

Research

Open Access

## OVEX1, a novel chicken endogenous retrovirus with sex-specific and left-right asymmetrical expression in gonads

Danièle Carré-Eusèbe\*<sup>1,2</sup>, Noëlline Coudouel<sup>3</sup> and Solange Magre<sup>3</sup>

Address: <sup>1</sup>Endocrinologie et Génétique de la Reproduction et du Développement, INSERM, U782, 32 rue des Carnets, F-92140, Clamart – France, <sup>2</sup>Univ. Paris-Sud, UMR-S0782, Clamart, F-92140 and <sup>3</sup>Physiologie de l'Axe Gonadotrope, Unité de Biologie Fonctionnelle et Adaptative (BFA), Univ. PARIS 7 – CNRS, 4 rue MA Lagroua Weill-Hallé, 75205 Paris CEDEX 13 – France

Email: Danièle Carré-Eusèbe\* - [daniele.carre-eusebe@u-psud.fr](mailto:daniele.carre-eusebe@u-psud.fr); Noëlline Coudouel - [noelline.coudouel@univ-paris-diderot.fr](mailto:noelline.coudouel@univ-paris-diderot.fr); Solange Magre - [solange.magre@univ-paris-diderot.fr](mailto:solange.magre@univ-paris-diderot.fr)

\* Corresponding author

Published: 17 June 2009

Received: 26 January 2009

*Retrovirology* 2009, **6**:59 doi:10.1186/1742-4690-6-59

Accepted: 17 June 2009

This article is available from: <http://www.retrovirology.com/content/6/1/59>

© 2009 Carré-Eusèbe et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** In chickens, as in most birds, female gonad morphogenesis is asymmetrical. Gonads appear first rather similarly, but only the left one undergoes full differentiation and gives rise to a functional ovary. The right gonad, in which the cortex does not develop, remains restricted to the medulla and finally regresses. Opportunity was taken of this left-right asymmetry to perform a suppression subtractive hybridization screening to select for transcripts preferentially expressed in the developing left ovary as compared to the right one, and thus identify genes that are potentially involved in the process of ovarian differentiation.

**Results:** One of these transcripts, named *Ovex1* according to its expression profile, corresponds to an endogenous retrovirus that has not been previously characterized. It is transcribed as full-length and singly spliced mRNAs and contains three uninterrupted open reading frames coding potentially for proteins with homology to Gag and Pro-Pol retroviral polyproteins and a third protein showing only a weak similarity with Env glycoproteins. *Ovex1* is severely degenerated; it is devoid of typical long terminal repeats and displays some evidence of recombination. An orthologous *Ovex1* locus was identified in the genome of zebra finch, a member of a different bird order, and similar sequences were detected in turkey, guinea fowl, and duck DNA. The relationship between these sequences follows the bird phylogeny, suggesting vertical transmission of the endogenous retrovirus for more than 100 million years.

*Ovex1* is transcribed in chicken gonads with a sex-dependent and left-right asymmetrical pattern. It is first expressed in the cortex of the left indifferent gonads of both sexes. Expression is transient in the left testis and absent in the right one. In developing ovaries, *Ovex1* transcription increases sharply in the left cortex and is weakly detected in the medulla. After folliculogenesis, *Ovex1*-expressing cells constitute the follicular granulosa cell layer. *Ovex1* expression highlights a striking desquamation process that leads to profound cortical remodeling associated with follicle morphogenesis.

**Conclusion:** Evidence for a selection pressure at the protein level suggests that this endogenous retrovirus, expressed in the ovarian supporting cell lineage, might play an active role in bird ovarian physiology.

## Background

In chickens, as in most birds, gonad differentiation is characterized by left-right (L/R) asymmetry. In the female, only the left gonad becomes a functional ovary. The right one fails to fully differentiate and ultimately disappears. By contrast, both male gonads, initially asymmetrical, become functional testes. Bird sex determination is not fully understood. The heterogametic sex, with a ZW karyotype, is female and the homogametic sex, with two Z chromosomes, is male. Unlike in mammals, the initial genetic sex trigger is not clearly identified (reviewed in [1,2]).

Gonad organogenesis begins around the 4<sup>th</sup> day of incubation (E4), with identification of the genital ridge, a thickening of the coelomic epithelium on the medial aspect of the mesonephros, in which primordial germ cells, migrating from the germinal crescent, are going to settle [3]. In both sexes, the gonad epithelium is characterized by the expression of the Lim homeodomain-containing protein Lhx9 [4]. The subjacent mesenchyme, which will give rise to the so-called "medulla", expresses the steroidogenic factor SF-1 (Nr5a1/Ad4BP), Wnt-4, and by day 5, anti-Müllerian hormone (AMH) [4-7]. Early L/R asymmetry is observed in both sexes. The left gonad is larger, has a thicker epithelium, usually called "cortex", and is colonized by a greater number of primordial germ cells. This asymmetry, more pronounced in females than in males, is considered an early sign of some sex differentiation, prior to other morphological changes [8]. L/R asymmetrical expression of several genes (*estrogen receptor ER- $\alpha$* , *FET-1*, *Bmp7*, *R-Spondin1*) in female gonads had been reported and was related to the asymmetry of their differentiation [9-13]. However, the mechanism responsible for this asymmetry has been identified only recently [14-16]. It depends on the bicoid type homeobox gene *Pitx2* (Pituitary homeobox 2), a general actor in early embryo L/R differentiation [17]. In both sexes, *Pitx2c* is asymmetrically expressed in the left lateral plate mesoderm, and later in the epithelium of the left coelomic cavity in the region of gonad formation. This factor is sufficient to induce the differentiation of gonads along a "left pathway". The action of *Pitx2c* results from its effect on retinoic acid (RA) signaling. A balance between RA-synthesizing enzymes (RALDH2) and RA-metabolizing enzymes (CYP26) controls RA levels. In both sexes, *Pitx2c* is expressed in the cortex of the left gonad. *Pitx2c* repression of RALDH2 synthesis prevents RA formation. In the cortex of the right gonad, where *Pitx2c* is not expressed, RA is formed and, as CYP26 is absent, the RA cascade is activated. RA suppresses transcription of *ER- $\alpha$*  and of *SF-1*, a factor required for the expression of the cell cycle mediator *cyclin D1*, thus limiting cortical cell proliferation. Conversely, in the left cortex, where RA is absent, *ER- $\alpha$*  and *SF-1* are expressed, and *cyclin D1*-stimulated cell proliferation is activated. In

the medulla of both gonads, RA is degraded by Cyp26A1, which allows some *ER- $\alpha$*  expression.

After 6.5 days of incubation, ovaries and testes can be distinguished histologically by the differentiation in the internal region of the male gonads of testicular cords delimited by a basement membrane, enclosing germ cells and supporting Sertoli cells. These somatic cells express high levels of SOX9, DMRT1 and AMH [7,18-22].

At the same time, aromatase (P450 arom), the enzyme converting androgens to estrogens, begins to be present in the female gonads but is not expressed in the male ones [9,10,23,24]. Estrogens are essential for ovarian differentiation [25-27]. The medulla is rather similar in both female gonads. It expresses the estrogen receptor *ER- $\alpha$* , aromatase and FoxL2, a forkhead transcription factor [28]. On the other hand, the expression of *ER- $\alpha$*  in the left cortex and its absence in the right one leads to an asymmetry in the estrogen-mediated differentiation of the gonads. Primordial germ cells and somatic cells multiply in the left ovarian cortex. Germ cells enter meiosis at E15.5 [29]. Folliculogenesis occurs after hatching at E21, with the progressive formation of follicles constituted of an oocyte surrounded by a layer of somatic granulosa cells. In contrast, the development of the right gonad is limited. The surface epithelium devoid of *ER- $\alpha$*  is not stimulated by estrogens and does not proliferate. The few germ cells do not enter meiosis, and the gonad becomes vestigial after hatching. The female right gonad does not have the physiological potential to form a functional ovary and even, in the case of left ovary castration, differentiates into a testis [30]. However, early *Pitx2c* ectopic expression is able to overcome the degenerative fate of the right gonad and to direct it toward an ovarian differentiation pathway [14].

With the purpose of finding yet unknown factors involved in the early steps of ovarian cortex differentiation, we took advantage of the differential fate of the two female gonads in chicken. Suppression subtractive hybridization screening (SSH) was used to select transcripts expressed in the left differentiating ovary and underrepresented in the right gonad. In the course of this study, we identified a new endogenous retroviral element that we named *Ovex1*, whose specific expression in gonads is characterized by a sexual dimorphism and a L/R asymmetry.

More than 30,000 endogenous retroviruses or ERVs, representing 2.9% of the DNA, are present in the chicken genome [31,32]. They are remnants of ancestral retroviruses that have gained access to the germ line of a host, leading to vertical transmission of the integrated provirus to the offspring in a Mendelian fashion. They are characterized by the presence of two long terminal repeats (LTRs), flanking sequences coding for structural and

enzymatic viral proteins. The genomic organization of simple retroviruses is 5'LTR-*Gag-Pro-Pol-Env*-3'LTR. Viral protein expression is controlled by the promoter and enhancer elements located in the 5' LTR. The polyprotein Gag (group specific antigen) is a structural component of the virus particle. Pro encodes an aspartyl protease required for processing of the Gag precursor. The polyprotein Pol contains domains for reverse transcriptase (RT), RNase H, and integrase (IN). The envelope protein (Env) is composed of two domains, a surface region (SU) and a transmembrane domain (TM). After their initial integration, ERVs can copy themselves to different locations within the genome, giving rise over long periods of time to a family of related ERV elements, most of them inactivated by mutations (for a review, see [33]). Retroviruses are divided into three major classes. Class I contains elements related to gammaretroviruses (such as Moloney murine leukemia virus, MMLV) and epsilonretroviruses (as Walleye dermal sarcoma virus, WDSV). Class II elements are related to alpharetroviruses (as the Avian leukosis virus, ALV), betaretroviruses (as Mouse mammary tumor virus, MMTV), deltaretroviruses (as Human T-lymphotropic virus, HTLV) and lentiviruses (as Human immunodeficiency virus, HIV). Class III contains elements related to spumaviruses (as Human foamy virus, HFV) and ERV-L elements [34]. Some fish viruses, as the Snakehead fish retrovirus, SnRV, have an intermediate position and are said to be epsilon-like [35]. Endogenous retroviruses of the chicken remain incompletely described in spite of intensive studies. The most studied chicken ERVs are class II elements specific for *Gallus* species. The existence of retroviral sequences, related to human ERVs and representative of the other classes of retroviruses has been reported [31,32,35,36], but little is known about these elements.

Chicken *Ovex1* contains three long open reading frames (ORF). The first two consecutive ORFs are similar to *Gag* and *Pro-Pol* retroviral sequences. The third ORF in a different frame is possibly related to *Env*. A sequence orthologous to chicken *Ovex1* was found in the genome of zebra finch. We also detected the presence of similar *Gag* and *Pol* sequences in the DNA of three other domestic birds. Expression of *Ovex1* was analyzed by RT-PCR and by *in situ* hybridization in chicken gonads from embryonic day-5 to adulthood. It depends both on the sexual determinism and on the L/R asymmetry pattern of gonad differentiation. This gene is specifically transcribed in somatic cells of the ovarian cortex involved in the formation of the follicles and in granulosa cells of the adult hen ovary. Its expression enlightens the profound remodeling of the ovarian cortex that occurs during follicle morphogenesis, an important phase of the ovarian differentiation which has not yet received much attention.

## Results

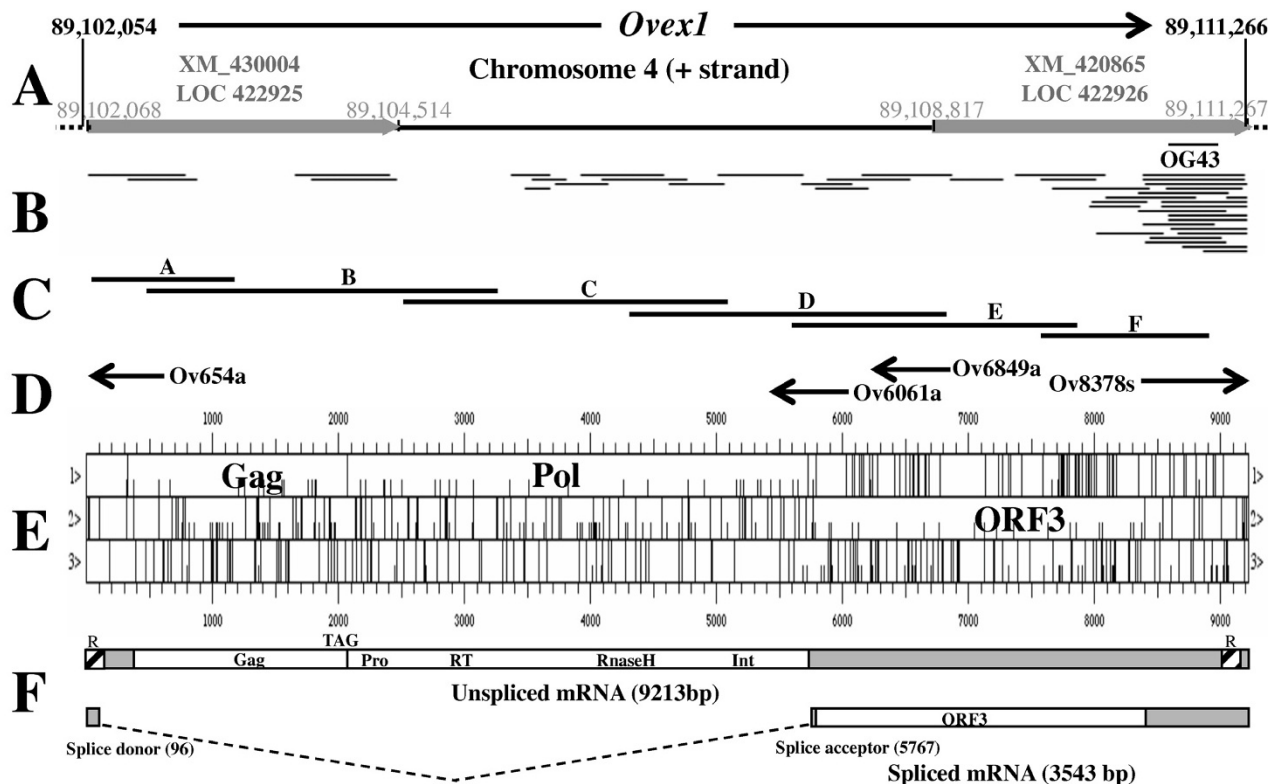
### Identification of differentially expressed genes in chicken embryonic ovaries using SSH

In order to identify unknown factors involved in early steps of the ovarian cortex differentiation, we performed a suppression subtractive hybridization screening (SSH) [37] to selected transcripts expressed in the chicken differentiating left ovary and underrepresented in the right gonad in which the cortex does not differentiate. The 8-day embryonic stage (E8) was chosen because the process of cortex development is at its beginning; meiosis has not yet started; and the gonads are easily dissected. We generated a left ovary cDNA-enriched library (LO-RO), by subtracting *RsaI*-digested cDNAs from the left ovary (LO) with those of the right ovary (RO), and a right ovary cDNA-enriched library (RO-LO) by the opposite screening. Two hundred and fifty clones from the (LO-RO) library were pick-up randomly and submitted to differential hybridization screening. Macroarrays established with the PCR-amplified inserts were hybridized with labeled cDNA probes prepared with either the (LO-RO) or the (RO-LO) subtracted libraries. Asymmetry of expression was quantified by the ratio of (LO-RO)/(RO-LO) counts. We report here the study of one clone, OG43, which had a differential hybridization-screening ratio of 154, and a high expression in gonads, preferentially the ovaries, with left-right asymmetry. This clone was shown to correspond to an endogenous retroviral element not yet identified that we named *Ovex1* (Ovary expressed, #1) in relation to its expression pattern.

### Sequence analysis of the *Ovex1* locus

Identification of the insert of clone OG43 was performed by BLAT screening of the May 2006, v2.1 draft assembly of the chicken genome (galGal3). It revealed a 99.7% identity with the 3'-untranslated region (UTR) of a hypothetical gene [GenBank:XM\_420865] (locus LOC422926), located on the chromosome 4 long arm (Fig. 1A). This gene, annotated as coding for a protein "similar to env", corresponds to the 3' region of the sequence given in Fig. 3, from nucleotide 6764.

