

UNIVERSITE DE MONTPELLIER II SCIENCES ET TECHNIQUES DU LANGUEDOC

THESE

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITE MONTPELLIER II

ECOLE DOCTORALE : SIBAGHE – SYSTEMES INTEGRES EN BIOLOGIE, AGRONOMIE,
GEOSCIENCES, HYDROSCIENCES ET ENVIRONNEMENT

DISCIPLINE : BIOLOGIE DES POPULATIONS & ECOLOGIE

**PHYLOGENIE MOLECULAIRE ET EVOLUTION DES TUNICIERS
(UROCHORDES) :
UNE APPROCHE PHYLOGENOMIQUE**

Présentée et soutenue publiquement
par

GEORGIA TSAGKOGEOGA

LE 7 DECEMBRE 2009

JURY:

M. David MOREIRA	Directeur de Recherche CNRS	UNIVERSITÉ PARIS XI	Rapporteur
M. Michaël MANUEL	Professeur	UNIVERSITÉ PARIS VI	Rapporteur
M. Patrick LEMAIRE	Directeur de Recherche CNRS	UNIVERSITÉ AIX-MARSEILLE II	Examineur
M. Xavier TURON	Directeur de Recherche CSIC	CENTRE D'ESTUDIS AVANÇATS DE BLANES	Examineur
M. Emmanuel DOUZERY	Professeur	UNIVERSITÉ MONTPELLIER II	CoDirecteur
M. Frédéric DELSUC	Chargé de Recherche CNRS	UNIVERSITÉ MONTPELLIER II	CoDirecteur

*« Once, we thought the relationships between all living things,
the branches of the tree of life,
were straightforward and simple.
Genetics is showing how wrong we were. »*

Henry Nicholls, journaliste “freelance”

Remerciements

*« Bonjour Mr. Douzery,
Je suis étudiante de Master 2 BEE, je viens de Grèce et je cherche un stage.
Je m'intéresse à la phylogénie. Auriez-vous un sujet à me proposer? ...»*

Mail, 18 Octobre 2005

Je voudrais commencer les remerciements en m'adressant à mon premier co-directeur de thèse, Emmanuel J.P. DOUZERY. Emmanuel, je te remercie d'avoir répondu à ce mail, qui constitua en effet un point de référence dans ma vie. Tu m'as ouvert la porte dans un domaine de recherche que j'ai toujours voulu découvrir. Je te suis très reconnaissante pour ceci et pour ton accueil tout au long de ces quatre années dans l'équipe de phylogénie moléculaire de l'ISEM.

Je dois tout particulièrement remercier mon deuxième co-directeur, principal encadrant, chef et tuteur, Frédéric DELSUC. Fred, ton travail rigoureux constitua sans doute un modèle à mes yeux. Je me rappelle que tu m'avais montré, pas à pas, comment reconstruire des arbres phylogénétiques à partir des données moléculaires. Près de quatre ans après, je suis contaminée par ton microbe : la passion de plonger dans des lots de données variées à la recherche d'un seul arbre, l'histoire évolutive la plus proche de la réalité. Je te remercie aussi de m'avoir appris comment se repérer dans la cascade des données moléculaires qui nous bombardent à l'heure actuelle, sans notamment se perdre dans la forêt des arbres qu'elles peuvent en produire. Je te suis particulièrement reconnaissante pour ta patiente immesurée, car au final c'est à toi que je dois la lisibilité de mes rapports, publications et de ce manuscrit de thèse.

Je tiens à remercier également les membres de mon jury, David MOREIRA, Michaël MANUEL, Patrick LEMAIRE et Xavier TURON, d'avoir accepté de consacrer une partie de leur temps précieux pour participer à l'évaluation de mes travaux de recherche.

Je souhaiterais adresser ma reconnaissance à Marie-Ka TILAK pour la confiance qu'elle m'a accordée dès mon arrivée au sein de l'équipe. Marie-Ka, tu as été et tu restes une chère amie, une deuxième mère pour moi. Un merci particulier à Andrew RODRIGUES pour m'avoir supporté durant un long voyage dans un si petit bureau - bateau de quelques mètres carrés. Je tiens ensuite à remercier Benjamin MARTEAUX, Cédric DELSOL et Suzanne JIQUEL pour les nombreuses pauses de « nicotine » partagées au quotidien. Je souhaiterais également remercier toute l'équipe de phylogénie moléculaire pour la bonne ambiance qui y règne, avec

notamment un continuum d'échanges d'idées, très prospère pour l'évolution de mon esprit critique.

Sur un plan moins professionnel, je voudrais tout d'abord remercier ma « carina », Céline SCORNAVACCA, qui m'a littéralement accompagné dans les hauts et les bas de ce défi de rédaction. Sans ta présence à la maison, je ne suis pas sûre d'avoir pu tenir le rythme. Puis, je remercie aussi Amine DEHBAOUI pour tout ce qui concerne un certain « placard ».

Un grand « namasté » à Arnaud KANCEL « Schwarzenegger » et son adorable famille « Schtroumpf ». Merci Adam KHIROUN « Stupidhead » et Jody FLETCHER « Poopy Head » pour les thés, les succulents chocolats – amaretto, les « mini cakes », les cocktails etc, mais avant tout pour vos sourires sincères et votre chaleureux accueil au 'vert anglais' où une grande partie de cette thèse fut rédigée.

Je remercie mes amies ou plutôt sœurs, Πόλα (Pola) et Μιμίνα (Mimina) BALTSOUKOU, ainsi que leurs maris, qui ont été parfois capables de traverser une très longue distance juste pour me faire savoir que je ne suis pas seule à l'autre bout du monde. Je remercie tous mes amis, et en particulier Κάλλια (Kallia) TRACHANA, dont l'intérêt manifesté par de mails, d'appels ou de sms m'a beaucoup soutenu au fil de ces années.

Enfin, je souhaiterais exprimer ma gratitude à ma famille et surtout ma mère, Πένα (Rena) VOKA. C'est grâce à elle que je suis arrivée jusque-là. Μαμά, je t'en suis profondément reconnaissante. Ευχαριστώ...

Table des matières

PROBLEMATIQUE	3
INTRODUCTION : LES TUNICIERS	7
1. CADRE HISTORIQUE	9
2. BIOLOGIE GENERALE	10
3. SYSTEMATIQUE ET CLASSIFICATION TRADITIONNELLE.....	13
<i>I. Ascidiacea</i>	13
<i>II. Thaliacea</i>	14
<i>III. Appendicularia</i>	15
4. LE REGISTRE FOSSILE.....	17
5. ÉCOLOGIE : LES ESPECES INVASIVES	18
6. METABOLITES SECONDAIRES ET COMPOSES CHIMIQUES ACTIFS NATURELS	20
7. L'ÉVOLEVO ET L'ÈRE GENOMIQUE	22
<i>I. Les gènes Hox</i>	24
<i>II. La voie de signalisation de l'acide rétinoïque</i>	25
CHAPITRE I :	
LES TUNICIERS AU SEIN DES CHORDES	27
1. L'ORIGINE DES CHORDES : DE LA SYSTEMATIQUE TRADITIONNELLE A LA PHYLOGENOMIQUE.....	29
2. LES DEBATS AUTOUR DE L'HYPOTHESE OLFACTORES	32
3. CORROBORATION DE LA NOUVELLE PHYLOGENIE DES CHORDES [ARTICLE 1]	34
CHAPITRE II :	
L'ARN RIBOSOMIQUE 18S ET LA PHYLOGENIE MOLECULAIRE DES TUNICIERS.....	39
1. PHYLOGENIE MOLECULAIRE DES TUNICIERS.....	41
2. LE STATUT PHYLOGENETIQUE DES GROUPES MAJEURS DE TUNICIERS.....	41
<i>I. Les Appendiculaires</i>	42
<i>II. Les Ascidiacés et la position controversée des Aplousobranches</i>	42
<i>III. Les Thaliacés</i>	44
3. MISE A JOUR DE LA PHYLOGENIE DES TUNICIERS SUR LA BASE DU GENE DE L'ARNr 18S EN UTILISANT DES MODELES DE MELANGE ET DE STRUCTURE SECONDAIRE [ARTICLE 2]	44
CHAPITRE III :	
LES GENOMES MITOCHONDRIAUX ET LA MITOGENOMIQUE DES TUNICIERS.....	49
1. LES GENOMES MITOCHONDRIAUX.....	51
2. LA DIVERGENCE MITOGENOMIQUE DES TUNICIERS AU SEIN DES CHORDES	52

I. Code génétique.....	52
II. Contenu en gènes.....	52
III. Distribution asymétrique des gènes sur les deux brins d'ADNmt	53
IV. Ordre des gènes.....	53
3. MITOGENOMIQUE ET RELATIONS PHYLOGENETIQUES DES ANIMAUX BILATERIENS	54
4. MITOGENOMIQUE DES TUNICIERS : LE GENOME MITOCHONDRIAL D' <i>HERDMANIA MOMUS</i> ET LA NOUVELLE PHYLOGENIE DES CHORDES [ARTICLE 3].....	55
CHAPITRE IV :	
L'ACCES AUX DONNEES GENOMIQUES ET L'EVOLUTION ACCELEREE DES TUNICIERS.....	59
1. L'ACCES AUX DONNEES GENOMIQUES DES TUNICIERS	61
2. LA REVOLUTION DU SEQUENÇAGE MASSIF A HAUT DEBIT	62
3. PROJET PILOTE DE L'APPLICATION DU SEQUENÇAGE 454 SUR LE TRANSCRIPTOME DE L'ASCIDIE INVASIVE <i>MICROCOSMUS SQUAMIGER</i>	62
3. ANALYSE EXPLORATOIRE DES DONNEES DE PYROSEQUENÇAGE DE <i>MICROCOSMUS SQUAMIGER</i>	63
I. Tuniciers.....	63
II. Deutérostomes	64
III. Métazoaires	65
4. CONTENU GENIQUE DU TRANSCRIPTOME PARTIEL DE <i>MICROCOSMUS SQUAMIGER</i>	65
5. ÉVOLUTION ACCELEREE DES TUNICIERS ET L'HETEROGENEITE DES TAUX D'EVOLUTION AU SEIN DES CHORDES [ARTICLE 4].....	67
SYNTHESE :	
CONCLUSIONS ET PERSPECTIVES	73
1. LA POSITION DES TUNICIERS DANS L'ARBRE DE LA VIE.....	74
2. REINTERPRETATION DE L'HISTOIRE EVOLUTIVE DES CHORDES : DISPARITES MORPHOLOGIQUE, GENETIQUE, ET GENOMIQUE DES OLFACTORES	78
I. Morphologie - anatomie.....	78
II. Génétique du développement embryonnaire	79
III. Génomique.....	80
3. PHYLOGENIE MOLECULAIRE DES TUNICIERS.....	82
4. DIVERGENCE MOLECULAIRE ET EVOLUTION ACCELEREE DES TUNICIERS	84
5. SEQUENÇAGE A HAUT DEBIT ET PHYLOGENOMIQUE DES TUNICIERS	86
REFERENCES BIBLIOGRAPHIQUES	91

Problématique

Les Tuniciers ou Urochordés sont des « invertébrés » marins qui regroupent les Ascidies, les Thaliacés (salpes et doliolles) et les Appendiculaires. Étant des animaux très diversifiés sur le plan des traits d'histoire de vie, il existe plusieurs types d'organisation (solitaire, colonial et social), de mode de vie (sédentaire, planctonique) ou encore de régimes de reproduction (asexué et sexué) au sein de la lignée (Grassé 1948). Du point de vue écologique, les Tuniciers présentent une répartition cosmopolite et se distribuent à des profondeurs allant de la surface jusqu'à la zone abyssale, disposant ainsi une aptitude à occuper des niches écologiques très variées (Lambert 2005b). De nombreuses espèces d'ascidies se sont en effet avérées invasives au cours des dernières années, ayant eu des impacts très nuisibles à la biodiversité autochtone des écosystèmes où elles se sont établies (Locke 2007).

Du point de vue systématique, les Tuniciers appartient au groupe des deutérostomiens, mais ce n'est qu'après les travaux pionniers d'Alexander Kowalevsky en 1867 sur l'embryologie comparée des Ascidies que les Tuniciers ont été classifiés parmi les Chordés (Kowalevsky 1868). Malgré leur apparence morphologique très simplifiée, la vaste majorité des espèces du groupe possède des larves pourvues d'une queue et d'une chorde nerveuse dorsale, qui régressent à l'âge adulte après une métamorphose qu'ils subissent au cours du passage à la maturité sexuelle (Lambert 2005a). Seuls les Appendiculaires parmi les Tuniciers maintiennent les caractères typiques des Chordés tout au long de leur vie.

Les Tuniciers revêtent un intérêt évolutif primordial pour l'étude des origines des Vertébrés de par leur position clé au sein des Chordés à la frontière entre les « invertébrés » et les Euchordés, rapprochant traditionnellement les Céphalochordés avec les Vertébrés. Dès lors, à l'aube des années 1990, certaines espèces des Tuniciers commencent peu à peu à s'établir parmi les organismes modèles en biologie évolutive du développement (EvoDevo) (Satoh 2003). Ainsi, le passage à l'ère génomique au début des années 2000 s'est accompagné du séquençage des génomes complets de trois Tuniciers modèles : *Ciona intestinalis* (Dehal *et al* 2002), *Ciona savignyi* (Small *et al* 2007) et *Oikopleura dioica* (Seo *et al* 2001).

L'accès aux données génomiques du groupe a eu des conséquences majeures sur la vision classique de l'évolution des Tuniciers. Tout d'abord, les analyses phylogénomiques réalisées à la faveur de l'essor des données génomiques dévoilent aux alentours de 2006 que, contre toute attente, les Tuniciers et non pas les Céphalochordés sont les plus proches parents actuels des Crâniates, en formant avec eux un groupe nommé Olfactores (Delsuc *et al* 2006).

En plus de réfuter le dogme des Euchordés, la phylogénomique couplée aux études de génomique comparative ont également mis en évidence le haut degré de divergences moléculaire et génomique distinguant les Tuniciers des autres lignées du phylum des Chordés (Seo *et al* 2004; Delsuc *et al* 2006; Holland 2007). Du point de vue moléculaire, cette forte divergence est parfaitement illustrée par les longues branches des membres du groupe observées dans les arbres phylogénomiques (Bourlat *et al* 2006; Delsuc *et al* 2008; Dunn *et al* 2008). Du point de vue de la structure génomique, elle se manifeste d'une part par l'ultra-compacité de leurs génomes due à la perte de nombreux gènes, et d'autre part par des remaniements géniques dramatiques associés principalement à des événements de translocation des introns et des régions intergéniques (Seo *et al* 2001; Dehal *et al* 2002; Small *et al* 2007).

L'émergence de ce projet de thèse se situe au moment où la phylogénomique vient inverser les positions respectives des Tuniciers et des Céphalochordés dans l'arbre de la vie, tandis que les premières ébauches génomiques soulignent l'évolution singulière du groupe vis-à-vis des autres Chordés. L'histoire évolutive des Tuniciers quant à eux reste obscure, puisqu'elle n'a été essentiellement étudiée que sur la base du seul marqueur phylogénétique universel que constitue l'ARN ribosomique 18S (ARNr 18S) (Tsagkogeorga *et al* 2009). Au même titre, la compréhension des modalités de l'évolution moléculaire du groupe restent quelque peu énigmatiques, puisqu'en dehors des trois Tuniciers dont les génomes sont séquencés, les données génomiques disponibles pour les autres groupes sont très restreintes. Celles-ci incluent d'une part des séquences d'ESTs (« Expressed Sequence Tags ») pour les trois ascidies, *Molgula tectiformis* où une large collection a été séquencée (Gyoja *et al* 2007), *Halocynthia roretzi* (Kim *et al* 2008) et *Diplosoma listerianum* (Blaxter & Thomas 2004) et d'autre part, de six génomes mitochondriaux (Gissi, Iannelli & Pesole 2008). Pour ces derniers, on compte un représentant des Thaliacea, *Doliolum nationalis*, et cinq ascidies, incluant *Ciona intestinalis* types A et B, *Ciona savignyi*, *Phallusia fumigata*, *Phallusia mammillata* et *Halocynthia roretzi* (Yokobori *et al* 1999; Yokobori, Watanabe & Oshima 2003; Gissi, Iannelli & Pesole 2004; Yokobori, Oshima & Wada 2005; Iannelli *et al* 2007a).

Dans ce contexte, l'axe principal de cette thèse s'est attaché au double objectif d'éclaircir l'histoire phylogénétique et la dynamique évolutive des Tuniciers sur le plan moléculaire. Quatre pistes de recherches ont été explorées à la lueur des différents types de marqueurs phylogénétiques à disposition :

(i) *La position des Tuniciers au sein des Chordés.*

Compte tenu du scepticisme affiché par rapport au positionnement des Tuniciers au sein des Chordés, tout d'abord, la robustesse de l'hypothèse Olfactores a été réévaluée à la lueur des différentes données moléculaires disponibles pour les Deutérostomiens. En utilisant des modèles d'évolution des séquences plus adéquats pour les analyses phylogénétiques qui prennent en compte l'hétérogénéité substitutionnelle dans les séquences (Lartillot & Philippe 2004) ou atténuent les biais de composition entre les taxons, la proche parenté phylogénétique des Tuniciers avec les Vertébrés a été corroborée (Delsuc *et al* 2008) [Article 1].

(ii) *L'histoire évolutive des Tuniciers.*

Dans le but de fournir un cadre phylogénétique plus fiable de l'histoire évolutive des Tuniciers, la phylogénie du groupe à partir de l'ARNr 18S a été mise en jour (Tsagkogeorga *et al* 2009) [Article 2]. Ceci a été réalisé par l'acquisition de 30 nouvelles séquences et par leur analyse phylogénétique en utilisant des modèles d'évolution des séquences plus adaptés à la dynamique évolutive de l'ARNr 18S, c'est-à-dire des modèles considérant l'hétérogénéité des pressions de sélection imposée par la structure secondaire de la molécule (Savill, Hoyle & Higgs 2001; Lartillot & Philippe 2004).

(iii) *La mitogénomique des Tuniciers.*

Visant à approfondir l'évolution mitogénomique des Tuniciers, le séquençage du génome de l'ascidie *Herdmania momus* a été réalisé. Avec les six génomes mitochondriaux déjà disponibles, ceci a permis de confirmer la dynamique évolutive particulière des Tuniciers par la mise en évidence de nombreux réarrangements de l'ordre des gènes au sein du groupe (Gissi, Iannelli & Pesole 2008). L'analyse phylogénétique des données mitogénomiques a corroboré d'une part le taux d'évolution élevé du groupe, et d'autre part a permis grâce à l'application de modèles d'évolution de séquences non-homogènes entre sites (Lartillot & Philippe 2004) et dans le temps (Blanquart & Lartillot 2008) d'obtenir pour la première fois un soutien significatif pour l'hypothèse Olfactores (Singh *et al*) [Article 3].

(iv) *La divergence moléculaire et l'évolution accélérée des Tuniciers.*

En ce qui concerne le génome nucléaire, l'évolution moléculaire du groupe a été appréhendée par une étude comparative portant sur 35 gènes orthologues conservés à l'échelle des métazoaires (« gènes de ménage ») qui ont été obtenus par séquençage 454 GS-FLX

standard (<http://www.454.com/>) d'une banque d'ADNc de l'ascidie invasive *Microcosmus squamiger*. Par une approche impliquant des modèles d'horloges locales (Yoder & Yang 2000) et d'auto-corrélation du taux d'évolution le long des branches (Thorne, Kishino & Painter 1998), l'analyse de ces marqueurs a révélé un taux d'évolution deux fois plus élevé en moyenne chez les Tuniciers que chez des autres Chordés, avec cependant de larges différences observées entre lignées et entre gènes. Une première évaluation des déterminants de cette accélération conduite par l'estimation du dN/dS sur des paires d'espèces au sein des Olfactores suggère que la composante principale de cette variation réside dans des changements du taux de mutation chez les Tuniciers (Tsagkogeorga *et al* En préparation) [Article 4].

En tant que suite logique aux travaux réalisés au cours de ces trois années, un projet à plus long terme a été initialisé visant à l'acquisition de nouvelles données phylogénomiques pour un nombre des espèces plus important de Tuniciers. Bénéficiant des avancées technologiques réalisées récemment dans le domaine de séquençage à haut-débit, les transcriptomes partiels de dix nouvelles espèces de Tuniciers ont été pyroséquencés en utilisant la technologie 454 GS-FLX Titanium. Ceci a permis l'accès à plus d'un million de séquences supplémentaires pour le groupe, dont les analyses préliminaires paraissent très prometteuses pour mener de futures études d'évolution moléculaire et de génomique comparative. Un prétraitement bioinformatique des données 454 acquises a déjà permis l'identification de 153 marqueurs orthologues à l'échelle des Métazoaires à partir desquels un cadre phylogénétique fiable a pu être établi pour les Tuniciers au sein des Chordés.

INTRODUCTION : LES TUNICIERS

1. Cadre historique

"Mr. Kowalevsky has lately observed that the larvae of ascidians are related to the Vertebrata, in their manner of development, in the relative position of the nervous system, and in possessing a structure closely like the chorda dorsalis of vertebrate animals;[...] Thus, if we may rely on embryology, ever the safest guide in classification, it seems that we have at last gained a clue to the source whence the Vertebrata were derived".

Charles Darwin (1871) cité dans Cañestro *et al.* (2003)

Les Tuniciers ou Urochordés (Appendiculaires, Salpes et Doliolés, Ascidies) sont des animaux deutérostomiens marins qui appartiennent au phylum des Chordés au même titre que les Céphalochordés et les Vertébrés. Déjà reconnus par Aristote sous le nom *Tethya* ou *Thalia*, leur position phylogénétique au sein des animaux resta controversée pour au moins les vingt trois siècles suivants, en fait jusqu'au début du 21^{ème} siècle et le passage à l'ère génomique.

La première tentative de classification du groupe date de l'époque de Lamarck qui, en 1809, les classa avec les mollusques (Grassé 1948). Il introduit aussi le nom « Tunicata » en référence à la présence chez tous les membres du groupe d'une « tunique », enveloppe externe semi-transparente de nature cellulosique qui les entoure. En 1828, Cuvier proposa, également, que les Tuniciers représentent un forme de mollusques « sans coquille », une opinion qui s'établit durant la première partie du 19^{ème} siècle (Katz 1983).

Alors que les Tuniciers été considérés comme des invertébrés protostomiens et donc très éloignés des Vertébrés du point de vue phylogénétique, l'embryologiste russe Alexander Kowalevsky fut le premier à démontrer que les larves des Ascidies possèdent en fait les caractères typiques du plan d'organisation des Chordés (Kowalevsky 1868). Elles possèdent en effet un système nerveux tubulaire dorsal, un tube intestinal ventral, ainsi qu'une notochorde axiale encadrée latéralement par du tissu musculaire (Kowalevsky 1868). Cette découverte éveilla Darwin, Huxley et d'autres défenseurs du courant darwinien, culminant avec l'introduction par Lankester en 1877 du nom « Urochordata » afin de dénoter la forme larvaire de type « têtard » des représentants du groupe [cité dans (Katz 1983)].

L'ensemble des travaux des Kowalevsky (1868), Willey (1893) et Conklin (1905) au détour du 20^{ème} siècle, établirent ainsi l'appartenance des Tuniciers aux Chordés, en leur accordant une position basale au sein du phylum. Les Céphalochordés comme l'amphioxus

qui possèdent une morphologie superficiellement similaire à celle des larves de poissons forment avec les Vertébrés le clade Euchordata, *i.e.* les véritables Chordés.

Dans les années 2000, le début de l'ère génomique incita les projets de séquençage des génomes de trois Tuniciers : *Ciona intestinalis*, *Ciona savignyi* et *Oikopleura dioica*. Du fait de leur position clé à la frontière entre « invertébrés » et Vertébrés, l'étude approfondie des génomes de Tuniciers promet une meilleure compréhension des mécanismes évolutifs et développementaux à l'origine des Vertébrés actuels. Bénéficiant du contexte, Delsuc *et al.* (2006) menèrent une étude phylogénétique à l'échelle génomique où contre toute attente, ils démontrèrent que les Tuniciers, et non pas les Céphalochordés, constituent les plus proches parents des Vertébrés (Delsuc *et al* 2006).

Le bouleversement de la vision traditionnelle de l'évolution des Chordés par l'inversion des positions respectives des Céphalochordés et des Tuniciers fut accueillie avec un certain scepticisme. Cependant, aujourd'hui, l'accumulation de travaux phylogénomiques (Bourlat *et al* 2006; Delsuc *et al* 2008; Dunn *et al* 2008) culminant avec le séquençage du génome complet de l'amphioxus (Putnam *et al* 2008), corroborent la chute définitive du dogme Euchordata en faveur d'un nouveau scénario évolutif pour les Chordés : l'hypothèse Olfactores groupant Vertébrés et Tuniciers.

2. Biologie générale

Les Tuniciers sont des animaux filtreurs dont la source alimentaire principale est constituée du plancton et de la matière organique en suspension dans le milieu marin. La plupart des espèces ressemblent morphologiquement à un sac semi-transparent pourvu de deux orifices : un siphon inhalant (buccal) et un siphon exhalant (cloacal). L'eau passe successivement par le siphon buccal jusqu'aux filtres des branchies et la cavité de l'atrium, aboutissant au siphon cloacal par où elle est expulsé. Les particules alimentaires retenues au niveau du pharynx sont ensuite transportées vers l'estomac. Le pharynx sert aussi à la respiration. La circulation est en partie lacunaire, en réseau non fermé, l'appareil circulatoire se réduisant à un cœur simple et tubuleux. Le système nerveux est lui représenté par un ganglion placé entre les orifices d'entrée et de sortie, et par des nerfs très fins qui suivent la ligne médiane dorsale.

En tant que membres des Chordés, les tuniciers présentent à l'état larvaire quatre structures typiques du plan d'organisation du phylum :



Figure 1 : L'Appendiculaire *Oikopleura labradorensis*

- 1) Une notochorde ;
- 2) Une corde nerveuse tubulaire dorsale ;
- 3) Des fentes pharyngiennes ;
- 4) Une queue post-anale.

Chez les Ascidies ainsi que les Thaliacés la corde et la queue régressent à l'état adulte au cours d'une métamorphose qui conduit à

des adultes fortement remaniés. En revanche, les Appendiculaires ne suivent pas ce programme de développement car, bien qu'ils passent par une métamorphose moins poussée, les espèces appartenant à cette classe maintiennent les caractères typiques des Chordés tout au long de leur vie (Fig. 1).

Les Ascidies constituent la lignée la plus diversifiée de Tuniciers, avec plus de 2500 espèces décrites à ce jour. Ces espèces ont un cycle de vie biphasique où la larve pélagique, suite à une métamorphose radicale, adopte à l'état adulte un mode de vie sessile ou sédentaire. Au sein du groupe, il existe trois grands types d'organisation : « solitaire » lorsque les individus vivent isolés, et « social » et « colonial » lorsqu'ils constituent des colonies interconnectées. Les Thaliacés et les Appendiculaires, qui comprennent environ une centaine d'espèces chacun, sont eux exclusivement planctoniques.

La majorité des Tuniciers sont hermaphrodites avec cependant certains cas de protandrie, c'est-à-dire d'hermaphrodisme successif par l'acquisition des caractères sexuels dans l'ordre mâle puis femelle. Dans ce contexte, l'auto-incompatibilité est un phénomène présent chez la plupart des espèces. Durant la reproduction sexuée, les Ascidies solitaires, ainsi que les Appendiculaires, relâchent les gamètes directement dans le milieu marin, la fécondation étant externe. Les espèces coloniales comme les Botryllidés possèdent en revanche des poches d'incubation des ovules dans l'atrium, seuls les gamètes mâles étant expulsés dans l'eau. La reproduction asexuée alternative s'effectue elle par bourgeonnement (Fig. 2). Les Thaliacés sont en revanche caractérisés par des cycles de reproduction plus complexes comportant notamment des générations alternantes (Lambert 2005a).

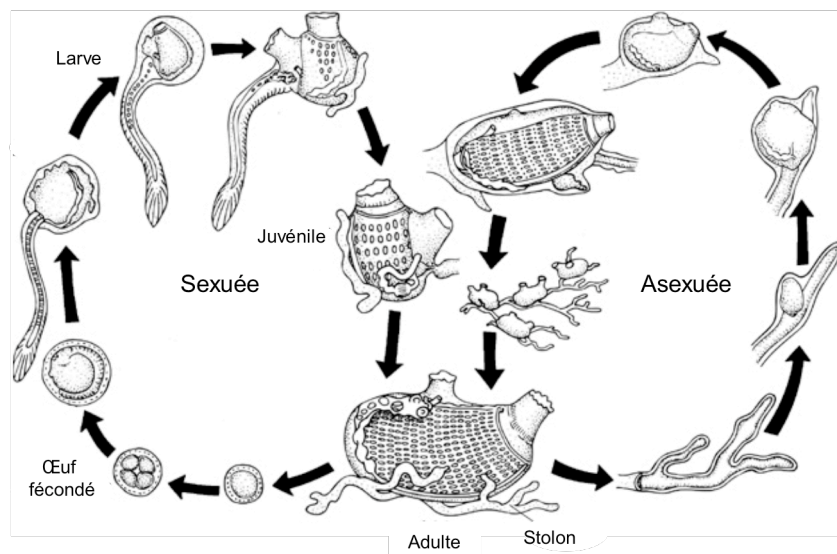


Figure 2 : Représentation schématique de la reproduction sexuée et asexuée chez les Ascidiées (modifiée après Swalla, 2006)

L'embryogenèse succédant à la fécondation est de type « déterminé », tous les Tuniciers suivant en effet un mode de développement en mosaïque (Lemaire, Smith & Nishida 2008). La différenciation cellulaire se réalise très tôt durant le développement embryonnaire et la segmentation de l'œuf s'effectue selon un plan prédéfini et inaltérable. Chaque blastomère atteint une position précise et possède un potentiel développemental limité en constituant chacun le précurseur d'un tissu spécifique du futur embryon.

Néanmoins, ce programme général de développement présente d'importantes variations entre espèces, bien que le phénomène ait été peu étudié jusqu'à présent, et uniquement chez les Ascidiées. Un exemple représentatif est fourni par le fort contraste existant entre *Ecteinascidia turbinata* et *Ciona intestinalis* en termes d'intervalles de temps entre la fécondation et l'état larvaire (Jeffery 2007). Chez *Ecteinascidia turbinata*, le développement de l'œuf fécondé à la formation d'une larve de type « têtard » dure 10 à 15 jours, tandis que l'œuf fécondé de l'ascidie *Ciona intestinalis* atteint cet état en moins de 18 heures. Restant un à deux jours dans la zone planctonique, la larve subit ensuite une métamorphose aboutissant à l'adulte sédentaire et la maturité sexuelle est atteinte à environ 3 mois.

3. Systématique et classification traditionnelle

La forte diversité des Tuniciers en termes de traits d'histoire de vie, et plus précisément du point de vue du développement, constitua le noyau de la classification traditionnelle. Par conséquent, le mode de vie, le plan d'organisation, ou encore la structure des branchies et des spermatozoïdes furent parmi les caractères à partir desquels l'histoire évolutive du groupe a été reconstruite. A haut niveau taxonomique, la classification traditionnelle distingue trois grandes classes au sein des Tuniciers: les Ascidiacea (Ascidies), les Thaliacea (Salpes, Pyrosomes et Doliolles), et les Appendicularia (Appendiculaires ou Larvacés).

I. Ascidiacea

Les Ascidiacés constituent sans doute la classe la plus diversifiée en incluant approximativement 2500 à 3000 espèces appartenant à environ 20 familles. La classification actuellement établie est essentiellement basée sur l'étude systématique de Lahille (1888-1890), qui divisa les Ascidies en trois ordres selon la structure de leurs branchies :

- I. *Aplousobranchia* : les Tuniciers de cet ordre présentent une branchie à trémas complets et le plus souvent à côtes transverses bien développées, mais en aucun cas des côtes longitudinales ni des sinus anastomotiques longitudinaux ne sont observés.
- II. *Phlebobranchia* : les seules Ascidies dont les branchies portent des sinus anastomotiques longitudinaux.
- III. *Stolidobranchia* : toutes les Ascidies ayant des branchies présentant des côtes longitudinales ou même des replis internes longitudinaux (plis méridiens).

