

Relaciones evolutivas ligadas a orígenes independientes de la bioluminiscencia en múltiples taxones

Estefanía Ceballos Benítez

Sara Agudelo Castrillón

Introducción

La bioluminiscencia, o la capacidad de un organismo vivo para generar luz visible, hace referencia al resultado de una reacción bioquímica en que una enzima luciferasa específica, cataliza la oxidación de un sustrato de molécula pequeña denominada luciferina. Esta, ha sido adaptada para satisfacer objetivos de comunicación visual dentro y entre especies, supervivencia, depredación y apareamiento (Tsarkova, 2021).

En contraste, actualmente se han encontrado especies bioluminiscentes en aproximadamente 700 géneros pertenecientes a múltiples linajes evolutivos como bacterias, dinoflagelados, artrópodos, moluscos, anélidos, equinodermos, urocordados o vertebrados (Delroisse, *et al.*, 2021). No obstante, la bioluminiscencia es un fenómeno principalmente marino y debido a la gran diversidad de organismos que utilizan este proceso, se ha utilizado para detectar actividades biológicas en las profundidades del océano, la presencia de animales pelágicos, y para evaluar biomasa para estudios oceanográficos (Martini & Abadejo, 2017).

En esencia, a pesar de que la reacción base de bioluminiscencia se considere genérica en todos los organismos que presentan dicha característica, estudios recientes establecen existen al menos 94 orígenes independientes de bioluminiscencia en todo el árbol de la vida. Sin embargo, mejoras futuras en la resolución filogenética, la reconstrucción del estado ancestral e investigaciones en nuevos taxones luminosos u orígenes independientes probablemente aumentarán esta estimación (Lau & Oakley, 2020).

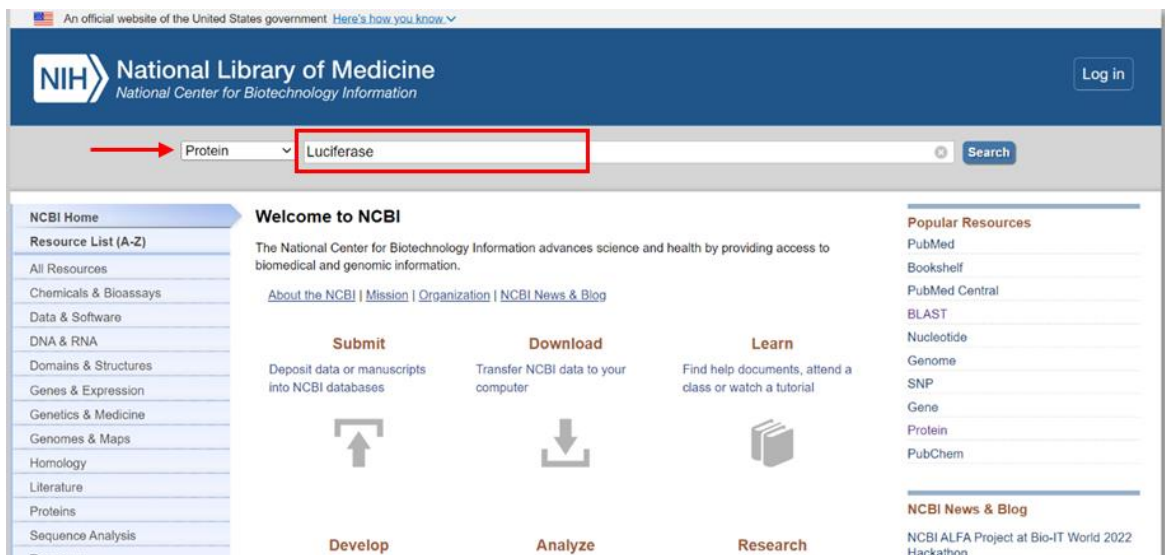
Del mismo modo, se han descrito múltiples luciferinas y luciferasas en una amplia variedad de taxones. Por lo que, las luciferasas suelen ser consideradas "específicas de un taxón", cabe aclarar que los organismos relacionados filogenéticamente a veces pueden depender de enzimas no homólogas para la fotogénesis, lo que respalda la evolución convergente de la bioluminiscencia. De acuerdo con lo anterior, no existe un ancestro luminoso común para todas las especies bioluminiscentes (Delroisse, *et al.*, 2021). Asimismo, se sugiere que la evolución multiconvergente

de la bioluminiscencia demuestra la existencia de intensas presiones selectivas que respaldan la aparición de mecanismos de bioluminiscencia durante la evolución del organismo (Hardock *et al.*, 2010).

