氏名（本籍）	古川　龍太郎(静岡県)
学位の種類	博士(生命科学)
学位記番号	博　第１００号
学位授与の日付	平成２９年３月１５日
学位授与の要件	学位規則第 4 条第 1 項該当
学位論文題目	Quest for early evolution of life based on phylogenetic analyses of aminoacyl tRNA synthetases
論文審査委員	（主査）　山岸　明彦　教授
	都筑　幹夫　教授
	井上　英史　教授
	渡邉　一哉　教授

論文内容の要旨

Searching for ancestor of Eukarya based on aminoacyl tRNA synthetase

Introduction

The three-domain phylogenetic system of life has been challenged, particularly with regard to the position of Eukarya. The recent increase of known genome sequences has allowed phylogenetic analyses of all extant organisms using concatenated sequence alignment of universally conserved genes; these data supported the two-domain hypothesis, which place eukaryal species as ingroups of the Domain Archaea. However, the origin of Eukarya is complicated: the closest archaeal species to Eukarya differs in single gene phylogenetic analyses depending on the genes. In this report, we performed molecular phylogenetic analyses of 23 aminoacyl-tRNA synthetases (ARS).

Materials and Methods

We selected two or three typical species from each order to reduce taxonomic bias. All protein sequences of 282 selected organisms (Archaea: 76, Bacteria: 142, Eukarya: 64) were collected from the National Center for Biotechnology Information. Protein sequences of 23 ARSs were searched with BlastP. Amino acid sequences of each ARS were aligned using alignment program and edited manually. The well-aligned regions of each alignment were selected from the final alignment using Trimming program. We reconstructed trees for 23 ARSs using Maximum Likelihood (RAxML) and Bayesian Inference (PhyloBayes) analyses.

Result and Discussion

Cytoplasmic ARSs in 12 trees showed a monophyletic Eukaryotic branch. One ARS originated from TACK superphylum. One ARS originated from Euryarchaeota and three originated from DPANN superphylum. Four ARSs originated from different bacterial species. The other 8 cytoplasmic ARSs were split into two or three groups in respective trees, which suggested that the cytoplasmic ARSs were replaced by secondary ARSs and the original ARSs have been lost during evolution of Eukarya. In these trees, one original cytoplasmic ARS was derived from Euryarchaeota and three were derived from DPANN superphylum.

Conclusion and Proposal

Our results strongly support the two-domain hypothesis. We discovered that rampant independent lateral gene transfers from several Archaeal species of DPANN superphylum have contributed to the formation of Eukaryal cells. Based on our phylogenetic analyses, we proposed a model for the establishment of Eukarya.

Evolution of aminoacyl tRNA synthetase based on composite tree analysis

Introduction

Expansion of amino acid repertory in early translation system is one of the largest scientific mysteries in early evolution of life. Many hypotheses regarding the evolution of genetic code have proposed on the expansion of amino acid repertory. Though the order of recruitment of amino acids into the protein synthesis has been proposed, no experimental evidences has been obtained. Aminoacyl-tRNA synthetase (ARS) is essential enzyme that attaches amino acid to cognate tRNA in translation system. The expansions of ARS might have contributed the extant translation system in early evolution before appearance of the last common ancestor *Commonote commonote*. To challenge early evolution of translation, I reconstructed composite trees of aminoacyl-tRNA synthetase. Tracing back to the ancestor of ARS of each class will lead us to the primitive translation system, when protein is emerging in RNA world.

Composite trees of each class have been reconstructed (Nagel and Doolittle 1991; 1995). Structure dendrogram of each class was reconstructed (Donoghue et al. 2003). These analyses provided important information that we can trace back to the class I ARS and the class II ARS ancestors. Aravind et al. have suggested that the catalytic domain of class I ARS is conserved as Rossmann-like topology, and the ancestor of class I ARS is diverged from primitive protein in RNA world (2003). Though increasing number of ARS data are available, the detail composite trees of ARS of each class have not been reported. Although Andam and Gogarten have reported composite tree of class II ARS, they have used limited number of species (2011). To clarifying the detailed evolutionary history, phylogenetic analysis using abundant taxonomical species is needed. I focused the root of each ARS in the composite tree of each subclass to reveal the position of *C.commonote* in the tree of life.

