Dear Author,

Here are the proofs of your article.

- You can submit your corrections **online**, via **e-mail** or by **fax**.
- For **online** submission please insert your corrections in the online correction form. Always indicate the line number to which the correction refers.
- You can also insert your corrections in the proof PDF and **email** the annotated PDF.
- For fax submission, please ensure that your corrections are clearly legible. Use a fine black pen and write the correction in the margin, not too close to the edge of the page.
- Remember to note the **journal title**, **article number**, and **your name** when sending your response via e-mail or fax.
- **Check** the metadata sheet to make sure that the header information, especially author names and the corresponding affiliations are correctly shown.
- **Check** the questions that may have arisen during copy editing and insert your answers/ corrections.
- **Check** that the text is complete and that all figures, tables and their legends are included. Also check the accuracy of special characters, equations, and electronic supplementary material if applicable. If necessary refer to the *Edited manuscript*.
- The publication of inaccurate data such as dosages and units can have serious consequences. Please take particular care that all such details are correct.
- Please **do not** make changes that involve only matters of style. We have generally introduced forms that follow the journal's style.
  Substantial changes in content, e.g., new results, corrected values, title and authorship are not allowed without the approval of the responsible editor. In such a case, please contact the Editorial Office and return his/her consent together with the proof.
- If we do not receive your corrections **within 48 hours**, we will send you a reminder.
- Your article will be published **Online First** approximately one week after receipt of your corrected proofs. This is the **official first publication** citable with the DOI. **Further changes are, therefore, not possible.**
- The **printed version** will follow in a forthcoming issue.

Please note

After online publication, subscribers (personal/institutional) to this journal will have access to the complete article via the DOI using the URL: http://dx.doi.org/[DOI].
If you would like to know when your article has been published online, take advantage of our free alert service. For registration and further information go to: http://www.link.springer.com.

Due to the electronic nature of the procedure, the manuscript and the original figures will only be returned to you on special request. When you return your corrections, please inform us if you would like to have these documents returned.

Metadata of the article that will be visualized in OnlineFirst

ArticleTitle	The evolutionary landscape of the Rab family in chordates	
Article Sub-Title		
Article CopyRight	Springer Nature Switzerland AG (This will be the copyright line in the final PDF)	
Journal Name	Cellular and Molecular Life Sciences	
Corresponding Author	Family Name	**Albalat**
	Particle	
	Given Name	**Ricard**
	Suffix	
	Division	Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio)
	Organization	Universitat de Barcelona
	Address	Av. Diagonal 643, 08028, Barcelona, Spain
	Phone	
	Fax	
	Email	ralbalat@ub.edu
	URL	
	ORCID	
Corresponding Author	Family Name	**D'Aniello**
	Particle	
	Given Name	**Salvatore**
	Suffix	
	Division	
	Organization	Biology and Evolution of Marine Organisms
	Address	Stazione Zoologica Anton Dohrn Napoli, Villa Comunale 1, 80121, Naples, Italy
	Phone	
	Fax	
	Email	salvatore.daniello@szn.it
	URL	
	ORCID	
Author	Family Name	**Coppola**
	Particle	
	Given Name	**Ugo**
	Suffix	
	Division	
	Organization	Biology and Evolution of Marine Organisms
	Address	Stazione Zoologica Anton Dohrn Napoli, Villa Comunale 1, 80121, Naples, Italy
	Division	Molecular Cardiovascular Biology Division
	Organization	Cincinnati Children's Hospital Medical Center

	Address	Cincinnati, USA
	Phone	
	Fax	
	Email	
	URL	
	ORCID	
Author	Family Name	**Ristoratore**
	Particle	
	Given Name	**Filomena**
	Suffix	
	Division	
	Organization	Biology and Evolution of Marine Organisms
	Address	Stazione Zoologica Anton Dohrn Napoli, Villa Comunale 1, 80121, Naples, Italy
	Phone	
	Fax	
	Email	
	URL	
	ORCID	

Schedule	Received	17 December 2018
	Revised	29 March 2019
	Accepted	10 April 2019

Abstract

Intracellular traffic amongst organelles represents a key feature for eukaryotes and is orchestrated principally by members of Rab family, the largest within Ras superfamily. Given variations in Rab repertoire have been fundamental in animal diversification, we provided the most exhaustive survey regarding the Rab toolkit of chordates. Our findings reveal the existence of 42 metazoan conserved subfamilies exhibiting a univocal intron/exon structure preserved from cnidarians to vertebrates. Since the current view does not capture the Rab complexity, we propose a new Rab family classification in three distinct monophyletic clades. The *Rab* complement of chordates shows a dramatic diversification due to genome duplications and independent gene duplications and losses with sharp differences amongst cephalochordates, tunicates and gnathostome vertebrates. Strikingly, the analysis of the domain architecture of this family highlighted the existence of chimeric calcium-binding Rabs, which are animal novelties characterized by a complex evolutionary history in gnathostomes and whose role in cellular metabolism is obscure. This work provides novel insights in the knowledge of *Rab* family: our hypothesis is that chordates represent a hotspot of *Rab* variability, with many events of gene gains and losses impacting intracellular traffic capabilities. Our results help to elucidate the role of Rab members in the transport amongst endomembranes and shed light on intracellular traffic routes in vertebrates. Then, since the predominant role of Rabs in the molecular communication between different cellular districts, this study paves to way to comprehend inherited or acquired human disorders provoked by dysfunctions in *Rab* genes.

Keywords (separated by '-') Metazoan Rab - Calcium-binding Rab chimeras - Small GTPase superfamily - Amphioxus *Branchiostoma* - Ascidian *Ciona* - Larvacean *Oikopleura*

Footnote Information **Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s00018-019-03103-7) contains supplementary material, which is available to authorized users.

Cellular and Molecular Life Sciences
https://doi.org/10.1007/s00018-019-03103-7

ORIGINAL ARTICLE

The evolutionary landscape of the Rab family in chordates

Ugo Coppola[1,3] · Filomena Ristoratore[1] · Ricard Albalat[2] · Salvatore D'Aniello[1]

Received: 17 December 2018 / Revised: 29 March 2019 / Accepted: 10 April 2019

Abstract

Intracellular traffic amongst organelles represents a key feature for eukaryotes and is orchestrated principally by members of Rab family, the largest within Ras superfamily. Given variations in Rab repertoire have been fundamental in animal diversification, we provided the most exhaustive survey regarding the Rab toolkit of chordates. Our findings reveal the existence of 42 metazoan conserved subfamilies exhibiting a univocal intron/exon structure preserved from cnidarians to vertebrates. Since the current view does not capture the Rab complexity, we propose a new Rab family classification in three distinct monophyletic clades. The *Rab* complement of chordates shows a dramatic diversification due to genome duplications and independent gene duplications and losses with sharp differences amongst cephalochordates, tunicates and gnathostome vertebrates. Strikingly, the analysis of the domain architecture of this family highlighted the existence of chimeric calcium-binding Rabs, which are animal novelties characterized by a complex evolutionary history in gnathostomes and whose role in cellular metabolism is obscure. This work provides novel insights in the knowledge of *Rab* family: our hypothesis is that chordates represent a hotspot of *Rab* variability, with many events of gene gains and losses impacting intracellular traffic capabilities. Our results help to elucidate the role of Rab members in the transport amongst endomembranes and shed light on intracellular traffic routes in vertebrates. Then, since the predominant role of Rabs in the molecular communication between different cellular districts, this study paves to way to comprehend inherited or acquired human disorders provoked by dysfunctions in *Rab* genes.

Keywords Metazoan Rab · Calcium-binding Rab chimeras · Small GTPase superfamily · Amphioxus *Branchiostoma* · Ascidian *Ciona* · Larvacean *Oikopleura*

Introduction

Intracellular membrane-bounded organelles are a distinguishing feature of eukaryotic cells with a crucial role in most of their biological processes [1]. As a consequence, the sophisticated transportation of cargo by different carriers and vesicles amongst internal compartments needs to be finely regulated. In the intracellular traffic, the Ras-related in brain (Rab) proteins, originally discovered in yeasts [2], are key players in the control of membrane transport in *Eukarya* [3]. Rabs represent by far the largest family within the small GTPase superfamily, comprising more than 60 members in humans [4].

At the structural level, Rabs consist approximately of 200 amino acids and are generally connected to lipid bilayers via a long hypervariable domain with a prenyl group on two cysteine residues [5]. Rabs have the same organization of other GTPases with the P-loop domain, which is a widespread nucleotide-binding motif necessary for

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00018-019-03103-7) contains supplementary material, which is available to authorized users.

✉ Ricard Albalat
ralbalat@ub.edu

✉ Salvatore D'Aniello
salvatore.daniello@szn.it

[1] Biology and Evolution of Marine Organisms, Stazione Zoologica Anton Dohrn Napoli, Villa Comunale 1, 80121 Naples, Italy

[2] Departament de Genètica, Microbiologia i Estadística and Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain

[3] Present Address: Molecular Cardiovascular Biology Division, Cincinnati Children's Hospital Medical Center, Cincinnati, USA

cycling between GTP and GDP forms, and the Switch I and Switch II domains, which are responsible for the Rab folding and the interaction with different effectors [6]. At the functional level, Rab proteins govern the recruitment of vesicle tethering factors, motor proteins and receptors in a highly ordered manner [7, 8]. With the help of effectors, Rab family members are considered the fundamental regulators of Golgi organization and functioning [9]. Like other GTPases, they function by shifting between an inactive GDP state and an active GTP state catalysed by guanine nucleotide exchange factors (GEF). Specific GTPase activator proteins (GAP) act blocking the Rabs [10] and are successively recycled by GDP inhibitor proteins (GDI) located on membranes [11]. According to their functional role as key regulators of membrane transport and vesicular trafficking in the cellular endomembrane system, Rabs are involved in a myriad of basic biological processes being expressed in a wide range of tissues and developmental stages. It is not, therefore, surprising that Rab malfunctions are implicated in a plethora of human pathologies including Parkinson's disease [12], several inherited genetic disorders [13], neuroblastoma differentiation [14, 15] and invasive growth and metastasis of different tumours [16]. Actually, given that Rab proteins are deregulated in cancer, they have been selected for novel therapies [17].