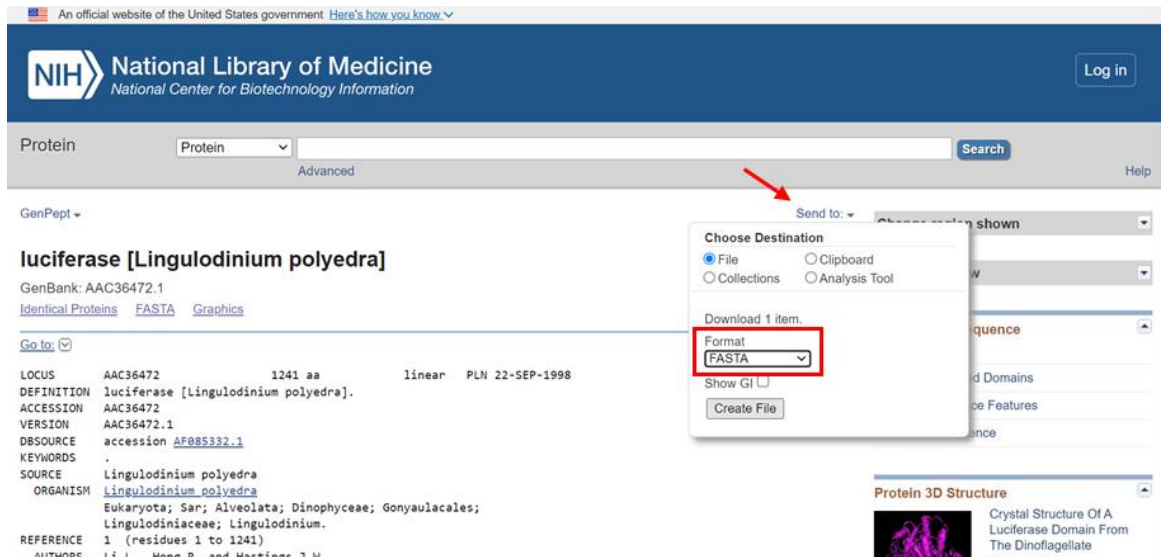
Ahora bien, en el presente estudio desarrollaremos un alineamiento de un total de 24 secuencias de luciferasa pertenecientes diferentes especies y pensadas de manera que 8 filos estuvieran presentes en el estudio, para posteriormente crear un mapa filogenético con el fin de evidenciar las relaciones evolutivas tanto entre taxones como entre especies que producen luciferasa, enfocándonos en la identificación de un posible ancestro común luminoso u origen independiente.

Metodología

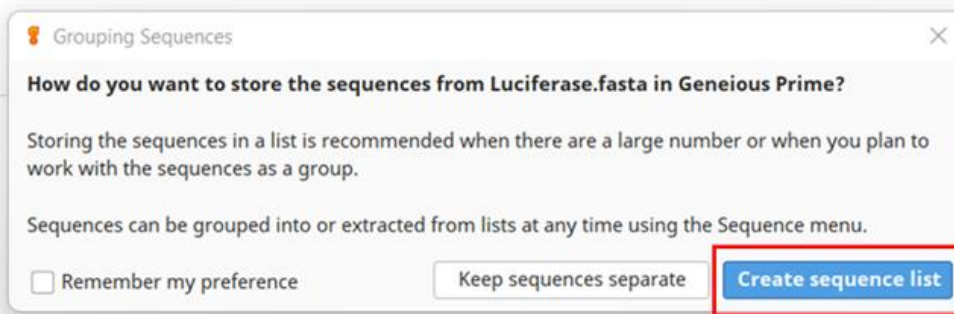
1. Luego de revisar la literatura, procedimos a buscar de manera manual cada una de las proteínas que conformarían nuestra reconstrucción filogenética. En este punto utilizamos entre 1 y 5 especies de diferentes géneros para representar los principales filos que albergan organismos productores de luciferasa (Ver Tabla 1 en Material Complementario). Los datos los obtuvimos de la base de datos de proteínas del NCBI, <https://www.ncbi.nlm.nih.gov/>.



