Published in final edited form as:

Gene. 2015 June 1; 563(2): 180–189. doi:10.1016/j.gene.2015.03.027.

Lineage-specific loss of *FGF17* within the avian orders Galliformes and Passeriformes

John Abramyan*

Life Sciences Institute, University of British Columbia, Vancouver BC, CANADA

Abstract

The genomic and developmental complexity of vertebrates is commonly attributed to two rounds of whole genome duplications which occurred at the base of the vertebrate radiation. These duplications led to the rise of several, multi-gene families of developmental proteins like the Fibroblast growth factors (FGFs); a signaling protein family which functions at various stages of embryonic development. One of the major FGF assemblages arising from these duplications is the FGF8 subfamily, which includes *FGF8*, *FGF17*, and *FGF18* in tetrapods. While *FGF8* and *FGF18* are found in all tetrapods and are critical for embryonic survival, genomic analyses suggest putative loss of *FGF17* in various lineages ranging from frogs and fish, to the chicken. This study utilizes 27 avian genomes in conjunction with molecular analyses of chicken embryos to confirm the loss of *FGF17* in chicken as a true, biological occurrence. *FGF17* is also missing in the turkey, black grouse, Japanese quail and the northern bobwhite genomes. These species, along with chicken, form a monophyletic clade in the order Galliformes. Four additional species, members of the clade Passeroidea, within the order Passeriformes, are also missing *FGF17*. Additionally, analysis of intact *FGF17* in other avian lineages reveals that it is still under strong purifying selection, despite being seemingly dispensable. Thus, *FGF17* likely represents a molecular spandrel arising from a genome duplication event and due to its high connectivity with *FGF8/FGF18*, and potential for interference with their function, is retained under strong purifying selection, despite itself not having a strong selective advantage.

Keywords

FGFD; FGF8; FGF18; gene loss; genome duplication; bird genome

1. Introduction

In metazoans (kingdom Animalia), embryonic development involves the organization of cells to undergo morphogenesis and subsequently, organogenesis. These processes take

© 2015 Published by Elsevier B.V

*Author for Correspondence: John Abramyan, Life Sciences Institute, UBC, 2350 Health Sciences Mall, Vancouver, BC, V6T 1Z3, CANADA, Phone: 604 822 0749, abramyan@dentistry.ubc.ca.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

place through a variety of molecular signals from gene families such as fibroblast growth factors (FGFs – Itoh and Ornitz, 2004), bone morphogenetic proteins (BMPs – Bragdon et al., 2011), and wingless-related MMTV integration sites (WNTs – Niehrs, 2012). In vertebrates, arguably the most complex metazoan group, these gene families often display extensive duplication and diversification, resulting in large, multi-gene families. Vertebrate developmental complexity is often cited as being the result of such gene expansions (reviewed by Cañestro et al., 2007, 2013). These proteins function as molecular signals in critical cell-cell/tissue communication in order to direct the formation of the developing embryo. In addition to large copy numbers (~19 WNTs -Niehrs, 2012; ~18 BMPs - Bragdon et al., 2011; ~ 22 FGFs – Itoh and Ornitz, 2004; Popovici et al., 2005), these gene families display an exceptional level of intraspecific homology and functional overlap. These attributes implicate them as having arisen through numerous gene duplications followed by functional diversification (Hughes, 2002).

Duplication events can occur either as small-scale duplications (SSD), which involve individual genes or chromosomal regions, or as whole-genome duplications (WGD) (Conant and Wolfe, 2008). In his canonical work *Evolution by Gene Duplication* (Ohno, 1970), Susumu Ohno established the idea of the successful vertebrate diversification and innovation being facilitated by two rounds of whole-genome duplications (2R–WGD) which are now presumed to have occurred around the cephalochordate - vertebrate divergence ~ 600 – 550 million years ago (mya) (Holland et al., 1994; Makalowski, 2001; Garcia-Fernandez, 2005). In fact, paralogous genes which have duplicated via WGDs are called “ohnologs” in his honor (Wolfe, 2000). As a result of the 2R–WGD, phylogenetic analyses of typical gene families in vertebrates often reveal a “four-to-one” relationship to a single invertebrate proto-ortholog (Feiner et al., 2013).

FGFs are a particularly good example of a gene family which has undergone the abovementioned processes on several occasions (Ornitz and Itoh, 2001; Van de Peer et al., 2001; Wotton et al., 2008), from at least 22 members (divided into 8 subfamilies) in vertebrates to 3 *Fgf*-related genes in *Drosophila melanogaster* and 2 in *Caenorhabditis elegans* (Itoh and Ornitz, 2004; Popovici et al., 2005). Members of FGF family are distantly positioned within the genome, indicating that they were generated by WGDs (Itoh and Ornitz, 2004; Popovici et al., 2005), and not by SSDs, which would result in tandem positioning within a chromosome (Abramyan and Stajich, 2012). Additionally, since they are products of WGDs, the 22 members of the FGF family can be grouped into subfamilies based on greater sequence similarity due to certain members sharing a more recent common ancestor than others (Itoh and Ornitz, 2004; Popovici et al., 2005).

FGFs are secreted proteins performing diverse roles during embryonic development ranging from patterning-related functions such as regulating cell proliferation, migration and differentiation (Xu et al., 2000), to the development of skeletal elements such as the perichondrium (Liu et al., 2002) and bones (Xu et al., 1999). Furthermore, they function through the binding of specific receptor tyrosine kinases (fibroblast growth factor receptors (*Fgfrs*) 1 – 4), and by necessity, due to the small number of receptors, each category of *Fgfr* can bind several different FGF ligands (Basilico and Moscatelli, 1992; Reifers et al., 1998; Popovici et al., 2005). Their ability to bind the same receptors, coupled with often

overlapping patterns of expression, results in functional redundancy between FGFs (Ornitz and Itoh, 2001). In turn, loss-of-function mutations often display weaker phenotypes than anticipated (Reifers et al., 1998).

This aforementioned redundancy is demonstrated particularly well by the members of the FGF8 subfamily (also referred to as the FGF-D Family by Cañestro et al., 2007 and Popovici et al., 2005), which is comprised of *Fgf8*, *Fgf17*, and *Fgf18* in amniotes (Xu et al., 1999; Ornitz and Itoh, 2001). Members of the FGF8 subfamily exhibit high sequence identity and are involved in early craniofacial development as well as patterning of the midbrain and cerebellum through regulating the midbrain (mesencephalon) - hindbrain (metencephalon) junction, (Maruoka et al., 1998; Reifers et al., 1998; Xu et al., 2000; Liu and Joyner, 2001; Ornitz and Itoh, 2001). All three ligands also share similar receptor specificity towards *Fgfr1–3* and display overlapping expression patterns in several tissues, with slight spatiotemporal differences (Xu et al., 1999; Xu et al., 2000). Their overlap in expression, and their ability to bind the same receptors, suggests that these FGFs may be redundant in function and explains the cumulative affects on target tissues, as reported by previous studies (Xu et al., 1999; Xu et al., 2000).

In addition to similarity in sequence and spatiotemporal expression, the FGF8 subfamily members also display extensive synteny of their respective chromosomal regions, as described in detail by Wotton et al., (2008). This pattern suggests that their appearance likely resulted from a 2R–WGD event (Itoh and Ornitz, 2004; Popovici et al., 2005; Cañestro et al., 2007). The ancestral chromosome containing the proto *Fgf8/17/18/24* gene was duplicated in one WGD to create the proto *Fgf8/17* and proto *Fgf18/24* genes. These genes duplicated yet again in a second WGD, giving rise to the four genes found in vertebrates (*Fgf8*, *Fgf17*, *Fgf18*, and *Fgf24*) (Fig. 1). *Fgf24* was subsequently lost in tetrapods (Popovici et al., 2005; Cañestro et al., 2007; Wotton et al., 2008; Jovelín et al., 2010).

