1	Genome sequence of white sturgeon herpesvirus 1 isolated from farmed white sturgeon
2	(Acipenser transmontanus)
3	
4	Thomas B. Waltzek ^{a,b,#} , Kuttichantran Subramaniam ^a , Andor Doszpoly ^c , Joseph Hughes ^d , Matej
5	Vučak ^d , Andrew J. Davison ^d
6	
7	^a Department of Infectious Diseases and Immunology, College of Veterinary Medicine, University of
8	Florida, Gainesville, Florida, USA
9	^b Present address: Washington Animal Disease Diagnostic Laboratory and Department of
10	Veterinary Microbiology and Pathology, College of Veterinary Medicine, Washington State
11	University, Pullman, Washington, USA
12	^c Veterinary Medical Research Institute, Budapest, Hungary
13	^d MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom
14	
15	Running title: Genome sequence of white sturgeon herpesvirus 1

- 16
- 17 #Address correspondence to Thomas Waltzek, tomwaltzek@gmail.com

18 ABSTRACT

19 The genome sequence of white sturgeon herpesvirus 1, which was isolated from farmed white

- 20 sturgeon (Acipenser transmontanus), was determined. Comparative analyses suggest the
- 21 classification of this virus as a new species in a new genus in the family *Alloherpesviridae*.
- 22

23 Words = 39

- 24
- 25 **Keywords:** *Alloherpesviridae, Acipenser transmontanus,* sturgeon, genome.
- 26

27 ANNOUNCEMENT

White sturgeon herpesvirus 1 (WSHV1) was isolated on the white sturgeon skin (WSSK-1) cell 28 line from juvenile white sturgeon experiencing elevated mortality on a commercial farm in 29 30 California, USA in 1991 (1). Additional herpesviruses were isolated subsequently from farmed white sturgeon experiencing elevated morbidity and mortality (2). Based on differences from 31 WSHV1 in serological properties and growth characteristics in cell culture, these isolates were 32 33 designated white sturgeon herpesvirus 2 (WSHV2). Phylogenetic analyses based on 12 core genes 34 present in all members of the family Alloherpesviridae revealed that WSHV1 is distantly related to WSHV2 (3) and that its closest relative is lake sturgeon herpesvirus (LSHV) (4). 35 36 The WSHV1 isolate (strain UC Davis) was passaged six times in WSSK-1 cells (1), and DNA was 37 purified from the culture medium using a Blood & Cell Culture DNA mini kit (Qiagen). A sequencing 38 library prepared using a Nextera XT kit (Illumina) was sequenced on a MiSeq instrument (Illumina) 39 using a v3 600-cycle reagent kit, yielding 11,301,304 300 nt paired-end reads. The data were 40 analyzed using default options except where stated otherwise. The reads were trimmed using Trim 41 Galore v.0.6.6 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with the -q 25, --42 length 50 and --paired options. The trimmed reads were deduplicated using FastUniq v.1.1 (5) and subsampled using Seqtk v.1.3 (<u>https://github.com/lh3/seqtk</u>) with the sample, -s34 and 0.03 (to 43 discard 97% of reads) options. The subsampled reads were assembled into contigs using SPAdes 44 v3.14.1 (6) with the --careful, --cov-cutoff auto and -k 33,77,127 options. Contigs related to the 45 genomes of members of the family *Alloherpesviridae* were identified by using BLASTX (7) against 46 47 the NCBI non-redundant protein database. These contigs were joined to others by extracting reads