2. Una vez tuvimos identificadas las secuencias de aminoácidos a utilizar, descargamos cada una de estas como archivos FASTA, y las identificamos con el nombre del organismo al que pertenecen para facilitar la posterior reconstrucción del árbol filogenético.



- Utilizamos el software *Geneious* para unir las 24 secuencias totales que obtuvimos en una lista de secuencias de formato FASTA y hacer más fácil el alineamiento y la manipulación de los datos.



- Utilizamos el algoritmo blastp del NCBI para encontrar proteínas homólogas a nuestras 24 secuencias base, esto con el fin de ampliar el set de datos y evaluar que tanta información podían proporcionarnos las proteínas iniciales.

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastp suite

blastn **blastp** blastx tblastn tblastx

Standard Protein BLAST

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>AAV35378.1 luciferase [Alexandrium tamarense]
MAELVRFLLDEANIDASVAAYVTKQLQLESVADFANYVTRDEYERGMQDDI
ISQLELFRSKPAGKLVAR
LRAAWQGAQQQMGAAPALKEAAPVAPADPQMGCIEKTGLEAGGTSKGGT

Query subrange [?](#)
From
To

Or, upload file Ninguno archivo selec. [?](#)

Job Title
Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Databases ☒ Standard databases (nr etc.) New ☐ Experimental databases [Try experimental clustered nr database](#) [For more info see What is clustered nr?](#) [Q](#)

Compare ☐ Select to compare standard and experimental database [?](#)

Standard

Database [?](#)

Organism Optional ☐ exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Optional ☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm ☐ Quick BLASTP (Accelerated protein-protein BLAST)
☒ blastp (protein-protein BLAST)
☐ PSI-BLAST (Position-Specific Iterated BLAST)
☐ PHI-BLAST (Pattern Hit Initiated BLAST)
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm [?](#)

BLAST Search database nr using Blastp (protein-protein BLAST)
☐ Show results in a new window

5. Filtramos nuestros resultados de acuerdo con el porcentaje de identidad y E value, para limitarnos principalmente a las coincidencias más “significativas”. Luego, descargamos el conjunto de datos asociado a cada una de las 24 secuencias base.

An official website of the United States government [Here's how you know](#)

NIH National Library of Medicine
National Center for Biotechnology Information

BLAST® » blastp suite

blastn **blastp** blastx tblastn tblastx

Standard Protein BLAST

BLASTP programs search protein databases using a protein query. [more...](#)

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [Clear](#)

>AAV35378.1 luciferase [Alexandrium tamarense]
MAELVRFLLDEANIDASVAAYVTQLOLESVADFANYWTRDEYERGMQDDI
ISQLELFRSKPAKLOVAR
LRAAIWQGAQQOMGAAPALKEAAPVAPADPQMIGICEKTGLEAGGTSKGGT

Query subrange [?](#)
From
To

Or, upload file ?

Job Title
Enter a descriptive title for your BLAST search [?](#)

☐ Align two or more sequences [?](#)

Choose Search Set

Databases ☒ Standard databases (nr etc.): [New](#) ☐ Experimental databases [Try experimental clustered nr database](#) [?](#)
For more info see What is clustered nr?

Compare ☐ Select to compare standard and experimental database [?](#)

Standard

Database ?

