



Universiteit
Leiden
The Netherlands

Fish genomes : a powerful tool to uncover new functional elements in vertebrates

Stupka, E.

Citation

Stupka, E. (2011, May 11). *Fish genomes : a powerful tool to uncover new functional elements in vertebrates*. Retrieved from <https://hdl.handle.net/1887/17640>

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/1887/17640>

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

**Fish genomes:
a powerful tool to uncover
new functional elements
in vertebrates**

Elia Stupka

This work was carried out with support from the European Commission Framework VI grant TRANSCODE (LSHG-CT-2004-511990) as well support from A-STAR Singapore and Temasek Life Sciences Laboratory, Singapore

Fish genomes: a powerful tool to uncover new functional elements in vertebrates

PROEFSCHRIFT

ter verkrijging van de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 11 Mei 2011
klokke 16.15 uur

door
Elia Stupka

door

Geboren te Quartu Sant'Elena, Italy in 1977

PROMOTIE COMISSIE

Promotor

Prof. Dr. J.N. Kok

Co-promotor

Dr. Ir. F.J. Verbeek

Overige Leden

Prof. Dr. H.P. Spaik

Prof. Dr. J. Den Hertog

Dr. P. Sordino (Stazione Zoologica Anton Dohrn, Naples, Italy)

To the two shining stars in my life, Ann and Anais

To my guiding light, my grandmother Giuliana

To my grandfather Aurelio and his free spirit

Chapter 1: Introduction	8
Introduction	8
Fish as model organisms.....	8
Fish genomes.....	9
Comparative Genomics.....	10
Transcriptomics.....	12
Organization of the thesis	12
Bibliography	14
Chapter 2: Whole-Genome Shotgun Assembly and Analysis of the Genome of Fugu rubripes	16
Abstract	17
Introduction	18
Methods	19
Sequencing Methods.....	19
Assembly.....	22
Repeats and assembly.....	25
Annotation methods.....	29
Results	36
Whole-Genome Shotgun Sequencing and Assembly of the Fugu rubripes Genome	36
Preliminary Annotation and Analysis of the Fugu Genome.....	40
Introns in Fugu.....	54
Structuring of the Fugu Genome over Evolutionary Time.....	58
Comparison of Fugu and Human Predicted Proteomes.....	68
Conclusions	78
References and Notes	81
Chapter 3: Shuffling of cis-regulatory elements is a pervasive feature of the vertebrate lineage	90
Abstract	91
Introduction	92
Results	96
Identification of mammalian regionally conserved elements.....	96
Shuffling of conserved elements is a widespread phenomenon.....	99
Shuffled conserved regions cast a wider net of nongenic conservation across the genome.....	103
The proximal promoter region is a shuffling 'oasis'.....	105
Shuffled conserved regions are able to predict vertebrate enhancers.....	109
Shuffled conserved regions act as enhancers in vivo.....	110
Discussion	115
Widespread shuffling of cis-regulatory elements in vertebrates.....	115
Conservation versus function.....	117
Toward improved detection of cis-regulatory elements.....	124
In vivo transient assays.....	126
Mechanisms for genome-wide shuffling.....	127
Conclusion	129
Materials and methods	129
Selection of genes and sequences.....	129
Identification of mammalian regionally conserved elements.....	130
Identification of shuffled conserved regions.....	131
Gene Ontology analysis.....	131
Mapping of conserved elements.....	132

BLAST versus CHAOS comparison	132
Overlap analysis	133
Identification of control fragments	133
Zebrafish embryo injections	133
Analysis of transgene expression	134
Acknowledgements	136
References	136
Chapter 4: The TATA-binding protein regulates maternal mRNA degradation and differential zygotic transcription in zebrafish	146
Abstract	147
Introduction	148
Results	150
TBP regulates specifically a subset of mRNAs in the dome-stage embryo	150
Most TBP activated genes are dynamically regulated during zebrafish ontogeny	154
TBP dependence of transcription from isolated zebrafish promoters	156
TBP is required for degradation of a large number of maternal mRNAs	158
Identification of TBP-dependent maternal transcripts	160
TBP regulates a zygotic transcription-dependent mRNA degradation process	162
Degradation of maternal mRNA by the miR-430 microRNA is specifically affected in TBP morphants	164
Redundant and specific function of TBP in the activation of subsets of genes at MBT	167
Discussion	168
Redundant and specific function of TBP in the activation of subsets of genes at MBT	169
TBP limits certain gene expression activities in the zebrafish embryo	172
The mRNA degradation machinery active during maternal to zygotic transition requires TBP function	173
Materials and methods	175
Embryo injection experiments	175
Whole-mount in situ hybridisation and immunostaining	177
RT-PCR analysis of maternal mRNA degradation	177
Gene identification and statistical analysis of EST microarray data	177
Annotation of ESTs of the TBP microarray, in relation to the stage-dependence array and to the zebrafish genome	178
Degradation pattern of maternal transcripts	179
Identification of miR-430 targets among the genes of the TBP microarray	179
Acknowledgements	180
References	180
Chapter 5: Assembly of the carp genome	184
Abstract	185
Introduction	186
Results	187
Initial Dataset: pseudo-tetraploid material	187
Preliminary Genome Assembly	188
Haploid material assembly	189
Varying the K parameter in SOAPdenovo	190
Varying the L parameter in SOAPdenovo	193
Testing read trimming strategies	195
Testing combination of assembly softwares	198
Adding BAC end reads	198
Assembly Statistics	199

Largest scaffolds	201
Quality Assessment.....	203
Coverage of existing BAC clones	203
Coverage of all carp Genbank sequences	204
Gap Filling	207
Mitochondrial genome	208
RNA-Seq Analysis	209
Methods	212
Genome Assembly	212
QC Analysis.....	214
Graphical Reporting	215
Discussion	215
Initial pseudo-tetraploid ABYSS based assembly.....	215
Evaluation of ABYSS	216
Haploid DNA CLC Bio and SOAP de novo based assembly	216
CLC Bio Contig Assembly	217
The K parameter	218
Other SOAPdenovo parameters	218
BAC end reads.....	219
Assembly Assessment and QC.....	220
References	222
Chapter 6: Discussion	224
Impact of next-generation sequencing on genome research	224
Searching for regulatory elements	225
Transcriptomics.....	227
Genome Assembly	228
References	230