According to phylogenetic data, Rab proteins have been classified into six supergroups, each predominantly localized in distinct cell compartments and controlling specific cellular trafficking steps [18, 19]. From this large diversity, a set of five Rabs (Rab1, Rab5, Rab6, Rab7 and Rab11) has been defined as the "core" Rabs [20, 21], which might represent the minimal protein trafficking machinery compatible with free life (not parasitic) [21]. These Rabs would regulate the basic secretory and endocytic pathways common to all eukaryotes [21, 22]. Initial phylogenetic analyses suggested that, however, the last eukaryotic common ancestor (LECA) had already a complex repertoire of Rabs, including the core Rabs plus other Rab subfamilies (e.g. Rab2, Rab8 and Rab18) [22], and following analyses increased the ancestral Rab repertoire up to 23 members, which is coherent with the presence of a functioning Golgi apparatus and the capability for both endocytosis and phagocytosis [22].

Lineage-specific duplications and losses of the primeval LECA repertoire have led to the current variable Rab number in the different eukaryote species, from approximately 20 in the majority of protists and unicellular algae [22] up to more than 60 in many multicellular species [18, 24–27]. Interestingly, although unicellular organisms traditionally have limited Rab complements, there are species with large repertoires of Rabs, due to the presence of many novel and divergent subfamilies: this unusual peculiarity has been indicated as a requirement for their life cycles [23–25].

Changes in Rab family have been associated with major events in the evolution of the eukaryotic cells and the unicellular to multicellular transition, especially for metazoan multicellularity [18, 24, 28]. Despite the invertebrate to vertebrate transition is also considered a major event in evolution for which variations in Rab complement have been important [18], a systematic comparison of the Rab components in non-vertebrate chordates versus vertebrate chordates is lacking. To fill this gap, we have studied the *Rab* family of five selected species of the chordate phylum: three non-vertebrate chordates—one cephalochordate and two tunicate (also known as urochordate) species—and two vertebrate chordates. We have identified and classified more than 243 Rabs, creating the most comprehensive catalogue of chordate Rabs and unravelling different patterns of evolution in each chordate subphyla: a conservative pattern in cephalochordates retaining the ancestral repertoire of chordate Rab subfamilies, a liberal pattern in tunicates characterized by numerous gene losses and an expansive pattern in vertebrates heavily impacted by the two rounds (1R, 2R) of whole-genome duplications (WGD) of this lineage. Finally, our analysis has improved the classification of metazoan Rabs and provided the first evolutionary reconstruction of the poorly studied calcium-binding Rab chimeras.

Results

Phylogenetic analysis of the Rab family

To understand the evolution of the chordate *Rab* family in the context of metazoans, we analysed the Rab complement of five selected species representing the three chordate subphyla: the amphioxus *Branchiostoma lanceolatum*, which belongs to the early branching, slow-evolving cephalochordate subphylum [29]; the ascidian *Ciona robusta* (formerly *C. intestinalis* [30]), and the larvacean *Oikopleura dioica,* both belonging to the fast-evolving tunicate subphylum [31]; and two vertebrate species: the mammal *Homo sapiens*, and the reptilian *Anolis carolinensis.* We selected this reptilian species because it is distantly related to mammals within the vertebrate subphylum, but it has not been affected by the extra whole-genome duplication event characteristic of teleost fish species (e.g. zebrafish, medaka or fugu). To provide a wider evolutionary framework to our study, we included several additional animal species on the basis of their position in the metazoan phylogeny. Two ambulacrarian species (non-chordate deuterostomes), the hemichordate *Saccoglossus kowalevskii* and echinoderm *Strongylocentrotus purpuratus*; three protostome species: the nematode *Caenorhabditis elegans* (Ecdysozoa), the annelid *Capitella teleta* and the mollusc *Lottia gigantea* (Lophotrochozoa); and one non-bilaterian species, the cnidarian *Nematostella vectensis*.

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

We conducted an exhaustive and systematic survey of *Rab* genes in the available databases of these 11 selected animal species, retrieving 485 Rab sequences, which is the most comprehensive catalogue of metazoan *Rab* genes compiled thus far. We aligned 457 protein sequences to build the phylogenetic tree shown in Fig. 1 (Supplementary files 1 and 2), excluding 28 partial or highly divergent Rabs (Supplementary file 3).

The phylogenetic reconstruction recovered 42 distinct metazoan Rab subfamilies (support values from approximate likelihood ratio test (aLRT) and from the Bayesian-like transformation of aLRT (aBayes) higher than 90% for 34 out of the 42 subfamilies), most of them with representatives from cnidarians to vertebrates (Figs. 1, S1). Phylogeny consistently supports the orthology of each *Rab* member from *N. vectensis* to *H. sapiens*, underpinning the scenario in which the last metazoan common ancestor (LMCA) already possessed most of the Rab subfamily diversity [18, 24]. In fact, our results summarized in Fig. 2 supported the existence of at least 38 Rab subfamilies in the LMCA, considering two losses in *N. vectensis* (Rab32/38 and RabX1) since they are present in the sponge

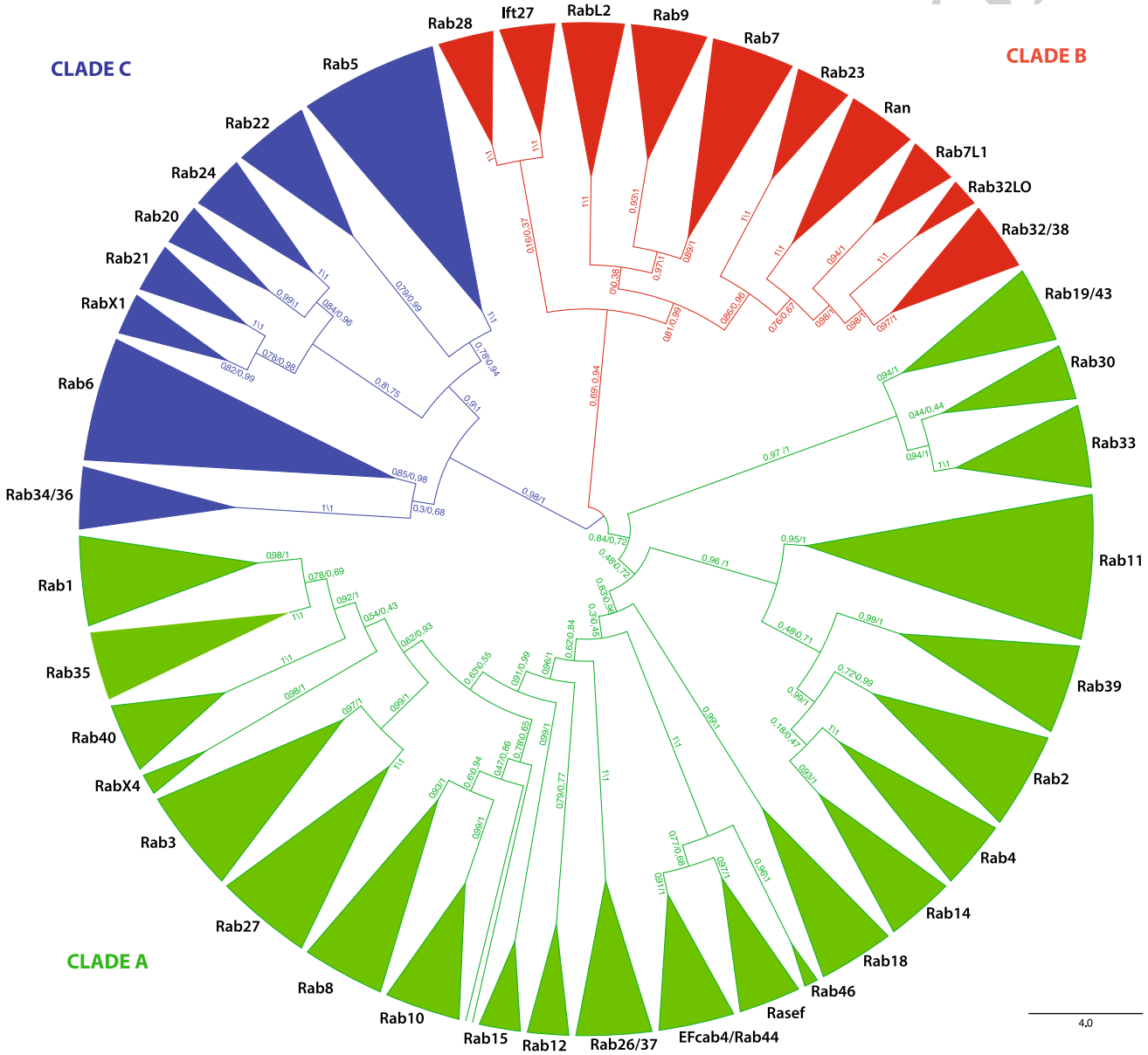


Fig. 1 Rab phylogeny in Metazoa. The cladogram shows three monophyletic Rab clades highlighted with different colours (A green, B red and C blue) encompassing Rab proteins of *Nematostella vectensis, Caenorhabditis elegans, Capitella teleta, Lottia gigantea, Saccoglossus kowalevskii, Strongylocentrotus purpuratus, Branchiostoma lanceolatum, Ciona robusta, Oikopleura dioica, Anolis carolinensis, Homo sapiens.* Values at the branches represent replicates obtained using aLRT and aBayes methods

