Virology 438 (2013) 14–19

Contents lists available at SciVerse ScienceDirect

Virology

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

Sequencing approach to analyze the role of quasispecies for classical swine fever

Armin Töpfer [a,b], Dirk Höper [c], Sandra Blome [c], Martin Beer [c], Niko Beerenwinkel [a,b], Nicolas Ruggli [d], Immanuel Leifer [d,*]

[a] Department of Biosystems Science and Engineering, ETH Zurich, Mattenstrasse 26, CH-4058 Basel, Switzerland
[b] SIB Swiss Institute of Bioinformatics, CH-4058 Basel, Switzerland
[c] Institute of Diagnostic Virology, Friedrich-Loeffler-Institut, Südufer 10, 17493 Greifswald-Insel Riems, Germany
[d] Institute of Virology and Immunoprophylaxis (IVI), Sensemattstrasse 293, CH-3147 Mittelhäusern, Switzerland

ARTICLE INFO

Article history:
Received 8 October 2012
Accepted 28 November 2012
Available online 13 February 2013

Keywords:
Classical swine fever virus
Pestivirus
Virulence
Quasispecies estimation
Haplotype reconstruction
ShoRAH
QuRe
QuasiRecomb

ABSTRACT

Classical swine fever virus (CSFV) is a positive-sense RNA virus with a high degree of genetic variability among isolates. High diversity is also found in virulence, with strains covering the complete spectrum from avirulent to highly virulent. The underlying genetic determinants are far from being understood. Since RNA polymerases of RNA viruses lack any proof-reading activity, different genome variations called haplotypes, occur during replication. A set of haplotypes is referred to as a viral quasispecies. Genetic variability can be a fitness advantage through facilitating of a more effective escape from the host immune response. In order to investigate the correlation of quasispecies composition and virulence *in vivo*, we analyzed next-generation sequencing data of CSFV isolates of varying virulence. Viral samples from pigs infected with the highly virulent isolates "Koslov" and "Brescia" showed higher quasispecies diversity and more nucleotide variability, compared to samples of pigs infected with low and moderately virulent isolates.

Introduction

Classical swine fever (CSF) is an important contagious disease of pigs causing major economical damage in the pig industries (Edwards, Fukusho et al., 2000;Vandeputte and Chappuis, 1999). The causative agent is the classical swine fever virus (CSFV), a member of the genus *Pestivirus* grouped into the family *Flaviviridae* (Fauquet and Fargette, 2005, King, Lefkowitz et al., 2011). *Pestiviruses* possess a single-stranded positive-sense RNA genome of approximately 12,300 nucleotides, with 5′-terminal and 3′-terminal non-translated regions (5′-NTR and 3′-NTR, respectively) (Meyers and Thiel, 1996). The genome contains one open reading frame encoding a polyprotein that is processed by cellular and viral proteases (Meyers and Thiel, 1996) to the structural (C, E^{rns}, E1, and E2) and the non-structural proteins (N^{pro}, p7, NS2, NS3, NS4A, NS4B, NS5A, and NS5B) (Rumenapf, Unger et al., 1993;Tautz, Elbers et al., 1997;Thiel, Stark et al., 1991).

The envelope glycoprotein E2 is highly immunogenic and essential for replication (Van Gennip, Bouma et al., 2002). Moreover, it was shown that it plays a role in viral adsorption to host cells together with other surface proteins, namely E^{rns} and E1 (Hulst and Moormann, 1997;Wang, Nie et al., 2004, Reimann, Depner et al., 2004). The E2 protein shows a high variability among the different CSFV isolates. Despite of cross neutralizing antibodies, an E2 escape from the host immune response seems possible, as it was demonstrated previously by *in vitro* studies (Leifer, Blome et al., 2012). The NS5B protein is the RNA-dependent RNA polymerase (RdRp) and it is more conserved among CSFV isolates (Bjorklund, Lowings et al., 1999). The E2 and NS5B encoding regions were selected for quasispecies analyses to investigate differences between conserved and variable genome regions.

