The Avian XPR1 Gammaretrovirus Receptor Is under Positive Selection and Is Disabled in Bird Species in Contact with Virus-Infected Wild Mice

Carrie Martin, Alicia Buckler-White, Kurt Wollenberg, Christine A. Kozak

Laboratory of Molecular Microbiology, Bioinformatics and Computational Biosciences Branch, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA

Xenotropic mouse leukemia viruses (X-MLVs) are broadly infectious for mammals except most of the classical strains of laboratory mice. These gammaretroviruses rely on the XPR1 receptor for entry, and the unique resistance of laboratory mice is due to two mutations in different putative XPR1 extracellular loops. Cells from avian species differ in susceptibility to X-MLVs, and 2 replacement mutations in the virus-resistant chicken XPR1 (K496Q and Q579E) distinguish it from the more permissive duck and quail receptors. These substitutions align with the two mutations that disable the laboratory mouse XPR1. Mutagenesis of the chicken and duck genes confirms that residues at both sites are critical for virus entry. Among 32 avian species, the 2 disabling XPR1 mutations are found together only in the chicken, an omnivorous, ground-dwelling fowl that was domesticated in India and/or Southeast Asia, which is also where X-MLV-infected house mice evolved. The receptor-disabling mutations are also present separately in 5 additional fowl and raptor species, all of which are native to areas of Asia populated by the virus-infected subspecies Mus musculus castaneus. Phylogenetic analysis showed that the avian XPR1 gene is under positive selection at sites implicated in receptor function, suggesting a defensive role for XPR1 in the avian lineage. Contact between bird species and virus-infected mice may thus have favored selection of mouse virus-resistant receptor orthologs in the birds, and our data suggest that similar receptor-disabling mutations were fixed in mammalian and avian species exposed to similar virus challenges.

The transmission of retroviruses to new hosts can result in the emergence of new infectious diseases and can alter the host genomic architecture and gene-regulatory networks, but the factors that determine whether a pathogen can successfully infect a novel host are poorly understood (1). The xenotropic mouse leukemia viruses (X-MLVs) are gammaretroviruses originally isolated from laboratory mice that are unable to infect the cells of these mice (2). These viruses rely on the XPR1 cell surface receptor for entry (3–5), and receptor orthologs from other mammals, like humans and cats, as well as the various species of wild mice, all permit X-MLV entry (6, 7). The resistance of laboratory mouse cells to infection is due to mutations that alter two XPR1 residues, K500E in the putative third extracellular loop (ECL3) and T582Δ in the fourth loop (ECL4) (8).

There is substantial sequence variation in the receptor-determining regions of the mammalian XPR1 orthologs, and some of these receptor variants show altered ability to mediate entry of one or more of the 6 known host range variants in the XP-MLV family of xenotropic/polytropic viruses (6, 9). In addition to X-MLV, this family includes viruses first described as broadly polytropic (P-MLVs) because of their ability to infect mouse cells, as well as cells of other species (10, 11). Other variants, like xenotropic murine leukemia virus-related virus (XMRV) (12) and several wild mouse isolates, have atypically host ranges defined by differences in their abilities to infect cells with variant XPR1 receptors (6, 7). These infectivity differences are due to specific polymorphisms in the receptor and in the viral envelope glycoprotein, and at least 6 XPR1 residues that modulate receptor sensitivity to the various XP-MLVs have been identified (6, 9).

The various XP-MLVs are found as endogenous and infectious viruses in laboratory mouse strains and in the house mouse subspecies. Although the common inbred strains carry multiple copies of both X-MLVs and P-MLVs, these virus subgroups are largely segregated in the three lineages of the house mouse, subspecies with largely nonoverlapping natural ranges in Eurasia (13). Mus musculus domesticus, which carries exclusively P-MLVs, is found in western Europe and the Mediterranean. Mus musculus musculus populates the area from eastern Europe through northern Asia to the Pacific, and Mus musculus castaneus is native to Southeast Asia and India, where Mus originated and which is also the site of the house mouse radiation. M. m. musculus and M. m. castaneus both carry predominantly X-MLVs as germ line copies, and infectious virus has also been isolated from mice of both species. These 3 house mouse subspecies carry 3 different XPR1 variants, none of which corresponds to the disabled laboratory mouse Xpr1 variant (6). M. m. domesticus has a fully permissive XPR1, but the X-MLV-infected subspecies carry two different restrictive Xpr1 variants; these subspecies are resistant to P-MLVs, but the XPR1 receptor of M. m. castaneus is a more efficient X-MLV receptor than that of M. m. musculus. These Eurasian house mouse species are all human commensals and were passively transported by humans to the New World, where M. m. domesticus is now widespread.

Because XP-MLV-infected mice are distributed globally, because all mammals but the laboratory mouse have X-MLV-sus-
ceptible XPR1 receptors, and because XPR1 orthologs are found in all eukaryotes, we looked for naturally occurring restrictive receptors in a related set of species outside the mammalian lineage. We focused on avian species for several reasons. First, early studies on X-MLVs suggested that avian cells are variably susceptible to these viruses (14, 15). Second, some birds and mice share habitats, sometimes with a predator-prey relationship. Third, RNA viruses more readily jump species than other pathogens, and there is documented evidence of epizoonotic retroviral transmission from mammals to birds (16, 17).

**MATERIALS AND METHODS**

Cells, viruses, pseudoviruses, and infection assays. CAST-X is an X-MLV isolated from the spleen of a CAST/EiJ mouse (18). XMRV (12) was kindly provided by R. Silverman (Cleveland Clinic, Cleveland, OH). C5724 MLV was isolated from the spleen of a CZECHI/EiJ mouse 2 months after inoculation with Moloney ecotropic MLV. CasE#1 (19), FrM CF-P-MLV, MoMCF-P-MLV, AKR X-MLV, and NZB-UI-6 X-MLV were originally obtained from J. Hartley (NIAID, Bethesda, MD). LacZ pseudotypes were generated for these viruses by infection of the packaging sanctuary (Clontech), by J. Silver (NIAID, Bethesda, MD), Filtered media from the virus-infected cultures contained a mixture of infectious virus and LacZ pseudovirus.