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

Fig. 2 Rab toolkit. Gene duplications and losses shaped the extant metazoan Rab complement. The number of Rab members for each species is indicated in correspondence of the column (#); red numbers represent *Rab* genes considered to be LECA, according to the classification proposed by Elias et al. [22], plus the *Ran* gene and without *RabTitan* and *Rab50*. The colour of dots indicates the gene presence (black), gene absence (white), invertebrate lineage-specific duplications (orange), vertebrate whole-genome duplicates (green), vertebrate-specific gene duplicates (blue), reptile-specific duplicates (brown), mammalian-specific duplicates (magenta) and primate-specific duplicates (yellow)

Amphimedon queenslandica (Rab32/38: XP_003388475; RabX1: XP_011407131). Importantly, this implies that only four novel Rab subfamilies appeared during animal evolution: Rab40 and RabX6 in bilaterians, Rab46 in deuterostomes, and Rab12 in chordates (Figure S1). In addition, our phylogeny provided good support for grouping particular subfamilies and, thereby, for defining some Rab clusters: Rab1–Rab35–Rab40 (aLRT = 0.92/aBayes = 1.00); Rab2–Rab4–Rab14 (0.99/1.00); Rab3–Rab27 (0.99/1.00); Rab5/17–Rab22 (0.78/0.94); Rab7–Rab9 (0.97/1.00); Rab7L1–Rab32/38–Rab32LO (0.98/1.00); Rab8–Rab10 (0.78/0.94); Rab19/43–Rab30–Rab33 (0.97/1.00); Rab20–Rab24–Rab21–RabX1 (0.80/0.75); and Rab44/EFcab4−Rasef−Rab46 (1.00/1.00) (Figs. 1; S1).

Before this study, the 42 metazoan Rab subfamilies had been classified into six supergroups (I–VI) corresponding to distinct routes of membrane trafficking [18, 19], but the monophyly of some of them was, however, poorly supported [18] or even not recovered [24, 32]. Our phylogenetic reconstruction sustained only some of the previously identified supergroups. Hence, whereas monophyletic supergroups II, III, IV and V were supported by our analysis (aLRT = 0.90, 0.81, 0.96 and 0.85, aBayes = 1.00, 0.99, 1.00, 0.98, respectively), supergroup VI was poorly sustained, and the monophyly of supergroup I was broken (Fig. 1). We proposed, therefore, to reclassify the Rab family into three major clades, named A, B and C (highlighted in Fig. 1 with green, red and blue colours, respectively). Clade A (aLRT = 0.84/aBayes = 0.92) encompassed the Rab subfamilies of the former supergroups I (except Rab34/36 subfamily) and IV; clade B (aLRT = 0.69/aBayes = 0.94) grouped all Rabs belonging to the former supergroup III and supergroup VI and the intraflagellar transporter Ift27 [33], plus the RabL2 (previously called RTW, [22]) and Ran subfamilies, the latter previously considered a distinct family of the small GTPase superfamily [32] and utilized as outgroup for Rab phylogenies [24]; and clade C (aLRT = 0.98/aBayes = 1.00) comprised former supergroups II and V, plus the Rab34/36 subfamily previously included in supergroup I.

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

Finally, our phylogenetic analysis highlighted the strong impact of gene duplications and losses in modelling *Rab* family dimension and complexity in both protostome and deuterostome repertoires (Fig. 2), with a major relevance in chordates: 29 out of the 42 subfamilies retain duplicates in one or more analysed species, and at least 57 independent gene losses might be deduced from the comparison of the *Rab* repertoire of the selected organisms. Interestingly, duplication and loss patterns revealed some notable biases. Duplicates, for instance, appeared to be more abundant in certain subfamilies (e.g. in Rab3, Rab5 and Rab11 subfamilies) or lineages (e.g. in vertebrates) than in others, while gene losses have been particularly frequent in fast-evolving species such as *O. dioica* and *C. elegans* (Fig. 2).

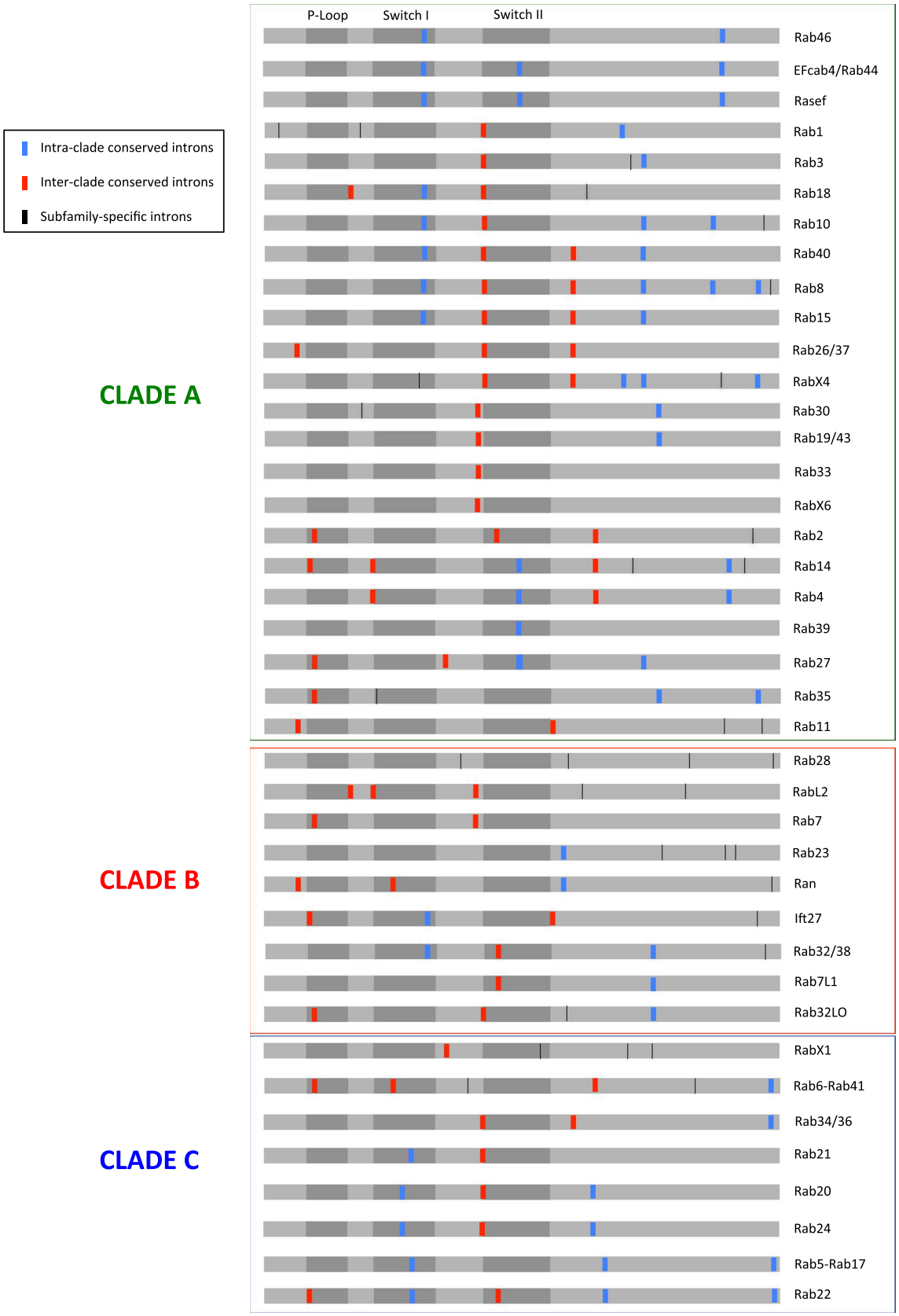
Rab evolution in chordates

The present work improved the knowledge regarding the evolutionary history of *Rab* family in metazoans compared to previous studies [18, 24, 34], identifying for the first time the Rab repertoire in five new organisms: the annelid *C. teleta*, the hemichordate *S. kowalevskii*, the cephalochordate *B. lanceolatum*, the tunicate *O. dioica* and the vertebrate *A. carolinensis* (Figs. 1, 2). We focused our attention on the chordate repertoires because we were interested in the Rabs of the chordate ancestor and in its changes during the transition from non-vertebrate to vertebrate chordates (Fig. 2). Our survey revealed that the set of Rabs of the chordate ancestor was made of at least 41 out of the 42 metazoan Rab subfamilies, of which *Rab12* was a chordate gain and *RabX6* was a chordate loss occurred before the diversification of the phylum. *Rab12* is strongly expressed in rat Sertoli cells [35] and in migrating neural crest cells [36], and regulates endosomes–lysosomes shift [37]; otherwise, RabX6 is present in neurons and in testis of male insect *Bombyx mori* [38]. With 41 Rab subfamilies, the current Rab toolkit of the cephalochordate amphioxus may be indeed very similar to the ancestral chordate one, corroborating the 'genomic stasis' attributed to cephalochordates [39–41] and supporting the idea that modern amphioxus resembles in many respects the ancestral chordate [42]. Amphioxus, however, also experimented gene expansions in three Rab subfamilies (Rab9, Rab11, Rab12), due to lineage-specific duplications during cephalochordate evolution (Figures S1; 2), as shown by their presence also in sibling species *B. floridae* (Rab9: XP_002599508 and XP_002600847; Rab11: XP_002605587 and XP_002602818; Rab12: XP_002588685 and XP_002588683).

