Online journal ISSN - 1676-5680
GMR Genetics and Molecular Research
EVOLUTION AND TECHNOLOGY

Avian *TAP* genes: detection of nucleotide polymorphisms and comparative analysis across species

L. Sironi[1], B. Lazzari[2], P. Ramelli[1], A. Stella[2] and P. Mariani[1]

[1]Livestock Genomics Unit, Parco Tecnologico Padano - CERSA, Lodi, Italy
[2]Statistical Genetics and Bioinformatics Unit,
Parco Tecnologico Padano - CERSA, Lodi, Italy

Corresponding author: P. Mariani
E-mail: paola.mariani@tecnoparco.org

Genet. Mol. Res. 7 (4): 1267-1281 (2008)
Received August 4, 2008
Accepted September 19, 2008
Published November 18, 2008

The nucleotide sequence data reported in this paper have been submitted to the EMBL Nucleotide Sequence Database (European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK), and have been assigned the following accession numbers: from AM420706 to AM420797, and from AM420901 to AM420999.

ABSTRACT. *TAP1* and *TAP2* genes code for the two subunits of the transporter associated with antigen processing (TAP), and in chicken they are located between the two MHC class I genes. Using primers based on chicken sequences, the genomic regions corresponding to chicken *TAP1* exons 6 to 7 and *TAP2* exons 4 to 6 (which encode portions of the chicken TAP1 and TAP2 molecules corresponding to the human peptide-binding regions) were amplified and sequenced from chicken (70 birds), turkey (24), pheasant (6), and guinea fowl (7). A total of 80 within-species single nucleotide polymorphisms (SNPs) were identified. None of the chicken SNPs detected here was present in public databases. The SNP frequencies in chicken were 9.57 SNP/kb in *TAP1* and 19.16 SNP/kb in *TAP2*, while turkey showed similar SNP frequencies in the two genes. Putative amino acid sequences were inferred to identify non-synonymous substitutions.

The alignment of the consensus polypeptide sequences showed that most of the amino acid variations were conserved or semi-conserved substitutions. In conclusion, a high variability in the level of nucleotide polymorphism was observed within the two genes, with chicken showing the highest polymorphism rate in both genes. Most of the SNPs identified were within introns, and a general conservation of both amino acid numbers and characteristics of residues among and within the species was found. These data underline the functional importance of these molecules, but also suggest their capacity to bind different antigenic peptides.

Key words: Poultry; Single nucleotide polymorphisms; Major histocompatibility complex; *TAP1*; *TAP2*

INTRODUCTION

The maintenance of immunity against pathogens and tumors is mostly due to the presentation of peptides bound to major histocompatibility complex (MHC) class I molecules to CD8+ cytotoxic T lymphocytes. Antigens located in the cell cytosol are cleaved into peptides, and these are then translocated across the endoplasmic reticulum (ER) membrane by the transporter associated with antigen processing (TAP) through an ATP-dependent process (Androlewicz et al., 1993), and are loaded onto MHC class I molecules, which are present on most nucleated cells (reviewed in Pamer and Cresswell, 1998 and Bouvier, 2003). The TAP molecule is a heterodimer constituted by the TAP1 and TAP2 proteins; the transporter is functional only when both subunits are present (Androlewicz et al., 1994). The human TAP heterodimer is located in the ER and *cis*-Golgi membranes and belongs to the ATP-binding cassette transporter super-family (see reviews in Schmitt and Tampè, 2000 and Abele and Tampè, 2004). Both TAP subunits are in contact with the peptide, and four regions (two in TAP1 and two in TAP2) have been identified that contribute to the peptide-binding site (Nijenhuis and Hämmerling, 1996; Nijenhuis et al., 1996). Human *TAP1* and *TAP2* genes are located on chromosome 6, within the classical class II region of the MHC - HLA (The MHC Sequencing Consortium, 1999; Beck and Trowsdale, 2000). Both *TAP1* and *TAP2* genes are made up of 11 exons (Schmitt and Tampè, 2000). Six human TAP1 and four human TAP2 alleles have been reported, and the molecules show a generally low level of polymorphism (McCluskey et al., 2004).

In chickens, the two genes are located next to each other, between the two MHC class I genes, on microchromosome 16 (GGA 16). The chicken *B-F/B-L* region is highly compact with the two *TAP* genes and the class I genes tightly linked to each other (Kaufman et al., 1999). *TAP1* and *TAP2* are in opposite transcriptional orientations and are thought to share a common central promoter region. The total length of the two genes is shorter than the corresponding human genes, the difference being mainly due to intron length and, for *TAP2*, number. Chicken *TAP1*, as for its human orthologue, is made up of 11 exons, but the first exon is much shorter in chickens than in humans (Walker et al., 2005), and the resulting polypeptide has a length of 590 amino acids. The two genes share 36% of identity. Because of the absence of the corresponding human introns 1 and 7, chicken *TAP2* gene has only 9 exons, coding for a 701-amino acid polypeptide, and shares 49% identity with its human orthologue. SSCP analysis showed that chicken *TAP* genes are highly polymorphic (Walker et al., 2005). For turkey,

by characterization of the chromosome bearing the MHC locus, sequence homology between chicken and turkey was identified, and the presence of *TAP1* gene was reported, but not *TAP2* (Chaves et al., 2007). Recently, the same authors have updated the MHC partial sequence and added the *TAP2* gene sequence data (Accession #DQ993255).

The genetic variation map described by the International Chicken Polymorphism Map Consortium (2004) suggested an average nucleotide substitution rate of 5 single nucleotide polymorphisms (SNP) *per* kilobase, estimated from the alignment of the partial genome sequences of three domestic chicken breeds (a broiler, a layer and a Chinese silkie) to the genome of the wild chicken ancestor, the red jungle fowl. The average SNP frequency across microchromosome 16 is greater (more than 12 SNP/kb), because of the presence of the highly variable MHC (International Chicken Polymorphism Map Consortium, 2004).

The aim of the present study was the comparison of the *TAP* genes in four avian species. The avian genomic regions examined are from exon 6 to exon 7 for *TAP1* and from exon 4 to exon 6 for *TAP2*, corresponding in humans to part of the region coding for the amino acid residues involved in peptide binding (Nijenhuis and Hämmerling, 1996; Nijenhuis et al., 1996), as shown by Walker and colleagues (2005). The presence of SNPs, the inter- and intra-breed polymorphism frequencies and the presence of synonymous/non-synonymous substitutions in these genes were investigated.

