TLT

Supplementary material for: Comparative genomics of firefly luciferase reveals complex evolutionary history

Chiara A. Thomas✉**, Adrian J. Hölzlwimmer, Noah Nussbaumer, Frank Andreas Weitz, and Rebecca S. Ellenrieder**

TUM School of Computation, Information and Technology, Technical University of Munich

✉ ge58rab@mytum.de

April 26, 2024

Abstract — Luciferase (LUC) enzymes are a broad family of proteins that catalyse light emitting reactions. Here we investigate the evolutionary history of LUC in fireflies (*Lampyridae*). We provide evidence for LUC to be a moonlighting 4-coumarate–CoA ligase (4CL), a protein involved in long chain fatty acid metabolism. These proteins not only exhibit close proximity in the investigated fireflies' genomes, but also share considerable sequence similarity, suggesting a common evolutionary origin. Moreover, we report that the region is not well conserved outside of members of the firefly family. We support our hypothesis by exploring the relationship of *LUC* and *4CL* genes using multiple sequence alignments, phylogenetic tree reconstruction methods and machine learning approaches (AlphaFold, ProtT5).

Supplementary Tab. 1 Short names for all proteins

| proteins NCBI id and location	Protein name	Protein short name
NW_022170249.1-28263852-28265083	transmembrane protein 50B	TMEM50B
NW_022170249.1-28265708-28279892	fanconi anemia group M protein	FANCM
NW_022170249.1-28280249-28282406	uncharacterized protein LOC116175580	unc1
NW_022170249.1-28282482-28283781	E3 ubiquitin-protein ligase RNF113A	RNF113A
NW_022170249.1-28284629-28286909	T-complex protein 1 subunit gamma	CCT3
NW_022170249.1-28287100-28287889	uncharacterized protein LOC116175589	unc2
NW_022170249.1-28288733-28303066	NACHT domain- and WD repeat-containing protein 1	NWD1
NW_022170249.1-28298883-28314673	structural maintenance of chromosomes 6	SMC6
NW_022170249.1-28317165-28319046	E3 ubiquitin-protein ligase TRAIP	TRAIP
NW_022170249.1-28318966-28319514	60S ribosomal protein L29	RPL29
NW_022170249.1-28326717-28329728	4-coumarate–CoA ligase 1-like	4CL1-like_1
NW_022170249.1-28332169-28336707	luciferin 4-monooxygenase-like	LUC-like
NW_022170249.1-28340362-28342515	luciferin 4-monooxygenase	LUC
NW_022170249.1-28345647-28347862	4-coumarate–CoA ligase 1-like	4CL1-like_2
NW_022170249.1-28350257-28353544	4-coumarate–CoA ligase 1-like	4CL1-like_3
NW_022170249.1-28355040-28356894	microsomal glutathione S-transferase 1-like	MGST1-like