In contrast with amphioxus stasis, tunicate *Rab* genes showed a dynamic evolution, in which high evolutionary rates and many gene gains and losses characterized their history. In the phylogenetic tree, tunicate Rab proteins had long branches (Figure S1) that, due to "long-branch attraction" artefacts, rarely clustered as the sister group of vertebrate Rabs within each subfamily as expected from their taxonomic relationships. Although some Rab subfamilies were expanded in tunicates (Rab11, Rab27 likely in the whole subphylum, Rab12 specifically in the ascidian *C. robusta*, and Rab5, Rab6, Rab7, Rab10 and Rab35 in the appendicularian *O. dioica*), tunicate Rab evolution was predominantly impacted by gene loss events. As a matter of fact, *C. robusta* and *O. dioica* shared the absence of 8 chordate Rab subfamilies, two of them, Rab32LO and RabX4, as Olfactores (tunicate + vertebrate) losses (Fig. 2). *O. dioica* lineage showed 11 additional losses, Rab4, Rab7L1, Rab9, Rab19/43, Rab21 Rab26/37, Rab28, Ift27, Rasef, EFcab4/Rab44 and RabX1. This species has, therefore, lost almost the 50% of the Rab toolkit, being the metazoan species with the smallest number of Rab subfamilies described so far.

The *Rab* repertoire in the vertebrate ancestor was also impacted by events of gene duplication and loss. The two rounds of whole-genome duplications at the root of vertebrate evolution [43, 44] twice duplicated the pre-vertebrate Rab complement (around 39 Rabs), and local duplication events might have further amplified the Rab catalogue in vertebrates. The ancestor of vertebrates might have had, therefore, more than 150 *Rab* genes, derived either from WGDs or local duplication events. To distinguish between these two types of duplicates, we examined the syntenic conservation (i.e. the tendency of neighbouring genes to retain their relative positions and orders on ohnologous chromosomes) in human chromosomes containing *Rab* genes using the Synteny Database [45]. Our work proved a strong impact of the WGDs in the rise of many vertebrate *Rab* genes since results showed ohnology for many duplicates (Figures S2; S3): 17 pairs, trios or quartets out of the 21 vertebrate subfamilies (81%) with more than one *Rab* gene appeared to have a WGD origin. The most parsimonious origin of the remaining non-ohnologous duplicates (Figure S2) is based on local duplications caused by either unequal recombination events or retrotranscription/retrotransposition processes along vertebrate evolution. Unequal recombination, for instance, was the most likely origin of tandem duplicates *Rab3A–Rab3D*, whereas the intron-less structure of *Rab9A–Rab9B* and in *Rab40A–Rab40AL* duplicates pointed to a retrotranscriptional origin. The presence of these non-ohnologous duplicates in both human and lizard species suggested an ancient origin for most of them during vertebrate evolution, likely before amniotes diversification. Further non-ohnologous pairs of duplicates would be *Rab18A−B*, *RabL2A−B* and *Rab40A−AL* pairs that seemed to be reptilian-, mammalian-, and primate-specific duplications, respectively, as suggested by their presence in other available genomes of these three groups of gnathostome vertebrates. Thus, we uncovered the presence of two *Rab18* protein-encoding genes in *A. carolinensis* and in the painted turtle

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019



Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

◂**Fig. 3** Rab intron code. Schematization of metazoan *Rab* gene structure showing the intron conservation code specific for each subfamily. Grey boxes represent the three canonical Rab domains: P-loop, Switch I and Switch II. The green, red and blue frames correspond to the three monophyletic clades resulted from the phylogenetic analysis. The intron/exon boundary of each Rab subfamily is shown by small vertical bars, in blue for intra-clade and in red for inter-clade conserved intron positions

Chrysemys picta bellii (Rab18a: ENSCPBP00000022122; Rab18b: ENSCPBP00000040164), two *RabL2* in human, goat *Capra hircus* (ENSCHIP00000000800 and ENSCHIP00000022194) and pig *Sus scrofa* (ENSSSCP00000055882 and ENSSSCP00000041172), and the duplet *Rab40A-AL* on chromosome X in human, orangutan *Pongo abelii* (ENSPPYP00000023039 and ENSPPYP00000023028) and crab-eating macaque *Macaca fascicularis* (ENSMFAP00000015171 and ENSMFAP00000011932). We did not find, however, these duplicates in related clades (i.e. *Rab18A−B* duplets in birds or mammals, *RabL2A−B* duplets in birds or reptiles, and *Rab40A-AL* duplets in other mammals).

Frequent events of gene loss have been occurred during *Rab* evolution in vertebrates (Fig. 2). Thus, from the twice-duplicated repertoire of the ancestral vertebrate Rabs, 1 copy was lost in at least 4 Rab subfamilies, 2 copies in 10 subfamilies, and 3 copies in 16 subfamilies, and there were no vertebrate Rab members for 4 subfamilies (excluding the RabX6 subfamily, lost in the chordate ancestor). In summary, at least 34 out of the 41 Rab subfamilies (83%) have experienced the loss of some (or all) members during vertebrate evolution.

Rab intron code

Conserved intron positions have been shown to support orthologous relationships in gene families providing value information about their evolutionary history [46, 47]. We compared the intron positions of the *Rab* genes, demonstrating that each subfamily exhibits a specific intron code retained in most of the analysed metazoan genes (Figs. 3; S4), as it had already been proposed for the Rab32/38 subfamily [48]. These intron codes were, therefore, useful for classifying *Rab* genes in distinct subfamilies, or for resolving orthologous relationships of highly divergent families. For instance, the fact that mollusc and ambulacrarian candidates of the fast-evolving *RabX6* subfamily shared one of such subfamily-specific introns (Figure S4) reinforced the orthologous relationship of these genes.

Besides subfamily conservation, we classified introns into two categories depending on their inter-subfamily conservation pattern: (1) inter-clade introns, which are intron positions shared by two or more genes in at least two *Rab* subfamilies of different clades (red lines in Fig. 3); (2) intra-clade introns, which were intron positions shared by two or more genes in at least two Rab subfamilies of the same clade (blue lines in Fig. 3). Our rationale was that the inter-clade introns could be informative about the ancestral pre-metazoan *Rab* gene structure, while intra-clade introns could provide support to the clades or to other levels of subfamily clustering. Our findings showed a remarkable abundance of inter-clade introns (14 conserved positions in Fig. 3), which suggested that the ancestral *Rab* gene had many introns unevenly retained and lost along *Rab* evolution. We found 7, 2 and 6 intra-clade introns in clades A, B and C, respectively, some of them supporting the phylogenetic grouping of some subfamilies. The cluster of Rab44/EFcab4−Rasef−Rab46 subfamilies, for instance, was supported by the fact that they share 3 out of the 7 clade A-specific introns (i.e. introns 1, 2 and 6, blue lines numbered from the left to the right in Fig. 3); Rab8–Rab10 group was supported by clade A introns 1, 3 and 5; Rab4–Rab14 group, by clade A introns 2 and 6; Rab19/43–Rab30 group, by clade A intron 4; Rab7L1−Rab32/38−Rab32LO group, by clade B-specific intron 2; Rab20–Rab24 group, by clade C-specific introns 1 and 3; Rab6–Rab34/36 group, by clade C-specific intron 5; and Rab5/17–Rab22 group, by clade C-specific introns 2, 4 and 6 (Fig. 3).

Evolution of Rab domain architecture

To gain more insights regarding Rab evolution, we investigated the changes in their domain architecture (Fig. 4). The majority of Rabs have the same three-domain organization that we define as the "canonical" organization (Fig. 4a): a P-Loop (from amino acid 18 to 25, referred to *H. sapiens* RAB1A), which is a nucleotide-binding motif fundamental for GTP/GDP cycling, and Switch I (from aa 36 to 47) and Switch II domains (from aa 66 to 78), necessary for the correct protein folding [6]. We found four exceptions to this canonical organization that implied the insertion of additional motifs in Rab structure. First, Rab32/38, Rab32LO and Rab7L1 subfamilies of clade B were characterized by an ultra-conserved amino acid stretch downstream Switch I domain (FALK, from aa 62 to 65, referred to *H. sapiens* RAB32) (Fig. 4b) with a possible role in protein folding linked to Switch I activity [48]. Second, Ran proteins, previously considered as an independent family of nuclear transporter [32, 49], showed a distinctive protein sequence (Fig. 4c) that is ultra-conserved in all the eukaryotes [50]. Third, Rab40 proteins in clade A exhibited an additional SOCS box at the C-terminal region (from aa 175 to 228, referred to *H. sapiens* RAB40A) (Fig. 4e) that is considered fundamental for lipid droplets biogenesis in *D. melanogaster* [51] and for Varp proteasomal degradation in mammalian melanocytes [52]. And fourth, the most striking motif