Post-duplication, limited functional diversification has occurred within the FGF8 subfamily members. While both *Fgf8* and *Fgf18* are critical for embryonic survival (Meyers et al., 1998; Ohbayashi et al., 2002), *Fgf17* has been cited as missing from the chicken genome (Wotton et al., 2008) as well as *Xenopus tropicalis* (Wotton et al., 2008) and medaka fish (Cañestro et al., 2007). The idea of *Fgf17* loss is not entirely surprising considering that mouse knockout studies show minimal effect on the organism. *Fgf17* mutant mice are born alive and phenotypically normal, except for mild defects observed in the cerebellum (Xu et al., 2000). However, in genome level analyses, it is not simple to distinguish between real mutations from genomic assembly artifacts. Thus, a thorough examination to assess whether *FGF17* is truly lost in the chicken, or whether it is a byproduct of sequencing mis-assembly, was undertaken.

In this study, investigation of 27 avian genomes, including all currently sequenced Passeriformes and Galliformes, reveals that *FGF17* was lost within two monophyletic lineages, supporting the loss of *FGF17* as a true biological phenomenon. This result is further supported through a molecular analysis in which Reverse-Transcriptase PCR (RT-PCR) detected *FGF17* expression both in mouse and pigeon, while none was detected in

chicken embryos at 5 developmental stages. Furthermore, maximum likelihood analysis for selection detected strong purifying selection on *FGF17*, indicating that it is still likely functional in lineages which have retained it.

2. Materials and Methods

2.1 Phylogenetic Analyses

A phylogenetic tree was produced for this study from www.birdtree.org using the “*Ericson All Species: a set of 10000 trees with 9993 OTUs each*” parameter based on (Ericson et al., 2006), comprised of 27 bird species from 18 families, 8 orders (Supplementary Table 1). From the 10,000 tree output, a consensus tree was generated using TreeAnnotator v.1.7.5 (Drummond et al., 2012). While the phylogeny assembled by Ericson et al., (2006) is based on five genomic regions comprised of a mixture of exons and introns, other independent studies using introns and exons from 19 independent loci (Hackett et al., 2008), retrotransposons (Suh et al., 2011), and whole genome data (Jarvis et al., 2014) all agree with this tree topology.

27 genomes (Supplementary Table 1) were downloaded from GenBank (Benson et al., 2013) and searched for *FGF8*, *FGF17*, *FGF18* using stand-alone, command line BLAST (Basic Local Alignment Tool) v 2.2.18 (Altschul et al., 1990). Of the 27 species analyzed, *FGF17* was detected in 18 species, of which 5 were incomplete. Incomplete sequence of *FGF17* was detected in the Collared flycatcher (*Ficedula albicollis*) (376bp of 648bp), White-tailed eagle (*Haliaeetus albicilla*) (178bp of 648bp), Hooded crow (*Corvus cornix*) (453bp of 648bp), Golden-collared manakin (*Manacus vitellinus*) (178bp of expected 648bp), and Adélie penguin (*Pygoscelis adeliae*) (178bp of 648bp).

Gene phylogeny was generated as described by Abramyan and Stajich, (2012). Briefly, coding sequence (CDS) alignments were generated using ClustalW (Thompson et al., 1994) and analyzed using a Bayesian Markov Chain Monte Carlo (MCMC) method implemented in BEAST v. 1.7.5 (Drummond et al., 2012). The program uses an MCMC method to sample tree and parameter space and can be summarized to generate a consensus tree with posterior probabilities for each clade representing confidence in the presence of the clade from the sampled phylogenies. The analysis was run with a minimum of 10 million generations, sampling every 100th tree. Subsequently, a consensus tree was generated using TreeAnnotator v.1.7.5 from the BEAST software package (Drummond et al., 2012) and viewed using FigTree v. 1.3.1 (<http://tree.bio.ed.ac.uk/software/figtree/>). Human *FGF21* was used as an outgroup. Both bird phylogeny and gene phylogeny figures were traced and modified using Adobe Illustrator CS4 (Adobe Systems).

2.2 Reverse Transcriptase PCR

Reverse transcriptase-polymerase chain reaction (RT-PCR) was performed on embryonic day (E) 13.5 mouse, stage 32 pigeon (stage according to Hamburger and Hamilton, 1951), as well as chicken embryonic tissues (stage 15 (whole body; n=2), stage 20 (whole body; n=1), stage 25 (head only; n=1), stage 29 (head only; n=1), stage 32 (head only; n=1) (stage according to Hamburger and Hamilton, 1951). RNA and cDNA were prepared from tissues

as described by Abramyan et al., (2009). Primers were designed from an alignment of avian *FGF17*, *FGF8*, and *FGF18* to ensure that the primers were specific to *FGF17* (FGF17-Forward: CCGTCTCCTAATTTTAACCA; Fgf17-Reverse: CAGATGTACTTCTCGCTCTC). cDNA fidelity was tested with primers specific to 18S ribosomal RNA (18S-Forward: GGAGGTAGTGACGAAAAATAAC; 18S-Reverse: GGAACACGACGGTATCTGA). DNA amplicons were gel extracted using QIAquick gel extraction kit (Qiagen, Canada) and sequenced by Genewiz, Inc. (www.genewiz.com). PCR products were confirmed to be *Fgf17* in both mouse and pigeon by using BLAST analysis in NCBI GenBank database (Benson et al., 2013).

2.3 Test for selection

Rates of synonymous (dS - silent) and nonsynonymous (dN - amino acid replacement) substitutions were analyzed within lineages of birds as well as for individual genes. The dN/dS ratio (ω) measures selective pressure on amino acids. An ω estimate significantly less than 1 (dN < dS) implies purifying selection, whereas an estimate of 1 (dS = dN) suggests neutral evolutionary pressure. Lastly, positive selection is indicated if estimated ω is greater than 1 (dN > dS). In other words, ω will exceed 1 if an amino acid change is selectively advantageous.

To test for selection on *FGF17*, a tree-based likelihood approach was implemented as described by Yang, (1998b). An unrooted version of the consensus tree depicted in Figure 2 was utilized for this analysis. Codon alignments were generated by creating an amino acid sequence alignment in ClustalW (Thompson et al., 1994) and subsequently converted to coding sequence, codon based alignment using PAL2NAL (Suyama et al., 2006 - <http://www.bork.embl.de/pal2nal/>). Codon alignment files as well as corresponding tree files were analyzed in the CodeML program in PAML v. 4.4 (Yang, 2007).

Within CodeML, branch model analyses were used to estimate selection along specific lineages/branches. The Free-ratio model is the most general, parameter rich model and allows for a different ω values for each branch. The One-ratio model is the simplest model and assumes the same ω for all branches. The Two-ratio models allows for two ω values, allowing a clade or branch of the tree (foreground) to differ from the average ω across the branches in the rest of the tree (background branches). Likelihood estimates assume the codon substitution model of Goldman and Yang (1994). The likelihood estimates for each were compared using a hierarchical likelihood ratio test (LRT) of twice the difference in the log likelihood value of the models being compared (2 $\ln$), with the result approximating a chi square distribution (Yang, 1998a). Only > 90% complete coding sequences were used for these analyses.

