The Evolution of Beetle Bioluminescence

BY

John C. Day

A thesis submitted in partial fulfilment of the requirements of Oxford Brookes University for the Degree of

Doctor of Philosophy

School of Life Sciences
Oxford Brookes University
September 2009

Appendix iii not scanned at the request of the awarding university

ABSTRACT

The research presented in this thesis examines the evolution of bioluminescence in beetles. The study utilises two enzymes implicated in the bioluminescent reaction of fireflies and glow-worms; the protein luciferase and the luciferin regenerating enzyme (LRE). A range of parallagous genes for both enzymes were identified in bioluminescent and non-bioluminescent taxa using degenerate primer PCR. Combined with genomic and database sequences phylogenies were reconstructed using Bayesian and Maximum Likelihood approaches. The LRE share identity with a large family of enzymes including the senescence marker protein 30 (SMP30), gluconolanctonases and LRE (SGL) which included enzymes involved in ascorbic acid biosynthesis. Orthologous primary sequence comparisons revealed key conserved residues that formed a putative active site when mapped onto the tertiary structure of the bacterial SGL orthologue. As a result a putative active site was predicted for the LRE and key residues identified that may be involved in luciferin binding. The phylogeny revealed two main coleopteran SGL clades with species-specific paralogues present in both clades, which suggests early duplication in the Coleoptera and potential functional divergence of SGL enzymes. The highest degree of divergence was found between a subclade of group I (Ia) containing all the lampyrid LREs and group II suggesting functional divergence between these two classes of enzymes in the Lampyridae.

Phylogenetic reconstruction of luciferase and luciferase like paralogues in the Lampyridae identified one principle luciferase clade containing multiple novel genes from single species and three luciferase-like clades. In *T. castaneum* six EST-associated sequences were identified none of which were present in the luciferase clade. Multiple luciferase genes were identified from members of the Photurinae and the Luciolinae. The majority of luciferase residues were identified to be under purifying selection as opposed to positive selection. No luciferase paralogues were found in *Cantharis* species suggesting a loss of this gene paralogue in non-bioluminescent members of the Cantharoidea.

Finally, the potential for paralogues to be present in a genomic tandem array was investigated by genome walking upstream of the luciferase gene in the glow-worm *Lampyris noctiluca*. Although no luciferase paralogues were identified in 7 kb of upstream region the remnants of a retrotransposase was determined. The evolutionary implications of these findings are discussed in light of recent published research.

ACKNOWLEDGEMENTS

This work was undertaken at the Centre for Ecology and Hydrology in Oxford. I wish to acknowledge all past and present colleagues for the good working atmosphere and the numerous scientific discussions.

I wish to sincerely thank and express my gratitude to my supervisors Mark Bailey, at the Centre for Ecology and Hydrology (Oxford) and Robert Possee at Oxford Brookes University. I am grateful to them for their support and encouragement throughout this project. Both of them have been always been enthusiastic about my work and for many aspects the help they provided was beyond their duties.

I would like to extend my personal gratitude to James Hogan at the Oxford University Museum for providing and identify all the non-bioluminescent taxa studied in this thesis. Thanks are also extended to J. Chaichi, Gonçalo Figueira, James Lloyd and Alan Stewart for the kind donation of Lampyridae for this work. During the time engaged in this research, I enjoyed fruitful discussions with many people regarding bioluminescent beetles. In particular, I wish to gratefully acknowledge, Donald Barclay, Al Carlson, Raphaël de Cock, Jonathon Copeland, Michael Geisthardt, Lynn Faust and John Tyler for their contributions.

Additionally I would like to thank my examiners Alfried Vogler and Shirley McCready for valuable discussions and comments that have greatly enhanced this thesis.

Finally I would like to thank Michelle for her many years of support, patience and encouragement, without which this volume would have never reached its completed state.

To my father and the
memory of my mother

CONTENTS

Abstract ... ii
Acknowledgement ... iii
Dedication ... iv
Contents ... v
List of Figures ... viii
List of Tables ... xi
List of Abbreviations ... x
Publications Arising from Research ... xi

1. LITERATURE REVIEW ... 1

1.1 INTRODUCTION ... 1
1.1.1 Why study the evolution of bioluminescent beetles? ... 1
1.1.2 How valuable are bioluminescent beetles for scientific studies? ... 2
1.2 THE TAXONOMIC EVOLUTION OF BEETLES ... 2
1.2.1 Phylogeography ... 6
1.3 THE FIRST BIOLUMINESCENCE ... 6
1.4 FIREFLY LUCIFERIN ... 7
1.5 THE LUCIFERIN REGENERATING ENZYME (LRE) ... 9
1.6 SENESCENCE MARKER PROTEIN 30 (SMP30) ... 10
1.6.1 SMP30 and ascorbic acid synthesis ... 11
1.6.2 The role of SMP30 homologues in insects ... 12
1.6.3 SGL Protein Family ... 12
1.7 FIREFLY LUCIFERASE ... 13
1.7.1 Railroad worm and click beetle luciferases ... 15
1.7.2 Evolution of click beetle bioluminescence ... 16
1.7.3 Beetle luciferases and bioluminescence spectra ... 17
1.7.4 Adenylate-forming protein family ... 18
1.8 THE FUNCTIONAL EVOLUTION IN FIREFLIES ... 20
1.8.1 The evolution of dialogue in fireflies ... 20
1.9 NATURAL SELECTION THROUGH PREDATION: DEFENCE AND APOSEMATISM IN LAMPYRIDS ... 23
1.9.1 Chemical defence ... 23
1.9.2 Larval defence ... 24
1.9.3 Evolution of Aposematism ... 25
1.10 SUMMARY ... 26
1.11 AIMS AND CONTENTS OF THE THESIS ... 27

2. **ORIGINS AND DIVERSIFICATION OF THE SGL GENE FAMILY AND THE ORIGINS OF THE LUCIFERIN REGENERATING ENZYME (LRE) IN BIOLUMINESCENT BEETLES............................29**

2.1 INTRODUCTION 29
2.1.1 Aims 30
2.2 MATERIALS AND METHODS..31
2.2.1 Sequence data retrieval...31
2.2.2 Beetle material...32
2.2.3 DNA and RNA extractions...34
2.2.4 PCR amplification of the LRE gene from *Photinus pyralis* 34
2.2.5 Cloning and sequencing 34
2.2.6 Degenerate primer design and PCR 36
2.2.7 Full length gene characterisation using genome walking..................................37
2.2.8 RT-PCR amplification..38
2.2.9 Sequence alignment..38
2.2.10 Phylogenetic analysis 39
2.2.11 Test of functional divergence 39
2.3 RESULTS...40
2.3.1 Identification and characterisation of members of the SGL protein family 40
2.3.2 Photinus pyralis LRE gene 45
2.3.3 Intron/exon site conservation 46
2.3.4 Amplification of GSL domain using degenerate primers from 21 beetle taxa 46
2.3.5 Full length LRE genes in the genomes of *P. pyralis* & *L. noctiluca* 50
2.3.6 GSL genes in genome of *Pachnoda m. peregrina*..52
2.3.7 Sequence conservation in the SMP30 protein family...54
2.3.8 Phylogenetic distribution and relationships of the SMP30 protein family 57
2.3.9 Descriptive analysis of the SGL protein family subgroups 57
2.3.10 Functional constraints and divergence 61
2.4 DISCUSSION 64
2.4.1 SGL protein family genes and early origins...64
2.4.2 SGL protein gene evolution in Diptera 65
2.4.3 SGL gene duplication in coleopteran evolution 65
2.4.4 Function of SGL genes in insects...68

3. **THE EVOLUTION OF THE ADENYLATE-FORMING PROTEIN FAMILY IN BEETLES: EVIDENCE FOR MULTIPLE LUCIFERASE GENE PARALOGUES IN THE LAMPYRIDAE 71**

3.1 INTRODUCTION 71
3.2 MATERIALS AND METHODS...73
3.2.1 Identification of beetle luciferase and luciferase-like sequences..........................73
3.2.2 Degenerate primer PCR of bioluminescent beetles 73
3.2.3 Phylogenetic analysis 75
3.2.4 Analysis of positive selection...75
3.3 RESULTS...76
3.3.1 Phylogenetics of luciferase and luciferase-like sequences in beetles 73
3.3.2 Selection analysis 81
3.4 DISCUSSION 82
3.4.1 Clade B..82
3.4.2 Clade A..83
3.4.3 Gene duplication and selection...86

4. **LUCIFERASE FLANKING REGION IN THE GLOW-WORM *LAMPYRIS NOCTILUCA*......88**
4.1 INTRODUCTION 88
4.2 MATERIALS AND METHODS......88
4.2.1 PCR, cloning and sequencing......89
4.2.2 Sequence Analysis......89
4.3 RESULTS......90
4.3.1 Luciferase gene of Lampyris noctiluca 90
4.3.2 Motif search in upstream region of luciferase gene......92
4.4 DISCUSSION 93

5. **CONCLUSIONS 95**
5.1 THE ORIGIN OF BIOLUMINESCENCE IN BEETLES 95
5.2 GENE DUPLICATION IN BIOLUMINESCENT BEETLES 96
5.3 THE RELATIONSHIP BETWEEN GENE TREES AND SPECIES TREES 98
5.4 PEROXISOMAL ORIGIN OF BIOLUMNESCENCE 101
5.5 THE ORIGIN OF LUCIFERASE 101
5.6 LUCIFERIN SYNTHESIS 103
5.7 EXPRESSION AND FUNCTIONAL STUDIES 104
5.8 CONTRIBUTION TO KNOWLEDGE OF THE EVOLUTION OF BIOLUMINESCENCE IN BEETLES 105
5.8 SUMMARY 107

6. **REFERENCES......108**

APPENDICES......109

Appendix I: PARTIAL AND FULL BEETLE SGL GENE SEQUENCES: PREDICTED GENE ARCHITECTURE & PUTATIVE OPEN READING FRAMES 121
Appendix II : PARTIAL AND FULL LENGTH BEETLE LUCIFERASE & LUCIFERASE-LIKE GENE SEQUENCES: PREDICTED GENE ARCHITECTURE & PUTATIVE OPEN READING FRAMES 138
Appendix III: PUBLICATIONS ARISING FROM RESEARCH 172

List of Figures

		Page
Figure 1.1	Phylogenetic tree of bioluminescence and related luciferin structures	3
Figure 1.2	Evolutionary relationships within and around the Cantharoidea	5
Figure 1.3	Oxyluciferin to luciferin regeneration pathway	10
Figure 1.4	Representations of the firefly luciferase protein	15
Figure 1.5	Bioluminescence and communication in fireflies	21
Figure 2.1	Position of degenerate SGL primers in relation to the LRE gene of *Photinus pyralis*	37
Figure 2.2	Microsynteny analysis between *Drosophila* SGL genes	41
Figure 2.3	Chromosomal location and gene architecture of *Tribolium castaneum* SGL paralogues	44
Figure 2.4	Predicted first intron position in TcSMP03 and TcSMP05	45
Figure 2.5	Intron/exon boundaries in insect SGL orthologues and paralogues	47-48
Figure 2.6	Intron/exon boundaries for lampyrid LRE genes and pseudogenes	49
Figure 2.7	*Pachnoda m. peregrina* SGL gene PCR & RT-PCR	52
Figure 2.8	Schematic of PCR and genome walking for the three loci PmSMP 1-3	53
Figure 2.9	Sequence logo representation of the SGL protein family	55
Figure 2.10	Crystal structure of *Agrobacterium tumefaciens* SGL protein	56
Figure 2.11	Phylogenetic analysis of the SGL domain-containing protein family in insects	58
Figure 2.12	Comparative analysis of coleopteran SGL protein family groups	60
Figure 2.13	Critical amino acid residues responsible for the functional divergence between beetle SGL groups	63
Figure 2.14	An illustration of functional divergence in SGL paralogues in beetles	64
Figure 3.1	Luciferase-like sequences in the genome of the red flour beetle *Tribolium castaneum*	77
Figure 3.2	Phylogenetic analysis of the AMP-binding domain-containing protein family in beetles	80
Figure 3.3	Output of the SELECTON analysis: representative amino acid sequences generated for each selected clade of Figure 3.2..........................	81
Figure 3.4	Luciferase amino acid sequence alignment of representative lamyprids from the four clades L1A-L1D	85
Figure 3.5	Schematic representation of predicted evolution of luciferase and luciferase-like genes from an ancestral acyl-CoA synthetase	87
Figure 4.1	Graphic representation of the PCR amplification and gene architecture of luc and flanking region genomic DNA from *L. noctiluca*.................	90
Figure 4.2	Alignment of the sequence upstream of the luciferase gene	91
Figure 4.3	Amino acid sequence alignment of transposase conceptual translations ...	93

List of Tables

Table 1.1	Cloned beetle luciferases and their characteristics	14
Table 2.1	List of coleopteran species studied	33
Table 2.2	Primers used in PCR and sequencing of beetle SGL genes	35
Table 2.3	SGL sequences retrieved from public databases	42
Table 2.4	Summary table of SGL partial gene characterisation from nineteen beetle taxa	50
Table 2.5	Quantitative survey of SGL protein features	59
Table 2.6	Maximum likelihood estimates of the coefficient of functional divergence (θ) from pairwise comparisons between SGL groups	61
Table 3.1	List of beetle specimens studied, collection details and gene nomenclature	74
Table 3.2	Coleopteran luciferase (LUC) and luciferase–like (LL) sequences retrieved from public databases	79

List of Abbreviations

AMP	Adenosine 5′-monophosphate
ATP	Adenosine 5′-triphosphate
DFP	Diisopropyl phosphorofluoridate
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
EST	Expressed sequence tag
kcal/mol	Kilocalories per mole
LRE	Luciferin regenerating enzyme
GLO	L-gulonolactone oxidase
GNL	Gluconolactonase
ORF	Open reading frame
SGL	SMP30 GNL LRE
SMP30	Senescence Marker Protein – 30
ML	Maximum Likelihood
MP	Maximum Parsimony
MYA	Million years ago
MYO	Million years old
pI	Isoelectric point
RNA	Ribonucleic acid
SMP30	Senescence Marker Protein-30

Publications arising from Research

1. Day, JC & Bailey, MJ (2003) Structure and evolution of the luciferin-regenerating enzyme (LRE) gene from the firefly *Photinus pyralis*. *Insect Molecular Biology* **12**, 365-372 [Chapter 2].

2. Day, JC, Tici, LC & Bailey, MJ (2004) Evolution of beetle bioluminescence: The origin of beetle luciferin. *Luminescence* **19**, 8-20. [Chapter 1].

3. Day JC (2005) Characterisation of the luciferase gene and the 5' upstream region in the European glow-worm *Lampyris noctiluca* (Coleoptera: Lampyridae). *European Journal of Entomology* **102** (4): 787–791 [Chapter 4].

4. Day, JC, Goodall, TI & Bailey, MJ (2009) The evolution of the adenylate-forming protein family in beetles: multiple luciferase gene paralogues in fireflies and glow-worms. *Molecular Biology and Evolution* **50**, 93-101. [Chapter 3].

1. LITERATURE REVIEW

1.1 Introduction

Charles Darwin's discovery of natural selection formed the foundation stone in the history of evolutionary theory. Darwin defined evolution as "descent with modification" and this elegant and simple idea has generated a vast science accelerating in growth with the advent of more powerful technology. Adaptation, a further key concept of evolution, refers to those properties that enable organisms to survive and reproduce in nature. Darwin noted on his travels a number of bioluminescent beetles and speculated on the role of luminescence as a form of sexual selection but questioned the role of luminescence in larva. Since then the evolution of beetle bioluminescence has been pondered by a number of scientists and three fundamental questions arise from such discourses: 1) when did bioluminescence arise in beetles? 2) What were the first bioluminescent beetles? and 3) How has beetle bioluminescence subsequently evolved?

1.1.1 Why study the evolution of bioluminescent beetles?

Because bioluminescent beetles are such unusual and distinctive insects, it might seem unlikely that they could have evolved multiple times from independent ancestors. Nonetheless on the basis of cladistical analysis a number of scenarios have been produced that present multiple origins of bioluminescence within the Coleoptera. Based upon physiological characters such multiple evolutionary events have been proposed in other animals such as bats, whales, pandas and crabs. Molecular data has been used to support or refute monophyletic and paraphyletic relationships between these groups. If beetle bioluminescence is of a genuine paraphyletic nature then this provides an important basis for understanding convergent evolution in insects.

Additionally the emission of light by these beetles is the result of enzymes coded for by a single gene. Such a linear relationship between phenotype and genotype provides an excellent straightforward model for understanding the evolution of this process thereby potentially elucidating the evolution of similar molecular mechanisms.

1.1.2 How valuable are bioluminescent beetles for scientific studies?

Today bioluminescent beetles are found across the globe on all the main land masses of the world with the exception of Antarctica. Over 3000 species have been described and, as with most insect groups, a much greater number potentially await discovery. The vast majority of these species are found in the families Lampyridae (fireflies and glow-worms), Phengodidae (railroad worms) and Elateridae (click beetles) and it is predicted that the bioluminescence reaction chemistry and the genetics is the same or similar for all beetles. Despite these similarities the actual phenotypic light emission can vary quite dramatically. In fireflies for example the emissions range from long steady glows through to a large range of interactive flash patterns. These are known or hypothesised to function primarily in bringing male and female together in courtship (Barber, 1951; Lloyd, 1966; Lloyd, 1972; McDermott, 1917; McDermott, Buck, 1959; Ohba, 1983). Fireflies thus present an almost ideal group of insects for neurobiological or behavioural studies of animal communication (Lloyd, 1983b). Firefly flash dialogue can be recorded easily at a distance as well as enabling the observer to participate in the exchange using flash communication equipment. Furthermore, each species of firefly produces a different courtship code (Lloyd, 1966) which provides the basis for examining the relationship of species evolution and the evolution of communication per se. Additionally, for the fireflies (Lampyridae) in particular, the study of bioluminescent beetles has attracted a wide range of researchers approaching the study of these organisms from a diverse range of scientific disciplines. These include such areas as ecology, behavioural biology, biochemistry, physical chemistry, genetics, physiology and taxonomy which ultimately provide a large corpus of literature as a solid foundation for future research.

1.2 The taxonomic evolution of beetles

Our understanding of bioluminescent mechanisms is based upon a few well established systems. Along with fireflies the process of luminescence has been extensively investigated in bacteria, crustacean, dinoflagellates and jellyfish as well as other coelenterates. A comparison of the chemistries of these mechanisms reveals no true homology and thus it has been proposed that these mechanisms arose independently indicating that beetle bioluminescence probably originated sometime after the terrestrial and marine arthropods diverged, around 400 million years ago (MYA) (Wood, 1995) (Figure 1.1).

FIGURE 1.1 Pylogenetic tree of bioluminescence and related luciferin structures. Phylogeny from Haddock, S (2006) "Luminous marine organisms" in *Photoproteins in Bioanalysis* Wiley.

Beetles, however, are not the only terrestrial arthropods to have developed bioluminescent capabilities. Luminescence is also evident in centipedes, millipedes, as wells as the fungoidal gnats (Diptera) (Meyer-Rochow, 2007). Well known in New Zealand for illuminating caves with an intense blue light they represent the closest known bioluminescent group of arthropods to the beetles. Although not fully characterised the mechanism for dipteran bioluminescence lacks homology and cross reactivity with beetle luciferin-luciferase systems (Lee, 1976; Shimomura et al., 1966; Viviani et al., 2002). As a result it has been proposed that bioluminescence arose less than 350 MYA when these two insect orders diverged (Wood, 1995).

Beetles diverged from their common ancestors probably during the Early Permian. The oldest primitive protocoleopteran record is from the Early Permian around 280 MYA but what are regarded as true Coleoptera are not considered to have arisen until the Triassic, around 230 MYA. The vast diversification of the Coleoptera was evident by the Late Jurassic, 155-160 MYA (Hunt et al., 2007). It is possible that such diversification led to the foundations of bioluminescence that may or may not have evolved in subsequent lineages.

Bioluminescent is evident in only a small proportion of beetles, around 3000 of the 370,000 known beetle species are luminescent. All bioluminescent beetles, with the exception of one dubious species, are contained within the Elateriformia. Subdivided into some forty extant families the series Elateriformia is one of the major groups of beetles within the suborder Polyphaga containing around 40,000 species. Bioluminescence in the series Elateriformia is currently classified as occurring in Elateridae and four other families: Omalisidae, Rhagophthalmidae, Phengodidae and Lampyridae which originally belonged to the superfamily Cantharoidea. Originally a part of the Malacoderma the superfamily Cantharoidea was revised by Crowson in 1972 and latter combined into the Elateroidea when Lawrence (1988) redefined Elateriformia (Crowson, 1972; Lawrence, 1988). Consequently all bioluminescent beetles, with the exception of the single staphilinid species, are now currently contained within a single superfamily the Elateroidea.

The interest in bioluminescence has inspired several studies to investigate the phylogenetic relationships of the Elateriformia (Beutel, 1995; Bocakova *et al.*, 2007; Branham, Wenzel, 2001; Branham, Wenzel, 2003; Crowson, 1972) (Figure 1.2). The relationship between the eighteen families of the Elateroidea has yet to be fully resolved as different authors present different associations in their communications. The only consensus that can be drawn at present is that bioluminescent beetle families appear not to have descended from a common ancestor and upon first inspection appear to have arisen independently in some cases (Figure 1.2).

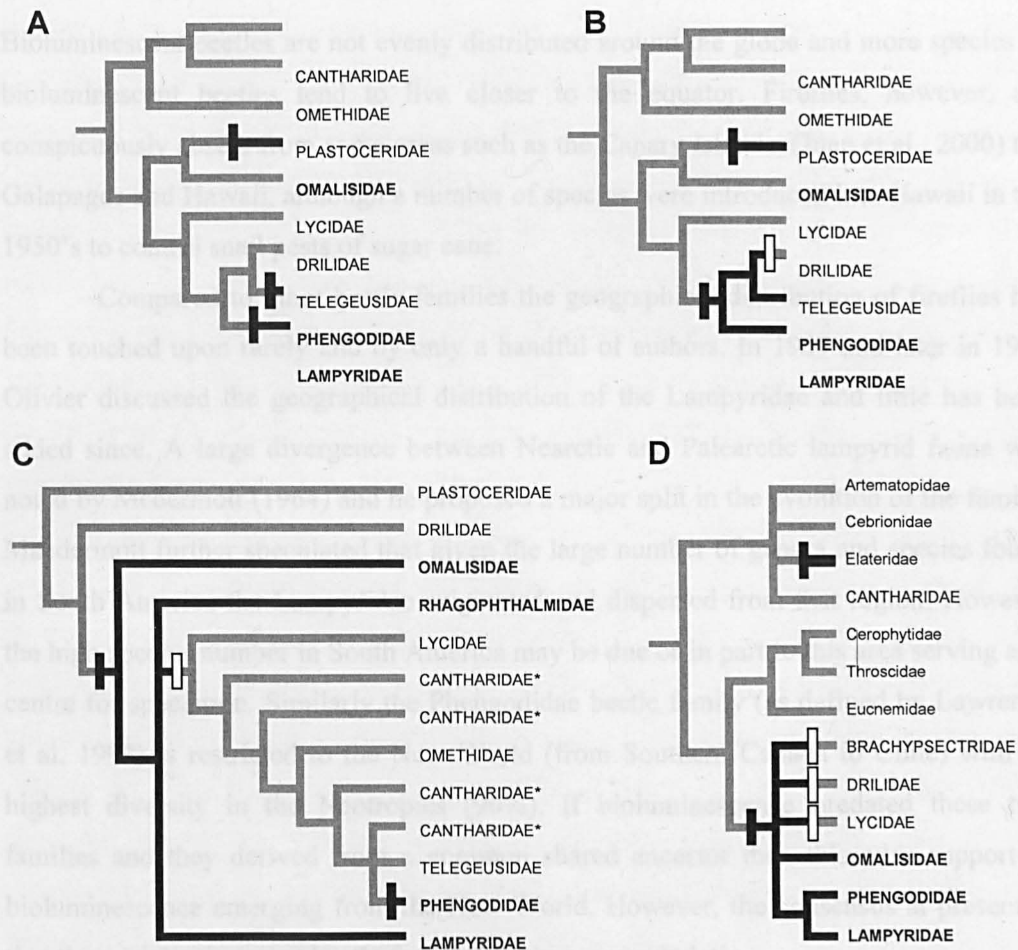


FIGURE 1.2 Evolutionary relationships within and around the Cantharoidea. Plotting luminescence onto different trees supports interpretations ranging from three origins A to one origin and three loses D. Non-bioluminescent lineages are shown in grey and bioluminescent lineages in black. Dark boxes denote an emergence and white boxes a loss of bioluminescence. A striped box and branch indicates the emergence of a non-cantharoid bioluminescent lineage. Cantharoidea families are shown in upper case and bioluminescent Cantharoidea families in bold. A and B are derived from a dendrogram of nine cantharoid families by Crowson (1972). This scheme predicts two character optimizations of three steps each. One optimization presents three origins of bioluminescence A, while the second presents two origins and one loss of bioluminescence B. C is a dendrogram derived from a condensed strict consensus tree based upon seventy-four morphological states in eighty-five exemplary taxa by Bramham and Wenzel (2001). This predicts one character optimization of three steps; two origins of luminescence and one loss. D is based upon the majority rule consensus tree of Beutel 1995 who proposed a phylogenetic analysis of Elateriformia based on twenty-seven larval characters which predicts a single topology of one origin and three losses. All of the cantharoid taxa were placed in a poorly resolved clade except for Cantharidae, which was placed close to Elateridae. A fourth study by Potatskaja (1983) proposed a dendrogram, based on larval characters, for the relationships between the cantharoid families Brachyspectridae, Cantharidae, Phengodidae, Drilidae, Omalisidae, Lycidae and Lampyridae. Potatskaja proposed two lineages, one termed “cantharid” (composed of Phengodidae, Drilidae, Omalisidae, Brachyspectridae, and Cantharidae) and the other “lycid” (composed of Lampyridae and Lycidae). According to Bramham and Wenzel (2001) this topology predicts two optimizations of three steps each: three separate origins, or two origins and one loss.

1.2.1 Phylogeography

Bioluminescent beetles are not evenly distributed around the globe and more species of bioluminescent beetles tend to live closer to the equator. Fireflies, however, are conspicuously absent from some areas such as the Canary Islands, (Juan et al., 2000) the Galapagos and Hawaii, although a number of species were introduced into Hawaii in the 1950's to control snail pests of sugar cane.

Compared to other beetle families the geographical distribution of fireflies has been touched upon rarely and by only a handful of authors. In 1907 and later in 1911 Olivier discussed the geographical distribution of the Lampyridae and little has been added since. A large divergence between Nearctic and Palearctic lampyrid fauna was noted by Mcdermott (1964) and he proposed a major split in the evolution of the family. Macdermott further speculated that given the large number of genera and species found in South America the Lampyridae originated and dispersed from that region. However the high species number in South America may be due or in part to this area serving as a centre for speciation. Similarly the Phengodidae beetle family (as defined by Lawrence et al. 1999) is restricted to the New World (from Southern Canada to Chile) with its highest diversity in the Neotropics (90%). If bioluminescence predated these two families and they derived from a common shared ancestor then this adds support to bioluminescence emerging from the New World. However, the consensus at present is that these bioluminescent beetle families are not monophyletic.

1.3 The first bioluminescence

Bioluminescent organisms are found in almost all phyla and all the systems involved in making light share a basic biochemical pathway that are luciferase-catalyzed reactions of the molecular oxygen with a substrate generically known as a luciferin. The term luciferase is also generic as luciferases from different groups of organisms are unrelated genetically as well as structurally. All bioluminescent reactions include a luciferase-bound peroxy-luciferin intermediate that breaks down to provide energy for excitation. Light is generated in a bioluminescent reaction very rapidly as a part of a multi-step chemical reaction. The penultimate step is the production of a molecule in an electrically excited state P* that has a very short lifetime, 1-10 nanoseconds.

Bioluminescent organisms range in diversity from bacteria, fungi, and algae through to earthworms, squid and fish and occur in 700 different genera (Harvey, 1952; Hastings, 1983). Although some mechanisms, such as fungal bioluminescence, remain elusive (Shimomura et al., 1993) a number of bioluminescent reactions have been well

characterised including those in bacteria, ctenophores, colenterates, dinoflagellates and fireflies (for a review of chemical mechanisms see Shimomura, 2006). However, when compared, the diversity of the chemistries and enzymes involved suggests these different organisms have evolved bioluminescent systems independently (Figure 1.1). The one common thread uniting these different systems at the molecular level is that these are all luciferase-catalyzed reactions of molecular oxygen with various luciferins. This requirement for oxygen has resulted in the hypothesis that a primitive antioxidant function is the evolutionary foundation for the emergence of bioluminescence (Seliger, 1975; Timmins et al., 2001). This is supported by evidence that in many marine animals the luciferin, a coelenterazine, is a proven antioxidant protecting cells against oxygen toxicity (de Wergifosse et al., 2004; Rees et al., 1998).

This theory has been extended to beetles; it has been proposed that bioluminescence could have first originated as an oxygen detoxification mechanism through the consumption of oxygen by the luciferase-catalysed reaction (Barros, Bechara, 1998). Evidence to support this is three-fold. Firstly the strong increase in luciferin biosynthesis observed in the prothorax of bioluminescent elaterid larvae upon exposure to hyperoxia may be an adaptive response to hyperoxidative conditions (Barros, Bechara, 1998). Secondly, photocytes are rich in tracheal terminals and are densely packed with mitochondria at the interface with the tracheal system (Ghiradella, 1998; Ghiradella, Schmidt, 2004) and the formation of reactive species derived from oxygen is to be expected (Barros, Bechara, 2001). Thirdly, it has been shown that firefly luciferin is endowed with antioxidant properties against oxidative and nitrosative stress thereby helping photocytes to cope with the hyperoxidant conditions resulting from light emission (Dubuisson et al., 2004).

1.4 Firefly luciferin

Luciferin, 2-(6-hydroxybenzothiazol-2-yl)-2-thiazoline-4-carboxylic acid, was first purified and crystalised from the North American firefly *Photinus pyralis* in 1957 (Bitler, McElroy, 1957). This process required 15,000 firefly lanterns to produce 9 mg of crystalline luciferin. Proof of its structure came from the successful chemical synthesis of enzymatically active luciferin in 1961 (White et al., 1961) and the structure confirmed by X-ray crystallography (Blank et al., 1971).

It is commonly regarded that all lampyrids emit light as larvae and many seem to produce light throughout all their life stages. Interestingly the non-luminescent adults of *Ellychnia corrusca* can be induced to elicit a flash of light indicating the expression of the enzyme and the synthesis of the substrate in the adult beetle (pers. comm. J.

Copeland). Luciferin appears to be conserved in structure between bioluminescent beetle species and even families irrespective of metamorphic stage or lantern location (Colepicolo *et al.*, 1988; Hadj-Mohammadi, Chaichi, 1996; Seliger, McELroy, 1965).

Biologically active natural products are usually homochiral e.g. proteins are predominantly composed of mainly L-amino acids. In general, most bioluminescent reactions employ a single chiral luciferin and are biosynthesised from peptides or L-amino acids as is postulated for the luminescent ostracod *Cypridina* (Kato et al., 2004). The firefly bioluminescent reaction is no exception. Only firefly D-luciferin contributes to bioluminescence and several researchers noted that no light is produced from L-luciferin (Branchini *et al.*, 1989; McElroy, Seliger, 1962). However, Lembert reported that L-luciferin produced weak light but extremely slowly (Lembert, 1996). As a result Lembert proposed that L-luciferin was racemized to give D-luciferin (Lembert, 1996). It has been recently presented that luciferase could be responsible for the stereoisomeric inversion of L-luciferin to D-luciferin which would explain the weak bioluminescence observed by Lembert (Nakamura et al., 2006). To further support this Niwa *et al.* 2006 measured levels of both D- and L-luciferins in the firefly *Luciola lateralis*; both forms were detectable in all firefly life stages, including the egg (Niwa et al., 2006). The enantiomeric excess of D-luciferin was highest at the adult stage, while it was lower during larval and pupal stages suggesting L-luciferin is converted to D-luciferin as the beetle matures.

Luminescence was originally thought to be produced by symbiotic bioluminescent bacteria in the firefly (Kuhnt, 1907; Pierantoni, 1914). This notion was dispelled when Harvey in 1929 demonstrated that the adult firefly develop new lanterns after the larval lanterns have been excised (Harvey, Hall, 1929). Seliger predicted that the adult lampyrids emerge with sufficient luciferin for luminescence during their lifetime (Seliger, 1973). Strause *et al.* studied the levels of luciferase and luciferin in larvae, pupae and adults of *Photuris pennsylvanica* (Strause et al., 1979). They calculated there would be sufficient luciferin (15 μmole) and luciferase in the newly emerged adult to provide 10,000 flashes, adequate for the lifespan of an average firefly. They concluded there would be no need for the adult to synthesise luciferin thereby supporting Seliger's hypothesis.

It has been suggested that, based upon the structure of beetle luciferin and its chemical synthesis, that the origin of the thiazoline ring is likely to be a cysteine (McCapra, Perring, 1985). Indeed studies using radiolabelled cysteine apparently confirm this. However, the origin of the benzothiazole portion is less clear but it has been suggested that cysteine is also a building block of this part of the luciferin molecule. Early attempts to chemically synthesise luciferin were based upon the

possibility that *in vivo* quinones, e.g. p-benzoquinone, known to be present in coleopterans, often as part of defence secretions, may react with cysteine to produce the 6-hydroxybenzothiazole moiety of beetle luciferin. The addition of a subsequent cysteine gives beetle luciferin. This proven chemical synthesis was seen as a model for the biosynthesis *in vivo* (McCapra, Razavi, 1975).

However the synthesis of luciferin is brought about it is apparent that this synthetic pathway is found in the three main bioluminescent beetle families and given the complexity of such processes would suggest a common ancestor with luciferin synthetic capabilities. The absence of D-luciferin in cantharids does not automatically rule out a common luciferin rich ancestor. It would seem to be more parsimonious that a biosynthetic pathway arose once and has been subsequently lost than to have arisen numerous times.

1.5 The Luciferin Regenerating Enzyme (LRE)

Few experimental studies have tried to examine the luciferin synthetic pathway in bioluminescent beetles. But in 1974 Okada and co-workers injected ^{14}C oxyluciferin and ^{14}C-2-cyano-6-hydroxybenzothiazole (2C6HB) into living fireflies and detected ^{14}C luciferin after a number of hours (Okada et al., 1974). They concluded that the luminescent product, oxyluciferin, is recycled back to the substrate luciferin for subsequent light emission. Okada *et al.* also identified that the addition of cysteine improved the yield of luciferin. Around the same time in *Pyrophorus pellucens* it was found that radiolabelled cysteine was incorporated into newly synthesised luciferin (McCapra, Razavi, 1975) These results have been explained in the following two-step reaction: (1) transformation of oxyluciferin to 2-cyano-6-hydroxybenzothiazole and (2) condensation of 2C6HB with D-cysteine to yield luciferin (Figure 1.3)

Derivatives of 2C6HB are now used for the large-scale chemical synthesis of beetle luciferin (Bowie, 1978; Branchini, 2000). However, it should be noted that it has been established that the reaction with cysteine and 2C6HB occurs non-enzymatically (Gomi, Kajiyama, 2001b; Okada *et al.*, 1974).

Nearly thirty years later Gomi and Kajiyama identified an enzyme that appeared to catalyze this reaction *in vitro* (Gomi, Kajiyama, 2001b). A protein fraction of three firefly extracts (*Photinus pyralis*, *Luciola lateralis* and *L. cruciata*) were originally found to exhibit an activity that enhanced bioluminescence *in vitro* (Gomi, Kajiyama, 2001b). This enzyme, the luciferin regenerating enzyme (LRE) was further characterised by obtaining the cDNA for all three fireflies which, when expressed and assayed, exhibited the same properties of enzymatically regenerating oxyluciferin back

into luciferin in the presence of D-cysteine (Gomi *et al.*, 2002; Gomi, Kajiyama, 2001) (Figure 1.3).

The three LRE cDNA sequences coded for between 307–309 amino acids with predicted molecular masses of 33.6–34.3 kDa. It is not yet known whether genes coding for LREs are found within the genome of click beetles and railroad worms. Sequence comparisons revealed that the three LREs shared significant identity with a group of proteins known as senescence marker protein-30 (SMP30) (Gomi *et al.*, 2002).

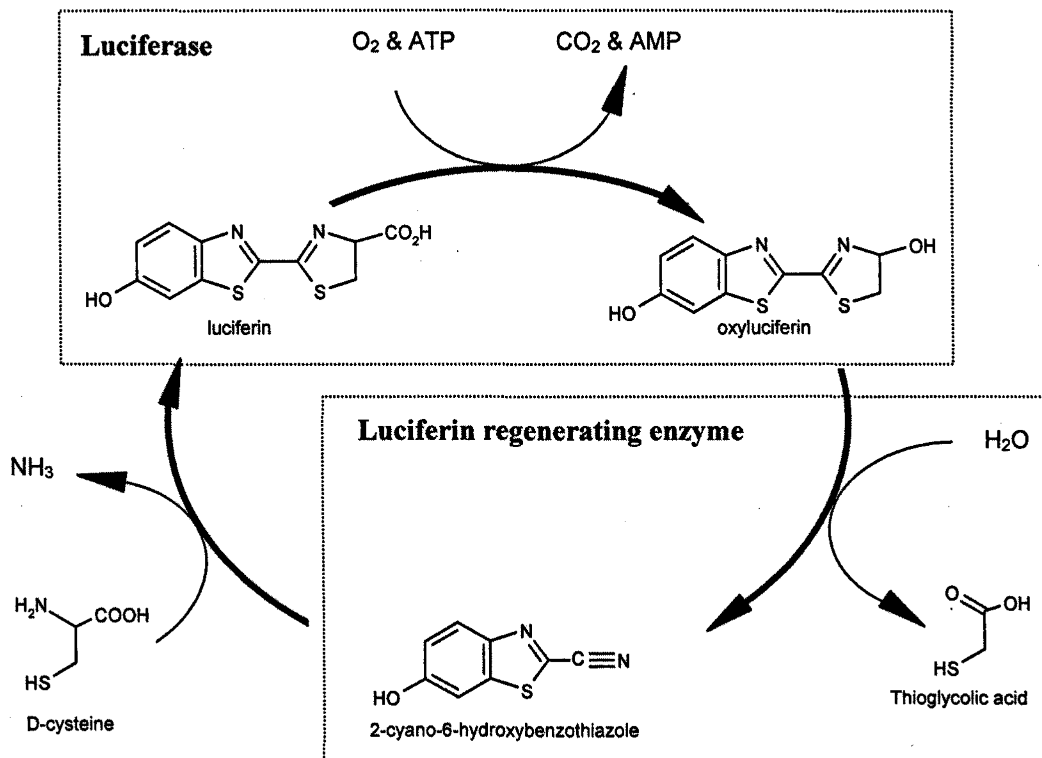


FIGURE 1.3 Theoretical pathway for the regeneration of luciferin from oxyluciferin.

1.6 Senescence Marker Protein 30 (SMP30)

To understand age-associated modifications at the genetic level Fujita *et al.* in 1992 surveyed differential levels of proteins produced in different aged rats. A novel hepatic protein was identified and found to be produced in reduced amounts (60-70% less) in older rats (Fujita et al., 1992). Due to the relationship with aging and the molecular mass of 30 kDa the protein exhibited, the protein was named senescence marker protein 30 (SMP30) (Fujita et al., 1992).

Molecular analysis of the rat SMP30 cDNA revealed a transcript coding for an enzyme composed of 299 amino acids with an estimated molecular weight of 33,387

(Fujita et al., 1992). Independently a Ca^{2+}-binding protein called regucalcin had been characterised and subsequently found to be the identical to SMP30 (Shimokawa, Yamaguchi, 1992; Shimokawa, Yamaguchi, 1993; Yamaguchi, Sugai, 1981; Yamaguchi, Yamamoto, 1978). Fujita *et al.* went on to characterise the cDNA encoding SMP30 from human RNA and the cDNA, gene and putative promoter sites in the mouse (Fujita et al., 1995 ; Fujita et al., 1996b). In the mouse the gene is composed of seven exons and spans approximately 17.5 kb (Fujita et al., 1996a). A number of promoter regions have been predicted including a cluster of motifs (Sp1) that, in age rats, decrease in binding efficiency (Fujita et al., 1996b). Although first identified in the liver, SMP30 transcripts have been found in other tissues such as the kidney, lung, testes and cerebrum (Mori et al., 2004). In mice the SMP30 has been shown to protect the liver, and potentially other organs, from apoptosis (Ishigami et al., 2002). In addition SMP30 potentially facilitates detoxification from harmful compounds classed as diisopropyl phosphorofluoridates (DFP) such as the chemical warfare nerve agents sarine, soman and tabun (Ishigami et al., 2002; Kondo et al., 2004). This evidence along with other studies proposing antioxidant properties, thereby protecting cells from oxidative stress, has led SMP30 to be regarded as an anti-aging molecule (Feng et al., 2004). Although the relationship between LRE and luciferase is unknown it is interesting to note that if LRE shares some metabolic similarity to SMP30 it could function as an antioxidant enzyme, a role originally given to luciferase. Additionally the presence of peroxisome targeting sequences (PTS) found on the LRE of *P. pyralis* and *L. cruciata* indicate that the like firefly luciferases these proteins could be localised in the peroxisomes.

1.6.1 SMP30 and ascorbic acid synthesis

Recently Kondo *et al.* (2006) showed that SMP30 is involved in the synthetic pathway of L-ascorbic acid or vitamin C (Kondo *et al.*, 2006). Ascorbic acid is a reducing agent and antioxidant and in mammals is synthesised in the liver whereas reptiles and fish produce it in the kidney (Chatterjee, 1973; Moreau, Dabrowski, 1998a). Gluconolactonase (GNL) is known to catalyse the penultimate reaction (Burns, 1960; Nishikimi *et al.*, 2003; Nishikimi, Yagi, 1996) and Kondo *et al.* (2006) showed that rat GNL and SMP30 are one and the same with regard to their catalytic activity (Kondo *et al.*, 2006). SMP30 exhibited catalytic acitvity with a range of aldonolactone substrates including L-gluconic acid, but the most activity was exhibited with d-glucono-δ-lactone. Furthermore, SMP30 knockout mice were prone to scurvy when subjected to an vitamin C deficient diet. Ascorbic acid is synthesised by the majority of vertebrates and the presence of a biosynthesis pathway in the sea lamprey suggests this appeared early in the

evolutionary history of fishes prior to the emergence of terrestrial vertebrates (Moreau, Dabrowski, 1998a; Moreau, Dabrowski, 1998b). This biosynthetic capability has been lost in a number of diverse organisms including teleost fishes (Dabrowski, 1990), passeriform birds (Chaudhuri, Chatterjee, 1969), bats (Birney et al., 1976), guinea pigs (Burns et al., 1956) and primates including humans (Stone, 1965). Enzyme studies in the late 1950's revealed that the inability to synthesise ascorbic acid was caused by the lack of L-gulonolactone oxidase (GLO) activity (Burns, 1957). Disrupted GLO genes have been characterised in humans (Nishikimi et al., 1994), primates (Ohta, Nishikimi, 1999), guinea pigs (Nishikimi et al., 1992) and some breeds of pig (Hasan et al., 1992).

1.6.2 The role of SMP30 homologues in insects

SMP30 homologues have been identified from four different insect families and a number of putative roles ascribed. In *Drosophila* a SMP30 orthologue, *Dca*, exhibited increased expression levels as a result of cold acclimatisation (Goto, 2000). Dipteran SMP30 orthologues have also been identified from the flesh fly *Sarcophaga peregrina* (Nakajima, Natori, 2000) which was found to be restricted almost exclusively to the anterior fat body (Nakajima, Natori, 2000) and was hence referred to as the anterior fat body protein (AFP). In late larvae of the blowfly, *Calliphora vicina*, this AFP was found to bind to the hexamerin receptor arylphorin-binding protein (ABP) and predicted to play a role in the regulation of hexamerin uptake by fat body cells along the anterior–posterior axis (Hansen et al., 2002). It has yet been established whether insect SMP30 homologues have a GNL function.

1.6.3 SGL Protein Family

Based upon sequence conservation and the recent evidence for conserved function the three classes of enzyme GNL, SMP30 and LRE have been designated as the SGL (SMP-30/Gluconolaconase/LRE-like) protein family (PFAM08450). However, despite the sequence identity, SMP30 in mice has been shown to lack any luciferin regenerating activity; conversely both LRE and SMP30 have been shown to hydrolyze diisopropyl phosphorofluoridate (DFP) (Kondo et al., 2004). Firefly LRE therefore stands apart as a enzyme with a unique catalytic activity, the regeneration of oxyluciferin into luciferin, an activity not exhibited by mammalian SMP30 (Kondo et al., 2004). LRE therefore may have arisen from an ancestral GNL in beetles which may have originally played a part in the synthesis of ascorbic acid. The range of activities that LRE plays *in vivo* has

yet to be carried out, in fact our knowledge of this system is slight when compared to the extensive information available on beetle luciferase.

1.7 Firefly Luciferase

The luciferase from the firefly *Photinus pyralis* was first purified, crystallised and partially characterised in 1956 (Green, McElroy, 1956). Purification based upon high-performance liquid chromatography (HPLC) was later reported for luciferase from *P. pyralis* and *Photinus macdermotti* (Branchini, Rollins, 1989). In 1984 Wood et al. cloned *P. pyralis* luciferase by *in vitro* translation and determined the molecular weight to be 62,000 (Wood et al., 1984). Wienhausen and DeLuca identified luciferases from other bioluminescent beetle species, including the click beetle *Pyrophorus plagiophthalamus*. These migrated at a similar position, although not identical, and exhibited extensive cross-reactivity with antibodies raised against *P. pyralis* luciferase (Wienhausen, DeLuca, 1985). Thus it was anticipated that luciferases from other bioluminescent beetles would have similar molecular weights.

The firefly *P. pyralis* was again used to provide the material for the first cloning of luciferase into a bacterial system. De Wet and co-workers in 1985 expressed the cDNA of *P. pyralis* luciferase in *Escherichia coli* providing the basis for mass production of luciferase *in vitro* and the further characterization of the enzyme through mutagenesis studies in the coming years. To date the luciferase cDNA has been characterised from over twenty bioluminescent beetle taxa and extensive information has been collated about these enzymes (Table 1.1). The first genomic luciferase sequence was characterised from *P. pyralis* and found to be composed of seven exons divided by six introns ranging in size from 48–58 bp (de Wet et al., 1985). To date no genomic luciferase sequences have been made publicly available for members of the Elateridae or Phengodidae. In fireflies the luciferase enzyme is composed of one polypeptide chain ranging in size from 545–552 highly conserved residues.. Over half are non-polar or ambivalent amino acids and the number of charged residues is virtually the same for all lampyrid species.

TABLE 1.1 Cloned beetle luciferases and their characteristics.

	Residues	Sequence identity (%)[†]	pI	λ_{max} (nm) *	Reference
LAMPYRIDAE (fireflies & glow-worms)					
Lampyrinae					
Cratomorphus distinctus	547	83	5.85	550	(Viviani *et al.*, 2004)
Diaphanes pectinealis	547	83	6.09	-	(Li *et al.*, 2006)
Lampyris noctiluca	547	84	6.08	550	(Sala-Newby *et al.*, 1996)
Nyctophila cf. caucasica [a]	547	84	6.19	-	(Said Alipour *et al.*, 2004)
Photinus pyralis	550	100	6.43	562	(de Wet *et al.*, 1985)
Pyrocoelia miyako	548	82	6.11	550	(Ohmiya *et al.*, 1995)
Pyrocoelia pygidialis	548	83	6.03	-	(Dong *et al.*, 2008)
Luciolinae					
Hotaria parvula	548	67	6.27	568	(Ohmiya *et al.*, 1995)
Hotaria unmunsana	548	67	6.10	-	(Choi *et al.*, 2002)
Lampyroidea maculata	548	64	5.99	-	(Emamzadeh *et al.*, 2006)
Luciola cruciata	548	67	7.17	562	(Masuda *et al.*, 1989)
Luciola italica	548	65	5.99	566	(Branchini *et al.*, 2006)
Luciola lateralis	548	67	6.52	552	(Tatsumi *et al.*, 1992)
Luciola mingrelica	548	67	6.24	570	(Devine *et al.*, 1993)
Luciola terminalis	548	65	6.47	-	*unpublished*
Photurinae					
Photuris pennsylvanica					
Ppe1	552	69	7.23	560	(Ye *et al.*, 1997)
Ppe2	545	59	8.29	538	(Ye *et al.*, 1997)
Phengodidae (railroad worms)					
Phrixothrix vivianii	545	55	6.39	548	(Viviani *et al.*, 1999a)
Phrixothrix hirtus	546	48	7.00	623	(Viviani *et al.*, 1999a)
Ragophthalmus ohbai	543	53	7.93	555	(Sumiya *et al.*, 1998)
Elateridae (click beetles)					
Pyrophorus mellifluus					
Green (dorsal)	543	47	6.92	549	(Stolz *et al.*, 2003)
Green (ventral)	543	47	7.63	554	(Stolz *et al.*, 2003)
Pyrophorus plagiophthalamus					
Green	543	47	6.71	546	(Wood *et al.*, 1989)
Yellow Green	543	47	6.71	560	(Wood *et al.*, 1989)
Yellow	543	47	6.39	578	(Wood *et al.*, 1989)
Orange	543	47	6.71	593	(Wood *et al.*, 1989)
Pyrearinus termitilluminans	543	46		538	(Viviani *et al.*, 1999b)

[†] Amino acid sequence identity to *Photinus pyralis* luciferase.
[a] Reported as *Lampyris turkistanicus* see (Day *et al.*, 2006).
* For comparative purposes the *in vitro* emission is reported. *In vivo* measurement of bioluminescence can be affected by a number of factors and does not necessarily depict the true light emission of the enzyme.

In 1996 the crystal structure of the *P. pyralis* luciferase was first determined at a resolution of 2.0 Å (Conti et al., 1996). The protein was found to be folded into two compact domains connected by a short flexible hinge (Figure 1.4). The large N-terminal domain being composed of a β-barrel and two β-sheets flanked by α-helices to form an αβαβα five-layered structure. The C-terminal portion of the molecule formed a distinct domain separated from the N-terminal domain by a wide cleft. Conti *et al.* proposed that the cleft was far too big to accommodate the substrate and the domains will close in the course of the reaction to sandwich the substrates. In 2006 the crystal structure of the

wild-type luciferase from *Luciola cruciata* complexed with a high-energy intermediate analogue of luciferin, 5'-O-[N-(dehydroluciferyl)-sulfamoyl] adenosine (DLSA) was determined at 1.3 Å resolution (Figure 1.4) (Nakatsu et al., 2006). It is apparent from the comparative structures that indeed the domains are closer when bound to a substrate (Figure 1.4).

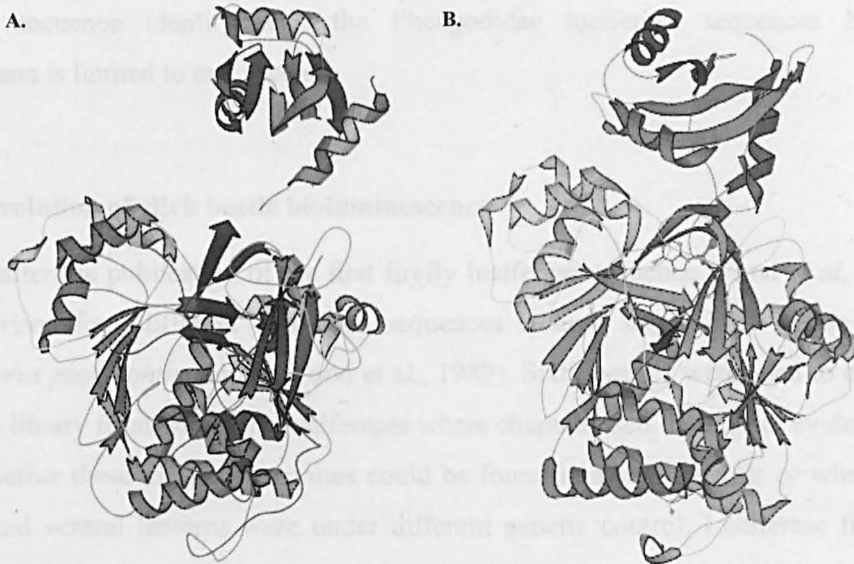


FIGURE 1.4 Representations of the firefly luciferase protein. (A) *Photinus pyralis* luciferase structure. (B) *Luciola cruciata* luciferase complexed with DLSA.

1.7.1 Railroad worm and click beetle luciferases

The bioluminescent mechanism in the Phengodidae and luminescent elaterids is considered to be the same as that found in fireflies (Lampyridae). Each mechanism is dependent upon ATP, luciferin, Mg^{2+} and the enzyme luciferase to create light. Beetle luciferin is regarded to be the same structure in Lampyridae, Phengodidae, and luminescent elaterids (Seliger, McElroy, 1964; Viviani, Becham, 1993). Despite these similarities the difference in colours of light produced in these families is quite dramatic. In lampyrids the light is limited in range from green to yellow (λ_{max} 538–584nm). However, bioluminescent click beetles have three light organs; a pair of dorsal oval light organs on the pronotum which emit a green light (λ_{max} 536–559nm) and a ventral organ located on the first abdominal segment which ranges in colour from green through to orange (λ_{max} 549 – 594 nm). In railroad worms the number of lanterns increases with eleven pairs of luminous organs located dorso-laterally along the abdominal and thoracic segments. These emit green through to orange light (λ_{max} 535 – 592 nm) and are present in both adults and larvae. In addition, some species such as the railroad worm

Phrixothrix have a luminous organ on the head which emits red light (λ_{max} 600–620nm). These colour differences are a result of amino acid differences in the luciferase protein.

In 1998 the luciferase from *Rhagophthalmus ohbai* was characterised (Sumiya et al., 1998). Although *R. ohbai* is currently classified in its own family the Rhagophthalmidae, opinion is still divided as to its placement. In the past it has been contained in the Phengodidae and the Lampyridae. The *R. obhai* luciferase shares greatest sequence identity with the Phengodidae luciferase sequences but this comparison is limited to one species.

1.7.2 Evolution of click beetle bioluminescence

Shortly after the publication of the first firefly luciferase sequence Wood *et al.* in 1989 characterised four different luciferase sequences from a single click beetle species *Pyrophorus plagiophthalamus* (Wood et al., 1989). Sixty beetles were used to construct a cDNA library from which the luciferases where characterised. It was not evident at the time whether these different enzymes could be found in a single beetle or whether the dorsal and ventral lanterns were under different genetic control. Luciferase from one further species of elaterid was carried out in 1999. One luciferase clone was characterised from *Pyrearinus termitilluminans* which produced a blue shifted bioluminescence which, *in vitro*, is the same as that found in *Photuris pennsylvanica Ppe2.* Although one clone was evaluated four other clones were bioluminescent but unfortunately were not characterised. Additional luciferase genes may exist in the genome of *P. termitilluminans* which have yet to be identified.

In 2003 Stolz *et al.* conducted a large study on the same species in Jamaica and found by comparing genomic clones with cDNA sequences there were two different genes controlling bioluminescence independently in the dorsal and ventral lanterns (Stolz *et al.*, 2003). Stolz *et al.* found that the luciferase sequence data seemed to imply an exchange event from the dorsal to the ventral luciferase locus on Jamaica. They extrapolated from this that the ancestral bioluminescent colour state in *P. plagiophthalamus* for the ventral organ was green (Stolz *et al.*, 2003). This exchange event was followed by a series of substitutions in the ventral luciferase locus of *P. plagiophthalamus* that selectively shifted the colour of the ventral organ from green toward longer wavelengths producing a recently derived ventral orange allele arising on the eastern side of the island (Stolz *et al.*, 2003).

This intergenic exchange was later examined in other bioluminescent click beetles and found to be a general phenomenon in *Pyrophorus* species (Feder, Velez, 2009).

1.7.3 Beetle luciferases and bioluminescence spectra

As previously mentioned beetle luciferases use the same luciferin substrate to naturally display light ranging in colour from green (λ_{max} ~ 530nm) to red (λ_{max} ~ 635nm). A specific property that differentiates firefly luciferase from those of the click beetles and railroad worms is pH sensitivity. The optimum pH for luminescence is around 7.8 but in 1964 Seliger and McElroy reported a strong pH dependence of the colour of the emitted light (Seliger, McElroy, 1964). In acidic ($pH < 6.5$) buffer solutions the intensity of the normal yellow-green emission of *P. pyralis* (λ_{max} 562nm) decreases markedly and a low intensity red emission was observed (λ_{max} 616nm). In addition, Seliger and McElroy found divalent heavy metal cations Cu^{+2} and Zn^{+2}, denaturants such as urea and an increase in temperature could illicit the same shift in the spectrum to the red (Seliger, McElroy, 1964).

Conversely, subsequent studies of click beetle and railroad worm luciferases exhibited no red shift in acidic conditions (Viviani, Bechara, 1995; Viviani *et al.*, 1999b; Wood *et al.*, 1989). As a result firefly luciferases have been described as 'pH sensitive' and both elaterid and phengodid luciferases as 'pH insensitive'. It is interesting to note that the original work carried by Seliger and McElroy evaluated the effect of pH and metal cations on the click beetle *Pyrophorus plagiophthalamus* and although they found no red shift in the ventral organ luciferase they did observe a small shift in spectra in the dorsal lantern luciferase suggesting some pH effect. Interestingly they found a blue shift in the dorsal organ luciferase in basic conditions and with the addition of metal cations (Seliger, McElroy, 1964).

To date, four main hypotheses have been presented to account for the range of colour emitted in the beetle bioluminescent reaction. The first explanation was presented in 1971 by White *et al.* who proposed that the excited state of the keto-form of the oxyluciferin anion can relax by emitting red light, whilst the excited state of the enol-form emits yellow-green light (White et al., 1971).

Later, McCapra proposed an alternative model, that colour variation is associated with conformations of the keto form of excited-state oxyluciferin (McCapra, 2000). McCapra proposed that all of the luminescent colours ranging from green to red are generated from twisted intra-molecular charge transfer (TICT) excited states of the keto form. The colour of the light emission should depend on the rotation around the C–C bond of the –N C–C N– moiety. Branchini and co-workers presented partial experimental support for McCapra's mechanism (Branchini et al., 2002).

The third hypothesis assumes that the colour of the bioluminescence is dependent upon the polarization of the oxyluciferin in the microenvironment of the luciferase–

oxyluciferin complex: the higher the polarization, the larger the red shift of bioluminescence (DeLuca, 1969; Gandelman *et al.*, 1993; Ugarova, Brovko, 2001).
The fourth hypothesis, published in 2006 by Nakatsu and co-workers proposed an energy loss control mechanism which is dependent upon the size of the cavity between the luciferase domains (Nakatsu et al., 2006). A non-relaxed form of the keto oxyluciferin should emit yellow-green light. Conversely after geometrical relaxation it should emit red light. The geometrical relaxation is determined by the size of the luciferase cavity.

There has been much discussion about the mechanism that can explain the effect of pH on bioluminescence and the precise nature of these emitters. However, irrespective of the molecular structure of the emitting forms, the organization of the protein environment of the emitter and the flexibility of key amino acid residues contribute significantly to the spectral parameters of beetle bioluminescence.

1.7.4 Adenylate-forming protein family

A number of amino acid residues in beetle luciferase have been found to be highly conserved in a range of related enzymes which are classified as belonging to a large superfamily of adenylate-forming enzymes (PFAM00501). The adenylate-forming proteins catalyze a two-step reaction converting an organic acid to a CoA thioester (Gulick et al., 2003; Reger et al., 2007). This mode of substrate activation is commonly used by adenylate-forming enzymes such as acyl-CoA ligases (Chang et al., 1997), acetyl-CoA synthetases (Gulick et al., 2003), non-ribosomal peptide synthetases (NRPSs) (Kleinkauf, Von Dohren, 1996) and aminoacyl-tRNA synthetases (Delarue, 1995), as well as luciferase. These enzymes are relatively large, ranging in size from 500 to 700 residues. Structurally they are composed of two domains, an N-terminal domain of 400-550 residues and a smaller C-terminal domain of 100-140 residues. An active site is situated at their interface. Members share limited sequence homology of 20-30%, however, several well-conserved sequence motifs have been identified between members and three principle motifs have been attributed with an adenylation function (Chang *et al.*, 1997; Morozov, N., 1997; Stuible *et al.*, 2000; Thompson *et al.*, 1997). Of particular note is the invariant residue K^{529} which was shown to be important in the adenylation step (Branchini et al., 2000).

These enzymes activate a variety of different substrates, including aromatic acids, acetic acid and long-chain fatty acids, to the corresponding enzyme-bound acyl-adenylates, which are then transferred to the thiol group of CoA. The two half-reactions occur in a ping-pong mechanism. A domain alternation mechanism has been proposed

for these enzymes. Upon completion of the initial adenylation reaction, the C-terminal domain of these enzymes undergoes a 140° rotation to perform the second thioester-forming half-reaction.

It has recently been speculated that beetle luciferase may have evolved from an ancestral fatty acyl-CoA synthetase as firefly luciferase retains this activity *in vitro* (Oba *et al.*, 2003; Oba *et al.*, 2006b). As such beetle luciferin may not itself have originally been the substrate for the ancestral luciferase, but rather a 'luciferin-like' molecule, with beetle luciferin appearing as a substrate later in evolution In support of this, dehydroluciferin, differing from luciferin by only two hydrogen atoms and inactive for chemiluminescence, can be efficiently ligated to CoA by firefly luciferase (Fontes et al., 1997). Luciferase may still function as a fatty acyl-CoA synthetase involved in the oxidation of fatty acids in the peroxisome of beetles. Interestingly, it was shown that firefly luciferase had a marked preference for fatty acids such as arachidonic acid (Oba *et al.*, 2003). This may be unsurprising as arachidonic acid, although typically occurring in very small amounts in the phospholipids of terrestrial insects, has been found in very high levels in the tissue lipids of adult fireflies (Nor Aliza et al., 2001).

Recently paralogous luciferase-like sequences have been identified from the Japanese firefly *L. cruciata* (Oba *et al.*, 2006b) suggesting gene duplication of luciferase-like sequences in bioluminescent beetle genomes. Despite extensive sequence identity of the *L. cruciata* luciferase-like genes to the *bona fide* luciferase, the two paralogous enzymes revealed no bioluminescence activity. Furthermore, only one gene product exhibited long-chain fatty acyl-CoA synthetic activity. It was subsequently proposed that luciferase has arisen from a gene duplication event of an ancestral fatty acyl-CoA synthetase and functionally diverged to acquire a novel bioluminescent function (Oba *et al.*, 2006b). Luciferase orthologues found in the non–bioluminescent mealworm *Tenebrio molitor* exhibited no bioluminescent activity and were reported to have acyl CoA synthetase activity (Oba et al., 2006a). Recently, however, a luciferase-like sequence identified in the non-bioluminescent meal-worm *Zophobas morio* was found to exhibit weak bioluminescent activity (Viviani et al., 2009).

Crowson proposed that luminescence must have originated as an accidental by-product of a chemical reaction serving an alternative purpose and that non-adaptive luminescence is unlikely to persist for long in the evolutionary time scale (Crowson, 1981). It seems feasible that luciferase, the only oxygenase in the adenylate-forming enzyme superfamily may have originally played a role in controlling levels of oxidative stress with or without a luciferin. The first light generated from the primaeval bioluminescent beetles would, as Crowson suggests, need to be adapted for a role that would ensure its survival as a trait over time.

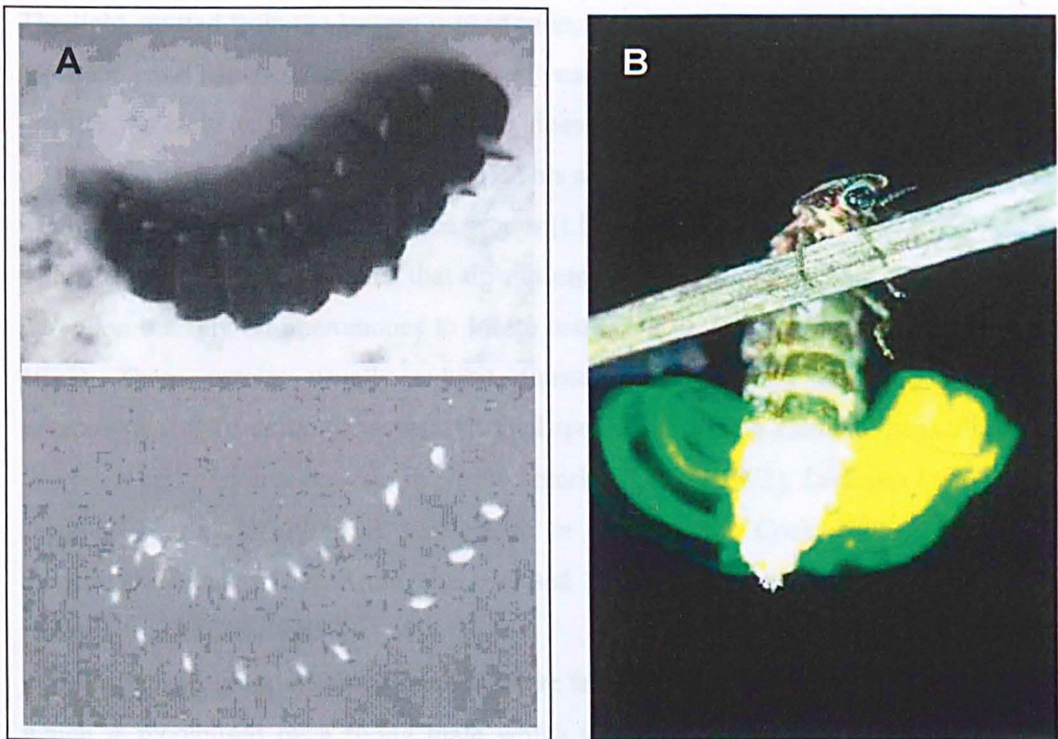
1.8 The Functional evolution in fireflies

It has been proposed that most cells of the primeval firefly would be luminescent (Buck & Case, 2002). This is evident today in the eggs of lampyrids that can glow steadily and even the pupae pass through a stage of body-wide luminescence. However, firefly adults and larvae predominantly emit light from localised areas of the body referred to as the lanterns. Nonetheless, body-wide light leakage is evident in larvae of the glow-worm *Lampyris noctiluca* (*pers. comm.* L. Tici) and other species (Viviani et al., 2008) (Figure 1.5).

Externally lanterns can be identified as white regions primarily found in the terminal abdominal segments of fireflies and are composed of specialised cells known as photocytes. The lantern can vary dramatically between lampyrid species in gross external morphology as well as internal physiology (Buck, 1948). Light organs range in complexity from the simplest form found in glowing species such as *Lamprohiza splendidula* and the larvae of *Phausis* (Bongardt, 1903; Bugnion, 1929; Wielowiejski, 1882) through to the complex rosette-like structures present in the flashing fireflies such as *Photuris* species (Ghiradella, Schmidt, 2004). These photocytes are supplied with fine respiratory and neural branches and thus early on in firefly evolution photocytes must have localised in a tissue which became a target for rapid development of neural and tracheal systems. The adult firefly light organ has been reported not to derive from the larval lantern but from the pupal fat body (Blum, Sannasi, 1974; Hess, 1920; Okada *et al.*, 1974; Williams, 1917). The fat body is found in the abdominal haemocoel and assists energy storage and metabolic processes thereby acting like an insect analogue of the liver.

1.8.1 The evolution of dialogue in fireflies

Although some adult firefly species exhibit no bioluminescence, those that do exhibit a dramatic array of luminescent courtship display ranging from a steady glow through to synchronised flashing (Figure 1.5). Although a large body of research has built up around this phenomenon but only recently have researchers begun to ask questions about the evolution of this process. Evidence so far supports two principle factors shaping the evolution of communication in fireflies, sexual selection through mate choice and natural selection through predation.



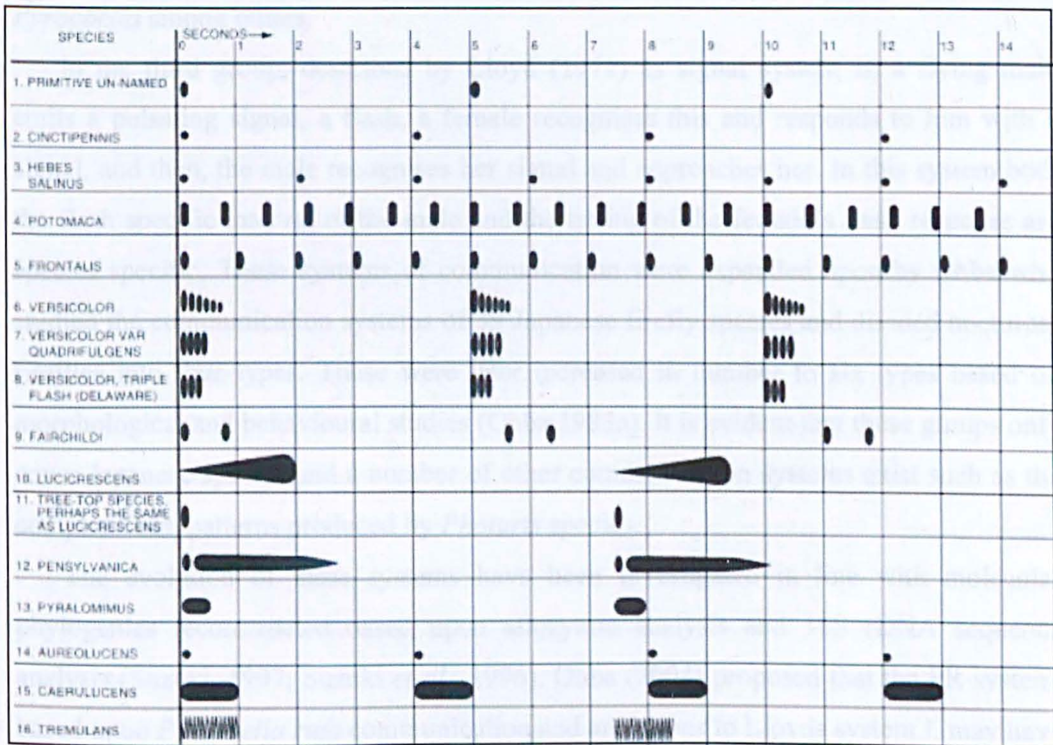


FIGURE 1.5 Bioluminescence and communication in fireflies. (A) *Lampyris noctiluca* light leakage from between sternites of larvae using 1hour exposure on a CCD camera (L. Tici) (B) Female glow-worm (*L. noctiluca*) emitting light from last abdominal segments (C) range of flash patterns recorded ni sixteen *Photuris* species.

The light emitted from the lantern is used to locate other individuals of the same species for mating and has evolved over millions of years into a vast array of different signalling profiles. As early as 1886 the use of bioluminescence in beetles was identified as a form of sexual communication (Emery, 1886). This sexual communication in the Lampyridae has been roughly classified into three groups (Lloyd, 1971).

The first group contain species that do not employ luminescence as a form of sexual attraction but rely on pheromones to locate mates (Hess, 1920; Lloyd, 1972; Williams, 1917). These species usually exhibit diurnal behaviour and evidence supporting pheromone signals exists for several diurnal species, including *Lucidota atra*, *Pyropyga nigricans* and *Photinus indictus* in North America (Lloyd, 1972), *Lucidina biplagiata* in Japan 102 and *Phosphaenus hemipterus* in Europe (De Cock, 2000). These and associated species in Japan are characterised by large antennae and small eyes with reduced facets (Ohba 2004).

The second group contains species where females elicit a continuous signal, a glow, which is recognised by a flying male which then approaches. This was described as signal system I by Lloyd (1971). These species include the genera *Lampyris* and *Pyrocoelia* among others.

In the third group, described by Lloyd (1971) as signal system II, a flying male emits a pulsating signal, a flash, a female recognises this and responds to him with a signal, and then, the male recognises her signal and approaches her. In this system both the flash specific interval of the male and the timing of the female's flash response are species specific. These systems of communication were expanded upon by Ohba who studied the communication systems of 39 Japanese firefly species and divided nocturnal fireflies into four types. These were later increased in number to six types based on morphological and behavioural studies (Ohba 1983a). It is evident that these groups only cover Japanese species and a number of other communication systems exist such as the complex flash patterns produced by *Photuris* species.

The evolution of these systems have been investigated in line with molecular phylogenies reconstructed based upon allozymic analysis and 16S rDNA sequence analysis (Suzuki, 1997; Suzuki *et al.*, 1996). Ohba (2004) proposed that the PR system, based upon *Pyrocoelia rufa* communication and analogous to Lloyds system I, may have been the most ancestral system for bioluminescent communication in the Lampyridae.