Susceptibility to replication-competent virus was quantitated by infecting cells with dilutions of virus stocks in the presence of Polybrene (4 μg/ml; Aldrich, Milwaukee, WI). After 4 days, cultures were UV irradiated and overlaid with mink S– L– cells (20). Foci of transformed S+ L+ cells that mark virus-producing infected cells were counted 5 to 7 days later, and titers were expressed as focus-forming units/200 μl.

Virus susceptibility was also assayed in a single-round assay using XP-MLV pseudoviruses to infect Mus dunni cells (21), Chinese hamster E36 cells (22), ferret MA139 cells obtained from J. Hartley, duck cells (ATCC CCL-141), quail cells (ATCC CRL-1708), and primary cell cultures of chicken embryo fibroblasts made from chick embryos. The cells were infected with 10-fold dilutions of pseudotype stocks in the presence of Polybrene. One day after infection, the cells were fixed with 0.4% glutaraldehyde and assayed for β-galactosidase activity using the substrate 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal) (2 mg/ml; ICN Biomedicals, Aurora, Ohio). Infectious titers were expressed as the number of blue cells per 50 μl. P values were determined from a 2-tailed Student’s t test using GraphPad Prism 6.

**Avian XPR1 sequencing.** DNA and RNA were isolated from tissue samples and cultured cells from various bird species (see Table S1 in the supplemental material). Tissue samples for DNA extraction were provided by J. Dean (Smithsonian National Museum of Natural History, Washington, DC) and S. Rasheed (University of Southern California, Los Angeles, CA). Additional DNA samples were provided by R. Kimball (University of Florida, Gainesville, FL) and G. Proudfoot (Vassar College, Poughkeepsie, NY).

Total RNA was extracted from duck, chicken, and quail cells using TRIReagent (Molecular Research Center, Cincinnati, OH). cDNA was synthesized using the SuperScript II reverse transcriptase kit (Invitrogen, Carlsbad, CA). Full-length XPR1 was amplified using primers designed from the sequenced chicken genome (23) (Table 1) and cloned into the expression vector pcDNA3.1/V5-His TOPO (Invitrogen). Our cloned chicken XPR1 differs from the Gallus sequence at one site, A85.

Segments of DNA including XPR1 exons 10 to 13 were amplified from genomic DNA using primers designed from coding or intron sequences based on the sequenced chicken genome or on an alignment of the sequenced chicken, turkey, and zebra finch genomes (Table 1). These PCR primers, as well as others not listed, failed to produce amplicons for some segments in some species, as noted (see Text S1 in the supplemental material). PCR products were cloned into pcR2.1-TOPO (Invitrogen, Carlsbad, CA) for sequencing. The sequenced segments varied from 407 bp (lark) to 1,468 bp (ostrich and rhea).

**Generation of XPR1 mutants.** Six mutant variants of the duck and chicken XPR1 genes were generated by mutagenesis PCR using the QuikChange II site-directed mutagenesis kit (Stratagene, La Jolla, CA). The primers used for mutagenesis are listed in Table 1. A chimeric XPR1 backbone was generated by replacing a segment of the chicken XPR1 with a 590-bp EcoRI-Hpal segment of the duck gene containing the N-terminal end of ECL3. The PCR fragment used for the replacement was generated with the forward and reverse primers 5′-CCCTGATAGAATTCCTCTTCTCTCGTGG and 5′-CCAGAGTTAACCAGATGAAAGGC. Q496K and E579Q were introduced into this chimera. All mutants were confirmed by sequencing.

The recombinant plasmids were transfected into E36 Chinese hamster cells with Fugene (Roche, Indianapolis, IN). E36 has a functional XPR1 receptor, but virus entry through the endogenous receptor is blocked by

---

**TABLE 1** Primers used to amplify full-length XPR1 from duck, quail, and chicken cells and XPR1 exons from genomic DNA and primers used for mutation

<table>
<thead>
<tr>
<th>Primer or primer pair</th>
<th>Sequence (5′–3′)</th>
<th>Forward primer</th>
<th>Reverse primer</th>
<th>Amplicon size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a/c</td>
<td>ATGAAGTTGGCAAGCAGATCATCCTTC</td>
<td>GAATGTTGCAAATTTCCTTTTGTAGG</td>
<td>2,079</td>
<td></td>
</tr>
<tr>
<td>b/d</td>
<td>CGGGGGCAAGTAGTACCTGC</td>
<td>GCTGCAATGTTGCACTTCTTTTGTAGG</td>
<td>2,092</td>
<td></td>
</tr>
<tr>
<td>c/f</td>
<td>CAGAATGTTCAGTGCTGCTCCTCCTC</td>
<td>CTGATTTGTCGAGACAAGCCTC</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>g/h</td>
<td>CATCCTCCTGCTTTGTTGGG</td>
<td>GGAGATGGAGAATCTTCAGGGATCAGACG</td>
<td>891</td>
<td></td>
</tr>
<tr>
<td>g/j</td>
<td>CATTGCTGTTGCTGTTGGG</td>
<td>GTCTGTTGCAAGGGAGAT</td>
<td>902</td>
<td></td>
</tr>
<tr>
<td>k/l</td>
<td>GTAAAGACACACAGCGACACC</td>
<td>GTGGGTAACATTCTCTCTTGTACGAAAGAGG</td>
<td>163</td>
<td></td>
</tr>
<tr>
<td>m1/o1</td>
<td>GCATACTACATITGGCCTC</td>
<td>GTGAAACATCAGTGGGGGGGAAACAG</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>m2/o2</td>
<td>GCGGCTACTTATGGCCTC</td>
<td>GTCGAAATCAGTGGGGGGGAAACAG</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>m3/o3</td>
<td>GCCTGCTACTCTTGGCCTC</td>
<td>GTCGAAATCAGTGGGGGGGAAACAG</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>n/o3</td>
<td>CATTGCTGTTGCTGTTGGG</td>
<td>GTGAAACATCAGTGGGGGGGAAACAG</td>
<td>140</td>
<td></td>
</tr>
</tbody>
</table>

a/ Reverse primers were reverse complements.
glycosylation; these cells show marginal susceptibility to X-MLVs and resistance to all other XP-MLVs (9, 24). At least 2 pools of stable transfec-
tants were selected for each mutant with Geneticin (830 μg/ml; Invitro-
gen).