Organism Optional
 ☐ exclude [Add organism](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown. [?](#)

Exclude Optional
☐ Models (XM/XP) ☐ Non-redundant RefSeq proteins (WP) ☐ Uncultured/environmental sample sequences

Program Selection

Algorithm ☐ Quick BLASTP (Accelerated protein-protein BLAST)
☒ blastp (protein-protein BLAST)
☐ PSI-BLAST (Position-Specific Iterated BLAST)
☐ PHI-BLAST (Pattern Hit Initiated BLAST)
☐ DELTA-BLAST (Domain Enhanced Lookup Time Accelerated BLAST)
Choose a BLAST algorithm [?](#)

BLAST
☐ Show results in a new window

6. Creamos una lista de secuencias con todos estos datos en *Geneious* y obtuvimos un total de aproximadamente 350 secuencias de aminoácidos, número que descendió a 266 una vez descartamos las secuencias de vectores sintéticos. En el proceso de reconstrucción del árbol filogenético se descartó este set de datos por su gran tamaño, pues encontramos problemas al momento de ejecutarlos. En su lugar, utilizamos únicamente las 24 secuencias seleccionadas inicialmente. Sin embargo, esta búsqueda y filtrado de datos es importante para estudios posteriores que alcancen mayor profundidad en las relaciones evolutivas de la luciferasa en distintas taxa.
7. A partir del software MEGA, alineamos de las 24 secuencias con MUSCLE.



8. Utilizamos el paquete Modeltest-ng desde conda para evaluar los modelos evolutivos. En nuestro caso, con base a los resultados obtenidos, seleccionamos el modelo WAG+G4+F.

```
(base) root@SCAC:~# conda install -c bioconda modeltest-ng
```

```
(base) root@SCAC:~# modeltest-ng -i Luciferase_Alignment.fasta -d aa -o model-luciferase.txt
```

BIC	model	K	lnL	score	delta	weight
1	WAG+G4	1	-19663.0272	39702.6573	0.0000	0.9872
2	WAG+I+G4	2	-19663.2763	39711.3426	8.6853	0.0128
3	WAG+G4+F	20	-19630.7512	39793.6588	91.0015	0.0000
4	LG+G4	1	-19710.0783	39796.7595	94.1022	0.0000
5	WAG+I	1	-19710.9648	39798.5325	95.8752	0.0000
6	WAG+I+G4+F	21	-19631.0128	39802.3690	99.7117	0.0000
7	LG+G4+F	20	-19654.9153	39841.9869	139.3296	0.0000
8	BLOSUM62+G4	1	-19733.9894	39844.5818	141.9245	0.0000
9	BLOSUM62+I+G4	2	-19734.2356	39853.2611	150.6038	0.0000
10	JTT-DCMUT+G4	1	-19739.3571	39855.3172	152.6600	0.0000

AIC	model	K	lnL	score	delta	weight
1	WAG+G4+F	20	-19630.7512	39391.5024	0.0000	0.7793
2	WAG+I+G4+F	21	-19631.0128	39394.0256	2.5232	0.2207
3	WAG+G4	1	-19663.0272	39418.0543	26.5519	0.0000
4	WAG+I+G4	2	-19663.2763	39420.5526	29.0502	0.0000
5	LG+G4+F	20	-19654.9153	39439.8306	48.3281	0.0000
6	VT+G4+F	20	-19666.1258	39462.2516	70.7492	0.0000
7	VT+I+G4+F	21	-19666.3645	39464.7291	73.2267	0.0000
8	CPREV+G4+F	20	-19676.2480	39482.4960	90.9936	0.0000
9	WAG+I+F	20	-19678.0731	39486.1463	94.6438	0.0000
10	BLOSUM62+G4+F	20	-19689.7392	39509.4785	117.9761	0.0000

