

THE C-VALUE ENIGMA

A Thesis

Presented to

The Faculty of Graduate Studies

of

The University of Guelph

by

T. RYAN GREGORY

In partial fulfilment of requirements

for the degree of

Doctor of Philosophy

September, 2002

© T. Ryan Gregory, 2002

ABSTRACT

THE C-VALUE ENIGMA

T. Ryan Gregory
University of Guelph, 2002

Advisor:
Professor Paul D.N. Hebert

This thesis is an investigation of the evolution of genome size in animals. The haploid nuclear DNA contents (genome sizes, or “C-values”) of eukaryotes vary more than 200,000-fold, and bear no relationship to organismal complexity or the presumed number of coding genes. The basis for this enormous variation in genome size has remained an unsolved puzzle in evolutionary biology for more than 50 years. The initial “C-value paradox” was solved with the discovery of non-coding DNA, but a new and multi-faceted “C-value enigma” has emerged. The thesis begins with a brief review of the history and central concepts of the puzzle, followed by a thorough discussion of the relationship between DNA content and cell size and the presentation of a mechanistic model which may account for it. Patterns of genome size variation in mammals, birds, amphibians, and insects are discussed, and the implications of non-coding DNA for developmental, physiological, and ecological characteristics are explored in these key groups. A novel method of DNA quantification (Feulgen image analysis densitometry) is also presented and used to provide several hundred new invertebrate genome size estimates. Finally, the implications of the C-value enigma for evolutionary theory are considered in the context of the hierarchical theory of macroevolution developed by palaeontologists. A set of appendices comprising a compilation of roughly 3,000 animal genome sizes is also provided in an effort to facilitate the ongoing study of the C-value enigma.

ACKNOWLEDGMENTS

The research described in this thesis has been generously supported by a University of Guelph Graduate Entrance Scholarship, the University of Guelph Alumni Doctoral Scholarship, two NSERC Post-Graduate Scholarships (PGS A and PGS B), and the NSERC grant of Paul Hebert. My sincerest thanks are extended to all those who nominated me for these awards or kindly provided letters of recommendation for the applications.

This thesis has been several years in the making, and would not have been possible without the positive input of a great many individuals. It seems inevitable that I will fail to thank everyone, so please accept my apologies in advance if you have been left off the list. First, I would like to thank the staff members in the Department of Zoology, whom I think deserve far more explicit recognition than they usually receive, especially Mary Anne Davis, Lori Ferguson, Janice Hall, Yoko Imai, Susan Mannhardt, Ian Smith, and Irene Teeter. I would also like to thank the staff of TugDOC, CISTI Source, and Inter-Library Loan who helped me to track down and acquire even the most obscure references.

I wish to express my deepest gratitude to all those colleagues from around the world who contributed specimens for DNA quantification: Coenraad Adema, Maydianne Andrade, Shelley Ball, Enrico Bernard, Jim Bogart, Nina Brooke, Barbara Campbell, Craig Cary, Mary Ellen Clark, Melania Cristescu, Ron Davies, Jeremy deWaard, Brock Fenton, Marc Freeman, Milton Gallardo, Sean Gregory, Miodrag Grbic, Steve Haddock, Paul Hebert, Linda Holland, Petr Horak, Chris Jastrebski, Jean Joss, Richard Liversage, Steve Marshall, Mark Martindale, Yoshio Masui, Seanna McTaggart, Antonia Monteiro, Francesco Nardi, Oldich Nedvid, Elizabeth Neves, Frank Nilsen, Mike O'Donnell, Weste

Osbrink, Pierre Pontarotti, Edward Ross, Bill Starck, Jennifer Stewart, David Weisblat, Kelly Wells, Tony Wood, and the staff of Vermitech Systems Ltd. (Toronto), Natural Insect Control (Niagara Falls), and the Arkell Poultry Research Facility (Guelph).

Species identifications were a crucial part of the original genome size measurements reported in this thesis, and I am extremely grateful to Marvin Gunderman, Paul Hebert, Steve Marshall, and David Shorthouse for doing most of the work in this area.

I would like to thank all those who provided comments on manuscript drafts, shared unpublished data, offered technical assistance, or gave other advice: Coenraad Adema, Jaume Baguña, Mike Bennett, Roberto Bertolani, John Bickham, David Dixon, Scott Edwards, Matt Gage, Dan Graur, Joseph Hoffman, Spencer Johnston, Lloyd Kinzer, Lana Manley, Rachel Mueller, Eduard Petitpierre, Dmitri Petrov, Ellen Rasch, Robert Ricklefs, Alan Schulman, Adrian Sumner, Alexander Vinogradov, David Wake, Eric Waltari, Grace Wyngaard, and various journal editors and anonymous reviewers.