1.9 Natural selection through predation: Defence and aposematism in lampyrids

Darwin noted on his travels a number of bioluminescent beetles and speculated on the role of luminescence as a form of sexual selection but was perplexed by the role of luminescence in larva. As far as is known all lampyrid larvae are luminescent and this might be expected to make them conspicuous targets for both vertebrate and invertebrate predation. Little is known about the functions of glowing behaviour in luminous beetle larvae although a number of theories have been presented (Sivinski, 1981). The use of bioluminescence for prey attraction is evident in a number of non-beetle bioluminescent organisms (Buck, 1978). This mechanisms has been speculated as being employed by *Pyrearinus* species (Elateridae) where the larvae inhabit holes in termite mounds and possibly attract flying termites as prey (Redford, 1982). It can be argued that aggressive mimicry is another example of bioluminescence being used for prey attraction (see below). Illumination and communication have also been presented as alternative roles of larval bioluminescence in Lampyridae and Phengodidae (Lloyd, 1983b; Sivinski, 1981; Viviani, Bechara, 1997). But probably the most likely role of bioluminescence in larvae is defence.

1.9.1 Chemical defence

A number of studies have show that adult fireflies appear to be protected against predation and are known to be unpalatable to a number of vertebrate species (Eisner *et al.*, 1978; Lloyd, 1973; Sydow, Lloyd, 1975). Furthermore, in some animals, ingestion of fireflies can lead to toxicosis and ultimately death (Knight et al., 1999).

Steroidal pyrones, found in a range of adult firefly species, appear to be the primary component of this unpalatability (Eisner et al., 1978; Goetz et al., 1981; Goetz et al., 1979; Meinwald et al., 1979). As a result of a close structural relationship to the cardiotonic steroids found in venomous toads (e.g. bufalin) these firefly compounds were named lucibufagins (Eisner et al., 1978). A second type of chemical, the defensive betaine N-methylquinolinium 2-carboxylate is found in *Photuris* species (González et al., 1999)

When disturbed or attacked fireflies emit droplets of blood, primarily from the elytra and pronotum, and such 'reflex bleeding' is regarded as a defensive act (Blum, Sannasi, 1974; Williams, 1917). Similar reflex bleeding has been observed in a range of Japanese firefly species both in the field and laboratory (Ohba, Hidaka, 2002). In these species

reflex bleeding was evident on the margins of the elytra, the pronotum and the antennal sockets.

1.9.2 Larval defence

Defence compounds in both adults and eggs have been studied in depth but less is known of the defence mechanisms of lampyrid larvae. Bioluminescence appears to be a universal phenomenon in lampyrid larvae (Crowson, 1972), though not necessarily in adults (McDermott, 1964), and is usually induced by disturbance, although a number of species glow continually during periods of nocturnal movement (Dreisig, 1974). The presence of defensive compounds in larvae has only recently been investigated and preliminary studies indicate a number of volatile compounds may be present in the immature stages of lampyrids. Although the *Photuris* defence betaine, N-methylquinolinium 2-carboxylate, was found in the larvae as well as adults and eggs (González et al., 1999). It is not know whether this is actively used as a defence compound or whether this is an artefact of egg defence. Although reflex bleeding is rarely seen in immature lampyrids observation in the field indicate firefly larvae are capable of deterring predation from a range of invertebrate and vertebrate predators Experimental studies of the larvae of *Lampyris noctiluca* have shown unprofitability to birds (De Cock, Matthysen, 2001; Tyler, 2001a) and to wood ants (*Formica rufa (L.)* (Tyler, 2001b). Lampyrid larvae can survive an onslaught for many hours before being overwhelmed by ants (Fu et al., 2007); it is presumed that the defensive compounds released are eventually depleted. It has been observed that ants while avoiding bioluminescent larvae will eat dead larvae (Travers, 1924).

The presence of numerous organs situated along the larvae which evert when physically stimulated have been observed in *L. noctiluca* and have been implicated as a possible defence mechanism against predation (Tyler, Trice, 2001). These organs are situated on either side of abdominal segments 1 to 7 and evert when larvae were placed on wood ant nests and the larvae remained more or less immune from attack for over an hour (Tyler, 2001b). Similar defensive organs, described as bifid epidermal processes on the abdominal and thoracic segments, are present in a number of aquatic larvae belonging to the firefly genus *Luciola* (Fu *et al.*, 2007; Ohba *et al.*, 1994; Ohba, Hidaka, 2002; Okada, 1928). In additional recent studies have found homologous structures in *Diaphanes* and *Pyrocoelia* species (Fu et al., 2009). In *L. cruciata* larvae the organs are not confined to the abdomen, occurring on the meso- and meta-thorax as well as on abdominal segments 1 to 8. As with *L. noctiluca,* the defensive organs remain hidden while the larva is at rest, but upon agitation they are everted and emit an odour

reminiscent of 'resin and peppermint' giving early indications that a defensive chemical is being discharged (Ohba, Hidaka, 2002; Okada, 1928).

The larvae of numerous insect species have evolved a diversity of glandular structures for the conservation and discharge of a multitude of toxic, deterrent, and repellant secretions, which can act as effective defense (Blum 1981; Evans and Schmidt 1991). Although adult *Photinus* and *Photuris* fireflies possess quinolines and lucibufagins, which serve as cardiotonic and emetic agents (Eisner *et al.* 1978; 1997), it is unknown whether these compounds are present in their respective larvae. However, in *L. noctiluca* larvae a similar compound to lucibufigen was partially characterised and it has been speculated that this compound carried in the haemolymph is released from vesicles covering the outer surface of the larvae's plural organs (Tyler et al., 2008). In the glands of *L. leii*, two volatile terpenoids were identified: a large amount of terpinolene and a smaller component, γ-terpinene (Fu et al., 2007). In numerous invertebrate defensive secretions terpenes and in particular terpinolene, are established as toxic, deterrent or repellant agents (Moore 1968; Billen *et al.* 2000; Aldrich 1988; Krall *et al.* 1997). To date, however, no direct evidence links the release of defense chemicals by these glands in response to the threat from predation. The range of aromas emitted by different species suggests a range of chemicals may be involved in defending lampyrid larvae from attack.

1.9.3 Evolution of Aposematism

The idea of warning signals manipulating predator foraging behaviour was first developed by Wallace in response to Darwin's sexual selection theory (Wallace 1867). The production of a signal, which can be a distinctive colour, odour or behaviour, to the predator that the prey is unprofitable can, in turn, enable and encourage the predator to change to more profitable prey. The association between signal and unprofitability is known as *aposematism* (Poulton, 1890).

Because firefly larvae are flightless, slow-moving, soft-bodied, and feed in exposed locations, they represent highly vulnerable targets for predators. As early as 1874 it was speculated that luminosity served as warning coloration for an indication of inedibility (Belt, 1874). Although the function of luminescence in beetle larvae is still open to speculation aposematism appears to illicit the most favourable response (Cowles, 1959) (De Cock, Matthysen, 1999; De Cock, Matthysen, 2001; De Cock, Matthysen, 2003; Guilford, Cuthill, 1989; Lloyd, 1973; Sivinski, 1981; Underwood *et al.*, 1997). Whilst the function of aposematism is currently well understood only a few studies have been conducted directly in fireflies. It has been demonstrated that laboratory mice reject

Photuris versicolor larvae and learn to avoid luminescent artificial prey (Underwood et al., 1997). Toads have been shown to be reluctant to attack luminescent artificial prey and being presented with luminescence *L. noctiluca* larvae was sufficient for aversion learning (De Cock and Matthysen 2001; 2003). Furthermore, larvae exhibit the key components of an aposematic defensive syndrome: 'early-warning' signals of unpalatability i.e. glowing, emitting strong odours, poor escape behaviour and distastefulness or toxicity.

McDermott (1964) and Sivinski (1981) hypothesised that luminescence first evolved as an aposematic signal in larvae; this was later supported by phylogenetic studies (Branham and Wenzel 2003). Branham and Wenzel went on to hypothesise that luminescence in the adult appears to function as an aposematic warning display, which has been co-opted in many species to serve also as a sexual signal used in courtship (Branham, Wenzel, 2003).

The aposematic warning display in adults is likely to be linked to reflex bleeding which has been identified in a large number of nocturnal firefly species. In addition, reflex bleeding has been observed in a number of adult diurnal firefly species (Ohba, Hidaka, 2002). In general, diurnally active firefly species emit a very weak light or are not luminescent and vivid colouration in these species in association with reflex bleeding has been proposed as the basis for an aposematic response to diurnally active predators (Ohba, Hidaka, 2002). However, as McDermott noted most diurnal species are not brightly coloured (McDermott, 1964) which suggests that there may be a more complex evolutionary story. It has been suggested that some diurnal species have changed from nocturnal to diurnal behaviour in order to avoid predation (Gronquist *et al.*, 2006). Thus the gain of warning colouration by diurnal species may be a relatively recent event with luminescence being the ancestral warning signal in adult fireflies.

1.10 Summary

Primitive ancestors of the current bioluminescent beetles at their non-bioluminescent stage would have utilised a protoluciferase in the form of an ancestral fatty acyl CoA synthetase. Over time this would have evolved a novel oxygenase function through gene duplication and neofunctionalization. This protoluciferase may have functioned as an early oxygen detoxification enzyme possibly acting on a substrate other than luciferin. At some point luciferin was synthesised by these ancestral beetles which would ultimately give rise to the first luminescent reaction. Bioluminescence may have emerged in early ancestors of the current Phengodidae, Lampyridae and Elateridae or as most researchers speculate arising independently after these families diverged post 170

MYA after most beetle lineages diversified. Given the current distribution of species it seems likely that bioluminescent beetles first appeared in the Neotropics before spreading throughout the world as in the case of the Lampyridae.

`The first bioluminescence would have been weak, probably restricted to the larvae and non-localised. The first light would have probably been green given the abundance of this wavelength throughout the bioluminescent beetle families and the evolutionary evidence for the gradual change from green to other colours in the ventral lanterns of click beetles supports this.

This bioluminescent reaction may still have functioned as an oxygen detoxification pathway utilising luciferin with light emission as a secondary artefact of this essential role. Light emission would become stronger over time, the addition of a luciferin regenerating pathway would provide a more efficient reaction resulting in visible light being emitted from key regions around the larval body. This bioluminescence may have enhanced an initial aposematic system in larvae and thus been selected for in certain lineages. These first bioluminescent beetles would have been nocturnal in order for this aposematic response to bioluminescence to work efficiently. Lanterns would develop over time and give a more concentrated light that eventually would be carried through into other life stages as a defensive mechanism. Ultimately lanterns became more complex and under nervous control resulting in flash communication and the subsequent evolution of species specific flash patterns that we see today.

1.11 Aims and Contents of the Thesis

The aim of this thesis is to examine the evolution of enzymes involved in beetle bioluminescence from a genetic, enzymatic and phylogenetic perspective. Such a study is desirable in that it informs future mutagenic studies as well as providing novel data on functional residues, conserved domains and gene paralogues that may be advantageous when improving the efficiency of bioluminescent systems. In a broader context such studies provide important information on gene duplication events within beetle genomes and the evolutionary pressures exerted on paralogous genes in the Coleoptera, a vastly understudied subject.

The overarching aim of this thesis is ***to investigate the genetics of beetle bioluminescence in order to identify possible historical scenarios for the origin and evolution of beetle bioluminescence.***

The present work is organised into five chapters, the first being an introduction to the literature relating to the evolution of bioluminescent beetles.

Chapter 2 consists of a genomic evaluation of insect SGL genes followed by a phylogenetic examination of the paralogues present in beetle genomes. This dataset was augmented with partial gene sequences obtained by degenerate PCR of both bioluminescent and non-bioluminescent taxa along with full length gene sequence from key species using genome walking. Finally a number of possible evolutionary scenarios to explain the extent of gene duplication in insect genomes and the putative biological function of firefly LRE and its paralogues are discussed.

Chapter 3 examines the extent of gene duplication of the AMP-forming proteins (AFP) in beetle genomes of which beetle luciferase is a member. A bioinformatics study was carried out on the published genome of the red flour beetle *Tribolium castaneum*. Extensive gene duplication of AFP genes was evident; ten different loci were identified, three of which were found in close proximity on the same chromosome. Degenerate primers were used to amplify luciferase and luciferase-like sequences in fifteen bioluminescent and non-bioluminescent taxa. A phylogenetic study was conducted to examine the relationship of luciferase and luciferase-like sequences in the complete dataset. Bayesian analysis was used to examine different evolutionary pressures acting upon different clades of the tree and a predicted evolutionary scenario is presented to account for the luciferase paralogues in bioluminescent beetle genomes.

It was hypothesised that lampyrid paralogues may be present in a tandem array, as is evident in some *T. castaneum* loci. Chapter 4 presents an examination of the flanking region of the luciferase gene in the glow-worm *Lampyris noctiluca*. A 7 kb region upstream of the *L. noctiluca* luciferase gene was amplified and sequenced. No luciferase-like sequences were identified upstream of this luciferase gene. However, one open reading frame (ORF) was found, 686 bp upstream of the luciferase start codon. The ORF showed strong sequence identity to a retrotransposase sequence found in other insects. The identification of this sequence suggests that luciferase may have been duplicated from an ancestral fatty acyl CoA synthetase via retrotransposition.

Chapter 5 is a synthesis of the research presented in previous chapters. The findings of the thesis are discussed in the context of the evolution of bioluminescence beetles as a whole, conclusions are drawn and further research recommended.

2. Origins and diversification of the SGL gene family and the origins of the Luciferin Regenerating Enzyme (LRE) in bioluminescent beetles

2.1 Introduction

The ability to produce light in beetles has been studied for many years resulting in a comprehensive knowledge of the sequence, structure and enzymatic properties of beetle luciferase (Baldwin, 1996; Conti *et al.*, 1996; de Wet *et al.*, 1987; Deluca, 1969; McCapra *et al.*, 1994). Studied to a much lesser extent is the luciferin regenerating enzyme (LRE). The activity of regenerating luciferin from the luminescent product oxyluciferin was first identified in the protein fraction of a lantern extract from the firefly *Photinus pyralis* (Gomi, Kajiyama, 2001a). Oxyluciferin, which is the product of the luciferase reaction, has a strong inhibitory effect on the firefly luciferase in a manner competitive with firefly luciferin (Gates, DeLuca, 1975; Suzuki, Goto, 1971). The purified LRE protein from the firefly *Photinus pyralis* was found to be a single polypeptide with a molecular mass of 38 kDa and converted oxyluciferin to 2-cyano-6-hydroxybenzothiazole and thioglycolic acid (Gomi, Kajiyama, 2001a). In the presence of D-cysteine, 2-cyano-6-hydroxybenzothiazole was turned over into luciferin. LRE cDNA was characterised from the lantern of *P. pyralis* and was found to code for 308 amino acids (Gomi, Kajiyama, 2001a). LRE cDNA was further isolated from *L. cruciata* and *L. lateralis,* expressed and found to exhibit the same enzymatic properties (Gomi *et al.*, 2002). Although the protein was isolated from the lantern it was noted that LRE, unlike luciferase, is not exclusively localised in the adults lantern but found elsewhere in the beetle (Gomi *et al.*, 2002). In the study of Gomi et al. 2002 the LRE amino acid sequence was found to have identity to a group of enzymes known as Senescence Marker Protein-30 (SMP30) (Gomi *et al.*, 2002). SMP30 is a 34 kDa protein whose tissue levels in the liver, kidney, and lung decrease with aging (Fujita *et al.*, 1996b; Fujita *et al.*, 1992; Mori *et al.*, 2004). Once thought to be involved in calcium regulation SMP30 has been recently identified as the metazoan equivalent of the bacterial lactone-hydrolyzing enzyme gluconolactonase (GNL) (Kondo *et al.*, 2006). GNL is a key enzyme which is involved in vitamin C (L-ascorbic acid) biosynthesis and SMP30 knockout mice displayed symptoms of scurvy when fed a vitamin C-deficient diet (Kondo *et al.*, 2006). Despite the sequence identity, SMP30 in

mice has been shown to lack any luciferin regenerating activity but conversely both LRE and SMP30 have been shown to hydrolyse diisopropyl phosphorofluoridate (DFP) (Kondo et al., 2004). This suggests that LRE originated as a SMP30 and has evolved a novel role in bioluminescent beetles or even earlier in insect evolution. Based upon sequence conservation and the recent evidence for conserved function the three classes of enzyme GNL, SMP30 and LRE have been designated as the SGL (SMP-30/Gluconolaconase/LRE-like) protein family (Pfam - PF08450).

Firefly LRE stands apart as a enzyme with a unique catalytic activity, the regeneration of oxyluciferin into luciferin, an activity that is not exhibited by mammalian SMP30 (Kondo *et al.*, 2004). To date no studies have evaluated the evolution of LRE and other SGL enzymes in beetles as well as other insects. It is not established whether LRE has a dual function or has arisen from an ancestral SGL gene through gene duplication. Elucidating the evolutionary history of the SGL family may provide insights into the origins of SGL enzymes and, in particular, the origins of the LRE in the Lampyridae. This comparative analysis will also provide a foundation to experimentally test subfunctionalization in gene copies that originated after gene duplication events in beetle genomes.

2.1.1 Aims

This study investigates the evolution of the SGL gene family in beetles in order to establish the evolutionary origins of the LRE in bioluminescent beetles.

1. The full length LRE gene from *Photinus pyralis* will be characterised to determine intron and exon positions informing degenerate primer design (see below) as well as informing the prediction of intron/exon boundaries in other taxa.
2. Degenerate primers will be designed to amplify SGL genes from a range of beetle taxa. These will be used to construct a phylogeny of SGL genes in beetles and examine the diversification of SGL genes in coleopteran genomes.
3. Full length gene sequences along with flanking sequence will be obtained for the bioluminescent beetle *P. pyralis* and the non-bioluminescent beetle *Pachnoda marginata* ssp. *peregrina* to provide additional data for the aims 5-8.

4. RT PCR will be used to investigate the expression and pseudogene status of SGL genes in beetles. Degenerate primers and gene specific primers will be employed.
5. Genome databases will be mined in order to obtain a full sequence dataset of SGL genes. Phylogenetic analysis of this dataset will provide an indication of evolutionary relationships of the SGL genes in metazoans.
6. Virtual mapping will be used to investigate microsynteny of SGL genes along with an investigation into the presence of conserved promoter regions.
7. Using the crystal structure of the SGL protein from *Agrobacterium tumefaciens* (RCSB protein databank (PDB) accession code, 2GHS) the substrate recognition site, (SRS) will be predicted for GNL activity and key residues highlighted that may be involved in LRE activity.
8. The program DIVERGE will be used to examine the divergence of LRE genes from those of other SGL genes found in beetles.

2.2 Materials and Methods

2.2.1 Sequence data retrieval

Assembly 2 of the *Tribolium casteneum* genome sequence (Wang et al., 2007) at the Baylor College of Medicine Human Genome Sequencing Center, Tcas_2.0[1] and BeetleBase[2] was searched via NCBI with all available insect SGL/LRE proteins as queries using translated BLAST searches (tBLASTn) (Gertz *et al.*, 2006). *T. casteneum* SMP30 sequences retrieved were in turn employed in searches to find more genes in an iterative process. To find highly divergent, already annotated members of these families, multiple PSI-BLASTP searches were initiated. The genes were reconstructed manually in BioEdit VERS 7.0 (Hall, 1999) using the expected exon/intron boundaries from the *Photinus pyralis* LRE gene sequence as a guide (see below). In addition, protein alignments were used to identify irregularities and refine the gene structures. Partial sequence information has not been considered in the present investigation. Further searches in the *Tribolium castaneum* genome involved using SGL and LRE sequences as probes in PSI-BLAST searches (Altschul et al., 1997) against the NCBI (Wheeler et al., 2005), Swiss-Prot (Gasteiger et al., 2003) and Ensembl (Hubbard et al., 2005) protein databases and through tBLASTn against the full *T. castaneum* genomic

[1] http://www.hgsc.bcm.tmc.edu/projects/tribolium
[2] http://www.bioinformatics.ksu.edu/BeetleBase

sequences available at the NCBI website. In the case of proteins with possible splice variants the longest predicted isoforms were used.

Further genome assemblies mined included: *Ciona intestinalis* (Dehal et al., 2002); *Apis mellifera*, Baylor, Amel_2.0 (NHGRI and USDA); *Bombyx mori*: BGI, 2003-10-01 (Xia et al., 2004); *Anopheles gambiae*, *Anopheles* Genome Consortium, AgamP3 (Zdobnov et al., 2002); *Aedes aegypti*: *Aedes* Genome Consortium, aedes_aegyti_1 (Nene et al., 2007); *Drosophila pseudoobscura*: Flybase, release 1.04 (Drysdale et al., 2005); *D. melanogaster*: Flybase, release 4 (Drysdale et al., 2005) and[3]. All assemblies were searched with all available insect SMP30 protein sequences as queries using *Nasonia vitripennis* tBLASTn. The genes were edited manually in BioEdit vers. 7 (Hall, 1999) using the expected exon/intron boundaries from dipteran SGL and the firefly *Photinus pyralis* LRE as guides. In addition, protein alignments were used to identify irregularities and refine the gene structures. To find evidence for the conservation of synteny, we compared genomic regions neighbouring the *Tribolium castaneum* SGL genes with the genes neighbouring SGL genes from other insects.

2.2.2 Beetle material

Twenty three coleopteran species from one adephagan and eleven polyphagan families formed the basis of this study (Table 2.1). Material was collected from a range of locations provided by a number of entomologists. The majority of non-bioluminescent taxa were donated by the University Museum in Oxford. A significant proportion of these were acquired from public inquiries and collection locations have therefore not been cited. All specimens were preserved in ethanol and stored at -20°C until processed with the exception of *Pachnoda marginata* ssp. *peregrina* and *Necrophorus vespillo* which were stored at -70°C and -20°C respectively without ethanol. All non-bioluminescent beetles were identified by James Hogan at the University Museum in Oxford. Belgian and Portuguese specimens were identified by Raphaël de Cock, *Lampyroidea maculata* by Michael Geisthardt, *Photuris* species by James Lloyd and *Phausis reticulata* by Lynn Faust.

[3] http://www.hgsc.bcm.tmc.edu/projects/nasonia/

Classification		Source/Location
POLYPHAGA		
SUPERFAMILY ELATEROIDEA		
Lampyridae	*Photinus pyralis* Linnaeus, 1767	Sigma Aldrich/Knoxville, USA
	Photuris congener LeConte, 1851	Gainesville, Florida
	Lampyris noctiluca Linnaeus, 1758	Sevenoaks, Kent
	Phausis reticulata Say, 1825	Knoxville, USA
	Lampyroidea maculata Geisthardt & Day, 2004	Amol forest, Iran
	Photuris sp. 'AC' Lloyd	Gainesville, Florida
	Phosphaenus hemipterus Fourcroy, 1785	Antwerp, Belgium
	Lamprohiza splendidula Linnaeus, 1767	Antwerp, Belgium
Cantharidae	*Cantharis rufa* Linnaeus, 1758	University Museum, Oxford
	Cantharis rustica Fallén, 1807	University Museum, Oxford
SUPERFAMILY BOSTRICHOIDEA		
Dermestidae	*Dermestes ater* De Geer, 1774	University Museum, Oxford
Anobiidae	*Stegobium paniceum* Linnaeus, 1758	University Museum, Oxford
SUPERFAMILY HYDROPHILOIDEA		
Hydrophilidae	*Anacaena* sp. Thomson, 1859	University Museum, Oxford
	Helophorus grandis Illiger, 1798	Stow Park, Bucks
SUPERFAMILY STAPHYLINOIDEA		
Silphidae	*Necrophorus vespillo* Linnaeus, 1768	Cuckhamsley Hill, Oxon
SUPERFAMILY SCARABAEOIDEA		
Scarabaeidae	*Aphodius rufipes* Linnaeus, 1758	University Museum, Oxford
	Pachnoda marginata ssp. *peregrina* Kolbe, 1906	University Museum, Oxford
SUPERFAMILY CHRYSOMELOIDEA		
Cerambycidae	*Grammoptera ruficornis* Fabricius, 1781	University Museum, Oxford
	Strangalia melanura Fabricius, 1792	Burnham, Bucks
Chrysomelidae	*Plagiodera versicolora*Laicharting, 1781	University Parks, Oxford
SUPERFAMILY CLEROIDEA		
Cleridae	*Tillus elongatus* Linnaeus, 1758	Burnham, Bucks
Melyridae	*Anthocomus fasciatus* Linnaeus, 1758	Lytton Road, Oxford
SUPERFAMILY TENEBRIONOIDEA		
Pyrochroidae	*Pyrochroa serraticornis* Scopoli, 1763	University Museum, Oxford
Tenebrionidae	*Tenebrio molitor* Linnaeus, 1758	University Museum, Oxford
ADEPHAGA		
Carabidae	*Dromius quadrimaculatus* Linnaeus, 1758	St Catherines College, Oxford

TABLE 2.1 List of coleopteran specimens studied.

2.2.3 DNA and RNA extractions

For all specimens DNA was extracted using the High Pure PCR Template Preparation (Roche) kit according to the manufacturer's instructions. For larger specimens such as *Pachnoda marginata* ssp. *peregrina* and *Necrophorus vespillo*, abdominal tissue was used for extraction. For the remaining specimens all material was used in the extraction process with the exception of the head and elytron. Total RNA was isolated using SV Total RNA Isolation system (Promega) according to the manufacturer's protocol from 30 mg of *P. m. peregrina* abdominal tissue and the abdomen and thorax of *L. noctiluca* and *L. splendidula* (elytron and head were removed prior to extraction).

2.2.4 PCR amplification of the LRE gene from *Photinus pyralis*

Based upon the LRE mRNA sequence of Gomi and Kajiyama, 2001 (GenBank accession number AB062786), primers *Pp*LRE F1 and *Pp*LRE R1 (Table 2.2) were designed to amplify the entire LRE gene from *P. pyralis*. PCR was performed in a 50 µl reaction containing 1 µl DNA (10-100 ng), 1 x reaction buffer (Sigma Aldrich), 0.2 µM of each primer, 0.5 mM dNTP, 1µl of DMSO and 5 U of Accu*Taq* LA DNA polymerase (Sigma Aldrich).

2.2.5 Cloning and sequencing

Escherichia coli (strain DH5α) competent cells were prepared and transformed with the plasmid DNA using the standard procedure (Inoue et al., 1990). Transformed bacteria were then plated out onto LB agar plates (1% tryptone, 0.5% yeast extract, 1% NaCl and 1.5% L agar) containing 50 µg/ml ampicillin, 40 mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-gal) in dimethylformamide and 40 µg/ml isopropyl β-D-1-thiogalactopyranoside (IPTG) and incubated at 37°C for 12 hours. Recombinant clones were identified using blue/white selection and transferred to 10 ml of L-broth (1% tryptone, 0.5% yeast extract, 1% NaCl) supplemented with 50 µg/ml ampicillin. Cultures were incubated for 12 hours at 37°C with shaking and the plasmid DNA subsequently purified using the QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instructions. Plasmid DNA was sequenced in both directions using standard M13F and M13R primers and internal primers on an ABI 377 automated sequencer (Applied Biosystems) using the BigDye Terminator vers. 3.1 Cycle Sequencing Kit according to manufacturer's protocol.

Name	Oligonucleotide sequence (5' – 3')
1 F	GGNGAAGGYCCNCACTGGGAT
1 R	GCHACCCATAAATTACCNTCNGYGTC
2 R	AAACNGCHACCCATAAATTHCC
2 F	TTGGAGARGGYCCHCAYTGGGA
3 F	GGAGAAGGTCCWCACTGGGATSAYGAA
3R	GCNACCCATAAATTACCATCSKC
PpLRE F1	ATGGGGCCAGTTGTTGAAAAAATTG
PpLRE R1	TCATAGCTTCACTTTAACTCCCGCG
PpLRE1GW5'1	AGGTCCTCACTGGGATCATGAAACT
PpLRE1GW5'2	CTCCAACCGTATACTTGCCAAGT
PpLRE1GW3'1	CATCTGCTGGTCTTCAGCTTGACGA
PpLRE1GW3'2	GTCTACAGAGTAACAGGTTTAGGCGT
PpLRE2GW5'1	TGCCCACTTTTATATGAGTATGCT
PpLRE2GW5'2	CTTCTTTACGATATCCACCCAGTA
PpLRE2GW3'1	CTAATCGTCAACCACTATTTAGTCTGGA
PpLRE2GW3'2	CATTCCAGGCTTTCCGGATGGCCA
LnLRE2GW5'1	ATGGGTATGTCTTTTAAGAGATGGTACA
LnLRE2GW5'2	GTGAATAGATTTATCTACAAGGTCCACCAA
LnLRE2GW5'3	GAGGGAAGTAAGCTGTACTAGACAGA
LnLRE2GW5'4	ATCAAAGTAGGCCGATCATCCTGA
LnLRE2GW3'1	AATCGTCAAACCTTGTTTAGTCTTGA
LnLRE2GW3'2	CAAATCCCGGGATTTCCAGATGGTCA
LnLRE2GW3'3	TCGTGACGAGTAAACCGGGATAAAGA
LnLRE2GW3'4	GTCTAAGGTTATGGCTACTTAATGCT
PmSGL1GW5'1	AATGTATTTATGAATAGCTTGCCCCA
PmSGL1GW5'2	ATGTCCACGTAATACAAGCATTGAGA
PmSGL1GW3'1	TTTTGATTTGGATCACCATGACATCCCT
PmSGL1GW3'2	GGAGTACCTGATGGTATGACTATAGA
PmSGL2GW5'1	GTTTACCAACTAGAGCCCTCGCATGCAAGT
PmSGL2GW5'2	TGACACATATTTATGGATAGTATGCTCA
PmSGL2GW3'1	AGGTGATTTTTCGTTTGAGAGAAGCT
PmSGL2GW3'2	GTATCGATGGCTTTGCAGATGGACAA
PmSGL3GW5'1	GTATCGATAAATAGATCTATCGTTTGTAT
PmSGL3GW5'2	CACGAAATAAAGAGATTGTGTGTCCT
PmSGL3GW3'1	CTAGATTGTTTACTTTTGACGATCACGA
PmSGL3GW3'2	GGAGCGCCTGACGGTATGACAATAGA
PmSGL1F	ATGGCGCCAGTTGTAGAAGTTGT
PmSGL1R	CTTCAAATTGTCCAACCTAACT
PmSGL2F	ATGCCCGTAATTGTAGAACGTCT
PmSGL2R	CATCTTAAAACAATTTCCATGAACA
PmSGL3F	ATGGCGGCTCAAATCGAAGCCCTAGT
PmSGL3R	CATAAATTCTAAATTAACTCTATTCCCGGGA

TABLE 2.2 Primers used in PCR and sequencing of beetle SGL genes.

2.2.6 Degenerate primer design and PCR

For amplification of the LRE gene a gradient block was initially employed providing up to 12 variable temperatures. PCR conditions generating the greatest amplicion yield could therefore be rapidly assessed. DNA fragments were amplified from the *L. noctiluca* DNA by the polymerase chain reaction (PCR) using two degenerate primers based upon conserved sequences identified from an alignment of all three firefly LRE sequences. Primer 1F and Primer 1R (Primer set 1) were used to amplify part of the first exon through to the fourth exon (Table 2.2; Figure 2.1). PCR was carried in a total volume of 25 μl containing 0.2 μM of each of the forward and reverse primers, 10 mM of each deoxynucleoside triphosphate, 1 U of *Taq* DNA polymerase (Sigma Aldrich), and 2 mM $MgCl_2$ in 1x PCR buffer (Sigma Aldrich). Optimization of thermal cycling was carried out on gradient block of the PTC-225 Tetrad thermal cycler (MJ Research) using the following PCR thermal cycling profile: denaturation for 2 min at 94°C; followed by 35 cycles of: 94 °C for 30 s, 45 °C – 67 °C at 2 °C increments 30 s, 72 °C for 2 min; and a final extension cycle of 72 °C for 10 min. Twelve duplicate samples were prepared for each reaction and corresponding annealing temperature. PCR products were visualised, cloned and sequenced as described previously. All twenty three coleopteran species were amplified using Primer set 1 and the optimal PCR conditions.

Due to the problem of pseudogene amplification additional primers were designed to overcome this problem. Based upon a comparative alignment of lampyrid LRE and pseudogenes two primers were designed to attempt to amplify the genuine LRE from *L. noctiluca* and other bioluminescent taxa. The position and sequence of these primers denoted: Primer 2F and Primer 2R (Primer set 2) are shown in Figure 2.1 (Table 2.2). Reaction mixes were prepared as previously described. PCR was carried out using the following conditions: 94 °C, 1 min; 42 °C, 1 min; 72 °C, 1 min; 35 cycles. PCR products were visualised and prepared as described previously.

Finally a further three primers were designed in order to obtain LRE gene sequence from specimens that had failed to amplify anything with the previous two sets. These primers denoted Primer 3F and Primer 3R (Primer set 3) (Figure 2.1; Table 2.2) were used as described for the previous primer set.

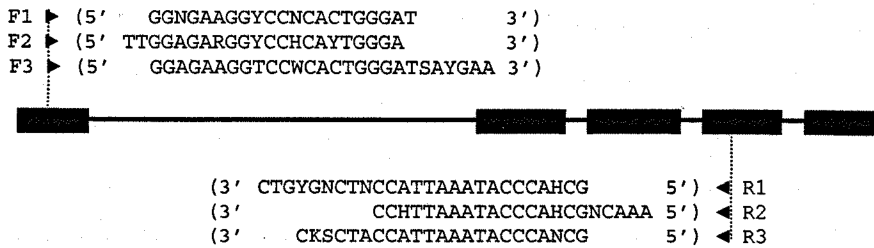


FIGURE 2.1 Position of degenerate SGL primers in relation to the LRE gene of *Photinus pyralis*.

2.2.7 Full length gene characterisation using genome walking

The genome walking method provides a systematic identification of unknown regions flanking a known DNA sequence. The use of PCR to isolate unknown flanking regions of known DNA sequence provides a rapid and economical approach that negates the need to construct a BAC library. Different PCR methods, such as inverse PCR (Ochman *et al.*, 1988), vectorette PCR (Arnold, Hodgson, 1991; Kondo *et al.*, 2006) , and adaptor-specific PCR (Siebert *et al.*, 1995) have been used to isolate unknown flanking regions. In all these examples, genomic DNA is digested with restriction enzyme and ligated to itself, to a vectorette or to adaptors respectively. The ligated product is then used as a template for amplifying flanking regions using PCR.

The Universal Genome Walking Kit (Clontech) employs the adaptor approach. In brief, separate DNA aliquots were digested to completion with four different blunt-end restriction endonucleases (*Dra*I, *Eco*RV, *Pvu*II and *Sca*I). To each digested DNA reaction GenomeWalker adaptors were added along with T4 DNA ligase and ligations carried out overnight at 14°C. An amine group on the lower strand of the adaptor blocks extension of the 3' end and thus prevents an adaptor primer site forming on the general population of adaptor modified DNA fragments. Unincorporated adaptors were removed by a phenol/chloroform extraction and after ethanol precipitation the DNA was re-suspended in TE buffer and used as a template for PCR. An adaptor primer and a gene specific primer were used in a first round of PCR. This PCR product, which is diluted 1:20, provides the template for a second round with nested, non-overlapping primers.

Genome walking was carried out on *P. m. peregrina*, *L. noctiluca* and *P. pyralis* DNA. The partial gene/transcript sequence obtained with the degenerate primers along with the full *P. pyralis* LRE gene sequence were used as the basis of genome walking.

Nested primers were designed for each successive walk both in a 5' and 3' direction. For *L. noctiluca* a modified protocol was employed (Rishi et al., 2004). Partial digestion was used with *Eco*RV and *Sca*I in order to generate larger fragments. Genomic DNA was only digested for 1 hour at 37°C as opposed to overnight digestion recommended by the Universal Genome Walking protocol. Furthermore, digested DNA was size selected by running on a 1% agarose gel and DNA ranging in size from 4 kb to 12 kb excised and gel purified using the QIAquick Gel Extraction Kit (Qiagen). The ligations and PCR amplification was carried out as described above.

2.2.8 RT-PCR amplification

RNA from the scarab beetle *P. m. peregrina* was reverse transcribed in a 20 µl reaction using the ThermoScript RT-PCR system (Invitrogen). 9ul of RNA was pretreated with Primer 1R (0.5 µM) and 10mM dNTPs by heating for 5 minutes at 65°C and then placing on ice. A sample mix was made in a total of 20 µl containing the previous reaction and 1x cDNA Synthesis Buffer, 5 mM DTT, 40 U RNaseOUT™ and 15 U ThermoScript™ RT. This sample was incubated for 60 min at 50°C then terminated by incubating at 85°C for 5 minutes. This cDNA was used as a template for the PCR along with a positive control of total genomic DNA from each respective species. PCR was carried out as described above using primer set 1 but using the following PCR thermal cycling profile: denaturation for 2 min at 94°C; followed by 35 cycles of: 94 °C for 30 s, 49°C for 30 s, 72 °C for 2 min; and a final extension cycle of 72 °C for 10 min. PCR products were visualised, cloned and sequenced as described previously.

Primers were designed to the predicted start and end of all *P. m. peregrina* SGL genes and used to amplify complimentary DNA from mRNA. Three sets of primers were employed PmSGL1 F & R, PmSGL2 F & R and PmSGL3 F & R for each respective gene (see Table 2.2). The conditions described above were used to produce full length cDNA product for each of the three *P. marginata* SGL genes.

2.2.9 Sequence alignment

All the SMP30 sequences were aligned with the software ClustalX v.1.81 (Thompson *et al.*, 1994) using the BLOSUM30 matrix. BIOEDIT vers 7.0.5.3 (Hall, 1999) was used for manual improvement of the alignment by eye. This led to an alignment of 51 SMP30 domain-containing proteins having 341 amino acid positions. A subset of this data was combined with the degenerate PCR products from 20 beetle taxa having 209 amino acid positions. Sequence logos for the protein alignments (whole dataset and

subsets) were created online using WebLogo Version 2.8.2 (http://weblogo.berkeley.edu/) (Crooks et al., 2004). All the multiple sequence alignments are available upon request.

2.2.10 Phylogenetic analysis

Phylogenetic inference was carried out on the combined dataset of 41 full length sequences obtained from databases and the aligned and trimmed to the partial SGL coleopteran datset of 29 sequences. The final amino acid alignment included 70 taxa and was composed of 214 characters. Maximum likelihood estimates of the topology and branch lengths were obtained using PhyML v2.4.4 (Guindon, Gascuel, 2003), via the online server (http://atgc.lirmm.fr/phyml/), with the $WAG + I + \Gamma$ model accounting for site-to-site rate variation using a discrete gamma distribution with four rate categories and invariable sites; the proportion of invariable sites was estimated from the data and support for individual branches was inferred by bootstrap analyses (100 replicates).

Phylogenetic relationships were investigated using Bayesian techniques as implemented in the computer program MrBayes v3.0b4 (Ronquist, Huelsenbeck, 2003). MrBayes estimates posterior probabilities of clade support using Metropolis-coupled Monte Carlo Markov Chain method (MC^3). Five independent runs were conducted using the same settings: starting with a random tree, using the WAG model of amino acid substitution (Whelan, Goldman, 2001) ($WAG + I + \Gamma$); as suggested by analysis of the alignment with ProtTest (v1.2.6) (Abascal et al., 2005), generating trees for 5 million generations with sampling every 100 generations, and with four chains (three with the heating parameter set to 0.2). The first 1,250,000 generations (12,500 trees) were discarded from every run and the remaining trees were concatenated. The remaining 187,500 trees were used to compute the final (consensus) tree, and to determine the posterior probabilities at the different nodes.

2.2.11 Test of functional divergence

We estimated the functional divergence among the three proposed insect groups (i.e., Diptera, beetle group I and beetle group II), functional divergence estimates were also conducted for the subclades Ia and Ib. The ML protein alignment and topology showing the relationship among SGL domain-containing proteins were used as input for the program DIVERGE v1.04 for calculating the coefficient of functional divergence (θ) between pairs of SGL groups (Pollock et al., 2002). The parameter θ is

a maximum likelihood estimator of type 1 functional divergence between two members of a protein family and is based on the detection of altered rates of amino acid replacements for a given position between the two paralogue groups. That is, when an amino acid configuration is very conserved in one of the paralogues and is highly variable or fast evolving in the other, it implies that this site-specific rate difference originated because the paralogues experienced different functional constraints at these residues (Wang, Gu, 2001). A Likelihood Ratio Test (LRT) was then carried out to test the null hypothesis $\theta = 0$ against the alternate hypothesis of $\theta > 0$. If the estimated θ was significantly greater than zero, functional divergence between the paralog pair was highly supported. A Bayesian approach was then followed to statistically predict which sites were likely to be responsible for the type I functional differences found between paralogues (Gu, 1999). For our analyses, those positions with posterior probabilities > 0.9 in the site-specific profile for each pairwise comparison were considered to be potential functional divergence-related amino acid sites.

2.3 Results

2.3.1 Identification and characterisation of members of the SGL protein family

Extensive similarity searches of the GenBank, Pfam and Swiss-Prot databases and genome sequence projects resulted in the identification of 24 insect SGL gene sequences encoding a complete open-reading frame from the eight completed and publicly available insect genomes. Two further SGL sequences described as anterior fat body proteins (AFP's) were identified from *Sarcophaga peregrina* (Diptera: Sarcophagidae) and *Calliphora vicina* (Diptera: Calliphoridae) (Hansen *et al.*, 2003; Hansen *et al.*, 2002; Nakajima, Natori, 2000). Also included in further analysis were the three firefly LRE sequences (Gomi *et al.*, 2002; Gomi, Kajiyama, 2001). All sequences collated from databases are shown in Table 2.3. A total of 29 full length insect SGL sequences retrieved from databases were employed in the subsequent analysis. Although two genes were identified in the *A. mellifera* genome with associated EST sequences the transcripts were significantly different in size and content, AmSMP01 coded for a 1090 bp transcript translating 313 amino acids and AmSMP02 codes for a 596 bp transcript that translates a truncated SMP30 protein of 94 amino acids, hence its exclusion from subsequent analysis. Other sequences identified but excluded were those found in the body louse *Pediculus humanus corporis* where strong identity to other insect SGL genes was identified but they

appeared to be pseudogenes. No complete full length ORF could be predicted for these genes although good homology was evident.

13 SGL paralogues and orthologues were identified from the dipteran genomes of *Drosophila melanogaster*, *D. pseudoobscura*, *Anopheles gambiae* and *Aedes aegypti* genomes. Goto identified a gene homologous to mammalian SMP30 in *D. melanogaster* which was up-regulated at the transcription level after acclimation to 15°C (Goto, 2000). Goto noted this gene was present as a single copy on chromosome 3R. Analysis of the *D. melanogaster* genome confirms the location of this gene (Figure 2.2). A further paralogue DmSGL2 is found on the X chromosome. Two orthologues are found in the *D. pseudoobscura* genome and DpSGL1 and DpSMP02 and extensive microsynteny was evident in the genes flanking these two loci (Figure 2.2).

In *Anopheles gambiae* and *Aedes aegypti* four and five paralogues were identified in these genomes respectively. In *Ae. aegypti* AeSGL3, 4 & 5 were located in a tandem array and in *An. gambiae* AgSGL2 was found adjacent to AgSGL4 on chromosome 3R (Figure 2.2). Unlike *Drosophila* no evidence of microsynteny was found in the genes flanking the *An. gambiae* and *Ae. aegypti* SGL gene orthologues.

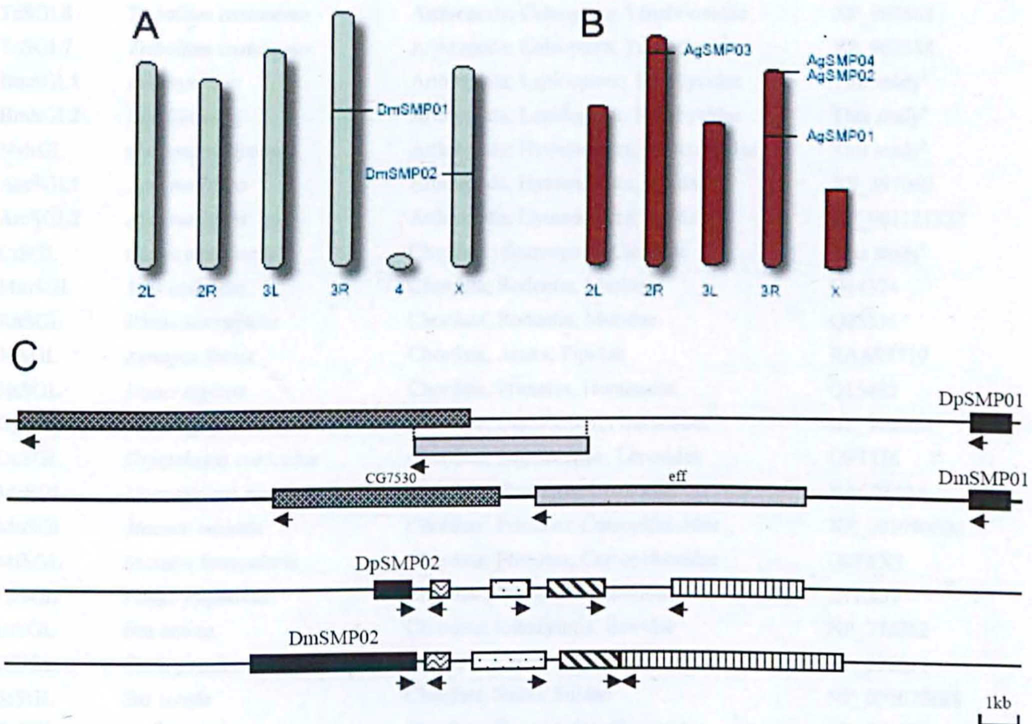


FIGURE 2.2 Microsynteny analyses between *Drosophila* SGL genes. SGL genes (black boxes) are abbreviated to DmSMP and DpSMP. Interspecific neighbouring homologues sharing at least 50% amino acid identity are shown shaded. The *D. melanogaster* neighbouring genes are labelled by the gene code. Gene orientation is indicated by arrows.

Locus	Species	Classification	GenBank Acc. No.
AtGHS	*Agrobacterium tumefaciens*	Proteobacteria; Rhizobiales; Rhizobiaceae	2GHS_A
AeSGL1	*Aedes aegypti*	Arthropoda; Diptera; Culicidae	EAT34066
AeSGL2	*Aedes aegypti*	Arthropoda; Diptera; Culicidae	EAT48186
AeSGL3	*Aedes aegypti*	Arthropoda; Diptera; Culicidae	EAT47890
AeSGL4	*Aedes aegypti*	Arthropoda; Diptera; Culicidae	EAT47888
AeSGL5	*Aedes aegypti*	Arthropoda; Diptera; Culicidae	EAT47891
AgSGL1	*Anopheles gambiae*	Arthropoda; Diptera; Culicidae	EAA14899
AgSGL2	*Anopheles gambiae*	Arthropoda; Diptera; Culicidae	EAA12326
AgSGL3	*Anopheles gambiae*	Arthropoda; Diptera; Culicidae	EAA01708
AgSGL4	*Anopheles gambiae*	Arthropoda; Diptera; Culicidae	EAA12283
DmSGL1	*Drosophila melanogaster*	Arthropoda; Diptera; Drosophilidae	AAF55095
DmSGL2	*Drosophila melanogaster*	Arthropoda; Diptera; Drosophilidae	AAF48128
DpSGL1	*Drosophila pseudoobscura*	Arthropoda; Diptera; Drosophilidae	EAL27359
DpSGL2	*Drosophila pseudoobscura*	Arthropoda; Diptera; Drosophilidae	EAL31737
CvSGL	*Calliphora vicina*	Arthropoda; Diptera; Calliphoridae	AAK26174
SpSGL	*Sarcophaga peregrina*	Arthropoda; Diptera; Sarcophagidae	BAA99282
LcLRE	*Luciola cruciata*	Arthropoda; Coleoptera; Lampyridae	BAB85479
LlLRE	*Luciola lateralis*	Arthropoda; Coleoptera; Lampyridae	BAB85478
PpLRE	*Photinus pyralis*	Arthropoda; Coleoptera; Lampyridae	AAO66431
TcSGL1	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_967986
TcSGL2	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_967905
TcSGL3	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_967825
TcSGL4	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_966689
TcSGL5	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_967747
TcSGL6	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_967668
TcSGL7	*Tribolium castaneum*	Arthropoda; Coleoptera; Tenebrionidae	XP_967588
BmSGL1	*Bombyx mori*	Arthropoda; Lepidoptera; Bombycidae	This study†
BmSGL2	*Bombyx mori*	Arthropoda; Lepidoptera; Bombycidae	This study†
NvSGL	*Nasonia vitripennis*	Arthropoda; Hymenoptera; Pteromalidae	This study†
AmSGL1	*Apis mellifera*	Arthropoda; Hymenoptera; Apidae	XP_397069
AmSGL2	*Apis mellifera*	Arthropoda; Hymenoptera; Apidae	XP_001121327
CiSGL	*Ciona intestinalis*	Chordata ; Enterogona; Cionidae	This study†
MmSGL	*Mus musculus*	Chordata; Rodentia; Muridae	Q64374
RnSGL	*Rattus norvegicus*	Chordata; Rodentia; Muridae	Q03336
XlSGL	*Xenopus laevis*	Chordata; Anura; Pipidae	BAA93719
HsSGL	*Homo sapiens*	Chordata; Primates; Hominidae	Q15493
GgSGL	*Gallus gallus*	Chordata; Galliformes; Phasianidae	NP_990060
OcSGL	*Oryctolagus cuniculus*	Chordata; Lagomorpha; Leporidae	Q9TTJ6
MaSGL	*Mesocricetus auratus*	Chordata; Rodentia; Cricetidae	BAC76714
MuSGL	*Macaca mulatta*	Chordata; Primates; Cercopithecidae	XP_001090600
MfSGL	*Macaca fascicularis*	Chordata; Primates; Cercopithecidae	Q2PFX5
PpSGL	*Pongo pygmaeus*	Chordata; Primates; Hominidae	Q5R837
BtSGL	*Bos taurus*	Chordata; Ruminantia; Bovidae	NP_776382
CfSGL	*Canis familiaris*	Chordata; Carnivora; Canidae	XP_538011
SsSGL	*Sus scrofa*	Chordata; Suina; Suidae	NP_001070688
DrSGL	*Danio rerio*	Chordata; Cyprinoidea; Cyprinidae	NP_991309
TnSGL	*Tetraodon nigroviridis*	Chordata; Tetraodontiformes; Tetraodontidae	CAG01328
OaSGL	*Ornithorhynchus anatinus*	Chordata; Monotremata; Ornithorhynchidae	XP_001505641
MdSGL	*Monodelphis domestica*	Chordata; Didelphimorphia; Didelphidae	XP_001365514

† Only DNA scaffold sequences available for these genomes at this time.

TABLE 2.3 SGL sequences retrieved from public databases.

There was no evidence in the mosquito genomes of microsyteny between any of the genes flanking SGL loci. However the four immediate genes flanking AgSMP01 were found to be conserved in both order and orientation in the genome of *A. aegypti* but in no close proximity to an SGL locus.

Data mining of the only beetle genome currently sequenced, *Tribolium castaneum*, revealed seven genes (TcSGL1–7) which were found in parallel all within a 13 kb region of the plus strand of chromosome 5 (Figure 2.3). Five genes correlated with identified transcripts of which TcSMP04 was found to have two different isoforms. TcSGL3 was found to be predicted to have a complete open reading frame coded for by a hypothetical transcript of 930 bp (GenBank acc no. XM_962732). However, no actual transcripts are associated with this locus and examination of the first predicted intron boundaries showed no evidence for a canonical GT/GA splice site. But it was possible using the first intron position of other paralogues to determine a complete open reading frame that conformed to the canonical positions of established splice sites (Figure 2.4). This modification provided a more robust amino acid alignment with other paralogues as well as a conserved intron position. This locus was still tentatively ascribed pseudogene status based upon a lack of functional transcript. The intron exon sites and open reading frame for TcSGL5 was also found to be incorrectly predicted. At the second intron site the position predicted was found to lack a GT position at the intron/exon boundary and upon alignment lacked thirteen residues present in all other predicted protein sequences. Using the second intron position from TcSMP03 as a template, a more parsimonious ORF was predicted (Figure 2.4). Of the seven paralogues on chromosome 5 two are regarded to be redundant due to a lack of data from EST libraries.

Three further SGL paralogues were identified on chromosome 9 (Figure 2.3). TcSGL8 was predicted to code for 189 amino acids by NCBI, which had greatest sequence identity to the C terminus of TcSGL7. The truncation of this sequence is a result of the computer prediction criteria. The open reading frame was predicted to begin with an internal methionine residue which is conserved in other paralogues.

TcSGL9 and TcSGL10 were predicted to code for 309 and 346 amino acids respectively. Both sequences showed more identity to themselves than any other locus. Like TcSMP08 no mRNA transcripts were found to be associated with these loci. Due to the doubtful functional status of these genes and their divergence from other SGL genes they have been excluded from further phylogenetic studies. Most sequence identity occurs between TcSGL2 and TcSGL3 as well as between TcSGL5 and TcSGL6. TcSGL1 and TcSGL7 are the most divergent of all these protein sequences with the least number of introns in their genes.

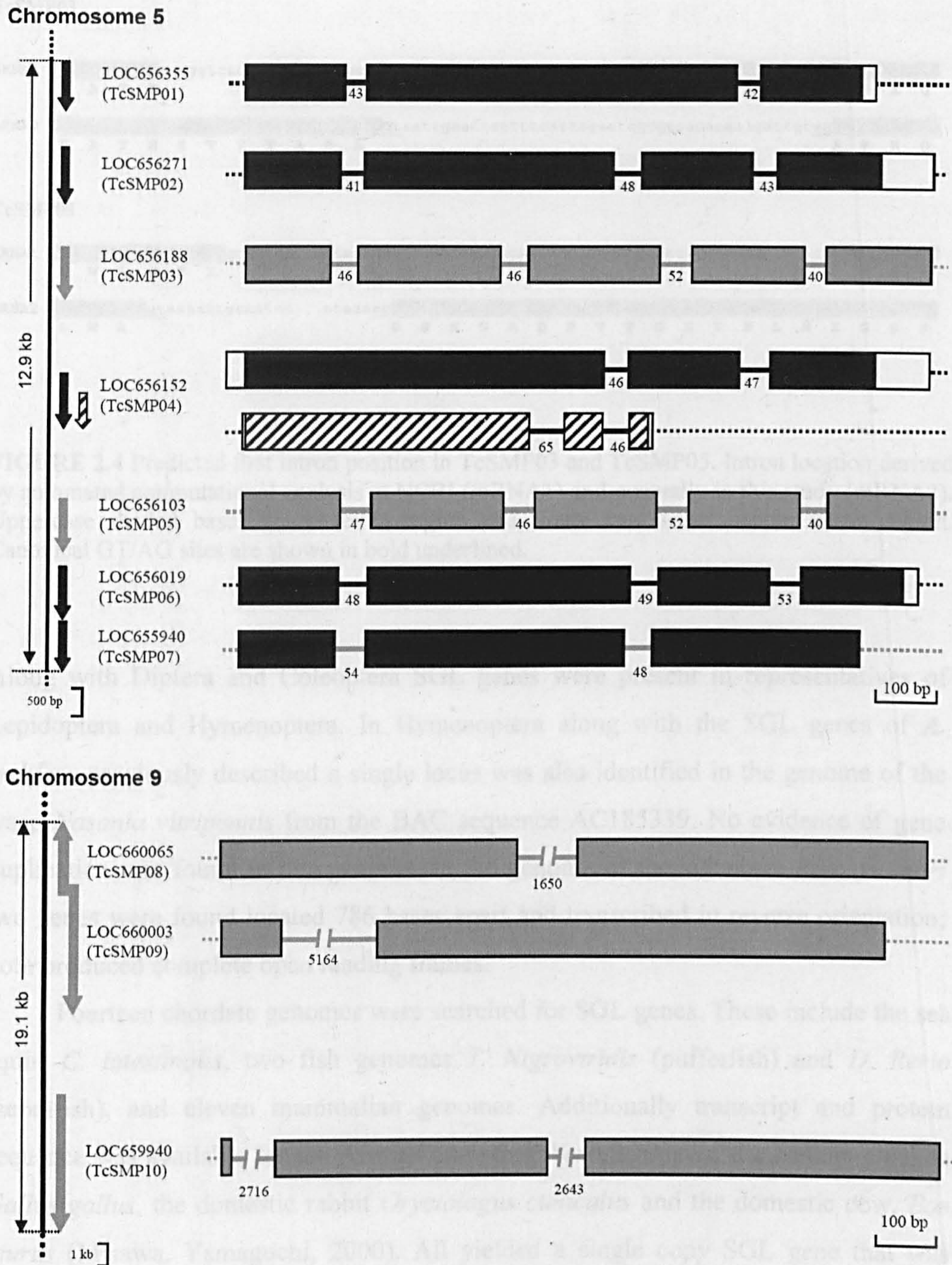


FIGURE 2.3 Chromosomal location and gene architecture of *Tribolium castaneum* SGL paralogues. Arrow direction denotes 5'-3' orientation of genes on the chromosome. Black arrows and genes are associated with full length transcripts whereas grey arrows and genes have no associated EST sequence. Solid boxes represent exons; solid lines define intron positions below which are shown the intron size. White boxes represent untranslated regions of associated EST sequences. Diagonal shaded boxes describe alternative transcription products.

TcSMP03

```
mRNA1 CCTGCGACGAAcaaggtcaccacagcatcggccggtaattgaattattttcattagactcGCGTCACAAAATGATTGTAGCACCAAAACAG
        P  A  T  K                                                  R  H  K  M  I  V  A  P  K  Q

mRNA2 CCTGCGACGAACAAGGTCACCACAGCATCGGCCGgtaattgaattattttcattagactcgcgtcacaaaatgattgtagCACCAAAACAG
        P  A  T  N  K  V  T  T  A  S  A                                                  A  P  K  Q
```

TcSMP05

```
mRNA1 CTATGGGCCGGTAAAATTgcaatca...ataaaagggtcaatgggggcacaaccagtcccgggacacgtggagctaaacaagGGCTCATTT
        L  W  A  G  K  I                                                                  G  S  F

mRNA2 CTATGGGCCGgtaaaattgcaatca...ataaaagGGTCAATGGGGGCACAACCAGTCCCGGGACACGTGGAGCTAAACAAGGGCTCATTT
        L  W  A                          G  S  M  G  A  Q  P  V  P  G  H  V  E  L  N  K  G  S  F
```

FIGURE 2.4 Predicted first intron position in TcSMP03 and TcSMP05. Intron location derived by automated computational analysis at NCBI (mRNA1) and manually in this study (mRNA2). Uppercase shaded bases denote exon region and lower case bases denote intron region. Canonical GT/AG sites are shown in bold underlined.

Along with Diptera and Coleoptera SGL genes were present in representatives of Lepidoptera and Hymenoptera. In Hymenoptera along with the SGL genes of *A. melifera* previously described a single locus was also identified in the genome of the wasp *Nasonia vitripennis* from the BAC sequence AC185339. No evidence of gene duplication was found in this genome. In the genome of the silkworm *Bombyx mori* two genes were found located 786 bases apart and transcribed in reverse orientation; both produced complete open reading frames.

Fourteen chordate genomes were searched for SGL genes. These include the sea squirt *C. intestinalis*, two fish genomes *T. Nigroviridis* (pufferfish) and *D. Rerio* (zebrafish), and eleven mammalian genomes. Additionally transcript and protein sequence was available for the African claw frog *Xenopus laevis,* the bantam chicken *Gallus gallus*, the domestic rabbit *Oryctolagus cuniculus* and the domestic cow, *Bos taurus* (Misawa, Yamaguchi, 2000). All yielded a single copy SGL gene that was highly conserved in sequence and gene architecture.

2.3.2 *Photinus pyralis* LRE gene

The 927 bp luciferin-regenerating gene cDNA sequence characterised by Gomi and Kajiyama (2001) from the firefly *Photinus pyralis* was shown to be subdivided into five almost equally sized exons. Sequencing of the 1985 bp gene (GenBank acc. no. AY197610) revealed the presence of five exons ranging in size from 167 bp to 211 bp divided by four introns where the exon-intron and intron-exon splice sites conform to

the canonical sequences GT-AG (Breathnach, Chambon, 1981) (Appendix I). Three of the introns were ≤ 54 bp in length while one is much larger, 904 bp. Only intron 4 was found in phase 0 the rest occurring in phase 1 (Appendix I). Eleven point mutations were identified between the LRE gene and the mRNA sequence (GenBank acc. no. AB062786) suggesting individual or population variation at this locus in *P. pyralis*. Four out of the five point mutations in the last two exons result in non-synonymous changes.

2.3.3 Intron/exon site conservation

In mammals all SGL sequences were found to be composed of six exons divided by five introns. Although the size of the introns varied dramatically all intron exon boundaries were found in the same location. This was also evident in the two fish species *D. rerio* and *T. nigroviridis*. Deviation from this was evident in the genome of *C. intestinalis*. The fifth intron was missing in the CiSGL gene and intron III was found at a position fifteen residues prior to the site conserved in insects, mammals and fish. Four out of the five intron positions were found to be, on the whole, highly conserved across the metazoan taxa with the most variation at the fourth intron site (Figure 2.5). Predominantly, intron/exon boundaries were found to be in the same phase across phyla with extensive intron loss was evident in the Diptera.

2.3.4 Amplification of GSL domain using degenerate primers from twenty one beetle taxa

Three sets of degenerate primer were employed to amplify GSL genes from beetle DNA. PCR amplification resulted in a range of amplicon sizes and in some cases multiple amplicons. A minimum size of gene was predicted and anything under 600 bp was disregarded. Primer set 1 produced, in many cases, a truncated product where the forward primer had annealed to a region of exon 2. These products were discarded as they were below 600 bp. Products greater than 600 bp were excised, cloned and at least four clones sequenced. Due to the length variation in primer sequence the final partial gene alignment was trimmed to produce a consensus alignment from amino acid position 26 through to 216 in the *Photinus pyralis* GenBank sequence (accession number AAO66431).

```
                                                                      I
DmSGL 1    MSY--KVE----AVPDSYAALGEGPHWDVDRQSLYYVDLESAGINRYDFKQNKVYRAKIEGEIFASFILPVENKPQEFAV
DmSGL 2    MSY--KVE----PLPDSYAGLGEGPHWDVARQSLYYVDLEAGSLLRYDYAQNKVYKTKIEGETLAGFVLPVEGRPQEFAV
DpSGL 1    MSY--KVE----ALPDSYAGLGEGPHWDVDRQSLYYVDLEVGFIHRYDFKQNKVYKAQIEGETFASFILPIENRPQEFAV
DpSGL 2    MSY--KVE----PLPDSYAGLGEGPHWDVATQSLYYVDLEAGKLLRYDFKQNKVYKTQIEGESFAAFVLPIEGKPQEFAV
AgSGL 1    MAS-YKVE----QLPSPLSVLGEGPHWDVERQSLYYNDIYGGSIHRYDYAENKTYNATIDGFPVISFIIPVKGNDRHFII
AgSGL 2    MAESYRVA----AIP-PYTELGEGPHWDIARQSLYYVSLTDAWIHRWDYREGKVYSASIDGIRFATFIVPVKGRSDCFVI
AgSGL 3    MAS-SDSSYQVVKLPSPRTKLGEGPVWDIDSQSLYYVDINTPAVLRYDYGENRTYSAKLAGANSISFIALVVGQPEHFVV
AgSGL 4    MAN-VQVD----VLPGPFLQLGEAPHWDGESQSLYYVCILSSTLHRYDWKENKTYSAKIEGSTYASFVIPVKGRKGEFVV
AeSGL 1    MDS-YKVE----QVPSPLNVLGEGPHWSIEQQCLYYNDIYGGTIHRYDYAENKTYTAKLDGYPVISFITPVKGRKTEFII
AeSGL 2    MAN-YKVE----EISAPKLDLGEGPHWDGKSQSLYFVDMFKAGVHRWDYHKNKTFSASVEGCTWVSFIIPVKGRSNEFVV
AeSGL 3    MAAEYSVK----QLPSPLSQLGEGPVWDVDTQSLYYVDINGAAILRYDYAENKTYSAKIDGVDPISPIILVQGKPGQYVL
AeSGL 4    MAAEHTVH----QLPSPLSIIGAKTVWDVDSQSLFYVDINVAAIRRYDYAENKTYSCTIDGVNPIAPVILVQGKPDHFVV
AeSGL 5    MKH--------------FQKSGEGPVWDIDRQYLYFVDIHECAILRYDPVQNRTYKAIIDGVPFVSAIVLIRDKPDHFVL
TcSGL 1    MAP--VIEI-----ISGRVTLGEGPHWDAPTQTLYYVDIFGQAIHKYVPSTNTHAKVVIEG-GPVTMVIPVEGTTDKFLV
TcSGL 2    M-H--TIER-----VTEGFSLGEGPHWDASTQSLYFVDVFGQSIVKYAPTTKKVTKASVAP-KTASFIIPVEGAKDQFVI
TcSGL 3    MSP--KIER-----LTESFLLGEGPHWDVATQSLYFLDCIRQNLVKYNPATNKVTTASAAP-KQPTFIIPVQGKSGQFII
TcSGL 4    MV---KIER-----LTDNIHLGEAPHWDAESESLFYVDILGMTIYKYTPATKKCTKASVGV-NLVSFIIPIEGEKNQFVV
TcSGL 5    MTL--TTRR-----VVESVELGEGPHWDPETQSLYFVDIYGKAIHKYVPATKKHTKAVIGI-NHVSLIVPVQGQKNKFLI
TcSGL 6    MAP--TIER-----IVDSVELGEGPHWDAATQSLYFVDIFGKTIHKYVPVTKKHTKAVIGT-NHVSLIIPVEGEKNKFLI
TcSGL 7    MVRITQIG--------DNLEVGTRIHWDEQTQNLYYVDVPTSTIYRYRPSTDEITQAQVGN-EPLAFAFPVDGKTDFFIA
PpLRE 1    MGP--VVEK---IAELGKYTVGEGPHWDHETQTLYFVDTVEKTFHKYVPSQKKYTFCKVD--KLVSFIIPLAGSPGRFVV
BmSGL 1    MAP--QLQ-----AVTEPVWLGEGPHWDHNEQALYFVSIFDETIHKYVPETGKHTRSKLD--GKPTFIIPIEGKKHHFVV
BmSGL 2    MSV--RIE-----KITEPLTLGEGPHWDERQQALYFVSIQDKTIHKYVPTTEKHTKTSLD--GRVGFILPVEGTTDQFVV
AmSGL 1    MSE-ITLE----PLVGP-YDLGEGPHWDPISQKLYYVDIYAQKVFRFDPASGIVTSVFIE-NGPVGFVVPLEGCTDKFVA
NvSGL      MS--VKVE-----KVAPAIELGEGPHWDVKTNKLYYVDINAQKILRLDPVTGNITSAYLK-DGPVGVVIPVEGTTDKLVV
MmSGL      MSS-IKVE----CVLRENYRCGESPVWEEASQSLLFVDIPSKIICRWDTVSNQVQRVAVD--APVSSVALRQ--LGGYVA
DrSGL      MSS-IKVE----CVIKEKNEVGESPVWEEKDSSLLYVDITGQKVSRWSSLTKQIESMNTE--KLVGCVVPRQ--AGGYVI
CiSGL      MSA-VKVE----LVHNYDCQLGEGPHWDDQTQTLLFVDIDNSAIHRWNPATKQTKTTIVKDSSIGAVVPRKSGDLMIAAG

                                                                 II
DmSGL 1    GCGLRTVIVQWDGVSA-VAKVTRTLFEVQP---DLKENRLNDAKTDPNGRFYGGTMAD--SGDIFT-QWKGELYSWQA--
DmSGL 2    GCGRRVVIVNWDGVSP-SAKVVRTLFEVQP---LMEKNRLNDAKVDPRGRFFGGTMRY--IGDEFE-FRHGELYRWEA--
DpSGL 1    GCNRRCVVVQWDGVAT-VAKVLRVLFEVQP---GLEDNRINDAKTDPRGRFVGGTMCC--SGDIFT-QWKGELYTWQA--
DpSGL 2    GCARRVVIVNWDGVSP-VAKVVRTLFEVQP---EMDKNRLNDAKADPRGRFFGGTMRY--IGDEFE-FRHGELYKWEA--
AgSGL 1    GTDRKVTLVDWDGRSE-KATFVRTVGEVEP---TMEDNRFNDAKVDSKGRFYGGTMRLEAKGDIFE-MRLGTFYRYDAK-
AgSGL 2    GDTIRLLVIRWDGKAS-KATIVRELACLGP---DHVDNRFNDGKVDPWGRLYVGSMLNESAGNPFE-KATGALWRYCDR-
AgSGL 3    GENNRVTLISWDGRSE-AASHVRVLADLGP---SQSHVRFNDGKIDPAGRLYAGTMQLESLGDLFA-QKEGQLFRYT---
AgSGL 4    GSGTRLLLVSWDGCSE-TATIVKVLTDLGE---EEADHRFNDGKVDGQGRLYAGTMLAEDSRNHFE-MDDGKLYRFDAG-
AeSGL 1    GTDKKITLIHWDGVSE-QAKFLKTIGEVEF---DLPNNRFNDAKTDFKGRFYGGTMRLEAKGDIFE-VRLGSFYRYDAK-
AeSGL 2    GDGTRLVLITWNGTSD-KAKIVKVIADLGE---AGQSNRFNDGKADSKGRLFTGTMKAEPFGNPFE-INSGKFFRFDAR-
AeSGL 3    GTGNKIVLVNWDGRSE-RGTLVKTLYDLGE---SEKHVRFNDGKVDPQGRLYAGTMQLETLGDVMQ-QKEGKLFRFDGKA
AeSGL 4    GSGNKLLLVSWDGRSE-KGTLVKTVYDLGE---SEKHVRFNEGKVDPKGRLYAGTMQLESLGDIFQ-QKEGKLYRFDGKV
AeSGL 5    GTGTKLSLVQWDGLSE-KASLVQTVADLGD---SESHVRFNDAKVDSQGRLYIGTMLRETVGEPLGNSMVGKLYRFDGRI
TcSGL 1    SIGRKLVIVTWDGTSD-KISNSELLVEVENKS-GYFNNRFNDGKADPTGRLWAGTMGPEPEIGKLE-KEKGALYTLVGK-
TcSGL 2    SLNNELVIILWDGESD-SAKIVEKLASVD--------NKFNDAKCDSTGRLWAGGHTLNESDFMNS-GPLGHLFSLDSN-
TcSGL 3    SLDKELAIINWDGQSD-KFSIVRKLCVADGGP-GTAQNKFNDGKCDSSGRLWAGTLNIDKEDEDKT-LPLGTLYSFDSK-
TcSGL 4    SLGREIVRIFWNSETE-DVKVVEKLAEVEDSP-EFADNRFNDGKCDPSGRLWAGTLNTKAEKPLVY-PPKGTLYSLDAE-
TcSGL 5    SIGRDLNIVTWDGESE-TVSGMEKIYEIDNTP-DTLQNRFNDGKCDALGRLWAGSMGAQPVPGHVE-LNKGSFISLESG-
TcSGL 6    TIGRQLVTVKWDGSSE-KVSEITKIGEVDDDP-ETLDNRFNDGKCDPTGRLWAGTMGGEPINGHVK-PNKGGLFSLGPN-
TcSGL 7    GLGRKIVLLKWDGESE-SVCSCTTIAEVDREP-HLAKNRLNGAKVDPYGRLWAGTMGAQDANGQTI-PKQGSLYSLT-N-
PpLRE 1    SLEREIAILTWDGVSA-APTSIE--AIVNVEP-HIKNNRLNDGKADPLGNLWTGTMAIDAGLPVG--PVTGSLYHLGAD-
BmSGL 1    GLDRLVVEIQWTGED-QTARLVRTVAEVDQ---DNPNNRFNDAKADPRGRLFAGTMGHEYEPGKFD-LKKGSLYRIDPD-
BmSGL 2    GVERKFLFIQWDGEDGSKVAVLKELGEVDK---DRPNNRINDGKADPRGRLFAGTMGHEDPPGNFE-RNKASLYKLDSAK
AmSGL 1    GCGIDFVLFSWNSEKSLENCTAQVLVSADS---DRIETRLNDGKVDSSGRLWAGTMGHEKN-GIFP-PNIGSLYSIGND-
NvSGL      GCGRDVVLVTWDGENDTTSPPVKKLLSLDT---DRTDTRINDGKCDPAGRFWLGTMALEVN-DAIE-PDRGTFYAVDQD-
MmSMP      TIGTKFCALNWE------NQSVFVLAMVDE---DKKNNRFNDGKVDPAGRYFAGTMAEETAPAVLE-RHQGSLYSLFPD-
DrSGL      AEGTRFAFVDWV------KRSITAVAEVNE----KPNTRFNDGKVDPAGRFFAGTMSMDMKPDVVD---AALYNLQPD-
CiSGL      HRFASWNENTGE-------SETFKEVNLEF-----PTSRFNDGKCDPAGRFWAGTMGREKVAASPD RLQGKLYCLDID-
```