2.4 Chromosome synteny

Chromosomal synteny was depicting as an alignment of chicken (chromosome 22), zebra finch (chromosome 22), human (chromosomes 8, 20 and 2), western painted turtle (*Chrysemys picta*) (scaffolds JH584511 and JH584489; GenBank: AHGY00000000.2; Shaffer et al., 2013) and American alligator (*Alligator mississippiensis*) (scaffolds 11658, 7655, 5002, 354, 3606, 12563, 3707, 7441, 5615; GenBank: AKHW00000000.1; Green et

al., 2014). Scaffolds positions in alligator are arranged according to the chicken since they are each other's closest phylogenetic relatives. Turtle scaffolds are subsequently arranged according to the alligator and chicken alignment. Human chromosomes are arranged, in turn, to all of the reptiles. Genes were identified in each genome using BLAST v 2.2.18 (Altschul et al., 1990) in the same manner as described for FGF genes and subsequently aligned and illustrated using Adobe Illustrator CS4 (Adobe Systems).

3. Results

3.1 Bird Phylogeny

Since the first avian genome was published by the Chicken Genome Sequencing Consortium (2004), a large number of bird species have been sequenced, including a recent “flock” of 48 genomes representing all major extant avian clades (Zhang et al., 2014), providing an ideal sample set to assess whether the absence of *FGF17* is common to all birds or whether it is specific to the chicken. Of the 27 genomes analyzed in this study (Supplementary Table 1), *FGF17* was missing in five members in the order Galliformes: chicken, turkey, Japanese quail, black grouse, and northern bobwhite, as well as four species from the clade Passeroidea (Cracraft and Barker, 2009), nested within the order Passeriformes: zebra finch (family – Estrildidae), white-throated sparrow (family – Emberizidae), medium ground finch (family – Thraupidae), and Atlantic canary (family – Fringillidae) (Fig. 2). This non-random pattern of *FGF17* loss is indicative of at least two independent loss events, and furthermore, lends strong support to the loss being a true biological phenomenon.

3.2 RT-PCR

In addition to *in silico* analyses, molecular confirmation of the loss of *FGF17* in chicken was obtained through RT-PCR on 5 stages of chicken embryos with stage 32 pigeon embryo as well as E13.5 mouse embryo cDNA used as positive controls. PCR analysis revealed a band of expected size (~250bp) in both mouse and pigeon, while no band was visible in the chicken at any of the stages analyzed (Fig. 3). This result confirms that the loss of *FGF17* in chicken is indeed biological event rather than a byproduct of sequencing or assembly error. RT-PCR analysis also confirms that *FGF17* is still expressed in pigeon, a species of bird with an intact ortholog.

3.3 Gene Phylogeny

Of the 27 bird genomes analyzed, *FGF17* was detected in 18 species (13 complete and 5 partial – see section 2.3 for details). To check whether the avian orthologs of *FGF17* were identified correctly, a Bayesian phylogenetic analysis was conducted on orthologs of all three FGF8 subfamily members from birds, as well as orthologs from turtle (*C. picta*), alligator (*A. mississippiensis*) and human. All three genes were monophyletic with respect to each (Fig. 4). Firstly, this analysis confirms that the putative *FGF17* sequences from the avian genomes are correctly identified and are monophyletic with their non-avian orthologs with strong support (Bayesian posterior probabilities (Bpp) = 1). Of further note were the relatively longer branches in the *FGF17* clade as compared to *FGF8* and *FGF18*. This feature could indicate a higher rate of mutation, likely due to relaxed selection on the gene. Therefore, an analysis of synonymous and non-synonymous mutation rates was undertaken,

using a maximum likelihood (ML) approach (Yang, 2007), to test for relaxed selection on *FGF17* as compared to *FGF8* and *FGF18*.

3.4 Tests for selection in the FGF8 subfamily

In order to test for relaxed selection on *FGF17*, two types of analyses were performed on a dataset comprised only of avian sequences. The first analysis tests only the branch leading to each of the three FGF clades (branch analysis) while the second analysis tests the leading branch as well as all of the branches within each clade/gene (clade analysis) (illustrated in Supplementary Table 2). A series of ML branch-based models of selection were applied to the dataset comprised of *FGF8*, *FGF17*, and *FGF18*, in the CodeML program in PAML v. 4.7 (Yang, 2007). First, the Free-ratio model, which assumes different ω for each branch in an unrooted phylogenetic tree, was applied to the dataset (log-likelihood value (l_1) = -6498.9766) and compared to the One-ratio model (Model A), which assumes the same ω value for all branches in the tree (log-likelihood value (l_0) = -6650.1007) (Table 1). To test whether the free-ratio model fits the dataset better than the one ratio model, a LRT was applied [$2 \ln(1 - l_0) = 302.2481$] and compared to a χ^2 distribution with a $df = 122$. The difference between the two models is highly significant ($P < 0.01$) indicating that the Free-ratio model fits the data better than the One-ratio model and that branches within the trees are indeed evolving at different rates (Table 1; Table 2 - *Model A vs. Free*).

Subsequently, branch analysis was applied to assess selection on the branches immediately after the split from the ancestral lineage (for example, at the split between *FGF 17* and *FGF8*). LRT was applied to models A – E (Tables 1 and 2; Supplementary Table 2) ranging from all branches having the same ω value (model A – One-ratio) to each of the branches leading to each gene being in the foreground independent of each other as well as the background ω_0 (Model E – Four-ratio). Model E still proved to best fit the dataset when compared to Models B and D, which have *FGF8* in the foreground (Table 2). However, when a LRT was applied to Model C, where the branch leading to *FGF17* was labeled as foreground, it fit the data significantly better than either models A or E, indicating a different ω value for the *FGF17* branch (Table 2). Interestingly, under Model C, $\omega = 0.0021$ for the foreground *FGF17* branch while background $\omega = 0.0676$, indicating particularly strong purifying selection for the *FGF17* branch (Table 1). This result was contrary to what would be expected since *FGF17* was hypothesized to have relaxed selection pressure based on the long branches and lineage specific losses observed.

A second set of analyses, the clade analyses, tested whether the distinct ω value observed for the branch leading to *FGF17* has continued throughout the divergence of *FGF17* with avian speciation events. To perform these analyses, the branch leading to the clade, as well as all internal branches, were labeled as foreground for each model. However, when LRT was applied to the various models (Supplementary Table 2), model C no longer fit the data best, but rather Model B, where the branches of the *FGF8* clade were in the foreground and exhibited a significant difference in ω ($\omega_{f8} = 0.0928$) as compared to the background branches consisting of all *FGF18* and *FGF17* ($\omega_0 = 0.05187$) (Supplementary Table 2). These results indicate that while there was differential selection on the branch leading to *FGF17* initially, it was subsequently not maintained across all branches. More importantly,

both analyses indicate that all three genes are still under strong selective pressure, judging from the very low ω values, characteristic of functional genes.

3.5 Tests for ω variation in FGF8 and FGF18 for lineages with FGF17 loss

Study of *FGF17* null mice shows that when one copy of *FGF8* is also lost in mutant animals, the mice display a more severe phenotype (Xu et al 2000), thus *FGF8* (or possibly *FGF18*) may be able to functionally replace *FGF17* in birds which have lost it. In order to test this theory, a series of ML branch-based models of selection were once again applied to datasets comprised of *FGF8* or *FGF18* only, with the aim of detecting a shift in the ω value within the Galliformes (**g**) or Passeroidea (**p**) as compared to the taxa which have retained *FGF17*.

