

Recent progress in understanding the early evolution of eukaryotes based on phylogenomic data-analyses





Background/introduction

- ✓ 3domains of Life
 - Bacteria, Archaea and Eukaryotes
- \checkmark Motivation of our study
 - Biodiversity, Data & Methods
- Our achievements
 - ✓ Tsukubamonas globose
 - ✓ Palpitomonas bilix
 - ✓ CRuMs
 - ✓ Fronicata
- Summary
- Acknowledgements

Diversity of organisms on the Earth





- Three domains of life
 - ✓ Archaea, Bacteria & Eukarya
- Diversity of the three domains are too far underrepresented

http://blogs.discovermagazine.com/loom/files/2011/03/three-domains.jpg

Diversity of bacteria





Diversity of archaea



Zaremba-Niedzwiedzka et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 2017 541:353-358

We have overlooked a large proportion of archaeal diversity

MOL EVOL MICROBES

Two domains, not three domains



Diversity of eukaryotes





- Diversity of eukaryotes have been underestimated
- Interested in reconstructing the tree of eukaryotes
 - ✓ Number of major groups
 - ✓ Branching order of the major groups

http://blogs.discovermagazine.com/loom/files/2011/03/three-domains.jpg

Our goal is



Cover the diversity of eukaryotes

Well resolved tree of eukaryotes

Robust methods for phylogeny

Generate large sequence data

Mol Evol Microbes UNIV.TSUKUBA

Blind men (researchers who are investigating the tree of eukaryotes) and the elephant (the diversity of eukaryotes)





hemoglobin α chains from four vertebrates

60

	* ** ** * * * * * * * * * * * * * * * *
carp	$\verb"mslsdkdkaavkglwakispkaddigaealgrmltvypqtktyfahwadlspgsgpvkkh"$
chicken	$\verb"mvlsaadknnvkgiftkiaghaeeygaetlermfttypptktyfphf-dlshgsaqikgh"$
dog	${\tt mvlspadktnikstwdkigghagdyggealdrtfqsfpttktyfphf-dlspgsaqvkah}$
Human	${\tt mvlspadktnvkaawgkvgahageygaealermflsfpttktyfphf-dlshgsaqvkgh}$

120

human	gkkv	adalt	nava	hvddmp	nals	also	ilhał	nklrvdp	vnfk	Llsh	cllvtla	aahlpaeft
dog	gkkv	adalt	tava	hlddlp	galsa	also	ilhay	klrvdp	vnfk	llsh	cllvtla	achhpteft
chicken	gkkv	vaali	eaan	hiddia	gtls	klsc	llhai	nklrvdp	vnfk	Ll <mark>g</mark> q	cflvvva	aihhpaalt
carp	gkvi	mgavg	davs	kiddlv	ggla	alse	lhai	Eklrvdp	anfk:	ilah	nvivvi	mlypgdfp
	**	*	*	**	*	**	***	*****	***	*	*	*

human	pavhasldkflasvstvltskyr
dog	pavhasldkfftavstvltskyr
chicken	pevhasldkflcavgtvltakyr
carp	pevhmsvdkffqnlalalsekyr
	* * * * * * * * * * * *

*identical residues among the four organisms

Amino acid residues vary among the four vertebrates ⇒substitutions occurred during vertebrate evolution

X*h* : time reversible Markov process, independent evolution for each branch

Based on the formula of Chapmann-Kolmogorov,

 $f(x_1, x_2, x_3, x_4 | \theta) \leftarrow \text{probability for getting data of } h'\text{th site}$ with a given $P_{ij}(t)$ and a given tree topology

$$= \sum_{i} \sum_{j} P\{X_{0} = i\} P\{X_{t_{1}} = x_{1}, X_{t_{2}} = x_{2}, X_{t_{5}} = j \mid X_{0} = i\}$$

$$\times P\{X_{t_{5}+t_{3}} = x_{3}, X_{t_{5}+t_{4}} = x_{4} \mid X_{t_{1}} = x_{1}, X_{t_{2}} = x_{2}, X_{t_{5}} = j, X_{0} = i\}$$

$$= \sum_{i} \left\{ \pi_{i} Pix_{1}(t_{1}) Pix_{2}(t_{2}) \sum_{j} Pij(t_{5}) Pjx_{3}(t_{3}) Pjx_{4}(t_{4}) \right\}$$

