Ancient Evolution of Mammarenaviruses: Adaptation via Changes in the L Protein and No Evidence for Host–Virus Codivergence

Diego Forni[1,*], Chiara Pontremoli[1], Uberto Pozzoli[1], Mario Clerici[2,3], Rachele Cagliani[1], and Manuela Sironi[1]

[1]Bioinformatics, Scientific Institute IRCCS E. MEDEA, Bosisio Parini, Italy

[2]Department of Physiopathology and Transplantation, University of Milan, Italy

[3]Don C. Gnocchi Foundation ONLUS, IRCCS, Milan, Italy

*Corresponding author: E-mail: diego.forni@bp.lnf.it.

Accepted: March 2, 2018

Abstract

The *Mammarenavirus* genus includes several pathogenic species of rodent-borne viruses. Old World (OW) mammarenaviruses infect rodents in the Murinae subfamily and are mainly transmitted in Africa and Asia; New World (NW) mammarenaviruses are found in rodents of the Cricetidae subfamily in the Americas. We applied a selection-informed method to estimate that OW and NW mammarenaviruses diverged less than ~45,000 years ago (ya). By incorporating phylogeographic inference, we show that NW mammarenaviruses emerged in the Latin America-Caribbean region ~41,400–3,300 ya, whereas OW mammarenaviruses originated ~23,100–1,880 ya, most likely in Southern Africa. Cophylogenetic analysis indicated that cospeciation did not contribute significantly to mammarenavirus–host associations. Finally, we show that extremely strong selective pressure on the viral polymerase accompanied the speciation of NW viruses. These data suggest that the evolutionary history of mammarenaviruses was not driven by codivergence with their hosts. The viral polymerase should be regarded as a major determinant of mammarenavirus adaptation.

Key words: Arenaviridae, mammarenavirus, positive selection, phylogeography, molecular dating.

Introduction

Arenaviruses are enveloped negative-sense RNA viruses belonging to the Arenaviridae family. Several arenaviruses that infect mammals have been known for years. More recently, the identification of divergent arenaviruses in alethinophidian snakes led to the definition of three genera in the Arenaviridae family: the *Reptarenavirus* and *Hartmanivirus* genera, which currently includes six species for reptilian viruses, and the *Mammarenavirus* genus, that comprises several species for viruses isolated in mammals (Radoshitzky et al. 2015) (https://talk.ictvonline.org/; last accessed February 10, 2018).

The *Mammarenavirus* genus is further divided into two large monophyletic groups. The New World (NW) or Tacaribe complex includes viruses distributed in the Americas, while the Old World (OW) complex comprises several African viruses, the ubiquitous lymphocytic choriomeningitis arenavirus (LCMV), as well as some viruses recently isolated in Asia (Wēnzhōu virus, WENV and Loei River virus, LORV) (Li et al. 2015; Radoshitzky et al. 2015; Blasdell et al. 2016).

With the exception of Tacaribe virus (TCRV), which was detected in phyllostomid bats (Downs et al. 1963), all mammarenaviruses have natural reservoirs in rodent hosts (Gonzalez et al. 2007). Viruses of each species establish acute or persistent infections in rodents of few species. In particular, NW mammarenaviruses preferentially infect rodents from the subfamilies Sigmodontinae (viruses from Latin America and the Caribbean) and Neotominae (Northern American viruses), whereas OW mammarenaviruses are found in rodents from the subfamily Murinae (Gonzalez et al. 2007).

Mammarenavirus infection is often asymptomatic in rodents, an observation that suggests long-standing coevolution (Gonzalez et al. 2007). In humans, several mammarenaviruses cause disease. NW mammarenaviruses are divided into four groups (A to D, with this latter also referred to as RecA due its possible origin following an ancient recombination event) (Radoshitzky et al. 2015). In addition to nonpathogenic viruses, group B includes several pathogens (e.g., Guanarito virus, Junín virus, and Machupo virus), which cause severe hemorragic fevers in different areas of Latin America and

© The Author(s) 2018. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

the Caribbean (Kerber et al. 2015). Among OW arenaviruses, Lassa virus (LASV) infects ~100,000–300,000 people annually in Western Africa and causes ~5,000 deaths (https://www.cdc.gov/vhf/lassa/index.html; last accessed February 10, 2018). As for LCMV, infection is particularly dangerous when contracted *in utero* (Charrel and de Lamballerie 2010) or during immunosuppresive treatments (Charrel and de Lamballerie 2010).

Most mammarenavirus infections are caused by rodent-to-human transmission, via direct contact with infected animals or their fomites (Charrel and de Lamballerie 2010). Human-to-human transmission is rare, indicating that humans are dead-end hosts for mammarenaviruses (Charrel and de Lamballerie 2010). Thus, the geographic range and evolutionary dynamics of these viruses are mainly determined by their natural hosts, although anthropogenic factors may also play a role (Albariño et al. 2010; Jones et al. 2013).

Despite the relevance of mammarenaviruses for human health, little is known about their origin and long-term evolutionary history. Herein we used molecular dating, cophylogenetic analysis, and phylogeography to fill this gap.

Materials and Methods

Sequence Alignments, Recombination, Substitution Saturation, and Gene Trees

Coding sequences were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/; last accessed February 10, 2018). Lists of accession numbers are reported in supplementary tables S1 and S2, Supplementary Material online. For all analyses, sequences were only included if the isolation procedure indicated fewer than seven passages in cell lines or mouse brain.

We used MAFFT (Katoh and Standley 2013) to generate multiple sequence alignments. We next used GUIDANCE2, a tool that allows the automated removal of poorly aligned codons from a multiple sequence alignment (Sela et al. 2015), for filtering codons with a score <0.90 (Privman et al. 2012).

Alignments were screened for the presence of recombination using GARD (Kosakovsky Pond et al. 2006), a genetic algorithm from the HYPHY suite. GARD is based on the concept that recombination results in regions having different evolutionary histories and thus uses phylogenetic incongruence among segments in the alignment to detect recombination breakpoints. The statistical significance of putative breakpoints is evaluated through Kishino–Hasegawa (HK) tests. No significant breakpoint ($P < 0.05$) was detected.

To evaluate the level of substitution saturation at the third codon position, we applied the Xia's index implemented in DAMBE (Xia 2013).

Phylogenetic trees were reconstructed using the phyML program with a maximum-likelihood approach, a General Time Reversible (GTR) model plus gamma-distributed rates and 4 substitution rate categories(Guindon et al. 2009).

Time Estimates

For molecular dating, we analyzed 188 mammarenavirus RdRp sequences with known isolation dates (supplementary table S1, Supplementary Material online), plus a reptarenavirus (University of Helsinki virus, UHV-1, NCBI ID: KR870020) (Hepojoki, Salmenperä, et al. 2015) to root the phylogeny.

Regression of root-to-tip genetic distances against sequence sampling times was performed using dedicated R scripts, as previously described (Murray et al. 2016). A method that minimizes the residual mean squares of the model was applied, as suggested (Murray et al. 2016). The *P* value was calculated by performing 1,000 clustered permutations of the sampling dates (Murray et al. 2016).

Estimates of the time to the most recent common ancestor (tMRCA) were obtained by calculating branch lengths using the aBS-REL (adaptive branch-site random effects likelihood) (Smith et al. 2015) model. aBS-REL is a model of molecular evolution that accounts for variation in selection pressures both across sites and across the phylogeny (Smith et al. 2015). The aBS-REL tree was used as the input tree for the LSD (least-squares dating) software (v0.2) (To et al. 2016) to obtain divergence dates. Confidence intervals were estimated by using a latin hypercube sampling scheme (LHC) from the aBS-REL parameter distributions (Wertheim and Kosakovsky Pond 2011). Briefly, 500 samples were drawn from aBS-REL analyses to estimate branch length variance, 500 trees were generated, and then used as input trees for LSD. The upper and the lower 95% bounds were used as confidence intervals.

