

Article

The Genomic Diversity and Phylogenetic Relationship in the Family *Iridoviridae*

Heather E. Eaton, Brooke A. Ring and Craig R. Brunetti *

Trent University, Peterborough, ON, K9J 7B8, Canada, E-Mails: heathereaton@trentu.ca (H.E.E.); brooking@trentu.ca (B.A.R.)

* Author to whom correspondence should be addressed; E-Mail: craigbrunetti@trentu.ca; Tel.: +1-705-748-1011; Fax: +1-705-748-1205.

Received: 8 June 2010; in revised form: 12 July 2010 / Accepted: 13 July 2010 /

Published: 15 July 2010

Abstract: The *Iridoviridae* family are large viruses (~120-200 nm) that contain a linear double-stranded DNA genome. The genomic size of *Iridoviridae* family members range from 105,903 bases encoding 97 open reading frames (ORFs) for frog virus 3 to 212,482 bases encoding 211 ORFs for Chilo iridescent virus. The family *Iridoviridae* is currently subdivided into five genera: *Chloriridovirus*, *Iridovirus*, *Lymphocystivirus*, *Megalocytivirus*, and *Ranavirus*. Iridoviruses have been found to infect invertebrates and poikilothermic vertebrates, including amphibians, reptiles, and fish. With such a diverse array of hosts, there is great diversity in gene content between different genera. To understand the origin of iridoviruses, we explored the phylogenetic relationship between individual iridoviruses and defined the core-set of genes shared by all members of the family. In order to further explore the evolutionary relationship between the *Iridoviridae* family repetitive sequences were identified and compared. Each genome was found to contain a set of unique repetitive sequences that could be used in future virus identification. Repeats common to more than one virus were also identified and changes in copy number between these repeats may provide a simple method to differentiate between very closely related virus strains. The results of this paper will be useful in identifying new iridoviruses and determining their relationship to other members of the family.

Keywords: *Iridoviridae*; evolution; repetitive sequences

1. Introduction

Iridoviruses are large double stranded DNA viruses (~120 – 200 nm) that replicate in the cytoplasm of infected cells. They are characterized by a distinct icosahedral capsid and range in size from 102 to 212 kbp [1,2]. Iridoviruses are subdivided into five genera that infect a wide range of vertebrate (*Ranavirus*, *Megalocytivirus*, *Lymphocystivirus*) and invertebrate (*Iridovirus*, *Chloriridovirus*) hosts [3]. Specifically, iridoviruses have been found to infect fish, amphibians, reptiles, crustaceans, molluscs, and insects [4]. Following attachment and entry of iridoviruses via receptor-mediated endocytosis, virus particles are uncoated and translocated to the nucleus where the first stage of DNA replication occurs via a virus-encoded DNA polymerase [5,6]. Viral DNA then exits the nucleus to the cytoplasm for the second stage of DNA replication where the formation of DNA concatemers occurs [5]. Iridoviruses are unique among eukaryotic virus genomes because they are described as both circularly permuted and terminally redundant [7-9].

Clinical manifestations of iridovirus infections vary greatly between species and can range from an absence in observable symptoms to death [10-14]. The high morbidity and mortality associated with some iridovirus members has demonstrated their high ecological and economical significance, specifically in aquatic organisms [15-19]. With the continuing isolation of new iridovirus-like viruses from a variety of species worldwide it is imperative to gain a greater understanding of iridovirus pathogenesis.

Fifteen iridoviruses have currently been sequenced including at least one from each genus. They include frog virus 3 (FV3; *Ranavirus*) [20], soft-shelled turtle iridovirus (STIV; *Ranavirus*) [21], tiger frog virus (TFV; *Ranavirus*) [1], epizootic haematopoietic necrosis virus (EHNV; *Ranavirus*) [22], *Ambystoma tigrinum* virus (ATV; *Ranavirus*) [23], grouper iridovirus (GIV; *Ranavirus*) [24], Singapore grouper iridovirus (SGIV; *Ranavirus*) [25], lymphocystis disease virus 1 (LCDV-1; *Lymphocystivirus*) [26], lymphocystis disease virus China (LCDV-C; *Lymphocystivirus*) [27], infectious spleen and kidney necrosis virus (ISKNV; *Megalocytivirus*) [28], rock bream iridovirus (RBIV; *Megalocytivirus*) [29], red sea bream iridovirus (RSIV; *Megalocytivirus*) [30], orange-spotted grouper iridovirus (OSGIV; *Megalocytivirus*) [31], invertebrate iridescent virus 6/Chilo iridescent virus (IIV-6/CIV; *Iridovirus*) [2], and invertebrate iridescent virus 3/mosquito iridescent virus (IIV-3/MIV; *Chloriridovirus*) [32]. These 15 sequenced genomes represent iridoviruses isolated from a variety of hosts including fish, amphibians, reptiles, and insects. The diverse host range of iridoviruses is reflected in the diverse gene content found between iridoviruses of different genera.

Little is currently known about the molecular biology of this family of viruses and much about the evolutionary biology of iridoviruses still contains many unanswered questions. Despite many recent advances in molecular phylogenetics, there is much to learn about the relationship of iridoviruses within the family itself. In order to gain a greater understanding of iridovirus evolutionary history, we will use previously identified iridovirus core genes and repetitive DNA sequences to explore the evolutionary links between iridoviruses. An increased knowledge about the evolutionary biology of iridoviruses may lead to a better understanding of the functional biology of these viruses, specifically in the understanding of iridovirus pathogenesis.

2. Results and Discussion

2.1. Phylogenetic analysis

Whether representing a single gene or a consensus, the *Iridoviridae* family genera branching order is often inconsistent between genomic papers [1,21,23,29,31,33]. The discrepancies between papers might be a result of different sequence alignment methods or comparison of insufficient data sets. In order to clarify the evolutionary relationships of the family *Iridoviridae*, a phylogenetic analysis was constructed using sequence alignments.

The open reading frames of 26 conserved iridovirus genes shared by 14 iridoviruses (representing all five genera; Figure 1A) were aligned using ClustalW in BioEdit 7.0.5. The sequence for the RSIV genome is not available and was therefore not included in the analysis. The alignments were then transferred to MEGA4.1 and fused together to create a consensus tree (Figure 1B). The consensus tree is a phylogenetic branching diagram of the *Iridoviridae* family that shows the evolutionary relationships between 14 iridovirus species (Figure 1B). The nodes of the tree demonstrate shared ancestry, and the length of the branches represent an estimation in time, although they can also reflect evolutionary pressures. The tree shows one main common ancestor, which shares 26 conserved genes with modern *Iridoviridae* (Figure 1B). The tree divides into two branches, the first branch consists of the *Iridovirus* and *Lymphocystivirus* genera and the second branch consists of the *Chloriridovirus*, *Megalocytivirus*, and *Ranavirus* genera (Figure 1B). Jakob *et al.* [2] sequenced the IIV-6 genome and found through comparison of 10 core gene products that the *Iridovirus* genus was most closely related to the *Lymphocystivirus* genus, which is consistent with the results of our study [2,27]. Many other studies have argued that the *Iridovirus* genus is more closely related to the *Chloriridovirus* genus however, this may be a result of insufficient data sets which do not include all genera, insufficient numbers of core genes used in the analysis, or different alignment methods [21,28,32,33]. Individual phylogenies showed that the *Iridovirus* genus clustered closer to the *Lymphocystivirus* genus than the *Chloriridovirus* genus in 22 out of the 26 core genes (data not shown). Another phylogeny based on the aligned genomes of IIV-6, IIV-3, LCDV-1 and LCDV-C also clustered the *Iridovirus* genus closer to the *Lymphocystivirus* genus than the *Chloriridovirus* genus (data not shown).