MATERIAL AND METHODS

Samples

Blood samples from unrelated animals of different local chicken (*Gallus gallus*) breeds (6 Pepoi, 8 Ermellinato di Rovigo, 4 Robusto Maculato and 7 Robusto Lionato) and three commercial lines (18 layers, 18 broilers and 9 light breeders) were collected in tubes containing CTAD. Additional blood samples were collected from 6 wild pheasants (*Phasianus colchicus*), from 7 guinea fowls (*Numida meleagris*) of the Camosciata breed, from one commercial turkey (*Meleagris gallopavo*) line (19 samples), and two local turkey breeds (3 Ermellinato di Rovigo and 2 Comune Bronzato). Genomic DNA was extracted using a commercial kit (Gentra Puregene, Qiagen).

Primer design, PCR amplification and sequencing

TAP1 and *TAP2* specific primers were designed based on the contig sequence Accession #AL023516 available in GenBank. *TAP1* primers (5'-CGAAGAGCCCACAGCCTTC-3', forward; 5'-ACCGTCAGCACTGGGGACC-3', reverse) were designed starting from base 64 of exon 6 up to base 48 of exon 7: the primer sequences are in the opposite orientation to the direction of transcription, and for this reason the subsequent analysis was carried out on the reversed-complemented sequences. *TAP2* primers (5'-AGGGCCATTTTTCTCCTCAT-3', forward; 5'-TGCATTGCTCAGAAGGTCAC-3', reverse) amplify the sequence included between base 172 of exon 4 and base 39 of exon 6.

Polymerase chain reactions (PCR) were carried out in a total volume of 15 µL with 50-100 ng genomic DNA, 0.4 pmol/µL each of forward and reverse primer, 1.5 µL 10X buffer, 1.5 mM $MgCl_2$, 0.2 mM deoxyribonucleotide triphosphates and 1.5 U AmpliTaq Gold (Applied

Biosystems), on an Eppendorf Mastercycler gradient (Eppendorf). For *TAP1*, the PCR protocol was as follows: 95°C for 10 min, 32 cycles of 94°C for 30 s, annealing at 61°C for 30 s, and 72°C for 30 s, followed by a final extension step at 72°C for 5 min. For *TAP2* the PCR protocol was 95°C for 10 min, 35 cycles of 94°C for 30 s, annealing at 53°C for 30 s, and 72°C for 30 s, followed by a final extension step at 72°C for 5 min. The resulting PCR products were separated by electrophoresis on 2% ethidium bromide-stained 1X Tris-acetate-EDTA agarose gels and purified with GenElute PCR Clean-up Kit (Sigma-Aldrich). PCR using the *TAP1* primer pair did not produce a single product from the turkey samples, and therefore, the fragments obtained were excised from the gel and eluted using the GenElute Gel Extraction Kit (Sigma-Aldrich).

Sequencing reactions were performed with BigDye (Applied Biosystems) terminator chemistry and resolved on an ABI PRISM 3100 DNA Analyzer (Applied Biosystems), according to the protocols from the manufacturer. To minimize sequencing errors, all fragments were sequenced in both directions.

Sequence analysis

The electropherograms obtained were tested for quality with the Phred program using the pipeline available at: http://asparagin.cenargen.embrapa.br/phph/ (Togawa and Brigido, 2003; Togawa et al., 2006): at least 30% of the bases analyzed for each gene reached an average Phred quality score >30. Forward-reverse consensus sequences were obtained for each sample by manual editing, using Chromas 1.45 (http://www.technelysium.com.au/chromas.html), followed by alignment and conversion to FASTA format. All sequences were then clustered in species-specific contigs with ClustalW (Chenna et al., 2003). Interspecific sequence alignments were generated with MAP2 (Ye and Huang, 2005). Database searches (GenBank) were carried out with BLASTn (Altschul et al., 1997). BLASTn was also used to compare the sequences produced within the present study with the EST data available at http://www.chick.manchester.ac.uk/ (Boardman et al., 2002). The same database was searched to ascertain if the same SNPs had already been identified in previous studies.

SNP identification

SNP discovery was performed using PolyBayes (Marth et al., 1999). The PhredPhrap script was used for automatic electropherogram reading (Phred) (Ewing and Green, 1998; Ewing et al., 1998) and sequence assembly (Phrap) (Green P, 1994, http://www.phrap.org/). The Phrap sequence assembly program was fed with the multifasta file containing the sequences produced in the present study, and with the selected parameters (-forcelevel 6, -repeat_stringency .8) sequences were split in species-specific groups, assembled in fully overlapping contigs, and a consensus sequence was determined for each contig. Consed (Gordon, 2004) was used for alignment visualization and editing. In the PolyBayes analysis, the expected *a priori* polymorphism frequency was set at 0.005 (5 SNP/kb; International Chicken Polymorphism Map Consortium, 2004). All polymorphisms identified were manually inspected.

Amino acid comparison

The coding regions of chicken genomic sequences were inferred from alignments

to homologous chicken cDNA sequences retrieved from GenBank (AJ843261 for TAP1 and AJ843262 for TAP2) and supported by Splice Predictor analysis (Brendel et al., 1998). The coding sequences for turkey TAP1 and turkey, pheasant and guinea fowl TAP2 were assumed to have the same gene organization as in chicken. The results of the analysis of putative splicing sites with the Splice Predictor program was in accordance with these assumptions. The deduced coding sequences were translated with Transeq (EMBOSS: http://emboss.sourceforge.net/apps/) and compared to chicken TAP1 and TAP2 sequences retrieved from GenBank (CAH58737 for TAP1 and CAH58738 for TAP2).

Nucleotide diversity analysis

The estimation of nucleotide diversity was carried out using the program DnaSP 4.10 (Rozas et al., 2003). The analysis was conducted using only the SNPs described in the present study as polymorphic sites. The polymorphism frequencies are indicated as $\pi \times 10^3$, thus giving the number of SNP/kb.

