



Universiteit  
Leiden  
The Netherlands

## **Spatio-temporal gene expression analysis from 3D in situ hybridization images**

Welten, M.C.M.

### **Citation**

Welten, M. C. M. (2007, November 27). *Spatio-temporal gene expression analysis from 3D in situ hybridization images*. Leiden Institute of Advanced Computer Science, group Imaging and Bio-informatics, Faculty of Science, Leiden University. Retrieved from <https://hdl.handle.net/1887/12465>

Version: Corrected Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/12465>

**Note:** To cite this publication please use the final published version (if applicable).

## Chapter 6

# **Limb - fin heterochrony: a case study analysis of molecular and morphological characters using frequent episode mining.**

R.Bathoorn<sup>3 (a)</sup>, M.C.M.Welten<sup>1,2 (b)</sup>, A.P.J.M Siebes<sup>3</sup>, M.K. Richardson<sup>2</sup> and F.J. Verbeek<sup>1</sup>

1 Imagery and Media, Leiden Institute of Advanced Computer Science,  
Leiden University, Niels Bohrweg 1, 2333 CA Leiden, the Netherlands

2 Institute of Biology Leiden, Leiden University, Wassenaarseweg 64, 2311 GP Leiden,  
the Netherlands

3 Department of Information and Computing Sciences, Utrecht University, Padualaan 14,  
3584CH Utrecht, the Netherlands

(a and b equally contributed)

Submitted.

## **Case study Late zebrafish and cross-species development**

## ABSTRACT

Developmental biologists use a wide range of vertebrate model species including chicken, mouse, axolotl, clawed toad and zebrafish. In this framework, zebrafish can be considered as the “fruitfly” of the vertebrates as experimental findings are often to be extrapolated to higher vertebrates. To that end, it is important to consider phylogenetic relationships between zebrafish and other model species in order to determine whether gene networks observed in one species are indeed representative for the other. Reconstruction of phylogenetic relationships can also reveal insights into the evolution of developmental mechanisms. One approach is to look at differences in the timing of developmental events, i.e. heterochrony. Analysis of heterochrony can render important information about both development and evolution.

In this paper we address the problem of limb-fin heterochrony by analysing both molecular and morphological characters at the same time and cross species. To that end we focus on a new method for computational analysis, i.e. frequent episode mining in development analysis (FEDA), for reconstruction of phylogenetic trees based on heterochrony in gene expression; through existing methods this could not be solved. We illustrate our approach using spatio-temporal gene expression timing data from limb and fin development. Therefore we have made extensive series of *in situ* hybridizations of a panel of developmental genes in the zebrafish pectoral fin, and compared these with our data from the chick limb as well as with (gene expression) data from other species from the literature. In support of our analysis, a relative time scale of embryonic development is introduced and made part of our computational framework.

This paper explores the application of FEDA based on gene expression patterns and the results provide evidence for the limb-fin heterochrony as well as a much wider scope of application. From our case studies we demonstrate that the method provides an extensive application to complex datasets of increasing complexity. Moreover, for the case studies presented we obtained new results and insights. It can be expected that further application of FEDA will support knowledge discovery in developmental and evolutionary biology.

**Keywords:** *in situ* hybridization, frequent episode mining, heterochrony, gene expression.

## INTRODUCTION

To analyze development, molecular genetics, and genetic networks in zebrafish and other model species, it is important to (a) study evolutionary relationships and consider the model species in the right phylogenetic context, and (b) study to what extent they are representative for each other and for human (Hanken, 1993; Metscher and Ahlberg, 1999). Zebrafish is increasingly popular as a vertebrate model species. It is a lower vertebrate model and thus interesting as a key species. A workflow in developmental biology inspired research could be that initial experiments are carried out in zebrafish, and henceforth extrapolated to other model species. We have taken this route in one of our case studies.

One way to analyze evolutionary relationships between species is based on differences in developmental timing, i.e. heterochrony. In itself, heterochrony is considered as one of the mechanisms producing evolutionary changes. Heterochrony manifests as species difference in growth patterns, in changes of developmental sequences, and changes in temporal patterns of gene expression (Jeffery et al, 2002; Richardson, 2002; Smith, 2002). Several methods have been developed to analyze heterochrony in a phylogenetic framework. These methods are based on heterochrony in sequences of morphological and developmental events (Bininda-Emonds et al, 2003; Jeffery et al, 2005; Schlosser, 2001; Schulmeister and Wheeler, 2002). In this paper we explore Frequent Episode mining in Developmental Analysis (FEDA). The theory of FEDA as a new method to analyze data with a focus on heterochrony in gene expression in an evolutionary and developmental context and with computational numerical evidence is discussed in Bathoorn et al (in preparation). The principle of the FEDA is to analyze sequences of developmental characters to find episodes; these are small ordered sets that are frequent over all developmental sequences considered. These episodes are used to determine differences between developmental sequences (Bathoorn et al., 2007). Two distinct case studies of frequent episode mining are described. These case studies focus on limb and fin development in vertebrates, but differ in the organization and focus of the data as well as model species:

- 1) A comparative study of gene expression in Zebrafish (*Danio rerio*) fin development and known orthologues in four other model species: African clawed toad (*Xenopus laevis* and *tropicalis*), Mexican axolotl (*Ambystoma mexicanum*), mouse (*Mus musculus*) and chicken (*Gallus gallus*). This employs our own new gene expression data and data from the literature.
- 2) An analysis of gene expression timing in morphological structures, carried out on data on digit formation in chicken fore and hindlimb (Welten et al, 2005). Here, we show that a heterochronic difference in gene expression can lead to developmental arrest and finally, evolutionary loss of a structure. Forelimb (derived structure) and hindlimb (ancestral condition) are analyzed in the same individual (n = 3-8; c.f. Materials and Methods). Using FEDA we do a reanalysis and evaluation of expression data from our previous study on chicken wing digit homology (Welten et al, 2005).

These studies will provide relevant biological results and at the same time they display possibilities for the applications of frequent episode mining (using FEDA) in specific areas.