CSF disease can vary from acute hemorrhagic fever to chronic or unapparent infection, which is dependent essentially on the virulence of the viral isolate. Accordingly, CSFV can be classified in highly virulent, moderately to low virulent, and avirulent strains, the latter being mainly vaccine strains (Floegel-Niesmann, Blome et al., 2009;Li, Xu et al., 2006;Mittelholzer, Moser et al., 2000;Tao, Dai et al., 2009). A number of studies have been undertaken to

* Corresponding author. Fax: +41 31 848 9271.
E-mail addresses: Armin.Toepfer@bsse.ethz.ch (A. Töpfer), Dirk.Hoeper@fli.bund.de (D. Höper), Sandra.Blome@fli.bund.de (S. Blome), Martin.Beer@fli.bund.de (M. Beer), Niko.Beerenwinkel@bsse.ethz.ch (N. Beerenwinkel), Nicolas.Ruggli@ivi.admin.ch (N. Ruggli), Immanuelleifer@hotmail.com, Immanuel.Leifer@ivi.admin.ch (I. Leifer).

0042-6822/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.virol.2012.11.020

investigate genetic markers that may account for the observed differences in CSFV virulence (Risatti, Borca et al., 2005;Risatti, Holinka et al., 2005;Risatti, Holinka et al., 2006;Risatti, Holinka et al., 2007b;Risatti, Holinka et al., 2007a;Sainz, Holinka et al., 2008;Tamura, Sakoda et al., 2012;Van Gennip, Vlot et al., 2004). Despite all these efforts, the viral determinants of virulence are not well understood yet.

Because the RdRp of RNA viruses lack any proof-reading activity, their error rate during genome replication is high. As long as the random mutations are not self-limiting, viruses with genetic differences are maintained in the same host during viral replication. This phenomenon was first described in 1971 and defined as quasispecies (Eigen, 1971). High genome diversity can be a fitness advantage through enabling viral escape from the immune response or from antiviral drug treatment (Thiel, Peters et al., 2002). The influence of viral quasispecies on pathogenesis was reported for several viruses, for instance picornaviruses, foot-and-mouth disease virus, and West Nile fever virus (Jerzak, Bernard et al., 2007;Sanz-Ramos, Diaz-San et al., 2008;Vignuzzi, Stone et al., 2006). This raises the question if the quasispecies composition of CSFV differs between isolates, and if so, whether these differences can be related to fitness or virulence of selected strains.

At present, almost nothing is reported on quasispecies composition of pestiviruses and the possible relevance for virulence. Therefore, the aim of this study was to analyze the quasispecies composition of RNA samples of CSFV isolates differing in virulence in the E2 and NS5B protein encoding genome regions. Sequence data were generated by next-generation sequencing (NGS). Single-site and local diversity estimates can be obtained reliably from NGS data, and it has been shown that local reconstruction can be a good measure of the global diversity, as the majority of the underlying diversity can often be observed in a single window (Zagordi, Däumer et al., 2012). For quasispecies analysis, three different open-source haplotype reconstruction tools, namely ShoRAH (Zagordi, Bhattacharya et al., 2011), QuRe (Prosperi and Salemi, 2012), and QuasiRecomb (Töpfer et al., 2012) were used. Haplotype reconstruction was performed on three different spatial levels: (i) at single sites of the DNA sequence, (ii) in a sliding window of a 300 bp width, called local reconstruction, and (iii) across the full-length of each of the two proteins, referred to as global reconstruction. The diversity of each inferred quasispecies was quantified as the entropy of the respective mutant distribution.

Results and discussion

In this study, we analyzed the intra-strain genomic diversity and estimated the quasispecies composition of the following five CSFV isolates: "Brescia" and "Koslov" (both highly virulent), and "Uelzen", "Paderborn", and "Hennef" (all three moderately virulent). For reliable quasispecies estimation in the E2 and NS5B protein coding regions, including the detection of low-frequency variants, we used a high sequencing coverage of 5333 ± 3244 (mean $\pm$ sd) reads per sequence position for E2 and 5002 ± 2090 for NS5B. Independent of the sample, the coverage varied greatly within the sequenced regions (Fig. 1). During library generation and emulsion PCR, the effectiveness of random primers could have caused these fluctuations, since the primers bind with a different affinity to the corresponding DNA regions. The average read length obtained was 363 bp, except for the isolate "Uelzen" with an average length of 271 bp.