AICc	model	K	lnL	score	delta	weight
1	WAG+G4+F	20	-19630.7512	39393.5024	0.0000	0.7793
2	WAG+I+G4+F	21	-19631.0128	39396.0256	2.5232	0.2207
3	WAG+G4	1	-19663.0272	39419.0543	25.5519	0.0000
4	WAG+I+G4	2	-19663.2763	39421.5526	28.0502	0.0000
5	LG+G4+F	20	-19654.9153	39441.8306	48.3281	0.0000
6	VT+G4+F	20	-19666.1258	39464.2516	70.7492	0.0000
7	VT+I+G4+F	21	-19666.3645	39466.7291	73.2267	0.0000
8	CPREV+G4+F	20	-19676.2480	39484.4960	90.9936	0.0000
9	WAG+I+F	20	-19678.0731	39488.1463	94.6438	0.0000
10	BLOSUM62+G4+F	20	-19689.7392	39511.4785	117.9761	0.0000

Summary:

Partition 1/1:

	Model	Score	Weight
BIC	WAG+G4	39702.6573	0.9872
AIC	WAG+G4+F	39391.5024	0.7793
AICc	WAG+G4+F	39393.5024	0.7793

9. Finalmente, hicimos la reconstrucción del árbol en MEGA por el método *Maximum Likelihood* y teniendo en cuenta los parámetros anteriores.



M11: Analysis Preferences

Phylogeny Reconstruction

Option	Setting
ANALYSIS	
Statistical Method	→ <i>Maximum Likelihood</i>
PHYLOGENY TEST	
Test of Phylogeny	→ <i>None</i>
No. of Bootstrap Replications	→ <i>Not Applicable</i>
SUBSTITUTION MODEL	
Substitutions Type	→ <i>Amino acid</i>
Model/Method	→ <i>WAG with Freqs. (+F) model</i>
RATES AND PATTERNS	
Rates among Sites	→ <i>Gamma Distributed (G)</i>
No of Discrete Gamma Categories	→ <i>4</i>
DATA SUBSET TO USE	
Gaps/Missing Data Treatment	→ <i>Use all sites</i>
Site Coverage Cutoff (%)	→ <i>Not Applicable</i>
TREE INFERENCE OPTIONS	
ML Heuristic Method	→ <i>Nearest-Neighbor-Interchange (NNI)</i>
Initial Tree for ML	→ <i>Make initial tree automatically (Default - NJ/BioNJ)</i>
Initial Tree File	→ <i>Not Applicable</i>
Branch Swap Filter	→ <i>None</i>
SYSTEM RESOURCE USAGE	
Number of Threads	→ <i>4</i>

