**) Germline IGHV allele usage for isolated mAbs from study participants. The germline IGHV alleles used, as well as their corresponding IMGT allele, nucleotide, and amino acid substitutions between the two alleles, are shown. (**B**) Alignment of germline IGHV allele usage for isolated mAbs from two study participants. Highlighted in yellow are the differences in amino acids at position 62: G (glycine) and S (serine). Amino acid changes not highlighted in the figure are likely a result of SHM during Ab maturation. (**C**) Germline IGHV allele usage for selected bNAbs. The published IMGT predicted allele and alternate allele(s) predicted from IGHV alleles in 28 South African individuals are shown. (**D**) Alignment of predicted germline IGHV allele usage for PGT130 and PGT131. Highlighted in yellow are the differences at the second amino acid: V (valine) and L (leucine). Amino acid changes not highlighted in the figure are likely a result of SHM during Ab maturation.

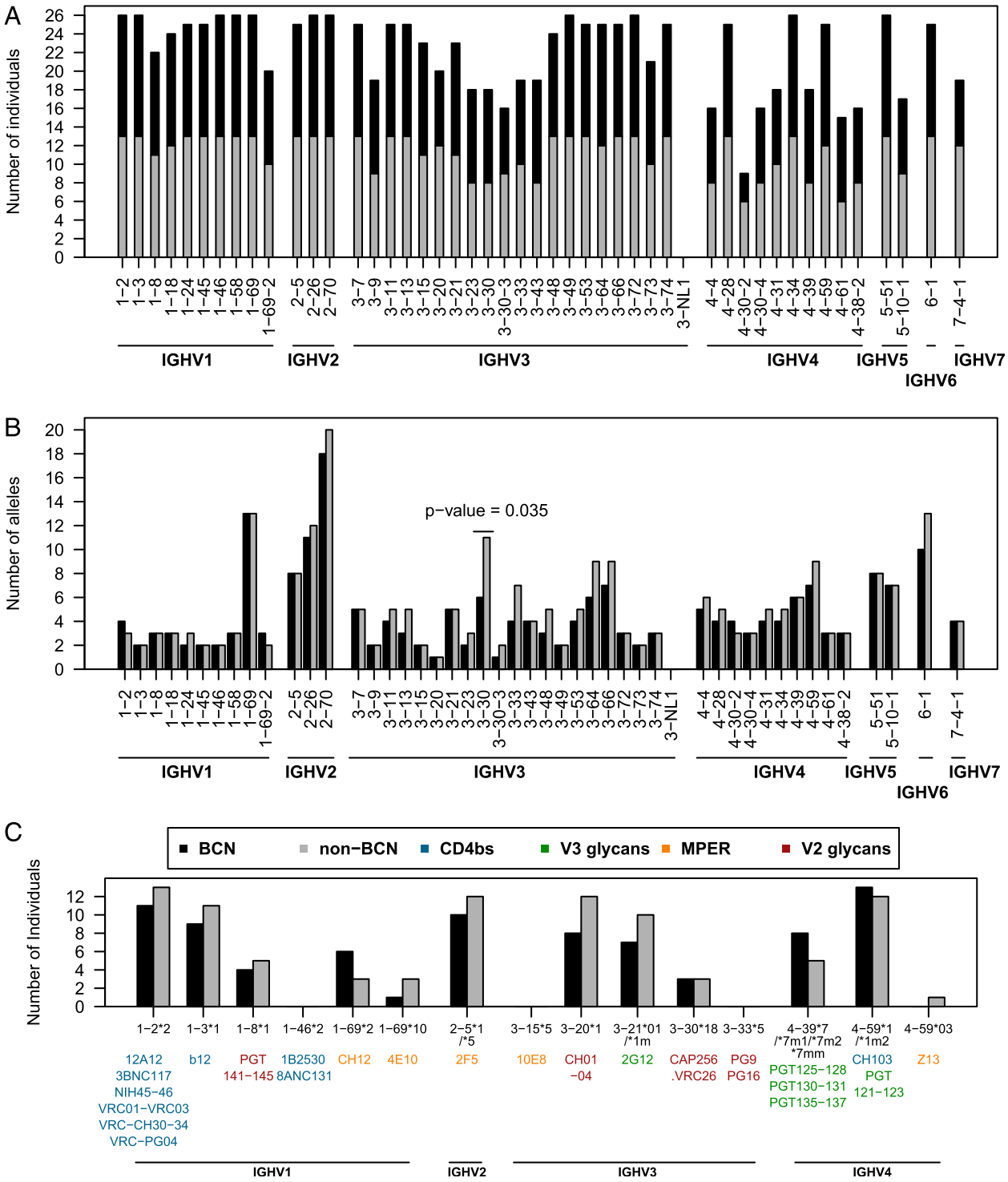


FIGURE 4. Comparison of germline IGHV repertoires between BCN and non-BCN individuals. (**A**) Number of BCN and non-BCN individuals with each of the functional IGHV genes in their IGHV repertoire. (**B**) Number of alleles observed for each of the functional IGHV genes in BCN and non-BCN individuals. (**C**) Number of BCN and non-BCN individuals with the same IGHV alleles in their repertoire as those used by monoclonal bNAbs.

statistical significance between the two groups was IGHV3-30 (p = 0.035, although significance was lost following Bonferroni multiple testing correction); non-BCN individuals had five more alleles (n = 11) compared with BCN individuals (n = 6) (Fig. 4B).