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

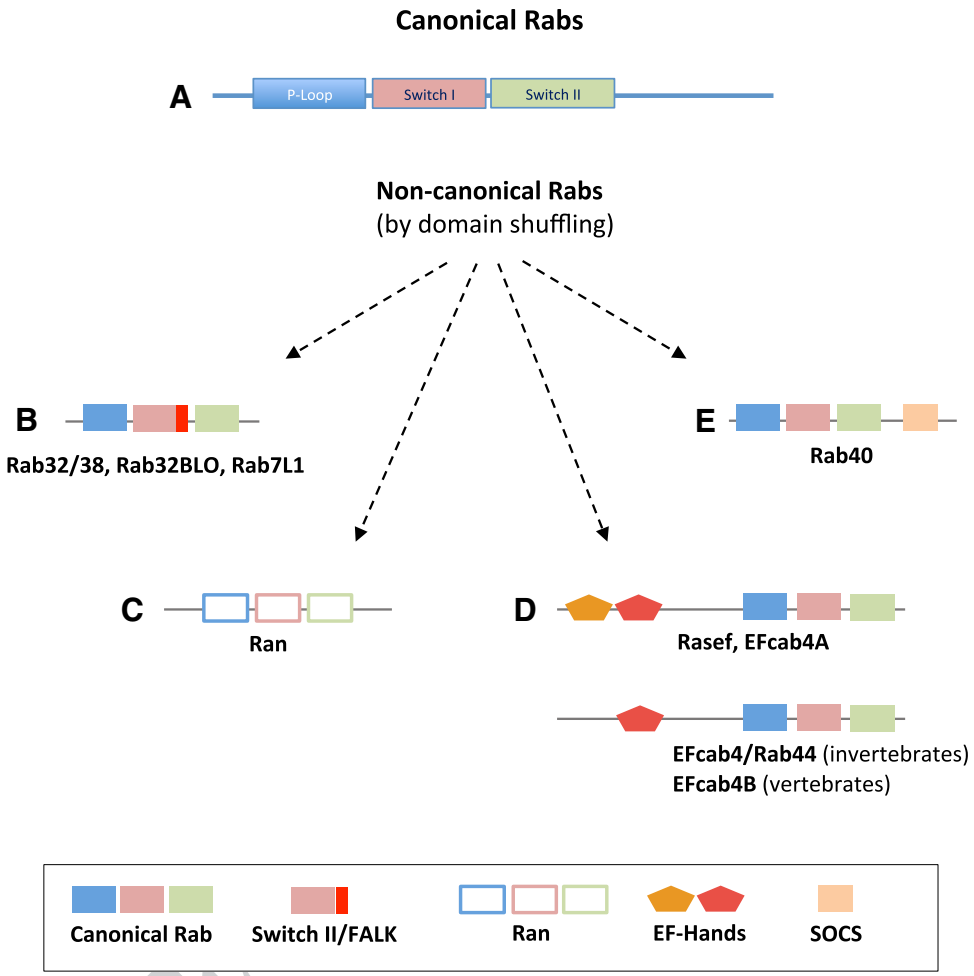


Fig. 4 Modular domain organization. Distinct Rab domain architectures found in metazoans. **a** Most of Rabs contain "canonical" Rab domains made of P-Loop (bluebox), Switch I (pink box) and Switch II (green box). **b** The FALK stretch (in red) is exclusive of Rab32, Rab38, Rab32LO and Rab7L1. **c** Ran proteins possess a divergent amino acid composition in P-Loop, Switch I and Switch II domains (empty boxes). **d** One or two EF-Hand domains at N-terminus (orange and red pentagons) characterize the chimeras Rab (Rasef, EFcab4/Rab44, EFcab4, and Rab44). **e** The unique case of Rab40 protein, which contains an SOCS box at the C terminus (light orange box)

novelty during the evolution of Rab domain architecture was the rise of what we named Rab chimeras in clade A.

Rab chimeras seem to be the result of the fusion of a canonical Rab at C-terminus with one or two calcium-binding EF-hand motifs [53] at N-terminus (from aa 8 to 40, and from aa 42 to 77, respectively, referred to *H. sapiens* RASEF; Fig. 4d). In phylogenetic analyses, Rab chimeras were grouped into two Rab subfamilies as sister clades, named Rasef and EFcab4/Rab44 (Fig. 1), which previously were known as Rab45 and Cracr2, respectively [54, 55]. The scarcity of information about Rasef and EFcab4/Rab44 chimeras led us to further investigate their evolutionary origins. A thorough genome search in many eukaryotic genomes, including eight unicellular species (*Monosiga brevicollis, Capsaspora owczarzaki, Saccharomyces cerevisiae, Dictyostelium discoideum, Trichomonas vaginalis, Chlamydomonas reinhardtii, Trypanosoma cruzi, Arabidopsis thaliana*), led us to propose that chimeric Rabs were an animal innovation. The pervasive presence of Rasef and EFcab4/Rab44 subfamilies in animals, from sponges (*A. queenslandica* Rasef: XP_019848798; EFcab4/Rab44: XP_011403598) to humans, dated their origin back to the LMCA. In addition, our analysis revealed a new subfamily of Rab chimeras in tunicate, cephalochordate and echinoderm genomes that we named Rab46 (Fig. 1). The absence of Rab46 sequences in cnidarian and protostome genomes indicated this subfamily as a deuterostome innovation independently lost in vertebrate and hemichordate lineages.

To further understand the evolution of Rasef and EFcab4/Rab44 chimeras in metazoans, it has been generated a manually curated database of 29 proteins that included the entire sequence of Rab chimeras, comprising the C-terminal Rab domains and the N-terminal EF-hand motifs (Supplementary file 4), and carried out a dedicated phylogenetic analysis including additional vertebrate and invertebrate species (Fig. 5a; Rab46 sequences were excluded from this phylogeny because of their high sequence divergence and limited representation). The tree topology showed that members of the Rasef subfamily (Fig. 5a, orange background) have been maintained as single-copy genes in all metazoan lineages analysed, whereas the members of the EFcab4/Rab44 subfamily (Fig. 5a, blue background, including EFcab4A, EFcab4B and Rab44 sequences) were duplicated from an ancestral *EFcab4/Rab44* gene during vertebrate evolution

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

(Fig. 5a). Syntenic conservation of human chromosomes 6, 11 and 12 suggested that the 3 paralogs, *EFcab4A*, *EFcab4B* and *Rab44*, derived from the WGD events of vertebrates [43, 44], and that the lost fourth paralog may have been located on human chromosome 1 (Figure S6).

Remarkably, the number of EF-hand motifs appeared to be diagnostic for discriminating amongst Rab-chimera members (Fig. 5b). Noteworthy, although mammalian and amphibian EFcab4A lacked a C-terminal Rab domain (Fig. 5b), their two EF-hand motifs together with the conservation of the genomic environment of mammal (*H. sapiens*), reptile (*A. carolinensis*), amphibian (*Xenopus tropicalis*) and cartilaginous fish (*Callorhinchus milii*) *EFcab4A* genes clearly supported their orthologous relationship (Fig. 5c), and suggested lineage-specific C-terminal domain losses. In light of these findings, the most parsimonious explanation is that the ancestor of Rab chimeras was constituted by canonical Rab motifs plus two EF-hands, which have been differentially lost during evolution (Fig. 5a–c).

Discussion

The evolutionary landscape of the Rab gene family in chordates

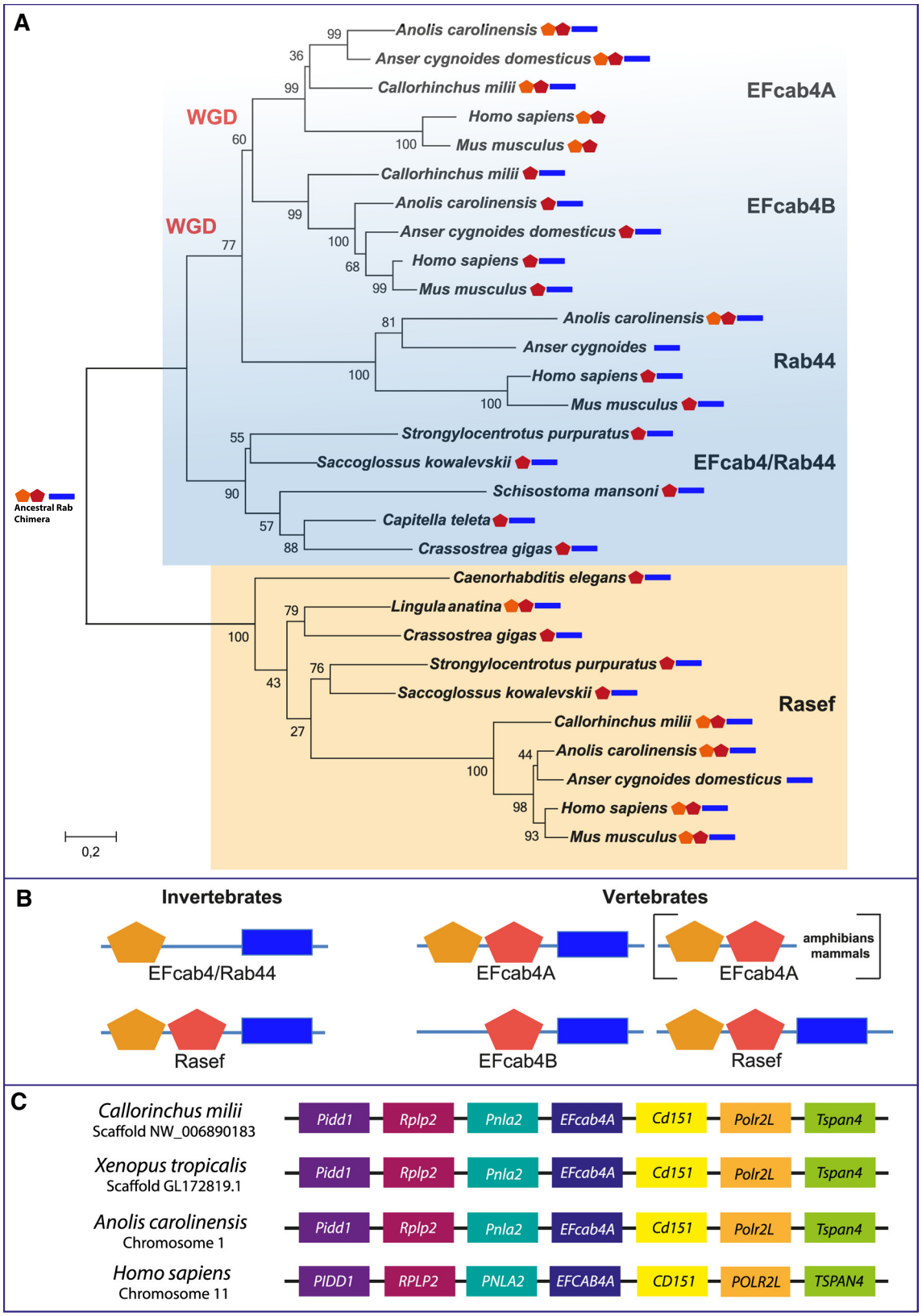
We have identified and classified 243 Rabs belonging to the three chordate subphyla, revealing distinct patterns in the evolution of Rab subfamilies in each of them (Figs. 1, 2). The 41 Rab subfamilies of the cephalochordate *B. lanceolatum* would likely represent the prototypical chordate Rab repertoire that has been preserved in the 'conservative' amphioxus genome. With respect to the metazoan complete toolkit, the ancestral chordate (and thereby, the amphioxus lineage) would have lost a unique gene, the *RabX6*, which is expressed in neurons and in testis of male insects [38]. The functional causes that may justify the retention of *RabX6* in some protostome and non-chordate deuterostome lineages (i.e. *Ambulacraria*), but its loss in chordates, remain to be elucidated.