Lahille inclut cependant les Salpes et les Doliolles dans l'ordre des *Aplousobranchia* (Lahille 1886; Lahille 1887; Lahille 1890), la monophylie des Ascidies fut définie quelques années plus tard.

Perrier (1898) et Garstang (1928) proposèrent également une classification alternative en divisant les Ascidiacea en deux Ordres, *Pleurogona* et *Enterogona*, qui diffèrent entre eux quant à la position des gonades. Les deux schémas de classification sont à présent rencontrés en étant même parfois combinés: *Enterogona* comprenant les *Aplousobranchia* et *Phlebobranchia*, et *Pleurogona* étant constitué des *Stolidobranchia*.

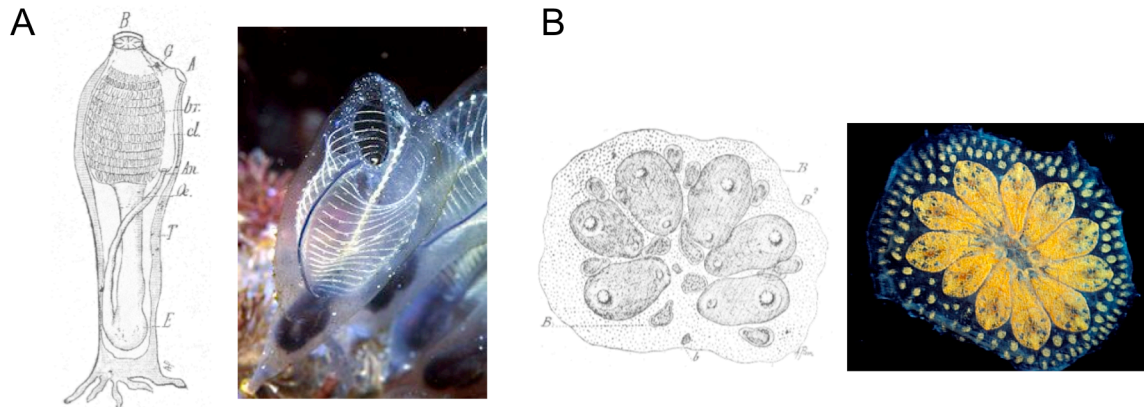


Figure 3 : LES ASCIDIÉS

(A) Une Claveline : B, bouche ; br, sac branchial ; oe, oesophage ; F, estomac ; An, anus qui débouche dans le cloaque cl ; T, tunique ; G, ganglion nerveux.

(B) Colonie de *Botryllus* : B², bourgeons ; B, restes granuleux ; b, restes granuleux de la génération précédente.

II. Thaliacea

La classe des Thaliacés comprend trois ordres, les Doliolida, les Pyrosomatida et les Salpida. L'ensemble de leurs représentants sont morphologiquement très similaires, cependant plusieurs auteurs attribuent cette ressemblance à une évolution convergente [Uljanin 1884, Holland 1988, Franzen 1992 cité dans (Holland 1990; Godeaux 1998; Holland 1999)].

L'ordre Pyrosomatida comprend environ dix espèces. Le nom du genre *Pyrosoma* a été proposé par Péron (1804) et fait référence à la bioluminescence de ces organismes pélagiques. Des espèces de Pyrosomes furent également décrites par Lesueur (1815) et Savignyi (1816), la systématique du groupe étant basée sur les révisions de van Soest (1979a, 1981) après Neumann (1913), Metcall et Hopkins (1919). Les caractères utilisés

pour la classification du groupe sont nombreux et incluent entre autres le plan d'organisation, la forme des colonies, ainsi que la position et le nombre de stigmates.

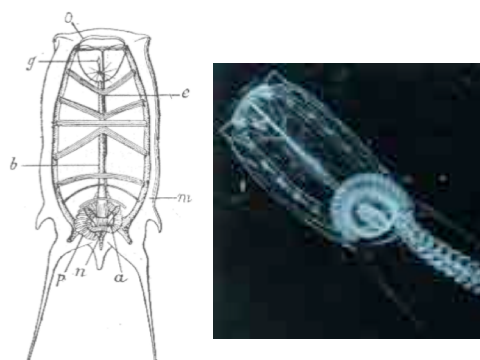


Figure 4 : Les Thaliacés

Une salpe : o, boucle ; a, siphon cloacal ; b, branchie ; g, pavillon cilié ; m, manteau ; n, nucleus (masse viscérale) ; p, jeune chaîne de Salpes en voie de développement ; e, endostyle.

L'ordre Doliolida inclut qu'en à lui trois familles : les Doliolidae, les Doliopsoididae et les Doliopsidae. La systématique de ce groupe a été établie à partir de caractères variés, parmi lesquels figurent la forme du tube intestinal, le nombre et la taille des anneaux musculaires et la position des gonades. La plupart des Doliolidae furent décrites par Garstang (1933), des centaines d'espèces étant identifiées par Rathke (1835), Krüger (1939), Tokioka et Berner (1958) et Braconnot (1970).

Le troisième ordre Salpida comprend environ une quarantaine d'espèces appartenant à une seule famille, les Salpidae incluant les sous-familles Cycloalpinæ et Salpinæ (Yount, 1954). Ce schéma de classification initialement proposé par Ihle et Ihle-Landerberg (1933) après Mercalf (1918), est principalement basé sur la morphologie comparative de la musculature, l'orientation du canal alimentaire et la forme du stolon, *i.e.* l'arrangement des blastozoïdes (Godeaux 1998).

III. Appendicularia

Les Appendiculaires se distinguent parmi les Tuniciers de par leur taille minuscule et de leur plan d'organisation caractérisé d'un tronc et d'une queue mobile qu'ils retiennent tout au long de leur vie. Ce groupe peut être également identifié par l'absence d'une cavité péripharyngienne et des spiracles, *i.e.* une paire de perforations pharyngiennes entourées de cils que l'on trouve chez tous les autres Tuniciers.

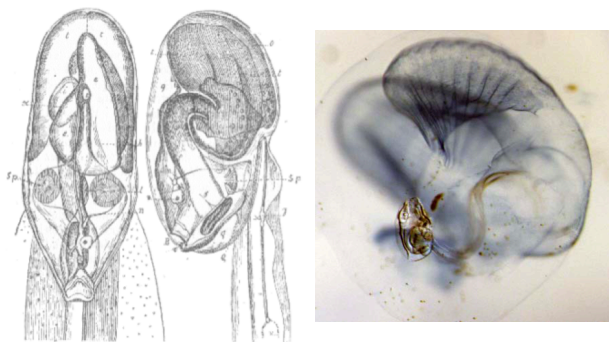


Figure 5 : Les Appendiculaires

Trois espèces d'*Oikopleura* : t, testicule; e, estomac; i, intestin; p, pylore; h, coeur; l, lignes vibratiles longitudinales; n, les nerfs branchiaux; o, ovaire ; Q, repli de l'épiderme sous la queue; B, bouche; vésicule auditive; D, granules terminaux; t, cirrhes que porte l'endostyle; sp, fentes branchiales.

mollusques (Mertens, 1830), ou encore des zoophytes (Quoy et Gaimard, 1833).

Les premières études des Appendiculaires datent de 1821, lorsque Chamisso décrivit l'Appendiculaire *Appendicularia flagellum*. Cependant les affinités phylogénétiques des Appendiculaires restèrent très longtemps controversées, et ils furent considérés tour à tour comme des médusoïdes (Chamisso 1821), des

Tableau 1 : Classification traditionnelle des Tuniciers

CLASSES	ORDRES	FAMILLES
APPENDICULARIA	Fritillariidae	
	Kowalevskiidae	
	Oikopleuridae	
ASCIDIACEA	Aplousobranchia	<i>Clavelinidae</i>
		<i>Didemnidae</i>
		<i>Polycitoridae</i>
		<i>Polyclinidae</i>
		<i>Holozoidae</i>
	Phlebobranchia	<i>Agnesiidae</i>
		<i>Asciidiidae</i>
		<i>Cionidae</i>
		<i>Corellidae</i>
		<i>Diazonidae</i>
		<i>Hypobythidae</i>
		<i>Octanemidae</i>
		<i>Perophoridae</i>
	Stolidobranchia	<i>Molgulidae</i>
		<i>Pyuridae</i>
		<i>Styelidae</i>
THALIACEA	Doliolida	<i>Doliolidae</i>
		<i>Doliopsoididae</i>
		<i>Doliopoidae</i>
	Pyrosomatida	<i>Pyrosomatidae</i>
	Salpida	<i>Salpidae</i>

Ce n'est qu'en 1851 que l'appartenance des Appendiculaires au groupe des Tuniciers fut proposée par Huxley, relation qui fut établie quelques années plus tard par Fol (1872), Herdman assignant au groupe le nom *Larvacea* en 1882 (Kott 2005).

La classification des Appendiculaires est à l'origine des travaux de Lohmann (1896a) sur les caractéristiques morphologiques du plan corporel, de l'endostyle, du pharynx, ainsi que de l'épithélium oikoplastique et de la queue. Nullement modifiée ultérieurement, la classification au sein du groupe inclut trois ordres/familles : les Oikopleurida(e) (Lohmann 1896a), les Fritillarrida(e) (Seeliger 1895) et les Kowalevskiida(e) décrits par Lahille (1888).

4. Le registre fossile

La fossilisation étant un processus particulièrement rare dans le cas des organismes à corps mous, le registre fossile des Tuniciers est très restreint. Le premier fossile décrit en tant que Tunicier est un spécimen partiel du genre *Cheungkongella* trouvé dans le Maotianshan Shale près du village Ercai d'Haikou (Kūnmíng) en Chine du Sud (Shu *et al* 2001a). Datant de la période du Cambrien Inférieur, ce spécimen fut cependant ultérieurement re-évalué par Chen *et al.* (2003), qui démontrèrent que *Cheungkongella* constitue en fait une espèce éteinte de Phlogites et non pas de Tuniciers.

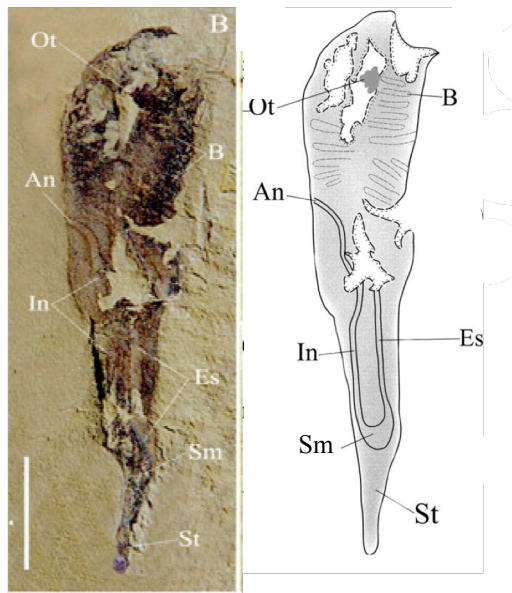


Figure 6 : *Shankouclava shankouense*

An, anus ; B, branchie ; Es, oesophage ; In, intestine ; Ot, appendice oral ; Sm, estomac ; St, stolon; (Echelles : 5 mm).

Chen *et al.* (2003) décrivent en revanche une autre espèce également issue du Cambrien Inférieur, *Shankouclava shankouense*, à partir de huit spécimens récoltés près du village Shankou d'Anning dans la même région (Kūnmíng, Chine). Portant des caractères représentatifs des Tuniciers, notamment un sac branchial perforé, deux siphons, un endostyle et un intestin en forme de U, *Shankouclava shankouense* (Fig. 6) constitue à présent le fossile le plus ancien du groupe, et est considéré comme étant morphologiquement très proche des Aplousobranchia actuels.

Enfin, une partie assez problématique

du registre fossile d'âge Cambrien du genre *Vetulicola*, fut proposée comme étant à l'origine des Tuniciers en tant que lignée basale des Larvacés (Dominguez & Jefferies 2003). Néanmoins, un certain débat subsiste quant à l'appartenance des Vetulicolians aux Appendiculaires, la position phylogénétique de ce groupe restant actuellement très controversée (Shu *et al* 2001b).

5. Écologie : les espèces invasives

Du point de vue écologique, le groupe des Tuniciers présente dans son ensemble une distribution cosmopolite s'étendant de l'Antarctique à l'Arctique et comprenant aussi bien les zones de marées côtières que des zones de grande profondeur supérieures à moins 2000 mètres (Lambert 2005b). Les Ascidies sont majoritairement cosmopolites et occupent la zone démersale proche des côtes mais il existe également des espèces abyssales appartenant à la famille des Octacnemidae (Stolidobranchia). Les Appendiculaires quant à eux occupent la zone épipélagique ou eutrophique (*i.e.* de la surface jusqu'à 200 mètres). Ils arrivent à la deuxième place parmi les organismes les plus abondants du plancton de tous océans du monde. Enfin, les Thaliacés sont également répandus dans la zone épipélagique de toutes mers du globe à l'exception de l'Arctique. Les Tuniciers de cet ordre effectuent cependant des migrations verticales journalières dans la colonne d'eau atteignant des profondeurs supérieures à moins 1000 mètres.

Les aspects écologiques et physiologiques à l'origine de l'aptitude des Tuniciers à occuper des niches écologiques aussi variées sont peu connus. Les facteurs régissant leurs aires de distribution comprennent la température, la salinité, la lumière et l'hydrodynamique du milieu. Certaines Ascidies sont connues pour afficher une tolérance remarquable à des concentrations élevées en éléments traces métalliques (ETM ou métaux lourds) du milieu comme l'arsénique, le cadmium, le cobalt, le cuivre, le fer, l'étain ou le zinc. Ces espèces présentent la particularité de comporter de fortes concentrations de vanadium et de lithium dans leur propre sang.

Des nombreuses Ascidies sont connues pour leur capacité envahissante et perturbent la biodiversité autochtone des écosystèmes où elles s'établissent. La liste des espèces invasives avérées est ainsi longue (Tableau 2). Des espèces appartenant aux genres *Ciona*, *Botryllus*, *Botrylloides*, *Nodostoma*, *Pyura*, et plus particulièrement aux genres *Styela* et *Diplosoma* comportent des populations invasives après leur introduction sur les côtes de différents pays,

notamment le Canada, les Etats-Unis, les Pays-Bas et la Nouvelle-Zélande (Locke & Carman 2009).

Tableau 2 : Liste des espèces d'Ascidies comportant des populations envahissantes et leur lieu d'introduction

ESPECE	LIEU D'INTRODUCTION
<i>Ascidiella aspersa</i>	Nouvelle-Zélande
<i>Ciona intestinalis</i>	Australie
<i>Corella eumyota</i>	France (Bretagne)
<i>Didemnum</i> sp.	États-Unis, Nouvelle-Zélande, France et Pays-Bas
<i>Molgula manhattensis</i>	Australie
<i>Perophora japonica</i>	France (English Channel), Angleterre, Californie
<i>Polycarpa zorritensis</i>	Italie du Nord, Californie
<i>Pyura praeputialis</i>	Chili
<i>Styela clava</i>	Répartition mondiale
<i>Styela plicata</i>	Australie

L'impact le plus nuisible d'une invasion documentée concerne l'introduction des espèces *Styela clava*, *Botrylloides violaceus*, *Botryllus schlosseri* et *Ciona intestinalis* sur l'île du Prince Edward au Canada. Ces introductions ont notamment eu un effet catastrophique sur la viabilité économique de l'industrie mytilicole locale (Locke 2007).

La capacité des Tuniciers à s'incruster et s'accroître rapidement sur une matière solide naturelle ou artificielle dans le milieu aquatique (« biofouling »), le transport anthropogénique par les coques des bateaux et les eaux de ballaste, ainsi que le réchauffement global des océans semblent être les moteurs principaux des nombreux épisodes d'invasion récemment rapportés. La croissance incessante des échanges maritimes mondiaux ainsi que les changements



Figure 7 : Une espèce d'Ascidie coloniale, *Didemnum lahillei*, attachée aux cages de mytiliculture en Colombie Britannique (Photo de Gordon King).

climatiques en cours risquent de renforcer ce problème en augmentant les introductions involontaires de Tuniciers dans des zones exogènes de leur répartition naturelle.

6. Métabolites secondaires et composés chimiques actifs naturels

Depuis les années 1990, les Tuniciers sont l'objet de nombreuses recherches dont l'objectif vise à l'identification de composés chimiques actifs toutefois naturels ayant un potentiel d'application en médecine et pharmaceutique. Parmi ceux déjà identifiés constituent des métabolites toxiques produits par les Ascidies qui sont utilisés dans la défense contre la prédation et dans la compétition territoriale intra- et inter-spécifique (Lambert 2005b).

Parmi les cas les plus connus, figurent la didemnine B, l'aplidine et l'ecteinascidine 743, qui ont été identifiés et isolés à partir des espèces *Trididemnum solidium*, *Aplidium albicans* et *Ecteinascidia turbinata*, respectivement (Rinehart 2000). Issus de recherches financées par l'Institut National du Cancer des Etats-Unis (US National Cancer Institut) et des grandes industries pharmaceutique (ex. PharmaMar), ces agents ont été sélectionnés pour leur action anti-tumeur et immunodépressive et ont passé le stade d'essais cliniques de type II réalisés sur l'homme à large échelle.

La recherche actuelle focalisant sur la découverte des composants chimiques chez des organismes marins est très active (Blunt *et al* 2004; Blunt *et al* 2006) et fait écho à l'intérêt de la communauté scientifique étudiant les Tuniciers. Le tableau 3 cite quelques exemples d'espèces chez lesquelles de nouveaux agents chimiques ont été caractérisés au cours des dernières années.

Tableau 3 : Exemples de composés chimiques isolés et caractérisés chez les Tuniciers depuis l'année 2002.

ESPECE D'ASCIDIE	COMPOSE CHIMIQUE
<i>Aplidium longithorax</i>	dimeric prenylated quinone (–)-longithorone A
<i>Aplidium meridianum</i>	aminopyrimidineindole alkaloids
<i>Atriolum robustum</i>	amino acid derived metabolites
<i>Botrylloides tyreum</i>	botryllamides E–H
<i>Botryllus leachi</i>	botryllazine B
<i>Ciona intestinalis</i>	sulfated steroid of sperm activation & attraction
<i>Clavelina lepadiformis</i>	lepadiformine
<i>Clavelina oblonga</i>	(2 <i>S</i> ,3 <i>R</i>)-2-aminododecan-3-ol
<i>Cystodytes</i> cf. <i>violatinctus</i>	violatinctamine
<i>Cystodytes dellechiajei</i>	sebastianines A & B
<i>Didemnum</i>	fascaplysin analogues
<i>Didemnum molle</i>	cyclodidemnamide B
<i>Didemnum obscurum</i>	lamellarins γ , α & ϵ
<i>Didemnum</i> , <i>Aplidium tabascum</i>	lepadins D, E & H63
<i>Ecteinascidia turbinata</i>	ecteinascidins, ETs 731, 745b, 808 & 815
<i>Halocynthia aurantium</i>	Halocidin
<i>Lissoclinum bistratum</i>	biselides A
<i>Lissoclinum bistratum</i>	bistratamides E, G & J
<i>Lissoclinum</i> cf. <i>badium</i>	lissoclibadin 1
<i>Microcosmus sulcatus</i>	sulcaceramide
<i>Perophora viridis</i>	1,2,3-trithiocane derivatives
<i>Polyandrocarpa</i> sp.	polyandrocarpamines A & B
<i>Polycarpa clavata</i>	polycarpine
<i>Pseudodistoma cereum</i>	1,3-dimethyl-8-oxoisoguanine
<i>Pycnoclavella kottae</i>	kottamides A–D
<i>Rhopalaea</i> sp.	rhopaladins A–D
<i>Styela plicata</i>	tunichrome <i>Sp</i> -1

7. L'EvoDevo et l'ère génomique

« A host of modern techniques and manipulations have been used to study ascidian development, including cell ablation, electroporation of eggs and embryos, antisense inactivation of mRNA, whole mount in situ hybridization, and overexpression of genes »

Ciona project information, Broad institute

<http://www.broadinstitute.org/annotation/ciona/>

Du point de vue historique, les Ascidies constituent les premiers animaux utilisés en tant que système modèle pour l'embryologie expérimentale, dans les travaux pionniers de Laurent Chabry et d'Edwin Grant Conklin à la charnière du 20^{ème} siècle. En effectuant des résections sur la progéniture des muscles chez les embryons d'*Ascidiella aspersa*, Chabry fut à l'origine de la théorie du développement en mosaïque (Chabry 1887). Quelques années plus tard, Conklin (1905) en examinant des embryons de *Styela partita* révéla la non-uniformité du contenu de l'œuf fécondé, ce qui conduit à la découverte primordiale du croissant jaune et de la ségrégation cytoplasmique après la fécondation (Lemaire 2009).

Après environ un siècle passé dans l'ombre, l'étude du développement des Ascidies ne reprit qu'à partir des années 1990, avec notamment les travaux du biologiste japonais Noriyuki Satoh de l'Université de Kyoto sur l'ascidie solitaire *Ciona intestinalis*. Pourvue du plan de développement de type « têtard » et étant en même temps facilement manipulable en laboratoire, la cione intestinale constitue un organisme expérimental particulièrement prometteur pour appréhender le développement plus complexe des Vertébrés (Satoh 2003). N. Satoh s'investit dès lors dans la génétique du développement de l'espèce, posant ainsi les fondements pour son futur établissement parmi les organismes modèles en biologie évolutive du développement (EvoDevo).

Dans le même temps, deux équipes californiennes indépendantes, celles de Michael Levine de l'Université de Berkeley et celle de William Smith de l'Université de Santa Barbara préconisent d'utiliser *Ciona intestinalis* comme organisme modèle pour l'étude de la régulation génique et de son impact sur le changement du programme de développement au cours de l'évolution des animaux (Pennisi 2002b). C'est dans ce contexte que le premier assemblage du génome complet de *Ciona intestinalis* est publié (Dehal *et al* 2002) alors que le génome de *Ciona savignyi* séquencé en parallèle est finalement publié plus tard (Small *et al* 2007). À ceci s'ajoute le séquençage en cours du génome miniature de l'Appendiculaire *Oikopleura dioica* (Seo *et al* 2001).

Ces projets génomiques ont révélé que les tuniciers disposent de génomes ultra-compacts, dont la taille chez les *Ciona* est de l'ordre de 160 Megabases (Mb), le dernier assemblage de référence de *Ciona savignyi* étant estimé à 174 Mb (<http://mendel.stanford.edu/sidowlab/ciona.html>). Ceci ne correspond qu'à environ 5% de la taille du génome humain, alors qu'en contenant approximativement 16 000 gènes, le génome de *Ciona intestinalis* porte l'essentiel de l'arsenal génique des Chordés (Satoh *et al* 2006). Étant dépourvu des duplications caractéristiques des génomes des Vertébrés, le contenu génique des Ascidiés couvre environ 50% de celui de l'homme (Dehal & Boore 2005). Le degré de compacité du génome d'*Oikopleura dioica* est encore plus frappant puisqu'une première estimation inférieure à 70 Mb a été avancée (Seo *et al* 2001), en faisant le plus petit génome animal séquencé jusqu'à présent.

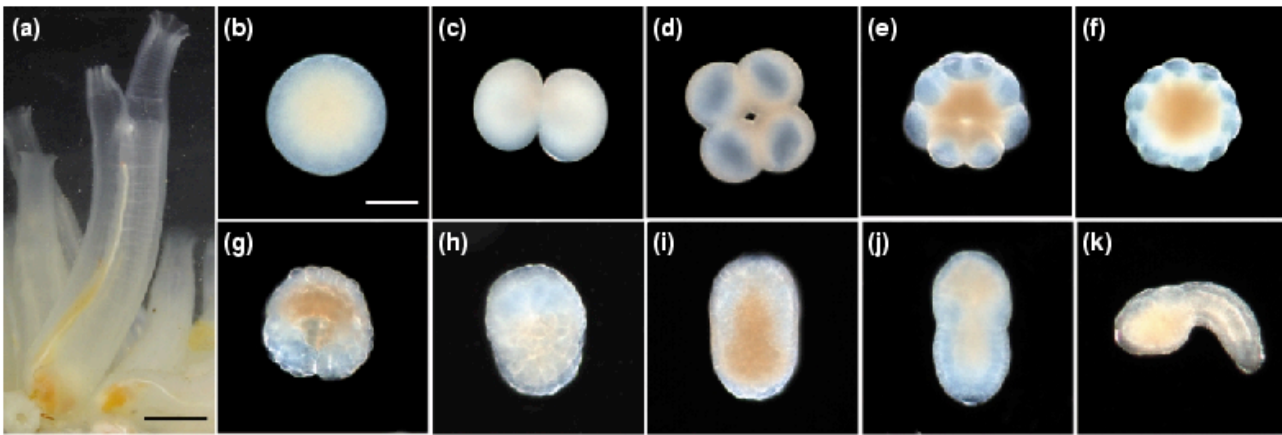


Figure 8 : L'embryogenèse chez *Ciona intestinalis*

(a) adulte, (b) œuf fécondé, (c) embryon à 2 cellules, (d) embryon à 4 cellules, (e) embryon à 16 cellules, (f) embryon à 32 cellules, (g) gastrula, (h -j) formation de la larve, (k) larve de type « têtard ».

Le passage à l'ère génomique déclencha un très fort regain d'intérêt envers les Tuniciers. Du point de vue du développement, le début des années 2000 fut marqué par une cascade de publications exhortant l'usage de *Ciona* et *Oikopleura* en tant que modèles d'étude pour appréhender les innovations développementales à l'origine des Vertébrés (Corbo, Di Gregorio & Levine 2001; Pennisi 2002a; Cañestro, Bassham & Postlethwait 2003; Holland & Gibson-Brown 2003; Passamanek & Di Gregorio 2005). Leur facilité d'élevage en laboratoire, leur simplicité anatomique et cellulaire, leur cycle de reproduction rapide, les avancées techniques de la manipulation génétique, s'ajoutant à la compacité génomique et l'accès aux données moléculaires, confèrent aujourd'hui aux Tuniciers un rôle central dans les études d'EvoDevo (Satoh 2003).

Au-delà d'une compacité inattendue, l'ère post-génomique dévoila également une divergence génomique surprenante des génomes de Tuniciers par rapport aux génomes de Vertébrés. Cette forte divergence se manifeste d'une part par la perte de nombreux gènes, mais aussi par l'expansion spécifique de certaines familles de gènes, et d'autre part par des remaniements géniques dramatiques liés principalement à la position mouvante des introns dans les gènes et des régions intergéniques particulièrement divergentes en termes de séquence (Holland 2007). Les paradigmes les plus éminents de cette divergence inattendue sont illustrés par des modifications des kits géniques considérés auparavant comme indispensables pour le développement des Chordés, à l'image de la désintégration du cluster de gènes Hox et l'absence de la voie de signalisation de l'acide rétinoïque chez *Oikopleura dioica*.

I. Les gènes Hox

Faisant partie des plus grandes découvertes du 20^{ème} siècle, les gènes Hox consistent en un groupe de gènes homéotiques dupliqués en tandem et codant pour des facteurs de transcription qui contrôlent le développement de l'axe antéro-postérieur (AP) chez la vaste majorité des animaux multicellulaires. Dans les génomes de bilatériens, y compris les Chordés, les gènes Hox forment généralement un « cluster », en étant physiquement agencés et ordonnés de manière à ce que l'ordre génique corresponde à la fois à l'expression séquentielle le long de l'axe antéro-postérieur (colinéarité spatiale), ainsi qu'à l'expression temporelle durant le développement embryonnaire (colinéarité temporelle).

Chez *Ciona intestinalis*, l'analyse post-génomique a démontré que le complément Hox comprend seulement neuf gènes (*Ci-Hox1-6*, *Ci-Hox10* et *Ci-Hox12-13*), parmi lesquels seuls sept se situent sur le même chromosome. L'ordre des gènes étant relativement coordonné par rapport à leur expression dans le système nerveux et l'intestin des juvéniles, la majorité des Hox chez *Ciona intestinalis* maintiennent donc partiellement la colinéarité spatiale, même si les deux autres gènes sont situés sur des chromosomes différents (Ikuta *et al* 2004).

Seo *et al.* (2004) en conduisant une marche génomique de l'ordre de 250 Kb de chaque côté des gènes Hox identifiés sur des clones BAC d'*Oikopleura dioica* ont révélé que les 9 gènes Hox ne sont pas agencés spatialement dans le génome, le cluster Hox étant fortement désintégré puisqu'ils sont situés sur des chromosomes différents. De façon surprenante, alors que le nombre de gènes Hox portés par *Oikopleura dioica* et *Ciona*

intestinalis soit le même, les deux Tuniciers présentent des compléments géniques très hétérogènes (Ikuta *et al* 2004). L'expression colinéaire étant globalement respectée malgré des déviations considérables, l'appendiculaire *Oikopleura dioica* présente le paradoxe de posséder des gènes Hox plus dispersés que tous les autres animaux examinés malgré le plus petit génome jusqu'à présent séquencé.

II. La voie de signalisation de l'acide rétinoïque

Chez les Chordés, la voie de signalisation par l'acide rétinoïque (« RA signaling ») joue un rôle central dans l'établissement de l'axe antéro-postérieur en régulant la colinéarité temporelle des gènes Hox au cours du développement (Marletaz *et al* 2006). Plus précisément, l'acide rétinoïque permet à la machinerie de transcription d'accéder aux gènes homéotiques de façon graduelle en induisant des changements dans la structure de la chromatine autour des gènes du cluster Hox (« chromatin remodelling »). Du point de vue génétique, ceci est réalisé par la transmission des signaux dans une voie métabolique impliquant des enzymes de synthèse, de dégradation, et de régulation de la distribution spatio-temporelle de l'acide rétinoïque durant l'embryogenèse. Des analyses conduites sur des mutants ont montré qu'un excès d'acide rétinoïque au cours du développement des Vertébrés et des Céphalocordés peut provoquer des irrégularités morphogénétiques majeures en déclenchant des transformations homéotiques chez les embryons (Cañestro & Postlethwait 2007).

Malgré le fait que les Appendiculaires suivent le même patron de développement que les autres Chordés, des études récentes ont mis en évidence que la formation de l'axe antéro-postérieur chez *Oikopleura dioica* est indépendante de la signalisation par l'acide rétinoïque (Cañestro, Yokoi & Postlethwait 2007). De plus, le génome d'*Oikopleura* semble avoir perdu des gènes fondamentaux de la régulation par l'acide rétinoïque, parmi lesquels les enzymes de synthèse ALDH1A et de dégradation CYP26, ainsi que le gène codant pour le récepteur RAR (Cañestro, Yokoi & Postlethwait 2007).

De plus, une étude systématique menée sur 76 gènes régulateurs du développement chez *Ciona intestinalis* a montré que bien que cette espèce comporte l'arsenal génique de la voie de signalisation par l'acide rétinoïque dans son génome, son développement embryonnaire n'est pas autant affecté par l'exposition à de hautes concentrations d'acide rétinoïque (Cañestro, Yokoi & Postlethwait 2007). Des résultats similaires ont également obtenus auparavant chez l'Ascidie *Herdmania curvata* (Hinman & Degnan 2001).

L'ensemble de ces découvertes illustre parfaitement le fait qu'au cours de l'histoire évolutive des Chordés, les Tuniciers ont subi des innovations évolutives qui les ont écartés de manière significative des modalités de développement caractéristiques des autres lignées du phylum.

CHAPITRE I :

LES TUNICIERS AU SEIN DES CHORDES

1. L'origine des Chordés : de la systématique traditionnelle à la phylogénomique

Les racines des Chordés remontent à plus de 500 millions d'années, à la période de l'explosion Cambrienne dénotant l'émergence fulgurante de la plupart des lignées majeures d'animaux multicellulaires actuels. Ayant engendré de nombreuses hypothèses et débats, l'histoire évolutive du groupe des Chordés occupa intensément la pensée scientifique pour plus de 200 ans (Swalla & Xavier-Neto 2008).