The *Megalocytivirus* genus is composed of very closely related sister taxa whose genomes differ only by 2.3%. The placement of the *Megalocytivirus* genus between the *Chloriridovirus* and *Ranavirus* genera has been previously observed [32]. It should be noted that the *Megalocytivirus* genus is more often observed to cluster with the *Lymphocystivirus* and *Ranavirus* genera, however, this outcome occurs in data sets that do not include the *Chloriridovirus* genus [21,23,28,31]. The relatedness of the *Megalocytivirus* and *Ranavirus* genera has been well documented in previous phylogenetic analysis between genomic papers [21,27,32]. Species within the *Ranavirus* genus are generally very closely related, however, SGIV and GIV are considered outliers as their genomes differ in sequence identity by approximately 30% from the genomes of FV3, STIV, TFV, ATV and EHNV, while the sequence identity within other *Ranavirus* genomes (FV3, STIV, TFV, ATV and EHNV) differ by less than approximately 10% [23,25,34]. The branching order of the individual viruses within the *Ranavirus* genus is consistent with many genomic papers [20,21,27,31,33].

2.2. G/C content

G/C content of specific genes can provide information about gene origins (*i.e.*, host-derived genes). Comparison of iridovirus gene G/C content with host species G/C content did not show evidence of apparent host gene transfer, but we did find evidence of specific trends in G/C content throughout iridovirus genomes. We used the VGO software to identify the G/C content of each coding and non-coding region of the sequenced iridoviruses (excluding STIV, EHNV, and RSIV which are not included in the VGO database) [35]. For some genera (*Ranaviruses*, *Chloriridovirus*), coding regions were found to exhibit higher G/C content than both the average G/C content of the virus and the G/C content of non-coding regions (Figure 2A). Figure 2B clearly shows that for coding regions of FV3, SGIV, and IIV-3, G/C content fluctuates around or above the average G/C content for the virus. G/C content in non-coding regions dips dramatically below the average G/C content (Figure 2B). This trend is most likely due to the fact that promoters are often A/T rich and therefore regions upstream of coding regions will contain a higher A/T content with respect to other locations in the genome. Specifically, large DNA virus promoters (*Poxviridae*, *Iridoviridae*, *Asfarviridae* families) have been described as A/T rich [36-39]. In contrast, viruses of the *Megalocytivirus* genus show the opposite trend of the *Ranavirus* and *Chloriridovirus* genera (Figure 2A and 2B). G/C content of non-coding regions is higher than both overall genome and coding region G/C content. Furthermore, some genera (*Lymphocystivirus*, *Iridovirus*) do not show any differences in G/C content throughout any part of their genome. This is most likely a result of a much lower overall G/C content than the *Ranavirus* and *Chloriridovirus* genera. These results demonstrate that variations in G/C content between non-coding and coding regions could provide an alternative method in annotating genomes, specifically in the *Ranavirus* and *Chloriridovirus* genera. The trends seen in G/C content between coding and non-coding regions may also help to classify new viruses or virus strains into their appropriate genus.

2.3. *Iridoviridae* repetitive sequences

While comparison of conserved genes and phylogenetics are valuable tools for exploring the evolution of viruses, the molecular differences between genomes is critical for understanding their evolutionary relationships. Repetitive sequences are key when assessing gene and genome diversity and provide a method for examining the evolution of viruses [40]. Simple sequence repeats are often widely abundant in genomes, and have been identified not only in a wide variety of eukaryotic and prokaryotic genomes, but also in several virus families. Specifically, dsDNA viruses including poxviruses, herpesviruses, baculoviruses, and adenoviruses have been found to contain extensive repeat sequences within their genomes [41-45]. Abundant numbers of repetitive sequences are also found throughout iridoviruses genomes [2,20,21,23-28,31,32], however no comparative approach has been taken to analyze the repeats between all iridoviruses. We are interested in determining the distribution and abundance of repeat elements throughout the sequenced genomes of the family *Iridoviridae*.

A DNA tandem repeat finder identified tandem repeats found within all available sequenced iridovirus genomes [46]. Repeats between 1–6 nucleotides in length are referred to as microsatellites and were found only rarely in iridoviruses (Table 1, Supplementary Tables 1–7). Repeats containing a minimum of 6–12 nucleotides are referred to as minisatellites. This type of tandem repeat was found

extensively through all iridovirus genomes (Table 1, Supplementary Tables 1–7). The copy number of the repeats also varied within and between genomes (Table 1, Supplementary Tables 1–7). Variability in repetitive sequences can occur through many mechanisms including recombination and slipped strand mispairing [47]. Inefficient polymerase activity and poor repair mechanisms during DNA replication can result in alterations within a repeat sequence [48,49].

Repetitive sequences can be grouped into categories based on several factors, one of those being their respective locations to coding regions in a genome. The repetitive sequences analyzed in the sequenced iridovirus genomes were widely dispersed and were located in both non-coding and coding regions (Figure 3). Furthermore, the majority of genomes contained individual repeats that transverse between both coding and non-coding regions (Figure 3). Repetitive sequences were also identified to occur within the core iridovirus genes of many viruses. While the current function of these repeats is unknown, simple sequence repeats have been found to influence gene regulation, transcription, and protein function (reviewed in [50]). Repetitive sequences can act as structural elements as well as binding sites for proteins [50]. This modulation in gene expression may be the result of changes in repeat number that alter the physical integrity of DNA domains as coding regions become modified or disrupted. Specifically, simple sequence repeat involvement in the modulation of gene expression has been recently identified in a number of microorganisms, including viruses [51–55]. In addition, changes in repeat number can cause quantitative changes in gene expression and function, which may lead to variation between similar viruses. For example, host range and pathogenicity could be affected by differences in repeat number between viruses. As more and more sequenced genomes become available it provides an opportunity to study the involvement of these repeat sequences and their effect on gene function and pathogenicity.

The number of total genomic repeats varied dramatically between iridoviruses and the number of repeats that were similar between viruses varied greatly (Figure 3). Large numbers of repeats were found in the *Ranavirus*, *Iridovirus*, and *Chloriridovirus* genera, while the *Lymphocystivirus* and *Megalocytivirus* genera exhibit relatively fewer repeats (Figure 3, Table 1, Supplementary Tables 1 – 7). Not only were fewer repeats present in the genomes, but also the copy numbers of the repetitive sequences were fewer than in other genera.

No repetitive sequences were shared between all iridoviruses or even between all members of one genus. In the *Lymphocystivirus* genus, no repeats were shared between LCDV-1 and LCDV-C, which is expected because the whole genome sequence identity between these two viruses is extremely low (14%; Supplementary Table 4). No repeats were shared between all members of the *Megalocytivirus* genus. However, all the repeats found within RBIV are shared with OSGIV, both in terms of sequence identity and copy number. This suggests that these viruses are extremely closely related and most likely strains of the same virus. ISKNV does not share any repetitive sequences in common with RBIV and OSGIV even though the sequence identity between ISKNV and RBIV or OSGIV is relatively high (97%; Supplementary Table 3). The relationship between viruses of the *Megalocytivirus* genus shown in Figure 1B is clearly supported using repetitive sequence data. While IIV-3 and IIV-6 contain many tandem repeats with varying sizes and copy numbers, the repeats are not similar to each other or to other iridoviruses (Supplementary Table 5, 6).

Figure 1. The phylogenic relationship of 14 iridoviruses as identified by comparison of 26 conserved iridovirus genes. **(A)** ORF’s of 26 conserved iridovirus genes were recorded from 14 iridoviruses. **(B)** The nucleotide sequence of the conserved genes were aligned and fused to create a neighbor-joining consensus tree. The numerical values adjacent to the branching nodes indicate bootstrap values (black) and the number of genes in common between genomes (red). This tree was constructed in MEGA4.1 using the p-distance model at 500 replicates.

