Genomic characterization of a novel iridovirus from redclaw crayfish \textit{Cherax quadricarinatus}: evidence for a new genus within the family \textit{Iridoviridae}

Fang Li, Limei Xu and Feng Yang

Abstract

A novel iridovirus, \textit{Cherax quadricarinatus} iridovirus (CQIV), was identified from diseased \textit{C. quadricarinatus} in 2014. This virus is considered as a new threat to crustacean aquaculture because it is lethal to both penaeid shrimp and crayfish. Here, we determined the complete genome sequence of CQIV. The double-stranded DNA genome is 165,695 bp in length with a G+C content of 34.6%. A total of 178 open reading frames (ORFs) have been predicted, encoding hypothetical proteins ranging from 50 to 1327 amino acids. Forty-seven of these exhibit similarities to proteins of known functions. Phylogenetic analysis based on multiple alignments of conserved proteins shows that CQIV clusters with the members of the family \textit{Iridoviridae}, but is placed in a distinct clade from all the five known genera. It indicates that CQIV may represent a new genus in the family \textit{Iridoviridae}, for which we propose the name \textit{Cheraxvirus} based on the host organism.

Iridoviruses (IVs) are large nucleocytoplasmic DNA viruses that are icosahedral in shape, and about 120–200 nm in diameter. IVs have been isolated from cold-blooded vertebrates including fishes, amphibians and reptiles, and various invertebrates such as insects, arachnids, cephalopods, crustaceans, molluscs, nematodes and polychaetes [1, 2]. The family \textit{Iridoviridae} is currently divided into five genera, two of which (\textit{Iridivirus} and \textit{Chloriridivirus}) include invertebrate-infecting members, whereas the other three (\textit{Rana-virus}, \textit{Lymphocystivirus} and \textit{Megalocytivirus}) represent viruses that infect only poikilothermic vertebrates [3].

All viruses within the family possess linear double-stranded DNA (dsDNA) genomes with circular permutation and terminal redundancy, and the replication of the viral genomes comprises separated nuclear and cytoplasmic phases [1, 3]. The size of IV genomes ranges from 102 to 220 kb (unique portion or non-redundant portion), encoding more than 92 open reading frames (ORFs) [3–5]. Complete genome sequences have been determined for viruses representing all five genera of the family [4–20]. Analysis of sequenced IV genomes identified 26 conserved core genes shared by all IVs [10, 21], which help to infer the phylogenetic relationship among the species.

IVs are considered as members of the nucleocytoplasmic large DNA viruses (NCLDVs), a monophyletic clade of giant viruses that includes the families \textit{Poxviridae}, \textit{Phycodnaviridae}, \textit{Asfarviridae} and \textit{Ascoviridae}, as well as the newly defined families \textit{Mimiviridae} and \textit{Marseilleviridae} isolated from amoeba [22–25]. The clade shares a total of 47 putative common ancestral genes encoding structural components and proteins participating in DNA packaging, replication and transcription [24, 25]. The 47 common ancestral genes are present in at least one species of each of the seven families, but are not shared by all genera/species.

Vertebrate iridoviruses (VIVs) are major pathogens affecting fish, amphibian and reptile aquaculture. They can lead to considerable morbidity and mortality in the animals concerned [26–30]. On the contrary, invertebrate iridoviruses (IIVs) usually cause either patent or covert infection in insects. Patent infections are often fatal in the larval or pupal stages of some insects, whereas covert infections are not lethal [31–35]. Some IVs from fully aquatic invertebrates

Received 13 June 2017; Accepted 27 July 2017

Author affiliations: 1State Key Laboratory Breeding Base of Marine Genetic Resources; Fujian Key Laboratory of Marine Genetic Resources; Fujian Collaborative Innovation Center for Exploitation and Utilization of Marine Biological Resources; Key Laboratory of Marine Genetic Resources of StateOceanic Administration, Third Institute of Oceanography, SOA, Xiamen 361005, PR China; 2Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Qingdao, PR China.

*Correspondence: Fang Li, lifang@tio.org.cn; Feng Yang, yangfeng@tio.org.cn

Keywords: \textit{Cherax quadricarinatus} iridovirus (CQIV); genome sequence; new genus; iridovirus.

Abbreviations: CQIV, \textit{Cherax quadricarinatus} iridovirus; IV, iridovirus; IIV, invertebrate iridovirus; NCLDV, nucleocytoplasmic large DNA virus; ORF, open reading frame; VIV, vertebrate iridovirus.

The GenBank accession number for the genomic sequence of CQIV is MF197913.