Durant le 19^{ème} siècle, à l'aube de la systématique traditionnelle, Geoffroy Saint-Hilaire (1822), Dohrn (1875) et Semper (1875) proposèrent les Articulata, réunissant les Annélides et les Arthropodes, comme la lignée ancestrale menant aux Vertébrés. Patten (1890) et Gaskell (1890) confièrent ensuite la descendance des Vertébrés aux Arachnides, tandis que Hubrecht (1883), et plus tard Jensen (1960), invoquèrent les Némertiens aux racines évolutives du groupe (cités dans (Stach 2008)). Des théories encore plus invraisemblables, voire extrêmes, peuvent également être retracées comme l'exemple de l'hypothèse de l'émergence du plan d'organisation des Vertébrés suite à des épisodes de fusion des différentes espèces de salpes (Lacalli 1999).

Après les remarquables travaux de Kowalevsky (1868) sur l'embryogenèse comparée des Ascidies, la majorité de ces scénarios fut aussitôt rejetée en faveur d'une nouvelle théorie avancée par Garstang en 1928 puis modifiée ultérieurement par Berill (1955) (Winchell *et al* 2002; Mallatt & Winchell 2007). Garstang proposa que les Chordés soient issus d'un ancêtre sessile et filtreur portant des tentacules, dont la perte au cours de l'évolution permit l'apparition de la notochorde à l'état larvaire et des fentes pharyngiennes au stade adulte. Dès lors, les Vertébrés seraient apparus à la faveur d'une série d'événements de néoténie ou paedomorphose ayant eu lieu au sein des Ascidies ancestrales, débouchant sur la rétention des caractères larvaires à l'état adulte via l'atteinte précoce de la maturité sexuelle (Garstang 1928).

Se focalisant sur les Appendiculaires et en se référant de manière générale aux Tuniciers pélagiques, la vision de Garstang fut très influente à l'époque en plaçant les Tuniciers dans une position basale au sein des Chordés en tant que Protochordés, *i.e.* formes de Chordés « primitifs ». La vision traditionnelle de l'évolution des Chordés commença ainsi à s'édifier, rapprochant les Vertébrés avec les Céphalochordés dans un groupe monophylétique reconnu sous le nom d'Euchordata (littéralement « Chordés vrais »).

Avec l'avènement de la phylogénie moléculaire dans les années 1990, la question de l'origine des Chordés connut un net regain d'intérêt. Sur la base du gène de l'ARN ribosomique 18S (ARNr 18S) codant pour la petite sous-unité du ribosome, l'image des relations de parenté au sein des Deutérostomiens fut très tôt remodelée. Les Echinodermes (ex. oursins, étoiles de mer) et les Hémichordés (Entéropeustes et Ptérobranches) furent regroupés dans un clade monophylétique (les Ambulacraria) au sein des Deutérostomes, alors que les Chaetognathes et Pogonophores furent déplacés au sein des Protostomiens (Turbeville, Schulz & Raff 1994; Wada & Satoh 1994). Les premiers arbres phylogénétiques incluant des Tuniciers proposèrent une étroite affinité phylogénétique entre Vertébrés et Céphalocordés à l'exclusion des Tuniciers, cependant la question de la monophylie des Chordés resta controversée (Cameron, Garey & Swalla 2000; Swalla *et al* 2000).

L'analyse combinée des ARNr 18S et 28S codant respectivement pour la petite (SSU) et la grande sous-unité (LSU) du ribosome renforça un peu plus l'hypothèse des Euchordés (Winchell *et al* 2002; Mallatt & Winchell 2007). L'origine unique des Chordés demeurant dans un premier temps irrésolue, les auteurs attribuèrent la manque de soutien statistique pour la monophylie du phylum à des artéfacts de reconstruction dus au taux d'évolution remarquablement élevé caractérisant les séquences d'ARN ribosomique de Tuniciers (Swalla *et al* 2000).

L'ère post-génomique permis l'émergence de la phylogénomique, définie comme la reconstruction phylogénétique à partir de données moléculaires à large échelle (Delsuc, Brinkmann & Philippe 2005). Ce contexte motiva des études qui entreprirent la quête de l'histoire évolutive des Bilatériens à la lueur des analyses portant sur de nombreux gènes, issus à la fois des génomes séquencés, ainsi que des projets de séquençage d'ESTs (Expressed Sequence Tags) (Philippe, Lartillot & Brinkmann 2005). Malgré le faible échantillonnage taxonomique des Chordés dans ces études, des indices furent pour la première fois fournis pour la remise en cause du clade des Euchordés (Philippe, Lartillot & Brinkmann 2005).

Une étude s'appuyant sur 146 gènes conservés à l'échelle des animaux Bilatériens et incluant treize représentants du groupe des Chordés dont quatre Tuniciers fut menée l'année suivante (Delsuc *et al* 2006). Contre toute attente, les analyses phylogénomiques dévoilèrent l'étroite affinité évolutive entre Vertébrés et Tuniciers, en fournissant pour la première fois un soutien statistique solide pour l'apparement des deux lignées dans un clade monophylétique nommé Olfactores (Fig. 9).

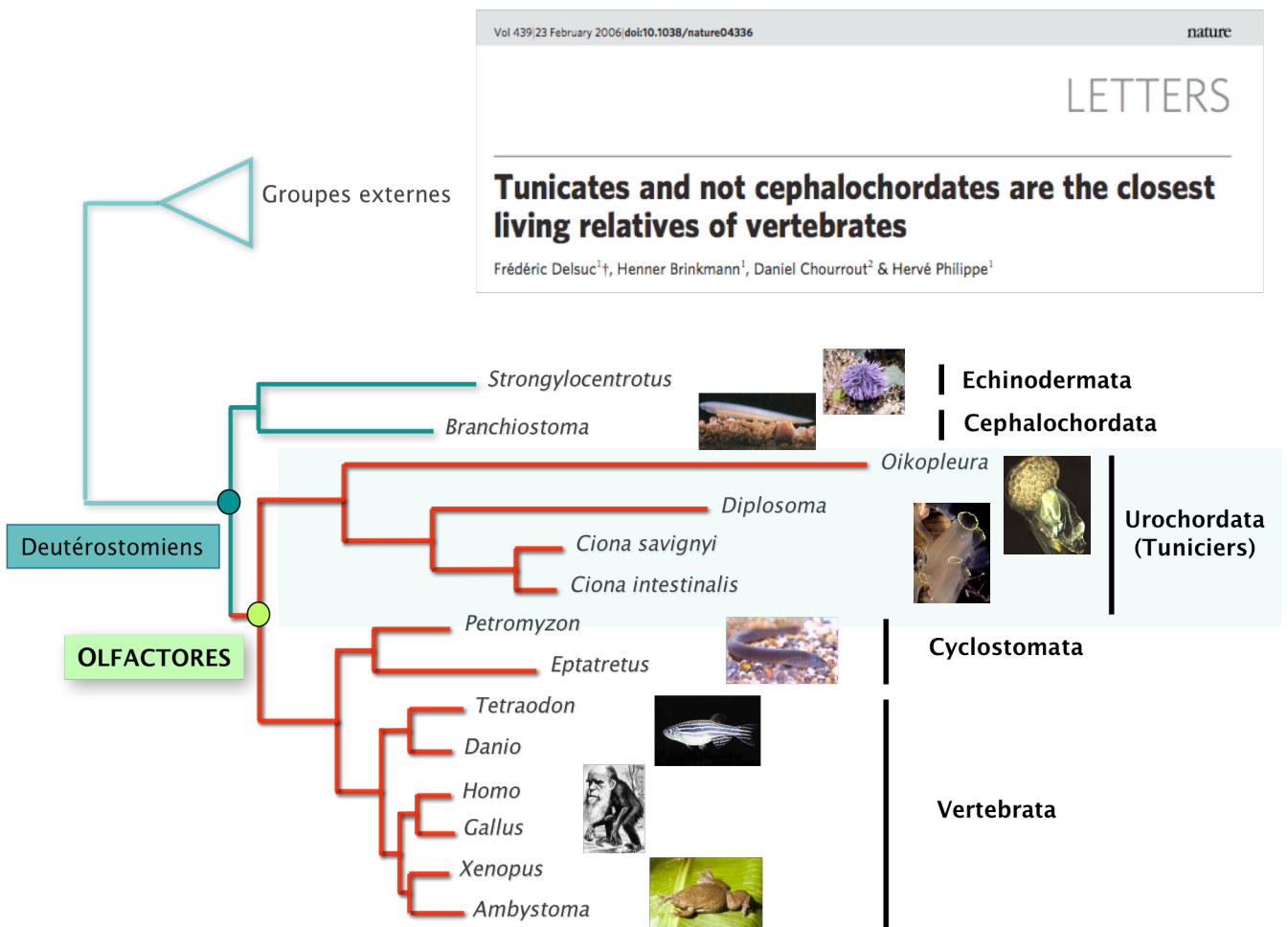


Figure 9 : Illustration de l'hypothèse Olfactores.

Arbre de maximum de vraisemblance obtenu à partir de l'analyse d'un jeu de données phylogénomique comprenant 146 gènes (33800 acides aminés) pour 38 taxa (modèle : WAG + Γ). Le regroupement des Tuniciers avec les Vertébrés est soutenu par de très fortes valeurs de soutien statistique (BP = 100, PP = 1.0). (Modifiée après Delsuc *et al*, 2006)

En dépit de cela, la monophylie des Chordés ne trouva pas de soutien dans cette étude, puisque curieusement le représentant des Céphalochordés (l'amphioxus *Branchiostoma*) apparut en tant que groupe-frère du représentant des Echinodermes (l'oursin *Strongylocentrotus*) (Delsuc *et al* 2006). Bien que le soutien statistique pour cette dernière relation de parenté fût relativement modéré, ce point particulier, associé à la reconnaissance des Olfactores allant contre le dogme général des Euchordés, contribua à susciter la controverse.

2. Les débats autour de l'hypothèse Olfactores

Issu de la théorie des Calcichordata sur l'origine des Vertébrés (Jefferies 1997), le nom Olfactores fut introduit par Richard Jefferies dès les années 1990 (Jefferies 1991). Ce nom fait référence à l'hypothèse d'homologie de l'appareil olfactif chez les Carpoïdes qui représente une partie du registre fossile considérée par l'auteur comme membres de la lignée ancestrale menant aux Tuniciers et aux Crâniates (Jefferies 1986). La théorie des Calcichordés étant très controversée (Peterson 1995), l'hypothèse Olfactores suggérant une proche parenté entre Tuniciers et Vertébrés trouva bien peu de supporters. Ainsi de manière analogue, sa résurrection par la phylogénomique (Delsuc *et al* 2006) fut accueillie avec un certain scepticisme.

La source principale de ce scepticisme réside dans le manque de soutien pour la nouvelle phylogénie des Chordés au sens de la systématique traditionnelle. En effet, les caractères anatomiques sont traditionnellement interprétés comme étant en faveur des Euchordés sur la base du caractère dérivé partagé que constituerait la présence de somites chez les Céphalochordés et les Craniates (Rowe 2004). À l'inverse les caractères partagés entre Tuniciers et Vertébrés à l'exclusion des Céphalochordés sont extrêmement limités. Il n'existe qu'un seul trait morpho-anatomique unanimement accepté comme étant propre aux deux groupes : la présence des cellules migratrices de la crête neurale, d'abord identifiées chez les larves géantes de l'ascidie *Ecteinascidia turbinata* (Jeffery, Strickler & Yamamoto 2004) puis chez de nombreuses autres espèces de Tuniciers (Jeffery 2007). Dans le cadre d'une étude portant sur les relations de parenté des Hémichordés au sein Deutérostomiens les Chordés, Ruppert proposa néanmoins six synapomorphies potentielles pour le clade d'Olfactores (Ruppert 2005). Parmi celles-ci figurent le remplacement des jonctions septées par des jonctions étroites, des similarités quant à la musculature et aux structures néphridiales, ainsi que l'origine ecto- et endodermique de l'adénohypophyse.

Une deuxième source de controverse pris naissance dans la position non orthodoxe des Céphalochordés qui étaient regroupés avec les Echinodermes dans les premiers arbres phylogénomiques (Delsuc *et al* 2006) (Fig. 9). Bien que cette hypothèse alternative ne soit pas statistiquement soutenue et basée sur un faible échantillonnage taxonomique (Delsuc *et al* 2006), la phylogénomique semblait donc contredire l'origine unique des Chordés, elle aussi considérée comme incontestable, puisque les trois lignées du phylum partagent des nombreuses synapomorphies dont certaines constituent des innovations morphologiques telles que la notocorde et la chorde nerveuse dorsale (Rowe 2004).

Une deuxième étude phylogénomique, incluant notamment un Hémichordé et un Echinoderme supplémentaire fut conduite la même année (Bourlat *et al* 2006). La question centrale concernait le positionnement de l'énigmatique taxon *Xenoturbella bocki* (Xenoturbellides) au sein de la phylogénie des Deutérostomiens. L'analyse de 170 gènes nucléaires avec ce nouvel échantillonnage retrouva la monophylie des Chordés, démontrant ainsi que la proximité phylogénétique obtenue précédemment entre l'amphioxus et l'oursin (Delsuc *et al* 2006) était vraisemblablement artéfactuelle à cause d'un échantillonnage trop limité du groupe des Ambulacraria (Echinodermes + Hémichordés) (Bourlat *et al* 2006). Les résultats de cette nouvelle étude vinrent néanmoins corroborer le clade Olfactores qui apparut toujours aussi solidement soutenu par les analyses (Bourlat *et al* 2006).

Ces travaux de phylogénomique remirent ainsi en cause la prévalence de l'hypothèse des Euchordata qui figure dans tous les traités zoologiques du dernier siècle et dans les manuels actuels. Ils invalident ainsi la vision de l'histoire évolutive des animaux deutérostomiens comme une marche progressive vers la complexité morphologique sous-tendant l'hypothèse traditionnelle de Garstang avec acquisition successive des fentes branchiales, de la chorde, puis des somites au travers d'une suite d'évènement de paedomorphose (Garstang 1928). Au contraire, la nouvelle phylogénie des deutérostomiens révélée par la phylogénomique suggère l'hypothèse que l'ancêtre des Deutérostomiens ait pu posséder des fentes branchiales et des somites qui auraient été perdues secondairement chez les Echinodermes et les Tuniciers, respectivement (Fig. 10).

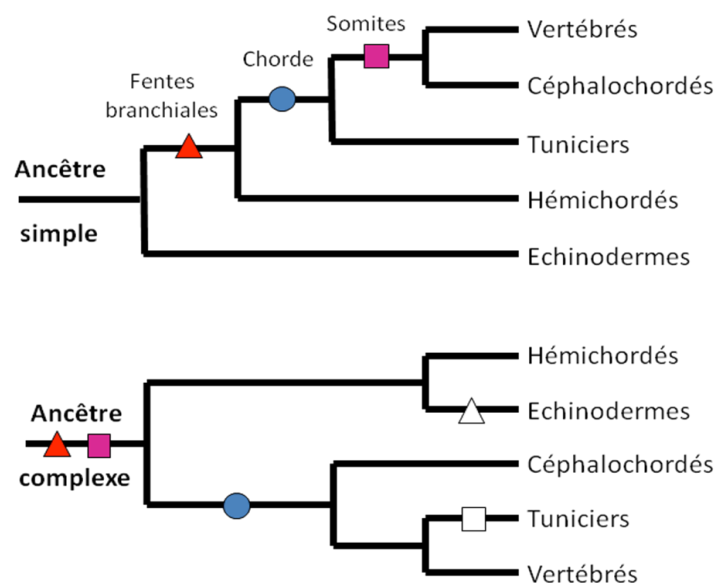


Figure 10 : Evolution des Deutérostomiens : du complexe au simple ?

3. Corroboration de la nouvelle phylogénie des Chordés [Article 1]

Compte tenu du scepticisme affiché par rapport au positionnement des Tuniciers au sein des Chordés, un projet fut initié en 2008 afin d'évaluer la nouvelle phylogénie à la lumière des différentes données moléculaires disponibles. L'objectif se porta sur plusieurs axes, le premier consista à re-analyser le jeu de données phylogénomiques assemblé en 2006 (Delsuc *et al* 2006) à la lueur du modèle d'évolution des séquences CAT développé par (Lartillot & Philippe 2004; Lartillot, Brinkmann & Philippe 2007) et désormais implémenté dans le logiciel PhyloBayes (Lartillot, Lepage & Blanquart 2009). Ce modèle de mélange en permettant de mieux estimer la distribution de l'hétérogénéité substitutionnelle dans les séquences offre une meilleure adéquation que les modèles classiques et permet ainsi de réduire certains artefacts de reconstruction tels que l'artéfact d'attraction des longues branches (Lartillot, Brinkmann & Philippe 2007).

En outre, un jeu des données étoffé en termes de gènes et de taxons a été assemblé en bénéficiant de la croissance des données génomiques disponibles depuis trois ans. Cette mise à jour a permis d'inclure 179 marqueurs phylogénétiques pour 51 Métazoaires, y compris 25 Deutérostomiens dont six tuniciers (Delsuc *et al* 2008). Cette supermatrice totalise 53 799 de sites d'acides aminés dont l'alignement est non-ambigu avec un pourcentage de données manquantes de l'ordre de 32%. Ce nouveau jeu de données phylogénomique a également été analysé sous le modèle CAT en testant la robustesse des topologies obtenues à la quantité de données manquantes et à l'échantillonnage génique par une par une méthode de Jackknife de gènes. Enfin, le signal phylogénétique porté par un jeu de données combinant les ARNr 18S et 28S (Mallatt & Winchell 2007) a également été étudié en détails avec une attention particulière portée aux biais de composition nucléotidiques.

Les résultats obtenus sur le jeu de données de 2006 avec l'utilisation du modèle CAT retrouvèrent la monophylie des Chordés, démontrant ainsi que la proche parenté observée entre Echinodermes et Céphalocordés résultait probablement d'une inadéquation du modèle couplée à un faible échantillonnage taxonomique des Ambulacraria. Par contre le fort soutien statistique obtenu pour les Olfactores ne fut pas affecté. L'analyse du nouveau jeu de données phylogénomique corrobora la monophylie des Chordés et l'étroite parenté entre Tuniciers et Vertébrés, des résultats qui apparurent également très robustes à l'échantillonnage taxonomique et au choix des gènes utilisés. Ces résultats renforcèrent la crédibilité de la nouvelle phylogénie des Chordés et Deuterostomiens.

Finalement, la re-évaluation des données ribosomiques (ARNr 18S et 28S) suggéra que le fort soutien accordé auparavant à l'hypothèse des Euchordés (Céphalocordés + Vertébrés) résidait sans doute dans l'hétérogénéité compositionnelle marquant les séquences des Chordés pour ces gènes. En effet, les Tuniciers présentant une proportion en AT élevée par rapport aux Vertébrés et à l'Amphioxus dont la composition est plus riche en GC. Cette hypothèse se renforça d'autant plus que les arbres reconstruits après un recodage des données en RY qui permet d'atténuer les biais compositionnels, retrouvent une topologie compatible avec l'hypothèse Olfactores où les Tuniciers apparaissent en tant que groupe-frère des Vertébrés malgré de faibles valeurs de soutien statistique.

ARTICLE I

ARTICLE PUBLIE DANS *GENESIS*

ADDITIONAL MOLECULAR SUPPORT FOR THE NEW CHORDATE PHYLOGENY

FRÉDÉRIC DELSUC^{1,2}, **GEORGIA TSAGKOGEOGA**^{1,2}, NICOLAS LARTILLOT³
& HERVE PHILIPPE³

¹*Université Montpellier II, CC064, Place Eugène Bataillon, Montpellier Cedex 5, France*

²*CNRS, Institut des Sciences de l'Evolution (UMR5554), CC064, Place Eugène Bataillon,
Montpellier Cedex 5, France*

³*Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal,
Québec, Canada*

Additional Molecular Support for the New Chordate Phylogeny

Frédéric Delsuc,^{1,2*} Georgia Tsagkogeorga,^{1,2} Nicolas Lartillot,³ and Hervé Philippe³

¹Université Montpellier II, CC064, Place Eugène Bataillon, Montpellier Cedex 5, France

²CNRS, Institut des Sciences de l'Évolution (UMR5554), CC064, Place Eugène Bataillon, Montpellier Cedex 5, France

³Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal, Québec, Canada

Received 6 May 2008; Revised 16 September 2008; Accepted 22 September 2008

Summary: Recent phylogenomic analyses have suggested tunicates instead of cephalochordates as the closest living relatives of vertebrates. In direct contradiction with the long accepted view of Euchordates, this new phylogenetic hypothesis for chordate evolution has been the object of some skepticism. We assembled an expanded phylogenomic dataset focused on deuterostomes. Maximum-likelihood using standard models and Bayesian phylogenetic analyses using the CAT site-heterogeneous mixture model of amino-acid replacement both provided unequivocal support for the sister-group relationship between tunicates and vertebrates (Olfactores). Chordates were recovered as monophyletic with cephalochordates as the most basal lineage. These results were robust to both gene sampling and missing data. New analyses of ribosomal rRNA also recovered Olfactores when compositional bias was alleviated. Despite the inclusion of 25 taxa representing all major lineages, the monophyly of deuterostomes remained poorly supported. The implications of these phylogenetic results for interpreting chordate evolution are discussed in light of recent advances from evolutionary developmental biology and genomics. *genesis* 46:592–604, 2008. © 2008 Wiley-Liss, Inc.

Key words: phylogenomics; deuterostomes; chordates; tunicates; cephalochordates; olfactores; ribosomal RNA; jackknife; evolution

INTRODUCTION

Besides its fundamental role in systematics, phylogenetic reconstruction is a prerequisite for understanding the evolution of organisms. The essential contribution of phylogenetics for understanding morphological diversity has perhaps been best exemplified in the case of animal evolution (Telford and Budd, 2003). The Cambrian explosion has produced a bewildering diversity of body plans whose origins and evolution can only be apprehended by undertaking an integrative approach through evolutionary developmental biology (Evo-Devo) (Conway-Morris, 2003). The knowledge of phylogenetic relationships, by allowing the polarization of character

transformations, sheds light on the extent of morphological convergence and reversal. A phylogenetic framework is therefore required for distinguishing ancestral characters from those representing morphological innovations. Comparative genomics is now providing the opportunity to track these morphological innovations back to the molecular level by revealing the patterns of gene acquisition/loss and giving clues to the molecular adaptations that underline the evolution of body plans (Cañestro *et al.*, 2007).

Animal taxonomy has deep roots. The study of morphological and embryological characters has allowed the definition of the major phyla but left their interrelationships almost unresolved (Nielsen, 2001). The advent of molecular data during the 1990s has revolutionized the traditional classification through a series of phylogenetic analyses of the 18S ribosomal RNA (rRNA) gene for an ever increasing number of key taxa (Aguinaldo *et al.*, 1997; Halanych *et al.*, 1995). This period culminated with the proposition of a new view of animal phylogeny at odds with the traditional paradigm of a steady increase toward morphological complexity, and revealing instead the major role played by secondary simplification from complex ancestors (Adoutte *et al.*, 2000; Lwoff, 1944). Despite these undeniable achievements, the resolving power provided by 18S rRNA and other single genes is nevertheless limited, and a number of open questions in animal phylogeny remained to be answered (Halanych, 2004).

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: Frédéric Delsuc, CC064, Institut des Sciences de l'Évolution, UMR5554-CNRS, Université Montpellier II, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France. E-mail: frederic.delsuc@univ-montp2.fr

Contract grant sponsor: Contribution ISEM 2008-062 of the Institut des Sciences de l'Évolution (UMR5554-CNRS); Contract grant sponsor: Research Networks Program in Bioinformatics from the High Council for Scientific and Technological Cooperation between France and Israel

Published online 10 November 2008 in

Wiley InterScience (www.interscience.wiley.com).

DOI: 10.1002/dvg.20450

The most recent advances in animal phylogeny have come from phylogenomics (Delsuc *et al.*, 2005), which considerably increases the resolving power by considering numerous concatenated genes from expressed sequence tags (ESTs) and complete genome projects (Philippe and Telford, 2006). Despite some troubled beginnings due to the shortcomings of using only a restricted set of taxa (Philippe *et al.*, 2005a), phylogenomics has provided strong corroborating support for the new animal phylogeny, essentially confirming the monophyly of Protostomia, Ecdysozoa, and Lophotrochozoa (Baurain *et al.*, 2007; Dunn *et al.*, 2008; Lartillot and Philippe, 2008; Philippe *et al.*, 2005b). Phylogenomics has also helped solving some longstanding mysteries such as the position of chaetognaths, which finally appear to belong to Protostomia (Marletaz *et al.*, 2006; Matus *et al.*, 2006) and also proposed unexpected phylogenetic affinities for enigmatic taxa such as *Buddenbrockia plumatellae* recently unmasked as a cnidarian worm (Jimenez-Guri *et al.*, 2007), or *Xenoturbella bocki*, representing a fourth deuterostome phylum on its own (Bourlat *et al.*, 2006).

Among the most groundbreaking results from recent phylogenomic studies was the identification of tunicates (or urochordates) as the closest living relatives of vertebrates, instead of cephalochordates as traditionally accepted (Delsuc *et al.*, 2006). Some hints of this unexpected result had been observed in previous large-scale phylogenetic studies including a single tunicate representative (Blair and Hedges, 2005; Philippe *et al.*, 2005b; Vienne and Pontarotti, 2006). However, a substantial increase in taxon sampling turned out to be required for recovering convincing support in favor of such an unorthodox relationship. In particular, the fact that the inclusion of the divergent appendicularian tunicate *Oikopleura dioica* did not disrupt the sister-group relationship between tunicate and vertebrates gave a good indication about the strength of the phylogenetic signal in its favor (Delsuc *et al.*, 2006). The grouping of tunicates and vertebrates had already been proposed on morphological grounds by Richard P.S. Jefferies who coined the name Olfactores after the presence a putatively homologous olfactory apparatus in fossils that were proposed to be precursors of tunicates and vertebrates (Jefferies, 1991). This phylogenetic result has nevertheless been the object of some skepticism. One reason for this maybe that it further invalidates the traditional textbook view of chordate evolution as a steady increase toward morphological complexity culminating with vertebrates, as betrayed by the use of the term Euchordates (literally “true chordates”) for denoting the grouping of cephalochordates and vertebrates (Gee, 2001). The lack of obvious morphological synapomorphies for Olfactores, apart from the presence of migratory neural crest-like cells (Jeffery, 2007; Jeffery *et al.*, 2004), and the apparent conflict with analyses of rRNA data which tend to favor Euchordates (Cameron *et al.*, 2000; Mallatt and Winchell, 2007; Winchell *et al.*, 2002) might also partly

explain the caution with which this result has been considered at first.

Phylogenomics, despite being a powerful approach, is however not immune to potential reconstruction artifacts. The possible pitfalls associated with phylogenomic studies include systematic errors that can be traced back to some kind of model misspecifications (Philippe *et al.*, 2005a) and caused mainly by heterogeneity of evolutionary rates among taxa (Lartillot *et al.*, 2007; Philippe *et al.*, 2005b) and compositional bias (Blanquart and Lartillot, 2008; Jeffroy *et al.*, 2006; Lartillot and Philippe, 2008; Phillips *et al.*, 2004). Empirical protocols have been designed to detect and reduce the impact of systematic error in genome-scale studies (Rodríguez-Ezpeleta *et al.*, 2007) but the ultimate solution lies in the development of improved models of sequence evolution (Felsenstein, 2004; Philippe *et al.*, 2005a; Steel, 2005). The reliance of current phylogenomic studies on a relatively limited number of highly expressed genes (Philippe and Telford, 2006) and the potential impact of missing data on phylogenomic inference (Hartmann and Vision, 2008; Philippe *et al.*, 2004) are also regularly cited as limitations of the phylogenomic approach.

The aim of this article is to evaluate the current evidence for the new chordate phylogeny by: (1) reanalyzing previous phylogenomic data using improved models of amino-acid replacement, (2) assembling and analyzing an updated phylogenomic dataset with more genes and more taxa, (3) assessing the impact of missing data and gene sampling on phylogenomic results, and (4) performing new analyses of rRNA data taking compositional bias into account.

MATERIALS AND METHODS

Phylogenomic Dataset Assembly

We built upon previous phylogenomic datasets assembled in the Philippe lab (Delsuc *et al.*, 2006; Jimenez-Guri *et al.*, 2007; Lartillot and Philippe, 2008; Philippe *et al.*, 2004, 2005b) to select a set of 179 orthologous markers showing sufficient conservation across metazoans to be useful for inferring the phylogeny of metazoans. Alignments were built and updated with available sequences downloaded from the Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/>) and the EST Database (<http://www.ncbi.nlm.nih.gov/dbEST/>) of GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>) using the program ED from the MUST package (Philippe, 1993). Unambiguously aligned regions were identified and excluded for each individual gene using the program GBLOCKS (Castresana, 2000) with a few manual refinements using NET from the MUST package. The complete list of genes with corresponding final numbers of amino-acid sites is available as Supporting Information.

The concatenation of the 179 genes was constructed with the program SCAFOS (Roure *et al.*, 2007) by defining 51 metazoan operational taxonomic units (OTUs)

including 25 deuterostomes representing all major lineages. When several sequences were available for a given OTU, the slowest evolving one was selected according to their degree of divergence using ML distances computed by TREE-PUZZLE (Schmidt *et al.*, 2002) under a WAG+F model (Whelan and Goldman, 2001) within SCAFOS. The percentage of missing data per taxon was reduced by creating some chimerical sequences for species belonging to the same OTU. The complete alignment consists of 179 genes and 51 taxa for 53,799 unambiguously aligned amino-acid sites with 32% missing data. To study the potential impact of missing data on phylogenetic inference (Hartmann and Vision, 2008; Philippe *et al.*, 2004; Wiens, 2006), a concatenation of the 106 genes with sequences available for at least 41 of the 51 OTUs was also constructed with SCAFOS. This reduced alignment consists of 106 genes and 51 taxa for 25,321 amino-acid sites and contains only 20% of missing data. The list of defined OTUs, chimerical sequences and percentages of missing data are available as Supporting Information. Individual gene alignments and their concatenations are available upon request.

Phylogenomic Analyses

Bayesian phylogenetic analyses of the two phylogenomic datasets were conducted using the program PHYLOBAYES 2.3c (<http://www.atgc-montpellier.fr/phylobayes/>) under the CAT+ Γ_4 site-heterogeneous mixture model (Lartillot and Philippe, 2004). For each dataset, four independent Monte Carlo Markov Chains (MCMCs) starting from a random topology were run in parallel for 20,000 cycles (1,500,000 generations), saving a point every cycle, and discarding the first 2,000 points as the burnin. Bayesian posterior probabilities (PP) were obtained from the 50% majority-rule consensus tree of the 18,000 MCMC sampled trees using the program READPB of PHYLOBAYES.

Maximum likelihood (ML) reconstruction on the new phylogenomic dataset was also performed using the program TREEFINDER version of March 2008 (Jobb *et al.*, 2004) under the empirical WAG+F+ Γ_8 model of amino-acid substitution. The α shape parameter of the Γ distribution was estimated along with the topology and the branch lengths. Reliability of nodes was estimated by bootstrap resampling with 100 pseudo-replicate datasets generated by the program SEQBOOT of the PHYLIP package (Felsenstein, 2001). The 100 corresponding ML heuristic searches were run in parallel on a computing cluster and the majority-rule consensus of the 100 resulting trees was computed using TREEFINDER.

