Origins of lineage-specific elements via gene duplication, relocation, and regional rearrangement in *Neurospora crassa*

Zheng Wang¹, Yen-Wen Wang¹, Takao Kasuga², Hayley Hassler¹, Francesc Lopez-Giraldez³, Caihong Dong⁴, Oded Yarden⁵, and Jeffrey Townsend⁶

¹Yale University School of Public Health
²University of California Davis
³Yale School of Medicine
⁴Institute of Microbiology Chinese Academy of Sciences
⁵Hebrew University of Jerusalem Robert H Smith Faculty of Agriculture Food and Environment
⁶Yale University

July 14, 2023

Abstract

The origin of new genes has long been a central interest of evolutionary biologists. However, their novelty evades reconstruction by the classical tools of evolutionary modeling. This evasion of insight from deep ancestral investigation necessitates intensive study of model species within well-sampled, recently diversified, clades. One such clade is the model genus *Neurospora*, members of which lack recent gene duplications, yet harbor clusters of lineage-specific genes (LSGs) adjacent to the telomeres. Several *Neurospora* species are comprehensively characterized organisms apt for studying the evolution of LSGs. Using gene synteny, we documented that 78% of *Neurospora* LSGs clusters are located in chromosomal regions featuring extensive tracts of non-coding DNA and duplicated genes. Here we report several instances of LSGs that are likely from regional rearrangements and potentially from gene rebirth. To broadly investigate functions of LSGs, we assembled transcriptomics data from 68 experimental data points and identified co-regulatory modules using Weighted Gene Correlation Network Analysis, revealing that LSGs are widely but peripherally involved in known regulatory machinery for diverse functions. The ancestral status of mas-1 and its neighbors was investigated in detail, suggesting that it arose from an ancient lysophospholipase precursor that is ubiquitous in lineages of the Sordariomycetes; mas-1 plays a role in cell-wall integrity and cellular sensitivity to antifungal toxins. Our discoveries illuminate a “rummage region” in the *N. crassa* genome that enables formation of new genes and functions to arise via gene duplication and relocation, followed by fast mutation and recombination facilitated by tandem repeats and deconstrained non-coding sequences.

Hosted file

N. discreta (107,819 bp)
N. tetrasperma (123,799 bp)
N. crassa (152,243 bp)

Non-LSG gene (single-copy)
Non-LSG gene (multi-copied)
LSGs only in N. discreta
LSGs only in N. tetrasperma
LSGs only in N. crassa
Non-coding with rearrangement
het NCU07230
het NCU07241

S1
S2
S3
Rearrangement and gene duplication
Function loss and became a non-coding region
Origin of a LSG from the non-coding region

Expression in fold-change
conidial germination on Bird medium
conidial germination on maple sap medium

WT
cordial-1