The vertebrate limb and fin are well-studied structures in developmental biology and recent studies reveal that a seemingly subtle event like change in gene expression timing can produce gross macro-evolutionary changes (Smith, 2003). In several studies, examples of evolutionary changes through heterochrony in gene expression are given. One example of change in duration of gene expression and effect on outgrowth, and therefore a final structure, is the dolphin flipper (Richardson and Oelschlaeger, 2004). Other examples of heterochrony in gene expression can be found in Blanco et al. (1998); where difference in duration of *hoxa11* expression in *Xenopus* limbs affects the identity of a limb region; in Shapiro et al (2003), describing digit number affected by *sonic hedgehog* (*shh*) expression (Smith, 2003), and in Zakany et al (1997), showing that a slight time delay in *hox* gene expression (transcriptional heterochrony) in mutant mice leads to caudal transposition of the sacrum. Another example is loss of pelvic spines in freshwater populations of three-spine stickleback. Marine three-spine sticklebacks possess a pelvic skeleton comprising bilateral pelvic spines that articulate with the pelvic girdle (Cole et al. 2004, Shapiro et al. 2004). Several freshwater stickleback populations however, show partial or complete loss of the pelvic skeleton. The marine populations show expression of *pitx1* in the future pelvic region, while *pitx1* expression was absent in freshwater populations of three-spine stickleback (Cole et al. 2004; Forster and Baker 2004; Shapiro et al. 2004). Here, the loss of *pitx1* expression leads to reduction of pelvic structures in two different populations of three-spine stickleback.

During embryonic development many genes are expressed and therefore, analysis of heterochrony in gene expression in embryos can provide useful information concerning evolutionary changes and phylogenetic relationship of vertebrate model organisms (Richardson 1995, Jeffery et al, 2002, Smith 2003). Recently, large amounts of gene expression data from different model species became available from functional genomics, clinical research and molecular developmental research (e.g. zebrafish: <http://zfin.org>, (<http://bio-imaging.liacs.nl/>), <http://cegs.stanford.edu.search.isp>; *Xenopus* : <http://www.xenbase.org/>; for mouse <http://genex.hgu.mrc.ac.uk/>), for chicken: <http://geisha.biosci.arizona.edu/data/>, and KEGG [www.kegg.org](http://www.kegg.org)). These online sources can be used to compare gene expression data and gene networks in the different model species.

#### *Case study 1: Comparative study of heterochrony in gene expression*

In this case study we want to illustrate the evolutionary relationships among different vertebrates, with a focus on differences in timing of gene expression during limb and fin development. Development of limb and fin is particularly interesting as a case study since paired appendages are unique to jawed vertebrates. Data from the fossil record suggest that the fin-limb transition occurred approximately 410 million years ago (Shubin et al, 2006). Though teleost fins and tetrapod limbs show important differences, the same, highly conserved patterns and genetic pathways are found both in teleost (main subgroup of ray-finned fish) fin and in tetrapod limb development (Coates and Cohn, 1998). For a comparative study of these highly conserved genetic pathways a selection of genes involved in zebrafish paired fin development, of which known orthologues exist in

tetrapod model species, is extracted from literature (Coates and Cohn, 1998, Hinchliffe, 2002; Tickle, 2002). For these genes, probes for Fluorescent *in situ* hybridization (FISH; Welten et al, 2006) and *in situ* hybridization (ISH; Thisse et al, 1993) are synthesized from clones or from cDNA. To obtain additional data, FISH and ISH are applied to developmental series of zebrafish embryos.

For chicken, gene expression data from a previous study were used (Welten et al, 2005). Supplemental data from chicken as well as data for mouse, *Xenopus* and axolotl are extracted from literature.

For *case study 1*, we address and assemble data from five species to apply FEDA, in order to analyse heterochrony in gene expression in five species, in a numeric and evo – devo (evolutionary-developmental) context. Though the evolutionary relationships of the model systems are evident from literature, we use this evidence to validate our algorithm. We expect that FEDA will be able to reconstruct phylogeny based on gene expression data.

#### *Case study 2: Heterochrony within one species*

Heterochrony can occur in one single species. To analyze heterochrony in gene expression in one single species, we used data from a study of early molecular markers of chondrogenesis in relation to a rudimentary digit I in the bird wing embryo (Welten et al, 2005; cf. **Chapter 5**). Other examples of heterochrony in gene expression within one species can be found in Blanco et al (1998), describing heterochronic differences in *hoxa-11* expression in *Xenopus* fore and hindlimb that affects the regional identity of tarsal bones within one individual.

In a previous study, chicken was used as a model to study digit loss (Welten et al., 2005; cf. **Chapter 5**) since chicken limb is a well-studied structure and many probes are available for developmental analysis. A *sox9* expression domain was found for a vestigial digit I (Welten et al., 2005). In the presumptive digit I domain no expression is found for subsequent markers from the molecular cascade of chondrogenesis. The onset of *sox9* in the presumptive digit I domain is relatively late compared to the other digits (Welten et al, 2005).

In *case study 2*, gene expression data of markers from the molecular cascade of chondrogenesis in chicken in hind limb and wing (Welten et al, 2005) are processed with frequent episode mining algorithm. In the chicken embryo, five digit primordia are found in the foot; in the wing 3 digit primordia are found. In the adult however, the chicken foot is four-toed while the wing is only three-fingered. In this analysis the five-toed foot of the chicken embryo is used as a reference for digit formation timing; timing data of the wing - which has three digits in the adult - are thus compared to timing data of the foot. In this particular application, FEDA is used to analyze heterochrony in gene expression in fore and hindlimb within one species.

#### *Relative time scale of development*

The onset of limb formation is variable among the different model species (Richardson, 1995). In chicken and mouse, limbs are formed during somitogenesis. In Axolotl and zebrafish, the formation of the forelimbs and pectoral fins respectively, occurs long after the end of somitogenesis (Bordzilowskaya, 1989; Kimmel et al, 1995). In *Xenopus*,

hindlimbs develop before forelimbs, relatively late, after completion of somitogenesis (Richardson 1995).

