



Universiteit
Leiden
The Netherlands

Spatio-temporal framework for integrative analysis of zebrafish development studies

Belmamoune, M.

Citation

Belmamoune, M. (2009, November 17). *Spatio-temporal framework for integrative analysis of zebrafish development studies*. Retrieved from <https://hdl.handle.net/1887/14433>

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/1887/14433>

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

CHAPTER 1

GENERAL INTRODUCTION

1.1 Introduction

The specific result of vertebrate embryonic development is the progression of structures over time, from a first apparition during the developmental process to mature structures (complex organs). Throughout such developmental process genes are expressed in complex and constantly changing anatomical patterns. For anatomists, it is critical to understand how such anatomical structures function, how they change to complex shapes and which genes are involved in such changing patterns. Bioinformatics is the science that focuses on the development and application of computational methods to organize, integrate, and analyze biological-related data to facilitate the workflow for biologists. In this context we developed a spatio-temporal framework for developmental studies. A spatio-temporal reference framework of standard anatomical information and patterns of genes expression is an important tool for any experimental organism in which form and function are of interest for developmental biology. The study of anatomy is an essentially three-dimensional (3D) attempt. Therefore, to increase the value of such spatio-temporal framework, data should describe the complex relationship between tissues in three-dimensional (3D) format.

The aim of the research described in this thesis is to establish an integrative 3D spatio-temporal framework with standard anatomical information (3D digital atlas) and gene expression information (3D *in situ* patterns of marker genes) for developing zebrafish embryo; this framework has to be designed in such a way to be transposed to other model systems.

The 3D atlas of zebrafish development is a digital representation of zebrafish embryo anatomy. It provides a standardized coordinate system to analyze patterns of gene expression. The 3D atlas contains 3D digital embryos resulting from 3D reconstruction from serial sections, i.e. section images at representative stages of zebrafish development. Each of the section images is segmented in anatomical domains. Each anatomical domain

(or structure) is annotated with a graphical contour (graphical annotation). This graphical annotation enables to detect the 3D outline of the annotated structures. Furthermore, to each structure in the 3D atlas an anatomical name is assigned (textual annotation) (cf. chapter 3).

The process of gene expression refers to the event that transfers the information content of the gene into the production of a functional product, usually a protein. To be valuable for developmental studies, a gene expression information resource should be documented by its temporal (when) and spatial (where) information. The experimental conditions (how) must also be part of the documentation process for an accurate interpretation of experimental observations. We followed this workflow to manage zebrafish 3D patterns of gene expression in the Gene Expression Management System (GEMS, cf. chapter 4).

We established the GEMS that contains gene expression patterns organized and published to be readily accessed. Efforts are also ongoing in other model systems yielding to a large selection of gene expression databases such as MEPD (Henrich et al, 2005) for medaka and ZFIN (Zebrafish Information Network; <http://zfin.org>) for zebrafish. In the work presented here, we focused on 3D patterns of gene expression of zebrafish. This data is 3D with a spatio-temporal characteristic that provides the relation between gene expression (at a molecular level) and tissue differentiation (at an anatomical level). Such 3D representation of gene expression patterns gives molecular definitions for developmental components.

Patterns of gene expression are generated by *in situ* experiments, i.e. ZebraFISH experimental protocol (Welten et al, 2006) and the Confocal Laser Scanner Microscopy (CLSM). The patterns are 3D images basically serial optical sections carrying the spatio-temporal information of the expressed genes. The images are initially submitted as raw data to the GEMS. This enables new raw data to be readily added and integrated with other information in the database. Moreover, raw data can always be processed for

presentation according to the user's needs. Furthermore, the 3D format of the patterns enables a detailed visualization and analysis of the spatial information of the expression patterns. For valuable framework, the challenge is to map gene expression data into the atlas. A key element will be a standard anatomical nomenclature for data description in both the 3D atlas and *in situ* gene expression data.

Bioinformatics has successfully demonstrated new approaches by computationally integrating various data sets such as by using standard descriptions, e.g. ontologies to annotate collected data. Data integration is defined as the process that combines data residing at different database systems and providing users with a unified view of these data (Lenzerini, 2005). Data integration has proven to be an effective strategy to extract biological meaning from heterogeneous data sets in both developmental research and other fields. In our research we applied this principle of data integration and we developed the Developmental Anatomy Ontology of Zebrafish (DAOZ, cf. chapter 2).