A

Gene Name	Ranavirus							Lymphocystivirus		Megalocytivirus			Iridovirus	Chloriridovirus
	FV3	STIV	TFV	EHN	ATV	SGIV	GIV	LCDV-I	LCDV-C	ISKNV	RBIV	OSGIV	IIV-6	IIV-3
1 Putative replication factor and/or DNA binding/packing	1R	1R	105R	100R	91R	116R	79R	162L	181R	61L	57L	60L	282R	79L
2 Myristilated membranc protein	2L	2L	2L	1L	1L	19R	4R	160L	38R	90.5L	85L	88.5L	337L	47R
3 DNA-dep RNA pol-II largest subunit	8R	10R	8R	7R	6R	104L	71L	16L	191R	28L	29L	31L	176R, 343L	90L
4 Putative NTPase I	9L	11L	9L	8L	7L	60R	36R	132L	75L	63L	59L	63L	22L	87L
5 Unknown	12L	14L	12L	95R	87R	118R	80R	108L	100L	96L	89.5L	93L	287R	56L
6 ATPase-like protein	15R	16R	16R	92L	83R	134L	90L	54R	114L	122R	116R	119R	75L	88R
7 Serine-threonine protein kinase	19R	21R	19R	89L	80L	39L	21L	10L	45R	55L	53L	56L	380R	10L
8 Helicase family	21L	24L	21L	86R	78R	54R	32R	6L	7L	56L	54L	57L	67R	4R
9 D5 family NTPase involved in DNA replication	22R	25R	22R	85L	77L	52L	31L	128L	80L	109L	101L	106L	184R	121R
10 Putative tyrosine kinase/lipopolysaccharide modifying enzyme	27R	31R	29R	62R	58R	78L, 81L	52L	195R	173R	114L	106L	111L	179R, 439L	35R
11 NIF-NLI interaction factor	37R	41R	40R	72R	64R	61R	37R	82L	148L	5L	6L	6L	355R	104L
12 Unknown	41R	45R	45R	77R	69R	57L	35L	163R	235R	76L	72L	75L	295L	16R
13 Myristilated membrane protein	53R	55R	55R	53L	51L	88L	59L	67L	158R	7L	8L	8L	188L, 458R	6R
14 Serine-threonine protein kinase	57R	60R	59R	58L	47L	150L	100L	143L	178L	13R	13R	15R	98R	98L
15 DNA pol Family B exonuclease	60R	63R	63R	44L	44L	128R	87R	135R	203L	19R	20R	22R	37L	120L
16 DNA-dep RNA pol-II second largest subunit	62L	64R	65L	43R	43R	73L	46L	25L	25R	34R	33R	36R	428L	9R
17 Ribonucleotide reductase small subunit	67L	71L	71L	38R	38R	47L	26L	27R	41L	24R	26R	27R	376L	48L
18 Ribonuclease III	80L	87L	85L	24R	25R	84L	55L	137R	187R	87R	83R	85R	142R	101R
19 Transcription elongation factor TFHS	81R	88R	86R	23L	24L	85R	56R	171R	115R	29L	29.5L	32L	349L	55R
20 Proliferating cell nuclear antigen	84R	91R	90R	19L	20L	68L	41L	3L	197L	112R	103R	109R	436R	60L
21 Deoxynucleoside kinase	85R	92R	91.5R	18L	19L	67L	40L	136R	27R	32R	31R	34R	143R	29R
22 ErvI/Alr family	88R	94R	94R	16L	16L	70R	43R	106L	142L	43L	43.5L	45L	347L	96R
23 Major capsid protein	90R	96R	96R	14L	14L	72R	45R	147L	43L	6L	7L	7L	274R	14L
24 Immediate early protein ICP-46	91L	97R	97R	13L	13L	162L	108L	47L	162R	115R	108.5R	112R	393L	39R
25 Hypothetical protein - Clostridium tetani	94L	99L	100R	11L	11L	98R	67R	19R	153L	86R	82.5R	84.5L	307L	33L
26 Putative XPPG-RAD2 type nuclease	95R	100R	101R	10L	10L	97L	66L	191R	169R	27L	28L	30L	369L	76L

Figure 1. Cont.

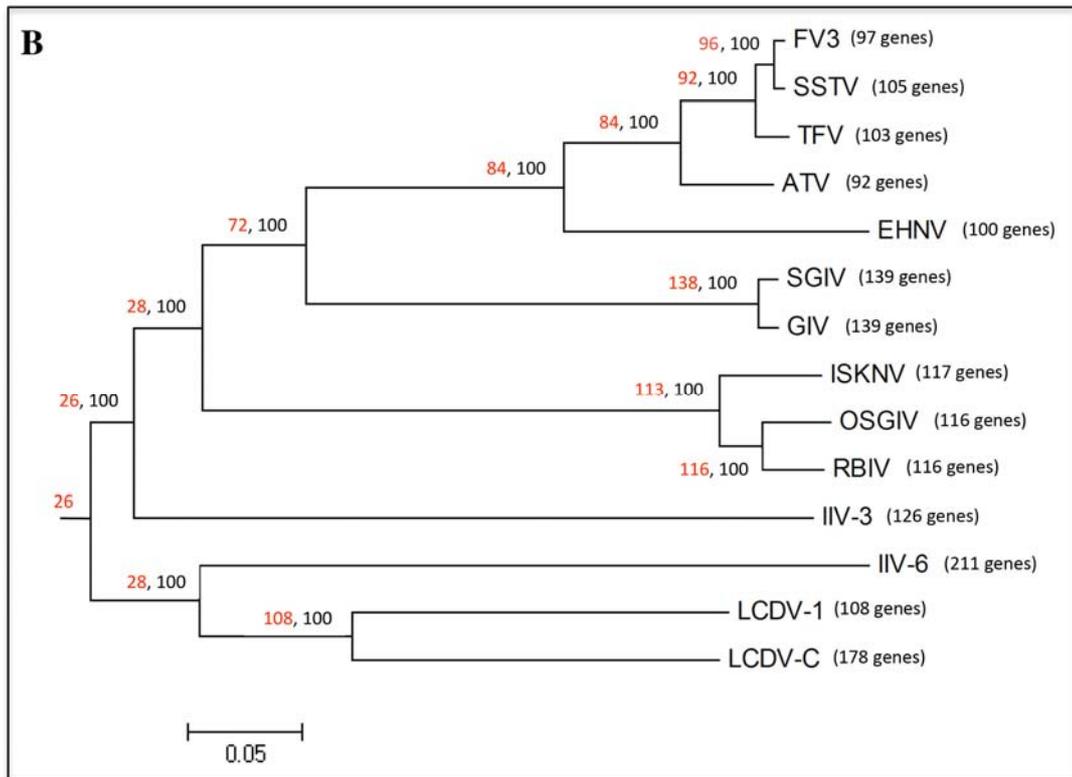


Figure 2. Different iridoviruses exhibit different trends in G/C content of coding *versus* non-coding regions. G/C content of coding regions was determined using the VGO software. The length and G/C content of the coding regions and overall length and G/C content was used to determine the G/C content of the non-coding regions. **(A)** Overall G/C content is shown in blue, coding region G/C content is shown in red, and non-coding G/C content is shown in green. **(B)** Images for one representative virus from each genus are displayed showing changes in G/C content (red line) in selected coding (blue bars) and non-coding regions of the genome. Average genome G/C content is displayed in blue.

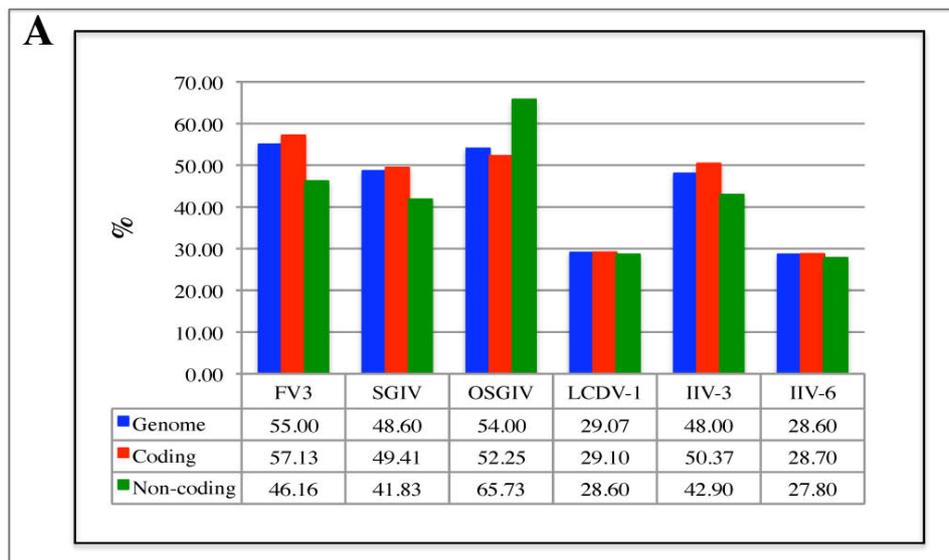


Figure 2. Cont.