Taking these differences into consideration as well as to enable comparison, a relative time scale of development during which limb and fin development occur is computed for every single model species (cf. Materials and Methods). For the application of FEDAs to gene expression data in the chicken wing and hindlimb, a different relative time scale is specified for limb development (cf. Materials and Methods).

*The remaining part of this paper is structured as follows: in the Materials and Methods section we describe the pre-processing, aggregation and analysis of data, and the calculation of the relative time scale of development. In the Results section, we describe the outcome of the analysis. In the Discussion and Conclusion sections, we address the results of our experiments and we present conclusions as well as directions for future research.*

## MATERIALS AND METHODS

For both applications, the preprocessing of gene expression data to make them suitable for FEDAs, as well as the different relative time scales are described in the next ten subsections.

### *Case study 1: Comparative study of heterochrony in gene expression*

#### *Selection of genes*

Data regarding genes, indicated in the literature as markers genes in limb and fin development, myogenesis, cartilage and joint formation, as well as *Hox* genes involved in limb and fin development, are extracted from literature (Coates and Cohn, 1998; Duprez, 2002; Hinchliffe, 2002; Karsenty and Wagner, 2002; Tickle, 2002). In Table 1, the selected genes, a description of the expression site in limb or fin, and corresponding Gene Ontology term (<http://www.geneontology.org/>) are summarized. Gene Ontology terms are included as these are based on a controlled vocabulary, such providing unique identifiers for further research and interoperable search in databases. GO is applicable to a range of species, combining biological processes and corresponding molecular and cellular functions of genes (The Gene Ontology Consortium, 2000; Camon et al, 2004). Moreover, we are aware that in zebrafish duplication of some of the genes has occurred (Akimenko et al. 1995; Amores et al.1998; Yan et al. 2004). Zebrafish isoforms of the genes used in *case study 1* are summarized in Table 1.

#### *Obtaining new and existing gene expression data*

Zebrafish were maintained under standard conditions (<http://zfin.org>) in our facilities. Zebrafish embryos were harvested and staged to hours post fertilization, as described by Kimmel et al. (1995) and processed for *in situ* hybridization (ISH) and fluorescent *in situ* hybridization (FISH; Welten et al, 2006) during subsequent stages of pectoral fin development (24 -120 hpf; Kimmel et al, 1995). Imaging was performed as previously described (Welten et al., 2006).

Chicken (*Gallus gallus domesticus*) gene expression data were obtained from ISH experiments for *sox9*, *bmpr1b* and *wnt-14*. Supplementary to these experimental data a selected panel of genes involved in limb development (cf Table 1) was assembled by literature search.

In addition to the experimental data, gene expression data of the relevant model species (mouse (*Mus musculus*), clawed toad (*Xenopus laevis / tropicalis*), and axloltl (*Ambystoma mexicanum*)) were collected through literature search (Pubmed <http://www.ncbi.nlm.nih.gov/entrez/>, Google Scholar <http://scholar.google.nl/>).



**Table 1.**

Selection of genes used for heterochrony analysis in five model species. Gene function, name, and expression site in fin or limb are given. The corresponding Gene Ontology term refers to the biological process described in Gene Ontology

| Gene function   | Gene name and expression site  | Gene Ontology Term                     |
|---|--|--|
| Limb initiation and identity                            | <i>tbx5</i> ; expressed in forelimb mesenchyme   | GO 0030326                             |
| Limb patterning and outgrowth                           | <i>fgf8</i> ; expressed in Apical Ectodermal Ridge.  | GO 0030326<br>GO 0051216               |
|   | <i>sonic hedgehog</i> ; expressed in zone of polarizing activity   | GO 0030326                             |
|   | <i>msx2</i> , ( <i>msx-b</i> for zebrafish), expressed in Apical Ectodermal Ridge                                      | GO 0030326                             |
| Cartilage and joint formation                           | <i>sox9</i> ( <i>sox9a</i> and <i>sox9b</i> , zebrafish), expressed before prechondronic condensation                  | GO 0030326<br>GO 0001501<br>GO 0051216 |
|   | <i>runx2</i> ( <i>runx2a</i> and <i>runx2b</i> , zebrafish), expressed in chondrogenic condensation and in osteoblasts | GO 0001501<br>GO 0051216<br>GO 0001503 |
|   | <i>bmpr1b</i> ; expressed in chondrogenic condensation   | GO 0001501<br>GO 0051216               |
|   | <i>Wnt-14</i> ( <i>wnt9a</i> ): expressed in joints  | GO 0030326                             |
|   | <i>Chondromodulin-1</i> ; expressed in chondrogenic cells  | GO 0006029                             |
| Myogenesis  | <i>myoD</i> , expressed in myogenic tissue   | GO 0007517<br>GO 0007519               |
| Posterior <i>hox</i> genes involved in limb development | <i>Hoxa9</i> ; tail and pectoral fin mesenchyme  | GO 0007275                             |
|   | <i>hoxb13</i> ; tail mesenchyme  | GO 0009611<br>GO 0040008<br>GO 0008544 |
|   | <i>hoxc12</i> ; tail mesenchyme fin mesenchyme   | GO 0007275                             |
|   | <i>hoxd9</i> ; tail and pectoral fin mesenchyme  | GO 0007275                             |
|   | <i>hoxd11</i> ; tail and posterior region of fin / limb bud  | GO 0007275<br>GO 0001501<br>GO 0001759 |

### *Aggregation of spatio-temporal data*

We focused on gene expression data that have been obtained from whole mount *in situ* experiments. These give both a spatial and a temporal clue. Both ISH data from literature and data generated by “in house” FISH and ISH (zebrafish and chicken, respectively) are organized in a spreadsheet table. Data are categorized according subsequent developmental stages, from limb bud stages to juvenile.