The DAOZ is a key component of our information systems. It is a dictionary of anatomical terms derived from the staging series of (kimmel et al, 1993). Terms from the DAOZ are assigned to anatomical domains in the 3D atlas and are used by the GEMS as the standard nomenclature for data annotation and retrieval. This assignment represents the critical link between the atlas and the gene expression database (cf. chapter 5). The anatomical terms in the DAOZ are modeled hierarchically in different degrees of granularity. This data modeling enables complex queries to be readily performed for an intuitive data access and analysis.

Patterns of gene expression are organized in the GEMS with a standardized and structured manner and GEMS database was coupled to a 3D atlas of zebrafish development and to the DAOZ. Additionally, we added another component to our framework, i.e. the 3D visual query system (cf. chapter 5) that we developed to link the other components, i.e. the 3D atlas, GEMS and DAOZ (cf. Figure 1).

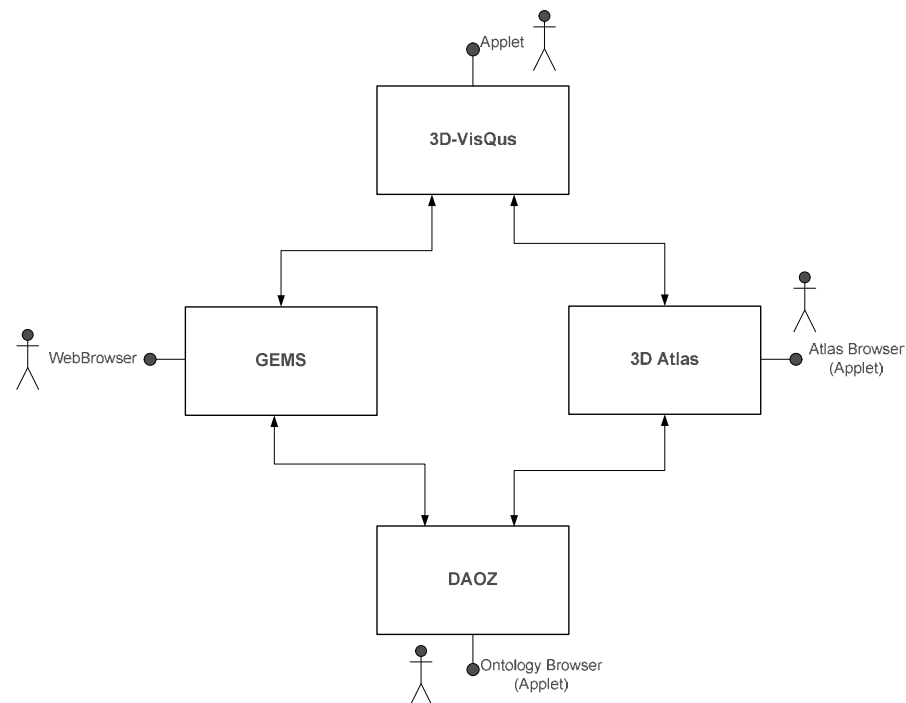


Figure 1: Diagram of the different components of our framework to study zebrafish development. Users can interact with each component through a user interface.

The role of bioinformatics is extended to uncover the wealth of biological information hidden in the mass of produced biological data and to obtain a clearer insight into the biology of organisms and to use this information to enhance the scientific benefit. In this context mining techniques were applied on gene expression data stored in the GEMS (cf. chapter 6).

With our spatio-temporal framework we introduced novel mechanisms for anatomical information storage and retrieval by using their spatio-temporal characteristics and we

make these mechanisms available to the research community in the form of novel bioinformatics tools. These tools, database systems, enable patterns of gene expression to be analyzed within a spatial and temporal context consistent with the spatial and temporal developmental concept of the organism. These resources should be seen as a tool for the developmental research community to put gene expression data into the proper biological and analytical context, so that the developmental dilemma can successively be understood.

1.2 The Zebrafish Spatio-Temporal Framework

In this part we will present the zebrafish model organism and the different components of our spatio-temporal framework for the embryonic development of zebrafish.