that extended their ends until overlaps were found. The genome terminI were assigned as 48 outlined below. The sizes of seven tandem reiterations were not resolved. The integrity of the 49 50 sequence was verified by aligning the trimmed reads using Bowtie 2 v2.3.1 (8) and Samtools v1.3 (9) and inspecting the alignment using Tablet v1.21.02.08 (10). The alignment incorporated 98% of 51 52 reads at an average coverage depth of 14,704 reads/nt. The WSHV1 genome (195,072 bp; 40% G+C) consists of a long unique region (U_1 ; 145,962 bp) 53 flanked by an inverted repeat (TR_L/IR_L ; 26 bp), linked to a short unique region (U_S ; 42,630 bp) also 54 flanked by an inverted repeat (TR_s/IR_s; 3,214 bp), yielding the overall structure 55 $TR_1 - U_1 - IR_1 - IR_5 - U_5 - TR_5$ (Fig. 1). The left terminus was assigned from a large set of reads sharing an 56 57 end, and the right terminus from the adjacent sequence at the IR_L-IR_S junction. A total of 129 open reading frames (ORFs) encoding functional proteins were predicted using approaches described 58 previously (4) (Fig. 1). 59 60 Although WSHV1 and LSHV (4) are each other's closest relative, the core genes are only 76.8% identical. This suggests the creation of a new genus in the family Alloherpesviridae incorporating 61 two new species. A second partially sequenced lake sturgeon herpesvirus (11) is more closely 62 related to LSHV than is WSHV1; the core genes are 84.3% identical. 63 64 Words = 497 65 66 Data availability 67 68 The WSHV1 genome sequence is available in GenBank under accession number OR001786. The sequence reads are available under BioProject accession number PRJNA1004073. 69 70 71 ACKNOWLEDGMENTS 72 TBW was supported by the College of Veterinary Medicine, Washington State University. AD was supported by the Bolyai János Research Scholarship. AJD was supported by Medical 73 Research Council grant MC UU 12014/3. The authors thank Dr. Ronald P. Hedrick for providing 74

75 WSHV1 strain UC Davis.

76

77 **REFERENCES**

- Hedrick RP, McDowell TS, Groff JM, Yun S, Wingfield WH. 1991. Isolation of an epitheliotropic
 herpesvirus from white sturgeon *Acipenser transmontanus*. *Dis Aquat Org* 11:49–56.
- 2. Watson LR, Yun SC, Groff JM, Hedrick RP. 1995. Characteristics and pathogenicity of a novel
- herpesvirus isolated from adult and subadult white sturgeon *Acipenser transmontanus*. *Dis Aquat Org* 22:199–210.
- Waltzek TB, Kelley GO, Alfaro ME, Kurobe T, Davison AJ, Hedrick RP. 2009. Phylogenetic
 relationships in the family *Alloherpesviridae*. *Dis Aquat Org* 84:179–194.
- 4. Walker L, Subramaniam K, Viadanna PHO, Vann JA, Marcquenski S, Godard D, Kieran E, Frasca
- 86 S Jr, Popov VL, Kerr K, Davison AJ, Waltzek TB. 2022. Characterization of an alloherpesvirus
- 87 from wild lake sturgeon *Acipenser fulvescens* in Wisconsin (USA). *Dis Aquat Org* 149:83–96.
- 5. Xu H, Luo X, Qian J, Pang X, Song J, Qian G, Chen J, Chen S. 2012. FastUniq: a fast *de novo*duplicates removal tool for paired short reads. *PLoS One* 7:e52249.
- 90 6. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI,
- 91 Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.
- 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.
 J Comput Biol 19:455–477.
- 7. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J
 Mol Biol 215:403–410.
- 8. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 97 9:357–359.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000
 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and
 SAMtools. *Bioinformatics* 25:2078–20792.
- 101 10. Milne I, Stephen G, Bayer M, Cock PJA, Pritchard L, Cardle L, Shaw PD, Marshall D. 2013. Using
- 102 Tablet for visual exploration of second-generation sequencing data. Brief Bioinform 14:193–
- 103 202.
- 104 11. Johnston AE, Shavalier MA, Scribner KT, Soto E, Griffin MJ, Waldbieser GC, Richardson BM,
- 105 Winters AD, Yun S, Baker EA, Larson DL, Kiupel M, Loch TP. 2022. First isolation of a
- 106 herpesvirus (family *Alloherpesviridae*) from Great Lakes lake sturgeon (*Acipenser fulvescens*).
- 107 *Animals* 12:3230.

