

Establishing the Taxa and Functional profile of Microbiota implicated in West Nile Fever

Jemy Ratna Jovita^{1,2*}, Shylesh Murthy IA² and Preenon Bagchi³

¹Padmashree Institute of Management and Sciences, Bengaluru, India. *E-mail: jemy2798@gmail.com ²Vasishth Academy of Advanced Studies and Research, Bengaluru, India.

³Institute of Biosciences and Technology, MGM University, Aurangabad, India

Abstract: West Nile fever causing microbiome is taken in this work. *Culex nigripalpus* mosquito is the causative factor for West Nile Virus. Using Metatranscriptomic sequencing, identified the taxa and functional profile of the microbiome is identified.

Keywords: West Nile fever, microbiome, Metatranscriptomics, taxonomy, functional profile, gene receptors, docking.

1. INTRODUCTION

West Nile fever is an infection caused by the West Nile virus, which is spread by mosquitoe, *Culex nigripalpus. Culex nigripalpus* is a primary vector for causing West Nile fever; they proliferate in the nutrient rich media and colonize in freshwater aquatic habitats. The microbiota of mosquitoes provides nutrition for the development and transmission of the pathogen [1, 2]. The virus replicates affecting the brain causing neuro-virulence, it turns to be neuroinvasive when it gains access to the central nervous system. The viral serine protease, NS2B-NS3 plays a crucial role in viral replication. The distribution of the virus is seen throughout Africa, the Middle East, southern Europe, western Russia, southwestern Asia, and Australia [3].

The Src family kinase c-Yes was recently reported to be important for maturation of West Nile virus particles. Individuals with West nile fever recover within days to months In acute flaccid paralysis individuals develop symptoms headache, fever, malaise, gastrointestinal upset, skin rash, and some patients have neck rigidity and changes in mental status [4].

2. MATERIALS AND METHODS

Metatranscriptomic analysis

Next-generation sequencing (NGS) is an advanced version of non-Sanger-based sequencing technology that offers ultra-high throughput, scalability, and speed. Galaxy is an open source, web-based platform for next generation computational biomedical research [5]. Metatranscriptomics analysis enables understanding of how the microbiome responds to the environment by studying the functional analysis of genes expressed by the microbiome [6, 7].

Structure based drug designing technique is used here to build, display, simulate and analyze the molecular structure. Here we have used SWISS-MODEL tool [8] for modelling the proteins (gene receptors) responsible for West Nile fever CCR5, CLEC4M, IFITM2, IRF3. Selected models from homology modelling [9] output are docked with selected phytocompound from medicinal herbs.

Selection of phytocompounds was done using Lipinski rule for drug which is based on the ADME properties. Molecular docking [10] was done using Patchdock tool and best interacting phytocompounds with the gene receptors can be selected as ligands.

West Nile viruses' fastq sequences SRR10017187.1.1 and SRR10017187.1.2 were retrieved from SRA database.

Sequences' quality was checked using FASTQC [11]. MultiQC [12] was done to aggregate results from FASTQC analyses into a single report.

Sequences were trimmed using cutadapt.

FASTQC followed by MultiQC was re-run using the results of cutadapt.

Next, using SortMeRNA tool [13-17] any reads identified as rRNA in dataset was removed. Next, using FASTQ INTERLACE tool [18] paired end FASTQ reads from two separate files were joined.

© The Author(s) 2023

R. Somashekhar et al. (eds.), *Proceedings of the International Conference on Advances in Nano-Neuro-Bio-Quantum (ICAN 2023)*, Advances in Intelligent Systems Research 69, https://doi.org/10.2991/978-94-6463-294-1_16 MetaPhlAn tool [19] was used for profiling the composition of microbial communities (Bacteria, Archaea and Eukaryotes) from our microbiota.

Krona tool [20, 21] was used to visualize the results of a metagenomic profiling as a zoomable pie chart and GraPhlAn tool [22] for visualizing high-quality circular representations of taxonomic and phylogenetic trees.