We also compared the frequency of the specific germline IGHV alleles used by broadly neutralizing mAbs between the non-BCN and BCN groups. Of the 57 monoclonal bNAbs targeting one of four different epitopes on the HIV envelope glycoprotein, most made use of IGHV1, IGHV3, and IGHV4 germline genes (Fig. 4C), and, as reported above, some may use novel/non-IMGT alleles (Fig. 3C). This included IGHV1-46*02, used by some CD4bs Abs; however, there were two other alleles for this gene, including IGHV1-46*01, which were found in all individuals. Similarly, IGHV3-15*05 (used by bNAb 10E8) and IGHV3-33*05 (used by bNAbs PG9 and PG16) were not observed in any of the individuals in this study, but other alleles were more commonly observed,

including some novel/non-IMGT alleles. The IGHV1-2*02 allele, which is also used preferentially by CD4bs Abs (shown in blue in Fig. 4C), occurred at high frequency in both the non-BCN and BCN groups. The most prevalent IGHV alleles in both groups were IGHV4-59*01 (used by the CD4bs bNAb CH103 and the V3/glycan bNAbs PGT121–123), IGHV1-2*02 (used by CD4bs mAbs), IGHV1-3*01 (used by the CD4bs bNAb IgG1b12), and IGHV2-5*01/05 (used by bNAb 2F5 that targets the MPER). Importantly, there was no difference in the frequency of any of these alleles between the 13 BCN and 13 non-BCN individuals. Thus, the inability of all HIV-infected individuals to develop bNAbs is not due to a paucity of the relevant alleles in their IGHV germline gene repertoires.

Discussion

We analyzed the germline IgH variable gene repertoire encoded in the genomic DNA of individuals from KwaZulu-Natal, South Africa, and noted that ∼48% of the alleles seen in this population are not reported in IMGT. Some of these alleles (non-IMGT alleles) were described in published studies of rearranged Abs, although most alleles were novel and are described in this article for the first time, to our knowledge. Further analysis of these IGHV repertoires revealed there to be no differences between individuals who developed bNAbs to HIV and those who did not, despite equivalent antigenic load. Because the induction of these types of Abs is considered essential for an effective HIV vaccine, these data suggest that the ability to develop bNAbs is not restricted by the IGHV repertoire.