Jackknife Procedure

The robustness of our phylogenomic inference with respect to gene sampling was assessed by applying a jackknife procedure. Fifty jackknife replicates of 50 genes drawn randomly from the full pool of 179 genes were generated. The only condition we imposed to this jackknife procedure was to require that each taxon is

represented by at least one gene in each replicate. The 50 jackknife supermatrices ranging from 11,163 to 17,181 amino-acid sites with 27 to 35% missing data were then analyzed using PHYLOBAYES under the CAT+ Γ_4 model. To ensure correct convergence of the MCMC on each replicate, an automated stopping rule was used. Specifically, for each jackknife replicate, two independent parallel (and synchronous) MCMC were run, until the posterior probability discrepancy between the two chains was less than 0.15 (maximum discrepancy over all bipartitions), and after removing the first 1,000 sampled trees of each chain as the burnin. A global majority-rule consensus tree was obtained from the 50 replicates as follows: for each jackknife replicate (D_r) taken in turn, we computed the frequency-table of all bipartitions (splits) observed in the sample collected from the posterior distribution $p(T|D_r)$. The frequencies associated to each bipartition were then averaged over the 50 replicates, and the resulting frequency table was used to build a consensus tree. The support values displayed by this Bayesian consensus tree are thus jackknife-resampled posterior probabilities (PP_{JK}). High PP_{JK} values indicate nodes that have high posterior probability support in most jackknife replicates.

Phylogenetic Analyses of Ribosomal RNA

The 46-taxa dataset of combined 18S+28S rRNA genes assembled by Mallatt and Winchell (2007) for studying deuterostome phylogeny was reanalyzed. This alignment contains a total of 3,925 unambiguously aligned nucleotide sites. A principal component analysis (PCA) of nucleotide composition was realized using the R statistical package (R Development Core Team, 2007). The best fitting model of nucleotide sequence evolution was evaluated using MODELTEST 3.7 (Posada and Crandall, 1998). The TIM+ Γ_4 +I transitional model (Posada and Crandall, 2001) was selected according to the Akaike information criterion. ML phylogenetic analysis of this nucleotide dataset was conducted with PAUP* 4.0b10 (Swofford, 2002) using a heuristic search with Tree Bisection Reconnection (TBR) branch swapping starting from a Neighbor-Joining (NJ) tree.

The nucleotide dataset was RY-coded by pooling puRines (AG = R) and pYrimidines (CT = Y) in an attempt to alleviate both compositional heterogeneity and substitutional saturation of transition events. This RY-coded dataset was then analyzed by conducting a ML heuristic search with TBR branch swapping on a NJ starting tree using PAUP* under the CF+ Γ_8 model for discrete characters (Cavender and Felsenstein, 1987). The α shape parameter of the Γ distribution was previously estimated during a ML heuristic search on the nucleotide dataset conducted with TREEFINDER under the GTR2+ Γ_8 two-state model.

Reliability of nodes was estimated for each dataset by nonparametric bootstrap resampling using 100 pseudo-replicates generated by SEQBOOT. The 100 corresponding ML heuristic searches using PAUP* with the previ-

ously estimated ML parameters, NJ starting trees, and TBR branch swapping were parallelized on a computing cluster. ML bootstrap percentages were obtained from the 50% majority-rule consensus tree of the 100 bootstrap ML trees using TREEFINDER.

RESULTS AND DISCUSSION

Effect of an Improved Model of Sequence Evolution

Our initial assessment of deuterostome phylogenetic relationships was based on a phylogenomic dataset encompassing 146 nuclear genes (33,800 amino-acids) from 38 taxa including 14 deuterostomes (Delsuc *et al.*, 2006). ML and Bayesian phylogenetic analyses conducted under the standard WAG+F+ Γ_4 model provided strong support for grouping tunicates with vertebrates (including cyclostomes), but also disrupted chordate monophyly because cephalochordates grouped with echinoderms, albeit with nonsignificant statistical support (Delsuc *et al.*, 2006). The limited taxon sampling available at the time for Ambulacraria (echinoderms and hemichordates), that is, a single echinoderm, prompted us to be cautious about this result and to call for the inclusion of xenoturbellidans, hemichordates, and more echinoderms before drawing definitive conclusions. In fact, a subsequent phylogenomic study did exactly what we pleaded for by adding a representative species for each of these three groups (Bourlat *et al.*, 2006). The inclusion of these taxa allowed retrieving the monophyly of chordates in Bayesian analyses using standard amino-acid models, although the alternative hypothesis of chordate paraphyly was still not statistically rejected by ML nonparametric tests (Bourlat *et al.*, 2006). Importantly, the strong statistical support for the monophyly of Olfactores was unaffected by taxon addition (Bourlat *et al.*, 2006).

Models accounting for site-specific modulations of the amino-acid replacement process, such as the CAT mixture model (Lartillot and Philippe, 2004), seem to offer a significantly better fit to real data than empirical substitution matrices currently used in standard models of amino-acid sequence evolution. Accounting for site-specific amino-acid propensities has also been shown to induce a significant improvement of phylogenetic reconstruction in difficult cases such as long-branch attraction (Baurain *et al.*, 2007; Lartillot *et al.*, 2007; Lartillot and Philippe, 2008). This improvement essentially lays in the ability of the CAT model to detect multiple conservative substitutions more efficiently than standard amino-acid models (Lartillot *et al.*, 2007).

To test for an eventual effect of model misspecification on previous phylogenomic analyses, we reanalyzed our previous dataset (Delsuc *et al.*, 2006) under the CAT+ Γ_4 model. This analysis provides strong corroborating support for the grouping of tunicates and vertebrates (see Fig. 1). However, in contrast with previous analyses using empirical amino-acid replacement matrices,

which favored a sister-group relationship between cephalochordates and echinoderms, the use of the CAT+ Γ_4 mixture model strongly supports the classical view of monophyletic chordates and deuterostomes (see Fig. 1). The fact that chordate polyphyly is disrupted both by a richer taxon sampling (Bourlat *et al.*, 2006), or upon the use of a more elaborate model, suggests that the previously observed grouping of cephalochordates and echinoderms (Delsuc *et al.*, 2006) was probably a phylogenetic reconstruction artifact. On the other hand, the fact that the grouping of tunicates and vertebrates is insensitive to the model used, adds further credence to the Olfactores hypothesis.

An Updated Phylogenomic Dataset

The continuously growing genomic databases allowed us to build an updated phylogenomic dataset that includes both more genes and more taxa than previously considered to address the question of deuterostome phylogeny. This new dataset of 179 genes for 51 taxa includes 25 deuterostomes representing all major lineages: Xenoturbellida (1 taxon), Hemichordata (1), Echinodermata (5), Cephalochordata (1), Tunicata (6), Cyclostomata (2) and Vertebrata (9), plus 26 selected slow evolving metazoan taxa including Cnidarians and Poriferans as the most distant outgroups. Chordates are particularly well sampled with the inclusion, for the first time, of six tunicate species covering the four major clades evidenced by 18S rRNA studies (Swalla *et al.*, 2000). This diverse taxon sampling is essential to further test the new chordate phylogeny recently revealed by phylogenomics (Bourlat *et al.*, 2006; Delsuc *et al.*, 2006).

Bayesian (CAT+ Γ_4) and ML (WAG+F+ Γ_8) phylogenetic reconstructions conducted on this updated dataset (179 genes, 53,799 amino-acid sites, 51 taxa) resulted in a highly resolved tree (Fig. 2a). These analyses provided strong support for Ambulacraria (PP_{CAT+ Γ_4} = 1.0/BP_{WAG+F+ Γ_8} = 97), chordates (1.0/69) and olfactores (1.0/100). Xenambulacraria (*Xenoturbella* + Ambulacraria) and a basal position for the chaetognath *Spadella* among Protostomia were also moderately supported by our analyses (Fig. 2a). These results are compatible with a recent phylogenomic analysis, which also found strong support for Ambulacraria, chordates, and Olfactores when using the CAT mixture model (Dunn *et al.*, 2008). However, the monophyly of Deuterostomes is unresolved in both Bayesian and ML phylogenetic reconstructions (Fig. 2a).

The complete dataset obtained by concatenating all 179 genes contains 32% missing data. Previous studies of the impact of missing data on the accuracy of phylogenetic inference have concluded that probabilistic methods are relatively tolerant to missing data (Hartmann and Vision, 2008; Philippe *et al.*, 2004; Wiens, 2003, 2005), the most important factor being the absolute amount of available data for a given taxon. In phylogenomics, even incomplete taxa are usually represented by thousand of sites, and the impact of missing data on accuracy is

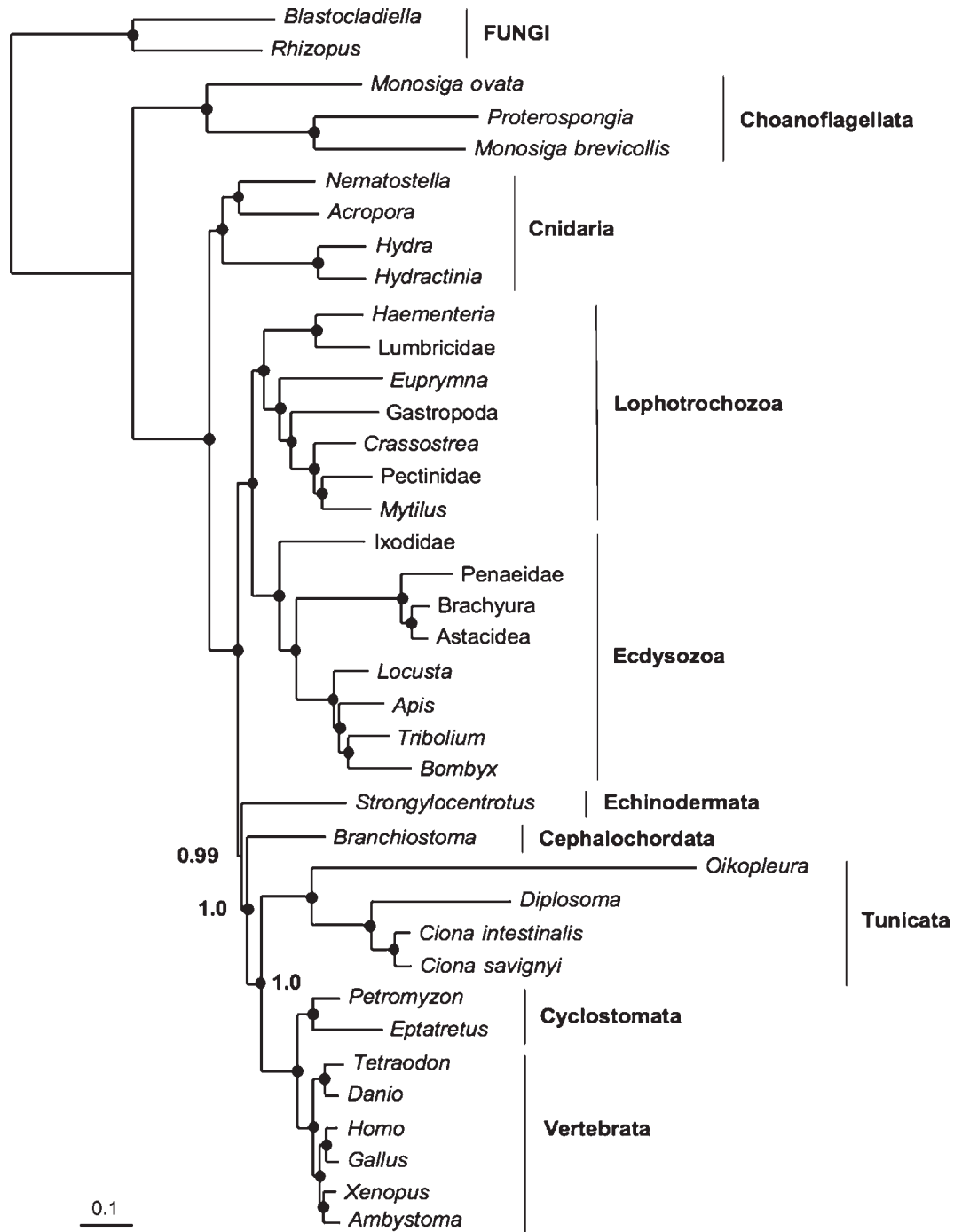


FIG. 1. Reanalysis of previous phylogenomic data using an improved model of sequence evolution. The Delsuc *et al.* (2006) phylogenomic dataset of 146 genes (38 taxa and 33,800 sites) was analyzed under the CAT+ Γ_4 site-heterogeneous mixture model of amino-acid replacement. Values at nodes represent Bayesian posterior probabilities (PP). Circles indicate nodes with maximal support PP = 1.0. The scale bar represents the estimated number of substitutions per site.

therefore relatively limited (Philippe *et al.*, 2004). Nonetheless, incomplete taxa might still be difficult to place with confidence especially when they represent isolated lineages such as *Xenoturbella* (65% missing data) and *Spadella* (75%) in our dataset. To control for a potential effect of missing data on our phylogenomic results, we

restricted our dataset to the 106 genes with sequences available for at least 41 of the 51 taxa. The concatenation of these 106 genes produces a matrix with 25,321 amino-acid sites that contains only 20% of missing data.

Bayesian and ML phylogenetic inference on this reduced dataset produced a tree fully congruent with

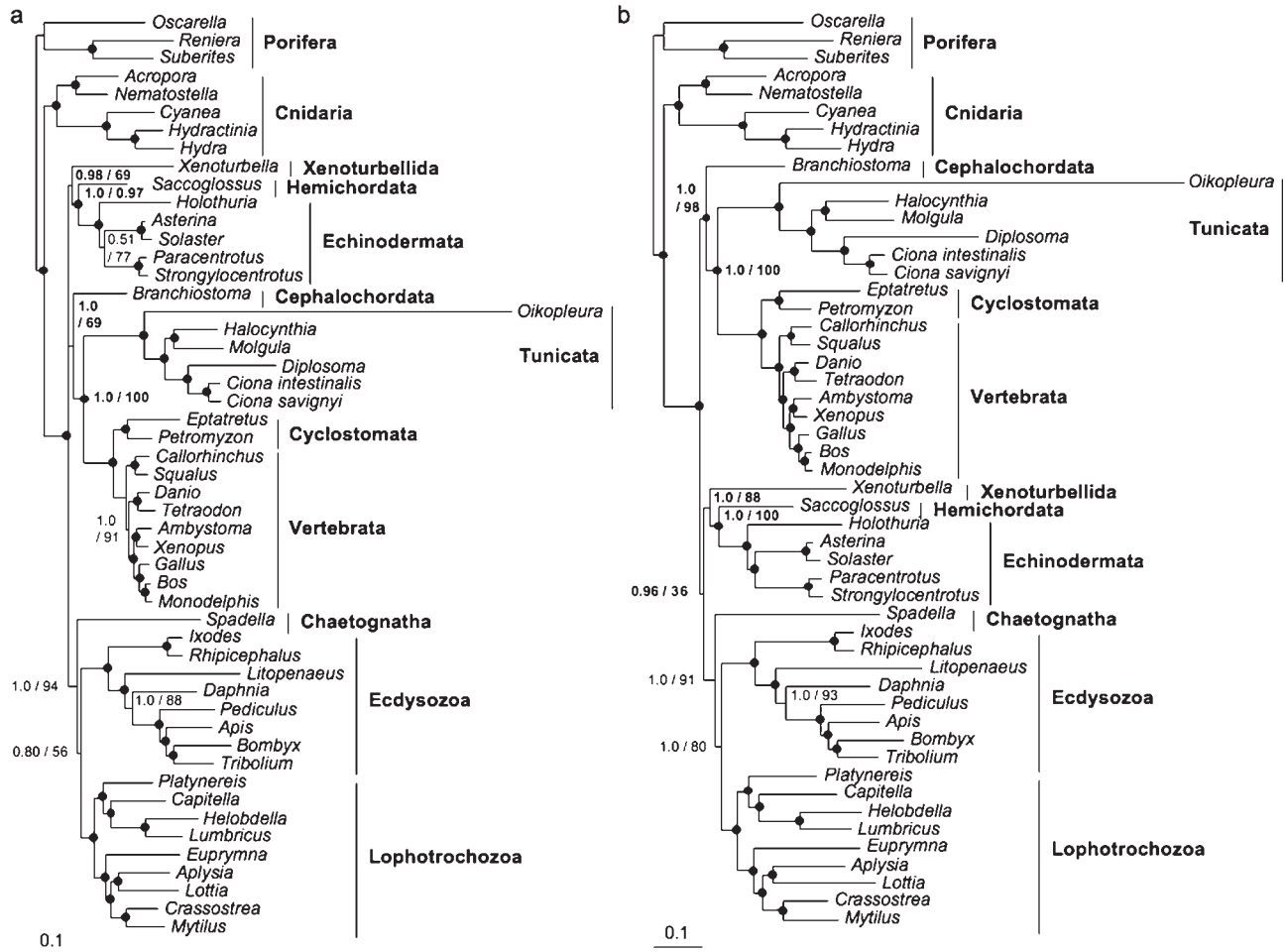


FIG. 2. Phylogenetic analyses of an updated phylogenomic dataset. (a) Bayesian consensus tree obtained using the CAT+ Γ_4 mixture model on the complete dataset based on the concatenation of 179 genes (51 taxa and 53,799 amino-acid sites) containing 32% missing data. (b) Bayesian inference using the CAT+ Γ_4 mixture model on the dataset reduced to the concatenation of the 106 genes for which sequences were available for at least 41 of the 51 taxa (25,321 amino-acid sites) containing only 20% of missing data. Values at nodes indicate Bayesian posterior probabilities (PP)/Maximum-likelihood bootstrap percentages (BP; 100 replicates) obtained under the WAG+ Γ_8 . Circles indicate strongly supported nodes with PP ≥ 0.95 and BP ≥ 95 . The scale bar represents the estimated number of substitutions per site.

the phylogenetic picture given by the complete dataset (Fig. 2b). In particular, the support for the monophyly of chordates was still maximal in terms of PP_{CAT+ Γ_4} , but BP_{WAG+F+ Γ_8} increased from 69 to 88%. The monophyly of Olfactores received maximal support in both cases and appeared not affected by missing data. Statistical support in terms of PP_{CAT+ Γ_4} and BP_{WAG+F+ Γ_8} was generally increased especially for locating incomplete taxa such as the *Xenoturbella* as the sister-group to Ambulacraria (from 0.98/69 to 1.0/88) and *Spadella* at the base of Protostomia (from 0.80/56 to 1.0/80). Altogether, reducing the amount of missing data, despite also reducing the total number of available sites, seems to result in a slight increase in bootstrap proportions. The only real noticeable difference between the two methods concerns the monophyly of deuterostomes. The Bayesian inference under the CAT mixture model suggests deuterostome paraphyly by supporting a basal position of chor-

dates within Bilateria (Fig. 2b) as previously reported (Lartillot and Philippe, 2008). However, ML retrieved the monophyly of deuterostomes, but with BP_{WAG+F+ Γ_8} of only 50%, leaving the monophyly of deuterostomes unresolved by our data.

Robustness of Phylogenomics to Gene Sampling

A legitimate question that can be directed to the phylogenomic approach is the degree to which the results are robust to the sample of genes used to infer phylogenetic trees. This potential concern was addressed by applying a jackknife statistical resampling protocol: fifty datasets were assembled by randomly drawing 50 genes from the total 179 genes, and subjected to Bayesian phylogenetic reconstruction using the CAT+ Γ_4 mixture model (see Methods). The resulting majority-rule consensus tree shows that the vast majority of inferred phy-

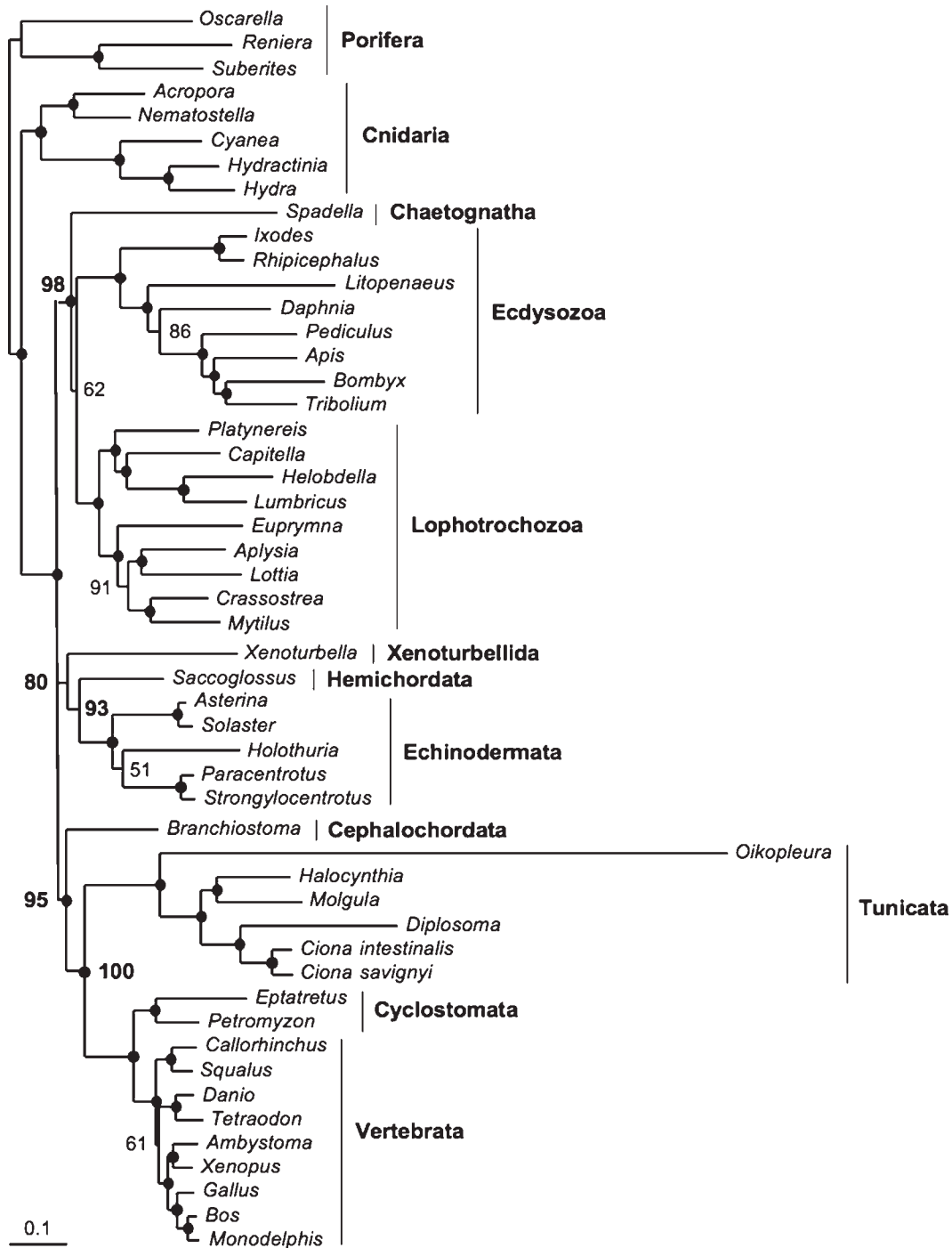


FIG. 3. Assessing the robustness of phylogenetic results to gene sampling using a jackknife procedure. The Bayesian phylogenetic inference was conducted under the CAT+ Γ_4 mixture model on 50 jackknife replicates of 50 genes over a total of 179. The tree presented is the weighted majority-rule consensus of all trees sampled every 10 cycles across the 50 replicates after removing the first 1000 trees in each MCMC as the burnin. Values at nodes represent corresponding jackknife-resampled posterior probabilities indices (PP_{JK}). Circles indicate highly repeatable nodes with PP_{JK} ≥ 95%. The scale bar represents the number of substitutions per site.

logenetic relationships are highly repeatable across the 50 jackknife replicates (see Fig. 3). Olfactores, Chordata, and Ambulacraria all received PP_{JK} of more than 90% indicating that phylogenetic support is not dependent upon a particular gene combination. Xenambulacraria

appears slightly more affected by gene sampling (PP_{JK} = 80%), but this relative instability might be explained by the poor gene representation available for *Xenoturbella* with only 98 genes over 179 (55%). The same kind of reasoning could apply to the relatively unstable posi-

tions of the chaetognath *Spadella* within protostomes ($PP_{JK} = 62\%$) and of *Holothuria* within echinoderms ($PP_{JK} = 51\%$) (see Fig. 3).

In fact, the only major clade whose monophyly appears to be influenced by gene sampling is deuterostomes for which PP_{JK} was less than 50% (see Fig. 3). In practice, this means that depending on the particular combination of 50 genes considered, deuterostomes might appear either monophyletic or paraphyletic, with the three possible topological alternatives retrieved in similar proportions: Deuterostomes (38%), basal chordates (28%), and basal Xenambulacraria (22%). Despite the inclusion of 25 taxa representing all major lineages in our dataset, these results confirm deuterostomes as one of the most difficult groups to resolve in the animal phylogeny despite its wide acceptance (see Lartillot and Philippe, 2008).

New Analyses of Ribosomal RNA Genes

The sister-group relationship between tunicates and vertebrates (Olfactores) observed in phylogenomics is in conflict with most (if not all) analyses of rRNA which favor cephalochordates as the closest relatives of vertebrates (Euchordates) (Cameron *et al.*, 2000; Mallatt and Winchell, 2007; Swalla *et al.*, 2000; Wada and Satoh, 1994; Winchell *et al.*, 2002). However, the statistical support for Euchordates in rRNA-based phylogenetic studies is moderate. Indeed, the first 18S rRNA study, based on a limited taxon sampling of deuterostomes, reported a bootstrap value of only 45% for Euchordates (Wada and Satoh, 1994). A subsequent 18S rRNA study considering only slowly evolving sequences for 16 deuterostomes found only a moderate bootstrap support of 71% for grouping cephalochordates and vertebrates (Cameron *et al.*, 2000). A study focused on tunicates also obtained moderate support for Euchordates (58 to 85% depending on the dataset and reconstruction method) but failed to support chordate monophyly likely because tunicate 18S rRNA sequences are rapidly evolving (Swalla *et al.*, 2000).

The next studies used the combination of 18S and 28S rRNAs. An investigation using 28 taxa for the two rRNA subunits found strong bootstrap support (89 to 97% depending on the method) for Euchordates (Winchell *et al.*, 2002). However, this study again failed to support chordate monophyly. Detailed analyses confirmed that tunicate genes have evolved rapidly and showed that they are compositionally biased toward AT, rendering tunicates virtually impossible to locate convincingly in the tree on the basis of rRNA data (Winchell *et al.*, 2002). Finally, increasing the sampling to 46 taxa for this 18S+28S rRNA data did not help in further resolving the relationships among the major groups of deuterostomes and even decreased the ML bootstrap support for Euchordates from 97% in the previous study to 50% (Mallatt and Winchell, 2007).

To gauge the extent to which the rRNA data conflicts with our phylogenomic results, we reanalyzed the 46-

taxa dataset of Mallatt and Winchell (2007). The heterogeneity of base composition in this dataset is well illustrated by the PCA presented in Figure 4a. At one extreme, tunicates (especially *Oikopleura*) are particularly AT-rich, and at the other extreme, Myxinidae (*Myxine* and *Eptatretus*) and the pterobranch hemichordate *Cephalodiscus* are highly GC-rich. We therefore compared phylogenetic reconstructions conducted on nucleotides and on RY-coded data, a coding scheme allowing reducing both substitutional saturation and nucleotide compositional bias (Fig. 4b). The two inferred ML trees appear mostly congruent except for two major topological shifts.

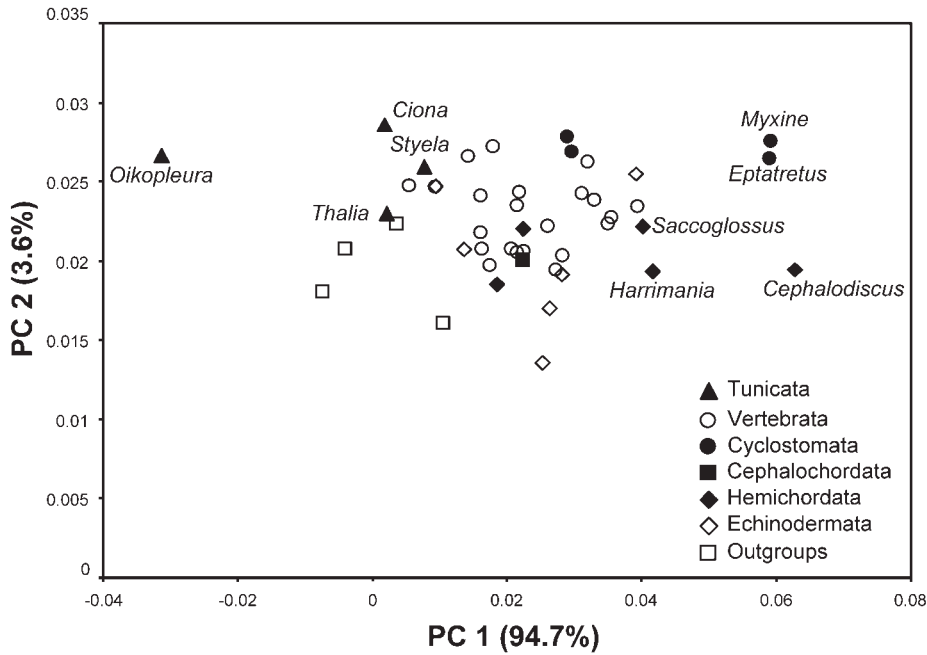
The strongest topological change occurred within hemichordates (Fig. 2b). Although the use of a standard DNA model strongly supports the paraphyly of enteropneusts by grouping the pterobranch *Cephalodiscus* with *Saccoglossus* and *Harrimania* (BP = 95), RY-coding allows recovering the monophyly of enteropneusts with high bootstrap support (BP = 90). This helps in understanding the conflict between 18S rRNA that supports enteropneust paraphyly (Cameron *et al.*, 2000; Halanych, 1995) and 28S rRNA that rather favors their monophyly (Mallatt and Winchell, 2007; Winchell *et al.*, 2002). This result is of particular importance because it potentially invalidates the controversial hypothesis that pterobranchs evolved from an enteropneust (Cameron *et al.*, 2000; Halanych, 1995) by suggesting that it is likely an artifact of 18S rRNA-based phylogenetic reconstructions due to shared nucleotide compositional bias between pterobranchs and Harrimaniidae (Fig. 4a).

Second, the support for the monophyly of Euchordates observed with nucleotides (BP = 84) disappeared in favor of the monophyly of Olfactores in the RY-coding dataset, albeit with no statistical support (BP = 44). This nevertheless strongly suggests that the high compositional bias of tunicate sequences has blurred the phylogenetic signal for Olfactores in previous analyses. Thus, according to our interpretation, reducing compositional bias and substitutional saturation by RY-recoding allows recovering a limited signal for Olfactores in agreement with our phylogenomic analysis of amino-acid data. It is worth noting however that rRNA does not statistically support chordate monophyly in both cases (Fig. 2b).