Most of the expressed sequence tags (EST) corresponding to this locus have an ovarian origin. Some of them extend 5' from the locus (Fig. 1B), suggesting that transcription might start further upstream. Interestingly, the next locus upstream from LOC422926 in the galGal3 draft assembly, LOC422925, displays also a strict ovarian expression. It corresponds to a predicted gene, [GenBank:XM\_430004], coding for a hypothetical protein of unknown nature (Fig. 1A) and extends from nucleotide 15 to 2461 of the sequence given in Fig. 2. None of the already published ESTs overlaps with the two loci. To examine if there might be a relationship between these two neighboring loci that have the same orientation and display the same specificity

**Figure 1**

**Structure of the chicken *Ovex1* endogenous retrovirus.** (A) Genomic position of *Ovex1*. LOC422925 and LOC422926 (May 2006 *Gallus gallus* v 2.1, galGal3 genome draft assembly) are indicated. The SSH-cloned OG43 cDNA fragment is shown. (B) Published ESTs (8-26-2007). (C) Overlapping RT-PCR fragments amplified from ovarian RNA used for cDNA sequencing. (D) Determination of mRNA ends by 5' and 3'RACE PCR (primers used and direction of the synthesis). (E) ORF map. (F) Transcription profile. Two types of mRNA are produced: a mRNA corresponding to the complete genomic sequence and a mRNA with one splicing event. Position of the splice donor and acceptor sites relative to the transcription start site is indicated between brackets, and the spliced out domain represented by a dotted line. Untranslated regions are shown in gray and the 5' and 3' imperfect repeats marked by (R). Encoded protein domains are indicated. Primer sequences are given in additional file 9 (Table SI, Primers and PCR conditions).

of expression, we tried to amplify overlapping cDNA fragments from one locus to the other by RT-PCR using embryonic ovary mRNA. The series of fragments obtained demonstrates that the two loci constitute in fact a single transcription unit (Fig. 1C).

The initiation cap site of this mRNA was determined by rapid amplification of the 5'cDNA-end method (5'RACE) using, in a first experiment, Ov6849a as antisense primer (Fig. 1D). Two sequences were amplified, indicating the existence of two types of mRNA: a genomic mRNA similar to DNA and a spliced subgenomic transcript lacking the 97–5766 sequence (Fig. 1F). Additional 5'RACE experiments confirmed this result, one with a primer (Ov6061a) located downstream from the acceptor splice site which allows only amplification of the short spliced transcript,

and the second with a primer (Ov654a) located in the intron to amplify the unspliced mRNA. Both experiments gave the same 5'-terminal sequences, indicating that the cap site of the two mRNAs is presumably G<sub>+1</sub> or A<sub>+4</sub> (Fig. 2), a few bases upstream from the putative start of LOC422925. The cap site is located 23 nucleotides after a consensus TATA box (Fig. 2).

The mRNA polyadenylation site was identified by 3' RACE (Fig. 1D), using a forward primer common to both mRNAs (Ov8378s). The longest sequence obtained was polyadenylated at position 9213 (Fig. 3). Shorter sequences, polyadenylated at positions 9203 to 9211, were also found. The polyadenylation site is preceded by a consensus polyadenylation signal, AAUAAA (nt 9190 to 9195). The maximum size of the unspliced mRNA is 9213



4681 ACAGAGCTGGTGGCTCTGGCTCAGCAGGGGCATTTGGGAATAGATAAGACCTTGGCAAGGCTACAGGGTCTGGTGGTGGCCGAGATGAGAGAGGATGTAGACAGGTATGTCAAC  
 T E L V A L A L E Q Q H L G I D K T L A R L Q G A G W W P E M R E D V D R Y V N 1477

4801 AACTGCCTCACTGGCAGCAAAATCTCGTATGCAAAAGCCCTCTGGCCACCAAGGATTTCTGGCCCTGGAGCAAGCTACAGATGGATTTATTTGGTCTTGGCC  
 N C L T C A A N N P D A K A K A L L G H Q R I S G P W S K L Q M E F I F G C P L P 1517

4921 CAAACGCCCAAGGGAACAAGTACTGCCTGGTATCAGTGATAAATCCACCAAGTGGTGAAGCATTTCAGCTCGAAACACACAGCCAAACAACCTGCCAAGATCTTGGTAGAGCAT  
 Q T A Q G N K Y C L V I S D N F T K W V E A F P A R N N T A N T A K I L V E H 1557

5041 GTCTTCTCACAGCCTCCCAAAAGAGGTCAGCTCAGACCATGGCAACGCTTCGTTGGAGAAGTCTCAAAGGTGTGGCCAGTACTGGCCATTAAGCAAAAGCTTCATATCATG  
 V F S R W G I P K E V D S D H G Q R F V G E V S K G V C Q V L G I K Q K L H I M 1597

5161 GGGCATTCCCAAGCCTGGCTGTAGGGGGCCCAACCAACTGAAGATGGCACTGAAGAGATCATAAACCAAGGAAGGACTGGGATCAGAACTCCCACTTGTCTCTCTA  
 G H F H K P G L M E G P N Q T L K M A L K K I I N Q Q R K D W D Q K L P L V L L 1637

5281 GATTACGGGAGCTGGCTCAGCTACCTTCTCCCAAAATGGCTCTGCAAGAGCCAAAGATGCTGGAGCACTGGTGTGAGAGGGCCAGCCAGACAGCTACGAGCC  
 A L R G A V A S H V P S P P K M A P A R D P K M L E H W C Q R G A P P D E L Q P 1677

5401 CGCTGTGACAGACAGTGGTTCAGGACATCTGAGGACGCTGTCAATACATACACCAACTCGCCCTCAGTCAGGAGGCCAATATCCCAAGATGGATAGCAGTAGGATGGT  
 R V L T D R W V Q D M L R T V S D T Y H Q L A S V Q E A N I H K M D K Q L G V L 1717

5521 CTGCGCCCATGGAGTGAACACAGGGGACCTGGTATCATCTGGGGGTTAGGGAAGAAGTACAGAGCTGGGACCCAGTGGATGGGACCCGTTAGGTTGTGAACAAGCCAGCTCC  
 L R P M E W N T G D L V S Y R G V R E K N Q E L G P Q W M G P V R V N K A S S 1757

5641 TCCATGTACCAAGTAGAAATCAGAAAGGGGCGAAAAGACAGGAAATGGTCCATCTTCCACAGTAAAGACATGAAGGAAATTAACGGGGCTGGAGACCCCGCTTGTATATGT  
 S M T Q V E L R K G A K R Q E K W F L S Q L K A W K G N \* 1786

5761 TTGCAGTGAAGAAATGGAAGTGATTTGACAGCGGGAAAGCGAAGTCTGGCAGTACACCTCCCTATGTGTTCTCGAGGAAATGCTTCAAATTTGGTGTGACGATTTGC  
 M E G D C D Q R E R R T S W H H H P Y V F C R E N A F K F V V A G F A 35

5881 TTTGATTTCTGGTATGTCTATAACCTGAGCATGCAACATGCCAACAAACCCACAGCTCCAAAGAAAAGTCACTTCAACTTAGACACATACCCAGCAAAAGGGCCACAGCCCA  
 L I S G M S I T L S M Q H A Q Q T H Q L H Q E K V S S N L D T Y P S K R A T A Q 75

6001 GCACAGGTCAGGCGGTATAAAAATTCACAGATTTGGCTTGGTCTCAAGATAGTAAATATACAGGTAGTATGGGTTAAATTTCTAACAGAGGCTTAATTTCAACAGTAGT  
 H R V R A R Y V K I D T D W P W S Q A H V K Y T G S M G L N S N R G L N L S T V V 115

6121 AATGCTGGGACTGAAGTGTATTTGGAAAATGAGTGGAGCTGGGATAGCAAGCAAGCTCCCTGAGTGTGGGAAAGTGGGGCAAGAAATTAAGTAGAGTACAGGTAATTAATG  
 M H G T E V Y L E N E W S W D S T S R L P Q L L G K V G Q E I K V G C R V I N G 155

6241 TTCACCTCATCAGCAGTAACTCAAAATCTCCATACTGAATAAACCAGAAAGAGTGCAGAAAACAACTGGCCCTCGAAAACCTGGATTTGGTGCACCTTTACATTTCCAGCC  
 S T H Q Q V T Q C I S I T E I K T K K S Q K K N C A L E K L D C W C N F T L V Q P 195

6361 AGTCTTTGGTCTCTCTGGGCGCAACAGTGTGGGTTGCTTAAATTAAGATTTGACCATGATGCATTCATTTGCTGCCCTTAAGTTCGAAGTCCATTTCTGGCCGT  
 V F V V C L W A Q N S V G L S P F K F K I V T M M H S F A A L K V P R C H F L A W 235

6481 GGTGCGACCAATATGCTGAAGTGGGAACCAAGTAAATAGAAATAGTACTCTCAAGAAAGTGAATGATGCAAAAGCAAAATTAATGGCACCCTGTGACTCTCAGTCTCTAA  
 R A E A Q V E A G N Q V K L E I R Y S Q E S V T D A K Q I N G T P V T L T A P N 275

6601 CAGCCGTAAAGTGGATTTGTCAGTAAATGCCCGGAGAGCAAAATGCTTAGTAGATGATGAATAAGTGCAGAGACATCATGGTATAGCCAGGATTAAGTATCACTGTTCAAGTCCACT  
 S R K M I V S V N C P P E S R K M L S R S E L N A E T S W Y S Q D Y G Y C S S P L 315

6721 GAAAACCTCCAGGTGTGGTGTCAAGAAAGCTGAGTGTAGAAATGTCCTCCCTGAGCTGGAGAAAGTGGTGTAGGAGGCCCGCAAGGATTTCAAAAGAGTTTCCCACTGCTGT  
 K N L Q V W C Q G K L S V E M S P E L G G K W L I G G P E G F Q K E F P I A V 355

6841 CTGCGTCCCTTTGTTTCAAAAATAGGCCATATGATGTCAGCAAAATACATCAAGAGTCTGACAGCCCTGTGAGATCATTAAGAAAGTGGTGTCTCCCTCCACTGCCAA  
 L R P F V S K I G P Y V V K Q N H I Q E L L T S P V R S L K K V V L S L S T A N 395

6961 CATTTCATCATCAGACACATGCTCCCTTCTCCCTGACCTGCTGACCTGGCTGGCCCTGGCTCAGCCGCTCCAGCAGGAGCCCGCCAGGAGAGCTTCTGGCCAC  
 I S S I H R P H C A P F L S T L Y T G W L A W L H S R S M Q E A R T R R 435

7081 AGCCCTGGGAGGAGGTCTGGGCTGGGATGCTCAACAGCATGAACGCTGAGGTTCTGGCCAAAGTGGAGGCTGCTACTCGGGTGTGAGCCCTCTCAACCCCTTGAATTC  
 A L G G G G A G L G V L N S M N A E V L A N K L E A V T S G V Q G L L N P L N S 475

7201 ATCCCTGACAGCCTGGATGGGGCAGTGGCTGTCTGAGGTCTGCCTACGTGGGAACACATAGTGAAGAACACCAGGCTGCTGTTGCAAGCGTGGGAATCGAGCAAAATAA  
 S L T S L G M G Q W L V S E V L P T W E H I S E K D H Q V L L Q A L G I E Q N N 515

7321 TGTCTCCCTTGTCTTAGTTGATCCAGGCCAGATGTGGTGCAGAGTGTATTGACAGTGTCTGAGGATGGCCACAAGGATCCTCCCACTGAGATCCGAAAGATCGTGTGGGA  
 V S L A L S C I Q A Q M W V Q S V I A G I L R D G D N G I L P T E I R K I V W D 555

7441 CGCAGCCAGGAGAGGAGCCAGCTCAAGCTGGTGGAGGCTGGTGAATTCACCCACAGCAGGCTCTCAACTCAGTCTCCGCCAGCTCCTCAGCTGGTGGAGCCGACATGA  
 A A T E K E R Q L Q A W W R L V N F T H D Q V L N S V I A H V L T V V E A R I E 595

7561 AAAGCTTACCCCTGGCCCTGGGAGTAAACAATAAGTCTGTGGTCTATCCGCTGACCCAGGATGTGGCAAGAGTGTGAGACAGAAATGGCAATCTCTAGATTTGGAAGC  
 K V Y P I V A L G I N T N G S V V Y P L D H R M W A R V S D R K W Q S V D L E A 635

7681 CTGCATTTGGAACAGGGGCTGGGATTCATATGCAAGATGATGCCCTTAAGGCAAGTGTGCTTGTGATACAGCAGGAGGGGTGTGCCATTTTGAGATCAATCTCCGAGCAATAA  
 C I L E Q D L F I C E D L A L K A S D V C T F D T S E G V C H F E I N P R S N N 675

7801 CAAAACCTAGTATGTTACGTAGGAAAGGTTGCTGTCTTTAGAACATCTGTAATAACCTGCAAAATTAATGACATCTACAATCAGACCCGATTAATCAATGATTAATGATGCTGTCTG  
 K T M L V Y V Y G K C V C F R T M C K Y V Q I N D I Y N Q T V F N D S T C A C 715

7921 CAATGATGATTTATGAGGCTGTGATTTGTTGATAGCCACCACTTTTACTAGCCAAATTAATCAGAAACTATAGCCGTATGCTGATTAACCCGACCCCAATCGATGATGA  
 N V A I I R G C D F V Y K P P V F T S Q L I R N Y S L Y R S I T P T P I G M D 755

8041 TTTGCTGCTAGTAAAGAAATTTAGAGCATGCAAAATTTACAACAGCTTTAGAAACCGCTAAGCCAGAACTAAAGATGCTTAATTAATCTGATCATATGATGATGATTAATGA  
 L S L V K E M L E H A N L Q L L E N A K A E A K K I L I T V H H D G N V I K Q 795

8161 GATAGTGGARCGGATTAAGAGGTTGGGAGCACCCTGGTGGAAATTTCTTTGGCTGGTCCCAACAGCGACCGGCATCTTCAACCGCTCTGCATCTCTGATCACTACTGCT  
 I V E R I A R V G E H H W W E I F F G W S P T A T G I F N A L L H P V V I I L L 835

8281 AATGCAAAATGGTGGTCTTTCGATGTTGCTACTTGTACTGATAGCCAGTGGAGTCTGCTGATGGAACCACTGAAAGGACTGGGATGGCCAAAGCCCTCTACCATGATC  
 M Q M C V C F A M V A T C Y W I R Q V R L C I E N Q L K G L G M A K R L L P \* 873

8401 AAGGCAAGACTGGTGGCTCTGTTGTCACCAAGAAGAACCTTTAGCTCCGCTGTGGCTGCAACCCAGTGGCTGTGAGCGTGGGGACAACAAGCCATACAGCTTGTATCGG  
 AAGGGGCTGCCTCTGATCAAGAGACCATCCAGGAGCAGAGGCGCGTGCAGGCTGCAGGTGGTGAACCAAGAAATGGTGGTGGAAAGGGCCCTCTGAGTACTCTG  
 CCAATTTGCTATTCAGGCGAGTACTAGAGCAGATATCTGGGACTGTATCCAGCTGGGTTTGGCTCTCCCTGAGTAGAGACTCCCAACTCTCTGGCAACCTGTTCCAGT  
 CTTGACTACTCCCAAGTAAAGTAAAGTAAAGTAACTCTCTCTGAAATCTCCTCCCAAAATCAATGACAAAAAGCTGCTTTTGTGTGATTTAAGACTCTTAAGCAGAGGAGAA  
 GGTTAGGAAAGCCAGGACAGTCTCTTGAAGGCTCAGAGGCAATGTAACACCTGAGTGCAGAAAAGCAGACACTGAAGATGGCAGAGATGTTATCAGCTCAGAGACTTGGAAAG  
 CTGTAGAGGAGCAGAACAAATGATCTGTGATCACTTGGACTTAAAGGAGCTGAATTCACATGGAACCTTTGCAGCAACCTCTCCCTGGGTTCCCAAGACTCCCAAGAGCTTGGCCAG  
 GGATTTGTTGATCTTCTTCTTGGAGCTGGACATCACTGCTTCTGCTTGTAGTGTGATGATGAACCAATAATCTGTTTGTAGAGGAT 9213

Integrase

CR1-Y4

Polyadenylation signal

**Figure 3**  
**Complete nucleotide sequence of chicken *Ovex1* and predicted translation products (second part).** Nucleotides are numbered on the left, from the most 5' cap site of the mRNA as position +1. The genomic sequence preceding the start of transcription is taken from the chicken genome draft assembly (galGal3). The 5' and 3' imperfect direct repeats are indicated by dashed underlines. Sequences identified as GGLTR11 and CR1-Y4 by RepeatMasker are on a gray background. The TATA box and polyadenylation signals (AATAAA and ATAAAA) are underlined. SD and SA indicate the splice donor and acceptor sites. The conceptual translation of *Gag*, *Pol* and *ORF3* is given under the nucleotide sequence, and amino-acid positions indicated on the right. Putative translation start codons are boxed. Conserved motifs are underlined: the *Gag* nuclear localization signals (NLS), leucine zipper motif and major homology region (MHR). In *Pro*, the active site is underlined. In *RT*, position of the usually conserved aspartate residues is indicated in bold underlined characters. In the integrase, amino acids involved in the zinc-binding finger are in bold characters and residues of the degenerated DD35E motif bold and underlined. In the *ORF3* protein, potential N-glycosylation sites are in bold letters, potential cleavage sites indicated by arrows, and the transmembrane domain underlined. In the 3' UTR, the 13-base polypurine tract is doubly underlined. The position of the SSH cDNA fragment is indicated by a double arrow and its *RsaI* terminal sites (GTAC) printed in bold characters.

bp and that of the spliced transcript 3543 bp (not taking into account the polyA tail). No other splicing was detected by RT-PCR amplification using various pairs of primers.