In the table, gene expression data are attributed a quadruplet consisting of:

- 1) Normal stage of embryonic development,
- 2) Corresponding age in developmental units, i.e. embryonic or larval day,
- 3) A corresponding relative time scale from fertilization to juvenile,
- 4) Presence at a particular stage.

As an example an excerpt from our spreadsheet table for *sox9a* expression in zebrafish pectoral fin, is given in Table 2.

**Table 2.** Spreadsheet table of *sox9a* expression in zebrafish pectoral fin.

| Gene          | Developmental stage (Kimmel et al, 1995) | Age (days) | % of development | Presence |
|---------------|--|------------|------------------|----------|
| <i>sox9a</i>  | Prim 5                                   | 1 day      | 20%              | no       |
| <i>sox9a</i>  | High pec                                 | 2 days     | 40%              | yes      |
| <i>bmpr1b</i> | High pec                                 | 2days      | 40%              | yes      |

### *Relative time scale of development in comparative study of heterochrony*

For the comparative study of the five species, the relative time scale is computed for every single model species. The relative time scale indicates a ratio of complete development, but also relative to limb development. For all species the onset, i.e. 0%, is set up at fertilization. For practical reasons the 100% stage was chosen differently for every model species. As a criterion for 100% of development, independent feeding (zebrafish, mouse, chicken) or juvenile stages (clawed toad, axolotl) was chosen.

For chicken, hatching, and for mouse, birth is considered as 100% of development (Hamburger and Hamilton, 1951; Theiler, 1972 respectively). For zebrafish, the relative time scale is calculated up to 96 hpf, as it reaches the early larval stage, though the pectoral fins are still developing during this stage. The pelvic fins develop during late larval stages (Kimmel et al, 1995, Grandel and Schulte - Merker, 1998). For clawed toad and axolotl, the relative time scale of development is calculated up to juvenile stages, since limb development occurs during larval stages (for *Xenopus*: Nieuwkoop and Faber, 1967; for Axolotl: Bordzilowskaya et al, 1989; Nye et al, 2003).

Gene expression data are organized according to subsequent developmental stages. In Table 3 the relative time scale computation, relative time point of appearance of fore and hind limb buds, as well as references for stages of embryonic development are summarized.

**Table 3.** Extraction of features used in data collection.

| Species        | Stage,<br>100% of<br>development           | First appearance of<br>limb/fin bud |                     | % of development,<br>first appearance<br>limb/fin bud |                 | Reference<br>(cf.<br>References)                  |
|----------------|--|-------------------------------------|---------------------|---|-----------------|---|
|                |  | fore/<br>pectoral                   | hind/<br>pelvic     | fore/<br>pectoral                                     | hind/<br>pelvic |   |
| Zebrafish      | 96 hpf<br>(Kimmel,<br>1995)                | prim 18                             | larva (>21<br>days) | 31%   | 100%            | Kimmel et al,<br>1995.                            |
| <i>Xenopus</i> | Juvenile, 58<br>days (N & F<br>66)         | NF 48                               | NF 46               | 13%   | 8%              | Nieuwkoop and<br>Faber, 1967                      |
| Axolotl        | Juvenile, ><br>36 days (Nye<br>57)         | B. 37                               | Nye 51              | 19%   | 82%             | Bordzilowskaya<br>et al, 1989;<br>Nye et al, 2003 |
| Mouse          | Newborn<br>mouse,<br>19-20 days<br>(Th 27) | Th 15                               | Th 16               | 50%   | 53%             | Theiler, 1972                                     |
| Chicken        | Hatching, 21<br>days (H&H<br>46)           | H& H 16                             | H&H 17              | Ca.10%  | Ca. 11%         | Hamburger and<br>Hamilton, 1951                   |

*Data analysis*

The datasets are collected in spreadsheet tables and subsequently converted to comma separated values (CSV) files; these files are the input for analysis with FEDAs (Bathoorn et al 2007). After initialization in the Frequent Episode tree, the Jaccard distance is computed and next a clustering is applied. From the clustering a cladogram can be derived and visualized. In this cladogram, species are expressed in clusters, depending on the rate of difference i.e. Jaccard distance, between the species (Bathoorn et al, in preparation).

*Case study 2: Heterochrony in gene expression in the chicken wing and hindlimb**Preprocessing of data*

For chicken (*Gallus gallus domesticus*), gene expression data of chondrogenesis markers (*sox9*, *bmpr1b*) and alcian blue staining for cartilage formation from stages 24 – 34 (Hamburger and Hamilton, 1951; Murray and Wilson, 1994) are recorded for the skeletal elements in the fore and hindlimb (Welten et al, 2005). All skeletal elements are recorded in a proximo-distal order: from humerus and femur (upper arm or leg) to the phalanges in the autopodium (hand and foot).

*Aggregation of spatio-temporal data*

Skeletal elements are organized according to proximal (humerus / femur) to distal position (phalanges) and plotted against onset and duration of gene expression.

*Relative time scale of development in context of wing and hind limb development*

For the application of FEDA to gene expression data in the chicken wing and hindlimb, a relative time scale of limb development is implemented. The onset of cartilage formation in the stylopodium i.e., upper arm or leg, is considered as 0 % of development (Hamburger and Hamilton stage 24). Chondrification of the digits is considered as 100% of development (Hamburger and Hamilton stage 34), all total a time span of four days (Hamburger and Hamilton, 1951; Murray and Wilson, 1994). All stages between HH 24 and 34 represent a ratio of development.

In this application FEDA is used to analyze heterochrony in gene expression in morphological structures, to show that morphological differences between fore and hindlimb within one species can be the result of heterochrony in gene expression in these different regions.

In like manner to the previous case study the data are imported to the FEDA as CSV-files.

*Data analysis*

Application of FEDA in this case study required a different organization of the data with respect to the previous that was concerning more species. Skeletal elements and future skeletal elements are organized according proximo-distal sequence along the limb. Presence of gene expression at a particular developmental stage (Hamburger and Hamilton, 1951) is scored for proximal, carpal and digit I t/m digit V region. In Table 4, an example is presented for *sox9* expression in the chicken wing.