Expression of XPR1 was confirmed by Western analysis. Proteins were
extracted from transfected cells with M-PER Mammalian Protein Extrac-
tion Reagent (Pierce, Rockford, IL). The expression vector used for XPR1
inserts a V5 epitope at the C terminus; XPR1 expression was detected in
Western blots using anti-V5 antibody (Invitrogen), followed by goat anti-
mouse IgG conjugated with horseradish peroxidase (HRP) (Invitrogen).
The membrane was then striped and incubated with mouse anti-
a-tubulin (Sigma, St. Louis, MO) and goat anti-mouse IgG conjugated with
HRP (Invitrogen).

Selection analysis. DNA sequences were aligned using CLUSTALW2
and improved manually. Two phylogenies were produced, one for a set of
23 sequences that included segments of XPR1 exons 10 to 13 and one for
a set of 31 sequences representing only the receptor-determining regions
in exons 11 and 13. Both phylogenies were cladograms that were manually
constructed to correspond to the consensus Aves phylogeny (25–28).

The codeml program of the PAML4 package (29) was used for maxi-

mum-likelihood analysis of positive selection on codons. This analysis
was performed using two different codon frequency models: the F3x4
model (with codon frequencies estimated from the nucleotide frequen-
cies at each codon site in the data) and the F61 model (with codon frequen-
cies of each of the 61 nonstop codons calculated from the data). The neu-
tral-selection/negative-selection model of codon selection (M7) was com-
pared to the corresponding positive-selection model (M8), which includes
a category for ratios of nonsynonymous to synonymous evolutionary
rates [dN/dS ratios] of >1). The significance of this additional codon
selection category was assessed using a likelihood ratio test of the phylog-
eny likelihoods under the neutral- and positive-selection models. The
significance of the test statistics was calculated using a chi-squared dis-
tribution with 2 degrees of freedom. The Bayes empirical Bayes algorithm
was used to calculate the posterior probability of individual codons expe-
riencing dN/dS ratios of >1.

Nucleotide sequence accession numbers. The GenBank accession
numbers for the sequenced segments are JX294430 to JX294458.

RESULTS
Susceptibility of avian cells to XP-MLVs. Duck and chicken cells
were assessed for susceptibility to infectious XP-MLVs. Cells were
infected with serial dilutions of two X-MLVs, NZB-9-1 and the
wild mouse-derived CAST-X, and two P-MLVs, FrMCF and
MoMCF (Fig. 1A), and foci of infected cells were detected using the mink S”- L– indicator cell line (20). Duck cells were efficiently
infected with both X-MLVs, as were cells of the Asian mouse M.
dunni, which carries the fully permissive Mus Xpr1allele receptor
variant. Duck cells were considerably less susceptible to the P-
MLVs, showing a 1,000- to 10,000-fold titer reduction compared
to that of M. dunni cells (Fig. 1A). In contrast, chicken cells were
completely resistant to 3 of the 4 viruses and showed poor infec-
tivity by CAST-X X-MLV.

We then tested chicken, duck, and quail cells for virus suscep-
tibility in a single-round infection assay using XP-MLV pseudovir-
uses (Fig. 1B). This virus set included 3 X-MLVs, 1 P-MLV, and
2 wild-mouse isolates with novel host ranges, CasE#1 and Cz524
(9). These 6 viruses were chosen because they have distinctively
different host ranges on nonrodent mammalian cells and on cells
expressing mutated variants of Mus Xpr1 (6, 9). The viruses rely
on different but overlapping sets of determinants on Mus Xpr1 for
entry. The pattern of pseudovirus infection for each cell line was
consistent with the results obtained with replicating virus (Fig.
1A). M. dunni cells are efficiently infected by all pseudoviruses, but
the 3 avian cells show three different susceptibility patterns (Fig.
1B). Chicken cells are inefficiently infected by X-MLVs and are
resistant to other XP-MLV pseudoviruses. The quail and duck
cells were efficiently infected by X-MLVs, and both showed re-
suced susceptibility to P-MLV, but quail cells are also poorly sus-
ceptible to the 2 wild-mouse viruses. These results indicate that
infection with XP-MLVs is restricted in chickens at an early stage
in the virus life cycle and that there is no postentry restriction of
X-MLV replication in duck cells.