RESULTS

Based on the data published by Walker and colleagues (2005), the *TAP1* fragment analyzed in the present study codes for part of the corresponding second region (Q453-R487) that contributes to peptide binding in human TAP1 (Nijenhuis and Hämmerling, 1996; Nijenhuis et al., 1996). On the other hand, the *TAP2* region investigated codes for part of the corresponding first region (P301-M389) and for the corresponding complete second region (I414-M433) involved in peptide binding in human TAP2 (Nijenhuis and Hämmerling, 1996). The PCR fragments obtained for chicken had the expected sizes (431 bp for *TAP1* and 399 bp for *TAP2*), whereas for turkey, 206-bp and 404-bp products were obtained, respectively. Pheasant and guinea fowl samples gave *TAP2* fragments of 399 and 437 bp, respectively. The use of *TAP1* primers produced sequences showing no homology with any *TAP1* sequence present in GenBank for pheasant, whereas guinea fowl sequences showed homology to putative *TAP1* genes, but had different sizes for each sample (results not shown).

Homology analysis of species-specific consensus sequences was carried out using BLASTn against the GenBank nr database. With regard to *TAP1*, chicken consensus sequence produced in the present study showed 99% homology with *Gallus gallus* MHC sequence (Accession #AB268588), whereas our turkey sequences showed 97% homology to turkey MHC partial sequence (Accession #DQ993255). For the *TAP2* consensus sequences, chicken showed 97% homology with *Gallus gallus* MHC accessions (#AL023516, and #AB268588), while turkey had 99% homology with turkey MHC partial sequence (DQ993255). Pheasant had similar homology with chicken (92%), turkey (91%), and quail (89%) accessions; guinea fowl had homology with chicken (82%), and quail (79%), whereas with turkey an 88% homology was identified only for a limited region of the sequence (bases 3-341).

Interspecific variations, for both *TAP1* and *TAP2*, were identified considering only the positions that gave an intraspecifically monomorphic result or that carried a specific polymorphism with the two possible nucleotides different from that observed in the species compared. The interspecies alignment of *TAP1* chicken and turkey sequences revealed a total of 10 transitions and 5 transversions. As shown in Figure 1, the difference in length is explained by the

presence of an additional 225-bp sequence block in chicken (nucleotides 112-336). Interspecies *TAP2* alignments (Figure 2) revealed that chicken and pheasant sequences - both having the same length of 399 bp - contained 11 transitions and 10 transversions. Compared to these species, turkey sequences had an insertion of 5 nucleotides between bases 343 and 344. Apart from this difference, a comparison of sequences between chicken and turkey revealed the presence of 11 transitions and 7 transversions, whereas there were 13 transitions and 9 transversions between pheasant and turkey. Guinea fowl sequences (437 bp in length) showed a repeated block of 39 nucleotides between bases 296 and 297 when compared to the other three species. The alignment of guinea fowl sequences with chicken sequences highlighted 21 transitions, 6 transversions and 1 indel, with pheasant sequences 21 transitions, 13 transversions and 1 indel, and with turkey sequences 25 transitions, 10 transversions and 1 indel. Furthermore, turkey had 5 additional bases (nucleotides 344-348) located between bases 381 and 382 of the guinea fowl sequence.

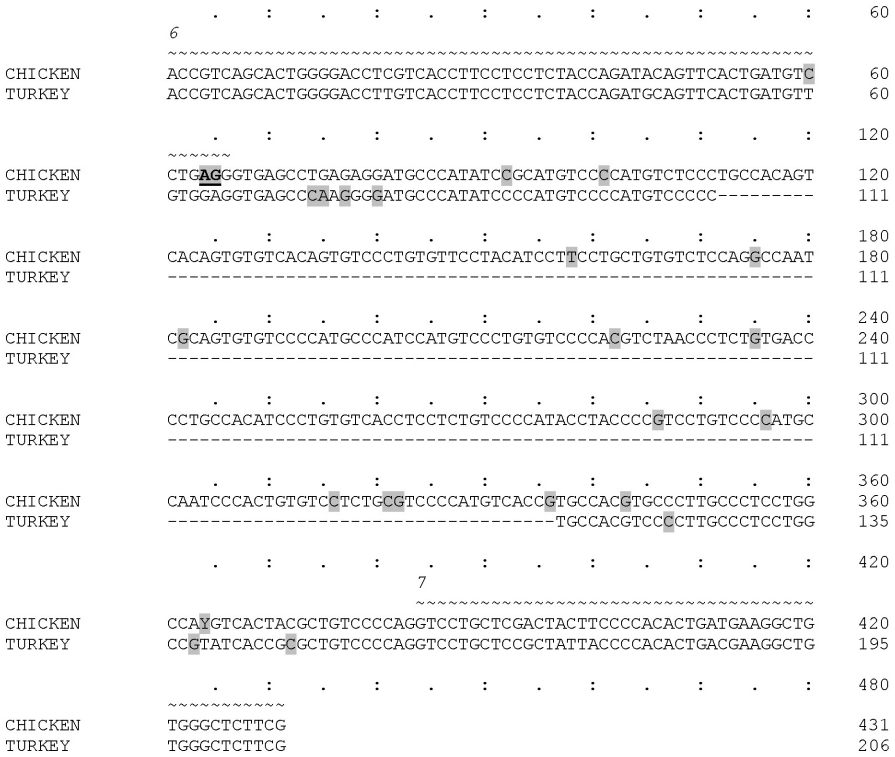

Figure 1. Interspecies alignment of the consensus *TAP1* genomic sequences. Shaded nucleotides are the single nucleotide polymorphism (SNP) sites identified with PolyBayes (18 chicken SNPs, 7 turkey SNPs); underlined and bolded SNPs (chicken SNPs 64-65) are the non-synonymous substitutions; tildes above the alignment indicate the partial chicken exons (exon number is above the tildes).