Three supplementary tables are available with the online Supplementary Material.
Fig. 1. Linear map of the CQIV genome. ORFs and their transcription directions are indicated with arrows. Green arrows represent genes involved in DNA replication, modification and processing. Yellow, blue and pink arrows represent genes involved in nucleic acid metabolism, gene transcription and protein processing/modification, respectively. Grey arrows represent genes with other functions. ORFs with homology to proteins of unknown function are in black, while ORFs with no homology to genes in the databases are in white.
have also been reported [36–41], but their relationships to other IVs are unknown due to the lack of biological and genetic information.

A new crustacean IV, *Cherax quadricarinatus* iridovirus (CQIV), was isolated from diseased redclaw crayfish in China in 2014 [42]. This virus has now received considerable attention, since its infection is lethal to crayfish and penaeid shrimp. A preliminary phylogenetic analysis based on a small fragment of the major capsid protein suggested that CQIV was distantly related to the five known genera of the family *Iridoviridae*.

In this study, we determined the complete genome sequence of CQIV to illustrate its gene content, and to refine its phylogenetic relationship to known viruses.

**OVERALL GENOME STRUCTURE OF CQIV**

CQIV was purified from diseased redclaw crayfish, *C. quadricarinatus*, collected in Fujian, China. The viral DNA was prepared from purified virions as described previously [42]. The viral genome was sequenced using 454 sequencing technology by Shanghai Majorbio Bio-pharm Biotechnology Co., Ltd., and assembled into a 165 695 bp linear molecule (GenBank accession number MF197913), with a G+C content of 34.6 % using the GS *de novo* assembler software (Version 2.8). Based on the available genomic information of IVs, the genomes of CQIV are smaller than those of the species in the genera *Iridovirus* (~200 kb), *Chloriridovirus* (~200 kb), and some members of the genus *Lymphocystivirus* (~200 kb), but are larger than the genomes of species
belong to the genera *Ranavirus* (105–140 kb) and *Megalocytivirus* (~110 kb). Moreover, the G+C content of CQIV is similar to those of the species belonging to the genera *Iridovirus*, *Chloriridovirus* and *Lymphocystivirus* (28–35%).

A total of 178 putative ORFs encoding hypothetical proteins ranging from 50 to 1327 amino acids (Table S1, available in the online Supplementary Material) were identified in the CQIV genome using Geneious Pro 9.1.5, corresponding to a

![Fig. 3. Phylogenetic analysis of CQIV. Phylogenetic relationships of 25 conserved proteins from 16 completely sequenced IV genomes (a), and six conserved proteins from 21 completely sequenced NCLDV genomes (b), were analysed using the maximal likelihood method. Bootstrap percentages for 500 replicates are shown for each node. Branch lengths were also estimated, and the scale bar denotes the number of substitutions per site.](image-url)
theoretical coding density of 91.9%. These putative protein coding genes were consecutively numbered starting from the conserved IV major capsid protein gene (ORF001R), where right (R) or left (L) refers to the orientations of the ORFs. Functional annotation was carried out by a PSI-BLAST search against the non-redundant database of NCBI. The location, orientation, size and possible function of each predicted protein coding gene are shown in Table S1 and Fig. 1.

**GENE HOMOLOGY BETWEEN CQIV AND OTHER IVS**

Among the 178 putative ORFs, 90 ORFs have orthologues in IVs, 17 ORFs have orthologues in other organisms including viruses, prokaryotic and eukaryotic species, whereas 71 ORFs show no homology with genes in the databases (Fig. 2). Among the 90 homologous genes of IVs, 40 genes have orthologues exclusively in IIVs and 10 genes have orthologues exclusively in IVs. The remaining 40 genes have orthologues in species of both IIVs and IVs, including 25 core genes common in all genera of the family *Iridoviridae* [21], and 9 core genes present in all families of the NCLDVs [24, 25] (Fig. 2). The ribonucleotide reductase small subunit gene, previously known as a core gene of IVs, was absent in CQIV, so that it should no longer be considered an IV core gene. Notably, the putative gene products of CQIV ORFs share relatively low homologies with their orthologues in IVs. Among the 90 hypothetical proteins homologous to proteins of other IVs, only nine exhibit ≥50% identity with their orthologues, whereas 30 share <30% identities with their orthologues (Table S1). This finding implies that CQIV is not closely related to other known IVs.