Phylogeographic Analysis

For the phylogeographic analysis of the RdRp segments, sequence locations were assigned on the basis of the United Nation geographical subregions (http://unstats.un.org/unsd/methods/m49/m49regin.htm; last accessed February 10, 2018). Due to the small sample size, Eastern Asia and South-eastern Asia were merged in a single region (supplementary table S1, Supplementary Material online).

Inferences of geographical origin of internal nodes in the mammarenavirus phylogeny were obtained using the discrete model (Lemey et al. 2009) implemented in BEAST (version 2.4.4) (Bouckaert et al. 2014).

Geographic origin was also inferred using the BBM (Bayesian Binary MCMC) method implemented in RASP (Reconstruct Ancestral State in Phylogenies) (Ronquist and Huelsenbeck 2003; Yu et al. 2015).

BEAST analyses were performed using the Bayesian Markov Chain Monte Carlo (MCMC) method with a General Time Reversible (GTR) substitution model and a gamma distribution (G) rate with 4 categories among sites. The GTR + G model was selected using the "ModelTest" utility (Posada and Crandall 1998) implemented in the HYPHY

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

package. A strict molecular clock was used. Two different runs, 100 million iterations each, were performed and sampled every 10,000 steps with a 10% burn-in. Runs were then combined after checking for convergence. Maximum clade credibility trees were summarized using TreeAnnotator (Bouckaert et al. 2014). Trees were visualized with FigTree (http://tree.bio.ed.ac.uk/; last accessed February 10, 2018).

For BBM analysis, 10,000 BEAST-generated trees and consensus tree were used as topology input. Two BBM chains were run for 100,000 generations with estimated state frequencies (F81), a gamma distributed among-site rate variation, sampling every 100 generations, and null character state for the outgroup (UHV-1 reptarenavirus strain).

Cophylogenetic Analysis

We compiled a list of 49 mammarenavirus–host association from available literature sources (supplementary table S3, Supplementary Material online). The natural reservoirs for Flexal, Chapare, Sabiá, and Lujo viruses are unknown and these viruses were thus not included in the analyses.

Cophylogeny between mammarenaviruses and their natural hosts was investigated with two event-based methods, Jane 4.0 (Conow et al. 2010) and CoRe-PA version 0.5.1(Merkle et al. 2010).

The phylogenetic tree of rodents was obtained from a previous work (Steppan and Schenk 2017) and derives from the concatenation of 5 nuclear genes. The phylogeny also included two hosts from different orders (a shrew and a bat). For mammarenavirus phylogenetic reconstruction (RdRp region), we used RAxML through the web server T-REX (Boc et al. 2012) using the GTR + G model and a reptarenavirus sequence as the outgroup. Node support was estimated with 1,000 bootstrap replicates and was >90% for all nodes. The tree was manually edited to include Gbagroube and Kodoko viruses, for which only small fragments of the *L* gene were sequenced (Lecompte et al. 2007; Coulibaly-N'Golo et al. 2011). The phylogenetic relationships of these two viruses with other OW mammarenaviruses was obtained by RAxML analysis of the available *L* region and was in line with previous reports (Coulibaly-N'Golo et al. 2011).

Jane 4.0 and CoRe-Pa assign costs to four coevolutionary events (cospeciation, duplication, sorting/loss, and host switching). Jane 4.0 also assigns a cost to failure-to-diverge events (a diversification event in the host but not in the parasite). Cost values must be set *a priori* in Jane 4.0, whereas CoRe-PA uses a parameter-adaptive approach to search for optimal cost values. Both methods compute minimal-cost reconstruction of the evolutionary history between hosts and viruses.

For Jane 4.0, two different sets of costs were applied. Set 1 corresponds to the default settings (cospeciation = 0, host shift = 2, all other events = 1). For set 2, cospeciation was assigned a cost of −1 (a negative cost maximizes the number of inferred cospeciations), whereas all other costs were set at 1. The vertex cost model was used and default parameters were set for generations ($n = 100$) and population size ($n = 100$). Statistical significance was established by 500 random tip mapping permutations and 500 random parasite tree permutations with Yule beta parameter equal to −1.

CoRe-PA analysis was performed with automatic estimation of the optimal cost setting and computed reconstructions of 10,000 random cost sets. Statistical significance was assessed with 1,000 random virus–host associations.

Detection of Positive Selection

To investigate whether positive selection acted on the internal branches of the NW arenavirus phylogeny, we applied the branch-site tests from the PAML suite (Zhang et al. 2005) and BUSTED (branch-site unrestricted statistical test for episodic diversification) (Pond et al. 2005; Murrell et al. 2015). The branch-site test compares a model (MA) that allows positive selection on one or more lineages (foreground lineages) with a model (MA1) that does not allow such positive selection. Twice the difference of likelihood for the two models (ΔlnL) is then compared with a χ^2 distribution with one degree of freedom (Zhang et al. 2005). To ensure consistency, analyses were run using two different codon frequency models (F3x4 and F61) and different initial omega values. Very similar results were obtained in all analyses.

BUSTED is designed to detect the action of positive selection that is acting on a subset of branches in the phylogeny in at least one site within the alignment (Murrell et al. 2015). In analogy to the MA/MA1 models, branches can be specified *a priori*, but an unrestricted branch-site random effects model is applied (Kosakovsky Pond et al. 2011), which allows dN/dS to vary from branch to branch across the entire phylogeny.

Selected sites on specific branches were identified with the BEB procedure implemented in PAML (with *P* value cutoff of 0.95) and with MEME (with the default cutoff of 0.1) (Murrell et al. 2012).

The overlap among sites selected in OW and NW arenaviruses was evaluated by aligning the LASV AV (AY179171) and the MACV reference (AY624354) sequences.

Results

Time Frame of Mammarenavirus Divergence and Speciation

The preferential association of OW and NW mammarenaviruses with rodents from different subfamilies (Murinae and Sigmodontinae/Neotominae) led some authors to suggest that these viruses coevolved and possibly cospeciated with their hosts (Gonzalez et al. 2007). Because the Murinae and Cricetidae families diverged ~20 Ma (Steppan and Schenk 2017), the hypothesis of cospeciation implies that the OW and NW mammarenavirus lineages also separated in very