Molecular Phylogenetic Conclusions

Our aim was to reanalyze the phylogenetic relationships among chordates. The revision of the position of tunicates proposed by recent phylogenomic studies (Bourlat *et al.*, 2006; Delsuc *et al.*, 2006; Dunn *et al.*, 2008) by concluding in favor of the monophyly of Olfactores, has not yet been considered as totally convincing, essentially because it is at odds with both the traditional view based on embryological and morphological characters (Rowe, 2004; Schaeffer, 1987), and with earlier molecular phylogenetic analyses based on rRNA (Cameron *et al.*, 2000; Mallatt and Winchell, 2007; Swalla *et al.*, 2000; Wada and Satoh, 1994; Winchell *et al.*, 2002). The

a



b

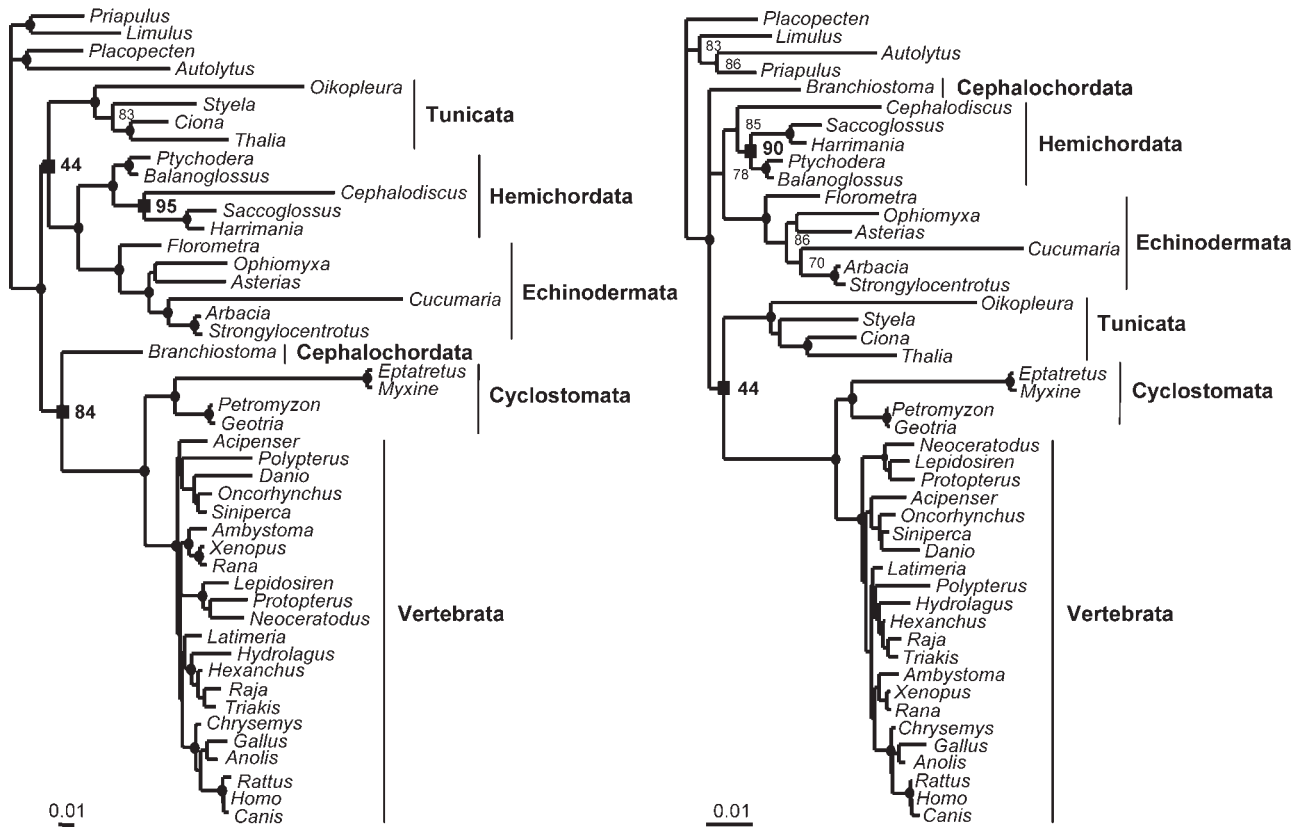


FIG. 4. New phylogenetic analyses of ribosomal RNA genes. **(a)** Principal component analysis of nucleotide composition of the combined 18S+28S rRNA dataset. The graph represents the projection of individuals on the first two axes, which explain more than 98% of the total variance. **(b)** Maximum-likelihood analyses of the 18S+28S dataset using the best-fitting standard DNA model (TIM+ Γ_4 +I) on nucleotides (left) and a two-state model (CF+ Γ_8) after RY-coding of nucleotides (right). ML bootstrap percentages are given at nodes when greater than 70 except within vertebrates. Circles indicate strongly supported nodes with BP ≥ 95 . Squares point to shifting nodes of interest between the two ML trees. Scale bars represent the number of substitutions per site.

unexpected sister-group relationship between echinoderms and cephalochordates observed in one of these studies (Delsuc *et al.*, 2006) may also have suggested the possibility that the monophyly of Olfactores was due to an artefactual attraction of cephalochordates with echinoderms (Bourlat *et al.*, 2006).

In this analysis, we have tried to address these points, essentially by reanalyzing both phylogenomic and rRNA data, under better taxonomic sampling and using more elaborate methods and probabilistic models. First, we demonstrate that, although the grouping of echinoderms and cephalochordates was indeed a probable artifact, disappearing upon the addition of several taxa or using an improved model of sequence evolution, the monophyly of Olfactores appears to be robust with respect to taxon sampling and model choice. Second, our reanalysis of rRNA data using RY-recoding also reveals a weak signal in favor of Olfactores, and suggests that the grouping of vertebrates and cephalochordates in former studies may have been an artifact driven by compositional biases. Altogether, our analyses allows a coherent interpretation of all empirical results observed thus far concerning chordate phylogeny, yielding further evidence in favor of the monophyly of both chordates and Olfactores.

At larger scale, however, we observe an overall lack of support for the monophyly of deuterostomes. Deuterostomes have nearly unanimously been considered as an unquestionable monophyletic group, a hypothesis backed up by traditional comparative analyses of embryological characters such as the fate of the blastopore (Nielsen, 2001), and morphological traits such as gill slits (Schaeffer, 1987). However, in our analyses, the status of deuterostomes seems to be sensitive to the model used, with CAT slightly favoring the paraphyletic configuration, and WAG the more traditional monophyly. In either case, the support measured by nonparametric resampling procedures (site-wise bootstrap or gene-wise jackknife) is weak.

Other phylogenomic studies (Dunn *et al.*, 2008; Lartillot and Philippe, 2008) also failed to obtain strong support for the relative phylogenetic positions of chordates and Ambulacraria. Moderate support for the monophyly of Deuterostomes was only obtained under empirical matrix models, the support disappearing when the CAT model was used instead (Dunn *et al.*, 2008; Lartillot and Philippe, 2008). Although profile mixture models such as CAT, whereas having a better fit than empirical matrices such as WAG, may have some inherent weaknesses as to their phylogenetic accuracy, the WAG empirical matrix fails in many cases, particularly when confronted to a high level of saturation (Lartillot *et al.*, 2007). This casts doubts on results that seem to receive support exclusively under this model, as is the case for deuterostome monophyly. More observations are needed to better gauge the relative merits of either type of model. Overall, although deuterostome monophyly still remains a reasonable working hypothesis to date, more work is needed before the question can be settled.

Corroborative Evidence for Olfactores Monophyly

The monophyly of Olfactores receives strong support from sequence-based phylogenomic inference. Rare genomic changes have also provided some evidence in its favor: the domain structure of cadherins (Oda *et al.*, 2002), a unique amino-acid insertion in fibrillar collagen (Wada *et al.*, 2006), and the distribution of micro RNAs (miRNAs) (Heimberg *et al.*, 2008). Finally, the *Branchiostoma floridae* genome helps confirming the sister-group relationship between tunicates and vertebrates in offering additional evidence from analyses of intron dynamics (Putnam *et al.*, 2008).

Cadherins are a superfamily of highly conserved adhesion molecules mediating cell communication and signaling that are pivotal for developmental processes of multicellular organisms. Their recent detection in the closest unicellular relatives of metazoans, the choanoflagellates, has highlighted their potential role in the origin of multicellularity (Abedin and King, 2008). Comparative studies on the classic cadherin subfamily has revealed that the structural element called Primitive Classic Cadherin Domain (PCCD) complex, otherwise termed non-chordate classic cadherin domain, is also present in cephalochordates, but has been lost in both tunicates and vertebrates (Oda *et al.*, 2002). The most parsimonious scenario is that this particular protein domain complex has been lost in the common ancestor of tunicates and vertebrates and constitutes a synapomorphy of Olfactores. However, cephalochordates possess two classic cadherin genes which originated by lineage-specific tandem duplication and that have a particular structure in lacking extracellular repeats found in all other investigated metazoans (Oda *et al.*, 2004). This derived state renders difficult to ascertain domain homology among chordate classic cadherin genes and casts doubt on its phylogenetic significance.

Further potential evidence for the clade Olfactores has been inferred from the evolution of fibrillar collagen genes within chordates. These genes represent important components of the notochord, the cartilage and mineralized bones in vertebrates. Phylogenetic analyses suggested that three ancestral fibrillar collagens gave rise to the gene diversity observed in living deuterostomes (Wada *et al.*, 2006). Comparative sequence analyses showed that tunicates and vertebrates share a molecular signature in the form of a six to seven amino-acid insertion in the C-terminus noncollagenous domain of one type of fibrillar collagens, that is absent in cephalochordates and echinoderms (Wada *et al.*, 2006). This insertion was interpreted as supporting the idea that vertebrates are more closely related to tunicates than to cephalochordates (Wada *et al.*, 2006). The homology of the insertion appears nevertheless difficult to assert with certainty given the high degree of sequence divergence observed in this region of the molecule. More tunicate fibrillar collagen sequences might help in better understanding the dynamics of this peculiar amino-acid insertion and the phylogenetic signal it conveys.

The comparison of miRNA repertoires in metazoans has also recently unearthed some potential signatures for the sister-group relationship of tunicates and vertebrates (Heimberg *et al.*, 2008). miRNAs are small non-coding RNAs involved in regulation of gene expression in eukaryotes and play an important role in the development of metazoans. Comparative genomic studies of miRNAs underlined that, during the evolution of metazoans, major body-plan innovations seemed to coincide with dramatic expansions of miRNA repertoires, suggesting a potential role in the increase of morphological complexity (Hertel *et al.*, 2006; Sempere *et al.*, 2006). The most recent study unveiled that three miRNA families (*mir-126*, *mir-135*, and *mir-155*) were likely acquired in the common ancestor of tunicates and vertebrates (Heimberg *et al.*, 2008). Taking into consideration that miRNAs might be only rarely secondarily lost once they have been recruited, this finding provides corroborative evidence for the clade Olfactores. It should be noted however that, of these three miRNA families, only *mir-126* constitutes an exclusive synapomorphy for Olfactores without subsequent secondary loss in descendant taxa confirmed by Northern analysis. Moreover, the profound reorganization of miRNA repertoire undergone by tunicates requires being cautious when interpreting acquisition of miRNAs as potential signatures for reconstructing their phylogenetic relationships (Fu *et al.*, 2008).

Additional sequence-based phylogenomic reconstructions and analyses of rare genomic changes have been issued along with the recently published draft sequence of a cephalochordate (*Branchiostoma floridae*) genome (Putnam *et al.*, 2008). The phylogenetic analysis of a concatenation of 1,090 orthologs from 12 complete genomes retrieved maximal Bayesian support for Olfactores and chordates, whereas the corresponding bootstrap support was maximal for Olfactores but of only 78% for chordate monophyly (Putnam *et al.*, 2008). Moreover, the analysis of individual gene phylogenies revealed twice more cases where Olfactores was favored over Euchordates than the reverse (Putnam *et al.*, 2008). Further evidence was obtained by analyzing the phylogenetic signal deduced from the dynamics of intron gain and loss among chordate genomes. Despite extensive intron losses along the tunicate lineage, a number of shared intron gain/loss events can be identified as a signature of tunicates and vertebrates common ancestry (Putnam *et al.*, 2008). Overall, the new evidence brought by the analysis of the *Branchiostoma floridae* genome essentially corroborates our phylogenetic results.

Implications for Chordate Evo-Devo

The additional evidence presented for the new chordate phylogeny provides a robust phylogenetic framework for (re)interpreting the evolution of morphological characters and developmental features. Inverting the phylogenetic position of tunicates and cephalochordates within monophyletic chordates highlights the preva-

lence of morphological simplification with characters that are likely ancestral for chordates, such as metameric segmentation, being lost secondarily in the tunicate lineage. On the other hand, the loss of preoral kidney and the presence of multiciliated epithelial cells might in fact constitute morphological synapomorphies for olfactores (Ruppert, 2005). The new chordate phylogeny further portrays tunicates as highly derived chordates with specialized lifestyles and developmental modes, whereas cephalochordates might have retained more ancestral chordate characteristics. We will use two examples to illustrate the importance of considering the new phylogenetic status of tunicates as the sister-group of vertebrates in the context of evolutionary developmental biology.

The first illustration concerns the evolutionary origin of such fundamental structures as the neural crest and olfactory placodes. Migratory neural crest cells and sensory placodes have long been considered as vertebrate innovations. Implicated respectively in the development of major tissues and sensory organs, their origin is generally correlated with the increase in morphological complexity of vertebrates. However, recent molecular developmental studies have revealed the presence in tunicates of migratory neural crest-like cells (Jeffery, 2006; Jeffery *et al.*, 2004) and olfactory placodes (Bassham and Postlethwait, 2005; Mazet *et al.*, 2005). When reinterpreted in light of the new chordate phylogeny, these results implied that both of these features did not evolve *de novo* in the vertebrate lineage, but rather evolved from specialized pre-existing structures in the common ancestor of vertebrates and tunicates.

The second example illustrates how the new phylogenetic context helps in understanding the genomic and developmental peculiarities of tunicates within chordates. The new phylogenetic picture implied that tunicate genomes have undergone significant genome reduction from the ancestral chordate genome (Holland, 2007). This genome compaction is also associated with a high rate of genomic evolution at the levels of both primary sequences (Delsuc *et al.*, 2006; Edvardsen *et al.*, 2004) and genome organization (Holland and Gibson-Brown, 2003). One of the most spectacular rearrangements of tunicate genomes is the loss of several Hox genes, the disintegration of the Hox cluster, and the loss of temporal colinearity in Hox gene expression during development (Ikuta *et al.*, 2004; Seo *et al.*, 2004). These observations raise the question of how tunicates, with their altered Hox clusters, are still able to develop a chordate body plan. In chordates, and deuterostomes more generally, temporal colinearity is regulated by the Retinoic-Acid (RA) signaling pathway which controls the antero-posterior patterning of the embryo (Cañestro *et al.*, 2006; Marlétaz *et al.*, 2006). However, axial patterning in tunicates seems to have become independent of RA-signaling, with the genes of the RA machinery even being lost in *Oikopleura* (Cañestro and Postlethwait, 2007). Functional studies have shown that if "*Oikopleura* can be considered as a classical RA-signal-

ing knock-down mutant naturally produced by evolution," it is still capable of developing a typical chordate body plan (Cañestro and Postlethwait, 2007). With cephalochordates, which possess the RA genomic toolkit, being basal among chordates, RA-signalling must have been present in the tunicate ancestor and secondarily lost in *Oikopleura* suggesting that appendicularians use alternative mechanisms for the development of chordate features (Cañestro *et al.*, 2007; Holland, 2007).

The new chordate phylogeny strengthens the view that tunicates and cephalochordates represent complementary models for studying vertebrate Evo-Devo (Schubert *et al.*, 2006). Tunicates are phylogenetically closer to vertebrates but appear both morphologically and molecularly highly derived. The diversity of their developmental modes offers the opportunity to study the evolution of alternative adaptive solutions to the typical chordate development. In having retained more ancestral features, cephalochordates provide an ideal outgroup for polarizing evolutionary changes that occurred in tunicates and vertebrates. With the cephalochordate *Branchiostoma floridae* genome (Putnam *et al.*, 2008) and the upcoming genome sequence of the appendicularian *Oikopleura dioica*, the newly established phylogenetic framework makes chordate comparative genomics appearing full of promises for the Evo-Devo community as exemplified in a recent work on the origin and evolution of the Pax gene family (Bassham *et al.*, 2008).

ACKNOWLEDGMENTS

The authors thank the associate editors Billie Swalla and José Xavier-Neto for giving the opportunity to write this article. The authors also thank John Mallatt for kindly providing his 18S-28S rRNA alignment, Julien Claude for help in using R statistical package, and two anonymous reviewers for their comments. The extensive phylogenomic calculations benefited from the ISEM computing cluster.

LITERATURE CITED

- Abedin M, King N. 2008. The premetazoan ancestry of cadherins. *Science* 319:946–948.
- Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R. 2000. The new animal phylogeny: Reliability and implications. *Proc Natl Acad Sci USA* 97:4453–4456.
- Aguinaldo AM, Turbeville JM, Linford LS, Rivera MC, Garey JR, Raff RA, Lake JA. 1997. Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* 387:489–493.
- Bassham S, Cañestro C, Postlethwait JH. 2008. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol* 6:35.
- Bassham S, Postlethwait JH. 2005. The evolutionary history of placodes: A molecular genetic investigation of the larvacean urochordate *Oikopleura dioica*. *Development* 132:4259–4272.
- Baurain D, Brinkmann H, Philippe H. 2007. Lack of resolution in the animal phylogeny: Closely spaced cladogeneses or undetected systematic errors? *Mol Biol Evol* 24:6–9.
- Blair JE, Hedges SB. 2005. Molecular phylogeny and divergence times of deuterostome animals. *Mol Biol Evol* 22:2275–2284.
- Blanquart S, Lartillot N. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* 25:842–858.
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444:85–88.
- Cameron CB, Garey JR, Swalla BJ. 2000. Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. *Proc Natl Acad Sci USA* 97:4469–4474.
- Cañestro C, Postlethwait JH. 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol* 305:522–538.
- Cañestro C, Postlethwait JH, Gonzalez-Duarte R, Albalat R. 2006. Is retinoic acid genetic machinery a chordate innovation? *Evol Dev* 8:394–406.
- Cañestro C, Yokoi H, Postlethwait JH. 2007. Evolutionary developmental biology and genomics. *Nat Rev Genet* 8:932–942.
- Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552.
- Cavender JA, Felsenstein J. 1987. Invariants of phylogenies in a simple case with discrete states. *J Classif* 4:57–71.
- Conway-Morris S. 2003. The Cambrian "explosion" of metazoans and molecular biology: Would Darwin be satisfied? *Int J Dev Biol* 47:505–515.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965–968.
- Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* 6:361–375.
- Dunn CW, Hejnal A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martinale MQ, Giribet G. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745–749.
- Edwardsen RB, Lerat E, Maeland AD, Flat M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo HC, Chourrout D. 2004. Hypervariable and highly divergent intron-exon organizations in the chordate *Oikopleura dioica*. *J Mol Evol* 59:448–457.
- Felsenstein J. 2001. PHYLIP (Phylogenetic Inference Package) version 3.6. (distributed by the author). Seattle: Department of Genetics, University of Washington.
- Felsenstein J. 2004. *Inferring phylogenies*. Sunderland, MA, USA: Sinauer Associates, Inc. 645 p.
- Fu X, Adamski M, Thompson EM. 2008. Altered miRNA repertoire in the simplified chordate, *Oikopleura dioica*. *Mol Biol Evol* 25:1067–1080.
- Gee H. 2001. Deuterostome phylogeny: the context for the origin and evolution of the vertebrates. In: Ahlberg PE, editor. *Major events in early vertebrate evolution: Palaeontology, phylogeny, genetics, and development*. London: Taylor and Francis. pp 1–14.
- Halanych KM. 1995. The phylogenetic position of the pterobranch hemichordates based on 18S rDNA sequence data. *Mol Phylogenet Evol* 4:72–76.
- Halanych KM. 2004. The new view of animal phylogeny. *Annu Rev Ecol Syst* 35:229–256.
- Halanych KM, Bacheller JD, Aguinaldo AM, Liva SM, Hillis DM, Lake JA. 1995. Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. *Science* 267:1641–1643.
- Hartmann S, Vision TJ. 2008. Using ESTs for phylogenomics: Can one accurately infer a phylogenetic tree from a gappy alignment? *BMC Evol Biol* 8:95.
- Heimberg AM, Sempere LF, Moy VN, Donoghue PC, Peterson KJ. 2008. MicroRNAs and the advent of vertebrate morphological complexity. *Proc Natl Acad Sci USA* 105:2946–2950.
- Hertel J, Lindemeyer M, Missal K, Fried C, Tanzer A, Flamm C, Hofacker IL, Stadler PF. 2006. The expansion of the metazoan microRNA repertoire. *BMC Genomics* 7:25.
- Holland LZ. 2007. Developmental biology: A chordate with a difference. *Nature* 447:153–155.
- Holland LZ, Gibson-Brown JJ. 2003. The *Ciona intestinalis* genome: When the constraints are off. *Bioessays* 25:529–532.
- Ikuta T, Yoshida N, Satoh N, Saiga H. 2004. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. *Proc Natl Acad Sci USA* 101:15118–15123.

- Jefferies RPS. 1991. Two types of bilateral symmetry in the Metazoa: Chordate and bilaterian. In: Bock GR, Marsh J, editors. *Biological asymmetry and handedness*. Chichester: Wiley. pp 94–127.
- Jeffery WR. 2006. Ascidian neural crest-like cells: Phylogenetic distribution, relationship to larval complexity, and pigment cell fate. *J Exp Zool B Mol Dev Evol* 306:470–480.
- Jeffery WR. 2007. Chordate ancestry of the neural crest: New insights from ascidians. *Semin Cell Dev Biol* 18:481–491.
- Jeffery WR, Strickler AG, Yamamoto Y. 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. *Nature* 431:696–699.
- Jeffroy O, Brinkmann H, Delsuc F, Philippe H. 2006. Phylogenomics: The beginning of incongruence? *Trends Genet* 22:225–231.
- Jimenez-Guri E, Philippe H, Okamura B, Holland PWH. 2007. Buddenbrockia is a Cnidarian worm. *Science* 317:116–118.
- Jobb G, von Haeseler A, Strimmer K. 2004. TREEFINDER: A powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18.
- Lartillot N, Brinkmann H, Philippe H. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 7 Suppl 1:S4.
- Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21:1095–1109.
- Lartillot N, Philippe H. 2008. Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos Trans R Soc Lond B Biol Sci* 363:1463–1472.
- Lwoff A. 1944. *L'évolution physiologique: Étude des pertes de fonctions chez les microorganismes*. Paris: Hermann. 308 p.
- Mallatt J, Winchell CJ. 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: More cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol Phylogenet Evol* 43:1005–1022.
- Marlétaz F, Holland LZ, Laudet V, Schubert M. 2006a. Retinoic acid signaling and the evolution of chordates. *Int J Biol Sci* 2:38–47.
- Marlétaz F, Martin E, Perez Y, Papillon D, Caubit X, Lowe CJ, Freeman B, Fasano L, Dossat C, Wincker P, Weissenbach J, Le Parco Y. 2006b. Chaetognath phylogenomics: A protostome with deuterostome-like development. *Curr Biol* 16:R577–R578.
- Matus DQ, Copley RR, Dunn CW, Hejnol A, Eccleston H, Halanych KM, Martindale MQ, Telford MJ. 2006. Broad taxon and gene sampling indicate that chaetognaths are protostomes. *Curr Biol* 16:R575–R576.
- Mazet F, Shimeld SM. 2005. Molecular evidence from ascidians for the evolutionary origin of vertebrate cranial sensory placodes. *J Exp Zool B Mol Dev Evol* 304:340–346.
- Nielsen C. 2001. *Animal evolution, interrelationships of the living phyla*. Oxford, UK: Oxford University Press.
- Oda H, Akiyama-Oda Y, Zhang S. 2004. Two classic cadherin-related molecules with no cadherin extracellular repeats in the cephalochordate amphioxus: Distinct adhesive specificities and possible involvement in the development of multicell-layered structures. *J Cell Sci* 117:2757–2767.
- Oda H, Wada H, Tagawa K, Akiyama-Oda Y, Satoh N, Humphreys T, Zhang S, Tsukita S. 2002. A novel amphioxus cadherin that localizes to epithelial adherens junctions has an unusual domain organization with implications for chordate phylogeny. *Evol Dev* 4:426–434.
- Philippe H. 1993. MUST, a computer package of management utilities for sequences and trees. *Nucleic Acids Res* 21:5264–5272.
- Philippe H, Delsuc F, Brinkmann H, Lartillot N. 2005a. Phylogenomics. *Annu Rev Ecol Syst* 36:541–562.
- Philippe H, Lartillot N, Brinkmann H. 2005b. Multigene analyses of bilaterian animals corroborate the monophyly of ecdysozoa, lophotrochozoa, and protostomia. *Mol Biol Evol* 22:1246–1253.
- Philippe H, Snell EA, Baptiste E, Lopez P, Holland PWH, Casane D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. *Mol Biol Evol* 21:1740–1752.
- Philippe H, Telford MJ. 2006. Large-scale sequencing and the new animal phylogeny. *Trends Ecol Evol* 21:614–620.
- Phillips MJ, Delsuc F, Penny D. 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol Biol Evol* 21:1455–1458.
- Posada D, Crandall KA. 1998. MODELTEST: Testing the model of DNA substitution. *Bioinformatics* 14:817–818.
- Posada D, Crandall KA. 2001. Selecting the best-fit model of nucleotide substitution. *Syst Biol* 50:580–601.
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064–1071.
- R Development Core Team. 2007. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Rodriguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF, Philippe H. 2007. Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst Biol* 56:389–399.
- Roure B, Rodriguez-Ezpeleta N, Philippe H. 2007. SCAFoS: A tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol* 7 Suppl 1:S2.
- Rowe T. 2004. Chordate phylogeny and development. In: Cracraft J, Donoghue MJ, editors. *Assembling the tree of life*. Oxford: Oxford University Press. pp 384–409.
- Ruppert EE. 2005. Key characters uniting hemichordates and chordates: Homologies or homoplasies? *Can J Zool* 83:8–23.
- Schaeffer B. 1987. Deuterostome monophyly and phylogeny. *Evol Biol* 21:179–235.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: Maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18:502–504.
- Schubert M, Escriva H, Xavier-Neto J, Laudet V. 2006. Amphioxus and tunicates as evolutionary model systems. *Trends Ecol Evol* 21:269–277.
- Sempere LF, Cole CN, McPeck MA, Peterson KJ. 2006. The phylogenetic distribution of metazoan microRNAs: Insights into evolutionary complexity and constraint. *J Exp Zool B Mol Dev Evol* 306:575–588.
- Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaot M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* 431:67–71.
- Steel M. 2005. Should phylogenetic models be trying to “fit an elephant”? *Trends Genet* 21:307–309.
- Swalla BJ, Cameron CB, Corley LS, Garey JR. 2000. Urochordates are monophyletic within the deuterostomes. *Syst Biol* 49:52–64.
- Swofford DL. 2002. PAUP*: Phylogenetic analysis using parsimony and other methods version 4.0b10. Sunderland, MA: Sinauer.
- Telford MJ, Budd GE. 2003. The place of phylogeny and cladistics in Evo-Devo research. *Int J Dev Biol* 47:479–490.
- Vienne A, Pontarotti P. 2006. Metaphylogeny of 82 gene families sheds a new light on chordate evolution. *Int J Biol Sci* 2:32–37.
- Wada H, Okuyama M, Satoh N, Zhang S. 2006. Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol Dev* 8:370–377.
- Wada H, Satoh N. 1994. Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. *Proc Natl Acad Sci USA* 91:1801–1804.
- Whelan S, Goldman N. 2001. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18:691–699.
- Wiens JJ. 2003. Missing data, incomplete taxa, and phylogenetic accuracy. *Syst Biol* 52:528–538.
- Wiens JJ. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Syst Biol* 54:731–742.
- Wiens JJ. 2006. Missing data and the design of phylogenetic analyses. *J Biomed Inform* 39:34–42.
- Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J. 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol Biol Evol* 19:762–776.

CHAPITRE II :

L'ARN RIBOSOMIQUE 18S ET
LA PHYLOGENIE MOLECULAIRE DES TUNICIERS

1. Phylogénie moléculaire des Tuniciers

Au cours des dernières années, la position phylogénétique des Tuniciers au sein des Chordés a été radicalement bouleversée. La reconnaissance de leur statut de plus proches parents des Vertébrés a renforcé leur rôle de modèle pour la compréhension des étapes fondamentales de l'évolution des Chordés (Bourlat *et al* 2006; Delsuc *et al* 2006; Delsuc *et al* 2008; Dunn *et al* 2008; Putnam *et al* 2008). Malgré cela, certaines relations de parenté à l'intérieur des Tuniciers demeurent débattues à l'heure actuelle, notamment la position de groupes majeurs tels que les Appendiculaires, les Aplousobranches et les Thaliacés (Zeng & Swalla 2005).

L'inférence de l'histoire évolutive des Tuniciers à partir des données moléculaires ne commença qu'aux alentours de 1998 (Wada 1998). Dans cette étude pionnière ainsi que dans les études suivantes, le gène de l'ARNr 18s constitua le marqueur majoritairement utilisé pour la détermination des affinités phylogénétiques du groupe (Swalla *et al* 2000; Yokobori *et al* 2006; Zeng, Jacobs & Swalla 2006). Seules deux études ont également utilisé le gène mitochondrial codant pour la Cytochrome Oxidase I (COI) (Stach & Turbeville 2002; Turon & Lopez-Legentil 2004).

De manière générale, l'image moléculaire obtenue par ces premières études s'écarta considérablement de la classification traditionnelle, en ayant jusqu'ici établi un nouveau cadre phylogénétique dont les points principaux se résument à la paraphylie des Ascidiées (Ascidiacea), avec la mise en évidence de trois clades majeurs au sein des Tuniciers : (i) Phlebobranchia + Thaliacea, (ii) Appendicularia, et (iii) Stolidobranchia.

2. Le statut phylogénétique des groupes majeurs de Tuniciers

Malgré les fortes valeurs de soutien statistique observées pour ce schéma général, des incertitudes perdurent quant aux relations de parenté entre les grandes lignées des Tuniciers et également en leur sein. L'instabilité persistante de la position des Appendicularia, le statut énigmatique de l'ordre des Aplousobranchia, et les affinités évolutives confuses des Thaliacea au sein des Phlebobranchia figurent parmi les questions les plus résistantes dans l'ensemble des phylogénies obtenues jusqu'à présent.

I. Les Appendiculaires

Les Appendiculaires ou Larvacés constituent le seul groupe de Tuniciers dont les membres retiennent le plan d'organisation typique des Chordés tout au long de leur vie, en vertu d'une métamorphose moins sévère qu'ils subissent au cours du passage à l'état adulte (Grassé 1948). Par conséquent, l'éclaircissement de leurs relations de parenté vis-à-vis des Ascidiacés et des Thaliacés revêt un intérêt primordial du point de vue développemental et évolutif (Nishino & Satoh 2001).

D'après la systématique traditionnelle, les Appendicularia incluent trois ordres / familles: Oikopleurida(e), Fritillarrida(e), et Kowalevskiida(e). Cependant la représentation du groupe dans les phylogénies moléculaires est restreinte à des membres du genre « modèle » *Oikopleura*. Les familles des Kowalevskiidae et Fritillarridae n'ont jusqu'ici jamais été explorées moléculairement.

Dans la plupart des études, les Appendiculaires se situèrent à la base des arbres phylogénétiques, constituant ainsi le premier groupe à émerger au sein des Tuniciers (Wada 1998; Swalla *et al* 2000). Cette position phylogénétique est conforme à l'image fournie par des données phylogénomiques plaçant *Oikopleura dioica* systématiquement à la base des autres Tuniciers inclus (Delsuc *et al* 2006; Delsuc *et al* 2008). Néanmoins, les analyses phylogénétiques les plus récentes portant sur l'ARNr 18S ont remis en question cette hypothèse en proposant une position plus interne des Appendiculaires au sein des Tuniciers en tant que group-frère potentiel des Stolidobranches (Zeng & Swalla 2005; Zeng, Jacobs & Swalla 2006).

La position phylogénétique des Appendiculaires dans la phylogénie des Tuniciers a suscité de nombreux débats car certains auteurs proposèrent que leur position basale pourrait refléter un artéfact de reconstruction phylogénétique (Swalla *et al* 2000; Winchell *et al* 2002). En effet, les représentants du genre *Oikopleura* présentent la particularité de posséder taux d'évolution rapide par rapport aux autres Tuniciers, les rendant ainsi prônes à subir les effets potentiels du phénomène d'attraction des longues branches. L'évolution rapide des Appendiculaires semble être un attribut génomique puisque le même phénomène est relevé dans les arbres phylogénomiques (Delsuc *et al* 2006; Delsuc *et al* 2008).

II. Les Ascidiacés et la position controversée des Aplousobranches

Comprenant trois grands Ordres (Aplousobranchia, Phlebobranchia et Stolidobranchia), les Ascidiacés s'avèrent très tôt être un groupe paraphylétique dans

l'ensemble des arbres phylogénétiques précédemment reconstruits (Wada 1998; Swalla *et al* 2000; Stach & Turbeville 2002; Zeng & Swalla 2005; Yokobori *et al* 2006; Zeng, Jacobs & Swalla 2006). Dès les premiers ébauches phylogénétiques, une forte affinité entre l'Ordre des Phlebobranchia et la Classe des Thaliacea a été mise en évidence (Wada 1998).

