Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Author's personal copy

Infection, Genetics and Evolution 12 (2012) 461–466

Contents lists available at SciVerse ScienceDirect

Infection, Genetics and Evolution

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

Norovirus GII.4 and GII.7 capsid sequences undergo positive selection in chronically infected patients

Dieter Hoffmann [a,*], Martin Hutzenthaler [b], Judith Seebach [a], Marcus Panning [d], Andreas Umgelter [e], Helge Menzel [f], Ulrike Protzer [a,c], Dirk Metzler [b]

[a] Institute of Virology, Technische Universität München, Munich, Germany
[b] LMU BioCenter, Department of Evolutionary Biology, Ludwig-Maximilians Universität, Munich, Germany
[c] Institute of Virology, Helmholtz Zentrum München, Munich, Germany
[d] Department of Virology, University of Freiburg, Germany
[e] 2nd Medical Department, Klinikum rechts der Isar, Technische Universität München, Munich, Germany
[f] 3rd Medical Department, Klinikum rechts der Isar, Technische Universität München, Munich, Germany

ARTICLE INFO

Article history:
Received 9 August 2011
Received in revised form 20 January 2012
Accepted 21 January 2012
Available online 30 January 2012

Keywords:
Norovirus
Intraindividual evolution
Selection
Genotypes
Coalescent-based computation
Bayesian sampling

ABSTRACT

Norovirus has become an important cause for infectious gastroenteritis. Particularly genotype II.4 (GII.4) has been shown to spread rapidly and causes worldwide pandemics. Emerging new strains evade population immunity and lead to high norovirus prevalence. Chronic infections have been described recently and will become more prevalent with increasing numbers of immunocompromized patients. Here, we studied norovirus evolution in three chronically infected patients, two genotypes II.4 and one II.7.

A 719 and 757 nt region was analyzed for GII.4 and GII.7, respectively. This covers the entire hypervariable P2 domain of the VP1 capsid gene. Genetic variability at given and between different time points was assessed. Evolutionary adaptation was analyzed by Bayesian sampling of genealogies. This analysis clearly demonstrated positive selection rather than incidental drift for all three strains. The GII.7 and one GII.4 strain accumulated on average 5–9 mutations per 100 days, most of them non-synonymous. This is a much higher evolutionary rate than observed for noroviruses on a global level.

Our data demonstrate that norovirus quasispecies are positively selected in chronically infected patients. The numbers of intraindividual amino acid mutations acquired in the capsid gene are similar to those separating consecutive GII.4 epidemic strains. Evolution in a given, chronically infected individual may thus generate novel genotypes at risk to expedite global evolution particularly for slowly evolving genotypes, as GII.7.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Noroviruses are small non-enveloped, positive-sense RNA viruses belonging to the *Caliciviridae* family. Over the recent years, they have become the most prevalent gastroenteritis viruses in adults (Lamhoujeb et al., 2007).

The noroviral genome is ≈7.5 kb long and contains three open reading frames. The first encodes non-structural proteins (including the RNA polymerase), the second the major capsid protein and the third a minor structural protein (Glass et al., 2000). The capsid protein (VP1) encompasses the N-terminal S (shell) and the C-terminal P (protruding) domain (Nilsson et al., 2003). The S domain is conserved whereas the P domain includes the hypervariable region P2.

Based on their VP1 and/or pol sequence five genogroups are distinguished among norovirus strains, illustrating their high genetic diversity. Only GI, II, and to a much lesser extent IV are clinically important for humans (Ando et al., 2000; Koopmans et al., 2002). Currently, 14 subtypes are differentiated in GI and 17 in GII, based on complete capsid sequences (Zheng et al., 2006). Worldwide, GII is more prevalent than GI. GII.4 was detected in 80% of outbreaks over the last decade (Pang et al., 2010) but also in sporadic cases (Kroneman et al., 2006; Siebenga et al., 2007). The consecutive GII.4 strains have been well characterized: 2002, Farmington Hills; 2004, Hunter; 2006, 2006a virus; and 2007, 2006b virus (Bull et al., 2010; Siebenga et al., 2007; Tu et al., 2008). These variants, which replace each other after a certain time span, illustrate that herd immunity forces the GII.4 strains to evolve continuously (Boon et al., 2011; Lindesmith et al., 2011). While the nucleotide evolutionary rate has been reported to be similar in GII.4 and GII.3 the number of nonsynonymous mutations was found to be much higher in GII.4 (Boon et al., 2011), indicating evolution under positive selective pressure.

Noroviral infections persisting over weeks and months have recently been described in immunocompromized patients (Nilsson

* Corresponding author. Tel.: +49 89 4140 6825; fax: +49 89 4140 6858.
E-mail address: Hoffmann@virologie.med.tum.de (D. Hoffmann).