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

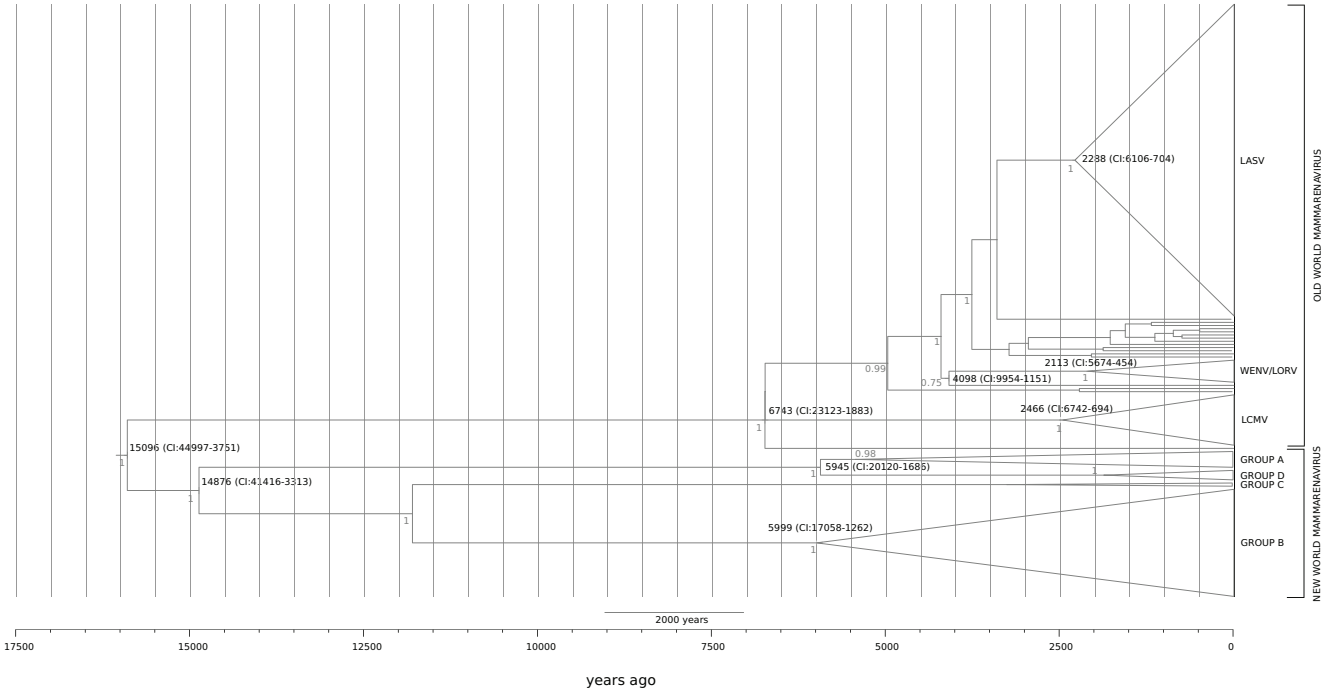

Fig. 1.—Mammarenavirus timescaled phylogenetic tree. The timescaled phylogenetic tree was estimated using the RdRp region. Branch lengths represent time (expressed in years ago) and the scale bar is shown at the tree base. The tMRCA of selected nodes is reported with 95% confidence intervals. Posterior probability of relevant nodes is also reported.

ancient times. We thus decided to formally test this possibility using molecular dating. A common problem associated with molecular dating, especially for fast-evolving viruses, is the time-dependent variation in evolutionary rates (Duchene et al. 2014). In fact, for most viruses (and other organisms), evolutionary rates appear to scale negatively with the time-frame of measurement. This phenomenon often results in severe underestimation of the age of viral lineages and is strongly associated with purifying selection (i.e., the elimination of deleterious mutations) and substitution saturation (i.e., the occurrence of multiple substitutions at the same position) (Duchene et al. 2014). This is because transient deleterious mutations inflate short-term rate estimates, whereas saturation (which decreases the measured substitution rate) is more likely to occur over long timeframes.

Therefore, we applied a selection-informed method which allows for site- and branch-specific variation in selective pressure and thus accounts for the effect of purifying selection (Wertheim and Kosakovsky Pond 2011; Wertheim et al. 2013).This approach was previously applied (Wertheim et al. 2013) to show that the divergence time of coronaviruses is in the range of millions of years, broadly consistent with the hypothesis of cospeciation of mammal-infecting and bird-infecting viruses with their hosts (Wertheim et al. 2013).

We retrieved sequence information for the RdRp fragment (in the *L* gene) of 188 mammarenaviruses sampled over 45 years. No evidence of substitution saturation was detected in the alignment (1,580 nucleotides) (supplementary table S4 and fig. S1, Supplementary Material online).

To determine whether the RdRp phylogeny had sufficient temporal structure, we calculated the correlation coefficient (*r*) of a regression of root-to-tip genetic distances against sequence sampling times (Murray et al. 2016). Evidence for temporal structure was obtained (supplementary fig. S2, Supplementary Material online).

For dating, we calculated tree branch lengths using the selection-informed aBS-REL model (Smith et al. 2015). Branch lengths were converted into time estimates using LSD (least-squares dating) (To et al. 2016). Confidence intervals were obtained by estimating the variance in branch lengths produced by aBS-REL. Using this approach, we obtained a tMRCA for mammarenaviruses of 15,906 years ago (ya) (CI: 44,996–3,759 ya) (fig. 1). This result clearly is not consistent with the hypothesis of host–virus cospeciation. tMRCAs of 41,416–3,313 and 23,123–1,883 ya were obtained for the NW and OW lineages, respectively (fig. 1). Analysis of the internal nodes of the RdRp phylogeny indicated that extant LASV and LCMV originated ~6,000–700 ya (fig. 1). Similarly, Asian OW mammarenaviruses (WENV/LORV) appeared 5,674–454 ya. Finally, the Northern American NW mammarenaviruses separated from the other NW mammarenaviruses 20,120–1,686 ya (fig. 1).

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

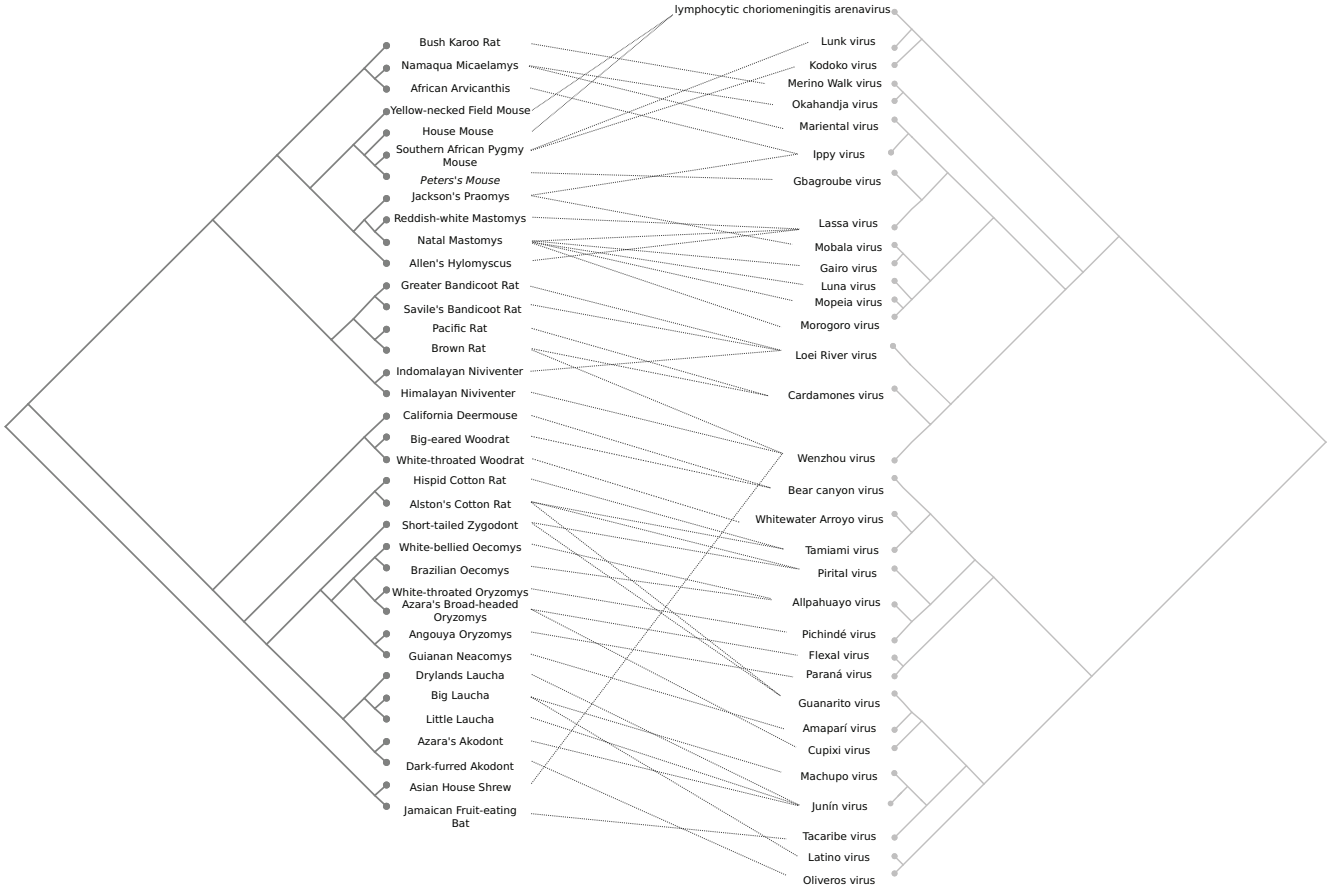

Fig. 2.—Associations between mammarenaviruses and their known hosts. A total of 49 associations are drawn. When data for a natural host were not available (i.e., Argentine Akodont; *Necromys benefactus* and Long-tailed Field Mouse; *Apodemus sylvaticus*), the most closely related available rodent was used (Dark-furred Akodont; *Necromys obscurus* and Yellow-necked Field Mouse; *Apodemus Flavicollis*). In the case of Flexal virus, which was isolated from unidentified members of the oryzomyini tribe (Radoshitzky et al. 2015), association was drawn with Azara's Broad-headed Oryzomys (*Hylaeamys megacephalus*, also known as *Oryzomys capito*).