Further, HUMAnN [23] pipeline was used for efficiently and accurately profiling the presence/absence and abundance of microbial pathways in our microbiota.

3. RESULTS AND DISCUSSION

Metagenome, having accession number SRR10017187, for West Nile virus was downloaded from SRA database.

As, per Per base sequence quality results of FASTOC and MultiOC, the sequence quality is not good hence we go ahead with trimming the sequence.

CUTADAPT tool [24] is used for trimming. It finds and removes adapter sequences, primers, poly-A tails, and other types of unwanted sequence from our data. It searches for the adapter in all reads and removes it when it finds it. Further, sequence quality of the cutadapt output is checked using FASTOC and MultiOC and it is found within the range.

SortMeRNA tool removes any reads identified as rRNA from our dataset. Fastq Interlace tool joins paired end FASTQ reads from two separate files. Taxonomic profiling [25] was done using MetaPhlAn tool.

The output is visualized using Krona and Graphlan.



Fig. 1: Generation, personalization and annotation of tree: Tree in PhyloXML

Viruses unclassified

kingdo m	kingdom_i d	phylum	class	order	family	family_i d	genus	genus_i d	species	species_i d	strain s	strains_i d
Viruses	10239					100.0						
Viruses	10239	Viruses unclassified					100.0					
Viruses	10239	Viruses unclassified	Viruses unclassified					100.0				
Viruses	10239	Viruses unclassified	Viruses unclassified	Viruses unclassified					100.0			
Viruses	10239	Viruses unclassified	Viruses unclassified	Viruses unclassified	Flavivirida e	11050				100.0		
Viruses	10239	Viruses	Viruses	Viruses	Flavivirida	11050	Flaviviru	11051			100.0	

Table 1: MetaPhlAn: Predicted taxon relative abundances at each taxonomic levels

After generation of taxonomy, we move to functional information of our microbiome. Functional information of the above microbiome community [28] was done using HUMAnN pipeline.

11050

11051

100.0

11082

# Gene Family	humann_Abundance-RELAB
UNMAPPED	0.888769
UniRef90_D9MXB1	0.0073362
UniRef90_D9MXB1 unclassified	0.0073362
UniRef90_UPI0001DD3770	0.00495637
UniRef90_UPI0001DD3770 unclassified	0.00495637
UniRef90_A0A1B1WW31	0.00484378
UniRef90_A0A1B1WW31 unclassified	0.00484378
UniRef90_B9DF73	0.00459488
UniRef90_B9DF73 unclassified	0.00459488
UniRef90_A0A067ZQ74	0.00422115
UniRef90_A0A067ZQ74 unclassified	0.00422115
UniRef90_A0A077EZ14	0.00417222
UniRef90_A0A077EZ14 unclassified	0.00417222
UniRef90_A0A067ZS80	0.00390475

Viruses unclassified

Viruse

10239

Viruses unclassified

UniRef90_A0A067ZS80 unclassified	0.00390475
UniRef90_UPI0003F054D7	0.00376104
UniRef90_UPI0003F054D7 unclassified	0.00376104
UniRef90_F1BA45	0.0036982
UniRef90_F1BA45 unclassified	0.0036982
UniRef90_A0A1X3CH90	0.00318336
UniRef90_A0A1X3CH90 unclassified	0.00318336
UniRef90_UPI000D0C92C0	0.00302072
UniRef90_UPI000D0C92C0 unclassified	0.00302072
UniRef90_K9L2G3	0.00265002
UniRef90_K9L2G3 unclassified	0.00265002
UniRef90_A0A1B1WWN4	0.00258533
UniRef90_A0A1B1WWN4 unclassified	0.00258533
UniRef90_K7ET80	0.00256502
UniRef90_K7ET80 unclassified	0.00256502

