Immunogenetics (2004) 56: 170–177
DOI 10.1007/s00251-004-0675-0

ORIGINAL PAPER

David J. Lynn · Rowan Higgs · Susan Gaines · Joanna Tierney · Tharappel James · Andrew T. Lloyd · Mario A. Fares · Grace Mulcahy · Cliona O'Farrelly

Bioinformatic discovery and initial characterisation of nine novel antimicrobial peptide genes in the chicken

Received: 30 January 2004 / Revised: 22 March 2004 / Accepted: 22 March 2004 / Published online: 18 May 2004

Abstract Antimicrobial peptides (AMPs) are essential components of innate immunity in a range of species from*Drosophila* to humans and are generally thought to act by disrupting the membrane integrity of microbes. In order to discover novel AMPs in the chicken, we have implemented a bioinformatic approach that involves the clustering of more than 420,000 chicken expressed sequence tags (ESTs). Similarity searching of proteins—predicted to be encoded by these EST clusters—for homology to known AMPs has resulted in the in silico identification of full-length sequences for seven novel gallinacins (Gal-4 to Gal-10), a novel cathelicidin and a novel liver-expressed antimicrobial peptide 2 (LEAP-2) in the chicken. Differential gene expression of these novel genes has been demonstrated across a panel of chicken tissues. An evolutionary analysis of the gallinacin family has detected sites—primarily in the mature AMP—that are under positive selection in these molecules. The functional implications of these results are discussed.

Keywords Chicken · Defensins · Antimicrobial peptide · Innate immune system

D. J. Lynn · R. Higgs · S. Gaines · T. James · A. T. Lloyd · C. O'Farrelly (✉)
Education and Research Centre, St. Vincent's University Hospital,
4 Dublin, Ireland
e-mail: cliona.ofarrelly@ucd.ie
Tel.: +353-1-2094940
Fax: +353-1-2838123

D. J. Lynn · R. Higgs
Department of Medicine, University College Dublin,
4 Belfield, Dublin, Ireland

T. James
Moyne Institute of Preventive Medicine, Trinity College Dublin,
2 Dublin, Ireland

A. T. Lloyd
Department of Genetics, Trinity College Dublin,
2 Dublin, Ireland

J. Tierney · G. Mulcahy
Department of Microbiology and Parasitology, Faculty of Veterinary Medicine, University College Dublin,
4 Belfield, Dublin, Ireland

M. A. Fares
Biology Department, National University of Ireland,
Maynooth, Ireland

Conway Institute, University College Dublin,
4 Belfield, Dublin, Ireland

Introduction

Antimicrobial peptides (AMPs), essential components of innate host defence in species as diverse as plants, flies and mammals are generally thought to act by disrupting the membrane integrity of microbes (Kagan et al. 1990; Satchell et al. 2003). In an age when antibiotic resistance is an increasing problem, these peptides are of interest as potential novel pharmaceutical agents. In vertebrates, there are two major families of AMPs: defensins and cathelicidins. In mammals, α-defensins and β-defensins are two structurally distinct cationic, cysteine-rich AMPs, which differ in size and in the spacing of a six-cysteine structural motif (Liu et al. 1997). α-defensins are unique to mammals, but β-defensins are much more widely distributed and, in the chicken, four β-defensins, known as gallinacins, have been described to date (Evans et al. 1994; Harwig et al. 1994; Zhao et al. 2001). Homologous peptides have also been described in the turkey (Evans et al. 1994; Zhao et al. 2001). Gal-1, Gal-1α and Gal-2 were isolated from chicken heterophils, while Gal-3 was shown to be constitutively expressed in the epithelia of a range of tissues and to be inducible in the trachea following infection (Zhao et al. 2001). Gallinacins exhibit variable activity against a number of Gram-positive and Gram-negative bacteria, and Gal-1 and Gal-1α are active against the yeast *Candida albicans* (Evans et al. 1995; Harmon 1998).

Cathelicidins are a family of highly diverse AMPs but are all encoded by prepropeptides containing highly

conserved cathelin domains. To date, cathelicidins have only been described in mammals, including humans, monkeys, horses, cattle, sheep, goats, pigs, rabbits, mice and guinea pig (for review, see Ramanathan et al. 2002). Each species has variable numbers of cathelicidin genes; artiodactyls in particular have high copy numbers (Scocchi et al. 1997), whereas humans and mice have only one gene copy each (Agerberth et al. 1995; Cowland et al. 1995; Gallo et al. 1997; Larrick et al. 1995). Neutrophils are a particularly rich source of cathelicidins in a variety of species. In humans, cathelicidins have also been found to be expressed in several other tissues, including the testis (Agerberth et al. 1995; Malm et al. 2000), squamous epithelia (Frohm Nilsson et al. 1999; Nizet et al. 2001), airway epithelia (Bals et al. 1998), sweat glands (Murakami et al. 2002b), salivary glands (Murakami et al. 2002a) and colon (Hase et al. 2002). Cathelicidins have a wide spectrum of antimicrobial activity and have been shown to be active against Gram-negative and Gram-positive bacteria (Travis et al. 2000), fungi (Shin et al. 2000) and enveloped viruses (Tamamura et al. 1995). Activity against a particular microbe depends on the type of mature peptide and the species from which it originates.