**Table 4.** Spreadsheet table of *sox9* gene expression data in the chicken wing. Onset of *sox9* gene expression in future skeletal elements is scored in proximal, carpal and digit regions. Onset of gene expression is scored at stages according to Hamburger and Hamilton (1951) and corresponding percentage of limb development.

| Skeletal element | Proximal <i>sox9</i> expression | Carpal <i>sox9</i> expression | Digit I <i>sox9</i> expression | Digit II <i>sox9</i> expression |
|------------------|---------------------------------|-------------------------------|--------------------------------|---------------------------------|
| Ulna             | HH 24 / 15%                     |                               |                                |                                 |
| Ulnare           |                                 | HH 25 / 30%                   |                                |                                 |
| Metacarpal       |                                 |                               | HH 30 / 50%                    | HH 30 / 50%                     |

In this application, the occurrence of frequently found patterns, i.e. (future) skeletal elements, is analyzed for the chosen chondrogenesis markers: *sox9* and *bmpr1b* as well as cartilage formation, visualized with alcian blue staining. The computed pattern shift results in a so –called pattern shift diagram. In this diagram, shift in timing of gene expression in different locations in one single organism is shown in relation to percentage of limb development (cf. Figure 3).

## RESULTS

*Comparative study of heterochrony in gene expression*

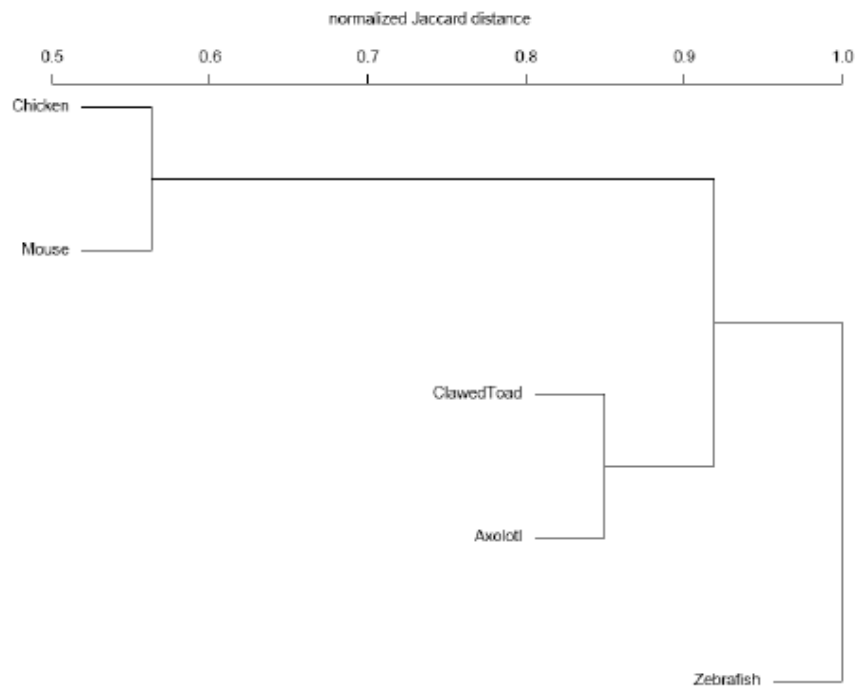
Application of FEDA to gene expression data in different model species has resulted in a cladogram which is depicted in Fig.1. This tree, based on gene expression data, is

completely consistent with the cladogram described in Metscher and Ahlberg (1999) as well as with cladograms described in the Tree of Life website (<http://tolweb.org/tree/phylogeny.html>); ergo phylogenetic reconstruction based on gene expression in limb / fin is conform phylogenetic reconstructions based on morphological characters.

From the cladogram it is immediately visible that zebrafish is separated from the tetrapods, while the tetrapod cluster itself contains two groups: the amphibians (*Xenopus* and axolotl) and the amniotes (chicken and mouse).

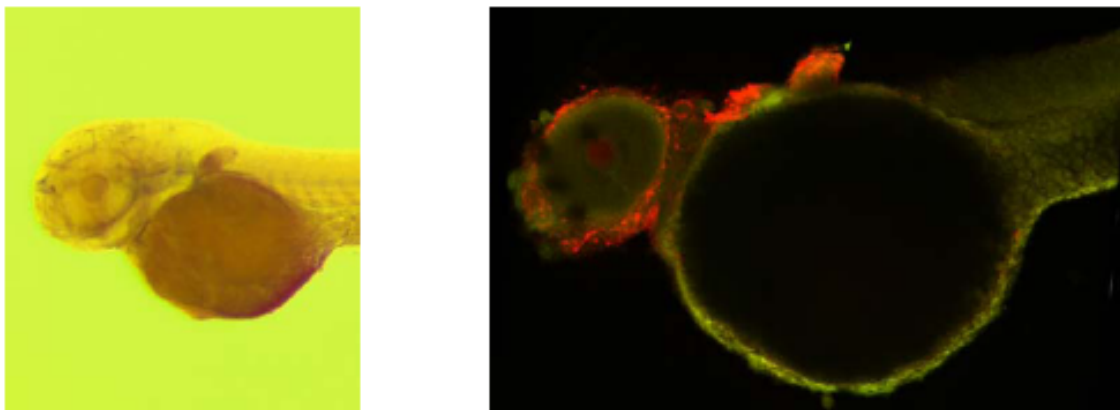
Analysis of the differences at the nodes of the cladogram is in correspondence with literature. For instance, the genetic pathway responsible for limb initiation, patterning and outgrowth (*tbx5*, *fgf8*, *msx2*, *sonic hedgehog*) is found in all 5 model species, though relative timing and duration of gene expression is different for the 5 model species.

The axis in Fig.1 shows Jaccard distance i.e., the rate of dissimilarity between species. (Bathoorn et al, in preparation). In this cladogram, it does not relate to mutation speed or evolutionary divergence time.



