



Database tool

The HIV oligonucleotide database (HIVoligoDB)

João Carneiro, Adriana Resende and Filipe Pereira*

Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), University of Porto, Terminal de Cruzeiros do Porto de Leixões Av. General Norton de Matos s/n 4450-208 Porto

*Corresponding author: Tel: (+351) 22 340 18 00; Fax (+351) 22 339 06 08; Email: fpereirapt@gmail.com

Citation details: Carneiro, J., Resende, A., and Pereira, F. The HIV oligonucleotide database (HIVoligoDB). *Database* (2017) Vol. 2017: article ID bax005; doi:10.1093/database/bax005

Received 3 September 2016; Revised 12 January 2017; Accepted 16 January 2017

Abstract

The human immunodeficiency virus (HIV) is associated with one of the most widespread infectious disease, the acquired immunodeficiency syndrome (AIDS). The development of antiretroviral drugs and methods for virus detection requires a comprehensive analysis of the HIV genomic diversity, particularly in the binding sites of oligonucleotides. Here, we describe a versatile online database (HIVoligoDB) with oligonucleotides selected for the diagnosis of HIV and treatment of AIDS. Currently, the database provides an interface for visualization, analysis and download of 380 HIV-1 and 65 HIV-2 oligonucleotides annotated according to curated reference genomes. The database also allows the selection of the most conserved HIV genomic regions for the development of molecular diagnostic assays and sequence-based candidate therapeutics.

Database URL: <http://portugene.com/HIVoligoDB>

Introduction

The human immunodeficiency virus (HIV) is a significant threat to public health, being associated with the acquired immunodeficiency syndrome (AIDS), one of the most widespread infectious diseases at global scale (1). Two HIV types have been identified. HIV-1 is responsible for most cases of AIDS while HIV-2 is confined mainly to West Africa, although several HIV-2 infections can be found in Mozambique, Angola, France and Portugal (2). HIV-2 infection are less virulent and transmissible than HIV-1, although HIV-2 is known to also cause AIDS (3). The HIV genome contains three major genes encoding essential enzymes and structural proteins: *gag*, *pol* and *env*. The single-stranded RNA genome is converted by the reverse transcriptase into a double-stranded DNA that is

integrated into the genome of infected individuals (4). The process of genome replication is a source of genetic variability due to the absence of proofreading ability of the reverse transcriptase, resulting in a high mutation rate (5, 6).

A considerable number of nucleic acid techniques (NATs) have been developed for clinical diagnosis, genetic characterization of isolates, identification of genotypes, determination of viral loads, detection of drug resistance mutations and epidemiological studies (7, 8). In recent years, synthetic oligonucleotides complementary to HIV-1 RNA have been tested as possible antiviral or virucidal agents. Oligonucleotide-based therapeutics include antisense RNA, ribozymes, decoy RNAs, aptamers and small interfering RNA (9). Despite the abundance of DNA or RNA molecules targeting HIV, no organized repository of