In this study, we describe a bioinformatics approach to the identification of novel AMPs in the chicken. This method involves homology searching of clustered chicken expressed sequences tags (ESTs) by BLAST (Altschul et al. 1997) and by the more sensitive hidden Markov model (HMM) profile searching (Eddy 1998). Our approach has identified eight novel antimicrobial peptides, seven gallinacins and one cathelicidin in the chicken. We show that all these peptides as well as liver-expressed antimicrobial peptide 2 (LEAP-2)—an AMP we have previously identified in chicken (Lynn et al. 2003)—are expressed at the mRNA level in a panel of chicken tissues. Furthermore, an evolutionary analysis of the gallinacin family has provided evidence that certain amino acid sites in the active peptide are subject to positive selection.

Materials and methods

From the dbEST Web site (http://www.ncbi.nlm.nih.gov/dbEST/), 422,426 chicken ESTs were downloaded. Prior to the clustering step, repeat sequences in the ESTs were masked using RepeatMasker (Smit and Green, unpublished), and contaminant sequences were removed using SeqClean (http://www.tigr.org/tdb/tgi/software/) to minimize false clustering due to spurious EST similarity.

The ESTs were then clustered using The Institute for Genomic Research Gene Indices clustering tools (TGICL) (Pertea et al. 2003) (available from http://www.tigr.org/tdb/tgi/software/). ESTs were clustered if they shared more than 30 bp of at least 95% identity. The clustered contig sequences are available at http://www.binf.org/immunogenetics/contigs.fa.

The sequences for each cluster were post-processed with ESTScan (Iseli et al. 1999). ESTScan detects and reconstructs potential coding regions in ESTs, using a novel HMM method that can automatically correct for frame-shift errors. From the coding regions, the predicted protein can be determined (available at http://www.binf.org/immunogenetics/estscanpredictions.fa). A database of known AMPs, as identified by an SRS search (http://srs.ebi.ac.uk/) of the Swissprot-Trembl protein database was constructed and is available at http://ercbinfo1.ucd.ie/APPDb/. All AMPs from this database were searched against the clustered EST contigs and the database of proteins predicted by ESTScan from the EST clusters using the BLAST programs, with an E-value cut-off of 0.001 (Altschul et al. 1997). Molecular weight and net charge for the complete chicken proteins were predicted using the ProtParam tool (http://www.expasy.org/tools/protparam.html).

To search for novel antimicrobial proteins that could have been missed by BLAST, we constructed HMMs for a number of AMP families. All the sequences annotated as either α-defensins, cathelicidins or hepcidins in the Pfam database (Bateman et al. 2002)—a manually curated and annotated collection of protein families (http://www.sanger.ac.uk/Software/Pfam)—were extracted. The gallinacin family is not represented in the Pfam database, so all known gallinacin sequences were extracted from the National Center for Biotechnology Information GenPept protein database. These sequences included Gal-1 (P46156), Gal-1α (Q9DG59), Gal-2 (P46158), Gal-3 (Q9DG58), turkey heterophil peptide-1 (THP1) (P80391), THP2 (P80392) and turkey β-defensin (TBD) (Q9DG57). To create an HMM profile for each family, all constituent sequences of AMPs were aligned using the T-Coffee program (Notredame et al. 2000), and these alignments were used as input for the HMMER, version 2.1.1, suite of programs (Eddy 1998) (http://hmmer.wustl.edu/). The HMM for each family was constructed using the

Table 1 PCR primer sequences and predicted product lengths

Target mRNA[a]	5′ Primer	3′ Primer	Product size (bp)
Gal-1	5′-GAAATGCTCAAGATTTCACCTCTG-3′	5′-CCTTTATTCAGCAGAGAAAAGCAG-3′	231
Gal-2	5′-GCATAAACACTTCATGAGTCCATC-3′	5′-GAAGAAAGGCAGTGCAGAAGATA-3′	166
Gal-3	5′-CCTTCTTCCTCTTGTTTCTCCAG-3′	5′-ATCAACCTCATATGCTCTTCCAC-3′	158
Gal-4	5′-GATCCTTTACCTGCTGCTGTCT-3′	5′-TCCTCACACAGCAAGATTTTAGTC-3′	185
Gal-5	5′-GATCCTTTACCTGCTGCTGTCT-3′	5′-AGCAAGAGCCTATTCCATTGTTAC-3′	176
Gal-6	5′-ATGAGAATCCTTTTCTTCCTTGTTGC-3′	5′-TTAGGAGCTAGGTGCCCATTTGCAGC-3′	201
Gal-7	5′-ATCGTGCTCCTCTTTGTGGCAGTTCA-3′	5′-CTACAACCATCTACAGCAAGAATACT-3′	171
Gal-8	5′-CTGTTCTCCTCTTCCTCTTCCAG-3′	5′-AATCTTGGCACAGCAGTTTAACA-3′	170
Gal-9	5′-ATGCAGATCCTGCCTCTCCTCTTTGCT-3′	5′-TCAGGAATACCATCGGCTCCGGCAGCAGAA-3′	201
Gal-10	5′-ATGAGGAACCTTTGTTTCGTGT-3′	5′-TCAGGTCTTGGTGGGAGTTGGTG-3′	198
LEAP-2	5′-CACCATGCACTGTTTGAAAATTATGGCA-3′	5′-TCACTCGGAGGCCGTTCTAAGGAA-3′	235
Cathelicidin	5′-CACCATGCTGAGCTGCTGGGTGCTGCTG-3′	5′-TCACTTCTTCTTGATCGCCCGGTA-3′	451
β-actin	5′-GCGCTCGTTGTTGACA-3′	5′-TCATCCCAGTTGGTGACA-3′	206

[a]*Gal*-Gallinacin, *LEAP-2* liver-expressed antimicrobial peptide 2

hmmbuild program, and hmmcalibrate was used to calibrate E-value scores. The HMM profiles were then used to search against the database of proteins predicted by ESTScan from the EST clusters using the hmmsearch program.