Surprisingly, in addition to the final polyadenylation signal, *Ovex1* contains one AAUAAA and three AUUAAA hexamers followed by U and GU-rich elements, clustered in the region of nucleotides 6082 to 6669, which might constitute polyadenylation signals leading to premature transcription arrest. However, the efficiency of these signals *in vivo* is low as demonstrated by the RT-PCR amplification of cDNA fragments D, E, and F (Fig. 1C) and the existence of ESTs overcoming the signals (Fig. 1B). To verify if polyadenylation at these internal signals would affect differently the spliced and unspliced mRNAs we performed RT-PCRs using sense primers located upstream from the polyadenylation signals, either into the intron to amplify the unspliced mRNA, or across the splice site to amplify the spliced product. Several antisense primers, upstream or downstream from the four polyadenylation signals, were tested in both cases. The amplification of cDNAs overrunning the signals indicates that at least a substantial fraction of both types of mRNAs is transcribed unto the final polyadenylation site (additional file 1: Figure S2, Effect of the internal polyadenylation signals).

The complete sequence of chicken *Ovex1*, obtained by sequencing the cDNA fragments and the cloned RACE products, is given in Figs. 2 and 3. It is 99.9% identical to a region of the chicken genome draft assembly version v2.1 (galGal3, chr. 4, + strand, 89,102,054 to 89,111,266). Differences between these two sequences determined on different strains of *Gallus gallus*, the wild type Red Jungle fowl for the DNA and the domestic White Leghorn strain in our study, are exclusively single nucleotide substitutions.

Blat search in the recently sequenced genome of a passerine bird *Taeniopygia guttata* (zebra finch) revealed the presence of a sequence with 83% identity to chicken *Ovex1* (taeGut1, chr. 4, + strand, 65,939,815 to 65,948,592). The sequence is incomplete, due to the presence of a sequencing gap of some 600 bp in this first version of the genome. This *Ovex1* homolog is located on zebra finch chromosome 4, syntenic to chicken chromosome 4 long arm, between the *CD8* and *SMYD1* loci as for chicken *Ovex1*. Such conservation indicates that this sequence is the ortholog of chicken *Ovex1*. The sequence of zebra finch *Ovex1* is given in additional file 2 (Figure S3, Partial zebra finch *Ovex1* sequence). Conservation of the TATA box, the polyadenylation signal, and of sequences surrounding the donor and acceptor splice sites suggests that zebra finch *Ovex1* is transcribed and spliced as chicken *Ovex1*.

The first two thirds of the *Ovex1* sequence appear to be present as a single copy in chicken (galGal3) and zebra finch (taeGut1) genomes. In contrast, in both species the last third of *Ovex1* contains sequences identified by RepeatMasker as the internal part of a multi-copy LTR-containing element, GGLTR11-int [36]. In addition, an imperfect antisense 3'-fragment of CR1-Y4, member of the large family of chicken CR1 non-LTR retrotransposons (LINE) [31,38], is included in the 3'-UTRs. These sequences are indicated in Fig. 3 and in additional file 2 (Figure S3, Partial zebra finch *Ovex1* sequence).

#### Repeated sequences

Endogenous retroviruses are usually bordered by two LTRs classically divided in U3-R-U5 regions. The 5' U3 region contains promoter and enhancer elements. Transcription is initiated at the 5' U3-R junction and terminates at the 3' R-U5 junction. We looked for repeated sequences in the chicken *Ovex1* gene and flanking sequences. Direct repeats of 120–122-bp were found on each side of the gene, from nucleotide -6 to +114, 17-bp after the TATA box, on the 5'-side, and from nucleotide 9023 to 9144, 46-bp before the polyadenylation signal, on the 3'-side (Figs. 2 and 3). These repeats, which have only 73% identity, might be degenerated forms of the R region of previous LTRs. Identity of the zebra finch repeats is even lower (69%). No other repeated sequences corresponding to the U3 and U5 regions were found in the genomic sequence in and around the gene. No sequence complementary to the 3'-end of a chicken tRNA [39] that might be the primer tRNA binding site (PBS) required for reverse transcription was identified, thus precluding classification of *Ovex1* according to this usual criterion. The 3'-UTR contains a 13-purine sequence that may correspond to the polypurine tract that precedes the 3'-U3 region in retroviruses (Fig. 3). Conservation of the 5'-UTRs between chicken and zebra finch is globally low (65%), but the sequence surrounding the splice donor site is well conserved. By contrast, the 3'-UTRs are very similar (80% identity score).

Chicken and zebra finch *Ovex1* 5'-proximal DNA sequences were screened to detect transcription factor binding sites, using the MatInspector program. Among putative responsive elements conserved with a similar position in the 5'-flanking regions, we identified a TATA box and sites for Ras-responsive element binding proteins, forkhead domain factors, estrogen-related receptors, GATA binding factors, and bicoid-like homeodomain transcription factors, in particular Pitx2 [40]. These sites are shown in additional file 3 (Figure S4, Alignment of chicken and zebra finch 5'-proximal DNA sequences).

### Open reading frames

As seen in Fig. 1E, the unspliced chicken mRNA contains three large uninterrupted open reading frames (ORF). The first two ORFs are contiguous in frame and separated by a stop-codon. The third one is in a different frame and non-overlapping. The spliced mRNA contains only the third ORF (Fig. 1F). The conceptual translation is given in Figs. 2 and 3 for chicken and in additional file 2 for zebra finch (Figure S3, Partial zebra finch *Ovex1* sequence). Protein similarity queries using Blastp have shown that the unspliced mRNA potentially encodes Gag and Pro-Pol polyproteins and that the spliced mRNA encodes a protein that might be an envelope protein. Chicken and zebra finch *Ovex1* display the same organization, which is typical of numerous retroviruses. Alignment of these proteins is shown in additional file 4 (Figure S5, Alignment of *Ovex1* putative proteins).

### Gag polyprotein

The *Gag* initiation codon is usually the first AUG after the cap site. In the chicken unspliced mRNA, the first AUG, in position 313, initiates potentially a 5-amino-acid hydrophobic peptide. Initiation may also occur at the next AUG downstream, by reinitiation or leaky scanning [41]. The next AUG is at position 370, in a correct context for initiation of the translation (CUCACUAUGAUGU). If translation starts effectively at this codon, the ORF would encode a 565-residue protein till the first stop-codon (nucleotide position 2065 to 2067) (Fig. 2). However, a second AUG, immediately following this initiation codon in a seemingly less favorable context, has been chosen as initiator for the hypothetical protein encoded by LOC422925. The putative protein was analyzed using Blastp. Gag polyproteins are constituted by the matrix, the capsid and the nucleocapsid domains. The C-terminal region (from the amino-acid residue 248) has 22% identity with the capsid domain of an epsilon-like retrovirus, Snakehead fish retrovirus (SnRV) [42]. The *Ovex1* protein contains a sequence EKQNPEERPDAYVNRKKMLY corresponding to the major homology region (MHR) with the three underlined fundamental residues [43]. This region, which provides the interface for capsid binding and dimerization, is present in all retroviruses with the exception of the spumaviruses. No zinc finger domain was identified here, unlike in SnRV and most retroviruses except the spumaviruses. In the N-terminal region, no sequence similarity was detected with other Gag proteins, even with that of SnRV, but they do have some homology. As in SnRV, the protein N-terminus is presumably not myristylated, and the N-terminal region contains a polybasic sequence (residues 50 to 61) with two successive consensus nuclear localization signals, KKPKPTNKCKKR [44]. As in SnRV, the protein has a region of strong probability of coiled-coil structure [45], between residues 62 and 114, like in the rod domain of myosin-type proteins. In addition, the

*Ovex1* protein contains a L(X)<sub>6</sub>L(X)<sub>6</sub>L(X)<sub>6</sub>L leucine zipper motif (residues 88–109) seen in numerous regulatory proteins. Thus, this protein may be classified as a Gag polyprotein by analogy with that of the SnRV and from the position of its coding sequence preceding a *Pol* sequence, but its functionality remains to be demonstrated. The uninterrupted zebra finch Gag polyprotein is well conserved, with 90% sequence identity to chicken Gag, and contains the same functional domains (see additional file 4 (Figure S5, Alignment of *Ovex1* putative proteins)).

### Pro-Pol polyprotein

The second long ORF immediately follows the *Gag* amber codon in the same translational frame and is presumably translated by suppression of this codon as seen for SnRV, gamma- and epsilon-retroviruses [46]. Translation results in a 1786-amino-acid Gag-Pro-Pol polyprotein unto the next stop-codon. Analysis of the conceptual translation of this ORF allows the identification of retroviral aspartyl protease (Pro) and Pol polyprotein domains: reverse transcriptase (RT), *Rnase H*, and integrase (IN) (Figs. 2 and 3). In zebra finch, the *Pol* region is only partially sequenced and most of the *Rnase H* domain is unknown. Chicken and zebra finch Pol sequences display 81% identity in their common regions and the same characteristics (see additional file 4 (Figure S5, Alignment of *Ovex1* putative proteins)).

The aspartyl protease domain, approximately between residues 583 and 686, contains the active site MLIDT-CASYSIL and presents 29 to 26% identity with the proteases of *Sphenodon* endogenous virus SpeV [47], immunodeficiency viruses and porcine endogenous retrovirus (PERV). As for SpeV, we note the absence of the conserved GRD/N motif present in all retroviral proteases, except for spumaviruses [48].

The RT domain, from residue 767 to 965, displays a similarity with RTs of foamy viruses, fish viruses: Walleye epidermal hyperplasia viruses WEHV 1 and 2 [49], Walleye dermal sarcoma virus (WDSV), SnRV (33–30% identity), and with the partially sequenced reptilian SpeV. However in chicken as in zebra finch *Ovex1*, several fundamental RT residues are not conserved, in particular two of the three aspartates that make up the catalytic active site [50], a difference resulting presumably in the loss of the enzyme activity.

The chicken *Rnase H* core domain, between amino-acid residues 1211 and 1357, is similar to those of SnRV (35% identity) and Moloney murine leukemia virus, MMLV (32%) with, in particular, the conserved WFVDGSN and FSDS motifs [51]. The distance between this sequence and the RT domain is consistent with the presence of the tether



region that separates RT and Rnase H domains in vertebrate retroviral *Pol* genes [52].

The integrase (INT) has a similarity with those of MMLV, WEHV2, foamy viruses and SnRV. It contains a putative zinc finger: H(X)<sub>3</sub>H(X)<sub>29</sub>C(X)<sub>2</sub>C, as most retroviral integrases. The core domain, *rve*, between residues 1500 and 1647, is only partially conserved. Where other integrases have the catalytic triad D(X)<sub>55-60</sub>D(X)<sub>35</sub>E, the motif is E(X)<sub>59</sub>D(X)<sub>35</sub>E in chicken and E(X)<sub>59</sub>E(X)<sub>35</sub>E in zebra finch *Ovex1*. The difference is important since substitution of the first aspartate by a glutamate drastically impairs the integrase activity of Rous sarcoma and HIV viruses [53]. The C-terminus does not contain the consensus GPY/F motif but a WMGPVVRV sequence that might be the degenerated remnant of this motif [35].

#### *ORF3: an envelope protein?*

The *Ovex1* third ORF (*ORF3*) is located downstream from *Gag-Pol*. It is entirely contained in the second exon of the singly spliced transcript (Fig. 1F). A similar organization occurs for the envelope protein of gammaretroviruses. The ORF starts 11 bp after the splice site and encodes a putative protein of 873 amino acids. This ORF is potentially interrupted by the presence of the internal polyadenylation signals that do not appear very efficient *in vivo*, as demonstrated above. In zebra finch, the third ORF encodes an 874-amino-acid putative protein with 81% identity with the chicken protein and similar characteristics (see additional file 4 (Figure S5, Alignment of *Ovex1* putative proteins)).

This region is quite complex. Segments of sequence on gray background in Fig. 3 were identified by RepeatMasker as weakly similar to GGLTR11-int, a consensus sequence quoted in Rebase as internal part of an ERV1 chicken endogenous retrovirus [36]. Nucleotide conservation with GGLTR11 is rather low (60%) and better-conserved copies of ERV1 can be found in the chicken genome. Furthermore, the first two GGLTR11-like sequences are separated by 30 nucleotides in GGLTR11 while located 960 nucleotides apart in *Ovex1*. The intervening region, not identified by RepeatMasker, is locally similar to sequences found in chromosome Z. A similar presence of GGLTR11-related sequences is detected in zebra finch *ORF3*, as seen in additional file 2 (Figure S3, Partial zebra finch *Ovex1* sequence). The conceptual translation of GGLTR11 is given in additional file 5 (Figure S6, Structure and conservation of *ORF3*-encoded proteins). Unlike *Ovex1 ORF3*, it contains stop-codons and frameshift mutations.

Significant similarity of the *ORF3* putative protein with other proteins was found only in an avian retrovirus, the Tetraonine endogenous retrovirus (TERV). This defective ERV of the ruffed grouse, *Bonasa umbellus*, a member of

the galliform order as chicken, contains 5' and 3' LTRs, a Gag-like protein and a truncated region considered as an envelope TM domain [54]. The similarity between *Ovex1* and TERV is limited to this envelope domain, with 52% identity over a 92-amino-acid region (residues 754 to 845) that can be extended upstream to amino acid 645, with stop-codons and frameshifts in the TERV sequence. This TERV region is closer to the consensus GGLTR11 than is the *Ovex1* sequence. The main arguments sustaining the identity of the TERV envelope are its position just before the 3' LTR and some similarity with envelope proteins of MMLV and PERV. Identity of *ORF3* with these retroviral envelopes is even more limited.

The protein potentially encoded by *ORF3* is larger than prototypical retroviral envelope proteins. It contains two hydrophobic regions shown in additional file 5 (Figure S6, Structure and conservation of *ORF3*-encoded proteins). One is close to the N-terminus, and the second, located in the C-terminal part, corresponds to the identified transmembrane anchor and is preceded by a sequence susceptible to adopting a coiled-coil structure (residues 753 to 780). The cytoplasmic domain, C-terminal to the TM region is short. Retroviral envelopes are glycosylated. Four conserved potential N-glycosylation sites are present in the sequence. However, the protein does not seem to possess a signal peptide. Retroviral envelope precursors are cleaved between the surface and transmembrane domains at specific RX(R/K)R sites. The proteins encoded by *ORF3* in chick and zebra finch contain conserved potential cleavage sites. However, neither these proteins nor TERV's contain the conserved CWLC sequence [55], the consensus immunosuppressive motif CKS17, and the disulfide bonded loop C(X)<sub>6</sub>CC that precede the TM region in many retroviruses [43,56]. Thus, despite its position and its homology with the TERV predicted envelope, the identity of the *ORF3* product as an envelope protein is still uncertain, given its limited similarity with other retroviral envelope proteins.