HIVoligoDB - The HIV oligonucleotide database

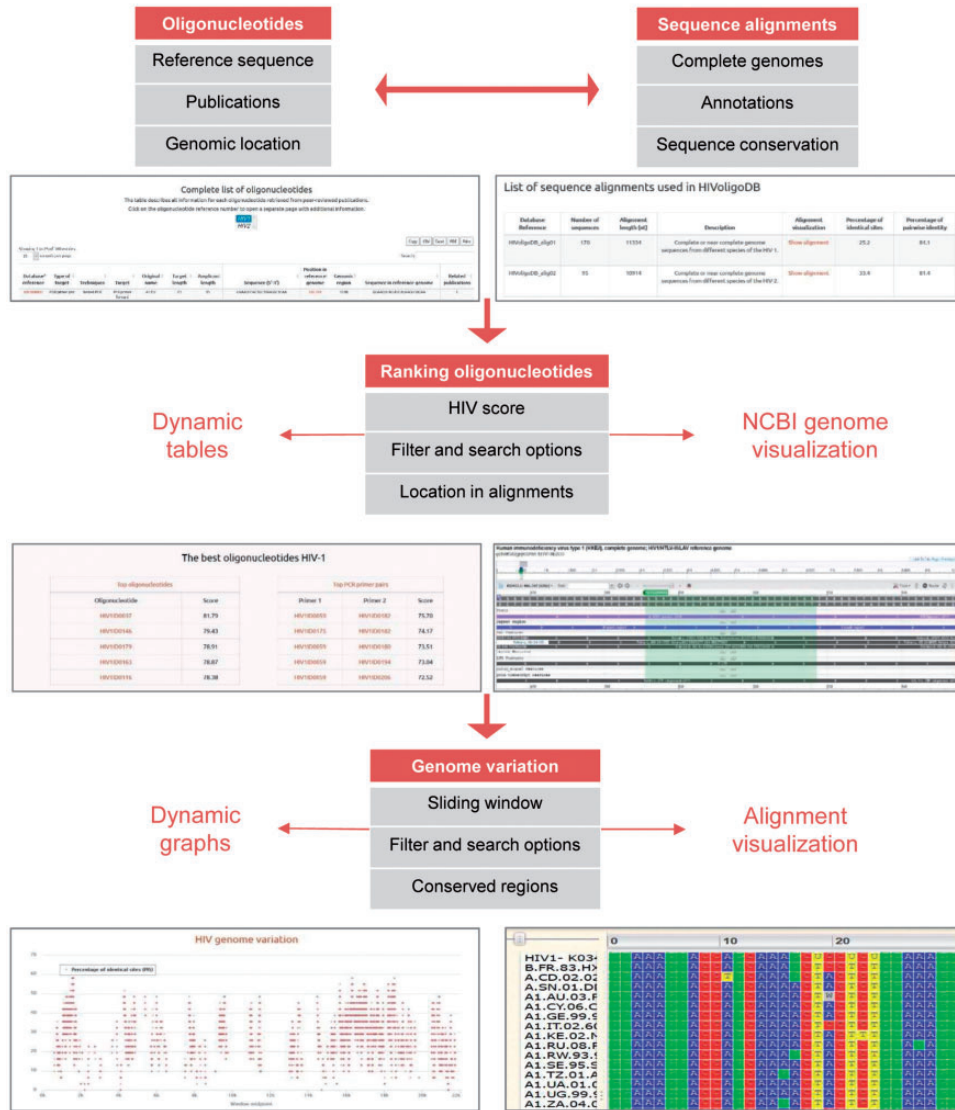


Figure 1. Screenshot of the data and tools included in the HIVoligoDB.

sequences is available. Some databases include genomic information for HIV-1 and HIV-2 (see online supplementary material for Figure S1). For example, the NCBI Probe Database (<https://www.ncbi.nlm.nih.gov/probe>) includes oligonucleotides for HIV, but lack any data on sequence conservation. The Los Alamos National Laboratory (<http://www.hiv.lanl.gov/content/index>) and the Stanford University (<http://hivdb.stanford.edu/>) HIV databases include information on genomic sequences and drug resistance mutations, but no information on oligonucleotides (10).

Here, we present the HIV oligonucleotide database (HIVoligoDB) describing a set of oligonucleotides and genomic regions that can be used to improve the efficiency of

HIV epidemiological studies and nucleic acid-based assays (Figure 1).

Materials and Methods

The HIVoligoDB (<http://portugene.com/HIVoligoDB>) includes 380 HIV-1 (236 primers and 144 probes) and 65 HIV-2 oligonucleotides (57 primers and 8 probes) retrieved from 54 peer-reviewed publications and the NCBI Probe Database. Each oligonucleotide has a specific database code (for example, HIV1ID0001) and is associated with a reference genome (K03455.1 for HIV-1 and M15390 for HIV-2). The database provides descriptive webpages (e.g. type of target, related publications) for each

oligonucleotide and a search engine to access dynamic tables with numeric data and multiple sequence alignments with complete HIV genomes. The multiple sequence alignments were retrieved from the Los Alamos National Laboratory (<http://www.hiv.lanl.gov/content/index>). The HIV-1 alignment includes sequences from all HIV-1 described sub-types (including non-M groups) (11). The HIV-2 alignment includes a carefully chosen subset of HIV-2 and SIV/SMM (simian immunodeficiency virus) sequences from the HIV Sequence Compendium (12).

The database ranks all oligonucleotides considering three main measures of sequence conservation: (a) percentage of identical sites (PIS), calculated by dividing the number of equal positions in the alignment for an oligonucleotide by its length; (b) percentage of identical sites in the last five nucleotides at the 3' end of the oligonucleotide (3'PIS)—the most critical regions for an efficient binding to the template DNA during PCR and (c) percentage of pairwise identity (PPI), calculated by counting the average number of pairwise matches across the positions of the alignment, divided by the total number of pairwise comparisons. The ranking score ('HIV score') considers the mean value of the three different measures (PIS, 3'PIS and PPI). In theory, high PIS and PPI values indicates that an oligonucleotide binds to a conserved genomic region, increasing the probability of successful binding in a PCR or other technique. Calculations of melting temperature (T_m) of each PCR primer were determined using the BioPython Melting Temperature module with Santa Lucia parameters (13) and the dimer free energy with the method used in Vienna Package (14). The melting temperatures and self-dimer free energies are described in the table 'Oligonucleotides properties' that are available through the menu 'Search'.