For the *FGF8* analysis, only the branches leading to the **g** or **p** was labeled first (Table 3 – Branch Analysis), followed by labeling of all of the branches within the clade stemming from that branch (Table 3 – Clade Analysis). When a LRT was applied to the various models in the branch analysis, Model B ($\omega_0 = \omega_g, \omega_p$), where the **p** clade was in the foreground and all other branches (including **g**) were in the background, did not fit the data better than Model A (One ratio) (LRT = 5.4E-05, P=54.2877, df=1) (Supplementary Table 2). However, Model C ($\omega_0 = \omega_p, \omega_g$) where **g** is in the foreground and Model D ($\omega_0, \omega_p = \omega_g$), where both **p** and **g** are in the foreground together, fit the data significantly better than either models A (One-ratio) or E (Three-ratio) (Table 4). In the clade analysis, Model D proved to best fit the dataset (Table 4). These analyses imply that for *FGF8*, ω is significantly smaller (signifying stronger purifying selection) within the lineages exhibiting *FGF17* loss, particularly within Galliformes. The same analysis applied to *FGF18* showed no significant differences between the **g** or **p** clades and all other avian lineages (Supplementary Table 2).

Lastly, when all three of the FGF8 subfamily members were analyzed under the Free-ratio model, ML estimates of the number of synonymous and nonsynonymous substitutions across all branches showed that *FGF17*, while not exhibiting a substantial number of non-synonymous substitutions, did exhibit a higher number of total substitutions per site (average of 6.421 nonsynonymous and 21.12 synonymous across all branches in the tree) compared to *FGF8* (3.783 non-synonymous and 7.985 synonymous) and *FGF18* (1.116 nonsynonymous and 5.036 synonymous) (Supplementary Table 3; Fig. 4B). These results indicate that while the gene is under strong purifying selection, it is on average amassing a higher number of substitutions than either *FGF8* or *FGF18*, which would explain the long branches observed in Figure 4. Taking into account that birds seem to exhibit either an intact *FGF17* or a complete loss, in conjunction with very low ω estimates, both deleterious substitutions as well as amino acid replacement seem to be strongly selected against.

3.6 Chromosome Synteny and FGF17 loss in chicken

In order to characterize the loss of *FGF17* from the avian chromosome 22, an analysis of chromosomal synteny was performed in comparison to human chromosome 8 in conjunction with an assembly of syntenic scaffolds from alligator and turtle, which contain *FGF17* and neighboring genes (Fig. 5). Between turtle, human and archosaurs, several inversions were

observed. Additionally, *FGF17* was not the only gene missing, but rather what appears to be an ~1.2 megabase (mb) gap when measured from the beginning of *Fgf17* in human chromosome 8 until the beginning of *Chmp7*. This region encodes ~35 genes in the human genome and would likely encode a similar number in reptiles.

In order to assess whether genes neighboring *FGF17* were also lost or had translocated to different parts of the genome, 7 of these genes from the “missing” region were searched for in all 9 of the specimens which exhibited *FGF17* loss as well as a subset of 9 bird species which have retained *FGF17*. All genes were also identified in turtle (*C. picta*), alligator (*A. mississippiensis*) and human. In chicken, only *DMTN* (location -chrUn_JH375617) and *LGI3* were identified. Of the other species analyzed, *LGI3* was identified in zebra finch, which is located on yet another unassembled scaffold (chrUn_EQ842673). Several of the other genomes with missing *FGF17* did seem to have retained the neighboring genes, however, since they are not assembled into large enough scaffolds, their synteny and location within the genome remain unknown (Supplementary Table 4). This analysis confirms disintegration of this chromosomal region in lineages where *FGF17* was lost.

4. Discussion

Gene duplication and subsequent functional diversification is one of the most important mechanisms for the evolution of novel genes and functions within a genome (Kimura and Ohta, 1974; Zhang et al., 1998). Indeed much of the vertebrate genome is thought to be the result of an intricate mixture of both WGDs and SSDs, which have provided vertebrates with the genetic foundation to evolve the complexity we observe today (Cañestro et al., 2007; Cañestro et al., 2013). However, not all genes are retained after a duplication event, with some being lost even after millions of years of retention, as is the case with the conspicuous absence of the *FGF17* from several avian genomes. This study examines the lineage-specific loss of *FGF17* in two avian clades and examines the circumstances under which it is lost in some lineages while retained in others.

4.1 Bioinformatic and molecular analyses of *FGF17* loss in birds

Of the 27 avian species examined in this study, *FGF17* was identified in 18 genomes (13 complete sequence and 5 incomplete sequence). However, even within the incomplete sequences, the portions that were identified exhibited no substantial substitutions, nor any insertions or deletion (indels) causing frame shifts in the coding sequence. This indicates that partial identification was a sequence assembly byproduct rather than a true deletion or pseudogenization of *FGF17* in these species. In the 9 species which exhibit a complete loss of *FGF17*, this study provides overwhelming evidence that this loss is indeed a true biological event. Furthermore, several species exhibiting *FGF17* loss, such as the Atlantic canary, were added to the study after the recent release of 45 newly sequenced genomes (Zhang et al., 2014), which only served to strengthen the conclusions reached.

When a gene fails to map to a genome assembly, it is difficult to distinguish between true gene loss and an unresolved state due to low quality or missing sequencing product (Hiller et al., 2012). This problem is often exacerbated by the use of previously mis-assembled genomes to assemble new genome sequences. However, this does not seem to be the case

with the avian *FGF17*. While the zebra finch and Japanese quail were partly assembled using the chicken genome as a reference, the northern bobwhite (Halley et al., 2014) and the turkey (Dalloul et al., 2010), as well as several of the other species missing *FGF17*, were assembled *de novo*. In fact, the manuscript introducing the turkey genome describes it as the first eukaryotic genome completely sequenced and assembled *de novo* (Dalloul et al., 2010). Thus, it seems unlikely that genome mis-assembly has led to the pattern of *FGF17* loss described here. Additionally, the distant phylogenetic relationship of Galliformes and Passeriformes, in conjunction with the fact that species which have lost *FGF17* group into two monophyletic clades, indicates that the loss of *FGF17* occurred at least twice in birds. It is certainly possible that the pattern of *FGF17* loss and retention described here is much more complex and more losses within other clades have occurred.

Additionally, despite bioinformatic analysis revealing the presence of a gene within a genome, its functionality is only inferred. Thus, an RT-PCR analysis using cDNA from five developmental stages of embryonic chicken, with pigeon and mouse as positive controls, further confirmed this loss to be a real biological phenomenon. While *FGF17* is not expressed at embryonic stages of chicken development, it is present and active in the pigeon genome.

Lastly, while this study has focused on *FGF17* itself, synteny analysis revealed a large region ~1.3 mb which was missing from the chicken and zebra finch chromosomes 22, despite being present in crocodilians, turtles and human. Moreover, when a select number of genes from this region were searched for in the genomes of birds missing *FGF17*, most of the genes were missing, while almost all of the genes were identified in birds with intact *FGF17*. This pattern suggest that this region was not deleted entirely in all birds, but is likely a fragile chromosomal site and the genes found within this region have moved onto different chromosomes through various recombination processes. Further characterization of this region will have to wait until more genomes are assembled into chromosomes.