FIGURE 2.5 Intron/exon boundaries in insect SGL orthologues and paralogues. Intron/exon sites are shown shaded. Protein names are abbreviated to SGL or LRE with suffixes: (Ae) *Aedes aegypti*; (Ag) – *Anopheles gambiae*; (Am) – *Apis mellifera*; (Bm) – *Bombyx mori*; (Dm) – *Drosophila melanogaster*; (Dp) – *Drosophila pseudoobscura*; (Nv) - *Nasonia vitripennis*; (Pp) – *Photinus pyralis*; (Sp) - *Sarcophaga peregrina*; (Tc) – *Tribolium castaneum*. The conserved chordate intron/exon structure are represented by the mouse SGL sequence (MmSGL) and the zebrafish *D. rerio* (DrSGL). Also shown is the sequence for the sea squirt *C. intestinalis* (CiSGL).

```
                              16                                  IIIi                 IIIii
DmSGL 1   GGQPNAIRSKVGISNGLAWDVKAK--KFYFIDTNNHEVLAYDYNQSTGAV--SNPKVIFDLRKIRPEGPLFPDGMTVDTD
DmSGL 2   GGQVSVIKGDVGISNGLAWDEKAK--KFYYIDTTDYEVKSYDYDFETGVA--SNPKVIFNLRKNSPKDHLLPDGLTIDTE
DpSGL 1   GGQVTKIRGEVGISNGLAWDVKAK--KFYFIDTNTHEVVAYDYNVDTGAV--ANPKVVFDLRKIRPNTPLYPDGMTIDTE
DpSGL 2   GGQVSIVKGDVGISNGLAWDEKAK--KFYYIDTTDYEVKSYDYDFETGVS--SNPKVIFNLRKTSPKDHLLPDGLTIDTE
AgSGL 1   QGKFVTLKEKIGVSNGLCWNEAGN--LFYYIDSCDLDVKEYQVDAN-GDI--SGERVVIDFRVNGERPPFVPDGMTIDAN
AgSGL 2   TGQMVEQDRNIYISNGLAWNRATN--KFYFVDSGANHIKEYDIDLD-GNL--INPRIWYDFKPDGADPGYFGDGMTIDSE
AgSGL 3   NGTMVVQKRNVSISNGLTWDEPFNPLRMYYIDSAALDVKAFDVDAN-GDL--KNETVFYDLRVNGAHPGYVPDGMTSDAE
AgSGL 4   RGQMVPLKSKVHISNGLTWSARTG--KFYYIDSFAFDIKEYTVDAE-GNLGKCGETVLIKLKDDEAATEFIADGMTSDAD
AeSGL 1   QKKFVVLKKNIGVSNGLCWNETGN--LLYYIDSCDLDVKEYHVDEN-GDL--SNERKVIDFTVDGARPPFVPDGMTIDTE
AeSGL 2   GVKFVEQFDKVFISNGLAWNDKTK--KFYYADTGAYDVKVFDFDDN-GDL--SNRKTFYDMTANTTDPKEAPDGMTIDTD
AeSGL 3   GGEFHLLKAGVSISNGLTWPEKSN--KFYYIDSVASDIKEYDVLEN-GDL--QNETVFYDLRVDGKSPGYVADGMTHDND
AeSGL 4   GGEFHMMKDGISISNGLAWIESTN--KLYYIDTAALDIKEFDVLEN-GDL--QNETVMYDWRVNGEGPGYFGDGMTNDAD
AeSGL 5   GGQLIVQRSGVGISNGITWNEKLG--KFYYIDSLALNIKEYDVAAN-GDL--YNEQVLLSFVVNGTYPGFYPDGMTCDSE
TcSGL 1   -HQVKTHLTKVSIANGLAWDLELK--KMYYIDSPRRTVDEYDNNLEKGEI--CNRKVVFNLDVHD--IPGVPDGMTMDTD
TcSGL 2   -KQLKKCLDKIRVANGLAFNDKVK--KMYYIDSLAGTVDWFDFDVNSGTI--SNRQVLFTLKKHN--VTGIADGMTIDTD
TcSGL 3   -RGLKGHVNQVRLTNGIAFNDQTK--KMFYIDTLKGTVDQFDFDVTNGEI--SNRKVWFTLTKNN--ISGKPDGMTIDTD
TcSGL 4   -RNVKNHVSSLRISNGMAFNPKLK--AMYFIDSAKGTVDHYNFDMAAGTI--SNPRPIFTLSNHG--IKGIPDGMTIDID
TcSGL 5   -RKAKTHLTKVGISNGLAWSLDNK--YLYYIDSQKNTLDRYDFNHEKGTI--SNGESIFSLKKAN--LSGILDGMTIDTD
TcSGL 6   -QQIRKHLSNVSISNGLAWSKDLK--KMYYIDSPKRTIDEYEVDMKYGTL--SNGRPIFTLEKHN--IQGFPDGMAIDTD
TcSGL 7   -GRLKREFKEVGISNGIAFDLDAN--KMYYVDTLIPAISAFDYDGESGEI--SNKTTVFKLECSC--INGLPDGLTIDTD
PpLRE 1   -KKVKMHESNIAIANGLAWSNDLK--KMYYIDSGKRRVDEYDYDASTLSI--SNQRPLFTFEKHE--VPGYPDGQTIDAE
BmSGL 1   -GSVHRLESNIDISNGLCWDLQRS--AFYFADSFEYTIRRYDYDVETGSI--ANAKTVFKYSDHG--LEGIVDGMTIDTD
BmSGL 2   DGKLEKIIETVSLSNGLAWDLKEK--AFYYTDSMQFSITKFDYDVDTGEI--SNPRNIFDFKQRG--LQGIPDGTTIDTD
AmSGL 1   -FMLKKQISPVSISNGLAWNPNND--IFYYIDSLSYQIVAYNYNSQTGII--SNKKIVFDFLKNN--IPGLPDGMTIDTN
NvSGL     -LNLRKIISPVSISNGLAWSLQND--VMYYIDSMSYQIWAYDYNHKDGAI--NNKRVIFDLKKNN--INGLPDGMTIDAD
MmSGL     -HSVKKYFDQVDISNGLDWSLDHK--IFYYIDSLSYTVDAFDYDLQTGQI--SNRRIVYKMEKDE----QIPDGMCIDAE
DrSGL     -HSVVRHFDQVHLSNGLDWSLDHR--VFYYIDSLAFMVEAFDYDIQTGGL--SNRRTVYKMEKDE----GIPDGMCIDTE
CiSGL     -HSVKTKVYPVDISNGLSWTSNT----MYYCDSLKLTIDAYDYDVTTGEI--KNMREVVKFDREK---EGVPDGHCIDTD

                        IVi   IVii            IVIii                                        V
DmSGL 1   GNIYVATFNGGTVFKVNPSTGKILLEIKI--PTTQITSVAFGGPNLDILYVTTANK---FDQPKP---------AGTTFQV
DmSGL 2   GNLYVATFNGATIYKVNPNTGKILLEIKF--PTKQITSAAFGGPNLDILYVTTAAK---FDQPAP---------AGTTYKV
DpSGL 1   GNIYVATFNGGSVFKVNPSSGKILLEIKI--PTTQITSVAFGGPQLDILFVTTANK---FDQPVP---------AGTTYKV
DpSGL 2   GNLYVATFNGATIFKVNPNTGKVLLEIKF--PTKQITSAAFGGPNLDILYVTTAAK---FDQPAP---------AGTTYKV
AgSGL 1   GSLYVATFGGSTVYKVNSKTGKVELEIKL--PCEQVTSAAFGGPNLDILYVTTAAKEFKSPQPAP---------AGALFAV
AgSGL 2   GNLYVACFNGYKVVKISPE-KKILAEYKV--PAKQVTSASFGGPKLDDLFVTTAAKNLTGPQEEP---------AGAVFKI
AgSGL 3   GNLYVATWGGSKVMKIDKGSQKLVLEIKI--PAEQVTSVAFGGPQMDELFVTTSSN---GDKPAP---------AGELFKV
AgSGL 4   GNLYVAVFAGSKIIKINPETAKVVQEIPL--PVAQVTSVAFGGPNLDVLFVTTAAKELSVPQEPP---------AGAVFKI
AeSGL 1   GFLYVATFGGSTVFKINPRNGKIELEIKL--PCEQVTSAAFGGPNLDVLFVTTAAKEFKTPQPPP---------AGALFKI
AeSGL 2   GNIFLAVFHGSKVLKISPS-GQLLQEIMV--PAKQVTSVAFGGPNLDELFVTTAGQPFLEPQGPQ---------AGATFKV
AeSGL 3   GNLYVATWGGSKLLKINPKSKKIEQEIAI--PAKQVTSAAFGGPQMDELFVTTAHT---GNQDPP---------AGALFKV
AeSGL 4   GNLYVATWGGSKVQKINPRTKKVELEIQI--PAKFVTTLAFGGPQLDELFVATAHT---DTQDPP---------AGALFKV
AeSGL 5   GNLYVATWFGSRVLKINAKEKRVEQEMWI--PTPQVTSVSFGGPCLDQLYVTTAKTNAFDNDLPPGAVLSRPPIAGALFKL
TcSGL 1   GNLWVAVFDGGCLLHVNPRTSE-LLNTINF-PARQITSAAFGGPNLDELYVTSAQLVVKGATQPH--------PAGALFKV
TcSGL 2   GNLWVAVFGGNRVLKIDGNKSETLLDTINM-PAEQVTSVAFGGRNLDELFVTTGRFETDEKKLPA--------PVNGATYRV
TcSGL 3   GNLWVAVFMGSRVIKIDGHTAETLLDTVEI-PAQQVTSVAFGGQNLDELYVTTASFEYKGRTPPP-------PVNGALYRI
TcSGL 4   GNLWVAIFNGGKIIRIDPRRPETLLDAIEM-PVRQVTSLAFGGAYLDELYVTTGAYKIDDQDLPP-------PDNGALYKI
TcSGL 5   GNLWIANFGGGRVLKVDPRKPQTLLQIIQM-PAHQVTSVAFGGPNLDELYVTTARLTTDGVAYPP-------PDNGATYVI
TcSGL 6   GNLWVAVFNGYRVIKIDPRKPETLLQTIQL-PAKQVTSVAFGGPDLDELYVTSAAFTVDGVELPP-------PDHGATFRI
TcSGL 7   GNLWLAVIFGSLVLKIDPRTGQ-LLKKIQM-PTPQITAMTWGNKNCDVLYVTSAKMAVNGKVPEK--------PAGCTFKL
PpLRE01   GNLWVAVFQGQRIIKISTEQPEVLLDTVKI-PDPQVTSVAFGGPNLDELYVTSAGLQLDDSSFDK------SLVNGHVYRV
BmSGL 1   GNLWVANFDGRQI---DPRGGKLLQKIPI--PALQVTSVTFGGPDLDQLYVTSASMNRGEEQLPP---------CGSTFRL
BmSGL 2   GNLWVAVFGGSCVLKINPKNGEILQKLAI--PAEQVTSATFGGPNLDILFVTSACCNVGKEQLPP---------SGATFMV
AmSGL 1   GNLWVAVYGGGGILNINPKTGELLRFVKIN-NAKNITSVAFGGPNLDILYVTSARTGLNENQLKE------QLHAGYLFAI
NvSGL     GNLWVALFNGGAVIQIDPKNGKLLRKVEL--PVDRITSVAFGDPQFDTLYVTTAHVGMTAEEKKS------KPNSGSLYAI
MmSGL     GKLWVACYNGGRVIRLDPETGKRLQTVKL--PVDKTTSCCFGGKDYSEMYVTCARDGLNAEGLLR------QPDAGNIFKI
DrSGL     GKLWVACFNGGRVLRIDPQTGKRLQTVKL--PAERITSCCFGGKDYSDLYITSAYIGMDAEALAK------QPEAGCTFKV
CiSGL     GNLWVAMFFTGQVIKVDPRTGEKLQYVKVSDIALKTTSVCFGGPNLDEMYVTSARNKPLIEPTEN-------GISGALFKA

DmSGL 1   TGLNA-KGYAGVNLKI----
DmSGL 2   TGLNA-TGYPGVNLKV----
DpSGL 1   TGLNA-KGLPGVPLKI----
DpSGL 2   TGLNA-TGYPGAYLK-----
AgSGL 1   TGLGV-KGTPMYPVDLS---
AgSGL 2   SGIGA-KGLAMNEVVLKD--
AgSGL 3   TGLGV-KGKPMHKMVL----
AgSGL 4   TGLDS-RGAPMNEIALP---
AeSGL 1   TGLNA-KGTPMYSVDLS---
AeSGL 2   SGLGV-RGFPMTDVDL----
AeSGL 3   TGLGA-RGKPMDKMILRD--
AeSGL 4   TGLGV-RGLPMTKMVLKD--
AeSGL 5   DGFFV-PEYFWKWCGAETGL
TcSGL 1   TGLGV-KGYPGVKVKIP---
TcSGL 2   TGTGA-KGLPGVSFKLN---
TcSGL 3   TGTGS-KGLPANNFKLD---
TcSGL 4   TGLGT-RGLPAAKFRLVNM-
TcSGL 5   TGTEA-KGLPGIGFKL----
TcSGL 6   KGIGA-KGYPGVNVKL----
TcSGL 7   ENLGHSKGLPGDRYKM----
PpLRE 1   TGLGV-KGFAGVKVKL----
BmSGL 1   DGLEV-KGHPNVNVRLQ---
BmSGL 2   TGLGV-KGLPNVSVRL----
AmSGL 1   KGLGV-CGFPANSFKLPKIN
NvSGL     KGLGV-KGCPPTNFKYSY–
MmSMP     TGLGV-KGIAPYSYAG----
DrSGL     TGLGV-KGIPPYSYTG----
CiSGL
```

FIGURE 2.5 continued.

Gene sequences with an uninterrupted open reading frame along with canonical intron/exon positions were regarded as GSL genes, others were designated pseudogenes. At least one GSL gene sequence was obtained from 21 beetle taxa representing twelve coleopteran families (Table 2.4). All three primer sets failed to amplify GSL sequences from four taxa, *D. ater*, *H. grandis*, *A. rufipes* and *D. quadrimaculatus*. However, representatives of each of their respective families were amplified with the exception of the adephagan *D. quadrimaculatus*. For the non-bioluminescent specimens that were successfully amplified one gene was detected for the majority; two loci were characterised for *C. rustica* and the *Anacaena* sp. and three genes were obtained from *P. m. peregrina*.

From the Lampyridae at least one LRE sequence (or LRE pseudogene sequence) was characterised from all eight species. Three pseudogenes were identified in the Lampyridae, *L. noctiluca*, *P. pyralis* and *P. hemipterus*; these were putatively reduced to 'pseudo' status due to all three sequences exhibited a 'GC' as opposed to a canonical 'GT' at the first intron/exon boundary (Appendix I; Figure 2.6).

Three sequences were obtained from *P. pyralis*, the LRE gene (PpLRE) described by Gomi and Kajiyama (2001), one pseudogene (PpPsG) and a third locus, PpLRE2 which exhibited only 43.2% protein sequence identity to PpLRE compared to a protein sequence identity of 61.1%. PpLRE2 showed more protein sequence identity (71.6%) to an LRE obtained from *P. congener* (PcLRE). For the other four bioluminescent beetles, sequences with high identity to the functional LRE sequences described by Gomi and Kajiyama (2001) were obtained.

```
        E1   *           E2                  E3                   E4                   E5
LmLRE ..TTTGgtaa...ttagAAAA...GCAGgtat...ttagGTAC...ATATgtaa...tcagCAAA...
ACLRE ..ATGGgtaa...acagATAA...ACAGgtat...taagGATC...GTTTgtaa...ttagCAAA...
PpLRE ..GTAGgtaa...tcagATAA...ACAGgtat...ttagGTAC...ATCAgtaa...tcagGCAA...TCAGgtaa...tcagGTCA..
PpPsG ..GTGGgcaa...tcagATAA...GCAGgtaa...tcagGCAC...ATCAgtac...atagCTAA...CCAGgtac...ttagATAA..
LnPsG ..GTAGgcaa...tcagATAA...GCAGgtga...tcagGCAC...ATCAgtat...tcagGTAA...GCTGgtga...ttagATAA..
PhPsG ..TTCGgcaa...ttagATAA...GCAGgtaa...tcagGCTC...ATCCgtat...ttagGTAA...
```

FIGURE 2.6 Intron/exon boundaries for lampyrid LRE genes and pseudogenes. *L. maculata* (Lm); *Photuris* sp. 'AC' (AC); *P. pyralis* (Pp); *L. noctiluca* (Ln) and *P. hemipterus* (Ph). Exons are show shaded grey and denoted E1-E5. Intron boundaries are shown in lower case. A conserved transitional substitution is present in the second base of intron 1 highlighted with an asterisk and shown shaded in black.

Taxa	Gene	Gene Length (bp)[1]	Primer set	Amino acids[1]	Intron size (bp) 1	2	3
Photinus pyralis	PpLRE	1587	-	191	904	54	53
	PpPsG1	803	3	193	120	57	47
	PpSGL	725	3	193	53	44	49
Phausis reticulata	PrLRE	1232	2	192	550	53	51
Phosphaenus hempiterus	PhPsG	732	1	193	55	46	52
Photuris sp 'AC'	ACLRE	864	1	190	153	92	49
Photuris congener	PcSGL	755	3	194	45	44	50
Lampyroidea maculata	LmLRE	726	1	193	46	47	54
Lampyris noctiluca	LnPsG	799	1	192	117	54	52
Lamprohiza splendidula	LsLRE	884	2	193	98	60	50
Cantharis rufa	CfSGL	1927	1	191	979	322	53
Cantharis rustica	CsSGL	1958	1	191	979	353	53
	CsSGL2	1227	1	192	125	463	63
Stegobium paniceum	SpSGL	688	3	193	54	-	55
Anacaena sp.	AnSGL1	815	1	194	105	128	-
	AnSGL2	668	1	199	71	-	-
Necrophorus vespillo	NvSMP	750	1	188	62	50	48
Pachnoda m. ssp. peregrina	PmSGL1	1370	-	195	666	55	58
	PmSGL2	768	-	195	88	58	60
	PmSGL3	579	-	193	-	-	-
Grammoptera ruficornis	GrSGL	910	1	191	206	72	59
Strangalia melanura	SmSGL	1483	1	190	77	763	73
Anthocomus fasciatus	AfSGL	758	1	194	57	55	64
	AfPsG	903	1	195	56	210	53
Pyrochroa serraticornis	PsSGL	724	2	195	49	44	46
Tenebrio molitor	TmSGL	688	3	195	47	-	50

[1] Due to primer position variation, a conserved region was taken for comparison spanning residues 26 to 216 in the *Photinus pyralis* LRE (Genbank acc no. AA066431).

TABLE 2.4 Summary Table of SGL partial gene characterisation from nineteen beetle taxa.

Although no lampyrids failed to produce a LRE or LRE-like sequence amplification was not consistent across all primer sets and products were obtained primarily with only one primer set (Table 2.4). All ten sequences exhibited full open-reading frames, composed of four exons and three introns. Complete conservation of intron/exon sites was seen with the most size variation occurring in the first intron (Table 2.4).

2.3.5 Full length LRE genes in the genomes of *Photinus pyralis* & *Lampyris noctiluca*

Using the degenerate primers SMPdegF4 and SMPdegR1 two different PCR products were visible on the gel and denoted LnLREU (upper) and LnLREL (lower). These PCR products were cloned and 8 recombinant plasmids sequenced for each band.

LnLREL sequences (480bp – 584bp) were all non-specific showing no identity to LRE orthologues in GenBank.[4] Sequenced clones of LnLREU, however, showed good identity to LRE and were denoted LnLRE2. From the eight clones four were successfully sequenced. The remainder suffered from secondary structure and subsequent attempts at sequencing with the addition of DMSO and a higher annealing temperature failed to resolve this problem.

Using a BLASTX search, LnLRE2 was identified as having the highest sequence identity (E value: 9e-57) to the LRE from *Photinus pyralis* (GenBank accession no. AAO66431). Extensive amino acid similarity to *P. pyralis* LRE was seen throughout the length of the LnLRE2 conceptual translation. Both upstream and downstream genome walking was carried out resulting in a 440 bp 5' flanking product and a 748 bp 3' flanking product. A contiguous sequence of 1913 bp was generated that contained a complete putative ORF for LRE divided by four introns (Appendix I). However, two point mutations, one in the putative start codon and a second at the first intron/exon position would render this gene potentially redundant and therefore LnLRE2 has been ascribed a pseudogene status (See Appendix I for full sequence).

Degenerate primers were used to amplify *P. pyralis* DNA as a positive control for the *L. noctiluca* PCR. Although a PCR product was amplified and sequenced, this did not correlate with the established LRE gene sequence. Nine cloned sequences were produced, all 856 bp long and were referred to as the same locus PpLRE2.

To obtain a full open reading frame genome walking was carried out both in a 5′ and 3′ direction. Products were sequenced and a contiguous sequence of 1698 bp was obtained which included a complete ORF with 276 bp of upstream sequence and 221 bp of downstream sequence (Appendix I). This sequence lacked the large intron found in PpLRE1 but the conceptual translation of PpLRE2 shared 60.9% amino acid sequence identity with that of PpLRE1. Furthermore PpLRE2 exhibited 61.6% amino acid sequence identity to the conceptual translation of LnLRE2. All four intron-exon boundaries fell in the same location in PpLRE1, PpLRE2 and LnLRE2. In addition, like LnLRE2, PpLRE2 has a non-canonical intron-exon boundary at intron I (Figure 2.6). All other intron/exon boundaries conform to the GT/AG rule.

---

[4] Two sequences failed to give a readable chromatogram. No flanking vector or insert was sequenced only 600bp of the beta-lactamase gene in pGEM-T Easy vector downstream region. One sequence shown no identity to any insect orthologue the highest identity was to an unknown protein from the long-tailed macaque (E value=1e-05). Two sequences used in blastx searches returned insect orthologues, one 497bp sequence showed strong identity to CG11414-PA in *Drosophila*, a Zn-finger-containing protein, and one 582bp sequence with similarity to CG13868 in *Drosophila* coding for a protein of unknown origin.

In order to ascertain whether PpLRE1 and PpLRE2 are in close proximity, genome walking primers for both loci were combined to attempt intragenic PCR. Both combinations of primers were used PpLRE1 5′ – PpLRE2 3′ and PpLRE1 3′ – PpLRE2 5′ (assuming both genes would be in the same orientation). The same PCR conditions were used as in the genome walking. Both failed to amplify intergenic regions.

2.3.6 GSL genes in genome of *Pachnoda m. peregrina*

PCR amplification of *P. m. peregrina* DNA using the degenerate primer set 1 and a gradient block produced three different sized amplicons (Figure 2.7). These bands were excised, cloned and sequenced. The smallest amplicon (327 bp) was found to be a non-specific product, with no sequence identity to any coding gene in GenBank. The other two amplicon sequences were both found to have high sequence identity to beetle SGL and LRE gene sequences. The two genes were designated PmSGL1 and PmSGL2 and referred to the 1417 bp and 821 bp products respectively.

In order to investigate the presence of functional transcripts coded for by the scarab genome, RT PCR was carried out using degenerate primer set 1. RT PCR produced three PCR products (984 bp, 638 bp and 389 bp) from *P. m. peregrina* total RNA (Figure 2.7). A tblastn revealed that both the 984 bp and the 386 bp were non-specific products. Six clones of the 638 bp product were sequenced of which all, with the exception of one, gave 100% sequence identity to PmSGL2 gene sequence. The RNA sequence confirmed the predicted intron and exon position for PmarSGL-2. The sixth clone, a PCR product of 632 bp showed, extensive sequence divergence from the other transcript (sequence identity 51.6%).

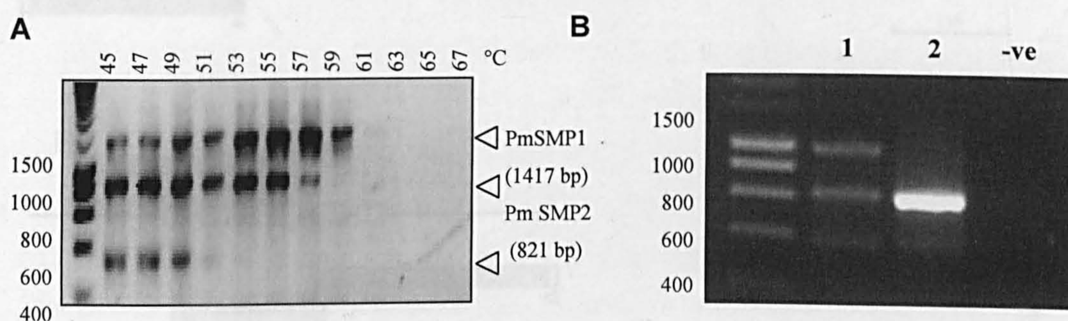


FIGURE 2.7 *Pachnoda m. peregrina* SGL gene PCR & RT-PCR. A. Agarose gel showing gradient block PCR products of *P. m. peregrina* DNA using degenerate primer set 1. 12 samples were amplified on a gradient block ranging in temperature from 45°C - 67°C at 2°C increments. B. Agarose gel showing RT PCR products of *P. m. peregrina* DNA using SMP30 degenerate primer set 1. Lane 1 shows RT PCR products amplified from total RNA. Lane 2 shows secondary PCR from the purifed 600 bp band from previous RT PCR samples.

From a TBLASTN search the sequence gave the greatest identity to the firefly LREs confirming the gene product to be a SGL orthologue and was designated PmSGL3.

Due to the tandem array nature of paralogues in the beetle genome of *T. castaneum* it seemed reasonable to assume that the three loci characterised so far could fall in close proximity to each other on the same chromosome. Using genome walking primers, designed to amplify upstream and down stream of the three loci, PCR was carried out to attempt the amplification of genomic regions between the PmSMP loci. All permutations of nested PCR were attempted of which one product was obtained, an 1163 bp product using primers SGL-1GW3'1 and SGL-3GW5'1. Sequencing confirmed this product contained 388 bp of the 3' region of PmSMP1 gene and 120 bp of the 5' region of PmSGL3 interspersed by 655 bp of intergenic sequence.

Three sets of genome walking primers were designed to amplify upstream and downstream of the three SGL loci characterised in the *P. m. peregrina* genome. The 5' walk from PmSGL1 produced an 1144 bp PCR product. The 3' walk from PmSGL3 produced an 1117 bp product. Combined with the intergenic PCR sequence, the genome walking PCR provided information for a contiguous sequence of 5215 bp that contained the full ORF for PmSGL1 and PmSGL3 (Figure 2.8; Appendix I). The genome walking from the PmSGL2 locus also provided sufficient sequence to identify the full ORF (Figure 2.8; Appendix I).

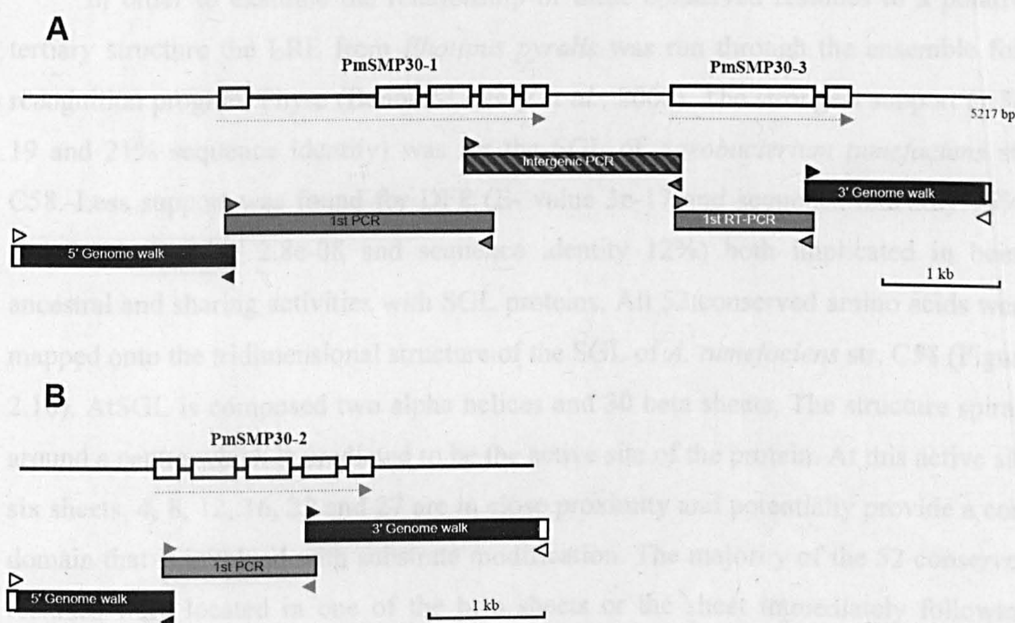


FIGURE 2.8 Schematic of PCR and genome walking for the three loci PmSMP 1-3. Grey filled boxes show the position of degenerate PCR amplified products. Black boxes denote genomic walks from the degenerate PCR products; the diagonal shaded box denotes the position of the intergenic PCR. White boxes denote the gene architecture based upon the contiguous sequence. Dotted arrows denote transcription direction.

Based upon BLASTX results the putative ORF and intron/exon sites were predicted for each of the PmSGL genes (Appendix I). PmSGL1 and PmSGL2 show conserved intron/exon positions composed of five exons divided by five introns averaging 50 bp with the exception of the first intron in PmSGL1 that has a much larger intron size of 666 bp. In PmSGL3 however, only one intron/exon site (IV) is present suggesting extensive intron loss in this gene. The intron exon number and location are not conserved with *T. castaneum* paralogues. It seems evident that extensive intron loss has occurred relatively recently since the two species diverged.

2.3.7 Sequence conservation in the SMP30 protein family

The protein sequence logo representing the alignment of all 45 SGL sequences (Figure 2.9) confirmed the presence of 52 highly conserved residue positions extending from the amino (N) through to the carboxyl (C) end. Of these 15 were glycine conserved residues and six aspartic acid conserved residues. Other residues common to SGLs included: the Try (W^{92}) which was the only residue conserved in a large variable region and Phe (F^{281}) which precedes two conserved glycine residues. Also notable was the high conservation of the residues Lys (K^{125}), Trp (W^{28}), Trp (W^{92}), Met (M^{137}), Pro (P^{26}), Pro (P^{272}), Thr (T^{292}) and Val (V^{291}).

In order to examine the relationship of these conserved residues to a putative tertiary structure the LRE from *Photinus pyralis* was run through the ensemble fold recognition program Phyre (Bennett-Lovsey *et al.*, 2008). The strongest support (1.3e-19 and 21% sequence identity) was for the SGL of *Agrobacterium tumefaciens* str. C58. Less support was found for DFP (E- value 3e-17 and sequence indentity 15%) and PON (E-value 2.8e-08 and sequence identity 12%) both implicated in being ancestral and sharing activities with SGL proteins. All 52 conserved amino acids were mapped onto the tridimensional structure of the SGL of *A. tumefaciens* str. C58 (Figure 2.10). AtSGL is composed two alpha helices and 30 beta sheets. The structure spirals around a centre which is predicted to be the active site of the protein. At this active site six sheets, 4, 8, 12, 16, 22 and 27 are in close proximity and potentially provide a core domain that is involved with substrate modification. The majority of the 52 conserved residues were located in one of the beta sheets or the sheet immediately following (Figure 2.10).

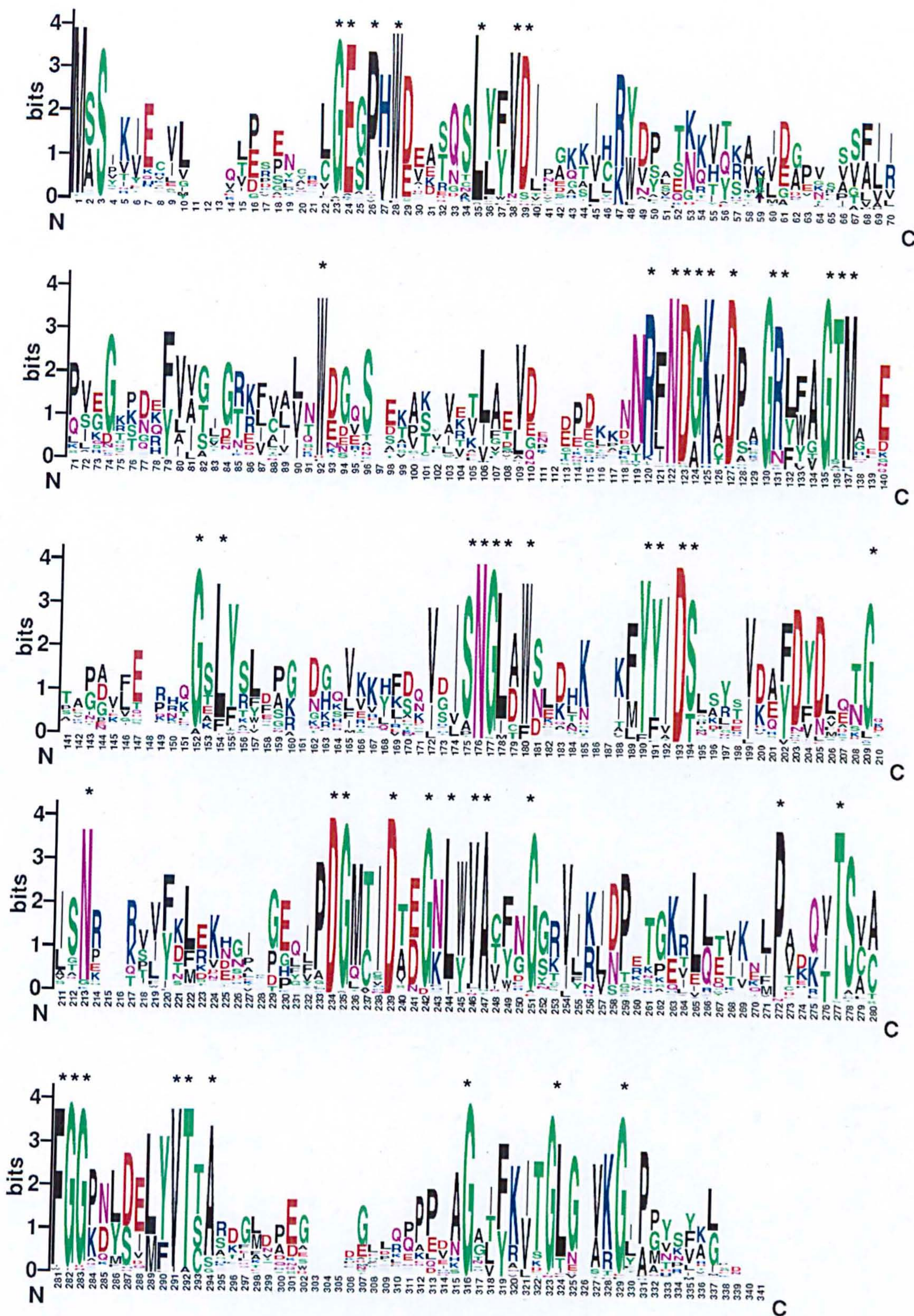


FIGURE 2.9 Sequence logo representation of the SGL protein family. Overall comparison among all the 52 SGL protein sequences used in this study in which the height of a given letter (amino acid residue) represents its frequency of occurrence at that particular site. High variability is observed in the amino (N) and carboxyl (C) terminals. Asterisks (*) above the letters denote invariable or almost invariable residues. Empty sites represent regions of the alignment where gaps were introduced in most of the sequences.

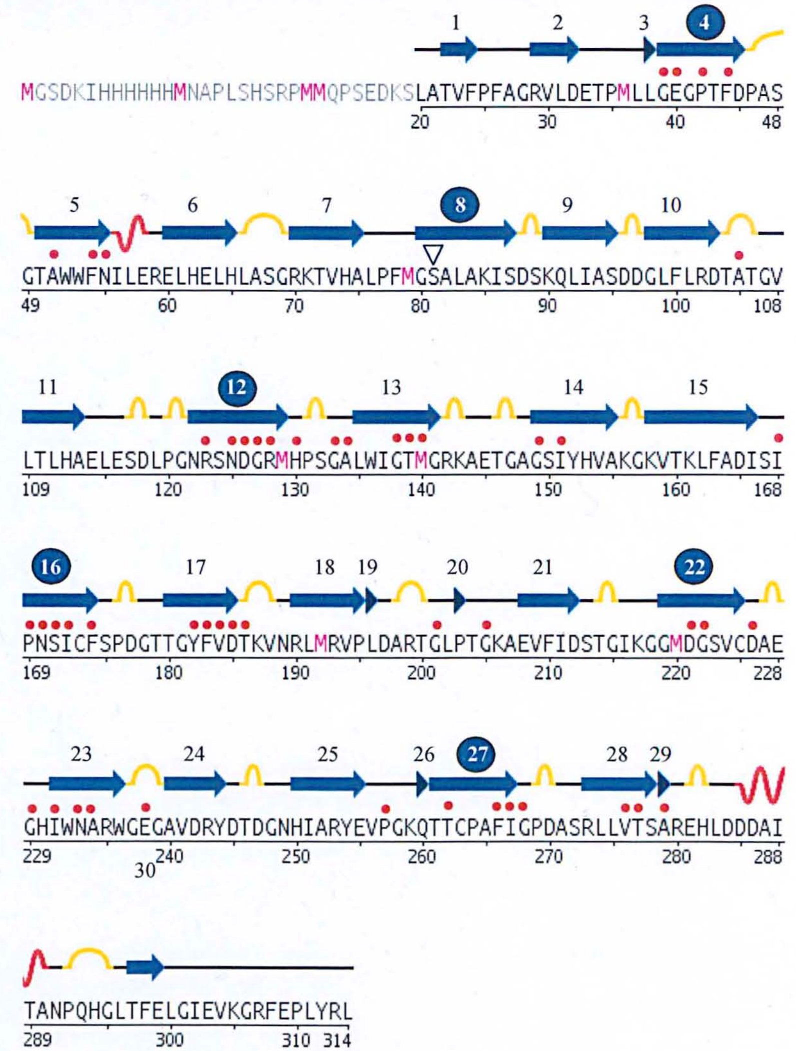


FIGURE 2.10 Crystal structure of SGL protein from *Agrobacterium tumefaciens*. Resolution 1.55 A°; image sourced from the RCSB protein databank (PDB) (accession code 2GHS). Conserved residues identified in Figure 2.9 are plotted as a red circle. The six beta sheets found at the centre of the protein are highlighted with a circle in the secondary structure and numerically marked on the tertiary structure.

2.3.8 Phylogenetic distribution and relationships of the SMP30 protein family

Since previously published metazoan phylogenies included only six insect sequences and four chordate sequences (Gomi *et al.*, 2002) the dataset of 71 family members represents a substantially more comprehensive sample of SGL diversity, albeit incomplete in length. Results using various methods for phylogenetic inference (Bayesian and Maximum Likelihood) showed essentially the same topology (Figure 2.11). Two main coleopteran clade groups could be inferred from the phylogenetic analyses (Figure 2.11). The first group composed of the majority of coleopteran SGL sequences was denoted group I and contained the LRE sequences from the Lampyridae. Also contained in this group were the *T. castaneum* paralogues TcSGL2-6. The group I clade was subsequently divided into subclade groups for further anlaysis. The LRE sequences were grouped as IA and were nested with a larger group, IB, which included: the *Cantharis* SGL sequences, which as expected were found basal to the Lampyridae; *P. m. pergrina* SGL3; *T. castaneum* SGL 5 & 6; two paralogues from *Anacaena* sp.; *S. paniceum* and *A. fasciatum.* A third subclade was denoted IC with reasonable support and consisted of *T. castaneum* SGL 2-4, *S. melanura* and *G. ruficornis.*

The second group was denoted group II and was found basal to the coleopteran SGL phylogeny with strong support and is composed of seven SGL proteins including the first *T. castaneum* SGL paralogue and the orthologue from *P. m. peregrina.* Also found in this group was a paralogue of *A. fasciatus.* The second LRE paralogue not regarded as a pseudogene from *P. pyralis* was found in this group along with another lampyrid sequence from *P. congener.* Four that have SGL genes in group I also have a potentially linked copy in group II. In the phylogenetic results, the bootstrap value and posterior probability that support group II as a monophyletic clade are very high, this is less so for group I. It is predicted that as more SGL sequences become available, the support value for this group will raise dramatically.

2.3.9 Descriptive analysis of the SGL protein family subgroups

To analyze the particular characteristics of each SGL group only complete sequences with the initial methionine and the final stop codon were selected. These included 9 SGLs from beetle group I, 2 from beetle group II, 15 from the Diptera, 11 from the mammalian SMP30's and the bacterial GNL sequence was used as a basal lineage. Table 2.5 shows the major findings of these comparisons. Using the bacterial group to represent the putative original protein it serves as a point of comparison on how the other groups have diverged.

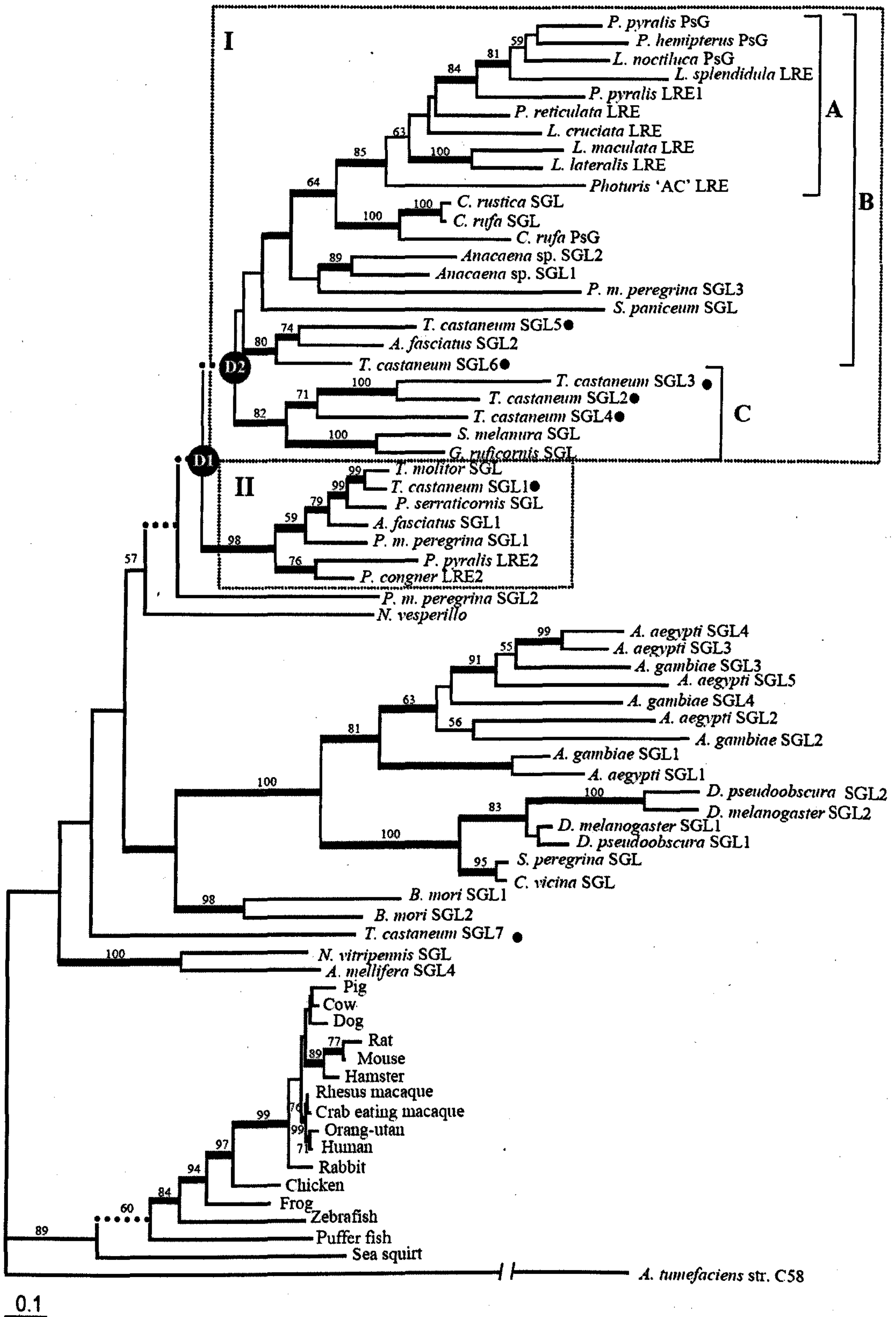


FIGURE 2.11 Phylogenetic analysis of the SGL domain-containing protein family. The results from a Maximum Likelihood bootstrap analysis are shown above the branches; only bootstrap values >50% are shown. The dashed and thick branches represent 91-95% and >95% Bayesian posterior probability respectively. The black circles indicate putative duplication events (D_1 and D_2). The tree has been rooted at the proposed Basal group of the *A. tumefaciens* SGL sequence. Proposed coleopteran groupings are shown boxed; I & II. Proposed subclades (post- D_2) are shown bracketed and denoted A-C. The seven *T. castaneum* paralogues are highlighted with a circle. Analyses were executed using unrooted trees. Branch lengths are proportional to the number of substitutions per site (see scale bar in the figure).

SGL Group	Length[a]	pI[b]	A[c]	B[d]	A/B[e]	AI[f]	Cys[g]	His[g]
Group I	307.1 ± 1.5	6.0 ± 0.2	37.1 ± 0.7	32.8 ± 0.7	1.1 ± 0.0	87.0 ± 1.2	2.4 ± 0.4	6.2 ± 0.6
Group II	309.5 ± 1.5	6.2 ± 0.2	37.5 ± 2.5	34.5 ± 3.5	1.1 ± 0.0	93.4 ± 2.1	2.0 ± 0.0	7.5 ± 0.5
Diptera	307.9 ± 1.0	5.6 ± 0.1	38.9 ± 0.8	33.6 ± 0.6	1.2 ± 0.0	79.3 ± 1.5	2.1 ± 0.4	4.1 ± 0.3
Mammalian	299.0 ± 0.0	5.5 ± 0.1	40.1 ± 0.3	34.7 ± 0.3	1.2 ± 0.0	78.3 ± 0.4	9.9 ± 0.1	4.1 ± 0.3
Basal	295.0 ± 0.0	5.5 ± 0.0	38.0 ± 0.0	29.0 ± 0.0	1.3 ± 0.0	80.8 ± 0.0	3.0 ± 0.0	10.0 ± 0.0

[a] Number of amino acid residues
[b] pI, Theoretical isoelectric point
[c] A, Number of acidic residues
[d] B, Number of basic residues
[e] A/B, Ratio of acidic to basic residues
[f] AI, Aliphatic index
[g] Number of residues in the proteins

TABLE 2.5 Quantitative survey of SGL protein features (Average ± SE).

When compared to the Basal group, SGLs from the other four groups show an increase in size. This increase is particularly evident in Group II with an average increase in size of 14 residues (or about 5%) over the Basal group. Our sequence comparisons show that it is in terms of the amino acid composition that the groups show highly significant differences. The predicted isoelectric points of the group range from 5.5 – 6.2. However, the bacterial sequence shows an acidic isoelectric point of around 5.5, while the isoelectric point of the beetle groups I and II is 6.0 and 6.2 respectively. The different isoelectric points are mainly due to a larger number of acidic residues and a decrease in basic residues found in the bacterial sequence.

Other differences among the SGL groups are noticeable. Beetle group II have a significantly higher aliphatic index than the other groups. However, at the amino acid level, it is surprising that the number of Cys residues is three to five times greater in the mammalian SMP30's in comparison to other groups. This is astonishing given that Cys is the rarest amino acid in the SGL dataset.

Eight Cys residues are completely conserved in the mammalian sequences and most are present in at the active site in close proximity to other Cys residues suggesting interactions between Cys residues in mammalian SGLs. However, these residues were not predicted by the Disulfind server (Vullo, Frasconi, 2004) to participate in disulphide bond formation suggesting these residues do not play a part in protein folding.

Similarly, the number of His residues in the molecules shows significant differences among the groups. The Basal species has 10 residues and the Diptera SGLs along with the mammalian SMP30s have an average of 4.1 His residues in their sequences, which increases to 6.2 in group I and to 7.5 in group II of the beetle SGLs.

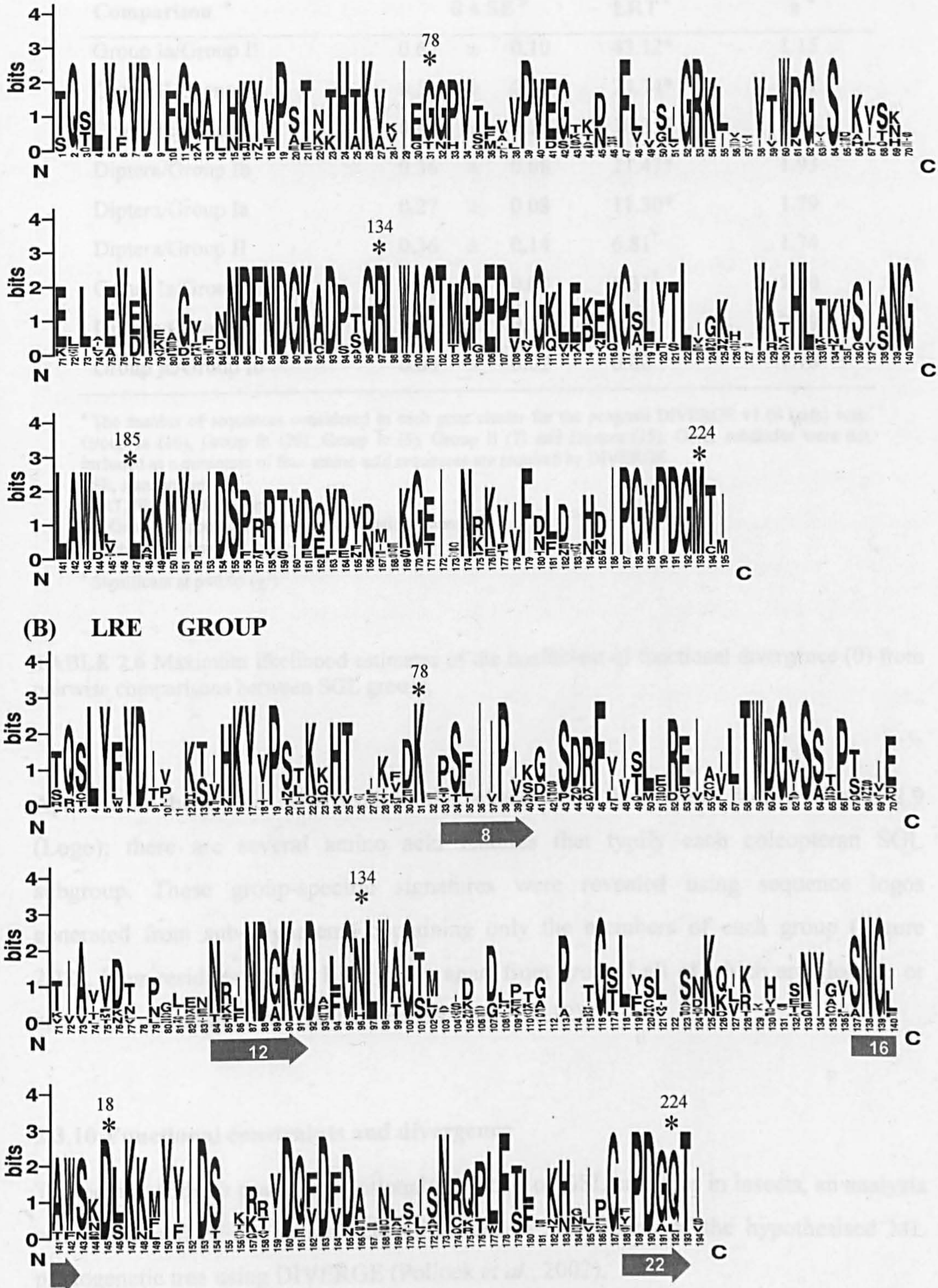


FIGURE 2.12 Comparative analysis of coleopteran SGL protein family groups. WebLogos were created from partial gene alignments that only included the sequences belonging to each SGL group: **(A)** Group I (8 sequences); **(B)** LRE group IIA (10 sequences). Diagnostic residues are denoted with an asterisk and the position in the tertiary structure of *A. tumefaciens* identified above the asterisk (see Figure 2.8). The position of the active domain beta sheets are shown by grey arrows.

Comparison [a]	θ ± SE [b]	LRT [c]	a [d]
Group Ia/Group II	0.67 ± 0.10	43.12*	1.15
Group II/Group Ib	0.51 ± 0.10	24.74*	1.47
Group Ic/Group II	0.47 ± 0.11	18.35*	1.13
Diptera/Group Ib	0.36 ± 0.08	21.47*	1.93
Diptera/Group Ia	0.27 ± 0.08	11.30*	1.79
Diptera/Group II	0.36 ± 0.14	6.81†	1.74
Group Ia/Group Ic	0.20 ± 0.09	5.31†	1.50
Diptera/Group Ic	0.20 ± 0.10	3.72	2.09
Group Ic/Group Ib	0.00 ± 0.02	0.00	1.78

[a] The number of sequences considered in each gene cluster for the program DIVERGE v1.04 (refs) was: Group Ia (10), Group Ib (20), Group Ic (5), Group II (7) and Diptera (15). Other subclades were not included as a minimum of four amino acid sequences are required by DIVERGE.
[b] SE, standard error.
[c] LRT, likelihood ratio test.
[d] α, Gamma shape parameter of rate variation among sites.
* Significant at $p<0.01$ (χ^2)
† Significant at $p<0.05$ (χ^2)

TABLE 2.6 Maximum likelihood estimates of the coefficient of functional divergence (θ) from pairwise comparisons between SGL groups.

Apart from the amino acid residues that characterise the protein family (Figure 2.9 (Logo); there are several amino acid features that typify each coleopteran SGL subgroup. These group-specific signatures were revealed using sequence logos generated from sub-alignments containing only the members of each group (Figure 2.12). Four residues set the LRE group apart from group I all of which are close to or contained within the 6 central beta sheets (Figure 2.12).

2.3.10 Functional constraints and divergence

To begin to address possible functional evolution of GSL enzymes in insects, an analysis was performed of the amino acid alignment in the context of the hypothesised ML phylogenetic tree using DIVERGE (Pollock *et al.*, 2002).

DIVERGE establishes differing site-specific rates of amino acid substitution after gene duplication events, thereby implying altered functional constraints. This is achieved by comparing site specific evolutionary rates in amino acid sequences among subclades within a phylogenetic tree. DIVERGE analyses were based upon pair-wise comparisons amongst four coleopteran GSL subclades and one basal dipteran subclade.

Detectable differences in the site-rate of amino acid replacement between SGL paralogue groups can give an indication of the grade of functional divergence generated since the duplicated genes diverged, splitting ancestral functions or generating new ones, and consequently succeeded avoiding pseudogenization. The coefficient of evolutionary functional divergence (θ) obtained for each SGL paralogue pair comparison was significantly greater than zero (Table 2.6), with the exception of subgroup IC compared to Diptera and subgroup IB, indicating that there is significant heterogeneity in the amino acid site-specific rate of evolution among SGL paralogue in beetles. The comparison between the LRE group (Ia) and group II showed the highest value for θ (0.67 ± 0.10), suggesting that these two groups have diverged considerably more at the functional level within the Coleoptera.

The predicted functional divergence between group II and all the other coleopteran subgroup ($\theta = 0.47 \pm 0.11$ - 0.67 ± 0.10) is higher than that for Diptera vs. subgroups ($\theta = 0.20 \pm 0.10 - 0.36 \pm 0.08$). The predicted functional divergence of group II from group I and its subgroups is clearly shown in the posterior probability analysis (Figure 2.13). The most dramatic difference occurs between group IA (the LREs) and group II with large P(S1/X) values for a number of residues. Key residues conserved in each group compared to the other subset were found to predominate in or in close proximity to the six β sheets forming the putative active group of the protein (Figure 2.14).

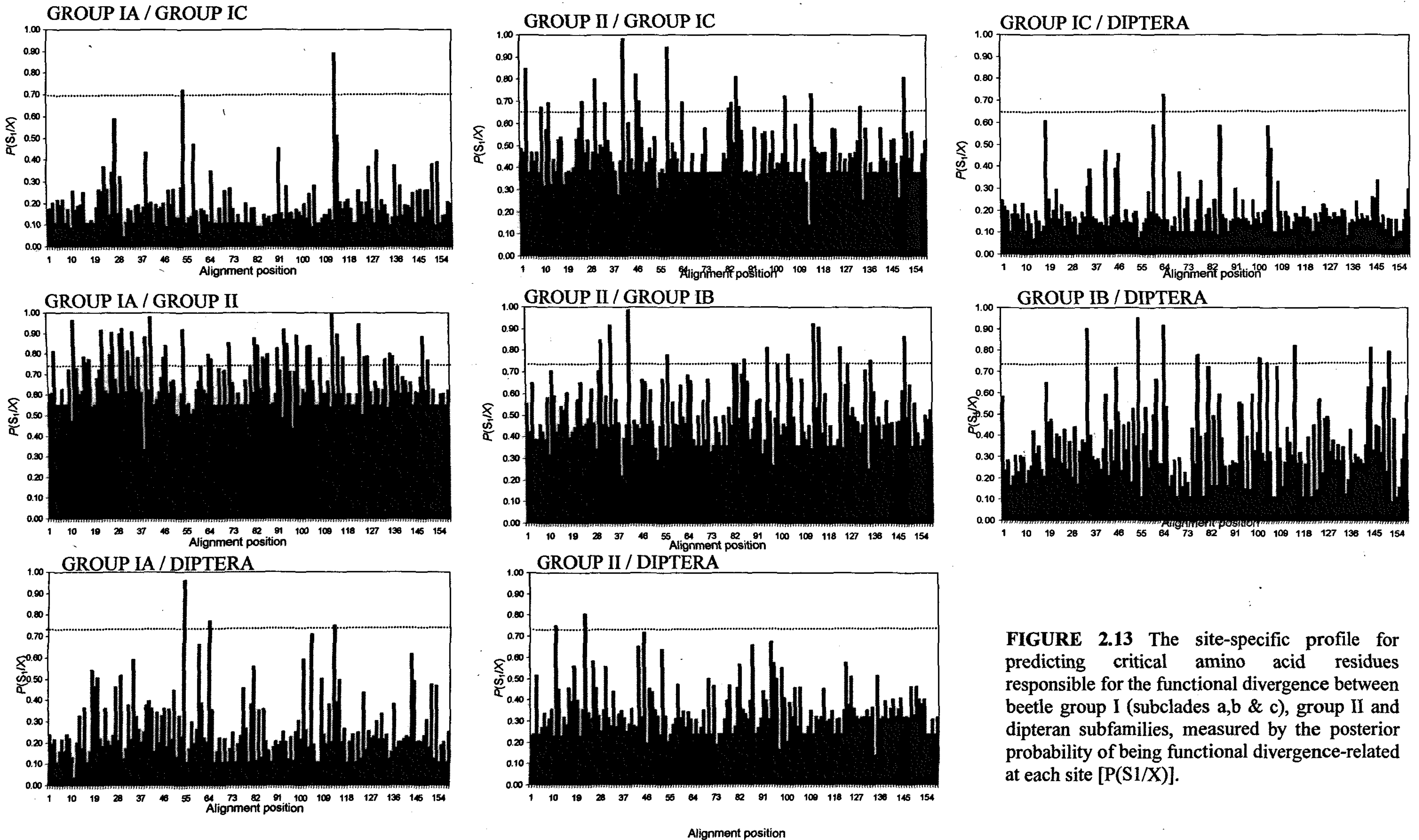


FIGURE 2.13 The site-specific profile for predicting critical amino acid residues responsible for the functional divergence between beetle group I (subclades a,b & c), group II and dipteran subfamilies, measured by the posterior probability of being functional divergence-related at each site [P(S1/X)].

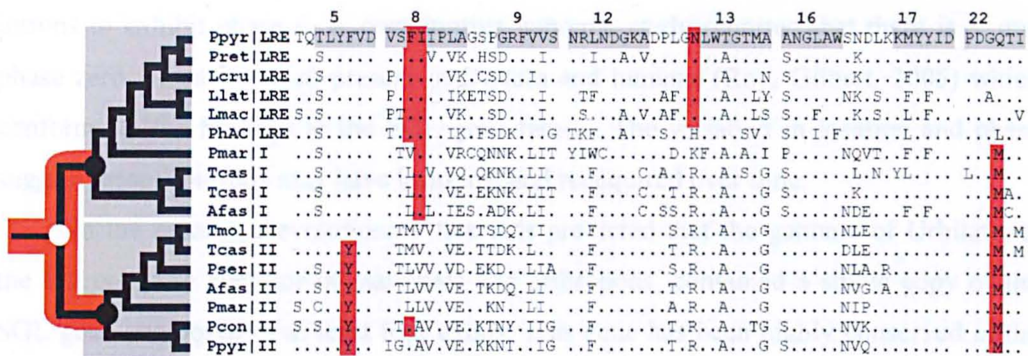


FIGURE 2.14 An illustration of functional divergence in SGL paralogues in beetles. The phylogeny of the genes is shown on the left with the location of the cluster duplication indicated with an open circle and speciation events indicated with a closed circle. The amino acid sequence of SGL genes from *P. pyralis* (Ppyr), *P. reticulata* (Pret), *L. cruciata* (Lcru), *L. lateralis* (Llat), *L. maculata* (Lm), *Ph. Sp. 'AC'* (PAC), *P. m. peregrina* (Pmar), *T. castaneum* (Tcas), *A. fasciatus* (Afas), *T. molitor* (Tmol), *P. serraticornis* (Pser) and *P. congener* (Pcon) are shown on the left with divergent sites highlighted in red.

2.4 Discussion

2.4.1 SGL protein family genes and early origins

Taking advantage of recently sequenced genomes it was possible to collate 29 full length insect SGL sequences for comparative analysis with 17 vertebrate sequences and one bacterial sequence. Although other bacterial sequences were available it was beyond the scope of this study to incorporate prokaryotic evolutionary comparisons. Sequence searches failed to identify SGL domain containing protein coding genes in the genome of *Caenorhabditis elegans*. However, this is not surprising given the recent proposals that *C. elegans* along with other model species have suffered extensive gene loss over time and as a result may hinder large scale genomic comparisons (Kortschak *et al.*, 2003). Although SGL genes were identified in the body louse *Pediculus humanus corporis* no complete full length ORF could be predicted and they were ascribed pseudogene status. The absence of a complete EST library made prediction problematic and due to the early stages of sequencing this genome, coverage may be limited and sequencing errors present. It is therefore necessary to treat the absence of functional SGL genes in the genome of *P. h. corporis* with caution.

The first three SGL exon/intron sites are conserved between the majority of arthropods and the vertebrate SGL gene sequences. Although present in all arthropods

and vertebrates the fourth and fifth (in mammals) intron/exon position are more variable. Thus the least conserved intron positions are those found in the 3′ region and the only introns to exhibit phase 0. A comparative genomic study reported that there is 3′ and phase zero intron loss bias present in Diptera and humans (Roy, Gilbert, 2005) which conforms to the findings in the SGL gene family. The variation in location and phase suggest these 3′ introns may have been lost and reacquired over time.

In the metazoan evolutionary tree it is predicted that the genome of Urbilateria, the last common ancestor of mammals and arthropods, contained a single copy of the SGL gene composed of at least four exons. This gene has been highly conserved in the Deuterostomia but has undergone extensive duplication and modification during arthropod evolution. The monophyly of dipteran and coleopteran SGL protein sequences supports the hypothesis that gene duplication did not occur early in arthropod evolution but has subsequently arisen as the major orders diverged.

2.4.2 SGL protein gene evolution in Diptera

Due to the monophyletic nature of the both the true flies and the mosquitoes the phylogeny (Figure 2.1) supports the hypothesis that one ancestral gene was present prior to the divergence of *Drosophila* and *Anopheles* (the divergence between the Brachycera (*Drosophila*) and the Nematocera (*Anopheles/Aedes*) lineages has been estimated to the Middle Permian (the Anisian; about 245 - 237 MYA) (Blagoderov *et al.*, 2002). The species *D. melanogaster* and *D. pseudoobscura* diverged 25 to 55 MYA (Russo *et al.*, 1995) and evident that the duplication event producing the two SGL paralogues occurred prior to this split. Conversely in mosquitoes neither the *Anopheles* nor *Aedes* gene arrays were evident in the other species and have arisen from a duplication event occurring after the two mosquito genera diverged. SGL diversity between these two mosquitoes indicates that many genus-specific differences have arisen in the 140–200 MYA since the Anophelinae/Culicinae divergence (Krzywinski *et al.*, 2001). Prior to the *Aedes/Anopheles* divergence three SGL genes (SGL01-03) appear to have been present in the ancestral mosquito and the subsequent duplications have occurred after this division.

2.4.3 SGL gene duplication in coleopteran evolution

Members of a gene family may arise from gene duplication events and gene duplication must be postulated whenever the gene tree results to be incompatible with the species tree and can be inferred even when some genes are lost or missing. The tree shown in

Figure 2.11 suggests the occurrence of at least two duplication events along the main lineages (denoted D1 and D2). It is noted that bootstrap values and posterior probabilities were low for certain branches. However, an assumption for the bootstrapping method is an even distribution of the phylogenetic signal throughout the data set (Felsenstein, 1985). As has been shown for the SGL sequences there is complete conservation of some sites and large divergence in others across the phylogeny, and therefore this assumption is not met. Thus, high bootstrap values are not always expected to be obtained for many nodes, because the sites supporting the existence of the principle clades may differ from the sites that are useful for resolving the relationships among more derived groups (Thornton, DeSalle, 2000).

With only one complete coleopteran genome to mine it is not possible to make firm predictions about the extent of gene duplication in all beetles however certain inferences can be made. At least one SGL gene was amplified from each polyphagan superfamily present in this study supporting the ubiquity of this protein in the Polyphaga. Although one adephagan specimen was available for analysis this failed to provide any SGL PCR products and it was not therefore possible to make inferences about ancestral genes present prior to the divergence on the Polyphaga and the Adephaga.

However, seven paralogues present on chromosome 5 of *T. castaneum* suggest extensive gene duplication with this and potentially other coleopteran genomes. This evidence supports at least three duplication events and one potential gene loss. However, a more complex scenario emerges based upon the phylogenetic evidence. The first duplication arose very early in polyphagan evolution, marking the divergence of the SGL group I from that of group II (D1); SGL genes for four beetle taxa were present in both groups: *T. castaneum*, *P. m. peregrina*, *P. pyralis* and *A. fasciatus*. It is probably that other coleopteran species will exhibit similar gene paralogues. Less clear is a second duplication which has led to the segregation of SGL Ib and SGL Ic (D2). All the other duplications evident in *T. castaneum* occurred during the diversification of groups Ib and Ic, disclosing a great complexity in the evolution of these genes.

PCR and RT-PCR amplification provided evidence for at least three functional genes present in the genome of *P. m. peregrina*. 45-70% amino-acid sequence identity is found between the three SMP30 gene products in *P. m. peregrina* and their orthologues in *T. castaneum*. For *Drosophila melanogaster* and *Anopheles gambiae*, which are considered to have diverged approximately 250 MYA (Gaunt, Miles, 2002; Zdobnov *et al.*, 2002), the mean value for sequence identity between orthologues is calculated at between 62% and 56% (Bolshakov *et al.*, 2002; Zdobnov *et al.*, 2002). The families Tenebrionidae (superfamily Cucujoidea) and Scarabaeoidae (Superfamily

Scarabaeoidea) both belong to the suborder Polyphaga. These true Coleoptera did not appear until the Triassic around 230 MYA (Ponomorenko, 2002). The oldest specimen of Cucujiformia was found to be 100 MYO and the oldest scarab known is *Holocorobeus nigrimontanus* (ca. 152 MYO) (Grimaldi, Engel, 2005). A parallel degree of gene duplication is evident in the genome when compared to that of *T. castaneum*. Although the complete compliment of paralogues has not been established it is clear that PmSGL1 & 3 correspond to TcSGL1 & 2 due to their chromosomal proximity and phylogenetic relationship. It is interesting to note the position of the TcSGL7 and PmSGL2 (Figure 2.11). PmSGL2 falls basal to the Coleoptera and *T. castaneum* is found basal to the Diptera and Leidoptera. The presence of a non-SGL ORF (cytochrome p450) downstream of PmSGL2 suggests this may in fact be the 3′ end of the SGL array. Long PCR failed to amplify between PmSGL2 and PmSGL3 but it is highly probable that other SGL paralogues exist between these two genes. No cytochrome p450 gene was found immediately downstream of the TcSGL7 gene suggesting the microsynteny of the SGL gene cluster does not extend beyond this region. This putative position in this gene cluster is also supported by the fact this gene is expressed in the scarab beetle, as is TcSGL7 in the red flour beetle. There seems little conservation in these sequences with other coleopteran SGL proteins and could easily be regarded as pseudogenes, yet both produce functional transcripts.

It seems likely that in coleopteran genomes SGL genes are arranged in an array that extends from a highly conserved gene at the 5′ end through to a highly divergent gene at the 3′ end. The highly conserved first gene has been shown to produce a well supported monophyletic clade with a ML bootstrap value of 97% and posterior probability of 100%. This SGL group contains members of Tenebrionidae, Scarabaeoidae, Cleroidea and Elateroidea representing 25% of polyphagan superfamilies suggesting this gene was present in primeval polyphagans.

The genes downstream present a more complex scenario in that a more divergent clade is composed of the lampyrid LREs, PmSGL3, TcSGL2-6 and other taxa sequences. This clade was divided into two subclades, Ib and Ic. Subclade Ib consisted of all the LRE sequences PmSGL3, TcSGL5 & 6 and other taxa. Subclade Ic was composed of two Chrosomeloidea species and TcSGL2, 3 & 4. TcSGL2 shares more sequence identity to TcSGL3 & 4 than to its syntengenic orthologue PmSGL3. This scenario, on the surface implies recent duplication events within the Tenebrionidae. However, this hypothesis does not always hold true due to the phenomenon of "concerted evolution" (reviewed in (Li, 1997; Ohta, 1980; Ohta, 1983). Under concerted evolution, the level of divergence between two duplicated genes is maintained very low, so that the observed divergence is usually much lower than the expectation when the

molecular clock is assumed. Gene conversion has been considered as the most important mechanism for this homogenization in duplicated genes (*i.e.*, a small multigene family with copy number of 2), although unequal crossing over could also be important for large- or middle-size multigene families. Support for gene conversion is seen when DNA polymorphism data are available for both of the duplicated genes as gene conversion produces "shared polymorphic sites" (Innan, 2003), at which both of the two corresponding sites in the duplicated genes are polymorphic. Unfortunately, no such data is presently available for the *P. m. peregrina* orthologues of TcSMP3 & 4.

2.4.4 Function of SGL genes in insects

This study has identified that the SGL protein family (Pfam - PF08450) is composed of protein sequences bearing a core SGL domain. The architecture and signature residues defining this domain are very distinctive and these signature residues map to a predicted active sight in the bacterial SGL tertiary structure from *A. tumefaciens*. These conserved residues fall within beta sheets that form a propeller structure around the centre of the protein and may provide an active site for substrate binding and modification. The conservation of these residues across phyla support the relationship between the SGL protein and the tertiary structure of *A. tumefaciens* compared to other proposed relatives such as PONS and RNA polymerases (Ishigami *et al.*, 2003; Kondo *et al.*, 2004). However no SGL structure from a eukaryote is available so it is necessary to treat functional and structural predictions with caution.

The large number of conserved glycine residues suggests this amino acid may play an important part in the architecture or function of the SGL family. It has long been recognised that a large proportion of glycine residues are highly conserved among the members of individual structural superfamilies (Branden, Tooze, 1999). The reason for such a high conservation has mainly been attributed to the unique role that glycine residues play in the structure of folded proteins (Branden, Tooze, 1999; Creighton, 1993; Guo *et al.*, 2003). A small size and minimal steric hindrance of side chains means that glycine residues can adopt a range of conformations that are rarely exhibited with any of the other naturally occurring amino acids. As a result, it has been proposed that there has been positive evolutionary pressure to maintain glycine residues at specific positions in the sequences of structurally related proteins in order to preserve their overall architecture (Branden, Tooze, 1999).

By means of BLAST searches Kondo *et al.* (2006) identified homology of rat SMP30 to two kinds of bacterial gluconolactonases (GNLs). The total amino acid sequence of rat SMP30 (299 aa) shares 32% homology with that of *N. punctiforme* GNL

(292 aa) and a part of the amino acid sequence of rat SMP30 (222 aa, residues 9–230) shares 26% homology with that of *Z. mobilis* GNL (247 aa, residues 67–313) (Kondo *et al.*, 2006). Kondo therefore speculated that the protein characterised previously as SMP30 in several animals was a GNL. The *A. tumefaciens* SGL sequence exhibited 24% identity to rat SMP30 and *P. pyralis* LRE and it is therefore entirely plausible that all SGL enzymes have a primary GNL function.

Kondo *et al.* (2006) illustrated that SMP30 is involved in the synthetic pathway of L-ascorbic acid or vitamin C. It is this penultimate step in the ascorbic acid synthesis pathway that is catalysed by mammalian SMP30 whose absence in mice leads to vitamin C deficiency (Kondo *et al.*, 2006). Ascorbic acid is synthesised by the majority of vertebrates and the presence of a biosynthesis pathway in the sea lamprey suggests this appeared early in the evolutionary history of fishes prior to the emergence of terrestrial vertebrates (590-500 MYA) (Moreau & Dabrowski, 1998a; 1998b). This biosynthetic capability has been lost in a number of diverse organisms including teleost fishes (Dabrowski, 1990), passeriform birds (Chaudhuri, Chatterjee, 1969), bats (Birney *et al.*, 1976), guinea pigs and primates including humans. Early studies predicted insects did not produce endogenous ascorbic acid but were limited in their scope. Interestingly Rousell 1958 reported that homogenates of the fat body of the cockroach, *Periplaneta americana* can synthesise L-ascorbic acid from D-mannose. Pierre 1962 considered this synthesis to be carried out by symbiotic bacteria. These studies were criticised by Gupta *et al.* 1972 who conducted independent studies on a number of insect taxa. He used a range of precursors but found no evidence of ascorbic acid synthesis in these species. It is possible that the pathway for ascorbic acid synthesis arose early on in the development of multicellular organisms and has been subsequently lost in insects. It is possible that in the insect SGL protein family another function has evolved.