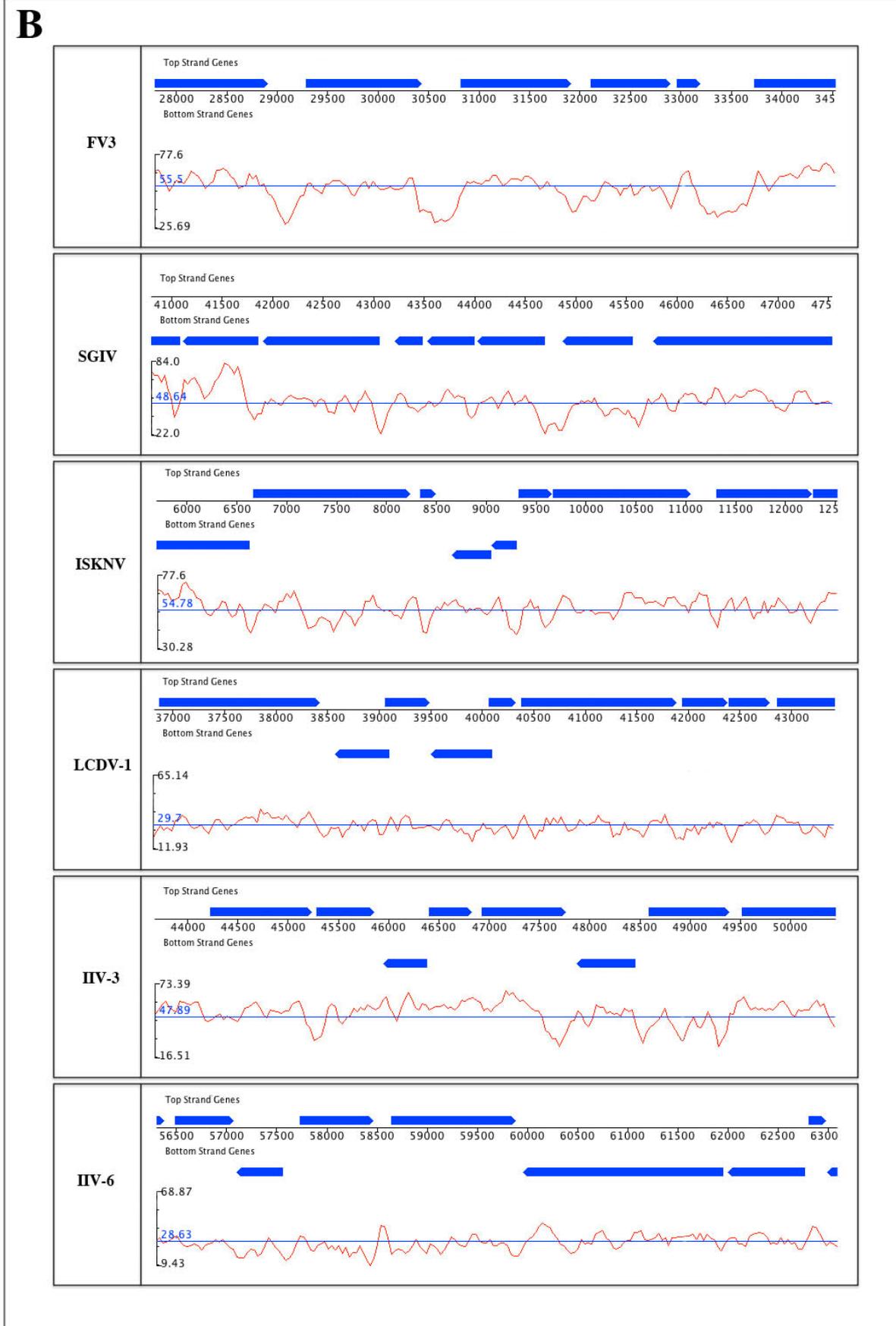
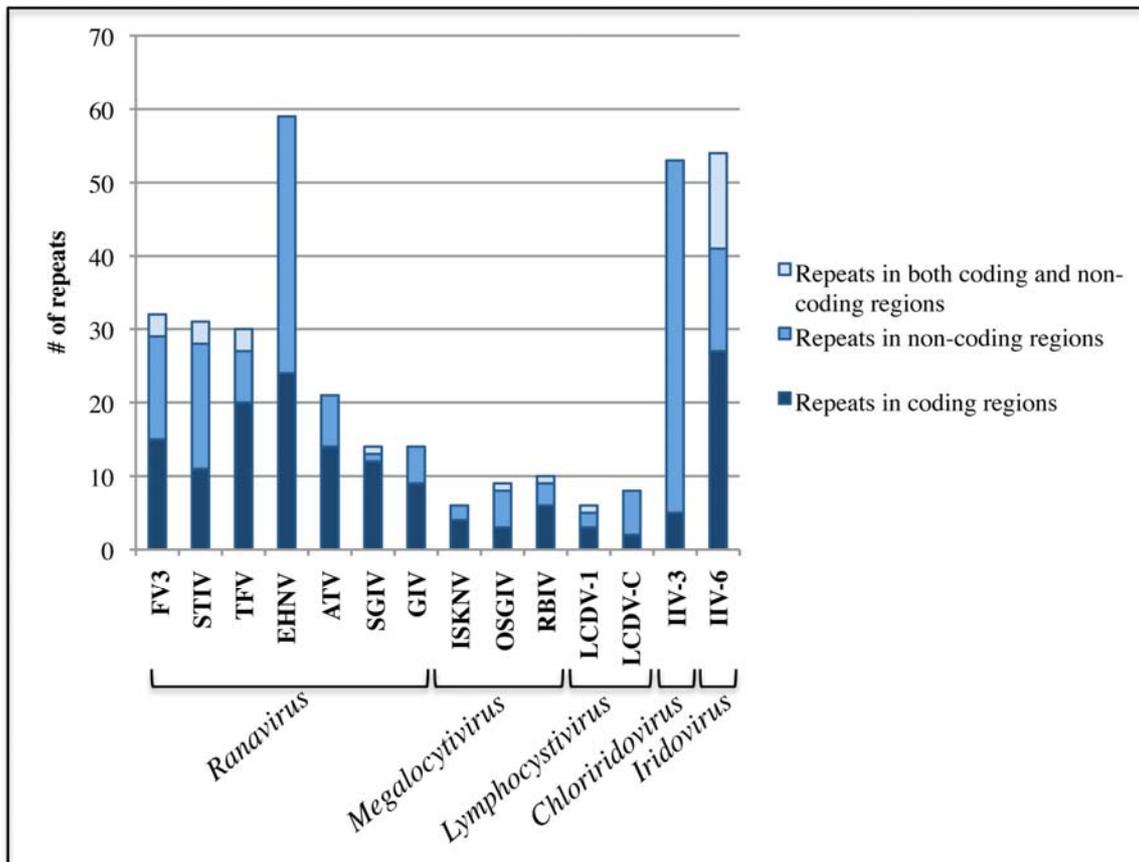


Table 1. Conserved repeats found in FV3, STIV, TFV, EHNV, and ATV. Repetitive sequences highlighted in yellow represent a repeat that differs in one nucleotide between viruses.