4.2 Developmental consequences of *FGF17* loss

Studies have shown that a significant proportion of the developmental tool kit can be lost without major changes to the body plan (Cañestro et al., 2007; Holland, 2007). The differential loss of ohnologs in particular has been implicated as a potential mechanism of lineage divergence (Postlethwait, 2007). The identification of *FGF17* loss as being a true biological event, begs the question of how these lineages cope without it. In the mouse, when the *Fgf8* or *Fgf18* are rendered non-functional through genetic manipulation, embryonic lethality is 100% (Meyers et al., 1998; Ohbayashi et al., 2002). Homozygous *Fgf17* null mice, on the other hand, have no gross abnormalities and generally appear healthy (Xu et al., 2000), with closer investigation revealing tissue loss specific to the caudal-most part of the midbrain, the inferior colliculus and the vermis cerebellum, responsible for maintaining motor coordination and body equilibrium (Xu et al., 2000). In birds, the cerebellum (the *corpus cerebelli*) is thought to be homologous to the vermis in mammals, thus there is no identifiable “vermis” tissue (Marzban et al., 2010). Considering the lack of a true vermis, in conjunction with their highly derived, bipedal gait, *FGF17* may not perform the same functions in birds as it does in mice.

Furthermore, when one copy of the *Fgf8* gene is mutated in *Fgf17* null mice, a more severe phenotype is observed, indicating that *Fgf17* and *Fgf8* have partially overlapping functions in development of the anterior hindbrain (Xu et al., 2000; Liu and Joyner, 2001). Studies have further revealed that duplicates tend to diverge more commonly in regulation and least commonly in their biochemical function, thereby retaining a substantial degree of functional overlap (Conant and Wolfe, 2008). Based on *in vitro* studies, the biochemical properties of the members of FGF8 subfamily are also similar (Xu et al., 1999). In fact, this pattern of FGF interaction is not uncommon and loss-of-function mutations for FGFs often display weaker phenotypes than anticipated (Reifers et al., 1998). Taking these facts into account, it is possible that *FGF8* has overtaken the roles of *FGF17* in lineages where it is lost.

4.3 Retention of intact FGF17 in select lineage

In theory, once a gene is duplicated, such as the proto*FGF8/17*, one copy should be free from selection pressure ($dS \approx dN$), which allows for mutations to be tolerated, while the intact copy can still perform the original function (Ohno, 1973). This generally either results in pseudogenization of one of the copies, through random, deleterious mutations which do not get purged (Kimura, 1973; Demuth and Hahn, 2009), or in rare cases, the evolution of novel function. Gene loss through pseudogenization is exemplified particularly well by the tooth-specific genes associated with enamel production in edentulous lineages such as birds and turtles (Sire et al., 2008; Meredith et al., 2013; Shaffer et al., 2013). However, recent studies report that null mutations accumulate at a rate that is lower than expected in the abovementioned scenario (Hughes and Hughes, 1993; Hughes, 1994). *FGF17* falls into this category in containing no discernable indels and being under strong purifying selection despite being theoretically dispensable. Why has pseudogenization not taken place in *FGF17*?

Studies of gene loss and maintenance after duplication often take an adaptationist approach in assuming that genes that are retained are adaptive for the organism (Krylov et al., 2003; Hahn, 2009; Hoffmann et al., 2011). However, *FGF17* may not have an adaptive advantage, but may simply be retained due to its high connectivity with its ohnologs. *FGF17* is a ligand with an intact promoter in animals where the gene is present in the genome (as revealed in this study by the RT-PCR analysis of pigeon). Thus, pseudogenization of the coding sequence could risk the production of a “mutant” protein which could still bind receptors and therefore, compete and interfere with normal *FGF8* and/or *FGF18* function (both critical genes for embryonic survival), due to the extensive functional and spatiotemporal overlap between these genes.

Thus, *FGF17* may simply be an unintentional byproduct, itself not being particularly adaptive, but stemming from a WGD event which did increase fitness as a whole and was retained. Subsequent to the duplication, due to the shared ligand - receptor relationship described above, having one pseudogenized and one intact copy may be more harmful than having two intact copies of a gene. This scenario would make *FGF17* a genetic version of a biological ‘spandrel’, as described by Gould and Lewontin, (1979), which cannot be purged through pseudogenization, and must be retained under strong selective pressure despite not being particularly useful or adaptive. The idea of a “molecular spandrel” has been proposed

by other authors in different contexts (Van de Peer et al., 2009; Barrett and Hoekstra, 2011), but this study describes a specific incidence of this process.

5. Conclusion

The results of this study reveal that *FGF17* has been lost in at least two avian lineages. In birds where *FGF17* is still retained, it is still intact and still expressed. Selection analysis further shows that *FGF17* is still under strong purifying selection within the avian genome, indicating that mutations within the coding sequence are deleterious and selected against. This forms an interesting circumstance where a gene is theoretically “dispensable” but, if it is present, it is kept intact and functional. Further investigation of this phenomenon is warranted. A theory stemming from this study suggests that genes like *FGF17* are genomic ‘spandrels’ that do not necessarily increase fitness themselves, but are simply byproducts from WGD events. Once present in the genome, due to its high functional overlap with other, more critical FGFs such as *FGF8* and *FGF18*, *FGF17* pseudogenization risks interfering with the function of its sister genes, thus it is kept under strong purifying selection until the entire gene is deleted. This scenario would create the all-or-nothing pattern of *FGF17* presence observed in the genomes of the class Aves.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The author would like to thank the laboratories of Joy Richman for chicken cDNA and molecular reagents (funded by the Canadian Institutes of Health Research award no: MOP-123536), the laboratory of Jane Roskams for molecular reagents, Aaron Bogutz and the laboratory of Louis Lefebvre for mouse embryonic cDNA, and Mr. Valeri Bogdanov of the Canadian Pigeon Fanciers Association for donation of pigeon tissue. The author would also like to thank Mr. Zack Demirtshyan for illustrations of chicken, turtle, alligator, zebra finch and human included in Figure 5. The Shaffer Lab as well as Darrin Hulsey and Theresa Grieco for insightful discussions. The author would also like to thank Katherine Fu for critically reading the manuscript. JA is a recipient of a NIH *Ruth L. Kirschstein* NRSA Postdoctoral Fellowship (award no: 1 F32 DE022999-03). Additionally, the author would like to thank two anonymous reviewers whose comments substantially improved the manuscript. The funding source played no role in study design, collection, analysis, and interpretation of data or writing of this report.

Abbreviations

BMP	Bone morphogenetic protein
Bpp	Bayesian posterior probabilities
cDNA	DNA complementary to RNA
CDS	coding sequence
FGF	Fibroblast Growth Factors
FGFR	Fibroblast growth factor receptors
indels	insertions/deletion
LRT	likelihood ratio test

mb	megabase
MCMC	Markov chain Monte Carlo
ML	maximum likelihood
mya	million years ago
RT-PCR	Reverse transcriptase-polymerase chain reaction
SSD	small-scale duplications
WGD	whole genome duplication
WNT	Wingless-related MMTV integration sites
2R-WGD	two rounds of whole-genome duplications