SMP30 in mammals was initially thought to function in the regulation of calcium. Nakajima and Natori found that an SGL orthologue in the flesh fly *Sarcophaga peregrina* isolated from the anterior fat body (reported as the anterior fat body protein, AFP) did not bind calcium (Nakajima, Natori, 2000). In late larvae of the blowfly, *Calliphora vicina*, the AFP was found instead to bind to the hexamerin receptor arylphorin-binding protein (ABP) and was predicted to play a role in the regulation of hexamerin uptake by fat body cells along the anterior–posterior axis (Hansen *et al.*, 2002). The metabolism of proteins, carbohydrates and lipids occur in the fat body and biochemically, it is the most active organ in insects. The fat body corresponds functionally, in part, to the liver of vertebrates where SMP30 proteins are found. Conversely, in *Apis mellifera* both proteins were only found present in EST libraries

made from the head suggesting a localisation and specialisation of this enzyme in *A. mellifera* and potentially other Hymenoptera.

In bioluminescent beetles the LRE transcripts were not reported to be localised to the lantern, as is evident in the case of luciferase but present in other parts of the body. It is possible that the LRE may also be a fat body protein and it is interesting to note that the lanterns of fireflies are considered to have developed from the fat body. The sequence logo profiles and prior probability analysis of LRE compared to the highly conserved group I (Figure 2.12 & 2.13) clearly show a divergent pattern of amino acid usage and conservation among these two groups. This pattern suggests differences in selective constraints, likely arising from divergence in structural and functional aspects of the proteins phenotype. The highlighted residues found conserved in the LRE were close or contained within the active site beta sheets predicted to be involved in substrate interaction. It is therefore proposed that subfunctionalization in gene copies resulting from duplication events has arisen in the Coleoptera.

Combining the phylogenetic evidence presented and the difference in substrate specificity of LREs and SMP30 enzymes (Kondo *et al.*, 2004) a evolutionary scenario of duplication and subfunctionalization may be present in this group of genes. Thus this protein family in beetles is a suitable target to experimentally test subfunctionalization in gene copies resulting from duplication events (Bush, 2001; Winkler *et al.*, 2003). It is reiterated that protein structural information and more experimental data including functional mutagenesis studies remains to be obtained from representative members of the insect SGL protein groups described here.

Group I and group II full length transcripts are now available for the non-bioluminescent beetle *P. m. peregrina* and work is ongoing to produce full length transcripts for the firefly *Photinus pyralis*. Once expressed it will be possible to test comparatively the activity of these enzymes thereby providing a basis for mutagenesis studies. Thus, it is hoped that these findings provide useful guidelines to further research on how evolution shaped the SGL protein family, their conserved and divergent gene functions.

3. The evolution of the adenylate-forming protein family in beetles: evidence for multiple luciferase gene paralogues in the Lampyridae

3.1 Introduction

Bioluminescence plays a crucial role in a number of aspects of insect life including prey attraction exhibited by the fungoidal gnats (Meyer-Rochow, 2007), aposematic responses in glow-worms (De Cock, Matthysen, 2003) and the most well documented of all, the elaborate courtship displays of fireflies (Lloyd, 1983a). However, the ability to produce light is found only in a limited number of terrestrial taxa and is most evident in beetles.

Three coleopteran families contain the majority of bioluminescent species, the Elateridae (click-beetles), the Phengodidae (Railroad-worms) and the Lampyridae (fireflies and glow-worms). Crowson (1972) collected all the bioluminescent beetles into the series Elateriformia with the exception of a single report of a luminescent staphylinid (Costa *et al.*, 1986). The largest family of bioluminescent beetles is the Lampyridae. In his revision of E. Olivier's Lampyridae catalogue of 1910, McDermott in 1966 listed 1891 lampyrid species and 92 genera. Mcdermott divided these genera into seven subfamilies: Lampyrinae, Photurinae, Luciolinae, Ototretinae, Mathetinae, Pterotinae and Rhagophthalminae. Species belonging to the latter subfamily have undergone considerable taxonomic revision. Olivier (1910) erected the family Rhagophthalmidae for the genus *Rhagophthalmus* and other related genera. Crowson (1955) and McDermott (1964), on the other hand, included these genera among the Lampyridae as a subfamily Rhagophthalminae, but more recent authors moved this subfamily to another bioluminescent family, the Phengodidae (Crowson, 1972; Lawrence *et al.,* 2000). Although Wittmer and Ohba (1994) revised the status of Rhagophthalmidae as a separate family, the relationship between Rhagophthalmidae and Lampyridae remains controversial. Eight subfamilies are currently recognised based upon the revisions of Crowson (1972) and Lawrence and Newton (1995): Amydetinae, Cyphonocerinae, Lampyrinae, Luciolinae, Ototretadrilinae, Ototretinae, Photurinae, and Pterotinae.

The bioluminescence reaction in fireflies (Lampyridae) is catalyzed by the enzyme luciferase (monooxygenase, [EC 1.13.12.7]). The cDNA encoding luciferase was initially isolated from the North American firefly *Photinus pyralis* (de Wet *et al.*, 1987) and has subsequently been characterised from over twenty species of bioluminescent beetle (see supplementary data). Luciferase belongs to a much larger

adenylate-forming family of proteins. Oba *et al.* (2003) presented evidence that firefly luciferase had two catalytic functions; it could act as a monooxygenase (luciferase) but could also synthesise a long-chain fatty acyl-CoA from various long-chain fatty acids (R-COOH) in the presence of ATP, coenzyme A (CoA) and Mg^{2+}. Recently paralogous luciferase-like sequences have been identified from the Japanese firefly *Luciola cruciata* (Oba *et al.*, 2006b). This suggests gene duplication of luciferase-like sequences in bioluminescent beetle genomes. Despite extensive sequence identity of the *L. cruciata* luciferase-like genes to the *bona fide* luciferase, the two paralogous enzymes revealed no bioluminescence activity. Furthermore, only one gene product exhibited long-chain fatty acyl-CoA synthetic activity. It was subsequently proposed that luciferase has arisen from a gene duplication event of an ancestral acyl-CoA synthetase and functionally diverged to acquire a novel bioluminescent function (Oba *et al.*, 2006b).

Duplications in particular gene families are often regarded as an important source of evolutionary novelties that contribute to innovative phenotypic traits and biological functions specific to certain groups of organisms (Ohno, 1970). Currently, the divergence of two paralogues after duplication is considered to follow one of three routes. The most likely outcome of a duplication event is nonfunctionalization when one copy first becomes a pseudogene and eventually becomes extinct (Nei, Roychoudhury, 1973), whereas the second copy retains the original function. The other, less frequent but nonetheless essential evolutionary scenarios are neofunctionalization and subfunctionalization. In the event of neofunctionalization, one paralogue retains the original function, whereas the other evolves a new function during a period of rapid, nearly neutral evolution (Ohno, 1970). Under the subfunctionalization model multiple functions of the ancestral gene are separated between paralogues, both of which evolve under purifying selection (Force *et al.*, 1999; He, Zhang, 2005; Lynch, Force, 2000).

Based upon the extensive knowledge of beetle luciferase, these enzymes and their luciferase-like orthologues present a rare opportunity to investigate the evolution of functional diversification in an insect gene family. However, despite the number of luciferase phylogenies published, very little is known about the extent of gene duplication within both bioluminescent and non-bioluminescent beetle genomes. In the present study, by making use of the recent availability of large scale genomic data from the red flour beetle *Tribolium castaneum* supplemented with additional sequence data it is possible to report the presence of a large family of luciferase-like genes that arose in the evolution of the Coleoptera.

This study identifies extensive gene duplication of lampyrid luciferase genes and presents the first evaluation of luciferase along with its paralogues in an

evolutionary framework, thereby providing a novel insight into the evolution of these genes in relation to the emergence of bioluminescence in the Coleoptera.

3.2 Materials and Methods

3.2.1 Identification of beetle luciferase and luciferase-like sequences

Assembly 2 of the *Tribolium casteneum* genome sequence at the Baylor College of Medicine Human Genome Sequencing Center, Tcas_2.0 and BeetleBase was searched via NCBI with all three paralogues found in *Luciola cruciata* and *Drosophila* luciferase-like proteins as queries using translated BLAST searches (TBLASTN) (Gertz *et al.*, 2006). *T. castaneum* luciferase-like sequences retrieved were in turn employed in searches to find more genes in an iterative process. A 30% amino acid cutoff level was applied to all paralogues. Errors in automated intron/exon site prediction can result in erroneous predicted open reading frames and therefore protein alignments were used to check for irregularities and refine the gene structures. GenBank and the literature were further searched to compile a dataset of full length coleopteran luciferase and luciferase-like sequences.

3.2.2 Degenerate primer PCR of bioluminescent beetles

Twenty coleopteran species from seven polyphagan families formed the basis of this study (Table 3.1). For all specimens DNA was extracted using the High Pure PCR Template Preparation (Roche) kit. The polymerase chain reaction (PCR) was employed following the method of Ohba et al. (2006). PCR was carried out using primers LH1 (5′ GGW-WCH-ACY-GGN-YTN-CCN-AA 3′) and LH3 (5′ AC-YTG-RTA-NCC-YTT-RTA-YTT 3′) using the following PCR conditions: 94 °C, 1 min; 40 °C, 1 min; 72 °C, 2 min; 35 cycles). This was followed by a nested PCR with primers LH2 (5′ ACY-GGN-YTN-CCN-AAR-GGN-GT 3′) and LH4 (5′ TG-RTA-NCC-YTT-RTA-YTT-DAT 3′) (94 °C, 30 s; 40 °C, 30 s; 72 °C, 2 min; 35 cycles). The nested primers flank the third base of codon 208 and codon 441 in the *Photinus pyralis* luciferase sequence (GenBank acc. no. M15077). PCR was carried in a total volume of 25 μl containing 0.2 μM of each of the forward and reverse primers, 10 mM of each deoxynucleoside triphosphate, 1 U of *Taq* DNA polymerase (Sigma Aldrich), and 2 mM $MgCl_2$ in 1x PCR buffer (Sigma Aldrich) using a Tetrad thermal cycler (MJ Research). The presence or absence of PCR product along with the detection of multiple amplicons was determined by running samples on a 1 % agarose gel stained with ethidium bromide.

Species	Location	Individual identifier	Gene	GenBank Acc. no.
Lampyridae				
Photinus pyralis Linnaeus, 1767	Tennessee, USA	PhpKnx1	PhpLUC[2]	EU684088
			PhpLL1	EU684089
Phosphaenus hemipterus Fourcroy, 1785	Antwerp, Belgium	PphAnt1	PphLUC	EU684086
			PphLL1	EU684087
Nyctophila reichii Jacquelin du Val, 1859	Rio de Onor, Portugal	NyrRdO1	NyrLUC	EU684103
			NyrLL1	EU684105
			NyrLL2	EU684104
Lampyris iberica Figueira et al. 2008	Rio de Onor, Portugal	LaiRdO1	LaiLL1	EU684099
Lampyris sardiniae Geisthardt, 1987	Sardinia, Italy	LasSar1	LasLUC	EU684097
			LasLL1	EU684098
Lampyris noctiluca Linnaeus, 1758	Sussex, U.K.	LanDiB1	LanLUC[2]	EU684100
			LanLL1	EU684102
			LanLL2	EU684101
Phausis reticulata Say, 1825	Tennessee, USA	ParKnx1	ParLUC	EU684090
Luciola italica Linnaeus, 1767	Sardinia, Italy	LuiSar1	LuiLUC1[2]	EU684094
			LuiLUC2	EU684096
			LuiLL1	EU684095
Lampyroidea maculata Geisthardt & Day, 2006	Amol forest, Iran	LdmAmF1	LdmLUC1[2]	EU684092
			LdmLUC2	EU684093
Lamprohiza splendidula Linnaeus, 1767	Antwerp, Belgium	LzsAnt1	LzsLUC	EU684091
Photuris congener LeConte, 1851	Florida, USA	PrcFl01	PrcLUC1	EU684081
			PrcLUC2	EU684082
			PrcLUC3	EU684083
Photuris species 'A'	Florida, USA	PrAFl01	PrALUC1	EU684084
			PrALUC2	EU684085
Cantharidae				
Cantharis rufa Linnaeus, 1758	Oxford, UK[1]	CafOxM1	CafLL1	EU684065
			CafLL2	EU684066
			CafLL3	EU684067
Cantharis rustica Fallén, 1807	Oxford, UK[1]	CasOxM1	CasLL1	EU684068
Anobiidae				
Stegobium paniceum Linnaeus, 1758	Oxford, UK[1]	StpOxM1	StpLL1	EU684069
Hydrophilidae				
Anacaena sp. Thomson, 1859	Oxford, UK[1]	AnsOxM1	AnsLL1	EU684070
			AnsLL2	EU684071
Silphidae				
Necrophorus vespillo Linnaeus, 1768	Oxfordshire, UK	NevCuH1	NevLL1	EU684072
			NevLL2	EU684073
			NevLL3	EU684074
			NevLL4	EU684075
Scarabaeidae				
Aphodius rufipes Linnaeus, 1758	Oxford, UK[1]	AfrOxM1	AfrLL1	EU684076
			AfrLL2	EU684077
Pachnoda marginata (Drury 1773) ssp. *peregrina* Kolbe, 1906	Oxford, UK[1]	PmpOxM1	PmpLL1	EU684078
			PmpLL2	EU684079
Cerambycidae				
Grammoptera ruficornis Fabricius, 1781	Oxford, UK[1]	GrrOxM1	GrrLL1	EU684080

1. Material provided by the University Museum, Oxford – collection details unknown.
2. Luciferase sequences *PhpLUC, LanLUC, LuiLUC1* and *LdmLUC1* were identified as homologous to functional, published luciferase sequences AAA29795 , CAA61668, ABA03040 and AAZ74651, respectively.

TABLE 3.1 List of beetle specimens studied, collection details and gene nomenclature.

All visible bands above 700 bp were excised and cleaned using a QIAquick Gel Extraction Kit (Qiagen). PCR products were cloned into the pGEM Easy T vector (Promega) and between eight and sixteen positive colonies for each species sequenced in one direction using the M13F primer with the BigDye version 3 Terminator Cycle Sequencing Kit (Applied Biosystems), and analysed on an ABI PRISM 3100 multicapillary automatic sequencer (Applied Biosystems). Clones containing luciferase and luciferase-like inserts were subsequently sequenced in the reverse direction using the M13R primer. Sequences were edited and aligned using Sequencher 4.5 (Gene Codes Corp.).

3.2.3 Phylogenetic analysis

All luciferase and luciferase-like protein sequences were aligned with the software ClustalX v.1.81 (Thompson *et al.*, 1994) using the BLOSUM30 matrix. BIOEDIT vers 7.0.5.3 (Hall, 1999) was used for manual improvement of the alignment by eye. Sequence logos for the protein alignment was created online using WebLogo Version 2.8.2 (Crooks *et al.*, 2004) (http://weblogo.berkeley.edu/). Based upon the amino acid alignment, maximum likelihood (ML) estimates of the topology and branch lengths were obtained using PhyML v2.4.4 (Guindon, Gascuel, 2003), via the online server (http://atgc.lirmm.fr/phyml/) with the WAG + I + Γ model (as recommended by alignment analysis with ProtTest (v1.2.6) (Abascal *et al.*, 2005)) accounting for site-to-site rate variation using a discrete gamma distribution with four rate categories and invariable sites; the proportion of invariable sites was estimated from the data and support for individual branches was inferred by bootstrap analyses (100 replicates).

Further phylogenetic relationships were investigated using Bayesian techniques as implemented in the computer program MrBayes v3.0b4 (Ronquist, Huelsenbeck, 2003). MrBayes estimates posterior probabilities of clade support using Metropolis-coupled Monte Carlo Markov Chain method (MC3). Five independent runs were conducted using the same settings: starting with a random tree, using the WAG model of amino acid substitution (Whelan, Goldman, 2001) (WAG + I + Γ); as suggested by analysis of the alignment with ProtTest (v1.2.6) (Abascal *et al.*, 2005), generating trees for 5 million generations with sampling every 100 generations, and with four chains (three with the heating parameter set to 0.2). The first 1,250,000 generations (12,500 trees) were discarded from every run and the remaining trees were concatenated. The remaining 187,500 trees were used to compute the final (consensus) tree, and to determine the posterior probabilities at the different nodes.

3.2.4 Analysis of positive selection

Four lampyrid containing clades of the phylogeny were selected for positive selection analysis. The criteria for this selection were primarily their monophyly, number of sequences present and, for clade L1A, the topology was congruent with the phylogeny of the species (Stanger-Hall *et al.*, 2007). Each clade was treated as a sub-tree and submitted as nucleotide alignments to the Selecton server (Doron-Faigenboim *et al.*, 2005) for an analysis of the synonymous *versus* non-synonymous substitution rate ratio.

3.3 Results

3.3.1 Phylogenetics of luciferase and luciferase-like sequences in beetles

Luciferase is a part of a larger AMP-forming superfamily with a conserved core domain suggesting a common ancestor. To obtain an insight into the potential range of AMP-forming protein coding gene paralogues in beetles the genome of the red flour beetle *Tribolium castaneum* was searched for luciferase-like sequences. Sequences from the *T. castaneum* genome bearing high similarity with luciferase and luciferase–like proteins were analysed, ESTs identified, start codons verified and intron/exon boundaries confirmed. In total eight full length CDS sequences were obtained from the *T. castaneum* genome located on four different chromosomes and designated *TcLL1-TcLL8* (Figure 3.1).

Two paralogues had no associated EST and were regarded as pseudogenes (*TcLL7* and *TcLL8*). At the amino acid level *TcLL1–TcLL3*, *TcLL5* and *TcLL6* showed most identity (33.1 – 48.1%) to the *L. cruciata LcLL1* whilst *TcLL4* showed the greatest sequence identity to luciferase (35.2%). *T. castaneum* luciferase–like genes are divided into 4-6 exons, with only a small number of splice sites being conserved amongst the majority of paralogues. Ten different positions are present in the six *T. castaneum* paralogues. Although all five intron/exon sites are found at analogous positions in 2-6 of the *T. castaneum* genes, no individual gene exhibited all five conserved sites (Figure 3.1). *TcLL1* & *TcLL3* exhibit four out of the five intron sites found in *Photinus pyralis*.

The sequence of the AMP-binding domain can be readily targeted by degenerate PCR primers which have been shown here to work for a broad range of beetle taxa spanning six coleopteran superfamilies. Forty one different sequences were obtained by PCR from 12 bioluminescent and 8 non-bioluminescent taxa (Appendix II).

A

Chromosome 2 · Chromosome 4 · Chromosome 7 · Chromosome 8

6 · 5 · 1 2 3 · p1(7) p2(8) · 4

B

XP_974050 TcLL1 SKL
XP_967226 TcLL2 PKL
XP_974076 TcLL3 SKL
XP_967599 TcLL4 MLE
XP_966640 TcLL5 INA
XP_973874 TcLL6 SKL

C

```
                                                     I
TcLL1    MVQEDDRFVI HGPPPLQPLP RLSLGKLVYD SLL--ANIKK GDAFVDAASG ETISYLDLLK KSCFLAESLL K-SGYGRDTI VSISSENNVQ 90
TcLL2    M-LEDDKYVI SGPPPLQPLL KISLGKLVYD SLL--ANFHK GEALVDVSTG ESISYREIFQ KSCSLAETLH R-LGYGQNTV VAISSENNLQ
TcLL3    M-VEDNTFII KGLPPLAPIP DTPIGKLLYD QLL--ANCDN NPALIDAMSG QTLTYRELLD KTCTLAENLR K-SGFGKTTN IAICCQNSVD
TcLL4    M----ESRII TAPLKNIKIP YESVGKLLHD RFN--SFPEN ATALVKVKAS VTWTYHELAT KSKNLAVNLQ EQMKIAKNDV IAIVSGNSGE
TcLL5    M---SAGIVQ SPLGPSKKIP NENLVEYIYK NND---KWIE EPAMTCGVSG RSYTYGMLRM LINRCAQALL GHCGMKPREV VGLLLPNIPE
TcLL6    M-AQTDPNVI VG-PEVERFI EGSLGELLLL LLK--THCDN -VLQVDAATD EELPANLLLS RSIQLAKWLR S-IGVKEGDS ISVNSENRLE
PpyrLUC  M--EDAKNIK KGPAPFYPLE DGTAGEQLHK AMKRYALVPG TIAFTDAHIE VNITYAEYFE MSVRLAEAMK R-YGLNTNHR IVVCSENSLQ

                         II                                                                      +
TcLL1    FYIPVIACLY IGAVVAPINH NYTEYETTHS LNICKPRIVF CSKAVAQKFV QLKNRLGFTE KIVIIDDDFC DIYGTETLAH FIKTGLRGFA 180
TcLL2    FYIPIISCFF IGAIVAPINQ NYTESETIHS LKICEPKIIF CSKAVSHKYI QIKRKFKFSE TIVIIDDQIG -VKGAETLDN FTRK------
TcLL3    FFTPIIAALY IGATVVPINH NYTETELGHA LRVVKPQIIF CSELTRPKFA KLQQRFDF-- LIFLIENLPR ---------- ----------
TcLL4    FWVVTLAALY LGAPVHLLNP RYTTYELKRY FELSRPKLIF CVSEALDKVQ EVGKECHFIE KIVLFDEAPD ASRGTTRL-G DLLKN---PC
TcLL5    YAVVCHGAIE AGLVVTFVNP LYTPDEIKRQ FENAGVKMIV TVPQLLEVAL TIAPQLQEYR TTICIGGEDD PSKNVNGLQS MLMAG-----
TcLL6    FAVVTVATFF VGAVFAPLNP EYTPGELNHV LKLSKPKVIF CSPQTIQTMT KVFADHPNLT HLVLFGAQKR -NESYVIMHE DIIRGA-TGD
PpyrLUC  FFMPVLGALF IGVAVAPAND IYNERELLNS MNISQPTVVF VSKKGLQKIL NVQKKLPIIQ KIIIMDSKTD ----YQGFQS MYTFVTSHLP

                                                         III      +
TcLL1    PLCRFPGV-- DFDPDK-QIA FIMCSSGTTG LPKGVMQTHT NLMVRYMHTI DPRY-VQKAD -----TFLGI LPFFHGFGLV -TNFFALVQG 270
TcLL2    -INSVRFKLI EFDPES-QVA FIMCSSGTTG LPKGVMITHT NVMVRYMHTI DPRY-VTKSD -----NFLAL LPQFHCYGLL -SNFFALVEG
TcLL3    -NGLYRCCLE EVDVGD-HVA FILFSSGTTG LPKGVMITHR NVLTRFAHAD DPRL-VLRKD ---GQSILGL LPFYHAYGLF -VSLACIQKR
TcLL4    SI-FEFETI- -EDLED-QVA FICHSSGTTG LPKGAMITHA NVWLNLCHSD DDDLYPKSPN -----PIVNV VPVYHVHGFS -LSYTSLYQG
TcLL5    ---HEAE-LP GINPRE--IA ILPYSSGTTG LPKGVMLSHY NLVANLVQGE HPALEDLETK DGKRHTMLTV LPFFHIYGFN GILNLCLKNG
TcLL6    NIDESFEATP -VDPKD-AIA TILMSSGTTG LPKGVMCTHE SMTTYVDIMR VTMA--QIIE NDDPSDAMMG LAPFFHSMGF MLMFLNLLRG
PpyrLUC  PGFNEYDFVP ESFDRDKTIA LIMNSSGSTG LPKGVALPHR TACVRFSHAR DPIFGNQIIP D---TAILSV VPFHHGFGMF -TTLGYLICG

                                                                                IV
TcLL1    EKIVVIKRFE EKLFLKAVQD YKIPSLWLAP PLVVLLAKSP LVDQYDLSCI REVTSGAAPL SKETEELVMK RLKIK--GIR QGYGLTEATL 360
TcLL2    QRLILMKKFD EEIFLQTIQN YQISSLFLVS PLIVLLAKSP LVGKYDLSCV KDIVGGAAPL SKETEEAVIT RLKIP--SIR QGYGLTEATL
TcLL3    VKIIVLQKFD ENIYLQCIEK YKITSLTLVP PLAIFLAKSP LAAKYDLSSV QEVGCGAAPL SKNIEELLKR RLKIS--NIT QAYGLTETTL
TcLL4    VKIVIMDNFQ PKIYLENVQN HGVRKLFLVP SLGDFLANSP LVDQYDLSSV KEIYLAAGVL RKNTEEKILD KFKIV--TIR TVYGLTELAA
TcLL5    AHIITIPRFT PEDYLKTLVE YKPSFIFVVP SLLLFLASHP AVTKEHLSSI EAVQSGAAPL TEGLLQKFRQ KVGRDDILIR QGYGMTESSP
TcLL6    KKMVVLSRFK TKIFLDAIIK YKISRLVVPP PVMLVLLKHP LTKQYDLSGI KEIRTGAAPM GKDMERELKN RFKVG--HVS QGYGMTETTL
PpyrLUC  FRVVLMYRFE EELFLRSLQD YKIQSALLVP TLFSFFAKST LIDKYDLSNL HEIASGGAPL SKEVGEAVAK RFHLP--GIR QGYGLTETTS

                                                     ++
TcLL1    GVIMMSVGD- -IKHGSSGKV ATYMKCKIRD PETGKSLGPG KVGELCFKGP MVMPGYYNNE EATRNSFTSD GWLLTGDLGY YDQDEYFYIV 450
TcLL2    CVLMMNVGD- -SKPGSCGKV VSYVTCKVRD PETGKSLGPG KVGELCFKGP LLMPGYYKNE EATRNSFTSD GWLLTGDLGY YDQDEYFYIV
TcLL3    AVMGVPTGE- -TKPGSCGKL YPHLLCKIRD PESRKSLGPN QVGELCVKGP IVMKGYYRDE EATKGAFTSD GWLLTGDLGY YDHDGYFFIT
TcLL4    AIFIIPVNG- -GKSGSCGRV TPGHQVKIVD PETGNPLGCN QTGEICVKG- FAMKGYVNDA GKSREAFDSD GFVRTGDLGY YDQDLYFFIV
TcLL5    VTFCMPKLTP PSKIATIGLP YPGTEAKVIS LSNGEPQGTH KSGELLVRGP QIMMGYLNNE QATAETVDEE GWLHTGDVAY YDEDFYFYIV
TcLL6    GILVSPLGK- -TKVGSVGKI VPGMMAKVID -DTGKALGPY KEGEVCFKGP LIMKGYVGDP VATANTIDQD GWIHTGDVAY YDEDGYFFIV
PpyrLUC  AILITPEGD- -DKPGAVGKV VPFFEAKVVD LDTGKTLGVN QRGELCVRGP MIMSGYVNNP EATNALIDKD GWLHSGDIAY WDEDEHFFIV

           V                                                            +
TcLL1    DRLKELIKYK GFQVAPAELE AVILSHPKVQ DVGVVGLPDE SSGELPVAFV VKKPGANLTE KEIIDFVAGK VSSQKRLRGG VIFVPAIPKN 540
TcLL2    DRLKELIKYK GFQVAPAELE AVILSHPKVQ DVGVVGLPDE SSGELPVAFV VKKPGAKLTE QEIINFVAGK VSSQKRLRGG VIFVPSIPKN
TcLL3    GRLKELIKYK GLQVPPAELE AILLTHPKIK DVGVIGIPDE EAGELPLAFI VRNE-DDLTE DQVKSFLDGK VSPHKRLRGG VIFLEEIPKN
TcLL4    DRMKDLIKYK SFQVPPLEVE QVLLMFPGVA DAAVVGRPDE RCGELPVAFV VREKGAEVDE SELVEHVGRF LTKEKHLHGG VRFIEGIPRN
TcLL5    DRCKELIKVK GNQVSPTELE NLLLEMPGVA DCAVVGIPDA LAGEVPRAFV VRQPGSSLSE DDILLYINPK VAHYKKIAGG VKFVESIPRN
TcLL6    DRIKELIKYK GYQVAPAELE ALLITHPAVA DAAVIGLPDE RAGELPLAFV VKKPNHETTD KELEKFVADN VSSQKQLRGG VVFIDAIPRN
PpyrLUC  DRLKSLIKYK GYQVAPAELE SILLQHPNIF DAGVAGLPDD DAGELPAAVV VLEHGKTMTE KEIVDYVASQ VTTAKKLRGG VVFVDEVPKG

TcLL1    PSGKILRREL RKMLSE---- FKSKL                                                                 565
TcLL2    PSGKILRREL RNTLIN---- RRPKL
TcLL3    PSGKILRRKL HELFHRY--- IRSKL
TcLL4    EIGKILRKKL REMLE----- -----
TcLL5    PSGKILRNEL KINA------ -----
TcLL6    PSGKILRRHL KQHAIT---- LKSKL
PpyrLUC  LTGKLDARKI REILIKAKKG GKSKL
```

FIGURE 3.1 Luciferase-like sequences in the genome of the red flour beetle *Tribolium castaneum.* A. Chromosome location of paralogues, transcription direction represented by arrows; striped arrows denote genes with no associated EST. B. Inton positions in CDS sequences of 6 paralogues. Peroxisome targeting sequences are highlighted in bold large font at the 3' terminus. C. Amino acid alignment of *T. castaneum* paralogues (TcLL1-TcLL6) and *Photinus pyralis* (PyrLUC). Intron exon boundaries are shown highlighted in grey and black for *T. castaneum* and *P. pyralis* respectively. Roman numerals refer to intron numbers in *P. pyralis* and + denotes addition splice positions found in *T. castaneum* paralogues.

Up to four paralogous gene sequences were identified in a single DNA extract (Appendix II). All sequences showed conserved intron/exon boundaries for both introns III and IV with the exception of the two *Necrophorus vespillo* paralogues *NevLL2* and *NevLL3* where intron IV was missing. The typical GT-intron-AG splice sites were conserved in most sequences but were absent in the *Cantharis rustica* sequence and the *Pachnoda m. peregrina* sequence *PmpLL2* resulting in a disrupted ORF suggesting potential pseudogene status for these loci (Appendix II).

The datset for analysis included a total of 73 sequences of the AMP-binding domain of bioluminescent and non-bioluminescent beetles, 41 of which are newly published sequences (Table 3.2). 32 full length beetle sequences were obtained from Genbank (Table 3.2) of which 4 were removed from the datset having been duplicated in the PCR analysis. The full alignment was trimmed to the partial luciferase and luciferase-like datset of 41 sequences. The final amino acid alignment was composed of 248 residues, representing 45% of the complete protein.

The resulting phylogeny presents two principle clades a luciferase clade 'A' and a luciferase-like clade B (Figure 3.2). Lampyrid paralogues were not monophyletic but distributed throughout the tree in four clades denoted L1-L4. Only one non-bioluminescent beetle taxa was included in clade A, the luciferase-like sequence from the click beetle *Agrypnus binodulus binodulus*. Interestingly, no evidence of a gene from the two cantharids was found despite repeated attempts at cloning and sequencing.

All known functional lampyrid luciferases formed a monophyletic group, L1, composed of four subclades L1a-d. L1a was composed of all Lampyrinae luciferases along with *Phausis reticulata*. Subclade L1b contained the *Photuris pennsylvanica* luciferase *Ppe2* (Ye et al. 1997) along with luciferase sequences from the two *Photuris* species in this study. However, what is surprising is the presence of two *Photuris* luciferase paralogues in subclade L1a. One further *Photuris* paralogue identified by PCR is present in subclade L1d along with a *P. pennsylvanica* luciferase paralogue obtained from GenBank and designated the same as *Ppe1* (see Ye et al., 1997).

Subclade L1c was composed of all Luciolinae luciferases. The two sequences obtained in this study for *Luciola italica* and *Lampyroidea maculata* were represented in GenBank by full length luciferase sequences exhibiting bioluminescent activity. One further paralogue was found in *L. maculata* (*LdmLUC2)* and two for *L. italica* (*LuiLUC2* & *LuiLL1*). *LdmLUC2* and *LuiLUC2* were present in subclade L1c along with the functional *P. pennsylvanica* luciferase and basal to this clade was the *Lamprohiza splendidula* luciferase sequence.

Taxa	Locus	PTS	GenBank Acc. No. & Reference
TENEBRIONIDAE			
Tribolium castaneum	TcLL1	PKL	XP_974050 (unpublished)
	TcLL2	SKL	XP_967226 (unpublished)
	TcLL3	SKL	XP_974076 (unpublished)
	TcLL4	-	XP_967599 (unpublished)
	TcLL5	-	XP_966640 (unpublished)
	TcLL6	SKL	XP_973874 (unpublished)
	TcLL7	-	XP_966820 † (unpublished)
	TcLL8	-	XP_966909 † (unpublished)
Tenebrio molitor	TmLL1	SKL	BAE95689 (Oba *et al.*, 2006a)
	TmLL2	SKL	BAE95690 (Oba *et al.*, 2006a)
	TmLL3	-	BAE95691 (Oba *et al.*, 2006a)
ELATERIDAE			
Agrypnus binodulus binodulus	AbLL1	SKL	BAF96580 (Oba *et al.*, 2008)
Pyrophorus mellifluus	PymLUC1	SKL	AAQ19142 [dorsal] (Stolz *et al.*, 2003)
	PymLUC2	SKL	AAQ19141 [ventral] (Stolz *et al.*, 2003)
Pyrophorus plagiophthalmus	PypLUC1	SKL	AAQ11733 [dorsal] (Stolz *et al.*, 2003)
	PypLUC2	SKL	AAQ11693 [ventral] (Stolz *et al.*, 2003)
PHENOGIDAE			
Rhagophthalmus ohbai	RhoLUC	SKL	BAF34360 (unpublished)
Phrixothrix hirtus	PxhLUC	SKL	AAD34543 [red] (Viviani *et al.*, 1999a)
Phrixothrix vivianii	PvLUC	SKL	AAD34542 [green] (Viviani *et al.*, 1999a)
LAMPYRIDAE			
Luciola cruciata	LucLUC	-	BAE80731 (Oba *et al.*, 2006b)
	LucLL1	AKL	BAE80728 (Oba *et al.*, 2006b)
	LucLL2	SKL	BAE80729 (Oba *et al.*, 2006b)
Luciola lateralis	LulLUC	-	Q01158 (Tatsumi et al., 1992)
Luciola mingrelica	LumLUC	-	AAB26932 (Devine *et al.*, 1993)
Luciola italica	LuiLUC	-	ABA03040 (Branchini *et al.*, 2006)
Lampyroidea maculata	LdmLUC	-	AAZ74651 (Emamzadeh *et al.*, 2006)
Photuris pennsylvanica	PupLUC	SKL	AAB60897 (Ye *et al.*, 1997)
	PupLUC2	SKL	BAA05006 (unpublished)
Lampyris noctiluca	LanLUC	SKL	CAA61668 (Sala-Newby *et al.*, 1996)
Photinus pyralis	PhpLUC	SKL	AAA29795 (de Wet *et al.*, 1987)
Diaphanes pectinealis	DipLUC	SKL	ABD66580 (Li *et al.*, 2006)
Nyctophila cf. caucasica	NycLUC	SKL	AAY99776 (Day *et al.*, 2006)
Cratomorphus distinctus	CrdLUC	SKL	AAV32457 (Viviani et al., 2004)
Hotaria unmunsana	HouLUC	-	AAN40975 (Choi et al., 2003)
Hotaria tsushimana	HotLUC	-	AF486804 (Choi et al., 2003)
Hotaria papariensis	HosLUC	-	AAN40978 (Choi et al., 2003)
Hotaria parvula	HopLUC	-	AAC37253 (Ohmiya *et al.*, 1995)
Pyrocoelia rufa	PyrLUC	SKL	AAG45439 (Lee et al., 2001)
Pyrocoelia pectoralis	PypLUC	SKL	ABM21578 (unpublished)
Pyrocoelia miyako	PymLUC	SKL	AAC37254 (Ohmiya *et al.*, 1995)

† Putatively denoted as a pseudogene due to unassociated EST and excluded from further analysis.

TABLE 3.2 Coleopteran luciferase (LUC) and luciferase-like (LL) sequences retrieved from public databases.

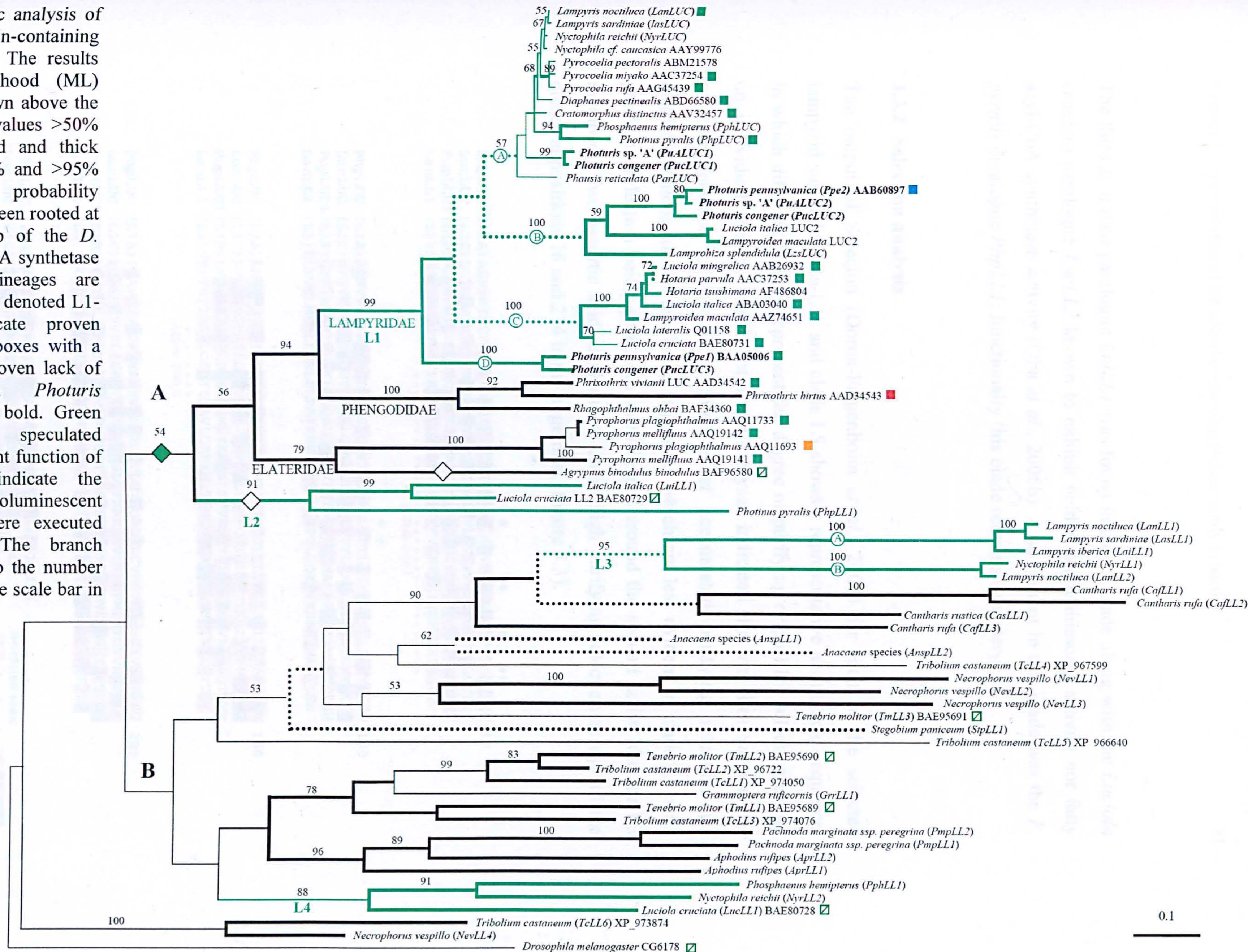


FIGURE 3.2 *Phylogenetic analysis of the AMP-binding domain-containing protein family in beetles. The results from a maximum likelihood (ML) bootstrap analysis are shown above the branches, only bootstrap values >50% are indicated. The dashed and thick branches represent 91-95% and >95% Bayesian posterior probability respectively. The tree has been rooted at the proposed Basal group of the* D. melanogaster *fatty acyl CoA synthetase sequence. Lampyridae lineages are shown in green and clades denoted L1-L4. Filled boxes indicate proven biolum-inescent activity; boxes with a line through denoted a proven lack of bioluminescent activity.* Photuris *paralogues are shown in bold. Green diamonds indicate the speculated emrgence of bioluminescent function of and white diamonds indicate the predicted loss of bioluminescent activity. All analyses were executed using unrooted trees. The branch lengths are proportional to the number of substitutions per site (see scale bar in the figure).*

The third *L. italica* paralogue *LuiLL1* was found in the L2 clade along with the *Luciola cruciata* paralogue *LucLL2* known to exhibit neither bioluminescent activity nor fatty acyl-CoA synthetase activity (Oba *et al.*, 2006b). Also present in this clade was the *P. pyralis* paralogue *PhpLL1*: functionally this clade remains a mystery.

3.3.2 Selection analysis

The output of Selecton (Doron-Faigenboim *et al.*, 2005) for each of the selected lampyrid subclades L1a-L1c and clade L3 shows a representative amino acid sequence, in which different colours represent the degree of purifying or positive selection acting on individual codons. Results of these analyses indicated that no sites were under positive selection (Figure 3.3). Sites under accentuated purifying selection are distributed along all the luciferase sequences which are less evident in clade L3 with purifying selection being greatly reduced in and around the second active site for L3 members. Two aspartic acid residues are under high purifying selection across all the clades at positions 16 and 214 in the alignment (Figure 3.3).

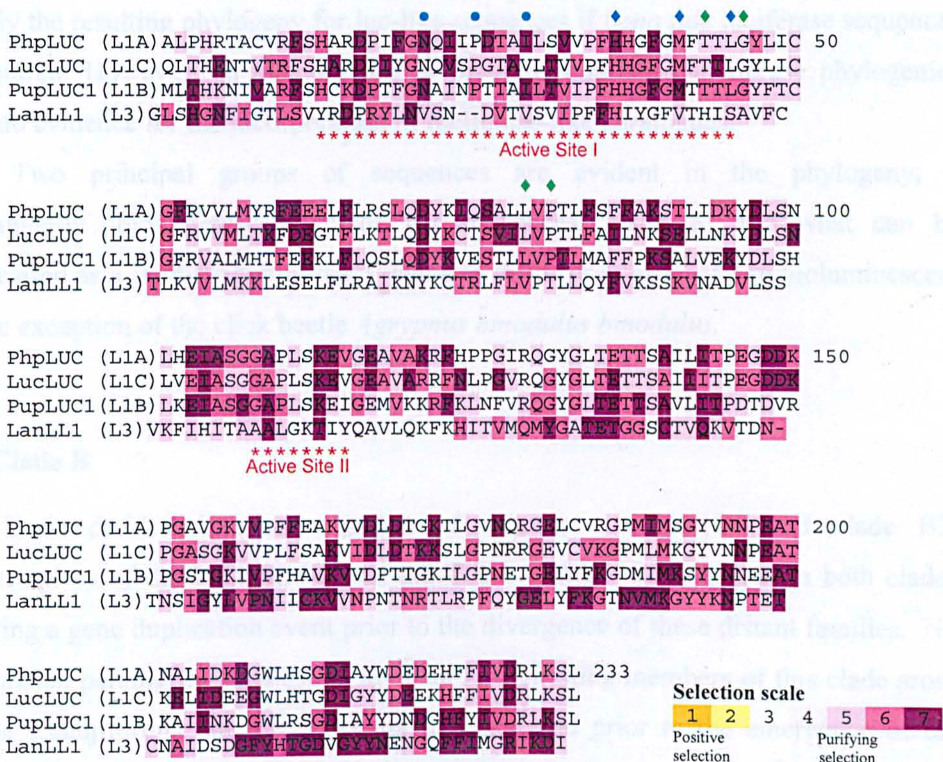


FIGURE 3.3 Output of the SELECTON analysis: representative amino acid sequences generated for each selected clade of Figure 3.2. Intensity of shading indicates likelihood of purifying selection estimated for the amino acid residues (white, pale pink and carmine: moderate purifying selection; mauve: strong purifying selection). Active sites are indicated by asterisks. Putative blue-shift residues are highlighted with a blue diamond, and those considered in involved in colour in *Luciola cruciata* are highlighted with a geen diamond.

3.4 Discussion

Recent studies revealed that the firefly *Luciola cruciata* expresses three paralogues, a *bona fide* luciferase and two luciferase-like sequences where only the luciferase exhibits bioluminescent activity (Oba *et al.*, 2006b). Other than this taxon the extent of AMP-forming protein duplications in the Lampyridae and beyond is unknown. The purpose of this study was to compile available sequence data in order to provide a phylogenetic framework for further studies including: beetle luciferase sequences; homologues in the *Tribolium castaneum* genome and a PCR based sequence dataset of paralogous genes from both bioluminescent and non-bioluminescent beetles. Luciferase sequences have in the past been predominantly identified from cDNA libraries derived from firefly and glow-worm lanterns where luciferase is expressed in much higher levels than the rest of the organism (Strause *et al.*, 1979). It is predicted that luc-like paralogues will not be expressed in the lantern so there is less chance of luc-like sequence being mistaken for true luciferases. Furthermore, luc-like sequences are unlikely to have a bioluminescent function and therefore will not be picked up through the luminescent screening of the cDNA library. The use of degenerate primers, as is seen in this study and that of Oba *et al.* (2006b), successfully detect luc-like genes. It is therefore necessary to evaluate carefully the resulting phylogeny for luc-like sequences if *bona fide* luciferase sequences are required. However, an examination of previous published luciferase phylogenies shows no evidence for the incorporation of luciferase-like paralogues.

Two principal groups of sequences are evident in the phylogeny; a predominantly 'bioluminescent' clade off sequences, clade A, and what can be hypothesised as a 'non-bioluminescent' clade B. All taxa in group A are bioluminescent with the exception of the click beetle *Agrypnus binodulus binodulus.*

3.4.1 Clade B

Clade B is divided into two groups of sequences, clade B1 and clade B2. Representatives of the Lampyridae and the Tenebrionidae were present in both clades suggesting a gene duplication event prior to the divergence of these distant families. No *T. castaneum* paralogues were found in clade A suggesting members of this clade arose after the Cucujiformia and Elateriformia diverged and prior to the emergence of the three main bioluminescent families, Elateridae, Phengodidae and the Lampyridae.

Clade B1 is composed of a monophyletic Elateroidea group with *Cantharis* sequences basal to this group. The close relationship between the Cantharidae and the Lampyridae is evident in other phylogenies (Bocakova *et al.*, 2007). The Tenebrionidae

sequences in clade B1 were not monophyletic and there is no clear relationship between the *T. molitor* sequence *TmLL3* and any *Tribolium* paralogue. Weak support for deeper branches, long branch lengths and limited taxa sampled is evident in clade B2 compared to clade A and B2. The long branch lengths could suggest the possibility of Tenebrionidae B1 paralogues being pseudogenes or on their way to becoming one. However, *TmLL3* is known to be expressed (Oba *et al.*, 2006a) and *TmLL1-6* all have associated EST sequences. In contrast TcLL7 and TcLL8 are regarded as pseudogenes based upon disrupted ORFs and lack of associated EST sequence.

In clade B2 the three *T. castaneum* paralogues *TcLL1-3* are found in a tandem array and form a monophyletic grouping along with two *Tenebrio molitor* paralogues (*TmLL1* and *TmLL2*) and surprisingly a *Grammoptera ruficornis* (Chrysomeloidea: Cerambycidae) paralogue *GrrLL1*. Oba *et al.* 2006a identified three paralogues in *T. molitor* that exhibited fatty acyl-CoA synthetase activity but no bioluminescent activity. From the phylogeny, given the close proximity of the *TcLL1-3* genes in the *T. castaneum* genome and the close relationship of *Tribolium* to *Tenebrio* it is anticipated that *TmLL1* and *TmLL2* are a part of a tandem array of three genes with a further uncharacterised *T. molitor* gene upstream of *TmLL2*. Four genes from the Scarabaeoidea form a sister clade to the Tenebrionoidea and Chrysomeloidea; basal to this are the Elateriformia sequences. This grouping is more consistent with the Coleoptera species tree and shorter branch lengths evident in clade B2 suggest more concerted evolution upon this group of genes and potential functional constraints.

3.4.2 Clade A

L2 forms the basal group of clade A and contains the *L. cruciata* paralogue *LucLL2* which Oba *et al.* (2006b) found exhibited neither bioluminescent nor long chain fatty acyl-CoA synthetase activity. As a result we hypothesise the ancestral gene for group A contained both an oxygenase and a long chain fatty acyl-CoA synthetase activity both of which have been subsequently lost in members of clade L2 as a result of subfunctionalization and place the origins of a bioluminescent protoluciferase prior to the divergence of the Lampyridae, Elateridae and the Phengodidae (Figure 3.2).

The strong support for the Lampyridae clade L1 (ML = 0.99; Bayesian = 1.0) and L2 (ML = 91; Bayesian = 1.0) provides a basis for differentiating the lampyrid *bona fide* luciferase sequences from luciferase-like paralogues without the need to express the protein. All lampyrid luciferases formed a monophyletic group L1 subdivided into four subcades L1a-L1d all with strong Bayesian/ML bootstrap support with the exception of L1a were ML support was only 0.57 despite high Bayesian support (0.90). *Photuris*

congener luciferase paralogues were identified and distributed in three of the four L1 subclades. *PucLL1* and *PucLL2* corresponded with *P. pennsylvanica* cDNA sequences isolated from adults (denoted herein as *Ppe1* and *Ppe2*) (Kutuzova *et al.*, 1997; Ye *et al.*, 1997). Further evidence of paralogous luciferase genes were found in Luciolinae specimens, *Luciola italica* and *Lampyroidea maculata.* Interestingly, paralogues *LuiLUC2* and *LdmLUC2* formed a subclade with *Ppe2.* Cho *et al.* (1999) identified the possibility of three luciferase alleles in *Luciola lateralis* indicating at least two luciferase genes within this species supporting our findings. Despite extensive sequencing of multiple clones no *Lampyris* or *Nyctophila* sequences were identified that were paralogous with L1b genes. With the exception of *Lamprohiza splendidula* all other species in subclade L1b are flash communicators. Conversely clade L3 is composed of glowing lampyrids and the two clades may be a result of early paralogue divergence of flashing and non-flashing species. However, without the presence of complete genomic sequence it is impossible to determine the true extent of paralogous genes for a species. It is evident in the genome of *T. castaneum* that luciferase-like paralogues are located in close proximity on a chromosome; a similar scenario may be present in lampyrid genomes.

Ppe1 and *Ppe2* exhibit different properties, the former is a pH sensitive green emitter (λ_{max} = 558 nm) and the latter is a pH insensitive blue-green emitter (λ_{max} = 538 nm) (Kutuzova *et al.*, 1997). Blue-green emission is an unusual phenomena in bioluminescent beetles and recent studies identified key residues involved in generating blue-shifted light in firefly luciferase (Branchini *et al.*, 2007). Residue comparisons of *Photuris* sequences for these sites support their findings in that *Photuris* species 'A' exhibited Ile241 and Ser250 which illicit a shift from green to blue in *Photinus pyralis* when replacing the wild type Val241 and Phe250 (Branchini *et al.*, 2007). However, it is apparent that other residues are involved in creating this blue shift; Branchini et al. achieved only a shift to 548 nm with a triple mutation compared to 538 nm of *Ppe2.* Consequently other residues may be involved in the active site and putative amino acids have been indicated in Figure 3.4.

The firefly *Phausis reticulata,* common in the South Eastern states of the USA, emits a blue-green glow resulting in the moniker the "blue ghost". One gene was obtained from this firefly species which was present in the clade L1a. Examination of the putative residues responsible for blue light revealed a predominance of 'green' residues and it is predicted that, despite the failure to identify further paralogues a second paralogue exists in the genome of *P. reticulata* related to sequences in clade L1b.

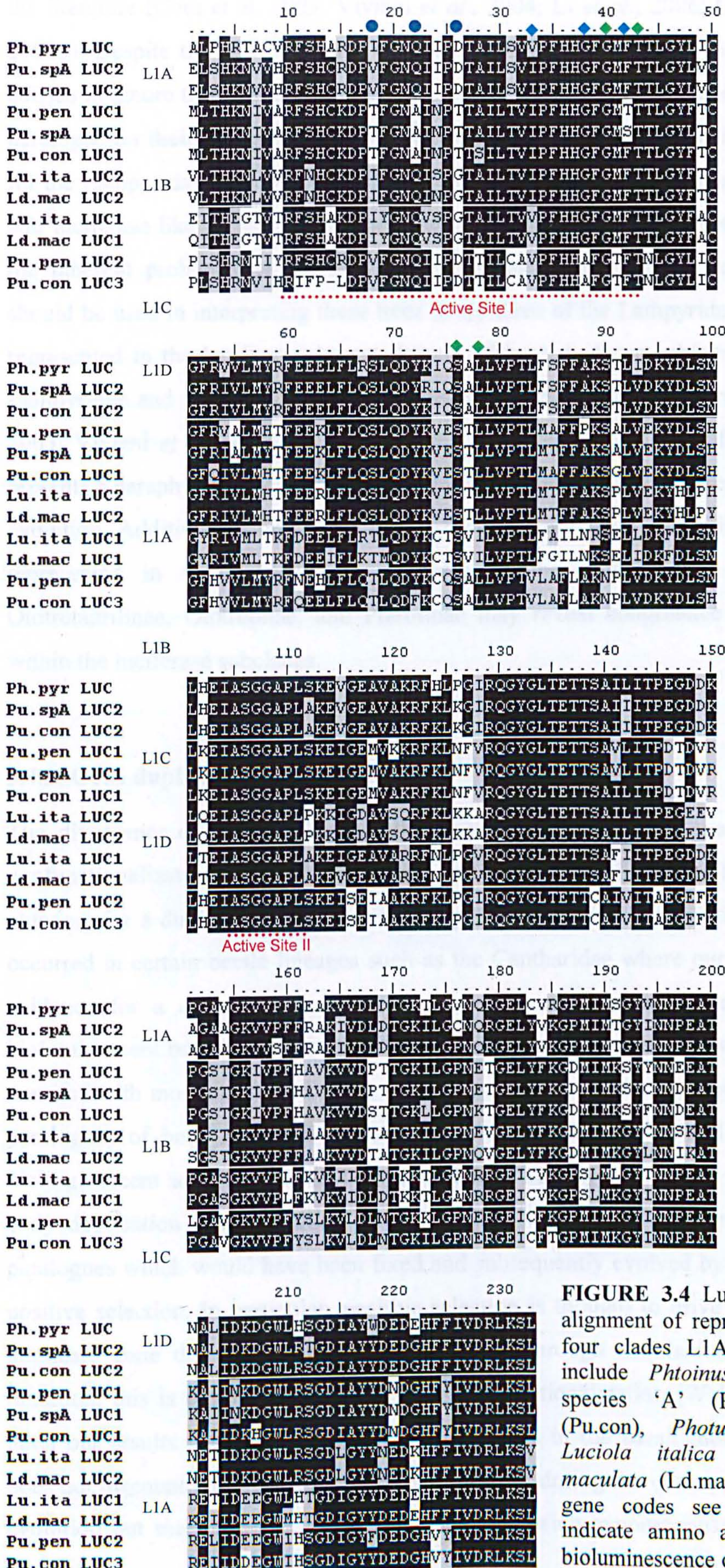


FIGURE 3.4 Luciferase amino acid sequence alignment of representative lamyprids from the four clades L1A-L1D. Representative species include *Phtoinus pyralis* (Ph.pyr), *Photuris* species 'A' (Pu.spA), *Photuris congener* (Pu.con), *Photuris pennsylvanica* (Pu.pen), *Luciola italica* (Lu.ita) and *Lampyroidea maculata* (Ld.mac). For accession numbers to gene codes see table 3.2. Green diamonds indicate amino acids considered important in bioluminescence colour, blue diamonds highlight residues generating blue-shifted mutants and blue circles putative blue-shifted residues.

A number of phylogenies reconstructed from beetle luciferase sequences are present in the literature (Choi et al. 2003; Viviani *et al.*, 2004; Li *et al.*, 2006; Oba *et al.*, 2006a, 2006b). Despite the reports of two or more luciferases in *Photuris* many authors have chosen to ignore this situation and only seen fit to select one *P. pennsylvanica* luciferase paralogue for their phylogenies. It is now evident that a single gene luciferase phylogeny for the Lampyridae is redundant and a complex evolutionary pattern exists of luciferase and luciferase-like paralogues in lampyrid genomes. The results present here illustrate the inherent problems of evolutionary inferences made with this locus and caution should be used in interpreting these trees. Only three of the Lampyridae subfamilies are represented in the luciferase dataset. Prior to this study these subfamilies, Luciolinae, Lampyrinae and Photurinae were monophyletic for the known luciferases (Choi *et al.*, 2003; Viviani *et al.*, 2004; Li *et al.*, 2006). However the addition of the current data presents a paraphyletic scenario for the Lampyridae subfamilies with regard to luciferase evolution. Additional lampyrid taxa, in particular those from subfamilies currently not represented in the dataset, i.e. members of the Amydetinae, Cyphonocerinae, Ototretadrilinae, Ototretinae, and Pterotinae may reveal congruence of species trees within the luciferase subclades.

3.4.3 Gene duplication and selection

The divergence of two paralogues after duplication can follow one of three routes, nonfunctionalization, neofunctionalization and subfunctionalization. The most expected outcome for a duplicated gene is nonfunctionalization. Nonfunctionalization may have occurred in certain beetle lineages such as the Cantharidae where our studies show no evidence for a clade A paralogue. Thus the ancestral clade A paralogue in non-bioluminescent beetles may have been become silenced by degenerative mutations as is the case with most duplicated genes (Lynch and Connery, 2000). However, individual paralogues of both A and B clades have proven functionality and exhibit either a bioluminescent activity or an acyl-CoA synthetase activity (Oba *et al.*, 2006b). Thus the early duplication event resulting in both clades would have given rise to functional paralogues which would have been fixed and subsequently evolved by genetic drift and positive selection. In particular, positive selection is thought to drive the fixation of a duplicate gene that has gained a new function through acquisition of a beneficial mutation, this is the process referred to as neofunctionalization (Walsh, 1995). Based upon our results positive selection was not evident in the paralogues examined. This does not discount the presence of positive selection during the course of the paralogues evolution but means this is no longer detectable using synonymous/non-synonymous

substitution rate comparisons given the data provided. Only part of the full ORF was analysed and there may be evidence of positive selection flanking the region investigated. Had positive selection been detected in these clades the residues under positive selection would have been identified, thereby providing important information about key amino acids and their role in bioluminescence or acyl-CoA synthesis.

Oba *et al.* (2006b) presented a putative schematic for the evolutionary descent of firefly luciferase from an ancestral fatty acyl-CoA synthetase. They speculated a single gene duplication event from an ancestral gene resulting in three paralogues one of which subsequently evolved a new function of bioluminescence. This scenario we classify as the luciferase neofunctionalization scenario and we present an alternative scenario based upon subfunctionalization illustrated in Figure 3.5.

Under the subfunctionalization model multiple functions of the ancestral gene are separated between paralogues, both of which evolve under purifying selection (Force *et al.*, 1999; He, Zhang, 2005; Lynch, Force, 2000). The detection of purifying selection, acting on selected lampyrid clades, is evident from the analysis. This suggests these paralogues result from rather ancient duplication events and their AMP-binding domain entered the phase of purifying selection long ago in their evolutionary history. Extensive purifying selection supports the model of subfunctionalization evident in the potential loss of bioluminescent activity in the *Agrypnus* lineage of the Elateridae clade and members of clade L2. However, not all paralogues are accounted for and this study can only be regarded as a first step in the understanding of beetle luciferase evolution.

FIGURE 3.5 Schematic representation of predicted evolution of luciferase and luciferase-like genes from an ancestral acyl-CoA synthetase. A. Schematic based upon Oba et al. 2006b. B. Schematic based upon present study. 'D' denotes duplication event, 'N' denotes neo-functionalization of paralogue, 'S' denotes subfunctionalization and 'Non' denotes nonfucntionalization.

4. Luciferase flanking region in the glow-worm *Lampyris noctiluca*

4.1 Introduction

Along with fireflies the European glow-worm *Lampyris noctiluca* (Linnaeus) is one of over 2000 bioluminescent beetle species belonging to the family Lampyridae. The geographical distribution of *L. noctiluca* is extensive, ranging from Portugal in the West through Europe to China in the East (Tyler, 1986). This range and also abundance probably makes the European glow-worm the most studied of all Paleartic lampyrids.

Although numerous beetle luciferases have been characterised, little is known about the regulatory region upstream of the luciferase gene (*luc*). Genomic clones coding for *luc* have been published from a number of different genera of Lampyridae but little or no investigation has extended into these flanking regions (de Wet et al., 1987; Cho et al., 1999; Choi et al., 2003). One investigation into the nature of the *luc* upstream region in the Japanese firefly *Luciola lateralis* revealed allelic variation but described no significant open reading frames (Cho et al., 1999). Although the luciferase amino acid sequence has been reported a large number of beetle species spanning three coleopteran families and extensive mutagenic studies have been conducted on the enzyme, little information has been provided on the *luc* upstream region, no neighbouring genes reported and no identification of control elements, such as promoters and enhancer sequences has been attempted.

The aim of this chapter is to present an investigation into the *luc* flanking region from the European glow-worm *L. noctiluca* in order to identify the presence of paralogous genes identified in chapter 3 upstream and downstream of the *luc* gene.

4.2 Materials and Methods

Specimen details and DNA preparation

A single adult female of the European glow-worm, *Lampyris noctiluca,* was collected from an established colony in Sevenoaks, England and stored at -70°C prior to use. Using the High Pure PCR Template Preparation Kit (Roche) total genomic DNA was extracted from the whole beetle and the majority used to construct an inverse PCR genome library using the GenomeWalker™ Universal Kit (Clontech) (see section 2.2.1).

4.2.1 PCR, cloning and sequencing

Genome walking PCR primers were designed to sequentially walk out from the *L. noctiluca* luciferase gene in both a 5' and a 3' direction. Using the *L. noctiluca* luciferase cDNA sequence previously characterised (GenBank acc. No. X89479) gene specific nested primer pairs were designed for upstream and downstream amplification (genome walking primer sequences are available from the author). Using each successive walker sequence the next primers were designed. The PCR was carried out in a 50 µL mixture containing 15 pmol of each primer (first gene specific primer and first adaptor primer), 75 mM Tris–HCl (pH 8.8), 2.5 mM magnesium chloride, 0.01% Tween 20, 1.3 M betaine, 400µM of each dNTP. *Taq-Pfu* DNA polymerases mixture (15:1 units) was used. The cycling conditions were as follows: 7 cycles of 25 sec at 94°C and 3 min at 72°C followed by 32 cycles of 25 sec at 94°C and 3 min at 67°C. The last cycle was followed by the extension step for 7 min at 67°C. 1µl of a 1:50 dilution of PCR mixture from the first round of PCR was used for second round PCR with the nested gene primer specific to the particular region and the second nested adaptor primer. PCR, the majority of the time, produced a single product that could be excised from the agarose gel and purified. In the result of multiple bands the whole PCR product was purified to remove small molecular weight products and the whole reaction cloned. The PCR products were cloned into pGEM Easy T (Promega), plasmid prepared using a plasmid mini prep kit (Qiagen) and the insert sequenced using a CEQ sequencing kit (Beckman Coulter) with M13F and M13R primers and any internal oligonucleotides where necessary.

Based up the sixth sequential walk and the *luc* sequence, primers *LnocLUC5'F* 5' AGA GAT ACG AAG ATA GAT ATG GAC ACG AC 3' and *LnocLUC5'R* 5' ATT TTT TTG CAG CGC TCT TTT GGA ACA GGA TAC 3' were designed to amplify a contiguous flanking region fragment extending over the length of six genomic walks through to the first 513 bp of *luc*. PCR amplification, cloning and sequencing was carried as described above. The fragment was primer sequenced in its entirety in both directions.

4.2.2 Sequence Analysis

Putative promoter sites were determined using the Neural Network Promoter Prediction software via the Berkeley Drosophila Genome Project website. Translations of *LnocLUC5'1* in all six frames were used in an rpsBLAST search against a conserved domain database (CDD) at the NCBI website. tBLASTn searches were used to

investigate the presence of open reading frames within flanking regions of *Luciola lateralis*.

4.3 Results

4.3.1 Luciferase gene of *Lampyris noctiluca*

Genome walking was carried out from the luciferase gene of *L. noctiluca* in both 5' and 3' directions. Only two walks out from the 3' end of the gene were successful before PCR failed to return contiguous product. However, 5' walking was successful for six overlapping walks. The initial genomic walks in both 5' and 3' directions gave sufficient sequence information to design primers (*lnocLUC F & R*) for the amplification of *LnocLUC* (GenBank accession number: AY748894), a PCR product 2960 bp in length composed of the entire *luc* gene along with upstream (540 bp) and downstream sequence (439 bp). The entire gene sequence was 1981 bp in length and showed *luc* to be composed of seven exons divided by six small introns (Figure 4.1). When compared to the *luc* sequences from two other species of Lampyridae, *Photinus pyralis* (deWet et al., 1987) and *Luciola lateralis* (Cho *et al.*, 1999) exon/intron sites in *luc* from *L. noctiluca* were found to be completely conserved in both number and position.

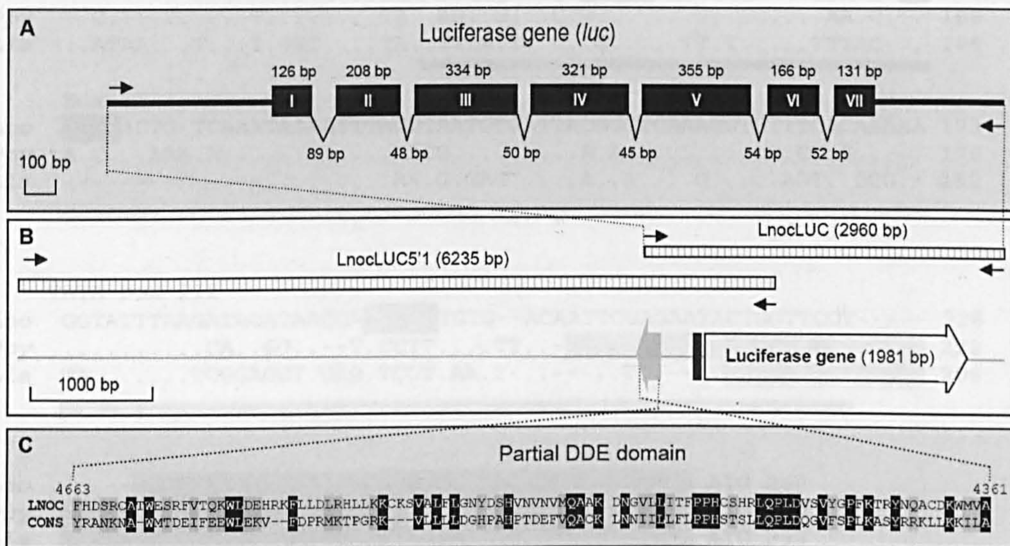


FIGURE 4.1 Graphic representation of the PCR amplification and gene architecture of *luc* and flanking region genomic DNA from *L. noctiluca*. A. Gene structure of *luc* based upon 2960 bp PCR product, sizes of exons (black boxes) and introns (inverted triangles) are shown. B. Schematic representation of 8 kb sequenced genomic region containing *luc* based upon two PCR products (shaded boxes, arrows indicate primer positions). Positions of the core promoter region (CPR) and the DDE transposase domain (DDE) are indicated. Open reading frame directions are indicated with open arrows. C. Amino acid translation of putative DDE domain from *L. noctiluca* (LNOC) aligned with the consensus sequence from pfam03184, DDE superfamily endonuclease.

Comparisons of the predicted mRNA from *LnocLUC* with the *L. noctiluca* luciferase cDNA originally identified by Sala Newby *et al.* (1996) showed eleven substitutions within the coding region and one substitution in the untranscribed regions. Extensive population variation is evident at this locus as four out of the eleven coding positions were found to be nonsynonymous mutations. Recently the luciferase gene from *L. noctiluca* was reported from a Korean glow-worm specimen (Li et al., 2003). However, it was not possible to make comparisons with the GenBank deposited sequence (AAR20794) as the sequences contained a number of errors both at the gene level and at the protein level.

A search of conserved promoter motifs found in insect genomes revealed a core promoter region (CPR) in the *L. noctiluca* sequence *LnocLUC* 32 bp upstream of the *luc* start codon (Figure 4.2). Three TATA boxes and one CAAT box were identified in this *L. noctiluca* CPR.

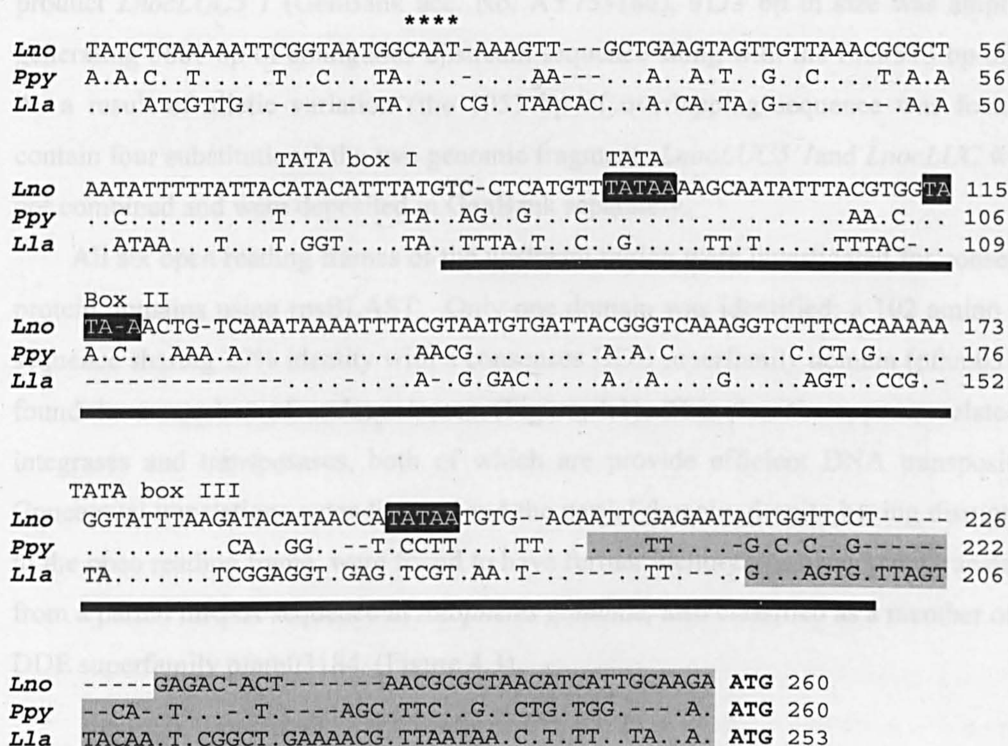


FIGURE 4.2 Alignment of the *luc* upstream sequence from *Lampyris noctiluca* (*Lno*), *Photinus pyralis* (*Ppr*, GenBank accession number: M15077) and *Luciola lateralis* (*Lla*, GenBank accession number: U49182). The ATG *luc* start codon is shown at the terminus of the alignment. Gray shaded regions indicate 5′UTR from cDNA sequences. TATA boxes are shown blocked in black and a CAAT box is highlighted with asterisks. The predicted Core Promotor Region is indicated with a bold line.

Upstream sequence of *luc* has been characterised from two other species of firefly, *P. pyralis* and *L. lateralis* (deWet et al., 1987; Cho et al., 1999). 500 bp was identified from the former and up to 1980 bp from the latter. An alignment of the first 500 bases for *L. noctiluca, P. pyralis* and *L. lateralis* reveals a conserved CPR region present within all three species but with variable TATA box sites. There is a conservation of the CAAT box between all three species and one TATA box (TATA box I) between *P. pyralis* and *L. noctiluca* (Figure 4.2).

4.3.2 Motif search in upstream region of luciferase gene

Genome walking was continued upstream of *luc* until a problematic region was encountered generating multiple sequences with core sequence homology to the previous walk. In total six unambiguous genomic walks were carried out generating approximately 7 kb of overlapping upstream sequence. Primers were designed to amplify 6173 bp of upstream region along with the start of *luc* (Figure 4.1). A PCR product *LnocLUC5'1* (GenBank acc. No. AY753186), 6173 bp in size was amplified generating 5661 bp of contiguous upstream sequence along with the first 513 bp of *luc*. As a result of allelic variation (the 1053 bp of overlapping sequence was found to contain four substitutions) the two genomic fragments *LnocLUC5'1*and *LnocLUC* where not combined and were deposited in GenBank separately.