**Fig. 1.** Phylogenetic tree, based on gene expression data, constructed with Frequent episode mining. The x-axis in the tree depicts Jaccard distance (Bathoorn et al, 2006), which can be explained as the numerical rate of dissimilarity between species. Numbers 0 -1 on the axis indicate increase of dissimilarity. According to the computations (Bathoorn et al, in preparation), the amniote taxa (chicken and mouse) are closest together and are clustered in one group. The amphibians, *Xenopus* and axolotl, showing a high rate of dissimilarity, are clustered in a different group. The zebrafish, which is the lowest vertebrate, show highest dissimilarity to the other species.

In *case study 1*, gene expression data for several markers of chondrogenesis were used (*sox9*, *runx2*, *bmpr-1b* and *chondromodulin-1*). Gene expression data for chondrogenesis markers were available from literature for nearly all model systems. For zebrafish however, *bmpr1b* expression has only been described at pre-limb bud stages (Nikaido et al, 1999). For FEDA we concluded that *bmpr1b* should be part of a certain episode. This could be confirmed by doing a FISH / ISH experiment in zebrafish at 36-72 hpf. Indeed expression of *bmpr1b* was found in the pectoral fins and in skeletal elements at these stages of development (Fig.2). This illustrates that FEDA provides possibility to predict gene activity in a particular species, at absence of gene expression data for this species. From the FEDA output as well as from pilot experiments, functional studies can be carried out to further characterize the role of the gene in this particular species (cf. Discussion).



**Fig. 2.** *bmpr-1b* expression in zebrafish pectoral fin and branchial arches at 48 hpf. Anterior is to the left, dorsal to the top. Left panel : ISH result from AP detection method; Right panel: FISH result. The picture is the result of the projection of a 2 channel 3D confocal image. The expression is in red, i.e. the red channel of the CLSM image, the green depicts a standard staining of the cell nuclei.

#### *Heterochrony in gene expression in the chicken wing and hindlimb*

The aim of this application was to show that the apparent differences in fore and hind limb of chicken may be the result of a shift in gene expression timing. Our method clearly visualizes timing shifts at different levels, i.e. on the level of gene expression and on the level of cartilage formation. In the heterochrony study of gene expression in chicken wing and hindlimb, timing differences for *sox9*, *bmpr1b* expression and cartilage formation are visualized in the pattern shift diagram in Fig.3. Data for wing and hindlimb are displayed in the relative time scale of cartilage formation in wing and hindlimb (cf. Materials and Methods). The onset of cartilage formation in the proximal part of the limb bud is determined as 0 % of development, where completion of cartilage formation in the digits is determined as 100% of development (Murray and Wilson 1994; own experiments). The heterochrony is considered and represented in the form of a so-called pattern shift diagram. In the pattern shift diagram, the relative position of red (wing) and green (hindlimb) along the relative time scale of limb development indicates the onset and duration of gene expression in wing or hindlimb. The onset of *sox9*, *bmpr1b* and cartilage formation in the metacarpals (mc) and metatarsals (mt), is first observed in mc

and mt IV and V, followed by mc / mt III and II. *sox9* gene expression is present in mc of digit I of both wing and hindlimb. The pattern shift diagram clearly illustrates that *sox9* gene expression in metacarpal I in the wing appears late compared to hindlimb metatarsal I. From the pattern shift diagram is also visible that no subsequent gene expression for *bmpr1b* and no cartilage formation are observed in metacarpal I in the wing. In the primordia of wing digit II – IV, foot digit I-IV and rudimentary foot digit V, *bmpr1b* expression is found; these primordia develop into fully ossified digits. This is in full agreement with previous studies (e.g. Burke and Feduccia 2002; Larsson and Wagner 2002; Kundrat et al. 2002; Welten et al. (2005)).