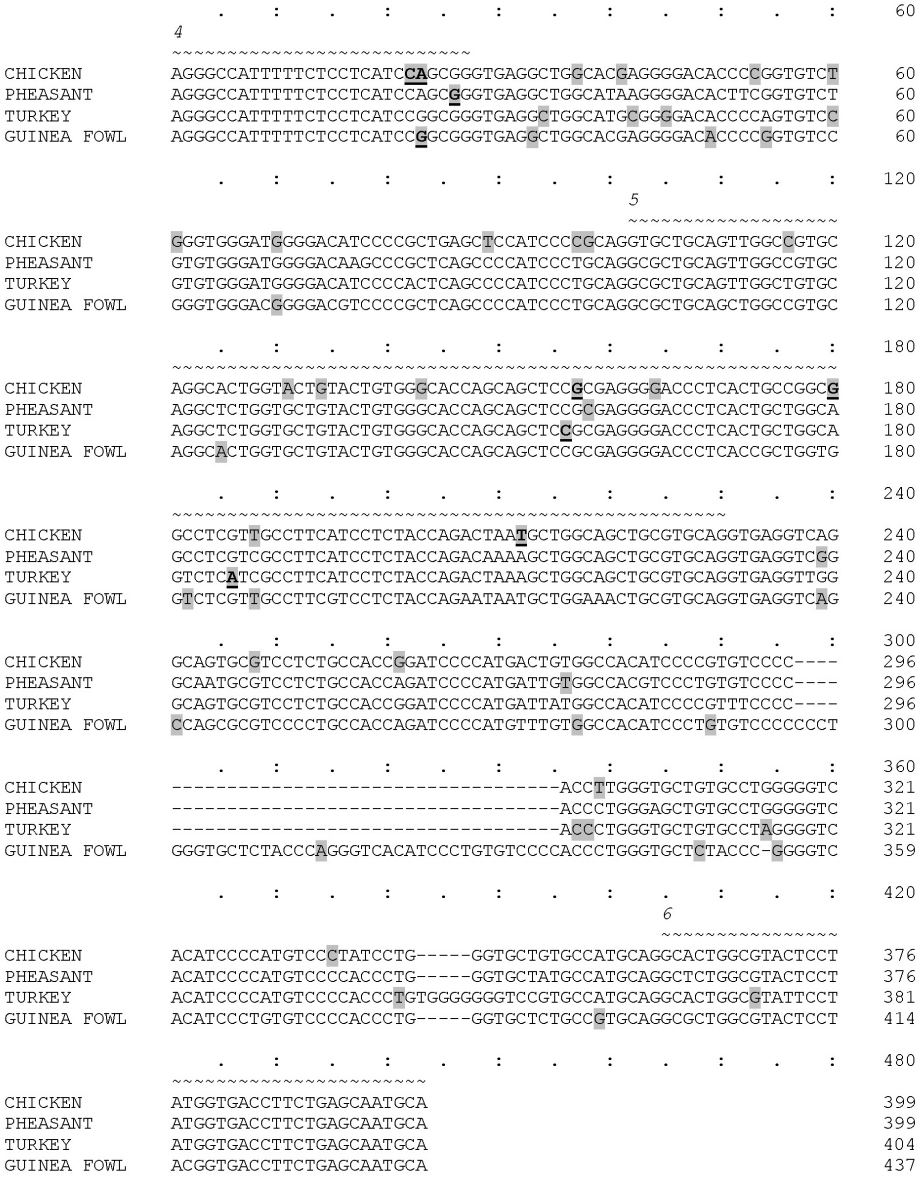

Figure 2. Interspecies alignment of the consensus *TAP2* genomic sequences. Shaded nucleotides are the single nucleotide polymorphism (SNP) sites identified with PolyBayes (24 chicken SNPs, 4 pheasant SNPs, 11 turkey SNPs, 16 guinea fowl SNPs); underlined and bolded SNPs (chicken SNPs 22-23-157-180-212, pheasant SNP 26; turkey SNPs 156-186; guinea fowl SNP 23) are the non-synonymous substitutions; tildes above the alignment indicate partial or complete chicken exons (exon number is above the tildes).

As indicated in Figures 1 and 2, all inserted/deleted blocks identified by interspecies alignment for both *TAP1* and *TAP2* were located within introns. The deduced consensus amino acid sequences (Figure 3) had the same TAP1 polypeptide length for chicken and turkey (38 residues). The same length was also found for the consensus sequence for the deduced TAP2 polypeptides from all four species (65 residues). Only 1 of 5 identified interspecies variations in TAP1 (position 15, Figure 3), and 2 of 7 in TAP2 (positions 36 and 41) were identified outside sites corresponding to those involved in peptide binding in human (Nijenhuis and Hämmerling, 1996; Nijenhuis et al., 1996; Walker et al., 2005). All interspecific amino acid substitutions were conserved or semi-conserved, with the exception of two non-conserved substitutions (D26R and M32T) in TAP1.

```
                                                    *     *    * *   *
                                                     \\\\\\\\\\\\\\\\\\\\\\\
CAH58737                                 268 TVSTGDLVTFLLYQIQFTDVLEVLLDYFPTLMKAVGSS 305
GaGa_TAP1                                  1 ......................R................. 38
MeGa_TAP1                                  1 ..............M.....V....R.Y...T...... 38

CAH58738    371 RAIFLLIQRVLQLAVQALVLYCGHQQLHEG.LTA.G..A.I...TKAGSCVQA.AYSYGD.LSNA    435
GaGa_TAP2     1 .......*.....................R...LTA.G..A.I...TNAGSCVQA.AYSYGD.LSNA    65
MeGa_TAP2     1 .......R.A...................R...LTA.S.IA.I...TKAGSCVQA.AYSYGD.LSNA    65
PhCo_TAP2     1 ........A.A...................R...LTA.S..A.I...TKAGSCVQA.AYSYGD.LSNA    65
NuMe_TAP2     1 .......R.A...................R...LTA.G..A.V...NNAGNCVQA.AYSYGD.LSNA    65
                ////////////////                          ////////////////////
                       * *                          *    *   **  *
```

Figure 3. Interspecies comparison of the TAP1 and TAP2 deduced amino acid consensus sequences. CAH58737 and CAH58738 are chicken TAP1 and TAP2, respectively, sequences retrieved from GenBank; dots indicate residues identical to CAH58737 in the overlapped fragment, and to CAH58738 for the first 30 TAP2 residues; shaded residues on the consensus sequences are the positions subjected to species-specific non-synonymous substitutions; asterisks above and under the alignment indicate the positions of the TAP1 (above) and TAP2 (under) interspecific amino acid variations; back slashes above the alignment indicate part of the corresponding second peptide-binding region (Q453-R487) in human TAP1 (Nijenhuis and Hämmerling, 1996; Nijenhuis et al., 1996); slashes below the alignment indicate part of the corresponding first region (P301-M389) and the second region (I414-M433) contributing to peptide-binding in human TAP2 (Nijenhuis and Hämmerling, 1996): the location of the corresponding human peptide-binding regions has been set according to Walker and colleagues (2005). GaGa = *Gallus gallus*; MeGa = *Meleagris gallopavo*; PhCo = *Phasianus colchicus*; NuMe = *Numida meleagris*.