#### **Single nucleotide polymorphisms (SNPs)**

Differences between the sequence of chicken *Ovex1* transcripts and the genomic galGal3 draft assembly are exclusively point substitutions reported in additional file 6 (Table S7, Polymorphisms in chicken *Ovex1* sequences). Among 14 substitutions in *Gag*, *Pol* and *ORF3* coding sequences, 12 are silent. These differences reflect the polymorphism between two types of *Gallus gallus* derived from a common ancestor: the wild Red Jungle fowl and the domestic White Leghorn strain. In addition, direct sequencing of RT-PCR products from pools of gonads allowed the detection of the presence of two nucleotides in variable proportions at some positions of the sequence. This heterogeneity is presumably due to polymorphism within the Leghorn chicken population. Some of these

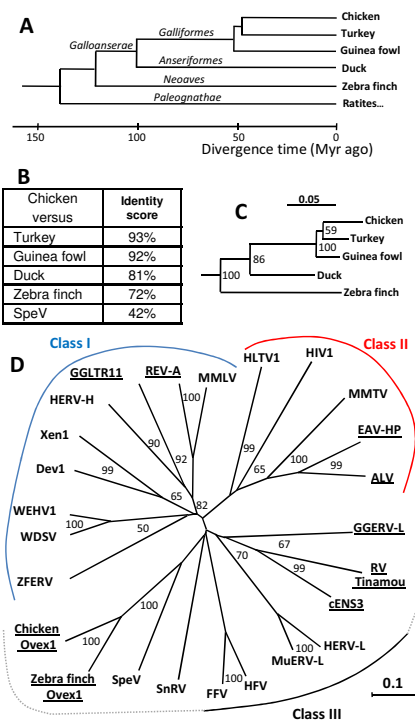
differences are also found in ESTs. In the sequence coding for Gag-Pol, these SNPs correspond to silent differences. In ORF3, conservative and non-conservative substitutions are observed.

#### Presence of *Ovex1*-related sequences in other birds

To investigate the presence of sequences related to *Ovex1* in the genome of other birds, we performed PCR amplifications with primers corresponding to *Ovex1* Gag and RT regions, using DNA from turkey (*Meleagris gallopavo*), guinea fowl (*Numida meleagris*) and duck (*Anas platyrhynchos*). Direct sequencing of the fragments gave unique sequences, corresponding to ORFs highly similar to those obtained from chicken. The 132-bp-long Gag fragment, which contains the nuclear localization signals, has no analog in nucleic-acid and protein databases. Conservation for turkey, guinea fowl and duck with respect to chicken is 98%, 96% and 92% at the nucleotide level and 100%, 100% and 98% at the protein level. The 400-bp-long RT fragments present 94%, 92% and 84% of nucleotide conservation. Although it is not proven that the amplified Gag and Pol sequences are linked in the DNA of turkey, guinea fowl and duck, the result suggests that *Ovex1* orthologs exist in these birds. Alignment of the protein sequences is shown in additional file 4 (Figure S5, Alignment of *Ovex1* protein sequences). Amino-acid identity scores for the RT fragments range from 93% between chicken and turkey to 76% between chicken and zebra finch (Fig. 4B). By comparison, identity with the closest known retrovirus, SpeV, is only 42%. The neighbor-joining tree based on the alignment of the five *Ovex1* RT sequences (Fig. 4C) follows the bird phylogenetic relationship shown in Fig. 4A.

#### Comparison with other retroviral elements

To classify *Ovex1* amongst retroviral sequences according to the criteria defined by Jern [35], we may recall general characteristics. (i) The Gag-Pro-Pol coding sequences are in the same frame and translation of the Pro-Pol polyprotein results likely from the translational suppression of the Gag stop-codon, as in gamma-, epsilon- and intermediate epsilon-like retroviruses. (ii) The putative Gag protein contains no zinc finger, as in spumaviruses and spuma-related HERV-L and MuERV-L [43,57]. This is in contrast to the SnRV, which displays some similarity with *Ovex1* Gag but has one zinc finger. (iii) No dUTPase domain was detected, unlike in MuERV-L. The absence of the integrase GPY/F motif is not discriminating as for spuma-like viruses, since a degenerated sequence may be present. A single splicing event and no obvious accessory ORFs were found in *Ovex1*, unlike in complex retroviruses like SnRV, WEHV and spumaviruses. In the analysis, it is necessary to distinguish the Gag-Pol and the ORF3 regions of *Ovex1*.



**Figure 4**

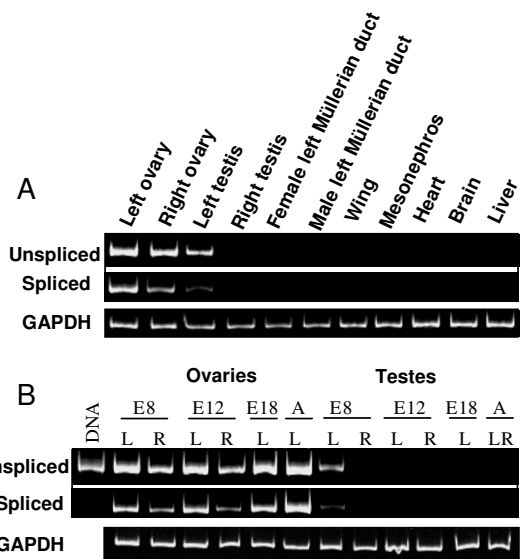
**RT-based phylogenetic analyses.** (A) Simplified phylogenetic relationship of modern bird lineage. Bayesian estimates of the divergence times (in million years, Myr) are based on mitochondrial gene data [68,91]. (B) Protein identity scores of avian *Ovex1* RT fragments. The 133-amino-acid sequences of the RT fragments obtained by PCR from turkey, guinea fowl and duck were aligned with the homologous regions of chicken and zebra finch (Gag-Pol residues 768 to 900). The corresponding sequence of SpeV was added for comparison. The alignment, performed with ClustalW2, is shown in additional file 7 (Figure S8, Alignment of *Ovex1* RT and homologous sequences). (C) Relationship of the five *Ovex1* RT-fragment sequences. The phylogenetic tree and bootstrap values (1,000 iterations) were calculated by the neighbor-joining (NJ) method [85] using the QuickTree software and the tree drawn with NJplot. Branch lengths are proportional to substitution levels. The scale bar represents 5% divergence. Bootstrap percentages are indicated at the branch forks. (D) RT-based unrooted dendrogram of *Ovex1* and representative retroviral elements. The alignment, corresponding to chicken *Ovex1* Gag-Pol amino-acid residues 759 to 917, was performed with ClustalW2 with default settings and adjusted manually. It is given in additional file 7 (Figure S8, Alignment of *Ovex1* RT and homologous sequences). The relationship and bootstrap values (1,000 iterations) were calculated by the neighbor-joining method with QuickTree and the unrooted dendrogram drawn with NJplot. Abbreviations and database accession numbers are given in Materials and methods. RTs of avian origin are underlined. Branch lengths are proportional to substitution levels. The scale bar represents 10% divergence. Bootstrap percentages are indicated at the branch forks when at least equal to 50%.

RT is the most conserved retroviral domain generally used for phylogenetic analysis, allowing detection of the relationship between distant elements [58]. We performed the alignment of a 159-amino-acid sequence of chicken and zebra finch putative RTs (Gag-Pol polyprotein residues 759 to 917) to that of representative retroviral elements and retroviruses, using ClustalW2 (see Additional file 7 (Figure S8, Alignment of *Ovex1* RT and homologous sequences)). The derived neighbor-joining unrooted dendrogram presented in Fig. 4D displays three groups of sequences. They correspond to the class I and class II of retroviral elements and to a third group more dispersed (similar to group I in [59]) that contains class III elements and the intermediate epsilon-like retrovirus SnRV [34,35]. *Ovex1* RT is not closely related to any known avian retroviral sequence. On the basis of this analysis, *Ovex1* does not belong to class II. Despite some similarity with the epsilon retroviruses WEHV and WDSV, it is not included either in class I elements. It appears more related to the third group of sequences. However, its relation with members of this group is not well defined. *Ovex1* does not cluster with spumaviruses or spuma-related elements MuERV-L and HERV-L. It is not either similar to the Repbase GGERV-L element [36] or to avian sequences like ENS-3 (Chicken Embryonic Normal Stem cell gene 3) [60] and the bird Tinamou retrovirus [59]. *Ovex1*'s closest relative is the *Sphenodon* endogenous virus, SpeV, found in an archaic reptile [47]. So far, only the *Pro* and *RT* domains of this endogenous virus are known, and they are the most similar to *Ovex1*. *Ovex1* and SpeV constitute a distinct branch of the RT-based phylogenetic tree, close to the branch point of SnRV and spumaviruses. However, *Ovex1* and SpeV are distantly related since their RT identity score is only 42%, which is not higher than between some members of different classes.

The second region of chicken and zebra finch *Ovex1*, corresponding to the third ORF, is partially related to GGLTR11, a class I ERV, which is not the case for *Pol*. Similarity with *Bonasa* TERV is limited to the transmembrane domain of the putative TERV's envelope. TERV is a defective endogenous retrovirus devoid of *Pol* that has been classified as an alpharetrovirus on the basis of its LTRs and *Gag* region but, according to the authors, the truncated *Env* might have a different origin [54]. Similarly in *Ovex1*, *ORF3* and *Gag-Pol* might originate from different retroviruses.

#### Analysis of chicken *Ovex1* expression by RT-PCR

Semi-quantitative RT-PCR amplification of *Ovex1* transcripts in various 8-day chicken embryo tissues shows that the unspliced *Gag-Pol* mRNA and the spliced *ORF3*-containing transcript are expressed in a similar manner (Fig. 5A). Expression is higher in the left ovary than in the right one, lower in the left testis, and absent in the right testis.



**Figure 5**

**Semi-quantitative RT-PCR analysis of *Ovex1* expression in chicken tissues.** (A) Expression of *Ovex1* in 8-day embryo tissues. RT-PCR was performed with pairs of primers designed to amplify specifically the unspliced mRNA (*Gag-Pol*) or the spliced transcript (*ORF3*). Amplification of *GAPDH* transcripts was used as control for mRNA amounts. Primers and RT-PCR conditions are given in additional file 9 (Table S1, Primers and PCR conditions). A similar pattern of expression is observed in both cases. The level of transcripts is higher in the left ovary than in the right one and in the left testis. Expression is null in the right testis and in the other tissues tested, but for occasional traces in the wing. (B) Expression of *Ovex1* in chicken gonads, as a function of age. The unspliced and spliced mRNAs were amplified from 8- and 12-day embryo left (L) or right (R) gonads, from 18-day embryo left gonads, and from adult (7 weeks) left ovaries or mixed (LR) testes. The identity of the RT-PCR amplified product was verified by sequencing. A control with chicken genomic DNA was performed in order to verify the absence of the deleted form in the genome and the specificity of amplification. The level of both types of mRNAs is up-regulated in the left ovary, weak in the right one, down-regulated after 8 days in the left testis, and null in the right one.

Amplification is negative for the other tissues investigated (mesonephros, heart, brain, liver), but for traces in the wings. In female embryos, expression of both types of transcripts is asymmetrical at 8 and 12 days (Fig. 5B). The highest expression is found in the adult (left) ovary. The expression observed in the left testis at 8 days is down-regulated at 12 days and after, and the right testis remains negative.

#### In situ hybridization

Expression of *Ovex1* in chickens was examined by *in situ* hybridization using a probe corresponding to the *Pol*

region and compared with that of other genes expressed in the gonads. In both sexes, the region of the presumptive gonad can be first identified at four days of incubation (E4) (Hamburger and Hamilton stage 24, HH24) by the expression of the Lim homeobox gene, *Lhx9*, in a restricted area of the mesonephros coelomic epithelium [4,61]. At this stage, neither the transcripts of *Ovex1* nor those of the estrogen receptor alpha *ER-α* are detected in this region (not shown).

At E5 (HH27) (Fig. 6A), male and female gonads, morphologically indistinguishable, are protrusions at the surface of the mesonephroi. They comprise two territories: the outer epithelial area or "cortex", negative for fibronectin, and the inner region, or "medulla", containing irregular groups of cells separated by strands of fibronectin-positive material [4,7]. The two gonads are not identical: the left one is larger and has a thicker cortex. The pattern of expression of the studied markers is the same for male and female embryos. *Ovex1* starts to be transcribed with a L/R asymmetry. Transcripts are present in the apical region of the cortex of left gonads, whereas they are not detected in right ones or in the mesonephros and surrounding tissues. *Lhx9* is transcribed in the totality of the cortex of both gonads, and in part of the dorsal mesentery epithelium. At the same time, *ER-α* starts to be expressed asymmetrically in the cortex of left gonads and symmetrically in the medulla of the both, as reported [14,15].

At E6 (HH28) (Fig. 6B) in both sexes, the cortex is thicker in left gonads. In the female (ZW), *Ovex1* is now detected in both gonads, but with a very dissimilar distribution. In the left ovary, it is expressed in the columnar cortical cells of the medioventral region, whereas in the right gonad a few *Ovex1*-expressing cells are scattered throughout the medulla. In the male (ZZ), expression of *Ovex1* is visible in the medioventral region of the left testis cortex, whereas no expression is detected in the right gonad.

At E7.5 (HH31), the morphological L/R asymmetry of the female gonads is even more evident. In the left ovary (Fig. 6C), the thickening cortex, constituted of multiplying oogonia and somatic cells, is bordered by a single epithelial-cell layer. The medulla contains loosely connected cords of epithelial cells. *Ovex1* is highly expressed in the cortical region including the surface epithelial cells, and also to a lower extent in the medullar cords. Likewise *Lhx9* and *ER-α* are expressed in the cortex and more faintly in the medulla, but the patterns are not strictly identical. *Ovex1* and *ER-α* transcripts are not detected at the lateral ends of the gonad, where the thickness of the cortex diminishes. Oogonia, identified by the chicken vasa homolog, *Cvh*, a germ cell-specific factor of the DEAD-box family [62], are mostly located in the depth of the cortex, although rarely some are dispersed in the medulla. If a