REFERENCES

- Abramyan J, Feng CW, Koopman P. Cloning and expression of candidate sexual development genes in the cane toad (*Bufo marinus*). *Dev. Dyn.* 2009; 238:2430–2441. [PubMed: 19653320]
- Abramyan J, Stajich JE. Species-specific chitin-binding module 18 expansion in the amphibian pathogen *Batrachochytrium dendrobatidis*. *MBio*. 2012; 3:e00150–E00112. [PubMed: 22718849]
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990; 215:403–410. [PubMed: 2231712]
- Barrett RD, Hoekstra HE. Molecular spandrels: tests of adaptation at the genetic level. *Nat Rev Genet.* 2011; 12:767–780. [PubMed: 22005986]
- Basilico C, Moscatelli D. The FGF family of growth factors and oncogenes. *Adv Cancer Res.* 1992; 59:115–165. [PubMed: 1381547]
- Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res.* 2013; 41:D36–D42. [PubMed: 23193287]
- Bragdon B, Moseychuk O, Saldanha S, King D, Julian J, Nohe A. Bone morphogenetic proteins: a critical review. *Cell Signal.* 2011; 23:609–620. [PubMed: 20959140]
- Cañestro C, Albalat R, Irimia M, Garcia-Fernandez J. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin Cell Dev Biol.* 2013; 24:83–94. [PubMed: 23291262]
- Cañestro C, Yokoi H, Postlethwait JH. Evolutionary developmental biology and genomics. *Nat Rev Genet.* 2007; 8:932–942. [PubMed: 18007650]
- Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. *Nat Rev Genet.* 2008; 9:938–950. [PubMed: 19015656]
- Consortium ICGS. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature.* 2004; 432:695–716. [PubMed: 15592404]
- Cracrafta, J.; Barkerb, FK. Passerine birds (Passeriformes). In: Hedges, SB.; Kumar, S., editors. *The timetree of life*. Oxford University Press New York; 2009.
- Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K, Blomberg Le A, Bouffard P, Burt DW, Crasta O, Crooijmans RP, Cooper K, Coulombe RA, De S, Delany ME, Dodgson JB, Dong JJ, Evans C, Frederickson KM, Flicek P, Florea L, Folkerts O, Groenen MA, Harkins TT, Herrero J, Hoffmann S, Megens HJ, Jiang A, de Jong P, Kaiser P, Kim H, Kim KW, Kim S, Langenberger D, Lee MK, Lee T, Mane S, Marcas G, Marz M, McElroy AP, Modise T, Nefedov M, Notredame C, Paton IR, Payne WS, Perte G, Prickett D, Puiu D, Qioa D, Raineri E, Ruffier M, Salzberg SL, Schatz MC, Scheuring C, Schmidt CJ, Schroeder S, Searle SM, Smith EJ, Smith J, Sonstegard TS, Stadler PF, Tafer H, Tu ZJ, Van Tassell CP, Vilella AJ, Williams KP, Yorke JA, Zhang L, Zhang HB, Zhang X, Zhang Y, Reed KM. Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS Biol.* 2010:8.

- Demuth JP, Hahn MW. The life and death of gene families. *Bioessays*. 2009; 31:29–39. [PubMed: 19153999]
- Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol*. 2012; 29:1969–1973. [PubMed: 22367748]
- Ericson PG, Anderson CL, Britton T, Elzanowski A, Johansson US, Kallersjo M, Ohlson JI, Parsons TJ, Zuccon D, Mayr G. Diversification of Neoaves: integration of molecular sequence data and fossils. *Biol Lett*. 2006; 2:543–547. [PubMed: 17148284]
- Feiner N, Murakami Y, Breithut L, Mazan S, Meyer A, Kuraku S. Saltatory evolution of the ectodermal neural cortex gene family at the vertebrate origin. *Genome Biol Evol*. 2013; 5:1485–1502. [PubMed: 23843192]
- Garcia-Fernandez J. The genesis and evolution of homeobox gene clusters. *Nat Rev Genet*. 2005; 6:881–892. [PubMed: 16341069]
- Goldman N, Yang Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol*. 1994; 11:725–736. [PubMed: 7968486]
- Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci*. 1979; 205:581–598. [PubMed: 42062]
- Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandeweghe MW, St John JA, Capella-Gutierrez S, Castoe TA, Kern C, Fujita MK, Opazo JC, Jurka J, Kojima KK, Caballero J, Hubley RM, Smit AF, Platt RN, Lavoie CA, Ramakodi MP, Finger JW Jr, Suh A, Isberg SR, Miles L, Chong AY, Jaratlerdsiri W, Gongora J, Moran C, Iriarte A, McCormack J, Burgess SC, Edwards SV, Lyons E, Williams C, Breen M, Howard JT, Gresham CR, Peterson DG, Schmitz J, Pollock DD, Haussler D, Triplett EW, Zhang G, Irie N, Jarvis ED, Brochu CA, Schmidt CJ, McCarthy FM, Faircloth BC, Hoffmann FG, Glenn TC, Gabaldon T, Paten B, Ray DA. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science*. 2014; 346:1254449. [PubMed: 25504731]
- Hackett SJ, Kimball RT, Reddy S, Bowie RC, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han KL, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T. A phylogenomic study of birds reveals their evolutionary history. *Science*. 2008; 320:1763–1768. [PubMed: 18583609]
- Hahn MW. Distinguishing among evolutionary models for the maintenance of gene duplicates. *J Hered*. 2009; 100:605–617. [PubMed: 19596713]
- Halley YA, Dowd SE, Decker JE, Seabury PM, Bhattarai E, Johnson CD, Rollins D, Tizard IR, Brightsmith DJ, Peterson MJ, Taylor JF, Seabury CM. A draft de novo genome assembly for the northern bobwhite (*Colinus virginianus*) reveals evidence for a rapid decline in effective population size beginning in the Late Pleistocene. *PLoS One*. 2014; 9:e90240. [PubMed: 24621616]
- Hamburger V, Hamilton HL. A series of normal stages in the development of the chick embryo. *J Morphol*. 1951; 88:49–92. [PubMed: 24539719]
- Hiller M, Schaar BT, Indjeian VB, Kingsley DM, Hagey LR, Bejerano G. A “forward genomics” approach links genotype to phenotype using independent phenotypic losses among related species. *Cell Rep*. 2012; 2:817–823. [PubMed: 23022484]
- Hoffmann FG, Opazo JC, Storz JF. Differential loss and retention of cytoglobin, myoglobin, and globin-E during the radiation of vertebrates. *Genome Biol Evol*. 2011; 3:588–600. [PubMed: 21697098]
- Holland LZ. Developmental biology: a chordate with a difference. *Nature*. 2007; 447:153–155. [PubMed: 17495912]
- Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. Gene duplications and the origins of vertebrate development. *Dev Suppl*. 1994:125–133. [PubMed: 7579513]
- Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci*. 1994; 256:119–124. [PubMed: 8029240]
- Hughes AL. Adaptive evolution after gene duplication. *Trends Genet*. 2002; 18:433–434. [PubMed: 12175796]
- Hughes AL, Hughes MK. Adaptive evolution in the rat olfactory receptor gene family. *J Mol Evol*. 1993; 36:249–254. [PubMed: 8483163]

