Stanford HIV RT and Protease Sequence Database
http://hivdb.stanford.edu
Data Dictionary
June 2002

If you have questions or comments regarding this document please contact:
Soo-Yon Rhee syrhee@stanford.edu
Matthew Gonzales mgonzales@stanford.edu

Core tables

Table name	Table definition	Field name	Field definition
tblPatients	Individuals from whom virus isolates were obtained	PtID	Automatically generated integer ID
		PseudoName	Name assigned to individual in GenBank or published reference. Name may also be created by HIVRT&PrDB staff
		Region	Country or region of individual at the time virus was isolated
		DateEntered	Date when patient entered
tblIsolates	Virus isolate obtained from an individual ('Type' = 'Clinical') or used in a laboratory experiment ('Type' = 'Lab')	IsolateID	Automatically generated integer ID
		PtID	Link to tblPatients
		IsolateDate	Date of virus isolation
		DateMatch	'<','<='or'='; represents uncertainty in the date
		IsolateName	Name assigned to isolate in published reference or GenBank. Name may also be created by HIVRT&PrDB staff
		Gene	RT or PR (protease)
		Type	'Clinical' or 'Lab'
		DateEntered	Date when isolate entered
tblClinIsolates	Contains data on isolates obtained directly from a person	IsolateID	Link to tblIsolates
		Source	Body source from which the isolate was obtained
		Culture	Whether the isolate was cultured before sequencing
		SeqTemplate	Source of nucleic acid used for sequencing. Usually PCR-amplified cDNA or proviral DNA
		CloneMethod	Method used for cloning or 'Direct PCR' for uncloned isolates
		SeqMethod	'Dideoxy' vs 'DNAChip' vs 'Unknown'
tblLabIsolates	Contains data on isolates modified in the laboratory	IsolateID	Link to tblIsolates
		Parent	Name of Laboratory isolate
		MutationList	List of mutations in the laboratory isolate
		SDM	Whether site-directed mutagenesis was done: yes vs. no
		Passage	Drug used in in vitro passage experiment
tblSequences	Contains sequences and	SequenceID	Automatically generated integer ID

	data needed for rapid sequence alignment	AccessionID	GenBank accession number
		IsolateID	Link to tblIsolates
		SeqType	'Sequence' vs 'Consensus' of mulitple clones vs 'IncompleteSequence'
		CloneName	Name or number of individual clones according to published reference or GenBank
		Firstaa	Amino acid position at which the sequence begins
		Lastaa	Amino acid position at which the sequence ends
		NASeq	Raw nucleotide sequence data
		AASeq	Amino acid sequence data - either submitted primarily or derived from the nucleotide sequence
tblInsertions	Contains nucleic acid and amino acid insertion data	SequenceID	Link to tblSequences
		CodonPos	Position of the insertion
		AA	Amino acid(s)
		NA	Nucleotides
tblRxHistory	Contains anti-HIV drug treatment histories of persons whose HIV-1 isolates have been sequenced	RxHistoryID	Automatically generated integer ID
		PtID	Link to tblPatients
		StartDate	Date at which regimen was begun
		StopDate	Date at which regimen was stopped. When StartDate and StopDate are not available, StartDate and StopDate are estimated on Year of publication
		DateMatch	'<','<='or'='; represents uncertainty in the StopDate
		Weeks	Number of weeks that regimen was received
		DurationUnknown	'Yes' or 'No'; represents uncertainty in the Weeks
		RegimenName	A list of the drugs received as part of the regimen automatically generated from tblDrugRegimens
tblDrugRegimens	Contains data on each drug treatment regimen	RxHistoryID	Link to tblRxHistory
		DrugName	Antiretroviral drug
		Dose	Drug dose
		Freq	Dose frequency (number of times per day)
tblDrugs	Contains data on anti-HIV drugs	DrugName	Drug abbreviation
		DrugClass	NRTI vs NNRTI vs PI vs RTI vs Unknown
		Approved	Yes vs No
tblRNA	Contains number of RNA copies obtained from a patient.	RNAID	Automatically generated integer ID
		PtID	Link to tblPatients
		RNADate	Date of Virus Load taken
		DateMatch	'<','<='or'='; represents uncertainty in the RNADate
		VLoad	Number of RNA copies per ml of plasma in log10
tblSuscResults	Contains drug susceptibility results	SuscID	Automatically generated integer ID
		IsolateID	Link to tblIsolates
		RefID	Link to tblReferences
		Method	Name of the susceptibility test method
		DrugName	Drug used for susceptibility testing
		ResultMatch	'=' vs '>' vs '<'