Les Stolidobranchia quant à eux apparurent solidement monophylétiques et divisés en deux grands clades : 1) les Molgulidae et 2) Pyuridae, Styelidae et Brotryllidae (Swalla *et al* 2000). L'application du test du taux relatif révéla par ailleurs que les Molgulidae évoluent significativement plus vite par rapport aux autres familles du groupe (Swalla *et al* 2000). En ce qui concerne les relations de parenté à l'échelon taxonomique inférieur, l'image phylogénétique dégagée se montra beaucoup plus imprécise, avec notamment une incertude totale quant aux affinités interfamiliales au sein des Puyridae et des Styelidae (Zeng & Swalla 2005; Zeng, Jacobs & Swalla 2006).

Cependant, le point le plus controversé quant au statut phylogénétique des Ascidies concerne sans doute le positionnement du troisième Ordre : les Aplousobranchia. Les membres de ce groupe se distinguent de par leur plan d'organisation, en comprenant exclusivement des espèces coloniales où les individus prennent la forme de zooïdes très simplifiés et de taille diminuée (Lahille 1887).

L'intégration des Aplousobranches dans un contexte phylogénétique moléculaire conduit à privilégier au départ un étroit apparentement du groupe avec les Appendiculaires dans une position basale au sein des Tuniciers (Stach & Turbeville 2002). Néanmoins, ce scénario évolutif fut complètement remis en question puisque les séquences d'ARNr 18S d'Aplousobranchia rapportées dans cette étude s'avérèrent quelques années plus tard être des contaminations par divers espèces de protistes (Yokobori *et al* 2006).

Une deuxième hypothèse fut également proposée en rapprochant les Aplousobranchia avec les Cionidae et les Diazonidae (Phlebobranchia) d'après l'analyse phylogénétique du gène mitochondrial COI (Turon & Lopez-Legentil 2004). Cette relation en accord avec les données morphologiques (Kott 2005) a néanmoins été contredite par les travaux de Yokobori *et al.* (2006) qui furent les premiers à publier des séquences authentiques d'ARNr 18S d'Aplousobranches. Dans cette dernière étude, les Aplousobranchia apparurent regroupés avec les Thaliacés, avec la particularité toutefois de posséder un taux d'évolution extraordinairement rapide conférant au groupe des branches excessivement longues dans la phylogénie inférée (Yokobori *et al* 2006). Étant source d'artéfacts de reconstruction potentiels, le taux d'évolution extrêmement élevé des séquences d'Aplousobranchia ne permit

pas de conclure quant à leur positionnement robuste au sein des Tuniciers (Zeng, Jacobs & Swalla 2006).

III. Les Thaliacés

Un dernier point faible de la phylogénie des Tuniciers concerne la position des Thaliacés au sein des Ascidiés, ainsi que les relations entre les trois Ordres que comprend cette Classe : Doliolida, Pyrosomatida et Salpida. Les Thaliacés ont été positionnés à l'intérieur des Phlebobranchia dans la vaste majorité des arbres à présent reconstruits (Swalla *et al* 2000; Yokobori *et al* 2006; Zeng, Jacobs & Swalla 2006). En dépit de cela, leur position exacte au sein du clade Phlebobranchia demeura considérablement instable puisqu'elle apparut variable selon l'échantillonnage taxonomique, la méthode de reconstruction suivie, ou encore le modèle d'évolution utilisé dans les analyses phylogénétiques. L'échantillonnage taxonomique a été notamment limité qu'à trois espèces représentant les trois Ordres constitutifs des Thaliacés qui présentent de longues branches dans les phylogénies inférées (Swalla *et al* 2000; Stach & Turbeville 2002; Zeng & Swalla 2005; Yokobori *et al* 2006; Zeng, Jacobs & Swalla 2006).

3. Mise à jour de la phylogénie des Tuniciers sur la base du gène de l'ARNr 18S en utilisant des modèles de mélange et de structure secondaire [Article 2]

Tirant parti des phylogénies précédentes, un projet de recherche fut conduit avec comme objectif principal de mettre à jour la phylogénie des Tuniciers à partir de l'ARNr 18S (Tsagkogeorga *et al* 2009). L'objectif étant de fournir une image phylogénétique robuste incluant tous les groupes majeurs de Tuniciers, cette étude fut abordée par deux approches différentes mais toutefois complémentaires. La première approche consista à incorporer un échantillonnage taxonomique plus représentatif des grandes lignées, avec une emphase sur les groupes dont le statut était le plus controversé. La deuxième approche reposa sur l'utilisation de modèle d'évolution plus élaborés pour conduire l'analyse du signal phylogénétique contenu dans les séquences acquises, ceci en prenant en compte la dynamique d'évolution particulière des ARN ribosomiques.

Un total de trente nouvelles séquences d'ARNr 18S furent acquises pour les Tuniciers assurant ainsi un nombre de séquences par clade proportionnel au nombre d'espèces comprises dans chaque lignée majeure. Un jeu de données fut ainsi assemblé, incluant 110

taxons (73 Tuniciers + 15 groupes-externes) au sein desquels figurent pour la première fois des espèces appartenant à des groupes faiblement ou pas du tout représentés précédemment tels que les Appendiculaires, les Aplousobranchees et les Thaliacés.

Considérant l'hétérogénéité des pressions de sélection s'exerçant sur les séquences d'ARNr 18S imposée par la structure secondaire de la molécule, ce jeu de données a été aligné manuellement par rapport à une structure consensus issue de l'European Ribosomal Database (Wuyts, Perriere & Van De Peer 2004). Les analyses phylogénétiques ont été conduites à la fois sous des modèles standards d'évolution des séquences et en utilisant des modèles plus adaptés à la dynamique évolutive de l'ARNr 18S. Ces dernières incluent des analyses Bayésiennes réalisées sous deux types de modèles d'évolution de séquences : le modèle de mélange site-hétérogène CAT-GTR (Lartillot & Philippe 2004) et les modèles ARN spécifiques (Savill, Hoyle & Higgs 2001). Le premier accommode l'hétérogénéité des profils substitutionnels entre les sites nucléotidiques, tandis que les seconds considèrent l'occurrence de substitutions compensées dans les régions double-brin (« tiges ») des ARN ribosomiques (Telford, Wise & Gowri-Shankar 2005; Lartillot, Brinkmann & Philippe 2007). Des analyses supplémentaires ont été également menées afin d'estimer l'adéquation des modèles d'évolution de séquences sélectionnés (CAT vs ARN vs standard), et d'approfondir les conséquences de la divergence étonnante des séquences d'Aplousobranchees en termes de structure secondaire.

En ce qui concerne l'évolution des grandes lignées de Tuniciers, les différentes approches de reconstruction convergèrent vers une image phylogénétique non-équivoque, divisant ce phylum en trois clades majeurs : 1) Phlebobranchia + Aplousobranchia + Thaliacea, 2) Appendicularia, et 3) Stolidobranchia, la position des Appendiculaires restant toutefois ambiguë. Les résultats soutiennent également la monophylie des Thaliacés au sein desquels une relation de parenté entre les ordres Salpida et Pyrosomatida est suggérée par le modèle CAT. Les séquences d'Aplousobranchees évoluant extraordinairement vite n'ont par conséquent pas pu être positionnées avec confiance. Cependant, leur appartenance au clade unissant les Phlebobranchia et les Thaliacea a été clairement mise en évidence.

Au sein des Aplousobranchia eux-mêmes, les membres de la famille Clavelinidae, échantillonnée pour la première fois, montrèrent un taux d'évolution plutôt lent s'écartant ainsi très significativement de celui des autres Aplousobranchees. Ces résultats ont permis d'identifier une accélération drastique de taux d'évolution à l'intérieur des Aplousobranchees

ayant entraînée plusieurs changements significatifs au niveau de la structure secondaire de l'ARNr 18S chez les Aplousobranchez à taux rapide (ex. *Diplosoma ooru*).

Enfin, du point de vue méthodologique, nos résultats ont montré que le modèle CAT semble offrir une meilleure adéquation que les modèles standards suggérant ainsi une dynamique évolutive de l'ARNr 18S beaucoup plus complexe que celle classiquement considérée par les modèles ARN-spécifiques.

ARTICLE II

ARTICLE PUBLIÉ DANS *BMC EVOLUTIONARY BIOLOGY*

AN UPDATED 18S rRNA PHYLOGENY OF TUNICATES BASED ON MIXTURE AND SECONDARY STRUCTURE MODELS

GEORGIA TSAGKOGEOGA^{1,2}, XAVIER TURON³, RUSSELL R HOPCROFT⁴,
MARIE-KA TILAK^{1,2}, TAMAR FELDSTEIN⁵, NOA SHENKAR^{5,6}, YOSSI LOYA⁵,
DOROTHEE HUCHON⁵, EMMANUEL JP DOUZERY^{1,2} & FREDERIC DELSUC^{1,2}

¹Université Montpellier II, CC064, Place Eugène Bataillon, Montpellier Cedex 5, France

²CNRS, Institut des Sciences de l'Evolution (UMR5554), CC064, Place Eugène Bataillon,
Montpellier Cedex 5, France

³Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala S. Francesc 14, 17300
Blanes (Girona), Spain

⁴Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, Alaska, USA

⁵Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University,
Tel Aviv, 69978, Israel

⁶Department of Biology, University of Washington, Seattle WA 98195, USA

Research article

Open Access

An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models

Georgia Tsagkogeorga^{1,2}, Xavier Turon³, Russell R Hopcroft⁴, Marie-Ka Tilak^{1,2}, Tamar Feldstein⁵, Noa Shenkar^{5,6}, Yossi Loya⁵, Dorothee Huchon⁵, Emmanuel JP Douzery^{1,2} and Frédéric Delsuc^{* 1,2}

Address: ¹Université Montpellier 2, Institut des Sciences de l'Evolution (UMR 5554), CC064, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France, ²CNRS, Institut des Sciences de l'Evolution (UMR 5554), CC064, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France, ³Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala S. Francesc 14, 17300 Blanes (Girona), Spain, ⁴Institute of Marine Science, University of Alaska Fairbanks, Fairbanks, Alaska, USA, ⁵Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, 69978, Israel and ⁶Department of Biology, University of Washington, Seattle WA 98195, USA

Email: Georgia Tsagkogeorga - Georgia.Tsagkogeorga@univ-montp2.fr; Xavier Turon - xturon@ceab.csic.es; Russell R Hopcroft - hopcroft@ims.uaf.edu; Marie-Ka Tilak - Marie-Ka.Tilak@univ-montp2.fr; Tamar Feldstein - tfeldste@post.tau.ac.il; Noa Shenkar - nshenkar@u.washington.edu; Yossi Loya - yosiloya@post.tau.ac.il; Dorothee Huchon - huchond@post.tau.ac.il; Emmanuel JP Douzery - Emmanuel.Douzery@univ-montp2.fr; Frédéric Delsuc* - Frederic.Delsuc@univ-montp2.fr

* Corresponding author

Published: 5 August 2009

Received: 16 October 2008

BMC Evolutionary Biology 2009, 9:187 doi:10.1186/1471-2148-9-187

Accepted: 5 August 2009

This article is available from: <http://www.biomedcentral.com/1471-2148/9/187>

© 2009 Tsagkogeorga et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Tunicates have been recently revealed to be the closest living relatives of vertebrates. Yet, with more than 2500 described species, details of their evolutionary history are still obscure. From a molecular point of view, tunicate phylogenetic relationships have been mostly studied based on analyses of 18S rRNA sequences, which indicate several major clades at odds with the traditional class-level arrangements. Nonetheless, substantial uncertainty remains about the phylogenetic relationships and taxonomic status of key groups such as the Aplousobranchia, Appendicularia, and Thaliacea.

Results: Thirty new complete 18S rRNA sequences were acquired from previously unsampled tunicate species, with special focus on groups presenting high evolutionary rate. The updated 18S rRNA dataset has been aligned with respect to the constraint on homology imposed by the rRNA secondary structure. A probabilistic framework of phylogenetic reconstruction was adopted to accommodate the particular evolutionary dynamics of this ribosomal marker. Detailed Bayesian analyses were conducted under the non-parametric CAT mixture model accounting for site-specific heterogeneity of the evolutionary process, and under RNA-specific doublet models accommodating the occurrence of compensatory substitutions in stem regions. Our results support the division of tunicates into three major clades: 1) Phlebobranchia + Thaliacea + Aplousobranchia, 2) Appendicularia, and 3) Stolidobranchia, but the position of Appendicularia could not be firmly resolved. Our study additionally reveals that most Aplousobranchia evolve at extremely high rates involving changes in secondary structure of their 18S rRNA, with the exception of the family Clavelinidae, which appears to be slowly evolving. This extreme rate heterogeneity precluded resolving with certainty the exact phylogenetic placement of

Aplousobranchia. Finally, the best fitting secondary-structure and CAT-mixture models suggest a sister-group relationship between Salpida and Pyrosomatida within Thaliacea.

Conclusion: An updated phylogenetic framework for tunicates is provided based on phylogenetic analyses using the most realistic evolutionary models currently available for ribosomal molecules and an unprecedented taxonomic sampling. Detailed analyses of the 18S rRNA gene allowed a clear definition of the major tunicate groups and revealed contrasting evolutionary dynamics among major lineages. The resolving power of this gene nevertheless appears limited within the clades composed of Phlebobranchia + Thaliacea + Aplousobranchia and Pyuridae + Styelidae, which were delineated as spots of low resolution. These limitations underline the need to develop new nuclear markers in order to further resolve the phylogeny of this keystone group in chordate evolution.

Background

For more than a century, it has been known that tunicates (or urochordates) belong to chordates [1]. Traditionally occupying a basal position within chordates, they quickly became model organisms in evolutionary developmental studies aimed at understanding the origins of chordates. The complete genome sequence of the ascidian *Ciona intestinalis* revealed the essential toolkit of vertebrate genes in a small and compact genome that had not undergone the round of vertebrate-specific genome duplication events [2,3]. However, the *Ciona intestinalis* genome has also been evolving rapidly, with a number of lineage-specific innovations pointing to tunicates as a particularly interesting group for comparative genomic studies [4]. Meanwhile, recent phylogenomic analyses have refuted the classical view by demonstrating that tunicates, not cephalochordates, are the closest living relatives of vertebrates [5-9].

Despite their key position in the tree of life, our comprehension of the phylogenetic affinities within the tunicates is still limited. Traditionally, tunicates have been classified into three major classes – Ascidiacea, Appendicularia and Thaliacea – with distinct life-history traits and developmental modes. Among these classes, the Ascidiacea, commonly referred to as ascidians, is by far the most diversified group, with more than 2500 species classified on the basis of the structure of their branchial sac into three orders: Phlebobranchia, Aplousobranchia and Stolidobranchia [10-12]. Ascidians begin their lives as tadpole-like swimming larvae that later undergo a metamorphosis resulting in morphologically modified, sac-like sessile adults that can be either solitary or colonial. In contrast, the Thaliacea and Appendicularia classes consist of exclusively planktonic species. Appendicularia are peculiar in departing from the common developmental program of metamorphosis. They retain larval characteristics for their entire lifespan and are consequently of pivotal interest from an evolutionary developmental point of view [13].

Over the last decade, molecular phylogenetic analyses have been introduced to shed light on tunicate relationships. Following the pioneering study of Wada [14], the 18S rRNA gene had been the main marker used in reconstructing tunicate relationships at different taxonomic scales [15-19]. In all these studies, molecular phylogenies have contradicted the traditional classification, especially when higher taxonomic levels are considered. Molecular data refuted the division into the three classes (Ascidiacea, Thaliacea and Appendicularia), arguing instead for more complex evolutionary relationships among members of these morphologically distinct groups [20]. In overview, 18S rRNA data have clearly supported the paraphyletic nature of the class Ascidiacea by dividing the tunicates into the following three clades: Phlebobranchia + Thaliacea, Appendicularia, and Stolidobranchia, the last comprising Molgulidae and Pyuridae + Styelidae [15-19]. However, while this scheme is generally accepted, several major phylogenetic questions remain unanswered, among which at least three are prominent.

The first involves the long-standing question of the position of Appendicularia, which is crucial for understanding the evolution of body plans and developmental modes in tunicates [13]. One problem with positioning appendicularians in tunicate phylogeny is that the two 18S rRNA sequences from the genus *Oikopleura*, the only representatives currently sampled, evolved at an elevated rate [14]. This high evolutionary rate seems to be a genome-wide characteristic in *Oikopleura dioica* and leads to problematic long branches in inferred molecular phylogenies [6]. Consequently, in the majority of phylogenetic studies conducted so far, Appendicularia occupy a basal position in being the sister-group of all other tunicates [14,17,21]. This basal position, which might reflect a long-branch attraction artefact [17], has recently been challenged by analyses, including an improved taxon sampling, that instead favoured a sister-group relationship of Appendicularia with Stolidobranchia [19,20].

The second irresolution concerns the placement of the ascidian order Aplousobranchia within the tunicates. All

aplousobranch species are colonial and constituted by small zooids often embedded in a common tunic. In early 18S rRNA studies, it was proposed that Aplousobranchia form the sister-group of Appendicularia and that together they occupy the most basal position among tunicates [16]. Later on, phylogenies based on analyses of the mitochondrial cytochrome oxidase subunit I gene (*cox1*) suggested a close affinity of Aplousobranchia with Cionidae and Diazonidae (Phlebobranchia) [22], which is also supported by morphological characters [23]. Recently, it was shown that the spurious relationship of Aplousobranchia with Appendicularia [16] was an artefact due to the contamination of aplousobranch 18S rRNA sequences by protistan symbionts [18], and analyses of authentic 18S rRNA sequences clustered Aplousobranchia with Thaliacea [18]. However, the 18S rRNA sequences of Aplousobranchia seemed to evolve at extraordinarily high rates, yielding extremely long branches in the inferred phylogenetic trees [18]. This renders it particularly difficult to obtain a reliable placement of Aplousobranchia within the tunicates [19,20].

The third uncertainty pertains to the phylogenetic relationships of the thaliaceans within the ascidians. Despite the fact that rRNA studies have always placed Thaliacea and Phlebobranchia together [17-20], the exact phylogenetic position of Thaliacea remains unclear. This is mainly due to poor taxonomic sampling, with only three representatives of Thaliacea included to date in 18S rRNA phylogenies. The high evolutionary rate of 18S rRNA sequences in Thaliacea also precluded firmly resolving the phylogenetic relationships among its three constitutive orders.

Most of our knowledge of tunicate molecular phylogeny comes from analyses of 18S rRNA sequences using standard models of nucleotide sequence evolution [18,19]. Although among site-rate heterogeneity can be efficiently modelled by fitting a gamma-distribution [24] or by

assuming a proportion of invariable sites [25], these standard models make several assumptions, such as the independent evolution of nucleotide sites, and the spatial homogeneity of the substitution process across sites. In this respect, it has been demonstrated that neglecting the co-evolving paired sites in stems affects the estimation of bootstrap values [26,27] and also influences topological inference [28-30]. The spatial substitution pattern heterogeneity of sequence evolution may also produce misleading phylogenetic signals [31,32] which can severely affect phylogenetic reconstruction in cases of high substitutional saturation [33]. Our aim was thus to establish an updated phylogenetic framework for tunicate evolution based on analyses of the 18S rRNA gene, using the most realistic models that account for secondary structure and across-site heterogeneities in the evolutionary process. Phylogenetic analyses of a dataset incorporating 30 new complete 18S rRNA tunicate sequences allowed us to compare for the first time the use of RNA-specific paired-site substitution models [29,30,34] and the CAT mixture model [31], relaxing the hypothesis of a uniform substitution process by letting patterns of substitution be distinct at different sites.

Results and discussion

Improved phylogenetic models and 18S rRNA evolution

Table 1 summarises the selected models and methods used for phylogenetic inference of the two assembled datasets: a 110-taxon dataset including all tunicate sequences and a reduced 88-taxon dataset excluding the highly divergent aplousobranch species (see Material and Methods). The tree-reconstruction approaches we followed can be divided into three categories based on the type of substitution models used: (1) standard independent-site DNA models (e. g. GTR+ Γ +I); (2) doublet or paired-site substitution models (e. g. RNA6A+ Γ +I); and (3) the CAT-GTR non-parametric mixture model.

Table 1: Phylogenetic approaches and best-fitting models.

Datasets	Phylogenetic Reconstruction Approaches				
	Maximum Likelihood Standard DNA model	Standard DNA model	Bayesian Inference		
			Doublet model	Mixture model	
Complete dataset 110 taxa 1326 sites	GTR + Γ + I	GTR + Γ + I	Loops: GTR + Γ + I	Stems: RNA6C + Γ + I RNA7B + Γ RNA16A + Γ + I	CAT-GTR+ Γ
Reduced dataset 88 taxa 1650 sites	TN93 + Γ + I	GTR + Γ + I	Loops: TN93 + Γ + I	Stems: RNA6CA + Γ + I RNA7B + Γ RNA16A + Γ	CAT-GTR+ Γ

Contrary to standard models which assume that all sites evolve independently, doublet or paired-site models include a wide range of substitution models developed to overcome this assumption in rRNA molecules [35]. Ribosomal RNAs are a mosaic of two main structural motifs: single-stranded domains (loops) and double-stranded helices (stems) formed by Watson-Crick base pairing between nucleotides. During rRNA evolution, compensatory substitutions occur regularly in stems in order to maintain the counterpart-paired nucleotides. Doublet models, by considering pairs of nucleotides as states in the substitution matrices, account for this dependence.

There are three classes of doublet models – 6, 7 and 16 state models – differing in their treatment of non-Watson-Crick pairs (mismatches) [35]. For determining the most adequate model, we first selected the best-fitting standard DNA model for the unpaired loop partition, which we subsequently kept fixed in likelihood estimations, varying only the model for the stem partition [see Additional file 1]. For each of the three classes of doublet models, we used the AIC criterion to determine the best-fitting substitution model (i.e., the number of parameters of the transition matrix). However, we could not test whether the 6, 7 or 16 state models were the most appropriate because there is currently little agreement on whether it is possible to compare models with different matrix dimensions. Trees obtained under the best fitting model were almost identical among the three model classes (data not shown). We therefore only discuss those reconstructed using 6-state RNA models, which offer the best compromise between the increase in likelihood and the lowest number of free parameters incorporated.

In the third reconstruction approach, we used a mixture-model analysis to relax the assumption of a homogeneous evolutionary process across sites of the 18S rRNA molecule. The non-parametric mixture model CAT-GTR [31] was chosen for two reasons. First, previous studies using the CAT mixture model showed that it handles the substitutional saturation of the data better than classical homogeneous models [31], hence alleviating phylogenetic artefacts [33]; and second, because in contrast to other mixture models [32], it does not specify the number of mixture components *a priori*. More precisely, the CAT-GTR model assumes an infinite mixture (based on a Dirichlet process) of GTR matrices that differ only in their equilibrium base frequencies. Thus, the CAT model assigns to each site a frequency vector or so called "profile", with the number of different profiles (K) being a free parameter.

In statistical analyses, the selection of the model that best fits the data is a prerequisite. Currently, there is no generally accepted way to evaluate with confidence which of the afore-mentioned strategies (standard homogeneous

model, doublet models or CAT model) offers the best statistical fit for a given dataset. An adequate evaluation would require calculating the Bayes factor between each pair of models using thermodynamic integration [36], but such a strategy is not currently affordable computationally. Nevertheless, several studies have shown that doublet models outperform standard DNA models in phylogenetic reconstructions based on rRNA data [28,29,34,37].

With regards to the CAT-GTR mixture model versus standard site-homogeneous models, however, evaluations can be made directly from the posterior distribution of the number of profiles (K). If there is no heterogeneity across sites in the data, the CAT model will preferentially assign one single frequency vector for all sites [31]. In fact, the standard GTR model is nested within the CAT-GTR model, so that when the number of profiles is equal to one ($K = 1$), all sites evolve under a homogeneous GTR model. The distribution of the number of inferred profiles (K) for our two datasets shows that the frequency of the $K = 1$ class is zero in both cases (Figure 1). This implies that the standard GTR model has not been visited through the MCMC runs for either dataset. This means by extension that the site-heterogeneous CAT-GTR mixture model is statistically better at explaining the data than the standard homogeneous GTR model. Furthermore, the minimum number of predicted profiles is always higher than 20 (Figure 1), which suggests that 18S rRNA sequences are subject to more complex evolutionary pressures than those implied by the simple stem/loop partition.

This observation seems biologically relevant considering that beyond the pairing of nucleotides in stems, the rRNA spatial conformation is also affected by the base composition outside of these regions [38]. In this context, the maintenance of other structural motifs (T-loops, hook turns, GNRA tetraloops, bulged-G motifs, k-turns and A-minor motifs) is likely to be selectively constrained. Indeed, these motifs play an essential role in the enzymatic activity of rRNA and in its interaction with ribosomal proteins. Obviously, even if doublet models allow us to capture a major part of the constraints acting on rRNA, they are still far from modelling the overall evolutionary complexity of these molecules.

Evolutionary shifts in 18S rRNA sequences and phylogenetic relationships of Aplousobranchia

Most 18S rRNA sequences from species belonging to the divergent order Aplousobranchia are about 300 nucleotides longer than the other typical tunicate ones [18]. That is, sequences belonging to Didemnidae, Polyclinidae, and Polycitoridae, including the newly obtained *Cystodytes* sp., possess large insertions in multiple parts of the molecule when compared to the remaining tunicate

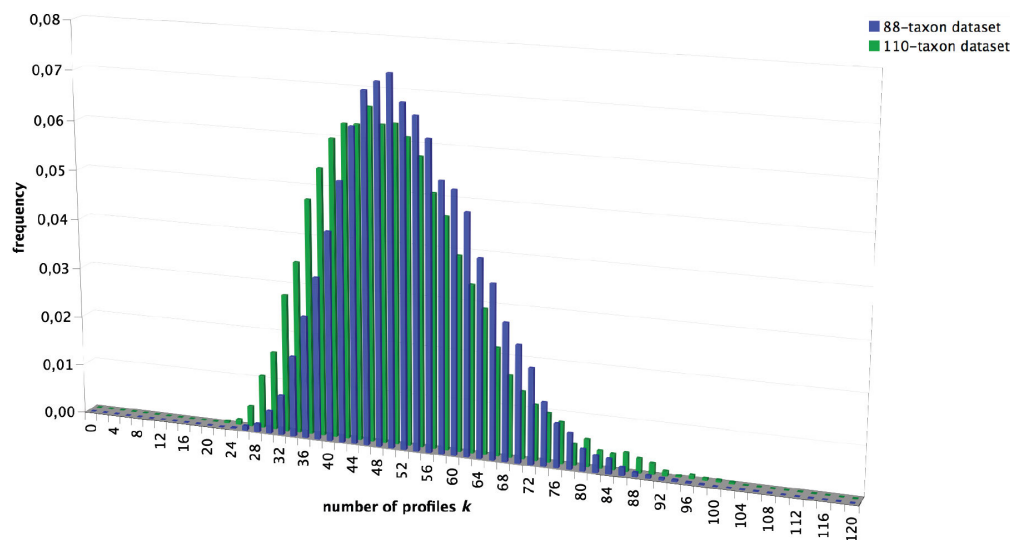


Figure 1

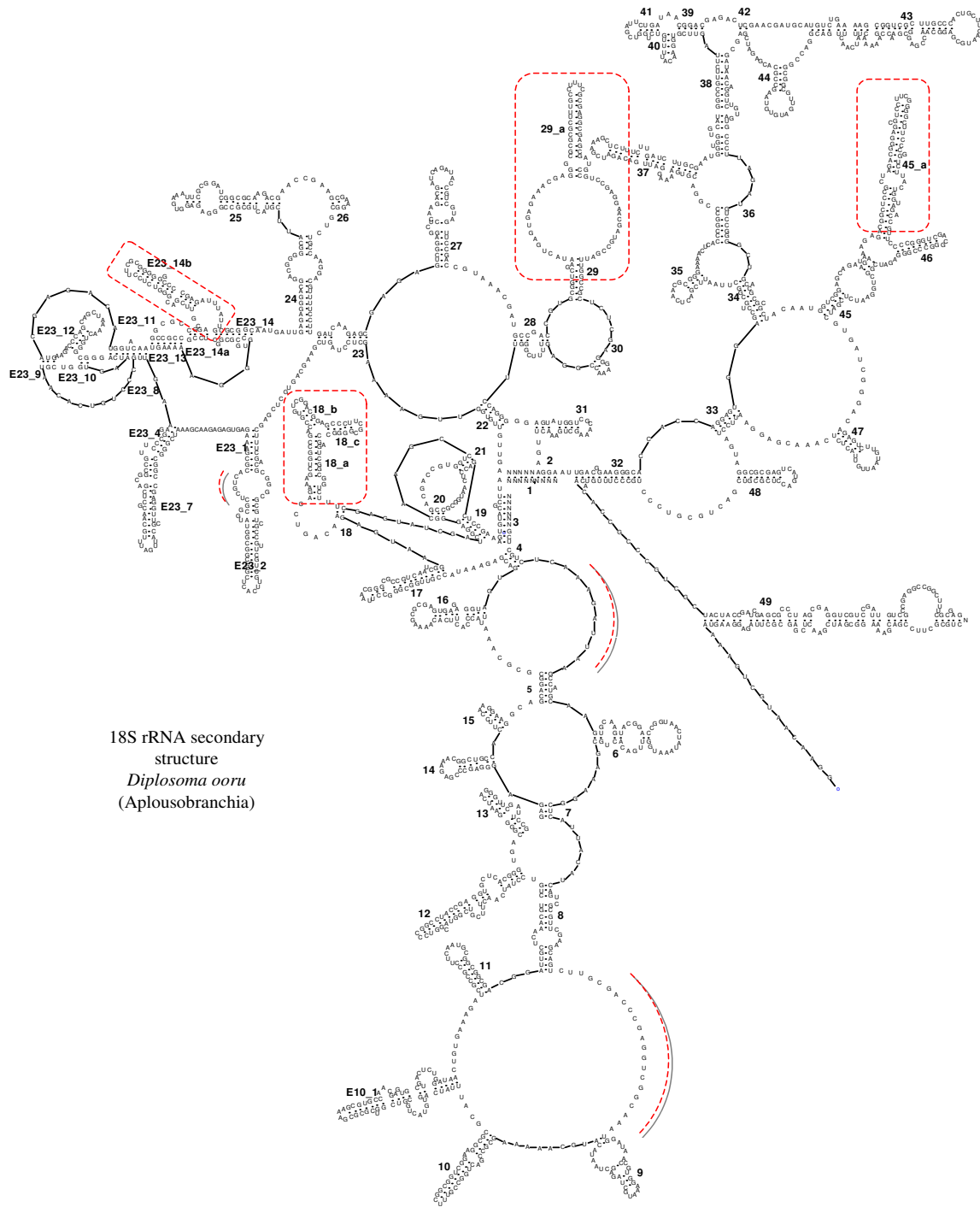
Estimated number of profiles K under the CAT-GTR+ Γ mixture model. The frequencies of the value (numbers of different profiles K), as estimated through the MCMC runs at the stationary stage for both the 110-taxa (green) and the 88-taxa (blue) datasets.

and outgroup species. These insertions take the form of elongations of already existing loops. More specifically, they occur in the unpaired helical regions between stems 4 and 5, and between stems 8 and 9, as well as within stems 18a-c, 29, 45a, E23_14b (Figure 2). Some of these insertions also induce changes to the secondary structure of 18S rRNA. Despite their high degree of divergence in terms of evolutionary rate and secondary structure, these unusual aplousobranch sequences are likely to be functional. Indeed, new helical structures are predicted in the elongated regions corresponding to helices 18, E23_14, 29 and 45 of the 18S rRNA molecule (Figure 2). Similar observations have also recently been reported in haplosclerid sponges [30] showing that a certain degree of freedom exists relative to the consensus 18S rRNA structure model. Knowing exactly which families belong to Aplousobranchia has been a matter of debate [23,39]. Here, we obtained sequences from two species belonging to the previously unsampled aplousobranch family Clavelinidae *sensu* Perez-Portela *et al.* [40] (*Clavelina meridionalis* and *Pycnoclavella aff. detorta*). Interestingly, these two sequences lack all the Aplousobranchia-specific structures described above, and thus show 18S rRNA sequence lengths more similar to other tunicate families.