Many people have contributed to my academic development over the years, and I want to thank all the teachers, from pre-school through to graduate school, who encouraged me, had faith in my abilities, and allowed me to go about things in my own idiosyncratic way (which, I am well aware, seldom makes their jobs easier). I single out for special mention M. Luc Mayer at École Samuel de Champlain, Mme. Mudry, Mrs. Russell, and Mr. Powell at O.D.C.V.I., Bennett Galef at McMaster University, and Ron Brooks and Michael Ruse at the University of Guelph. Chris Wood gave me my first shot at independent research and my first introduction to the world of scientific investigation, something I will never forget. Thanks also to Roy Danzmann and Judy Wearing-Wilde for

giving me the opportunity to present my first set of lectures.

Special thanks are also due to those who have served on the various committees required for passage through the graduate school gauntlet. My graduate committee, Jim Bogart, Teri Crease, Paul Hebert, and Brian Husband reminded me that “you had better figure out what you will do for a thesis”, and then paid the price in the form of an initially herculean tome! My qualifying examination committee, Brian Husband, Teri Crease, Paul Hebert, Denis Lynn, and John Roff made the experience enjoyable, especially when one of them asked the one question no one has been able to answer: “Is there a question in there somewhere, Prof. Hebert?”. And finally, my defence committee, Ron Brooks, Teri Crease, Paul Hebert, Donal Hickey, and Denis Lynn provided inputs which are very much responsible for the final format of this thesis. I especially wish to thank Teri Crease for all that she has done for me.

Thank you to all my friends and labmates, especially Shelley Ball, John Colbourne, Andrea Cox, Melania Cristescu, Alison Derry, Jeremy deWaard, Rob Dooh, Seanna McTaggart, Erin Penton, Heather Pollard, Cheryl Prokopowich, James Rhydderch, Chad Rowe, Kelly Wells, Jonathan Witt, Tony Wood, and the other past and present members of the Hebert lab.

It has been a privilege and a pleasure to get to know Grace Wyngaard and Ellen Rasch, both of whom deserve my most heartfelt thanks for their friendship, advice, support, and extraordinary hospitality.

David Hardie has been a close colleague and friend, and without his contributions much of the work in this thesis would have been impossible. Thanks for everything, Zeke!

Paul Hebert first inspired me with a talk given at McMaster, and introduced me to

the topic of genome size evolution. He has provided invaluable guidance and support throughout my time in Guelph while simultaneously giving me an essential level of freedom. This is not the thesis either one of us would have planned from the outset, but I hope it has become something of which he can be proud.

Sarah Adamowicz has been there to support me through the good times and the bad. I thank her for her patience and her wit, for always listening, and for inspiring me in ways she probably doesn't even realize. I could not ask for a better friend and partner.

This thesis is dedicated to my family – genetic and otherwise – who have always provided me with the highest degree of love and support. This includes my biological parents (Bob Gregory and Michele Davis) as well as my extended set of parents (Marilyn Gregory and Frank Brewster), my identical twin brother born four years later (Sean Gregory), and my feline children (Moe and Minnie).

**TABLE OF CONTENTS
(VOLUME ONE)**

Acknowledgments i

Table of contents v-xii

List of Tables xiii

List of Figures xiii-xv

Chapter One: Introduction to the C-value enigma

Abstract 2

Introduction 3

Genome size constancy and the C-value concept 4

Genome size variation and the C-value paradox 5

From paradox to puzzle 7

Mutation pressure theories 8

 Junk DNA 8

 Selfish DNA 11

Optimal DNA theories 12

 The nucleoskeletal theory 12

 The nucleotypic theory 13

Outline of the thesis 13

Concluding Remarks 15

Chapter Two: DNA content and the cellular phenotype

Abstract	20
Introduction: the importance of cell size	21
DNA content and cell size	22
The karyoplasmic ratio	22
Polyploidy and cell size	22
Genome size and cell size	23
Reptiles	24
Amphibians	25
Fishes	26
Birds	28
Mammals	31
Coincidence?	37
Size-dependent threshold	38
Overall increase in DNA content	38
Inability to delete extra DNA	39
Coevolution?	43
Nuclear size and the nucleoskeleton	43
Nuclear pores and RNA transport	45
Return and reversal of the karyoplasmic ratio hypothesis	47
The pros and cons of coevolution	51
Causation?	54
Challenges to the nucleotypic theory	54
In support of the nucleotype	57
Previous models of nucleotypic influence	61
The nucleotide sequestration model	61
The division-initiation model	62
The gene-nucleus interaction model	65
Eukaryotic cell cycle regulation	66
DNA content and cell cycle length	68
DNA content and cell cycle control	71
Issues awaiting resolution	78
Concluding remarks	80
Effect versus function	80