ORF1 ORF3 ORF2	35 RF23 ORF24	ORF8 40 ORF25 ORF2 0RF25 ORF2 70 2 ORF46 0RF43 ORF44	26 ORF27 OF	ORF13 45 7 ORF30 RF28 ORF3 ORF29 O 75 75 48 ORF51	31 DRF32	ORF15 OR 50 F33 ORF34	ORF18 ORF17 ORF ORF35		:20 	ORF2: 11(11; 11; 11; 0RF37 11; 11; 11; 11; 11; 11; 11; 11; 11; 11
OI 01 01 01 01 01 01 01 01 01 01 01 01 01	RF23 ORF24	ORF25 ORF2 70 2 ORF46 0RF43 ORF44	26 ORF27 OF 	7 ORF30 RF28 ORF3 ORF29 O 75 75 48 ORF51	31 DRF32	F33 ORF34	ORF35	ORF3	· · · ·	111 0RF37 114 114
539 ORF39 ORF39A ORF39B	65 • • • • •	70 70 2 ORF46 0RF43 0RF44	OF	RF28 ORF3 ORF29 C 75 X X 48 ORF51	31 DRF32	ORF34	ORF35		36 	11. ORF37 114 111
539 ORF39 ORF39A ORF39B	65 • • • • •	70 70 2 ORF46 0RF43 0RF44	OF	RF28 ORF3 ORF29 C 75 X X 48 ORF51	31 DRF32	ORF34	ORF35		36	ORF37
F39 ORF39 ORF39A ORF39B	ORF4	2 ORF46 DRF43 ORF44	ORF4	48 ORF51		80		85		, ,
739 ORF39 ORF39A ORF39B	OC ORF4	2 ORF46 DRF43 ORF44						ĵ C		1
739 ORF39 ORF39A ORF39B	OC ORF4	2 ORF46 DRF43 ORF44								
· · · · · ·		ORF45		RF49 ORI ORF50	F52	RF53 ORF8			ORF6	
	95	100)	105		110		115		'11
ORF67 C ORF68	ORF70 ORF72		ORF75 ORF76			DRF79 C	DRF83 O	RF85 ORF86		12
	125	130)	135		140		145	-	12
)∎ (•	
	RF92 ORF94 ORF93	ORF95	ORF96 ORF97 ORF9	OR	OR RF100	F101 OR	F102 ORF1	03	O	RF103/ 12 12
)	155	160	, , , , , , , , , , , , , , , , , , ,	165	,	170		175		12
										12
ORF109A ORF10	ORF110 OF A ORF111 09B ORF111A 109C	RF113 ORF1 ²		ORF ⁷ DRF117	 118 O	RF119 ORF1		22 OR ORF123	EF124 ORF12	12
)	185	190)	 195 kbp	o					13
				•						13
F126 ORF12		OR	F130A							13
ORF128	ORF130		ORF103/	A						13
ORF0	-	IgD	⁼ 39A famil family ⁼ 64 family	у	ORF	77 family 81 family 98 family			00 family noncore	- 13

FIG 1. WSHV1 genome map. Unique regions (U_L and U_S) are shown in a thinner format than inverted repeats
 (TR_L/IR_L and IR_S/TR_S). Predicted functional ORFs are named to correspond with orthologues in the LSHV
 genome and are indicated by arrows colored according to the key as belonging to gene families (sets of
 paralogous genes), core ORFs (conserved among alloherpesviruses) and other noncore ORFs. Some ORFs are
 shown by narrow arrows to make their locations clearer. Introns connecting ORFs are shown as narrow white

141 bars. Small, black rectangles indicate tandem repeats of undetermined size.