UniRef90_Q91CD9	0.00251161
UniRef90_Q91CD9 unclassified	0.00251161
UniRef90_UPI00018E17CF	0.00240202
UniRef90_UPI00018E17CF unclassified	0.00240202
UniRef90_G7PXF9	0.00234066
UniRef90 G7PXF9 unclassified	0.00234066
UniRef90 F6Z663	0.00218081
UniRef90 F6Z663 Junclassified	0.00218081
UniRef90 G7MFM1	0.00215974
UniRef90 G7MFM1lunclassified	0.00215974
UniRef90 A0A2K6NB76	0.0020891
UniRef90 A0A2K6NB76 unclassified	0.0020891
UniRef90 UPI000181CCFE	0.00203554
UniRef90 UPI000181CCFE unclassified	0.00203554
UniRef90_C9WPK8	0.00195453
UniRef90_C9WPK8 unclassified	0.00195453
UniRef90_P14335	0.00180304
UniRef90_P14335 unclassified	0.00180304
UniRef90 F7HJ82	0.00180269
UniRef90 F7HJ82 unclassified	0.00180269
UniRef90 A0A0D3MDX0	0.00160595
UniRef90 A0A0D3MDX0lunclassified	0.00160595
UniRef90 UPI0000E69E2F	0.00153697
UniRef90 UPI0000E69E2Flunclassified	0.00153697
UniRef90 I3VPR6	0.00149368
UniRef90 I3VPR6/unclassified	0.00149368
UniRef90 G8E0A9	0.00146336
UniRef90 G8E0A9lunclassified	0.00146336
UniRef90 D3X888	0.0013421
UniRef90 D3X888 unclassified	0.0013421
UniRef90 UPI00003BECBC	0.00130469
UniRef90 UPI00003BECBClunclassified	0.00130469
UniRef90 UPI0002C39258	0.00127138
UniRef90 UPI0002C39258 unclassified	0.00127138
UniRef90 Q6VXX6	0.00120319
UniRef90 Q6VXX6 unclassified	0.00120319
UniRef90 D3X8F1	0.0011535
UniRef90 D3X8F1 unclassified	0.0011535
UniRef90 A0A1S6WNB4	0.00110115
UniRef90 A0A1S6WNB4 unclassified	0.00110115
UniRef90 D3X8F7	0.00102768
UniRef90 D3X8F7 unclassified	0.00102768
UniRef90 D3X8D8	0.000998791
UniRef90 D3X8D8 unclassified	0.000998791
UniRef90 Q91KP3	0.000989086
UniRef90 091KP3/unclassified	0.000989086
UniRef90 D9IFF5	0.000952321
UniRef90 D9IFF5 unclassified	0.000952321
UniRef90_F7EY45	0.000912381
UniRef90_F7EY45 unclassified	0.000912381
UniRef90_D5K1B9	0.00090469
UniRef90 D5K1B9 unclassified	0.00090469
UniRef90_Q67428	0.000861045
UniRef90_Q67428 unclassified	0.000861045
UniRef90_Q91AB0	0.000854561
UniRef90_Q91AB0 unclassified	0.000854561
UniRef90_D3X875	0.000850089
UniRef90_D3X875 unclassified	0.000850089