**FUNCTIONAL PREDICTION OF CQIV ORFS**

The putative products of 47 CQIV ORFs show significant homologies with functional characterized proteins. These include a set of proteins involved in DNA replication, repair and nucleotide metabolism, such as DNA polymerase (ORF042L), DNA helicase (ORF004L, 048L, 082L, 088L and 170R), DNA primase (ORF037L), DNA topoisomerase (ORF075L and 150R), proliferating cell nuclear antigen (ORF114L), deoxyuridine 5'-triphosphate nucleotidohydrolase (ORF012L), thymidylate kinase (ORF055L and 118R), ribonuclease (ORF107L and 127L), DNA ligase (ORF087L), DNA endonuclease (ORF125R) and DNA exonuclease (ORF133L, 110R and 116L). CQIV also encodes its own DNA endonuclease (ORF125R) and DNA exonuclease (ORF133L, 110R and 116L). CQIV also encodes its own DNA endonuclease (ORF125R) and DNA exonuclease (ORF133L, 110R and 116L). CQIV also encodes its own RNA polymerase (Rpb) subunits, Rpb1 (ORF097L), Rpb2 (ORF083L), Rpb5 (ORF065L) and Rpb7 (ORF134L), a transcription elongation factor TFIIS (ORF063L) and a potential transcription factor (ORF154R). Interestingly, the RNA polymerase subunit Rpb7 has not previously been found in any other NCLDV. Moreover, there are proteins potentially involved in protein modification/procession, including a RING finger protein (ORF014L) and an ubiquitin (ORF078R) participating in protein ubiquitination, as well as serine/threonine kinases (ORF057R, 061L, 117R, 139R and 144L) and a phosphatase (ORF038R) involved in protein phosphorylation (Table S1 and Fig. 1).

**CQIV REPRESENTS A NEW GENUS OF THE FAMILY IRIDOVIＲＩДАＥ**

In our previous study, a preliminary phylogenetic study based a short fragment of the major capsid protein (ORF001) suggested an independent branching of CQIV in the phylogenetic tree of IVs. However, due to limited genetic information, this tree is poorly supported [42]. To refine the phylogeny of CQIV within the family *Iridoviridae*, a maximal likelihood phylogenetic analysis was carried out using the concatenated sequences of 25 core proteins of IVs (Table S2). Each protein was aligned separately using Muscle Web Server (http://www.ebi.ac.uk/Tools/msa/muscle/) [43]. The alignments were trimmed using TrimAl 1.3 (Automated 1 method with default parameters) (http://phy lemon2.bioinfo.cif.es/utilities.html) [44] to remove less conserved positions. The trimmed alignments of the 25 conserved proteins were then concatenated using Mesquite (http://mesquiteproject.wikispaces.com/) [45]. An unrooted maximal likelihood tree was generated using PhyML 3.0 [46] combined with Smart Model Selection (SMS) [47] (http://www.atgc-montpellier.fr/phyml-sms/). The LG+G+I+F model was identified as the optimal for maximal likelihood analysis in this case. A bootstrap test was carried out with 500 replicates. The results show that CQIV is located on a distinct branch in the unrooted tree, independent of the five known genera of IVs, *Iridovirus*, *Chloriridovirus*, *Chloriridivirus*, *Ranavirus*, *Lymphocystivirus* and *Megalocytivirus* (Fig. 3a). The separated branch of CQIV is highly supported by a bootstrap value of 100, suggesting that CQIV has an ancestral relation with other IVs but only distantly related to the five known genera. Moreover, it is notable that IIV-22, IIV-30 and IIV-9, which belong to the genus *Iridovirus*, cluster in a strongly supported branch with IIV-3, the type species of the genus *Chlorovirus*. This is concordant with previous findings [9, 11].

IVs are considered as members of NCLDVs, which include the families *Poxviridae*, *Phycodnaviridae*, *Asfarviridae*, *Ascoviridae*, *Mimiviridae* and *Marseilleviridae* [24, 25]. Therefore, the phylogenetic relationship of CQIV to 20 known NCLDV's was further investigated using PhyML 3.0 with SMS as described above. Although CQIV contains orthologues of nine NCLDV core proteins, only six of these are shared by all the species analysed in this study. Therefore, the six core proteins of NCLDV's were used in the phylogenetic analysis (Table S3). A strongly supported clustering of CQIV with the type species of the five genera of IVs was observed, confirming that CQIV is a member of the family *Iridoviridae*, although it is distantly related to known IVs. In addition, CQIV was placed on a branch rooted with species of the families *Ranavirus*, *Lymphocystivirus* and *Megalocytivirus*, suggesting a common ancestor for CQIV and VIVs (Fig. 3b). Moreover, the two genera of the family *Ascoviridae* cluster with IVs as previously described [48].
In conclusion, here we determined the complete genomic sequence of CQIV and refined its phylogeny. Together with its morphological features and mode of replication [42], we demonstrate that CQIV is a new member of IV. It may represent a previously uncharacterized genus of the family Iridoviridae, for which we propose the name Cheravirus based on the host organism. These findings broaden our knowledge of IVs, and provide useful information for the research of this new virus.

**Funding information**
This work was supported by National Natural Science Foundation of China (No. 31672691), China Agriculture Research System (No. CARS-48).

**Conflicts of interest**
The authors declare that they have no conflicts of interest.

**References**


---

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as ‘excellent’ or ‘very good’.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.