Functional evaluation and mutational analysis of avian
XPR1 receptors. To determine if susceptibility differences are re-
ceptor mediated, we sequenced the XPR1 receptor genes from the
3 avian cell lines (duck, quail, and chicken). These genes are 94
to 97% identical to one another (see Fig. S1 in the supplemental
material) and are 92 to 99% identical to the XPR1 genes from the
sequenced turkey and zebra finch genomes (NC_015020 and
NW_002198997). The avian XPR1 protein is ~83% identical to the
permissive M. dunni XPR1, although identity drops to 38% in
the receptor-determining ECL4. Comparison of the 3 avian genes
identified 27 replacement mutations (see Fig. S1 in the supplemental
material), but only three substitutions distinguish the vi-
rus-resistant chicken XPR1 from both of the more permissive
duck and quail receptors. Two of the three substitutions are at sites

FIG 1 Infectivity of XP-MLVs in avian cells. (A) Susceptibility of cells to
infectious XP-MLVs. Cells were infected with virus dilutions, UV irradiated 4
days later, and overlaid with mink S”- L– cells. Infected, virus-producing avian
cells induced transformed foci in the indicator cells. The titers represent focus-
forming units in 200 μl plus standard errors of the mean (SEM). (B) Suscep-
tibility to LacZ pseudovirus. Cells were stained to detect the reporter gene 24 h
after infection with pseudovirus dilutions. The titers represent blue cells in 50
μl of virus stock and are presented as the means of 3 to 5 tests plus SEM. In
some cases, there were no positive cells after infection with undiluted virus
stock. The asterisks indicate significant P values determined from 3 or 4 inde-
pendent tests using Student’s t test (*, P ≤ 0.05; **, P ≤ 0.01; ***, P ≤ 0.001;
****, P ≤ 0.0001); infectivity of the avian cells was compared to the permissive
M. dunni (M.d.) mouse line. ch, chicken; qu, quail; du, duck.
that are homologous to the two substitutions that disable the receptor function of laboratory mouse XPR1, namely, K496Q in ECL3 (K500E in mouse) and Q579E in ECL4 (T582/H9004 in mouse, which deletes one of two adjacent Thr residues) (Fig. 2A; see Fig. S1 in the supplemental material). The third substitution that differentiates chicken XPR1 from the other avian gene products, E261D, is in ECL1, a region that has not been implicated in XPR1 receptor function. Previous analysis of Mus Xpr1 variants identi-
fied 6 sites in addition to 500 and 582 that influence entry by the various XP-MLVs (6, 8, 9, 18); all of these sites are found in the C-terminal end of ECL3 or in ECL4, and all 6 of the sites are invariant in the 3 bird genes.

The chicken and duck genes were cloned into an expression vector and transfected into virus-resistant Chinese hamster E36 cells (Fig. 2B). Transfectants were infected with pseudoviruses representative of 4 distinct host range types: an X-MLV, a P-MLV, and the two wild-mouse isolates. The transfected cells largely recapitulated the susceptibility patterns of the chicken and duck cells. Transfectants expressing the chicken gene showed very poor susceptibility to X-MLVs and were resistant to the other XP-MLV pseudoviruses. E36 cells expressing the duck XPR1 were highly susceptible to most XP-MLVs, although susceptibility to P-MLVs, which is marginal in duck cells (Fig. 1), was not detected in these transfectants.

We introduced reciprocal mutations into the cloned chicken and duck genes at the two sites implicated by sequence comparisons (496 and 579) and expressed these constructs in E36 cells (Fig. 2B). Mutations at either of the sites in the defective laboratory mouse Xpr1 gene generate functional X-MLV receptors (8), and improvement of receptor function was also noted for the chicken gene following introduction of these separate mutations, Q496K (P < 0.0001) and E579Q (P = 0.0406). The reciprocal changes in the duck XPR1, K496Q and Q579E, each reduced receptor function (Fig. 2B). The two mutations were not, however, equivalent; both receptors with the permissive K496 residue showed better receptor function than those with Q579. The introduction of both mutations into the chicken gene produced highly functional receptors, suggesting that these residues cooperate to produce an efficient receptor.

The chicken and duck genes also differ substantially in sequence at the N-terminal end of ECL3, residues 423 to 434 (Fig. 2A). This region of XPR1 was shown to be under positive selection in the rodent gene (6) but has not been evaluated for a role in receptor function. Therefore, we created a chimeric avian gene (termed ch/du) containing duck ECL3 codons 419 to 474 on the chicken XPR1 backbone. This substitution alone did not affect receptor function (Fig. 2B), but the addition of the Q496K and E579Q mutations to this chimeric XPR1 produced a better CasE#1 receptor than the chicken XPR1 with only 2 mutations (P = 0.0004) (Fig. 2B). This suggests that residues in the N-terminal end of ECL3 can modulate receptor function.

To determine if other XP-MLVs display similar infectivity patterns and show additive effects with the double mutation, we used pseudoviruses for three additional X-MLVs (NZB-IU-6, AKR6, and XMRV) and an additional P-MLV (MoMCF). These viruses have sequence differences in the receptor-determining VRA region of the viral env and show infectivity differences in mammalian species (6, 19). The three X-MLVs showed infectivity patterns in the transfected cells that resembled each other and CAST-X, and MoMCF, like FrMCF P-MLV, failed to infect any of the transfected cells (data not shown). The additive effects observed for the two chicken constructs with double mutations are thus restricted to the wild-mouse viruses, CasE#1 and Cz524.

Comparative analysis of avian XPR1 genes. To determine the species and geographic distributions of these disabling mutations in birds, we sequenced receptor-determining segments of XPR1 in 30 avian species trapped on 5 continents (Fig. 3; see Table S1 and Text S1 in the supplemental material) and extracted the corresponding segments from the sequenced genomes of the turkey and zebra finch (30, 31). Overrepresented in this data set are Galliformes species related to domestic chickens and species of avian raptors. The mouse and chicken XPR1 proteins are encoded by 15 exons spanning 142.9 kb in the mouse genome and 77.4 kb in the sequenced chicken genome (Fig. 3). The putative 85-residue receptor-determining ECL3 is contained in 3 exons, and the shorter 13-residue ECL4 is encoded by exon 13.