For the detection of species-specific SNPs, two contigs of *TAP1* homologous sequences (reflecting the species of origin of the samples: chicken and turkey) and four contigs of *TAP2* homologous sequences (corresponding to the four species analyzed) were created by the Phrap sequence assembly program and analyzed using PolyBayes for SNP detection. In *TAP1*, 18 specific putative SNP sites were identified in chicken and 7 in turkey (Table 1A), while for *TAP2* 24 specific putative SNPs were identified in chicken, 11 in turkey, 4 in pheasant, and 16 in guinea fowl (Table 1B). As indicated in Table 1, each polymorphic position was characterized by only 2 possible species-specific variants. The

only species-specific indel identified was at position 97 of the chicken *TAP2* fragment, where all samples showed amplicons of the expected size of 399 bp, except for 1 Ermellinato di Rovigo bird, which was homozygous for a 398-bp allele, and 3 other birds (2 Ermellinato di Rovigo, and 1 Robusto Lionato), that were heterozygous 398/399 bp. These last 3 sequences obtained from clones were not used in the present analysis.

Table 1. List of the detected species-specific single nucleotide polymorphisms (SNP) and indication of the non-synonymous substitutions in the *TAP1* (A) and *TAP2* (B) genes.

A.

	Chicken				Turkey			
	POS	NT	PSNP	AA	POS	NT	PSNP	AA
Exon 6	60	C/T	0.99	V20	60	T		V20
	64	A-G	1	*	64	G		E22
	65	A-G	1		65	A		E22
Intron 6	74	T			74	C/T	0.99	
	75	G			75	A/G	0.99	
	77	G			77	C/G	0.58	
	80	G			80	A/G	1	
	92	C/T	0.99		92	C		
	101	C/T	1		101	C		
	158	C/T	1		~	-		
	175	A/G	0.99		~	-		
	182	A/G	1		~	-		
	222	C/T	0.99		~	-		
	235	A/G	0.99		~	-		
	286	A/G	1		~	-		
	296	C/T	0.99		~	-		
	316	A/C	0.99		~	-		
	321	C/T	1		~	-		
	322	A/G	0.99		~	-		
	336	A/G	1		~	-		
	343	A/G	1		118	G		
	347	C			122	C/T	0.99	
	363	A			138	A/G	0.99	
	364	C/T	1		139	T		
	372	C			147	C/T	0.99	

B.

	Chicken				Turkey				Pheasant				Guinea fowl			
	POS	NT	PSNP	AA	POS	NT	PSNP	AA	POS	NT	PSNP	AA	POS	NT	PSNP	AA
Exon 4	22	A-C	1	**	22	C		R8	22	C		Q8	22	C		R-Q8
	23	A-G	1		23	G		R8	23	A		Q8	23	G-A	0.99	
	26	G		R9	26	G		R9	26	G-A	0.55	R-Q9	26	G		R9
Intron 4	33	G			33	G			33	G			33	C/G	0.99	
	34	C			34	C/T	0.95		34	C			34	C		
	37	A/G	1		37	G			37	G			37	G		
	41	A/G	1		41	G			41	A			41	G		
	42	A			42	C/T	0.99		42	A			42	A		
	45	G			45	A/G	0.99		45	G			45	G		
	49	A			49	A			49	A			49	A/G	0.94	
	53	C/T	0.76		53	C			53	C			53	C		
	54	G			54	A			54	G			54	A/G	0.82	
	60	C/T	1		60	C/T	1		60	T			60	C		
	61	A/G	1		61	G			61	G			61	G		
	70	C/G	0.75		70	G			70	G			70	A/G	0.98	
	89	C/T	1		89	C			89	C			89	C		
	97	C/-	1		97	T			97	T			97	T		
	98	A/G	1		98	G			98	G			98	G		

Continued on next page

Table 1. Continued.

B.

	Chicken				Turkey				Pheasant				Guinea fowl			
	POS	NT	PSNP	AA	POS	NT	PSNP	AA	POS	NT	PSNP	AA	POS	NT	PSNP	AA
Exon 5	116	C/T	0.99	A14	116	T		A14	116	C		A14	116	C		A14
	125	A		A17	125	T		A17	125	T		A17	125	A/G	0.99	A17
	131	A/G	1	V19	131	G		V19	131	G		V19	131	G		V19
	134	A/G	1	L20	134	G		L20	134	G		L20	134	G		L20
	143	A/G	1	G23	143	G		G23	143	G		G23	143	G		G23
	156	C			156	C-T	1		156	C			156	C		
	157	G-A	1	R-H28	157	G		R-C28	157	G		R28	157	G		R28
	158	C			158	C			158	C/T	0.61		158	C		
	164	A/G	0.55	G30	164	G		G30	164	G		G30	164	G		G30
	180	G-A	1	G-S36	180	A		S36	180	A		S36	180	G		G36
	182	C			182	T			182	C			182	C/T	0.99	
	186	G		V38	186	A-G	0.99	I-V38	186	G		V38	186	G		V38
	188	C/T	1		188	C			188	C			188	C/T	0.99	
	212	T-A	1	N-K46	212	A		K46	212	A		K46	212	T		N46
Intron 5	239	A			239	G			239	G/T	0.12		239	A/G	0.99	
	241	G			241	G			241	G			241	C/G	0.99	
	248	A/G	1		248	G			248	G			248	G		
	261	C/G	0.99		261	G			261	A			261	A		
	276	T			276	T			276	A/T	0.99		276	T		
	277	G			277	G			277	G			277	A/G	0.99	
	289	G			289	G			289	G			289	G/T	0.99	
	~	-			~	-			~	-			314	A/G	0.99	
	298	C			298	C/T	0.30		298	C			337	C		
	299	C			299	C/T	1		299	C			338	C		
	300	C/T	1		300	C			300	C			339	C		
	309	G			309	G			309	G			348	C/G	0.99	
	315	G			315	A/G	1		315	G			~	-		
	316	G			316	G			316	G			354	A/G	0.99	
	336	C/G	1		336	C			336	C			374	C		
	342	T			342	C/T	0.99		342	T			380	T		
	355	A			360	A			355	A			393	A/G	0.49	
Exon 6	369	G		A55	374	C/G	0.86	A55	369	G		A55	407	G		A55