Materials and Methods

All protein sequences of 118 selected organisms (Archaea: 23, Bacteria: 57, Eukarya: 38) were collected from the National Center for Biotechnology Information. Protein sequences of 23 ARSs were searched with BlastP. Collected amino acid sequences were classifies into each class of ARS (class Ia [MetRS, ValRS, LeuRS, IleRS, CysRS and ArgRS], class Ib [GluRS, GlnRS and LysRS-class I], class Ic [TyrRS and TrpRS], class IIa [SerRS, ThrRS, GlyRS-α_2, ProRS and HisRS], class IIb [AspRS, AsnRS and LysRS-class II], and class IIc [PheRS-α and PheRS-β] and class IId [AlaRS and GlyRS-$\alpha_2\beta_2$]). Amino acid sequences of each subclass set of ARS were aligned using alignment program and edited manually. The well-aligned regions of each alignment were selected from the final alignment using Trimming program. We reconstructed seven composite trees of each subclass of ARS using Maximum Likelihood (RAxML) and Bayesian inference (PhyloBayes) analyses.

Result and Discussion

Seven composite trees of each subclass of ARS were reconstructed from seven composite alignment of each subclass (class Ia, class Ib, class Ic, class IIa, class IIb, class IIc and class IId).

In composite tree of class Ia, the root was placed between ArgRS and other ARSs. CysRS diverged earliest in both ML and BI tree. However, the position of *C. commonote* differs significantly depending on the method as well as the class Ia RS species. Accordingly, further analysis is needed to determine the root position in class Ia RS.

In composite tree of class Ib, monophyly of each ARS (LysRS-class I and GluRS/GlnRS) was supported. The root position of LysRS-class I was in archaeal group. Bacterial group has another LysRS-class II, the root position of LysRS-class I is not related to the position of *C. commonote*. The root position of GluRS was between Bacteria and Archaea/Eukarya group in both ML and BI analyses, supporting the position of *C. commonote* between Bacteria and Archaea/Eukarya group. GlnRS was ingroup of archaeal GluRS and was a sister group of Eukaryal GluRS, which shows that GlnRS was late invention evolved from GluRS.

In composite tree of class Ic, monophyly of each ARS (TyrRS, TrpRS) was supported. The root position of TyrRS was between Bacteria and Archaea/Eukarya group in both analysis. On the other hand, the root position of TrpRS was different in both analysis. Since the resolution of deep branch of TrpRS in BI analysis was very low, the root position of ML tree is more reliable, supporting the position of *C. commonote* between Bacteria and Archaea/Eukarya group.

The root of the tree was placed between HisRS and other ARSs in class IIa. The root position of GlyRS-1 was in archaeal group in both analyses, because bacterial group has another types GlyRS-2. The root position of ThrRS and SerRS is between Bacteria and Archaea/Eukarya group in ML and BI analyses, supporting the position of *C. commonote*

between Bacteria and Archaea/Eukarya group.

In composite tree of class IIb, monophyly of each ARS (LysRS-class II and AspRS/AsnRS) was supported. The root position of LysRS-class II was in bacterial group in both analyses, because archaeal group has LysRS-class I. Since the resolution of deep branch of AspRS in BI analysis was very low, the root position of ML tree is more reasonable, supporting the position of *C. commonote* between Bacteria and Archaea/Eukarya group. AsnRS was ingroup of archaeal AspRS, which showed that AspRS was late invention evolved from AspRS.

In composite tree of class IIc, monophyly of each ARS (PheRS-α, PheRS-β) was supported. The root position of PheRS-β was in bacterial group in both analyses with low resolution. The root position of PheRS-α was between Bacteria and Archaea/Eukarya group in both ML and BI analyses, supporting the position of *C. commonote* between Bacteria and Archaea/Eukarya group.