Evolutionary analysis of the gallinacins

A multiple-sequence alignment of the gallinacin family of AMPs (including the novel sequences and the homologous turkey sequences) was constructed using the T-Coffee program (Notredame et al. 2000). A neighbor-joining phylogenetic tree was inferred from the protein alignment using MEGA, version 2.1, with the Poisson corrected model implemented (Kumar et al. 2001). One thousand bootstrap replicates were carried out to test the significance of each node in the tree. To construct an alignment of the coding sequences, the protein alignment was used as a template and a 'copygaps' Perl script was used to align the DNA, maintaining the gaps that were present in the protein alignment. Any columns in the DNA alignment that had more than three gap characters were removed. The topology of the neighbor-joining tree and the DNA alignment were used as input to the CODEML and CODEMLSITES programs from the PAML package, version 3.12 (Yang 1997), to test for evidence of positive selection during the evolution of the gallinacins.

The principle involved in such tests is to compare the rates of synonymous (d_S) and non-synonymous (amino acid changing: d_N) changes among the DNA sequences. If amino acid changes are selectively neutral (i.e. mutations that are neither advantageous or deleterious), they will be fixed at the same rate as synonymous mutations and ω ratio (d_N/d_S)=1. ω values >1 are taken to indicate that amino acid changes are accumulating at a faster rate than is acceptable under a neutral mutation model. That is to say, the rate of amino acid changes (d_N) significantly exceeds the rate of synonymous changes (d_S) at the DNA level. The CODEML program tests for variable selective pressures among lineages in the phylogeny by looking for significant differences in ω ratios. To test for variable selective pressures among phylogenetic lineages, the one-ratio model, which assumes an equal ω ratio for all branches in the phylogeny, was compared to the free-ratios model, which allows an independent ω ratio for each branch (Yang 1998; Yang and Nielsen 1998). The result of this program is a log-likelihood value for each model. To test which is the favoured model, the log-likelihood values for each model are compared by a likelihood ratio test (LRT). Twice the log-likelihood difference between the two models is compared to a χ^2 distribution with $n-1$ *df*, where n is the number of branches of the phylogeny. If a significant *P*-value is obtained, it can be concluded that the free-ratios model is the favoured model, and branches on the phylogeny with ω values >1 are subject to positive selection.

Positive selection in amino acid sites

Another way of looking for positive selection is to look for significant variability in ω ratios among amino acid sites in the multiple-sequence alignment (Nielsen and Yang 1998). The CODEMLSITES program determines whether any of six progressively more complex models of evolution are significantly better at explaining the observed variation in the dataset (Yang et al. 2000).

The first test compares the models M0 and M3. Model M0 is an evolutionary model whereby all the amino acid sites have a single ω value. This model is compared to M3, which classifies the amino acid sites into one of three classes, with the proportion of sites belonging to a particular class and the ω values for each class of site estimated by CODEMLSITES from the data. M3 is a test of amino acid sites subject to positive selection, as it allows for the presence of sites with ω>1.

The second test compares the models M1 and M2. M1 is a model of neutral evolution where amino acid sites can be conserved (ω=0) or neutrally evolving (ω=1). Model M2 is a test of selection, as it allows for the presence of sites where ω is a free parameter and as such can have a value >1.

The final test, which compares the models M7 and M8, is the most stringent test. M7 allows for sites with ω values that follow a β distribution of values between ω= 0 and ω=1. Model M8 is the same as M7 but allows for the presence of sites with ω>1, and comparing these two models is a test of selection.

As with the test of positive selection among lineages, CODEMLSITES estimates a log-likelihood value for each model. To test which are the favoured models, the log-likelihood values for M0 versus M3, M1 versus M2 and M7 versus M8 were compared by LRTs. Posterior Bayesian probabilities were calculated to determine which amino sites belong to which site classes (Nielsen and Yang 1998). If significant variability is revealed, then those sites, which have ω>1 and high posterior probabilities, are likely to be under positive, diversifying selection.

Expression of AMPs in chicken tissue

One-day-old male chickens (Cobb 500 broiler) were purchased from the Knocknagarm Hatchery, then housed in a floor pen in the Biomedical Facility, University College Dublin, Belfield, Dublin, Ireland. Environmental temperature was kept at a constant 25°C. Animals were fed commercial coccidiostat-free starter/grower ration and water ad lib. One bird was sacrificed at 3 weeks of age by intravenous pentobarbitone sodium inoculation. The tissues were quickly dissected, squeezed between Whatman filter paper to remove excess blood, rinsed in saline and snap frozen in liquid nitrogen. Tissues were stored at −80°C until processed further. Following pulverisation of the tissues using a Mikro-Dismembrator U (B. Braun Biotech International), total cellular RNA was purified using the RNeasy Kit (Qiagen, West Sussex, UK) according to the manufacturer's recommendations. Spectrophometric analysis was performed in order to assess the quantity and quality of total RNA. Single-stranded cDNA was synthesised from 1 μg RNA using oligo-dT primer (Promega, Madison, Wis.) and Omniscript (Qiagen). The AMP-specific cDNAs were amplified by PCR using *Taq* polymerase (Qiagen) and primers designed internally from the coding sequence of Gal-1 to Gal-10, LEAP-2, cathelicidin and β-actin. Thirty cycles (94°C for 30 s, 55°C for 30 s and 72°C for 30 s) were used for amplification. PCR products were separated by electrophoresis on ethidium bromide-stained 2% agarose gels and visualised using Eagle Eye (Stratagene, La Jolla, Calif.). A list of PCR primer sequences and product lengths are shown in Table 1.

cDNA cloning

Positive tissues from the above expression panel were chosen as source material for each specific gene. The cDNA was amplified as described above but using *Pfu* DNA polymerase (Promega) and gene-specific primers with CACC overhangs upstream of the start codon, thus providing the complementary sequence necessary for directional cloning. The amplified cDNA was purified (GenElute PCR Clean-up Kit, Sigma, St. Louis, Mo.), and ligated into the pcDNA 3.1 cloning vector (Invitrogen, Groningen, The Netherlands). Cloned plasmids were sequenced (Advanced Biotechnology Centre, London, UK), using vector-specific primers and compared to the EST consensus sequence.