All six open reading frames of the upstream region were investigated for conserved protein domains using rpsBLAST. Only one domain was identified: a 102 amino acid sequence sharing 29% identity with a consensus DDE superfamily domain (pfam03184) found in a number of endonucleases (Figure 4.1). This domain appears related to integrases and transposases, both of which are provide efficient DNA transposition. Conceptual translations extending beyond the partial domain, despite having disruptions to the open reading frame, were found to have further identity to a conceptual translation from a partial mRNA sequence in *Anopheles gambiae,* also classified as a member of the DDE superfamily pfam03184 (Figure 4.3).

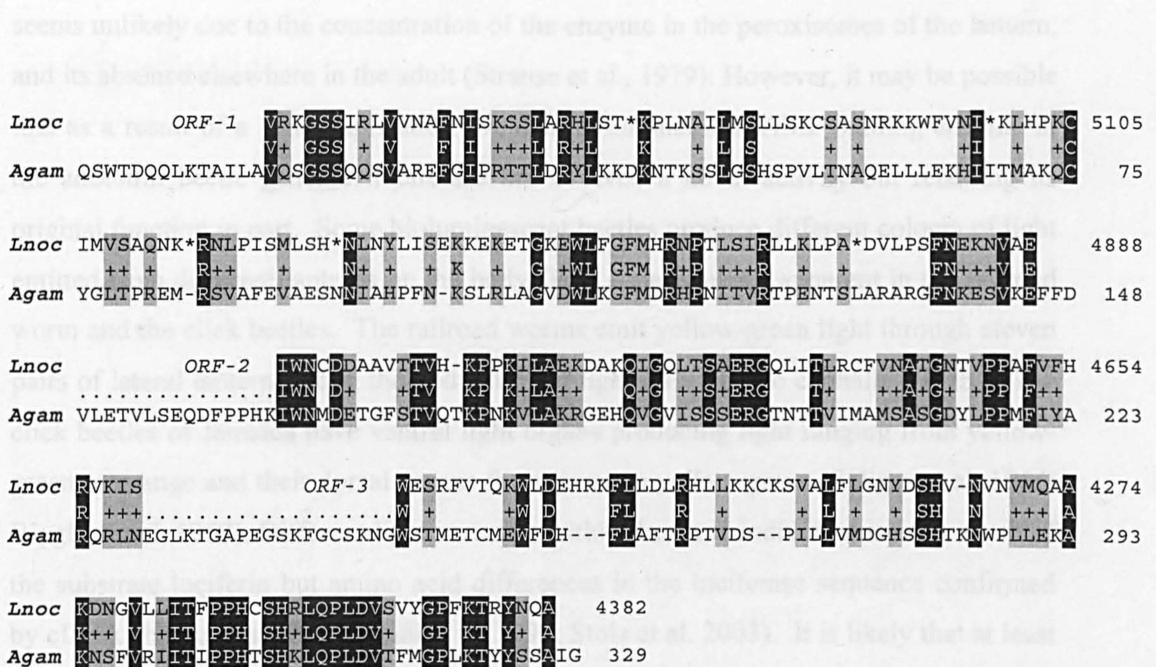


FIGURE 4.3 Amino acid sequence alignment of transposase conceptual translations. *Lampyris noctiluca* (*Lnoc*) is a composite of three conceptual translations in different open reading frames running from positions 4382 to 5283. *Agam* shows a conceptual translation of a partial mRNA sequence from *Anopheles gambiae* (GenBank accession number: XM554483).

4.4 Discussion

This partial domain in *L. noctiluca*, identified 686 bp upstream of the start codon of *luc*, presents the first evidence for a transposition event possibly occurring in bioluminescence evolution. Furthermore, tBLASTn searches of the 1980 bp upstream sequence of *L. lateralis* revealed regions with identity to transposases found in other insects (results not shown). A lack of strong identity with these transposase sequences combined with the interrupted open reading frame found in *L. noctiluca* suggests these elements are ancient and inactive but their close proximity to the luciferase gene in beetles indicates that they may have initially served to mobilise a luciferase precursor.

Based upon its catalytic properties firefly luciferase can be classified as a member of the adenylate-forming enzyme group which includes amino plant *p*-coumarate:CoA ligases, acyl-tRNA synthetases and long-chain acyl-CoA synthetases (McElroy *et al.*, 1967; Conti et al., 1996). Recently, a novel catalytic function of firefly luciferase was identified - the ability to synthesise a long-chain fatty acyl-CoA from various long-chain fatty acids in the presence of ATP, coenzyme A (CoA) and Mg^{2+} (Oba et al., 2003). *In vivo* luciferase functions as a mono-oxygenase in the bioluminescent reaction but it is not yet known whether luciferase acts as an acyl-CoA synthetase within the beetle. It

seems unlikely due to the concentration of the enzyme in the peroxisomes of the lantern, and its absence elsewhere in the adult (Strause et al., 1979). However, it may be possible that as a result of a gene duplication event from an ancestral AMP-binding enzyme in the ancestral beetle genome, luciferase has evolved a novel activity but retaining its original function in part. Some bioluminescent beetles produce different colours of light emitted from different lanterns on the body. This is particularly apparent in the railroad worm and the click beetles. The railroad worms emit yellow-green light through eleven pairs of lateral lanterns along the body and red light through two cephalic lanterns. The click beetles of Jamaica have ventral light organs producing light ranging from yellow-green to orange and their dorsal organs from green to yellow-green (Seliger et al. 1964; Biggley et al. 1967). Different light emission within the same individual is not a result of the substrate luciferin but amino acid differences in the luciferase sequence confirmed by cDNA characterisation (Viviani et al. 1999; Stolz et al. 2003). It is likely that at least two copies of the luciferase gene exist in the genomes of click beetles, railroad worms and lampyrids resulting from an ancestral duplication event enabling diverse colour production in localised lanterns. Evidence for two or three potential luciferases in lampyrids is presented in Chapter 3. The findings in this chapter suggest that this gene duplication may have been augmented by a transposition event facilitated by the putative transposase described upstream of the luciferase gene. The lack of any evidence of a paralogous luciferase gene upstream of the glow-worm *luc* gene suggests either only one gene in glow-worm (no evidence of a second paralogue in the luciferase clade was identified from *L. noctiluca*), the presence of a paralogue in the downstream region of the *luc* gene or a second paralogue present elsewhere in the genome.

5. Conclusions

5.1 The origins of bioluminescence in beetles

Bioluminescence as a phenomenon exists in a small percentage of extant beetles, less than 1%. The closest bioluminescent system is found in a small Dipteran family. Although little is known about the latter there appears to be little cross reactivity between the two and it is proposed these arose independently. Thus beetle bioluminescence would have arisen after these Orders diverged. Unless bioluminescence was an attribute of the primeval beetles it is, in accord with the neoDarwinian view of evolution, necessary to speculate the ability to emit light would have originated by mutation in the germ line of a previously non-luminous species. A minimal system would have been required; in addition to oxygen and ATP, at least a specific activating enzyme (a luciferase) and a luciferin synthetic pathway would have been present. The origination of a system this complex in one step is unlikely so it is predicted that the components of the system would have existed in the ancestral non-luminous cell prior to the final mutation resulting in the luminescent reaction. This final mutation would have been the modification of an enzyme in the luciferin biosynthetic pathway producing an excitable substrate with an already active enzyme or conversely the origination of a luciferase activity with the luciferin already synthesized in the cell. Although a great deal is already known about beetle luciferase much less is known about the synthetic pathway of luciferin. The only recognized enzyme to be associated with the latter is the luciferin regenerating enzyme (LRE). This study has focused on these two important markers involved, or implicated, in beetle bioluminescence in order to understand the evolution of this phenotype from a genotypic perspective.

The origin of bioluminescence in beetles is the basis of much discussion. Derived from morphologically based phylogenies a number of evolutionary scenarios have been presented that illustrate multiple gain and loss of bioluminescence within the Elateroidea (Bocakova *et al.*, 2007; Branham & Wenzel, 2001; Branham & Wenzel, 2003). Recent molecular phylogenies have supported the theory that bioluminescence has arisen multiple times in beetles (Hunt *et al.*, 2007).

Bocakova and co-workers (2007) established two independent origins for the evolution of bioluminescence in Lampyridae and Phenogdidae thereby arguing against a single origin of bioluminescence in elateroids. The species tree of Bocakova *et al.* provided the most comprehensive and robust phylogeny to date of the Elateroidea. However, bioluminescent elaterids were absent and no discussion of this third major bioluminescent family was made. Based upon the luciferase phylogeny, this current study has speculated a single origin of bioluminescence in an ancestral lineage of Elateroidea predating the divergence of the major bioluminescent families. This origin placed at the base of luciferase clade A (Figure 3.2) indicates a loss in bioluminescence of potentially non-functional paralogues in clade L2. It is potentially more parsimonious that this origin of bioluminescence occurred after this duplication in the Lampyridae, with bioluminescence still arising prior to the divergence of the three main bioluminescent families. This thesis presents strong evidence to support that mutations leading to a bioluminescent function in ancestral genes arose prior to the establishment of the Lampyridae, Phengodidae and most importantly the Elateridae. This latter family exhibits both bioluminescent and non-bioluminescent beetles. If proto-bioluminescence arose in the early Elateroidea it clearly has not been maintained by many members of this family. To say that these are not bioluminescent does not exclude the possibility that they exhibit one or more of the basics components required for light production. Only one non-bioluminescent elaterid has been studied so far, *Agrypnus binodulus binodulus*, from which a luciferase-like sequence was found to have no luciferase activity but functioned as a fatty-acyl CoA synthetase (Ohba et al., 2008). Based upon this sole finding a loss of bioluminescence has been added to this single branch potentially denoting a non-bioluminescent elaterid clade. However, no further paralogues have been examined in this non-bioluminescent elaterid and, as is evident in bioluminescent members, gene duplication is found in a number of species (Stolz et al., 2003; Feder & Velez, 2009). It may be seen in the future that non-bioluminescent elaterids have functional luciferases but lack luciferin biosynthetic capabilities. The converse could also be true or that both elements have been lost in non-bioluminescent elaterids.

5.2 Gene Duplication in Bioluminescent Beetles

Gene duplication and diversification is believed to be the main source or new gene function in evolution. However, it is often the case that these newly duplicated genes become non-functional pseudogenes. Intergenic exchange between related loci can

prevent pseudogenes arising. Furthermore intergenic exchange can shuffle blocks of differentiated nucleotides between paralogues to create new alleles and phenotypes. Such exchange events have been proposed to have occurred between luciferase genes in click beetles shifting the colour of the ventral light organ from green towards longer wavelengths culminating in the recent derived ventral orange allele (Stolz et al. 2003). The ratios of polymorphism to divergence (P/D) for the two paralogous genes in the click beetles were compared using McDonald–Kreitman (M–K) test for selection. For comparative purposes the gene was partitioned into colour/recombination-based region (CR) and non colour/recombination (NCR) based region. Stolz *et al.* found the M–K test was significant for the phenotypically relevant CR region but not the NCR region. However, the P/D ratios for the CR region were atypical of directional selection, which was accounted for by the basal intergenic recombination event in *P. plagiophthalamus* purging the CR region of dorsal/ventral differences and ventral polymorphism, followed by the rapid fixation of a series of selectively driven ventral mutations. It is possible that such exchange may have occurred in lampyrids enabling the luciferase paralogues from becoming pseudogenes and generating different wavelengths as is evident in the two luciferases identified in *Photuris pennsylvanica.*

For the luciferase paralogue studies in chapter 3 a different approach was employed than that used by Stoltz *et al.* Lampyrid luciferases were examined using a maximum-likelihood based analysis of synonymous and non-synonymous mutations. It was not possible to follow the M-K test as applied to the click beetle sequences as the full gene was not identified for our paralogues and it was not possible to identify the colour based sites as no expression data was available for the paralogues identified in the study. Spectral information is available for the two *Photuris pennsylvanica* luciferase paralogues and it may be possible to apply a similar approach to *Photuris* species as has been applied to *P. plagiophthalamus.* In the future it is anticipate that the third *Photuris* sequence will be fully characterised and expressed in order to provide a complete suite of luciferases from *Photuris* fireflies.

The absence of clade A paralogues in the genome of *T. castaneum* combined with the evidence of a retrotransposase upstream of the *L. noctiluca* indicates the mobilization of group A genes in ancestral Elateroidea beetles. It is possible to hypothesize the presence of an ancestral Elateroidea bifunctional gene that exhibited bioluminescent properties that has been subsequently lost in non-bioluminescent taxa, such as the Cantharidae, but maintained in bioluminescent taxa through extensive purifying selection. Luciferases have clearly diversified through gene duplication providing a basis for bioluminescent colour evolution as is evident in the Elateridae and Phengodidae. However, the limited knowledge of beetle genomes, despite the importance of this group

of insects, impedes the investigation of paralogue evolution and ultimately the emergence of bioluminescence in beetles.

The implications of gene duplication of LRE and related genes are evolutionary diversification and the attainment of new functionality in SGL paralogues. The greatest level of duplication occurs in the beetles implying those new functions arose after the main insect orders diverged. The research presented in chapter 2 identifies a large gene family of SGL paralogues in beetle genomes that share extensive conservation to the predicted active site of the bacterial gluconolactonase from *A. tumefaciens.* Beetle SGL proteins formed two principle clades, group I and group II. Lampyridae sequences were present in both groups and suggest an early duplication event in ancestral lampyrids. It is unknown whether both sequences function as LREs but prior probability analysis of LRE compared to the highly conserved group I clearly show a divergent pattern of amino acid usage and conservation among these two groups. This pattern suggests differences in selective constraints, likely arising from divergence in structural and functional aspects of the proteins phenotype. Combining the phylogenetic evidence presented and the difference in substrate specificity of LREs and SMP30 enzymes (Kondo *et al.*, 2004) a evolutionary scenario of duplication and subfunctionalization may be present in this group of genes. Thus this protein family in beetles is a suitable target to experimentally test subfunctionalization in gene copies resulting from duplication events (Bush, 2001; Winkler *et al.*, 2003). It is reiterated that protein structural information and more experimental data including functional mutagenesis studies remains to be obtained from representative members of the insect SGL protein groups described here.

5.3 The relationship between gene trees and species trees

Most luminescent groups are soft-bodied beetles and are mostly contained within the Cantharoidea. Conversely the bioluminescent Elateridae, which belong to the Elateroidea are fully sclerotized and as a result the two are considered to be distantly related. Consequently previous authors hypothesised independent origins of elaterid and cantharoid bioluminescent taxa (Crowson, 1972; Colepicolo-Neto *et al.*, 1986).

Within the cantharoid group the Lampyridae, Phengodidae and Rhagophthalmidae have often been regarded as closely related. This is illustrated by the history of species within the Rhagophthalmidae. Olivier (1910) erected the family Rhagophthalmidae for *Rhagophthalmus* Motschulsky, 1853 and the similar genera *Dioptoma*

Pascoe, 1860 and *Ochotyra* Pascoe, 1862 (now synonymized with *Rhagophthalmus*). Subsequently Olivier, without any explanation, placed *Rhagophthalmus* in the Lampyridae (Olivier, 1911). McDermott, accepting the placement, contained the members of the Rhagophthalmidae in a new subfamily of Lampyridae the Rhagophthalminae (McDermott, 1964, 1966). However Crowson in 1972 and placed Rhagophthalminae in Phengodidae which was accepted by Lawrence and Newton (1995) and supported by recent phylogenetic data (Bocakova et al., 2007). *Rhagophthalmus* research was largely ignored until Wittmer and Ohba examined this taxon and Rhagophthalmidae was again resurrected as a separate family (Wittmer & Ohba, 1994). This was supported by morphological observations (Branham & Wenzel, 2001; 2003) and luciferase gene data (Ohmiya *et al.* 2000). However, analysis of the mitochondrial 16S gene sequence (Suzuki, 1997b) and embryonic data (Kobayashi *et al.*, 2001; 2002; 2003) support McDermott's proposal that Rhagophthalmidae should be treated as a subfamily of Lampyridae. However, Suzuki excluded any members of the Phengodidae and was also dependent on a single locus for his phylogeny.

Despite the extensive revision of Rhagophthalmidae it is undisputed there is a close relationship between these three families. Consequently a single origin of bioluminescence has been proposed for these three families (Crowson, 1972). Beutel (1995) conducted a phylogenetic study of the Elateriformia and found the Cantharidae closely related to the Elateridae and the other cantharoids largely unresolved except for a clade of Phengodidae and Lampyridae. Subsequently Branham and Wenzel (2000) found evidence to support the hypothesis of two origins of bioluminescence within the Cantharoidea one for the Phengodidae and an independent origin for the Lampyridae and Rhagophthalmidae. Pototskaja (1983) was the first to consider the Lampyridae and Phengodidae as distantly related groups and present two independent origins of bioluminescence. Most recently, Bocakova *et al.* (2007) presented a molecular phylogenetic study of the Elateriformia and found that the Lampyridae and Phengodidae + Rhagophthalmidae were never monophyletic and concluded that bioluminescence as well as soft-bodiedness and neoteny all result from multiple origins and considered bioluminescence to have arisen at least four times in the Elateroidea.

No evidence of a luciferase-like sequence was found in members of the Cantharidae, a family closely related to the Lampyridae. However, the searches were not exhaustive and primer site indels or substitutions may have resulted in the bias toward group B paralogue amplification. Alternatively luciferase orthologues may truly be absent from the Cantharidae. This presents a contradiction between gene tree and species tree. A possible scenario is the loss of luciferase orthologues from the Cantharidae. Early on this family may have lost any potential bioluminescent traits. An alternative scenario

is that these families were originally more divergent that currently predicted and the current opinion is a result of subsequent convergence between these families. The range of relationships proposed by different authors does suggest there are problems inherent within the Cantharoidea as far as phylogenetic reconstruction are concerned.

The extensive information available on the luciferase gene from a large number of bioluminescent beetle species makes it an appealing candidate locus for phylogenetic studies and inferences about the relationship between bioluminescent families. However the paralogous nature of luciferase combined with the putative incomplete representation of each individual taxon in each paralogous luciferase clade limits the utility of this locus in the phylogenetic reconstruction of bioluminescent species. Consequently, in the future, alternative nuclear loci should be sourced for lampyrid phylogenetic reconstruction to supplement the mitochondrial genes commonly employed. In the past coleopteran phylogenetic studies have exploited: wingless and phosphoenolpyruvate carboxykinase (Sota & Vogler, 2001); long-wavelength rhodopsin (Zhang & Sota, 2007); Mp20 gene (Pons *et al.*, 2004); nuclear elongation factor 1α and nuclear phosphoglycerate mutase (Hughes & Vogler, 2004); and the most frequently employed of all the 18S and 28S rDNA genes (Gómez-Zurita *et al.*, 2005; Hunt *et al.*, 2007; Stanger-Hall *et al.*, 2007)

The concordance between the SGL gene tree and the genomic tree of holometabolan insects strongly supports an evolutionary history parallel to that of the organism at a deeper level. As a result LRE could be presented as a good candidate gene for deeper phylogenetic studies in insects. The restriction of luciferase and related sequences analysis to coleopteran genes limits the extrapolation of this observation to the broader syndrome of bioluminescence. For adenylate-forming proteins the scenario is considerably more complex. The latter protein family is large in number and diverse in function thereby rendering it unwieldy for phylogenetic reconstruction. For SGL these enzymes radiate from a single common ancestor whereas in the case of adenylate-forming enzymes numerous examples can be found in individual organisms. It is apparent that luciferase is most closely related to the fatty acyl-CoA synthetases, in fact luciferase exhibits this aforementioned activity *in vitro*, and a broader phylogeny could be constructed to include this group of enzymes in holometabolan insects.

Limiting the outgroup taxa to a single *Drosophila* species for the luciferase phylogeny caused problems in the rooting of the tree, with the resulting topology suggesting the divergence of clades A and B occurred prior to the divergence of Diptera from Coleoptera. To produce a more accurate reflection of deeper evolution it would have been sensible to include a broader range of taxa such as that included in the SGL phylogeny and rooting the tree with a bacteria orthologue. As only the ATP-binding

domain was used in the phylogeny alignment of this highly conserved region would not have been a problem.

5.4 Peroxisomal origin of bioluminescence

For a long time luciferase has been associated with the peroxisome organelle. With the exception of Luciolinae luciferases, all beetle luciferases have a C-terminal three-peptide signal which directs them to the peroxisomes. The absence of the peroxisomal targeting sequence in Luciolinae is unexpected given the density of the peroxisomes in photocytes and the concentration of luciferase therein. Full length sequencing of the second luciferase sequences may reveal a peroxisomal targeting sequence and ultimately present an evolutionary scenario whereby two luciferases are expressed in the lantern but only one is directed to the peroxisome; or alternatively, the original luciferase is expressed and is a part of a different cellular pathway and the second luciferase is a redundant paralogue. Full length gene characterization and RT PCR of the second *Luciola* luciferase gene will illuminate the evolution of luciferase paralogues in fireflies. The genomic proximity of luciferase paralogues will also be of interest and the evidence of retrotransposition as a mechanism of luciferase mobilization could be used in interpreting the organization of adenylate-forming enzymes in firefly genomes.
Similarly peroxisome targeting sequences are also evident in the SGL dataset, most notably the LRE's and some *T. castaneum* SGL genes. This intracellular localisation between LRE's and luciferases adds support to a bioluminescent function for LRE *in vivo.*

There is therefore strong evidence that in the first luminescent beetle species bioluminescence originated in the peroxiosome of cells. It was here that luciferase, the LRE and theoretically luciferin must have been localized to enable the bioluminescent reaction to occur. With no information on the actual mechanism of luciferin biosynthesis it is only possible to speculate that luciferin is synthesized in the peroxisome and enzymes involved may exhibit a peroxisome targeting sequence. Given a complete cDNA library of a bioluminescent beetle it may be possible to identify potential enzymes involved in the biosynthesis of luciferin. Sadly no such libraries are currently available.

5.5 The Origins of Luciferase

As currently understood, there are two types of process that could have given rise to the gene duplication of protoluciferase genes: unequal crossing-over and transposition. Transposition can give rise to gene duplication by two mechanistically distinct processes. Retrotransposition is a process whereby DNA copies of mRNAs are made and inserted into the genome, whereas active DNA transposons can potentially carry copies of genes with them when they transpose, resulting in new copies arising elsewhere in the genome. These are incorporated in a comparatively random fashion and are unlikely to be organized in tandem. Unequal crossing-over typically occurs between identical or nearly identical sequences which are predominantly arranged in a tandem array. The remnants of a retrotransposase found upstream of the *Lampyris noctiluca* luciferase gene suggest early mobilization of the bioluminescent paralogue by transposition.

The proximity of group A and group B gene paralogues is not yet established in bioluminescent beetle genomes. No evidence of a paralogue was found in 7 kb upstream of the glow-worm luciferase gene and in regions flanking the luciferase gene of *Luciola lateralis* (Cho *et al.*, 1999). The absence of group A paralogues in the genome of *T. castaneum* combined with the evidence of a retrotransposase upstream of the *L. noctiluca* indicates the mobilization of group A genes in ancestral Elateriformia beetles. Furthermore, evidence for gene translocation is the number of paralogues found on different chromosomes in *T. castaneum*, a similar scenario way be evident in bioluminescent beetles, with luciferase paralogues arranged in a tandem array on one chromosome and luciferase-like paralogues arranged on another chromosome. However, no evidence for luciferase parlaogues was evident in 7kb of upstream sequence from the *L. noctiluca* luciferase gene. However, the same range was not studied down-stream due to problems genome walking. It is therefore possible that the *L. noctiluca* luciferase gene is the first in a number of paralogues. An alternative possibility is that there has been no gene duplication of luciferase in the *L. noctiluca* genome. No evidence of luciferase paralogues were found in the glow-worm using degenerate primers. All lampyrids exhibiting luciferase paralogues were flash communicators and it is possible that the evolution of more complex lantern morphology is correlated with a gene dosage mechanism. Further studies are required in order to find strong evidence to support this observation.

Luciferase has been shown to be a bifunctional enzyme functioning as an oxygenase in the bioluminescent reaction and as a fatty acyl-CoA synthetase. Other

lampyrid paralogues have been shown to lack the former function presenting two evolutionary scenarios, either neofunctionalisation of the luciferase genes or subfunctionalisation of the luciferase-like paralogues

Subfunctionalization is most easily identifiable by the partitioning of the original expression pattern. For example, if the original gene was expressed in regions A and B, the differential loss of regulatory regions might cause one duplicate to be expressed in region A and the other in region B so that both paralogues are kept under selective pressure and become fixed in the genome. Although present in low levels throughout the larva, firefly luciferase is localised in the lanterns during pupation and almost absent elsewhere in the adult body (Strause *et al.*, 1979). It is not know yet the expression pattern of luciferase paralogues in lampyrids. However, Oba *et al.* (2006b) characterised *LucLL1* from the anterior half of a *L. cruciata* larva and luciferase from the posterior half. Furthermore, *LcLL2* was isolated and characterised from whole adults. Although not explicit in paper by Oba *et al.* differential expression patterns may be present in different life stages for the three paralogues as is evident for luciferase in *Photuris pennsylvannica* (Strause *et al.*, 1979). During *Photuris* development the larval light organ regress and are replaced by the adult lantern (Strause *et al.*, 1979). Strause and DeLuca identified a luciferase isozyme in larval *P. pennsylvanica* that is distinct from the adult luciferase (Strause & DeLuca, 1981). Thus, a third luciferase expressed in the larvae but not in the adult may exist; a hypothesis supported by the three *P. congener* luciferase genes. These cases illustrate the possibility of partitioning of the original expression profile. However, these examples are circumstantial and more detailed empirical studies are required before our proposed model of subfunctionalisation of luciferase paralogues is substantially supported.

5.6 Luciferin synthesis

An often neglected aspect of bioluminescent beetle evolution is the substrate luciferin. This is an essential component of bioluminescence and little work has been conducted upon the evolution of luciferin. Beetle luciferin is regarded to be the same structure in the Phengodidae and Elateridae as the Lampyridae (Seliger & McElroy, 1964; Viviani & Becham, 1993). It is therefore possible that luciferin biosynthesis was a mechanism present in the early Elateriformia. If multiple emergences of bioluminescence are proposed in the Elateriformia then it would be sensible to conclude that luciferin biosynthesis arose independently in these bioluminescent lineages. It seems highly

plausible to identify genetic drift as a mechanism of protein evolving a new function but less so to suggest a complete catalytic pathway has arisen independently producing the same novel phenotye. Recently the contents of firefly luciferin in luminous and non-luminous beetles have been determined by using HPLC with fluorescence detection and the luminescence reaction kinetics (Oba *et al.*, 2008b). Luminous cantharoids and elaterids contained varying amounts of luciferin but no luciferin was detected in the non-luminous cantharoids and elaterids. Such non-luminous cantharoids include members of the Cantharidae. No evidence for a luciferase sequences were found in *Cantharis* species despite the presence of luciferase-like sequences in clade B (Chapter 3). Despite the close relationship between the Cantharidae and other bioluminescent families contained within the Cantharoidea it is apparent that this family exhibits no evidence for bioluminescence. Although the relationship between the Cantharoidea families is debated it seems logical to present the hypothesis that all members of this group had protobioluminescent capabilities, the ability to synthesis luciferin and a functional oxygenase. For those taxa that are no longer exhibit bioluminescence it seems the most parsimonious evolutionary explanation is to propose non-functionalisation of genes involved in the luciferin biosynthetic pathway. An example of such evolutionary redundancy is seen in the vitamin C pathway discussed in chapter 3. An absence of luciferin in certain lineages of protocoleoptera would thereby render the luciferase activity of the early luciferases redundant and hence their absence from non-bioluminescent genomes.

5.7 Expression and functional studies

The majority of the research was focused on gene characterisation with only a small proportion of the studies evaluating the expression of these genes. For the duration of the study specimens were mostly collected in from the field and preserved in ethanol. From this material there would have been little chance of recovering good quality mRNA to amplify a full length transcript. The exception was the work conducted on the *P. p. marginata* material donated by the Oxford University Museum from a maintained colony. Group I and group II full length transcripts are now available for the non-bioluminescent beetle *P. m. peregrina* and work is ongoing to produce full length transcripts for the firefly *Photinus pyralis*.

Time restrictions prevented expression studies of the three SGL cDNA paralogues. Initial attempts at expression failed to yield a single product of the correct

size. More time and effort is required to optimize the expression and purification of these proteins before conducting functional assays. It is hoped that recently collected *Lamprohiza splendidula* and *Luciola italica* material preserved in RNAse later will enable the characterization of lampyrid SGL sequences enabling comparative studies between non-bioluminescent and bioluminescent SGL paralogues. Once expressed, it will be possible to comparatively test the activities of these enzymes, providing a basis for mutagenesis studies. Thus, it is hoped that these findings present useful foundations to further research on how evolution shaped the SGL protein family, their conserved and divergent gene functions.

A similar situation is evident with the luciferase studies. Full length characterization of the luciferase paralogues followed by the RT PCR amplification of transcripts may be attempted in the future. This will provide an insight into luciferase evolution, in particular paralogues may exhibit similar properties to those of the *Ppe1* and *Ppe2* genes in *Photuris pennsylvanica,* generating light at different wavelengths and potentially blue-shifted. It will be interesting to see if this phenomenon is present in different genera of lampyrids. Ultimately the most interesting question arising from the luciferase work is the presence of a third gene in *Photuris* which may be functional and only expressed in larvae. A collaborative project is currently being finalised to develop this area further by characterizing luciferase paralogues in the firefly *Photuris versicolor*. In the future it is hope that these functional studies and full gene molecular phylogenies, augmented by the application of molecular clocks to both luciferase and LRE trees, will provide a clearer understanding of when bioluminescence in beetles arose and how it evolved to the present day.

5.8 Contribution to knowledge of the evolution of bioluminescence in beetles

Concentrating on the genes involved in beetle bioluminescence as a direct correlation between genotype and phenotype this thesis presents a novel study into the relationship of gene duplication and the evolution of bioluminescence. This research presents new evidence showing the potential for multiple luciferases existing in a range of bioluminescent beetle taxa. This multiplicity is also reflected in the LRE scenario. A much greater level of gene duplication is a novel finding and suggests such duplications have driven the emergence of bioluminescence in beetles. Evidence that retrotransposition may have driven this is seen in the flanking region of the luciferase

gene from *Lampyris noctiluca* (Chapter 4). To date only one study has revealed different transcripts from a single lampyrid species. This study reveals such a situation is widespread across numerous lampyrid genera. Furthermore, a range of luciferase-like sequences are also evident in these species providing a solid foundation for future evolutionary studies into luciferase precursors.

This study is the first to collate all SGL sequences from available genomes and present a comprehensive phylogeny which is not only strongly supported but reflects the species tree of higher level phylogenetic relationships in holometabolan insects. This preliminary evidence that SGL evolution reflects species evolution suggests this group of genes could be an excellent candidate marker for broader evolutionary studies.

Similarly this study presents the largest dataset published to date of luciferases and related gene sequences for beetles. Prior to this study little was know about the LRE gene in beetles and its relationship to more distant orthologues. Three LRE sequences from fireflies had been studied in depth at a functional level and strong sequence identity to dipteran paralogues had been noted (Gomi *et al.*, 2002). However, no other beetle orthologues had been studied at that time. It was unknown: whether LRE was a single or multi-copy gene in firefly genomes; whether LRE orthologues were present in other beetle genomes; and the relationship of LRE to more distant orthologues.

For beetle luciferase more is know; studies have identified a range of luciferase-like orthologues in non-bioluminescent beetles as well as flies that have shown no evidence of luminescent activity but have exhibited fatty acyl CoA synthetase activity (Oba *et al.*, 2005 & 2006a). In addition luciferase-like sequences had been identified in *Luciola cruciata* (Oba *et al.*, 2006b). However, these had been seen to be quite distant to the *bona fide* luciferases and a comprehensive phylogeny of orthologues and paralogues had yet to be established prior to this study, both in fireflies and more distant beetle taxa. Such related sequences include those found in the genome of the red flour beetle *Tribolium castaneum* which has, in its complete form, only recently been available for study. This study presents a detailed evaluation of these bioluminescent enzyme precursors. This contribution is important both in providing an insight into these enzymatic precursors but also informs on the genome evolution of *T. castaneum*.

This study identified up to three luciferase paralogues in *Photuris* and two paralogues in *Luciola*. Only single luciferase genes were found for glowing species and this is the first study that has indicated there may be a relationship between lantern physiology and luciferase copy number. Consequently the findings presented in this study provide preliminary evidence for a possible model evolutionary system that links the gene copy number with physiological complexity.

5.9 Summary

With regards to the broad syndrome of bioluminescence in beetles this study illustrated such a phenomenon has originated and evolved through a series of gene duplications of existing enzymes resulting in the potential for the creation of a new function selected for in primaeval beetles. This novel function has been maintained in the current spectrum of bioluminescent taxa. Such gene duplications are relatively ancient in the history of beetle evolution but the neofunctionalisation of certain paralogues occurred more recently and probably predated the divergence of the main bioluminescent families which has been predicted to occur sometime during the Jurassic. Ancestral beetle bioluminescence would have ultimately benefited from subsequent gene duplication illustrated by the number of luciferase paralogues found in the Lampyridae. Such duplications may be correlated to the complexity of lantern physiology with at least three paralogues in the *Photuris* firefly and only one in the glow-worm. Thus with the adaptation of lantern physiology positive selection has also driven an increase in luciferase paralogues.

6. REFERENCES

Abascal F, Zardoya R, Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.

Altschul SF, Madden TL, Schaffer AA, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389-3402.

Arnold C, Hodgson IJ (1991) Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl.* **1**, 39-42.

Baldwin TO (1996) Firefly luciferase: the structure is known, but the mystery remains. *Structure* **4**, 223-228.

Barber HS (1951) North American fireflies of the genus *Photuris*. *Smithsonian Miscellaneous Collection* **117**, 1-58.

Barros M, Bechara E (1998) Bioluminescence as a possible auxiliary oxygen detoxification mechanism in elaterid larvae. *Free Radic. Biol. Med.* **24**, 767-777.

Barros M, Bechara EJH (2001) Daily variations of antioxidant enzyme and luciferase activities in the luminescent click-beetle *Pyrearinus termitilluminans*: cooperation against oxygen toxicity. *Insect Biochemistry & Physiology* **31**, 393-400.

Belt T (1874) *The Naturalist in Nicaragua*, 3rd edn. University of Chicago Press, Chicago.

Bennett-Lovsey R, Hebert A, Sternberg M, Kelley L (2008) Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins: structure, function, bioinformatics.* **70**, 611-625.

Beutel RG (1995) Phylogenetic analysis of Elateriformia (Coleoptera: Polyphaga) based on larval characters. *Journal of Zoological Systematics & Evolutionary Research* **33**, 145-171.

Birney EC, Jenness R, Ayaz KM (1976) Inability of bats to synthesise L-ascorbic acid. *Nature* **260**, 626-628.

Bitler B, McElroy WD (1957) The preparation and properties of crystalline firefly luciferin. *Arch. Biochem. Biophys.* **72**, 358-368.

Blagoderov VA, Lukashevich ED, Mostovski MB (2002) Order Diptera Linné, 1758. The True Flies.

. In: *History of Insects.* (eds. Rasnitsyn A, Quicke D), pp. 227-241. Kluwer Academic Publishers, Dordrecht.

Blank GE, Pletcher J, Sax M (1971) The molecular structure of firefly D-(-)-luciferin: a single crystal x-ray analysis. *Biochem. Biophys. Res. Commun.* **42**, 583-588.

Blum MS, Sannasi A (1974) Reflex bleeding in the lampyrid *Photinus pyralis*: defensive function. *Journal of Insect Physiology* **20**, 451-460.

Bocakova M, Bocak L, Hunt T, Teraväinen M, Vogler AP (2007) Molecular phylogenetics of Elateriformia (Coleoptera): evolution of bioluminescence and neoteny. *Cladistics* **23**, 477-496.

Bolshakov VN, Topalis P, Blass C, *et al.* (2002) A Comparative Genomic Analysis of Two Distant Diptera, the Fruit Fly, Drosophila melanogaster, and the Malaria Mosquito, Anopheles gambiae. *Genome Res.* **12**, 57-66.

Bongardt J (1903) Beiträge zur kenntnis der leuchtorgane einheimischer Lampyriden. *Zeitschrift für wissenschaftliche Zoologie* **75**, 145 pls. 141-143.

Bowie LJ (1978) Synthesis of firefly luciferin and structural analogs. *Meth. Enzymol.* **57**, 15-28.

Branchini BR (2000) Chemical synthesis of firefly luciferin analogs and inhibitors. *Meth. Enzymol.* **305**, 188-195.

Branchini BR, Ablamsky DM, Rosenman JM, *et al.* (2007) Synergistic mutations produce blue-shifted bioluminescence in firefly luciferase. *Biochemistry* **46**, 13847 - 13855.

Branchini BR, Hayard MM, Bamford S, Brennan PM, Lajiness EJ (1989) Naphthyl- and quinolylluciferin: green and red light emitting firefly luciferin analogues. *Photochem. Photobiol.* **49**, 689–695.

Branchini BR, Murtiashaw MH, Magyar RA, Anderson SM (2000) The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* **39**, 5433-5440.

Branchini BR, Murtiashaw MH, Magyar RA, *et al.* (2002) Yellow-green and red firefly bioluminescence from 5,5-dimethyloxyluciferin. *J. Am. Chem. Soc.* **124**, 2112-2113.

Branchini BR, Rollins CB (1989) High-performance liquid chromatography-based purification of firefly luciferases. *Photochem. Photobiol.* **50**, 679-684.

Branchini BR, Southworth TL, DeAngelis JP, Roda A, Michelini E (2006) Luciferase from the Italian firefly Luciola italica: Molecular cloning and expression. *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **145**, 159-167.

Branden C, Tooze J (1999) *Introduction to Protein Structure*, Second Edition edn. Garland Publishing Taylor and Francis Group, New York

Branham MA, Wenzel JW (2001) The evolution of bioluminescence in cantharoids (Coleoptera: Elateroidea). *Florida Entomologist* **84**, 565-586.

Branham MA, Wenzel JW (2003) The origin of photic behavior and the evolution of sexual communication in fireflies (Coleoptera: Lampyridae). *Cladistics* **19**, 1-22.

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. *Annual Review of Biochemistry* **50**, 349-383.

Buck JB (1978) Function and evolution of bioluminescence. In: *Bioluminescence in Action* (ed. Herring PJ). Academic Press.

Bugnion Ep (1929) *Le ver-luisant provençal et la Luciole niçoise*, Association des naturalistes de Nice et des Alpes-Maritimes (Suppl. Riviera Scientifique 1929) pp. 1-131

Burns JJ (1957) Missing Step in Man, Monkey and Guinea Pig required for the Biosynthesis of L-Ascorbic Acid. *Nature* **180**, 553-553.

Burns JJ (1960) Ascorbic acid. In: *Metabolic Pathways* (ed. Greenberg DM), pp. 341-356. Academic, New York.

Burns JJ, Moltz A, Peyser P (1956) Missing step in guinea pigs required for the biosynthesis of L-ascorbic acid. *Science* **124**, 1148-1149.

Bush RM (2001) Predicting adaptive evolution. *Nat Rev Genet* **2**, 387-392.

Chang KH, Xiang H, Dunaway-Mariano D (1997) Acyl-adenylate motif of the acyl-adenylate/thioester-forming enzyme superfamily: A site-directed mutagenesis study with the *Pseudomonas* sp. strain CBS3 4-chlorobenzoate:Coenzyme A ligase. *Biochemistry* **36**, 15650-15659.

Chatterjee IB (1973) Evolution and the Biosynthesis of Ascorbic Acid. *Science* **182**, 1271-1272.

Chaudhuri CR, Chatterjee IB (1969) L-Ascorbic Acid Synthesis in Birds: Phylogenetic Trend. *Science* **164**, 435-436.

Choi Y, Bae J, Lee K, *et al.* (2003) Genomic structure of the luciferase gene and phylogenetic analysis in the *Hotaria*-group fireflies. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **134**, 199-214.

Choi YS, Lee KS, Bae JS, *et al.* (2002) Molecular cloning and expression of a cDNA encoding the luciferase from the firefly, *Hotaria unmunsana. Comp. Biochem. Physiol. Biochem. Mol. Biol.* **132**, 661-670.

Colepicolo P, Pagni D, Bechara EJH (1988) Luciferin biosynthesis in larval *Pyrearinus termitilluminans* (Coleoptera: Elateridae). *Comp. Biochem. Physiol. B, Biochem. Mol. Biol.* **91**, 143-147.

Conti E, Franks NP, Brick P (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure* **4**, 287-298.

Cowles RB (1959) *Zulu Journal* University of California Press, Berkeley.

Creighton TE (1993) *Proteins: Structures and Molecular Properties* W.H. Freeman, New York

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Research* **14**, 1188-1190.

Crowson RA (1972) A review of the classification of Cantharoidea (Coleoptera), with the definition of two new families, Cneoglossidae and Omethidae. *Revista de la Universidad de Madrid* **21**, 35-77.

Dabrowski K (1990) Gulonolactone oxidase is missing in teleost fish. The direct spectrophotometric assay. *Biol. Chem. Hoppe-Seyler* **371**, 207-214.

Day JC, Chaichi MJ, Najafil I, Whiteley AS (2006) Genomic structure of the luciferase gene from the bioluminescent beetle, *Nyctophila* cf. *caucasica. J. Insect Sci.* **6**, 1-8.

De Cock R (2000) Rare, or simply overlooked? Practical notes for survey and monitoring of the small glow-worm *Phosphaenus hemipterus* (Coleoptera: Lampyridae). *Belgian Journal of Zoology* **130**, 93-101.

De Cock R, Matthysen E (1999) Aposematism and bioluminescence: experimental evidence from glow-worm larvae (Coleoptera: Lampyridae). *Evol Ecol* **13**, 619-639.

De Cock R, Matthysen E (2001) Aposematism and bioluminescence: experimental evidence from glow-worm larvae (Coleoptera: Lampyridae). *Evolutionary Ecology* **13**, 619-639.

De Cock R, Matthysen E (2003) Glow-worm larvae bioluminescence (Coleoptera: Lampyridae) operates as an aposematic signal upon toads (*Bufo bufo*). *Behav. Ecol.* **14**, 103-108.

de Wergifosse B, Dubuisson M, Marchand-Brynaert J, Trouet A, Rees J-F (2004) Coelenterazine: a two-stage antioxidant in lipid micelles. *Free Radical Biology and Medicine* **36**, 278-287.

de Wet JR, Wood KV, deLuca M, Helinski DR, Subramani S (1987) Firefly luciferase gene: structure and expression in mammalian cells. *Molecular and Cellular Biology* 7, 725-737.

de Wet JR, Wood KV, Helinski DR, DeLuca M (1985) Cloning of firefly luciferase cDNA and the expression of active luciferase in *Escherichia coli. Proc. Natl Acad. Sci. USA* **82**, 7870-7873.

Dehal P, Satou Y, Campbell RK, *et al.* (2002) The Draft Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins. *Science* **298**, 2157-2167.

Delarue M (1995) Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* **5**, 48-55.

DeLuca M (1969) Hydrophobic nature of the active site of firefly luciferase. *Biochemistry* **8**, 160-166.

Devine JH, Kutuzova GD, Green VA, Ugarova NN, Baldwin TO (1993) Luciferase from the east European firefly *Luciola mingrelica*: cloning and nucleotide sequence of the cDNA, overexpression in *Escherichia coli* and purification of the enzyme. *Biochim. Biophys. Acta* **1173**, 121-132.

Dong P, Hou Qi, LI X, Liang X (2008) Cloning, expression and sequence analysis of a luciferase gene from the Chinese firefly *Pyrocoelia pygidialis. Zool. Res.* **29**, 477-484

Doron-Faigenboim A, Stern A, Mayrose I, Bacharach E, Pupko T (2005) Selecton: a server for detecting evolutionary forces at a single amino-acid site. *Bioinformatics* **21**, 2101-2103.

Dreisig H (1974) Observations on the Luminescence of the Larval Glowworm, *Lampyris noctiluca* L. (Col. Lampyridae). *Entomologica Scandinavica* **5**, 103-109.

Drysdale RA, Crosby MA, The FlyBase C (2005) FlyBase: genes and gene models. *Nucl. Acids Res.* **33**, D390-395.

Dubuisson M, Marchand C, Rees J-F (2004) Firefly luciferin as antioxidant and light emitter: the evolution of insect bioluminescence. *Luminescence* **19**, 339-344.

Eisner T, Wiemer DF, Haynes LW, Meinwald J (1978) Lucibufagins: Defensive steroids from the fireflies *Photinus ignitus* and *P. marginellus* (Coleoptera: Lampyridae). *Proceedings of the National Academy of Sciences of the United States of America* **75**, 905-908.

Emamzadeh AR, Hosseinkhani S, Sadeghizadeh M, *et al.* (2006) cDNA cloning, expression and homology modeling of a luciferase from the firefly *Lampyroidea maculata*. *Journal of Biochemistry & Molecular Biology* **39**, 578-585.

Emery C (1886) La luce negli amori delle lucciole. *Boll. Soc. Ent. Ital.* **18**, 406-411.

Feder J, L., Velez S (2009) Intergenic exchange, geographic isolation, and the evolution of bioluminescent color for *Pyrophorus* click beetles. *Evolution* **63**, 1203-1216.

Felsenstein J (1985) Confidence Limits on Phylogenies: An Approach Using the Bootstrap. *Evolution* **39**, 783-791.

Feng D, Kondo Y, Ishigami A, *et al.* (2004) Senescence Marker Protein-30 as a Novel Antiaging Molecule. *Ann NY Acad Sci* **1019**, 360-364.

Fontes R, Dukhovich A, Sillero A, Sillero MAG (1997) Synthesis of dehydroluciferin by firefly luciferase: effect of dehydroluciferin, coenzyme A and nucleoside triphosphates on the luminescent reaction. *Biochem. Biophys. Res. Commun.* **237**, 445-450.

Force A, Lynch M, Pickett FB, *et al.* (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531-1545.

Fu X, Meyer-Rochow BV, Tyler J, Suzuki H, De Cock R (2009) Structure and function of the eversible organs of several genera of larval firefly (Coleoptera: Lampyridae). *Chemoecology*.

Fu X, Vencl FV, Ohba N, *et al.* (2007) Structure and function of the eversible glands of the aquatic firefly *Luciola leii* (Coleoptera: Lampyridae). *Chemoecology* **17**, 117-124.

Fujita T, Mandel JL, Shirasawa T, *et al.* (1995) Isolation of cDNA clone encoding human homologue of senescence marker protein-30 (SMP30) and its location on the X chromosome. *Biochim Biophys Acta.* **1263**, 249-252.

Fujita T, Shirasawa T, Maruyama N (1996a) Isolation and characterization of genomic and cDNA clones encoding mouse senescence marker protein-30 (SMP30). *Biochim Biophys Acta* **1308**, 49-57.

Fujita T, Shirasawa T, Uchida K, Maruyama N (1996b) Gene regulation of senescence marker protein-30 (SMP30): coordinated up-regulation with tissue maturation and gradual down-regulation with aging. *Mechanisms of Ageing and Development* **87**, 219-229.

Fujita T, Uchida K, Maruyama N (1992) Purification of senescence marker protein-30 (SMP30) and its androgen-independent decrease with age in the rat liver. *Biochim Biophys Acta* **1116**, 122-128.

Gandelman OA, Brovko LY, Ugarova NN, Chikishev AY, Shkurimov AP (1993) Oxyluciferin fluorescence is a model of native bioluminescence in the firefly luciferin--luciferase system. *J. Photochem. Photobiol. B Biol.* **19**, 187-191.

Gasteiger E, Gattiker A, Hoogland C, *et al.* (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucl. Acids Res.* **31**, 3784-3788.

Gates BJ, DeLuca M (1975) The production of oxyluciferin during the firefly luciferase light reaction. *Arch. Biochem. Biophys.* **169**, 616-621.

Gaunt MW, Miles MA (2002) An insect molecular clock dates the origins of insects and accords with palaeontological and biogeographic landmarks. *Mol. Biol. Evol.* **19**, 748-761.

Gertz EM, Yu Y-K, Agarwala R, Schaffer A, Altschul S (2006) Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biology* **4**, 41.

Ghiradella H (1998) The anatomy of light production: the fine structure of the firefly lantern. In: *Microscopic Anatomy of Invertebrates 11A, Insects,* (eds. Locke M, Harrisson F), pp. 363-381. Wiley-Liss, New York.

Ghiradella H, Schmidt JT (2004) Fireflies at one hundred plus: a new look at flash control. *Integrative & Comparative Biology* **44**, 203-212.

Goetz MA, Meinwald J, Eisner T (1981) Lucibufagins, IV. New defensive steroids and a pterin from the firefly *Photinus pyralis* (Coleoptera: Lampyridae). *Experientia* **37**, 679-680.

Goetz MA, Wiemer DF, Haynes LW, Meinwald J, Eisner T (1979) Lucibufagines. Partie III). Oxo-11- et oxo-12-bufalines, stéroïdes défensifs des lampyres *Photinus ignitus* et *P. marginellus* (Coleoptera: Lampyridae). *Helvetica Chimica Acta* **62**, 1396-1400.

Gomi K, Hirokawa K, Kajiyama N (2002) Molecular cloning and expression of the cDNAs encoding luciferin-regenerating enzyme from *Luciola cruciata* and *Luciola lateralis*. *Gene* **294**, 157-166.

Gomi K, Kajiyama N (2001) Oxyluciferin, a luminescence product of firefly luciferase, Is enzymatically regenerated into luciferin. *J. Biol. Chem.* **276**, 36508-36513.

González A, Schroeder F, Meinwald J, Eisner T (1999) N-methylquinolinium 2-carboxylate, a defensive betaine from *Photuris versicolor* fireflies. *Journal of Natural Products* **62**, 378-380.

Goto SG (2000) Expression of *Drosophila* homologue of senescence marker protein-30 during cold acclimation *Journal of Insect Physiology* **46**, 1111-1120

Green AA, McElroy WD (1956) Crystalline firefly luciferase. *Biochim. Biophys. Acta* **20**, 170-176.

Grimaldi D, Engel MS (2005) *Evolution of the Insects* Cambridge University Press, Cambridge.

Gronquist M, Schroeder F, Ghiradella H, *et al.* (2006) Shunning the night to elude the hunter: diurnal fireflies and the "femmes fatales". *Chemoecology* **16**, 39-43.

Gu X (1999) Statistical methods for testing functional divergence after gene duplication. . *Molecular Biology and Evolution* **16**, 1664-1674.

Guilford T, Cuthill I (1989) Aposematism and bioluminescence. *Animal behavior* **34**, 286-288.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696-704.

Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC (2003) The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry* **42**, 2866-2873.

Guo ZY, Tang YH, Wang S, Feng YM (2003) Contribution of the absolutely conserved B8Gly to the foldability of insulin. *Biol. Chem.* **384**, 805-809.

Hadj-Mohammadi MR, Chaichi MJ (1996) Separation, identification and determination of luciferin in the Iranian firefly, *Lampyris turkestanicus* by HPLC and spectroscopic methods. *Photochem. Photobiol.* **64**, 821-822.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.* **41**, 95-98.

Hansen IA, Gutsmann V, Meyer SR, Scheller K (2003) Functional dissection of the hexamerin receptor and its ligand arylphorin in the blowfly Calliphora vicina. *Insect Molecular Biology* **12**, 427-432.

Hansen IA, Meyer SR, Schafer I, Scheller K (2002) Interaction of the anterior fat body protein with the hexamerin receptor in the blowfly Calliphora vicina. *European Journal of Biochemistry* **269**, 954-960.

Harvey EN (1952) *Bioluminescence* Academic Press, New York.

Harvey EN, Hall RT (1929) Will the adult firefly luminesce if its larval organs are entirely removed? *Science* **69**, 253-254.

Hasan L, Vögeli P, Stoll P, *et al.* (1992) Intragenic deletion in the gene encoding L-gulonolactone oxidase causes vitamin C deficiency in pigs. *J Biol Chem.* **267**, 21967-21972.

Hastings JW (1983) Biological diversity, chemical mechanisms, and the evolutionary origins of bioluminescent systems. *J. Mol. Evol.* **19**, 309-321.

He X, Zhang J (2005) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157-1164.

Hess WN (1920) Notes on the biology of some common Lampyridae. *Biol. Bull.* **38**, 39-76.

Hubbard T, Andrews D, Caccamo M, *et al.* (2005) Ensembl 2005. *Nucl. Acids Res.* **33**, D447-453.

Hunt T, Bergsten J, Levkanicova Z, *et al.* (2007) A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* **318**, 1913-1916.

Innan H (2003) The Coalescent and Infinite-Site Model of a Small Multigene Family. *Genetics* **163**, 803-810.

Inoue H, Nojima H, Okayama H (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **96**, 23-28.

Ishigami A, Fujita T, Handa S, *et al.* (2002) Senescence marker protein-30 knockout mouse licer is highly susceptible to TNF-α- and Fas-mediated apoptosis. *Am. J. Pathol.* **161**, 1273-1281.

Ishigami A, Handa S, Maruyama N, Supakar PC (2003) Nuclear localization of Senescence Marker Protein-30, SMP30, in cultured mouse hepatocytes and Its similarity to RNA polymerase. *Bioscience, Biotechnology, and Biochemistry* **67** 158-160

Juan C, Emerson BC, Oromí P, Hewitt GM (2000) Colonization and diversification: towards a phylogeographic synthesis for the Canary Islands. *Trends in Ecology & Evolution* **15**, 104-109.

Kato S-i, Oba Y, Ojika M, Inouye S (2004) Identification of the biosynthetic units of Cypridina luciferin in Cypridina (Vargula) hilgendorfii by LC/ESI-TOF-MS. *Tetrahedron* **60**, 11427-11434.

Kleinkauf H, Von Dohren H (1996) A nonribosomal system of peptide biosynthesis. *Eur. J. Biochem.* **236**, 335-351.

Knight M, Glor R, Smedley SR, *et al.* (1999) Firefly toxicosis in lizards. *J Chem Ecol* **25**, 1981–1986.

Kondo Y, Inai Y, Sato Y, *et al.* (2006) Senescence marker protein 30 functions as gluconolactonase in L-ascorbic acid biosynthesis, and its knockout mice are prone to scurvy. *Proc. Natl Acad. Sci. USA* **103**, 5723-5728.

Kondo Y, Ishigami A, Kubo S, *et al.* (2004) Senescence marker protein-30 is a unique enzyme that hydrolyzes diisopropyl phosphorofluoridate in the liver. *FEBS Letters* **570**, 57-62.

Kortschak RD, Samuel G, Saint R, Miller DJ (2003) EST analysis of the cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Current Biology* **13**, 2190-2195.

Krzywinski J, Wilkerson RC, Besansky NJ (2001) Toward understanding Anophelinae (Diptera, Culicidae) phylogeny: insights from nuclear single-copy genes and the weight of evidence. *Syst Biol* **50**, 540-556.

Kuhnt P (1907) Das Leuchten der Lampyriden. *Entomol. Wochenblatt* **24**, 3-4.

Kutuzova GD, Hannah RR, Wood KV (1997) Bioluminescence color variation and kinetic behavior relationships among beetle luciferases. In: *Bioluminescence and Chemiluminescence: Molecular Reporting with Photons, Proceedings of the International Symposium on Bioluminescence and Chemiluminescence, 9th, Woods Hole, MA, Oct 4-8, 1996* (eds. Hastings JW, L.J K, Stanley PE), pp. 248-252. John Wiley, Chichester.

Lawrence JF (1988) Rhinorhipidae, a new beetle family from Australia, with comments on the phylogeny of the Elateriformia. *Invertebrate Taxonomy* **2**, 1-53.

Lee J (1976) Bioluminescence of the Australian glow-worm, *Arachnocampa richardsae* Harrison. *Photochemistry & Photobiology* **24**, 279-285.

Lee KS, Park HJ, Bae JS, *et al.* (2001) Molecular cloning and expression of a cDNA encoding the luciferase from the firefly, *Pyrocoelia rufa. Journal of Biotechnology* **92**, 9-19.

Lembert N (1996) Firefly luciferase can use L-luciferin to produce light. *Biochem. J.* **317**, 273-277.

Li W-H (1997) *Molecular Evolution* Sinauer Associates, Sunderland, MA.

Li X, Yang S, Liang X (2006) Phylogenetic relationship of the firefly, *Diaphanes pectinealis* based on the DNA sequence and gene structure of luciferase. *Zool. Res.* **27** 367-374

Lloyd JE (1966) *Studies on the flash communication system in Photinus Fireflies* Museum of Zoology, University of Michigan, Ann Arbor.

Lloyd JE (1971) Bioluminescent communication in insects. *Ann. Rev. Entomol* **16**, 97-112.

Lloyd JE (1972) Chemical communication in fireflies. *Environmental Entomology* **1**, 265-266.

Lloyd JE (1973) Firefly parasites and predators. *Coleopterists' Bulletin* **27**, 91-106.

Lloyd JE (1983a) Bioluminescence and communication in insects. *Annual Review of Entomology* **28**, 131–160.

Lloyd JE (1983b) Bioluminescence and communication in insects *Annual Review of Entomology* **28**, 131-160.

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459-473.

Masuda T, Tatsumi H, Nakano E (1989) Cloning and sequence analysis of cDNA for luciferase of a Japanese firefly, *Luciola cruciata. Gene* **77**, 265-270.

McCapra F (2000) Chemical generation of excited states: The basis of chemiluminescence and bioluminescence. In: *Methods in Enzymology*, pp. 3-47. Academic Press.

McCapra F, Gilfoyle DJ, Young DW, Church NJ, Spencer P (1994) The chemical origin of colour differences in beetle bioluminescence. In: *Bioluminescence and Chemiluminescence: Fundamental and Applied Aspects* (ed. Campbell AK, Kricka, L. J., Stanley, P.E.), pp. 387-391. Wiley, Chichester.

McCapra F, Razavi Z (1975) A model for firefly luciferin biosynthesis. *J.C.S. Chem. Comm.* **2**, 42-43.

McCapra FM, Perring KD (1985) Luciferin Bioluminescence. In: *Chemiluminescence and Bioluminescence* (ed. Burr JG), pp. 359-386. Marcel Dekker, New York.

McDermott FA (1917) Observations on the light-emission of American Lampyridae: The photogenic functions as a mating adaptation; 5th paper. *Canadian Entomologist* **49**, 53-61.

McDermott FA (1964) The taxonomy of the Lampyridae (Coleoptera). *Trans. Amer. Entomol. Soc.* **90**, 1-72.

McDermott FA, Buck JB (1959) The lampyrid fireflies of Jamaica. *Transactions of the American Entomological Society* **85**, 1-112.

McElroy WD, Seliger HH (1962) *Fed. Proc. Fed. Am. Soc. Exp. Biol.*, 1006–1012.

Meinwald J, Wiemer DF, Eisner T (1979) Lucibufagins. 2. Esters of 12-oxo-2.beta.,5.beta.,11.alpha.-trihydroxybufalin, the major defensive steroids of the

firefly *Photinus pyralis* (Coleoptera: Lampyridae). *Journal of the American Chemical Society* **101**, 3055 - 3060.

Meyer-Rochow VB (2007) Glowworms: a review of *Arachnocampa* spp. and kin. *Luminescence* **22**, 251-265.

Misawa H, Yamaguchi M (2000) The gene of Ca2+-binding protein regucalcin is highly conserved in vertebrate species. *Int J Mol Med* **6**, 191-196.

Moreau R, Dabrowski K (1998) Body pool and synthesis of ascorbic acid in adult sea lamprey (*Petromyzon marinus*): An agnathan fish with gulonolactone oxidase activity. *Proc. Natl Acad. Sci. USA* **95**, 10279-10282.

Moreau R, Dabrowski K (1998b) Fish acquired ascorbic acid synthesis prior to terrestrial vertebrate emergence. *Free Radic. Biol. Med.* **25**, 989-990

Mori T, Ishigami A, Seyama K, *et al.* (2004) Senescence marker protein-30 knockout mouse as a novel murine model of senile lung. *Pathology International* **54**, 167-173.

Morozov VM, N. UN (1997) Conserved motifs in superfamily of enzymes catalyzing acyl adenylate formation from ATP and compounds with carboxyl group. In: *Bioluminescence and Chemiluminescence: Molecular Reporting with Photons, Proceedings of the 9th International Symposium, pp.* (eds. Hastings JW, Kricka LJ, Stanley PE), pp. 236-239. Wiley, Chichester.

Nakajima Y, Natori S (2000) Identification and characterizatin of an anterior fat body protein in an insect. *J. Biochem* **127**, 901-908.

Nakamura M, Niwa K, Maki S, *et al.* (2006) Construction of a new firefly bioluminescence system using l-luciferin as substrate. *Tetrahedron Lett.* **47**, 1197-1200.

Nakatsu T, Ichiyama S, Hiratake J, *et al.* (2006) Structural basis for the spectral difference in luciferase bioluminescence. *Nature* **440**, 372-376.

Nei M, Roychoudhury AK (1973) Probability of fixation of nonfunctional genes at duplicate loci *Am. Nat.* **107**, 590-605.

Nene V, Wortman JR, Lawson D, *et al.* (2007) Genome Sequence of *Aedes aegypti*, a Major Arbovirus Vector. *Science* **316**, 1718-1723.

Nishikimi M, Fukuyama R, Minoshima S, Shimizu N, Yagi K (1994) Cloning and chromosomal mapping of the human nonfunctional gene for L- gulono-gamma-lactone oxidase, the enzyme for L-ascorbic acid biosynthesis missing in man. *J. Biol. Chem.* **269**, 13685-13688.

Nishikimi M, Kawai T, Yagi K (1992) Guinea pigs possess a highly mutated gene for L-gulono-gamma-lactone oxidase, the key enzyme for L-ascorbic acid biosynthesis missing in this species. *J. Biol. Chem.* **267**, 21967-21972.

Nishikimi M, Okamura M, Ohta Y (2003) Recent advances in biochemistry of L-ascorbic acid biosynthesis. In: *Recent Research Developments in Biophysics and Biochemistry* (ed. Pandalai SG), pp. 531-545. Research Signpost, Kerala, India.

Nishikimi M, Yagi K (1996) Biochemistry and molecular biology of ascorbic acid biosynthesis. In: *Subcellular Biochemistry - Ascorbic Acid: Biochemistry and Biomedical Cell Biology.* (ed. Harris JR), pp. 17-39. Plenum Press, New York.

Niwa K, Nakamura M, Ohmiya Y (2006) Stereoisomeric bio-inversion key to biosynthesis of firefly d-luciferin. *FEBS Lett.* **580**, 5283-5287.

Nor Aliza AR, Bedick JC, Rana RL, *et al.* (2001) Arachidonic and eicosapentaenoic acids in tissues of the firefly, *Photinus pyralis* (Insecta: Coleoptera). *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **128**, 251-257.

Oba Y, Iida K, Ojika M, Inouye S (2008) Orthologous gene of beetle luciferase in non-luminous click beetle, *Agrypnus binodulus* (Elateridae), encodes a fatty acyl-CoA synthetase. *Gene* **407**, 169-175.

Oba Y, Sato M, Ojika, M., Inouye S (2005) Enzymatic and genetic characterization of firefly luciferase and *Drosophila* CG6178 as a fatty acyl-CoA synthetase. *Bioscience, Biotechnology & Biochemistry* **69**, 819-828.

Oba Y, Ojika M, Inouye S (2003) Firefly luciferase is a bifunctional enzyme: ATP-dependent monooxygenase and a long chain fatty acyl-CoA synthetase. *FEBS Lett.* **540**, 251-254.

Oba Y, Sato M, Inouye S (2006a) Cloning and characterization of the homologous genes of firefly luciferase in the mealworm beetle, *Tenebrio molitor*. *Insect Mol. Biol.* **15**, 293-299.

Oba Y, Sato M, Ohta Y, Inouye S (2006b) Identification of paralogous genes of firefly luciferase in the Japanese firefly, *Luciola cruciata*. *Gene* **368**, 53-60.

Ochman H, Gerber AS, Hartl DL (1988) Genetic applications of an inverse polymerase chain reaction. *Genetics* **120**, 621-623.

Ohba N (1983) Studies on the communication system of Japanese fireflies. *Science Report of the Yokosuka City Museum* **30**, 1–62.

Ohba N, Azuma S, Nishiyama K, Goto Y (1994) Morphology, life history and behavior of the firefly, *Luciola owadai* (Coleoptera: Lampyridae) *Sci. Rep Yokosuka City Mus* **42**, 13-26.

Ohba N, Hidaka T (2002) Reflex bleeding of fireflies and prey-predator relationship. *Sci Rept Yokosuka City Mus* **49**, 1-12.

Ohmiya Y, Ohba N, Toh H, Tsuji F (1995) Cloning, expression and sequence analysis of cDNA for the luciferases from the Japanese fireflies, *Pyrocoelia miyako* and *Hotaria parvula*. *Photochem. Photobiol.* **62**, 309-313.

Ohno S (1970) *Evolution by gene duplication* Springer, New York.

Ohta T (1980) *Evolution and Variation of Multigene Families* Springer-Verlag, Berlin/New York.

Ohta T (1983) On the evolution of multigene families. *Theoretical Population Biology* **23**, 216-240.

Ohta Y, Nishikimi M (1999) Random nucleotide substitutions in primate nonfunctional gene for -gulono-[gamma]-lactone oxidase, the missing enzyme in -ascorbic acid biosynthesis. *Biochimica et Biophysica Acta (BBA) - General Subjects* **1472**, 408-411.

Okada K, Iio H, Kubota I, Goto T (1974) Firefly bioluminescence III. Conversion of oxyluciferin to luciferin in firefly. *Tetrahedron Lett.* **15**, 2771-2774.

Okada YK (1928) Two Japanese aquatic glowworms. *Transactions of the Entomological Society, London* **76**, 101-108.

Pierantoni U (1914) La luce degli insetti luminosi e la simbiosi ereditaria. *R. C. Accad. Napoli* **20**, 15-21.

Pollock DD, Zwickl DJ, McGuire JA, Hillis DM (2002) Increased Taxon Sampling Is Advantageous for Phylogenetic Inference. *Systematic Biology* **51**, 664 - 671.

Ponomorenko AG (2002) Superorder Scarabaeidea Laicharting, 1781. Order Coleoptera Linné, 1758. The Beetles. Pp. 164-176. In: *History of Insects* (eds. Rasnitsyn AP, Quicke DJL), p. xii+517pp. Kluwer Academic Publishers, Dordrecht, the Netherlands.

Poulton EB (1890) *The Colours of Animals: Their Meaning and Use Especially Considered in the Case of Insects* Kegan Paul, Trench,

Trubner and Co.

Redford KH (1982) Prey attraction as a possible function of bioluminescence in the larvae of *Pyrearinus termitilluminans* (Coleoptera: Elateridae). *Revta Bras Zool S. Paulo* **1**, 31-34.

Rees J, de Wergifosse B, Noiset O, *et al.* (1998) The origins of marine bioluminescence: recycling oxygen defence mechanisms into deepsea communication tools. . *Journal of Experimental Biology* **201**, 1211-1221.

Reger AS, Carney JM, Gulick AM (2007) Biochemical and crystallographic analysis of substrate binding and conformational changes in acetyl-CoA synthetase. *Biochemistry* **46**, 6536-6546.

Rishi AS, Nelson ND, Goyal A (2004) Genome walking of large fragments: an improved method. *Journal of Biotechnology* **111**, 9-15.

Ronquist F, Huelsenbeck J (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.

Roy SW, Gilbert W (2005) The pattern of intron loss. *Proc Natl Acad Sci U S A.* **102**, 713-718.

Russo CA, Takezaki N, Nei M (1995) Molecular phylogeny and divergence times of drosophilid species. *Mol Biol Evol* **12**, 391-404.

Said Alipour B, Hosseinkhani S, Nikkhah M, *et al.* (2004) Molecular cloning, sequence analysis, and expression of a cDNA encoding the luciferase from the glow-worm, Lampyris turkestanicus. *Biochem. Biophys. Res. Commun.* **325**, 215-222.

Sala-Newby GB, Thomson CM, Campbell AK (1996) Sequence and biochemical similarities between the luciferases of the glow-worm *Lampyris noctiluca* and the firefly *Photinus pyralis*. *Biochem. J.* **313**, 761-767.

Seliger HH (1973). In: *Chemiluminescence and Bioluminescence* (eds. Cormier MJ, Hercules DM, Lee J), p. 335. Plenum Press, New York.

Seliger HH (1975) The origin of bioluminescence. *Photochemistry and Photobiology* **21**, 355-361.

Seliger HH, McElroy WD (1964) The colours of firefly bioluminescence: enzyme configuration and species specificity. *Proc. Natl Acad. Sci. USA* **52**, 75-81.

Seliger HH, McELroy WD (1965) *Light - Physical and Biological Action* Academic Press, New York.

Shimokawa N, Yamaguchi M (1992) Calcium administration stimulates the expression of calcium-binding protein regucalcin mRNA in rat liver. *FEBS letters* **305**, 151-154

Shimokawa N, Yamaguchi M (1993) Molecular cloning and sequencing of the cDNA coding for a calcium-binding protein regucalcin from rat liver. *FEBS Letters* **327**, 251-255.

Shimomura O, Johnson FH, Haneda Y (1966) Observations on the biochemistry of luminescence in the New Zealand glowworm Arachnocampa luminosa. In: *Bioluminescence in Progress* (eds. Johnson FH, Haneda Y), pp. 487-494. Princeton University Press, Princeton.

Shimomura O, Satoh S, Kishi Y (1993) Structure and non-enzymatic light emission of two luciferin precursors isolated from the luminous mushroom *Panellus stipticus*. *J. Biolumin. Chemilumin.* **8**, 201-205.

Siebert PD, Chenchik A, Kellogg DE, Lukyanov KA, Lukyanov SA (1995) An improved method for walking in uncloned genomic DNA. . *Nucl. Acids Res.* **23**, 1087-1088.

Sivinski J (1981) The nature and possible functions of luminescence in Coleoptera larvae. *Coleopterists Bulletin* **35**, 167-179.

Stanger-Hall KF, Lloyd JE, Hillis DM (2007) Phylogeny of North American fireflies (Coleoptera: Lampyridae): Implications for the evolution of light signals. *Molecular Phylogenetics and Evolution* **45**, 33-49

Stolz U, Velez S, Wood KV, Wood M, Feder JL (2003) Darwinian natural selection for orange bioluminescent color in a Jamaican click beetle. *Proc. Natl Acad. Sci. USA* **100**, 14955-14959.

Stone I (1965) Studies of a mammalian enzyme system for producing evolutionary evidence on man. *American Journal of Physical Anthropology* **23**, 83-85.

Strause LG, DeLuca M, Case JF (1979) Biochemical and morphological changes accompanying light organ development in the firefly *Photuris pennsylvanica*. *J. Insect Physiol.* **25**, 339-348.

Stuible H-P, Büttner D, Ehlting J, Hahlbrock K, Kombrink E (2000) Mutational analysis of 4-coumarate:CoA ligase identifies functionally important amino acids and verifies its close relationship to other adenylate-forming enzymes. *FEBS Letters* **467**, 117-122.

Sumiya M, Viviani VR, Ohba N, Ohmiya Y (1998) Cloning and expression of a luciferase from the Japanese luminous beetle *Ragophthalmus ohbai*. In:

Bioluminescence and Chemiluminescence. Proceeding of 10th International Symposium (eds. Roda A, Pazzagli M, Kricka LJ, Stanley PE), pp. 433-436, Bologna, Italy.

Suzuki H, . (1997) Molecular phylogenetic studies of Japanese fireflies and their mating systems (Coleoptera: Cantharoidea). *Tokyo Metropolitan University Bulletin of Natural History* **3**, 1-53.

Suzuki H, Sato Y, Fujiyama S, Ohba N (1996) Biochemical systematics of Japanese fireflies of the subfamily Luciolinae and their flash communication systems. *Biochemical Genetics* **34**, 1996.

Suzuki N, Goto T (1971) Firefly Bioluminescence II. Identification of 2-(6'-hydroxybenzothiazol-2'-yl)-4-hydroxythiazole as a product in the bioluminescence of firefly lanterns and as a product in the chemiluminescence of firefly luciferin in DMSO. *Tetrahedron Lett* **12**, 2021-2024.

Sydow SL, Lloyd JE (1975) Distasteful fireflies sometimes emetic, but not lethal. *Florida Entomologist* **58**, 312.

Tatsumi H, Kajiyama N, Nakano E (1992) Molecular cloning and expression in *Escherichia coli* of a cDNA clone encoding luciferase of a firefly, *Luciola lateralis*. *Biochim. Biophys. Acta* **1131**, 161-165.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

Thompson JF, Geoghegan KF, Lloyd DB, *et al.* (1997) Mutation of a Protease-sensitive Region in Firefly Luciferase Alters Light Emission Properties. *J. Biol. Chem.* **272**, 18766-18771.

Thornton JW, DeSalle R (2000) GENE FAMILY EVOLUTION AND HOMOLOGY: Genomics Meets Phylogenetics. *Annual Review of Genomics and Human Genetics* **1**, 41-73.

Timmins G, Jackson S, Swartz H (2001) The evolution of bioluminescent oxygen consumption as an ancient oxygen detoxification mechanism. *Journal of Molecular Evolution* **52**, 321-332.

Tyler J (2001a) Are glow-worms *Lampyris noctiluca* (Linnaeus) (Lampyridae) distatesful? *The Coleopterist* **9**, 148.