1567-1348/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2012.01.020

Author's personal copy

et al., 2003; Schorn et al., 2010) and even in individuals without obvious immune suppression (Obara et al., 2008). In patients after allogenic stem cell transplantation, norovirus gastroenteritis has been associated with prolonged morbidity and higher mortality (Roddie et al., 2009). With a high mutation rate and evolutionary rate (Bok et al., 2011; Bull et al., 2010), noroviruses can be expected to evolve within one host resulting in variable and diverse quasi-species. In fact, intra-individual evolution under immune selective pressure has been observed (Nilsson et al., 2003; Schorn et al., 2010). Siebenga et al. reported that the more a patient's immune function was suppressed the less amino acid mutations per time were selected in intraindividual quasi-species. Thus, the individual immune system seems important for positively selecting mutants even during chronic infections. The authors also emphasize the role of chronically infected persons as viral reservoirs (Siebenga et al., 2008).

As discussed above, global epidemiology is substantially different for GII.4 compared to the other norovirus subtypes. We therefore examined the evolution of one GII.7- and two GII.4 2006b strains applying population genetic methods. The variant GII.4 2006b has been very successful at a global level being the dominant norovirus genotype from 2006 to 2010 and causing increased norovirus prevalence in various settings and geographical regions (Motomura et al., 2008). On the other hand, GII.7 had a low prevalence for years (Phan et al., 2007) and closely related sequences to our GII.7 strain A date back to 1990 (Hoffmann et al., 2010).

For most population genetic datasets it is not possible to reconstruct genealogy of sequence data with high confidence. For our datasets, however, we can infer a certain aspect of genealogy: the number of branches that separates the ancestral lineages of the latest samples from the early ones. Applying Bayesian tree reconstruction, we could estimate these numbers with high credibility and subsequently used them as test statistics for the null hypothesis of neutral evolution.

2. Material and methods

All three patients studied were immunocompromized. Patients A and R underwent allogenic bone marrow transplantation; patient P had received a small bowel transplant. Patients A and P were treated at the university hospital, Technische Universität München, samples from patient 3 originated from the Institute of Virology in Freiburg.

Noroviral strains analyzed in this study have been submitted to GenBank and were assigned the following accession numbers: A_03-2006 JQ417292–300; A_06-2006 JQ417301–308; P_04-2009 JQ417309–316; P_09-2009 JQ417317–326; P_02-2010 JQ417327–336; R_09-2009 JQ417337–344; R_02-2010 JQ417345–353.

2.1. RNA extraction, quantitative PCR, and sequencing

After nucleic acid extraction with the High Pure Nucleic Acid Kit (Roche, Mannheim, Germany) noroviral load was determined with an in house real time PCR assay as recently described in detail (Hoffmann et al., 2010). The amplicons for sequence analysis encompassing VP1 capsid sequences were generated with a semi-nested PCR using primers S1b AGCTATGCTCTACACCCCTTT, S2b GTCCTCACAA AACCGTCACC, and Ab TGTGGCAGCTTCCTGATAGA. PCR products were cloned into vector pCR4 (Invitrogen, Karlsruhe, Germany) by TA cloning. Two-strand sequencing was carried out at GATC (Konstanz, Germany). The analyzed sequences included 757 bp for isolate A and 719 bp for strains P and R.

2.2. Nucleic acid and amino acid alignments, assessment of pairwise distances

The sequences were aligned using the Clustal algorithm (Thompson et al., 1994) and Mega version 5 (Tamura et al., 2007). Pairwise distances between clones were also assessed with Mega. Consensus sequences were determined with Consensus Maker http://www.hiv.lanl.gov/content/sequence/CONSENSUS/consensus.html. Consensus sequences were used to determine genetic distances between different time points. Related published sequences were identified by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

3. Theory and calculation

For the Bayesian sampling of genealogies we used the Markov chain Monte Carlo (MCMC) phylogeny reconstruction program BEAST, version 1.5.4 (Drummond and Rambaut, 2007). We used the HKY nucleotide substitution model with rate heterogeneity (Yang, 1993) and the SRD06 model for the third codon position. We chose this model since it covers essential aspects of sequence evolution and has not too many parameters in consideration of the size of our datasets. However, we made sure that our results are robust with respect to the choice of the substitution model by repeating the BEAST analyses with three other widely used models (HKY, GTR, GTR + GAMMA with 4 categories) without SD06. For each of the (in total four) models and each three datasets we launched four independent BEAST runs. Each of these $4 \times 3 \times 4 = 48$ runs resulted in trees that contained the branches that we use in our subsequent analyses, and in each of the 48 Program runs the posterior probabilities of these branches were always higher that 99%, and most times higher than 99.9%.