Results

In the absence of complete genome sequence, ESTs are a rich source of novel sequence information. By definition, ESTs are short, error-prone sequences. Clustering of ESTs that are likely to be encoded by the same mRNA reduces the redundancy in the EST database, improves the

Table 2 GenBank accession numbers and properties of novel AMPs

Name	Accession number	Length (aa)	M_r	Net charge
Gal-4	AY534892	67	7.5	+8
Gal-5	AY534893	67	7.6	+7
Gal-6	AY534894	67	7.3	+3
Gal-7	AY534895	63	7.2	+8
Gal-8	AY534896	68	7.1	+2
Gal-9	AY534897	66	7.4	+4
Gal-10	AY534898	65	7.2	+2
LEAP-2	AY534899	76	8.8	+8
Cathelicidin	AY534900	148	16.1	+2

sequence quality and increases the sequence coverage for a particular cluster.

By implementing a bioinformatic approach that involves the clustering of more than 420,000 chicken ESTs, we have identified eight novel AMPs in the chicken. Clustering of these ESTs resulted in the generation of 34,819 chicken contigs and from this, 29,344 coding sequences were predicted. Given estimates of the gene number in human of 30,000 (Pennisi 2003), we expect to have at least partial sequence information for most chicken genes. TBLASTN searches (which searches protein queries against a nucleotide database) of known AMPs against the clustered EST contigs identified five contigs with homology to the β-defensins, which we have named Gal-4 to Gal-8. These searches also identified a novel chicken cathelicidin. BLASTP searches (which searches protein queries against a protein database translated in all six reading frames) of proteins, predicted by ESTScan to be encoded by these EST clusters, failed to identify any other gallinacins or cathelicidins.

Due to their small size and poor sequence conservation, searching for novel AMPs by conventional homology search tools such as BLAST (Altschul et al. 1997) may mean that significant hits are missed. However, the presence of conserved motifs makes these peptides good candidates for HMM profile searching. An HMM profile is a probabilistic model of a protein family multiple-sequence alignment, which uses position-specific scores to indicate the likelihood of each amino acid occurring in each position in the alignment (Eddy 1998). To search for novel AMPs that could be missed by BLAST, we constructed HMM profiles for a number of AMP families, including α-defensins, gallinacins, cathelicidins and hepcidins. The HMM profiles were then used to search against the database of proteins predicted by ESTScan from the EST clusters. This method identified two additional gallinacins (Gal-9 and Gal-10). Furthermore, this approach also led to the identification of a gallinacin-like sequence that has unusual cysteine spacing. There is evidence from other species that β-defensins with alternative cysteine motifs are still active as AMPs (Maxwell et al. 2003).

We have cloned and sequenced all the novel AMPs identified in this study and submitted the sequences to GenBank. The accession numbers and properties of the predicted encoded proteins are summarised in Table 2. All the sequences were in agreement with the EST predictions, except for Gal-10, which had a single synonymous change at base position 159 from C to T. We have examined the expression of the known gallinacins (Gal-1–Gal-3), the novel gallinacins (Gal-4–Gal-10), cathelicidin and LEAP-2 in a panel of 21 different tissues from a healthy 3-week-old chicken (Fig. 4). These tissues cover the digestive system, the respiratory system, the genito-urinary system and several other areas of the chicken anatomy.

The known gallinacins Gal-1 and Gal-2 are expressed strongly in the bone marrow and the lung, as has been previously shown (Zhao et al. 2001). However, we have also shown strong expression of Gal-1 and Gal-2 in the testis, moderate expression in the bursa and intestine and low expression in the cloaca, gall bladder, brain and pancreas. Gal-2 is also expressed at low levels in the trachea, air sacs and spleen. The third known gallinacin, Gal-3, was expressed in the tongue and bone marrow, as has previously been shown (Zhao et al. 2001), however, we found no Gal-3 expression in other tissues.

The novel gallinacins (Gal-4–Gal-10) exhibit variable expression across most of the tissues examined, with different gallinacins being expressed in different tissues. The phylogenetically related Gal-4 and Gal-5 show a similar pattern of expression, as does Gal-7, with all three being highly expressed in the bone marrow and testis. Gal-8 is also strongly expressed in the testis, and along with Gal-6 shows very strong expression in the liver, gall bladder and kidneys. Gal-9 is the only novel gallinacin to be expressed in the tongue and also shows low expression in the oesophagus, trachea, brain and bone marrow whilst Gal-10 shows low expression in the large intestine, kidneys and testis.