POS = position of the SNP indicated for each species analyzed (numeration according to the interspecies alignments, Figures 1 and 2, respectively); NT = nucleotide(s) at each position (indication of one or two bases if the site is monomorphic or polymorphic); PSNP = SNP probability score (indicated only for SNP sites); AA = amino acid residue(s) (followed by position, as defined by Figure 3) coded by the corresponding nucleotide(s) (in case of non-synonymous substitution there is the indication of the two possible residues, except for * and **); shaded NTs are the non-synonymous substitutions; tildes indicate a gap; dashes indicate lack of nucleotide (due to gap). *There are 4 possible base-combinations corresponding to 4 residues, but only two are observed in homozygous state (in parentheses the two non-observed in homozygosity): AG R22, GA E22 (AA K, GG G). **There are 4 possible base-combinations corresponding to 3 residues, but only two are observed in homozygous state (in parentheses the two non-observed in homozygosity): AG R8, CA Q8 (CG R, AA K).

Chicken was the species with the highest number of SNPs in both *TAP1* and *TAP2*; among the 18 *TAP1* polymorphic sites, those at positions 92 and 175 were specific to the Pepoi breed (3 heterozygous animals) and to the broiler line (4 heterozygous samples), respectively. The *TAP2* SNP at position 164 was also specific (in heterozygous form) to the same 4 broiler samples.

Regarding turkey species, the SNPs detected in the *TAP1* and *TAP2* gene fragments were mainly found within the local breeds or from the comparison of the local breeds and commercial turkey sequences. The commercial line was monomorphic in most of the SNPs identified: only 1 site (position 77) of 7 for *TAP1*, and 3 (positions 156, 186, 342) of 11 for *TAP2* were polymorphic in this line. Finally, pheasant *TAP2* polymorphisms

were only found in the heterozygous state, and in a limited number of samples, whereas 5 of the 7 guinea fowl *TAP2* sequences shared 100% identity, and 2 samples had 12 transitions and 4 transversions in the homozygous or heterozygous state.

Chicken consensus sequences for both genes were compared to the EST sequences present in the ChickEST Database using BLASTn: no significant hits were found for *TAP1*, while a single EST similar to the *TAP2* sequences was recovered (Accession #344535.3); no high confidence SNP was associated with that sequence in ChickEST.

For *TAP1*, 3 of the 18 identified chicken SNPs were located within exon 6 (polymorphic sites 60, 64 and 65). The remaining 15 SNPs were in the intronic region, as were all 7 SNPs identified in turkey (see Figure 1). Chicken polymorphisms at positions 64 and 65 were non-synonymous substitutions: the 4 base combinations corresponded to 4 possible amino acids (Table 1A), but only 2 combinations (64A-65G and 64G-65A) were observed in homozygous samples. With regard to the *TAP2* fragment, 11 of 24 SNPs in chicken, 3 of 11 in turkey, 2 of 4 in pheasant, and 4 of 16 in guinea fowl were located within exons (Figure 2). In chicken, polymorphic sites 22, 23, 157, 180, and 212 were identified as non-synonymous. The SNPs at positions 22 and 23 could give rise to 3 different amino acids, although only two of the possible variants were observed in homozygous samples. For turkey, 2 SNPs were non-synonymous (sites 156 and 186), and pheasant and guinea fowl had non-synonymous substitutions at positions 26 and 23, respectively (Table 1B). The majority of the non-synonymous nucleotide substitutions in both *TAP1* and *TAP2* resulted in conserved or semi-conserved amino acid substitutions; the exception was the substitution at position 28 (R/C) of turkey TAP2 (Table 1). Based on the data published by Walker and co-workers (2005), 3 non-synonymous TAP2 substitutions (amino acids 28 of chicken and turkey, 36 of chicken and 38 of turkey; see Figure 3) were identified outside the partial regions corresponding to those reported to be involved in peptide binding in human (Nijenhuis and Hämmerling, 1996; Nijenhuis et al., 1996).

Previous study showed an SNP frequency of 9.57 SNP/kb in chicken for *TAP1* (Sironi et al., 2007), and these data were confirmed in the present study. Within the same gene, SNP frequency of 5.90 SNP/kb was found in turkey. For *TAP2* the average SNP rates were 19.16 SNP/kb in chicken, 5.73 SNP/kb in turkey, 2.70 SNP/kb in pheasant, and 13.05 SNP/kb in guinea fowl. In the chicken *TAP2* analysis, only 23 polymorphic sites of 24 were considered, as the polymorphism at position 97 is not a simple base substitution but an indel, and as such is not taken into account by the selected analysis program (DnaSP 4.10). The results, together with the polymorphism frequency calculated within commercial lines for each species, are shown in Table 2.

DISCUSSION

In the present study, regions of the *TAP* genes (from exon 6 to exon 7 for *TAP1* and from exon 4 to exon 6 for *TAP2*) in chicken, turkey, pheasant, and guinea fowl were compared. Important information was obtained from both the interspecies analysis and the species-specific identification of SNPs. The interspecies analysis at the nucleotide level allowed the identification of differences in both length and base composition of the sequences analyzed. Among the observed differences, the most relevant was that found in *TAP1* amplicon size between turkey and chicken (206 and 431 bp, respectively; Figure

1). The similarity of turkey *TAP1* sequence with quail *TAP1* (in particular a similar sequence length in this region) supports previous phylogenetic analyses (Pimentel-Smith et al., 2000), which suggested that turkey and Japanese quail are more closely related than chicken and turkey or chicken and quail.