For the prior of the genealogy we used the neutral coalescent, taking the different sampling times into account. For all other options we used the default settings of BEAST's graphical input tool BEAUTY. Each MCMC procedure performed 10,000,000 steps, storing a tree and a set of parameter values every 1000 steps. We considered the first 10% of each MCMC run as "burn-in" and removed it from the sample. This resulted in 9001 sampled trees per MCMC run. For each dataset we launched five independent MCMC runs with random initial trees. We used the program Tracer, version 1.4 (http://tree.bio.ed.ac.uk/software/tracer/), to assess convergence of the MCM chains. To compute and to visualize the maximum credibility trees from the samples we used the programs treeannotator and Figtree version 1.1.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Using the genealogies reconstructed with BEAST we applied the following tests for evolutionary adaptation of viruses between the two or three sampling times. The null-hypothesis is that the virus population in a patient evolved neutrally by genetic drift and random mutations on the locus under consideration. As test statistic we use the number of branches in the reconstructed genealogy that separate the ancestral lineages of the newly sampled sequences from the sequences from the first sampling. More precisely, the test statistic S is the minimal number of branches that must be removed from the genealogy to prune the new samples from the old ones. Thus, our test should only be applied when this number can be estimated with high credibility, e.g. from MCMC analyses of the sequences. The p-value of the observation $\{S = k\}$ is the probability under the null-hypothesis that S takes a value in $\{1, 2, ..., k\}$. To compute p-values it is necessary to assume values for the effective population size and for the number of virus generations between the two samplings. To achieve conservative p-values we made sure not to overestimate the population size or to underestimate the number of generations. We assumed a generation time of 24 h.

Author's personal copy

The effective population size is hard to estimate since population structure may play an important role. However, we assume that 1000 is a rather conservative estimation for the effective size of a virus population in a patient. Our computation of the p-value is based on a Wright-Fisher model (Ewens, 2004) with the simplifying assumption of discrete populations. To compute the probability $p_{t,a,b,k}$ of $\{S = k\}$, where t is the number of generations between the sampling times, a is the number of sequences in the first and b is the number of sequences in the later sample, we have to compute for all m in $\{k, k+1, \ldots, b\}$ the probability $q_{t,b,m}$ that sequences of the second sample have exactly m ancestors in the generation of the first sample and the probability $v_{k,m,a}$ that these m ancestors are separated from the a individuals from the first sample by k branches in their common genealogy (in the same sense as in the definition of S). We can then compute $p_{t,a,b,k}$ by

$$p_{t,a,b,k} = \sum_{m=k}^{b} q_{t,b,m} \cdot v_{k,m,a}$$

To compute $q_{t,b,m}$ we use a formula from coalescent theory (Tavaré, 1984):

$$q_{t,b,m} = \sum_{m=k}^{b} e^{-\binom{k}{2}^{t/N_e}} \frac{(-1)^{k-m}(2k-1)\cdot(m+k-2)!\cdot b!\cdot(b-1)!}{m!\cdot(k-m)!\cdot(b+k-1)!\cdot(m-1)!\cdot(b-k)!}$$

For the effective population size, N_e, we use the conservative estimate 1000. We assume that the effective population virus size in a patient is much larger, and the values of $q_{t,b,m}$ and thus also the p-values are much lower. The values of $v_{k,m,a}$ are only needed for $k = 1$ and $k = 2$, and we compute them by $v_{1,1,a} = v_{1,m,1} = 1$, $v_{2,1,a} = v_{2,m,1} = 0$, $v_{2,2,2} = 2/3$ and the recursions

$$v_{1,m,a} = \frac{m\cdot(m-1)\cdot v_{1,m-1,a} + a\cdot(a-1)\cdot v_{1,m,a-1}}{(m+a)\cdot(m+a-1)}$$

and

$$v_{2,m,a} = \frac{m\cdot(m-1)\cdot v_{2,m-1,a} + a\cdot(a-1)\cdot v_{2,m,a-1} + m\cdot a\cdot v_{1,m-1,a-1}}{(m+a)\cdot(m+a-1)}$$

4. Results

4.1. Sample characteristics

We tested two longitudinal samples from patient A, three from P, and two from patient R. The label of each sample includes the letter for patient identification followed by month and year the respective specimen was taken. Table 1 summarizes the characteristics of the tested samples. The time periods covered by positive samples are shown in the left column. Per sample we analyzed 8–10 clones.

4.2. Closest published sequences

The closest sequence related to strain A was a GII.7 strain from Australia identified in 2008 (GQ849130). Thirty-two nucleic acid mutations, only 3 being non-synonymous clearly illustrate pronounced negative selection on a global level. For strain P we found sequences from Japan (GII-4/Shimane2/2008/JP; AB541348) and France (GGII.4/Dijon/E3808/2009/FRA; GQ246789) differing by 19 and 21 nucleic acid mutations, 8 and 11 of which were non-synonymous. This argues for positive selection. Strain R, however, was very close to previously published sequences, e.g. GII.4/Hong Kong/CU060024/2006/CHN (HM802538) bearing only nine mutations (3 non-synonymous). According to these alignments P and R were assigned to GII.4 2006b.

Table 1
Characteristics of the analyzed patients, samples, and strains.

Strain	Norovirus positive	Tested samples (strains)	Clones per sample
A	>81 days	A_03-2006; A_06-2006	10; 9
P	>302 days	P_04-2009; P_09-2009; P_02-2010	8; 10; 10
R	>144 days	R_09-2009; R_02-2010	8; 9

Table 2
Mean pairwise differences between clonal sequences originating from one sample. Sequenced samples are specified in the left column. The alphabetic label is followed by the month and year of the respective sample.