Cophylogenetic Analysis of Mammarenaviruses and Their Hosts

We next tested the hypothesis of mammarenavirus–host cospeciation more directly by using reconciliation analysis. We retrieved 49 mammarenavirus–host association from literature sources (fig. 2 and supplementary table S3, Supplementary Material online) and the topology of the viral (RdRp region) and mammalian phylogenies were compared using two event-based methods for cophylogenetic analysis, Jane (Conow et al. 2010) and CoRe-Pa (Merkle et al. 2010).

Event-based cophylogeny methods apply cost schemes to different evolutionary events to test the level of congruence between the host and parasite (virus) trees. Minimization of the overall costs is used to infer the most parsimonious scenario to explain the observed host–parasite associations. Both Jane and CoRe-Pa allow analysis of viruses with multiple hosts.

Jane was run using two different cost schemes that assign the cheapest cost to cospeciations. In both cases, evidence for significant cophylogeny was obtained, as the total costs resulting from random permutations were always greater than those of the reconstructed isomorphic solutions (permutation P values < 0.002) (table 1).

For both cost sets, several isomorphic solutions were retrieved, all of them with similar numbers of events and the same cost (table 1). Despite the use of cost sets that favored cospeciations, these latter did not explain the cophylogenetic patterns, as other events, including host shifts, were more common (table 1).

Because the outcome of event-based cophylogenetic analyses strongly depends on the adopted cost scheme (Merkle et al. 2010), we also performed cophylogenetic analyses with CoRe-PA. This method applies several random cost schemes to the data and ranks cophylogenetic reconstructions based on their quality values (q_c). CoRe-Pa yielded 52 reconstructions, the preferred one (q_c=0.0227) had 23 codivergence events, 63 sortings, 17 duplications, and 8 host switchings. The second-best solution had a similar number of events

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

Table 1
Results of Cophylogenetic Analyses

Jane 4.0

Cost Set	Total Cost	Cospeciations	Duplications	Host Shifts	Losses	Failure to Diverge	*P* Value (tip mapping)[a]	*P* Value (parasite tree)[b]
1	72	14–15	4	20–21	17–19	9	<0.002	<0.002
2	37	13–15	3–4	20–23	15–19	9	<0.002	<0.002

CoRe-Pa

Reconstruction (q_c)	Total Cost	Cospeciations (cost)	Duplications (cost)	Host Shifts (cost)	Sortings (cost)	*P* Value[c]
1 (0.0227)	16.75	23 (0.192)	17 (0.239)	8 (0.499)	63 (0.068)	0.189
2 (0.0460)	16.24	23 (0.180)	19 (0.258)	6 (0.503)	71 (0.059)	0.189

[a]*P* value obtained using the tip mapping permutation method.
[b]*P* values obtained using the parasite tree permutation method.
[c]*P* value for cospeciation events.

(table 1) and a q_c of 0.046. All other reconstructions had considerably worse quality values (q_c>0.075). Although CoRe-Pa inferred a higher number of cospeciation events (and fewer host shifts) compared with Jane, analysis of random mammarenavirus–host associations indicated that 18.9% of these led to a reconstruction with 23 or more cospeciation events (table 1). Thus, both Jane and CoRe-Pa indicated that cospeciation did not play a major role in the observed associations between mammarenaviruses and their hosts.

Phylogeography of Mammarenaviruses

We next investigated the geographic origin of mammarenaviruses. To this aim, we analyzed the RdRp alignment using two methods for phylogeographic reconstruction, namely the discrete phylogeography analysis in BEAST (Lemey et al. 2009; Bouckaert et al. 2014) and the BBM (Bayesian Binary MCMC) method (Ronquist and Huelsenbeck 2003). Sequences were assigned to geographic areas (supplementary table S1, Supplementary Material online), which represent the character states for ancestral state reconstruction.

The geographic origin of the mammarenavirus ancestor could not be confidently assigned, as the two methods yielded different results with relatively low posterior probabilities (fig. 3). Conversely, both methods assigned the origin of NW mammarenaviruses to Latin America and the Caribbean with high confidence. OW mammarenaviruses were inferred to have originated in Southern Africa, although the probability obtained with BEAST was not very high (fig. 3). However, the second most likely location obtained with both BEAST and BBM was also in Africa, indicating this continent as the original location of OW mammarenaviruses.

Both LCMV and the lineage of WENV/LORV were found to have separated in Southern Africa from other OW arenaviruses, and the MRCA of extant LCMV sequences was inferred to be of Northern American origin (fig. 3).

Evolution of NW Mammarenavirus Coding Sequences

We previously showed that the speciation of OW arenaviruses was accompanied by intense positive selection that mainly acted on the L protein and, in the case of LCMV, on the NP sequence (Pontremoli et al. 2017). Positive selection is characterized by an accumulation of favorable amino acid-replacing substitutions (e.g., changes that favor host adaptation or allow immune evasion), which results in more nonsynonymous changes than expected under neutrality. Positive selection is thus defined by a nonsynonymous/synonymous substitution rate ratio (dN/dS) >1 and analysis of dN/dS variation from site to site and from branch to branch in a phylogeny can provide important insight into selective events.

We thus investigated whether positive selection also drove the evolution of NW arenavirus coding sequences. To this purpose we retrieved information of all available NW arenavirus complete *L* and *S* segments (supplementary table S2, Supplementary Material online) and we separately aligned the four coding sequences. Because poor alignment quality represents a major source of false positives in evolutionary inference, we filtered unreliably aligned codons. This procedure resulted in the filtering of minor portions of codons in all genes (table 2). Filtered alignments did not show substantial substitution saturation (supplementary table S4, Supplementary Material online) and no evidence of recombination. Previous studies indicated that recombination events at the 3′ end of *GPC* occurred in the ancestor of Northern American NW viruses (Grande-Pérez et al. 2016). Our failure to detect recombination in this region may derive from the fact that we filtered some codons or from the choice to separately analyze *GPC* and *NP*, thus reducing the power to detect phylogenetic inconsistency. Indeed, we obtained different phylogenies for *GPC* and *NP* (fig. 4*A*). However, GARD analysis was performed to avoid the inflation of positive selection inference caused by unrecognized recombination. Because we