- Itoh N, Ornitz DM. Evolution of the Fgf and Fgfr gene families. *Trends Genet.* 2004; 20:563–569. [PubMed: 15475116]
- Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldon T, Capella-Gutierrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MP, Prosdocimi F, Samaniego JA, Vargas Velazquez AM, Alfaro-Nunez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jonsson KA, Johnson W, Koepfli KP, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alstrom P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, et al. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science.* 2014; 346:1320–1331. [PubMed: 25504713]
- Jovelín R, Yan YL, He X, Catchen J, Amores A, Canestro C, Yokoi H, Postlethwait JH. Evolution of developmental regulation in the vertebrate FgfD subfamily. *J Exp Zool B Mol Dev Evol.* 2010; 314:33–56. [PubMed: 19562753]
- Kimura, M. The neutral theory of molecular evolution. Cambridge University Press; 1973.
- Kimura M, Ohta T. On some principles governing molecular evolution. *Proc Natl Acad Sci U S A.* 1974; 71:2848–2852. [PubMed: 4527913]
- Krylov DM, Wolf YI, Rogozin IB, Koonin EV. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Res.* 2003; 13:2229–2235. [PubMed: 14525925]
- Liu A, Joyner AL. Early anterior/posterior patterning of the midbrain and cerebellum. *Annu Rev Neurosci.* 2001; 24:869–896. [PubMed: 11520921]
- Liu Z, Xu J, Colvin JS, Ornitz DM. Coordination of chondrogenesis and osteogenesis by fibroblast growth factor 18. *Genes Dev.* 2002; 16:859–869. [PubMed: 11937493]
- Makalowski W. Are we polyploids? A brief history of one hypothesis. *Genome Res.* 2001; 11:667–670. [PubMed: 11337465]
- Maruoka Y, Ohbayashi N, Hoshikawa M, Itoh N, Hogan BL, Furuta Y. Comparison of the expression of three highly related genes, Fgf8, Fgf17 and Fgf18, in the mouse embryo. *Mech Dev.* 1998; 74:175–177. [PubMed: 9651520]
- Marzban H, Chung SH, Pezhouh MK, Feirabend H, Watanabe M, Voogd J, Hawkes R. Antigenic compartmentation of the cerebellar cortex in the chicken (*Gallus domesticus*). *J Comp Neurol.* 2010; 518:2221–2239. [PubMed: 20437525]
- Meredith RW, Gatesy J, Springer MS. Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC Evol Biol.* 2013; 13:20. [PubMed: 23342979]
- Meyers EN, Lewandoski M, Martin GR. An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat Genet.* 1998; 18:136–141. [PubMed: 9462741]
- Niehrs C. The complex world of WNT receptor signalling. *Nat Rev Mol Cell Biol.* 2012; 13:767–779. [PubMed: 23151663]
- Ohbayashi N, Shibayama M, Kurotaki Y, Imanishi M, Fujimori T, Itoh N, Takada S. FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis. *Genes Dev.* 2002; 16:870–879. [PubMed: 11937494]
- Ohno, S. Evolution by gene duplication. London, Berlin, Heidelberg, New York: Springer-Verlag; 1970.
- Ohno S. Ancient linkage groups and frozen accidents. *Nature.* 1973; 244:259–262. [PubMed: 4200792]
- Ornitz DM, Itoh N. Fibroblast growth factors. *Genome Biol.* 2001; 2:REVIEWS3005.
- Popovici C, Roubin R, Coulier F, Birnbaum D. An evolutionary history of the FGF superfamily. *Bioessays.* 2005; 27:849–857. [PubMed: 16015590]

- Postlethwait JH. The zebrafish genome in context: ohnologs gone missing. *J Exp Zool B Mol Dev Evol.* 2007; 308:563–577. [PubMed: 17068775]
- Reifers F, Bohli H, Walsh EC, Crossley PH, Stainier DY, Brand M. Fgf8 is mutated in zebrafish acerebellar (ace) mutants and is required for maintenance of midbrain-hindbrain boundary development and somitogenesis. *Development.* 1998; 125:2381–2395. [PubMed: 9609821]
- Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, Valenzuela N, Abramyan J, Amemiya CT, Badenhorst D, Biggar KK, Borchert GM, Botka CW, Bowden RM, Braun EL, Bronikowski AM, Bruneau BG, Buck LT, Capel B, Castoe TA, Czerwinski M, Delehaunty KD, Edwards SV, Fronick CC, Fujita MK, Fulton L, Graves TA, Green RE, Haerty W, Hariharan R, Hernandez O, Hillier LW, Holloway AK, Janes D, Janzen FJ, Kandathil C, Kong L, de Koning AP, Li Y, Literman R, McGaugh SE, Mork L, O’Laughlin M, Paiz RT, Pollock DD, Ponting CP, Radhakrishnan S, Raney BJ, Richman JM, St John J, Schwartz T, Sethuraman A, Spinks PQ, Storey KB, Thane N, Vinar T, Zimmerman LM, Warren WC, Mardis ER, Wilson RK. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* 2013; 14:R28. [PubMed: 23537068]
- Sire JY, Delgado SC, Girondot M. Hen’s teeth with enamel cap: from dream to impossibility. *BMC Evol Biol.* 2008; 8:246. [PubMed: 18775069]
- Suh A, Paus M, Kieffmann M, Churakov G, Franke FA, Brosius J, Kriegs JO, Schmitz J. Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat Commun.* 2011; 2:443. [PubMed: 21863010]
- Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 2006; 34:W609–W612. [PubMed: 16845082]
- 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. [PubMed: 7984417]
- Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet.* 2009; 10:725–732. [PubMed: 19652647]
- Van de Peer Y, Taylor JS, Braasch I, Meyer A. The ghost of selection past: rates of evolution and functional divergence of anciently duplicated genes. *J Mol Evol.* 2001; 53:436–446. [PubMed: 11675603]
- Wolfe K. Robustness—it’s not where you think it is. *Nat Genet.* 2000; 25:3–4. [PubMed: 10802639]
- Wotton KR, Weierud FK, Dietrich S, Lewis KE. Comparative genomics of Lbx loci reveals conservation of identical Lbx ohnologs in bony vertebrates. *BMC Evol Biol.* 2008; 8:171. [PubMed: 18541024]
- Xu J, Lawshe A, MacArthur CA, Ornitz DM. Genomic structure, mapping, activity and expression of fibroblast growth factor 17. *Mech Dev.* 1999; 83:165–178. [PubMed: 10381577]
- Xu J, Liu Z, Ornitz DM. Temporal and spatial gradients of Fgf8 and Fgf17 regulate proliferation and differentiation of midline cerebellar structures. *Development.* 2000; 127:1833–1843. [PubMed: 10751172]
- Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 1998a; 15:568–573. [PubMed: 9580986]
- Yang Z. On the best evolutionary rate for phylogenetic analysis. *Syst Biol.* 1998b; 47:125–133. [PubMed: 12064232]
- Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 2007; 24:1586–1591. [PubMed: 17483113]
- Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Odeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, Zeng Y, Xiong Z, Liu S, Zhou L, Huang Z, An N, Wang J, Zheng Q, Xiong Y, Wang G, Wang B, Wang J, Fan Y, da Fonseca RR, Alfaro-Nunez A, Schubert M, Orlando L, Mourier T, Howard JT, Ganapathy G, Pfenning A, Whitney O, Rivas MV, Hara E, Smith J, Farre M, Narayan J, Slavov G, Romanov MN, Borges R, Machado JP, Khan I, Springer MS, Gatesy J, Hoffmann FG, Opazo JC, Hastad O, Sawyer RH, Kim H, Kim KW, Kim HJ, Cho S, Li N, Huang Y, Bruford MW, Zhan X, Dixon A, Bertelsen MF,

Derryberry E, Warren W, Wilson RK, Li S, Ray DA, Green RE, O'Brien SJ, Griffin D, Johnson WE, Haussler D, Ryder OA, Willerslev E, Graves GR, Alstrom P, Fjeldsa J, Mindell DP, Edwards SV, Braun EL, Rahbek C, Burt DW, Houde P, Zhang Y, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 2014; 346:1311–1320. [PubMed: 25504712]

Zhang J, Rosenberg HF, Nei M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc Natl Acad Sci U S A*. 1998; 95:3708–3713. [PubMed: 9520431]

Highlights

- *FGF17* has been independently lost in at least two avian lineages.
- In bird species which retain *FGF17*, it is intact and functional.
- *FGF17* is under purifying selection in birds which retain it.
- *FGF17* is likely prevented from pseudogenization because of its high connectivity.

Fig. 1.