Chapter Three: Genome size evolution in mammals and birds

Abstract	104
Introduction	105
Summary of the dataset	106
Statistical analyses	107
Patterns of variation in mammals and birds	109
Genome size and chromosome number	112
Genome size, cell size, and metabolism in mammals	114
Mammalian metabolism: analysis and results	116
Mammalian metabolism: discussion	118
Genome size, cell size, and metabolism in birds	119
Avian metabolism: analysis and results	121
Avian metabolism: discussion	122
Genome size and flight	125
A causal connection?	125
Genomic baggage: lost or never loaded?	126
Genome size and developmental parameters	130
Sources of data	133
Avian development: dataset #1	134
Avian development: dataset #2	135
Mammalian development	138
Genome size, development, and body size in homeotherms	139
Concluding remarks	141

Chapter Four: Genome size evolution in amphibians

Abstract	155
Introduction	156
Summary of the dataset	156
Statistical analyses	157
Patterns of variation	157
Genome size and chromosome number	159
Genome size, cell size, and metabolic rate in amphibians	162
Cell size and cellular metabolism	162
Genome size and metabolic rate	163
Cell size and metabolic rate	165
Genome size and developmental complexity	167
The threshold concept: genome size and development in plants	168
Developmental complexity in amphibians	172
Developmental rate and constant complexity	172
Developmental process: direct development	175
Developmental process: neoteny	178
Developmental products: big genomes and simple brains	184
Development in amphibians: a summary	187
Amphibian genome size and the hierarchy of evolution	189

Chapter Five: DNA quantification by Feulgen image analysis densitometry

Abstract	202
Introduction	203
DNA quantification: past and present	204
Densitometry	204
Fluorometry	206
Image analysis densitometry	209
Basic concepts	209
Guidelines for specimen preparation: vertebrates	211
Fishes	212
Amphibians	213
Reptiles	214
Birds	214
Mammals	215
Guidelines for specimen preparation: invertebrates	216
Crustaceans	217
Insects	219
Arachnids	223
Myriapods	224
Annelids	224
Molluscs	226
Echinoderms	228
Flatworms	228
Nematodes	229
Cnidarians	229
Miscellaneous invertebrates	230
Staining methods	231
The Feulgen reaction	231
Stain preparation	232
Staining protocol	233
Measurement protocol	234
Hardware and software	234
Microscope set-up and image capture	234
Choice of standards and calculation of genome size	236
Concluding remarks	236

Chapter Six: Genome size evolution in terrestrial arthropods

Abstract	252
Introduction	253
Summary of the dataset(s)	254
Insects: previously published data	254
Insects: new genome size estimates	254
Spiders	255
Other arachnids	257
Myriapods	257
Patterns of variation in insects	258
Blattaria	258
Coleoptera	259
Collembola	264
Dermaptera	264
Diptera	264
Embiidina	268
Ephemeroptera	268
Hemiptera	268
Hymenoptera	271
Isoptera	272
Lepidoptera	272
Mantodea	274
Odonata	274
Orthoptera	274
Phasmida	276
Plecoptera	278
Siphonaptera	278
Thysanura	279
Trichoptera	279
Patterns of variation in arachnids	279
General patterns in spiders	279
A note on the spider mite	281
Patterns of variation in myriapods	281
Developmental complexity in insects: more on metamorphosis	282
Concluding remarks	286

Chapter Seven: Macroevolution, hierarchy theory, and the C-value enigma

Abstract	300
Introduction	301
Macroevolutionary theory for neontologists	302
What is macroevolution?	302
Critiques of the Modern Synthesis	304
Reductionism in biology	309
Group selection: new and improved and no longer naïve	312
The concept of individuality	315
Punctuated equilibria and species as individuals	318
Species selection in principle and in practice	321
Aggregate versus emergent characters	323
Emergent fitness versus the effect hypothesis (the Lloyd-Vrba debate)	325
Selection versus sorting	328
Hierarchical macroevolutionary theory: a summary	330
Molecular macroevolution	332
Are genomes “individuals”?	332
The necessity of hierarchy theory for understanding genomes (and vice versa)	337
Group selection and the origin of the genome	337
Selfish DNA and the necessity of hierarchy	339
The C-value enigma from a hierarchical perspective	342
Genomes as phenotypes and genotypes	343
The evolution and ecology of transposable elements	346
Selection, sorting, and genome size	351
Stasis and DNA constancy	353
Genome-level processes and the major transitions in evolution	356
Concluding remarks	365
References for Volume One	368

TABLE OF CONTENTS
(VOLUME TWO)