1.2.1. Zebrafish as a model organism

During the last decades, the zebrafish (*Danio rerio*) has become an important model organism in scientific research. Zebrafish offers a powerful combination of low cost, transparent embryo that develops rapidly outside the mother's body. Moreover, the study of zebrafish developmental genetics has proven valuable results in determining many aspects of vertebrate development. Further using this model organism promises to generate many interesting and useful data. Increasingly, it is recognized as a useful organism for human genetic and diseases modelling. The increasing use of zebrafish as a model system to study human disease has necessarily generated interest in the anatomy of this species at different developmental stages to map the many key aspects of organ morphogenesis that take place.

1.2.2 Developmental Anatomy Ontology of Zebrafish

All along of our research was the principle of data integration applied. An aspect of integration is to make databases integrated. This integration can be achieved if data in different databases are annotated with a common terminology. Ontological concepts are usually applied to provide such common terminology for data annotation in a structured way. These concepts enable therefore information sharing among different information systems.

The Developmental Anatomy Ontology of Zebrafish (DAOZ) is the anatomical ontology that we developed on behalf of our project. The anatomy is modelled hierarchically from body region to organ to structure in order to fit with the different degrees of abstraction in data capture and analysis. To the anatomical and temporal concepts we introduced new concepts of spatial and functional characteristics. In addition, we used different relationships to link DAOZ concepts with each other, i.e. aggregation, composition and association relationships. These relationships provide the opportunity for more complex queries to be performed. The anatomical terminology of the DAOZ is the same as this used inside the zebrafish community. Therefore, data annotated with concept from the DAOZ ontology can be linked to each other and to other resources and more importantly to the ZFIN resources.

1.2.3 The 3D atlas of zebrafish

Core to our efforts is the 3D digital atlas of zebrafish development. The atlas is representing standard embryonic development of the zebrafish. For a number of canonical developmental stages, 3D models are generated and organized in a database system. This database contains three kinds of information: digital images, referred to as section images, graphical annotation of anatomical domains in these section images and text-based descriptions of the anatomical domains. The 3D digital atlas of zebrafish is unique because of its 3D data which is represented within a spatio-temporal context. Each

3D model is the result of a 3D reconstruction from serial sections, i.e. 2D section images. Each section image is segmented in anatomical structures that are annotated graphically and semantically.

Users can access the atlas 3D models through a web-application. This application provides a portal interface to access complex anatomical data of the 3D atlas. Users can query the 3D atlas database without a prior knowledge of the exact anatomical terms. 3D models are, on the fly, assembled and presented according to the user requested queries.

1.2.4 The gene expression data

Following the atlas, we designed and implemented the GEMS. Results from gene expression experiments sometimes redefine anatomical borders through patterns of gene expression. GEMS is a key element of our framework to study embryonic development. It is a database system for managing, linking and mining spatio-temporal patterns of gene expression in zebrafish. The patterns of gene expression are obtained from, but not restricted to, 3D images generated with the zebraFISH protocol (Welten et al, 2006) combined with the Confocal Laser Scanning Microscope (CLSM). The CLSM images show the expression domain, some surrounding tissues and the outline of the embryo. These images offer a precise approach to define gene expression domains based on reference models. These patterns of gene expression are therefore, intended to be mapped to models of the 3D digital atlas. Consequently, we applied systematic methods to manage patterns of gene expression within an integrative spatio-temporal context. The GEMS is publically accessible for data submission and inspection. Hence, integration with other resources is a key issue. The GEMS provides some level of integration with other bioinformatics resources on the Internet such as ZFIN. Moreover, integration with other model system is easier to realize.

1.3 Outline of the thesis

The work presented in this thesis is based on a number of publications in scientific journals and international conferences. Here is an overview of the chapters discussed in this thesis and their related publications.

Chapter 2 describes the Developmental Anatomy Ontology of Zebrafish. This ontology contains anatomical description of the zebrafish over time. In this chapter we will discover how the anatomical concepts have been organized as an ontology. We will also shed light on how this ontology has been translated into a database to facilitate its presentation but more importantly to facilitate its task for annotation. This ontology was initially presented in:

- Y. Bei, M. Belmamoune and F. J. Verbeek, Ontology and image semantics in multimodal imaging: submission and retrieval, Proc. of SPIE Internet Imaging VII, Vol. 6061, 60610C1 C12, 2006.