A great variability was also observed in the species-specific identification of SNPs (Table 2). The pheasant showed low nucleotide diversity in the *TAP2* gene (2.70 SNP/kb), while the SNP frequency for guinea fowl in the same region was second only to chicken with 13.05 SNP/kb and the SNPs identified define two main haplotypes in the *TAP2* region analyzed. Chicken exhibited the highest number of polymorphisms both in *TAP1* and *TAP2*, but very few SNPs were specific to a commercial line or a local breed. The frequency of polymorphic sites in both genes was higher than the estimated genome-wide average of 5 SNP/kb, with *TAP2* having a polymorphism rate even higher than the published GGA 16 mean frequency (International Chicken Polymorphism Map Consortium, 2004). These data confirmed the high level of polymorphism in the chicken *TAP* genes described by Walker and colleagues (2005) and Shiina and co-authors (2007). In most public databases, information was available only for the *TAP1* gene and none of the SNPs identified in the present study was present in the Ensembl Genome Browser - *Gallus gallus*, http://www.ensembl.org/Gallus_gallus/index.html, or in the USCS Genome Browser, http://genome.ucsc.edu/. In the ChickEST database, no *TAP1* sequences for the region analyzed here are currently present, and it is noteworthy that none of the chicken *TAP2* SNPs identified in the present study is, up to now, present in this database. Moreover, no polymorphism has been described by the International Chicken Polymorphism Map Consortium (2004 - Supplementary Information) in the regions of *TAP1* and *TAP2* analyzed in the present study. In both genes, broilers and layers showed different polymorphism frequencies (Table 2).

Table 2. Nucleotide diversity values.

Samples	*TAP1*			*TAP2*		
	L	S	[a]π x 10^3	L	S	π x 10^3
Chicken						
All	431	18	9.57 (0.34)	398	23	19.16 (0.50)
Bro	431	17	9.15 (0.99)	399	23	18.51 (1.28)
Lay	431	10	6.55 (0.66)	399	15	11.35 (1.22)
Turkey						
All	206	7	5.90 (1.57)	404	11	5.73 (0.85)
T. Com.	206	1	1.66 (0.38)	404	7	3.20 (0.62)
Pheasant	-	-	-	399	4	2.70 (0.69)
Guinea fowl	-	-	-	437	16	13.05 (4.24)

L = number of sites (excluding sites with gaps); S = number of polymorphic sites (SNPs identified with PolyBayes); All = all chicken or turkey samples analyzed at each locus; Bro = broilers; Lay = layers; T. Com. = turkey commercial line. [a]Nucleotide diversity, according to Nei (1987), equations 10.5 or 10.6. Standard deviation in parentheses.

Compared to the average chicken nucleotide diversity, broilers had similar SNP frequency, whereas layers had lower polymorphism frequency in both genes. The same pattern of polymorphism frequencies between layers and broilers had already been reported for *Tapasin* (Sironi et al., 2006, 2007). In the *TAPs* and *Tapasin*, layers had a relatively low level of polymorphism, whereas broilers had a very high intra-line frequency of polymor-

phism, carrying the majority of the species-specific SNPs identified in this study. This may be important as breeding and selection in commercial populations lead to a reduced diversity among animals. The layers analyzed in the present study show, as expected, a reduced level of polymorphism, while broilers have maintained a greater diversity in spite of selection. Moreover, this trend is in contrast to what was described by the International Chicken Polymorphism Map Consortium (2004), which showed that broiler and layer had a comparable mean SNP frequency (4 SNP/kb), lower than the estimated genome-wide average of 5 SNP/kb. In contrast to the chicken data, the mean nucleotide diversity in turkey was almost the same in the two genes, with the commercial line showing the lowest value for *TAP1* (Table 2) with little variability among turkeys belonging to this line, even if, it must be reported that one of these birds was polymorphic at the majority of the *TAP2* identified SNPs. Recent published data (Reed et al., 2006) reported an average of 1 SNP every 240 bp in this species. In the present study, the polymorphism level of turkey *TAP* genes was slightly higher (Table 2).

Our analyses of the chicken data showed that the majority of the SNPs identified were located within intron sequences (*TAP1*: 15 of 18, *TAP2*: 13 of 24), that 2 SNPs in TAP1 and 5 in TAP2 were non-synonymous substitutions, that there were no specific differences in the number of residues in the coded putative peptides, and that the specific amino acid substitutions were mostly conserved or semi-conserved. Based on these features, the high level of polymorphism observed seems to only partially affect the structure of these molecules, and indirect selection might be the major driving force in these genes' variability, as suggested by Shiina and colleagues (2007). Shiina and colleagues (2006, 2007) hypothesized that the chicken *TAP* genes may display a high level of polymorphism because of indirect selection due to their position within the MHC, adjacent to the highly selected *BF* genes. Moreover, the fundamental structure of the TAP molecules seems to be generally conserved in the four species analyzed here: indeed the features listed for chicken were also found in turkey, pheasant and guinea fowl. At an interspecific level, the structure of TAP molecules was also preserved: all the identified insertions/deletions were located within intron sequences (Figures 1 and 2); the number of both TAP1 and TAP2 amino acid residues in the regions analyzed did not change in the different species (Figure 3), and the amino acid substitutions were mostly conserved or semi-conserved. Nevertheless, Figure 3 shows also that most of the intra- and interspecies amino acid substitutions are located in residues corresponding to those, which in humans are involved in peptide binding. This could indicate that the gene polymorphisms potentially affect the function of the TAP molecules, thereby causing differential binding of peptides, and hence may be under direct selection.

In conclusion, the results presented here provide new structural information about the avian *TAP* genes and the polymorphism level of these genes. Significant levels of nucleotide polymorphism were found, and general conservation of both amino acid numbers and characteristics of residues among and within the species was observed. This may be attributed to the need of this specific structure for the fundamental function of the TAP molecules. At the same time, the presence of substitutions in the region reported to be important for peptide binding in humans could explain the capacity of these proteins to bind different antigenic peptides.

ACKNOWLEDGMENTS

The authors thank Francesco Salamini (Parco Tecnologico Padano, Lodi, Italy),

John Williams (Parco Tecnologico Padano, Lodi, Italy) and Sem Genini (University of Pennsylvania, Philadelphia, USA) for valuable discussion, and Martino Cassandro (University of Padua, Padua, Italy) and Veneto Agricoltura for providing blood samples from local Italian breeds. Research supported by the Italian Ministry of Education, University and Research (#RBNE01SFXY) and by Cariplo Foundation (#2001.2489/11.8094).