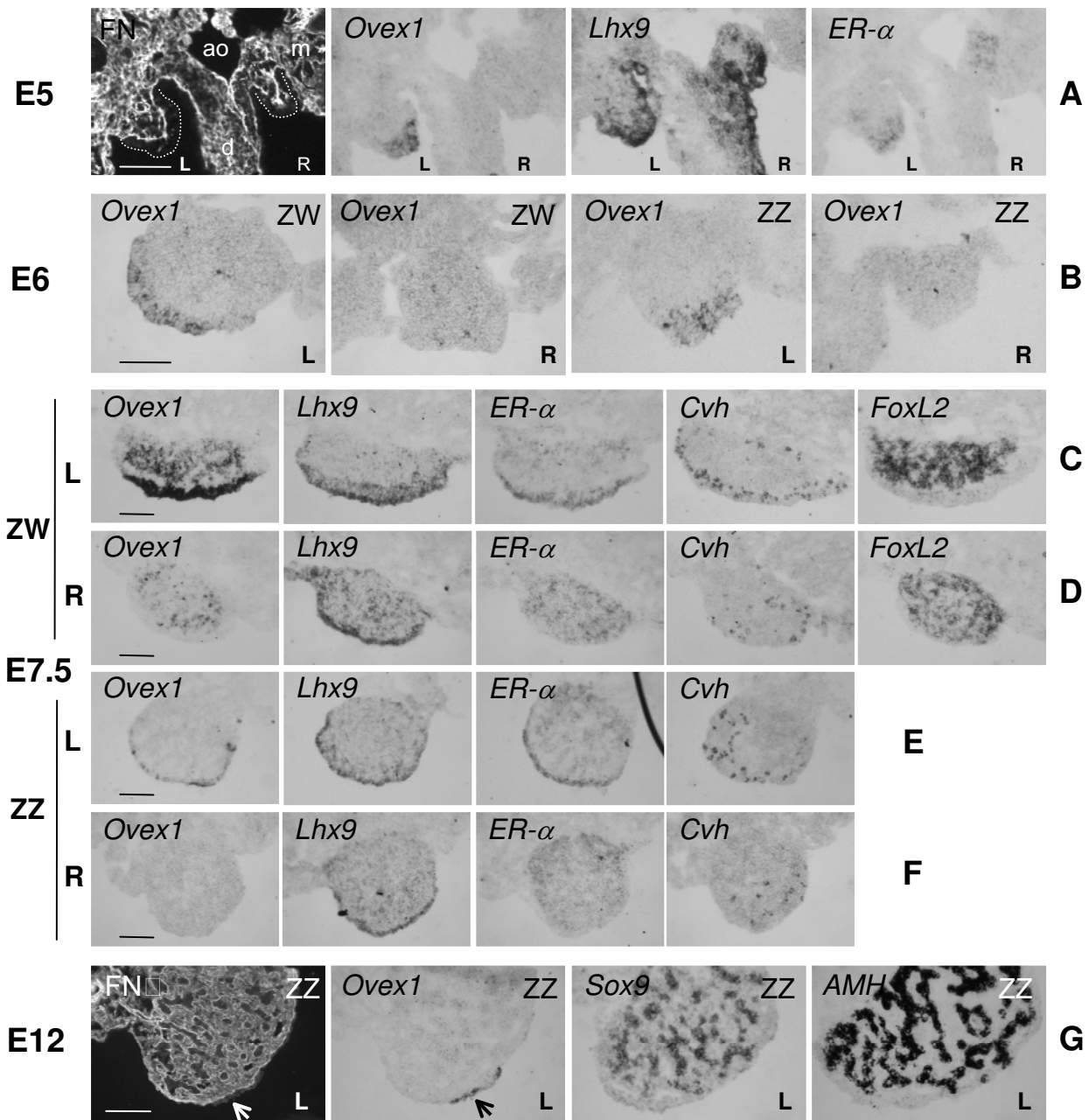
germ-cell expression of *Ovex1* cannot be fully excluded, it is clear at this stage that most of the cells that express *Ovex1* in the left gonad are somatic cells and not germ cells. *FoxL2*, the female-specific forkhead transcription factor [28,63], is transcribed exclusively in the medulla in cordonal cells.

In the smaller right ovary (Fig. 6D), the cortex has not undergone the same development and the medulla represents the major part of the gonad. *Ovex1* transcripts are completely absent from the cortical region but are present in some dispersed medullar cells. Similarly, *ER-α* is only expressed in the medulla as reported [9,10]. *Lhx9* is expressed in the thin cortical region and in dispersed medullar cells, whereas *FoxL2* is expressed in patches in the major part of the gonad but absent from the surface region. A few germ cells are present in the gonad.

At this stage, the morphological L/R asymmetry of the gonads is less obvious in males. Testicular differentiation becomes morphologically visible in both gonads. Sertoli cells, that start to express *Sox9* and a high level of *AMH*, are clustering to form the sex cords [7]. A weak expression of *Ovex1* is observed in the thin surface epithelium of the left testis (Fig. 6E), but none in the right one (Fig. 6F). No expression is detected in the medullas. *Lhx9*, by contrast, is expressed in the cortex of both gonads (Figs. 6E and 6F). Expression of *ER-α* is asymmetrical. Only the left testis displays a cortical expression whereas a faint symmetrical presence of transcripts is detected in both medullas (compare Figs. 6E and 6F), as reported [14,15]. Germ cells, more numerous in the left testis than in the right one, are dispersed in the gonads (Figs. 6E and 6F). At this stage, germ cells display a mostly peripheral distribution, a rather puzzling situation because Sertoli cells are clustering in the central part of the medulla to form the sex cords in which these germ cells are to be enclosed, as previously observed [4].

In the male, *Ovex1* expression remains limited. At E12 (HH38), left and right testes appear morphologically rather similar. In both gonads, testis cords have formed into the medulla, containing germ cells and supporting Sertoli cells surrounded by a basal membrane. The male-specific marker *Sox9* identifies the Sertoli cells that express also *AMH*. In the left testis, *Ovex1* transcription is restricted to a small apical zone where the epithelium is thicker (Fig. 6G). It is still absent in the right testis (not shown). At E18, *Ovex1* transcripts are no longer detected in the left testis (not shown). Disappearance of *Ovex1* transcription in the left testis is coincident with that of *Pitx2c* [14] and *ER-α* [10].

At E14 (Fig. 7A), the well-developed cortex of the left ovary, overlaid by a single epithelial-cell layer, is com-

**Figure 6****Expression profiles of *Ovex1* and gonad markers in embryonic gonads before and after sex determination.**

Fibronectin (FN) is revealed by immunofluorescence and gene expression detected by *in situ* hybridization of digoxigenin-labeled RNA probes on transverse cryostat sections. **(A)** Immunostaining of fibronectin (FN) and expression of *Ovex1*, *Lhx9* and *ER-α* in serial sections of a chick embryo (female) at embryonic day 5 (E5), (Hamburger and Hamilton stage HH27). Male embryos gave similar results. Gonads are outlined by a dashed line on the fibronectin image that corresponds to a double labeling of the section hybridized with *Ovex1*. m, mesonephros; d, dorsal mesentery; ao, aorta. **(B)** *Ovex1* expression in left (L) and right (R), female (ZW) and male (ZZ) gonads at E6 (HH28). **C-F:** Expression of *Ovex1*, *Lhx9*, *ER-α*, *Cvh* and *FoxL2* at E7.5 (HH32). **(C)** Left ovary. **(D)** Right female gonad. **(E)** Left testis. **(F)** Right testis. *FoxL2* was not detected in testes. **(G)** Immunostaining of fibronectin (FN) and expression of *Ovex1*, *Sox9* and *AMH* in E12 (HH38) left testis. Arrows indicate the epithelial region expressing *Ovex1*. Scale bars = 100 μm.

posed of fibronectin-negative regions separated by strands of fibronectin-positive material, often in continuity with the medulla, infiltrating sometimes up to the surface cell layer. The fibronectin-positive medulla is formed of groups of loosely connected cells and contains lacunae in its inner region. The interface between cortex and medulla is underlined by a thick deposit of fibronectin. The cortical fibronectin-negative nests contain clusters of germ cells and somatic cells. Germ cells are identified by the expression of *Cvh* and of the pre-meiotic factor *Stra8* [64]. Somatic cells express *Lhx9*. *Ovex1* is highly transcribed in cells located at the inner side of these cortical nests and faintly expressed in the gonad surface cell layer. At this stage, the high density of germ and somatic cells intermingled in the cortical nests does not allow to exclude that *Ovex1* might be expressed in both cell types. Groups of *Ovex1*-positive cells are also dispersed within the subcortical region of the medulla. *Lhx9* transcripts are totally absent from the medulla. The dispersed *FoxL2* expression is limited to the medulla and *AMH* transcripts are located in the subcortical region of the medulla and in the fibronectin-positive cortical strands, as previously reported [4]. The degenerating right female gonad, relatively small and devoid of cortex, has no longer been studied.

Around hatching (P1) (Fig. 7B) occurs a dramatic remodeling of the left ovarian cortex. This is a prelude to folliculogenesis. The gonad surface cell layer, which has become negative for *Ovex1*, undergoes local disruptions leading to a sort of peeling. Exfoliation is initiated by apoptosis of superficial cells, as seen by TUNEL labeling (Additional file 8 (Figure S9, Detection of apoptotic cells in chicken ovary at hatching)). The desquamated surface of the gonad becomes jagged, with irregular cracks and protuberances. The phenomenon is emphasized by the intense expression of *Ovex1* in cells that form a nearly continuous but irregular layer, especially visible in the bottom of the cracks. This layer appears to form a barrier resisting the desquamation process. At this stage, *FoxL2* is expressed as previously in the medulla, more specifically in the juxtacortical region, but now also in some cells located on the cortical side of the fibronectin deposit that delimits cortex and medulla (shown by arrows). *Wnt4* expression in the gonad is almost undetectable.

Folliculogenesis starts in the left ovary after hatching, and is completed by 22 days [65]. One week after hatching, at P7 (Fig. 7C), the limit between cortex and medulla becomes undefined. Small follicular structures expressing high levels of *Ovex1* are observed near or at the surface of the desquamated ovary. It is also in these structures that a low *FoxL2* expression is now visible and where *Wnt4* starts to be expressed. These cells express also *AMH*, as illustrated in [4]. At previous stages, expression of this hor-

mone was restricted to cells of the fibronectin-positive regions, as seen in Fig. 7A.

At P14 (Fig. 7D), follicles of various sizes are present at the periphery of the ovary, the smaller ones (the last formed) being the most external. They are constituted of a single oocyte surrounded by a layer of somatic cells. These follicular cells (also called granulosa cells) express *Ovex1* at a high level, *FoxL2*, *AMH*, and in addition *Wnt4* as shown in [4]. At this stage, folliculogenesis appears uncompleted because a layer of *Ovex1*-positive cells remains close to the surface of the gonad.

At P28 (Fig. 7E), the ovary that has acquired an important development is highly folded. Cells expressing *Ovex1*, *FoxL2*, and *AMH*, as well as *Wnt4* (not shown) are essentially the granulosa cells of the follicles.

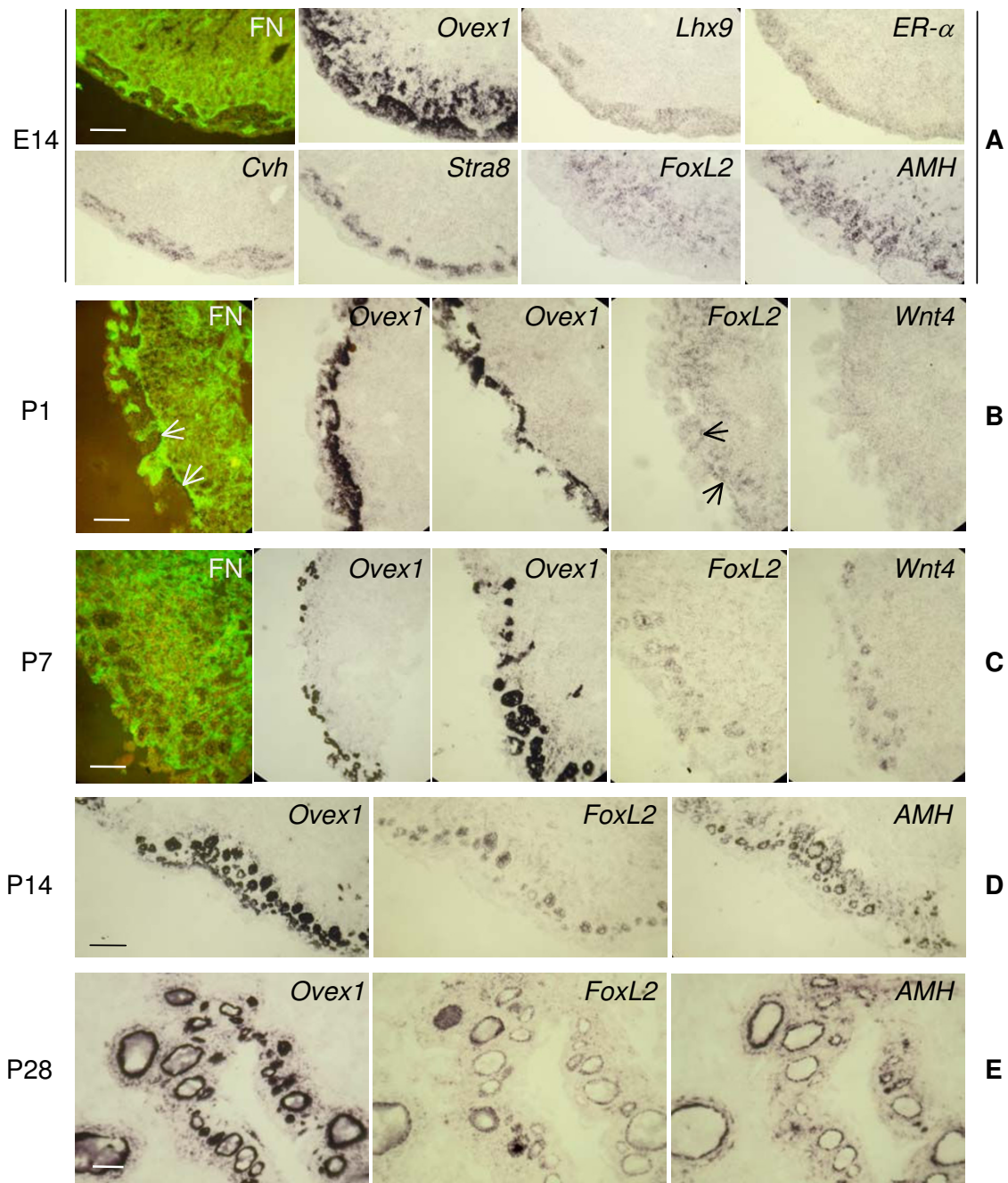
#### **Effect of the inhibition of estrogen synthesis on *Ovex1* expression**

Estrogens play an essential role in ovarian differentiation. Aromatase, the key enzyme that converts androgens to estrogens, is expressed in female gonads from E5-E6. A possible involvement of estrogens in *Ovex1* up-regulation after E6 in the left ovarian cortex and in the medulla of both female gonads, where the estrogen receptor ER- $\alpha$  is present, was thus investigated. Fadrozole (CGS 16949A), a non-steroidal aromatase inhibitor, was injected into the eggs at E4 to prevent estrogen synthesis, and the embryonic gonads examined at E14. Fadrozole treatment has been shown to lead to female to male sex reversal. The right gonad differentiates into a testis and the left one into an ovotestis or a testis. This is characterized by impairment of cortex development, decrease of aromatase synthesis, and formation of testicular cords expressing the testicular marker *SOX9* and elevated levels of *AMH* and *DMRT1* transcripts [22,27,66,67]. By comparison with the E14 control ovary (Fig. 7A), it is evident in Fig. 8 that the left gonad of the fadrozole-treated female is masculinized, with a thin cortical region and important medulla containing epithelial cords and lacunae. *Ovex1* expression is extremely reduced in this gonad in the medulla as in the remnants of the cortical region. Conversely, *Sox9*, which is not normally transcribed in female gonads, and *AMH* are expressed in internal epithelial structures that resemble testis cords.

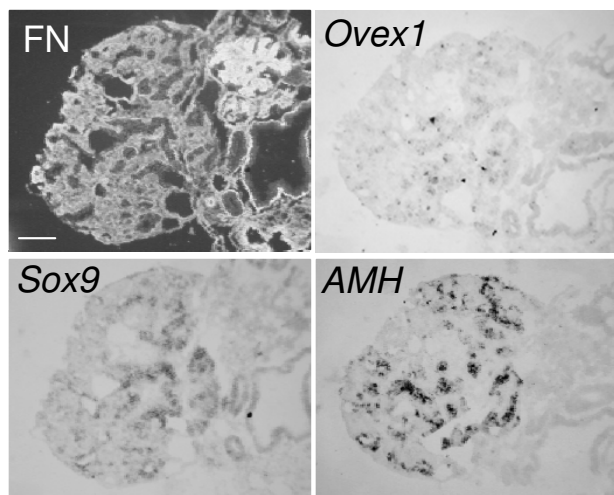
## **Discussion**

### **Characterization of *Ovex1* as endogenous retrovirus**

The SSH technique used in this study was designed to identify genes transcribed preferentially in the chicken embryo left ovary and thus potentially involved in the differentiation of this gonad. It revealed the existence of a locus that we called *Ovex1* according to its expression pattern. *Ovex1* has a structure similar to that of simple endog-

**Figure 7**

**Expression profiles of *Ovex1* and gonad markers in the left ovary before and after folliculogenesis.** Fibronectin (FN) is revealed by immunofluorescence and gene expression detected by *in situ* hybridization of digoxigenin-labeled RNA probes on transverse cryostat sections. **(A)** Left ovary at E14. Immunostaining of fibronectin (FN) and expression of *Ovex1*, *Lhx9*, *ER-α*, *Cvh*, *Stra8*, *FoxL2* and *AMH* in gonad serial sections. Fibronectin and *Cvh* are double labeling of the same section. **(B)** Left ovary at hatching (P1). Immunostaining of fibronectin (FN) and expression of *Ovex1* (two aspects of the cortex desquamation process are presented), *FoxL2* and *Wnt4*. Fibronectin and *FoxL2* are double labeling of the same section. Arrows indicate *FoxL2*-positive cells located on the cortical side of the fibronectin cortico-medullar interface. **(C)** Left ovary 7 days after hatching (P7). Immunostaining of fibronectin (FN) and expression of *Ovex1* (two aspects of the folliculogenesis are presented), *FoxL2* and *Wnt4*. Fibronectin and *FoxL2* are double labeling of the same section. **(D)** Left ovary at P14. Expression of *Ovex1*, *FoxL2* and *AMH* in serial sections. **(E)** Left ovary at P28. Expression of *Ovex1*, *FoxL2* and *AMH* in serial sections. Scale bars = 100 μm.



