

# Comparative Genomics of Cyclin-Dependent Kinases Suggest Co-Evolution of the RNAP II C-Terminal Domain and CTD-Directed CDKs

Zhenhua Guo<sup>1</sup> and John W. Stiller<sup>1,2</sup>

<sup>1</sup> Howell Science Complex N108, Department of Biology, East Carolina University, Greenville, NC 27858

<sup>2</sup> Author for correspondence. Tel: 252-328-2738, Fax: 252-328-4178  
email: [stillerj@mail.ecu.edu](mailto:stillerj@mail.ecu.edu), [guoz@mail.ecu.edu](mailto:guoz@mail.ecu.edu)

## Abstract

We identify 101 Cyclin-dependent kinases (CDKs) family members from animals, plants, yeasts and three protists with complete genome sequences. Comparative analyses suggest that cell-cycle CDKs are present in all organisms sampled in this study. In contrast, no clear orthologs of transcription-related CDKs are detected in the most putatively ancestral eukaryotes, *Trypanosoma* or *Giardia*. Kinases involved in C-terminal domain (CTD) of RNA polymerase II phosphorylation, CDK7, CDK8 and CDK9, all are recovered as well-supported and distinct orthologous families. Significantly, clear orthologs of CDK7 and CDK8 are restricted to those organisms belonging to groups in which the RNAP II CTD is strongly conserved suggesting co-evolution of the CTD and these CTD-directed CDKs. Alternatively, extensive CTD phosphorylation may occur in only a subset of eukaryotes and, when present, this interaction results in greater stabilizing selection on both CTD and CDK sequences. Overall, our results suggest that transcription-related kinases originated after cell-cycle related CDKs.

## Introduction

Cyclin-dependent kinases (CDKs) are a large family of proteins that function in the cell cycle and gene transcription[1]. Among the most important and broadly studied of these roles is reversible phosphorylation of the C-terminal domain (CTD) of RNA polymerase II, part of a complex array of CTD/protein interactions that coordinate the RNAP II transcription cycle [2,3]. To date at least five of CDKs (CDK1, 2, 7, 8 and 9) have been shown to phosphorylate the CTD *in vitro*; they all have been referred to as ‘CTD kinases’ [4,5]. The RNAP II CTD is strongly conserved in some groups of eukaryotes, but highly degenerate or absent in others [6]; the reasons for these differences in stabilizing selection on CTD structure are not clear.

## Results and discussion

We identified 101 CDK family members from animals, plants, yeasts and three protists from which genome sequences have been completed. The 50% majority rule consensus tree of 4,000 likelihood trees, sampled from the posterior probability distribution from Bayesian phylogenetic inference, is shown in Figure 1.

Evolutionary investigations of RPB1 sequences show that canonical CTD heptads are conserved strongly in only a subset of eukaryotic groups, all apparently descended from a single common ancestor [6]. This “CTD-clade” is composed of animals, plants, fungi, and several related protistan groups, such as chytridiomycetes, choanoflagellates and slime molds. In our analyses, members of this “CTD-clade” are precisely the same eukaryotes to which clear orthologs of CDK7 and CDK8 are restricted. These findings suggest that the RNAP II CTD has undergone a co-evolutionary process with CDK7 and CDK8. In