Beside the greater gene length and the modified secondary structure, a clear compositional bias was detected for the Aplousobranchia by a Principal Component Analysis

(PCA) of base composition (Figure 3). The PCA revealed that the 18S rRNA sequences of the non-clavelinid Aplousobranchia have a different base composition than other tunicates and the outgroup species. That is, the sequences of Didemnidae, Polycitoridae and Polyclinidae are GC-rich. Conversely, Appendicularia and Molgulidae, are the most AT-rich according to the PCA results (Figure 3). The exact same compositional differences were evidenced when the loop and stem regions of the sequences were analysed separately (data not shown). Once again, the Clavelinidae sequences differ from those of other Aplousobranchia and have a base composition like that of typical tunicates and the outgroups (Figure 3).

The phylogenetic analysis of the complete tunicate dataset (110 taxa, including 95 tunicate species, and 1326 unambiguously aligned nucleotide sites: Table 1) revealed particularly long branches for the three Aplousobranchia families Polyclinidae, Polycitoridae and Didemnidae (Figure 4), as reported previously [18]. This dramatic rate acceleration was found regardless of the probabilistic method and the evolutionary model used. Furthermore, the newly obtained sequence from *Cystodytes sp.* (Polycitoridae) also had an elevated evolutionary rate and fell squarely within Aplousobranchia with high support. In striking contrast, the two sequences of the Clavelinidae evolved remarkably slowly compared to other Aplousobranchia species (Figure 4).



18S rRNA secondary structure
Diplosoma ooru
(Aplosobranchia)

Figure 2
Predicted 18S rRNA secondary structure of a divergent aplousobranch sequence (*Diplosoma ooru*). New predicted structures unique to such divergent Aplousobranchia species (and absent in the conserved *Pycnoclavella aff. detorta* and *Clavelina meridionalis* sequences) are boxed in red. Red dotted lines indicate additional loop regions where major elongations occurred in other divergent aplousobranchs.

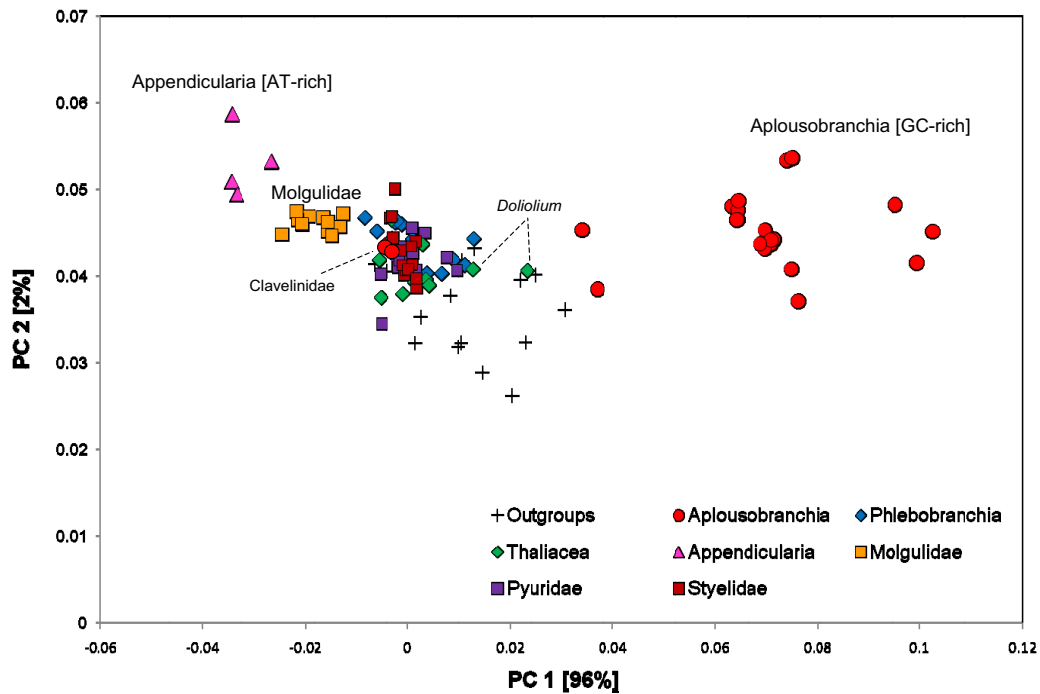


Figure 3

Analysis of base-composition heterogeneity. Principal component analysis (PCA) of the base composition of 18S rRNA from the 110-taxa dataset considering all nucleotide sites. The graph shows the first two principal components (PC), which contribute 96% and 2% of the total variance, respectively. The main component represents the variance along the AT versus GC axis, with the AT-rich Appendicularia, and the GC-rich Aplousobranchia at the two extremes.

From the phylogenetic point of view, the two slowly evolving Clavelinidae grouped together as a sister-group to the three divergent Aplousobranchia families in Bayesian analyses under the CAT mixture model and in Maximum Likelihood (ML) analyses, but with a lack of nodal support (Figure 4). This result is in agreement with what was recently proposed on the basis of the mitochondrial gene *cox1* [22]. However, monophyly of Aplousobranchia, including Clavelinidae, was not obtained in Bayesian analyses under secondary-structure models or site-homogeneous DNA models. These models instead yielded a tree topology where the divergent aplousobranch families (Polyclinidae, Polycitoridae and Didemnidae) clustered within Thaliacea to the exclusion of the Clavelinidae species. More precisely, a sister-group relationship to species of the genus *Doliolum* was obtained, but also with poor support values (data not shown).

Several arguments suggest that the previously proposed branching of Aplousobranchia within Thaliacea [18] might be a long-branch attraction (LBA) artefact [41] combined with a compositional effect. First, the topology is unstable and this grouping is not longer supported by the clavelinid sequences when the divergent aplousobranch species are removed (see Figure 5). Also, when the

Doliolum species were removed from the analyses, the Aplousobranchia affinity with the remaining Thaliacea was not recovered (data not shown). Second, *Doliolida* is the second fastest-evolving group in the Phlebobranchia + Thaliacea + Aplousobranchia clade and this could promote LBA to the fastest evolving Aplousobranchia. Third, the PCA of base composition revealed that the slightly GC-rich sequences from the *Doliolum* species are the closest in composition among tunicates to the deviant aplousobranch sequences (Figure 2), another potential cause of the observed phylogenetic artefact.

In conclusion, our results are consistent with the traditional view that the aplousobranch families Clavelinidae, Didemnidae, Polycitoridae and Polyclinidae form a monophyletic group which implies the occurrence within this order of drastic shifts in secondary structure, base composition, and evolutionary rate. Our phylogenetic analyses also provide firm support for their clustering with Phlebobranchia and Thaliacea species in a common clade. On the other hand, because of the peculiarities of 18S rRNA evolution in most species of Aplousobranchia, this gene does not allow establishing the phylogenetic relationships of this order with certitude. Also, the taxonomic sampling of 18S rRNA sequences is still insufficient



Figure 4
Phylogeny of tunicates inferred from the complete 18S rRNA dataset (110 taxa and 1373 sites). Bayesian majority-rule consensus tree obtained under the CAT-GTR+Γ mixture model implemented in PhyloBayes. Support values at nodes represent: Bayesian Posterior Probabilities (PP) obtained under: 1. PPI = CAT-GTR+Γ (PhyloBayes)/2. PP2 = RNA6C+Γ+I and GTR+Γ+I (Phase)/3. PP3 = GTR+Γ+I (MrBayes)/4. BP = Maximum likelihood bootstrap percentages (BP) under GTR+Γ+I (PAUP*). Support values are indicated for the main tunicate clades, and within Aplousobranchia, when PP ≥ 0.95 and BP ≥ 65. Newly sequenced Aplousobranchia species are underlined. Among the Stolidobranchia, a newly obtained sequence from *Botryllus schlosseri* is marked with an asterisk. The red triangle indicates the evolutionary shift in secondary structure of the 18S rRNA molecule within Aplousobranchia.

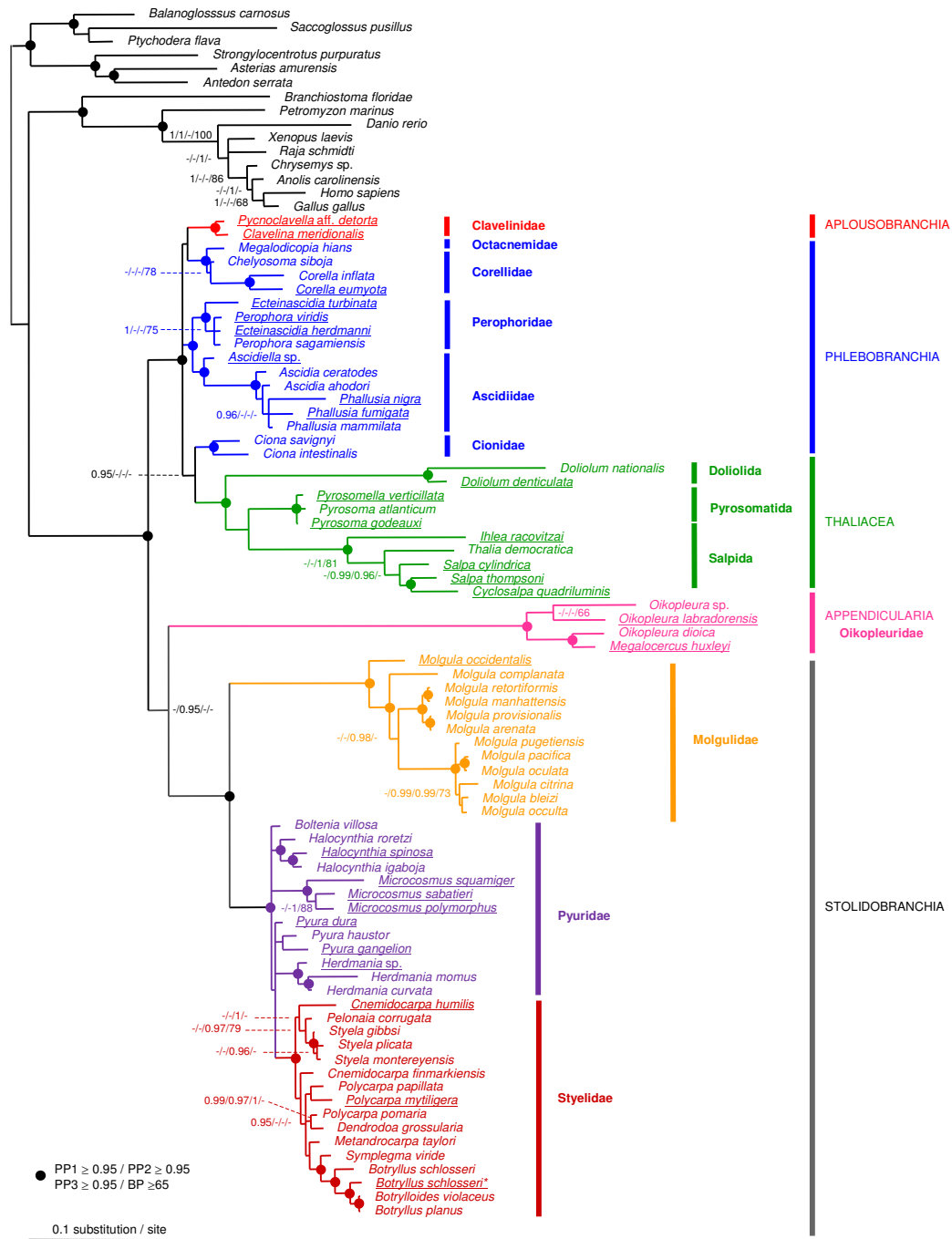


Figure 5
Phylogeny of tunicates inferred from a reduced 18S rRNA dataset (88 taxa and 1675 sites). Bayesian majority-rule consensus tree obtained under the CAT-GTR+ Γ mixture model implemented in PhyloBayes after exclusion of the fast-evolving Aplousobranchia species. Support values obtained using different reconstruction approaches are indicated at nodes in the following order: Bayesian posterior probabilities (PP) under: 1. PPI = CAT-GTR+ Γ (PhyloBayes)/2. PP2 = RNA6A+ Γ +I and TN93+ Γ +I (Phase)/3. PP3 = GTR+ Γ +I (MrBayes)/and 4. BP = Maximum Likelihood bootstrap percentages (BP) under TN93+ Γ +I (PAUP*). Support values are displayed when PP \geq 0.95 and BP \geq 65. Dots indicate nodes for which all four reconstruction methods agree and provide PP \geq 0.95 and BP \geq 65. Newly obtained sequences are underlined, including an additional one from *Botryllus schlosseri* marked with an asterisk.

within Aplousobranchia, because the diversity of this group is huge and may comprise more than the five families here included [23,42]. Further work is needed for gaining an accurate picture of Aplousobranchia evolution within tunicates.

An updated 18S rRNA phylogenetic picture for tunicates

The growing interest in tunicates, along with the realisation that they are the sister-group of vertebrates [5-9], reinforces the importance of reconstructing a reliable phylogenetic framework for their evolutionary history. Previous molecular studies have revealed that tunicate phylogenetic relationships are difficult to resolve, mainly because of marked differences in evolutionary rates among lineages [16-19,43]. Among others, Appendicularia, Aplousobranchia, and Thaliacea have been the typical examples of rapidly evolving groups. In this respect, we aimed at revisiting tunicate 18S-rRNA phylogeny by using a wider taxonomic sampling targeted at these crucial groups, and by applying a range of phylogenetic-reconstruction approaches in a probabilistic framework. Under this general scheme, we re-evaluated previously proposed relationships and the resulting evolutionary scenarios.

I. Ascidiacea paraphyly and tunicate systematics

The different reconstruction methods applied to the complete dataset unambiguously supported the division of tunicates into the following groups: Phlebobranchia + Thaliacea + Aplousobranchia, Appendicularia and Stolidobranchia (Figure 4). Bayesian analyses provided maximal branch support (PP = 1.0) for the respective monophyly of each group, and good ML bootstrap support was obtained for the monophyly of the clade Phlebobranchia + Thaliacea + Aplousobranchia (BP = 86). A monophyletic Stolidobranchia is also retrieved with strong branch support values grouping monophyletic Molgulidae with Pyuridae + Styelidae (Figure 4). Interestingly, this dataset strongly favoured Appendicularia as the sister-group of Stolidobranchia (BP = 90) (Figure 4).

However, the inclusion of the divergent aplousobranch sequences in this 110-taxon analysis precludes a reliable examination of the phylogenetic relationships among the slower-evolving taxa. In order to obtain a reliable picture of tunicate relationships, the divergent aplousobranch sequences were removed from subsequent phylogenetic analyses, keeping only the slowly evolving Clavelinidae as representatives. This resulted in a dataset with 88 taxa, including 73 tunicates, but encompassing an expanded set of 1650 unambiguously aligned sites (Table 1). Probabilistic analyses of this dataset under standard DNA models using ML or a Bayesian approach yielded congruent results. Comparing these results to the ones obtained in Bayesian analyses using mixture and partitioned-doublet

models, some topological differences were observed (compare the different support values in Figure 5).

Concerning the relationships among the major tunicate lineages, all probabilistic analyses of this reduced dataset yielded a phylogenetic pattern similar to that for the complete data set (compare Figure 5 with Figure 4). The two ascidian orders Aplousobranchia and Phlebobranchia again grouped strongly with Thaliacea. Appendicularia was unambiguously monophyletic, as were also Molgulidae, and the grouping of Pyuridae with Styelidae. The monophyly of Stolidobranchia was also strongly recovered (Figure 5). However, in contrast with analyses of the complete dataset, the support for Appendicularia as the sister-group of Stolidobranchia was decreased (Figure 5). Indeed, only the partitioned Bayesian analysis using RNA secondary-structure models provided some support for this relationship (PP = 0.95).

Our results therefore confirm that Ascidiacea, as currently defined, is a paraphyletic group that includes Thaliacea and, most likely, Appendicularia, as previously suggested [14,16-20]. Traditionally, ascidians were subdivided morphologically in two different ways (or combinations thereof): 1) based on the structure of the branchial sac, dividing ascidians into Aplousobranchia, Phlebobranchia and Stolidobranchia, as established by Lahille [10-12]; and 2) based on the position of the gonads, which separates ascidians into Enterogona and Pleurogona, following Perrier [44] as modified by Garstang [45]. Enterogona includes the Aplousobranchia and Phlebobranchia, while Pleurogona consists of Stolidobranchia (Figures 4 and 5). The finding that 18S rRNA joins Phlebobranchia and Aplousobranchia into a common clade supports the classification scheme of Enterogona and Pleurogona. Yet, accepting this view, Thaliacea branched in our trees within Enterogona, and Appendicularia with Pleurogona. The gonad position in the thaliaceans indeed conforms to the enterogonid type (associated with the gut), while in the appendicularians the highly modified metamorphosis renders any inference about the gonad position difficult.

II. Phlebobranchia and the sister-group Cionidae

Within Phlebobranchia, the five represented families – Ascidiidae, Perophoridae, Octacnemidae, Corellidae and Cionidae – appeared reciprocally monophyletic with high support values, except for Corellidae (Figure 5). At the inter-familial level, strong support was obtained for grouping Octacnemidae with Corellidae, and Perophoridae with Ascidiidae within which *Ascidia* appears divergent from the closely related genera *Phallusia* and *Ascidia* (Figure 5). The inclusion of Cionidae within Aplousobranchia instead of in its traditional position within Phlebobranchia has been suggested on the basis of both morphological characters [23] and analyses of *cox1* [22].

However, this grouping was not recovered in any of our trees, which are consistent with the traditional phlebobranch position of Cionidae. Our results support the paraphyly of Phlebobranchia, with a close phylogenetic affinity between the *Ciona* species and Thaliacea in Bayesian analyses using the CAT-GTR mixture model (PP = 0.95). Although not supported by more conventional reconstruction methods (Figure 5), such a relationship provides a good working hypothesis for future phylogenomic studies aiming at determining the sister-group of model species from the *Ciona* genus.

III. Thaliacea relationships

The Thaliacea are a class of exclusively planktonic tunicates and consist of three main orders: Doliolida, Pyrosomatida and Salpida. They are distinct from the other tunicates, because all members of this class form specialized colonies and have complex life cycles, characterized by the alternation between sexual and asexual generations. There were two traditional views on the origin of the class: the first suggesting that Thaliacea derived from an Appendicularia-like ancestor (Herdman 1888; Seeliger 1893 – 1911; Newman 1933 cited in [46]), and the second deriving Thaliacea from an ascidian-like ancestor [47–49] or even placing the split more basally, with the ascidians and thaliaceans evolving from a common ancestor [45]. In parallel to this controversy, there has been considerable debate on whether the thaliaceans are monophyletic or not [46]. Within Thaliacea, however, the commonly accepted view places the pyrosomatids in a basal position, with the doliolids and salps being sister-groups, following a trend toward branchial sac simplification and muscle-band development [47].

Meanwhile, early molecular phylogenies supported a close relationship between Thaliacea and Phlebobranchia thereby suggesting Thaliacea arose from a sessile ascidian-like ancestor [14]. Then, Thaliacea was also reported to be monophyletic in more recent studies [16,19,20]. On the phylogenetic affinities among its three constitutive orders, the phylogenetic evidence appears more ambiguous. In fact, in half of the previous 18S-rRNA tunicate phylogenies, Doliolida were found to be the sister-group of Salpida [14,19], whereas in others a closer relationship was recovered between Pyrosomatida and Salpida to the exclusion of Doliolida [16,17].

All previous sampling of Thaliacea was limited to the species *Pyrosoma atlanticum*, *Doliolum nationalis*, and *Thalia democratica*, among which the latter two were characterized by relatively long branches. This evolutionary-rate heterogeneity is probably the reason for the reported conflicting relationships among the three orders. In our study, we subdivided these long branches by tripling the taxonomic sampling which rendered the strong statistical sup-

port for the clustering of Thaliacea with Phlebobranchia and Aplousobranchia even more reliable (Figure 5). Although we still lack some representative families, the monophyletic nature of Thaliacea and the respective monophyly of its three constitutive orders were also firmly supported in all reconstructed trees (Figure 5). This is in agreement with previous taxonomic views (e.g., [47,48]).

Regarding thaliacean interrelationships, our phylogeny suggests a sister-group relationship between Pyrosomatida and Salpida, with Doliolida species arising first (Figure 5). Such a relationship was obtained in Bayesian analyses using mixture (PP = 0.69) and secondary structure (PP = 0.56) models whereas analyses conducted under standard DNA models (PP = 0.69; BP = 57) tend to group the two fast-evolving groups Salpida and Doliolida as reported in the most recent study [19]. These contrasting results suggested a possible effect of model misspecification leading to a potential LBA artefact grouping the fast evolving Salpida and Doliolida within Thaliacea in analyses conducted under standard DNA models. If confirmed, this particular case probably represents an example where the use of the most complex and best fitting models of sequence evolution allows improving phylogenetic inference [29,33].

Our proposed phylogenetic hypothesis seems at first counterintuitive since both doliolids and salps have solitary generations, in contrast to pyrosomes, which are exclusively colonial and are generally considered as the least specialized thaliaceans. However, the notion of a basal Pyrosomatida from which Doliolida and Salpida derive is at odds with the fact that only Doliolida in this group retain the larval features of tail and notochord. In contrast, our phylogenetic picture implies a single event of tail loss within the thaliacean lineage. In agreement with this view of a basal Doliolida, comparative studies of sperm morphology have shown that doliolid sperm is more plesiomorphic than that of the pyrosomes or salps [46,50]. Likewise, the simple arrangement of gill slits and the particular morphogenesis and structure of stolons in pyrosomes and salps may represent synapomorphic characters of the two groups [51].

IV. The phylogenetic position of Appendicularia

Appendicularians or Larvaceans are small planktonic tunicates that have short generation times and characteristic tadpole body plans, consisting of a short trunk and a posterior tail with a notochord and a dorsal neural tube. From a developmental point of view they are unique among tunicates, since all species belonging to this class retain the tail and the notochord in adulthood [52]. For this reason, the relationships of Appendicularia to the other tunicate lineages have always been considered as

crucial for understanding the evolution of chordate body plans [14,16]. The traditional classification distinguishes three families within the class Appendicularia: Fritillariidae, Kowalevskiidae and Oikopleuridae [53]. However, all previous molecular phylogenies have been taxonomically restrained to species from the model genus *Oikopleura* (Oikopleuridae). While early studies of 18S rRNA indicated that Appendicularia emerged first among tunicates [14,17], more recent analyses suggested a sister-group relationship with Stolidobranchia [19]. Their phylogenetic position remains controversial mainly because *Oikopleura* 18S rRNA sequences are highly evolving [17,43].

Our results, including two additional species belonging to Oikopleuridae (*Oikopleura labradorensis* and *Megalocercus huxleyi*), did not seem to break up the long appendicularian ancestral branch enough to firmly establish the position of this order within tunicates. Indeed, whereas the complete dataset strongly favoured Appendicularia as the sister-group of Stolidobranchia (Figure 4), in the reduced but less heterogeneous dataset (Figure 5), only the partitioned Bayesian analysis using RNA-secondary structure models provided support for this relationship (PP = 0.95). These results do not support the basal emergence of Appendicularia [14,16,17], suggesting instead their placement as a sister-group to Stolidobranchia, as recently proposed [19]. The most parsimonious interpretation of this internal position would imply that the planktonic Appendicularia derived from a sessile ascidian-like ancestor.

Still, a reliable placement for Appendicularia based on 18S rRNA data is hampered by two potentially confounding factors: the persisting long branches characterizing the group, and the similarity of nucleotide composition in Appendicularia and Molgulidae (Figure 2). In both, 18S rRNA sequences are AT-rich and experienced accelerated evolutionary rates, leaving open the possibility of an artefactual attraction of the two groups due to systematic biases. Finally, phylogenomic studies, although based on just a few taxa, strongly support the basal position of Appendicularia within tunicates [6,7]. However, *Oikopleura dioica*, the sole member of the class with currently available genomic data, evolves at such an elevated rate that any conclusion concerning its exact phylogenetic position based on these data would be premature.

V. Stolidobranchia

Stolidobranchia appears as unambiguously monophyletic in all phylogenetic reconstructions performed (Figures 4 and 5). Within Stolidobranchia, Molgulidae formed a monophyletic group characterized by accelerated evolutionary rates compare to other stolidobranchs (Figure 5). The basal emergence of *Molgula occidentalis* was significantly supported (0.97/1.0/1.0/88), followed by *Molgula*

complanata as the sister-group of two reciprocally monophyletic groups that include other molgulids, in good agreement with a previous study of the group based on partial sequences [15].

Overall, phylogenetic affinities within Molgulidae were rather well resolved and stable, whereas Pyuridae, and especially, Styelidae showed greater topological instability (Figure 5). Despite the addition of 10 species belonging to both Pyuridae and Styelidae, phylogenetic relationships among and within the two families were crippled by low resolution, with two multifurcations occurring within Pyuridae and poor support values at several nodes within this clade (Figure 5). In our trees, the family Styelidae is monophyletic whereas Pyuridae appears paraphyletic (Figure 5). These results contrast with those from a recent study of the stolidobranchs based on partial 18S rRNA and *cox1* sequences where the reverse situation was observed [54]. This incongruence might stand in a rooting problem of the Pyuridae-Styelidae group whereby its fast-evolving, and distant molgulid sister-group, created LBA effects [54]. In fact, only the relationships within Botryllinae were well-supported by all reconstruction methods. Furthermore, the inclusion of Botryllinae within Styelidae was strongly supported, upholding the view that the latter is in fact a subfamily of Styelidae [47,55]. Yet, little evidence is provided for the exact position of the Botryllinae, because its proposed sister-group relation to the colonial *Metandrocarpa taylori* [19] was not significantly supported (Figure 5).

Conclusion

Our work has improved tunicate phylogeny based on 18S rRNA, in order to provide a reliable phylogenetic framework for the evolution of this key group of chordates. This was accomplished by a wider taxonomic sampling than in previous studies, yielding a more thorough representation of major tunicate lineages, and subdividing previously long branches. Our results showed that the non-parametric CAT-GTR mixture model has a better fit than standard DNA models on tunicate 18S rRNA data. The high number of site profiles inferred under the CAT-GTR model also suggested that the *a priori* partition of 18S rRNA data into stem and loop regions is likely an oversimplification of the complex evolutionary constraints acting on tunicate 18S rRNA sequences.

The 18S rRNA gene provided a clear view of the evolution of major tunicate lineages, but appeared less informative for relationships at lower taxonomic levels. This was reflected by the weak support values accorded to several nodes in the reconstructed trees, and by the unstable position of certain groups between phylogenies inferred using different reconstruction approaches. Among the spots of low resolution, the relationships at inter- and intra-famil-

ial levels within Phlebobranchia and within the Pyuridae – Styelidae group proved difficult to clarify. The development of new phylogenetic markers is thus necessary for the further comprehension of tunicate evolutionary history. An obvious candidate is the 28S rRNA gene, which is already available for some species and whose resolving power in combination with the 18S rRNA has been widely demonstrated [43,56]. Finally, the increasing availability of genomic information for tunicates also constitutes a promising source of future nuclear protein-coding markers.

Methods

Taxon sampling

The tissue samples used in this study were collected from the following species: class Ascidiacea: Phlebobranchia: *Perophora viridis*, *Ecteinascidia herdmanni*, *Ecteinascidia turbinata*, *Asciidiella* sp., *Corella eumyota*, *Phallusia nigra*, *Phallusia fumigata*; Aplousobranchia: *Pycnoclavella* aff. *detorta*, *Clavelina meridionalis*, *Cystodytes* sp.; Stolidobranchia: *Molgula occidentalis*, *Halocynthia spinosa*, *Pyura dura*, *Pyura gangelion*, *Herdmania* sp., *Microcosmus squamiger*, *Microcosmus sabatieri*, *Microcosmus polymorphus*, *Polycarpa mytiligera*, *Botryllus schlosseri*, *Cnemidocarpa humilis*, class Thaliacea: *Doliolum denticulata*, *Pyrosomella verticillata*, *Pyrosoma godeauxi*, *Ihlea racovitzai*, *Salpa cylindrica*, *Salpa thompsoni*, *Cyclosalpa* sp.; and class Appendicularia: *Oikopleura labradorensis*, *Megalocercus huxleyi*.

DNA extraction, amplification and sequencing

Genomic DNA was extracted from 95% ethanol-preserved tissue samples using either the QIAamp DNA or the DNeasy Plant Mini kits following the manufacturer's protocols or following the procedure of Bernatzky and Tanksley [57]. The 18S rRNA gene was PCR-amplified either in one fragment (using the primer sets 18S1/18S2 or 18S1/18S_Herdm_R1) or in two overlapping fragments of approximately 1 kb, 18S1/18S4 and 18S3/18S2, using the following primers: 18S1 (Fwd) 5'-CCTGGTTGATCCTGCCAG-3', 18S2 (Rev) 5'-TAATGATCCATCTGCAGG-3', 18S3 (Fwd) 5'-TTAGAGTGTCAAAGCAGGC-3', 18S4 (Rev) 5'-GATTAAGAAAACATTCTTGCC-3' and the newly designed 18S_Herdm_R1 (Rev) 5'-GATTRACCCGAGACCGCMATTYGCRIT-3'. PCR products were purified from 1.2% agarose gel using Gel Extraction Kit (Millipore) or using polyethylene glycol (PEG) in saline (NaCl). Most of the products were directly sequenced using Big Dye Terminator v1.1 (Applied Biosystems) on an ABI 310 sequencer. Some PCR products were ligated into pGEM-T Easy vector for cloning into One Shot TOP10 Competent Cells (Invitrogen), and five clones per species were sequenced. All chromatograms were manually corrected and assembled using the software Sequencher. The 30 new sequences have been deposited

in the EMBL database under Accession Numbers [FM244840](#) to [FM244869](#).

Data assembly and alignments

In addition to the 30 sequences obtained for this study, more tunicate 18S rRNA sequences with > 85% length coverage were recovered from GenBank [see Additional file 2]. Fifteen outgroups were chosen to evenly sample the diversity of the other phyla of Deuterostomes, with representative species belonging to Vertebrata and Cephalochordata (Chordata), and Hemichordata and Echinodermata (Ambulacraria). In all analyses, the outgroup was used for rooting trees *a posteriori*. Primary multiple alignments were performed using MAFFT [58] and were subsequently adjusted by eye. Following secondary structure models for the 18S rRNA molecule, available in the European Ribosomal Database [59], two partitions or character groups were assigned to the sequences: (1) paired *stem* – characters forming helices in the secondary structure and (2) unpaired *loop* – characters forming single strands. Further alignment adjustments were made manually and by using the script Xstem [29] in order to ensure that all sites in predicted helices form AU, GC Watson-Crick or GU wobble bonds with their partner nucleotide. Finally, ambiguously aligned sites as well as sites including gaps in more than 50% of the sequences were excluded using Gblocks [60] set to the following parameters: Minimum Number Of Sequences For A Conserved Position = $n/2 + 1$ (where n = number of taxa), Minimum Number Of Sequences For A Flanking Position $\approx 0.85 \times n$, Maximum Number Of Contiguous Nonconserved Positions = 8, Minimum Length Of A Block = 2 or 5, Allowed Gap Positions = With Half.

Two final datasets were assembled. The first one maximised taxonomic representation by including the Aplousobranchia sequences obtained by Yokobori *et al.* [18] which yielded a matrix of 110 taxa (95 Tunicates + 15 outgroups) and 1373 nucleotide sites. The inclusion of the divergent Aplousobranchia sequences had hindered the alignment of the 18S rRNA regions corresponding to helices E23_1-E23_1', E23_2-E23_2', E23_7-E23_7' and 49-49'. So, the corresponding sites were removed from subsequent phylogenetic analyses. In the second dataset, in order to reduce potential reconstruction artefacts associated with divergent sequences due to long-branch attraction effects [41], the fast-evolving Aplousobranchia sequences were removed. This second dataset had 88 taxa (73 Tunicates + 15 outgroups) and 1675 nucleotide sites.