In addition to LEAP-2 being highly expressed in the liver, similar levels of expression were found in the

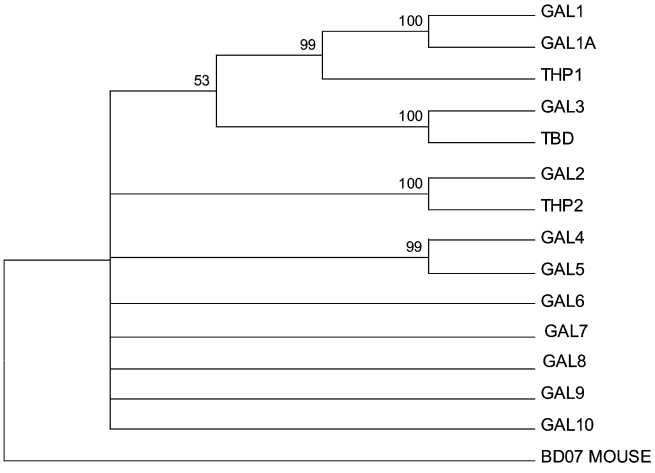


Fig. 1 Neighbor-joined tree of the gallinacin family of AMPs. Constructed using MEGA, version 2.1 (Poisson corrected model, 1,000 bootstrap replicates). Branches with less than 50% bootstrap support have been collapsed. *GAL1–10* Gallinacins (Gal-) 1–10, *GAL1A* Gal-1α, *THP* turkey heterophil peptide, *TBD* turkey β-defensin, *BD07 MOUSE* mouse β-defensin 7 (Q91V70)

Table 3 Evidence of adaptive evolution among sites in chicken gallinacins. ℓ Log-likelihood value for model, d_N/d_S Ratio of rates of synonymous (d_S) and non-synonymous (amino acid changing: d_N) changes among the DNA sequences

Model	ℓ	d_N/d_S	Positively Selected Sites
M0—one-ratio	−2,662.94	0.5834	
M1—neutral	−2,581.48	0.8923	
M2—selection	−2,541.32	2.5815	25,29,30,33,34,39,42,45,47,49,52,54,55,57,58,63,65
M3—discrete	−2,516.42	1.2094	21,23,24,25,26,27,28,29,30,32,33,34,35,39,42,45,46, 47,48,49,52,54,55,56,57,58,61,62,63,64,65
M7—β	−2,531.96	0.6170	
M8—βand ω	−2,516.42	1.2987	25,29,30,42,45,52,57,63,65

intestine, gall bladder and kidneys. These results are consistent with studies of human LEAP-2, which is also expressed in the liver, kidney and colon (Krause et al. 2003). Chicken cathelicidin is expressed across a wide variety of tissues, but shows particularly high levels of expression in the bursa, testis and bone marrow, and is the only novel AMP to show expression in the gizzard.

Evolutionary analysis of the gallinacins

We have performed an evolutionary analysis of the gallinacin family and have detected sites that are under positive selection in these molecules. A neighbor-joining phylogenetic tree was reconstructed from the amino acid alignment (Fig. 1). This tree topology was used in the subsequent analyses to detect adaptive evolution. To test for variable ωratios among phylogenetic lineages, the one-ratio model (Goldman and Yang 1994)—which assumes the same ω ratio for all lineages—was compared using the LRT to the free-ratio model (Yang 1998), which assumes an independent ω ratio for each branch. The free-ratio model is not significantly better than the one-ratio model ($P>0.5$) but does, however, predict variable ωvalues among lineages, some of which are greater than 1 (Fig. 2). Since the LRT did not reveal a significant difference, we cannot conclude that there is evidence of positive selection among the gallinacin lineages.

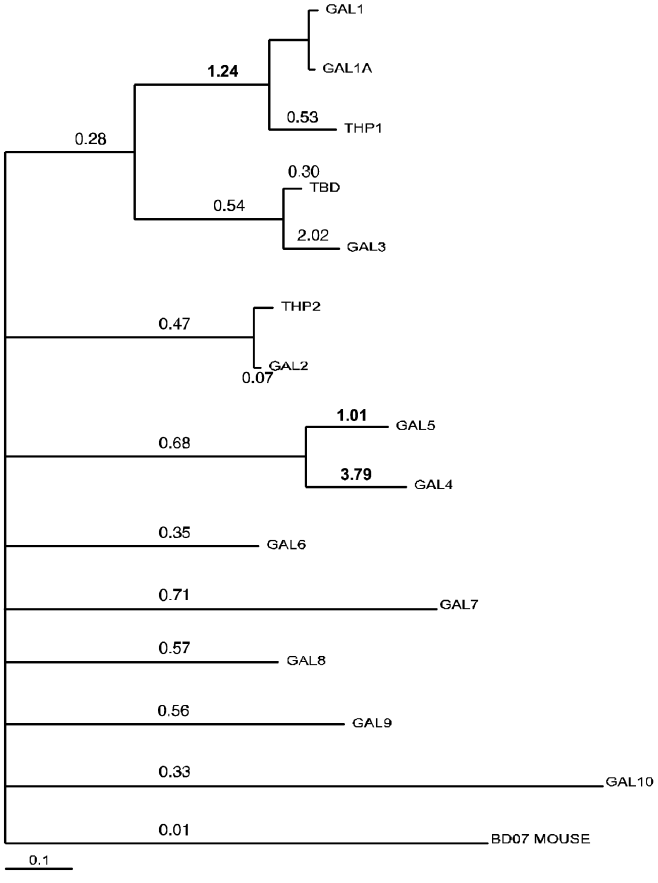


Fig. 2 Phylogeny of gallinacins. Branch lengths were estimated by maximum likelihood under the free-ratio model, which assumes an independent ω value for each branch. Branches with no ω values shown had values=∞. ω values >1 are shown in *boldface*