Tyler J (2001b) A previously undescribed defence mechanism in the larval Glow-worm *Lampyris noctiluca* (Linnaeus) (Lampyridae)? *The Coleopterist* **10**, 38.

Tyler J, McKinnon W, Lord GA, Hilton PJ (2008) A defensive steroidal pyrone in the glow-worm *Lampyris noctiluca* L. (Coleoptera: Lampyridae). *Physiol Entomol* **33**, 167-170.

Tyler J, Trice E (2001) A description of a possible defensive organ in the larvae of the European Glow-worm *Lampyris noctiluca* (Linnaeus) (Lampyridae). *The Coleopterist* **10**, 75-78.

Ugarova NN, Brovko LY (2001) Relationship between the structure of the protein globule and bioluminescence spectra of firefly luciferase *Russ. Chem. Bull.* **50**, 1752-1761.

Underwood TJ, Tallamy DW, Pesek JD (1997) Bioluminescence in firefly larvae: a test of the aposematic display hypothesis (Coleoptera: Lampyridae). *Journal of Insect Behavior* **10**, 365-370.

Viviani VR, Arnoldi FGC,. Brochetto-Braga M, Ohmiya Y (2004) Cloning and characterization of the cDNA for the Brazilian *Cratomorphus distinctus* larval firefly luciferase: similarities with European *Lampyris noctiluca* and Asiatic *Pyrocoelia* luciferases. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **139**, 151-156.

Viviani VR, Becham EJH (1993) Biophysical and biochemical aspects of phengodid (railroad-worm) bioluminescence. *Photochem. Photobiol.* **58**, 615-622.

Viviani VR, Bechara EJ, Ohmiya Y (1999a) Cloning, sequence analysis, and expression of active *Phrixothrix* railroad-worms luciferases: relationship between bioluminescence spectra and primary structures. *Biochemistry* **38**, 8271-8279.

Viviani VR, Bechara EJH (1995) Bioluminescence of Brazilian fireflies (Coleoptera): spectral distribution and pH effect on luciferase-elicited colors. Comparison with elaterid and phengodid luciferases. *Photochem. Photobiol.* **62**, 490-495.

Viviani VR, Bechara JH (1997) Bioluminescence and biological aspects of Brazilian railroad- worms (Coleoptera: Phengodidae). *Annals of the Entomological Society of America* **90**, 389-398.

Viviani VR, Hastings JW, Wilson T (2002) Two bioluminescent diptera: the North American *Orfelia fultoni* and the Australian *Arachnocampa flava*. Similar niche, different bioluminescence systems. *Photochem. Photobiol.* **75**, 22-27.

Viviani VR, Okawachi FM, Scorsato V, Abdalla FC (2008) CCD imaging of basal bioluminescence in larval fireflies: clues on the anatomic origin and evolution of bioluminescence. *Photochemical & Photobiological Sciences* **7**, 448 - 452.

Viviani VR, Silva ACR, Perez GLO, *et al.* (1999b) Cloning and molecular characterization of the cDNA for the Brazilian larval click-beetle *Pyrearinus termitilluminans* luciferase. *Photochemistry & Photobiology* **70**, 254-260.

Vullo A, Frasconi P (2004) Disulfide connectivity prediction using recursive neural networks and evolutionary information. *Bioinformatics* **20**, 653-659.

Wang L, Wang S, Li Y, Paradesi MS, Brown SJ (2007) BeetleBase: the model organism database for *Tribolium castaneum. Nucleic Acids Research* **35(Database issue)**, D476-D479..

Wang Y, Gu X (2001) Functional Divergence in the Caspase Gene Family and Altered Functional Constraints: Statistical Analysis and Prediction. *Genetics* **158**, 1311-1320.

Wheeler DL, Barrett T, Benson DA, *et al.* (2005) Database resources of the National Center for Biotechnology Information. *Nucl. Acids Res.* **33**, D39-45.

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Molecular Biology & Evolution* **18**, 691-699.

White EH, McCapra FM, Field GC, McElroy WD (1961) The structure and synthesis of firefly luciferin. *J. Am. Chem. Soc.* **83**, 2402-2403.

White EH, Rapaport E, Seliger HH, Hopkins TA (1971) The chemi- and bioluminescence of firefly luciferin: An efficient chemical production of electronically excited states. *Bioorg. Chem.* **1**, 92-122.

Wielowiejski HR (1882) Studien uber die Lampyriden. *Zeitschrift fur wissenschaftliche Zoologie* **73**, 354-428.

Wienhausen G, DeLuca M (1985) Luciferases from different species of fireflies are antigenically similar. *Photochem. Photobiol.* **42**, 609-661.

Williams FX (1917) Notes on the life-history of some North American Lampyridae. *Journal of the New York Entomological Society* **25**, 11-33.

Winkler C, Schafer M, Duschl J, Schartl M, Volff J-N (2003) Functional Divergence of Two Zebrafish Midkine Growth Factors Following Fish-Specific Gene Duplication. *Genome Res.* **13**, 1067-1081.

Wood KV (1995) The chemical mechanism and evolutionary development of beetle bioluminescence. *Photochem. Photobiol.* **62**, 662-673.

Wood KV, de Wet JR, Dewji N, DeLuca M (1984) Synthesis of active firefly luciferase by in vitro translation of RNA obtained from adult lanterns. *Biochem. Biophys. Res. Commun.* **124**, 592-596.

Wood KV, Lam YA, Seliger HH, McElroy WD (1989) Complementary DNA coding click beetle luciferases can elicit bioluminescence of different colors. *Science* **244**, 700-702.

Xia Q, Zhou Z, Lu C, *et al.* (2004) A Draft Sequence for the Genome of the Domesticated Silkworm (*Bombyx mori*). *Science* **306**, 1937-1940.

Yamaguchi M, Sugai K (1981) Properties of calcium-binding protein isolated from the soluble fraction of normal rat liver. *Chem. Pharm. Bull.* **29**, 567-570.

Yamaguchi M, Yamamoto T (1978) Purification of calcium binding substance from soluble fraction of normal rat liver. *Chem. Pharm. Bull.* **26**, 1915-1918.

Ye L, Buck LM, Schaeffer HJ, Leach FR (1997a) Cloning and sequencing of a cDNA for firefly luciferase from *Photuris pennsylvanica. Biochim. Biophys. Acta* **1339**, 39-52.

Ye L, Buck LM, Schaeffer HJ, Leach FR (1997b) Cloning and sequencing of a cDNA for firefly luciferase from *Photuris pennsylvanica. Biochim. Biophys. Acta* **1339**, 39-52.

Zdobnov EM, von Mering C, Letunic I, *et al.* (2002) Comparative Genome and Proteome Analysis of Anopheles gambiae and Drosophila melanogaster. *Science* **298**, 149-159.

APPENDIX I

PARTIAL AND FULL BEETLE SGL GENE SEQUENCES: PREDICTED GENE ARCHITECTURE AND PUTATIVE OPEN READING FRAMES

Predicted exon and intron sequences are denoted in upper and lower case respectively. Primer sequences are shown underlined. Non-canonical intron/exon boundaries are shown highlighted in grey. Predicted open reading frames are shown below the gene sequence. Square brackets contain the primers used in the PCR – Primer sequences are shown underlined.

CONTENTS

A. Bioluminescent Beetles (Lampyridae)

a. *Photinus pyralis* Linnaeus, 1767
b. *Lampyris noctiluca* Linnaeus, 1758
c. *Phausis reticulata* Say, 1825
d. *Phosphaenus hemiptera* Fourcroy, 1785
e. *Lamprohiza splendidula* Linnaeus, 1767
f. *Lampyroidea maculata* Geisthardt & Day, 2006
g. *Photuris 'AC'*
h. *Photuris congener* LeConte, 1851

B. Non-Bioluminescent Beetles

a. *Cantharis rufa* Linnaeus, 1758
b. *Cantharis rustica* Fallén, 1807
c. *Stegobium paniceum* Linnaeus, 1758
d. *Anacaena sp.* Thomson, 1859
e. *Necrophorus vespillo* Linnaeus, 1768
f. *Grammoptera ruficornis* Fabricius, 1781
g. *Strangalia melanura* Fabricius, 1792
h. *Plagiodera versicolora* Laicharting, 1781
i. *Tillus elongatus* Linnaeus, 1758
j. *Anthocomus fasciatus* Linnaeus, 1758
k. *Pyrochroa serraticornis* Scopoli, 1763
l. *Tenebrio molitor* Linnaeus, 1758
m. *Pachnoda marginata* (Drury 1773) ssp. *peregrina* Kolbe, 1906

Bioluminescent Beetles (Lampyridae)

a. *Photinus pyralis* Linnaeus, 1767

i. PpLRE

```
ATGGGGCCAGTTGTTGAAAAAATTGCAGAACTTGGCAAGTATACGGTTGGAGAAGGTCCTCACTGGGATCATGAAACTCA   80
 M  G  P  V  V  E  K  I  A  E  L  G  K  Y  T  V  G  E  G  P  H  W  D  H  E  T  Q

GACCTTATATTTCGTCGACACCGTAGAGAAAACTTTTCATAAATATGTACCTTCTCAGAAAAAATACACGTTTTGTAAAG  160
  T  L  Y  F  V  D  T  V  E  K  T  F  H  K  Y  V  P  S  Q  K  K  Y  T  F  C  K

TAGgtaagaatcattgttcaaactatcacgcgtaaccgtttttacataagtacaatgtaaagtagctacgtacaccgtgg  240
V

atagtagcattgtatttataaacgaaatacaatggtagcattgtatttatatagtaggcactttatagttagaaataggg  320

gaagtattagccaagagtatacagggtggtgagttgcgtaacacattatagggatttcaatgtaaaattcataacatgcg  400

attattggctccccaaattatttagagcaaaaatgcattaaggtttttttgtagaaaaacatttttttacatcaatcaacc  480

gtattagttttaaaattttaagcaagtcaccaccctgtatatactctttggtattagtacaagatcccacataaggccaa  560

cctcgctcatctgtttactgagtgaaatacatctttgttagtaatcttttatgtgtaaacaaacacactaaaaaacttaa  640

gtcaagactatccttaaaatttctgtctgcttctattagatagagaggagacacacggaacaacacacgtttctctcccg  720

cgcgtttcttaaaattacgacaaccgtgtgaaacaggacgttactcgtgacggcggtgtacgcattaatgttaagagagt  800

gcggcgcactgaattttaattgagggtagtcttgacttaatttttttaggcaaccatttgcaatgtgtttgattacatat  980

taaaattgctaaaacatgtaattcacttggtaaacagtagtcggtagggcgaataatgctgctatctattttctcttcta 1060

tagcttattgagctgagtatgaagggtagaatagcatttagattagctaaagaatggcgtgagaatttaacccatctgta 1140

cataaactgaagttgatgccatttcagATAAACTGGTTTCATTCATTATTCCCCTTGCTGGATCCCCTGGCCGTTTTGTA 1220
                             D  K  L  V  S  F  I  I  P  L  A  G  S  P  G  R  F  V

GTCAGTTTGGAACGTGAAATAGCCATTCTCACGTGGGATGGCGTTAGTGCTGCACCTACGAGCATAGAAGCTATTGTTAA 1300
 V  S  L  E  R  E  I  A  I  L  T  W  D  G  V  S  A  A  P  T  S  I  E  A  I  V  N

TGTCGAACCCCACATTAAAAATAACAGACTCAATGATGGCAAAGCAGATCCTCTTGGCAATCTATGGACAGgtataactg 1380
  V  E  P  H  I  K  N  N  R  L  N  D  G  K  A  D  P  L  G  N  L  W  T

aaatcgactaaccaaattataataattctagctcgaattctttagGTACAATGGCTATTGACGCTGGTCTCCCCGTAGGA 1460
                                               G  T  M  A  I  D  A  G  L  P  V  G

CCGGTCACTGGCAGTTTATATCATTTAGGGGCTGATAAAAAGGTAAAAATGCACGAGAGCAACATAGCTATAGCAAATGG 1540
 P  V  T  G  S  L  Y  H  L  G  A  D  K  K  V  K  M  H  E  S  N  I  A  I  A  N  G

GCTCGCGTGGAGTAATGATTTGAAGAAAATGTATTATATTGATTCGGGAAAAAGAAGAGTAGACGAGTACGATTATGATG 1620
  L  A  W  S  N  D  L  K  K  M  Y  Y  I  D  S  G  K  R  R  V  D  E  Y  D  Y  D

CTTCTACATTATCCATCAgtaaacactatttaaatatattgcactatttttttgttagatgaatattttcagGCAATCAAC 1700
A  S  T  L  S  I                                                         S  N  Q

GGCCATTATTTACTTTTGAAAAGCATGAAGTGCCTGGATATCCAGATGGTCAAACAATTGATGCGGAGGGTAATTTATGG 1880
R  P  L  F  T  F  E  K  H  E  V  P  G  Y  P  D  G  Q  T  I  D  A  E  G  N  L  W

GTTGCCGTTTTCCAAGGACAGCGAATTATTAAAATCAGTACCGAACAACCGGAAGTGTTACTGGATACCGTAAAAATACC 1960
 V  A  V  F  Q  G  Q  R  I  I  K  I  S  T  E  Q  P  E  V  L  L  D  T  V  K  I  P

AGATCCTCAGgtaaattaagtaactttgtgtattacattaattacctttggagtcagGTCACCTCTGTTGCATTTGGCGG 2040
  D  P  Q                                                   V  T  S  V  A  F  G  G

TCCGAATTTGGATGAACTGTATGTAACATCTGCTGGTCTTCAGCTTGACGACAGTTCTTTTGACAAAAGTTTAGTTAATG 2120
  P  N  L  D  E  L  Y  V  T  S  A  G  L  Q  L  D  D  S  S  F  D  K  S  L  V  N

GGCACGTCTACAGAGTAACAGGTTTAGGCGTCAAAGGTTTCGCGGGAGTTAAAGTGAAGCTATGA
```

```
G H V Y R V T G L G V K G F A G V K V K L *
```

ii. PpLRE2 [DEGLRE F & R]

```
TTGGAGAGGGTCCACATTGGGATATTCCTACTCAAACCCTCTACTATGTGGACATATTGGGCCAAGCCCTTCATAAAAAT   80
                          I  P  T  Q  T  L  Y  Y  V  D  I  L  G  Q  A  L  H  K  N

GTACCTTCAACTGATACTCACGCAAAGGTTAGAATAGgtatgcgcctataataattttggggtaagatatttcgagtgca  160
 V  P  S  T  D  T  H  A  K  V  R  I

atttcagaagGAGGGCCGATTGGTTTTGCAGTACCAGTCGAAGGGAAAAAGAACACGTTTGTAATTGGGCTTGGTCGAAA  240
          G  G  P  I  G  F  A  V  P  V  E  G  K  K  N  T  F  V  I  G  L  G  R  K

AATTGAGGAAATTGTCTGGGACGGCGTCAGCAGTCATGTTCAACACACCAAAACGCTCGCAGCCGTTGATAACGAAGAGG  320
 I  E  E  I  V  W  D  G  V  S  S  H  V  Q  H  T  K  T  L  A  A  V  D  N  E  E

GTTTTACTAATAACAGATTTAACGATGGGAAAGCAGATCCCACGGGGAGACTGTGGCAGgtatgggacagctgcctaaa  400
G  F  T  N  N  R  F  N  D  G  K  A  D  P  T  G  R  L  W  A

tgttaaagtaacctaaagctttagGAACGATGGGCCCTGAACCTGTGGTCGGACAGCTTGAGCCATATAAAGGCATATTG  480
                        G  T  M  G  P  E  P  V  V  G  Q  L  E  P  Y  K  G  I  L

TACACTCTGGATTGTAACCAAGTCAAAGCCCATTTGAAAACCATAAGTATCTCAAATGGGCTCGCTTGGAACGTTCAGCT  560
 Y  T  L  D  C  N  Q  V  K  A  H  L  K  T  I  S  I  S  N  G  L  A  W  N  V  Q  L

GAAGAAGATGTACTACATTGACTCCCCGCTCAGAACCGTCGACCAATTTGACTATGACATGGTCAAAGGCGAAATATgta  640
 K  K  M  Y  Y  I  D  S  P  L  R  T  V  D  Q  F  D  Y  D  M  V  K  G  E  I

acagatcctcattctactgtaattgtttctaataaattacttctagGCAACAAAAAGGTCATTTTTGACTTTGAAAGGAA  720
                                              C  N  K  K  V  I  F  D  F  E  R  N

TAACATCCCTGGTATACCCGATGGAATGACTATCGACAGTGATGGTAATTTATGGGTTGC
 N  I  P  G  I  P  D  G  M  T  I
```

iii. PpPsG [DEGLRE F & R]

```
GGAGAAGGTCCACATTGGGATAACGAATCGCAAAGTTTATACTGGGTGGATATCGTAAAGAAGACGATTCATAAATACAT   80
                     N  E  S  Q  S  L  Y  W  V  D  I  V  K  K  T  I  H  K  Y  I

ACCATCTCTTAAAAAGCATACTCATATAAAAGTGGg[illegible]aagtagatcaaattatttaatacgtaactatgcctatctttac  160
 P  S  L  K  K  H  T  H  I  K  V

ttatacaatcaccaaaatttgagaaaaatgtttacccaacctacaggagtggatatctatgtgaattactttcagATAAA  240
                                                                           D  K

TTTCCATCTCTAATTGTGCCCATAAGTGGTTCGTCTGATCGTTTCATTATTACTTTGCAACGAGAAGTAGCCGTCCTTAC  320
F  P  S  L  I  V  P  I  S  G  S  S  D  R  F  I  I  T  L  Q  R  E  V  A  V  L  T

CTGGGATGGTGTTAGTTCCAGACCTACTAGTATAGAAACGGTCGCTATTGTTGATACCGACAACGCTTTAAAAAATAACA  400
 W  D  G  V  S  S  R  P  T  S  I  E  T  V  A  I  V  D  T  D  N  A  L  K  N  N

GAATAAATGATGGCAAAGTGGACCCTCTTGGTAACTTATGGGCAGgtaactgaaatagaactgaaagataagggaaggtt  480
R  I  N  D  G  K  V  D  P  L  G  N  L  W  A

aaatacgcacgttctctttcagGCACTATGGCCATTGACGGCGATACACCATACGGTCCAAGAACGGGCACTCTATTTTG  560
                      G  T  M  A  I  D  G  D  T  P  Y  G  P  R  T  G  T  L  F  C

CTTGAGCTCCAATAAAAAAATCAGATAYTTTGAAGAAAATGTGGGCATATCAAACGGGCTTGCTTGGAGTAGCGATTTAA  640
 L  S  S  N  K  K  I  R  Y  F  E  E  N  V  G  I  S  N  G  L  A  W  S  S  D  L

AAAAAATGTATTACATTGATTCGGTTAAAAGAAGAATAGATCAAWACGACTTTGATGCAACTAACGCGTCTATCAgtacg  720
K  K  M  Y  Y  I  D  S  V  K  R  R  I  D  Q  Y  D  F  D  A  T  N  A  S  I

cattaccacatcgttttccataactatattgtttatttatagCTAATCGTCAACCACTATTTAGTCTGGAAAAGCATAAC  800
                                          T  N  R  Q  P  L  F  S  L  E  K  H  N

ATTCCAGGCTTTCCGGATGGCCAAACTATTGATGCKGATGGKAATTTATGGGTWGC
 I  P  G  F  P  D  G  Q  T  I
```

b. *Lampyris noctiluca* Linnaeus, 1758

LnPsG [DEgSMPf4 –R1]

```
GGAGAAGGACCTCATTGGGATGATAAGACACAAAGTTTATATTTGGTGGACCTTGTAGATAAATCTATTCACAAATATGT   80
                      D  K  T  Q  S  L  Y  L  V  D  L  V  D  K  S  I  H  K  Y  V

ACCATCTCTTAAAAGACATACCCATATAAAAGTAGgtaagtattattacaacttgttatatccgagtcaaccaaataatg  160
  P  S  L  K  R  H  T  H  I  K  V

taaaaaatttacgtcagtcccactataagatttaagaaattcataaatataaatttctatgctcattttcagATAAAATA  240
                                                                        D  K  I

CCGTCTTTGATTATTCCCATAAGCGATTGTTCAAATCGTTTTCTTATTACTTTAGACCGAGAAATCGCCATCCTTACTTG  320
 P  S  L  I  I  P  I  S  D  C  S  N  R  F  L  I  T  L  D  R  E  I  A  I  L  T  W

GGATGGCATTAGTTCTACACCCACAAGTATAGAAACAATCGCTGTAGTTGACAATGAGCCAGGTCTTGAAAATAATAGAA  400
  D  G  I  S  S  T  P  T  S  I  E  T  I  A  V  V  D  N  E  P  G  L  E  N  N  R

TTAATGATGGTAAAGCAGATCATCTTGGTAATCTGTGGGCAGgtgaatgaagaaatctataaaactataacatgaaatca  480
I  N  D  G  K  A  D  H  L  G  N  L  W  A

aacgcaccttttcagGCACCATGGACGTAAATGGATATCCAACGGGACCAATAACCGGTAGATTGTTTAGTATAAATTC  560
                G  T  M  D  V  N  G  Y  P  T  G  P  I  T  G  R  L  F  S  I  N  S

TACTAAAGACGTTAGATACTACAGAACAAACATAGCGGTGTCAAACGGAATCGCTTGGAGTAAAGATTTGAAGAAAATGT  640
  T  K  D  V  R  Y  Y  R  T  N  I  A  V  S  N  G  I  A  W  S  K  D  L  K  K  M

ATTATATTGATTCCATGACTAAAGTAATAGATCAATACGATTTTGATGCTTCCAACAGATTGATCAgtatacagtaagtg  720
Y  Y  I  D  S  M  T  K  V  I  D  Q  Y  D  F  D  A  S  N  R  L  I

ataaacattcgttacatatttatgttgtatattttcagGTAATCGTCAAACCTTGTTTAGTCTTGAAAATAATCAAATCC  800
                                      S  N  R  Q  T  L  F  S  L  E  N  N  Q  I

CGGGATTTCCAGATGGTCAAACAATTGATACTGATGATAATTTATGGGTCGG
P  G  F  P  D  G  Q  T  I
```

c. *Phausis reticulata* Say, 1825

PrLRE [LamLRE1.1 F & R]

```
GGAGAAGGTCCTCACTGGGATGACGAAACACAAAGTTTATATTTGTGGACATCTTAGGAATGTCTATTCATAAATATGT   80
                            T  Q  S  L  Y  F  V  D  I  L  G  M  S  I  H  K  Y  V

ACCATCTACAAAGAAACATACACATGTTAAACTTGgtaagaattattattattacaatcacaatacaattttatacagtg  160
  P  S  T  K  K  H  T  H  V  K  L

gttagccattattacggttacaaatggggcaacaatcatgtttactgtgtgaacggaggattgccatttaactattaaga  240

ggcctcaaaactaatttggtataggtaatacaggatcgtcgaaggtacattttaaaatattcaggtgatcaacaaaaa  320

tatgccaaaccaaacttgtggttttaaaatagaatacctagtatatttgcgcatttatttaagtactgtcacattcaga  400

accaaaaatgtaataaaagttgtcggtatcaccaactgttctgaaaatattgtcattttggagtcacgaggtcagggctg  480

taatcaatttaaatctgtttataccaccgagcttaatactaaaaatgatacctcgaaatatattaaataggtacgtacct  560

acataccaattaggtatgaaatttttattatatctgcgtcacttgggtttgtaaaactgaccgttattccattataaatt  640

aataacacctatgaaattattgcagATAAAAAGGTTTCATTCATTGTTCCTGTAAAGGGGCATTCAGACCGTTTTGTGAT  720
                         D  K  K  V  S  F  I  V  P  V  K  G  H  S  D  R  F  V  I

CAGTATGGAAAGAGAAATAGTTATGATTACTTGGGATGGAGTTAGTTCTACTCTTGGTAAAACAGAAACGATTGCTATAG  800
  S  M  E  R  E  I  V  M  I  T  W  D  G  V  S  S  T  L  G  K  T  E  T  I  A  I

TTGATGAAGATTATGAAACTAATAGAATCAATGATGCAAAGGTCGATCCCTTGGGCAACCTATGGGCAGgtaatgtctaa  880
V  D  E  D  Y  E  T  N  R  I  N  D  A  K  V  D  P  L  G  N  L  W  A

acaaataatacgaaatataaataaagttatagtttttttaagGTACCATGGCCACAGACGCTGATCATGTAAAAGGTACA  960
                                          G  T  M  A  T  D  A  D  H  V  K  G  T

CATATAACTGGTAGCTTGTATAATTTAGGATCAGATAAACAAGTTAAGAAACACCTTAGTAATGTTTGTGTATCAAACGG 1040
 H  I  T  G  S  L  Y  N  L  G  S  D  K  Q  V  K  K  H  L  S  N  V  C  V  S  N  G

TCTTGCTTGGAGCAAAGATTTGAAAAAAATGTATTACATAGATAGTTTGCTTAGGCGAATTGACCAATTTGATTATGATT 1120
  L  A  W  S  K  D  L  K  K  M  Y  Y  I  D  S  L  L  R  R  I  D  Q  F  D  Y  D

CCAAAACATTGTCAATTTgtaagtgtagatagcaaaaacttagcaaggtactaacacaatattcggtagCAAGTCGTCAA 1200
S  K  T  L  S  I                                                     S  S  R  Q

ACGCTGTTTACTTTTGAAAAGCATCATGTTGAGGGTTACCCCGACGGACAAACAATTGACGCGGATGGTAATTTATGGGT 1280
 T  L  F  T  F  E  K  H  H  V  E  G  Y  P  D  G  Q  T  I  D
```

d. *Phosphaenus hemiptera* Fourcroy, 1785

PhPsG [[DEgSMPf4 –R1]

```
TTGGAGAAGGCCCGCATTGAGATGAAACAACGCAAAGTTTATATTTCGTAGATACAATTATGAAAACCGTCCACAAATAC   80
                       E  T  T  Q  S  L  Y  F  V  D  T  I  M  K  T  V  H  K  Y

ATCCCATCACTCAAGAAGCATACTTATATCAAATTCGgtaagtattatttaattcggcgtagaatattctttcacgttga   160
 I  P  S  L  K  K  H  T  Y  I  K  F

tggtcttttttagATAAATTCCCATCCATAATTATACCAATAAGGGGATTTTCCGATCGATTTATCATTACTTTAAATCGA   240
              D  K  F  P  S  I  I  I  P  I  R  G  F  S  D  R  F  I  I  T  L  N  R

GAAATAGTTATCCTTACGTGGAATGGTATTAGTTCTACACCTACAGATTATGACACAATTGCTGTGGTTGATACTGCACT   320
 E  I  V  I  L  T  W  N  G  I  S  S  T  P  T  D  Y  D  T  I  A  V  V  D  T  A  L

CGATCTTCAAAATAACAGAATTAATGATGGTAAAGCAGACGTTTTTGGGAACCTCTGGGCAGgtaaggttgatttgccga   400
  D  L  Q  N  N  R  I  N  D  G  K  A  D  V  F  G  N  L  W  A

cttaacaacaatataacactgttgctcagGCTCTATGTCCATTGATGCTGATCACCCAACGGCACCACAAATGGGCACTT   480
                                G  S  M  S  I  D  A  D  H  P  T  A  P  Q  M  G  T

TGTTTCGCATGAACCCTAATAAACAAATTACATATTATGAAAAAAACGTAGGCGTATCAAACGGAATTGCTTGGAGTGGC   560
L  F  R  M  N  P  N  K  Q  I  T  Y  Y  E  K  N  V  G  V  S  N  G  I  A  W  S  G

GATTTGAAGAAAATGTATTACATCGATTCGCTGGCTAAAAGAATAGACCAATTTGATTTTGATGCGAATAACGCGTCAAT   640
 D  L  K  K  M  Y  Y  I  D  S  L  A  K  R  I  D  Q  F  D  F  D  A  N  N  A  S  I

CCgtatggaataaaaaagctattcttattttatcgttacatcatatatttttagGTAATCGACAACCACTGTTTACTCTT   720
                                                         R  N  R  Q  P  L  F  T  L

CAAAAACACAATATTCCTGGATATCCTGATGGACAAACAATTGATACCGATGGTAATTTATGGGTGGC               788
 Q  K  H  N  I  P  G  Y  P  D  G  Q  T  I
```

e. *Lamprohiza splendidula* Linnaeus, 1767

LsLRE [LRE1.1 F&R]

```
GGAGAAGGTCCTCACTGGGATCACGAAAGGAGATGTTTATATTTTGTGGACATCCCAAAGAAAACTATTCATAAATATAT   80
                         R  R  C  L  Y  F  V  D  I  P  K  K  T  I  H  K  Y  I

ACCAAATATACAAAAACATACCTACCTAACATTAGgtaagtaattgtgaaagcatatggagaagtattataggtacctat   160
  P  N  I  Q  K  H  T  Y  L  T  L

atcataaaattacattaggatcaattaatacagtaaatatatagtacctacctagtaagtaagtctatagtttgttctat   240

ttgatccttggttattgtatgaacctgcctgtatgtgtggtaggtaattattttaaaatgttttcagACAAATCTTCCTC   320
                                                                    D  K  S  S  S

ATTGATTATTCCTGTGAGTGATCATTCTGACCAGTTTCTAATCACGTTAGAACGAGAAATAGCTGTTCTAACATGGGACG   400
  L  I  I  P  V  S  D  H  S  D  Q  F  L  I  T  L  E  R  E  I  A  V  L  T  W  D

GTTTTAGTTCAACTCCGGGAAATGTAAAAGTAATCGCAGTCATGGATACTGAACCCGGTCTTGAGACAAATAAAATTAAT   480
G  F  S  S  T  P  G  N  V  K  V  I  A  V  M  D  T  E  P  G  L  E  T  N  K  I  N

GACGGCAAGGCAGATGCTTTCGGTAACTTATGGGCAGgtgatatgagttttaattgacaacattgcatcaaaagtaacca   560
 D  G  K  A  D  A  F  G  N  L  W  A

catacattcgctaccagGTTCCGTAGGAAGCGATAGTGATCTGAAAAGAGGAGCATTGAAGGGTAGCTTATASAGTCTTG   640
                   G  S  V  G  S  D  S  D  L  K  R  G  A  L  K  G  S  L  X  S  L

ATTGTTGTAAACAACTTAAACGCCATGATACAAATATAGGAGTATCTAATGGAATTACATGGAGTAAAGATCTAAAGAAT   720
D  C  C  K  Q  L  K  R  H  D  T  N  I  G  V  S  N  G  I  T  W  S  K  D  L  K  N

ATGTATTACATTGATTCTTTCACAGGAAGGGTAGATCAATATGATTTTGACGCATCTAGCAATTCCCTCAgtgagaagta   800
 M  Y  Y  I  D  S  F  T  G  R  V  D  Q  Y  D  F  D  A  S  S  N  S  L

tttaatgaacattaattagagttacactgaaattgtttagGTAATCGCAAGCCATGGTTCACTTTTGCAACGCACAACAT   880
                                          S  N  R  K  P  W  F  T  F  A  T  H  N  I

CCCTGGGTTCCCAGATGGTCAAACTAGTGATGCCGATGGTAATTTATGGGTRC
  P  G  F  P  D  G  Q  T  S  D
```

f. *Lampyroidea maculata* Geisthardt & Day, 2004

LmLRE [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGCCCGCATTGAGACACTGAATCTCAAAGTTTGTATTTTGTAGATATTCCAGCAAAATCCGTACACAAATAT  80
                            T  E  S  Q  S  L  Y  F  V  D  I  P  A  K  S  V  H  K  Y

GTACCATCTACAAAGCAGCACACCAAGATCGTTTTTGgtaagttgttttatcactttcgcactttattaaatccgtactt 160
 V  P  S  T  K  Q  H  T  K  I  V  F

tagAAAAGTCCCCAACTTTTATAATACCTGTAAAGGGATCCTCTGACCGTTTTGTAATAAGTTTGCAGCGATATATTTGT 240
   E  K  S  P  T  F  I  I  P  V  K  G  S  S  D  R  F  V  I  S  L  Q  R  Y  I  C

GTTATCACTTGGGATGGAGTTAGTAGCAAACCAAGTCATGTAGAAACCATCGCTACAGTTGATACTCATCCTGGAGGAGA 320
 V  I  T  W  D  G  V  S  S  K  P  S  H  V  E  T  I  A  T  V  D  T  H  P  G  G  E

GGAAAACTCGTTAAATGACGCTAAAGTGGATGCTTTTGGAAATTTATGGGCAGgtatttgcaataatggttgtaaatatt 400
  E  N  S  L  N  D  A  K  V  D  A  F  G  N  L  W  A

gtttagacgcatacttttagGTACATTAAGTACTAAGGTAGATTTGGAGAAAGCATCACCAATAACTGGAAGCATATATT 480
                       G  T  L  S  T  K  V  D  L  E  K  A  S  P  I  T  G  S  I  Y

GTGTATCTAATAAGCAGTTAAAGAAATGCGTATCGGGTGTATGCGTATCGAATGGATTTGCTTGGAGTAAAGATTCTAAA 560
C  V  S  N  K  Q  L  K  K  C  V  S  G  V  C  V  S  N  G  F  A  W  S  K  D  S  K

AAGTTGTATTTCATTGATACTGTCAAGAAAACTGTTGATCAGTTTGATTATGATTTTGAAAATTTAGCAATATgtaagta 640
 K  L  Y  F  I  D  T  V  K  K  T  V  D  Q  F  D  Y  D  F  E  N  L  A  I

tctatgtaaaaacaatattaagtgcattacaaaaagtcttgtttcagCAAATTGCCAGCCACTTTTTTCTTTTAACAAAC 720
                                                 A  N  C  Q  P  L  F  S  F  N  K

ATGGAATTCTGGGTTATCCCGATGGACAAACAGTAGATACTGATGGTAATTTATGGGTGGC 781
H  G  I  L  G  Y  P  D  G  Q  T  V
```

g. *Photuris* sp. 'AC'

ACLRE [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGCCCTCATTGGGACGCAGACACTCAAAGTCTGTATTTTGTTGACATTCCCCAACACACTATTCATAAATAC  80
                            A  D  T  Q  S  L  Y  F  V  D  I  P  Q  H  T  I  H  K  Y

ACTCCATCAACTAAGCAACACGTTAGTGTAAAGATGGgtaagtcacgttagttaagtttttacaaattttgataacctttt 160
 T  P  S  T  K  Q  H  V  S  V  K  M

agtcaagttgtcataaaaaccttgacagtttctatatttttattttaaagaatatttcgattctacatttacattgttac 240

ttcgagttgtaatagtttggtttgttacagATAAATTAACTTCTTTTATAATTCCAATTAAAGGATTTTCCGATAAATTC 320
                                 D  K  L  T  S  F  I  I  P  I  K  G  F  S  D  K  F

GTTGTTGGAATGTGTGATGAAATTAACATTATTACTTGGGATGGCGTCAGTTCCCAACTGACCAACATAGAAGCACTTAC 400
 V  V  G  M  C  D  E  I  N  I  I  T  W  D  G  V  S  S  Q  L  T  N  I  E  A  L  T

AAAATTGAACAATGCCATCACAAAATTTAATGATGCAAAAGCAGATTCATTAGGACATTTATGGACAGgtatatactttt 480
  K  L  N  N  A  I  T  K  F  N  D  A  K  A  D  S  L  G  H  L  W  T

atataagaacgtttgtattatatttgttgctataacatttactaaaatacaaaacattttcttactattactcttttaag 560

GATCTGTGGCAAGCGATTTTGATTTAAAAACATTTACACCATCAATTGGCTCTTTGTACAGTTTAGGATCAGACAACAAA 640
G  S  V  A  S  D  F  D  L  K  T  F  T  P  S  I  G  S  L  Y  S  L  G  S  D  N  K

ATAAAAACGCATGCTGAAAATGTTGGAATGTCAAATGGTATTGCTTTTAGCAACGATTTGAAAAAGCTTTACTATATTGA 720
 I  K  T  H  A  E  N  V  G  M  S  N  G  I  A  F  S  N  D  L  K  K  L  Y  Y  I  D

TAGTGCTTTAAGAACAGTAGATCAATTTGATTACAACGTTCAAGATCAAACAGTTTgtaagtttataacaaacttaacac 800
  S  A  L  R  T  V  D  Q  F  D  Y  N  V  Q  D  Q  T  V

tatattttattacaacatttattagCAAACCGGCAAACACTCTTTACTTTAAAAAAAAATCACCTTGAAGGAGTTCCTGA 880
                          S  N  R  Q  T  L  F  T  L  K  K  N  H  L  E  G  V  P  D

TGGATTAACAATTGATACGGATGGTAATTTATGGGTGGC
  G  L  T  I
```

h. *Photuris congener* LeConte, 1851

PcLRE2 [DegLRE F & R]

```
TTGGAGAGGGTCCACATTGGGACATCCCTTCACAAAGTCTTTACTACGTTGATATATTTGGTCAAACCCTTAATAAATAC   80
                      I  P  S  Q  S  L  Y  Y  V  D  I  F  G  Q  T  L  N  K  Y

GAACCAGCAACTAAAACTCATACAAAAGTCAAAATAGgtaaggtatactttaaattagaatgtatcttatttagattttt  160
 E  P  A  T  K  T  H  T  K  V  K  I

agaagGGGGGCCCGTGACTTTTGCTGTACCAGTAGAAGGAAAAACTAATTATTTTATCATTGGACTTGGTCGAAAAATTG  240
    G  G  P  V  T  F  A  V  P  V  E  G  K  T  N  Y  F  I  I  G  L  G  R  K  I

CTGAAGTTATATGGGATGGCATAAGCAATGCAGTTTCTCATCTTGAAGTACTTTTAGAAGTCGATAATGAAGCAGGATAT  320
A  E  V  I  W  D  G  I  S  N  A  V  S  H  L  E  V  L  L  E  V  D  N  E  A  G  Y

TCTAATAATAGATTTAACGACGGAAAAGCTGATCCAACAGGAAGATTATGGGCAGgtaattgacgttaatcattgtttat  400
 S  N  N  R  F  N  D  G  K  A  D  P  T  G  R  L  W  A

ttgtatt-aatttaatttagGTACAACGGGTCCTGAGCCTGAAATTGGAAAACTGGAACCTGAAAAAGGTTCGCTTTACA  480
                    G  T  T  G  P  E  P  E  I  G  K  L  E  P  E  K  G  S  L  Y

CTTTAATTAGTAGACATGGCGTTAAAAAGCATTTAACAAAAGTAAGCATTTCAAATGGATTGGCATGGAACGTCAAACTT  560
T  L  I  S  R  H  G  V  K  K  H  L  T  K  V  S  I  S  N  G  L  A  W  N  V  K  L

AAAAAAATGTACTACATTGATTCACCATTGAGAACGGTTGATCAGTATGACTATAACATGGCTAAGGGAGAAATATgtaa  640
 K  K  M  Y  Y  I  D  S  P  L  R  T  V  D  Q  Y  D  Y  N  M  A  K  G  E  I

gtacctgctattttccattttacatctgaaaaagtgttatttccagCTAATAAAAAAGTCATATTCGATTTCGATAAGAA  720
                                              S  N  K  K  V  I  F  D  F  D  K  N

TGATATTCCTGGGGTACCTGACGGTATGACAATTGACAGTGATGGTAATTTATGGGTGGCTGTTT                 784
 D  I  P  G  V  P  D  G  M  T  I
```

C. Non-Bioluminescent Beetles

a. *Cantharis rufa* Linnaeus, 1758

CsfSMP [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGCCCTCATTGGGACGCCAAAACCCAAACTCTATACTACGTCGATATTTTTGGAGAATCCATACATAAATAT   80
                        A  K  T  Q  T  L  Y  Y  V  D  I  F  G  K  S  I  H  K  Y

GTACCGGCGACAAATTTTCACAGTAAAGCAAACCTAGgtaagaattttccatttatcaaaatacttaataaatcagaggt  160
 V  P  A  T  N  F  H  S  K  A  N  L

cgtcaacatttaagcttgaccatggtaaccaatatattttctcattaatagtttattgttaattgtatattacgaaataa  240

tcgtgataatctcaataataaattattatcaataaaataattattttatgatgtgataacgtagaaataacaataaaca  320

atgattggataacctttaaaaaaatgaccagatttcgataataaaaagaaaattattgagatttttttttaaattaaaa  400

ttccgtaaaatgaagtaaaagtttctaaaaaattatgcaaaattaaatttattttcacttggatgaatagaaagaaattt  480

aaaaaagacttaaaaataatgtactctttgttcaataytcctagcatttaatattatcaacattctttgtgttactttga  560

aaatcattaaaaatttaaatatgccttggtacattaaatatttaaaaaaagtttttacttaaagctcaaatgagggtatt  640

tcgatggtatcttatttacttattgaaaatgttgccggattattcgtttatcaacctataatggagttggagttccattt  720

agtggcttaaaatttaaattatccacttgaaattttaattaagcagcaataggctcattattcaatttttttaattaagt  800

tctaaatatttgtcgtttatcattcgttagcacaattttcaatattgaactgaaaccgggattgtatttcaaagtacttt  880

gttgttaagaattcgtattatttactttcacttttgtaaagaatgcatttagttgtaaaggatccaaattaagcggccgt  960

gcggagagacaataaaatagttgattaaaaaaacttcaaagaaaatctgtcgtcaagaagtatttaattttacttaaaat 1040

acaaaccttatttgtttaatttaaatttaattaattgctaattgtttaatttagGGGCGCACGTGGGGTTTATCATTC 1120
                                                       G  A  H  V  G  F  I  I

CAGTGGAAGGGAGCGATGATCGATTTGTTATAAGTATGGGCCGAGAAATCGTTTCGGTTAAGTGGGATGGAGTTAGCACA 1200
P  V  E  G  S  D  D  R  F  V  I  S  M  G  R  E  I  V  S  V  K  W  D  G  V  S  T

AGTGTTTCTGAAATAGAAAAAATTGCTGAAGTTGATACAAACTTGGATGGCAATAGAGTTAATGACGGAAAGGCCGATCC 1280
 S  V  S  E  I  E  K  I  A  E  V  D  T  N  L  D  G  N  R  V  N  D  G  K  A  D  P

TACAGGAAAACTTTGGGCAGgtagcaaaaacaacgtgaaatgacctaactttctgtaatcgtaattgagtaataaattta 1360
  T  G  K  L  W  A

tagaatgcttcgacttaaaatatcctaatttgaagttgttacattaattcataataacgattttaatttagataacgtaa 1440

caatatacgaggttacttcaaattaaaaaaacatttttttattctcacacgaactgatactggggtgagattgacccgccc 1520

catttttttgttttaaagctcccagttcaaaagagttgtttttttttatttgaattgacttggtatttcttatattaaattt 1600

taaaatgttatgacgattaaagGTACAATGTCAATCGAAAAAAACGGTAAAATTTTGAAAACCACCGGATCTTTCTACAG 1680
                      G  T  M  S  I  E  K  N  G  K  I  L  K  T  T  G  S  F  Y  S

TTTTGGAAAAAAAAATCAAGTTAAAAAACATCTGTCGAACGTGCATATTTCTAATGGTTTGGCTTGGAGTGCAGACGCTA 1760
  F  G  K  K  N  Q  V  K  K  H  L  S  N  V  H  I  S  N  G  L  A  W  S  A  D  A

AAAAATTCTATTACATCGATTCTGGAGAAGGCAGGGTTGACCAGTTTGATTATGATAAGGATACGGAAACAATTTgtaag 1840
K  K  F  Y  Y  I  D  S  G  E  G  R  V  D  Q  F  D  Y  D  K  D  T  E  T  I

cttttacgttatcagtgcagaaatttttactgtattcttttttttaagCAAACCGTCAACCGCTCTTTACCTTAGCGAAA 1920
                                                A  N  R  Q  P  L  F  T  L  A  K

CACAATATCGATGGTTTTGCCGATGGACAAGCTATCGATACCGATGGTAATTTATGGGTAGC                   1982
 H  N  I  D  G  F  A  D  G  Q  A  I
```

b. *Cantharis rustica* Fallén, 1807

i. CsSMP [DegSMPF4 – DegSMPR1]

```
TTGGGGAAGGCCCGCATTGAGACGCCAAAACTCAAACTTTATACTATGTCGACATTTTTGGGAAATCTATACATAAATAC   80
                        A  K  T  Q  T  L  Y  Y  V  D  I  F  G  K  S  I  H  K  Y

GTACCGGCGACGAATTTTCACAGTAAAGCAAACCTAGgtaagagttttcaatttttagacttgatctaaaattaattaat  160
 V  P  A  T  N  F  H  S  K  A  N  L

aaattagaagtcttcaacatttaaacttgatcttggtaaccaatttatttattcattaattgtttgttgttaattacacg  240

ttatattatgaaatatgttgtgataatattaataatgaattactatcaatgaaataattatttttgtgttacgataaaca  320

gtaatcggataacctttaaggatgaccagattacaataagcaaataagaattactgagattttctttaaagtaaaattc  400

cataaaataaaatgcaataagtaacagaagtttcgaaaaaattatgcaaaattaaatttattttcacttggatgaataga  480

aagaaatttaaaaaagacttaaaaaataatttactctttgttcaatattcctaacgtttaatattatcaacattctttgt  560

gttactttgaaaatcattaaaaatttaaatgtgcgttgctatattaaatatttagaatttttttctacttaatgcccaaa  640

ttggggtatttactgattaaccacagcaccacctgaatacctacaagggagatggtatcttatttacttattgaaaatgt  720

tgccggattattcgattagtggcttaaaatttaaattatccactttaaattttaattaaacagcaataggctcattatt  800

caatttttaattaagttctaaatatttgtcgtttatcattcgttagcacaattttcaatattgaactgaaaccgggatt  880

gtatttcaaagtactttgttgttaagaattcgtattatttactttcacttttgtaaagaatgcatttagttgtaaaggat  960

ccaaattaagcggccgtgcggagagacaataaaatagttgattaaaaaaacttcaaagaaaatctgtcgtcaagaagtat 1040

ttcattttacttaaaatacaaacctatttaattaattgctaattgtttaattttagGGGAGTACGTGGGGTTTATAATTC 1120
                                                         G  E  Y  V  G  F  I  I

CAGTGGAAGGGAGCGATGATCGATTCGTTATAAGTATGGGCCGAGAAATCGTTTCGGTTAAGTGGGACGGAGTTAGCACA 1200
P  V  E  G  S  D  D  R  F  V  I  S  M  G  R  E  I  V  S  V  K  W  D  G  V  S  T

AGTGTTTCTGAAATAGAAAAAATTGCTGAAGTTGATACAAACTTGGATGGCAATAGAGTTAATGACGGAAAGGCCGATCC 1280
 S  V  S  E  I  E  K  I  A  E  V  D  T  N  L  D  G  N  R  V  N  D  G  K  A  D  P

TACAGGAAAACTTTGGGCAGgtagcaaaaacaacgtgaaatgacctaactttctgtaatcgtaattgagtaataaattta 1360
  T  G  K  L  W  A

tagaatgcttagacttaaaataatcctaattttaagttgttacattaatttataataaagattttaatttagataacgtt 1440

acaatgtacgaagttatttcaaattaaaaaaaaacattttgaaaacaacttttctactaaaataaaacagatttcaacgt 1520

atattaggggtgagattgaccgccgcctaatttttttgttttaaagcttccgatccagaagaattaagtttttttatttgaa 1600

ttggcttcgtatttcttatgttcactaatttaaaaatgttctgacgtttaaagGAACAATGTCAATCGAAAAGGAAGGCA 1680
                                                     G  T  M  S  I  E  K  E  G

AAATTCTGGGAACTACGGGGTCTTTCTACAGTTTTGGAAAAAAGAATCAAGTTAAAAAACATCTGTCTAACGTGCATATT 1760
K  I  L  G  T  T  G  S  F  Y  S  F  G  K  K  N  Q  V  K  K  H  L  S  N  V  H  I

TCTAATGGTTTGGCTTGGAGTGCAGACTTAAAAAAATTCTATTACATTGATTCTGGCGAGGGCAGGGTAGACCAGTTTGA 1840
 S  N  G  L  A  W  S  A  D  L  K  K  F  Y  Y  I  D  S  G  E  G  R  V  D  Q  F  D

TTATGATAAGGATACAGAAACAATTTgtaagtctttacattgttcgttctgaattttgactgtaactctttttcaagC 1920
  Y  D  K  D  T  E  T  I                                                    A

AAACCGTCAACCGCTTTTTACCTTGGCAAAACACAAAATTGATGGTTTTGCCGATGGACAAGCTATCGATACCGATGGTA 2000
  N  R  Q  P  L  F  T  L  A  K  H  K  I  D  G  F  A  D  G  Q  A  I

ATTTATGGGTGGC                                                                    2013
```

ii. CsPsG [DegSMPF4 – DegSMPR1]

```
TGGAGAAGGCCCCCATTGAGACGCCAATACACAAACTTTGTACTTTGTTGATATGCTTGGAGAAACAATACATAAATACG  80
                              A  N  T  Q  T  L  Y  F  V  D  M  L  G  E  T  I  H  K  Y

TGCCTGCAACAAATCAACACTGCAAAGCCTACATGAGtggtaatttttcattaaaaactagtttatcattcatctctac 160
V  P  A  T  N  Q  H  C  K  A  Y  M

agttttgcttacttttatttacgtataccattgggctcaagtgttctaatgcaaaaattacaattttatataaatctt 240

gcagGTAAAGAAGTGAGTTTTATAATTCCAGTAGAAGGAAGCAATGATCGATTTATATTAAGTATGGGCAAAGAACTCGT 320
    G  K  E  V  S  F  I  I  P  V  E  G  S  N  D  R  F  I  L  S  M  G  K  E  L  V

CTCAGTAAAATGGGATGGGGTTAGTTCAAGTGTTTCTGAAATACAAAAAATTGTTAAGGTTGAACAAAACTTGCCCAACA 400
 S  V  K  W  D  G  V  S  S  S  V  S  E  I  Q  K  I  V  K  V  E  Q  N  L  P  N

ATAAATTTAACGATGGCAAAGCTGTTCCTACAGGAAAAGTTTGGGCAGgtaagtaaaaattgattaaatttaatttaaat 480
N  K  F  N  D  G  K  A  V  P  T  G  K  V  W  A

attggtcattagggggggggggggtcattttcacaataacgcacgtccgagttaattacaaaagttgtacaaaattatat 560

ggtatccatttaaataatttaaatatctttagaagtctgttttcacaaatgccgatttaattatgttttcaaattttatg 640

cctaactatttcacatgttttgggttacacttttacgtagcgtaaattacggtgtcgtctgcgtatcgcatcacgtgaat 720

tagaattttatacaaacatttaatatcacaggtaaagcctcccagttttatttttttttatcactttagagcagttccaa 800

aatggaacaatacgcatctgttgatttattaaattgtattttgcgattcttagcatttacgcattttggcaaattgta 880

tttaatgtgatgtttacttttatattgaaagGCACAACAGCAAAACTAGAGAATGGAAAAATTAACGCAATCACTGGAAC 960
                                 G  T  T  A  K  L  E  N  G  K  I  N  A  I  T  G  T

ATTGTATAGTTTTAGTGGAGCTAATCAAATTAAAAAACATGTTTCTAACATCCATATGGCAAATGGTATGGCATGGAATG 1040
 L  Y  S  F  S  G  A  N  Q  I  K  K  H  V  S  N  I  H  M  A  N  G  M  A  W  N

CAGAACTAAAAAAATACTATTATATTGATTCTGGTCAAGGCACAATCGATCAATTTGATTATGACCAAGATACAGAAACA 1120
A  E  L  K  K  Y  Y  Y  I  D  S  G  Q  G  T  I  D  Q  F  D  Y  D  Q  D  T  E  T

ATTTgtaagtatttatccaatttgttggttcaatctgaaaacaaaaacaaacttaaattttttttagCTAATCGTCAAAT 1200
 I                                                                    S  N  R  Q  I

ATTTTTTACCTTGGGCAAACATAATGTAGATGGTGCAGCAGATGGACAGGCAATCGATACGGATGGTAATTTATGGGTGG 1280
 F  F  T  L  G  K  H  N  V  D  G  A  A  D  G  Q  A  I

C
```

c. *Stegobium paniceum* Linnaeus, 1758

SpSMP [DegSMPF4 – DegSMPR1]

```
TGGAGAGGGTCCTCATTGGAGCCAAAATCGAAAACTTTGTATTATGTGGAGCTCATGGAGGGAACCGTCAACAAATATGC   80
                        Q  N  R  K  L  C  I  M  W  S  S  W  R  E  P  S  T  N  M

GCCTTCTTCTGGAGAGCAAACTCAAGCCGCAGTCGGTACgtcttttaaaaaaaataatagtgatgaatttattagaaaaa  160
R  L  L  L  E  S  K  L  K  P  Q  S  V

tgtaatttctaagGAAAACACGTGTCGTTCATCATCCCAATAAAGGGAACAACTAACGAATTTGTGGTCGGGATTCAGCG  240
                G  K  H  V  S  F  I  I  P  I  K  G  T  T  N  E  F  V  V  G  I  Q  R

ACAATTAGCCAGAATTACTTGGGATGGAGTAAGCGAGAAACCATCAAAAATTGAACAATTGCTGGAATTAGATGATCCCT  320
 Q  L  A  R  I  T  W  D  G  V  S  E  K  P  S  K  I  E  Q  L  L  E  L  D  D  P

CAGATGTTATTCAATTGAATGACGCAAAAGTGGACAGTTTCGGTAGATTGTGGTTCGGTACTGTAGCTTTCGACCGCCAG  400
S  D  V  I  Q  L  N  D  A  K  V  D  S  F  G  R  L  W  F  G  T  V  A  F  D  R  Q

AACCACAGATGGCTTGCCAACAAGGCATCGTTCTACAGTTATGCGAAAAAAGAAGGCCTCAAAACCCATCTGGACAACGT  480
 N  H  R  W  L  A  N  K  A  S  F  Y  S  Y  A  K  K  E  G  L  K  T  H  L  D  N  V

TACCATCTCAAATGGAATGGATTGGGACGTAAAGAGGAAGAAGTTTTATTACATCGACTCACCTCAGAGGCAAATCTTTC  560
 T  I  S  N  G  M  D  W  D  V  K  R  K  K  F  Y  Y  I  D  S  P  Q  R  Q  I  F

AATACGACTTCGATGGGAGTGAAGGAAAAAATATgtaagtaccttccggaaacttttctagaaaaccaacgaaaaccaaac  640
Q  Y  D  F  D  G  S  E  G  K  I

atccacagACAACCAACAAACGATATTCACCCTGGACAAACACGACATTCCGGGAATCCCCGATGGGCTAACCATCGACG  720
          Y  N  Q  Q  T  I  F  T  L  D  K  H  D  I  P  G  I  P  D  G  L  T  I

CGGATGGAAATTTATGGGTAGCTGTTT
```

d. *Anacaena sp.* Thomson, 1859

i. AnSMP1 SpSMP [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGTCCACATTGGGACGAGGCAACTCAAACTTTGTACTACATCGATATACTTGGAAAATCGATCCATAAATAT
                        E  A  T  Q  T  L  Y  Y  I  D  I  L  G  K  S  I  H  K  Y

GTACCTGCAAGCAACACTCATACTAAAGCAGTCCTTGgtaattaaatatttgtctagattagcacacatcttttatttc
 V  P  A  S  N  T  H  T  K  A  V  L

atatacaaaaatttaatacctacattcatatttgtaataatatgcttcattcgttttacagAAAAAAATGTGGCTCTTA
                                                                  E  K  N  V  A  L

TCGTACCTGTAGCTGGAACTAAGGACAAATTCTTAATTTCTCAAGACCGAGATCTCGCCATTGTAACATGGGATGGAGTA
I  V  P  V  A  G  T  K  D  K  F  L  I  S  Q  D  R  D  L  A  I  V  T  W  D  G  V

AGTCCAAAGGCAAGTAACATTAAAAAAATTGCAGAAGGTGATACAGCTCCTGGTCTTGAAGGAAACAGGTTCAATGATGG
 S  P  K  A  S  N  I  K  K  I  A  E  G  D  T  A  P  G  L  E  G  N  R  F  N  D  G

CAAAGCTGATCCATCTGGAAGACTGTGGGTTGgtaagtaaaatgctactactagtaataatacctgtgcttctttaatac
  K  A  D  P  S  G  R  L  W  V

tgttgttttctcttattgcacattattcttcaaatatgcaacatataaactctttaatttcattcgaactgaatttccag

GAACTATGGGCGCAGAACCCATTAGGGGCCAAATCGCGCAAAACATGGGAACTCTTTACAGCTTAGAGAAAAAAAAACCAA
G  T  M  G  A  E  P  I  R  G  Q  I  A  Q  N  M  G  T  L  Y  S  L  E  K  K  N  Q

TTAAAAGCCCACCTCTCTCCCGTGTCCATTTCCAATGGTTTAGCCTGGAATGCAGCCCTAAAGAAATTCTATTACATCGA
 L  K  A  H  L  S  P  V  S  I  S  N  G  L  A  W  N  A  A  L  K  K  F  Y  Y  I  D

CTCATCAACCTATGAAATACACCAATTCGATTTTGACATCGATAACGGAACAATTTCAAATAAACAAACCATTTTCACCT
  S  S  T  Y  E  I  H  Q  F  D  F  D  I  D  N  G  T  I  S  N  K  Q  T  I  F  T

TTGAAAAACACAATATTCCTGGACTACCCGATGGACAATGTATTGATACTGATGGTAATTTATGGGTGGC
F  E  K  H  N  I  P  G  L  P  D  G  Q  C  I
```

ii. AnSMP2 [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGTCCGCATTGAGACGCAGGAACTCAAAGCTTATACTATGTGGATATCCTCGGAAAAACCATTAACAAATAT   80
                        A  G  T  Q  S  L  Y  Y  V  D  I  L  G  K  T  I  N  K  Y
ACTCCATCTACCGAAACACACACCAAAGCAACTTTAGgtaatgtaataatataactttaatgtttgcattgaattgatta  160
 T  P  S  T  E  T  H  T  K  A  T  L

attttaactcttaaaaaaactcttccagATAAAACTGTATCAATAATCATACCAGTCGAAGGAACCACTGATAAATTCTT  240
                              D  K  T  V  S  I  I  I  P  V  E  G  T  T  D  K  F  L

AATCACACATGGTAGAGATGTTTTGATTATAACCTGGGACGGAGAAAGTGAGTGGATAAGCGACTCAAAGAAAATTGCCG  320
 I  T  H  G  R  D  V  L  I  I  T  W  D  G  E  S  E  W  I  S  D  S  K  K  I  A

AAGTCGATACGGAACCAGGTTTGGAAAGCAACCAAATAAATGATGGAAAAGTTGATCCTACCGGAAGATTATGGGCTGGa  400
E  V  D  T  E  P  G  L  E  S  N  Q  I  N  D  G  K  V  D  P  T  G  R  L  W  A  G

ACCATAGGTACTTTAATACCTGATGCTACAGACCTTCGGTATATAAACGAACCTGACATGGGTACATTATACAGTGTAGA  480
 T  I  G  T  L  I  P  D  A  T  D  L  R  Y  I  N  E  P  D  M  G  T  L  Y  S  V  D

TAAAGACTACACCGTCAAAGCCCAATTCTCTCCAGTAACAATTTCCAATGGAGTAGCATGGAACTTGGATTTAAAAAAGT  560
 K  D  Y  T  V  K  A  Q  F  S  P  V  T  I  S  N  G  V  A  W  N  L  D  L  K  K

TTTATTACATAGACTCTCCACAACGTAAAGTATTTCAATTTGACTTTGACATTGAAAAAGGAACAATCACTAATAAGGAG  640
F  Y  Y  I  D  S  P  Q  R  K  V  F  Q  F  D  F  D  I  E  K  G  T  I  T  N  K  E

ACAATATTTACATTCGATATACATAATGTGCCTGGAATACCTGATGGTCAATGCATTGATGCAGATGGTAATTTATGGGT  720
 T  I  F  T  F  D  I  H  N  V  P  G  I  P  D  G  Q  C  I

TGC                                                                               723
```

e. *Necrophorus vespillo* Linnaeus, 1768

NvSMP [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGCCCGCATTGAGACCAACGAACGCAGAGCCTTTACTACGTGGACATCTTCGGACAGACCATCAACAAGTAC 80
                        Q  R  T  Q  S  L  Y  Y  V  D  I  F  G  Q  T  I  N  K  Y

ACGCCGGCAACCGAATCCCACTCCAAAGCAAAACTAGgtaacatcgtcaattgattgagatgaaacgattactacaaaga 160
 T  P  A  T  E  S  H  S  K  A  K  L

taagagaatttgtttacagATGAGAACGTCACGTTCATCATCCCAAAAGATGGGGAACCGGACAAGTACATTATCACCGT 240
                   D  E  N  V  T  F  I  I  P  K  D  G  E  P  D  K  Y  I  I  T  V

AGGCAGCAAAATCGTCGAAATTGAATGGGACGGTTCCAGTCCGGACGTTAAGATACTGAAGATATTGACTGATTTTGATT 320
  G  S  K  I  V  E  I  E  W  D  G  S  S  P  D  V  K  I  L  K  I  L  T  D  F  D

CAAAGAAGATCAACGATGGAAAATGCGACCCAAAGGGTCGTCTGTGGTTCGgtaagttacacgagctcagtttcgaattg 400
S  K  K  I  N  D  G  K  C  D  P  K  G  R  L  W  F

aattgcattgatgattcacagGTACAATTGGCGAAGATCCACTCAACATTCAGAGCAAACCAGAGATGGGCAGTCTTTTC 480
                     G  T  I  G  E  D  P  L  N  I  Q  S  K  P  E  M  G  S  L  F

TGTTTAGATCCCATGGGCAATGTATCCATGCACGTGAGCAACGTCATGATCTCAAACGGTCTCACCTGGAACGCAGATCG 560
 C  L  D  P  M  G  N  V  S  M  H  V  S  N  V  M  I  S  N  G  L  T  W  N  A  D  R

CACAAAGTTCTATTACATCGACTCCCTGAAGTTCACCATCGATGAATTCGATTATGACGATGCAACCGGGCACATTGgta 640
  T  K  F  Y  Y  I  D  S  L  K  F  T  I  D  E  F  D  Y  D  D  A  T  G  H  I

ctatatcataactgttccattagttctttcattacgttgtttcagGAAACGGAAGATCGATCTTCTCGTTGCAAACGAAC 740
                                             G  N  G  R  S  I  F  S  L  Q  T  N

AAGATCGAAGGAATGCCGGATGGGATGACAATCGATACAGATGGTAATTTATGGGTGGC 799
 K  I  E  G  M  P  D  G  M  T  I
```

f. *Grammoptera ruficornis* Fabricius, 1781

GrSMP1 [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGTCCGCATTGGGACGTGGAGACCCAGTCTCTTTATTTTGTTGATATGTTAGACAATTCGATACACAAATAC 80
                        V  E  T  Q  S  L  Y  F  V  D  M  L  D  N  S  I  H  K  Y

GTACCTTCTACAAGAAAACATACCAAAGCAAATGTAGgtaagtttcctatttgttatactagctgttggcccaaaggtc 160
 V  P  S  T  R  K  H  T  K  A  N  V

atttgcgtggtcttcactttaataatttagtagcagtaggtaccaggcgtacaaaacaaatcaccaaactttcatcgcaa 240

tcgcatgaatggcatagggacgcatgcgggacagatggacaatcaaacatttttttagctgtttaaatgacatgtatttt 320

tagGTCATGCTTCTTTGATTATACCTGTACAAGGAAAAAATGACCAATTCGTGATAGGTTTGGATCGGGAATTGGTGTTA 400
   G  H  A  S  L  I  I  P  V  Q  G  K  N  D  Q  F  V  I  G  L  D  R  E  L  V  L

ATAACTTGGGATGGTGAGAGTGAAACAGTTTCAATCGTGGAAAAATTGTATAAAGTGGATGAAGATAAAAGTAATAATGC 480
 I  T  W  D  G  E  S  E  T  V  S  I  V  E  K  L  Y  K  V  D  E  D  K  S  N  N  A

ATTCAATGACGGAAAATGTGACCCGTCCGGGAGGTTGTGGGCAGgtaaaaatgattataattaattaaactcagttgttt 560
  F  N  D  G  K  C  D  P  S  G  R  L  W  A

attgagaaaataaattaaaatgtatcactgttttagGTACCTTATATCAAACACCGGACGGAAATATTTTAGATGAACAA 640
                                    G  T  L  Y  Q  T  P  D  G  N  I  L  D  E  Q

GGTAGTTTGTACAGTTTCCAAAATAAGAAAATCACTAAGCATGCGAGTAAAATTGGGATAGCGAATGGGCTTGCCTTTGA 720
 G  S  L  Y  S  F  Q  N  K  K  I  T  K  H  A  S  K  I  G  I  A  N  G  L  A  F  D

TACCAGTAGAAACAAATTCTATTATATCGATTCTTTTCGTGGCACCCTTGACCAATACGATTTTGATATTAAAAATGGAA 800
  T  S  R  N  K  F  Y  Y  I  D  S  F  R  G  T  L  D  Q  Y  D  F  D  I  K  N  G

CAATATgtaagtgcggtgttgattgtaaatataagatatattattgtaaaaaatcaaaattttagCGAACAAAAAAACCGA 980
T  I                                                              S  N  K  K  P

TCTTTACTTTAAAAGAGGGGGTAAGAAAAAACTTTCAAGTACTAGATGGTATGACCATCGATACMGATGGTAATTTATGG 1060
I  F  T  L  K  E  G  V  R  K  N  F  Q  V  L  D  G  M  T  I

GTRGC 1065
 V
```

g. *Strangalia melanura* Fabricius, 1792

SmSMP [DegSMPF4 – DegSMPR1]

```
TTGGAGAAGGCCCACATTGAGACGCAGAAACCCAATCTCTTTATTATGTGGATATATTTGGGCATTACATTCACAAATAT  80
                        A  E  T  Q  S  L  Y  Y  V  D  I  F  G  H  Y  I  H  K  Y

GTACCTTCCACAAAAAAACATACTAAGGCCCACGTAGgtaagtcgtaacctgcgataaacttaattttgctatatgtatt 160
 V  P  S  T  K  K  H  T  K  A  H  V

agctataggtttaaataacttactatgtttttagGAGCAAATGTTTCATTAATTATACCTGTAGAAGGAAAAACTGATCA 240
                                  G  A  N  V  S  L  I  I  P  V  E  G  K  T  D  Q

GTTTGTGATAACTCTAGGTCGGGAAGTAGCATTAATAACTTGGGATGGTGAGAGTGAGGAAGTTTCAAAAGTGGAAAAAT 320
  F  V  I  T  L  G  R  E  V  A  L  I  T  W  D  G  E  S  E  E  V  S  K  V  E  K

TGTATGCAGTGGATGATAACACGGATAATGTCTTCAATGATGGAAAATGTGATCCTTCTGGGAGGTTATGGGCAGgtaac 400
L  Y  A  V  D  D  N  T  D  N  V  F  N  D  G  K  C  D  P  S  G  R  L  W  A

gagcatacatagttgtatttattgcattacaaagtacggtttgcataaataaatggtcacattgttgttgctaacgctac 480

acgttaaagtgatagcgaacaaggctattgtcaaatagtttcatctgcttttgattaaaattaaataataataggcagac 560

ttttatttataccactgaaaattgtaatgggagtataatggaataggatttatataattgtttgggtgaagggccagcac 640

acttgacacgtcatgataacgactctcactcaggagttctaggttcgatttccagtcttggcactgggactttttagaag 720

aaaaaataatttcatataagtacccaaatattggatgtgcgtttttcatatctaaatgtgtaggcactcgcacagcaaca 800

atttgggtccaaataatgtaatgtgaattattatgaatggtctaccctgaataagtatggtgtgttcaaatagttaattc 880

gtccacctgagtcctagcgcaaagttgcttaacttccgtgatctagcgtctcttcagttcccacactgagttgtcttaac 960

gaaaaggataactcaacaactcatttagaattccgtatttttttttgcaaaaatataccaaattacttgcagccagagct 1040

gggcggccatagggattgaagaacagatatttatgtttgattcttgtttatttcaaaatttacagcagacaaaatttatt 1120

atagcaaaagaaaaaaattcaatataaaactgttttagGCACCTTGGCACCAATGAAGGATAATAAATTTGTAGATGAAC 1200
                                      G  T  L  A  P  M  K  D  N  K  F  V  D  E

AAGGTAGTTTATATAGTTTCCAAAACAAAAAAGTTGCTAAGCATGCAAGTAAAATTGGTGTTGCGAATGGACTTGCTTTT 1280
Q  G  S  L  Y  S  F  Q  N  K  K  V  A  K  H  A  S  K  I  G  V  A  N  G  L  A  F

GATGCTGATCGAAAGAAATTCTACTATATTGACACTCTTCGGCGCACACTTGATGAATATGATTTTGACATAAAAAATGG 1360
 D  A  D  R  K  K  F  Y  Y  I  D  T  L  R  R  T  L  D  E  Y  D  F  D  I  K  N  G

AAAAATATgtaagtgtaccgtttagttgcgggatttaattataattttcataacacgtcataaaaaaaatacaaatttca 1440
 K  I

gCAAACAAAAAACCAATCTTTACCTTAAAACGCAGTGCAGTAGAACACGACGGATTAGACGGTATGGCTATTGATGCCGA 1520
 S  N  K  K  P  I  F  T  L  K  R  S  A  V  E  H  D  G  L  D  G  M  A  I

TAATCTATGGGTGGC 1525
```

h. *Anthocomus fasciatus* Linnaeus, 1758

i. AfSMP [DegSMPF4 – DegSMPR1]

```
GGAGAAGGCCCACATTTGGGACGCCGCAACTCAGAGCCTCTATTTCGTCGACATTTTCGGGAAAGCAATTCACAAGTACG  80
                       A  A  T  Q  S  L  Y  F  V  D  I  F  G  K  A  I  H  K  Y

TTCCGGCTGAAAAGAAACACACGAAAGCGATCATCGgtaagttcacttttcattcacatttttgcaataataccaaatg 160
 V  P  A  E  K  K  H  T  K  A  I  I

atttttgctcagGAACCAACCATGTCTCTCTCATATTACCAATCGAAAGCTCCGCCGATAAATTTTTAATTTCCATTGG 240
             G  T  N  H  V  S  L  I  L  P  I  E  S  S  A  D  K  F  L  I  S  I  G

TAGAGAGTTGGCGGTAGTGACTTGGGATGGGAAAAGTGAAAAGGTATCCAATATAGAAAAGATCGCAGAAGTGGATAACG 320
  R  E  L  A  V  V  T  W  D  G  K  S  E  K  V  S  N  I  E  K  I  A  E  V  D  N

TTCCTGGAAAATTAGACAACAGATTCAATGATGGAAATGCGATTCTTCGGGGCGACTCTGGGCAGgttcggtaatttt 400
 V  P  G  K  L  D  N  R  F  N  D  G  K  C  D  S  S  G  R  L  W  A

tgggagatattataatttaatcctgtttgatataattatagGTACGATGGGTGGTGAGCCGGTCAATGGTCAAGTAAAAC 480
                                          G  T  M  G  G  E  P  V  N  G  Q  V  K

CGAATCAAGGTAGCTTTTTCAGTTTGGAAGGAAAGAAAGTTAGACAGCACTTGACTAATTTGGGAGTTTCGAATGGCCTC 560
 P  N  Q  G  S  F  F  S  L  E  G  K  K  V  R  Q  H  L  T  N  L  G  V  S  N  G  L

GCCTGGAATGATGAGTTGAAGAAATTCTACTTCATCGATTCTTTCAAATATTCGATTGAGCAATACGAATTTGATATCTC 640
  A  W  N  D  E  L  K  K  F  Y  F  I  D  S  F  K  Y  S  I  E  Q  Y  E  F  D  I  S

TTCAGGAACAATTAgtaagtgattattaaattaaattattacttctttatttatgcgtatgttattatttattttagAA 720
  S  G  T  I                                                                  K

AATCCTGAGACGATTTTTACTCTGAATAAGCACGATATACCAGGCGTACCAGATGGAATGTGTATAGATGCCGATGGTAA 800
  N  P  E  T  I  F  T  L  N  K  H  D  I  P  G  V  P  D  G  M  C  I

TTTATGGGTGGC 812
```

ii. AfSMP [DegSMPF4 – DegSMPR1]

```
TTGGTGAAGGCCCGCATTGAGACATTGAAACTCAATCCTTATATTACGTCGATATCTTCGGCCAGACCATCAACAAGTAC  80
                       I  E  T  Q  S  L  Y  Y  V  D  I  F  G  Q  T  I  N  K  Y

ACACCAGAAACCAAAACACACACCAAAGTTAAAATAGgtaagtaattaccagtaatattatttaacacaaaattttatt 160
  T  P  E  T  K  T  H  T  K  V  K  I

aaattgcttgtagAAGGAGGCCCCGTGACTCTAGTGGTACCAGTTGAAGGGACGAAGGATCAATTCTTGATCAGTGTTGG 240
              E  G  G  P  V  T  L  V  V  P  V  E  G  T  K  D  Q  F  L  I  S  V  G

ACGAAAACTATTAATCGTTACTTGGGATGGAGTCAGCGATAAGATCTCAAAATCTGAACTGCTGGTCGAAGTTGAAAATC 320
  R  K  L  L  I  V  T  W  D  G  V  S  D  K  I  S  K  S  E  L  L  V  E  V  E  N