UniRef90_A0A2K5WQL4	0.000795493
UniRef90_A0A2K5WQL4 unclassified	0.000795493
UniRef90_B1P6F1	0.000689513
UniRef90_B1P6F1 unclassified	0.000689513
UniRef90_P29984	0.000618302
UniRef90_P29984 unclassified	0.000618302
UniRef90_A0A0B5H4C4	0.000582165
UniRef90_A0A0B5H4C4 unclassified	0.000582165
UniRef90_Q6YFU6	0.00051542
UniRef90_Q6YFU6 unclassified	0.00051542
UniRef90_A0A0P0RST2	0.000505187
UniRef90_A0A0P0RST2 unclassified	0.000505187
UniRef90_A0A0D5CWD5	0.000489091
UniRef90_A0A0D5CWD5 unclassified	0.000489091
UniRef90_A0A068AXW2	0.000463204
UniRef90_A0A068AXW2 unclassified	0.000463204
UniRef90_A0A1X3CWX3	0.000451315
UniRef90_A0A1X3CWX3 unclassified	0.000451315
UniRef90_I1TEC3	0.000443267
UniRef90_I1TEC3 unclassified	0.000443267
UniRef90_A0A229QXW6	0.000390111
UniRef90_A0A229QXW6 unclassified	0.000390111
UniRef90_UP10000485C0B	0.000373856
UniRef90_UPI0000485C0B unclassified	0.000373856
UniRef90_Q6VXX7	0.000367549
UniRef90_Q6VXX7 unclassified	0.000367549
UniRef90_H2NBX3	0.000356109
UniRef90_H2NBX3 unclassified	0.000356109
UniRef90_UPI000642CDD8	0.000349761
UniRef90_UPI000642CDD8 unclassified	0.000349761
UniRef90_I3N2P9	0.000333919
UniRef90_I3N2P9 unclassified	0.000333919
UniRef90_D9IFG5	0.000299914
UniRef90_D9IFG5 unclassified	0.000299914
UniRef90_D3X874	0.000283626
UniRef90_D3X874 unclassified	0.000283626
UniRef90_Q9DP92	0.000282183
UniRef90_Q9DP92 unclassified	0.000282183
UniRef90_D3X8B2	0.000279068
UniRef90_D3X8B2 unclassified	0.000279068
UniRef90_A0A075FC35	0.000211602
UniRef90_A0A075FC35 unclassified	0.000211602
UniRef90_B7U2U9	0.000200363
UniRef90_B7U2U9 unclassified	0.000200363
UniRef90_C9WPM0	0.000194638
UniRef90_C9WPM0 unclassified	0.000194638
UniRef90_A0A0P7UTC2	0.000191691
UniRef90_A0A0P7UTC2 unclassified	0.000191691
UniRef90_F7FBV8	0.000191691
UniRef90_F7FBV8 unclassified	0.000191691
UniRef90_G3SQL6	0.000191691
UniRef90_G3SQL6 unclassified	0.000191691
UniRef90_H3BYH5	0.000191691
UniRef90_H3BYH5 unclassified	0.000191691
UniRef90_P60866	0.000191691
UniRef90_P60866 unclassified	0.000191691
UniRef90_D3X8C7	0.000186455
UniRef90_D3X8C7 unclassified	0.000186455

UniRef90_A0A2K6EGR5	0.000170037
UniRef90_A0A2K6EGR5 unclassified	0.000170037
UniRef90_A0A0B7JKQ7	0.000166418
UniRef90_A0A0B7JKQ7 unclassified	0.000166418
UniRef90_UPI0005215E94	0.000165638
UniRef90_UPI0005215E94 unclassified	0.000165638
UniRef90_G3I2D3	0.00016095
UniRef90_G3I2D3 unclassified	0.00016095
UniRef90_A0A2I0MVY6	0.000152329
UniRef90_A0A2I0MVY6 unclassified	0.000152329
UniRef90_L9LBB7	0.000143369
UniRef90_L9LBB7 unclassified	0.000143369
UniRef90_UPI0006B0C551	0.000133289
UniRef90_UPI0006B0C551 unclassified	0.000133289
UniRef90_A0A091DCD6	0.000130236
UniRef90_A0A091DCD6 unclassified	0.000130236

UniRef90_UPI000D3231B4	0.00012363
UniRef90_UPI000D3231B4 unclassified	0.00012363
UniRef90_UP1000359548B	0.000119308
UniRef90_UPI000359548B unclassified	0.000119308
UniRef90_UPI00051F13F9	0.000116856
UniRef90_UPI00051F13F9 unclassified	0.000116856
UniRef90_UPI0005216B30	0.000112244
UniRef90_UPI0005216B30 unclassified	0.000112244
UniRef90_K7P5A3	0.000108977
UniRef90_K7P5A3 unclassified	0.000108977
UniRef90_L5JTI1	8.00319e-05
UniRef90_L5JTI1 unclassified	8.00319e-05
UniRef90_Q6VYH3	7.08503e-05
UniRef90_Q6VYH3 unclassified	7.08503e-05
UniRef90_A0A2U3XJL7	5.7032e-05
UniRef90_A0A2U3XJL7 unclassified	5.7032e-05

Next, from the gene family information, we obtain the functional information of our microbiome using Superfamily server. The Functional information of 1st five families from Normalized gene families as detected by Superfamily (HMM library and genome assignments server) is given below.