The sequenced segments include the entire ECL4, the variable 5’ end and 3’ end of ECL3, and the 5’ end of intron 11 (see Text S1...
The sequenced exon segments show 93 to 99% identity across all avian species. The avian XPR1 ECL4 is much less variable than the comparable segment of the mammalian gene (2). Rare replacement mutations alter only 3 ECL4 codons in bird species (Fig. 4), whereas for mammalian XPR1 genes, there are substitutions affecting all but 3 ECL4 residues. The 3 conserved mammalian ECL4 sites do not contribute to the virus attachment site in *Mus Xpr1* (6). Although there is otherwise limited protein sequence identity between the ECL4 domains of birds and mammals (38% identity with mouse), two of the 3 ECL4 residues conserved in mice are also conserved in bird XPR1 (S574 and T576), suggesting they have been subjected to purifying selection.

Of the two receptor-inactivating mutations identified in the chicken, one, the ECL4 substitution Q579E, was found in one other Galliformes species, the Chinese bamboo partridge, *Bambusicola thoracica* (Fig. 5; see Text S1 in the supplemental material). Of the 23 additional species from 17 different avian families examined, one raptor, the Philippine hawk-owl, *Ninox philippensis*, but was not found in 8 other raptors tested, which included various owls, hawks, and a falcon. Although the same mutation, K496E, was also identified in the South American ratite, the rhea, these results show 93 to 99% identity across all avian species. The avian XPR1 ECL4 is much less variable than the comparable segment of the mammalian gene (2). Rare replacement mutations alter only 3 ECL4 codons in bird species (Fig. 4), whereas for mammalian XPR1 genes, there are substitutions affecting all but 3 ECL4 residues. The 3 conserved mammalian ECL4 sites do not contribute to the virus attachment site in *Mus Xpr1* (6). Although there is otherwise limited protein sequence identity between the ECL4 domains of birds and mammals (38% identity with mouse), two of the 3 ECL4 residues conserved in mice are also conserved in bird XPR1 (S574 and T576), suggesting they have been subjected to purifying selection.

The substitution with the greater impact on receptor function, K496E (Fig. 2B), was identified only in chicken and one other Gallus species, the green junglefowl, *Gallus varius* (Fig. 5). The 2 mutations that disable the chicken gene thus either predate domestication or were introduced into the wild through cross-breeding with chickens, a distinct possibility, as chickens and green junglefowl are interfertile (32) and are routinely interbred by the Javanese, producing a hybrid termed the Bekisar.

The sequence analysis also identified a novel substitution altering one of these two critical sites, K496E, in two Gallus species, the gray and Sri Lankan junglefowls, *Gallus sonnerati* and *Gallus lafayetti* (Fig. 5). In *Mus*, the homologous ECL4 substitution (K500E) helps disable the laboratory mouse XPR1 (8). To determine the impact of this mutation on avian receptor function, we introduced K496E into the duck clone; it results in significant impairment of receptor function (Fig. 2B). For the other Galliformes species examined, the XPR1 gene of the black francolin (*Francolinus francolinus*), turkey (*Meleagris gallopavo*), and quail (*Coturnix japonica*), resemble the permissive duck gene at both critical sites (Fig. 5).

With two exceptions, 23 additional species from 17 different avian families were found to carry the permissive residues at positions 496 and 579 (Fig. 5; see Text S1 in the supplemental material). One of the disabling mutations, K496E, was also identified in one raptor, the Philippine hawk-owl, *Ninox philippensis*, but was not found in 8 other raptors tested, which included various owls, hawks, and a falcon. Although the same mutation, K496E, was also identified in the South American ratite, the rhea, these results show 93 to 99% identity across all avian species. The avian XPR1 ECL4 is much less variable than the comparable segment of the mammalian gene (2). Rare replacement mutations alter only 3 ECL4 codons in bird species (Fig. 4), whereas for mammalian XPR1 genes, there are substitutions affecting all but 3 ECL4 residues. The 3 conserved mammalian ECL4 sites do not contribute to the virus attachment site in *Mus Xpr1* (6). Although there is otherwise limited protein sequence identity between the ECL4 domains of birds and mammals (38% identity with mouse), two of the 3 ECL4 residues conserved in mice are also conserved in bird XPR1 (S574 and T576), suggesting they have been subjected to purifying selection.

The substitution with the greater impact on receptor function, K496E (Fig. 2B), was identified only in chicken and one other Gallus species, the green junglefowl, *Gallus varius* (Fig. 5). The 2 mutations that disable the chicken gene thus either predate domestication or were introduced into the wild through cross-breeding with chickens, a distinct possibility, as chickens and green junglefowl are interfertile (32) and are routinely interbred by the Javanese, producing a hybrid termed the Bekisar.

The sequence analysis also identified a novel substitution altering one of these two critical sites, K496E, in two Gallus species, the gray and Sri Lankan junglefowls, *Gallus sonnerati* and *Gallus lafayetti* (Fig. 5). In *Mus*, the homologous ECL4 substitution (K500E) helps disable the laboratory mouse XPR1 (8). To determine the impact of this mutation on avian receptor function, we introduced K496E into the duck clone; it results in significant impairment of receptor function (Fig. 2B). For the other Galliformes species examined, the XPR1 gene of the black francolin (*Francolinus francolinus*), turkey (*Meleagris gallopavo*), and quail (*Coturnix japonica*), resemble the permissive duck gene at both critical sites (Fig. 5).

With two exceptions, 23 additional species from 17 different avian families were found to carry the permissive residues at positions 496 and 579 (Fig. 5; see Text S1 in the supplemental material). One of the disabling mutations, K496E, was also identified in one raptor, the Philippine hawk-owl, *Ninox philippensis*, but was not found in 8 other raptors tested, which included various owls, hawks, and a falcon. Although the same mutation, K496E, was also identified in the South American ratite, the rhea, these results show 93 to 99% identity across all avian species. The avian XPR1 ECL4 is much less variable than the comparable segment of the mammalian gene (2). Rare replacement mutations alter only 3 ECL4 codons in bird species (Fig. 4), whereas for mammalian XPR1 genes, there are substitutions affecting all but 3 ECL4 residues. The 3 conserved mammalian ECL4 sites do not contribute to the virus attachment site in *Mus Xpr1* (6). Although there is otherwise limited protein sequence identity between the ECL4 domains of birds and mammals (38% identity with mouse), two of the 3 ECL4 residues conserved in mice are also conserved in bird XPR1 (S574 and T576), suggesting they have been subjected to purifying selection.