**Figure 8**  
**Effect of fadrozole treatment on *Ovex1* expression in the female left gonad at E14.** Chicken embryos were injected with fadrozole *in ovo* at E4. Fibronectin (FN) was revealed by immunofluorescence, and expression of *Ovex1*, *Sox9* and *AMH* was detected by *in situ* hybridization of digoxigenin-labeled RNA probes on serial transverse cryostat sections of the masculinized left female gonad. Fibronectin and *Sox9* are double labeling of the same section. Scale bar = 100  $\mu$ m.

enous retroviruses, with ORFs coding for potential proteins related to retroviral Gag and Pro-Pol polyproteins and a third one (ORF3) which displays only a faint similarity with retroviral envelope proteins. In the chicken ovary a full-length genomic mRNA and a singly spliced subgenomic transcript encoding only the third protein are present. Translation of the *Pro-Pol* ORF results presumably from read-through of the *Gag* stop-codon, like in gamma, epsilon and intermediate epsilon-like retroviruses.

A homologous sequence was found in the genome of zebra finch, a distantly related bird species. The location of these sequences, at similar positions in syntenic chromosomes, indicates an orthologous relationship. Partial sequences corresponding to *Gag* and *Pol* regions were also detected in turkey, guinea fowl and duck, suggesting the presence of similar genes in these species. Protein similarity of the five *Ovex1* RT fragments follows the phylogenetic relationship of the bird species. This relation is indicative of the vertical transmission of an ERV integrated into the genome of a common ancestor of these birds, therefore before the split between Galloanserae (chicken, turkey, guinea fowl, duck...) and Neoaves (zebra finch...) some 122 Myr (million years) ago [68]. Investigation for similar sequences in Paleognathae (tinamou and ratites), which diverged from the lineage leading to other

birds 139 Myr ago, would be necessary to ascertain the precise time of insertion.

Available data indicate that only one copy of the *Ovex1* *Gag-Pol* region exists in the genome of chicken and zebra finch, in contrast to most ERVs that constitute families of related sequences derived from an initial integrated virus. This region is not closely related to other known avian retroviruses. According to the RT-based phylogenetic analysis, *Ovex1* is mostly related to SnRV and class III retroviral elements but its relationship with these elements is very distant. Its closest relative is SpeV, an endogenous retrovirus of the tuatara (*Sphenodon*), last representative of an archaic reptilian lineage that already existed 220 Myr ago [69]. However the identity is rather low. In SpeV, only a part of the *Pro* and *RT* domains have been sequenced and the comparison is limited. More information would be required to confirm the relationship of these ERVs.

In contrast, the region of *Ovex1* ORF3 and the 3'UTR contain sequences similar to an ERV I repeated element and a fragment of CR1 that are similarly located in chicken and zebra finch DNA. This means that *Ovex1* is likely a composite element and that the recombination events occurred in an ancestral genome before divergence of the two bird clades 122 Myr ago. Such a chimerism resulting from exchange of *Pol* and *Env* is frequently observed in the ERVs [56].

The obvious deficiency of chicken and zebra finch *Ovex1* in comparison with other ERVs is the absence of identifiable LTRs. However, two imperfect direct repeats, located in the regions at the start and end of the transcription unit are the presumable remnants of former LTRs. Retroviral LTRs, generated from a single template during reverse transcription of RNA into DNA, are identical at the time of integration into the host genome. Over time, they will diverge in sequence because of mutations occurring during cellular DNA replication. Differences between the two repeats can be used to estimate the time that has elapsed since the ERV integration [32]. The lack of full LTRs and the high divergence of the remaining 5' and 3' repeats are consistent with the ancient origin of *Ovex1*.

The great majority of ERVs, having suffered random mutations since their insertion, have accumulated deletions, frameshifts and stop-codons that interrupt their ORFs [70]. In chicken *Ovex1*, all three ORFs are uninterrupted (though small deletions or local frameshifts might have remained undetected) and in zebra finch at least two of them. This is a rare event, unless the ERV has a very recent origin, which is not the case. *In vivo* translation of *Ovex1* transcripts remains to demonstrate. However, conservation of the ORFs suggests the existence of a selection pressure acting to retain not only the RNA but also the



encoded proteins. The effect of selection can be revealed by comparing the rates of synonymous (dS) and nonsynonymous (dN) nucleotide substitutions in chicken and zebra finch coding sequences [71]. dN/dS ratios <1 indicate the effect of a purifying selection acting to eliminate divergent protein sequences. *Gag* is under a strong purifying selection, with a dN/dS ratio of 0.08, close to the mean value (0.085) for genes located on macrochromosomes [72]. *Pol* and *ORF3*, with dN/dS ratios of 0.18 and 0.16, display less constraint in the purifying selection. In addition, we detected a certain number of single nucleotide substitutions between the wild Red Jungle fowl and the White Leghorn chicken, a domestic strain derived from the same *Gallus gallus* ancestor after some 10,000 years of domestication [31]. Most of these substitutions are silent (as seen in additional file 6 (Table S7, Polymorphisms in chicken *Ovex1* sequences)). This is in support of an ongoing purifying selection that appears more stringent for *Gag* and *Pol* than for *ORF3*.

*Ovex1*, devoid of homologous LTRs, is likely unable to propagate its sequence. *Gag* and *Pol* have retained enough conservation to be identified, but are nevertheless considerably different from proteins of other retroviruses. Important amino-acid residues implicated in viral enzyme activity are not present in chicken and zebra finch *Ovex1* sequences whereas they are conserved in all retroviruses including *Ovex1*'s closest relative, SpeV. RT and Int, if translated, are presumably inactive. The protein encoded by *ORF3*, if it is an envelope protein, appears also highly defective. Conservation of uninterrupted retroviral ORFs during more than 100 Myr despite the presence of a high number of presumably invalidating mutations can hardly occur just by chance and might indicate that these proteins fulfill some unknown biological function and are useful to the host.

#### Expression of *Ovex1* in chicken gonads

The *in situ* hybridization study shows that expression of *Ovex1* is restricted to specific cells of the gonads and thus tightly controlled. This expression is characterized by L/R asymmetry and is dependent on sex differentiation. In both sexes, *Ovex1* transcription is initiated at E5 in the cortex of left gonads before the onset of evident sex differentiation. At this stage, asymmetry of the gonads is already determined by the asymmetrical expression of *Pitx2c* in the epithelium of the left coelomic cavity, including the genital ridge [14]. In both sexes, *Pitx2c* inhibition of retinoic acid synthesis allows expression of *ER-α* and *SF-1* and cortical cell proliferation in the left gonad [15]. The patterns of expression of *Pitx2c* [15] and *Ovex1* at E5 and their transient expression in the left testis are rather similar, and it is plausible that *Pitx2c* controls *Ovex1* expression, directly or indirectly. In this context, the presence of a conserved *Pitx2* responsive element in the promoter region of

chicken and zebra finch *Ovex1* is interesting and deserves further study.

From E6 thereon, aromatase, the key enzyme of estrogen synthesis, is present in female gonads, not in male ones. Consequently, in female embryos, the cortex of the left gonad and both medullas, where the estrogen receptor *ER-α* is present, become direct physiological targets of estrogens. After E6 in the female, an increased and sustained transcription of *Ovex1* occurs in the cortex of the left ovary and a low expression is observed in the medulla of both gonads. In the male, the cortical region of the left testis able to express *Ovex1* becomes more and more limited. As *Pitx2c* [14], *Ovex1* is no longer expressed at E18. The absence of *Ovex1* transcription in testicular medullas suggests that estrogen stimulation might be required for the medullar expression. Estrogen deprivation by fadrozole treatment leads to the masculinization of female gonads with an inhibition of the left ovarian cortex development and the apparition in the medulla of epithelial structures analogous to testis cords. *Ovex1* expression is strongly decreased in the medulla as well as in remnants of the cortical region, showing the requirement for estrogens to get a sustained expression. This effect might be due to a direct action of estrogens on the *Ovex1* promoter but reflects also the role of the hormone in proliferation and/or differentiation of cortical and medullar cells able to express *Ovex1*.

During the embryonic differentiation of the ovary, most of the *Ovex1*-expressing cells are tightly associated with germ cells in the left cortex (or in close vicinity in the sub-cortical medulla). After hatching, they constitute the granulosa cell layer surrounding the oocyte in the forming follicles. This is in contrast with other markers of the granulosa like *FoxL2*, *AMH* and *Wnt4* [4], which are not expressed in cortical cells during embryogenesis, but in the medullar compartment, and start to be expressed in granulosa cells only when follicles form. The continuity of expression of *Ovex1* makes this factor an interesting marker of the filiation between cortical somatic cells and granulosa cells.

The expression of *Ovex1* enlightens the dramatic remodeling of the ovarian cortical surface that occurs around hatching preceding the folliculogenesis. The previously compact cortex, which contained nests of tightly packed meiotic germ cells and somatic cells covered by a continuous surface cell layer, suffers an important morphological change. The surface cell layer loses *Ovex1* expression; patches of cells undergo apoptosis, and local disruptions lead to an exfoliation of the superficial region of the cortex. *Ovex1*-expressing cells appear to constitute a barrier resisting the desquamation process before follicle formation. A more limited phenomenon has been observed by

scanning electron microscopy in human and mouse ovaries. The extrusion of germ cells and satellite somatic cells through temporary breaks of the ovarian surface epithelium under the pressure of the subjacent highly proliferating tissue has been reported [73].

#### **Putative roles of *Ovex1* in the gonads**

*Ovex1* expression constitutes an interesting marker of ovarian morphogenesis in chicken, but can it be an actor in this process? In the male, the transient *Ovex1* transcription in the left testis is dispensable for testis function since the right gonad becomes a functional testis despite the absence of *Ovex1* expression. The same conclusion cannot be drawn for female gonads.

ERV transcription in vertebrates is not an exception and displays variable tissue specificity [33,74]. In mammals, expression of several ERV envelopes is detected in various cancer cells, in particular ovarian cancers [75] and in several normal tissues [55,76]. However, the role of these retroviral proteins in vertebrate cells – if they are translated – is not often known. The unique but prototypical example is the recruitment of the fusogenic properties of retroviral envelopes for the specific mammalian function of placenta morphogenesis. HERV-W and HERV-FRD envelope proteins, named syncytin 1 and 2, are responsible for trophoblast cell fusion in placenta syncytiotrophoblast formation in humans [77,78]. Envelope proteins of various retroviruses have been independently recruited to fulfill the same function in other mammals [79,80]. However, it does not seem that embryonic ovary differentiation involves cell fusion. Another positive role of ERVs for the vertebrate host is protecting cells against infection by viruses of the same family. Endogenous envelope proteins saturate viral cell receptors, thus preventing exogenous virus entry. In addition, the Gag protein of MuERV has been shown to be closely related to the murine Fv1 gene that controls the replication of MMLV and prevents disease in mice infected by this retrovirus [43]. Such a role might be useful in the ovary.

It is noticeable that *Ovex1* is not the only retroviral sequence expressed in the embryonic chicken ovary. *FET-1* (female expressed transcript 1), a W-chromosome gene asymmetrically expressed in the left ovarian cortex between E4.5 and E6 has also a retroviral origin [2,11]. We found that it is similar to the consensus GGLTR7A retrovirus (described in RepBase) [36] over 87% of its length and encodes potentially an Env-like protein with similarity to the HERV-FRD Env protein (syncytin 2). A common role and/or a common regulation of *FET-1* and *Ovex1* might be investigated. To speculate further about the function of *Ovex1* (and *FET-1*) in the ovary, it would be necessary first to establish whether the encoded proteins are actually translated, and if so, whether they retain viral

functions, in particular if the putative envelope possess fusogenic properties. Alternatively, these proteins might have acquired new functions. The presence of nuclear localization signals and of a leucine zipper motif in *Ovex1* Gag protein suggests a possible role in nuclear regulations.

In a recent screening for ESTs present in adult hen ovarian follicles, an EST [GenBank:EC912004] that corresponds to *Ovex1* has been found to be expressed at a rate 6-time lower in hen ovaries of a chicken strain selected for its high egg production, compared with another strain with low egg production [81]. *Ovex1*, which is expressed in supporting somatic cells in close contact with germ cells in the embryonic cortex and in follicles up to adulthood, might play a role in regulating hen fertility at different stages: germ cell proliferation, meiosis, folliculogenesis, follicular survival or rate of follicle recruitment.

#### **Conclusion**

*Ovex1* is an ERV present in both chicken and zebra finch genomes. Similar sequences are detected in three other domestic birds. Sequencing of more bird and reptile genomes will presumably enlarge the family of *Ovex1*-containing species. Unlike most ERVs that are represented in the host genome by a family of related sequences, most of the *Ovex1* sequence (*Gag* and *Pol* regions) exists as a unique copy in the genome of chicken and zebra finch. However, these genomes are not fully investigated and random divergence of elements inserted more than 100 Myr ago may impair their identification. The conserved copy of this ERV might have been preserved because of a specific function. Specifically expressed in the gonads, *Ovex1* constitutes an interesting marker of the granulosa cell lineage useful for the study of ovarian morphogenesis. The future objective will be to determine if it is only a witness or if it is an actor. The involvement of a retroviral element in crucial processes of the ovarian development would be very exciting.

#### **Methods**

##### **RNA preparation**

Commercial White Leghorn chicken eggs were incubated at 38°C. Development stages (HH) are defined according to Hamburger and Hamilton [82]. Embryonic left and right gonads were dissected, collected individually into RNA-later solution (Qiagen), and stored at -80°C. Young embryos (up to E8) were genetically sexed by PCR [4] on DNA purified from extragonadal tissues with the NucleoSpin tissue kit (Macherey-Nagel). Total RNA was purified from pools of male or female, left or right gonads (35 gonads, about 18 mg of tissues for E8 embryos), using the RNeasy RNA mini extraction kit (Qiagen), with DNase treatment.

### Suppression Subtractive Hybridization (SSH)

The suppression subtractive hybridization (SSH) technique [37] was used to select transcripts expressed at a higher level in the left ovary than in the right one of E8 female chicken embryos. The cDNAs were prepared and amplified from total RNA of left or right ovaries, using the SMART PCR cDNA synthesis kit (BD Biosciences Clontech) according to the manufacturer's instructions. To generate the first strand, 1 µg of total RNA was reverse-transcribed with an oligo-dT containing primer (CDS), using PowerScript reverse transcriptase (BD Biosciences Clontech). Bicatena cDNA was produced using the SMART II 5'-anchored primer and amplified by PCR for 15 cycles. SSH was performed with the PCR-Select cDNA Subtraction kit (BD Biosciences Clontech) according to the user's manual. The two amplified cDNA populations were digested with the restriction enzyme *RsaI*. After digestion, the left ovary cDNAs (LO) were subtracted against the right ovary cDNAs (RO), giving a cDNA pool enriched in left ovary transcripts (LO-RO). Conversely, the right ovary cDNAs were subtracted against the left ovary cDNAs, giving a cDNA pool enriched in right ovary transcripts (RO-LO). The efficiency of normalization and subtraction was assayed by comparing the abundance of the constitutively expressed glyceraldehyde 3-phosphate dehydrogenase cDNA in initial and subtracted cDNA pools, using chicken-specific PCR primers (cGAPDHs: 5'ACCACTGTCCATGCCATCAC3', cGAPDHα: 5'TCCACAACACGGTTGCTGTA3').