Appendix 1 – Reptile erythrocyte sizes	444
Appendix 2 – Amphibian erythrocyte sizes	447
Appendix 3 – Fish erythrocyte sizes	457
Appendix 4 – Bird erythrocyte sizes	468
Appendix 5 – Mammal erythrocyte sizes	485
Appendix 6 – Mammal genome sizes	507
Appendix 7 – Bird genome sizes	539
Appendix 8 – Amphibian genome sizes	555
Appendix 9 – Reptile genome sizes	602
Appendix 10 – Fish genome sizes	628
Appendix 11 – Insect genome sizes (database)	722
Appendix 12 – Insect genome sizes (original data)	745
Appendix 13 – Spider genome sizes	764
Appendix 14 – Crustacean genome sizes	772
Appendix 15 – Mollusc genome sizes	792
Appendix 16 – Annelid genome sizes	808
Appendix 17 – Flatworm genome sizes	819
Appendix 18 – Echinoderm genome sizes	823
Appendix 19 – Nematode genome sizes	827
Appendix 20 – Miscellaneous invertebrate genome sizes	829
References for Volume Two	837

LIST OF TABLES

Table 3.1 – Variance distribution of genome size in mammals and birds	144
Table 3.2 – Genome size, cell size, and metabolic rate correlations in mammals	145
Table 4.1 – Variance distribution of genome size in amphibians	192
Table 4.2 – Genome size and developmental rate in amphibians	193
Table 5.1 – Summary of slide preparation methods for animal specimens	238-240
Table 5.2 – Outline of Feulgen staining protocol	241
Table 5.3 – Explanation of steps in Feulgen staining protocol	242
Table 5.4 – Checklist of steps in image analysis protocol	243
Table 6.1 – Summary of hexapod C-value measurements	289-290

LIST OF FIGURES

Figure 1.1 – C-value ranges in eukaryotes	17
Figure 1.2 – Sequence composition of the human genome	18
Figure 2.1 – Red blood cells of vertebrates (Gulliver 1875)	82-83
Figure 2.2 – Red blood cells of polyploid salamanders	84
Figure 2.3 – Genome size versus erythrocyte size in vertebrates	85
Figure 2.4 – Red blood cells of vertebrates (photos)	86-87
Figure 2.5 – Genome size versus erythrocyte size in reptiles	88
Figure 2.6 – Genome size versus erythrocyte size in amphibians	89-91
Figure 2.7 – Genome size versus erythrocyte size in fishes (preliminary)	92
Figure 2.8 – Red blood cells of a teleost and a lungfish (photo)	93
Figure 2.9 – Genome size versus erythrocyte size in fishes (Hardie 2002)	94

Figure 2.10 – Genome size versus erythrocyte size in birds	95
Figure 2.11 – Genome size versus erythrocyte size in mammals	96
Figure 2.12 – Red blood cells of mammals (photos)	97
Figure 2.13 – Genome size versus epithelium cell size in mammals	98
Figure 2.14 – Genome size versus sperm cell size in mammals	99
Figure 2.15 – Cell cycle regulation in eukaryotes	100-101
Figure 2.16 – The Rab1 configuration of nuclear DNA	102
Figure 3.1 – Genome size distribution in mammal families	146
Figure 3.2 – Genome size distribution in bird families	147
Figure 3.3 – Genome size distribution in mammals and birds	148
Figure 3.4 – Genome size versus chromosome number in mammals and birds	149
Figure 3.5 – Genome size versus metabolic rate in birds	150
Figure 3.6 – Genome size distribution in bird families (flight ability)	151
Figure 3.7 – Genome size and flight ability in birds	152
Figure 4.1 – Genome size distribution in amphibian families	194
Figure 4.2 – Genome size versus chromosome number in amphibians	195
Figure 4.3 – Genome size versus metabolic rate in amphibians	196
Figure 4.4 – Erythrocyte size versus metabolic rate in amphibians	197
Figure 4.5 – Genome size and developmental lifestyle in plants	198
Figure 4.6 – Genome size and developmental lifestyle in amphibians	199
Figure 4.7 – Nucleated and enucleated erythrocytes of salamanders (photos)	200
Figure 5.1 – Blood smear preparation protocol	244

Figure 5.2 – Feulgen image analysis standard curve	245
Figure 5.3 – Hepatocyte, leukocyte, and erythrocyte nuclei (photos)	246
Figure 5.4 – Stain density versus DNA compaction level (photo)	247
Figure 5.5. – Sperm and haemocyte nuclei from insects (photos)	248
Figure 5.6 – Sperm bundle dispersion in moths	249
Figure 5.7 – Sperm and blood cell nuclei from annelids	250
Figure 6.1 – Genome size distribution in beetle families	291
Figure 6.2 – Genome size distribution in dipteran families	292
Figure 6.3 – Genome size distribution in hemipteran families	293
Figure 6.4 – Genome size distribution in lepidopteran families	294
Figure 6.5 – Genome size distribution in orthopteran families	295
Figure 6.6 – Genome size distribution in spider families	296
Figure 6.7 – Genome size and metamorphosis in insects (order means)	296
Figure 6.8 – Genome size and metamorphosis in insects (order ranges)	296
Figure 7.1 – Genomes in an expanded biological hierarchy	367