The substitution with the greater impact on receptor function, K496E (Fig. 2B), was identified only in chicken and one other Gallus species, the green junglefowl, *Gallus varius* (Fig. 5). The 2 mutations that disable the chicken gene thus either predate domestication or were introduced into the wild through cross-breeding with chickens, a distinct possibility, as chickens and green junglefowl are interfertile (32) and are routinely interbred by the Javanese, producing a hybrid termed the Bekisar.
indicate that the species that carry the mutations that disable receptor function are largely native to a defined geographic area that encompasses India and Southeast Asia (Fig. 6).

**Positive selection of avian XPR1.** We used the PAML4 suite of programs (29) to test for evidence of adaptive evolution and to identify possible sites of positive selection marked by an excess of nonsynonymous mutations in the sequenced segments of avian XPR1. Positive selection is a hallmark of genes involved in genetic conflicts and often identifies genes that serve a defensive function. Evidence of involvement in an evolutionary "arms race" has been found for other genes capable of restricting MLVs, like primate TRIM5α (33) and the mouse genes Fv1 (34) and Apobec3 (35). There is also evidence that rodent Xpr1 has been under positive selection at sites coding for ECL3 and ECL4 that have been implicated in receptor function (6). Rodent Xpr1 variation is coincident with exposure to MLVs; 4 of the 5 polymorphic variants of Mus Xpr1 restrict different subsets of XP-MLVs, and all 4 variants evolved in Southeast Asia, where mice were first exposed to XP-MLV infection and acquired MLV endogenous retroviruses (ERVs) (7).

Avian sequences of receptor-determining regions of exons 11 and 13 for 31 species and exons 10 to 13 for 23 species were used in this analysis. The phylogeny used was based on the accepted phylogenetic tree for Aves (25–28) (Fig. 5). Likelihood ratio tests indicate that XPR1 has experienced positive selection, and this was the case for both codon frequency models tested (Fig. 7A; see Table S2 in the supplemental material). The Bayes empirical Bayes calculation of posterior probabilities in PAML4 identified one XPR1 codon position, 496, as being under positive selection in both data sets (>95% posterior probability) (Fig. 7B; see Table S2 and Text S1 in the supplemental material). A second codon in the variable region at the N terminus of ECL3, 423, was also identified as being under positive selection in the 4-exon data set (posterior probability of >0.99). Thus, positive selection affects one codon critical for XPR1 receptor function (496) and a second codon in a region of ECL3 implicated in the modulation of receptor function (Fig. 2B).

Because one of the disabling mutations was found in the rhea, a species with no known exposure to infectious XP-MLVs, we reran the analysis without this species (see Table S3 in the supplemental material), and we also ran the analysis without the passerines and water birds (not shown). Significance levels of selection were reduced, but both sites remained under positive selection in at least one codon frequency model tested.

The analysis indicates that the evolution of this avian gene has been driven by genetic conflicts centered on codons implicated in virus restriction, and this is consistent with an antiviral function for the gene in birds.

**DISCUSSION**

There are many documented examples of the transspecies transmission of retroviruses (36). Such naturally occurring transmissions can produce disease in the new hosts (HIV-1), and in some cases the introduced viruses can be endogenized (koala endogenous retrovirus [KoRV]) (37). The fact that birds have acquired retroviruses from mammalian sources has been established by the observed phylogenetic discontinuities between birds and the endogenous viruses they carry (16, 17). The results presented here suggest that interorder transmission of X-MLVs from mice to birds may have driven fixation of resistant alleles in the avian ortholog of the X-MLV XPR1 receptor, a phenomenon also doc-
although recent studies suggest additional small contributions to the chicken genome from other junglefowl species (43). The domesticated chicken was subsequently introduced to the West through trade between the city states of the Indus Valley and the Middle East and may have been introduced to the New World by Polynesians (38). The Indian subcontinent is also where the 3 lineages of house mice carrying the various XP-MLVs originated before dispersing to their current Eurasian ranges and beyond (44, 45) (Fig. 6). These are *M. m. domesticus* (western Europe and the Mediterranean), which carries endogenous P-MLVs, and *M. m. castaneus* (Southeast Asia) and *M. m. musculus* (central Europe to China), both of which carry mainly X-MLVs (13). These commensal or house mice differ from their *Mus* progenitors by their dependence on humans; the animals live in close contact with humans in manmade structures, while their aboriginal progenitors inhabit various niches in the wild (46). Thus, red junglefowl and house mice both evolved in the same geographic area, and omnivorous chickens were domesticated in human habitats infested with X-MLV-infected *M. m. castaneus* house mice.

Four of 8 additional fowl and 1 of 9 raptors were found to carry 1 of the 3 XPR1 receptor-disabling mutations (K496Q, K496E, or Q579E). All 5 of these species are native to south central or Southeast Asia and thus, like *G. gallus*, have ranges that overlap the range of *M. m. castaneus* (Fig. 6) (45, 47). None of the 3 XPR1 mutations was found in 11 other fowl and raptor species native to the Middle East, Australia, or the Americas, areas populated largely by *M. m. domesticus*, or in the case of quail, by *M. m. musculus* and *M. m. molossinus*, which is a natural hybrid of *M. m. musculus* and *M. m. castaneus* (48). Our analysis included 14 species that are neither fowl nor raptors, such as water birds, passerines, and rodents, representing 100 million years of evolution. Only 1 of these 14 species, the South American rhea, carries mutations at either of the two key sites. The habitat of this bird is thus outside the *M. m. castaneus* range of Southeast Asia, and although *M. m. castaneus* mice have been found in the Americas, in California (13, 49), the few characterized South American house mice, all from Peru, have been typed as *M. m. domestica* (50). The presence of this mutation in the rhea could thus result from neutral processes or from an unknown nonneutral pressure. Removing the rhea sequence from the PAML4 analysis for positive selection reduces but does not eliminate evidence for positive selection.