Strong hits			Protein sets	ant a	
Second and a second and	DOM(XB)		Townson .		
Dennale Reminer 1	Benlow 206-430		3(804)	7921	
Classification Law	el Chamilication	E available	fargument in	12.	
Swerfamily	P. loss (potaloino ourteoside trioteophote liedrolage	3.454-55	Cartowne 12	CHARGE (CHARGE WAY) Restriction protects 1 (812) 03001/10986, 80404309-17 0A-11082 OG-Rest Nav	
Family	BMA beliense	0.00000358		ner (MW). GR- (XFlavorider, Revivus, Japanese enzykaltö etter grud.	
Forther Dotallie	Farth Bergin (Apparents) Encourse Apparents) Burger Cart	(Contraction)	Depieter .	Construction Control Contro	
ruriner Details:	Contraction of the owner owne			1.4.4. Starting and the first starting and the starting and the starting startin starting starting	
Bernelle Breneller B	Andrew 1 Mar			P 1 THE AT A VALUE AND AND A VALUE AND A VALUE AND	
Domain Number 2	Megion: 3-284	In looking		AND STRUCTURE AND ALCONOMIC PROPERTY ONLY	
Classification Leve	et Classification	E value		Sub-prime decisis	
Superiamity	e-toot presente nacienale provisionale nyorcase	0.0000000431	manyant (Def	
Fairsty	Print Delight Internet Course Supervised Press Con	0.0000000000021	Meeto.st 2	endi .	
Further Details:	Contraction (Contraction of Contraction		MAPLE	756, 994	
UniRef90_D	9MXB1 Abundance-0.0073362				
Strong hits			Prote	in sequence	
Sequence:	AGAIBLIWWII		Segue		
Domain Number 1	Region: 291-590		Seque		
Classification Lev	el Classification	E-value	nce	2584	
Superfamily	Viral phycocratein, central and dimeritation domain	1.26e-126	1		
Family	Viral physicistein, central and dimensation domain	0.000000634	length		
Further Details:	Banda Genetal (Algoreant) Common Assignments) (Danase Co	adjust processi	C		
122210000000000000000000000000000000000			Comm	(tr A0A1B1WW31 A0A1B1WW31 WNV)	
Channalm Humber 1	Realise: 1000-2134		ent	Genome nolyprotein	
Classification Leve	el Classification	E-scalue	om	oenome polyprotem	
Superfamily	E have restance northeastly triplesestate buried as	er 1.05e-54		{ECO:0000256 SAAS:SAAS00368684}	
Family	PAA helicate	0.00000358		OV-11082 OS-West Nile virus (WNW)	
Erest or Described	Rent Death) (descent) Dense Assessed) Dense ()	and a state of the		OA-11062 $OS-west fille vilus (winv).$	
ruriner betails:				GN=MZ11 60591gpGP3 OC=Flaviviridae;	
				Flouivinus, Iononoso, onconhalitis, vinus, group	
Domain Number	Region: 382-000			Flavivirus, Japanese encephantis virus group.	
Classification Lev	el Classification	E-value		OH=7158,8782,9606,34610,34627,34630,5352	
Supertainaly	E. set. domains	6.65e-64		7 200725 200727	
ramily	Clean II, you habiin proteim C-terminal domain	0.00000794		/,508/55,508/5/	
Further Details:	Zant Data) (Agreen) Course Automatic) (Desired)	PO(RATIO)			
Domain Number (Region: 1687-1970				
Classification Lev	el Classification	E-value			
Superfamily	P-loss containing successide triphosphate hydrolet	no 1.93e-60			
Family	RNA beligase	9.00000000117			
Forther Details	Farets Detail (Approxim) Events Acaptron (Datas De	formante.			
Domain Number	Region: 1510-1676				
Classification Lev	el Classification	E-value			
Superfamily	Trypsin-like serine protestes	0.27e-16			
Family	Viral protesses	0.000000259			
Further Betails	Fandy Deute Adapterent) General Acapterent Distant Co	enter automation and and and and and and and and and an			
Consider Pressure					
Domain Number (Region: 22-99				
Classification Lev	el Classification	E-value			
Superfamily	Environa causid trataio.C	1.44+-31			
Family	Findvitus consid anatoin C	0.0000192			
Further Details:	Faith Deale Separate Grane Anigeretti Datar (2	COLUMN PR			
UniRef90 A	0A1B1WW31 Abundance-0.00	484378			