Strain	Nucleic acid level	Amino acid level
A_03-2006; A_06-2006	1.67; 3.83	1.47; 1.94
P_04-2009; P_09-2009; P_02-2010	1.50; 4.47; 4.67	1.00; 2.60; 3.67
R_09-2009; R_02-2010	2,50; 3.89	1.25; 1.44

Table 3
Mean pairwise variation between longitudinal samples referred to 100 days.

Strain	Nucleic acid level	Amino acid level
A_03-2006; A_06-2006	8.76	6.77
P_04-2009; P_09-2009; P_02-2010	5.94; 5.91; 8.25	4.83; 4.96; 6.14
R_09-2009; R_02-2010	2.68	1.32

4.3. Pairwise differences between clonal sequences

Genetic variability between quasispecies within one sample increased over time for A, P, and R (Table 2). This was expected since the early samples represented acute infection with transmission-related genetic bottlenecks limiting variability. Clones of strain R were less variable than those of A and P.

4.4. Longitudinal mutations between consensus sequences

Consensus sequences of A and P accumulated similar numbers of mutations per 100 days (Table 3). Slightly more nucleic acid mutations (8.76) emerged in A compared to P (5.94; 5.91; 8.25), while the proportion of non-synonymous mutations was again higher for P. Strain R yielded less nucleic acid mutations per 100 days (2.68) and only half of them were non-synonymous.

4.5. Bayesian reconstruction of genealogies

For strain A there was one branch in the maximum credibility trees that separated the 10 sequences of the first sampling from the 9 sequences sampled approximately 80 generations later (Fig. 1). For P there was a branch in each maximum credibility tree that separated the 8 sequences of the first sample from the 10 sequences of the second sampling. Another branch separated the 10 sequences of the second sampling from the 10 sequences of the third sampling (Fig. 2). For R the 8 sequences of the first sample were separated by two branches from the 9 sequences of the second sample (Fig. 3). Figs. 1–3 show the maximum credibility trees from the first MCMC BEAST runs of each dataset, the branches are labeled with the posterior probabilities estimated in these runs. These branches are the basis of the statistical tests in Section 4.6. The posterior probabilities computed in the other four BEAST runs were always higher than 0.9975 for all branches separating longitudinal samples and thus rebut the possibility that these branches were wrong. The higher probabilities of isolate R compared to the

Author's personal copy

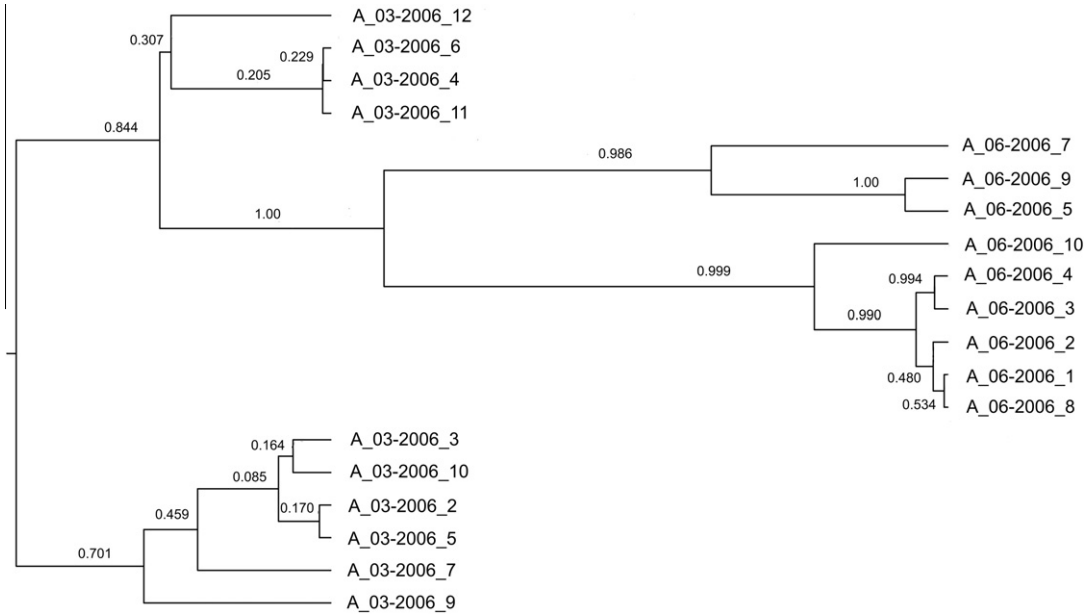

Fig. 1. Phylogram including 10 clonal sequences from the first time point A_03-2006 and 9 sequences from the second sample A_06-2006. Numbers on the branches indicate the posterior probabilities resulting from the first of five independent MCMC runs executed in BEAST.

other two strains parallels the lower number of nonsynonymous mutations and the fact that this strain evolved following two branches between the two time points. On the other hand, quasispecies of A and P required only one branch to evolve from one longitudinal sample to the next. This implies a selective advantage for the strain that gives rise to the quasispecies detected at the following time points.