Among wild mouse populations, X-MLVs are found in *M. m. molossinus and M. m. musculus*, as well as *M. m. castaneus*. These mice all carry endogenous X-MLVs, and infectious virus has been isolated from all 3 subspecies (51–53). This raises the question of why receptor mutations were not found in birds from other parts of Eurasia where these other X-MLV-infected mouse subspecies are found. There are several possible explanations. First, our collection of 32 species did not include raptors likely to have intimate contact with these other Eurasian mouse subspecies. Second, as shown for inbred laboratory strains, mice can vary significantly in their ability to produce infectious X-MLV, with some strains producing high levels of X-MLV throughout life (54, 55). We have no information, however, on the level of endogenous virus production in the 3 X-MLV-positive wild mouse populations, but we do know that the Xpr1 receptor allele of *M. m. castaneus* (*Xpr1*) is a more efficient receptor for X-MLVs than the Xpr1 variant shared by *M. m. musculus* and *M. m. molossinus* (*Xpr1") (6), and this difference could certainly influence endogenous virus levels.

There is no evidence for endogenous retroviruses closely re-

<table>
<thead>
<tr>
<th>Exons 11 and 13</th>
<th>Tree length</th>
<th>dN/dS</th>
<th>F3x4</th>
<th>p = 0.01</th>
<th>1.67283</th>
<th>3.84</th>
</tr>
</thead>
<tbody>
<tr>
<td>F61</td>
<td>p = 0.02</td>
<td>1.71391</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exons 10-13</th>
<th>Tree length</th>
<th>dN/dS</th>
<th>F3x4</th>
<th>p = 0.0001</th>
<th>1.24815</th>
<th>3.37</th>
</tr>
</thead>
<tbody>
<tr>
<td>F61</td>
<td>p = 0.0001</td>
<td>1.25117</td>
<td>3.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG 7** Avian *XPR1* sites under positive selection. (A) Likelihood ratio tests (LRTs) were used to test for positive selection. A neutral model (M7) was compared with a selection model (M8) using two different models of codon frequency (F3x4 and F61). *P* values of <0.05 provide evidence of positive selection, and *P* values of <0.01 indicate strong positive selection. The tree length is the average number of substitutions per codon along all branches. *dN/dS* ratios are the mean values estimated by PAML4 for the category *dN/dS* ratios of >1. (B) (Top) Diagram of a portion of the XPR1 gene showing the locations of ECL3 and ECL4 and the locations of the key codons involved in receptor function. The four alternating green and purple horizontal lines indicate the positions of the sequenced segments of exons 10 to 13. (Bottom) Graph showing the posterior probability of positive selection at each codon based on an analysis of 22 sequences using codon frequency model F3x4 and selection model 8. The horizontal red line marks a posterior probability greater than 0.95, and the 3 arrows identify the 2 codons over this threshold (and therefore under selection) and critical residue 579.

**TABLE 1** Summary of *dN/dS* values for *XPR1* differ ec genetic positions under positive selection.

<table>
<thead>
<tr>
<th>Exons</th>
<th>Tree length</th>
<th>dN/dS</th>
<th>F3x4</th>
<th>p = 0.01</th>
<th>1.67283</th>
<th>3.84</th>
</tr>
</thead>
<tbody>
<tr>
<td>F61</td>
<td>p = 0.02</td>
<td>1.71391</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exons</th>
<th>Tree length</th>
<th>dN/dS</th>
<th>F3x4</th>
<th>p = 0.0001</th>
<th>1.24815</th>
<th>3.37</th>
</tr>
</thead>
<tbody>
<tr>
<td>F61</td>
<td>p = 0.0001</td>
<td>1.25117</td>
<td>3.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG 7** Avian *XPR1* sites under positive selection. (A) Likelihood ratio tests (LRTs) were used to test for positive selection. A neutral model (M7) was compared with a selection model (M8) using two different models of codon frequency (F3x4 and F61). *P* values of <0.05 provide evidence of positive selection, and *P* values of <0.01 indicate strong positive selection. The tree length is the average number of substitutions per codon along all branches. *dN/dS* ratios are the mean values estimated by PAML4 for the category *dN/dS* ratios of >1. (B) (Top) Diagram of a portion of the XPR1 gene showing the locations of ECL3 and ECL4 and the locations of the key codons involved in receptor function. The four alternating green and purple horizontal lines indicate the positions of the sequenced segments of exons 10 to 13. (Bottom) Graph showing the posterior probability of positive selection at each codon based on an analysis of 22 sequences using codon frequency model F3x4 and selection model 8. The horizontal red line marks a posterior probability greater than 0.95, and the 3 arrows identify the 2 codons over this threshold (and therefore under selection) and critical residue 579.