 π_i : composition of the base (amino acid) *i*



MOL EVOL MICROBES

On internal nods (ancestral species), all possibilities for the base (amino acid) *i* or *j* are summed over In case of base, *i* or j = A,C,G,T



 S_1, \dots, S_4 : extant species x_1, \dots, x_4 : sequence data of *h*'th site for extant species

- *i*, *j* : sequence data for ancestral species
 - : branch lengths

 $t_1, ..., t_5$





- Larger the data, more accurate the phylogenetic inference
 - ✓ Genome data
 - ✓ Transcriptome data
- Technology is rapidly growing
 - ✓ A single or few cells are sufficient to generate genome/transcriptome data



Background/introduction

- ✓ 3domains of Life
 - Bacteria, Archaea and Eukaryotes
- \checkmark Motivation of our study
 - Biodiversity, Data & Methods

Our achievements

- ✓ Tsukubamonas globose
- ✓ Palpitomonas bilix
- ✓ CRuMs
- ✓ Fronicata

Summary

Acknowledgements



> 8 major groups have been recognized



"Orphans" can be members of the 8 groups or represent novel groups

Major groups of eukaryotes



8 major groups have been recognized



"Orphans" can be members of the 8 groups or represent novel groups

Represents a novel lineage in Discoba





University of Tsukuba: Home of Tsukubamonas globosa







Image taken from Google Maps

Isolated from Hyoutaro-pond

Maintained in UR-YT medium at 20°C since October 2002

Tsukubamonas: How it looks like



AF

Yabuki et al. 2011 J Eukaryot Microbiol.

Tsukubamonas: phylogenomic analyses



MOL EVOL MICROBES Palpitomonas bilix

Represents the earliest diverged lineage in Cryptista



Yabuki et al. 2010 Protist



MOL EVOL MICROBES





- Isolated from seawater sampled on Macharchar island, on July 2006
- > Maintained in EMS medium at 20°C







Yabuki et al. 2010 Protist

Palpitomonas: Phylogenomic analysis



157 genes, 41,372 aa positions ML method, LG4X model Bayesian method, CAT + Γ model

- Pb is a basal to a clade of kathablepharids, goniomonads and cryptophytes
- Pb is a novel member of Cryptista



MOL EVOL MICROBES

Yabuki et al. 2014 Sci Rep.

Phylogenomic study assembled multiple orphan species into a group



MOL EVOL MICROBES

Emergence of a new group tentatively named as 'CRuMs'

- Diphylleids, Rigifila and Mantamonas have been classified as "orphan" eukaryotes
- Phylogenomic study by collaboration with institutes in the US and Canada

diphylleids

Collodictyon tricilatum





Diphylleia rotans









Emergence of a new group tentatively named as 'CRuMs'



Brown et al. 2018 Genome Biol Evol

MOL EVOL MICROBES

I MIV TOURUP

Metamonade Discoba AR" Haptist -ryptis heteroloboseans kathabrepharids Tsukubamonas euglenozoans ^Dalpitomonas goniomonads preaxostylids orminiferans endomyxans parabasalids cryptophytes centrohelids haptophytes Hemiarma ornicates filosians PAP020 akobids **SRT308**

MOL EVOL MICROBES

- Fornicata is composed of mitochondrion-less organisms
 - ✓ Parasitic species have been studies
 - ✓ Free-living members have been overlooked
 - Possess reduced types of mitochondria

Free-living fornicates



Free-living species have been isolated from various anaerobic/microaerophilic environments

Ultrastructure and Ribosomal RNA Phylogeny of the Free-Living Heterotrophic Flagellate Dysnectes brevis n. gen., n. sp., a New Member of the Fornicata