In the E2 and NS5B polyprotein sequence of viable CSFV field isolates, single nucleotide deletions and insertions were never observed. The single-base gaps detected by NGS are likely to reflect sequencing errors. If real, each of these gaps would lead to a frame shift mutation in the polyprotein, which would result in non-functional viral proteins and premature translation stop disabling genome replication due to the *cis* requirement of protein translation for *pestivirus* RNA replication (Frolov, McBride et al., 1998). For CSFV, proteins expressed from alternative reading frames have never been described. Therefore, gaps in the raw sequencing data were not considered for haplotype estimation and quasispecies analysis. Importantly, no insertions or deletions of one or more entire codons have been observed in the data.

As a first level of analysis, Fig. 1 shows, for each isolate, the single-site entropy (*i.e.* diversity) landscape along the E2 and NS5B protein encoding sequences. In general, regions with higher nucleotide variability with clear-cut peaks within E2 and NS5B appear randomly distributed and not restricted to any particular

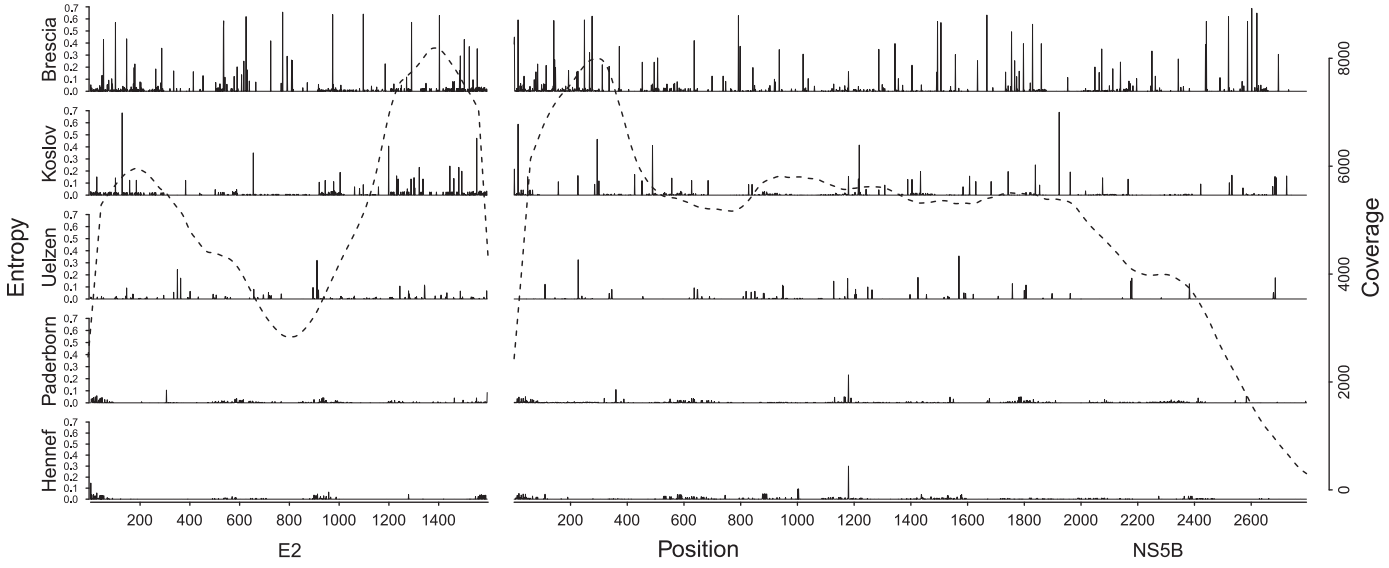

Fig. 1. Position-wise entropy of nucleotide distributions in the E2 and NS5B protein encoding regions. For each strain (Brescia, Koslov, Uelzen, Paderborn, Hennef), the entropy (left axis) computed from error-corrected reads is shown at each single genomic site of the E2 (left) and NS5B (right) encoding sequences. The mean coverage (right axis) of the next-generation sequencing of the five isolates data is plotted as a dotted lined.