Previous studies reported differences in the frequencies of IGHV genes between different populations, with Africans showing particularly unique profiles (16). The presence of IGHV3-64D, IGHV5-10-1, and IGHV7-4-1 was reported to be lower in African (Luyha, Maasai, and Yoruba) populations compared with Asian and European groups (16). We found a similarly low frequency of IGHV3-64D (14%) in this Zulu-speaking South African population. However, IGHV5-10-1 was observed at a higher prevalence (64%) compared with the studied African, Asian, and European groups (0.03–16, 20–21, and 34–48%, respectively) (16), whereas IGHV7-4-1 frequencies were similar (75%) to those seen in Asian (∼78%) and European (54–75%) groups (16). IGHV1-69-2, IGHV3-43D, and IGHV4-38-2 genes were reported to be common in African populations (16), which we corroborated in our study for IGHV1-69-2 and IGHV4-38-2 (75 and 61%, respectively). However, we observed IGHV3-43D*01 in only 29% (8/28) of individuals studied, which is lower than the prevalence reported in other African groups (45–65%) (16).

We showed that both novel and non-IMGT alleles are being used by mAbs isolated from individuals in this study, as well as some well-characterized anti-HIV bNAbs. This included CH103 and 12 Abs in the PGT121-137 series, all of which were isolated from African donors (34–36). Given that three of the novel and non-IMGT alleles (IGHV3-21*1m, IGHV4-39*1m2, and IGHV4-39*7m2) potentially being used by these bNAbs were fairly common in the South African individuals (39, 36 and 21%, respectively), it is perhaps not surprising that these were found to contribute to functional Abs in other African individuals. IGHV4-39*7m1 and IGHV4-39*7mm were less common in our study group (4 and 11%, respectively). The other novel and non-IMGT alleles used by mAbs isolated from the study participants also were fairly commonly observed, ranging from 14 to 36%. Only 2 of the 15 mAbs isolated from the study participants were able to bind HIV, highlighting that the use of novel and non-IMGT alleles by functional Abs is not HIV specific and, thus, could play a role in immune responses to other diseases or infections.

Despite the wide range in the number of IGHV alleles present in each individual, there were no differences in the overall germline IGHV repertoires between BCN and non-BCN individuals. This extended to a subanalysis of the genes and alleles used by known broadly neutralizing mAbs whose frequency differed within the cohort but not between BCN and non-BCN groups. The only gene that showed a difference between the two groups was IGHV3-30: non-BCN individuals had more alleles than did BCN individuals. IGHV3-30 is used by the CAP256-VRC26 family of broad and potent V1V2 Abs, which were isolated from CAPRISA donor CAP256, a participant in this study (37). We also noted that alleles IGHV1-46*02, IGHV3-15*05, and IGHV3-33*05, used by CD4bs, MPER, and V2 glycan bNAbs, respectively, were not found in this South African cohort. However, other alleles of these genes were observed in this cohort, as well as allelic variants used by other bNAbs targeting the same epitopes. Thus, although the specific alleles used by these bNAbs were not present, it did not preclude these individuals from generating Abs to these epitopes. In addition, CH103 and PGT121-123, which both use IGHV4-59*01 (or the non-IMGT allele IGHV4-59*1m2), target different epitopes (CD4bs and V3 glycans, respectively), demonstrating that a single germline allele can be used by Abs against numerous HIV envelope targets.

The germline IGHV gene makes the greatest contribution to the unmutated common ancestor (UCA) Ab encoding the entire CDRH1, CDRH2, FR1, FR2, and FR3 regions. Elegant studies on the UCAs of the VRC01 class of Abs, which use the VH1-2*02 allele, showed how two glycans in the V5 region of the HIV envelope obstruct binding of the UCA (9, 38). Such studies guide the design of suitable envelope immunogens able to trigger the VRC01 class of Abs (9). However, other studies aimed at identifying UCA-binding envelopes for PG9 and other bNAb lineages were less successful, primarily because the UCA approximations used in such studies were predicted from highly mutated Abs and using incomplete germline databases. The public availability of a comprehensive database of germline Ig genes, including data from this study (available from the authors), will significantly enhance the accuracy with which UCAs can be inferred.