The database works with major web browsers (e.g. Internet Explorer, Mozilla Firefox, Chrome). The SQLite local database is used for data storage and runs on an Apache web server. The dynamic HTML pages were implemented using CGI-Perl and JavaScript and the dataset tables using the JQuery plugin DataTables v1.9.4 (<http://datatables.net/>). The database uses java graphics and dynamic tables. Python and perl algorithms implemented with BioPerl were used to perform identity and pairwise calculations.

Results

The HIVoligoDB is a free resource with detailed records of oligonucleotides to study HIV-1 and HIV-2. The database has a core integration of multiple data from different sources, which facilitates the analysis of hundreds of records through search and filtering processes. The HIVoligoDB is

being constantly updated with new oligonucleotides added once they are described in peer-reviewed publications. We have used the HIVoligoDB to perform calculations over 445 oligonucleotides located in different HIV-1 and HIV-2 genomic regions (see online [supplementary material](#) for Figure S2). The multiple sequence alignment of HIV-1 (alig01) has a PIS of 15.26% and a PPI of 85.46%. The HIV-2 alignment (alig02) has a PIS of 29.09% and a PPI of 83.20%. The researcher can use the database to access the most conserved oligonucleotides (Table 1). Our analyses revealed that oligonucleotides HIV1ID0037 (HIV score of 81.79%), HIV1ID0146 (79.43%) and HIV1ID0179 (78.91%) are the most conserved for HIV-1. The HIV-1 PCR primer pair with the highest HIV score (80.35%) was HIV1ID0037–HIV1ID0179. The most conserved oligonucleotides for HIV-2 were HIV2ID0015 (95.14%), HIV2ID0061 (87.49%) and HIV2ID0057 (84.71%). The PCR primer pair with the highest HIV score (91.31%) for HIV-2 was HIV2ID0015–HIV2ID0061. Similar melting temperatures (T_m) in oligonucleotide pairs are important for the efficiency of PCR (15). We identified some PCR primer pairs with similar T_m and high HIV scores. For example, the primers HIV1ID0037 (T_m of 55.97°C) and HIV1ID0179 (T_m of 58.72°C) have similar T_m values and a high HIV score (80.35).

Extensive genetic diversity has been observed in HIV (16), and several works have used a multiplicity of techniques to study these viruses (16–23). The HIV score allows users to select the most accurate oligonucleotides for different techniques. The oligonucleotides can also be associated with the technique where they were used by the first time. Moreover, each oligonucleotide can be easily associated with the respective region in the reference genome.

The most conserved regions for HIV-1 were the *gag* and *pol*, while LTRs had the lowest HIV scores (see online [supplementary material](#) for Figure S3). The protease and p51 RT regions (located inside *pol* region) include 11% (43 oligonucleotides) of all oligonucleotides. The PCR primers' pairs for HIV-1 with the highest HIV score in these genomic regions were HIV1ID0073–HIV1ID0206 (score of 73.61 and T_m of 58.9°C and 60.9°C) and HIV1ID0035–HIV1ID087 (a score of 73.28, T_m of 53.9°C and 57.1°C). For HIV-2, the oligonucleotides with highest conservation scores were those located in the *pol* region, and the lowest located in the *tat* region (see online [supplementary material](#) for Figure 3B). The *pol* region included 21.5% of all oligonucleotides. The primers pairs for HIV-2 with the highest HIV score located in the *pol* region were HIV2ID0015HIV2ID0061 (a score of 91.31, T_m of 49.40°C and 58.80°C) and HIV2ID0057–HIV2ID0061 (a score of 86.1, T_m of 51.65°C and 58.80°C). By combining different set of PCR primers, a complete

