Identification of orphan genes and de novo gene birth in the evolution of plant parasitic nematodes

Ercan SEÇKIN[1,2], Dominique COLINET[1], Marc BAILLY-BECHET[1], Edoardo SARTI[2], Etienne GJ DANCHIN[1]

1 Institut Sophia Agrobiotech - INRAE, Université Côte d'Azur, CNRS, 400 Route des Chappes, 06903, Sophia Antipolis, France

2 Centre Inria d'Université Côte d'Azur, 2004 Route des Lucioles, 06902, Sophia Antipolis, France

Corresponding Author: ercan.seckin@inrae.fr

Keywords

evolution, orphan genes, de novo gene birth, plant parasitic nematodes.

Abstract

Genes lacking homology in other species are systematically found in genomes and have various important functions [1]. The presence of some of these so-called orphan genes can be explained by extensive divergence from pre-existing genes up to the point where no homologues can be identified. However, another possibility is de novo gene birth, which consists in the emergence of genes from non-genic regions [2]. Here, we present a bioinformatics pipeline designed to identify orphan genes and to investigate their emergence dynamics and possible de novo origin. The developed pipeline was applied to the genomes of worldwide distributed plant parasitic nematodes of the genus Meloidogyne, which represent an interesting model for studying the origin of orphan genes: indeed, (i) the majority of genes involved in plant parasitism by these nematodes have no recognizable homology in the rest of nematodes and other species [3], and (ii) several of these species have undergone whole genome duplications with the potential for high divergence between gene copies [4,5].

A complementary homology search was conducted using OrthoFinder [6], SonicParanoid [7], and Diamond [8] to identify proteins specific to the genus Meloidogyne. In order to accomplish that, the nr database was used together with the genomes of 85 nematode species representing the entire phylum Nematoda. This analysis revealed that 24.3% of the proteins predicted in the eight Meloidogyne proteomes studied lacked recognizable homology in the remaining species. Transcriptomic data supported the expression of 91% of these genes, which were therefore considered "transcribed" orphan genes, representing 21.9% of all genes in the genus Meloidogyne.

Furthermore, once the orphan genes have been identified, it was necessary to develop an approach that allowed the distinction to be made between highly diverged orphan genes from pre-existing genes and de novo genes. Therefore, to further study their origin, the ancestral sequences of the orphan genes were constructed by using gap-aware methods guided by the species tree. The distinction between highly diverged orphan genes from pre-existing genes and de novo emerged genes was then established by aligning these ancestral sequences with the genome of a closely related species. This was done with a particular focus on genes that are common to all Meloidogyne species but absent in other species. The conservation of these genes within the Meloidogyne genus suggests that they were retained during evolution after their emergence in a common ancestor of the genus and might play important roles. To date, results indicate that at least 7.5% of the highly conserved genes (corresponding to 1,757 genes) are the result of de novo gene birth.

Overall, this study introduces a pipeline that combines homology, transcriptomics, and ancestral sequence prediction for the purpose of identifying and investigating the origin of orphan genes in Meloidogyne species, with a particular focus on genes common to the entire genus. The pipeline allows us to predict the total number of orphan genes and their possible origin for genes common to the entire genus. However, several questions remain unanswered, prompting further investigation. The origin of the remaining orphan genes that are not common to whole Meloidogyne is still unclear. Subsequently, the approach used in the pipeline for distinguishing between highly diverged orphan genes from pre-existing genes and de novo emerged genes will be more generalized so that it can be applied to all orphan genes. Also, currently, there is still no establishment between the orphan genes and their functions in Meloidogyne. Therefore, we are investigating structural approaches for better characterizing them in the goal of elucidating their functions.

References

1. Tautz D, Domazet-Lošo T. The evolutionary origin of orphan genes. Nat Rev Genet. 2011 Oct;12(10):692–702.
2. Weisman CM. The Origins and Functions of De Novo Genes: Against All Odds? J Mol Evol. 2022 Aug;90(3–4):244–57.
3. Mitchum MG, Hussey RS, Baum TJ, Wang X, Elling AA, Wubben M, et al. Nematode effector proteins: an emerging paradigm of parasitism. New Phytol. 2013 Sep;199(4):879–94.
4. Castagnone-Sereno P, Danchin EGJ, Perfus-Barbeoch L, Abad P. Diversity and Evolution of Root-Knot Nematodes, Genus *Meloidogyne* : New Insights from the Genomic Era. Annu Rev Phytopathol. 2013 Aug 4;51(1):203–20.
5. Koutsovoulos GD, Poullet M, Elashry A, Kozlowski DKL, Sallet E, Da Rocha M, et al. Genome assembly and annotation of Meloidogyne enterolobii, an emerging parthenogenetic root-knot nematode. Sci Data. 2020 Oct 5;7(1):324.
6. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019 Dec;20(1):238.
7. Cosentino S, Iwasaki W. SonicParanoid2: fast, accurate, and comprehensive orthology inference with machine learning and language models [Internet]. Bioinformatics; 2023 May [cited 2024 Jan 17]. Available from: http://biorxiv.org/lookup/doi/10.1101/2023.05.14.540736
8. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015 Jan;12(1):59–60.