NW_022170249.1-28360457-28398731	lutropin-choriogonadotropic hormone receptor-like	LHCGR-like
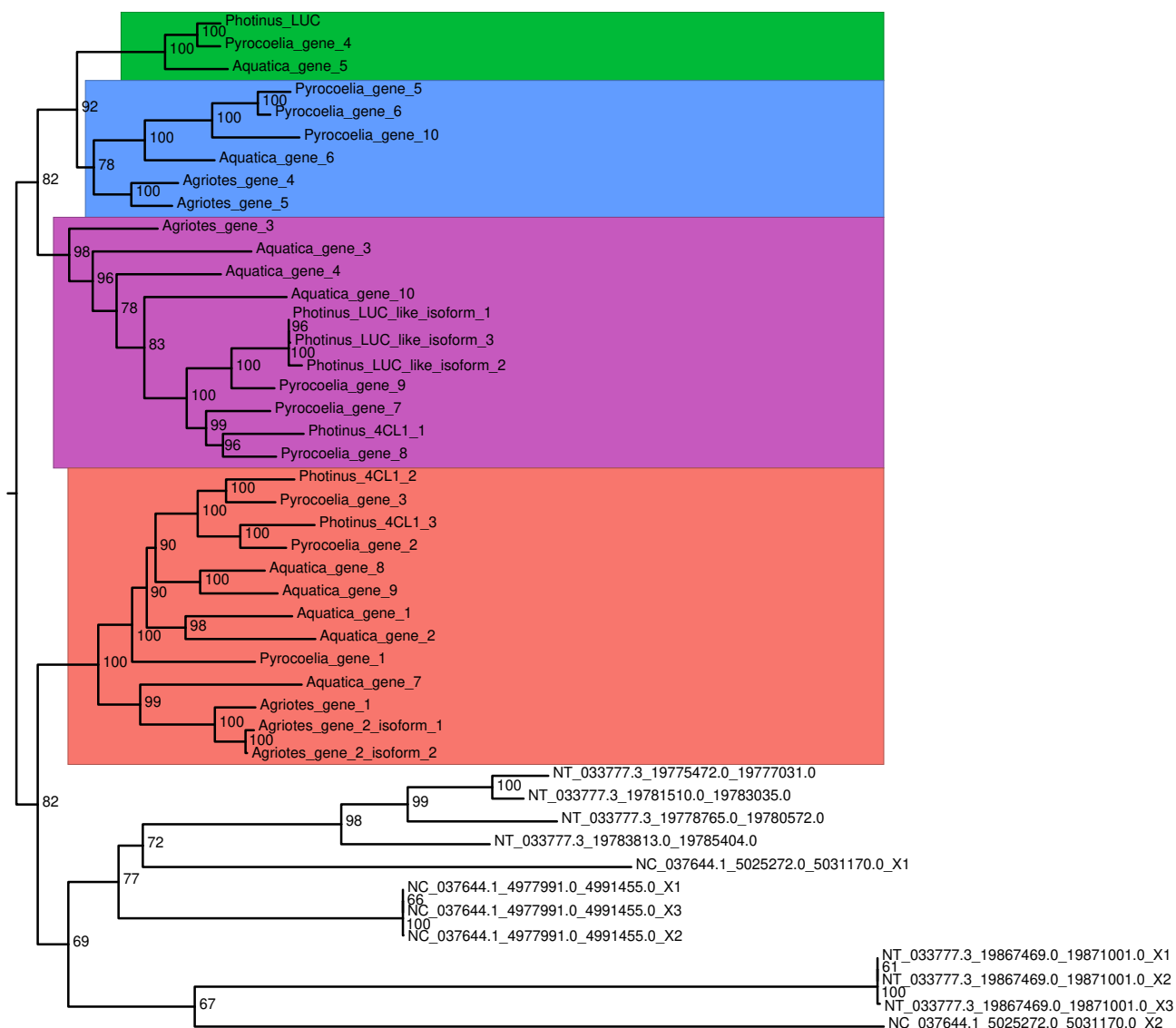
Supplementary Fig. 1 Maximum likelihood tree of LUC and LUC-related proteins in *Photinus pyralis*, *Pyrocoelia pectoralis*, *Aquatica leii* and *Agriotes lineatus* with outgroup (*Drosophila melanogaster*, *Apis mellifera*). Tip labels for *Photinus pyralis* are the gene names from the reference annotation. Numbers at the end of gene names correspond to their position on the track in Fig. 1. Nodes labelled with bootstrap values. LUCs in green, LUC-related groups 1 - 3 in blue, purple and red, respectively.
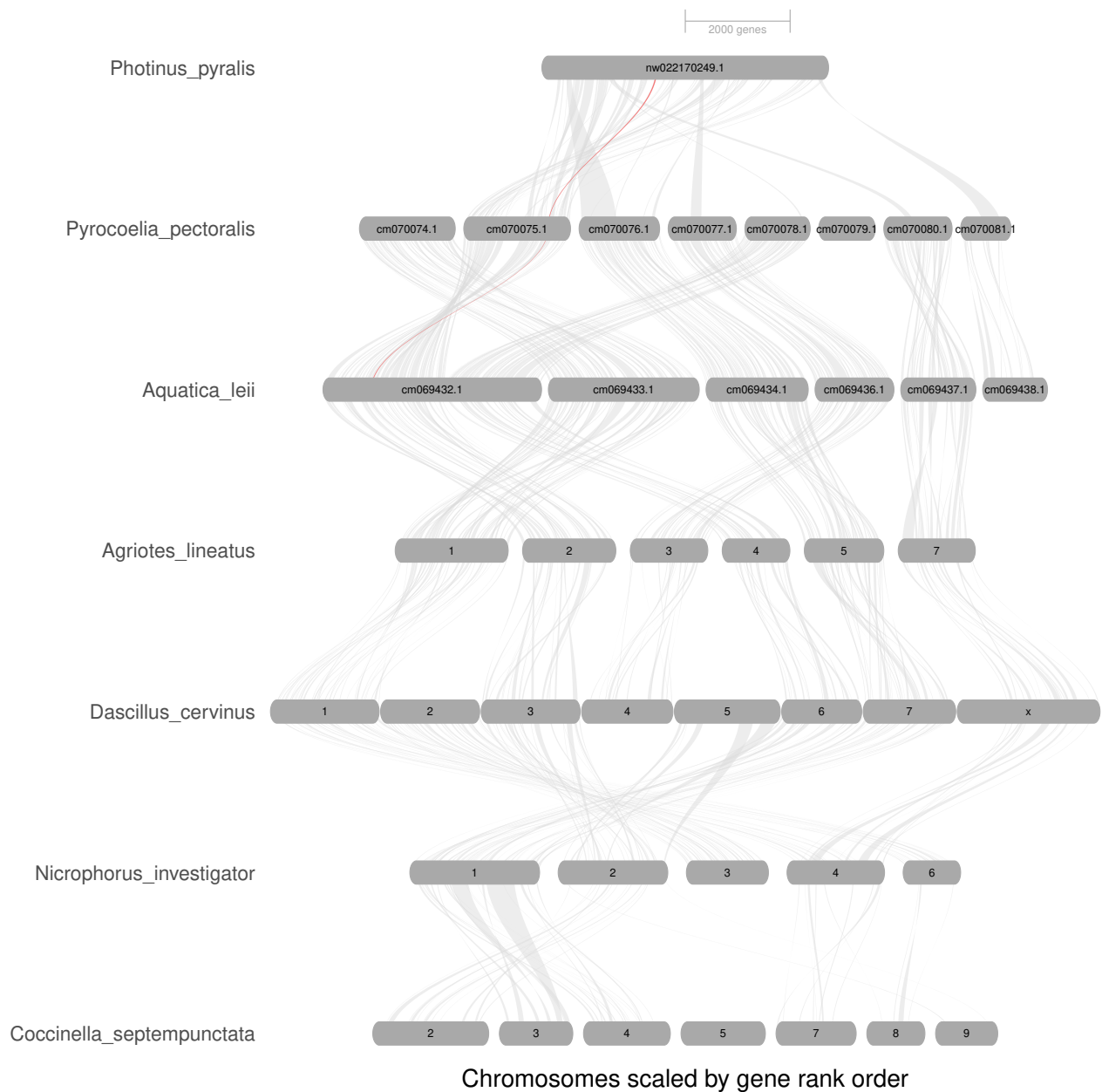
Supplementary Fig. 2 Result of a GENESPACE run with the scaffold containing the *LUC* gene in *Photinus pyralis* and the whole genomes of the other two fireflies and four non-luminous beetles. The whole surveyed region (*LUC*-related cluster + flanking genes) and its computationally determined syntenic region is highlighted in red.
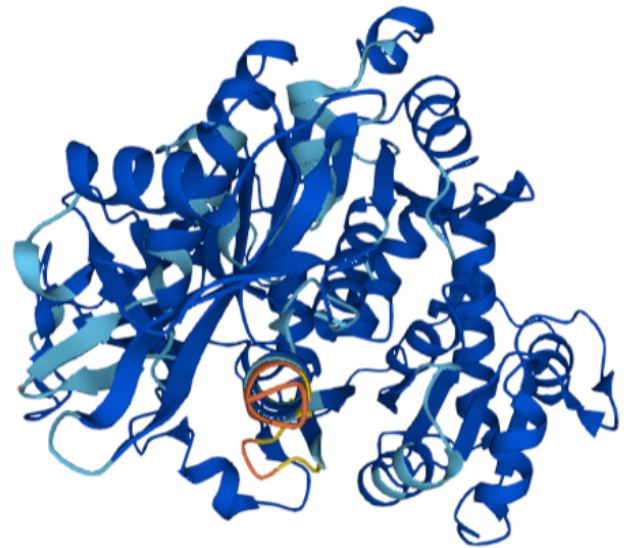
Supplementary Fig. 3 On the left side is the structure of the NWD protein in *Agrilus Planipennis*. On the right side is the LUC protein in *Photinus Pyralis*. Structures from Uniprot.