Table 1. Oligonucleotides with highest HIV score for HIV-1 and HIV-2

Virus	Code	Type	Original name	Target length	Sequence (5'-3')	Position in reference genome	Genomic region	HIV Score
HIV-1	HIV1ID0037	PCR primer reverse	NP174	21	CTACYGCCCTTYACCTTCC	4957-4977	<i>pol</i>	81.79
	HIV1ID0146	Degenerate PCR primer	env27F	19	CTGGYATAGTGCARCARCA	7861-7879	<i>env</i>	79.43
	HIV1ID0179	PCR primer forward	HPOL4235	23	CCCTACAATCCCCAAAGTCAAAGG	4653-4675	<i>pol</i>	78.91
	HIV1ID0163	PCR primer forward	intF	20	CCCTACAATCCCCAAAGTCA	4653-4672	<i>pol</i>	78.87
	HIV1ID0116	PCR primer forward	E180	20	GTCTGGTATAGTGCAACAGCA	7860-7879	<i>env</i>	78.38
	HIV1ID0182	PCR primer reverse	HPOL4481	21	GCTGTCCCTGTAATAAACCCCG	4899-4919	<i>pol</i>	77.9
	HIV1ID0283	siRNA antisense	si4652	21	UGACUUUGGGGAUUGUAGGGA	4652-4672	<i>pol</i>	
	HIV1ID0017	Probe	H-1 probe	23	AATGAGGAGGCTGCAGAAATGGGA	1405-1427	<i>gag</i>	77.24
	HIV1ID0087	PCR primer forward	G80	23	ATGAGAGAA CCAAGGGAAATGGA	1471-1493	<i>gag</i>	76.61
	HIV1ID0052	PCR primer reverse	gag R4	32	TTGATGGTCTCTTTAACATTTGCATGGCTGC	1375-1406	<i>gag</i>	76.18
	HIV2ID0015	PCR primer reverse	A-Loop B	21	AATTTAAAGAAGGGGAGGA	4607-4627	<i>pol</i>	95.14
	HIV2ID0061	PCR primer forward	POL OG 479 INV	23	GGAGCAGTCCCTAGTCAAGGTAGG	4796-4818	<i>pol</i>	87.49
	HIV2ID0057	PCR primer reverse	PFID INV	21	CTGCCTTCTCTGAAATAGACC	4729-4749	<i>pol</i>	84.71
HIV2ID0044	PCR primer forward	EB7	17	CCYAGGCAAGCATGGTG	7149-7165	<i>env</i>	83.47	
HIV2ID0017	PCR primer forward	Seq.ID.NO.19	26	CCTCAATTCTCTCTTTGGAAAAAGACC	2084-2109	<i>gag</i>	82.91	
HIV2ID0041	PCR primer reverse	EB6	20	CCATTRAAGCCAACCAWGT	6945-6964	<i>env</i>	81.32	
HIV2ID0060	PCR primer reverse	POL OG 479	23	CCTACCTTGACTAGGACTGCTCC	4796-4818	<i>pol</i>	80.82	
HIV2ID0043	PCR primer reverse	EB5	20	CTCCTCTGCAGTTAGTCCAC	7310-7329	<i>env</i>	80.61	
HIV2ID0003	PCR primer forward	SR64, envF	23	CTCCAGGCAAGATCACTGCTAT	7869-7891	<i>env</i>	80.44	
HIV2ID0055	PCR primer forward	X410	21	CACCTCAAATCTCTCTTTGGA	2082-2102	<i>gag</i>	80.09	

amplification of the protease and p51 RT regions can be performed for detection of drug resistant mutations.

We calculated the PIS and PPI values for all 30 nt sliding windows across the HIV-1 and HIV-2 alignments, which allow the identification of the most conserved genomic regions in these viruses (see online [supplementary material](#) for Figures S4 and S5). This information can be used to design new oligonucleotides.

Examples of Use

In order to select the oligonucleotides located in the most conserved HIV-1 genomic regions, the user can go to the ‘Search’ tab on the top menu bar, and open the ‘The best oligonucleotides’ tab. The table is automatically ordered by the ‘HIV Score’ column filter. The user can now access the oligonucleotide information by clicking in the oligonucleotide ID hyperlink, which shows the summary information for that record. The database can also be used to filter all columns using the search tool. For instance, to access the best oligonucleotide for HIV-1 located in *pol* region, the user can type ‘pol’ in the search box. The database table now shows only the records where the word pol is found. The oligonucleotide with highest ‘HIV score’ in the *pol* region is HIV1ID0037.

If the aim is to design a new oligonucleotide, the user can click on the ‘Genome variation’ tab on the top menu bar. The user can then select the alignment to display in the graph (HIV-1 or HIV-2), which describe a sliding window analysis of 30 nt windows with 1 nt of overlap of PIS and PPI values. The list of the most conserved genomic regions can be found in a table. For example, the genomic region 5524–5554 of the HIV-1 alignment has the highest PIS value (64.52%). This section of the alignment can be visualized by clicking on the 5524–5554 value in the table. The user can also visualize any window of the alignment by using the ‘Show window in alignment’ box.