In composite tree of class IId, monophyly of each ARS (AlaRS and GlyRS-2) was supported. The root position of GlyRS-2 was in bacterial group in both analyses, because archaeal group has GlyRS-1. The root position of AlaRS was between Bacteria and Archaea/Eukarya group in ML tree. However in BI tree, the root position of AlaRS was in archaeal group.

Conclusion and Perspective

I have reconstructed 14 composite trees. Among the composite trees in this thesis, the root position was not clear in some ARS. GlyRS and LysRS have each two types of RS, and cannot be used to determine the position of root. AsnRS and GlnRS have evolved from AspRS and GluRS, respectively, and *C. commonote* did not have AsnRS or GlnRS. The reliable root position in my composite trees showed the root position of the *C.commonote* between Bacteria and Archaea in 14 cases (Table 5).

The order of incorporation of amino acid species in protein synthesis has been proposed based on amino acid abundance in the history after *C. commonote*. Though it is possible to find the order of branching of each ARS species in my composite trees, it is not directly related to the amino acid species used at the branching point: Both amino acid species used after the divergence may be used at the branching point. However, it may be possible to check the amino acid specificity of the ancestral ARS corresponding the branching point of the two ARS species. The resurrection and analysis of the ancestral ARS is on going in other members in my lab.

研究結果掲載誌

R. Furukawa, M. Nakagawa,T. Kuroyanagi, S. Yokobori, A. Yamagishi, Quest for Ancestors of Eukaryal Cells Based on Phylogenetic Analyses of Aminoacyl-tRNA Synthetases. Journal of Molecular Evolution, published online 26 November 2016

審査結果の要旨

本学位申請者は、翻訳系の鍵酵素であるアミノアシル tRNA 合成酵素（ARS）を研究対象として生命初期進化に関わる二つの課題に関する研究を行った。生物界には 20 アミノ酸種に関わる 23 ARS が存在している。本申請者は全生物界３つのドメインから代表的生物種を選び出し、23ARS に関する分子系統解析を行った。その結果 ARS 系統樹はいずれも、これまで広く知られている Woese が提唱した 3 ドメイン（Bacteria、Archaea、Eukarya）ではなく、Bacteria と Archaea の２分岐を支持していた。また、真核生物 ARS は ARS 種によって様々な異なった生物種由来であることがわかった。しかし、23ARS の進化系統樹を総合するならば、真核生物は古細菌の中でも TACK 上門と呼ばれる古細菌グループから誕生し、それにユーリアーキオータ由来の ARS と DPANN 上門と呼ばれる古細菌グループ由来の ARS が水平伝播して誕生したと推定された。こうして複数の古細菌から成立した真核生物祖先に、さらにアルファプロテオバクテリアが細胞内共生してミトコンドリアとなる他に、多数の真正細菌からの ARS 遺伝子水平伝播がおきていたことも明らかとなった。本申請者はこうした ARS の解析に基づき、3 ドメインに代わる全生物分類体系として 2 ドメイン（ドメイン・アーキアとドメイン・バクテリア）と３つのサブドメイン（古細菌、真核生物、真正細菌）からなる分類体系を提案した。

第二の課題として、本申請者は 23ARS の複合系統樹を作製することから、全生物の共通祖先が系統樹上のどこに存在するのかという点と、全生物共通祖先以前の ARS 進化の分析を行った。その結果、全生物共通祖先は真正細菌と古細菌の間に存在していたことが確認された。さらに、ARS の進化に関して、セリル tRNA 合成酵素とトレオニル tRNA 合成酵素が分岐した後でプロリル合成酵素が誕生したと推定された。

二つの課題の内の第一の課題に関しては J. Mol. Evol. に出版済みである。研究の関連分野を含む質問に対しても概ね適切に回答しており、本申請者が十分な学力をもち博士の学位授与に値する成果を上げたと結論した。