The effective sample sizes computed by Tracer were above 400 for all parameters and all individual MCMC runs. When we pooled the results of the 5 runs for any of the three dataset we obtained effective sampling sizes over 2500 for all parameters. Also, Tracer's visualization of the log files did not yield any evidence for insufficient convergence of the MCM chains.

4.6. P-value based on a Wright-Fisher model

We tested the null hypothesis of neutral evolution assuming an effective population size of at least 1000. As test statistics we used the numbers of branches that separate the longitudinal samples according to the Bayesian analyses in Section 4.5. With an estimated 24 h generation time we assumed 80 generations between the samplings for strain A and 150 generations for P, and R, respectively. A_06-2006 originating from A_03-2006 yielded a p-value of 1.33×10^{-4}. P_09-2009 evolving out of P_04-2009 resulted in 8.94×10^{-4}, while P_09-2009 gave rise to P_02-2010 with 5.11×10^{-4}. For R_09-2009 evolving to R_02-2010 test statistics gave a p-value of 1.34×10^{-2}.

5. Discussion

In the presented study we examined the evolution of the norovirus capsid gene within three chronically infected patients. Comparing A_03-2006 to previous GII.7 strains, we found much less amino acid exchanges than P_04-2009 acquired compared to published GII.4 strains (Boon et al., 2011). However, on a within-host level, A and P accumulated a similar number of mutations per 100 days (Table 3). Between A_03-2006 and A_06-2006 we found 4 non-synonymous and no silent mutations in the P2 domain (Hoffmann et al., 2010). In contrast to GII.4 no modeling data about VP1 structure are available for GII.7, thus we can only speculate about the biological implications of these mutations.

Strain P (GII.4) showed 15 nucleic acid mutations in the P2 domain of the capsid protein. Thirteen of them were non-synonymous. This is in the range of 8 and 23 amino acid exchanges the epidemic strains 2006a und 2006b showed compared to their predecessors (Siebenga et al., 2007). Six of the 13 amino acid exchanges occurred at positions that had mutated already in GII.4 2006b compared to its predecessor (Fukuda et al., 2010): S296T, G340R, Y352S, A356V, G393S, N412D. The resulting genotype is most closely related to a cluster termed 04/05/AU/N, including strains epidemic in the Netherlands and Australia in 2004–2005 (Motomura et al., 2008; Siebenga et al., 2007). S296T and G393S were associated with the appearance of new strains and thus are considered to be part of variant-specific epitopes (Allen et al., 2008). This was confirmed by homology modeling. G340R was located in a conserved RGD motive (amino acid 339–341), which may play a role in ligand binding (Siebenga et al., 2007; Tan et al., 2003).

Strain R, even though clustering with GII.4 2006b, yielded only one non-synonymous (V333E) and one synonymous mutation. At position 333, a hypervariable position in GII.4 with valine being the dominant amino acid in 2006b strains (Fukuda et al., 2010), glutamate was found. This amino acid has not yet been described at amino acid 333 and its hydrophilic side chain and may involve a different function. Virtually all sequences closely related to R originated from 2006 or 2007. We speculate that the strain had a selection disadvantage because of its slow evolution, and thus decreased in prevalence over the recent years. On the other hand, patient R was bone marrow transplanted during norovirus infection and resulting massive immunosuppression probably lead to a weak selective pressure. All three strains, however, increased their genetic distance to published isolates during intra-individual evolution. Thus their evolution seems be patient-specific. Even though norovirus transmission has not been documented after long time evolution within a single host, this type of virus evolution may generate distinct genotypes driven by other factors than those driving global evolution.

Our statistical test for selection is based on the assumption that the branches that separate later sampled viruses from earlier

