Virology 521 (2018) 92–98

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

Host subspecific viral strains in European house mice: Murine cytomegalovirus in the Eastern (*Mus musculus musculus*) and Western house mouse (*Mus musculus domesticus*)

Dagmar Čížková[a,*], Stuart J.E. Baird[a], Jana Těšíková[a,b], Sebastian Voigt[c,d], Ďureje Ľudovít[a], Jaroslav Piálek[a], Joëlle Goüy de Bellocq[a]

[a] Institute of Vertebrate Biology of the Czech Academy of Sciences, Brno, Czech Republic
[b] Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
[c] Department of Pediatric Oncology/Hematology/Stem Cell Transplantation, Charité-Universitätsmedizin, Berlin, Germany
[d] Department of Infectious Diseases, Robert Koch Institute, Berlin, Germany

ARTICLE INFO

Keywords:
Murine cytomegalovirus
House mouse
Mus musculus musculus
Host-specificity
Introgression
Hybrid zone

ABSTRACT

Murine cytomegalovirus (MCMV) has been reported from house mice (*Mus musculus*) worldwide, but only recently from Eastern house mice (*M. m. musculus*), of particular interest because they form a semi-permeable species barrier in Europe with Western house mice, *M. m. domesticus*. Here we report genome sequences of EastMCMV (from Eastern mice), and set these in the context of MCMV genomes from genus *Mus* hosts. We show EastMCMV and WestMCMV are genetically distinct. Phylogeny splitting analyses show a genome wide (94%) pattern consistent with no West-East introgression, the major exception (3.8%) being a genome-terminal region of duplicated genes involved in host immune system evasion. As expected from its function, this is a region of maintenance of ancestral polymorphism: The lack of clear splitting signal cannot be interpreted as evidence of introgression. The EastMCMV genome sequences reported here can therefore serve as a well-described resource for exploration of murid MCMV diversity.

1. Introduction

Murine cytomegalovirus (MCMV) gained biomedical attention as a laboratory mouse model of the closely related betaherpesvirus Human cytomegalovirus (HCMV), which can cause congenital viral infection and severe disease in immunocompromised patients (reviewed in Boeckh and Geballe, 2011). MCMV infection has been reported in wild house mice (*Mus musculus*) worldwide with a prevalence of 60–90% (reviewed in Shellam et al., 2007, Becker et al., 2007), and these studies, together with functional assays in laboratory mice, provide good genomic and functional annotation resources.

The viral genome is linear double stranded DNA of about 230 kb in length, with approximately 180 predicted genes (Rawlinson et al., 1996; Smith et al., 2008) of which 78 show sequence homology with HCMV genes. Most of these have confirmed transcription or expression and some have been functionally characterized using mutant strains (Tang et al., 2006; Kattenhorn et al., 2004; Lacaze et al., 2011; Lisnic et al., 2013). Across viruses, betaherpesviruses such as MCMV have relatively high gene count, and this is thought due to their complex life cycle which includes tropism to specific organs during acute infection, long-term latency in salivary glands and periodic re-activation (Hudson, 1979). It is therefore perhaps unsurprising that the majority of the genes do not code for structural proteins but rather various regulatory elements implicated in life cycle control and evasion of the host immune system. MCMV genomes examined so far are conserved in structure and their sequence divergence is relatively low, with highest diversity found at the genome extremities, which harbour genes coding immunoevasins. The MCMV genomes seem to evolve mostly under purifying selection and recombination plays an important role in shaping MCMV genetic variability (Smith et al., 2013).

Recent research on arenaviruses, whose natural hosts are also rodents, suggests viral strains may be unable to spread across species range boundaries dividing subspecific taxa (Gryseels et al., 2017). This finding is of special interest for MCMV, which infects two subspecies of the house mouse that meet along a semi-permeable contact zone. The house mouse hybrid zone (HMHZ) stretches 2500 km north to south in Europe, dividing the Eastern and Western house mice (*M. m. musculus* and *M. m. domesticus*, respectively) (Boursot et al., 1993, Jones et al., 2010, Dureje et al., 2012). MCMV has only recently been looked for in the Eastern house mouse: Sanger sequencing of the M94 gene indicated

* Correspondence to: Institute of Vertebrate Biology of the Czech Academy of Sciences, Research Facility Studenec, Studenec 122, 675 02 Koněšín, Czech Republic.
E-mail address: dejsha@seznam.cz (D. Čížková).

https://doi.org/10.1016/j.virol.2018.05.023
Received 31 January 2018; Received in revised form 27 May 2018; Accepted 28 May 2018

that the two mouse subspecies carry divergent MCMV lineages that have codiverged with their hosts (Goüy de Bellocq et al., 2015).