By studying an underrepresented population group in southern Africa, this investigation has greatly expanded the repertoire of germline IGHV genes. We further hypothesize that the IGHD and IGHJ germline genes, which make up the remainder of the VH region of functional Abs, are also likely to be highly variable within this population and, thus, warrant investigation. Furthermore, we showed that some of these novel and non-IMGT alleles are functionally active in both HIV and non-HIV Abs. This knowledge will contribute to a better understanding of the Ab response to HIV, as well as other infections, immunizations, and B cell pathologies. Importantly, we also showed that the development of bNAbs against HIV is not restricted by the germline IGHV repertoire, which is significant for vaccine development, because it suggests that everyone has the potential to make Abs capable of neutralizing all strains and subtypes of HIV.

Acknowledgments

We thank the participants of the CAPRISA 002 and 004 studies from whom these samples were obtained and the staff involved in all of the sample collection and processing. We also thank Dr. Mashudu Madzivhandila and Tandile Hermanus for generating the neutralization data and Dr. Nono Mkhize and Dr. Elin Gray for generating mAbs from some of these participants.

Disclosures

The authors have no financial conflicts of interest.

References

1. Hraber, P., M. S. Seaman, R. T. Bailer, J. R. Mascola, D. C. Montefiori, and B. T. Korber. 2014. Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. *AIDS* 28: 163–169.
2. Stamatatos, L., L. Morris, D. R. Burton, and J. R. Mascola. 2009. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nat. Med.* 15: 866–870.
3. Kwong, P. D., and J. R. Mascola. 2012. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. *Immunity* 37: 412–425.
4. Scharf, L., J. F. Scheid, J. H. Lee, A. P. West, Jr., C. Chen, H. Gao, P. N. Gnanapragasam, R. Mares, M. S. Seaman, A. B. Ward, et al. 2014. Antibody 8ANC195 reveals a site of broad vulnerability on the HIV-1 envelope spike. *Cell Reports* 7: 785–795.
5. Breden, F., C. Lepik, N. S. Longo, M. Montero, P. E. Lipsky, and J. K. Scott. 2011. Comparison of antibody repertoires produced by HIV-1 infection, other chronic and acute infections, and systemic autoimmune disease. *PLoS ONE* 6: e16857.
6. Gorny, M. K., X. H. Wang, C. Williams, B. Volsky, K. Revesz, B. Witover, S. Burda, M. Urbanski, P. Nyambi, C. Krachmarov, et al. 2009. Preferential use of the VH5-51 gene segment by the human immune response to code for antibodies against the V3 domain of HIV-1. *Mol. Immunol.* 46: 917–926.
7. Gorny, M. K., J. Sampson, H. Li, X. Jiang, M. Totrov, X. H. Wang, C. Williams, T. O'Neal, B. Volsky, L. Li, et al. 2011. Human anti-V3 HIV-1 monoclonal antibodies encoded by the VH5-51/VL lambda genes define a conserved antigenic structure. *PLoS ONE* 6: e27780.
8. Gorny, M. K., R. Pan, C. Williams, X. H. Wang, B. Volsky, T. O'Neal, B. Spurrier, J. M. Sampson, L. Li, M. S. Seaman, et al. 2012. Functional and immunochemical cross-reactivity of V2-specific monoclonal antibodies from HIV-1-infected individuals. *Virology* 427: 198–207.