### Cloning of the subtracted cDNA fragments

(LO-RO) subtracted cDNA fragments were cloned into pGEMT-easy plasmid (Promega), to construct a (LO-RO) subtracted cDNA library. Two hundred and fifty individual colonies were randomly selected, grown in 96-well plates in 100 µL LB medium plus ampicillin for 14 hours at 37°C and frozen at -80°C after addition of 10% glycerol.

### Differential hybridization screening

Differential hybridization screening was performed to identify the cDNA clones displaying the greatest asymmetry of expression between left and right gonads. A procedure adapted from the PCR-Select differential screening kit user's manual (BD Biosciences Clontech) was used with the following differences. The (LO-RO) cDNA clone library was used to prepare macroarrays. The insert of each clone was PCR-amplified, using PCR-Select nested primers 1 (5'TCGAGCGGCCGCCCGGCGGAGGT3') and 2R (5'AGCGTGGTCGCGGCCGAGGT3'). Four identical macroarrays were prepared by spotting 2-µL aliquots of each PCR reaction mixture on nylon membranes (Hybond-N+, Amersham Pharmacia Biotech). Denaturation of the DNA was achieved by blotting the membranes onto 0.5 M NaOH, 1.5 M NaCl-impregnated Whatman 3

MM paper for 2 min, followed by neutralization with 0.5 M Tris-HCl pH7.4, 1.5 M NaCl for 5 min and 3xSSC for 5 min, under the same conditions. Dry membranes were then exposed to ultraviolet light (0.6 J/cm<sup>2</sup>) to perform DNA cross-linking. The membranes were hybridized with radioactive probes corresponding to the subtracted and non-subtracted cDNA pools: LO-RO, RO-LO, LO, and RO. <sup>32</sup>P labeled probes were prepared by random priming (Invitrogen Random primers DNA labeling system). Hybridization was performed as described previously [83]. Denatured herring sperm DNA (100 µg/mL) and 3 µg/mL each of PCR-Select nested primers 1 and 2R, anti-nested primer 1: (5'ACCTGCCCCGGCGGCCGCTCGA3') and anti-nested primer 2R (5'ACCTGCGCCGCGACCACGCT3') were added to the hybridization solution to prevent unspecific hybridization of the PCR-Select adaptor sequences. Membranes were subjected to autoradiography and the intensity of the spots quantified using the ImageJ software. The clones displaying the highest ratios of (LO-RO) versus (RO-LO) hybridization were selected. The inserts were sequenced by automated fluorescence sequencing (MWG Biotech France), using pGEMT-specific primers.

### Determination of the complete *Ovex1* sequence

Primers derived from the genomic sequence were used to amplify by RT-PCR overlapping cDNA fragments from E8 left ovary RNA. Reverse transcription was performed by random priming on 2 µg total RNA using SuperScript II Rnase H- reverse transcriptase (Invitrogen). PCR amplification was performed on 1/20th of the cDNA product (equivalent to 100 ng of RNA), with 10 µg/mL of the PCR primer pairs listed in additional file 9 (Table S1, Primers and PCR conditions). The mixture contained 67 mM Tris-HCl (pH8.8), 17 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 6.7 mM MgCl<sub>2</sub>, 10 mM 2-mercaptoethanol, 6.7 µM EDTA, 10% (v/v) dimethylsulphoxide, 0.5 mM each of the four dNTP, and 2 units of EurobioTaq polymerase (Eurobio) in a final volume of 50 µL. Hot start was performed by addition of the Taq polymerase after 5 min of preincubation at 80° and followed by 30 cycles of amplification (45 s at 95°, 45 s at 57°, and 180 s at 72°). After analysis for purity of the PCR product by polyacrylamide gel electrophoresis, the PCR reaction mixture was treated with ExoSap-It (USB) to remove primers and nucleotides and used for direct sequencing of the fragments in both directions (MWG Biotech France).

### Determination of cDNA ends

5' and 3' cDNA fragments were isolated by the rapid amplification of cDNA end methods (5'RACE and 3'RACE) using total RNA purified from E12 left ovaries. Amplification was performed with the SMART RACE cDNA amplification kit (BD Biosciences Clontech), according to the manufacturer's instructions, using

SMART universal primers and the gene specific primers indicated in additional file 9 (Table S1, Primers and PCR conditions). The amplification products were analyzed by electrophoresis on agarose gel and the main band, purified with NucleoSpin extract II (Macherey Nagel), was cloned into pGEMT-easy vector (Promega). At least 5 independent recombinant clones were sequenced in each experiment.

#### **Amplification of *Ovex1*-related sequences from other domestic fowls**

Turkey, guinea fowl and duck DNAs were prepared from muscular tissues. PCR amplification was carried out with primers and conditions indicated in additional file 9 (Table S1, Primers and PCR conditions). PCR fragments were sequenced directly.

#### **Computer sequence analysis**

DNA sequence searches in the chicken (*galGal3*) and zebra finch (*taeGut1*) genome databases were performed with BLAT program <http://genome.ucsc.edu/cgi-bin/hgGateway>. Nucleic acid and protein similarity was detected with the BLAST server at NCBI <http://www.ncbi.nlm.nih.gov>. Mobile elements were identified by RepeatMasker <http://www.repeatmasker.org> based on the RepBase database <http://www.girinst.org> [36]. Searches for the PBS were carried out using chicken tRNA sequence data <http://gtrnadb.ucsc.edu/Ggall/Ggall-align.html> [39]. Promoter screening for transcription factor responsive elements was carried out with the MatInspector program <http://www.genomatix.de/>. Sequence alignments were performed with ClustalW2 multialignment program [84], with default settings <http://www.ebi.ac.uk> and adjusted manually. The phylogenetic trees and bootstrap values were calculated using the Neighbor-Joining method [85] with QuickTree software [86] <http://mobyle.pasteur.fr> and drawn with the NJplot software [87]. The ORF map and the hydrophobicity plot calculated according to Kyte and Doolittle [88] were created with the DNA Strider program [89]. Protein structure was analyzed with Psort version II <http://mobyle.pasteur.fr> and N-glycosylation sites predicted by the NetNGlyc 1.0 Server <http://www.cbs.dtu.dk/>. Synonymous and non-synonymous substitution rates corrected for multiple substitutions were determined using the SNAP program <http://www.hiv.lanl.gov> [90]. Quantification of photographic spot intensity was carried out with Image processing and analysis in Java (ImageJ 1.33) <http://rsb.info.nih.gov/ij/>.

#### **Database accession numbers and abbreviations**

Complete nucleotide sequences of the chicken *Ovex1* unspliced and spliced mRNAs (splice variant 1) were deposited under accessions [GenBank: [FJ406461](#)] and [GenBank: [FJ406462](#)]. Partial *Gag* sequences of turkey,

guinea fowl and duck are respectively [GenBank: [FJ423166](#), GenBank: [FJ423167](#), GenBank: [FJ423168](#)] and partial *RT* sequences [GenBank: [FJ423169](#), GenBank: [FJ423170](#), GenBank: [FJ423171](#)].

Abbreviations and database accession numbers of other sequences referred to in the text and figures are as follows: ALV, Avian leukosis virus, [GenBank: [NC\\_001408](#)]; cENS3, Pol-like protein ENS-3, [GenBank: [NP\\_989963](#)]; Dev1, *Dendrobates ventrimaculatus* ERV 1, [EMBL: [X95795](#)]; EAV-HP, Avian endogenous retrovirus EAV-HP, [EMBL: [AJ292966](#)]; FFV, Feline foamy virus, [GenBank: [NP\\_056914](#)]; GGERV-L<sub>C</sub>, ERV3 endogenous retrovirus from chicken, RepBase [36], RT (nt 2312–2785); GGLTR11-int, ERV1 endogenous retrovirus from chicken, RepBase [36], RT (nt 1686–2183), Env (nt 4126–5472); HERV-H, Human endogenous retrovirus HERV-H/env62, [EMBL: [AJ289709](#)]; HERV-L, Human endogenous retroviral element HERV-L, [EMBL: [X89211](#)]; HFV, Human foamy virus, [GenBank: [NC\\_001736](#)]; HIV1, Human immunodeficiency virus 1, [GenBank: [NC\\_001802](#)]; HTLV1, Human T-lymphotropic virus 1, [GenBank: [NC\\_001436](#)]; MMLV, Moloney murine leukemia virus (Pol polyprotein), [Swiss-Prot: [P03355](#)]; MMLV, Moloney murine leukemia virus (Env polyprotein), [Swiss-Prot: [P03385](#)]; MMTV, Mouse mammary tumor virus, [GenBank: [NC\\_001503](#)]; MuERV-L, Mus musculus endogenous retroviral sequence MuERV-L, [EMBL: [Y12713](#)]; PERV, Porcine endogenous retrovirus, [EMBL: [AJ279057](#)]; REV-A, Reticuloendotheliosis virus isolate REV-A, [GenBank: [DQ237900](#)]; RV\_Tinamou, RV-Tinamou partial mRNA for polyprotein, [EMBL: [AJ225235](#)]; SnRV, Snakehead retrovirus, [GenBank: [NC\\_001724](#)]; SpeV, Sphenodon endogenous retroviral gene for protease and reverse transcriptase, [EMBL: [X85037](#)]; TERV, Tetraodon endogenous retrovirus, [GenBank: [AF289082](#)]; WDSV, Walleye dermal sarcoma virus, [GenBank: [NC\\_001867](#)]; WEHV1, Walleye epidermal hyperplasia virus type 1, [GenBank: [AF133051](#)]; WEHV2, Walleye epidermal hyperplasia virus type 2, [GenBank: [AF133052](#)]; Xen1, *Xenopus laevis* endogenous retrovirus Xen1, [EMBL: [AJ506107](#)]; ZFERV, *Danio rerio* endogenous retrovirus ZFERV, [GenBank: [AF503912](#)].

#### **Gene expression analysis by semi-quantitative RT-PCR**

Semi-quantitative RT-PCR was carried out to determine the specificity of expression of *Ovex1*. Experimental conditions were as described above, except for the use of 1/100<sup>th</sup> of the RT product (equivalent to 20 ng RNA). The number of PCR cycles was limited and optimized in each case. Primer pairs, hybridization temperatures and number of cycles are given in additional file 9 (Table S1, Primers and PCR conditions). Elongation at 72° was for 75 s. PCR products were analyzed by electrophoresis on

10% polyacrylamide gels. Controls without reverse transcriptase were negative.

### In situ hybridization

*In situ* hybridization was performed on Leghorn gonad cryostat sections as described previously [7], using anti-sense digoxigenin-labeled riboprobes. *Lhx9* and *Cvh* probes have already been described [4]. Other probes were prepared from cloned RT-PCR fragments, according to the same procedure. The *FoxL2*, *Stra8* and *ER-α* probes correspond respectively to [GenBank:AY155534] (nucleotides 1 to 330), [GenBank:XM\_416179] (nt 121 to 825) and [EMBL:X03805] (nt 827 to 1752). For *Ovex1*, a *Pol* probe corresponding to nucleotides 2511 to 3245 was used in the experiments reported. Hybridization of a 3'-UTR probe (nt 8633 to 8919) gave similar results. Hybridization with sense probes was negative. Apoptotic cells were detected by the TUNEL method, using In Situ Cell Death Detection Kit, Fluorescein (Roche), on frozen sections previously treated for *in situ* hybridization. After PBS washing, sections were incubated for 1 hour at room temperature with the TUNEL Label Mixture containing terminal transferase.

### Fadrozole treatment

White Leghorn eggs, were treated on day 4 of incubation (HH stage 24) with a single injection of 1 mg fadrozole (CGS 16949A), a specific nonsteroidal aromatase inhibitor (Novartis, Basel, Switzerland), as described [66]. Embryos were sacrificed at 14 days of incubation.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

DCE conceived the study, carried out the SSH selection and molecular biology studies, and drafted the manuscript. *In situ* hybridization studies were designed by SM and DCE, carried out by SM and NC and analyzed by SM who participated in drafting the manuscript. All authors read and approved the final manuscript

### Additional material

#### Additional file 1

*Figure S2. Effect of the internal polyadenylation signals.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S1.pdf]

#### Additional file 2

*Figure S3. Partial zebra finch Ovex1 sequence.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S2.pdf]

#### Additional file 3

*Figure S4. Alignment of chicken and zebra finch Ovex1 5'-proximal DNA sequences.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S3.pdf]

#### Additional file 4

*Figure S5. Alignment of Ovex1 putative proteins.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S4.pdf]

#### Additional file 5

*Figure S6. Structure and conservation of ORF3-encoded proteins.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S5.pdf]

#### Additional file 6

*Table S7 - Polymorphisms in chicken Ovex1 sequences. Polymorphisms observed in the Ovex1 sequence between the Red Jungle Fowl (genomic sequence galGal3) and the white Leghorn chicken strain (this study).*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S6.pdf]

#### Additional file 7

*Figure S8. Alignment of Ovex1 RT and homologous sequences.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S7.pdf]

#### Additional file 8

*Figure S9. Detection of apoptotic cells in chicken ovary at hatching.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S8.pdf]

#### Additional file 9

*Table S1. Primers and PCR conditions.*

Click here for file

[http://www.biomedcentral.com/content/supplementary/1742-4690-6-59-S9.pdf]

### Acknowledgements

We thank J-Y Picard, director of INSERM U782, for supporting this work. We acknowledge Dr. C. Pieau's laboratory for providing us with fadrozole-treated chicken embryos. We are grateful to Dr. T. Heidmann for helpful discussion, to Dr. D. Ribet for discussion and critical reading of the manuscript and to Dr. S. Pereira for advice about bird phylogeny.

### References

1. Smith CA, Sinclair AH: **Sex determination: insights from the chicken.** *Bioessays* 2004, **26**:120-132.
2. Schmid M, Nanda I, Hoehn H, Schartl M, Haaf T, Buerstedde JM, Arakawa H, Caldwell RB, Weigend S, Burt DW, et al.: **Second report on chicken genes and chromosomes 2005.** *Cytogenet Genome Res* 2005, **109**:415-479.