Here we use high throughput sequencing to gain a genome-wide picture of 'EastMCMV' and set that in the context of MCMV variation in the genus *Mus*. The Western and Eastern house mice originated in South Central Asia 0.35–0.79 million years ago (Geraldes et al., 2011; Duvaux et al., 2011), they are synanthropic, and it is thought they colonised Europe after the agricultural revolution in two directions around the Mediterranean and Black Sea (Cucchi et al., 2012). Their sister taxon, the Steppe mouse *Mus spicilegus,* by contrast is not associated with humans and the 3-way most common recent ancestor lived approximately 1.1 – 2 million years ago (Macholán et al., 2012).

2. Methods

2.1. Sampling design

To characterize East- and West-MCMVs from Europe, we chose samples from a collection of dried house mouse saliva and frozen salivary glands collected in 2014 at localities inhabited by genetically 'pure' Eastern and Western house mice as measured by their hybrid indexes (HIs). HIs are estimates of the proportion of Eastern vs. Western mouse genome; individuals of very low or very high HI are categorised as 'pure' Western and Eastern mouse, respectively. At the same time, these localities were situated in close proximity to the house mouse hybrid zone (HMHZ), allowing for potential introgression of the virus across the HMHZ. We selected 20 and 10 mouse saliva samples from Eastern and Western localities for direct sequencing. We selected 10 Eastern and 5 Western salivary glands for virus isolation in cell culture, out of which we sequenced 2 East and 2 WestMCMV isolates. For the analyses in this report we used the four samples which resulted in sufficient MCMV sequencing reads and high-quality genome assemblies (i.e. a low number of contigs spanning the entire Smith MCMV reference). These were directly sequenced saliva samples s88_sk2733_KRASIK (hereafter s88), s02_sk2676_UNWE (s02) and s17_sk2742_SEYB (s17) and one cell-culture isolate s45_sk2703_VER15 (s45). The two Eastern localities KRASIK and VER15 ([49° 52′ 52"N, 12° 56′ 6″E and 50° 8′ 31″N, 13° 10′ 35″E, respectively], Czech Republic) are 27.5 and 48.9 km east of the estimated hybrid zone centre, while Western localities UNWE and SEYB ([50° 9′ 25″N, 12° 6′ 12″E and 49° 52′ 57″N, 11° 43′ 12″E, respectively], Germany) are 21.4 and 50.9 km west of contact zone centre. Mice from these localities, which were genotyped using the 630000-SNP Mouse Diversity Array (MDA, Yang et al., 2009) had HIs of at least 97.7% (KRASIK) and 98.7% (VER15) and at most 3.6% (UNWE) and 0.8% (SEYB), indicating that these localities are inhabited by 'pure' subspecies. The hosts of samples s88, s45, s02 and s17 were genotyped using 14 mitochondrial, sex-linked and autosomal markers fixed for alternative alleles in the Western and Eastern mice (Dureje et al., 2012; Macholán et al., 2011) and these more approximate estimates confirmed low genetic 'contamination' with the other subspecies i.e. HIs 92% and 100% and 10% and 8%. As an outgroup within the genus *Mus*, we sampled MCMV from a Steppe mouse s09_sk3086_spicSkDri1A, trapped in Drienovec (48° 37' 8"N, 20° 55' 50"E, Slovakia), approximately 610 km from the HMHZ. Its MCMV was sequenced from dried saliva.