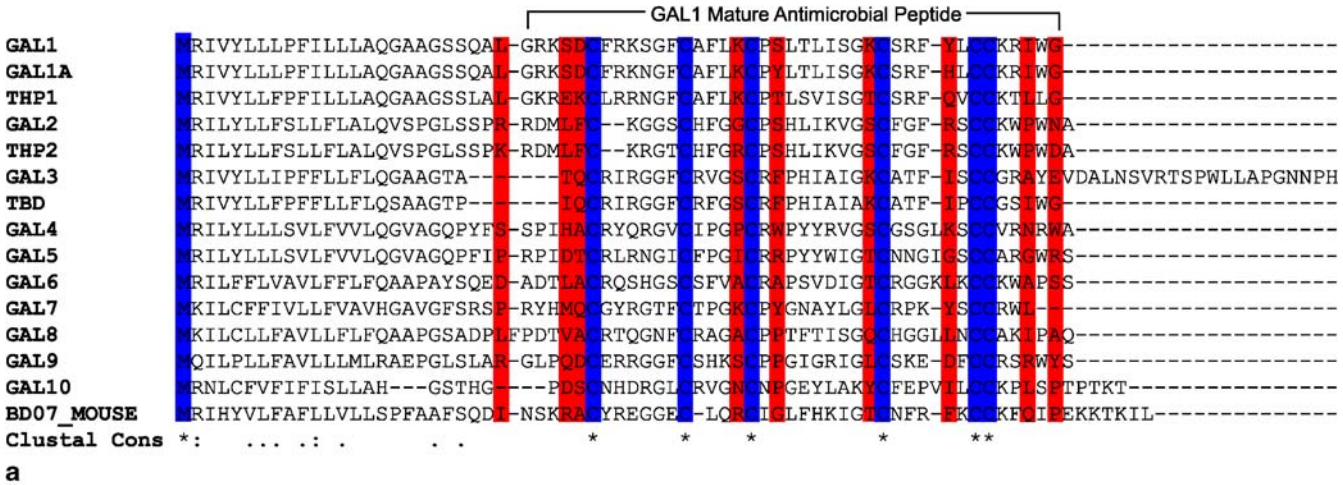
Table 4 Likelihood ratio test to detect adaptive evolution

Models	2Δℓ	χ^2value	*df*	*P*-value
M1 versus M2	2(−2581.48–2541.32)	80.32	2	<0.001
M0 versus M3	2(−2662.94–2541.32)	243.24	4	<0.001
M7 versus M8	2(−2531.96–2516.42)	31.08	2	<0.001

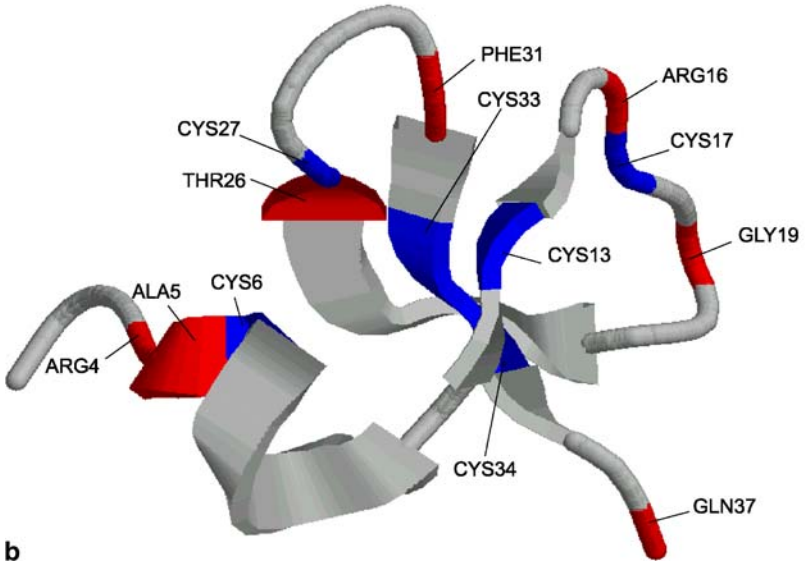
To test for positive selection at individual amino acid sites, LRTs were carried out between model M0 and M3, M1 and M2, and M7 and M8. All 3 models (M2, M3 and M8) which allow for selection (Table 3) are significantly favoured over the other models ($P<0.001$) in all cases (Table 4). Gallinacins are encoded as prepropeptides that are proteolytically cleaved to release the C-terminal AMP. All of the sites predicted to be subject to positive selection are located in the mature AMP and not in the prepropeptide region (Fig. 3a), suggesting functional significance. The sites predicted to be subject to positive selection have been displayed superimposed on the three-dimensional structure of mouse β-defensin 7, a related molecule for which a three-dimensional structure is available (Fig. 3b). The sites under positive selection occur throughout the molecule and no particular pattern or clustering of sites is discernable.

Discussion

We have applied a bioinformatics approach that involves the clustering of more than 420,000 ESTs to the identification of novel AMPs in the chicken. This approach has



a



b

Fig. 3 a Sites predicted to be under positive selection in the gallinacins. Sites predicted to be under positive selection are *highlighted* in the multiple-sequence alignment. Sites shown in *red* are those sites predicted to be under positive selection (model M8). Posterior probabilities for these sites are all greater than 0.95. Sites shown in *blue* are the sites that 100% conserved across all operational taxonomic units (OTUs). The mature AMP for Gal-1 is highlighted. **b** The structure of the mature mouse β-defensin 7 (PDB entry = 1E4T) was displayed using RasMol, version 2.7.2.1 (http://www.openrasmol.org/software/rasmol/). Sites shown in *red* are those sites predicted to be under positive selection in the gallinacins. Sites shown in *blue* are the sites that 100% conserved across all OTUs