3. Merchant-Larios HPL, Reyss-Brion M: **Early morphogenesis of chick gonad in the absence of mesonephros.** *Dev Growth Differ* 1984, **26**:403-417.
4. Oreal E, Mazaud S, Picard JY, Magre S, Carre-Eusebe D: **Different patterns of anti-Mullerian hormone expression, as related to DMRT1, SF-1, WTI, GATA-4, Wnt-4, and Lhx9 expression, in the chick differentiating gonads.** *Dev Dyn* 2002, **225**:221-232.
5. Kudo T, Sutou S: **Molecular cloning of chicken FTZ-F1-related orphan receptors.** *Gene* 1997, **197**:261-268.
6. Smith CA, Smith MJ, Sinclair AH: **Expression of chicken steroidogenic factor-1 during gonadal sex differentiation.** *Gen Comp Endocrinol* 1999, **113**:187-196.
7. Oreal E, Pieau C, Mattei MG, Josso N, Picard JY, Carre-Eusebe D, Magre S: **Early expression of AMH in chicken embryonic gonads precedes testicular SOX9 expression.** *Dev Dyn* 1998, **212**:522-532.
8. Van Limborgh J: **Le premier indice de la differenciation sexuelle des gonades chez l'embryon de poulet.** *Archives d Anatomie Microscopique et de Morphologie Experimentale* 1968, **57**:79-90.
9. Andrews JE, Smith CA, Sinclair AH: **Sites of estrogen receptor and aromatase expression in the chicken embryo.** *Gen Comp Endocrinol* 1997, **108**:182-190.
10. Nakabayashi O, Kikuchi H, Kikuchi T, Mizuno S: **Differential expression of genes for aromatase and estrogen receptor during the gonadal development in chicken embryos.** *J Mol Endocrinol* 1998, **20**:193-202.
11. Reed KJ, Sinclair AH: **FET-1: a novel W-linked, female specific gene up-regulated in the embryonic chicken ovary.** *Mech Dev* 2002, **119**(Suppl 1):S87-90.
12. Hoshino A, Koide M, Ono T, Yasugi S: **Sex-specific and left-right asymmetric expression pattern of Bmp7 in the gonad of normal and sex-reversed chicken embryos.** *Dev Growth Differ* 2005, **47**:65-74.
13. Smith CA, Shoemaker CM, Roeszler KN, Queen J, Crews D, Sinclair AH: **Cloning and expression of R-Spondin1 in different vertebrates suggests a conserved role in ovarian development.** *BMC Dev Biol* 2008, **8**:72.
14. Guioli S, Lovell-Badge R: **PITX2 controls asymmetric gonadal development in both sexes of the chick and can rescue the degeneration of the right ovary.** *Development* 2007, **134**:4199-4208.
15. Ishimaru Y, Komatsu T, Kasahara M, Katoh-Fukui Y, Ogawa H, Toyama Y, Maekawa M, Toshimori K, Chandraratna RA, Morohashi K, Yoshioka H: **Mechanism of asymmetric ovarian development in chick embryos.** *Development* 2008, **135**:677-685.
16. Rodriguez-Leon J, Rodriguez Esteban C, Marti M, Santiago-Josefat B, Dubova I, Rubiralta X, Izpisua Belmonte JC: **Pitx2 regulates gonad morphogenesis.** *Proc Natl Acad Sci USA* 2008, **105**:11242-11247.
17. Yoshioka H, Meno C, Koshiba K, Sugihara M, Itoh H, Ishimaru Y, Inoue T, Ohuchi H, Semina EV, Murray JC, et al.: **Pitx2, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry.** *Cell* 1998, **94**:299-305.
18. Morais da Silva S, Hacker A, Harley V, Goodfellow P, Swain A, Lovell-Badge R: **Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds.** *Nat Genet* 1996, **14**:62-68.
19. Kent J, Wheatley SC, Andrews JE, Sinclair AH, Koopman P: **A male-specific role for SOX9 in vertebrate sex determination.** *Development* 1996, **122**(9):2813-2822.
20. Shan Z, Nanda I, Wang Y, Schmid M, Vortkamp A, Haaf T: **Sex-specific expression of an evolutionarily conserved male regulatory gene, DMRT1, in birds.** *Cytogenet Cell Genet* 2000, **89**:252-257.
21. Raymond CS, Kettlewell JR, Hirsch B, Bardwell VJ, Zarkower D: **Expression of Dmrt1 in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development.** *Dev Biol* 1999, **215**:208-220.
22. Nishikimi H, Kansaku N, Saito N, Usami M, Ohno Y, Shimada K: **Sex differentiation and mRNA expression of P450c17, P450arom and AMH in gonads of the chicken.** *Mol Reprod Dev* 2000, **55**:20-30.
23. Yoshida K, Shimada K, Saito N: **Expression of P450(17 alpha) hydroxylase and P450 aromatase genes in the chicken gonad before and after sexual differentiation.** *Gen Comp Endocrinol* 1996, **102**:233-240.
24. Nomura O, Nakabayashi O, Nishimori K, Yasue H, Mizuno S: **Expression of five steroidogenic genes including aromatase gene at early developmental stages of chicken male and female embryos.** *J Steroid Biochem Mol Biol* 1999, **71**:103-109.
25. Scheib D: **Effects and role of estrogens in avian gonadal differentiation.** *Differentiation* 1983, **23**(Suppl):S87-92.
26. Elbrecht A, Smith RG: **Aromatase enzyme activity and sex determination in chickens.** *Science* 1992, **255**:467-470.
27. Vaillant S, Dorizzi M, Pieau C, Richard-Mercier N: **Sex reversal and aromatase in chicken.** *J Exp Zool* 2001, **290**:727-740.
28. Govoroun MS, Pannetier M, Pailhoux E, Cocquet J, Brillard JP, Couty I, Batellier F, Cotinot C: **Isolation of chicken homolog of the FOXL2 gene and comparison of its expression patterns with those of aromatase during ovarian development.** *Dev Dyn* 2004, **231**:859-870.
29. Smith CA, Roeszler KN, Bowles J, Koopman P, Sinclair AH: **Onset of meiosis in the chicken embryo; evidence of a role for retinoic acid.** *BMC Dev Biol* 2008, **8**:85.
30. Wallenburg J: **[Macroscopy, light and electron microscopy studies on the genesis and function of the gonads after experimental sex-reversal following left-side ovariectomy of hen chicks (Gallus domesticus)].** *Gegenbaurs Morphol Jahrb* 1982, **128**:463-529.
31. ICGSC: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.** *Nature* 2004, **432**:695-716.
32. Huda A, Polavarapu N, Jordan IK, McDonald JF: **Endogenous retroviruses of the chicken genome.** *Biol Direct* 2008, **3**:9.
33. Jern P, Coffin JM: **Effects of retroviruses on host genome function.** *Annu Rev Genet* 2008, **42**:709-732.
34. Gifford R, Kabat P, Martin J, Lynch C, Tristem M: **Evolution and distribution of class II-related endogenous retroviruses.** *J Virol* 2005, **79**:6478-6486.
35. Jern P, Sperber GO, Blomberg J: **Use of endogenous retroviral sequences (ERVs) and structural markers for retroviral phylogenetic inference and taxonomy.** *Retrovirology* 2005, **2**:50.
36. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase Update, a database of eukaryotic repetitive elements.** *Cytogenet Genome Res* 2005, **110**:462-467.
37. Diatchenko L, Lukyanov S, Lau YF, Siebert PD: **Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes.** *Methods Enzymol* 1999, **303**:349-380.
38. Haas NB, Grabowski JM, Sivitz AB, Burch JB: **Chicken repeat I (CR1) elements, which define an ancient family of vertebrate non-LTR retrotransposons, contain two closely spaced open reading frames.** *Gene* 1997, **197**:305-309.
39. Schattner P, Brooks AN, Lowe TM: **The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs.** *Nucleic Acids Res* 2005, **33**:V686-689.
40. Amendt BA, Sutherland LB, Semina EV, Russo AF: **The molecular basis of Rieger syndrome. Analysis of Pitx2 homeodomain protein activities.** *J Biol Chem* 1998, **273**:20066-20072.
41. Kozak M: **Pushing the limits of the scanning mechanism for initiation of translation.** *Gene* 2002, **299**:1-34.
42. Hart D, Frerichs GN, Rambaut A, Onions DE: **Complete nucleotide sequence and transcriptional analysis of snakehead fish retrovirus.** *J Virol* 1996, **70**:3606-3616.
43. Benit L, De Parseval N, Casella JF, Callebaut I, Cordonnier A, Heidmann T: **Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene.** *J Virol* 1997, **71**:5652-5657.
44. Hicks GR, Raikhel NV: **Protein import into the nucleus: an integrated view.** *Annu Rev Cell Dev Biol* 1995, **11**:155-188.
45. Lupas A, Van Dyke M, Stock J: **Predicting coiled coils from protein sequences.** *Science* 1991, **252**(5010):1162-1164.
46. Yoshinaka Y, Katoh I, Copeland TD, Oroszlan S: **Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon.** *Proc Natl Acad Sci USA* 1985, **82**:1618-1622.
47. Tristem M, Myles T, Hill F: **A highly divergent retroviral sequence in the tuatara (Sphenodon).** *Virology* 1995, **210**:206-211.
48. Doolittle RF, Feng DF, Johnson MS, McClure MA: **Origins and evolutionary relationships of retroviruses.** *Q Rev Biol* 1989, **64**:1-30.

49. LaPierre LA, Holzschu DL, Bowser PR, Casey JW: **Sequence and transcriptional analyses of the fish retroviruses walleye epidermal hyperplasia virus types 1 and 2: evidence for a gene duplication.** *J Virol* 1999, **73**:9393-9403.
50. Jacobo-Molina A, Ding J, Nanni RG, Clark AD Jr, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P, et al.: **Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 Å resolution shows bent DNA.** *Proc Natl Acad Sci USA* 1993, **90**:6320-6324.
51. Johnson MS, McClure MA, Feng DF, Gray J, Doolittle RF: **Computer analysis of retroviral pol genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes.** *Proc Natl Acad Sci USA* 1986, **83**:7648-7652.
52. Malik HS, Eickbush TH: **Phylogenetic analysis of ribonuclease H domains suggests a late, chimeric origin of LTR retrotransposable elements and retroviruses.** *Genome Res* 2001, **11**:1187-1197.
53. Kulkosky J, Jones KS, Katz RA, Mack JP, Skalka AM: **Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases.** *Mol Cell Biol* 1992, **12**:2331-2338.
54. Dimcheff DE, Krishnan M, Mindell DP: **Evolution and characterization of tetraonine endogenous retrovirus: a new virus related to avian sarcoma and leukosis viruses.** *J Virol* 2001, **75**:2002-2009.
55. de Parseval N, Lazar V, Casella JF, Benit L, Heidmann T: **Survey of human genes of retroviral origin: identification and transcriptome of the genes with coding capacity for complete envelope proteins.** *J Virol* 2003, **77**:10414-10422.
56. Benit L, Dessen P, Heidmann T: **Identification, phylogeny, and evolution of retroviral elements based on their envelope genes.** *J Virol* 2001, **75**:11709-11719.
57. Cordonnier A, Casella JF, Heidmann T: **Isolation of novel human endogenous retrovirus-like elements with foamy virus-related pol sequence.** *J Virol* 1995, **69**:5890-5897.
58. Xiong Y, Eickbush T: **Origin and evolution of retroelements based upon their reverse transcriptase sequences.** *Embo J* 1990, **9**:3353-3362.
59. Herniou E, Martin J, Miller K, Cook J, Wilkinson M, Tristem M: **Retroviral diversity and distribution in vertebrates.** *J Virol* 1998, **72**:5955-5966.
60. Acloque H, Risson V, Birot AM, Kunita R, Pain B, Samarut J: **Identification of a new gene family specifically expressed in chicken embryonic stem cells and early embryo.** *Mech Dev* 2001, **103**:79-91.
61. Mazaud S, Oreal E, Guigon CJ, Carre-Eusebe D, Magre S: **Lhx9 expression during gonadal morphogenesis as related to the state of cell differentiation.** *Gene Expr Patterns* 2002, **2**:373-377.
62. Tsunekawa N, Naito M, Sakai Y, Nishida T, Noce T: **Isolation of chicken vasa homolog gene and tracing the origin of primordial germ cells.** *Development* 2000, **127**:2741-2750.
63. Hudson QJ, Smith CA, Sinclair AH: **Aromatase inhibition reduces expression of FOXL2 in the embryonic chicken ovary.** *Dev Dyn* 2005, **233**:1052-1055.
64. Oulad-Abdelghani M, Bouillet P, Decimo D, Gansmuller A, Heyberger S, Dolle P, Bronner S, Lutz Y, Chambon P: **Characterization of a premeiotic germ cell-specific cytoplasmic protein encoded by Stra8, a novel retinoic acid-responsive gene.** *J Cell Biol* 1996, **135**:469-477.
65. Ch'in Suang H, Gaginskaia ER, Kalinina El: **[Characteristics of oogenesis in the chick. I. The extrafollicular period in the development of the oocytes].** *Ontogenez* 1979, **10**:340-349.
66. Vaillant S, Magre S, Dorizzi M, Pieau C, Richard-Mercier N: **Expression of AMH, SFI, and SOX9 in gonads of genetic female chickens during sex reversal induced by an aromatase inhibitor.** *Dev Dyn* 2001, **222**:228-237.
67. Smith CA, Katz M, Sinclair AH: **DMRT1 is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos.** *Biol Reprod* 2003, **68**:560-570.
68. Pereira SL, Baker AJ: **A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock.** *Mol Biol Evol* 2006, **23**:1731-1740.
69. Wang Z, Miyake T, Edwards SV, Amemiya CT: **Tuatara (Sphenodon) genomics: BAC library construction, sequence survey, and application to the DMRT gene family.** *J Hered* 2006, **97**:541-548.
70. Villesen P, Aagaard L, Wiuf C, Pedersen FS: **Identification of endogenous retroviral reading frames in the human genome.** *Retrovirology* 2004, **1**:32.
71. Nei M, Gojobori T: **Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions.** *Mol Biol Evol* 1986, **3**:418-426.
72. Mank JE, Axelsson E, Ellegren H: **Fast-X on the Z: rapid evolution of sex-linked genes in birds.** *Genome Res* 2007, **17**:618-624.
73. Motta PM, Makabe S, Nottola SA: **The ultrastructure of human reproduction. I. The natural history of the female germ cell: origin, migration and differentiation inside the developing ovary.** *Hum Reprod Update* 1997, **3**:281-295.
74. Shen CH, Steiner LA: **Genome structure and thymic expression of an endogenous retrovirus in zebrafish.** *J Virol* 2004, **78**:899-911.
75. Wang-Johanning F, Liu J, Rycak K, Huang M, Tsai K, Rosen DG, Chen DT, Lu DW, Barnhart KF, Johanning GL: **Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer.** *Int J Cancer* 2007, **120**:81-90.
76. Lee YK, Chew A, Phan H, Greenhalgh DG, Cho K: **Genome-wide expression profiles of endogenous retroviruses in lymphoid tissues and their biological properties.** *Virology* 2008, **373**:263-273.
77. Mi S, Lee X, Li X, Veldman GM, Finnerty H, Racie L, LaVallie E, Tang XY, Edouard P, Howes S, et al.: **Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis.** *Nature* 2000, **403**:785-789.
78. Blaise S, de Parseval N, Benit L, Heidmann T: **Genomewide screening for fusogenic human endogenous retrovirus envelopes identifies syncytin 2, a gene conserved on primate evolution.** *Proc Natl Acad Sci USA* 2003, **100**:13013-13018.
79. Dupressoir A, Marceau G, Vernochet C, Benit L, Kanellopoulos C, Sapin V, Heidmann T: **Syncytin-A and syncytin-B, two fusogenic placenta-specific murine envelope genes of retroviral origin conserved in Muridae.** *Proc Natl Acad Sci USA* 2005, **102**:725-730.
80. Dunlap KA, Palmarini M, Varela M, Burghardt RC, Hayashi K, Farmer JL, Spencer TE: **Endogenous retroviruses regulate periimplantation placental growth and differentiation.** *Proc Natl Acad Sci USA* 2006, **103**:14390-14395.
81. Yang KT, Lin CY, Huang HL, Liou JS, Chien CY, Wu CP, Huang CW, Ou BR, Chen CF, Lee YP, et al.: **Expressed transcripts associated with high rates of egg production in chicken ovarian follicles.** *Mol Cell Probes* 2008, **22**:47-54.
82. Hamburger V, Hamilton HL: **A series of normal stages in the development of the chick embryo. 1951.** *Dev Dyn* 1992, **195**:231-272.
83. Eusebe D, di Clemente N, Rey R, Pieau C, Vigier B, Josso N, Picard JY: **Cloning and expression of the chick anti-Mullerian hormone gene.** *J Biol Chem* 1996, **271**:4798-4804.
84. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice.** *Nucleic Acids Res* 1994, **22**:4673-4680.
85. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees.** *Mol Biol Evol* 1987, **4**:406-425.
86. Howe K, Bateman A, Durbin R: **QuickTree: building huge Neighbour-Joining trees of protein sequences.** *Bioinformatics* 2002, **18**:1546-1547.
87. Perriere G, Gouy M: **WWW-query: an on-line retrieval system for biological sequence banks.** *Biochimie* 1996, **78**:364-369.
88. Kyte J, Doolittle RF: **A simple method for displaying the hydrophobic character of a protein.** *J Mol Biol* 1982, **157**:105-132.
89. March C: **'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers.** *Nucleic Acids Res* 1988, **16**:1829-1836.
90. Korber B: *HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences* Edited by: Allen G Rodrigo, Gerald H Learn. Kluwer Academic Publishers. Dordrecht, Netherlands; 2000.
91. Pereira SL, Baker AJ: **A molecular timescale for galliform birds accounting for uncertainty in time estimates and heteroge-**

neity of rates of DNA substitutions across lineages and sites.  
*Mol Phylogenet Evol* 2006, **38**:499-509.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