GAGAAGGGTATTTCAACAACAGATTCAACGACGGTAAAGCGGATCCCACTGGAAGATTATGGGCAGgtaaatgttttttt 400
 R  E  G  Y  F  N  N  R  F  N  D  G  K  A  D  P  T  G  R  L  W  A

ttctgcattatattaatttgctcatacccaattttaaattggtaggtgattataatatctaatcgacgtgtcttgaaaag 480

taaatgaaatcgcctcgtcacccccatggcgtcttagagtctattattggcgtctggacgtcataatttatttcactgag 560

ttttacctttcaaattatgcttttacaaattttcagGAACTATGGGACCCGAGCCAGAAATCGGCAAATTAGAGAAGGAA 640
                                     G  T  M  G  P  E  P  E  I  G  K  L  E  K  E
                                    ^
AAAGGTTCGCTTTACACTCTAATCAACAAGTATAAGTAAAAACGCACTTGACCAAAGTGAGCATAGCAAACGGTCTAGCT 720
  K  G  S  L  Y  T  L  I  N  K  Y  K  *
                                    *  V  K  T  H  L  T  K  V  S  I  A  N  G  L  A

TGGAATGTGGGATTAGCGAAAATGTACTACATCGATTCACCCTTAAGGACCGTCGATCAATACGATTACGACATGGAAAA 800
  W  N  V  G  L  A  K  M  Y  Y  I  D  S  P  L  R  T  V  D  Q  Y  D  Y  D  M  E  K

GGGAGAAATTAgttagtatgttaatatgtgttagttattattattctgagaggcatctatttagGTAACAAGAAGGTCGT 880
  G  E  I                                                          S  N  K  K  V  V

TTTCGATTTGGACGAGCACGGCATTCCAGGTGTTCCAGATGGTATGACCATAGATACCGATGGTAATTTATGGGTGGC  958
  F  D  L  D  E  H  G  I  P  G  V  P  D  G  M  T  I
```

i. *Pyrochroa serraticornis* Scopoli, 1763

PsSMP [LamLRE1.1 F & R]

```
GAGAAGGTCCACACTGGGATGATGAAACCCAAAGTCTTTACTACGTCGATATTTTCGGGCAAGCAATCCACAGATATGTT   80
                           T  Q  S  L  Y  Y  V  D  I  F  G  Q  A  I  H  R  Y  V

CCATCTACCAACACCCACACCAAAGCCGTTATAGgtaagtttattcaaattagaaaatttctatataaaactgatttatg   160
 P  S  T  N  T  H  T  K  A  V  I

tagAGGGAGGTCCCGTTACTTTGGTCGTGCCAGTAGATGGGGAGAAAGACCGTTTTTGATCGCTCTAGGCCGCAAATTG   240
   E  G  G  P  V  T  L  V  V  P  V  D  G  E  K  D  R  F  L  I  A  L  G  R  K  L

ATGACAGTCACTTGGGATGGAATCAGCGATAAAGTTACTAAACCAGAATTATTAATTGAAGTCGAAAATAAAGCAGGATA   320
 M  T  V  T  W  D  G  I  S  D  K  V  T  K  P  E  L  L  I  E  V  E  N  K  A  G  Y

TTTTGATAATAGATTCAACGACGGAAAAGCCGATCCTTCCGGAAGATTATGGGCTGgtaagttgtaattatcgtaaaatt   400
  F  D  N  R  F  N  D  G  K  A  D  P  S  G  R  L  W  A

ttccactaaagttgaattagGAACGATGGGTCCTGAGCCGGAAATTGGAAAACTAGAGAAAGAAAAGGGAGCGTTGTACA   480
                    G  T  M  G  P  E  P  E  I  G  K  L  E  K  E  K  G  A  L  Y

CTTTAATCGGGAAACACCAAGTGAAAACTCATTTAACTAAAGTTAGTATCGCTAACGGTTTAGCCTGGAATTTGGCGTTG   560
T  L  I  G  K  H  Q  V  K  T  H  L  T  K  V  S  I  A  N  G  L  A  W  N  L  A  L

AGAAAGATGTACTATATCGATTCCCCAAGAAGAACTGTGGATGAATACGATTATAATGAAGAGAAAGGGGAAATTTgtga   640
 R  K  M  Y  Y  I  D  S  P  R  R  T  V  D  E  Y  D  Y  N  E  E  K  G  E  I

gtagttttcgaaaatttctttaattatttggtaattcacagGTAATCGTAAGGTAGTTTTCAATTTGGACCATCACGAT   720
                                          G  N  R  K  V  V  F  N  L  D  H  H  D

ATTCCTGGAGTTCCCGATGGTATGACCATTGACGACGATGGTAATTTATGGGTCGC
 I  P  G  V  P  D  G  M  T  I
```

j. *Tenebrio molitor* Linnaeus, 1758

TmSMP [DEGLRE F & R]

```
TTGGAGAGGGTCCTCATTGGGACGAAGCCACTCAGACTCTCTACTTCGTCGACATTTTCGGTCAAGCCATCCACAAATAC   80
                        E  A  T  Q  T  L  Y  F  V  D  I  F  G  Q  A  I  H  K  Y

GTCCCTTCGACCAACACGCACACCAAAGTGGTAATCGgtgagtacaagcaacaatctcttcaaacgtgatttgagtggtt   160
 V  P  S  T  N  T  H  T  K  V  V  I

tcagAGGGCGGCCCAGTCACTATGGTGGTCCCTGTGGAGGGCACCAGCGACCAGTTTGTGATCAGTATCGGGCGAAAGCT   240
    E  G  G  P  V  T  M  V  V  P  V  E  G  T  S  D  Q  F  V  I  S  I  G  R  K  L

AGTCCTGGTCACTTGGAATGGCGCCAGCGGAAAGATCTCCAATTCTGAGCTTCTGATCGAAGTCGAGAACAAAGCAGGAT   320
  V  L  V  T  W  N  G  A  S  G  K  I  S  N  S  E  L  L  I  E  V  E  N  K  A  G

ACTTCAACAATCGGTTCAACGATGGGAAAGCGGATCCCAGTGGGCGATTGTGGGCTGGTACGATGGGCCCCGAACCAGAA   400
Y  F  N  N  R  F  N  D  G  K  A  D  P  S  G  R  L  W  A  G  T  M  G  P  E  P  E

ATTGGCAAGCTAGAGAAGGAGAAAGGCGCTTTGTACACTTTGATTGGGAAGCACCAAGTGAAGACGCACCTCACCAAAGT   480
 I  G  K  L  E  K  E  K  G  A  L  Y  T  L  I  G  K  H  Q  V  K  T  H  L  T  K  V

GAGCATCGCTAATGGTTTAGCCTGGAATCTGGAGTTGAAGAAGATGTACTACATTGACTCACCGAGACGGACGATAGATG   560
  S  I  A  N  G  L  A  W  N  L  E  L  K  K  M  Y  Y  I  D  S  P  R  R  T  I  D

AGTACGACTACAACATCGACAAAGGAGAAATCTgtaagaactcgtcttcgattttaaccgtttggcaaaagactgtgtat   640
E  Y  D  Y  N  I  D  K  G  E  I

tagGTAACCGCAAGGTGGTTTTTAATTTGGACGTGCACAACATTCCGGGAGTCCCCGACGGCATGACTATGGACACAGAC   720
   C  N  R  K  V  V  F  N  L  D  V  H  N  I  P  G  V  P  D  G  M  T  M  D  T

GGTAATTTATGGGTTGCTGTTT                                                              742
```

k. *Pachnoda marginata* (Drury 1773) ssp. *peregrina* Kolbe, 1906

i. PmSMP1

```
ATGGCGCCAGTTGTAGAAGTTGTTACTGAACGCGTCACCTTGGGGAAGGACCAGCATTGGGACGTTGAATCTCAATGCTT    80
 M  A  P  V  V  E  V  V  T  E  R  V  T  L  G  K  D  Q  H  W  D  V  E  S  Q  C  L

GTATTACGTGGACATTTTGGGGCAAGCTATTCATAAATATGTTCCTTCTACAAATAAACACACTAAAGTGAAGATTGgta   160
  Y  Y  V  D  I  L  G  Q  A  I  H  K  Y  V  P  S  T  N  K  H  T  K  V  K  I

agtaagaaactaaaataatgatttattatctgcataaatgaaatcatcaaactaatatcgtctaaacatacctgatacga   240

tatgtatcatgcatgtgtttaatgaaaatatacctgcagaaagtcactcaaacataaagtaaattaaaaaaagtgtcgtt   320

ctttttgttacgctcagccgtagcagtgctggtgaatcagcgtaatagaaattttttgttttgattccgatattttttga   400

cgaataaaagcgtatttttgattgtattgtaaatagcgttttaattggaatgacgacgaagatgagatccatttaaatac   480

ctaacatttacgtcatcagttcctggtattcataaaaattgttaaacaattagtaatttttcgtttttattttacacaga   560

atatgatttttttcattgttttatcattttaccgataagctctagttaatacttttctcgtaaatcctttaatgatttaa   640

gtcacttttattatttgatgtaattctttgctttcattgtgataaagctatatataataacatacttaaaattaaacttt   720

taaacacatttcaaattattttattatgattgatgaacattttttttttctgttataataaaagttttatcactgtataa   800

tgtattgctcgtattacatttagAAGGGGGTCCAGTTTCTTATTGGTACCAGTCGAAGGTAGCAAGAACCGTTTCTTAA   880
                         E  G  G  P  V  S  L  L  V  P  V  E  G  S  K  N  R  F  L

TAAGCATCGGGAGAAAGTTAGTTGTGGTCACATGGGATGGAGTTAGCGAAAAACCGACAAAAATTGAAGAACTGTTGGAG   960
 I  S  I  G  R  K  L  V  V  V  T  W  D  G  V  S  E  K  P  T  K  I  E  E  L  L  E

GTAGAAAATGAACCGGATCTTATTGGAAATAGATTCAATGATGGAAAAGCTGATCCCGCTGGTCGACTATGGGCTGgtaa  1040
  V  E  N  E  P  D  L  I  G  N  R  F  N  D  G  K  A  D  P  A  G  R  L  W  A

attcattatttcgcgattacttaaaatttcacccacaggtgtttgtattagGGACAATGGGTCCGGAACCGGAAATTGGA  1120
                                                    G  T  M  G  P  E  P  E  I  G

AAACTGGAACCGGAAAAAGGGATCACTCTATACGTTGGTTGGGAAAAAGAGAGGCCAAGAAGCATTTGTCCAAAGTTAGCAT  1200
 K  L  E  P  E  K  G  S  L  Y  T  L  V  G  K  R  E  V  K  K  H  L  S  K  V  S  I

CGCTAACGGATTAGCTTGGAATATTCCCCTGAAAAAAAATGTATTATATTGATTCACCCAGAAGAACTGTGGATCAGTATG  1280
  A  N  G  L  A  W  N  I  P  L  K  K  M  Y  Y  I  D  S  P  R  R  T  V  D  Q  Y

ATTATGATATGGATAAAGGAACAATTGgtagttttttttttttgttttcaaaatctttctcgctccaatataattcctat  1360
 D  Y  D  M  D  K  G  T  I

tttagCCAATAGGAAACCAATTTTTGATTTGGATCACCATGACATCCCTGGAGTACCTGATGGTATGACTATAGATACCG  1440
      A  N  R  K  P  I  F  D  L  D  H  H  D  I  P  G  V  P  D  G  M  T  I  D  T

ATGGGAATCTTTGGGTGGCGGTCTTTGATGGAGGCTGTTTATTGCATATCGATCCAAGAACATCAAAATTGATATCGTTA  1520
 D  G  N  L  W  V  A  V  F  D  G  G  C  L  L  H  I  D  P  R  T  S  K  L  I  S  L

ATAAAGTTCCCTGCAAATCAAgtaacaaatcctaaactcataaaatctgccttttataaagtcattgtatttcaaatagA  1600
  I  K  F  P  A  N  Q

TTACTGCTGCCGCATTTGGAGGTCAAAATATGGACATCTTGTATGTAACCAGTGCACAATTGGTTGTGAAAGGAGTTCCT  1680
 I  T  A  A  A  F  G  G  Q  N  M  D  I  L  Y  V  T  S  A  Q  L  V  V  K  G  V  P

CAAGAAGAACCAGCTGGTTCTGTTTTTAAAGTAATCGGAACTGGAGCCAGAGGACTGCCGGGATATAAAGTTAGGTTGGA  1760
 Q  E  E  P  A  G  S  V  F  K  V  I  G  T  G  A  R  G  L  P  G  Y  K  V  R  L  D

CAATTTGAAG                                                                         1770
 N  L  K
```

ii. PmSMP2

```
ATGCCCGTAATTGTAGAACGTCTTACTGACTCCAATCAGTTTGGAGAAGGCCCACATTGGGACGAGAAGACTCAAAGTTT   80
 M  P  V  I  V  E  R  L  T  D  S  N  Q  F  G  E  G  P  H  W  D  E  K  T  Q  S

ATATTATGTTGATATTTTTGAGCATACTATCCATAAATATGTGTCATCGAAAAACTTGCATGCGAGGGCTCTAGTTGgta  160
L  Y  Y  V  D  I  F  E  H  T  I  H  K  Y  V  S  S  K  N  L  H  A  R  A  L  V

aactctttttctacaaaataatccaggagtacaaacagctgatattatgcatatctttaagGTGATGATACTGTTTCTC  240
                                                              G  D  D  T  V  S

TAATAATACCAATAAGTCAAGCCGTGGATAAATATTTAATTGCTTTTGGAAGACAAATAGGCATAGTAACATGGAATGGA  320
L  I  I  P  I  S  Q  A  V  D  K  Y  L  I  A  F  G  R  Q  I  G  I  V  T  W  N  G

GTAAATGATAAGGCAAGTGATTTCAAAAAACTTGCCGAGGTAGATATTGAAGAAGGCTCTCGTGAAAACAGAATAAATGA  400
 V  N  D  K  A  S  D  F  K  K  L  A  E  V  D  I  E  E  G  S  R  E  N  R  I  N  D

TGGAAAATGTGATGCATCTGGGAGATTGTGGTGTGgtaatagagaaatattaaatttgctctaaatctatcacagcattt  480
 G  K  C  D  A  S  G  R  L  W  C

attgtcctttagGTACGATGGGTCCAGAACCACAAATTGGACACATACAACCGAACAAGGGCTCACTGTATAGTATGGA  560
             G  T  M  G  P  E  P  Q  I  G  H  I  Q  P  N  K  G  S  L  Y  S  M  D

TCCGAATGGCAGAATTAAGAAGCATTTGGCGTCAATTACTTGTTCCAACGGTTTGGCTTGGAACGCTGATAATACCAAGC  640
 P  N  G  R  I  K  K  H  L  A  S  I  T  C  S  N  G  L  A  W  N  A  D  N  T  K

TATATTATATAGATTCTGGAAACTTAGAAATACATCAATATGATTTTAATTTAGTGCAAGGAGAAATAAgtctgtttatt  720
L  Y  Y  I  D  S  G  N  L  E  I  H  Q  Y  D  F  N  L  V  Q  G  E  I

cttatatcgttgcacgtatcataatattaatagagggaattttctctagGCAACAAAAAGGTGATTTTTCGTTTGAGAGA  800
                                                  S  N  K  K  V  I  F  R  L  R  E

AGCTGGTATCGATGGCTTTGCAGATGGACAAACTATCGATRCAGATGGTAATTTATGGGTAGCCATTTTCAACGGTTACA  880
 A  G  I  D  G  F  A  D  G  Q  T  I  D  X  D  G  N  L  W  V  A  I  F  N  G  Y

AGGTCATTAAAATCGACCCCAGAAGGTACAATGCCCTTCTTCAAACAGTGGAGATTCCTGCTAAACAGgtaaacgaatct  960
K  V  I  K  I  D  P  R  R  Y  N  A  L  L  Q  T  V  E  I  P  A  K  Q

aaacggttataaatcacttgcattcgtgtttcgtatttagGTCACTTCTGTTGCGTTTGGTGGTCCAAATCTGGATGAAT 1040
                                         V  T  S  V  A  F  G  G  P  N  L  D  E

TGTATGTTACTACAGCCAGTTTTACCGTTGACGGAGAAGTCCTGCCTCCTCCAGAACACGGGGCTCTGTATAAAGTTACA 1120
L  Y  V  T  T  A  S  F  T  V  D  G  E  V  L  P  P  P  E  H  G  A  L  Y  K  V  T

GGTTTAGGAGTGAAGGGTGTTCATGGAAATTGTTTTAAGATG                                      1162
 G  L  G  V  K  G  V  H  G  N  C  F  K  M
```

iii. PmSMP3

```
ATGGCGGCTCAAATCGAAGCCCTAGTTGGTGGCCTTATCCTACCTGAAGGCGTACATTGGGACCAGGACACACAATCTCT   80
 M  A  A  Q  I  E  A  L  V  G  G  L  I  L  P  E  G  V  H  W  D  Q  D  T  Q  S  L

TTATTTCGTGGATACAAACGATAGATCTATTTATCGATACGCAGTTGCAACGAAAACTTATGTGAAAGCCACTTTAGCGT  160
 Y  F  V  D  T  N  D  R  S  I  Y  R  Y  A  V  A  T  K  T  Y  V  K  A  T  L  A

TAAATAACGTTACTGTTATAATTCCCGTGCGTGGTCAAAATGATAAATTCTTGATTACCCACGGAAATAAATTATCTGTG  240
L  N  N  V  T  V  I  I  P  V  R  G  Q  N  D  K  F  L  I  T  H  G  N  K  L  S  V

ATAAGATGGGACGGGCAAAGCGAAAGGGTATCAAGTAGCGAACTGGTGGCATCCCTGGACGACCCAGACACTGGTTCATA  320
 I  R  W  D  G  Q  S  E  R  V  S  S  S  E  L  V  A  S  L  D  D  P  D  T  G  S  Y

TATTTGGTGCGATGGGAAAGCTGATCCCGATGGAAAATTTTGGGCAGGAGCTATGATAGTAGGATCCCGCGGTGGCTTTG  400
 I  W  C  D  G  K  A  D  P  D  G  K  F  W  A  G  A  M  I  V  G  S  R  G  G  F

CTGAAAAGACTGGGACTTTATATAATTTAGAATCAGATAGACGAATAAAAAGGCATTTTAACTCTTTAACAATACCAAAT  480
A  E  K  T  G  T  L  Y  N  L  E  S  D  R  R  I  K  R  H  F  N  S  L  T  I  P  N

GGATTAGCGTGGAATCAAGTCACAAAGAAATTTTATTTTATTGACAGTCCCACCAGGAGGATTGAACAATTCGATTATGA  560
 G  L  A  W  N  Q  V  T  K  K  F  Y  F  I  D  S  P  T  R  R  I  E  Q  F  D  Y  D

CCCCGACAACAGTATTATTACCAATAGAACTAGATTGTTTACTTTTGACGATCACGATATTCCTGGAGCGCCTGACGGTA  640
 P  D  N  S  I  I  T  N  R  T  R  L  F  T  F  D  D  H  D  I  P  G  A  P  D  G

TGACAATAGATACAGATGGAAATCTATGGGTTGCCTGTTTTGGCGGAGGAATGGTTATTAAAGTTGATCCAACCAAACGT  720
M  T  I  D  T  D  G  N  L  W  V  A  C  F  G  G  G  M  V  I  K  V  D  P  T  K  R

GAAACTTTATTAGAACAAATTAAATTACCAGCAGCGCAGgtatatataaatgttttattcattaattaaatgcgtgatac  800
 E  T  L  L  E  Q  I  K  L  P  A  A  Q

atgaaaatactcaaaatatagGTGTCTTCATTGGCGTTTGGCGGTATAAATATGGACGAACTATATATAGCAACGGCGAA  880
                       V  S  S  L  A  F  G  G  I  N  M  D  E  L  Y  I  A  T  A  K

AATTCAATTTGGACCAATGGCTGCGCCAAGCGGCCCAGAAAATGGCGTCATCTACAGGATAACTGGTTTAGGAGCGAAAG  960
 I  Q  F  G  P  M  A  A  P  S  G  P  E  N  G  V  I  Y  R  I  T  G  L  G  A  K

GTTATCCCGGGAATAGAGTTAATTTAGAATTTATG                                              995
G      Y      P      G      N      R      V      N      L      E      F      M
```

APPENDIX II

PARTIAL AND FULL LENGTH BEETLE LUCIFERASE AND LUCIFERASE-LIKE GENE SEQUENCES: PREDICTED GENE ARCHITECTURE AND PUTATIVE OPEN READING FRAMES

Predicted exon and intron sequences are denoted in upper and lower case respectively. Primer sequences are shown underlined. Non-canonical intron/exon boundaries are shown highlighted in grey. Predicted open reading frames are shown below the gene sequence. Square brackets contain the primers used in the PCR – Primer sequences are shown underlined. A microsatellite repeat present in the first intron is highlighted in red.

CONTENTS

A. Bioluminescent Beetles (Lampyridae)

a. *Photinus pyralis* Linnaeus, 1767
b. *Phosphaenus hemipterus* Fourcroy, 1785
c. *Nyctophila reichii* Jacquelin du Val, 1859
d. *Lampyris iberica* Figueira, Geistardt & Day 2008
e. *Lampyris sardiniae* Geisthardt, 1987
f. *Lampyris noctiluca* Linnaeus, 1758
g. *Phausis reticulata* Say, 1825
h. *Luciola italica* Linnaeus, 1767
i. *Lampyroidea maculata* Geisthardt & Day, 2006
j. *Lamprohiza splendidula* Linnaeus, 1767
k. *Photuris congener* LeConte, 1851
l. *Photuris* species 'A'

B. Non-Bioluminescent Beetles

a. *Cantharis rufa* Linnaeus, 1758
b. *Cantharis rustica* Fallén, 1807
c. *Aphodius rufipes* Linnaeus, 1758
d. *Grammoptera ruficornis* Fabricius, 1781
e. *Pachnoda marginata* (Drury 1773) ssp. *peregrina* Kolbe, 1906
f. *Stegobium paniceum* Linnaeus, 1758
g. *Necrophorus vespillo* Linnaeus, 1768
h. *Anacaena sp.* Thomson, 1859

A. Bioluminescent Beetles

a. *Photinus pyralis* Linnaeus, 1767

i. PhpLUC

```
GGYGTGGCCCTTCCGCATAGAACTGCCTGCGTCAGATTCTCGCATGCCAGgtatgtcgtataacaagagattaagtaatg   80
       A  L  P  H  R  T  A  C  V  R  F  S  H  A  R

ttgctacacacattgtagAGATCCTATTTTTGGCAATCAAATCATTCCGGATACTGCGATTTTAAGTGTTGTTCCATTCC   160
                   D  P  I  F  G  N  Q  I  I  P  D  T  A  I  L  S  V  V  P  F

ATCACGGTTTTGGAATGTTTACTACACTCGGATATTTGATATGTGGATTTCGAGTCGTCTTAATGTATAGATTTGAAGAA   240
H  H  G  F  G  M  F  T  T  L  G  Y  L  I  C  G  F  R  V  V  L  M  Y  R  F  E  E

GAGCTGTTTTTACGATCCCTTCAGGATTACAAAATTCAAAGTGCGTTGCTAGTACCAACCCTATTTTCATTCTTCGCCAA   320
 E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T  L  F  S  F  F  A  K

AAGCACTCTGATTGACAAATACGATTTATCTAATTTACACGAAATTGCTTCTGGGGGCGCACCTCTTTCGAAAGAAGTCG   400
  S  T  L  I  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L  S  K  E  V

GGGAAGCGGTTGCAAAACGgtgagttaagcgcattgctagtatttcaaggctctaaaacggcgcgtagCTTCCATCCTCC   480
G  E  A  V  A  K  R                                                     F  H  P  P

AGGGATACGACAAGGATATGGGCTCACTGAGACTACATCAGCTATTCTGATTACACCCGAGGGGGATGATAAACCGGGCG   560
  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  L  I  T  P  E  G  D  D  K  P  G

CGGTCGGTAAAGTTGTTCCATTTTTTGAAGCGAAGGTTGTGGATCTGGATACCGGGAAAACGCTGGGCGTTAATCAGAGA   640
A  V  G  K  V  V  P  F  F  E  A  K  V  V  D  L  D  T  G  K  T  L  G  V  N  Q  R

GGCGAATTATGTGTCAGAGGACCTATGATTATGTCCGGTTATGTAAACAATCCGGAAGCGACCAACGCCTTGATTGACAA   720
 G  E  L  C  V  R  G  P  M  I  M  S  G  Y  V  N  N  P  E  A  T  N  A  L  I  D  K

GGATGGATGGCTACATTCTGGARACATAGCTTACTGGGACGAAGACGAACACTTCTTCATAGTTGACCGCTTGAAGTCTT   800
  D  G  W  L  H  S  G  B  I  A  Y  W  D  E  D  E  H  F  F  I  V  D  R  L  K  S

TAATAAAATACAAAGGCTACCA
L
```

ii. PhpLL1

```
CCGGGTTGCCAAAAGGCGTTATGTTGACCGAGAAGAACATGATGATTAGATATAATCAATTCGTgtgagctacacaacat   80
                     M  L  T  E  K  N  M  M  I  R  Y  N  Q  F  V

acctacatcccaaattattatactgatattattttagGGATCCGCAGTGTATGTTTGGAAGGGACCTTAGACCAGGAGAC   160
                                       D  P  Q  C  M  F  G  R  D  L  R  P  G  D

GCTGTAATCAACTATTTGCCGTTTTTTCATGGTTACGGGTTTTCTATATTATTAGGGTACCTTATTATGGGCTTACATGT   240
 A  V  I  N  Y  L  P  F  F  H  G  Y  G  F  S  I  L  L  G  Y  L  I  M  G  L  H  V

TTTTATAATGGAGAGTTACAAGGAAGATTTGTTGCTAAAGTTTCTCGAGAAGTTTCAAGTAAAAAGCCTCTGTGTAGTTC   320
  F  I  M  E  S  Y  K  E  D  L  L  L  K  F  L  E  K  F  Q  V  K  S  L  C  V  V

CATCAATTCTTATATCCCTTGCAAAGAAAGAGACTTTGAACGACCGCAGTTTGTCAAAGCTAAACGAAGTTGTGTATGGT   400
P  S  I  L  I  S  L  A  K  K  E  T  L  N  D  R  S  L  S  K  L  N  E  V  V  Y  G

GCTGCACCTACTTCCAAGCAGATAGTTGTGCAAGCAAAGCAAAGgtgaccagtacttagataaaatgtgttttcttatat   480
 A  A  P  T  S  K  Q  I  V  V  Q  A  K  Q  R

atacctgcagGTTTCAAATTCAAGAGATGCGATCAGGATACGGATTAACGGAAGGAACTATTGTTTCGATTAGTACTCCA   560
            F  Q  I  Q  E  M  R  S  G  Y  G  L  T  E  G  T  I  V  S  I  S  T  P

TTGGGTTGTGTTAAGTATTCCTCAGTTGGAAAACTTTTACCATTTGTAGATGCAAAAATTGTTGATATAGTCACTCAAGA   640
 L  G  C  V  K  Y  S  S  V  G  K  L  L  P  F  V  D  A  K  I  V  D  I  V  T  Q  E

GCCTCTAGGCCCGAATCGAACCGGCGAATTGTGCATTAAAGGCGATACTGTAATGAAGGGATATATGGGAAATGCCAAAG   720
  P  L  G  P  N  R  T  G  E  L  C  I  K  G  D  T  V  M  K  G  Y  M  G  N  A  K

CAACAGAAGATACAATTGACCGAGACGGTTGGCTCCATACCGGCGACATGGCTTATTACGACACAGAAGAATACTTCTAC   800
A  T  E  D  T  I  D  R  D  G  W  L  H  T  G  D  M  A  Y  Y  D  T  E  E  Y  F  Y

ATAGTAGATAGGCTTAAAGAATTGATAAAATACAAGGGCTACCA
 I  V  D  R  L  K  E  L
```

b. *Phosphaenus hemipterus* Fourcroy, 1785

i. PphLUC

```
ACCGGTTTGCCTAAAGGGTAGAGCTTAACCACACTAGTGTTTGCGTGAGATTTTCGCATTGCGGgtacgtagtacttga    80
                     E  L  N  H  T  S  V  C  V  R  F  S  H  C  G

tcgaagaaatagaataatttgtaaatgtgcagAGATCCTGTTTTTGGGAATCAAATCATTCCCGATACTGCAATTTTAAG   160
                                  D  P  V  F  G  N  Q  I  I  P  D  T  A  I  L  S

CGTTATCCCATTCCATCATGGATTTGGAATGTTTACAACGCTAGGTTATTTAATATGCGGTTTTCGAGTTGTGCTGATGT   240
  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  I  C  G  F  R  V  V  L  M

ATAGGTTTGAAGAAGAACTATTTTTGAGATCCCTTCAAGATTATAAAATTCAAAGTGCGTTACTGGTACCCACGCTATTT   320
Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T  L  F

TCGTTCTTTGCTAAAAGCACTCTAATTGACAAGTACGATTTATCTAACTTACATGAAATTGCTTCTGGAGGTGCTCCCCT   400
 S  F  F  A  K  S  T  L  I  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L

TGCAAAAGAAGTTGGAGAAGCAGTGGCAAAACGgtgagtatttaccattttcaaaaggtttttataaggtgggtagATT   480
  A  K  E  V  G  E  A  V  A  K  R                                              F

TAACCTTCGCGGTATACGGCAAGGGTACGGTCTGACCGAAACTACATCGGCCGTTATTATTACACCCGAGGGAGATGATA   560
  N  L  R  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  V  I  I  T  P  E  G  D  D

AGCCAGGTGCAGTTGGTAAGGTTGTTCCCTTCTTTTCGGCAAAAGTTGTTGATCTCGATACCGGAAAAACTTTGGGACTT   640
K  P  G  A  V  G  K  V  V  P  F  F  S  A  K  V  V  D  L  D  T  G  K  T  L  G  L

AATCAAAGAGGCGAATTGTGTCTGAAAGGGCCCATGATTATGAAAGGTTATGTAAATAACCCTGAAGCAACAAATGCCTT   720
 N  Q  R  G  E  L  C  L  K  G  P  M  I  M  K  G  Y  V  N  N  P  E  A  T  N  A  L

GATCGACAAAGACGGGTGGCTACACTCTGGGGATATATCATACTGGGACGAAGACGGTCACTTCTTCATTGTTGATCGCT   800
  I  D  K  D  G  W  L  H  S  G  D  I  S  Y  W  D  E  D  G  H  F  F  I  V  D  R

TGAAATCTTTGATAAARTACAAAGGCTACCA
L  K  S  L
```

ii. PphLL1 [LH2-LH4]

```
CCGGGTTGCCGAAAGGTGTTATGTTAACTCATTTTAATTACAATACAACTTTGGGCGTACTAAAgtaagtgttgttatat
                     M  L  T  H  F  N  Y  N  T  T  L  G  V  L  K

gtgtgggtaaaagtttaacgtataattacatttgttgcagTTTCTCTTATCCAAACGGTCATCCACAAGTAGGTCTCCTA
                                          F  S  Y  P  N  G  H  P  Q  V  G  L  L

CTTCTTCCGCTGCACCATGCATATGGGTTGTATGGGGGCATTCAACGAATATTTATTGGAGGAAAGGGAATACTGTTAAA
 L  L  P  L  H  H  A  Y  G  L  Y  G  G  I  Q  R  I  F  I  G  G  K  G  I  L  L  K

GAAGTTCGATCAAAATGCCTATCTTAAAACGATTCAGGAATACAAAATCGAGCGTATATCTATAGTACCGTCGATGGCAA
  K  F  D  Q  N  A  Y  L  K  T  I  Q  E  Y  K  I  E  R  I  S  I  V  P  S  M  A

ATTTTTTGGCTAAAAGTGAGCTGGTTAACAAGTACGATCTTTCTAGCGTAAAGTCAATATCATGTGGTGCTGCACCTTTA
N  F  L  A  K  S  E  L  V  N  K  Y  D  L  S  S  V  K  S  I  S  C  G  A  A  P  L

AGTGATAATATCCAACACAGCGTGATGAAACGgtaaggattgtgcagatacaactgctgtactcccaaacgtttttgagt
 S  D  N  I  Q  H  S  V  M  K  R

acatcaaaactccaaaacctgcaccctggcaaatatgtgacatggtaacgattcaccgctttctcccagaatattgcagt

tttgacggatagacggacaaaactatttgcgaccgtttctcacctataggatataagctatcacatagcaaaaaccgaga

gtaactacgacatctaggggctgaatttagcgtataccgacggacagacagacggaaaaatgagaacataagcacccctt

gtgtgggttgtggtaacaatcttagttatttaaatcacgatatcacgtttatgatactgattgggatctttctcattagA

TTAAAAGTGGATTCCATTCGGCAGGGATATGGGATGACTGAATCAACCATTGCGATCTTAGTTCCTCCGTCCAATAAGGA
 L  K  V  D  S  I  R  Q  G  Y  G  M  T  E  S  T  I  A  I  L  V  P  P  S  N  K  E

AAACATTCCTCCTGGTTCATCTGGCTGTGCGTTGCCAAACATTTCAATAAAAATAGTAGATATTGATACAGGAAGATCAT
  N  I  P  P  G  S  S  G  C  A  L  P  N  I  S  I  K  I  V  D  I  D  T  G  R  S

TAGGTCCCATGAAAGCGGGAGAGTTATGGTGCCGTGGTGACATGGTTATGAAGGGATACTTTAATAATCCTGAGGCTACT
L  G  P  M  K  A  G  E  L  W  C  R  G  D  M  V  M  K  G  Y  F  N  N  P  E  A  T

CGAAATACAATTGATGAACATGGATGGTTGCATACAGGAGATATAGCTTATTATGATCACGATTGTCTGTTTTACATAGT
 R  N  T  I  D  E  H  G  W  L  H  T  G  D  I  A  Y  Y  D  H  D  C  L  F  Y  I  V

GGATAGGCTTAAAGAATTAATAAAGTATAAAGGCTACCA
  D  R  L  K  E  L
```

c. *Nyctophila reichii* Jacquelin du Val, 1859

i. NyrLUC

```
ACTGGATTTCCAAAGGGGGTTGAGCTTACTCACAAAAATGTTTGTGTTAGATTTTCTCACTGCAGgtacgtatggtcttg   80
                      E  L  T  H  K  N  V  C  V  R  F  S  H  C  R

aaactataaaaacgtaattgtattcaaaattttcagAGATCCTGTGTTTGGTAATCAAATTATCCCCGATACTGCGATT   160
                                      D  P  V  F  G  N  Q  I  I  P  D  T  A  I

TTAACAGTTATGCCATTHCATCATGGTTTTGGAATGTTTACAACACTAGGATATTTAACGTGTGGATTTCGTATTGTGCT   240
 L  T  V  M  P  X  H  H  G  F  G  M  F  T  T  L  G  Y  L  T  C  G  F  R  I  V  L

TATGTATAGATTTGAAGAGGAATTATTTTTACGATCACTTCAAGATTATAAAATTCAAAGTGCGTTGCTGGTACCTACCC   360
  M  Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T

TATTTTCATTCTTTGCCAAAAGCACCTTAGTCGACAAATACGATTTATCCAACTTACATGAAATTGCTTCTGGTGGAGCT   400
L  F  S  F  F  A  K  S  T  L  V  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A

CCCCTCGCGAAAGAAGTTGGAGAAGCTGTAGCAAAACGgtaagtcacgataccaagtactcagtttctattaaggctttg   480
 P  L  A  K  E  V  G  E  A  V  A  K  R

tagTTTTAAGCTACCGGGCATACGACAAGGGTATGGACTTACTGAAACTACCTCAGCTATTATAATTACACCAGAAGGGG   560
     F  K  L  P  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  I  I  T  P  E  G

ATGATAAACCAGGAGCATGTGGTAAAGTTGTTCCATTCTTTTCTGCCAAAATTGTTGATCTGGATACGGGTAAAACTTTG   640
D  D  K  P  G  A  C  G  K  V  V  P  F  F  S  A  K  I  V  D  L  D  T  G  K  T  L

GGTGTTAATCAGAGGGGGGAATTATGTGTGAAAGGTCCAATGATAATGAAGGGTTACGTAAACAWCCCAGAAGCAACAAG   720
 G  V  N  Q  R  G  E  L  C  V  K  G  P  M  I  M  K  G  Y  V  N  X  P  E  A  T  S

TGCATTGATAGACAAAGATGGATGGTTACACTCTGGTGACATAGCTTACTACGACAAAGATGGYCACTTCTTCATAGTAG   800
  A  L  I  D  K  D  G  W  L  H  S  G  D  I  A  Y  Y  D  K  D  G  H  F  F  I  V

ATCGTTTGAAATCGTTAATYAAATWCAAAGGTTAYCA
D  R  L  K  S  L
```

ii. NyrLL2

```
CTGGTTTTCCGAAAGGTGTCATGCTAACACATTTCAATTATAATGTATCTATTGCAATTTTTGAgtaagtatttattcac   80
                     M  L  T  H  F  N  Y  N  V  S  I  A  I  F  E

acgtcactagcaaggttttgctctacttttggctaactcggtaaaatctgaaattaattcatttttgtagTGACGTTAT   160
                                                                        D  V  I

CCAATCAAGAGCACGAAATACACCTGCCGTGGCTTTTGTTCCGTTATATCACGCATATGGTTTATTCCTAGTATCGTTAA   240
  Q  S  R  A  R  N  T  P  A  V  A  F  V  P  L  Y  H  A  Y  G  L  F  L  V  S  L

AGATTTTATGGGGTGGGATTGTTGTAATAATGAAAAAGTTTAATCCCGAACTCTATCTCAAAACTGTACAGGATTATAAA   320
K  I  L  W  G  G  I  V  V  I  M  K  K  F  N  P  E  L  Y  L  K  T  V  Q  D  Y  K

ATTGGTGACATCAATATTGTACCATCCATCGCACAATTTTTGGTAAAGAGTGATTTGGTAAATAAATATGATCTATCAAG   400
 I  G  D  I  N  I  V  P  S  I  A  Q  F  L  V  K  S  D  L  V  N  K  Y  D  L  S  S

TATCAAAGCAATATATTCGGGTGCAGCTCCCTTAAGTAAGGACGTTGAACTGGCTTTAATAGAACGgtgagttgacatga   480
  I  K  A  I  Y  S  G  A  A  P  L  S  K  D  V  E  L  A  L  I  E  R

tcagagagagcgctctcaaaaagcattaacaagtactaatattaatcgttttcaatagATTTAAAGTGAAAGATATTCAG   560
                                                            F  K  V  K  D  I  Q

CAAGGGTATGGAATGACAGAAACTACAGTTGGAGCAATAAGTCACTTACATAATACAATGGAGAATAGTCATGGTTCATG   640
 Q  G  Y  G  M  T  E  T  T  V  G  A  I  S  H  L  H  N  T  M  E  N  S  H  G  S  C

TGGTTGCATCTTACCAAGCCTTTCAGCAAAAATAGTAGACGTAGATACAGGAAAGTCGTTGGGTCCTATGCAAGCTGGTG   720
 G  C  I  L  P  S  L  S  A  K  I  V  D  V  D  T  G  K  S  L  G  P  M  Q  A  G

AATTATGGTGCCGTGGTGGAGTGGTAATGAAAGGATATTTAAACAATCCAGACGCAACAAAAGATGCTATAGATGAAGAA   800
E  L  W  C  R  G  G  V  V  M  K  G  Y  L  N  N  P  D  A  T  K  D  A  I  D  E  E

GGATGGCTACATACAGGCGATATCGCTTATTATGACAATAACTTTGTGTTTTACATAGTGGACAGACTTAAGGAAATAAT   880
 G  W  L  H  T  G  D  I  A  Y  Y  D  N  N  F  V  F  Y  I  V  D  R  L  K  E  I

AAACGNTNTAAGGGNACCA
```

iii. NyrLL1

```
ACCGGTCTGCCCAAGGGCGTAGGACTATCACATGGAAACTTTATTGCTACATTAAGCGTGTATAGgtaagcaagtaggtg   80
                      G  L  S  H  G  N  F  I  A  T  L  S  V  Y  R

gtgatttctaatctagattttgcattgtaacagAGATCCACGTTACCTAAATGTTTCCAACGATGTTACCGTTTCCGTGA   160
                                   D  P  R  Y  L  N  V  S  N  D  V  T  V  S  V

TTCCCTTTTTCCATATCTATGGTTTCGTCACATACATTTCGGCAGTCTTTTGTACTTTAAAAGTTGTACTGATGAAAAAG   240
I  P  F  F  H  I  Y  G  F  V  T  Y  I  S  A  V  F  C  T  L  K  V  V  L  M  K  K

CTAGAAAGCGAATTATTTCTGAGAGCGATCAAGGACTACAAGTGTACCAGATTGTTCCTTGTTCCGACTTTACTGCACTA   320
 L  E  S  E  L  F  L  R  A  I  K  D  Y  K  C  T  R  L  F  L  V  P  T  L  L  H  Y

CTTTGTGAAAAATTCAAAGGTTAATGCAGATGTTCTTTCAAGCGTAAAATTTATACATATAACTGCGGCTGCATTAGGCA   400
  F  V  K  N  S  K  V  N  A  D  V  L  S  S  V  K  F  I  H  I  T  A  A  A  L  G

AAACTGTATATCAGGCAGTCTTGCAAAAgtaagaagagtgaccaattttcatctaagttaagtaatattatttctagATT   480
K  T  V  Y  Q  A  V  L  Q  K                                                   F

TAAACACATAACCGTAATGCAGATGTATGGAGCAACTGAAACAGGTGGTTCATGCACTGTACAAAAAGTTACCGACAACA   560
  K  H  I  T  V  M  Q  M  Y  G  A  T  E  T  G  G  S  C  T  V  Q  K  V  T  D  N

CCAATACCATTGGATATCTTGTTCCAAACATCATTTGTAAAGTAGTAAATCCAAATACAAACAGAACTCTCCGCTCTTTT   640
T  N  T  I  G  Y  L  V  P  N  I  I  C  K  V  V  N  P  N  T  N  R  T  L  R  S  F

CAGTATGGCGAATTGTGTTTTAAGGGAACTAACGTAATGAAAGGGTACTATAAAAATCCTACGGAAACATGTACTGCAAT   720
 Q  Y  G  E  L  C  F  K  G  T  N  V  M  K  G  Y  Y  K  N  P  T  E  T  C  T  A  I

CGACAGTGAAGGATTTTACCATACTGGAGATATTGGTTATTACAACGAAAATGGGCAGTTTTTCATTATAGACAGAATTA   800
  D  S  E  G  F  Y  H  T  G  D  I  G  Y  Y  N  E  N  G  Q  F  F  I  I  D  R  I

AAGACATAATCAANTNCAAAGGGTACCA
K  D  I
```

d. ***Lampyris iberica*** **Figueira, Geistardt & Day 2008**

i. LaiLUC

```
ACTGGATTTCCAAAGGGGGTTGAGCTTACTCATCAAAATGTTTGTGTTAGATTTTCTCACTGCAGgtacgtggtcttgaa   80
                      E  L  T  H  Q  N  V  C  V  R  F  S  H  C  R

atgataaaaatgtaattgtattcaaaaatttgcagAGATCCTGTGTTTGGTAATCAAATTATTCCCGATACTGCGATTTT   160
                                     D  P  V  F  G  N  Q  I  I  P  D  T  A  I  L

AACAGTTATTCCATTTCATCATGGTTTTGGAATGTTTACAACACTAGGATATTTAACGTGTGGATTTCGTATTGTGCTTA   240
  T  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  T  C  G  F  R  I  V  L

TGTATAGATTTGAAGAGGAATTATTTTTACGATCACTTCAAGATTATAAAATTCAAAGTGCGTTGCTGGTACCTACCCTA   320
M  Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T  L

TTTTCATTCTTTGCCAAAAGCACCTTAGTCGATAAATACGATTTATCCAACTTACATGAAATTGCGTCTGGTGGAGCTCC   400
 F  S  F  F  A  K  S  T  L  V  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P

CCTCGCGAAAGAAGTTGGAGAAGCTGTAGCAAAACGgtaagtcacaataccaagtactcagtgcctattaaggctttgta   480
  L  A  K  E  V  G  E  A  V  A  K  R

gTTTTAAGCTGCCGGGAATACGACAAGGGTATGGACTTACTGAAACTACCTCAGCTATTATAATTACACCAGAAGGGGAT   560
   F  K  L  P  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  I  I  T  P  E  G  D

GATAAACCAGGAGCATGTGGTAAAGTTGTTCCATTCTTTTCTGCCAAAATTGTTGATCTGGATACGGGTAAAACTTTGGG   640
 D  K  P  G  A  C  G  K  V  V  P  F  F  S  A  K  I  V  D  L  D  T  G  K  T  L  G

TGTTAATCAGAGGGGGGAATTATGTGTGAAAGGCCCAATGATAATGAAGGGTTACGTAAACAACCCAGAAGCAACAAGTG   720
  V  N  Q  R  G  E  L  C  V  K  G  P  M  I  M  K  G  Y  V  N  N  P  E  A  T  S

CATTGATAGACAAAGATGGTTGG ... incomplete sequence
A  L  I  D  K  D  G  W
```

ii. LaiLL1

```
ACCGGTCTCCCTAAGGGTGTGAATTTAAGTCATGAAAACATATTTCCATTGATAAACGTTACAGAgtaaggttgctatcg   80
                         N  L  S  H  E  N  I  F  P  L  I  N  V  T  E

tgtgtgtgtgtgtgtgtgtgtgtgactttgagaaacctttcagGGACGAACGATACATAAACCTCACGGAGCACGATTG   160
                                             D  E  R  Y  I  N  L  T  E  H  D  C

CTTGGTATCGTTTTTACCCTTTTACCATATTTATGGTTTTTCTGCGCACTTGTCATCAATAACGGCTCGCGCAAAAATTA   240
 L  V  S  F  L  P  F  Y  H  I  Y  G  F  S  A  H  L  S  S  I  T  A  R  A  K  I

TAGTTATGGAAAGGTTTGTACCTGACACATATCTGAAATTGATCGAACAACACCATGCAACAAAGCTATTTGTCGTTCCT   320
I  V  M  E  R  F  V  P  D  T  Y  L  K  L  I  E  Q  H  H  A  T  K  L  F  V  V  P

AGCGTATTATTATTTTTAGTCAAGAATGAAATAGTCAATGAATTTAATCTTTCGAGTATTAAAAGTATTTTTGTGTCGGG   400
 S  V  L  L  F  L  V  K  N  E  I  V  N  E  F  N  L  S  S  I  K  S  I  F  V  S  G

TGCACCATTGGGAACCGATTTGTACAAAGCGGCTATAGCAAGgtttggcaatgtgcatcagccacaatttctttacttat   480
 A  P  L  G  T  D  L  Y  K  A  A  I  A  R

gctaacgtagACTTCAAGTGCCAATTCGACAAATGTATGGCTCTACCGAAACCGCTGCGGCATGTGCAATTCAAGATGTG   560
           L  Q  V  P  I  R  Q  M  Y  G  S  T  E  T  A  A  A  C  A  I  Q  D  V

GGCTACAAATATGAATCCGTTGGAGGTCTCATACCAAATTTATCATGTAAAGTATTAGATTTATCCAATCATAAATCAGT   640
 G  Y  K  Y  E  S  V  G  G  L  I  P  N  L  S  C  K  V  L  D  L  S  N  H  K  S  V

GGGCCCCTCTCACATTGGTGAATTGTGCTTTAAAGGCGTTAATGTCATGAAGGGTTACTACAACAACGAAACAGCTACAC   720
  G  P  S  H  I  G  E  L  C  F  K  G  V  N  V  M  K  G  Y  Y  N  N  E  T  A  T

GAAATACATTTGAGGAGGATGGTTTTTATCGTACTGGTGATGTTGGATATTATAATGATGAAGGAAATTTTTTTATTGTG   800
R  N  T  F  E  E  D  G  F  Y  R  T  G  D  V  G  Y  Y  N  D  E  G  N  F  F  I  V

GACAGAGTAAAAGATTTAATAAAGTACAAAGGCTACCA
 D  R  V  K  D  L
```

e. *Lampyris sardiniae* Geisthardt, 1987

i . LasLUC

```
ACYGGWYTWCCTAAGGGTGTTGAGCTTACTCACCTAAAYGTTTGTGTTAGATTTTCTCACTGCAGgtacgtatggtcttg   80
                         E  L  T  H  L  N  V  C  V  R  F  S  H  C  R

aaactataaaaatgtaattgtattcaaaaatttgcagAGATCCTGTGTTTGGTAATCAAATTATTCCCGATACTGCGATT   160
                                        D  P  V  F  G  N  Q  I  I  P  D  T  A  I

TTAACAGTTATACCATTTCATCATGGTTTTGGAATGTTTACAACACTAGGATATTTAACGTGTGGATTTCGTATTGTGCT   240
 L  T  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  T  C  G  F  R  I  V  L

TATGTATAGATTTGAAGAGGAATTATTTTTACGATCACTTCAGGATTATAAAATTCAAAGTGCGTTGCTGGTACCTACCC   320
  M  Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T

TATTTTCATTCTTTGCCAAAAGCACCTTAGTCGACAAATACGATTTATCAAACTTACATGAAATTGCTTCTGGTGGAGCA   400
L  F  S  F  F  A  K  S  T  L  V  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A

CCCCTCGCGAAAGAAGTTGGAGAAGCTGTAGCAAAACGgtaagtcacgataccaagtactcagtttctattaaggctttg   480
 P  L  A  K  E  V  G  E  A  V  A  K  R

tagTTTTAAGCTGCCGGGTATACGACAAGGGTATGGACTTACTGAAACTACCTCAGCTATTATAATTACACCAGAAGGGG   560
    F  K  L  P  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  I  I  T  P  E  G

ATGATAAACCAGGAGCATGTGGTAAAGTTGTTCCATTCTTTTCTGCCAAAATTGTTGATCTGGATACGGGTAAAACTTTG   640
D  D  K  P  G  A  C  G  K  V  V  P  F  F  S  A  K  I  V  D  L  D  T  G  K  T  L

GGTGTTAATCAGAGGGGGGAATTATGTGTGAAAGGGCCAATGATAATGAAGGGTTACGTAAACAACCCAGAAGCAACAAG   720
 G  V  N  Q  R  G  E  L  C  V  K  G  P  M  I  M  K  G  Y  V  N  N  P  E  A  T  S

TGCATTGATAGACAAAGATGGATGGTTACACTCTGGTGACATAGCTTACTACGACAAAGATGGTCACTTCTTCATAGTAG   800
 A  L  I  D  K  D  G  W  L  H  S  G  D  I  A  Y  Y  D  K  D  G  H  F  F  I  V

ATCGTTTGAAATCGTTAATAAAATCTCNAAGGNTACCA
D  R  L  K  S  L
```

ii. LasLL1

```
ACTGGGTTGCCTAAAGGTGTGAATTTAAGTCATGAAAACATATTTCCATTGATAAACATTATAGAgtaaggttcctatcg   80
                     N  L  S  H  E  N  I  F  P  L  I  N  I  I  E

tgtgtgtatgattgactttgagaaacctttttagAGACGAACGATACTTAAACCTCACGAAGCACGATTGCTTGGTATCG   160
                                  D  E  R  Y  L  N  L  T  K  H  D  C  L  V  S

TTTTTACCCTTTTACCATTTTTATGCTTTTGCTTTGCACTTGTCATCAATAACGGCTTGCTCAAAAATTACAGTTATGGA   240
 F  L  P  F  Y  H  F  Y  A  F  A  L  H  L  S  S  I  T  A  C  S  K  I  T  V  M  E

AAGGTTTGTACCTGACACATATCTGAAACTGATCGAACAACAGCACGCAACAAAGCTACTTGTCGTTCCTAGCGTATTAT   320
  R  F  V  P  D  T  Y  L  K  L  I  E  Q  Q  H  A  T  K  L  L  V  V  P  S  V  L

TATTTTTAGTCAAGAATGAAATAGTCAACCAATTTAATCTTTCGAGTATTAAAAGTATTTTTGTGGCGGGCGCACCATTG   400
l  f  l  v  K  N  E  I  V  N  Q  F  N  L  S  S  I  K  S  I  F  V  A  G  A  P  L

GGAACCGTTTTGTACAAAGAGGCGATAGCAAGgtttcgcaatgtgcatcagccacaatttctttacttattctaacgtag   480
 G  T  V  L  Y  K  E  A  I  A  R

ATTTCAAGTGCCAATTAGACAAATGTATGGGTCGACCGAAACCGGTGGAATATGTACAGTTCAAGATGTGGGCTGCAAAT   560
  F  Q  V  P  I  R  Q  M  Y  G  S  T  E  T  G  G  I  C  T  V  Q  D  V  G  C  K

ATGAATCCGTAGGAGGTCTCATACCAAATTTATCATGTAAAGTATTAGATTTATCCAATCATAAATCAGTGGGCCCCTCT   640
Y  E  S  V  G  G  L  I  P  N  L  S  C  K  V  L  D  L  S  N  H  K  S  V  G  P  S

CACATTGGTGAATTGTACTTTAAAGGCGTTAATGTTATGAAAGGTTACTACAACAACGAAACAGCTACACGAAAAACATT   720
 H  I  G  E  L  Y  F  K  G  V  N  V  M  K  G  Y  Y  N  N  E  T  A  T  R  K  T  F

TAAGGAGGATGGTTTTTATCGTACTGGTGATGTTGGATATTATAATGTCGAAGGAAATTTTTTTATTGTGGACAGAGTAA   800
  K  E  D  G  F  Y  R  T  G  D  V  G  Y  Y  N  V  E  G  N  F  F  I  V  D  R  V

AAGATTTAATTAAGNATNAAAGGCTACCAAATC
K  D  L
```

f. Lampyris noctiluca Linnaeus, 1758

i. LanLUC

```
CTGGTCTTCCGAAGGGGGTTGAGCTTACTCACCAAAATGTTTGTGTTAGATTTTCTCACTGCAGgtacgtggtcttgaaa
                    E  L  T  H  Q  N  V  C  V  R  F  S  H  C  R

ttataaaaatgtagttgaattcaaaaatttgcagAGATCCTGTGTTTGGTAATCAAATTATTCCCGATACTGCGATTTTA
                                  D  P  V  F  G  N  Q  I  I  P  D  T  A  I  L

ACAGTTATACCATTTCATCATGGTTTTGGAATGTTTACAACACTAGGATATTTAACGTGTGGATTTCGTATTGTGCTTAT
 T  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  T  C  G  F  R  I  V  L  M

GTATAGATTTGAAGAGGAATTATTTTTACGATCACTTCAAGATTATAAAATTCAAAGTGCGTTGCTGGTACCTACCCTAT
  Y  R  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T  L

TTTCATTCTTTGCCAAAAGCACCTTAGTCGATAAATACGATTTATCCAACTTACATGAAATTGCGTCTGGTGGAGCTCCC
F  S  F  F  A  K  S  T  L  V  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P

CTCGCGAAAGAAGTTGGAGAAGCTGTAGCAAAACGgtaagtcacgataccaagtactcagtgcctattaaggctttgtag
 L  A  K  E  V  G  E  A  V  A  K  R

TTTTAAGCTGCCGGGAATACGACAAGGGTATGGACTTACTGAAACTACCTCAGCTATTATAATTACACCAGAAGGGGATG
  F  K  L  P  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  I  I  T  P  E  G  D

ATAAACCAGGAGCATGTGGTAAAGTTGTTCCATTCTTTTCTGCCAAAATTGTTGATCTGGATACGGGTAAAACTTTGGGT
D  K  P  G  A  C  G  K  V  V  P  F  F  S  A  K  I  V  D  L  D  T  G  K  T  L  G

GTTAATCAGAGGGGGGAATTATGTGTGAAAGGCCCAATGATAATGAAGGGTTACGTAAACAACCCAGAAGCAACAAGTGC
 V  N  Q  R  G  E  L  C  V  K  G  P  M  I  M  K  G  Y  V  N  N  P  E  A  T  S  A

ATTGATAGACAAAGATGGTTGGTTACACTCTGGTGACATAGCTTACTACGACAAAGATGGTCACTTCTTCATAGTAGATC
  L  I  D  K  D  G  W  L  H  S  G  D  I  A  Y  Y  D  K  D  G  H  F  F  I  V  D

GTTTGAAATCGCTAATTAAGTACAAAGGTTACCA
R  L  K  S  L
```

ii. LanLL2

```
CTGGATTGCCTAAGGGTGTAGGACTATCACATGGAAACTTTATTGGTACATTAAGCGTGTATAGgtaagcaagtaggtgg
                      G  L  S  H  G  N  F  I  G  T  L  S  V  Y  R

tgatttctaatctaaattttgcattttaatagAGATCCACGTTACCTAAATGTTTCTAACGATGTTACCGTTTCCGTGAT
                                    D  P  R  Y  L  N  V  S  N  D  V  T  V  S  V  I

TCCCTTTTTCCATATCTATGGTTTCGTCACACACATTTCGGCAGTCTTTTGTACTTTAAAAGTTGTACTGATGAAAAAGC
  P  F  F  H  I  Y  G  F  V  T  H  I  S  A  V  F  C  T  L  K  V  V  L  M  K  K

TAGAAAGCGAATTATTTTTGAGAGCGATCAAGAACTACAAGTGTACCAGATTGTTCCTTGTTCCGACTTTACTGCAGTAC
L  E  S  E  L  F  L  R  A  I  K  N  Y  K  C  T  R  L  F  L  V  P  T  L  L  Q  Y

TTTGTGAAAAGTTCAAAGGTTAATGCAGATGTTCTTTCAAGCGTAAAATTTATACATATAACTGCGGCTGCATTAGGCAA
 F  V  K  S  S  K  V  N  A  D  V  L  S  S  V  K  F  I  H  I  T  A  A  A  L  G  K

AACTATATACCAGGCAGTCTTGCAAAAgtaagaagagtgatcaattttcatctaagttaagtaatattatttctagATTT
  T  I  Y  Q  A  V  L  Q  K                                                    F

AAACACATAACCGTAATGCAGATGTATGGAGCAACTGAAACAGGTGGTTCATGCACTGTACAAAAAGTTACCGACAACAC
 K  H  I  T  V  M  Q  M  Y  G  A  T  E  T  G  G  S  C  T  V  Q  K  V  T  D  N  T

CAATAGCATAGGATATCTTGTTCCAAACATCATTTGTAAAGTAGTAAATCCAAATACAAACAGAACTCTCCGCCCTTTTC
  N  S  I  G  Y  L  V  P  N  I  I  C  K  V  V  N  P  N  T  N  R  T  L  R  P  F

AGTATGGCGAATTGTATTTTAAGGGGACTAACGTAATGAAAGGGTATTATAAAAATCCTACGGAAACATGTAATGCAATC
Q  Y  G  E  L  Y  F  K  G  T  N  V  M  K  G  Y  Y  K  N  P  T  E  T  C  N  A  I

GACAGTGACGGATTTTACCATACTGGAGATGTTGGTTATTACAACGAAAATGGGCAGTTTTTCATTATGGGCAGAATTAA
 D  S  D  G  F  Y  H  T  G  D  V  G  Y  Y  N  E  N  G  Q  F  F  I  M  G  R  I  K

AGACATAATAAAGTACAAAGGCTACCA
  D  I
```

iii. LanLL1

```
CCGGGCTACCGAAGGGTGTGAATTTAAGTCATGAAAACATATTTCCATTGATAAACATTATAGAgtaaggttcctatcgt 80
                      N  L  S  H  E  N  I  F  P  L  I  N  I  I  E

gtgtgtgattgactttgagaaacctttcagAGACGAACGATACATAAACCTCACGAAGCACGATTGCTTGGTATCGTTT
                                  D  E  R  Y  I  N  L  T  K  H  D  C  L  V  S  F

TTACCCTTTTACCATTTTTATGCTTTTGCTTTGCACTTGTCATCAATAACGGCTAGCTCAAAAATTACAGTTATGGAAAG
 L  P  F  Y  H  F  Y  A  F  A  L  H  L  S  S  I  T  A  S  S  K  I  T  V  M  E  R

GTTTGTACCTGACACATATCTGAAACTGATCGAACAACAACATGCAACAAAGCTATTTGTCGTTCCTAGCGTATTATTAT
  F  V  P  D  T  Y  L  K  L  I  E  Q  Q  H  A  T  K  L  F  V  V  P  S  V  L  L

TTTTAGTCAAGAATGAAATAGTCAACCAATTTAGTCTTTCGAGTATTAAAAGTATTTTTGTGGCGGGCGCACCATTGGGA
F  L  V  K  N  E  I  V  N  Q  F  S  L  S  S  I  K  S  I  F  V  A  G  A  P  L  G

ACCGATTTGTACAAAGAGGCTATAACAAGgtttggcaatgtgcatcagccacaatttctttacttattctaacgtagATT
 T  D  L  Y  K  E  A  I  T  R                                                  F

TCAAGTGCCAATTAGACAAATCTATGGCTCGACCGAAACCGGTGGAATATGTGCAGTTCAAGATGTGGGCTGCAAATGTG
  Q  V  P  I  R  Q  I  Y  G  S  T  E  T  G  G  I  C  A  V  Q  D  V  G  C  K  C

AATCCGTAGGAGGTCTCATACCAAATTTATCATGTAAAGTATTAGATTTATCCAATCATAAATCAGTGGGCCCCTCTCAC
E  S  V  G  G  L  I  P  N  L  S  C  K  V  L  D  L  S  N  H  K  S  V  G  P  S  H

ATTGGTGAATTGTGCTTTAAAGGCGTTAATGTTATGAAAGGTTACTACAACAACGAAACAGCTACACGAAATACATTTGA
 I  G  E  L  C  F  K  G  V  N  V  M  K  G  Y  Y  N  N  E  T  A  T  R  N  T  F  E

GGAGGATGGTTTTTATCGTACTGGTGATGTTGGATATTATAATGACGAAGGAAATTTTTTTATTGTGGACAGAGTAAAAG
  E  D  G  F  Y  R  T  G  D  V  G  Y  Y  N  D  E  G  N  F  F  I  V  D  R  V  K

ATTTAATTAAATACAAGGGCTACCA
D  L
```

g. *Phausis reticulata* Say, 1825

i. ParLUC

```
ACYGGTCTGCCTAAGGGTGTAGCGTTATCCCATAAGAATGTAGTTGTAAGATTTTCTCATTGCAAgtaagtgtaattcac   80
                      A  L  S  H  K  N  V  V  V  R  F  S  H  C  K

taacgtgtatacgtgtatttaaaaattccatttacagAGACCCAGTGTTTGGTAATCAAATTGTACCAGATACCGCTATT  160
                                       D  P  V  F  G  N  Q  I  V  P  D  T  A  I

TTAACTGTTATTCCATTCCATCATGGATTTGGTATGTTTACCACATTAGGATATTTAACATGCGGATTCCGTATTGTGCT  240
 L  T  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  T  C  G  F  R  I  V  L

AATGAACAAATTTGAAGAAGAGTTATTTTTGCGTTCACTTCAAGATTACAAAATTCAAAGTGCACTACTTGTGCCAACTT  320
  M  N  K  F  E  E  E  L  F  L  R  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T

TGTTTTCGTTCTTCGCTAAAAGCACACTAGTCGATAAATATGATTTATCACATTTAGAAGAAATCGCTTCTGGTGGAGCT  400
L  F  S  F  F  A  K  S  T  L  V  D  K  Y  D  L  S  H  L  E  E  I  A  S  G  G  A

CCTCTTGCTAAAGAAGTGGGCGAAGCTGTTGCAAAACGgttggtactaattattaaatcttccactaataattattatgg  480
 P  L  A  K  E  V  G  E  A  V  A  K  R

catgttacaattttacagTTTCAAACTGCCCGGAGTGAGACAAGGTTATGGATTAACCGAAACAACATCAGCTTGTATTA  560
                    F  K  L  P  G  V  R  Q  G  Y  G  L  T  E  T  T  S  A  C  I

TTACTCCAGAAGGTGACGATAAACCTGGATCCACGGGAAAAGTGGTTCCTTTCTTTTCTGCTAAAATTATCGATCTTGAT  640
I  T  P  E  G  D  D  K  P  G  S  T  G  K  V  V  P  F  F  S  A  K  I  I  D  L  D

ACTCGCAAAAGTTTAGGCCCTAACCAACGCGGAGAATTATGCCTTAAAGGTGATATGATTATGAAAGGCTACGTTAATAA  720
 T  R  K  S  L  G  P  N  Q  R  G  E  L  C  L  K  G  D  M  I  M  K  G  Y  V  N  N

TCCAGAAGCCACAAATGCGTTGATTGACAAAGACGGGTGGTTGCATTCTGGTGACATCGCGTACTACGATGAGGACGGTC  800
  P  E  A  T  N  A  L  I  D  K  D  G  W  L  H  S  G  D  I  A  Y  Y  D  E  D  G

ACTTCTTCATTGTCGATCGTTTGAAATCTTTAATAAAGTACAAAGGCTACCAAATCAC
H  F  F  I  V  D  R  L  K  S  L
```

h. *Luciola italica* Linnaeus, 1767

i. LuiLUC1

```
ACTGGATTTCCGAAAGGAGTAGAGATTACCCACGAAGGAACAGTTACAAGATTCTCACACGCTAAgtaaacagtttttta   80
                      E  I  T  H  E  G  T  V  T  R  F  S  H  A  K

gaaacaaatctatgattcaatttgtttatttttagGGATCCAATTTACGGAAACCAAGTTTCACCTGGTACTGCTATTTT  160
                                     D  P  I  Y  G  N  Q  V  S  P  G  T  A  I  L

AACTGTCGTTCCGTTCCATCATGGATTTGGAATGTTTACCACTTTAGGATACTTTGCTTGTGGATACCGTATTGTAATGT  240
  T  V  V  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  F  A  C  G  Y  R  I  V  M

TAACAAAATTCGATGAAGAACTATTTTTGAGAACTTTGCAAGATTATAAGTGTACCAGTGTTATTCTTGTACCAACGTTA  320
L  T  K  F  D  E  E  L  F  L  R  T  L  Q  D  Y  K  C  T  S  V  I  L  V  P  T  L

TTTGCTATTCTCAACAGGAGTGAATTGCTCGATAAGTTCGATTTATCTAATCTAACTGAAATTGCTTCTGGTGGAGCTCC  400
 F  A  I  L  N  R  S  E  L  L  D  K  F  D  L  S  N  L  T  E  I  A  S  G  G  A  P

TTTGGCAAAAGAAATTGGTGAAGCAGTCGCTAGAAGgtaattttgagtatgcaccattattttgctctttctaattttg  480
  L  A  K  E  I  G  E  A  V  A  R  R

tttgtagATTTAATCTACCCGGTGTCCGTCAGGGTTACGGATTGACAGAAACGACATCTGCATTTATTATTACCCCAGAA  560
         F  N  L  P  G  V  R  Q  G  Y  G  L  T  E  T  T  S  A  F  I  I  T  P  E

GGTGATGATAAACCTGGAGCATCTGGAAAAGTAGTACCCTTATTCAAAGTAAAAATTATTGATCTTGACACTAAAAAAAC  640
 G  D  D  K  P  G  A  S  G  K  V  V  P  L  F  K  V  K  I  I  D  L  D  T  K  K  T

TTTGGGTGTCAACCGACGAGGAGAGATCTGTGTAAAAGGTCCGAGTCTTATGTTAGGCTACACAAACAATCCGGAAGCAA  720
  L  G  V  N  R  R  G  E  I  C  V  K  G  P  S  L  M  L  G  Y  T  N  N  P  E  A

CAAGAGAAACTATTGATGAAGAGGGTTGGTTGCACACCGGAGATATTGGATATTACGACGAAGACGAACATTTCTTCATT  800
T  R  E  T  I  D  E  E  G  W  L  H  T  G  D  I  G  Y  Y  D  E  D  E  H  F  F  I

GTAGATCGTTTGAAATCATTAATCAANTNCAAAGGGTACCA
 V  D  R  L  K  S  L
```

ii. LuiLL1

```
ACTGGATTTCCGAAAGGAGTTATGTTAACTCACAAACATCTAATGATTAGATTTAGGCATTGCAGgtacgttattaattt   80
                      M  L  T  H  K  H  L  M  I  R  F  R  H  C  R

ggacaagctcgtattaatgtttattttagAGAACCACAATTTGCCACTGCCTTGATGGTTAAAAAAGGCGAAGCAATATT  160
                               E  P  Q  F  A  T  A  L  M  V  K  K  G  E  A  I  L

GAATTTTATGCCTTTGTTCCATAATTTTGGTTTTATGATTACTCTCGGATATATTTCAATGGGTTTACATATTATTCAAA  240
  N  F  M  P  L  F  H  N  F  G  F  M  I  T  L  G  Y  I  S  M  G  L  H  I  I  Q

TGCAAAAATTTACCGAGGTGAAATTTTTAGAATCTATAGAAAAATATCAAGTTCAAAGCACGTTAGTGGTGCCTCCTATT  320
M  Q  K  F  T  E  V  K  F  L  E  S  I  E  K  Y  Q  V  Q  S  T  L  V  V  P  P  I

ATGATTTTTTTACTTAAAAATAATTTAGTAGAAAAGTACAATTTATCAAGTTTAAAAGAAATTGGTTGTGGAGCTGCGCC  400
 M  I  F  L  L  K  N  N  L  V  E  K  Y  N  L  S  S  L  K  E  I  G  C  G  A  A  P

TTTGTCCAATGATATTATTACAGAAGTTAAGAAAAAgtaatatccaacatttaaactcattttgttatgatttttaatt  480
  L  S  N  D  I  I  T  E  V  K  K  K

taagGTTAAATATAAAATCTGTTCGACAAGGTTATGGACTAACGGAAGTCACACTGTTAGTCTGTATGCACCCTACGAAT  560
      L  N  I  K  S  V  R  Q  G  Y  G  L  T  E  V  T  L  L  V  C  M  H  P  T  N

TCTAAGAAATACGAATCGTCAGGAATACTTATTCCATCACTATCTGCAAAATTTATTGATTTGGATACAGGAAARTCTTT  640
 S  K  K  Y  E  S  S  G  I  L  I  P  S  L  S  A  K  F  I  D  L  D  T  G  K  S  L

AGGGCCTTATCAGCCAGGTGAAATATGTGTAAAAGGAGATGTAGTTATGAAAGGTTATATGGATAATCTTGAAGCAACGC  720
  G  P  Y  Q  P  G  E  I  C  V  K  G  D  V  V  M  K  G  Y  M  D  N  L  E  A  T

AAAATGCGATAGACAAAGATGGCTGGTTTCACACTGGGGATGTTGGATACTACGACGATGATGAATATATCTATGTGGTA  800
Q  N  A  I  D  K  D  G  W  F  H  T  G  D  V  G  Y  Y  D  D  D  E  Y  I  Y  V  V

GATAGAATAAAAGAGCTCATCAANTNCAAAGGGTACCA
 D  R  I  K  E  L
```

iii. LuiLUC2

```
ACTGGATTGCCCAAAGGAGTCGTGCTGACTCACAAAAACCTCGTAGTTAGATTTAACCATTGCAAgtaagtactataact
                      V  L  T  H  K  N  L  V  V  R  F  N  H  C  K

caatgacgtttaatccaaattttcttaagGGATCCTATTTTTGGAAATCAAATTAGTCCGGGAACCGCGATCTTGACCG
                               D  P  I  F  G  N  Q  I  S  P  G  T  A  I  L  T

TCATTCCATTTCACCACGGCTTTGGCATGTTCACAACCTTGGGTTACTTCACATGCGGCTTCCGAATTGTACTAATGCAC
V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  F  T  C  G  F  R  I  V  L  M  H

ACATTTGAGGAGCGCTTGTTCTTGCAATCCCTGCAGGACTACAAGGTTGAAAGCACTTTATTAGTACCGACTTTGATGAC
 T  F  E  E  R  L  F  L  Q  S  L  Q  D  Y  K  V  E  S  T  L  L  V  P  T  L  M  T

GTTTTTTGCTAAAAGTCCTTTGGTGGAGAAATACCATTTGCCCCATTTACAGGAAATTGCGTCGGGTGGCGCTCCACTTC
  F  F  A  K  S  P  L  V  E  K  Y  H  L  P  H  L  Q  E  I  A  S  G  G  A  P  L

CAAAAAAAATAGGAGACGCCGTTTCGCAAAGgtaattattccattaacaattttggtaggtactcagtattgtttccagG
P  K  K  I  G  D  A  V  S  Q  R

TTCAAACTAAAAAAGGCCAGACAAGGTTATGGACTGACCGAAACTACTTCCGCTATTTTAATTACTCCGGAAGGAGAAGA
 F  K  L  K  K  A  R  Q  G  Y  G  L  T  E  T  T  S  A  I  L  I  T  P  E  G  E  E

AGTTTCTGGGTCAACAGGAAAAGTGGTGCCGTTTTTTGCGGCTAAAGTCGTAGATACGGCTACTGGTAAAATCTTAGGGC
  V  S  G  S  T  G  K  V  V  P  F  F  A  A  K  V  V  D  T  A  T  G  K  I  L  G