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

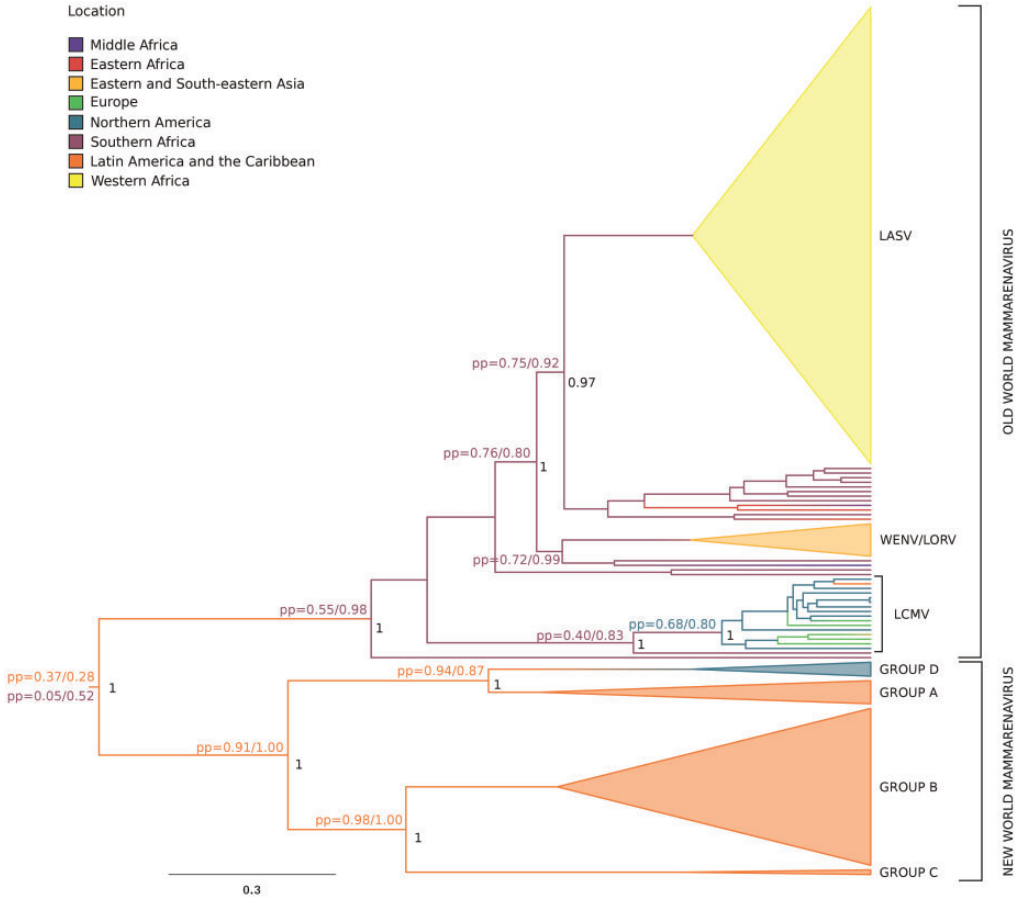

Fig. 3.—Phylogeographic analysis of the *Mammarenavirus* genus. Bayesian maximum clade credibility (MCC) tree for the RdRp region. Branches are colored according to inferred ancestral locations. Posterior probability support (pp) for relevant node locations are shown, as calculated with both the BEAST discrete model and BBM. Posterior probability of relevant nodes is also reported in black.

Table 2

Analysis of Positive Selection for NW Mammarenavirus Genomes

Region	Number of Sequences	Alignment Length (nt)	Filtered Codons (%)	Tree Length (substitutions/site)	Number of Positively Selected Sites
L	46	6,948	7.0	57.14	53[a]
Z	46	306	17.8	42.47	0
NP	63	1,734	2.7	59.19	0
GPC	63	1,671	11.8	60.22	0

[a]Sites identified by both BEB and MEME.

detected no positive selection acting on *GPC* (see below) the issue of recombination was not explored further.

Two branch-site methods, BUSTED and the *codeml* MA/MA1 models (Zhang et al. 2005; Murrell et al. 2015), were used to search for positive selection events along the internal branches of the phylogeny. These two approaches can detect selection along *a priori* specified branches in a phylogeny. Selection was declared if, after false discovery rate (FDR) correction, both methods provided statistical support for selection on a given branch.

No evidence of positive selection was detected for the *GPC*, *NP*, and *Z* regions, although in this latter power to detect selection was probably limited by the short alignment and by the codon filtering procedure (table 2). Conversely, strong evidence of positive selection was observed at all tested internal branches in the *L* gene phylogeny (fig. 4*A* and supplementary table S5, Supplementary Material online).

Specific sites that were positively selected on these branches were identified using the BEB analysis implemented in PAML and with MEME. A total of 53 positively selected sites were detected by both methods (table 2 and supplementary table S6, Supplementary Material online). Several of these sites were located on the branch that separates groups B and C from groups A and D (fig. 4*A*).

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

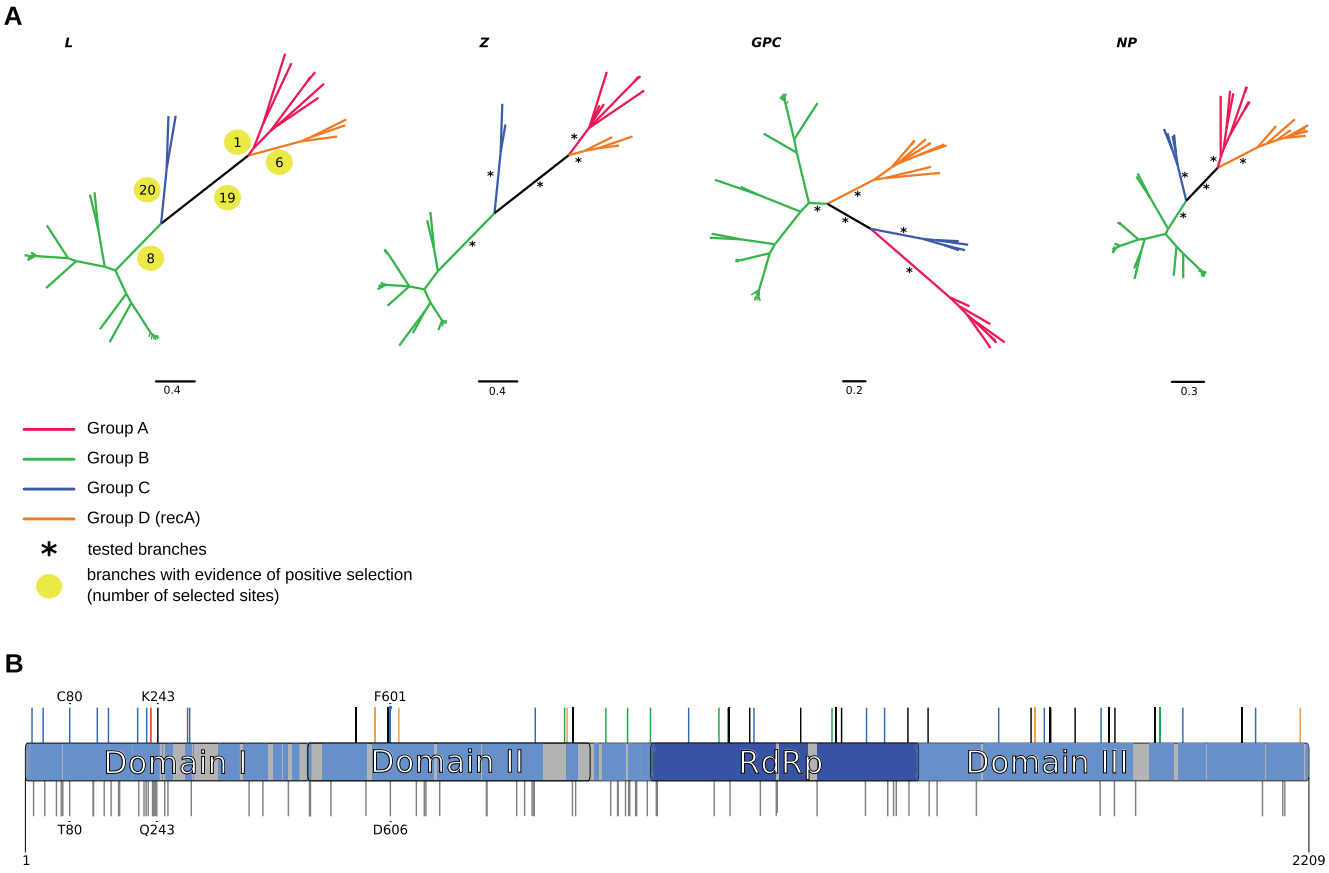

Fig. 4.—Positive selection in NW mammarenaviruses. (*A*) Phylogenetic trees for *L*, *Z*, *GPC*, and *NP*. Asterisks denote tested branches, the yellow highlight indicates evidence of positive selection detected using two methods (BUSTED and the PAML branch-site models). (*B*) Positively selected sites are mapped onto a schematic representation of the NW mammarenavirus L protein (numbers refer to the MACV sequence, GenBank ID: AY624354). Circles indicate sites that are positively selected on more than one branch; color codes denote branches, as in panel (*A*). Regions where codons were filtered by GUIDANCE2 are shown in light gray. As a comparison, sites that were detected as positively selected in the OW mammarenavirus L protein are shown in dark gray below the cartoon structure. Sites that are positively selected in both NW and OW mammarenaviruses (based on the alignment between MACV and the LASV AV strain) are marked with numbers. As limited information is available for NW mammarenaviruses, the domain structure is based on the OW mammarenavirus L protein.