? Help

✕ Cancel

✓ OK

Resultados

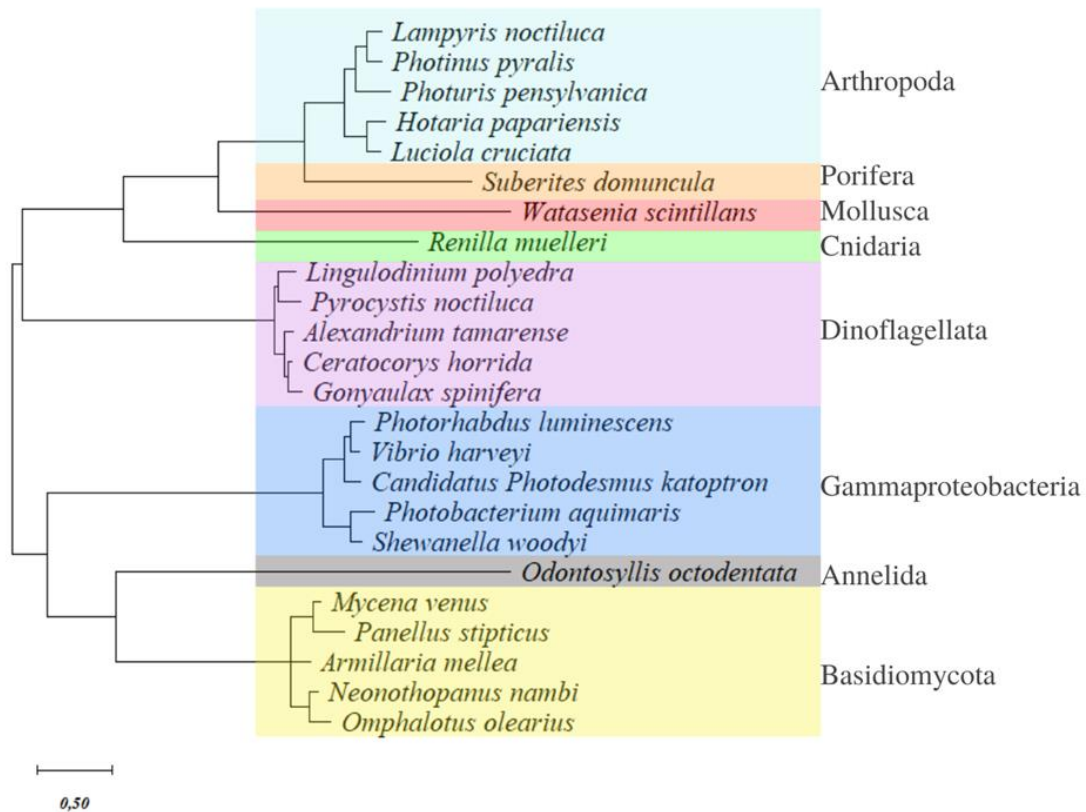


Figura 1. Maximum Likelihood Tree. Cada color corresponde a un filo.

Discusión

Los objetivos centrales de la biología evolutiva son comprender cómo se origina la complejidad, qué da forma a su evolución, y a su vez, cómo evolucionan repetidamente los rasgos complejos con funciones similares. En este caso, la bioluminiscencia, es un rasgo ecológicamente importante, un excelente sistema para estudiar los roles de la divergencia y la conservación en la evolución convergente debido a su complejidad biológica, biodiversidad y ubicuidad entre formas de vida (Lau & Oakley, 2020).

En primer lugar, en la Figura 1, como era de esperarse, podemos observar como el árbol construido relacionó los organismos de modo que las especies pertenecientes a un mismo filo quedaron agrupadas en un único clado, lo cual indica el uso de luciferasas probablemente homólogas.

Asimismo, en términos generales, la distancia de las ramas demuestra que hay una gran cantidad de sustituciones en las secuencias de aminoácidos que codifican para la misma proteína. Por lo que, es claro que múltiples proteínas bioluminiscentes aparecieron potencialmente durante la evolución a raíz de la manifestación independiente de nuevos genes. En consecuencia, sugerimos que la proteína luciferasa aparece como una convergencia, más que de una relación ancestro-descendiente.

Adicionalmente, en la Figura 1C del Material Complementario, es evidente que los puntajes de confianza indican que los dos grupos más basales son los que están menos soportados, mientras que la relaciones entre las especies del mismo filo presentan un mejor soporte, lo cual puede representar orígenes independientes entre subdivisiones.