Author's personal copy

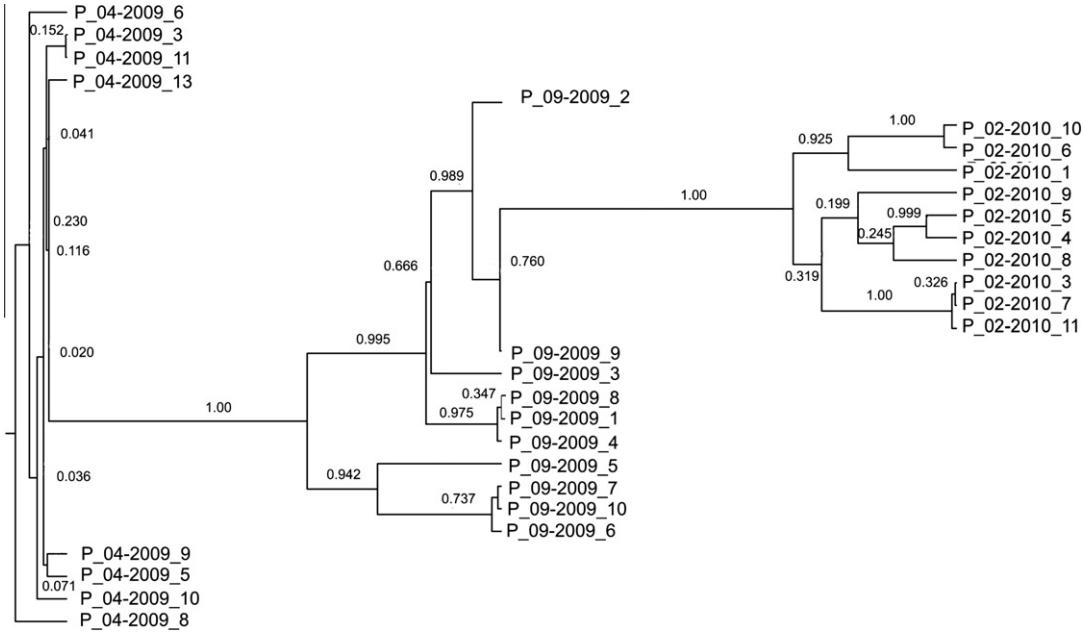

Fig. 2. Phylogram including 8 clonal sequences from the first (P_04-2009), 10 from the second (P_09-2009), and 10 from the third sample (P_02-2010), respectively.

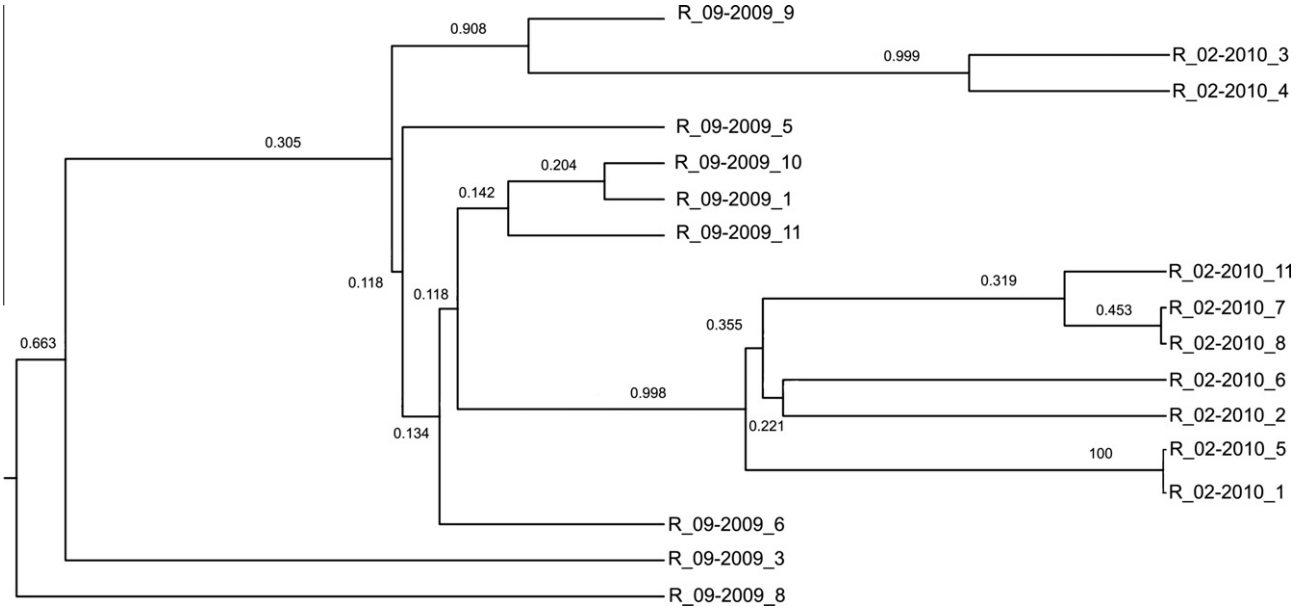

Fig. 3. Phylogram including 8 clonal sequences from the first (R_09-2009) and 9 clones from the second sample (R_02-2010). Compared to strains A and P two branches are required to separate the 09-2009 from the 02-2010 clones.

sampled ones in the genealogies given by BEAST are true. BEAST estimated the posterior probabilities of these branches above 0.9975. To take the uncertainty about these branches into account, we could add the total posterior probability of genealogies not having these branches to the *p*-values. Since these values are below $1 - 0.9975 = 0.0025$, the corrected *p*-values would still stay well below a significance level of 0.01. This argument mixes a Bayesian touch into the frequentistic test. From a frequentist perspective, a critical point in Bayesian statistics is the choice of a prior distribution. In our case, this not a problem because the prior that we used for the genealogy in the BEAST-analysis is the Kingman coalescent, which is also part of the null hypothesis of the genealogy-based test. In the tree reconstruction process we neglected the possibility of recombination within the locus under consideration. We did so because recombination events in noroviruses have mainly been described in the regions between the open reading frames 1 and 2 (Bull et al., 2007).