REFERENCES

Abele R and Tampè R (2004). The ABCs of immunology: structure and function of TAP, the transporter associated with antigen processing. *Physiology* 19: 216-224.

Altschul SF, Madden TL, Schaffer AA, Zhang J, et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389-3402.

Androlewicz MJ, Anderson KS and Cresswell P (1993). Evidence that transporters associated with antigen processing translocate a major histocompatibility complex class I-binding peptide into the endoplasmic reticulum in an ATP-dependent manner. *Proc. Natl. Acad. Sci. U. S. A.* 90: 9130-9134.

Androlewicz MJ, Ortmann B, van Endert PM, Spies T, et al. (1994). Characteristics of peptide and major histocompatibility complex class I/beta 2-microglobulin binding to the transporters associated with antigen processing (TAP1 and TAP2). *Proc. Natl. Acad. Sci. U. S. A.* 91: 12716-12720.

Beck S and Trowsdale J (2000). The human major histocompatibility complex: lessons from the DNA sequence. *Annu. Rev. Genomics Hum. Genet.* 1: 117-137.

Boardman PE, Sanz-Ezquerro J, Overton IM, Burt DW, et al. (2002). A comprehensive collection of chicken cDNAs. *Curr. Biol.* 12: 1965-1969.

Bouvier M (2003). Accessory proteins and the assembly of human class I MHC molecules: a molecular and structural perspective. *Mol. Immunol.* 39: 697-706.

Brendel V, Kleffe J, Carle-Urioste JC and Walbot V (1998). Prediction of splice sites in plant pre-mRNA from sequence properties. *J. Mol. Biol.* 276: 85-104.

Chaves LD, Krueth SB and Reed KM (2007). Characterization of the turkey MHC chromosome through genetic and physical mapping. *Cytogenet. Genome Res.* 117: 213-220.

Chenna R, Sugawara H, Koike T, Lopez R, et al. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res.* 31: 3497-3500.

Ewing B and Green P (1998). Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 8: 186-194.

Ewing B, Hillier L, Wendl MC and Green P (1998). Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8: 175-185.

Gordon D (2004). Viewing and Editing Assembled Sequences Using Consed. In: Current Protocols in Bioinformatics (Baxevanis AD and Davison DB, eds.). John Wiley and Co., New York, 11.2.1-11.2.43.

International Chicken Polymorphism Map Consortium (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* 432: 717-722.

Kaufman J, Milne S, Gobel TW, Walker BA, et al. (1999). The chicken B locus is a minimal essential major histocompatibility complex. *Nature* 401: 923-925.

Marth GT, Korf I, Yandell MD, Yeh RT, et al. (1999). A general approach to single-nucleotide polymorphism discovery. *Nat. Genet.* 23: 452-456.

McCluskey J, Rossjohn J and Purcell AW (2004). TAP genes and immunity. *Curr. Opin. Immunol.* 16: 651-659.

Nei M (1987). Molecular Evolutionary Genetics. Columbia University Press, New York.

Nijenhuis M and Hämmerling GJ (1996). Multiple regions of the transporter associated with antigen processing (TAP) contribute to its peptide binding site. *J. Immunol.* 157: 5467-5477.

Nijenhuis M, Schmitt S, Armandola EA, Obst R, et al. (1996). Identification of a contact region for peptide on the TAP1 chain of the transporter associated with antigen processing. *J. Immunol.* 156: 2186-2195.

Pamer E and Cresswell P (1998). Mechanisms of MHC class I - restricted antigen processing. *Annu. Rev. Immunol.* 16: 323-358.

Pimentel-Smith GE, Shi L, Drummond P, Tu Z, et al. (2000). Amplification of sequence tagged sites in five avian species using heterologous oligonucleotides. *Genetica* 110: 219-226.

Reed KM, Hall MK, Chaves LD and Knutson TP (2006). Single nucleotide polymorphisms for integrative mapping in the

turkey (*Meleagris gallopavo*). *Anim. Biotechnol.* 17: 73-80.
Rozas J, Sanchez-DelBarrio JC, Messeguer X and Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496-2497.
Schmitt L and Tampè R (2000). Affinity, specificity, diversity: a challenge for the ABC transporter TAP in cellular immunity. *Chembiochem* 1: 16-35.
Shiina T, Ota M, Shimizu S, Katsuyama Y, et al. (2006). Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. *Genetics* 173: 1555-1570.
Shiina T, Briles WE, Goto RM, Hosomichi K, et al. (2007). Extended gene map reveals tripartite motif, C-type lectin, and Ig superfamily type genes within a subregion of the chicken MHC-B affecting infectious disease. *J. Immunol.* 178: 7162-7172.
Sironi L, Lazzari B, Ramelli P, Gorni C, et al. (2006). Single nucleotide polymorphism discovery in the avian Tapasin gene. *Poult. Sci.* 85: 606-612.
Sironi L, Stella A, Lazzari B, Ramelli P, et al. (2007). Molecular characterization of genes involved in chicken MHC class I antigen presentation pathway. *Ital. J. Anim. Sci.* 6 (Suppl 1): 211-213.
The MHC Sequencing Consortium (1999). Complete sequence and gene map of a human major histocompatibility complex. *Nature* 401: 921-923.
Togawa RC and Brigido MM (2003). PHPH: Web based tool for simple electropherogram quality analysis. 1st International Conference on Bioinformatics and Computational Biology - IcoBiCoBi, 14th-16th May, Ribeirão Preto.
Togawa RC, Brigido MM, Santos CMR and Júnior MT (2006). The use of the PHPH tool to assembly the gene sequences that are candidate to the biotic and abiotic stress in *Musa acumitana*. XXXV Annual Meeting of the Brazilian Society of Biochemistry and Molecular Biology (SBBq), Águas de Lindóia.
Walker BA, van Hateren A, Milne S, Beck S, et al. (2005). Chicken TAP genes differ from their human orthologues in locus organisation, size, sequence features and polymorphism. *Immunogenetics* 57: 232-247.
Ye L and Huang X (2005). MAP2: multiple alignment of syntenic genomic sequences. *Nucleic Acids Res.* 33: 162-170.