Sequence	FV3				STIV				TFV				EHNV				ATV			
	Indices	Period size	Copy number	Match identity	Indices	Period size	Copy number	Match identity	Indices	Period size	Copy number	Match identity	Indices	Period size	Copy number	Match identity	Indices	Period size	Copy number	Match identity
CONSERVED REPEATS																				
tatettaaga taa	70421--70445	13	1.9	100	70191--70215	13	1.9	100	68845--68869	13	1.9	100	54319--54343	13	1.9	100				
ggtaa tctctgat	6616--6642	14	1.9	100	6390--6416	14	1.9	100					5374--5498	14	8.9	100	3534--3560	14	1.9	100
ga gaaagctcc tgc	22509--22564	15	3.7	97	22236--22276	15	2.7	100	20333--20403	15	4.7	96								
tcacaacag cctctcagg etgtcttc tctctgt	54981--55158	39	4.6	98	54757--54890	39	3.4	100	53458--53534	39	2	97								
c cactcagat ctacca	45990--46026	18	2.1	100	45776--45812	18	2.1	100					93390--93430	18	2.3	100				
ag agcccaca gaacategc					39783--39831	21	2.3	100	38459--38570	21	5.3	98	88097--88145	21	2.3	96				
agaaagt ctccc					38839--38885	12	3.9	100	37464--37498	12	2.9	100					69741--69775	12	2.9	100
a gagtctcc taca					800--845	15	3.1	96					125666--126323	15	43.9	100	105620--105650	15	2.1	100
ca	80838--80905	2	34	100	80609--80676	2	34	100												
ggagtg gg	1734--1804	9	7.9	98	1499--1578	9	8.9	95												
ttaaac ttatc	25034--25058	13	1.9	100	24766--24790	13	1.9	100												
atctctga aagaat	50452--50480	15	1.9	100	91029--91063	15	2.3	100												
tcttcagaa gatic	65831--65862	15	2.1	100	65561--65592	15	2.1	100												
gcacttg catctga	41973--42032	16	3.8	100	41855--41886	16	2	100												
tcttgagata ctatta	60581--60656	16	4.8	98	60328--60387	16	3.8	97												
gggact taccttca a	52445--52737	18	16.2	98	52268--52597	18	18.3	100												
aa gtctcagct gggagaaagt ctccatc	38699--38928	30	7.9	98	38536--38651	30	3.9	100												
agtac gtctcttc tgalgtgaga cattaanaact ttgt	50166--50245	39	2.1	100	49952--50070	39	3.1	100												
tcttaa gatatactt ctctcagaaa gttcttaga caagaaggag atgtaaacac attatag	1245--1372	64	2	95	1004--1190	64	2.9	97												
ctccc ctgaaactt catctctcg gccgaggaag ccgtgtgaga cgagtgctc acagctctcg gctgaaggg aaggacttt ctctcgc	82295--82505	93	2.3	96	82066--82276	93	2.3	96												
t gaggtgagta	100057--100090	11	3.1	100					98650--98683	11	3.1	100								
g ctgtgtctgg tgta	51310--51347	15	2.5	100					49710--49837	15	8.5	97								
aagatgce ctactgaca acg	18613--18679	21	3.2	95					16563--16608	21	2.2	96								
agaaagt ctccc					39061--39107	12	3.9	100	37674--37756	12	6.9	100								
tatct taagatatta ccactctgat attat					28871--28952	30	2.7	100	26936--26992	30	1.9	96								
a agatatctta tctctga													25889--25931	19	2.3	96	22300--22342	19	2.3	96

Figure 3. Repeats in the family *Iridoviridae* vary in number and dispersion throughout the genomes. Tandem repeats were identified using a tandem repeat finder [46]. For each sequenced iridovirus, the number of tandem repeats found within coding regions are shown in dark blue, repeats found within non-coding regions are shown in blue, and repeats that transverse between both coding and non-coding regions are shown in light blue. The height of the bar represents the total number of repeats found in the virus.



2.4. *Ranavirus repetitive sequences*

Several shared repeats were found within the subset of ranaviruses that includes FV3, STIV, TFV, EHNV, and ATV (Table 1). These viruses have relatively high sequence identity between one another (>91%) and are similar in terms of size and G/C content. While no repeat was shared between all of these viruses, some repeats were found to be identical between 2 to 4 of these viruses (Table 1). A single repeat was found to be common between FV3, STIV, TFV, and EHNV (Table 1: FV3 indices - 70421) while another single repeat was found to be common between FV3, STIV, EHNV, and ATV (Table 1: FV3 indices - 6616). Although these viruses have relatively high sequence identity, the commonality between their repetitive sequences was found to be low. Only 3 repeats were found to be in common between FV3, STIV, and TFV (Table 1; FV3 indices - 70421, 22509, 54981), which is surprising considering the sequence identity between these three viruses is greater than 97%. Overall, 6 repeats from FV3 (FV3 indices - 70421, 22509, 54981, 100057, 51310, 18613), 7 from STIV (STIV indices - 70191, 22236, 54757, 39783, 38839, 39061, 28871), 2 from EHNV (EHNV indices - 54319,

88097), and 1 from ATV (ATV indices – 60741) were found to be in common with TFV. Furthermore, only 2 repeats from FV3 (FV3 indices – 6616), 3 from STIV (STIV indices – 6390, 38839, 800), 1 from TFV (TFV indices - 37464, and 3 from EHNV (EHNV indices – 5374, 125666, 25889) were found to be similar to ATV. The lack of conserved repeats between ATV and other viruses of this subset of ranaviruses is not surprising due to the fact that ATV has a much lower sequence identity with FV3, STIV, EHNV, and TFV than they do with each other. While there are very few conserved repeats between ATV and FV3, STIV, TFV, and EHNV, there are no repeats in common between ATV and the second subset of ranaviruses (GIV, SGIV). This confirms the evolutionary relationship between the *Ranavirus* genus found in Figure 1B. It suggests that ATV is more closely related to FV3, STIV, TFV, and EHNV, than with GIV and SGIV, but that FV3, STIV, and TFV are more closely related to each other than with ATV. EHNV, a ranaviruses with highest sequence identity to ATV, contains a substantial number of repeats within its genome compared to other ranaviruses. Interestingly, most of these repeats are unique only to EHNV, even though EHNV shares relatively high sequence identity to the other ranaviruses (FV3, STIV, TFV, ATV; >97%).

Although FV3 and STIV share 99% sequence identity, they share just over 50% of repeats in common with each other (Table 1). This suggests that much of the difference between these two viruses lies within the repetitive regions. While the match identity between repeats found in these two viruses is high, a large change in copy number between repeats is evident (Table 1, Figure 4). Changes in copy number between matching repeats may provide a simple method to differentiate between what is most likely strains of the same virus. Furthermore, a dinucleotide repeat (microsatellite) was identified to be common only between FV3 and STIV. This specific microsatellite may be exploited as a rapid method to identify FV3 or viruses with extremely high similarity to FV3.

While GIV and SGIV share no repeats in common with the first subset of ranaviruses, they do share several repeats in common with each other (Supplementary Table 2). Similar to the sequence identity between FV3 and STIV, GIV and SGIV are 99% similar but share only 50% of repeats in common. This again suggests that the differences between these two viruses lies within the repetitive sequence regions. While the repeats that are common between SGIV and GIV share high sequence identity with each other, they do exhibit large changes in copy numbers.

Our analysis of repetitive regions within sequenced iridoviruses has successfully identified both unique and similar repeats. Identical repeats that exhibit differing copy numbers can be used in conjunction with unique repeats to quickly and effectively identify iridoviruses. This has previously been tested by Jancovich *et al.* [16], in which a 16 base pair preset in ATV was used to help distinguish ATV isolates from Arizona, Utah, Colorado, and Canada. This technique may specifically apply to the identification of several isolates of FV3-like species listed by Hyatt *et al.* [56]. In the past decade there has been a significant increase in the number of iridoviruses found in vertebrates, many of which have yet to be properly classified [1,2,4,15,21,23,56]. The use of repetitive sequences are predicted to be sufficient in determining otherwise undetermined isolates of FV3-like viruses that infect reptiles, amphibians and fish. Due to their high mutation rate in copy numbers, repetitive sequences are considered ideal genetic markers and may provide an efficient method to distinguish between highly similar virus strains and further clarify the evolutionary link between viruses of this family.