In chronic norovirus infection the host's immune system competes with the virus' ability to escape immune selective pressure by mutations. If the immune response suppresses viral replication quickly it does not give the virus enough time to generate and select mutations, and the infection is cleared. Without an immune response mutations would merely result from genetic drift and the only selection criterion would be virus fitness. The evolutionary rate on amino acid level will be highest when the immune response cannot eliminate the virus but still exerts a considerable selective pressure. Our genealogy-based tests clearly indicated positive selection in the strains A, P, and R. The proportion of acquired

Author's personal copy

non-synonymous mutations already indicated this for A and P. For R only the genealogy-based test provided statistical evidence for positive selection. The test statistics using the number of branches separating the longitudinal sampled sequences detects positive selection very sensitively. Our data (Table 3) indicate that the low prevalence subtype GII.7 can accumulate a similar number of nucleic acid mutations per time as GII.4 on an intra-individual level, confirming recent findings from other groups (Boon et al., 2011). However, we cannot rule out that variations in the immune status of the infected individuals alter the evolutionary rates of their respective norovirus populations. GII.7 strains may not be forced to evolve rapidly on a global level as they are not prevalent enough to generate effective herd immunity. Thus their slow evolution by negative selection seems to be the consequence rather than the cause of their low global prevalence. GII.7 strains may be less virulent or less stable in the environment, which then reduces their global transmission.

6. Conclusions

Chronic norovirus infections will become an increasing clinical problem in the future, because of the growing number of immune-compromised patients. Chronically infected patients have been reported to be involved in outbreaks, even though no transmission was documented after >17 days of viral shedding (Sukhrie et al., 2010). In this study we demonstrate that both GII.4 and GII.7 strains undergo positive selection during chronic infection. Because of the rather constant intra-individual selection pressure the evolutionary rate is much higher than that found when the virus spreads throughout a population. Thereby norovirus strains within weeks to months can acquire enough amino acid mutations to constitute novel epidemic subtypes. Particularly for GII.7 with a slow evolution over years on a global scale, intra-individual positive selection can give rise to substantially different genotypes that normally take years to evolve.

Genetic variability is one of the main reasons allowing noroviruses to persist in the human population at high prevalence. Thus, a better knowledge of norovirus evolution and its driving forces is not only of scientific but also of clinical and public health interest.

Acknowledgements

We thank Hermann Schätzl and Gerd Frösner for helping to establish Norvirus PCRs. We also thank the technicians of the Institute of Virology, Technische Universität München, as well as patients, nurses, and doctors for their support of this study. The study was partially funded by the "Förderverein zur Bekämpfung der Viruskrankheiten".

References

Allen, D.J., Gray, J.J., Gallimore, C.I., Xerry, J., Iturriza-Gomara, M., 2008. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. PloS One 3, e1485.

Ando, T., Noel, J.S., Fankhauser, R.L., 2000. Genetic classification of "Norwalk-like viruses". J. Infect. Dis. 181 (Suppl. 2), S336–S348.

Bok, K., Parra, G.I., Mitra, T., Abente, E., Shaver, C.K., Boon, D., Engle, R., Yu, C., Kapikian, A.Z., Sosnovtsev, S.V., Purcell, R.H., Green, K.Y., 2011. Chimpanzees as an animal model for human norovirus infection and vaccine development. Proc. Natl. Acad. Sci. USA 108, 325–330.

Boon, D., Mahar, J.E., Abente, E.J., Kirkwood, C.D., Purcell, R.H., Kapikian, A.Z., Green, K.Y., Bok, K., 2011. Comparative evolution of GII.3 and GII.4 norovirus over a 31-year period. J. Virol. 85, 8656–8666.

Bull, R.A., Eden, J.S., Rawlinson, W.D., White, P.A., 2010. Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS Pathog. 6, e1000831.

Bull, R.A., Tanaka, M.M., White, P.A., 2007. Norovirus recombination. J. Gen. Virol. 88, 3347–3359.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Ewens, W.J., 2004. Mathematical population genetics. 1„ .. Theoretical Introduction, second ed. Springer, New York.

Fukuda, S., Takao, S., Shigemoto, N., Tanizawa, Y., Seno, M., 2010. Transition of genotypes associated with norovirus gastroenteritis outbreaks in a limited area of Japan, Hiroshima Prefecture, during eight epidemic seasons. Arch. Virol. 155, 111–115.

Glass, P.J., White, L.J., Ball, J.M., Leparc-Goffart, I., Hardy, M.E., Estes, M.K., 2000. Norwalk virus open reading frame 3 encodes a minor structural protein. J. Virol. 74, 6581–6591.

Hoffmann, D., Seebach, J., Foley, B.T., Frosner, G., Nadas, K., Protzer, U., Schatzl, H.M., 2010. Isolated norovirus GII.7 strain within an extended GII.4 outbreak. J. Med. Virol. 82, 1058–1064.