Strong hits.	Protein sequences
Sequence: BD/72	Laborat (Mali) (and -)
Domain Number 1 Region: 670-996.1024-1247	Intel Line Control of Control (Control Control
Classification Level Classification E-value	VX - Facebook Paralities Indexes and adds also graph
Family RNA dependent RNA information 0.054	The complete control with the definition of the control of the con
further Details: East line) ingreen) Easter togeners) Barren tortenaner)	Contract Contract and Contract and Contract C
Domain Number 2 Region: 370-621	Marrier de provincia de la construcción de la construcción de la construcción de la construcción de la constru construcción de la construcción de
Classification E-value	- And Control and Control of Control and Control an
Superfamily Stationary Line Teams december and an and a set 1 1 le 12	
(and the first of the second s	Transland Intel
Forther Decares. Succession Subsection Statements Statements	ACTION INTERNAL AND A
UniRef00 B0DE73 Abundance-0 00459488	
Stream hits	- · ·
	Protein sequence
Neglence: AUADD/2024 Demain Number 1 Region: 2015-3157 1185-5408	External
Classification Level Classification E-value	$\frac{A0A067ZQ74}{1000000000000000000000000000000000000$
Superfamily DIAA-RIAA polymeraped 5.85e-135	link(s)
Zandy RNA-dependent SNA-information (0.004	Sequenc
Further Details: Contracting Contracting Contracting	a langth 3429
Domain Number 2 Region: 201-500	e length
Classification Level Classification E-value	Comme (tr A0A067ZO74 A0A067ZO74 WNV) Genome
Superfamily Vital disconstein, central and dimension domains 1.82e-126	$\begin{bmatrix} a_1 & a_1 & a_2 & a_3 & a_4 & a_1 & a_1 & a_2 & a_2 & a_2 & a_1 & a_1 & a_2 & a_2 & a_1 & a_2 & a_2 & a_1 & a_1 & a_1 & a_2 & a_1 & a_$
Family Vital decocotain, central and dimensation domains 0.0000000834	nt polyprotein {ECO:0000256 SAAS:SAAS00368684}
Further Details: Canada C	KW=Complete proteome OX=11082 OS=West Nile
Domain Number 1 Region: 1990-2124	virus (WNV), GN= OC=Flaviviridae: Flavivirus
Classification Level Classification E-value	Iononogo onconhalitia viena
Superfemily Prices containing inclosed at Industrials Industrial 101e-53	Japanese encephantis virus group.
Family Risk here and Dates Transmit Transmit Transmit	OH=7158,8782,9606,34610,34627,34630,53527,308
Further Decards:	735,308737
Domain Number # Region: 582-091	
Classification Level Classification E-value	
Supertamily East domesis 6.44e-44	
Control of the Contro	
Domain Humber 3 Region: 1687-1978	
Classification Level Classification I value	
Superfamily P. Non-continuum technicità 1 inforcitata tedicianes (8.14-45	
Further Details: Fact bank Research Groun Augurent Down Contention)	
Domain Number 1 Region: 1530-1676	
Classification Level Classification L value	
Superformity Trytolin Site sectors, scriptopes 1,44e-33	
Farther Details Park (Mar) (Agreen) (Second Assessed) [Sour (Second and	
Domain Number 7 Region: 22-99	
Classification E-value	
Factorianty Easterna cannot existen C 0.0002102	
Further Details: Exclusion) Surrent Course Augeners) Encon Connectors)	
Domain Number 8 Region: 2535-2782	
Clessification Level Classification Levelue	
Paraly Indiana Control and Con	
Further Details: East Intel Stateman Covers Surgearth Docus Coversion	
$I_{\text{Ini}} P_{\text{of}} 00 = A0A0677074 A hypdomas 0.00422115$	
Unixe190_AUAU0/ZQ/4 Abundance-0.00422115	Brothale appropriate
Strong inte	Present actions to
Sequence: AGA077E214	Tatamai Audol/10214
Classification Level Classification E-value	Separate 11
Superfamily 2-lose contamos michosida triphosidate hydrolases 2-024-55	Contrast 31/4/4/1782/4/4/8/78214.000/ Genera (physican 19/2 000/2015445 SAASID MARAD DX-1100 DX-Ben
Family RMA.helicase 0.00000385	Ne vita (900), OK+ 02-Planitilar, Planitar, Approve exceptedat mut prop. OH= 108 8783 9006 38023 2627 2629 10517 388771 388732
Further Details: (methodal (Marshell (Marshell)) (Marshell (Marshell))	Sejana automaticanona contrato trata
Bomain Number 3 Region: 183-471	ESERCE AND A RECEIPTION OF A R
Classification Level Classification E-value	2017 Lander Australia Concerning (Concerning and Concerning) (Concerning and Concerning and Conc
Superfamily P. loup containing nucleoside trubuschate hydrolanes 3.47e-41	A CONTRACT OF A
Tamity RNA helicase 0.000000317	ALTER PROVIDENT AND A DESCRIPTION OF A D
Forther Details: Control Contr	Downed (Las)
Domain Number 1 (begion: 5-171	New York Commence
Classification Level Classification E-value	HOMEN CONTRACTOR
Superfamily Trypsin like service proteines 8 09e-37	
Formary What professions 10.000000239	
Further Details:	
UniRet90_A0A077EZ14 Abundance-0.00417222	