		Result	Drug concentration required for inhibition in nM
		IC	% inhibition 50 vs 90 vs 95
		FoldMatch	'=' vs '>' vs '<'
		Fold	Fold-resistant compared to wildtype
tblSuscMethod	Contains reference and description of method for susceptibility testing	MethodName	Descriptive name for the method of susceptibility testing
		RefID	Link to tblReferences. The reference with the most descriptive data on the susceptibility test method
		Notes	Free-text description and notes about the susceptibility test method
tblIsolateFilters	Lists isolates which should be filtered from certain queries	IsolateID	Link to tblIsolates
		Filter	Reason for why this isolate should be excluded from certain queries. The complete description of each filter is found in tblLUIsolateFilters
tblSuscFilters	Lists susceptibility results which should be filtered from certain queries	SuscID	Link to tblSuscResults
		Filter	Reason for why the susceptibility results on an isolate should be excluded from certain queries. The complete description of each filter is found in tblLUSuscFilters
tblSpecies	Contains the species of the isolate	IsolateID	Link to tblIsolates
		Species	HIV1, HIV2, AGM
tblSubtypes	Contains the subtype of the isolate	IsolateID	Link to tblIsolates
		Subtype	Subtype classification of the RT or protease gene as determined by comparison to reference sequence. The subtype also includes group information (e.g. O and N) even though these are not subtypes.
tblReferences	Contains basic information about the references in which the isolate was described	RefID	Automatically generated integer ID
		Author	First author
		Title	Title of article
		Journal	Name of Journal
		RefYear	Year of Publication
		MedlineID	Unique PubMed ID. This allows CGI scripts to link to the online MEDLINE abstract
		Published	Yes vs no. Yes means that data from this reference can be obtained via web queries. No means that the data in the reference are temporarily embaroged.
tblRefLink	Links tblIsolates and tblReferences	RefID	Link to tblReferences
		IsolateID	Link to tblIsolates
		Priority	One isolate can be linked to multiple References and the Reference with Priority = 1 will be shown on the web
tblRefNoMedline	Contains data on references that are not cited by MEDLINE	RefID	Automatically generated integer ID
		Description	Detailed description of data set for references that are no cited by MEDLINE. These include references not yet cited by GenBank, data presented at scientific meetings, and submissions made directly to GenBank or HIVRT&PrDB

Look-up tables

Table name	Table definition	Field name	Field definition
tblLUConsensusPR	Contains the consensus PR sequence. One row per position	Codon	Codon position (1...99)
		AA	Consensus B reference sequence amino acid at the codon
tblLUConsensusRT	Contains the consensus RT sequence. One row per position	Codon	Codon position (1...560)
		AA	Consensus B reference sequence amino acid at the codon
tblLUCountries	Contains a list of acceptable country or region names	Region	Name of country or region
tblLUBodySource	Contains a list of body sources	Source	Body source of the sequenced isolate
tblLUCloningMethod	Contains a list and description of cloning methods	Method	Brief name of method used for cloning. 'None' indicates direct PCR sequencing
		Description	Detailed description of the cloning method
tblLUSeqTemplate	Contains a list and description of methods for preparing sequence template	Method	Brief name of DNA (or cDNA) used for sequencing. 'PCR-amplified DNA' is the most common entry.
		Description	Detailed description of the DNA or cDNA used for sequencing
tblLUIsolateFilter	Contains a list of reasons for filtering isolates from certain queries	Filter	Brief name for the filter
		Description	Detailed description of the filter. Explanation of why isolates linked to this filter may need to be excluded from certain queries
tblLUSuscFilter	Contains a list of reasons for filtering certain susceptibility results from certain queries	Filter	Brief name for the filter
		Description	Detailed description of the filter. Explanation of why the susceptibility results of isolates linked to this filter may need to be excluded from certain queries
tblLUMutationTypes PR	Assigns a mutation type to each residue in PR. Currently there are just 2 classification systems. Others may be added.	Position	Amino acid position or codon
		Class5way	5-way classification system. Each residue is assigned a mutation type ('Major', 'Minor', 'Common', 'Rare', 'Conserved')
		Class3way	3-way classification system. Each residue is assigned 'Major', 'Minor', 'Other'
tblLUMutationTypes RT	Assigns a mutation type to each residue in RT. Currently there are just 2 classification systems. Others may be added.	Position	Amino acid position or codon
		Class5way	5 way classification system. Each residue is assigned a mutation type ('NRTI', 'NNRTI', 'Common', 'Rare', 'Conserved')
		Class3way	3-way classification system. Each residue is assigned 'NRTI', 'NNRTI', 'Other'