A diagram depicting the establishment of the FGF8 subfamily members through two rounds of whole genome duplications (1R–WGD and 2R–WGD) (Cañestro et al., 2013; Itoh and Ornitz, 2004; Popovici et al., 2005; Wotton et al., 2008). After 1R–WGD, the putative ancestral gene, *Fgf8/17/18/24*, divides into two genes, proto*Fgf8/17* and proto*Fgf18/24*, which subsequently divided again in a second round of duplication (2R–WGD), giving rise to the 4 members of the FGF8 subfamily we currently identify. *Fgf24* is only retained in fishes and has presumably been lost in tetrapods.

Fig. 2.

Consensus cladogram based on Ericson et al., (2006), depicting the phylogenetic relationship of all 27 lineages included in this study. Clades that are missing FGF17, marked by thicker branches and in red, fall into two monophyletic clades. This pattern indicates only two losses of FGF17, within Passeriformes (in the clade Passeroidea) and Galliformes. The order of each lineage, is marked either on the root branch for each clade (in black) or on the left for Galliformes and Passeriformes. Common names are used for all species.

Fig. 3.

Expression of FGF17 in E13.5 mouse and stage 32 pigeon determined by RT-PCR and demonstrated lack of *FGF17* expression in chicken at five developmental stages ranging from stage 15 to stage 32 (top panel). 18S was used as a positive control (bottom panel). A product of ~250 base pair (bp) was produced from cDNA generated from E13.5 mouse and stage 32 pigeon, while no bands were generated from any of the chicken samples. 18S PCR proves cDNA used for the chicken analyses was competent for PCR amplification.

Fig. 4.

A) Rooted phylogram of *FGF8*, *FGF17* and *FGF18* constructed from phylogenetic analyses of gene CDS implemented in BEAST v. 1.6.1 (Drummond et al., 2012). Avian *FGF17* represent true orthologs to the *Fgf17* in other species (alligator, turtle and human). All three genes are respectively monophyletic. *FGF8* and *FGF17* also display monophyly, as proposed by previous studies (Itoh and Ornitz, 2004; Popovici et al., 2005). Branches are labeled with common name of species as well as corresponding gene name. Bayesian posterior probabilities (Bpp) are indicated on branches as either 1 or * (.75 < * < 1). Clades

with Bpp < .60 were collapsed. Human, alligator (*A. mississippiensis*) and turtle (*C. picta*) sequences were also included in this analysis. Human *Fgf21* was used as an outgroup to root the tree. Scale bar represents number of substitutions per site. B) Inset cladogram of FGF17-possessing birds with branch labels along each branch displaying the ML estimate of the numbers of nonsynonymous (N-Black) and synonymous (S - Red) substitutions for FGF17 along that branch. N/S estimates were generated using the “Free-ratio” model in the Branch Models analysis using CodeML program in PAML v. 4.4 (Yang, 2007).

Fig. 5.

Chromosomal synteny of the regions surrounding the FGF17 locus depicted in chicken, zebra finch, human (chromosomes), turtle and alligator (scaffolds). There is an ~1.2 mb gap in the chicken and zebra finch chromosome 22, within the region where *Fgf17* is located in alligator, turtle and human. Scaffold arrangement reveals no inversions within archosaurs (chicken, zebra finch, and alligator), with minimal local arrangement between turtle and archosaurs and human. Genes which are located within the “missing” region in chicken chromosomes 22 are identified in both alligator and turtle. Potentially “missing” avian genes are depicted in grey, while those identified are in black (DMTN and LGI13). LGI13 was also identified in an unplaced scaffold in zebra finch.

Table 1
Maximum Likelihood estimates of dN/dS (ω) on (branch analysis) ω_f indicates foreground branch

	Model	nP	Parameters	lnL	K (ts/tv)	ω_0	ω_f	
Free	Free-ratio	247	n/a	-6498.9766	1.8247			
A	One-ratio	125	ω	-6650.1007	1.8084	0.0655		
B	Two-ratio	125	ω_0, ω_f8	-6649.6589	1.8136	0.0658	0.026	
C	Two-ratio	125	ω_0, ω_f17	-6645.7568	1.8323	0.0676	0.0021	
D	Two-ratio	125	$\omega_0, \omega_f17 = \omega_f8$	-6649.2281	1.8238	0.0151	0.0668	
E	Four-ratio	128	$\omega_0, \omega_f18, \omega_f17, \omega_f8$	-6643.6155	1.8472	$\omega_f18 = 0.02$	$\omega_f17 = 0.0021$	$\omega_f8 = 0.0023$

Table 2

Likelihood-Ratio Statistics (2 l) for testing hypotheses

Test	2 l	df	P value
<i>Model A vs. Free</i>	302.2481	122	8.06E-18**
<i>Model A vs. E</i>	12.9705	3	0.002192252**
<i>Model B vs. A</i>	0.8836	1	0.2728348
<i>Model B vs. E</i>	12.0868	2	0.001186723**
<i>Model C vs. A*</i>	8.6878	1	0.001757598**
<i>Model C vs. E*</i>	4.2826	2	0.05875006
<i>Model D vs. A</i>	1.7452	1	0.1257139
<i>Model D vs. E</i>	11.2253	2	0.001825725**

*
P< .05**
P< .01

Table 3

Maximum Likelihood estimates of dN/dS (ω) in Fgf8

Branch Analysis	Model	nP	Parameters	lnL	K (ts/tv)	ω_0	ω_g	ω_p
Free	Free-ratio	83	n/a	-2567.8286	2.34279			
A	One-ratio	43	$\omega_0 = \omega_g = \omega_p$	-2627.8146	2.36653	0.0813	$= \omega_0$	$= \omega_0$
B	Two-ratio	44	$\omega_0 = \omega_g, \omega_p$	-2627.8145	2.36653	0.0813	$= \omega_0$	0.9417
C	Two-ratio	44	$\omega_0 = \omega_p, \omega_g$	-2624.3031	2.37727	0.0874	0.0001	$= \omega_0$
D	Two-ratio	44	$\omega_0, \omega_g = \omega_p$	-2624.3031	2.37727	0.0874	0.0001	$= \omega_0$
E	Three-ratio	46	$\omega_0, \omega_g, \omega_p$	-2624.3030	2.37727	0.0874	0.0001	0.9213
Clade Analysis	Model	nP	Parameters	lnL	K (ts/tv)	ω_0	ω_g	ω_p
Free	Free-ratio	83	n/a	-2567.8286	2.3428			
A	One-ratio	43	$\omega_0 = \omega_g = \omega_p$	-2627.8146	2.3665	0.0813	$= \omega_0$	$= \omega_0$
B	Two-ratio	44	$\omega_0 = \omega_g, \omega_p$	-2626.0528	2.3634	0.0847	$= \omega_0$	0.017
C	Two-ratio	44	$\omega_0 = \omega_p, \omega_g$	-2617.5731	2.3779	0.0974	0.0001	$= \omega_0$
D	Two-ratio	44	$\omega_0, \omega_g = \omega_p$	-2616.3431	2.3733	0.1023	0.0054	0.0054
E	Three-ratio	46	$\omega_0, \omega_g, \omega_p$	-2615.2624	2.3749	0.1027	0.0001	0.017

ω_0 = background; ω_g = Galliformes, ω_p = Passeroidea

Table 4

Likelihood-Ratio Statistics (2 l) for testing hypotheses

Test	2 l	df	P value
<i>Branch</i> - C vs A	7.023006	1	0.004493903**
<i>Branch</i> - C vs E	4.4E-05	2	0.499989
<i>Branch</i> - D vs A	7.022976	1	0.00449398**
<i>Branch</i> - D vs E	7.4E-05	2	0.4999815
<i>Clade</i> - D vs A	22.94283	1	8.68E-07**
<i>Clade</i> - D vs E	2.16153	2	0.1696679

*
P< .05**
P< .01