Funding

Portuguese Foundation for Science and Technology (FCT), European Regional Development Fund (ERDF) [IF/01356/2012 to F.P. and SFRH/BPD/100912/2014 to J.C.] and Direção-Geral da Saúde and Coordenação Nacional para a Infecção VIH/SIDA [VIH/SAU/0011/2011]. CIIMAR—FCT and ERDF [UID/Multi/04423/2013].

Supplementary data

Supplementary data are available at *Database* Online.

Conflict of interest: None declared.

References

1. Fauci, A.S. (2001) Infectious diseases: considerations for the 21st century. *Clin. Infect. Dis.*, 32, 675–685.
2. DeCock, K.M. *et al.* (1993) Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. *JAMA*, 270, 2083–2086.
3. Marlink, R. *et al.* (1994) Reduced rate of disease development after HIV-2 infection as compared to HIV-1. *Science*, 265, 1587–1590.
4. Briant, L. (2011) HIV-1 assembly, release and maturation. *World J. AIDS*, 1, 111–130.
5. Cann, A.J. (2005) *Principles of molecular virology* 4th Edition., Academic Press, United States, pp. 1–352.
6. Smyth, R.P., Davenport, M.P., and Mak, J. (2012) The origin of genetic diversity in HIV-1. *Virus Res.*, 169, 415–429.
7. Tang, Y.W., Procop, G.W., and Persing, D.H. (1997) Molecular diagnostics of infectious diseases. *Clin. Chem.*, 43, 2021–2038.
8. Barry, E. (1990) New molecular techniques for microbial epidemiology and the diagnosis of infectious diseases. *J. Infect. Dis.*, 161, 595–602.
9. Turner, J.J. *et al.* (2006) Targeting the HIV-1 RNA leader sequence with synthetic oligonucleotides and siRNA: Chemistry and cell delivery. *Biochim. Biophys. Acta: Biomembr.*, 1758, 290–300.
10. Kuiken, C., Korber, B., and Shafer, R.W. (2003) HIV sequence databases. *AIDS Rev.*, 5, 52–61.
11. Leitner, T. *et al.* (2005) HIV-1 subtype and circulating recombinant form (CRF) reference sequences, 2005. *HIV Seq. Compend.*, 3, 41–48.
12. Abecasis, A., Vandamme, A., and Lemey, P. (2006) Sequence alignment in HIV Computational analysis. *Reviews*, 2–16.
13. SantaLucia, Jr. J., and Hicks, D. (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct.*, 33, 415–440.
14. Tafer, H. *et al.* (2011) ViennaRNA Package 2.0. *Algorithms Mol. Biol.*, 6, 26.
15. Lorenz, T.C. (2012) Polymerase chain reaction: basic protocol plus troubleshooting and optimization strategies. *J. Vis. Exp.*, e3998.
16. Smyth, R.P., and Negroni, M. (2016) A step forward understanding HIV-1 diversity. *Retrovirology*, 13, 27.
17. Khodakov, D.A. *et al.* (2008) An oligonucleotide microarray for multiplex real-time PCR identification of HIV-1, HBV, and HCV. *Biotechniques*, 44, 241–248.
18. Udaykumar. *et al.* (1994) Enhanced diagnostic efficiency of the polymerase chain reaction by co-amplification of multiple regions of HIV-1 and HIV-2. *J. Virol. Methods*, 49, 37–46.
19. Rohrman, B. A., and Richards-Kortum, R.R. (2012) A paper and plastic device for performing recombinase polymerase amplification of HIV DNA. *Lab Chip*, 12, 3082.
20. Boyle, D.S. *et al.* (2013) Rapid detection of HIV-1 proviral DNA for early infant diagnosis using recombinase polymerase amplification. *MBio*, 4, pii: e00135–13
21. Wang, J.H. *et al.* (2013) An integrated chip capable of performing sample pretreatment and nucleic acid amplification for HIV-1 detection. *Biosens. Bioelectron.*, 41, 484–491.
22. Curtis, K.A. *et al.* (2012) Isothermal amplification using a chemical heating device for point-of-care detection of HIV-1. *PLoS One*, 7, e31432
23. Curtis, K.A. *et al.* (2014) Real-time detection of HIV-2 by reverse transcription-loop-mediated isothermal amplification. *J. Clin. Microbiol.*, 52, 2674–2676.