2.2. Isolation of MCMV in cell culture

We attempted to isolate 15 East- and West-MCMVs in cell culture. First, a small piece of salivary gland was placed on monolayers of mouse fibroblast 3T3 cells. After 1 h incubation the cells were maintained in Dulbecco´s Modified Eagle Medium (DMEM, ThermoFisher Scientific) supplemented with 2% fetal calf serum and 100 µg/ml Penicillin-Streptomycin (ThermoFisher Scientific) at 37 °C in CO_2 incubator. In the following weeks, supernatants from infected cell cultures were repeatedly harvested (once a cytopathic effect was clearly visible) and used to infect a new cell cultures to maximize virus multiplication. Virus particles present in the final supernatant (i.e. after 5 passages; without plaque purification) of 13 successfully reactivated viral samples were concentrated by 3 h centrifugation at 27,000 × *g* at 4 °C.

2.3. Sequencing and assembly

DNAs from the four viruses in cell culture were extracted from 100 µl of supernatant using JetQuick Tissue DNA Miniprep Kit (Genomed). For the remaining 30 samples we used mouse saliva dried on filter paper, which was extracted with DNeasy Blood and Tissue Kit (Qiagen). DNA was fragmented with Covaris to 300 bp, sequencing libraries were prepared using the Kapa HyperPrep kit (Kapa Biosystems) with custom sequencing adapters. Each library was first sequenced on 1/96- and, with the exception of s45, subsequently resequenced on 1/72 of a Nextseq High-Output run (Illumina) and 1/72 of a Nextseq Mid-Output run (Illumina) with 2 × 150bp paired end reads. The library preparations and sequencing were performed at the Core Genomics Facility, CEITEC, Brno, Czech Republic.

Remains of sequencing adapters were trimmed from the sequencing reads from each library, as were low quality bases and polyG stretches, using BBMap 36.76 in BBTools (https://jgi.doe.gov/data-and-tools/bbtools/). To isolate viral genomes from these metagenomic samples (i.e. mixtures of viral, bacterial and host DNA), we used a simplification of the "Blobology" approach described by Kumar et al. (2013). Importantly, this approach is based on de-novo assembly of metagenomes and therefore does not suffer from reference-mapping biases. In brief, we performed de-novo assembly of sequencing reads with Spades v3.9.0 (Nurk et al., 2013, 2017) (without any error correction), with varying kmer lengths and both with and without the 'meta' option. Taxonomic assignment and GC content were plotted for the resulting contigs and MCMV contigs identified using the Blobtools package (Laetsch and Blaxter, 2017; Laetsch et al., 2017). In the case of more than one MCMV contig in a sample, the set of contigs was aligned to the Smith reference to find their order and orientation and contig ends were edited if overlap was detected. The sequencing reads were mapped back to the de-novo assembled genomes with Geneious mapper (v9.1.4, Biomatters) and SNPs were called on the mapped reads with Geneious SNP caller to check for potential presence of multiple infections. The genomes were annotated using RATT (Otto et al., 2011) by transferring annotations of coding regions from the Smith reference. Before the transfer, the Smith annotation was updated based on Smith et al. (2008) and by including some splice variants and novel coding regions based on the current literature (mainly Lacaze et al., 2011, Lisnic et al., 2013).

2.4. Sequence analyses

Genomes of 12 MCMV strains are publically available. Two (Smith and K181) are highly passaged strains derived in 1954 and 1980 in US from laboratory mouse (Smith, 1954; Misra and Hudson, 1980) and WT1 is probably a hybrid between these two strains (Smith et al., 2013). The rest are low-passage wild-derived strains from mice trapped in Australia, and nearby islands or at Macquarie Island (Booth et al., 1993; Smith et al., 2008, 2013). The laboratory mouse was derived mostly from Western mouse stock (Yang et al., 2011) and the mice introduced to the Australian territory including subantarctic Macquarie Island (Gabriel et al., 2011; Hardouin et al., 2010) were also Western. For phylogeny splitting analyses described below we used the complete dataset of 2 European East MCMVs, 2 European WestMCMVs and 12 already available WestMCMVs. For the sake of clarity, the phylogenetic trees and network only included 6 of the 12 already available WestMCMVs that span the geographic range of the strains available.

Whole MCMV genomes were aligned with MAFFT online service (Mafft v7) using the FFT-NS-i algorithm (Katoh et al., 2017). To visualise evolutionary relationships between the genomes, we