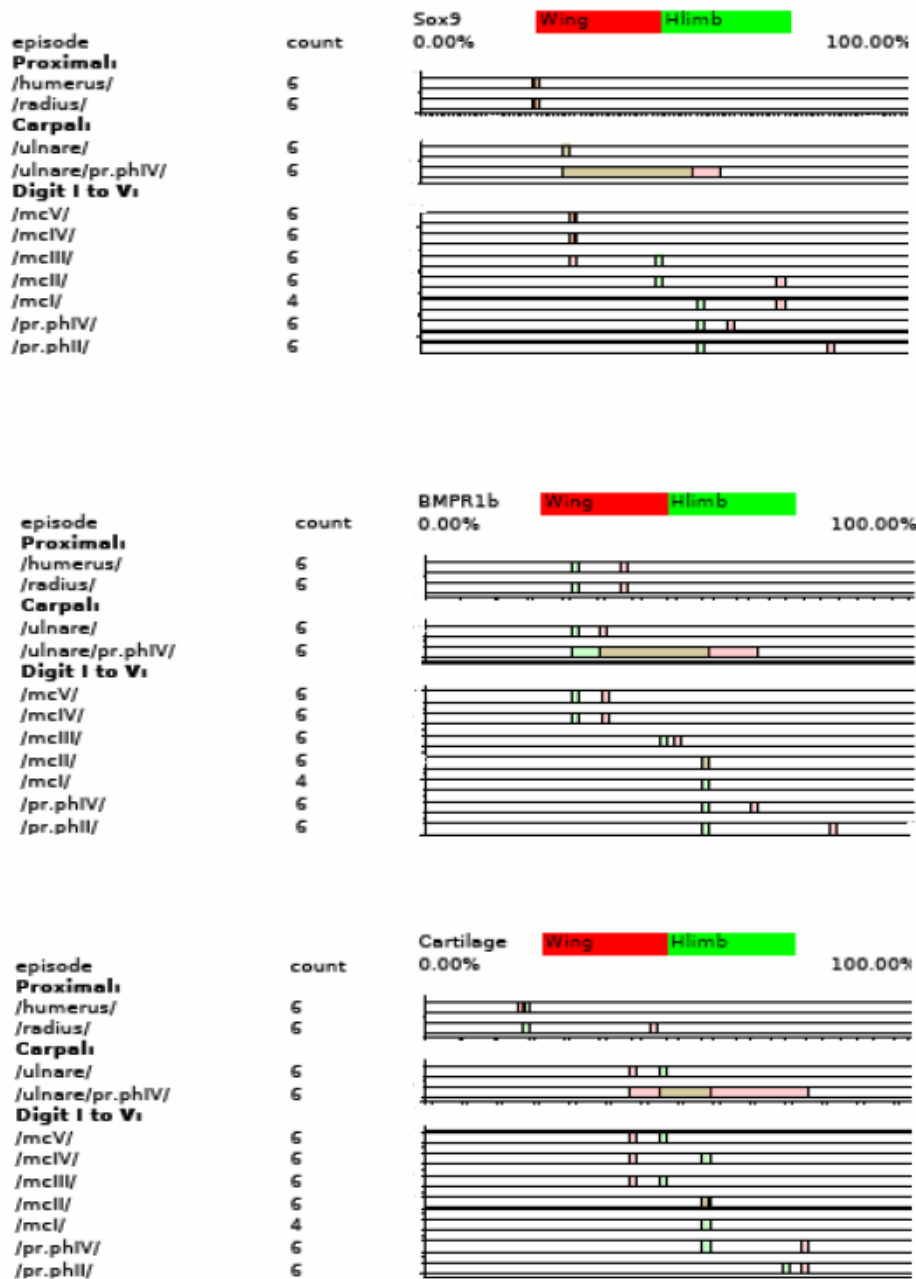


Fig.3

**Fig. 3.** Pattern shift diagram showing episodes, frequently found sequences (Bathoorn et al, in preparation) of skeletal elements plotted against duration of gene expression, using a relative time scale of chicken limb development. The diagram shows a selection from the total analysis of gene expression timing and cartilage formation in chicken wing and hind limb. Onset and duration of gene expression as well as alcian blue staining for cartilage formation are shown in the red bar (wing) and in the green bar (hind limb). Structures are organized in proximal, carpal and digit region. The pattern shift diagram shows that only *sox9* gene expression is found in the presumptive wing metacarpal (mc) I while no subsequent *bmpr1b* expression and cartilage formation are found in the presumptive wing mc I. The onset of gene expression in presumptive wing mc I is relatively late compared to metacarpals and proximal phalanges of digit II and IV. The pattern shift diagram clearly illustrates that the difference in number of digits between wing and hind limb skeleton is the result of a time shift in gene expression. mc metacarpal, mt metatarsal, ph phalanx, pr.ph proximal phalanx. Metacarpals / tarsals and digits are indicated with roman numerals (cf. Welten et al., 2005).

## DISCUSSION

Several methods have previously been developed for the analysis of heterochrony in a phylogenetic context. Methods include event pairing (Smith, 2002, Jeffery et al, 2002 and 2005), search – based character optimization (Schulmeister and Wheeler, 2004). These methods have been presented based on morphological criteria. In this study, frequent episode mining is used to show that evolutionary relations can be reconstructed on the basis of gene expression data, morphological criteria (Bathoorn et al, 2006), temporal data and combinations of all these data.

First, we present a tree showing the phylogenetic relations of five laboratory model organisms that is produced with our frequent episode-mining algorithm. The reconstruction is based on differences in timing of gene expression in the selected model organisms. The phylogenetic relationships of these model systems are evident from literature; from this evidence we were able to validate our algorithm. In order to be able to compare development, a relative time scale of development from fertilization (0%) to juvenile (100%) is made part of the analysis (Bathoorn et al, 2007). The phylogenetic tree, based on gene expression data and produced from the results of FEDA, is consistent with phylogenetic trees described in literature. The differences at the nodes of the cladogram produced by FEDA are consistent with literature. Although most genes are present in all five model species, onset and duration of expression of these genes show a wide variation between species. For instance, the genetic pathway responsible for limb initiation, patterning and outgrowth is found to be present in all 5 model species. In mouse, chicken, axolotl and zebrafish, *fgf8* expression occurs in pre-limb bud stages. In *Xenopus* however, *fgf8* is expressed at later stages (Christen and Slack, 1997). *Tbx5* expression, which occurs in the forelimb only, is delayed in *Xenopus* since forelimb develops later than hindlimb in *Xenopus* (Kahn et al, 2002).

Analysis of differences at the nodes of the cladogram provides the possibility to predict presence of gene activity (gene expression) in a model species, when no data are available for a particular species. In our study, markers from the molecular cascade of chondrogenesis (*sox9*, *bmpr1b*, *runx2*) are found in most of the species. For zebrafish,



*bmpr1b* data were absent in a certain episode. From literature, no zebrafish *bmpr1b* gene expression data are available in relation to chondrogenesis; only gene expression patterns from younger stages have been described (Nikaido et al, 1999). FISH experiments for our case study exhibit *bmpr1b* expression in the pectoral fin: most likely in the endoskeletal disc and other skeletal structures. Recent images from large – scale screening also display *bmpr1b* expression in the pectoral fin (Thisse et al.; [www.zfin.org](http://www.zfin.org)). Further experiments are needed to investigate the function of *bmpr1b* in zebrafish and its possible role in zebrafish chondrogenesis. In *case study 1* we demonstrate that FEDA is capable of revealing presence of gene expression if no data are available for a particular species. FEDA can be tuned to specifically output such prediction. In our application, the prediction was not focused on. The more gene expression data are available for FEDA, the higher the predictive power of our method will be.

In the phylogenetic tree in Fig. 1, the model species are clustered in different groups. The clustering in the cladogram (Fig. 1) is based on Jaccard distance expressing dissimilarity between species (Bathoorn et al., in preparation). The cladogram shows that chicken and mouse in the amniote cluster display lower numerical dissimilarity than the dissimilarity computed for *Xenopus* and axolotl in the amphibian cluster; according to existing established knowledge. Amphibians show more diverse body plans than amniotes in general. Moreover, amphibians hatch at a relatively early stage, so that juveniles are more subject to evolutionary changes (Bininda- Emonds et al, 2003).