**TABLE 1** Summary of *dN/dS* values for *XPR1* different genetic positions under positive selection.

<table>
<thead>
<tr>
<th>Exons</th>
<th>Tree length</th>
<th>dN/dS</th>
<th>F3x4</th>
<th>p = 0.01</th>
<th>1.67283</th>
<th>3.84</th>
</tr>
</thead>
<tbody>
<tr>
<td>F61</td>
<td>p = 0.02</td>
<td>1.71391</td>
<td>2.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exons</th>
<th>Tree length</th>
<th>dN/dS</th>
<th>F3x4</th>
<th>p = 0.0001</th>
<th>1.24815</th>
<th>3.37</th>
</tr>
</thead>
<tbody>
<tr>
<td>F61</td>
<td>p = 0.0001</td>
<td>1.25117</td>
<td>3.02</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
sembling MLVs in birds. Although one of the studies demonstrating phylogenetic discontinuities between bird species and their endogenous retroviruses identified those avian ERVs using degenerate primers designed from MLVs, the amplified ERVs cluster with mammalian retroviruses but are not MLVs (17). The sequenced \textit{G. gallus} genome lacks recognizable MLV-like ERVs (56). While the presence of MLVs in birds would lend support to a long-term evolutionary arms race, the absence of such sequences is not surprising. After all, endogenization is a rare consequence of virus infection; pathogenic retroviruses are not always endogenous in their susceptible hosts, and some MLVs isolated from wild mouse populations are not endogenous, like hortulanus murine leukemia virus (HoMLV) from eastern European mice (57) and amphotropic MLV in California wild mice (58).

Alternative explanations for the observed mutational changes that alter the receptor function of the avian XPR1s, for their positive selection, and for their species and geographic distribution include the possibility that other pathogens have converged on XPR1. This is reminiscent of the fact that a mutation in the HIV-1 coreceptor CCR5 is prevalent in humans in Eurasia but originated well before human exposure to HIV-1 (59, 60). The emergence of this CCR5 mutation and its geographic distribution are consistent with a strong selective event, like an epidemic, and although the responsible pathogen has not been identified, CCR5 is the receptor for a toxin produced by \textit{Staphylococcus aureus} (61). At this point, however, no infectious agents other than MLVs are known to interact with XPR1. A second alternative explanation for the species distribution of these mutations is that the XPR1 mutations may have arisen in ancestral Galliformes or are shared due to possible interbreeding between \textit{Gallus} species. However, examination of the pattern of substitutions in XPR1 exon and intron sequences does not suggest any recent recombination events, and analysis using the GARD program (reference 62 and data not shown) finds only weak evidence of recombination at a single site following codon 473 in ECL3 (P = 0.05) that comes from outside the galliform clade and therefore does not account for the presence of the two receptor-disabling mutations in chickens.

Domestication has a major impact on emerging diseases, and the antiviral XPR1 genotypes with the double ECL3/ECL4 mutations are found exclusively in domestic chickens and domesticated (laboratory) mice. For the mouse, domestication was begun by Asian hobbyists who produced the fancy mouse progenitors of inbred laboratory strains (63). These hobbyists inadvertently interbred subspecies with active XP-MLVs with subspecies carrying the permissive XPR1 receptor. Domestication also creates high-risk environments with increased opportunity for jumps between and within species, and gammaretroviruses can be transmitted horizontally, as has been shown for wild and laboratory mice (64, 65). The fact that the highly restrictive receptor of domesticated chickens is not also found in domesticated turkeys and ducks is not surprising, as turkeys evolved and were domesticated in North America and therefore were not exposed to X-MLV-infected mice. The virus-susceptible duck is a waterfowl with a wide distribution that was likely domesticated in China about 2,000 years ago from mallard progenitors indigenous to an area where \textit{M. m. musculus} likely predominates (44, 45, 66).

Virus-resistant XPR1 receptors are produced by a small number of rare mutations at the same two sites in birds (K496Q/E and Q576E) and in laboratory mice (K500E and T582Δ). One of these two sites is in an ECL (ECL4) that shows little homology between mice and birds (38%) and is also highly variable among mammals; there are numerous replacement mutations at and around the critical ECL4 site in mammals, most of which are compatible with receptor function (7) (Fig. 4). In contrast to mammalian XPR1, the avian XPR1 ECL4 is highly constrained, and positive selection was detected in the ECL4 of rodents, but not birds (6). On the other hand, for ECL3, the critical site for receptor function, K496/500, is under positive selection in birds, but not rodents. This codon spans the exon 11 consensus splice acceptor site, which imposes a bias on sequence variation for the second and third codon positions, although this bias does not disallow any of the 4 bases at the two sites (67). The restricted species and geographical distributions of these separate receptor-disabling mutations, the fact that species with restrictive receptors have been exposed to X-MLVs, the links to diet and domestication, and the observed positive selection all argue that the mutations did not occur coincidentally or through some alternative, undefined selection pressure but are the result of genetic conflicts characteristic of antiviral activity.

The observation that the laboratory mouse and various bird species in different lineages have evolved receptor-mediated resistance to X-MLVs raises the possibility that X-MLVs may be pathogenic in these species. Naturally occurring infectious and endogenous MLVs of laboratory mice and wild mice are linked to a variety of neoplastic, immunological, and neurological diseases (64, 68). In the XP-MLV family, recombinant P-MLVs derived from naturally occurring or laboratory MLV strains are pathogenic in mice (69), but there is limited work on the pathogenic potential of X-MLVs in mice because the belief that all mice are resistant to these viruses was only recently dispelled (70, 71). While it is not yet known if exposure of birds to X-MLVs results in infection and if these viruses are pathogenic in birds, our observations that duck cells support replication of infectious virus indicates that there are no significant postentry host blocks to replication. Further assessment of the possible adaptive advantage of the mutations described here should contribute to our knowledge of the host-pathogen interactions affecting transspecies transmissions in natural populations and in laboratory and agricultural stocks and should help define the evolutionary pathways that thwart those transmissions.

**ACKNOWLEDGMENTS**

This work was supported by the Intramural Research Program of the NIAID, NIH.

We thank Qingping Liu and Ronald Plishka for technical assistance and Devinka Banmussenighe for help preparing figures. We also thank Rebecca Kimball (University of Florida, Gainesville, FL), James Dean (National Museum of Natural History, Washington, DC), and Glenn Proudfoot (Vassar College, Poughkeepsie, NY) for avian DNAs and tissue samples.

**REFERENCES**