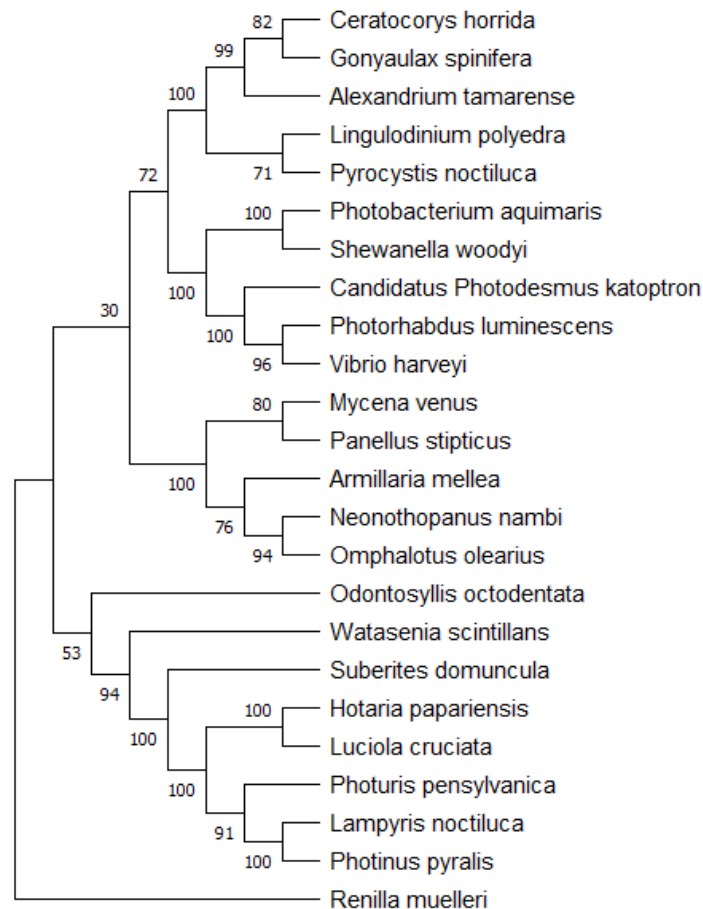
Conclusiones

Trás el análisis, es posible establecer que los datos obtenidos no reflejan la identificación de un posible ancestro común luminoso, más aún proponen una tendencia a orígenes independientes, lo cual concuerda con la literatura presentada. Del mismo modo, cabe mencionar que desarrollar la bioluminiscencia como un sistema para estudiar convergencia requiere el estudio no sólo de los diferentes niveles de la vida, sino de un enfoque interdisciplinario que incorpora biología molecular, computacional, ecología, evolución, bioquímica y demás.

Material Complementario

Especie	Filo
<i>Mycena venus</i>	Basidiomycota
<i>Armillaria mellea</i>	Basidiomycota
<i>Panellus stipticus</i>	Basidiomycota
<i>Omphalotus olearius</i>	Basidiomycota
<i>Neonothopanus nambi</i>	Basidiomycota
<i>Vibrio harveyi</i>	Gammaproteobacteria
<i>Photorhabdus luminescens</i>	Gammaproteobacteria
<i>Photobacterium aquimaris</i>	Gammaproteobacteria
<i>Candidatus Photodesmus katoptron Akat1</i>	Gammaproteobacteria
<i>Shewanella woodyi</i>	Gammaproteobacteria
<i>Luciola cruciata</i>	Arthropoda
<i>Photinus pyralis</i>	Arthropoda
<i>Photuris pensylvanica</i>	Arthropoda

Lampyris noctiluca	Arthropoda
Hotaria papariensis	Arthropoda
Lingulodinium polyedra	Dinoflagellata
Pyrocystis noctiluca	Dinoflagellata
Alexandrium tamarense	Dinoflagellata
Ceratocorys horrida	Dinoflagellata
Gonyaulax spinifera	Dinoflagellata
Odontosyllis octodentata	Annelida
Renilla muelleri	Cnidaria
Watasenia scintillans	Mollusca
Suberites domuncula	Porifera



Referencias

Delroisse, J., Duchatelet, L., Flammang, P., & Mallefet, J. (2021). Leaving the Dark Side? Insights Into the Evolution of Luciferases. *Front. Mar. Sci.* 8:673620. doi: 10.3389/fmars.2021.673620.

Lau, E. S., & Oakley, T. H. (2020). Multi-level convergence of complex traits and the evolution of bioluminescence. *Biological Reviews*. doi:10.1111/brv.12672

Haddock, S. H. D., Moline, M. A., and Case, J. F. (2010). Bioluminescence in the sea. *Ann. Rev. Mar. Sci.* 2, 443–493. doi: 10.1146/annurev-marine-120308-081028

Martini, S., & Haddock, S. (2017). Quantification of bioluminescence from the surface to the deep sea demonstrates its predominance as an ecological trait. *Sci Rep* 7, 45750. <https://doi.org/10.1038/srep45750>

Tsarkova, A. S. (2021). Luciferins Under Construction: A Review of Known Biosynthetic Pathways. *Front. Ecol. Evol.* 9:667829. doi: 10.3389/fevo.2021.667829.