9. Jardine, J., J.-P. Julien, S. Menis, T. Ota, O. Kalyuzhniy, A. McGuire, D. Sok, P.-S. Huang, S. MacPherson, M. Jones, et al. 2013. Rational HIV immunogen design to target specific germline B cell receptors. *Science* 340: 711–716.
10. Scheid, J. F., H. Mouquet, B. Ueberheide, R. Diskin, F. Klein, T. Y. Oliveira, J. Pietzsch, D. Fenyo, A. Abadir, K. Velinzon, et al. 2011. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science* 333: 1633–1637.
11. West, A. P., Jr., R. Diskin, M. C. Nussenzweig, and P. J. Bjorkman. 2012. Structural basis for germ-line gene usage of a potent class of antibodies targeting the CD4-binding site of HIV-1 gp120. *Proc. Natl. Acad. Sci. USA* 109: E2083–E2090.
12. Corti, D., A. L. Suguitan, Jr., D. Pinna, C. Silacci, B. M. Fernandez-Rodriguez, F. Vanzetta, C. Santos, C. J. Luke, F. J. Torres-Velez, N. J. Temperton, et al. 2010. Heterosubtypic neutralizing antibodies are produced by individuals immunized with a seasonal influenza vaccine. *J. Clin. Invest.* 120: 1663–1673.
13. Lingwood, D., P. M. McTamney, H. M. Yassine, J. R. Whittle, X. Guo, J. C. Boyington, C. J. Wei, and G. J. Nabel. 2012. Structural and genetic basis for development of broadly neutralizing influenza antibodies. *Nature* 489: 566–570.
14. Lefranc, M. P., and G. Lefranc. 2001. *The Immunoglobulin FactsBook*. Academic Press: London.
15. Watson, C. T., and F. Breden. 2012. The immunoglobulin heavy chain locus: genetic variation, missing data, and implications for human disease. *Genes Immun.* 13: 363–373.
16. Watson, C. T., K. M. Steinberg, J. Huddleston, R. L. Warren, M. Malig, J. Schein, A. J. Willsey, J. B. Joy, J. K. Scott, T. A. Graves, et al. 2013. Complete haplotype sequence of the human immunoglobulin heavy-chain variable, diversity, and joining genes and characterization of allelic and copy-number variation. *Am. J. Hum. Genet.* 92: 530–546.
17. van Loggerenberg, F., K. Mlisana, C. Williamson, S. C. Auld, L. Morris, C. M. Gray, Q. Abdool Karim, A. Grobler, N. Barnabas, I. Iriogbe, S. S. Abdool Karim; CAPRISA 002 Acute Infection Study Team. 2008. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PLoS ONE* 3: e1954.
18. Abdool Karim, Q., S. S. Abdool Karim, J. A. Frohlich, A. C. Grobler, C. Baxter, L. E. Mansoor, A. B. Kharsany, S. Sibeko, K. P. Mlisana, Z. Omar, et al; CAPRISA 004 Trial Group. 2010. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* 329: 1168–1174.
19. Gray, E. S., M. C. Madiga, T. Hermanus, P. L. Moore, C. K. Wibmer, N. L. Tumba, L. Werner, K. Mlisana, S. Sibeko, C. Williamson, et al; CAPRISA002 Study Team. 2011. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *J. Virol.* 85: 4828–4840.
20. Wang, Y., K. J. Jackson, B. Gäeta, W. Pomat, P. Siba, W. A. Sewell, and A. M. Collins. 2011. Genomic screening by 454 pyrosequencing identifies a new human IGHV gene and sixteen other new IGHV allelic variants. *Immunogenetics* 63: 259–265.
21. Shrestha, R. K., B. Lubinsky, V. B. Bansode, M. B. Moinz, G. P. McCormack, and S. A. Travers. 2014. QTrim: a novel tool for the quality trimming of sequence reads generated using the Roche/454 sequencing platform. *BMC Bioinformatics* 15: 33.
22. Zhang, J., K. Kobert, T. Flouri, and A. Stamatakis. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* 30: 614–620.
23. Wright, I. A., and S. A. Travers. 2014. RAMICS: trainable, high-speed and biologically relevant alignment of high-throughput sequencing reads to coding DNA. *Nucleic Acids Res.* 42: e106.