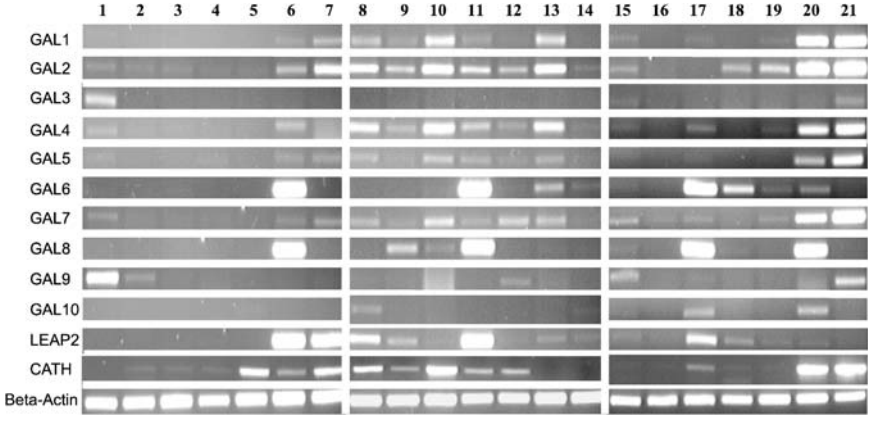


Fig. 4 Expression of Gal-1–Gal-10, cathelicidin (*Cath*) and liver-expressed antimicrobial peptide 2 (LEAP-2) in a panel of tissues from a healthy chicken. *1* tongue, *2* oesophagus, *3* proventriculus, *4* crop, *5* gizzard, *6* liver, *7* small intestine, *8* LARGE intestine, *9* cloaca, *10* bursa of Fabricius, *11* gall bladder, *12* trachea,*13* lung, *14* air sacs, *15* brain, *16* skin, *17* kidney, *18* spleen, *19* pancreas, *20* testis, *21* bone marrow

identified nine novel AMPs, seven of which are gallinacins, one a cathelicidin and one a LEAP-2 (Lynn et al. 2003). We have shown the differential expression of these genes in a panel of tissues from a single chicken and have cloned and sequenced the mRNAs encoding these novel AMPs. Identification of these AMPs in the chicken will aid in the study of the innate immune response of the chicken, economically an important species. Moreover, these novel AMPs may be exploited for the development of new therapeutic agents for economically significant chicken diseases such as coccidiosis, which results in a loss to the world poultry industry that is estimated at $700 million annually. These AMPs could be potentially developed as natural alternatives to the artificial antibiotics that are commonly fed to chickens and which are of growing public concern.

In this study, we have also detected positive selection at several amino acid sites located in the active antimicrobial peptide region of the gallinacin family of antimicrobial peptides. It is likely that as birds evolved to occupy new niches, they were faced with new ranges of microbial pathogens. Evolution of antimicrobial peptides with new sensitivities capable of targeting novel infectious agents would confer a selective advantage. There is experimental evidence that gallinacin peptides are diverse in their potency against different pathogens. For example, Gal-1 and Gal-1α are active against the yeast *Candida albicans*, whereas the others tested so far do not show activity (Evans et al. 1995; Harmon 1998). Our results indicate that gallinacins have been subject to adaptive evolution to increase the structural and functional diversity of this protein family. This is an effective response in an arms race against an increasing diversity of microbial pathogens.

Acknowledgements This research was supported by the Food Institutional Research Measure grant no. 01/R&D/D/135 from the Irish Department of Agriculture, Food and Rural Development. All experiments described in this manuscript comply with the current laws of the Republic of Ireland. Authors David Lynn and Rowan Higgs contributed equally to this work.

References

Agerberth B, Gunne H, Odeberg J, Kogner P, Boman HG, Gudmundsson GH (1995) FALL-39, a putative human peptide antibiotic, is cysteine-free and expressed in bone marrow and testis. Proc Natl Acad Sci USA 92:195–199

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Bals R, Wang X, Zasloff M, Wilson JM (1998) The peptide antibiotic LL-37/hCAP-18 is expressed in epithelia of the human lung where it has broad antimicrobial activity at the airway surface. Proc Natl Acad Sci USA 95:9541–9546

Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, Eddy SR, Griffiths-Jones S, Howe KL, Marshall M, Sonnhammer EL (2002) The Pfam protein families database. Nucleic Acids Res 30:276–280

Cowland JB, Johnsen AH, Borregaard N (1995) hCAP-18, a cathelin/pro-bactenecin-like protein of human neutrophil specific granules. FEBS Lett 368:173–176

Eddy SR (1998) Profile hidden Markov models. Bioinformatics 14:755–763

Evans EW, Beach GG, Wunderlich J, Harmon BG (1994) Isolation of antimicrobial peptides from avian heterophils. J Leukoc Biol 56:661–665

Evans EW, Beach FG, Moore KM, Jackwood MW, Glisson JR, Harmon BG (1995) Antimicrobial activity of chicken and turkey heterophil peptides CHP1, CHP2, THP1, and THP3. Vet Microbiol 47:295–303

Frohm Nilsson M, Sandstedt B, Sorensen O, Weber G, Borregaard N, Stahle-Backdahl M (1999) The human cationic antimicrobial protein (hCAP18), a peptide antibiotic, is widely expressed in human squamous epithelia and colocalizes with interleukin-6. Infect Immun 67:2561–2566

Gallo RL, Kim KJ, Bernfield M, Kozak CA, Zanetti M, Merluzzi L, Gennaro R (1997) Identification of CRAMP, a cathelin-related antimicrobial peptide expressed in the embryonic and adult mouse. J Biol Chem 272:13088–93