Selection was also strong on the branch leading to the group C lineage (fig. 4*A*).

Mapping of positively selected sites on the L protein indicated that they are scattered throughout the sequence (fig. 4*B*). Alignment of OW and NW mammarenavirus L proteins indicated that 3 sites that are positively selected in NW viruses were also previously detected in OW viruses (Pontremoli et al. 2017), indicating that selection independently targeted specific residues in the two viral groups (fig. 4*B*).

Discussion

The preferential association of OW and NW mammarenaviruses with specific subfamilies of rodents was proposed to derive from long-term evolutionary relationships and, possibly, cospeciation (Bowen et al. 1997; Gonzalez et al. 2007). This hypothesis is however controversial (Emonet et al. 2009) and previous works based on the comparison of the viral and host phylogenies produced different results (Hugot et al. 2001; Jackson and Charleston 2004; Coulibaly-N'Golo et al. 2011; Irwin et al. 2012). We estimated a low bound tMRCA for OW and NW mammarenavirus lineages at ~45,000 ya. This time-frame is clearly not consistent with the hypothesis of codivergence. Similarly, we estimated that the NW mammarenaviruses shared an MRCA 20,120–1,686 ya, again inconsistent with the divergence of the Sigmodontinae and Neotominae subfamilies (~15 Ma) (Steppan and Schenk 2017).

In line with these data, cophylogenetic reconciliation analyses estimated different numbers of cospeciation events, depending on the method, but indicated that such events do not represent the major determinant of observed mammarenavirus–host associations. Thus, as previously suggested for NW mammarenaviruses (Irwin et al. 2012), as well as for other rodent-infecting viruses (Ramsden et al. 2009), we

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

propose that the association of mammarenaviruses with specific host subfamilies is due to geographic factors.

Under the cospeciation hypothesis, mammarenaviruses were implicitly thought to have emerged in Asia, where rodents also originated, and to have migrated to Europe, Africa, and the Americas with their hosts (Gonzalez et al. 2007). Even accounting for virus extinction events, this scenario does not explain why mammarenaviruses other than LCMV have never been detected in European rodents (Blasdell et al. 2008; Ledesma et al. 2009; Yama et al. 2012; Forbes et al. 2014) and Asian mammarenaviruses were assigned an African origin in the analyses herein. Indeed, the inferred time of origin for the WENV/LORV lineage (ranges: 5,674–454 ya) is in theory consistent with human-mediated introduction: trade routes connecting Asia and the Eastern African coasts were already active around the first century CE, and resulted in the exchange of goods, plants, and animals, including rodents (Beaujard 2007; Boivin et al. 2013). Unfortunately, we were unable to reconstruct the geographic origin of mammarenaviruses and our data do not explain how these viruses came to infect rodents in Africa and the Americas. One possibility is that OW and NW mammarenaviruses originated from a common ancestor that was present in both continents. A reptilian arenavirus may be an appealing candidate, as some reptarenaviruses can infect mammalian cells in laboratory settings (Hetzel et al. 2013; Hepojoki, Kipar, et al. 2015; Abba 2016). However, data on the origin, diversity, and geographic range of reptarenaviruses and hartmaniviruses are presently missing and these viruses have only been detected in captive snakes. This latter observation raises the possibility that reptilian viruses originated by the recent cross-species transmission of mammarenaviruses via ingestion of supplied infected rodents. This is highly unlikely, though, as reptarenaviruses and hartmaniviruses differ considerably from each other and substantially from mammarenaviruses (Stenglein et al. 2012, 2015; Bodewes et al. 2013; Hetzel et al. 2013; Hepojoki, Salmenperä, et al. 2015). In the arenavirus phylogeny, reptarenaviruses and mammarenaviruses form sister clades, with hartmaniviruses creating the most basal lineage (Hepojoki, Salmenperä, et al. 2015). Thus, as previously noted (Hetzel et al. 2013), if transmission to snakes occurred from a mammalian host, this must have happened long ago, leading to the diversification of the three virus genera. An alternative possibility is that both mammalian and reptilian arenaviruses originated from multiple transmissions from one or more unknown reservoirs. Although mammarenaviruses have been mainly detected in rodents, TCRV was described in bats (Downs et al. 1963) and WENV was isolated from both rodents and shrews, suggesting that it has a relatively broad host range (Li et al. 2015). These observations imply that additional mammalian hosts (and arenaviruses) may exists and that the prevalent association with rodents may derive from biased ascertainment in rodents and undersampling of nonrodent mammals. Addressing this possibility will need extensive field work to characterize arenavirus diversity in different hosts and geographic areas.

Indeed, a limitation of the analyses presented here is that they are very sensitive to the exclusion of undescribed extant species or basal extinct lineages. Both time frames and geographic ranges were based on extant strains and may therefore be affected by sampling biases and incomplete knowledge of mammarenavirus diversity (e.g., from geographic areas that have been underinvestigated). These problems may be particularly severe for LCMV, as this is the only mammarenavirus to be detected in multiple continents, most likely by virtue of its association with commensal house mice. However, available LCMV sequences were mostly sampled in United States and Europe, with only two sequences from Asia (both from Japan). The tMRCA of LCMV strains we obtained (6,742–694 ya) is roughly consistent with previous estimates (5,000–3,000 ya; Albariño et al. 2010), and phylogeography placed the origin of this virus in Northern America. It is clearly difficult to imagine how an ancestral African mammarenavirus could reach the NW before human transatlantic travel was developed, unless a nonrodent host was involved. Thus, extensive sampling of LCMV sequences from different continents may reveal a different geographic origin for this virus. Notably, a recent search for LCMV in Gabon showed that the virus was introduced in the country, most likely with its rodent host, from America (N'Dilimabaka et al. 2015). Although LCMV sampling in Africa is still limited, these data suggest that an ancestral African mammarenavirus migrated and speciated via geographical isolation into LCMV, to be reintroduced into the continent more recently.

Another cautionary note on our tMRCA calculations relates to the problems associated with the reliable estimation of the age of viral lineages (Duchene et al. 2014). Although we applied a selection-informed approach that accounts for the effect of purifying selection (Wertheim and Kosakovsky Pond 2011; Wertheim et al. 2013) and we did not detect significant saturation in the alignment, we possibly failed to fully correct for time-dependent substitution rate variation. However, even if our estimates were one or two orders of magnitude too recent, the hypothesis of codivergence between mammarenaviruses and their rodent hosts would still be unsupported.

Finally, we mention that, although we only included viral isolates that underwent few passages (in cell lines or mouse brain), some variants may still have been introduced and become selected under laboratory growth conditions. Such variants are however expected to be few and they are thus unlikely to affect dating and phylogeographic inferences. Conversely, the possibility exists that variants selected as a consequence of culture adaptation have an effect on the inference of positive selection. However, we only tested the internal branches of the NW mammarenavirus phylogeny, meaning that the effect of selected variants on individual tip branches has minimal effect.

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

Similar considerations apply to the fact that the sequences analyzed herein were obtained using different techniques, which may in turn have resulted in different degrees of sequencing errors. In fact, sequencing errors are expected to be casual and most likely unrelated to geographic origin, viral species or other features. For these reasons, they are unlikely to affect phylogeographic analyses and inference of positive selection. Concerning molecular dating, it is worth noting that the date of sample isolation does not necessarily correspond to the time (and consequently technique) of sequencing. For instance, several old LCMV isolates were sequenced in 2010 from stocks (Albariño et al. 2010); likewise, PICV, AMAV, WWAV, and MACV isolates from the Sixties and Seventies were sequenced in 2007–2008 (Cajimat et al. 2007; Lan et al. 2008) or recently deposited in GenBank as part of a large BioProject (PRJNA257008). These observations suggest that heterogeneity in sequencing techniques does not affect the dating analyses.