CTAACGAAGTGGGTGAATTGTATTTTAAAGGCGATATGATTATGAAGGGATACTGTAACAACAGCAAAGCTACTAATGAG
P  N  E  V  G  E  L  Y  F  K  G  D  M  I  M  K  G  Y  C  N  N  S  K  A  T  N  E

ACTATTGACAAAGATGGCTGGTTGCGTTCAGGCGATCTCGGTTACTACAATGAAGATAAACATTTTTTTATTGTTGACCG
 T  I  D  K  D  G  W  L  R  S  G  D  L  G  Y  Y  N  E  D  K  H  F  F  I  V  D  R

GTTAAAATCGGTGATCAANTNCAAAGGGTACCA
 L  K  S  V
```

i. Lampyroidea maculata Geisthardt & Day 2004

i. LdmLUC1

```
ACTGGTTTACCTAAAGGGGTACAAATTACACACGAAGGTACAGTTACAAGATTCTCACACGCTAAgtaaataattgtttc   80
                      Q  I  T  H  E  G  T  V  T  R  F  S  H  A  K

agaaaaaaatctgtaattaatttgtttacttttagGGATCCAATTTACGGAAACCAAGTTTCACCTGGTACTGCTATTTT   160
                                     D  P  I  Y  G  N  Q  V  S  P  G  T  A  I  L

AACTGTCGTTCCGTTCCATCATGGATTTGGCATGTTTACCACTTTAGGATACTTTGCTTGTGGATACCGTATTGTCATGT   240
  T  V  V  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  F  A  C  G  Y  R  I  V  M

TAACAAAATTCGATGAAGAAATATTTTTGAAAACTATGCAAGATTATAAATGTACCAGTGTTATTCTTGTACCAACTTTA   320
L  T  K  F  D  E  E  I  F  L  K  T  M  Q  D  Y  K  C  T  S  V  I  L  V  P  T  L

TTTGGTATTCTCAACAAGAGTGAATTGATCGATAAGTTCGATTTATCTAATCTAACTGAAATTGCTTCCGGTGGAGCTCC   400
 F  G  I  L  N  K  S  E  L  I  D  K  F  D  L  S  N  L  T  E  I  A  S  G  G  A  P

TTTGGCCAAAGAAGTTGGTGAAGCAGTCGCTAGAAGgtaatttggaatatgcaccattattttgttctttctaattttg   480
  L  A  K  E  V  G  E  A  V  A  R  R

tttgtagATTTAATCTACCCGGTGTCCGCCAGGGTTACGGATTAACAGAAACGACATCTGCATTTATTATTACACCAGAA   560
         F  N  L  P  G  V  R  Q  G  Y  G  L  T  E  T  T  S  A  F  I  I  T  P  E

GGTGATGATAAACCTGGAGCATCTGGAAAAGTAGTACCCTTATTTAAAGTAAAAGTTATTGATCTTGATACTAAAAAAAC   640
 G  D  D  K  P  G  A  S  G  K  V  V  P  L  F  K  V  K  V  I  D  L  D  T  K  K  T

TTTGGGTGCCAACCGACGAGGAGAGATTTGTGTAAAAGGTCCCAGTCTTATGAAAGGGTACATAAACAATCCGGAAGCAA   720
  L  G  A  N  R  R  G  E  I  C  V  K  G  P  S  L  M  K  G  Y  I  N  N  P  E  A

CAAAAGAAATTATTGATGAAGAGGGTTGGATGCACACTGGAGATATTGGTTATTACGACGAAGACGAACATTTTTTCATT   800
T  K  E  I  I  D  E  E  G  W  M  H  T  G  D  I  G  Y  Y  D  E  D  E  H  F  F  I

GTAGATCGTTTGAAATCATTAATCAARTACAAAGGSTACCAAATCAC
 V  D  R  L  K  S  L
```

ii. LdmLUC2

```
GGGTGTCGTGCTGACTCACAAAAACCTCGTAGTTAGATTCAACCATTGCAAgtgagtgctttgattaaatgacgcttaaa   80
        V  L  T  H  K  N  L  V  V  R  F  N  H  C  K

ctaattttcttaagGGATCCCATTTTTGGAAATCAAATTAATCCTGGAACCGCGATTTTGACCGTCATTCCATTTCACC   160
                 D  P  I  F  G  N  Q  I  N  P  G  T  A  I  L  T  V  I  P  F  H

ATGGCTTTGGTATGTTCACAACCTTGGGCTACTTCACATGCGGTTTCCGAATTGTACTAATGCACACATTTGAGGAGCGC   240
H  G  F  G  M  F  T  T  L  G  Y  F  T  C  G  F  R  I  V  L  M  H  T  F  E  E  R

TTGTTCTTGCAATCCCTGCAGGACTACAAGGTTGAAAGCACTTTATTGGTACCGACTTTGATGACGTTTTTTGCTAAAAG   320
 L  F  L  Q  S  L  Q  D  Y  K  V  E  S  T  L  L  V  P  T  L  M  T  F  F  A  K  S

TCCTTTGGTGGAAAAATACCATTTGCCCTATTTACAGGAAATTGCGTCGGGTGGCGCCCCGCTTCCAAAAAAAATAGGAG   400
  P  L  V  E  K  Y  H  L  P  Y  L  Q  E  I  A  S  G  G  A  P  L  P  K  K  I  G

ACGCCGTTTCTCAAAGgtaattcttccattaaaatttggtactaagtattgtttacagGTTCAAACTAAAAAAGGCCAG   480
D  A  V  S  Q  R                                              F  K  L  K  K  A  R

ACAAGGTTATGGACTGACCGAAACTACTTCCGCTATTTTAATTACACCGGAAGGAGAAGAGGTTTCTGGGTCAACAGGAA   560
  Q  G  Y  G  L  T  E  T  T  S  A  I  L  I  T  P  E  G  E  E  V  S  G  S  T  G

AAGTGGTGCCGTTTTTTGCGGCTAAAGTCGTAGATACGGCTACTGGCAAAATTTTAGGACCTAACCAAGTGGGTGAATTG   640
K  V  V  P  F  F  A  A  K  V  V  D  T  A  T  G  K  I  L  G  P  N  Q  V  G  E  L

TATTTTAAAGGCGATATGATTATGAAAGGGTACTTGAACAACATCAAAGCCACTAATGAGACTATTGACAAAGATGGGTG   720
 Y  F  K  G  D  M  I  M  K  G  Y  L  N  N  I  K  A  T  N  E  T  I  D  K  D  G  W

GTTGCGTTCAGGCGATCTCGGTTACTACAATGAAGATAAACATTTTTTTATTGTTGACCGGTTAAAATCAGTGATAAAGT   800
  L  R  S  G  D  L  G  Y  Y  N  E  D  K  H  F  F  I  V  D  R  L  K  S  V

ACAAAGGCTACCAAATCAC
```

j. *Lamprohiza splendidula* Linnaeus, 1767

i. LzsLUC

```
GGTNTTCYCAAAGGKGTTATGTTGACACATAAAAATGTTGTTGTCAGATTTTGTCATTGCAAgtaagtgctagtgcatac   80
                 M  L  T  H  K  N  V  V  V  R  F  C  H  C  K

gttacgtaattaaatatatttaacattttattttaagGGATCCTACATTTGGAAATAAAATAAGTCCAGGTACTGCAATT   160
                                      R  D  P  T  F  G  N  K  I  S  P  G  T  A  I

TTAACAGTTATACCCTTTCATCACGGATTTGGAATGTTTACAACCTTAGGATATTTTACTTGCGGATTTCGAATTGTCTT   240
 L  T  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  F  T  C  G  F  R  I  V  L

AATGCACACATTCAATGAAAATCAATTTCTACAATCACTTCAAGATTATAAAGTTGAAAGCACTTTACTTGTACCAACTT   320
  M  H  T  F  N  E  N  Q  F  L  Q  S  L  Q  D  Y  K  V  E  S  T  L  L  V  P  T

TAATGACGTTTTTTGCAAAAAGCGAGCTTGTCGCAAAATATGATCTGTCAAATTTGCAAGAAATTGCGTCTGGAGGAGCT   400
L  M  T  F  F  A  K  S  E  L  V  A  K  Y  D  L  S  N  L  Q  E  I  A  S  G  G  A

CCTTTACCAAAGGAAGTTGGAGAAGCTGTTGCAAAAAGgtaatattttaaaataactttgtaggttactttacacgtc   480
 P  L  P  K  E  V  G  E  A  V  A  K  R

aaaccatagATTTAAACTGGACGCAATTAGACAAGGATTCGGACTAACAGAAACTACATCCGCTGTTTTAATAACACCAG   560
          F  K  L  D  A  I  R  Q  G  F  G  L  T  E  T  T  S  A  V  L  I  T  P

AAAGGGATGTACGGCCTGGATCAACCGGTCAAATAGTACCTTATTTAGCAGTTAAAGTTATTAACCCAAGTAATGGTGCA   640
E  R  D  V  R  P  G  S  T  G  Q  I  V  P  Y  L  A  V  K  V  I  N  P  S  N  G  A

ATATTAGGACCAAACCAGTCCGGAGAATTGTGTTTTAAAGGCGATATGATAATGAAAGGCTATTGTAATGATGTAGAAGC   720
 I  L  G  P  N  Q  S  G  E  L  C  F  K  G  D  M  I  M  K  G  Y  C  N  D  V  E  A

AACAAATGCCATTATTGATAAGGACGGCTGGTTGCATTCTGGAGATAGTGGTTATTTTGACGAAGATGGCCATTTTTACA   800
  T  N  A  I  I  D  K  D  G  W  L  H  S  G  D  S  G  Y  F  D  E  D  G  H  F  Y

TTGTGGACAGACTAAAATCGATCATAAAGTACAAAGGCTACCA
I  V  D  R  L  K  S  I
```

k. *Photuris congener* LeConte 1851

i.PrcLUC1

```
AGGGTGTAGAACTCTCTCATAAGAATGTTGTCCATCGATTTTCTCATTGCCGgtaatgttcattaaccacaatacttatg
       E  L  S  H  K  N  V  V  H  R  F  S  H  C  R

taataaatatcttttgcagAGATCCTGTGTTTGGCAATCAAATTATTCCAGACACAGCTATTTTATCTGTTATACCATTT
                    D  P  V  F  G  N  Q  I  I  P  D  T  A  I  L  S  V  I  P  F

CATCATGGATTTGGTATGTTTACCACCTTAGGGTATTTGGTGTGTGGATTTCGCATTGTCCTGATGTACAGATTTGAAGA
 H  H  G  F  G  M  F  T  T  L  G  Y  L  V  C  G  F  R  I  V  L  M  Y  R  F  E  E

AGAATTGTTTTTGCAATCACTTCAAGATTATAAAATTCAAAGTGCTTTACTTGTTCCTACATTATTTTCTTTCTTTGCTA
  E  L  F  L  Q  S  L  Q  D  Y  K  I  Q  S  A  L  L  V  P  T  L  F  S  F  F  A

AAAGTACTTTAGTTGACAAATATGATTTATCAAATTTACATGAAATTGCTTCTGGCGGTGCTCCACTTGCAAAAGAGGTT
K  S  T  L  V  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L  A  K  E  V

GGAGAAGCTGTGGCCAAACGgttagtataatattacagtataaacaaagtgtaaccttaagtaaaatcatacagTTTTAA
 G  E  A  V  A  K                                                           F  K

ACTAAAAGGAATTCGACAAGGGTATGGTCTCACAGAAACAACATCAGCTATTATAATTACCCCTGAGGGCGATGACAAAG
  L  K  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  I  I  T  P  E  G  D  D  K

CAGGAGCAGCAGGAAAAGTTGTTTCATTCTTTCGTGCAAAAATTGTTGATTTAGATACGGGTAAAATTTTGGGTCCTAAC
A  G  A  A  G  K  V  V  S  F  F  R  A  K  I  V  D  L  D  T  G  K  I  L  G  P  N

CAACGAGGTGAACTGTACGTTAAAGGTCCCATGATAATGACGGGTTACATAAATAATCCAGAAGCTACAAATGCATTAAT
 Q  R  G  E  L  Y  V  K  G  P  M  I  M  T  G  Y  I  N  N  P  E  A  T  N  A  L  I

TGACAAAGATGGGTGGTTGCGTTCTGGTGATATAGCGTATTACGATGAAGACGGTCACTTTTTCATTGTTGACCGGTTAA
  D  K  D  G  W  L  R  S  G  D  I  A  Y  Y  D  E  D  G  H  F  F  I  V  D  R  L

AATCATTGATCAAATACAAAGGCTACC
K  S  L
```

ii.PrcLUC2

```
CTGGTTGCGAAAGGGGTCATGCTAACCCATAAAAACATTGTTGCACGATTTTCTCATTGCAAgtatgtaattaattaatc
                    M  L  T  H  K  N  I  V  A  R  F  S  H  C  K

tatgtttagagcattaatcaattttgatcttcagAGATCCTACTTTTGGTAACGCAATTAATCCAACGACATCAATTTTA
                                   D  P  T  F  G  N  A  I  N  P  T  T  S  I  L

ACAGTAATACCTTTCCACCATGGTTTTGGTATGTTTACAACATTAGGATATTTTACTTGTGGATTCCAAATTGTTCTAAT
 T  V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  F  T  C  G  F  Q  I  V  L  M

GCACACGTTTGAAGAAAAACTATTTTTACAATCATTACAAGATTATAAAGTGGAAAGCACTTTACTTGTGCCAACATTAA
  H  T  F  E  E  K  L  F  L  Q  S  L  Q  D  Y  K  V  E  S  T  L  L  V  P  T  L

TGGCATTTTTTGCAAAAAGTGGGTTAGTTGAAAAATACGATTTATCGCACTTAAAAGAAATTGCATCTGGTGGTGCACCT
M  A  F  F  A  K  S  G  L  V  E  K  Y  D  L  S  H  L  K  E  I  A  S  G  G  A  P

TTATCAAAAGAAATTGGGGAAATGGTTGCAAAAAGgtaaactttaatccccattctgcaatttcatacttagctcatttt
 L  S  K  E  I  G  E  M  V  A  K  R

agGTTTAAATTAAACTTTGTCAGGCAAGGGTATGGATTGACAGAAACCACTTCGGCTATTTTAATTACACCGGATACTGA
    F  K  L  N  F  V  R  Q  G  Y  G  L  T  E  T  T  S  A  I  L  I  T  P  D  T  D

CGTCAGACCGGGATCAACTGGTAAAATAGTACCATTTCACGCTGTTAAAGTTGTAGATTCAACAACAGGAAAACTTTTGG
  V  R  P  G  S  T  G  K  I  V  P  F  H  A  V  K  V  V  D  S  T  T  G  K  L  L

GGCCCAACAAAACCGGAGAATTGTATTTTAAGGGTGACATGATAATGAAAAGTTATTTTAATAATGATGAAGCTACAAAA
G  P  N  K  T  G  E  L  Y  F  K  G  D  M  I  M  K  S  Y  F  N  N  D  E  A  T  K

GCAATTATTGATAAACACGGATGGTTGCGTTCTGGTGATATTGCTTATTATGACAATGATGGTCATTTCTATATTGTGGA
 A  I  I  D  K  H  G  W  L  R  S  G  D  I  A  Y  Y  D  N  D  G  H  F  Y  I  V  D

CAGGCTTAAGTCATTAATAAAGTACAAAGGCTACCA
  R  L  K  S  L
```

iii. PrcLUC3 (corrected intron position)

```
CCGGGCTCCCAAAAGGTGTACCACTATCGCACAGAAACGTCATTCATAGATTTTCCCATTGCAGgtattttcttgtaca
                        P  L  S  H  R  N  V  I  H  R  F  S  H  C  R

cacgctggctgtagatgtatagtaattatatgttacagAGATCCCGTATTTGGTAATCAAATTATTCCGGATACTACTAT
                                      D  P  V  F  G  N  Q  I  I  P  D  T  T  I

ACTATGTGCTGTTCCATTCCATCATGCGTTTGGCACTTTCACAAATTTAGGATATCTAATATGTGGGTTCCACGTAGTGC
  L  C  A  V  P  F  H  H  A  F  G  T  F  T  N  L  G  Y  L  I  C  G  F  H  V  V

TTATGTACAGATTCCAAGAAGAATTATTCTTACAAACACTTCAAGATTTTAAATGTCAAAGCGCGTTACTAGTACCTACA
L  M  Y  R  F  Q  E  E  L  F  L  Q  T  L  Q  D  F  K  C  Q  S  A  L  L  V  P  T

GTACTTGCATTTCTTGCTAAAAATCCTTTAGTTGATAAATATGATTTATCACATTTACATGAAATTGCTTCTGGTGGTGC
 V  L  A  F  L  A  K  N  P  L  V  D  K  Y  D  L  S  H  L  H  E  I  A  S  G  G  A

TCCACTTTCAAAAGAAATTTCAGAAATAGCTGCAAAACGgtttgtatcaatttcaatgttattttaatttctaatgattt
  P  L  S  K  E  I  S  E  I  A  A  K  R

tgtatagATTTAAACTACCAGGAATACGACAAGGGTATGGTCTAACTGAAACAACATGTGCTATTGTTATTACTGCTGAA
        F  K  L  P  G  I  R  Q  G  Y  G  L  T  E  T  T  C  A  I  V  I  T  A  E

GGAGAATTTAAACCTGGGGCAGTTGGAAAAGTTGTACCATTTTATTCCCTAAAAGTACTTGATCTTAATACAGGAAAAAT
 G  E  F  K  P  G  A  V  G  K  V  V  P  F  Y  S  L  K  V  L  D  L  N  T  G  K  I

TTTGGGACCAAACGAGCGGGGGGAAATATGTTTCACAGGACCTATGATCATGAAAGGTTATATAAATAACCCAGAAGCAA
  L  G  P  N  E  R  G  E  I  C  F  T  G  P  M  I  M  K  G  Y  I  N  N  P  E  A

CACGAGAGATAATTGACGATGAGGGATGGATACATTCTGGAGATATAGGATATTACGATGAGGATGGTCATGTATACATT
T  R  E  I  I  D  D  E  G  W  I  H  S  G  D  I  G  Y  Y  D  E  D  G  H  V  Y  I

GTTGATCGTTTGAAATCTTTGATTAAATACAAAGGCTACCA
 V  D  R  L  K  S  L
```

I. *Photuris* species 'A'

i. PrspALUC1

```
CCGGGTTCCCTAAAGGCGTAGAACTTTCCCATAAGAATGTTGTCCATCGCTTTTCCCATTGCAGgtaatgttcattaaac
                     E  L  S  H  K  N  V  V  H  R  F  S  H  C  R

acaatacttacgtaatgattattcttttgcagAGATCCTGTGTTTGGAAATCAAATTATACCAGACACAGCTATTCTATC
                                  D  P  V  F  G  N  Q  I  I  P  D  T  A  I  L  S

TGTTATACCATTTCATCATGGATTTGGTATGTTTACCACCTTAGGGTATTTAGTGTGTGGATTTCGCATTGTCCTCATGT
 V  I  P  F  H  H  G  F  G  M  F  T  T  L  G  Y  L  V  C  G  F  R  I  V  L  M

ATCGATTTGAAGAAGAATTGTTTTTACAATCACTTCAAGATTATAGAATTCAAAGTGCATTACTAGTCCCTACATTATTT
Y  R  F  E  E  E  L  F  L  Q  S  L  Q  D  Y  R  I  Q  S  A  L  L  V  P  T  L  F

TCTTTCTTTGCTAAAAGTACTCTAGTTGACAAATATGATTTATCAAATTTACATGAAATTGCTTCTGGTGGTGCTCCACT
 S  F  F  A  K  S  T  L  V  D  K  Y  D  L  S  N  L  H  E  I  A  S  G  G  A  P  L

TGCAAAAGAGGTTGGAGAAGCTGTGGCCAAACGgtttgtataatattacagtgtaaatgaagtgtaattttaaataaaat
 A  K  E  V  G  E  A  V  A  K  R

catacagTTTTAAACTAAAAGGAATTCGACAAGGGTATGGTCTCACAGAGACAACTTCGGCTATTATAATTACTCCTGAG
         F  K  L  K  G  I  R  Q  G  Y  G  L  T  E  T  T  S  A  I  I  I  T  P  E

GGCGATGACAAAGCAGGGGCAGCGGGCAAAGTTGTGCCATTCTTCCGTGCAAAAATTGTTGATTTAGATACGGGTAAAAT
 G  D  D  K  A  G  A  A  G  K  V  V  P  F  F  R  A  K  I  V  D  L  D  T  G  K  I

TTTGGGTTGTAATCAACGAGGTGAACTGTACGTTAAAGGTCCTATGATAATGACGGGTTATATAAATAATCCGGAAGCCA
 L  G  C  N  Q  R  G  E  L  Y  V  K  G  P  M  I  M  T  G  Y  I  N  N  P  E  A

CAAATGCATTAATTGACAAAGATGGGTGGTTGCGTACTGGTGATATAGCGTATTACGATGAAGATGGCCACTTTTTCATT
T  N  A  L  I  D  K  D  G  W  L  R  T  G  D  I  A  Y  Y  D  E  D  G  H  F  F  I

GTTGACCGTCTAAAATCATTGATAAAATACAAGGGCTATCA
 V  D  R  L  K  S  L
```

ii. PrspALUC2

```
ACTGGTTTGCCGAAGGGGTTATGCTAACTCACAAGAATATTGTTGCACGATTTTCTCATTGCAAgtatgtaattaattaa
         A  E  G  V  M  L  T  H  K  N  I  V  A  R  F  S  H  C  K

tccacgtttacaactttaatcaattttaattttagAGATCCTACTTTTGGTAACGCAATTAATCCAACGACAGCAATT
                                       D  P  T  F  G  N  A  I  N  P  T  T  A  I

TTAACAGTAATACCTTTCCACCATGGTTTTGGTATGTCTACCACATTAGGATACTTGACTTGTGGATTTCGAATTGCTCT
 L  T  V  I  P  F  H  H  G  F  G  M  S  T  T  L  G  Y  L  T  C  G  F  R  I  A  L

AATGTACACGTTTGAGGAAAAGCTATTTTTACAATCATTACAAGATTATAAAGTGGAAAGTACTTTACTTGTGCCAACAC
 M  Y  T  F  E  E  K  L  F  L  Q  S  L  Q  D  Y  K  V  E  S  T  L  L  V  P  T

TAATGACATTTTTTGCAAAAAGTGCGTTAGTTGAAAAGTACGATTTATCGCACTTAAAAGAAATTGCATCTGGTGGCGCA
L  M  T  F  F  A  K  S  A  L  V  E  K  Y  D  L  S  H  L  K  E  I  A  S  G  G  A

CCTTTATCAAAAGAAATTGGGGAGATGGTGGCAAAACGgtaaacattgacccccctcctgcaattttatatttagcttatt
 P  L  S  K  E  I  G  E  M  V  A  K  R

ttagGTTTAAATTAAACTTTGTCAGGCAAGGGTATGGATTAACAGAAACTACTTCGGCTGTTTTAATTACCCCGGACACT
      F  K  L  N  F  V  R  Q  G  Y  G  L  T  E  T  T  S  A  V  L  I  T  P  D  T

GACGTCAGACCTGGATCAACTGGTAAAATAGTACCATTTCACGCTGTTAAAGTTGTGGATCCTACAACAGGAAAAATTTT
 D  V  R  P  G  S  T  G  K  I  V  P  F  H  A  V  K  V  V  D  P  T  T  G  K  I  L

GGGGCCCAATGAAACTGGAGAATTGTATTTTAAAGGCGACATGATAATGAAAAGTTATTGTAATAATGATGAAGCTACTA
 G  P  N  E  T  G  E  L  Y  F  K  G  D  M  I  M  K  S  Y  C  N  N  D  E  A  T

AAGCAATTATTAACAAAGACGGATGGTTGCGCTCTGGTGATATTGCTTATTATGACAATGATGGCCATTTTTATATTGTG
K  A  I  I  N  K  D  G  W  L  R  S  G  D  I  A  Y  Y  D  N  D  G  H  F  Y  I  V

GACAGGCTGAAGTCATTAATTAAATACAAAGGTTACCA
 D  R  L  K  S  L
```

B. Non-bioluminescent Beetles

a. *Cantharis rufa* Linnaeus, 1758

i. CfLL1

```
ACCGGTCTTCCTAAGGGCGTAGTTTTAAACCATTTTTCTGTTGGAAGATTATTCGGATTATTCGAgtactaaattataaa    80
                      V  L  N  H  F  S  V  G  R  L  F  G  L  F  D

aataaccgatcaaaccttttgcattagctttaattgattttttttagGTCGTTGGACTATTGCTGTACTTCCTTAGTAAC   160
                                                 S  L  D  Y  C  C  T  S  L  V  T

TGTTCAGGTAGTTCCAAATTTTCACATATATGGATTTGTGGTTCAATCGGTTTCGATGTGTTCTGTAAAAACAATTCTGA   240
  V  Q  V  V  P  N  F  H  I  Y  G  F  V  V  Q  S  V  S  M  C  S  V  K  T  I  L

TGAAAAAGTTTACTCCAAATTTGTTTTTAAGGATGATAGAAAAATATAAGGCAAATTTATTATACATCGTACCATCAATA   320
M  K  K  F  T  P  N  L  F  L  R  M  I  E  K  Y  K  A  N  L  L  Y  I  V  P  S  I

TTAAATTATTTATGTAAAAATCCGTTGGTAGAAAACTATGACACCTCGAGTCTTCGAGATGTCGTGGTAGGTGCGGCTCC   400
 L  N  Y  L  C  K  N  P  L  V  E  N  Y  D  T  S  S  L  R  D  V  V  V  G  A  A  P

AGTAGGAGCAACTTTGTTAAAAGAGGCGAAAAATAAgtaagttaaatttatgtaaaacagcaacaaaaaatagttaatcc   480
  V  G  A  T  L  L  K  E  A  K  N  K

attgcgtgtgtagATTTAAAAGTATATTCGTGAGAGAGATGTACGGATGTACAGAGGTTGGAGGAATCTCCTGTGCCCAG   560
               F  K  S  I  F  V  R  E  M  Y  G  C  T  E  V  G  G  I  S  C  A  Q

ACACAAAAACTGTATAAACCAGAAAGTACTGGATTGCTTTGTCCTGGATATATTGCAAAAATATGTGATATAAATTCGAA   640
 T  Q  K  L  Y  K  P  E  S  T  G  L  L  C  P  G  Y  I  A  K  I  C  D  I  N  S  N

CAAAGTACTAGGTCCTTTCGAGAAAGGAGAAATATGTATTAAGACTAAACAAATTATGAACGGCTATCTGAGAAACGATA   720
  K  V  L  G  P  F  E  K  G  E  I  C  I  K  T  K  Q  I  M  N  G  Y  L  R  N  D

CAGCTACTCGTGAGAGTTTTGATGACGAAGGCTTCTATTATACTGGAGATTATGGTTATTACGATAACGACAAATGTTTC   800
T  A  T  R  E  S  F  D  D  E  G  F  Y  Y  T  G  D  Y  G  Y  Y  D  N  D  K  C  F

TACATATGCGACAGAATTAAAGAATTAATAAAGTACAAGGGCTACCA
 Y  I  C  D  R  I  K  E  L
```

ii. CfLL3

```
ACCGGACTTCCAAAGGGTGTTCTATTAAATCAACAAGGCATTCGACATTGTATTGGTTTTTTCTCgtaagtactaataga
                      L  L  N  Q  Q  G  I  R  H  C  I  G  F  F  S

acttgacaataaattttttataatttacataaaacaaattattatagGGATTCTAAATTAGGTCGCTTAGGTTCTACAGAC
                                                D  S  K  L  G  R  L  G  S  T  D

GTAACTCTCCAAGTTATTCCATTTTTTCACATTCTCGGCTTCATGATTCAATTATTTTGTTTTGCTTGTTCTGCAAAAGT
 V  T  L  Q  V  I  P  F  F  H  I  L  G  F  M  I  Q  L  F  C  F  A  C  S  A  K  V

TGTAGTTTTGACCAAATTCAAACCAGATGTGTATCTGAAAAATATTGAAGATCATAAGGTTACAAAACTATATTCGGTGC
  V  V  L  T  K  F  K  P  D  V  Y  L  K  N  I  E  D  H  K  V  T  K  L  Y  S  V

CTTCTCTTGTAGTATTTCTGGCAAAATCTCCTTTGGTTAGCAAATATGATATATCATCTGTTAATCGTATTGTAGTAGGA
P  S  L  V  V  F  L  A  K  S  P  L  V  S  K  Y  D  I  S  S  V  N  R  I  V  V  G

GGAGCTCCACTTAGTGTTGGAGTTATAGAAGAAGTTGAAGAAAAgtaagtccacgtaaaaacaatttcttataataaaac
 G  A  P  L  S  V  G  V  I  E  E  V  E  E  K

gaatttatttagATTAAAATCAGTAAAAATTTGTCAAGTGTATGGGATGACAGAATTAGGTGGAATGATGGCGATGCAAA
              L  K  S  V  K  I  C  Q  V  Y  G  M  T  E  L  G  G  M  M  A  M  Q

CAATAGTTGGAAATAATAAAGTTGGAAGTGTTGGAACAGTTCCACCAGGTGTATCTTCAAAAGTATGCAATCCAGATAGC
T  I  V  G  N  N  k  v  g  s  v  g  t  v  p  p  g  v  s  s  k  v  C  N  P  D  S

GGTTTAGCATTAAAACCGCTGGAAATGGGAGAATTATGCTTTAAAAGTAGCGCCATGATGAAAGGGTACATCAGCAATAC
 G  L  A  L  K  P  L  E  M  G  E  L  C  F  K  S  S  A  M  M  K  G  Y  I  S  N  T

TGGTTTAATTCGGAACTGCTTTGATGATGATGGTTTCTACCACACGGGAGATTTAGGTTATTACGACCACGATAATAATT
  G  L  I  R  N  C  F  D  D  D  G  F  Y  H  T  G  D  L  G  Y  Y  D  H  D  N  N

TTTCATTGTTGATAGAATTAAAGAATTAATCAAGTATAAAGGCTACCA
F  F  I  V  D  R  I  K  E  L
```

iii. CfLL2

```
ACCGGGCTGCCGAAGGGTGTAGTGTTAACTCACTTTTCTGTTGGAAGATTATTAGCATTGATAAAgtaagcattaaatta
                      V  L  T  H  F  S  V  G  R  L  L  A  L  I  K

caaaaaataatagttaaacattatattcatattatttagGTCGCTCGACTATTGCAGTTCTACCTTAGTTACTGTTCAGA
                                         S  L  D  Y  C  S  S  T  L  V  T  V  Q

TAGTTCCAAATTTTCACCTATTTGGATTTGCAATGCAACTTATATCAATGTTTGTAGTAAAAACGATTCTTATGAAAAAG
I  V  P  N  F  H  L  F  G  F  A  M  Q  L  I  S  M  F  V  V  K  T  I  L  M  K  K

TTCGTTCCGGATTTGTTTTTAAGTATAATAGAAAAATACAAGGTAAATATATTATATACCGTGCCATCAATATTACAGTA
 F  V  P  D  L  F  L  S  I  I  E  K  Y  K  V  N  I  L  Y  T  V  P  S  I  L  Q  Y

TTTATGCAAACACCCACTGGTTGAAAAATACGACGTGTCGAGTCTTCGAGATGTTGTGGTAGGTGCAGCTCCTGTAGGAG
  L  C  K  H  P  L  V  E  K  Y  D  V  S  S  L  R  D  V  V  V  G  A  A  P  V  G

CAACTTTGTTAAACGATGCTCAAAATAAgtaagtaatttttttaaaaccagttgtaacaaaatgtattgtataataatac
A  T  L  L  N  D  A  Q  N  K

ttgtagATTTCAAGATTTGTTTGTAATAGAACTATATGGGGCTACAGAAATTGGAATTTCATGTTGTCAAACACGCAAGT
        F  Q  D  L  F  V  I  E  L  Y  G  A  T  E  I  G  I  S  C  C  Q  T  R  K

TGTATAAACCAACCAGTACCGGCATGCTTTGTCCTGGATTAATTGCAAAAGTTTGTGATATAAATTTGAACAAACCTCTA
L  Y  K  P  T  S  T  G  M  L  C  P  G  L  I  A  K  V  C  D  I  N  L  N  K  P  L

GGTCCTTTCGAGCAAGGTGAATTATGCTTCAAGAGTAAACAAATTATGAAAGGCTACCTGAGAAATGATGCAGCTACTCG
 G  P  F  E  Q  G  E  L  C  F  K  S  K  Q  I  M  K  G  Y  L  R  N  D  A  A  T  R

AGATAGTTTTGACGAAGATGGCTTTTACCATACGGGAGATTTTGGATATTACGATAACGATAAATGTTTTTACATAACTG
  D  S  F  D  E  D  G  F  Y  H  T  G  D  F  G  Y  Y  D  N  D  K  C  F  Y  I  T

ACAGAATTAAAGAATTGATAAAGTACAAGGGCTACCA
D  R  I  K  E  L
```

b. *Cantharis rustica* Fallén, 1807

```
ACCGGGCTTCCTAAAGGBGTACTCCTTACTCATTCAAATTTAAGAGAGAGCATCGTATATTTCAGgtaagtaaagaaaaa   80
                      L  L  T  H  S  N  L  R  E  S  I  V  Y  F  R

gtgttggataacaatttattatttcggctcaagtgttgtgacaattattttagTAGCGAAAATTACTTCAATGTTGCCTC  160
                                                       S  E  N  Y  F  N  V  A  S

TACGACGACCATACAGGTGGTTCCGAATTTCCATATATTTGGTATGACGCTGCAACTGGCTTCGGTAGTTAGCGCCCTAA  240
  T  T  T  I  Q  V  V  P  N  F  H  I  F  G  M  T  L  Q  L  A  S  V  V  S  A  L

AAATTATACTGATGAAAAAGTTTACTCCCGATTTATTTTTAAGTTCGATTGAGAAATACGGAGTGGCGAAATTATTTGCA  320
K  I  I  L  M  K  K  F  T  P  D  L  F  L  S  S  I  E  K  Y  G  V  A  K  L  F  A

GTGCCGTCGCTTTTATTATTTTTAGTAAAAAGCCCAATGGTGACGCAATATAATTTATCGAGCGTTACCGATATTTTTGT  400
 V  P  S  L  L  L  F  L  V  K  S  P  M  V  T  Q  Y  N  L  S  S  V  T  D  I  F  V

CGGCGCTGCTCCCGTCAGCGAGAAAATTCAAAACGAAGCCAAACAACGgtacgcacaatactatgccatcgataatccaa  480
  G  A  A  P  V  S  E  K  I  Q  N  E  A  K  Q  R

taattatttttgcgcgtcttcagATTTCCCAACATATCGGTAAACGAATTTTATGGGGCCACAGAAATTAGTGGAGCTT  560
                          F  P  N  I  S  V  N  E  F  Y  G  A  T  E  I  S  G  A

GTACCGTACAAACAAAAAAATACAACAAACCCGGAAGCTCGGGTATTTTAATTTCCAAGACGATAGCCAAAGTAGCTAAG  640
C  T  V  Q  T  K  K  Y  N  K  P  G  S  S  G  I  L  I  S  K  T  I  A  K  V  A  K

CTGAATAACATTAAAGAAAAATTAGGACCGTGCGAGTTAGGAGAATTGTGCTTCAAAGGTCCACAACTTATGAAAGAATA  720
 L  N  N  I  K  E  K  L  G  P  C  E  L  G  E  L  C  F  K  G  P  Q  L  M  K  E  Y

TGTGGGCAATCCTATAGCAACCAAGGAAAGTTTCGACGACGATGGATTCTACCGTACGGGAGATTTGGGTTATTACGATA  800
  V  G  N  P  I  A  T  K  E  S  F  D  D  D  G  F  Y  R  T  G  D  L  G  Y  Y  D

ACGATAAGTTTTTTTATGTGGTCGATAGAATTAAAGAATTAATYAAGTACAARGGSTACCA
N  D  K  F  F  Y  V  V  D  R  I  K  E  L
```

c. *Aphodius rufipes* Linnaeus, 1758

i. AfrLL1

```
ACCGGGTTCCCTAAAGGTGTAATGATAACCCATAAGAACATAAACGCCAAGAACTTAATGTTATTgtaagttatcaacac
                     M  I  T  H  K  N  I  N  A  K  N  L  M  L  F

tcttgaatatcacttatctagattattcaaactgaaatgataataattgttttgcagTGATCCAGACTATGCTAGCGATT
                                                            D  P  D  Y  A  S  D

TTAAGATCAACGTTTCGTTAGGAGTTTTGCCATTCTTCCATGCGTTCGGGCTGCATTCAGGTCTGACCGGTTTAATACTA
F  K  I  N  V  S  L  G  V  L  P  F  F  H  A  F  G  L  H  S  G  L  T  G  L  I  L

GGCAAAAAAATTATCGTGCTTCCAGTGTTTAATCCAATACAGTACTTGGAGTGCATCGAGAAATACAAAATTCCCATATT
 G  K  K  I  I  V  L  P  V  F  N  P  I  Q  Y  L  E  C  I  E  K  Y  K  I  P  I  L

GGGCATGGTTCCACCGTTGGTTAACTTCTTGGCCAAAAGTCCGTTGGTGGATGGCTTCGACTTATCACATGTCGAAGAGC
  G  M  V  P  P  L  V  N  F  L  A  K  S  P  L  V  D  G  F  D  L  S  H  V  E  E

TCATAGTAGGCGCTGGGCCAATCGGTAAGGATCTACAATATGAAATTAAAAAGAAgtacatatttttatatgtcaacca
L  I  V  G  A  G  P  I  G  K  D  L  Q  Y  E  I  K  K  K

taaatggggttcaccatggtttcttttagATTTGGTATCAAACATATAACGCAGGGATATGGATTGACTGAAGTGACGAT
                               F  G  I  K  H  I  T  Q  G  Y  G  L  T  E  V  T  I

TGGTCTGACACTAGCTCCTAAAAACAAAGAAAAGATTGGATCTTGTGGAACGCCTATACCAGGGGCGTATCTAGTGATAA
  G  L  T  L  A  P  K  N  K  E  K  I  G  S  C  G  T  P  I  P  G  A  Y  L  V  I

AAGATTTGGAAACTGGGCGAAATTTAGGACCAAATCAGACGGGGGAGATCTGTTGTAAATCGGATTGTGTCATGAAGGGA
K  D  L  E  T  G  R  N  L  G  P  N  Q  T  G  E  I  C  C  K  S  D  C  V  M  K  G

TATTATAAGAATGAAGAGGCAACCAGAGAGTCTTTCACTGCAGATGGTTGGCTGAGGACTGGAGATATTGGGTACTACGA
 Y  Y  K  N  E  E  A  T  R  E  S  F  T  A  D  G  W  L  R  T  G  D  I  G  Y  Y  D

TGAAGAAAACTATTTCTACATCGTTGATAGATTGAAGGAACTTATCAAGTACAAAGGCTATCA
  E  E  N  Y  F  Y  I  V  D  R  L  K  E  L
```

ii. AfrLL2

```
ACCGGGYTKCCGAARGGCGTCATGTTAACACATTTGAGTGCGAACTCAACTATAACAACGCTACGgtaaatactggtgta
                     M  L  T  H  L  S  A  N  S  T  I  T  T  L  R

ttagaaaccaacactaataataataactaatttatgcagGGATCCAACCGTCAAATCTCCGTTTGAAGCGAATACAACTT
                                         D  P  T  V  K  S  P  F  E  A  N  T  T

TAGGAGTGTTACCTTTCTTCCATGTGTACGGCCTGTACGTCGTATTGTTATCAGTACTGGAGGGCAGAAGAATTATTACC
L  G  V  L  P  F  F  H  V  Y  G  L  Y  V  V  L  L  S  V  L  E  G  R  R  I  I  T

ATGAACAAATTTGACTTGGAGGAATATCTAAGCACCATCCAACGGTACAAGATTGAAAAATTAGCGCTAGTGCCACCGAT
 M  N  K  F  D  L  E  E  Y  L  S  T  I  Q  R  Y  K  I  E  K  L  A  L  V  P  P  I

TGTCCAATACCTAATTAAAAACCCCGTCGTGAATCAGTACGATTTATCCAGCGTTAAAGAAATAGGGTGCGGCGGCGCCC
  V  Q  Y  L  I  K  N  P  V  V  N  Q  Y  D  L  S  S  V  K  E  I  G  C  G  G  A

CAATAAGCGAAGCATCCATACAAACAATCCGGAAGAAgttcagtattacgcttctcgtcacaatcttcaatttcaatttc
P  I  S  E  A  S  I  Q  T  I  R  K  K

ttattttccagGCTAAAATTGAAAGAGGTGCGACAAGGTTACGGCTTGACCGAGTCAGGKTACGGCGTAAGTCTCACTC
             L  K  L  K  E  V  R  Q  G  Y  G  L  T  E  S  G  Y  G  V  S  L  T

CAATCGGTCACACCAGACCGGGTAGTGTTGGGAAACTTTACCCCGGTTTGTCAGCGGTTGTTAGAGATGTCCGGACAGGG
P  I  G  H  T  R  P  G  S  V  G  K  L  Y  P  G  L  S  A  V  V  R  D  V  R  T  G

GAAAACTTGGGCGTTTACGTCGAAGGTGAAATTTGCTTTAAAGGCGATACACTAATGAAAGGTTACTACGCGAATCCTAA
 E  N  L  G  V  Y  V  E  G  E  I  C  F  K  G  D  T  L  M  K  G  Y  Y  A  N  P  K

AGCGACACAGACGTCTTTTACATCGGATGGTTGGCTCAGAACTGGAGATATCGGTTATTACGATAATGATGGTTACCTTT
  A  T  Q  T  S  F  T  S  D  G  W  L  R  T  G  D  I  G  Y  Y  D  N  D  G  Y  L

ATATTGTCGATAGACTTAAAGACGTCATAAAGTACAAAGGCTACCA
Y  I  V  D  R  L  K  D  V
```

d. *Grammoptera ruficornis* Fabricius, 1781

i. GrfLL1

```
ACCGGATTACCTAAGGGGGTGTTGATCTCGCACCTTAATATTATGGTCAGGATCTTACAGTCGAGgtaaggtaggtcagt
                      L  I  S  H  L  N  I  M  V  R  I  L  Q  S  R

aaaaattaacaaaaagttgttttaagtagttgtagcatctaatttatctatagaattatttgttggtccatttaaacgtt

gttttttttttattgataaagtttcaataactctactatggtgacaaattacttcgaactattttaatgacaaaaattta

gcaatttttttaattttaaatctgactacgccactggattgatgatacctttaaatgggtttcataaaaaacttataaaa

aatattagttgacgataacagaaatatggcataattagagttctaattattttatcaaaatatttcccatagtcgtgcaa

atcacaattttgtaaataatattgtttaaaattatcctatagaccgctataagttaaactttaagaatttctaatcaata

ctatcaacagttattcgtaatattggcaatatttgttttatttacaaaattgtgattacatgacgactattccaaatatt

ttgataaaataacttcacctcgatcgatttcggaaagtgataattttacttcagGGACCCTCAGTACATTAACGAAGACC
                                                        D  P  Q  Y  I  N  E  D

ATGCCATCCTCGGACTGATGCCCTTTTTCCACGCATATGGTTTAAGTATTGGCATTTCCACAATTATCAACAAACAAAAA
H  A  I  L  G  L  M  P  F  F  H  A  Y  G  L  S  I  G  I  S  T  I  I  N  K  Q  K

ATCTTAATATTCAAGAAATTTGAGGAAGACAGTTTCTTGAAAGCGATACAGGACTACAAGATCAAAATATTAGCTATAGT
 I  L  I  F  K  K  F  E  E  D  S  F  L  K  A  I  Q  D  Y  K  I  K  I  L  A  I  V

TCCACCTCTAGCCGTTTTCCTGGAGAAGACCCCGAAACTATCTAATTATGATTTGTCCTGCGTTGAAAGGATTTATTGCG
  P  P  L  A  V  F  L  E  K  T  P  K  L  S  N  Y  D  L  S  C  V  E  R  I  Y  C

GAGCTGCACCGCTCAGTAAGAATACCGAGCTTGCCATCACGAAAAGgtaattattaaagtatattggtggactgaaaaa
G  A  A  P  L  S  K  N  T  E  L  A  I  T  K  R

cgaagaaagcaaaatctaaaaaatgtataacgccgcttttttcggtctactgacaaatgttggattgtaacatgaaaaaca

ataattttaagGTTAAAAAACCTGAGAGGGATACATCAAGCCTACGGCCTAACAGAAGCAACCTTAGCCGTAACAGTGCC
             L  K  N  L  R  G  I  H  Q  A  Y  G  L  T  E  A  T  L  A  V  T  V  P

AGATAAAAACGTGCTCAAATTTGGATCTTGTGGGAAAGTCGCCACGTACATGTCCTGCAAAGTCAGGGACCCAGAGACGA
  D  K  N  V  L  K  F  G  S  C  G  K  V  A  T  Y  M  S  C  K  V  R  D  P  E  T

GGAGATCCTTGAACGCTAATCAGGTTGGGGAGCTGTGCGTCAAGGGACCAATGGTGATGATGAGATATTATAACGATCAG
R  R  S  L  N  A  N  Q  V  G  E  L  C  V  K  G  P  M  V  M  M  R  Y  Y  N  D  Q

AAAGCTACTAGGGAATCTTTTACTCCCGATGGCTGGCTCAAGACGGGCGATCTAGGTTATTATGACGACGAGGGGTTCTT
 K  A  T  R  E  S  F  T  P  D  G  W  L  K  T  G  D  L  G  Y  Y  D  D  E  G  F  F

TTACATTGTGGACAGACTGAAGGAGCTCATAAAGTATAAGGCTACC
  Y  I  V  D  R  L  K  E  L
```

e. ***Pachnoda marginata*** **Drury 1773 ssp.** ***peregrina*** **Kolbe, 1906**

i. PmpLL1

```
ACCGGTCTTCCAAAGGGGGTAATTTTAACACATAAAAATATAAATTACAAAAACGCATCATTTAGgtaagtttaaacgaa 80
                        I  L  T  H  K  N  I  N  Y  K  N  A  S  F  R

attaaatggtttttgtgattctgaatttaatttgatcgtcaataagataaatttatcaactcttaaatttaatatctta 160

ttatccgtaagtaatatatgaaaaagaatcgcaaatagttgatgatagtaaaactaggttaaaaattgttgctaattttt 240

ttatttatttagAGCACCTTGGTTAAATGCGGAGAAACAATCACAGTACGTGTTAGGATTTCTGCCATATTTTCATGTGT 320
              A  P  W  L  N  A  E  K  Q  S  Q  Y  V  L  G  F  L  P  Y  F  H  V

ATGGGTTGCATGTTGTAATAGATGGGATTATCATGGGACGCACCGTCATTGTTGTAAATAAATTTGATTTTGAATTACAT 400
Y  G  L  H  V  V  I  D  G  I  I  M  G  R  T  V  I  V  V  N  K  F  D  F  E  L  H

TTAAAGTCCATTTCGAAGTATAAGATAACTCAGTTCGCCGTGGTACCACCGGTGTTGCAAATGTATGCCAAAAGCCCACT 480
 L  K  S  I  S  K  Y  K  I  T  Q  F  A  V  V  P  P  V  L  Q  M  Y  A  K  S  P  L

GACTGAYAAATATGATTTATCACATATAGAGGGTATTTTAGTTGGTGCAGCGCCGGTTAGCGAATCGTTGCGTAAAGCTA 560
  T  D  K  Y  D  L  S  H  I  E  G  I  L  V  G  A  A  P  V  S  E  S  L  R  K  A

TTTTACAACGgtaacctattcgccaattttgttagcgcayatttgtattacagatttttttctagTACTGGAATAAAAA 640
I  L  Q  R                                                         T  G  I  K

GTATATTTCAAGGATATGGACTGACTGAGATCACAGTAGCCGCAACAGTGACAGACGTAGGGCTGGATAAGCCTGAAACG 720
S  I  F  Q  G  Y  G  L  T  E  I  T  V  A  A  T  V  T  D  V  G  L  D  K  P  E  T

TGCGGCAAATTACTTCCGTATTTAACAGGGGTGGTTAGAGATTTGAAAACTGGTCGAAATCTAGGACCAAATCAGAGTGG 800
 C  G  K  L  L  P  Y  L  T  G  V  V  R  D  L  K  T  G  R  N  L  G  P  N  Q  S  G

CGAGCTATGTTTCAAGGGTGGCATGGTAATGAAAGGTTACTATAAAAACGAAGCTGCCACTAAAGATGTTTTCACTGAAG 880
  E  L  C  F  K  G  G  M  V  M  K  G  Y  Y  K  N  E  A  A  T  K  D  V  F  T  E

ACGGCTACTTAAGAACTGGTGACATAGGCTATTATGACAAAGAAGGGTACTTTTATATAGAAGATAGACTGAAAGATCTT 960
D  G  Y  L  R  T  G  D  I  G  Y  Y  D  K  E  G  Y  F  Y  I  E  D  R  L  K  D  L

ATCAAATACAAAGCT
```

ii PmpLL2

```
ACTGGTTTGCCAAAGGGAGTTATGTTAACACATAAGAACATTAACTATAAAAATGCTACATTTCGg[illegible]aagtttctaaaat 80
                        M  L  T  H  K  N  I  N  Y  K  N  A  T  F  R

gattgtatgagttttcgtctagaaatgctttttttgcacttgtttaaatcgctggaaacacactatttttacactataa 160

gcctaactaactaattaactacgctacaaaactatctataagccaacaaagaggtcactaaagcttgatatatctacatg 240

ctggttacagatttatattatatcgcagatttgattgttcaaatttaatagctgggtttaaatatgaaacaaaggtttgt 320

tgtctcaaaagtaaagtgtgtaagcaccaaaacattaacacttttatatttatatttactacgtcaccaaatgaaattat 400

tttttttagAGAACCTTGGTTAAATGCAGAGAACCAATCACAACATGTACTGGGCTTTGCTCCCTATTTCCACGTTTTCG 480
           E  P  W  L  N  A  E  N  Q  S  Q  H  V  L  G  F  A  P  Y  F  H  V  F

GATTGCACGTTATAATGAATTCCATTATCTTGGGGTGCACTGTCGTTGTTCTGGAGAAATTTGATTTCGAATTACATTTA 560
G  L  H  V  I  M  N  S  I  I  L  G  C  T  V  V  V  L  E  K  F  D  F  E  L  H  L

AATTGCATCCAAAAATATAAAATTACAACGCTTGCACTTCCACCGCCGGTATTGCAAATGTACGCTAAAAGTCCGTTAAT 640
 N  C  I  Q  K  Y  K  I  T  T  L  A  L  P  P  P  V  L  Q  M  Y  A  K  S  P  L  M

GGAGAAATACGACTTGTCGCATGTGGAATACGTTCTGGTCGGCGGAGCTCCTCTGGACGAATCACTGCGTAAAGCTATTT 720
  E  K  Y  D  L  S  H  V  E  Y  V  L  V  G  G  A  P  L  D  E  S  L  R  K  A  I

TAGAACGGtaagtctggtgtattggcatgcttgaaattcgagacttataagggattttaagtttagTGTTGGAATTAAAA 800
L  E  R                                                               V  G  I  K

GTATATGTCAAGGGTACGGCATGACGGAATTGACTCTTCCAGCAACCTTGGTGCCAGTAGGGCTAAGTCGCCCTGGAACT 880
S  I  C  Q  G  Y  G  M  T  E  L  T  L  P  A  T  L  V  P  V  G  L  S  R  P  G  T

TGTGGCAAGCTAATCCCGTATTTGACAGCGATTGTGAAGGATTTGAAAACCGGGCGAAATCTAGGACCAAACGAGAACGG 960
 C  G  K  L  I  P  Y  L  T  A  I  V  K  D  L  K  T  G  R  N  L  G  P  N  E  N  G

TGAATTGTGCTTTAAAGGCGATCTGGTGATGAAAGGTTACTACAAAAACGAAGCGGCCACTAAAGAAACTTTCACGGAAG 1040
  E  L  C  F  K  G  D  L  V  M  K  G  Y  Y  K  N  E  A  A  T  K  E  T  F  T  E

ACGGCTACTTGAAAACTGGCGACATTGGCCACTACGACGAAGAGAAATTTTTTTACGTTGAAGATAGATTAAAGGAACTA 1120
D  G  Y  L  K  T  G  D  I  G  H  Y  D  E  E  K  F  F  Y  V  E  D  R  L  K  E  L

ATCAAATACAAAGGCTACCA
```

f. *Stegobium paniceum* Linnaeus, 1758

```
ACTGGGTTGCCGAAGGGTGTGATAATAACTCAGGAACACTTCTTGGCGCAATTAGCCACTGCTGGgtaaaaataatattt
                       I  I  T  Q  E  H  F  L  A  Q  L  A  T  A  G

tagaaaaaaacatttaaaaaatattaatttgactttagTGACTCTCGGTATGCAAACATGCACGAAGATATGGTAGTTTT
                                          D  S  R  Y  A  N  M  H  E  D  M  V  V  L

GGGTCAATTACCATTCTTTCACATATTTGGACTCTTTCTTCTTTTGGGAAGTACGTTGTTTGGGATGAAGTTGGTAGTAC
  G  Q  L  P  F  F  H  I  F  G  L  F  L  L  L  G  S  T  L  F  G  M  K  L  V  V

TAAAAGCATTCAAGCCGAATACCTACTTAAATGCCTTAGAAAAATACAAGGTGCAGCAAATCTATCTCGTCCCTGCTCTT
L  K  A  F  K  P  N  T  Y  L  N  A  L  E  K  Y  K  V  Q  Q  I  Y  L  V  P  A  L

TTGCTTTTCCTGGTCAAAAGTGATCTGGTAGAAAATTATGATTTGTCTTTTGTTGAGGATATCCTTTGTGGCGGTGCACC
 L  L  F  L  V  K  S  D  L  V  E  N  Y  D  L  S  F  V  E  D  I  L  C  G  G  A  P

TTTAAGTGAGGAATTGCAACGCACCGCGCAGATGAAgtacatttgtgatgttatcttttaaaaaaaattttaataatgga
  L  S  E  E  L  Q  R  T  A  Q  M  R

ttttagATTAAACTGCGAAGTGAGACAAGTCTATGGACTTACAGAAGCTGGTGGATGTATATCATTTTTACCGAAAGGGT
       L  N  C  E  V  R  Q  V  Y  G  L  T  E  A  G  G  C  I  S  F  L  P  K  G

TTCAAAAATTTGCCTACTCTGGGAAATTAATACCTTTCGGCGAAGCAAAAATTTCTCATATAGACTCTGGAAAGAATTTG
F  Q  K  F  A  Y  S  G  K  L  I  P  F  G  E  A  K  I  S  H  I  D  S  G  K  N  L

GGTCCTAATGAATTTGGAGAAATTTGCGTAAGAATGCCGTCGTGCATGGAATATTACATCGACAACCAAAAAGCTACAAA
 G  P  N  E  F  G  E  I  C  V  R  M  P  S  C  M  E  Y  Y  I  D  N  Q  K  A  T  N

TGAGATATTTGATGCCGACAAATTCTTACACACAGGCGATGTGGGTTATTTTGATGAAGACGGAATTTTGCACGTTATTG
  E  I  F  D  A  D  K  F  L  H  T  G  D  V  G  Y  F  D  E  D  G  I  L  H  V  I

ATCGAATAAAAGAACTTATCAAATACAAGGGATACCAGGT
D  R  I  K  E  L
```

g. *Necrophorus vespillo* Linnaeus, 1768

i. NevLL1

```
AMCGGGTTTCCTAAAGGTGTCCAAATAACTCACGACAATTTGAAGTTTATAATAGGATACGTTGGgtaattatttaatag
                       Q  I  T  H  D  N  L  K  F  I  I  G  Y  V  G

tactgagattcattgagtaattgttataattttagGTCTTCTAAATTTATTGAACTACATTCCAATGATGTTGGCATTG
                                       S  S  K  F  I  E  L  H  S  N  D  V  G  I

CTATATGTCCTTTCTACCACTTATATGGATTTATAGTGTTTGTGTCCACTCTATTAACAGGATCGTTGTCAGTTGTAATG
A  I  C  P  F  Y  H  L  Y  G  F  I  V  F  V  S  T  L  L  T  G  S  L  S  V  V  M

TCAAAGTTTAAAAGGGAGCGTTACTTGGAGCTGATCGAAACATACCACGTGACCAAATTGTGGCTGGTCCCACCAATTGC
 S  K  F  K  R  E  R  Y  L  E  L  I  E  T  Y  H  V  T  K  L  W  L  V  P  P  I

GATATTTTTAGCCAAAAGTCCGATGGTCGATAATTACAAATTGGATAGTCTAAAGTCAATTATCTGCGGCGCCGCGGCTT
  I  F  L  A  K  S  P  M  V  D  N  Y  K  L  D  S  L  K  S  I  I  C  G  A  A  A

TGGGAATTGAAATTAAAAATATGGTGAGCAAACGTTTAGACGTCACCGTGCAACAAGTTTTCGGTATGACAGAACTGTCA
   G  I  E  I  K  N  M  V  S  K  R  L  D  V  T  V  Q  Q  V  F  G  M  T  E  L  S

GGTGTCGTTGTCGTGATGCCTACTGAAGCAACTGGAGAACTAGGTGGATGTATTGGTAAATTGTGTCCAGGTGTTGTTGG
 G  V  V  V  V  M  P  T  E  A  T  G  E  L  G  G  C  I  G  K  L  C  P  G  V  V  G

TATGATCAAAGATGTGGAAACTGATGAGATCCTTGGACCGTACCAGAATGGCGAAATCTGCTTCAAGGGAAATTTTGTTA
  M  I  K  D  V  E  T  D  E  I  L  G  P  Y  Q  N  G  E  I  C  F  K  G  N  F  V

TGAAAGGCTATTTGAATAATCATGCGGCAACGGCTTTAATTTTAGATGAGGATAAAGTCCTTCGAACTGGTGATATAGGT
   K  G  Y  L  N  N  H  A  A  T  A  L  I  L  D  E  D  K  V  L  R  T  G  D  I  G

TATTACGATGAAAATGGCTACTTCTTCGTCATGGACAGATTGAAGGAACTGATTAAATACAAAGGCTACCA
 Y  Y  D  E  N  G  Y  F  F  V  M  D  R  L  K  E  L
```

ii.NevLL2

```
ACCGGTTTACCTAAAGGTGTCGAGATGACTCATGAAAATTTGAGTAATTTAATTAACTTTTATGGgtatgtaattaagat
                         E  M  T  H  E  N  L  S  N  L  I  N  F  Y  G

ggaattatttgctccgattaattattgtaattagGTCTCCGAGGTTTATATCTTGCGGAGTGGATGACGTCAGCATTGCC
                                   S  P  R  F  I  S  C  G  V  D  D  V  S  I  A

GTTTGCCCTTTTCACCATCTGTATGGTATCATTATCTTTACGAATACTTTATTAACTGGTATTGTCAACGTTTTGATGAC
 V  C  P  F  H  H  L  Y  G  I  I  I  F  T  N  T  L  L  T  G  I  V  N  V  L  M  T

AAAATTCGAGAAAGAAACTTACTTGAAGCTTATTGAATCTCACAAAGCTACTGTACTATTTATAGTGCCACCTATTGCTA
  K  F  E  K  E  T  Y  L  K  L  I  E  S  H  K  A  T  V  L  F  I  V  P  P  I  A

CGCTCTTGGCAAAAAGTTCGATCGTCGATGATTATAATCTAAGCAGTCTTAAAACAGTTTTCTCCTGCGCTGCGCCTTTA
T  L  L  A  K  S  S  I  V  D  D  Y  N  L  S  S  L  K  T  V  F  S  C  A  A  P  L

GGAGGCGATATTCAAGATGTTTTAAATAAGAAATTGAATTTGCCTATTCAACAGCTTTACGGTATGACTGAGATGTCTGG
 G  G  D  I  Q  D  V  L  N  K  K  L  N  L  P  I  Q  Q  L  Y  G  M  T  E  M  S  G

AGCAATTACGGCATTTCCAAAAGATGCTTACGTGAAAAGAAAGGCAGGATGCGTTGGGATCTTACTTCCAGGTGTATTGG
  A  I  T  A  F  P  K  D  A  Y  V  K  R  K  A  G  C  V  G  I  L  L  P  G  V  L

GGATGGTCAAGGATCTGGGCGGCAATAAGGCTTTGAGTCCCAATCAACCTGGAGAGTTATGTTTCAAAGGAAAATTCATC
G  M  V  K  D  L  G  G  N  K  A  L  S  P  N  Q  P  G  E  L  C  F  K  G  K  F  I

ATGAAAGGGTACCTCAATAATACTGCTGCAAGTGAAATGCTATTAAATGAGGATGGCTTTCTCATAACAGGGGATTTGGG
 M  K  G  Y  L  N  N  T  A  A  S  E  M  L  L  N  E  D  G  F  L  I  T  G  D  L  G

GTATTATGATGAAGATGGTTACTTTTACGTCATCGATCGGCTGAAAGAGTTGATCAAATACAAAGGCTATCA
  Y  Y  D  E  D  G  Y  F  Y  V  I  D  R  L  K  E  L
```

iii.NevLL3

```
ACCGGCTTCCCTAAGGGTGTAGTTTTAACACAAAAGAATATCAAATACTGCTTCAATTATTTAAAgtattgtattttatt
                         V  L  T  Q  K  N  I  K  Y  C  F  N  Y  L  K

tttgtagcaacaacatttatgtataatttaatatttatctgatagATATCCGTATATAAATTTATGTCCAAATAATGTCC
                                              Y  P  Y  I  N  L  C  P  N  N  V

AGATTGTAATGATGCCCACCTTTCACGTATTCGGATTCTTACTGGGCATCGGAAACATCCACAATCTCTCGTTAATAGTA
Q  I  V  M  M  P  T  F  H  V  F  G  F  L  L  G  I  G  N  I  H  N  L  S  L  I  V

ATTCTGCAGAAATTCCAACCTAATCACTTTTGTGAAATAATTGAAGAGTACAATGTAACGGCTCTTCACATCGTTCCAAC
 I  L  Q  K  F  Q  P  N  H  F  C  E  I  I  E  E  Y  N  V  T  A  L  H  I  V  P  T

TATTGCGGTTTTTTTAGCTAAACATCCTACCGTGAAAAACTACGACTTTTCATCTGTCAAGGATATCATGTGTGGTGCTG
  I  A  V  F  L  A  K  H  P  T  V  K  N  Y  D  F  S  S  V  K  D  I  M  C  G  A

CTCCTTTGGGTACCGAAGTCCAAAGCATCTTAGAACAAAGgttagtgattaaagtactttcaatgtaaaaaaaaacaaat
A  P  L  G  T  E  V  Q  S  I  L  E  Q  R

agcatgtataatttgttaataaataaaagtgagaaaacattttatgatttagATTCAATTGCAAGATACGACAAATATA
                                                     F  N  C  K  I  R  Q  I  Y

TGGAATGACGGAAACTTGCGGTCTGATTACACTGATGCCAATGGGCGAAGAATATAAGATTGGATCTGCAGGAAAACCTT
  G  M  T  E  T  C  G  L  I  T  L  M  P  M  G  E  E  Y  K  I  G  S  A  G  K  P

TACCGTGGGTTGAAATCAAAATAATCGATATTGTTACCAACAAAGAGATGGCTCAGGGTGGCATTGGCGAAATATGCGTG
L  P  W  V  E  I  K  I  I  D  I  V  T  N  K  E  M  A  Q  G  G  I  G  E  I  C  V

AAGACCGAGCAAAATATGAAAGGTTATTTGGGTATGGTGGAGGAAACCCAAAATATGTATTTGGAAGATGGTTATATGCG
 K  T  E  Q  N  M  K  G  Y  L  G  M  V  E  E  T  Q  N  M  Y  L  E  D  G  Y  M  R

TACCGGAGATATTGGTTATTTGGACGAGGATGGTGAACTCTTCATCGTAGATAGACTAAAAGAATTGATCAAGTACAAGG
  T  G  D  I  G  Y  L  D  E  D  G  E  L  F  I  V  D  R  L  K  E  L

GCTACCA
```

iv.NevLL4

```
CCGGGTTACCTAAGGGTGTTATGACGACGCACGCTAATATGACCGCTTTCATTGAAGTTGCAAAgtctgtagaagtttaa
                     M  T  T  H  A  N  M  T  A  F  I  E  V  A  K

aatctaaatgcaattggattctatttcatcgaatgatttagGGTGCGCGTTCAAGAATTTTTAGAATCGGATTCTAATA
                                           V  R  V  Q  E  F  L  E  S  D  S  N

TTTCGGGGACTGTAACGTTTGGGTTGACGCCGTTCTTCCATTCCATGGGCTTCATGTCGATGTACATGAACTTCATAGGA
I  S  G  T  V  T  F  G  L  T  P  F  F  H  S  M  G  F  M  S  M  Y  M  N  F  I  G

GGCAATCTTGTGGTTGTGATGAAGAAGTTCAAGACTAAGCTCTTCCTTGAAGCCATTGCAAAGTACAAAGTCACCACCCT
 G  N  L  V  V  V  M  K  K  F  K  T  K  L  F  L  E  A  I  A  K  Y  K  V  T  T  L

TGTTGTGCCTCCACCAATCATCCTTATTCTCAATAAGCATCCGTTAGTTAAGAATTACGACTTGTCCAGTTTGAGGGATT
  V  V  P  P  P  I  I  L  I  L  N  K  H  P  L  V  K  N  Y  D  L  S  S  L  R  D

TGAGGTCTGGGGCTGCGCCGATGGGTAAGGAGATGGAGAGGGAAGTTAAGGATCGTCTGAAGTTGCACCACGTCTCTCAG
L  R  S  G  A  A  P  M  G  K  E  M  E  R  E  V  K  D  R  L  K  L  H  H  V  S  Q

AACTATGGAATGACCGAAACGACGCTCGGTGTTCTTATGACCAAATACAACGAGTGCAGGTTCGGCTCTGTTGGTCAGAT
 N  Y  G  M  T  E  T  T  L  G  V  L  M  T  K  Y  N  E  C  R  F  G  S  V  G  Q  I

CGTGCCTTCCATGAAAGTTAAAgtaagtggttatcatAgtcatcttgaatacattgaagtttattgggaatagATTGTAG
  V  P  S  M  K  V  K                                                     I  V

ATGAAGACGGTAGAGCCTTGGGACCAAACCAGGAGGGCGAGCTTTGCTTTAAGGGACCAATGATCATGAAGGGATACGTT
D  E  D  G  R  A  L  G  P  N  Q  E  G  E  L  C  F  K  G  P  M  I  M  K  G  Y  V

GGAAACCCAGTGTCCACAGCAAACACTATTGACAAAGACGGTTGGTTGCACACTGGAGACATCGGTTATTACGATAACGA
 G  N  P  V  S  T  A  N  T  I  D  K  D  G  W  L  H  T  G  D  I  G  Y  Y  D  N  D

TAAGTACTTCTTCATTGTGGACCGCATTAAGGAATTGATCAAATACAAAGGCTATCA
  K  Y  F  F  I  V  D  R  I  K  E  L
```

h. *Anacaena* sp. Thomson, 1859

i. AnspLL1

```
ACTGGGTTTCCTAAAGGCGTTATGCTCACGCATGAAAATTTGAGGGCTACTATGAATTACCTAAAgtaagtaaaaaaatc
                      M  L  T  H  E  N  L  R  A  T  M  N  Y  L  K

atatgaattaaattcatattaaaaattaatttcattttagAAGCCCAGATTTTATGAACGTTAACCCTGATGACATAACC
                                          S  P  D  F  M  N  V  N  P  D  D  I  T

ATCAGCGCTCTGCCCTTTTTCCACGTATATGGAGTAATGCTGATTTTTAATGTCTTAGCCAACGGAATGAAGTTGATTAA
 I  S  A  L  P  F  F  H  V  Y  G  V  M  L  I  F  N  V  L  A  N  G  M  K  L  I  N

CATGAAGCAGTTTCAACCAGAAACGTTCCTCAAAACTATTCAAGAAAAGAAAATTAACAAACTATTCCTTGTTCCATCCC
  M  K  Q  F  Q  P  E  T  F  L  K  T  I  Q  E  K  K  I  N  K  L  F  L  V  P  S

TAGCAGTATTTTTGGCCAAACATCCCTTGGTGGAGAAGTTTGATCTTTCTAGTGTTAAGTATGCGTATTGTGGAGGAGCA
L  A  V  F  L  A  K  H  P  L  V  E  K  F  D  L  S  S  V  K  Y  A  Y  C  G  G  A

CCTCTAGCTGACGAAGTAGAAAAGACATTATGGAATAAgtagagtattttattttattttgctattgagaattgcgta 480
 P  L  A  D  E  V  E  K  T  L  W  N  K

gaaggattaattacaaaaaaaaatattttcagATTTAGCCTGATCTCATTATTAAATGGTTATGGTTTGACTGAATGCGC 560
                                  F  S  L  I  S  L  L  N  G  Y  G  L  T  E  C  A

TGGACTAACTCATCTGGCACCAAGGAACGCTGAACCTAGGTTTGGTTCTTCTGGGATTCCCGTACCCCTCAGCGTATGCA
  G  L  T  H  L  A  P  R  N  A  E  P  R  F  G  S  S  G  I  P  V  P  L  S  V  C

AAGTGGTAGAACATTCGTCAGGAAAACCTCTAGGACCGAACAAAGCTGGAGAGCTTTGCTTCAAAGGTTGCCTTGTAATG
K  V  V  E  H  S  S  G  K  P  L  G  P  N  K  A  G  E  L  C  F  K  G  C  L  V  M

AAGGGTTACATTGACGATCCCGACTCAACAATACAAGCATTTGATGAGGAAGGCTTCCTCCACTCAGGGGGTTATGGTTA
 K  G  Y  I  D  D  P  D  S  T  I  Q  A  F  D  E  E  G  F  L  H  S  G  G  Y  G  Y

TTACGATGAGGACAATTATCTTTTCATAATAGACAGACTGAAGGACATCATCAAGTACAAAGGCTATCACCTA
  Y  D  E  D  N  Y  L  F  I  I  D  R  L  K  D  I
```

ii. AnspLL2

```
ACTGGTTTTCCCAAGGGTGTTATGCTCTCTCACGAAAATGTTCGATGCTGCTTGGAATATATGAGgtatttttatcatt
                      M  L  S  H  E  N  V  R  C  C  L  E  Y  M  R

atatttatataatgtaaactaattcagcaattaatcaaatttcttttcatttaagCGATCCAAATTATGCCAACATTTCA
                                                          D  P  N  Y  A  N  I  S

TCAACAGATATAACTGTTAGTGTCCTCCCTTACTTCCATATATTTGGACTTTACCTACTCGTAAATTCCATAGCACATGA
 S  T  D  I  T  V  S  V  L  P  Y  F  H  I  F  G  L  Y  L  L  V  N  S  I  A  H  D

TATGAAATTAATAGATCTGACAAAATTCAAGCCTGATGTTTATCTTAAATGCATTCAAGATTATAAAGCCACGAAACTTT
  M  K  L  I  D  L  T  K  F  K  P  D  V  Y  L  K  C  I  Q  D  Y  K  A  T  K  L

TCGTAGTTCCATCTTTAGCAGTATTCCTAGCGAAAAGTCCCTTATTAGACAATTACGATATATCAAGTATTAACGATATA
F  V  V  P  S  L  A  V  F  L  A  K  S  P  L  L  D  N  Y  D  I  S  S  I  N  D  I

CTTTGCGGAGCTGCACCATTAGGAAAAGATATTGAGGTTATTTTAAGAAACAAgtaagtaaatttgtaatacaattcaat
 L  C  G  A  A  P  L  G  K  D  I  E  V  I  L  R  N  K

ttttttacaatctatgaatgtttagATTAAAACTTGCTGCTATACGACAAGTATATGGCCTCACTGAATGTGCTGGGGC
                           L  K  L  A  A  I  R  Q  V  Y  G  L  T  E  C  A  G  A

TGCTACCATGTTCCCAACAGGAGTTGCGTCTAAACCTGGTAGTTCTGGAATATTGTTGCCCATGGCTAAATGTAAAGTTG
  A  T  M  F  P  T  G  V  A  S  K  P  G  S  S  G  I  L  L  P  M  A  K  C  K  V

TAGATCCAGAAACCAGAGAAACTCTTGGTCCCAACAAACCTGGAGAACTGTGCTTTAAAGGTCCTTTGGCCATGAAAGGA
V  D  P  E  T  R  E  T  L  G  P  N  K  P  G  E  L  C  F  K  G  P  L  A  M  K  G

TATATAGGAGACCCGGTGGCTACTACGAATACATTCGATGAAGATGGTTTTCTACATACAGGTGACATTGGATATTATGA
 Y  I  G  D  P  V  A  T  T  N  T  F  D  E  D  G  F  L  H  T  G  D  I  G  Y  Y  D

TGACAGTCAACACTTTTACATCGTAGATCGGTTGAAAGAACTGATTAAGTATAARGCTACC
  D  S  Q  H  F  Y  I  V  D  R  L  K  E  L
```

Appendix III

PUBLICATIONS ARISING FROM RESEARCH