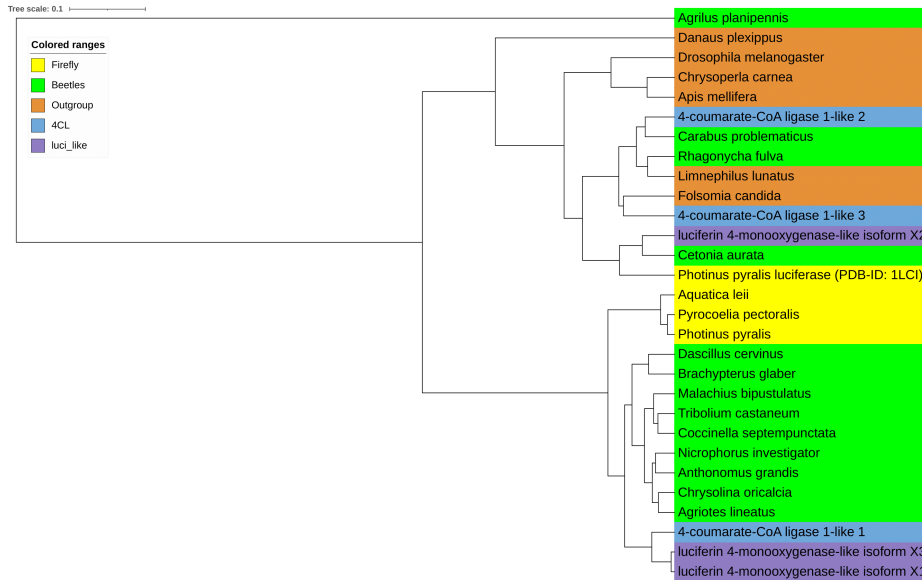
(a)
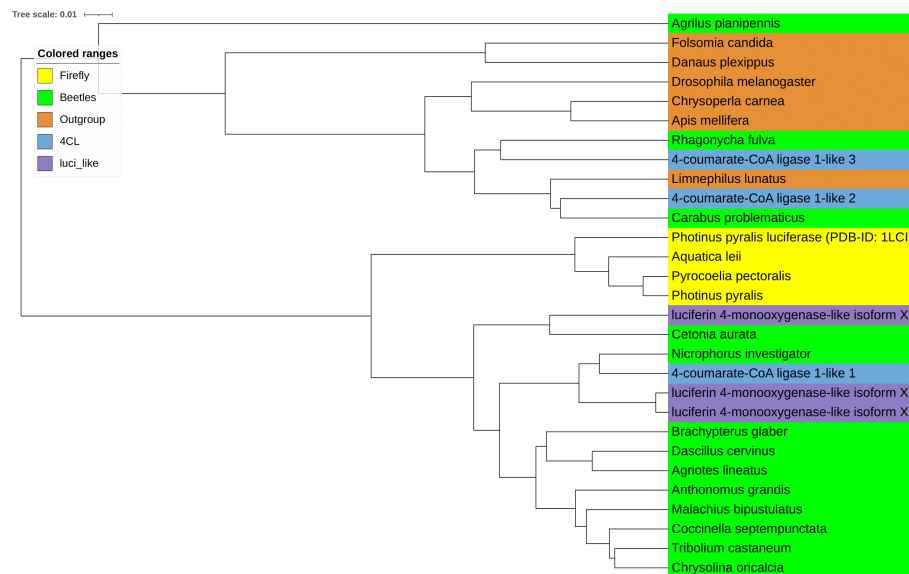
(b)

Supplementary Fig. 4 Results of the 3D structure comparison between the luciferase in different species and the other genes of the LUC-4CL group in Photinus pyralis **a** Dendrogram constructed from the TM scores from the lower triangular matrix of the heatmap with proteins grouped by their species and accordingly their type **b** Dendrogram constructed from the TM scores from the upper triangular matrix of the heatmap with proteins grouped by their species and accordingly their type

```
noah on noahs-thinkpad💻: ~ via ⭕ gobi took 10s
❯ python
Python 3.10.13 (main, Sep 11 2023, 13:44:35) [GCC 11.2.0] on linux
Type "help", "copyright", "credits" or "license" for more information.
>>> import numpy as np
>>> np.random.seed(2021)
>>> us = ['Adrian', 'Andreas', 'Chiara', 'Noah', 'Rebecca']
>>> np.random.shuffle(us)
>>> us
['Chiara', 'Adrian', 'Noah', 'Andreas', 'Rebecca']
>>> █
```

Supplementary Fig. 5 Screenshot of the Python code used to determine the order of authorship. Python version 3.10.13 with NumPy version 1.24.3 was used.