In sharp contrast to the conservative scenario of amphioxus, tunicates (i.e. ascidian *C. robusta* and larvacean *O. dioica*) exhibit a liberal pattern of evolution with many gene losses (up to 20) and duplications (up to 8). Especially remarkable is the case of *O. dioica* because with the loss of almost half (9) of the Rab subfamilies, this free-living (non-parasitic) animal is the metazoan species with the smallest number of subfamilies described so far. The tendency of *O. dioica* genome to lose genes and gene families, thought to be fundamental for key biological processes, is notorious [56, 57] and well documented [57, 58]. The loss of so many *Rab* genes in *O. dioica* implies that these became dispensable during species evolution due to either situation of mutational robustness or of environment-dependent conditional dispensability (reviewed in [56]). It can be argued, for instance, that the loss of *Ift27* genes important for cilia/flagella trafficking took place under a mutational robustness situation (e.g. functional compensation by other *Rabs*) because *O. dioica* has operative cilia and flagella [59, 60] and, therefore, that other Rabs (e.g. Rab8, Rab23) might compensate the loss by function shuffling [41]. This phenomenon may be, indeed, usual in *O. dioica*, and it would account for the preservation of the *Rab32/38*, typically involved in melanosome biogenesis, in a species lacking pigmented cells. In addition, because Rabs cooperate with many interacting effector proteins [4, 28], the identification of co-elimination patterns of the different Rabs in *O. dioica* may be a useful strategy for recognizing the Rab-associated machinery in other species. Regarding duplications, both surveyed tunicates have expanded *Rab11/25* and *Rab27* subfamilies, leading to suppose their origin predated the radiation of the subphylum. On the other hand, we detected a single *C. robusta*-specific duplication event (*Rab12*) and 5 independent duplications in *O. dioica* (*Rab5/17*, *Rab6*, *Rab7*, *Rab10*, *Rab35*).

In gnathostomes, our findings reveal that the two rounds of WGD (1R, 2R) impacted on the *Rab* repertoire (i.e. 40 *Rab* onhologs), which nowadays is the largest in metazoans (Figures S1, 2). It would be interesting in the future to analyse the Rab toolkit of fish species such as zebrafish, medaka or salmon, to evaluate the impact on Rab family of additional genome duplications (3R and Ss4R) that have involved teleost lineage [61–63]. Moreover, our results shed light on a number of local duplication events, some of them likely ancestral (*Rab3*, *Rab9*, shared by reptiles and mammals), and some affecting only some lineages, like *Rab18* in reptiles, *RabL2* in mammals, and *Rab40A/AL* in primates (Figure S2).

Interestingly, many duplicates originated by the WGDs were, however, lost during vertebrate evolution. The surveyed vertebrate species have retained *Rab* duplicates for only 21 subfamilies, have returned to singletons in 16 subfamilies, and have totally lost 2 subfamilies, Rab46 and RabX1, while the lizard *A. carolinensis* has additionally lost the Rab34/36 subfamily. Because preservation or loss of duplicates appear to depend on the duplication mode (genome versus local gene duplication) (reviewed in [64]), it can be argued that the mode of duplication (WGD) rather than the duplication itself was key to facilitate the *Rab* expansion in vertebrates and the subsequent increase in the functions of *Rabs* in this lineage. Although the functional implications of the *Rab* gains and losses need to be investigated, one can predict neofunctionalization or subfunctionalization processes amongst duplicated subfamilies or functional changes associated with the absence of some subfamilies. For example, the loss of *RabX1* could be related to modifications in the localization of E-cadherins

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019



Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

◂**Fig. 5** Evolutionary history of Rab chimeras. **a** Phylogenetic three of Rab chimeras. *EFcab4/Rab44* already present in invertebrates underwent two rounds of genome duplication in vertebrates generating *Rab44*, *EFcab4A* and *Efcab4B*. On the other hand, *Rasef* is a single-copy gene in both invertebrates and gnathostomes; Rab46 was excluded for their sequence divergence. First and second EF-Hand domains (orange and red pentagons, respectively) and Rab motifs (blue bar) are depicted. On the left, the ancestral organization of Rab chimeras is shown. Numbers at the branches are replicates obtained employing the ML estimation method. **b** Domain organization of invertebrate EFcab4/Rab44 and gnathostome EFcab4A and EFcab4B. EFcab4A in amphibians and mammals has lost the canonical Rab domain. Rab44 was excluded here for their variability in domain organization. **c** The orthology of gnathostome *EFcab4A* genes is demonstrated by the conserved neighbourhood on respective chromosomes

to the *zonula adherens* in vertebrates [65], while the loss of *Rab34/36* genes in lizard might be associated with changes in their late-endosomal and lysosomal trafficking machinery [66]. In summary, our results demonstrate that chordates represent a hotspot of Rab variability and highlight that the comprehension of the evolutionary history of the *Rab* gene family paves the way for future functional analyses. These analyses will be relevant not only for basic research in intracellular trafficking, but also for biomedical applications due to the numerous pathologies correlated to dysfunctions in organellar organization and transport [67].

Evolution of Rab chimeras

Amongst unconventional Rabs in terms of domain architecture (Fig. 4), Rab chimeras emerged as a set of poorly characterized Rabs of unusual length and domain composition with the capability to bind calcium ions through EF-hands. The survey in several key eukaryotic genomes, including unicellular and multicellular species, suggested the absence of such chimeric genes in unicellular eukaryotes and plants, but a pervasive presence in animals (Fig. 5). Phylogenetic analyses based on either the C-terminal Rab domain (Fig. 1), or including the N-terminal EF-hand motifs (Fig. 5a), classified Rab chimeras into three sister subfamilies: Rasef, EFcab4/Rab44 and Rab46. *Rasef* has been maintained as single-gene subfamily in all metazoans, from cnidarians to human, with the exception of *O. dioica*, where it was lost. EFcab4/Rab44 subfamily, also lost in *O. dioica,* was duplicated in vertebrates as result of the whole-genome duplications (Figure S6). In contrast, Rab46 subfamily likely arose in the stem of deuterostomes, but it was lost during the transition from non-vertebrate to vertebrate chordates.

Interestingly, the number of EF-hand motifs appeared to be variable amongst Rab-chimera members (Fig. 5b). Vertebrate and invertebrate Rasef as well as vertebrate EFcab4A and Rab44 had two EF-hand motifs, whereas invertebrate EFcab4/Rab44 and vertebrate EFcab4B had just one (Fig. 5b). In all the cases, the EF-hand that has been lost was the first (orange in Figs. 4 and 5). Mammalian and amphibian EFcab4A lacked a C-terminal Rab domain (Fig. 5b), but the sequence conservation of their two EF-hand motifs together with the preservation of the genomic environment of *EFcab4A* genes in genomes of distantly related vertebrates clearly supported their orthologous relationship (Fig. 5c) and suggested lineage-specific C-terminal domain losses. Overall, our results showed that Rab chimeras constitute a peculiar class of Rabs emerged in animals by the fusion of a canonical Rab with two EF-Hand motifs followed by a gene duplication. The implication of Rab chimeras in diverse human diseases, such as lung carcinoma [68] and melanoma [69, 70] and the scarcity of knowledge about this protein class, encourages further investigations on their cellular role and expression patterns in animals, with possible consequences related to frequent domain losses.

Conclusions

The Rabs of 11 metazoan species have been explored, with a focus on chordate species, under the hypothesis that Rabs might have been instrumental for the increasing complexity of intracellular traffic mechanisms in animals. We classified Rabs into 42 robust metazoan subfamilies with specific intron codes, and grouped them into 3 distinct Rab clades rather than the 5 or 6 supergroups proposed in previous studies. The analysis of the chordate Rab toolkit highlighted dramatic differences in the evolutionary patterns in the three chordate subphyla—conservative in cephalochordates, liberal in tunicates, and expansive in vertebrates—which most likely had a strong impact on their intracellular communication machineries and provide the first comprehensive evolutionary analysis of the Rab chimeras as an animal novelty.