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol Biol Evol 11:725–736

Harmon BG (1998) Avian heterophils in inflammation and disease resistance. Poultry Sci 77:972–977

Harwig SS, Swiderek KM, Kokryakov VN, Tan L, Lee TD, Panyutich EA, Aleshina GM, Shamova OV, Lehrer RI (1994) Gallinacins: cysteine-rich antimicrobial peptides of chicken leukocytes. FEBS Lett 342:281–285

Hase K, Eckmann L, Leopard JD, Varki N, Kagnoff MF (2002) Cell differentiation is a key determinant of cathelicidin LL-37/human cationic antimicrobial protein 18 expression by human colon epithelium. Infect Immun 70:953–963

Iseli C, Jongeneel CV, Bucher P (1999) ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proceedings of the international conference on intelligent systems molecular biology, pp 138–148

Kagan BL, Selsted ME, Ganz T, Lehrer RI (1990) Antimicrobial defensin peptides form voltage-dependent ion-permeable channels in planar lipid bilayer membranes. Proc Natl Acad Sci USA 87:210–214

Krause A, Sillard R, Kleemeier B, Kluver E, Maronde E, Conejo-Garcia JR, Forssmann WG, Schulz-Knappe P, Nehls MC, Wattler F, Wattler S, Adermann K (2003) Isolation and biochemical characterization of LEAP-2, a novel blood peptide expressed in the liver. Protein Sci 12:143–152

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Larrick JW, Hirata M, Zhong J, Wright SC (1995) Anti-microbial activity of human CAP18 peptides. Immunotechnology 1:65–72

Liu L, Zhao C, Heng HH, Ganz T (1997) The human β-defensin-1 and α-defensins are encoded by adjacent genes: two peptide families with differing disulfide topology share a common ancestry. Genomics 43:316–320

Lynn DJ, Lloyd AT, O'Farrelly C (2003) In silico identification of components of the Toll-like receptor (TLR) signaling pathway in clustered chicken expressed sequence tags (ESTs). Vet Immunol Immunopathol 93:177–184

Malm J, Sorensen O, Persson T, Frohm-Nilsson M, Johansson B, Bjartell A, Lilja H, Stahle-Backdahl M, Borregaard N, Egesten A (2000) The human cationic antimicrobial protein (hCAP-18) is expressed in the epithelium of human epididymis, is present in seminal plasma at high concentrations, and is attached to spermatozoa. Infect Immun 68:4297–4302

Maxwell AI, Morrison GM, Dorin JR (2003) Rapid sequence divergence in mammalian β-defensins by adaptive evolution. Mol Immunol 40:413–421

Murakami M, Ohtake T, Dorschner RA, Gallo RL (2002a) Cathelicidin antimicrobial peptides are expressed in salivary glands and saliva. J Dent Res 81:845–850

Murakami M, Ohtake T, Dorschner RA, Schittek B, Garbe C, Gallo RL (2002b) Cathelicidin anti-microbial peptide expression in sweat, an innate defense system for the skin. J Invest Dermatol 119:1090–1095

Nielsen R, Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929–936

Nizet V, Ohtake T, Lauth X, Trowbridge J, Rudisill J, Dorschner RA, Pestonjamasp V, Piraino J, Huttner K, Gallo RL (2001) Innate antimicrobial peptide protects the skin from invasive bacterial infection. Nature 414:454–457

Notredame C, Higgins DG, Heringa J (2000) T-Coffee: a novel method for fast and accurate multiple-sequence alignment. J Mol Biol 302:205–217

Pennisi E (2003) Human genome. A low number wins the GeneSweep Pool. Science 300:1484

Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J (2003) TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 19:651–652

Ramanathan B, Davis EG, Ross CR, Blecha F (2002) Cathelicidins: microbicidal activity, mechanisms of action, and roles in innate immunity. Microbes Infect 4:361–372

Satchell DP, Sheynis T, Shirafuji Y, Kolusheva S, Ouellette AJ, Jelinek R (2003) Interactions of mouse Paneth cell α-defensins and α-defensin precursors with membranes. Prosegment inhibition of peptide association with biomimetic membranes. J Biol Chem 278:13838–13846

Scocchi M, Wang S, Zanetti M (1997) Structural organization of the bovine cathelicidin gene family and identification of a novel member. FEBS Lett 417:311–315

Shin SY, Kang SW, Lee DG, Eom SH, Song WK, Kim JI (2000) CRAMP analogues having potent antibiotic activity against bacterial, fungal, and tumor cells without hemolytic activity. Biochem Biophys Res Commun 275:904–909

Tamamura H, Murakami T, Horiuchi S, Sugihara K, Otaka A, Takada W, Ibuka T, Waki M, Yamamoto N, Fujii N (1995) Synthesis of protegrin-related peptides and their antibacterial and anti-human immunodeficiency virus activity. Chem Pharm Bull (Tokyo) 43:853–858

Travis SM, Anderson NN, Forsyth WR, Espiritu C, Conway BD, Greenberg EP, McCray PB Jr, Lehrer RI, Welsh MJ, Tack BF (2000) Bactericidal activity of mammalian cathelicidin-derived peptides. Infect Immun 68:2748–2755

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang Z (1998) Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol 15:568–573

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431–449

Zhao C, Nguyen T, Liu L, Sacco RE, Brogden KA, Lehrer RI (2001) Gallinacin-3, an inducible epithelial β-defensin in the chicken. Infect Immun 69:2684–2691