Koopmans, M., von Bonsdorff, C.H., Vinje, J., de Medici, D., Monroe, S., 2002. Foodborne viruses. FEMS Microbiol. Rev. 26, 187–205.

Kroneman, A., Vennema, H., Harris, J., Reuter, G., von Bonsdorff, C.H., Hedlund, K.O., Vainio, K., Jackson, V., Pothier, P., Koch, J., Schreier, E., Bottiger, B.E., Koopmans, M., 2006. Increase in norovirus activity reported in Europe. Euro Surveill. 11, E061214, 061211.

Lamhoujeb, S., Charest, H., Fliss, I., Ngazoa, S., Jean, J., 2007. Phylogenetic analysis of norovirus isolates involved in some Canadian gastroenteritis outbreaks in 2004 and 2005. Can. J. Microbiol. 53, 1133–1140.

Lindesmith, L.C., Donaldson, E.F., Baric, R.S., 2011. Norovirus GII.4 strain antigenic variation. J. Virol. 85, 231–242.

Motomura, K., Oka, T., Yokoyama, M., Nakamura, H., Mori, H., Ode, H., Hansman, G.S., Katayama, K., Kanda, T., Tanaka, T., Takeda, N., Sato, H., 2008. Identification of monomorphic and divergent haplotypes in the 2006–2007 norovirus GII/4 epidemic population by genomewide tracing of evolutionary history. J. Virol. 82, 11247–11262.

Nilsson, M., Hedlund, K.O., Thorhagen, M., Larson, G., Johansen, K., Ekspong, A., Svensson, L., 2003. Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. J. Virol. 77, 13117–13124.

Obara, M., Hasegawa, S., Iwai, M., Horimoto, E., Nakamura, K., Kurata, T., Saito, N., Oe, H., Takizawa, T., 2008. Single base substitutions in the capsid region of the norovirus genome during viral shedding in cases of infection in areas where norovirus infection is endemic. J. Clin. Microbiol. 46, 3397–3403.

Pang, X.L., Preiksaitis, J.K., Wong, S., Li, V., Lee, B.E., 2010. Influence of novel norovirus GII.4 variants on gastroenteritis outbreak dynamics in Alberta and the Northern Territories, Canada between 2000 and 2008. PloS One 5, e11599.

Phan, T.G., Kaneshi, K., Ueda, Y., Nakaya, S., Nishimura, S., Yamamoto, A., Sugita, K., Takanashi, S., Okitsu, S., Ushijima, H., 2007. Genetic heterogeneity, evolution, and recombination in noroviruses. J. Med. Virol. 79, 1388–1400.

Roddie, C., Paul, J.P., Benjamin, R., Gallimore, C.I., Xerry, J., Gray, J.J., Peggs, K.S., Morris, E.C., Thomson, K.J., Ward, K.N., 2009. Allogeneic hematopoietic stem cell transplantation and norovirus gastroenteritis: a previously unrecognized cause of morbidity. Clin. Infect. Dis. 49, 1061–1068.

Schorn, R., Hohne, M., Meerbach, A., Bossart, W., Wuthrich, R.P., Schreier, E., Muller, N.J., Fehr, T., 2010. Chronic norovirus infection after kidney transplantation: molecular evidence for immune-driven viral evolution. Clin. Infect. Dis. 51, 307–314.

Siebenga, J.J., Beersma, M.F., Vennema, H., van Biezen, P., Hartwig, N.J., Koopmans, M., 2008. High prevalence of prolonged norovirus shedding and illness among hospitalized patients: a model for in vivo molecular evolution. J. Infect. Dis. 198, 994–1001.

Siebenga, J.J., Vennema, H., Renckens, B., de Bruin, E., van der Veer, B., Siezen, R.J., Koopmans, M., 2007. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. J. Virol. 81, 9932–9941.

Sukhrie, F.H., Siebenga, J.J., Beersma, M.F., Koopmans, M., 2010. Chronic shedders as reservoir for nosocomial transmission of norovirus. J. Clin. Microbiol. 48, 4303–4305.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.

Tan, M., Huang, P., Meller, J., Zhong, W., Farkas, T., Jiang, X., 2003. Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. J. Virol. 77, 12562–12571.

Tavaré, S., 1984. Line-of-descent and genealogical processes, and their applications in population genetic models. Theor. Popul. Biol. 26 (119–164), 119–164.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Tu, E.T., Bull, R.A., Greening, G.E., Hewitt, J., Lyon, M.J., Marshall, J.A., McIver, C.J., Rawlinson, W.D., White, P.A., 2008. Epidemics of gastroenteritis during 2006 were associated with the spread of norovirus GII.4 variants 2006a and 2006b. Clin. Infect. Dis. 46, 413–420.

Yang, Z., 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10, 1396–1401.

Zheng, D.P., Ando, T., Fankhauser, R.L., Beard, R.S., Glass, R.I., Monroe, S.S., 2006. Norovirus classification and proposed strain nomenclature. Virology 346, 312–323.