Collectively, our results set the grounds for future investigations on comparative analyses deputed to Rab functions in vertebrates, and an additional step to understand their involvement in human diseases.

Materials and methods

Genome database searches and phylogenetic reconstructions

Protein sequences of the Rab repertoire from vertebrate *H. sapiens* were used as queries in BLASTp and tBLASTn searches in NCBI or Ensembl genome databases of selected species. Orthologies of the Rab members were initially assessed by reciprocal best blast hit (RBBH) approach employing default parameters and corroborated by phylogenetic analyses. Phylogenetic reconstructions were based

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

on maximum-likelihood inferences calculated with PhyML v3.0 using automatic Akaike Information Criterion (AIC) for selection of the substitution model [71], which selected the LG + G+I model with discrete gamma distribution in four categories. All parameters (gamma shape = 0.846; proportion of invariants = 0.004) were estimated from the dataset. Protein alignments were generated with MUSCLE [72] and ClustalX [73] programs and reviewed by hand to exclude too short or divergent Rab sequences. Only the conserved parts of the proteins whose alignments appeared unambiguous were considered for the phylogenetic analysis, i.e. from codon D9 to G177 of human RAB1A.

Branch support was provided by aLRT [74] and aBayes methods [75]. Phylogenetic reconstruction of Fig. 5 was based on branch supports obtained using ML estimation method with a WAG matrix ($\gamma = 4$). Accession numbers and protein alignment for phylogenetic tree reconstructions of Fig. 1 are provided in Supplementary file 1, while those employed for phylogeny of Fig. 5 are listed in Supplementary file 4 (the entire sequences of chimeric Rabs were utilized).

Branchiostoma lanceolatum Rabs

Branchiostoma lanceolatum Rab genes were annotated in the genome draft version Bl71nemr, kindly provided by the "*Branchiostoma lanceolatum* Genome Consortium" [76].

Analysis of intron/exon structures and phases

Gene structures were deduced after merging the genomic sequences with ESTs when available, as previously described [46, 47]. Introns were classified as phase 0, phase 1, and phase 2 depending on their positions relative to the protein-reading frame. For all Rab subfamilies, we showed a schematic representation of 200 amino acid residues taking as reference the human RAB1A from aa 1 to 200, and manually mapped the conserved introns.

Synteny conservation

We evaluated the presence or absence of synteny conservation using the Syntenic Database developed by [46, 47]. Synteny Database is an automatic tool that provides gene clusters using several different sliding window sizes measured in terms of contiguous gene number. Smaller window sizes identify tightly conserved syntenic regions, while larger window sizes can accommodate chromosomal rearrangements. We used a sliding window size of 100 (default) or 200 genes. Synteny Database allowed us to perform genomic comparisons between the human genome and an outgroup genome that diverged prior to the two rounds of genome duplication (usually *C. robusta*) and visualize the regions of conserved synteny within the source genome, i.e. human paralogons.

Acknowledgements The authors are grateful to the *Branchiostoma lanceolatum* Genome Consortium that provided access to the European amphioxus genome [76]. We would like to thank Dr. Eva Jimenez-Guri for her critical reading of the manuscript, and three anonymous reviewers for their comments that helped to improve the manuscript. R.A. was supported by BIO2015-67358-C2-1-P grant from Ministerio de Economía y Competitividad (Spain) and by Grant SGR2017-1665 from Generalitat de Catalunya. U.C. was supported by a OU-SZN PhD fellowship.

References

1. Szathmary E, Smith JM (1995) The major evolutionary transitions. Nature 374(6519):227–232
2. Touchot N, Chardin P, Tavitian A (1987) Four additional members of the ras gene superfamily isolated by an oligonucleotide strategy: molecular cloning of YPT-related cDNAs from a rat brain library. Proc Natl Acad Sci USA 84(23):8210–8214
3. Diekmann Y, Seixas E, Gouw M, Tavares-Cadete F, Seabra MC, Pereira-Leal JB (2011) Thousands of rab GTPases for the cell biologist. PLoS Comput Biol 7(10):e1002217
4. Stenmark H (2009) Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol 10(8):513–525
5. Lee MT, Mishra A, Lambright DG (2009) Structural mechanisms for regulation of membrane traffic by rab GTPases. Traffic 10(10):1377–1389
6. Park HH (2013) Structural basis of membrane trafficking by Rab family small G protein. Int J Mol Sci 14(5):8912–8923
7. Bonifacino JS, Glick BS (2004) The mechanisms of vesicle budding and fusion. Cell 116(2):153–166
8. Sztul E, Lupashin V (2009) Role of vesicle tethering factors in the ER-Golgi membrane traffic. FEBS Lett 583(23):3770–3783
9. Goud B, Liu S, Storrie B (2018) Rab proteins as major determinants of the Golgi complex structure. Small GTPases 9(1–2):66–75
10. Barr F, Lambright DG (2010) Rab GEFs and GAPs. Curr Opin Cell Biol 22(4):461–470
11. Alexandrov K, Horiuchi H, Steele-Mortimer O, Seabra MC, Zerial M (1994) Rab escort protein-1 is a multifunctional protein that accompanies newly prenylated rab proteins to their target membranes. EMBO J 13(22):5262–5273
12. Shi CH, Zhang SY, Yang ZH, Yang J, Shang DD, Mao CY, Liu H, Hou HM, Shi MM, Wu J et al (2016) A novel RAB39B gene mutation in X-linked juvenile parkinsonism with basal ganglia calcification. Mov Disord 31(12):1905–1909
13. Banworth MJ, Li G (2018) Consequences of Rab GTPase dysfunction in genetic or acquired human diseases. Small GTPases 9(1–2):158–181
14. Nishimura N, Van Huyen Pham T, Hartomo TB, Lee MJ, Hasegawa D, Takeda H, Kawasaki K, Kosaka Y, Yamamoto T, Morikawa S et al (2011) Rab15 expression correlates with retinoic acid-induced differentiation of neuroblastoma cells. Oncol Rep 26(1):145–151
15. Pham TV, Hartomo TB, Lee MJ, Hasegawa D, Ishida T, Kawasaki K, Kosaka Y, Yamamoto T, Morikawa S, Yamamoto N et al (2012) Rab15 alternative splicing is altered in spheres of neuroblastoma cells. Oncol Rep 27(6):2045–2049
16. Hendrix A, De Wever O (2013) Rab27 GTPases distribute extracellular nanomaps for invasive growth and metastasis: implications for prognosis and treatment. Int J Mol Sci 14(5):9883–9892