2.5. Repetitive sequence flanking regions

Repetitive sequences may change in orders of magnitude faster than non-repetitive regions of the genome and are prone to deletions and duplications. In order to determine whether the regions flanking the repetitive sequences mutate slower than the repetitive sequences themselves, we compared the genetic sequence of the regions flanking the repeats that were common to more than one virus. If the repetitive regions mutate faster than other locations in the genome, then the flanking regions should exhibit fewer changes when compared to each other, even when the copy number of the repeat changes. Figure 4 shows the flanking regions (6 nucleotides on either side of the repeat) of each identified tandem repeat common between 2 or more viruses of the first subset of *Ranaviruses* (FV3, STIV, TFV, EHNV, ATV). The majority of flanking regions surrounding a single repeat were highly conserved between viruses. For repeats in which the copy number did not change between viruses, the flanking regions exhibited extremely high conservation with only a few small nucleotide changes being evident (Figure 4A). For repetitive sequences that exhibited changes in copy numbers, there still remained a high level of conservation in the flanking regions with only some small nucleotide changes evident (Figure 4B). Copy numbers that do not form an even number (*i.e.*, 7.9) indicate that the final repeat is not complete. Interestingly, when comparing some viruses in which the copy number of the repeat differed, there were a small number of poorly conserved nucleotides present prior to the conserved flanking region. This occurred only at flanking regions to the right of the repeat and this extra nucleotide sequence generally contained partial sequences from the actual repeat. This suggests that this region represents a former repeat copy that has been altered due to recent insertions, deletions, or recombination. Regions flanking repetitive sequences within the second subset of *Ranaviruses* (GIV, SGIV) and the genus *Megalocytivirus* showed almost perfect conservation (Supplementary Figure 1).

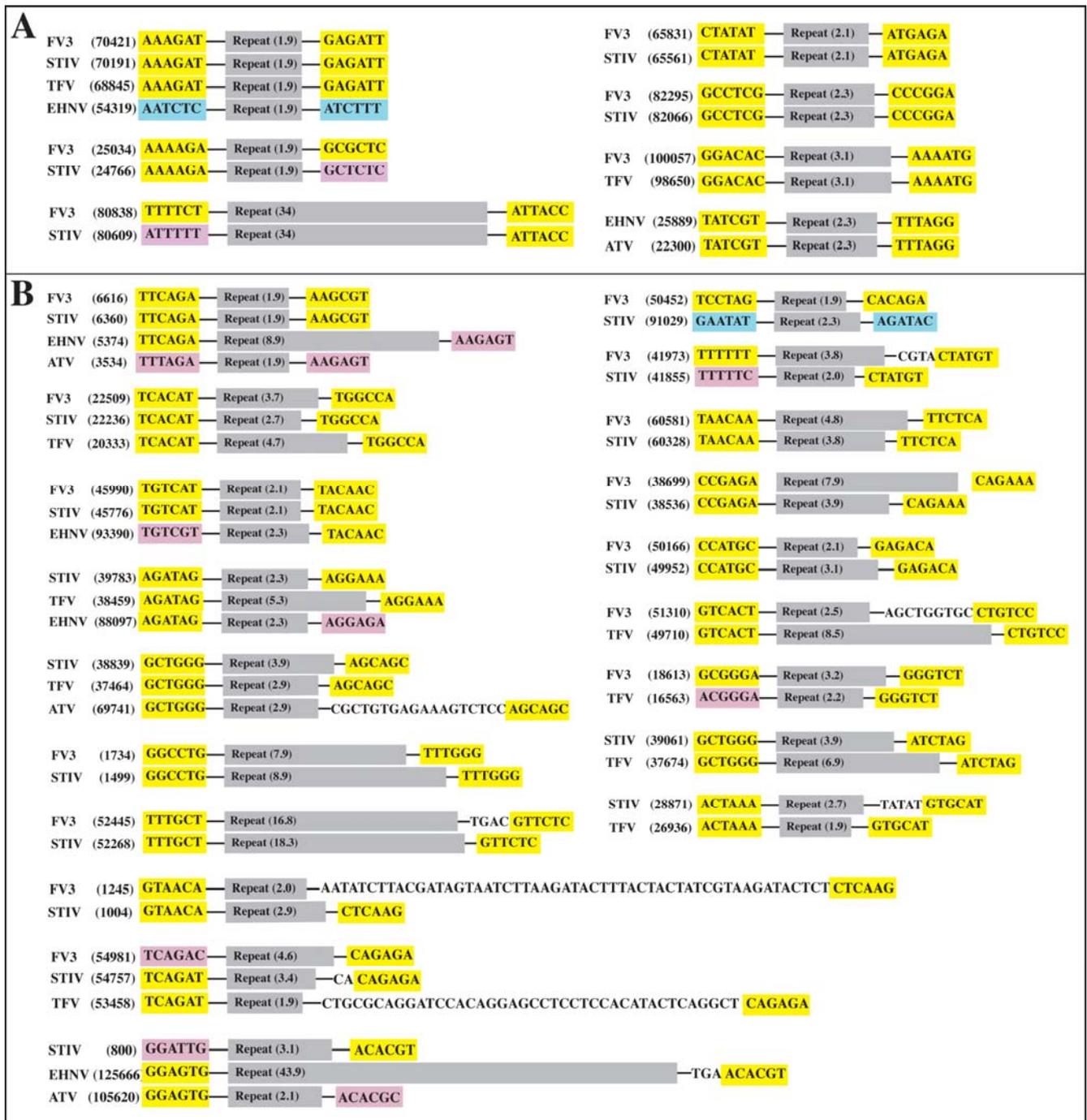
The flanking regions exhibited high levels of sequence conservation as compared to the repetitive sequences. Due to the fact that the repetitive sequences exhibit changes in copy numbers between identical repeats, they create polymorphisms that can easily be detected by PCR using flanking primers.

3. Experimental Section

Phylogenetic analysis: Nucleotide sequences for each gene were obtained for the 14 sequenced iridovirus genomes from the NCBI website [57]. Genes were blasted by BLASTn against each other to find the optimal homology and the ORFs were determined and recorded for the 26 conserved iridovirus genes. The nucleotide sequences of each gene were transferred to the program BioEdit 7.0.5 and aligned individually through ClustalW (multiple alignment). All gaps were striped to ensuring no false divergence conclusions. Each gene alignment was fused when transferred to MEGA4.1 and a neighbor-joining bootstrap consensus trees was constructed using a p-distance model at 500 replicates.

Orthologous gene analysis: Orthologous genes were identified using the Viral Orthologous Cluster (VOC) software [58,59]. Genome sequences were obtained from the VOC database. Orthologous genes identified by the VOC software were confirmed by a BLAST search.

Figure 4. Repeat flanking regions are highly conserved. Regions flanking the repeats common to 2 or more viruses from the first subset of *Ranaviruses* (FV3, STIV, TFV, EHNV, ATV) are shown. Flanking regions exhibiting perfect conservation are shown in yellow. Flanking regions that show high conservation but with one nucleotide change are shown in pink. Flanking regions demonstrating no conservation are shown in blue. Repeats are displayed in gray and the copy number of each repeat is displayed in black. (A) Repetitive sequences that display identical copy numbers and (B) repetitive sequences that display changes in copy numbers are shown.



G/C content analysis: G/C content analysis was completed using the Viral Genome Organizer (VGO) software [35,59]. The sequence for each genome was obtained from the VGO database. The software identified the G/C content of coding regions and this data was used to determine the G/C content of non-coding regions.