In contrast to methods previously described, application of the FEDA algorithm produces a numerical value to express the rate of difference between species i.e., the Jaccard distance. This distance measure helps to clearly quantify differences between species in the context of gene expression.

The relative time scales we implemented in this study are based either on the developmental stages in which the limbs or fins develop or on the stages in which cartilage develops in the skeletal structures of hind and forelimb. Alternatively, somite number can be used as a time scale indicator on which timing of limb or fin development can be mapped (Richardson, 1995). In our study we chose a relative time scale from fertilization to juvenile, since in some model species, limbs develop long after completion of somitogenesis (Richardson, 1995). FEDA has shown to be able to deal with the different relative timing schemes that we used in the case studies. In future applications this flexibility should be further explored.

In general, lack of data might influence the accuracy for construction of phylogenetic trees. For some species, spatio-temporal gene expression data available from literature are absent or sparse. FEDA is capable to find a solution despite of sparseness or complete lack of data, since it uses episodes, collections of events (Bathoorn et al, in preparation). Missing an event, only means that the episode cannot be found for a particular species, but it can still be found for the other species. This feature again emphasizes the flexibility in data arrangement of FEDA.

In the lizard genus *Hemiergis* digit reduction is found in several closely related species (Shapiro, 2002; Shapiro et al. 2003). The mechanism underlying this phenomenon cannot simply be explained by heterochrony of the process of chondrification and ossification alone but rather by the molecular cascades underlying these processes. When timing of cartilage formation was compared between lizard species, cartilage elements in the digits were formed at approximately the same stages until digits were fully developed (Shapiro,

2002). On the level of gene expression however, a difference in duration of exposure to *shh* expression was observed among the four *Hemiergus* species (Shapiro et al. 2003). The limb buds of the five-fingered *Hemiergus initialis* displayed a later offset of *shh* expression than those of *Hemiergus quadrilineatus*, which has only two digits. This indicates that heterochrony may be found at different levels of developmental processes (Shapiro et al. 2003). In like manner, we analyzed heterochrony in gene expression as well as in cartilage formation in chicken wing and hindlimb digits (cf. *case study 2*). In *case study 2* FEDA is used to analyze heterochrony at different views on developmental processes; i.e. gene expression, organogenesis, morphology. Therefore, FEDA was applied to investigate onset of gene expression in future skeletal structures in the autopodium (hand or foot) of one species. Additionally, timing of cartilage formation was analyzed. With the pattern shift diagram (Fig.3) we demonstrate that FEDA is capable to display heterochrony in various developmental processes.

Different morphological features in fore and hindlimbs of one species may be the result of a shift in gene expression timing in one of these regions in relation to the other. A wide range of genes is involved in digit patterning and specification of digit identity (Dahn and Fallon 2000, Litingtung et al. 2002; Shapiro et al. 2003; Tiecke et al. 2007). It has been shown that *shh* has a direct role in the regulation of *sox9* (Tavella et al. 2004). For our analysis, more gene expression data are needed to further investigate whether heterochrony in *sox9* expression observed in Fig. 3 is caused by changes in timing of gene expression of other genes such as *shh*.

In all cases, an existing dataset can still be used when adding more species to the dataset. Only a new computation for these new species needs to be performed. When more genes are added to the dataset, re-evaluation of the complete dataset is required. In this manner, our method is capable of dealing with large datasets. This is an important characteristic since the amount of gene expression data available from large genomic screens, functional genomics and molecular genetics, is still growing. Analyzing the gene expression patterns in the manner presented will render new insights.

#### *Conclusions and future work*

Our two case studies show that heterochrony in gene expression in limb-fin development can be studied very well using the computational approach provided by FEDA. We have shown that heterochrony in gene expression between species (case study 1) and within species (case study 2) was elucidated in both cases. The frequent episode mining on the data that we have aggregated for our studies was sufficient to analyze limb-fin heterochrony in gene expression in both a phylogenetic framework and in relation to different time scales. The development of a relative time scale was crucial to the outcomes presented for this studies; it supported heterochrony analysis of gene expression with FEDA and allowed flexibility in data preparation (Bathoorn et al., in preparation).

We have used standard visualization methods for the phylogenetic analysis; for the pattern shift study a new method was introduced to visualize the results of FEDA. With this visualization method, heterochrony can be revealed at different levels of developmental process. This is a potentially useful tool in evo-devo research problems.

FEDA has shown to be a very powerful and flexible method. Data can be rearranged such that several features can be compared in the analysis; i.e., gene expression, morphological

structures, or species. This flexibility contributes to utility in the analysis of developmental differences such as heterochrony. FEDA computations result in numerical values to express difference between species, i.e., Jaccard distance (cf. figure 1), which represents an objective measure of dissimilarity between entities in data.

FEDA can be engaged in the prediction of gene expression, when data are incomplete.

The computational framework of FEDA will be employed in future work in the analysis of spatiotemporal gene expression data available from structured repositories (Belmamoune and Verbeek, 2006). In the analysis presented in this paper data were based on *in situ* gene expression data; a next step is to include data from microarray in the analysis. This next step will provide new possibilities in the heterogeneous and comprehensive analysis of genetic networks and evolutionary relationships.

#### ACKNOWLEDGEMENTS

This project is partially supported by The Netherlands Research Council through the BioMolecular Informatics programme of Chemical Sciences (grant number # 050.50.213). We thank M.A.G. de Bakker and G.E.M. Lamers for their valuable advice, and O. Ayachi, E.M. Dondorp and R.T. Schoon for their help with the *in situ* hybridizations. We also thank M. Akimenko, P.W. Ingham, C. and B. Thisse, J. Postlethwait, N. Ueno, C. Shukunami and J. Bakkers for kindly providing us the cDNA clones for *msxb*, *tbx5*, *fgf8*, *sox9*, *bmpr1b*, *chondromodulin-1*, and *myoD*, respectively.