MOL EVOL MICROBES

Yubuki et al. J Eukaryot Microbiol 2007 57:191-200

Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures

Takishita et al. Extremphiles 2007 11:563-576

A wide diversity of previously undetected free-living relatives of diplomonads isolated from marine/saline habitats

Kolisko et al. Environ Microbiol 2010 12:2700-2710

Multigene Phylogenies of Diverse *Carpediemonas*-like Organisms Identify the Closest Relatives of 'Amitochondriate' Diplomonads and Retortamonads

Takishita et al. Protist 2010 163:344-355



- To model how eukaryotes adapted to anaerobic/microaerophilic environments, the robust relationship among fornicates
- Phylogenomic study by collaboration with institutes in the US, Canada, Czech Republic and Sweden

ecology & evolution

PUBLISHED: 13 MARCH 2017 | VOLUME: 1 | ARTICLE NUMBER: 0092

Organelles that illuminate the origins of *Trichomonas* hydrogenosomes and *Giardia* mitosomes

Michelle M. Leger^{1†‡}, Martin Kolisko^{1†‡}, Ryoma Kamikawa^{2†}, Courtney W. Stairs^{1‡}, Keitaro Kume³, Ivan Čepička⁴, Jeffrey D. Silberman⁵, Jan O. Andersson⁶, Feifei Xu⁶, Akinori Yabuki⁷, Laura Eme^{1‡}, Qianqian Zhang⁸, Kiyotaka Takishita⁷, Yuji Inagaki^{3,9}, Alastair G. B. Simpson¹⁰, Tetsuo Hashimoto^{3,9*} and Andrew J. Roger^{1*}

Mol Evol Microbes UNIV.TSUKUBA

Fornicata tree was fully resolved Proposed the evolution of reduced mitochondria



Summary



MOL EVOL MICROBES

- Identify new "orphans" in natural environments
- Clarify the placements of orphans by phylogenomic studies
- Challenge to resolve "the backbone"

Iterate the process from the field to the HPC









Culturing



2

https://www.forbes.com/

and the second s



https://assets.illumina.com/

Sequencing

Phylogenomic analyses



Acknowledgements





文部科学省

計算科学研究センター 学際共同利用

 Phylogenomicデータによる真核生物大系統解析 (稲垣)

科学研究費補助金

・基盤B H15-17年度 複数遺伝子配列情報に基づく真核生物の初期進 化の解明 (橋本)

Mol <mark>Evol</mark> Microbes

- ・基盤C H18-19年度 カタブレファリス類配列を含む連結データ系統解 析によるクロミスタ仮説の検討(稲垣)
- ・基盤B H21-24年度 ハプト・クリプト藻類を含む新奇巨大生物群の提唱とクロムアルベオラータ仮説の検証(稲垣)
- ・基盤A H23-26年度 新型分割イントロン切り出し因子同定に基づく真 核生物mRNA成熟機構進化の解明(橋本)
- ・挑戦的萌芽 H22-23年度 原始真核生物を求めて:新奇真核微生物 の形態・発現遺伝子・ミトコンドリアゲノム解析 (稲垣)
- •新学術領域研究計画研究 H23-27年度 ミトコンドリア・色素体以外の 共生オルガネラ成立過程の解明(稲垣)
- ・基盤B海外学術 H23-25年度 ミトコンドリアを持たない真核微生物群 フォルニカータの多様性の解明 (橋本)
- ・基盤B H26-28年度 新型分割イントロンのスプライシング機構と進化 多様性の解明(橋本)
- ・基盤B海外学術 H26-28年度 嫌気環境に生育する真核微生物の多 様性の解明 (橋本)
- ・基盤B H28-30年度 渦鞭毛藻細胞内に発見された新たな共生体痕跡 核ゲノムの解読 (稲垣)
- ・国際共同研究強化B H30-H34年度 海洋原生生物に共生する細菌多 様性の実態解明(稲垣)

大学からの研究経費

特別経費プロジェクト「生命の樹」研究機能の充実 – 多様な非モデル生物研究に基づく生命原理の再構築一



4 published works; highlighted in blue



3 unpublished works; highlighted in green