Repeat analysis: DNA tandem repeats were identified using a DNA tandem repeat finder [46]. The alignment parameters were 2, 7, 7 for match, mismatch, indel respectively. The minimum alignment score was 50 and the maximum period size was 500. Repetitive sequences with less than 95% match identity are provided but were not included in the analysis (Supplementary Table 7). Genomic sequences were obtained from the NCBI and downloaded in Fasta format [57]. The accession numbers are as follows: FV3 (AY548484), STIV (EU627010), TFV (AF389451), EHNV (FJ433873), ATV (AY150217), GIV (AY666015), SGIV (AY521625), LCDV-1 (L63545), LCDV-C (AY380026), ISKNV (AF371960), RBIV (AY532606), OSGIV (AY894343), IIV-6/CIV (AF303741), IIV-3/MIV (DQ643392). The sequence for RSIV is unavailable and was not included in analysis.

4. Conclusions

By aligning 26 core genes and establishing the shared genes between the complete genomes of 14 iridoviruses, the evolutionary phylogeny of each genera of the *Iridoviridae* family was determined. The phylogeny showed that the tree shared one common ancestor, which then split into two groups consisting of the *Iridovirus* and *Lymphocystivirus* genera and the *Chloriridovirus*, *Megalocytivirus*, and *Ranavirus* genera. The repetitive sequences confirm this phylogenetic relationship. They also demonstrate that much of the difference between viruses with high sequence identity lies within the repeat regions. Repeats both unique to each virus and repeats found within more than one virus were identified and provide a simple and effective method to explore the evolutionary relationship between this family of viruses. The repetitive sequences identified in this paper can be used to compare to newly isolated virus strains in order to find the relationship of that virus to known iridoviruses. Our analysis of the flanking regions suggests that these regions can be used to create primers required to detect changes in copy number between repeats shared between viruses. Once created, these primers will enable a fast and simple method to uniquely identify a specific virus or differentiate between closely related iridoviruses.

Acknowledgements

This work is supported by Discovery Grants (Natural Science and Engineering Research Council (NSERC) of Canada) to C.R.B. H.E.E. is the recipient of a NSERC postgraduate scholarship.

References and Notes

1. He, J.G.; Lu, L.; Deng, M.; He, H.H.; Weng, S.P.; Wang, X.H.; Zhou, S.Y.; Long, Q.X.; Wang, X.Z.; Chan, S.M. Sequence analysis of the complete genome of an iridovirus isolated from the tiger frog. *Virology* **2002**, *292*, 185-197.

2. Jakob, N.J.; Muller, K.; Bahr, U.; Darai, G. Analysis of the first complete DNA sequence of an invertebrate iridovirus: coding strategy of the genome of Chilo iridescent virus. *Virology* **2001**, *286*, 182-196.
3. Williams, T.; Chinchar, G.; Darai, G.; Hyatt, A.; Kalmakoff, J.; Seligg, V.: Family *Iridoviridae*. In *Virus Taxonomy - 7th report of the international committee on taxonomy of viruses*. van Regenmortel, M.H.V., Ed.; Academic Press: New York, U.S.A., 2000; pp. 167-182.
4. Williams, T. The iridoviruses. *Adv. Virus Res.* **1996**, *46*, 345-412.
5. Goorha, R. Frog virus 3 DNA replication occurs in two stages. *J. Virol.* **1982**, *43*, 519-528.
6. Goorha, R.; Murti, G.; Granoff, A.; Tirey, R. Macromolecular synthesis in cells infected by frog virus 3. VIII. The nucleus is a site of frog virus 3 DNA and RNA synthesis. *Virology* **1978**, *84*, 32-50.
7. Goorha, R.; Murti, K.G. The genome of frog virus 3, an animal DNA virus, is circularly permuted and terminally redundant. *Proc. Natl. Acad. Sci. U. S. A.* **1982**, *79*, 248-252.
8. Schnitzler, P.; Soltau, J.B.; Fischer, M.; Reisner, H.; Scholz, J.; Delius, H.; Darai, G. Molecular cloning and physical mapping of the genome of insect iridescent virus type 6: further evidence for circular permutation of the viral genome. *Virology* **1987**, *160*, 66-74.
9. Delius, H.; Darai, G.; Flugel, R.M. DNA Analysis of Insect Iridescent Virus 6: Evidence for Circular Permutation and Terminal Redundancy. *J Virol* **1984**, *49*, 609-614.
10. Ahne, W.; Schlotfeldt, H.J.; Thomsen, I. Fish viruses: isolation of an icosahedral cytoplasmic deoxyribovirus from sheatfish (*Silurus glanis*). *Zentralbl Veterinarmed B* **1989**, *36*, 333-336.
11. Hedrick, R.P.; McDowell, T.S. Properties of iridoviruses from ornamental fish. *Vet. Res.* **1995**, *26*, 423-427.
12. Hedrick, R.P.; McDowell, T.S.; Ahne, W.; Torhy, C.; De Kinkelin, P. Properties of three iridovirus-like agents associated with systemic infections of fish. *Dis. Aquat. Organ.* **1992**, *13*, 203-209.
13. Langdon, J.S.; Humphrey, J.D.; Williams, L.M.; Hyatt, A.D.; Westbury, H.A. First virus isolation from Australian fish: an iridovirus-like pathogen from redfin perch *Perca fluviatilis*. *J. Fish Dis.* **1986**, *9*, 263-268.
14. Pozet, F.; Morand, M.; Moussa, A.; Torhy, C.; De Kinkelin, P. Isolation and preliminary characterization of a pathogenic icosahedral deoxyribovirus from the catfish *Ictalurus melas*. *Dis. Aquat. Organ.* **1992**, *14*, 35-42.
15. Daszak, P.; Cunningham, A.A.; Hyatt, A.D. Infectious disease and amphibian population declines. *Divers. Distrib.* **2003**, *9*, 141-150.
16. Jancovich, J.K.; Davidson, E.W.; Parameswaran, N.; Mao, J.; Chinchar, V.G.; Collins, J.P.; Jacobs, B.L.; Storfer, A. Evidence for emergence of an amphibian iridoviral disease because of human-enhanced spread. *Mol. Ecol.* **2005**, *14*, 213-224.
17. Collins, J.P.; Storfer, A. Global amphibian declines: sorting the hypotheses. *Divers. Distrib.* **2003**, *9*, 89-98.
18. Daszak, P.; Berger, L.; Cunningham, A.A.; Hyatt, A.D.; Green, D.E.; Speare, R. Emerging infectious diseases and amphibian population declines. *Emerg. Infect. Dis.* **1999**, *5*, 735-748.

