



Library
Indian Institute of Science Education and Research
Mohali



DSpace@IISERMohali (/jspui/)

/ Thesis & Dissertation (/jspui/handle/123456789/1)

/ Master of Science (/jspui/handle/123456789/2)

/ MS-09 (/jspui/handle/123456789/393)

Please use this identifier to cite or link to this item: <http://hdl.handle.net/123456789/402>

Title: Analysis of Circulating Recombinant Forms (CRFs) of HIV-1 using Chaos Game Representation (CGR)

Authors: Bansiwala, Adhikar (/jspui/browse?type=author&value=Bansiwala%2C+Adhikar)

Keywords: Human Immunodeficiency Virus (HIV)
Acquired Immune Deficiency Syndrome (AIDS)
Chaos Game Representation
Circulating Recombinant Forms

Issue Date: 22-Jul-2014

Publisher: IISER M

Abstract: Human Immunodeficiency Virus (HIV) is the causative agent for Acquired Immune Deficiency Syndrome (AIDS). It exhibits very high genetic diversity with different variants and subtypes. Classification of these subtypes is thus essential for monitoring epidemic. Current methods of classification include specific genes-based phylogenetic analysis, but these methods showed certain inconsistencies in classification of subtypes in past. However, recent alignment free methods, like Chaos Game Representation (CGR), have been shown to be successful in classification of HIV subtypes at word length $k=6$ (Sinha, Pandit ; 2010). This method is not only computationally less intensive, but can also analyze whole genome variations. Problem with HIV classification becomes more complex as different HIV subtypes can recombine and form Circulating Recombinant Forms (CRFs). These CRFs continuously emerge over time and circulate into host population. They show variable susceptibility to drugs. In my 5 year MS project, my aim was to test if these CRFs could also be classified using the CGR method. Being recombinants of subtypes, the variation in the sequences are expected to be quite low. My studies are presented in this thesis in the following sections. Chapter -1 of this thesis is an introduction to HIV subtypes and CRFs. It also introduces basics of CGR plotting and classification using CGR method. Chapter-2 gives an overview of various software tools, algorithms and other computational methods used in this work. In Chapter-3, the results are shown for classification of the CRFs using the CGR method. I checked the effect of lowering word-length and it is shown that again $k=6$ is the minimum word-length required for correct classification. In cladograms generated it was reported that CRFs clustered with those parental subtypes that have the largest length in the genome. Chapter-4 deals with reduction in word-set, and it was seen that correct clustering can still be obtained even by selecting lesser number of words. Base composition analysis of these selected words was performed and it was reported that these words were mostly A-rich. Chapter-5 shows the use of certain HIV genes, instead of whole genome, to classify CRFs properly using CGR method. It shows the drawback of this method in analyzing short genomic sequences. Lastly, Chapter-6 discusses a simple software tool created in PHP and HTML to generate CGR and to calculate base composition of the given input sequence.


URI: <http://hdl.handle.net/123456789/402> (<http://hdl.handle.net/123456789/402>)

Appears in MS-09 (</jspui/handle/123456789/393>)
Collections:

Files in This Item:

File	Description	Size	Format	
MS-09007.pdf (/jspui/bitstream/123456789/402/3/MS-09007.pdf)		2.69 MB	Adobe PDF	View/Open (/jspui/bitstream/123456789/402/3/MS-09007.pdf)

[Show full item record \(/jspui/handle/123456789/402?mode=full\)](/jspui/handle/123456789/402?mode=full)

 </jspui/handle/123456789/402/statistics>

Items in DSpace are protected by copyright, with all rights reserved, unless otherwise indicated.