A previous analysis of the selective events that accompanied OW speciation indicated that the viral polymerase was the preferential target of selection (Pontremoli et al. 2017). Results herein show that the *L* gene evolved under strong positive selection in NW mammarenavirus, as well, indicating that changes in the activity, specificity or other features of this viral enzyme contributes to mammarenavirus adaptation.

An interesting possibility is that changes in L modulate mammarenavirus virulence, which may in turn contribute to the establishment of new reservoir host populations. Indeed, mammarenaviruses seem to display relatively broad host ranges. Several OW and NW mammarenaviruses infect humans and evidence of mammarenavirus infection was reported in animal species other than the natural reservoirs (Grande-Pérez et al. 2016). In addition, models for arenavirus infections have been established in rodents distinct from the natural hosts, as well as in nonhuman primates (Golden et al. 2015). However, at least in experimental settings, mammarenavirus infection of nonnatural hosts often result in a severe pathology and even lethality (Golden et al. 2015). Optimal natural reservoirs, though, tolerate infection with limited consequence, thus contributing to viral maintenance in the population. Thus, changes in the polymerase may be required to finely tune viral persistence and virulence after host switches. In OW mammarenaviruses, changes in the L protein were previously associated with differential replication efficiency and different disease phenotypes in rodents (Matloubian et al. 1993; Bergthaler et al. 2010; Ng et al. 2011). For instance, rapid viral replication in the early phases of infection was shown to be central for the establishment of a persistent infection for some LCMV strains (Bergthaler et al. 2010; Sullivan et al. 2015). In the case of Pichindé virus, a NW mammarenavirus used as a model for LASV pathogenesis, three amino acid substitutions in the polymerase account for the different phenotype of two laboratory strains (McLay et al. 2013). The P18 strain causes fatal disease in guinea pigs, whereas a related strain (P2) only causes a febrile disease with modest weight loss. The different virulence is a consequence of increased efficiency of P18 viral genomic replication (McLay et al. 2013).

However, these observations were obtained in experimental models, often inoculated intravenously or intraperitoneally at high viral doses. It thus remains to be evaluated whether similar effects would be evident following natural infection and how variation in the viral polymerase modulate viral phenotypes such as disease severity and duration in natural host populations.

Recently, Khamina et al. (2017) used a human cell line to define the interactome of the LCMV L protein. They found a consistent number of interactors, including components of the innate immune response such as TRIM21, DDX3X, and NKRF. Consistently, in vitro inhibition of these proteins affected LCMV propagation, and two weeks after intravenous inoculation, *Trim21*$^{-/-}$ mice had higher viral titers than wild-type animals (Khamina et al. 2017). Although the significance of these findings during natural infections remains to be evaluated, these results indicate that positive selection at mammarenavirus L proteins might result from interaction with the host immune system. In line with this view, a recent work also indicated that the L protein of Mopeia virus can activate the RLR/MAVS signaling pathway, possibly via the production of small RNAs (Zhang et al. 2016). Such activation results in the induction of type I IFN responses. Different mammarenaviruses and even different strains of the same virus elicit distinct immune responses (Pannetier et al. 2004; Hayes et al. 2012; Huang et al. 2012, 2015; Meyer and Ly 2016). Whereas the role of mammarenavirus NP and Z proteins as modulators of the IFN system are established (Martinez-Sobrido et al. 2007; Xing et al. 2015), whether and to which degree these differences are determined by adaptive changes in the L protein remain open issues.

Overall, the role of mammarenavirus L proteins as drivers of viral evolution deserve further exploration, both for shedding light on mammarenavirus biology and because virus attenuation is regarded as a promising approach for antiviral therapy and vaccination (Grande-Pérez et al. 2016).

Acknowledgment

The authors received no specific funding for this work.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Literature Cited

Abba Y. 2016. *In vitro* isolation and molecular identification of reptarenavirus in Malaysia. Virus Genes 52(5):640–650.

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

Albariño CG, et al. 2010. High diversity and ancient common ancestry of lymphocytic choriomeningitis virus. Emerg Infect Dis. 16(7): 1093–1100.
Beaujard P. 2007. East Africa, the Comoros Islands and Madagascar before the sixteenth century. Azania Archaeolog Res Afr 42(1):15–35.
Bergthaler A, et al. 2010. Viral replicative capacity is the primary determinant of lymphocytic choriomeningitis virus persistence and immunosuppression. Proc Natl Acad Sci U S A. 107(50):21641–21646.
Blasdell KR, Becker SD, Hurst J, Begon M, Bennett M. 2008. Host range and genetic diversity of arenaviruses in rodents, United Kingdom. Emerg Infect Dis. 14(9):1455–1458.
Blasdell KR, et al. 2016. Evidence of human infection by a new mammarenavirus endemic to Southeastern Asia. Elife 5: e13135.
Boc A, Diallo AB, Makarenkov V. 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Res. 40:W573–W579.
Bodewes R, et al. 2013. Detection of novel divergent arenaviruses in boid snakes with inclusion body disease in The Netherlands. J Gen Virol. 94(Pt 6):1206–1210.
Boivin N, Crowther A, Helm R, Fuller DQ. 2013. East Africa and Madagascar in the Indian Ocean world. J World Prehist. 26(3):213–281.
Bouckaert R, et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 10(4):e1003537.
Bowen MD, Peters CJ, Nichol ST. 1997. Phylogenetic analysis of the Arenaviridae: patterns of virus evolution and evidence for cospeciation between arenaviruses and their rodent hosts. Mol Phylogenet Evol. 8(3):301–316.
Cajimat MN, Milazzo ML, Hess BD, Rood MP, Fulhorst CF. 2007. Principal host relationships and evolutionary history of the North American arenaviruses. Virology 367(2):235–243.
Charrel RN, de Lamballerie X. 2010. Zoonotic aspects of arenavirus infections. Vet Microbiol. 140(3–4):213–220.
Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol Biol. 5(1):16.
Coulibaly-N'Golo D, et al. 2011. Novel arenavirus sequences in *Hylomyscus sp.* and *Mus (Nannomys) setulosus* from Cote d'Ivoire: implications for evolution of arenaviruses in Africa. PLoS One 6(6):e20893.
Downs WG, Anderson CR, Spence L, Aitken TH, Greenhall AH. 1963. Tacaribe virus, a new agent isolated from *Artibeus* bats and mosquitoes in Trinidad, West Indies. Am J Trop Med Hyg. 12(4):640–646.
Duchene S, Holmes EC, Ho SY. 2014. Analyses of evolutionary dynamics in viruses are hindered by a time-dependent bias in rate estimates. Proc Biol Sci. 281:20140732.
Emonet SF, de la Torre JC, Domingo E, Sevilla N. 2009. Arenavirus genetic diversity and its biological implications. Infect Genet Evol. 9(4): 417–429.
Forbes KM, et al. 2014. Serological survey of rodent-borne viruses in Finnish field voles. Vector Borne Zoonotic Dis. 14(4):278–283.
Golden JW, Hammerbeck CD, Mucker EM, Brocato RL. 2015. Animal models for the study of rodent-borne hemorrhagic fever viruses: arenaviruses and Hantaviruses. Biomed Res Int. 2015:1.
Gonzalez JP, Emonet S, de Lamballerie X, Charrel R. 2007. Arenaviruses. Curr Top Microbiol Immunol. 315:253–288.
Grande-Pérez A, Martin V, Moreno H, de la Torre JC. 2016. Arenavirus quasispecies and their biological implications. Curr Top Microbiol Immunol. 392:231–276.
Guindon S, Delsuc F, Dufayard JF, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol. 537:113–137.
Hayes MW, et al. 2012. Pathogenic Old World arenaviruses inhibit TLR2/Mal-dependent proinflammatory cytokines in vitro. J Virol. 86(13):7216–7226.
Hepojoki J, Kipar A, et al. 2015. Replication of boid inclusion body disease-associated arenaviruses is temperature sensitive in both boid and mammalian cells. J Virol. 89:1119–1128.
Hepojoki J, Salmenperä P, et al. 2015. Arenavirus coinfections are common in snakes with boid inclusion body disease. J Virol. 89(16): 8657–8660.
Hetzel U, et al. 2013. Isolation, identification, and characterization of novel arenaviruses, the etiological agents of boid inclusion body disease. J Virol. 87(20):10918–10935.
Huang C, et al. 2015. Highly pathogenic new world and old world human arenaviruses induce distinct interferon responses in human cells. J Virol. 89(14):7079–7088.
Huang C, et al. 2012. Junín virus infection activates the type I interferon pathway in a RIG-I-dependent manner. PLoS Negl Trop Dis. 6(5): e1659.
Hugot JP, Gonzalez JP, Denys C. 2001. Evolution of the Old World Arenaviridae and their rodent hosts: generalized host-transfer or association by descent? Infect Genet Evol. 1(1):13–20.
Irwin NR, Bayerlova M, Missa O, Martinkova N. 2012. Complex patterns of host switching in New World arenaviruses. Mol Ecol. 21(16): 4137–4150.
Jackson AP, Charleston MA. 2004. A cophylogenetic perspective of RNA-virus evolution. Mol Biol Evol. 21(1):45–57.
Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013. Genetic tracking of mice and other bioproxies to infer human history. Trends Genet. 29(5):298–308.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kerber R, et al. 2015. Research efforts to control highly pathogenic arenaviruses: a summary of the progress and gaps. J Clin Virol. 64:120–127.
Khamina K, et al. 2017. Characterization of host proteins interacting with the lymphocytic choriomeningitis virus L protein. PLoS Pathog. 13(12):e1006758.
Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol. 23(10):1891–1901.
Kosakovsky Pond SL, et al. 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol Biol Evol. 28(11): 3033–3043.
Lan S, McLay L, Aronson J, Ly H, Liang Y. 2008. Genome comparison of virulent and avirulent strains of the Pichindé arenavirus. Arch Virol. 153(7):1241–1250.
Lecompte E, ter Meulen J, Emonet S, Daffis S, Charrel RN. 2007. Genetic identification of Kodoko virus, a novel arenavirus of the African pigmy mouse (*Mus Nannomys minutoides*) in West Africa. Virology 364(1): 178–183.
Ledesma J, et al. 2009. Independent lineage of lymphocytic choriomeningitis virus in wood mice (*Apodemus sylvaticus*), Spain. Emerg Infect Dis. 15(10):1677–1680.
Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol. 5(9):e1000520.
Li K, et al. 2015. Isolation and characterization of a novel arenavirus harbored by Rodents and Shrews in Zhejiang province, China. Virology 476:37–42.
Martinez-Sobrido L, Giannakas P, Cubitt B, Garcia-Sastre A, de la Torre JC. 2007. Differential inhibition of type I interferon induction by arenavirus nucleoproteins. J Virol. 81(22):12696–12703.
Matloubian M, Kolhekar SR, Somasundaram T, Ahmed R. 1993. Molecular determinants of macrophage tropism and viral persistence: importance of single amino acid changes in the polymerase and glycoprotein of lymphocytic choriomeningitis virus. J Virol. 67:7340–7349.
McLay L, Lan S, Ansari A, Liang Y, Ly H. 2013. Identification of virulence determinants within the L genomic segment of the Pichindé arenavirus. J Virol. 87(12):6635–6643.