Fig. 2: Functional information of 1st five gene families as detected by Superfamily server [26]

CONCLUSION

The taxonomy information of West Nile fever microbiome are identified. The functional

information of the microbiome was identified using Superfamily database. This family information will be useful for further drug designing studies.

REFERENCE

- 1. Duguma, D., Hall, M. W., Smartt, C. T., Debboun, M., & Neufeld, J. D. (2019). Microbiota variations in *Culex nigripalpus* disease vector mosquito of West Nile virus and Saint Louis Encephalitis from different geographic origins. PeerJ, 6, e6168.
- Ezgimen, M., Lai, H., Mueller, N. H., Lee, K., Cuny, G., Ostrov, D. A., & Padmanabhan, R. (2012). Characterization of the 8-hydroxyquinoline scaffold for inhibitors of West Nile virus serine protease. *Antiviral research*, 94(1), 18-24.
- 3. Petersen, L. R., Brault, A. C., &Nasci, R. S. (2013). West Nile virus: review of the literature. Jama, 310(3), 308-315.
- 4. Kramer, L. D., Li, J., & Shi, P. Y. (2007). West nile virus. *The Lancet Neurology*, 6(2), 171-181.
- 5. Blankenberg, D., & Hillman-Jackson, J. (2014). Analysis of next-generation sequencing data using Galaxy. In *Stem cell transcriptional networks* (pp. 21-43). Humana Press, New York, NY.
- 6. Bashiardes S, Zilberman-Schapira G, and Elinav E, (2016), Use of Metatranscriptomics in Microbiome Research, Bioinform Biol Insights. ; 10: 19–25.
- 7. Handelsman J, (2004), Metagenomics: Application of Genomics to Uncultured Microorganisms, Microbiol Mol Biol Rev. 68(4): 669–685.
- 8. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, A P de Beer T, Rempfer C, Bordoli L, Lepore R and Schwede T, (2018), SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res.; 46(Web Server issue): W296–W303.
- 9. Vyas VK, Ukawala RD, Ghate M, and Chintha C (2012), Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives, Indian J Pharm Sci.74(1): 1–17.
- 10. Morris GM, Lim-Wilby M, (2008), Molecular docking, Methods Mol Biol, 443:365-82.
- 11. Andrews, S. (n.d.). FastQC A Quality Control tool for High Throughput Sequence Data. Retrieved from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
- Ewels, P., Magnusson, M., Lundin, S., &Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics, 32(19), 3047–3048. https://doi.org/10.1093/bioinformatics/btw354
- Kopylova, E., Noé, L., &Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics, 28(24), 3211–3217. https://doi.org/10.1093/bioinformatics/bts611