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

17. Qin X, Wang J, Wang X, Liu F, Jiang B, Zhang Y (2017) Targeting Rabs as a novel therapeutic strategy for cancer therapy. Drug Discov Today 22(8):1139–1147
18. Klopper TH, Kienle N, Fasshauer D, Munro S (2012) Untangling the evolution of Rab G proteins: implications of a comprehensive genomic analysis. BMC Biol 10:71
19. Stenmark H (2012) The Rabs: a family at the root of metazoan evolution. BMC Biol 10:68
20. Dunst S, Kazimiers T, von Zadow F, Jambor H, Sagner A, Brankatschk B, Mahmoud A, Spannl S, Tomancak P, Eaton S et al (2015) Endogenously tagged rab proteins: a resource to study membrane trafficking in *Drosophila*. Dev Cell 33(3):351–365
21. Cavalier-Smith T (2009) Predation and eukaryote cell origins: a coevolutionary perspective. Int J Biochem Cell Biol 41(2):307–322
22. Elias M, Brighouse A, Gabernet-Castello C, Field MC, Dacks JB (2012) Sculpting the endomembrane system in deep time: high resolution phylogenetics of Rab GTPases. J Cell Sci 125(Pt 10):2500–2508
23. Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL, Alsmark UC, Besteiro S et al (2007) Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. Science 315(5809):207–212
24. Saito-Nakano Y, Nakahara T, Nakano K, Nozaki T, Numata O (2010) Marked amplification and diversification of products of ras genes from rat brain, Rab GTPases, in the ciliates *Tetrahymena thermophila* and *Paramecium tetraurelia*. J Eukaryot Microbiol 57(5):389–399
25. Lal K, Field MC, Carlton JM, Warwicker J, Hirt RP (2005) Identification of a very large Rab GTPase family in the parasitic protozoan *Trichomonas vaginalis*. Mol Biochem Parasitol 143(2):226–235
26. Pereira-Leal JB, Seabra MC (2001) Evolution of the Rab family of small GTP-binding proteins. J Mol Biol 313(4):889–901
27. Rutherford S, Moore I (2002) The *Arabidopsis* Rab GTPase family: another enigma variation. Curr Opin Plant Biol 5(6):518–528
28. Brighouse A, Dacks JB, Field MC (2010) Rab protein evolution and the history of the eukaryotic endomembrane system. Cell Mol Life Sci 67(20):3449–3465
29. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK et al (2008) The amphioxus genome and the evolution of the chordate karyotype. Nature 453(7198):1064–1071
30. Pennati R, Ficetola GF, Brunetti R, Caicci F, Gasparini F, Griggio F, Sato A, Stach T, Kaul-Strehlow S, Gissi C et al (2015) Morphological differences between larvae of the *Ciona intestinalis* species complex: hints for a valid taxonomic definition of distinct species. PLoS One 10(5):e0122879
31. Berna L, Alvarez-Valin F (2014) Evolutionary genomics of fast evolving tunicates. Genome Biol Evol 6(7):1724–1738
32. Rojas AM, Fuentes G, Rausell A, Valencia A (2012) The Ras protein superfamily: evolutionary tree and role of conserved amino acids. J Cell Biol 196(2):189–201
33. Huet D, Blisnick T, Perrot S, Bastin P (2014) The GTPase IFT27 is involved in both anterograde and retrograde intraflagellar transport. Elife 3:e02419
34. Gallegos ME, Balakrishnan S, Chandramouli P, Arora S, Azameera A, Babushekar A, Bargoma E, Bokhari A, Chava SK, Das P et al (2012) The *C. elegans* rab family: identification, classification and toolkit construction. PLoS One 7(11):e49387
35. Iida H, Noda M, Kaneko T, Doiguchi M, Mori T (2005) Identification of rab12 as a vesicle-associated small GTPase highly expressed in Sertoli cells of rat testis. Mol Reprod Dev 71(2):178–185
36. Piloto S, Schilling TF (2010) Ovo1 links Wnt signaling with N-cadherin localization during neural crest migration. Development 137(12):1981–1990
37. Matsui T, Fukuda M (2011) Small GTPase Rab12 regulates transferrin receptor degradation: implications for a novel membrane trafficking pathway from recycling endosomes to lysosomes. Cell Logist 1(4):155–158
38. Uno T, Ozakiya Y, Furutani M, Sakamoto K, Uno Y, Kajiwara H, Kanamaru K, Mizoguchi A (2018) Functional characterization of insect-specific RabX6 of *Bombyx mori*. Histochem Cell Biol 151:187–198
39. Canestro C, Albalat R (2012) Transposon diversity is higher in amphioxus than in vertebrates: functional and evolutionary inferences. Brief Funct Genom 11(2):131–141
40. Paps J, Holland PW, Shimeld SM (2012) A genome-wide view of transcription factor gene diversity in chordate evolution: less gene loss in amphioxus? Brief Funct Genom 11(2):177–186
41. Somorjai IML, Martí-Solans J, Diaz-Gracia M, Nishida H, Imai KS, Escriva H, Cañestro C, Albalat R (2018) Wnt evolution and function shuffling in liberal and conservative chordate genomes. Genome Biol 19(1):98
42. Holland LZ, Albalat R, Azumi K, Benito-Gutierrez E, Blow MJ, Bronner-Fraser M, Brunet F, Butts T, Candiani S, Dishaw LJ et al (2008) The amphioxus genome illuminates vertebrate origins and cephalochordate biology. Genome Res 18(7):1100–1111
43. Abi-Rached L, Gilles A, Shiina T, Pontarotti P, Inoko H (2002) Evidence of en bloc duplication in vertebrate genomes. Nat Genet 31(1):100–105
44. Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. PLoS Biol 3(10):e314
45. Catchen JM, Conery JS, Postlethwait JH (2009) Automated identification of conserved synteny after whole-genome duplication. Genome Res 19(8):1497–1505
46. D'Aniello S, Irimia M, Maeso I, Pascual-Anaya J, Jimenez-Delgado S, Bertrand S, Garcia-Fernandez J (2008) Gene expansion and retention leads to a diverse tyrosine kinase superfamily in amphioxus. Mol Biol Evol 25(9):1841–1854
47. Irimia M, Roy SW (2008) Spliceosomal introns as tools for genomic and evolutionary analysis. Nucleic Acids Res 36(5):1703–1712
48. Coppola U, Annona G, D'Aniello S, Ristoratore F (2016) Rab32 and Rab38 genes in chordate pigmentation: an evolutionary perspective. BMC Evol Biol 16:26
49. Melchior F, Paschal B, Evans J, Gerace L (1993) Inhibition of nuclear protein import by nonhydrolyzable analogues of GTP and identification of the small GTPase Ran/TC4 as an essential transport factor. J Cell Biol 123(6 Pt 2):1649–1659
50. Coppola U, Caccavale F, Scelzo M, Holland ND, Ristoratore F, D'Aniello S (2018) Ran GTPase, an eukaryotic gene novelty, is involved in amphioxus mitosis. PLoS One 13(10):e0196930
51. Tan R, Wang W, Wang S, Wang Z, Sun L, He W, Fan R, Zhou Y, Xu X, Hong W et al (2013) Small GTPase Rab40c associates with lipid droplets and modulates the biogenesis of lipid droplets. PLoS One 8(4):e63213
52. Yatsu A, Shimada H, Ohbayashi N, Fukuda M (2015) Rab40C is a novel Varp-binding protein that promotes proteasomal degradation of Varp in melanocytes. Biol Open 4(3):267–275
53. Nakayama S, Moncrief ND, Kretsinger RH (1992) Evolution of EF-hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories. J Mol Evol 34(5):416–448
54. Shintani M, Tada M, Kobayashi T, Kajiho H, Kontani K, Katada T (2007) Characterization of Rab45/RASEF containing EF-hand domain and a coiled-coil motif as a self-associating GTPase. Biochem Biophys Res Commun 357(3):661–667

Journal : Large 18 | Article No : 3103 | Pages : 14 | MS Code : CMLS-D-18-01792 | Dispatch : 12-4-2019

55. Srikanth S, Jung HJ, Kim KD, Souda P, Whitelegge J, Gwack Y (2010) A novel EF-hand protein, CRACR55A, is a cytosolic Ca2 + sensor that stabilizes CRAC channels in T cells. Nat Cell Biol 12(5):436–446
56. Albalat R, Canestro C (2016) Evolution by gene loss. Nat Rev Genet 17(7):379–391
57. Denoeud F, Henriet S, Mungpakdee S, Aury JM, Da Silva C, Brinkmann H, Mikhaleva J, Olsen LC, Jubin C, Canestro C et al (2010) Plasticity of animal genome architecture unmasked by rapid evolution of a pelagic tunicate. Science 330(6009):1381–1385
58. Martí-Solans J, Belyaeva OV, Torres-Aguila NP, Kedishvili NY, Albalat R, Cañestro C (2016) Coelimination and survival in gene network evolution: dismantling the RA-signaling in a chordate. Mol Biol Evol 33(9):2401–2416
59. Flood PR, Afzelius BA (1978) The spermatozoon of *Oikopleura dioica* Fol (Larvacea, Tunicata). Cell Tissue Res 191(1):27–37
60. Onuma TA, Isobe M, Nishida H (2017) Internal and external morphology of adults of the appendicularian, *Oikopleura dioica*: an SEM study. Cell Tissue Res 367(2):213–227
61. Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59(2):190–203
62. Kuraku S, Meyer A (2009) The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. Int J Dev Biol 53(5–6):765–773
63. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A et al (2016) The Atlantic salmon genome provides insights into rediploidization. Nature 533(7602):200–205
64. Cañestro C, Albalat R, Irimia M, Garcia-Fernandez J (2013) Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. Semin Cell Dev Biol 24(2):83–94
65. Woichansky I, Beretta CA, Berns N, Riechmann V (2016) Three mechanisms control E-cadherin localization to the zonula adherens. Nat Commun 7:10834
66. Chen L, Hu J, Yun Y, Wang T (2010) Rab36 regulates the spatial distribution of late endosomes and lysosomes through a similar mechanism to Rab34. Mol Membr Biol 27(1):23–30
67. Olkkonen VM, Ikonen E (2006) When intracellular logistics fails–genetic defects in membrane trafficking. J Cell Sci 119(Pt 24):5031–5045
68. Oshita H, Nishino R, Takano A, Fujitomo T, Aragaki M, Kato T, Akiyama H, Tsuchiya E, Kohno N, Nakamura Y et al (2013) RASEF is a novel diagnostic biomarker and a therapeutic target for lung cancer. Mol Cancer Res 11(8):937–951
69. Kaplon J, Homig-Holzel C, Gao L, Meissl K, Verdegaal EM, van der Burg SH, van Doorn R, Peeper DS (2014) Near-genomewide RNAi screening for regulators of BRAF(V600E)-induced senescence identifies RASEF, a gene epigenetically silenced in melanoma. Pigment Cell Melanoma Res 27(4):640–652
70. Maat W, Beiboer SH, Jager MJ, Luyten GP, Gruis NA, van der Velden PA (2008) Epigenetic regulation identifies RASEF as a tumor-suppressor gene in uveal melanoma. Invest Ophthalmol Vis Sci 49(4):1291–1298
71. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59(3):307–321
72. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinform 5:113
73. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948
74. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55(4):539–552
75. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst Biol 60(5):685–699
76. Marletaz F, Firbas PN, Maeso I, Tena JJ, Bogdanovic O, Perry M, Wyatt CDR, de la Calle-Mustienes E, Bertrand S, Burguera D et al (2018) Amphioxus functional genomics and the origins of vertebrate gene regulation. Nature 564:64–70

Journal : **Large 18**	Article No : **3103**	Pages : **14**	MS Code : **CMLS-D-18-01792**	Dispatch : **12-4-2019**

Journal:	**18**
Article:	**3103**

Author Query Form

Please ensure you fill out your response to the queries raised below and return this form along with your corrections

Dear Author

During the process of typesetting your article, the following queries have arisen. Please check your typeset proof carefully against the queries listed below and mark the necessary changes either directly on the proof/online grid or in the 'Author's response' area provided below

Query	Details Required	Author's Response
AQ1	Kindly check and confirm the section heading. As per style conclusion should be at the end article but here it persents before materials and methods (introduction, materials and methods, results and conclusion).	
AQ2	There is a mismatch of ESM file captions between the manuscript and the ESM PDF files provided in the package. ESM Figure 1 caption is missing in the ESM PDF file provided in the package; therefore, we have processed it from the manuscript. All other figure captions are processed as given in the ESM PDF file.	