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019

Merkle D, Middendorf M, Wieseke N. 2010. A parameter-adaptive dynamic programming approach for inferring cophylogenies. BMC Bioinformatics 11(Suppl 1):S60.
Meyer B, Ly H. 2016. Inhibition of innate immune responses is key to pathogenesis by arenaviruses. J Virol. 90(8):3810–3818.
Murray GGR, et al. 2016. The effect of genetic structure on molecular dating and tests for temporal signal. Methods Ecol Evol. 7(1):80–89.
Murrell B, et al. 2012. Detecting individual sites subject to episodic diversifying selection. PLoS Genet. 8(7):e1002764.
Murrell B, et al. 2015. Gene-wide identification of episodic selection. Mol Biol Evol. 32(5):1365–1371.
N'Dilimabaka N, et al. 2015. Evidence of lymphocytic choriomeningitis virus (LCMV) in domestic mice in Gabon: risk of emergence of LCMV encephalitis in Central Africa. J Virol. 89:1456–1460.
Ng CT, Sullivan BM, Oldstone MB. 2011. The role of dendritic cells in viral persistence. Curr Opin Virol. 1(3):160–166.
Pannetier D, Faure C, Georges-Courbot MC, Deubel V, Baize S. 2004. Human macrophages, but not dendritic cells, are activated and produce alpha/beta interferons in response to Mopeia virus infection. J Virol. 78(19):10516–10524.
Pond SL, Frost SD, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21(5):676–679.
Pontremoli C, et al. 2017. Evolutionary analysis of Old World arenaviruses reveals a major adaptive contribution of the viral polymerase. Mol Ecol. 26(19):5173–5188.
Posada D, Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14(9):817–818.
Privman E, Penn O, Pupko T. 2012. Improving the performance of positive selection inference by filtering unreliable alignment regions. Mol Biol Evol. 29(1):1–5.
Radoshitzky SR, et al. 2015. Past, present, and future of arenavirus taxonomy. Arch Virol. 160(7):1851–1874.
Ramsden C, Holmes EC, Charleston MA. 2009. Hantavirus evolution in relation to its rodent and insectivore hosts: no evidence for codivergence. Mol Biol Evol. 26(1):143–153.
Ronquist F, Huelsenbeck JP. 2003. MrBayes 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19(12):1572–1574.
Sela I, Ashkenazy H, Katoh K, Pupko T. 2015. GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 43(W1):W7–14.
Smith MD, et al. 2015. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 32(5):1342–1353.
Stenglein MD, et al. 2012. Identification, characterization, and *in vitro* culture of highly divergent arenaviruses from boa constrictors and annulated tree boas: candidate etiological agents for snake inclusion body disease. MBio 3(4):e00180–e00112.
Stenglein MD, et al. 2015. Widespread recombination, reassortment, and transmission of unbalanced compound viral genotypes in natural arenavirus infections. PLoS Pathog. 11(5):e1004900.
Steppan SJ, Schenk JJ. 2017. Muroid rodent phylogenetics: 900-species tree reveals increasing diversification rates. PLoS One 12:e0183070.
Sullivan BM, Teijaro JR, de la Torre JC, Oldstone MB. 2015. Early virus-host interactions dictate the course of a persistent infection. PLoS Pathog. 11(1):e1004588.
To TH, Jung M, Lycett S, Gascuel O. 2016. Fast dating using least-squares criteria and algorithms. Syst Biol. 65(1):82–97.
Wertheim JO, Kosakovsky Pond SL. 2011. Purifying selection can obscure the ancient age of viral lineages. Mol Biol Evol. 28(12):3355–3365.
Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. 2013. A case for the ancient origin of coronaviruses. J Virol. 87(12):7039–7045.
Xia X. 2013. DAMBE5: a comprehensive software package for data analysis in molecular biology and evolution. Mol Biol Evol. 30(7):1720–1728.
Xing J, Ly H, Liang Y. 2015. The Z proteins of pathogenic but not nonpathogenic arenaviruses inhibit RIG-I-like receptor-dependent interferon production. J Virol. 89(11):6161–2955.
Yama IN, et al. 2012. Isolation and characterization of a new strain of lymphocytic choriomeningitis virus from rodents in southwestern France. Vector Borne Zoonotic Dis. 12(10):893–903.
Yu Y, Harris AJ, Blair C, He X. 2015. RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. Mol Phylogenet Evol. 87:46–49.
Zhang J, Nielsen R, Yang Z. 2005. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 22(12):2472–2479.
Zhang LK, et al. 2016. Activation of the RLR/MAVS signaling pathway by the L protein of Mopeia virus. J Virol. 90(22):10259–10270.

Associate editor: Chantal Abergel

Downloaded from https://academic.oup.com/gbe/article-abstract/10/3/863/4920864 by Università degli Studi di Milano user on 20 February 2019