24. Boyd, S. D., E. L. Marshall, J. D. Merker, J. M. Maniar, L. N. Zhang, B. Sahaf, C. D. Jones, B. B. Simen, B. Hanczaruk, K. D. Nguyen, et al. 2009. Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci. Transl. Med.* 1: 12ra23.
25. Jackson, K. J., Y. Liu, K. M. Roskin, J. Glanville, R. A. Hoh, K. Seo, E. L. Marshall, T. C. Gurley, M. A. Moody, B. F. Haynes, et al. 2014. Human responses to influenza vaccination show seroconversion signatures and convergent antibody rearrangements. *Cell Host Microbe* 16: 105–114.
26. Parameswaran, P., Y. Liu, K. M. Roskin, K. K. Jackson, V. P. Dixit, J.-Y. Lee, K. L. Artiles, S. Zompi, M. J. Vargas, B. B. Simen, et al. 2013. Convergent antibody signatures in human dengue. *Cell Host Microbe* 13: 691–700.
27. Wang, C., Y. Liu, L. T. Xu, K. J. Jackson, K. M. Roskin, T. D. Pham, J. Laserson, E. L. Marshall, K. Seo, J.-Y. Lee, et al. 2014. Effects of aging, cytomegalovirus infection, and EBV infection on human B cell repertoires. *J. Immunol.* 192: 603–611.
28. Wang, Y., K. J. Jackson, J. Davies, Z. Chen, B. A. Gaeta, J. Rimmer, W. A. Sewell, and A. M. Collins. 2014. IgE-associated IGHV genes from venom and peanut allergic individuals lack mutational evidence of antigen selection. *PLoS ONE* 9: e89730.
29. The International Immunogenetics Information System Database. Available at: http://www.imgt.org/. Accessed: June 2014.
30. Immunoglobulin Polymorphism Database. Available at: http://cgi.cse.unsw.edu.au/~ihmmune/IgPdb/. Accessed: June 2014.
31. National Center for Biotechnology Information dbSNP - Short Genetic Variation Database. Available at: http://www.ncbi.nlm.nih.gov/SNP/. Accessed: June 2014.
32. National Center for Biotechnology Information GenBank. Available at: http://www.ncbi.nlm.nih.gov/genbank/. Accessed: June 2014.
33. National Center for Biotechnology Information. Immunoglobulin BLAST Tool. Available at: http://www.ncbi.nlm.nih.gov/igblast/. Accessed: June 2014.
34. Simek, M. D., W. Rida, F. H. Priddy, P. Pung, E. Carrow, D. S. Laufer, J. K. Lehrman, M. Boaz, T. Tarragona-Fiol, G. Miiro, et al. 2009. Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *J. Virol.* 83: 7337–7348.
35. Walker, L. M., M. Huber, K. J. Doores, E. Falkowska, R. Pejchal, J.-P. Julien, S.-K. Wang, A. Ramos, P.-Y. Chan-Hui, M. Moyle, et al; Protocol G Principal Investigators. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. *Nature* 477: 466–470.
36. Liao, H.-X., R. Lynch, T. Zhou, F. Gao, S. M. Alam, S. D. Boyd, A. Z. Fire, K. M. Roskin, C. A. Schramm, Z. Zhang, et al; NISC Comparative Sequencing Program. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. *Nature* 496: 469–476.
37. Doria-Rose, N. A., C. A. Schramm, J. Gorman, P. L. Moore, J. N. Bhiman, B. J. DeKosky, M. J. Ernandes, I. S. Georgiev, H. J. Kim, M. Pancera, et al; NISC Comparative Sequencing Program. 2014. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. *Nature* 509: 55–62.
38. McGuire, A. T., J. A. Glenn, A. Lippy, and L. Stamatatos. 2014. Diverse recombinant HIV-1 Envs fail to activate B cells expressing the germline B cell receptors of the broadly neutralizing anti-HIV-1 antibodies PG9 and 447-52D. *J. Virol.* 88: 2645–2657.