19. Collins, J.P.; Brunner, J.L.; Miera, V.; Parris, M.J.; Schock, D.M.; Storfer, A.: Ecology and evolution of infectious disease. In *Amphibian Conservation*. Semlitsch, R., Ed.; Smithsonian Institution Press: Washington, U.S.A., 2003; pp. 137-151.
20. Tan, W.G.; Barkman, T.J.; Gregory Chinchar, V.; Essani, K. Comparative genomic analyses of frog virus 3, type species of the genus Ranavirus (family Iridoviridae). *Virology* **2004**, *323*, 70-84.
21. Huang, Y.; Huang, X.; Liu, H.; Gong, J.; Ouyang, Z.; Cui, H.; Cao, J.; Zhao, Y.; Wang, X.; Jiang, Y.; Qin, Q. Complete sequence determination of a novel reptile iridovirus isolated from soft-shelled turtle and evolutionary analysis of Iridoviridae. *BMC Genomics* **2009**, *10*, 224.
22. Jancovich, J.K.; Bremont, M.; Touchman, J.W.; Jacobs, B.L. Evidence for multiple recent host species shifts among the Ranaviruses (family Iridoviridae). *J. Virol.* **2010**, *84*, 2636-2647.
23. Jancovich, J.K.; Mao, J.; Chinchar, V.G.; Wyatt, C.; Case, S.T.; Kumar, S.; Valente, G.; Subramanian, S.; Davidson, E.W.; Collins, J.P.; Jacobs, B.L. Genomic sequence of a ranavirus (family Iridoviridae) associated with salamander mortalities in North America. *Virology* **2003**, *316*, 90-103.
24. Tsai, C.T.; Ting, J.W.; Wu, M.H.; Wu, M.F.; Guo, I.C.; Chang, C.Y. Complete genome sequence of the grouper iridovirus and comparison of genomic organization with those of other iridoviruses. *J. Virol.* **2005**, *79*, 2010-2023.
25. Song, W.J.; Qin, Q.W.; Qiu, J.; Huang, C.H.; Wang, F.; Hew, C.L. Functional genomics analysis of Singapore grouper iridovirus: complete sequence determination and proteomic analysis. *J. Virol.* **2004**, *78*, 12576-12590.
26. Tidona, C.A.; Darai, G. The complete DNA sequence of lymphocystis disease virus. *Virology* **1997**, *230*, 207-216.
27. Zhang, Q.Y.; Xiao, F.; Xie, J.; Li, Z.Q.; Gui, J.F. Complete genome sequence of lymphocystis disease virus isolated from China. *J. Virol.* **2004**, *78*, 6982-6994.
28. He, J.G.; Deng, M.; Weng, S.P.; Li, Z.; Zhou, S.Y.; Long, Q.X.; Wang, X.Z.; Chan, S. Complete Genome Analysis of the Mandarin Fish Infectious Spleen and Kidney Necrosis Iridovirus. *Virology* **2001**, *291*, 126-139.
29. Do, J.W.; Moon, C.H.; Kim, H.J.; Ko, M.S.; Kim, S.B.; Son, J.H.; Kim, J.S.; An, E.J.; Kim, M.K.; Lee, S.K.; Han, M.S.; Cha, S.J.; Park, M.S.; Park, M.A.; Kim, Y.C.; Kim, J.W.; Park, J.W. Complete genomic DNA sequence of rock bream iridovirus. *Virology* **2004**, *325*, 351-363.
30. Kurita, J.; Nakajima, K.; Hirono, I.; Aoki, T. Complete genome sequencing of red sea bream iridovirus. *Fisheries Sci.* **2002**, *68*, 1113-1115.
31. Lu, L.; Zhou, S.Y.; Chen, C.; Weng, S.P.; Chan, S.M.; He, J.G. Complete genome sequence analysis of an iridovirus isolated from the orange-spotted grouper, *Epinephelus coioides*. *Virology* **2005**, *339*, 81-100.
32. Delhon, G.; Tulman, E.R.; Afonso, C.L.; Lu, Z.; Becnel, J.J.; Moser, B.A.; Kutish, G.F.; Rock, D.L. Genome of invertebrate iridescent virus type 3 (mosquito iridescent virus). *J. Virol.* **2006**, *80*, 8439-8449.
33. Wu, G.A.; Jun, S.R.; Sims, G.E.; Kim, S.H. Whole-proteome phylogeny of large dsDNA virus families by an alignment-free method. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12826-12831.

34. Eaton, H.E.; Metcalf, J.; Penny, E.; Tcherepanov, V.; Upton, C.; Brunetti, C.R. Comparative genomic analysis of the family Iridoviridae: re-annotating and defining the core set of iridovirus genes. *Viol. J.* **2007**, *4*, 11.
35. Upton, C.; Hogg, D.; Perrin, D.; Boone, M.; Harris, N.L. Viral genome organizer: a system for analyzing complete viral genomes. *Virus Res.* **2000**, *70*, 55-64.
36. Garcia-Escudero, R.; Vinuela, E. Structure of African swine fever virus late promoters: requirement of a TATA sequence at the initiation region. *J. Virol.* **2000**, *74*, 8176-8182.
37. Moss, B.: Poxviridae: The viruses and their replication. In *Fields Virology*. Fields, B.N., Knipe, D.M., Howley, P.M., Ed.; Lippincott Raven: Philadelphia, U.S.A., 1996; pp. 2637-2671.
38. Nalcacioglu, R.; Marks, H.; Vlak, J.M.; Demirbag, Z.; van Oers, M.M. Promoter analysis of the Chilo iridescent virus DNA polymerase and major capsid protein genes. *Virology* **2003**, *317*, 321-329.
39. Rodriguez, J.M.; Salas, M.L.; Vinuela, E. Intermediate class of mRNAs in African swine fever virus. *J. Virol.* **1996**, *70*, 8584-8589.
40. Levinson, G.; Gutman, G.A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. *Nucleic Acids Res.* **1987**, *15*, 5323-5338.
41. Wittek, R.; Moss, B. Tandem repeats within the inverted terminal repetition of vaccinia virus DNA. *Cell* **1980**, *21*, 277-284.
42. Wadsworth, S.; Jacob, R.J.; Roizman, B. Anatomy of herpes simplex virus DNA. II. Size, composition, and arrangement of inverted terminal repetitions. *J. Virol.* **1975**, *15*, 1487-1497.
43. Arrand, J.R.; Roberts, R.J. The nucleotide sequences at the termini of adenovirus-2 DNA. *J. Mol. Biol.* **1979**, *128*, 577-594.
44. Hayakawa, T.; Ko, R.; Okano, K.; Seong, S.I.; Goto, C.; Maeda, S. Sequence analysis of the Xestia c-nigrum granulovirus genome. *Virology* **1999**, *262*, 277-297.
45. Shoemaker, C.; Goff, S.; Gilboa, E.; Paskind, M.; Mitra, S.W.; Baltimore, D. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. *Proc. Natl. Acad. Sci. U. S. A.* **1980**, *77*, 3932-3936.
46. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573-580.
47. Strand, M.; Prolla, T.A.; Liskay, R.M.; Petes, T.D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **1993**, *365*, 274-276.
48. Levinson, G.; Gutman, G.A. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **1987**, *4*, 203-221.
49. Hancock, J.M. The contribution of slippage-like processes to genome evolution. *J. Mol. Evol.* **1995**, *41*, 1038-1047.
50. van Belkum, A.; Scherer, S.; van Alphen, L.; Verbrugh, H. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 275-293.
51. Zaia, J.A.; Gallez-Hawkins, G.; Churchill, M.A.; Morton-Blackshire, A.; Pande, H.; Adler, S.P.; Schmidt, G.M.; Forman, S.J. Comparative analysis of human cytomegalovirus a-sequence in multiple clinical isolates by using polymerase chain reaction and restriction fragment length polymorphism assays. *J. Clin. Microbiol.* **1990**, *28*, 2602-2607.

52. Jackson, P.J.; Walthers, E.A.; Kalif, A.S.; Richmond, K.L.; Adair, D.M.; Hill, K.K.; Kuske, C.R.; Andersen, G.L.; Wilson, K.H.; Hugh-Jones, M.; Keim, P. Characterization of the variable-number tandem repeats in *vrrA* from different *Bacillus anthracis* isolates. *Appl. Environ. Microbiol.* **1997**, *63*, 1400-1405.
53. Glew, M.D.; Baseggio, N.; Markham, P.F.; Browning, G.F.; Walker, I.D. Expression of the pMGA genes of *Mycoplasma gallisepticum* is controlled by variation in the GAA trinucleotide repeat lengths within the 5' noncoding regions. *Infect. Immun.* **1998**, *66*, 5833-5841.
54. Field, D.; Wills, C. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 1647-1652.
55. Saunders, N.J.; Peden, J.F.; Hood, D.W.; Moxon, E.R. Simple sequence repeats in the *Helicobacter pylori* genome. *Mol. Microbiol.* **1998**, *27*, 1091-1098.
56. Hyatt, A.D.; Gould, A.R.; Zupanovic, Z.; Cunningham, A.A.; Hengstberger, S.; Whittington, R.J.; Kattenbelt, J.; Coupar, B.E. Comparative studies of piscine and amphibian iridoviruses. *Arch. Virol.* **2000**, *145*, 301-331.
57. National Center for Biotechnology Information Homepage. <http://www.ncbi.nlm.nih.gov/> (accessed 3 April 2010).
58. Ehlers, A.; Osborne, J.; Slack, S.; Roper, R.L.; Upton, C. Poxvirus Orthologous Clusters (POCs). *Bioinformatics* **2002**, *18*, 1544-1545.
59. Viral Bioinformatics Resource Center Homepage. www.virology.ca (accessed 26 March 2010).

© 2010 by the authors; licensee MDPI, Basel, Switzerland. This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).