A complete description of the ontology was published in:

- M. Belmamoune and F.J. Verbeek. Developmental Anatomy Ontology of Zebrafish: an Integrative semantic framework. Journal of Integrative Bioinformatics, 4(3):65, 2007.

In Chapter 3 the 3D digital atlas of zebrafish development is presented. This chapter is partially published in:

- S.A. Brittijn, S.J. Duivesteijn, M. Belmamoune, L. F.M. Bertens, W.B., J.D. de Bruijn, D.L. Champagne, E. Cuppen, G. Flik, C.M. Vandenbroucke-Grauls, R.A.J. Janssen, I.M.L. de Jong, E.R. de Kloet, A. Kros, A.H. Meijer, J.R. Metz, A.M. van der Sar, M.J.M. Schaaf, S. Schulte-Merker, H.P. Spaink, P.P. Tak, F. J.

Verbeek, M.J. Vervoordeldonk, F.J. Vonk, F. Witte, H. Yuan and M.K. Richardson. Zebrafish development and regeneration: new tools for biomedical research. *Int. J. Dev. Biol.* (2009) 53: 835-850.

An advanced description of the 3D digital atlas of zebrafish development is presented in:

- M. Belmamoune, L. Bertens, D. Potikanond, R. v.d. Velde and F. J. Verbeek. The 3D digital atlas of zebrafish: 3D models visualization through the Internet. (Submitted, 2009).

In Chapter 4 we will present the Gene Expression Management System (GEMS). During embryonic development of the zebrafish, patterns of gene expression of marker genes are visualized from, but not restricted to, *in situ* hybridization experiments in combination with Confocal Laser Scanner Microscopy (CLSM). In this chapter we provide information about mechanisms of these patterns storage and retrieval. We will also give more details about the system design and implementation. The work presented here was initially published in:

- M. Belmamoune and F. J. Verbeek. Heterogeneous Information Systems: bridging the gap of time and space. Management and retrieval of spatio-temporal Gene Expression data. InSCit2006 (Ed. Vicente P. Guerrero-Bote), Volume I "Current Research in Information Sciences and Technologies. Multidisciplinary approaches to global information systems", pp 53-58, 2006.

The complete work has been published in:

- M. Belmamoune and F. J. Verbeek, Data Integration for Spatio-Temporal Patterns of Gene Expression of Zebrafish development: the GEMS database. *Journal of Integrative BioInformatics*, 5(2):92, 2008.

We will present the 3D Visual query system (3D-VisQus) in Chapter 5. This system maps standard phenotype data in the 3D digital atlas of zebrafish with genotypic data in the Gene Expression Management System. The 3D-VisQus enables 3D models of the zebrafish embryo to be viewed, browsed and queried. From a visualized element in a 3D model, a user can send a visual query to the GEMS. Questions in the kind of how this system works and how it has been designed and implemented could be further answered in Chapter 5. This chapter is based on an early publication:

- M. Belmamoune, E. Lindoorn and F. J. Verbeek. 3D-VisQuS: A 3D Visual Query System integrating semantic and geometric models. InSCit2006 (Ed. Vicente P. Guerrero-Bote), Volume II "Current Research in Information Sciences and Technologies. Multidisciplinary approaches to global information systems", pp 401-405, 2006.

To further analyze gene expression data that are present in GEMS, mining workflows have been developed. We choose for association rules techniques to investigate the mining workflow services offered by the GEMS framework. Association rules techniques have been applied to uncover possible relations between genes. Association patterns are extracted from the GEMS database and could be directly integrated with each other for a primary comparison and analysis. The uniform annotation of the gene expression data with formal ontological metadata enables cross-reference with other resources. Therefore, cross-model system comparative studies and analysis of gene expression patterns is facilitated. For more details refer to Chapter 6. This chapter is based on the following paper:

- M. Belmamoune and F. J. Verbeek. Mining the zebrafish 3D patterns of gene expression database for association rules. (Submitted, 2009).

In Chapter 7 discussions and conclusions are presented. Also a summary in Dutch is presented.