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research, 41(D1), D590–D596. https://doi.org/10.1093/nar/gks1219
- Burge, S. W., Daub, J., Eberhardt, R., Tate, J., Barquist, L., Nawrocki, E. P., ... Bateman, A. (2012). Rfam 11.0: 10 years of RNA families. Nucleic Acids Research, 41(D1), D226–D232. https://doi.org/10.1093/nar/gks1005
- 16. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460–2461. https://doi.org/10.1093/bioinformatics/btq461
- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., &Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. Nature Biotechnology, 30(5), 434–439. https://doi.org/10.1038/nbt.2198
- Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., & and, A. N. (2010). Manipulation of FASTQ data with Galaxy. Bioinformatics, 26(14), 1783–1785. https://doi.org/10.1093/bioinformatics/btq281
- Beghini, F., McIver, L. J., Blanco-Míguez, A., Dubois, L., Asnicar, F., Maharjan, S., ... Segata, N. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. ELife, 10. https://doi.org/10.7554/elife.65088
- Ondov, B. D., Bergman, N. H., &Phillippy, A. M. (2011). Interactive metagenomic visualization in a Web browser. BMC Bioinformatics, 12(1). https://doi.org/10.1186/1471-2105-12-385

- Cuccuru, G., Orsini, M., Pinna, A., Sbardellati, A., Soranzo, N., Travaglione, A., ... Fotia, G. (2014). Orione, a web-based framework for NGS analysis in microbiology. Bioinformatics, 30(13), 1928–1929. https://doi.org/10.1093/bioinformatics/btu135
- 22. Asnicar, F., Weingart, G., Tickle, T. L., Huttenhower, C., &Segata, N. (2015). Compact graphical representation of phylogenetic data and metadata with GraPhlAn. PeerJ, 3, e1029. https://doi.org/10.7717/peerj.1029
- 23. Abubucker, S., Segata, N., Goll, J., Schubert, A. M., Izard, J., Cantarel, B. L., ... Huttenhower, C. (2012). Metabolic Reconstruction for Metagenomic Data and Its Application to the Human Microbiome. PLoS Computational Biology, 8(6), e1002358. https://doi.org/10.1371/journal.pcbi.1002358
- 24. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17(1), 10. https://doi.org/10.14806/ej.17.1.200
- 25. Jagtap P, Mehta S, Sajulga R, Batut B, Leith E, Kumar P, Hiltemann S, 2021 Metatranscriptomics analysis using microbiome RNA-seq data (Galaxy Training Materials).
- Gough, J., Karplus, K., Hughey, R. and Chothia, C. (2001). Assignment of Homology to Genome Sequences using a Library of Hidden Markov Models that Represent all Proteins of Known Structure. J. Mol. Biol., 313(4), 903-919

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