Secondary structure prediction and visualisation

The 18S rRNA consensus secondary structure was obtained by following eukaryotic models available in the European Ribosomal RNA Database [59]. Potential structures for specific aplousobranch insertions were predicted

using the RNAfold Web Server [61] by choosing the minimum free energy algorithm and the option to avoid isolated base pairs. The predicted secondary structures were visualised using the program RNAviz [62].

Phylogenetic analyses

Phylogenetic analyses were conducted using Maximum Likelihood (ML) and Bayesian Inference (BI) approaches. In analyses using standard DNA models of sequence evolution, for the entire dataset as well as for the unpaired *loop* partition, the best fitting model was selected using MODELTEST for ML [63] and MrMODELTEST for BI [64] based on the Akaike Information Criterion (AIC) [65]. The best-fitting RNA-specific model for the paired *stem* partition was selected by calculating the AIC from log-likelihoods values previously estimated using the program Optimizer of the PHASE 2.0 package [66]. The selected models for each dataset and for the two different phylogenetic-reconstruction strategies are shown in Table 1.

All ML analyses were performed using PAUP* 4.0b10 [67] using a three round successive-approximation approach for estimating model parameters [68]. Starting from a neighbour-joining (NJ) tree, model parameters were estimated under the likelihood criterion and further kept fixed for heuristic searches with Tree Bisection Reconnection (TBR) branch swapping. The ML tree topology was then kept fixed and model parameters were re-estimated, with the whole process being repeated twice. Statistical support for the nodes was obtained by Bootstrap resampling with 100 pseudo-replicates generated by the program SeqBoot 3.5 of the PHYLIP package [69]. In all replicates, ML analyses were conducted in parallel on a computing cluster using PAUP* through the same heuristic search approach, with model parameters fixed at values estimated previously. Bootstrap Percentages (BP) were obtained from the 50% majority rule consensus of the 100 reconstructed trees using the program TREEFINDER [70]. In all ML analyses, the 6 Ambulacriaria species were declared as outgroup taxa.

Bayesian analyses were conducted using the programs MRBAYES 3.1.2 [71], mcmcphase from the PHASE package [66] and PHYLOBAYES 2.3 <http://www.atgc-montpellier.fr/phylobayes/> [31]. Although both MRBAYES and mcmcphase permit partitioned analysis using RNA secondary structure models, a more complete list of doublet models is implemented in PHASE. Thus in the present study, MRBAYES was principally used for conducting Bayesian phylogenetic reconstructions under a single DNA model of sequence evolution, while mcmcphase was used when both DNA and RNA secondary structure models were considered. Finally, PHYLOBAYES was used for

reconstructions under the mixture model CAT, allowing for a general substitution process (CAT-GTR) [31].

In BI analyses conducted with MRBAYES, two independent runs of four incrementally heated Metropolis-coupled Markov chains Monte Carlo (MCMCMC) were launched and run for 5,000,000 generations. In BI analyses conducted with mcmcphase, a single MCMC was launched for 10,000,000 generations. In both cases, parameters and trees were sampled every 100 generations. In PHYLOBAYES, two independent MCMCs were launched for 16,000 to 20,000 cycles with parameters and trees being sampled every cycle (20,000 cycles correspond to about 1,500,000 generations).

In all BI analyses, priors were set to default values and the convergence of chains was double checked. First, values of the marginal likelihood and independent run discrepancies were monitored through generations. Second, the posterior probabilities of all splits were plotted at 20 cycle increments over a run, using the AWTY system [72] [see Additional file 3]. The burnin value was determined when both indicators entered a stationary phase. This typically involved eliminating about 25% of the run samples. Bayesian Posterior Probabilities (PP) were obtained from the 50% majority-rule consensus of the trees sampled during the stationary phase.

Authors' contributions

FD, XT and EJPD conceived and initiated the study. XT, RRH, YL and NS collected samples and identified specimens. GT, MKT and TF carried out PCR amplifications and sequencing. EJPD, FD, DH and YL supervised the work in their respective labs. GT and FD performed the phylogenetic analyses and drafted the manuscript. All other authors assisted in revising the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Secondary structure model selection. The table presents the selection of the best-fitting model of rRNA sequence evolution for the 88-taxon and 110-taxon datasets based on the Akaike Information Criterion (AIC). The best-fitting models within each class (6, 7, and 16 states) of doublet models are bold-faced.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-187-S1.pdf>]

Additional file 2

Species sampling, taxonomy and sequence accession numbers. The table indicate the taxonomy and the species sampling used in the present study with associated sequence Accession Numbers. The 30 new sequences obtained in this study are indicated with a star ().*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-187-S2.pdf>]

Additional file 3

Monitoring the convergence of MCMC in Bayesian analyses. The figure illustrates the post-analysis of chain convergence in Bayesian analyses under the CAT-GTR+ Γ_4 model for the 88-taxon dataset, using the AWTY system (Nylander et al. 2007). A. Cumulative plot of clade posterior probabilities of the 20 more variable splits over a run of 20,000 cycles (i.e. 1,500,000 MCMC generations) sampled at every cycle. The vertical red line indicates the determined burn-in value of 5,000. B. Comparisons of clade posterior probabilities between the two independent MCMC runs.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-9-187-S3.pdf>]

Acknowledgements

We would like to thank: Corinna Dubischar, Camille Martinand-Mari and Gabriel Gorsky for their help in obtaining precious tissue samples; Nicolas Galtier and Hervé Philippe for their helpful comments on this work; and Naomi Paz for English editing. We especially thank Nicolas Lartillot for his guidance on the use of the CAT model for this work. We also like to thank four anonymous reviewers for helpful comments and detailed suggestions that greatly improved the manuscript. This study is part of a Research Networks Program in Bioinformatics funded by the High Council for Scientific and Technological Cooperation between France and Israel (Grant # 3-3449 to DH, EJP and YL). The present work was carried out as part of the Galathea 3 expedition under the auspices of the Danish Expedition Foundation (Dansk Ekspeditionsfond). This is Galathea 3 contribution no. P42 and also contributes to the Census of Marine Zooplankton. This work benefited from the bioinformatics cluster of the Institut des Sciences de l'Evolution de Montpellier. This is contribution ISEM 2009-88 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 – CNRS).

References

- Kowalevski A: **Beiträge zur Entwicklungsgeschichte der Tunicaten.** *Nachrichten Gesellschaft Wissenschaften Göttingen* 1868, **19**:401-415.
- Cañestro C, Bassham S, Postlethwait JH: **Seeing chordate evolution through the *Ciona* genome sequence.** *Genome Biol* 2003, **4**:208.
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, et al.: **The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins.** *Science* 2002, **298**:2157-2167.
- Holland LZ, Gibson-Brown JJ: **The *Ciona intestinalis* genome: when the constraints are off.** *Bioessays* 2003, **25**:529-532.
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, et al.: **Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida.** *Nature* 2006, **444**:85-88.
- Delsuc F, Brinkmann H, Chourrout D, Philippe H: **Tunicates and not cephalochordates are the closest living relatives of vertebrates.** *Nature* 2006, **439**:965-968.
- Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H: **Additional molecular support for the new chordate phylogeny.** *Genesis* 2008, **46**:592-604.
- Dunn CW, Hejnal A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al.: **Broad phylogenomic sampling improves resolution of the animal tree of life.** *Nature* 2008, **452**:745-749.
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al.: **The amphioxus genome and the evolution of the chordate karyotype.** *Nature* 2008, **453**:1064-1071.
- Lahille MF: **Sur la classification des Tuniciers.** *C R Acad Sci Paris* 1886, **102**:1573-1575.
- Lahille MF: **Etude systématique des Tuniciers.** *C R Assoc Fr Adv Sci* 1887, **16**:667-677.
- Lahille MF: **Recherche sur les Tuniciers.** Paris: Faculté des Sciences de Paris; 1890.
- Nishino A, Satoh N: **The simple tail of chordates: phylogenetic significance of appendicularians.** *Genesis* 2001, **29**:36-45.
- Wada H: **Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny.** *Mol Biol Evol* 1998, **15**:1189-1194.
- Huber JL, da Silva KB, Bates WR, Swalla BJ: **The evolution of anural larvae in molgulid ascidians.** *Semin Cell Dev Biol* 2000, **11**:419-426.
- Stach T, Turbeville JM: **Phylogeny of Tunicata inferred from molecular and morphological characters.** *Mol Phylogenet Evol* 2002, **25**:408-428.
- Swalla BJ, Cameron CB, Corley LS, Garey JR: **Urochordates are monophyletic within the deuterostomes.** *Syst Biol* 2000, **49**:52-64.
- Yokobori S, Kurabayashi A, Neilan BA, Maruyama T, Hirose E: **Multiple origins of the ascidian-Prochloron symbiosis: molecular phylogeny of photosymbiotic and non-symbiotic colonial ascidians inferred from 18S rDNA sequences.** *Mol Phylogenet Evol* 2006, **40**:8-19.
- Zeng L, Jacobs M, Swalla B: **Coloniality and sociality has evolved once in Stolidobranch ascidians.** *Integr Comp Biol* 2006, **46**:255-268.
- Zeng L, Swalla B: **Molecular phylogeny of the protochordates: chordate evolution.** *Can J Zool* 2005, **83**:24-33.
- Kurabayashi A, Okuyama M, Ogawa M, Takeuchi A, Jing Z, Naganuma T, Saito Y: **Phylogenetic position of a deep-sea ascidian, *Megalodicopia hians*, inferred from the molecular data.** *Zool Sci* 2003, **20**:1243-1247.
- Turon X, Lopez-Legentil S: **Ascidian molecular phylogeny inferred from mtDNA data with emphasis on the Aplousobranchiata.** *Mol Phylogenet Evol* 2004, **33**:309-320.
- Kott P: **The Australian Ascidiacea. Part 2, Aplousobranchiata (1).** *Mem Qld Mus* 1990, **29**:1-299.
- Yang Z: **Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods.** *J Mol Evol* 1994, **39**:306-314.
- Lockhart PJ, Larkum AVW, Steel M, Waddell PJ, Penny D: **Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis.** *Proc Natl Acad Sci USA* 1996, **93**:1930-1934.
- Dixon MT, Hillis DM: **Ribosomal RNA secondary structure: compensatory mutations and implications for phylogenetic analysis.** *Mol Biol Evol* 1993, **10**:256-267.
- Galtier N: **Sampling properties of the bootstrap support in molecular phylogeny: influence of nonindependence among sites.** *Syst Biol* 2004, **53**:38-46.
- Jow H, Hudelot C, Rattray M, Higgs PG: **Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution.** *Mol Biol Evol* 2002, **19**:1591-1601.
- Telford MJ, Wise MJ, Gowri-Shankar V: **Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria.** *Mol Biol Evol* 2005, **22**:1129-1136.
- Voigt O, Erpenbeck D, Worheide G: **Molecular evolution of rDNA in early diverging Metazoa: first comparative analysis and phylogenetic application of complete SSU rRNA secondary structures in Porifera.** *BMC Evol Biol* 2008, **8**:69.
- Lartillot N, Philippe H: **A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process.** *Mol Biol Evol* 2004, **21**:1095-1109.
- Pagel M, Meade A: **A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data.** *Syst Biol* 2004, **53**:571-581.

33. Lartillot N, Brinkmann H, Philippe H: **Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model.** *BMC Evol Biol* 2007, **7**(Suppl 1):S4.
34. Erpenbeck D, Nichols SA, Voigt O, Dohrmann M, Degnan BM, Hooper JN, Worheide G: **Phylogenetic analyses under secondary structure-specific substitution models outperform traditional approaches: case studies with diploblast LSU.** *J Mol Evol* 2007, **64**:543-557.
35. Savill NJ, Hoyle DC, Higgs PG: **RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods.** *Genetics* 2001, **157**:399-411.
36. Lartillot N, Philippe H: **Computing Bayes factors using thermodynamic integration.** *Syst Biol* 2006, **55**:195-207.
37. Dohrmann M, Janussen D, Reitner J, Collins AG, Worheide G: **Phylogeny and evolution of glass sponges (porifera, hexactinellida).** *Syst Biol* 2008, **57**:388-405.
38. Wagner GP, Pavlicev M, Cheverud JM: **The road to modularity.** *Nat Rev Genet* 2007, **8**:921-931.
39. Monniot C, Monniot F, Laboute P: **Coral Reef Ascidiaceans of New Caledonia.** Paris: Orstom; 1991.
40. Perez-Portela R, Turon X: **Phylogenetic relationships of the Clavelinidae and Pycnoclavellidae (Ascidiacea) inferred from mtDNA data.** *Invertebr Biol* 2008, **127**:108-120.
41. Felsenstein J: **Cases in which parsimony or compatibility methods will be positively misleading.** *Syst Zool* 1978, **27**:401-410.
42. Moreno TR, Rocha RM: **Phylogeny of the Aplousobranchiata (Tunicata: Ascidiacea).** *Revista Brasileira de Zoologia* 2008, **25**:269-298.
43. Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J: **Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data.** *Mol Biol Evol* 2002, **19**:762-776.
44. Perrier E: **Note sur la Classification des Tuniciers.** *C R Acad Sci Paris* 1898, **124**:1758-1762.
45. Garstang W: **The morphology of the Tunicata, and its bearings on the phylogeny of the Chordata.** *Quart J Microsc Sci* 1928, **72**:51-187.
46. Holland LZ: **Fine structure of spermatids and sperm of *Doliolletta gegenbauri* and *Doliolum nationalis* (Tunicata: Thaliacea): implications for tunicate phylogeny.** *J Mar Biol* 1989, **110**:83-95.
47. Berrill NJ: **The Tunicata. With an Account of the British Species.** London: The Ray Society; 1950.
48. Brien P: **Embranchement des Tuniciers.** In *Traité de Zoologie, Tome XI Echinodermes, Stomocordés, Procordés* Edited by: Grassé PP. Paris: Masson; 1948.
49. Ujjanin BN: **Die Arten der Gattung *Doliolum* in Golfe von Neapel und den angrenzenden Meeresabschnitten.** Leipzig: Verlag von Wilhelm Engelmann; 1884.
50. Holland LZ: **Spermatogenesis in *Pyrosoma atlanticum* (Tunicata: Thaliacea: Pyrosomatida): Implications for tunicate phylogeny.** *J Mar Biol* 1990, **105**:451-470.
51. Lacalli TC: **Tunicate tails, stolons, and the origin of the vertebrate trunk.** *Biol Rev Camb Philos Soc* 1999, **74**:177-198.
52. Stach T, Winter J, Bouquet JM, Chourrout D, Schnabel R: **Embryology of a planktonic tunicate reveals traces of sessility.** *Proc Natl Acad Sci USA* 2008, **105**:7229-7234.
53. Fenaux R: **The classification of Appendicularia.** In *The biology of tunicates* Edited by: Bone Q. Oxford: Oxford University Press; 1998:295-306.
54. Perez-Portela R, Bishop JD, Davis AR, Turon X: **Phylogeny of the families Pyuridae and Styelidae (Stolidobranchiata, Ascidiacea) inferred from mitochondrial and nuclear DNA sequences.** *Mol Phylogenet Evol* 2009, **50**:560-570.
55. Kott P: **The Australian Ascidiacea. Part I, Phlebobranchia and Stolidobranchia.** *Mem Qld Mus* 1985, **23**:1-440.
56. Mallatt J, Winchell CJ: **Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star.** *Mol Phylogenet Evol* 2007, **43**:1005-1022.
57. Bernatzky R, Tanksley SD: **Genetics of actin-related sequences.** *Theor Appl Genet* 1986, **72**:314-321.
58. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment.** *Nucleic Acids Res* 2005, **33**:511-518.
59. Wuyts J, Perriere G, Peer Y Van De: **The European ribosomal RNA database.** *Nucleic Acids Res* 2004, **32**:D101-103.
60. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Mol Biol Evol* 2000, **17**:540-552.
61. Gruber AR, Lorenz R, Bernhart SH, Neubeck R, Hofacker IL: **The Vienna RNA websuite.** *Nucleic Acids Res* 2008, **36**:W70-74.
62. De Rijk P, Wuyts J, De Wachter R: **RnaViz 2: an improved representation of RNA secondary structure.** *Bioinformatics* 2003, **19**:299-300.
63. Posada D, Crandall KA: **MODELTEST: testing the model of DNA substitution.** *Bioinformatics* 1998, **14**:817-818.
64. Nylander JAA: **MrModeltest v2.3.** Evolutionary Biology Centre, Uppsala University.: Program distributed by the author; 2004.
65. Akaike H: **A new look at the statistical model identification.** *IEEE Trans Autom Contr AC* 1974, **19**:716-723.
66. Jow H, Gowri-Shankar V: **PHASE: a Software Package for Phylogenetics and Sequence Evolution.** University of Manchester, Manchester, UK: Distributed by the author; 2003.
67. Swofford D: **PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods).** In *Version 4.0b10 edn* Sunderland, Massachusetts: Sinauer Associates; 2002.
68. Sullivan J, Abdo Z, Joyce P, Swofford DL: **Evaluating the performance of a successive-approximations approach to parameter optimization in maximum-likelihood phylogeny estimation.** *Mol Biol Evol* 2005, **22**:1386-1392.
69. Felsenstein J: **PHYLIP (PHYLogeny Inference Package).** In *Version 3.66 edn* Seattle: Department of Genome Sciences. University of Washington; 2001.
70. Jobb G, von Haeseler A, Strimmer K: **TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics.** *BMC Evol Biol* 2004, **4**:18.
71. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572-1574.
72. Nylander JA, Wilgenbusch JC, Warren DL, Swofford DL: **AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics.** *Bioinformatics* 2008, **24**:581-583.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp



CHAPITRE III :

LES GENOMES MITOCHONDRIAUX ET
LA MITOGENOMIQUE DES TUNICIERS

1. Les génomes mitochondriaux

Les génomes mitochondriaux des animaux Bilatériens sont des molécules d'ADN circulaire de taille variant de 14 à 16 kb dont le contenu en gènes est extrêmement conservé (Gissi, Iannelli & Pesole 2008). Ces génomes contiennent typiquement les 13 gènes protéiques, deux ARN ribosomiques et 22 ARNt. Dépourvus d'introns, ils ne contiennent que des régions intergéniques courtes qui sont impliquées dans la régulation de la réplication et de la transcription (Boore 1999).

De par leurs caractéristiques particulières, les génomes mitochondriaux ont constitué un modèle de choix pour l'étude des processus déterminant l'évolution génomique (Boore & Brown 1998; Curole & Kocher 1999; Saccone *et al* 1999). Au cours de ces dernières années, les projets du séquençage du génome mitochondrial ont ainsi été particulièrement abondants et ils fournissent à l'heure actuelle une base de données conséquente (<http://www.ncbi.nlm.nih.gov/genomes/OrganelleResource>). En particulier, les génomes mitochondriaux disponibles dépassent les 1700 pour les Métazoaires dont 1244 Chordés !

L'échantillonnage est fortement biaisé en faveur des Crâniates qui dénombrent aujourd'hui 1229 génomes complets, alors que huit génomes de Céphalochordés ont été séquencés pour ce groupe à la diversité spécifique réduite à une trentaine d'espèces décrites. En ce qui concerne les Tuniciers, la représentation taxonomique du groupe apparaît très restreinte puisqu'il n'y a que six espèces pour lesquels on dispose du génome complet. Parmi ceux-ci, on compte un représentant des Thaliacea (*Doliolum nationalis*) et cinq ascidies (Yokobori *et al* 1999; Yokobori, Watanabe & Oshima 2003; Gissi, Iannelli & Pesole 2004; Yokobori, Oshima & Wada 2005; Iannelli *et al* 2007a; Iannelli *et al* 2007b) incluant quatre représentants des Phlebobranches (*Ciona intestinalis* type A et B, *Ciona savignyi*, *Phallusia fumigata*, *Phallusia mammillata*) et une espèce de Stolidobranche (*Halocynthia roretzi*).

Malgré la représentation disproportionnée des Tuniciers par rapport aux autres Chordés, l'étude comparative de ces génomes a déjà fourni un nombre d'éléments importants pointant vers le fait que les Tuniciers possèdent une dynamique évolutive singulière (Gissi, Iannelli & Pesole 2008). Une des principales caractéristiques de ces génomes est de casser la règle générale de la conservation de l'ordre des gènes dans le génome mitochondrial des Métazoaires. En effet, l'étude des premiers génomes de Tuniciers disponibles a révélé l'occurrence de réarrangements extensifs de l'ordre des gènes dans leurs génomes (Gissi, Iannelli & Pesole 2004), même entre espèces congénériques (Iannelli *et al* 2007a; Gissi,

Iannelli & Pesole 2008). De plus, les arbres phylogénétiques issus de l'analyse mitogénomique ont montré un taux d'évolution rapide des Tuniciers qui constituent d'ailleurs un réel écueil pour déterminer leur positionnement dans la phylogénie des Bilatériens de façon fiable (Bourlat *et al* 2006; Bourlat *et al* 2008; Bourlat *et al* 2009).

2. La divergence mitogénomique des Tuniciers au sein des Chordés

Dés la détermination du génome mitochondrial de l'Ascidie *Halocynthia roretzi* (Yokobori *et al* 1999) et celui de l'amphioxus *Branchiostoma floridae* (Boore, Daehler & Brown 1999) en 1999, l'évolution mitogénomique des Chordés commença se dessiner, avec notamment la mise en évidence d'une importante déviation quant aux caractères structuraux et fonctionnels des ADNmt de Tuniciers. Si les génomes mitochondriaux de Vertébrés et Céphalochordés apparaissent structurellement similaires (Boore, Daehler & Brown 1999), la divergence mitogénomique des Tuniciers par rapport à ce patron général recouvre de nombreux aspects : le code génétique, le contenu génique, l'orientation unidirectionnelle de la transcription, et les réarrangements de l'ordre des gènes (Gissi, Iannelli & Pesole 2008).

I. Code génétique

Le code génétique mitochondrial au sein des Chordés est globalement conservé avec des variations propres à chaque lignée, les différences portant sur l'interprétation du codon AGR (Gissi, Griggio & Iannelli 2009). Chez les Vertébrés, AGA et AGG signalent la terminaison de la traduction en tant que codon-stops, en revanche chez les Céphalochordés, les deux triplets correspondent à l'acide aminé Glycine (Gly) et l'usage du codon AGG reste controversé. Les Tuniciers quant à eux représentent encore un variant, en utilisant les triplets AGR pour coder l'acide aminé Serine (Ser).

II. Contenu en gènes

Les génomes mitochondriaux des Vertébrés et des Céphalochordés comportent typiquement 37 gènes : 13 codant pour des protéines de la chaîne respiratoire [sous-unités 6 et 8 de l'ATP-synthase (atp6 et atp8), les cytochrome-oxidases 1 à 3 (cox1, cox2 et cox3), le cytochrome b (cytb) et les sous-unités de la NADH-déshydrogénase (nd1, nd2, nd3, nd4, nd5,

nd6 et nd4L)], deux codant pour les ARN ribosomiques (ARNsr et ARNlr) et 22 gènes d'ARN de transfert qui constituent la machinerie de traduction de la mitochondrie.

Le répertoire génique est assez homogène entre les Vertébrés et les Céphalochordés avec quelques variations documentées en particulier pour des groupes d'Amphibiens et de Lépidosauriens. Les Tuniciers se distinguent par le fait qu'il disposent de deux ARNs de transfert additionnels : (i) l'ARNt-Gly, correspondant aux codons AGR en accord avec leur code génétique modifié, et (ii) l'ARNt-Met pour le triplet AUA, dont la présence semble contribuer à la diminution des conflits liés à l'utilisation de la Méthionine durant les phases d'initiation et d'élongation de la traduction (Xia *et al* 2007).

III. Distribution asymétrique des gènes sur les deux brins d'ADNmt

Chez la plupart des Métazoaires, ainsi que chez les Chordés, les gènes mitochondriaux sont distribués de manière asymétrique sur les deux brins d'ADN du génome. Ainsi, chez les Vertébrés on utilise les annotations « H (heavy)-strand » et « L (light)-strand » correspondant aux brins génomiques portant la majorité et la minorité des gènes, respectivement. Chez les Céphalochordés et les Vertébrés, le gène codant pour la sous-unité 6 de la NADH déshydrogénase (nd6) est codé sur le brin L. En revanche, chez les Tuniciers l'ensemble des gènes se situent sur le même brin et la transcription est unidirectionnelle (Gissi, Griggio & Iannelli 2009).

IV. Ordre des gènes

Les remaniements de l'ordre de gènes dans les génomes mitochondriaux des Vertébrés et des Céphalochordés sont modestes en impliquant quasi-uniquement des translocations des ARNs de transfert. En fort contraste avec ceci, l'ADN mitochondrial des Tuniciers présente une variabilité extrême quant à son architecture génomique. L'arrangement des gènes leurs génomes est en effet très différent de ceux des autres Chordés et même Deutérostomiens (Gissi, Iannelli & Pesole 2008). Plus précisément, il n'existe que deux gènes (cox2 et cytb) qui sont agencés et ordonnés de manière similaire entre les six génomes mitochondriaux de Tuniciers disponibles. Les génomes mitochondriaux de Tuniciers sont ainsi caractérisés par un taux de réarrangements de l'ordre des gènes extraordinairement élevé, la disposition de

gènes différant même entre espèces phylogénétiquement proches appartenant au même genre (ex. *Ciona* et *Phallusia*) (Iannelli *et al* 2007a).

3. Mitogénomique et relations phylogénétiques des animaux bilatériens

En vertu de certaines propriétés telles que sa taille relativement réduite, ses multiples copies par cellule, son absence de recombinaison, et son mode d'hérédité uniparental, le génome mitochondrial a été très largement utilisé en phylogénie des animaux (Curole & Kocher 1999). Certains auteurs ont même proposé que la comparaison de l'ordre des gènes dans les génomes mitochondriaux des bilatériens porterait un signal phylogénétique particulièrement intéressant (Boore 1999). Cependant l'essentiel des études a porté sur la concaténation des gènes protéiques, la mitogénomique dévoilant ainsi un signal phylogénétique assez informatif pour retracer l'histoire évolutive des animaux (Bourlat *et al* 2008).

Les arbres issus de l'analyse des génomes mitochondriaux semblent ainsi corroborer la nouvelle phylogénie des animaux (Adoutte *et al* 2000) avec la monophylie des Protostomiens et leur division en deux lignées majeures les Lophotrochozoa et les Ecdysozoa (Rota-Stabelli, Yang & Telford 2009). En revanche, les relations de parenté au sein des Deutérostomiens apparaissent beaucoup plus difficiles à déduire principalement à cause du positionnement problématique des Tuniciers. En effet, malgré les fortes valeurs de soutien statistique obtenues pour les Ambulacraria, la vaste majorité des arbres mitogénomiques retrouvent une proche affinité entre Vertébrés et Céphalochordés, alors que les Tuniciers apparaissent systématiquement à la base des Bilatériens (Bourlat *et al* 2006; Bourlat *et al* 2008). Seule une étude récente retrouve une topologie dans laquelle les chordés apparaissent monophylétiques mais où les Tuniciers occupent une position basale par rapport à un clade réunissant les Céphalochordés et les Vertébrés (Bourlat *et al* 2009) lorsque les données sont analysées sous le modèle site-temps hétérogène CAT-BP (Blanquart & Lartillot 2008).

Comme c'est le cas dans les jeux de données phylogénomiques et d'ARNr 18S, les Tuniciers présentent un taux d'évolution mitochondrial significativement plus élevé que celui des autres lignées de Deutérostomiens. L'occurrence d'erreurs systématiques en reconstruction phylogénétique étant souvent liée à l'hétérogénéité des processus substitutionnels, la divergence génomique extrême et le taux d'évolution rapide des génomes mitochondriaux de Tuniciers sont les deux facteurs soupçonnés d'être à l'origine du placement erroné du groupe dans les études mitogénomiques (Bourlat *et al* 2009).

4. Mitogénomique des Tuniciers : le génome mitochondrial d'*Herdmania momus* et la nouvelle phylogénie des Chordés [Article 3]

La disponibilité de six génomes mitochondriaux avec une dynamique évolutive particulière et le positionnement problématique des Tuniciers dans les arbres issus de la mitogénomique, furent les points de départ d'une étude plus approfondie. Le projet se développa autour d'un axe central, qui s'attacha à déterminer la séquence du génome mitochondrial d'un second représentant des Stolidobranchia : l'Ascidie solitaire *Herdmania momus*. Appartenant à la famille des Pyuridae, l'espèce *Herdmania momus* figure actuellement parmi les espèces signalées comme invasives du fait de son introduction récente en Méditerranée, alors qu'elle est native de l'Indopacifique (Shenkar & Loya 2008).

Comportant 13 gènes codant pour des protéines, 2 ARNr, et 24 ARNt, le génome nouvellement séquencé présente le contenu en gènes typique des Tuniciers. De plus, le fait que l'ensemble des gènes soient distribués sur le même brin d'ADN conforte l'hypothèse que la transcription unidirectionnelle des gènes mitochondriaux constitue probablement une caractéristique propre des génomes de Tuniciers. En revanche, l'architecture génomique révéla encore un nouvel ordre des gènes, *Herdmania momus* ne partageant qu'un seul bloc conservé de deux gènes avec *Halocynthia roretzi*, malgré leur proche parenté phylogénétique au sein des Pyuridae. De plus, contrairement à tous les autres Tuniciers séquencés, les gènes *cox2* et *cytb* ne sont pas contigus dans le génome d'*Herdmania momus* renforçant ainsi l'image d'une plasticité extraordinaire de l'ordre des gènes au sein du groupe (Gissi, Iannelli & Pesole 2008).

Afin d'approfondir la problématique du positionnement aberrant des Tuniciers dans les arbres mitogénomiques, un jeu de données phylogénétique fut assemblé à partir de la concaténation des 13 gènes protéiques mitochondriaux pour 54 taxons représentant les principaux phylums à l'échelle des Bilatériens. Des analyses phylogénétiques furent ensuite réalisées, en utilisant à la fois des modèles standards d'évolution des séquences, ainsi que le modèle de mélange site-hétérogène CAT dans ses versions stationnaire (Lartillot & Philippe 2004) et non-stationnaire CAT-BP (Blanquart & Lartillot 2008).

Les arbres phylogénétiques Bayésiens reconstruits sous des modèles classiques, tels que GTR ou MtREV, aboutirent à une image phylogénétique où les Tuniciers étaient regroupés de façon artéfactuelle avec *Xenoturbella bocki*. Le résultats obtenus en Maximum

de vraisemblance furent identiques à ceux des études précédentes où les Tuniciers apparaissaient en position basale au sein des Bilatériens. En plus d'un taux d'évolution accéléré, les génomes mitochondriaux de Tuniciers examinés présentèrent un biais de composition en acides aminés par rapport aux autres Bilatériens échantillonnés. L'utilisation des modèles de mélange site-hétérogènes stationnaire (CAT) et non-stationnaire (CAT-BP), permit néanmoins de retrouver la monophylie des Chordés, ainsi que le clade Olfactores regroupant Tuniciers et Vertébrés avec de fortes probabilités Bayésiennes associées. Ces résultats illustre l'importance du choix du modèle sur la reconstruction phylogénétique, en fournissant pour la première fois un soutien statistique pour la nouvelle phylogénie des Chordés (Bourlat *et al* 2006; Delsuc *et al* 2006; Delsuc *et al* 2008) à la lueur des données mitogénomiques.

ARTICLE III

ARTICLE ACCEPTE DANS **BMC GENOMICS**

SOUS RESERVE DE REVISIONS MINEURES

TUNICATE MITOGENOMICS AND PHYLOGENETICS: PECULIARITIES OF THE *HERDMANIA MOMUS* MITOCHONDRIAL GENOME AND SUPPORT FOR THE NEW CHORDATE PHYLOGENY

TIRATHA RAJ SINGH^{1*}, GEORGIA TSAGKOGEOGA^{2*}, FRÉDÉRIC DELSUC²,
SAMUEL BLANQUART³, NOA SHENKAR^{1,4}, YOSSI LOYA¹,
EMMANUEL J. P. DOUZERY² & DOROTHEE HUCHON¹

¹*Department of Zoology, George S. Wise Faculty of Life Sciences, Tel-Aviv University,
Tel Aviv 69978, Israel*

²*Institut des Sciences de l'Evolution (UMR 5554), CNRS-Université Montpellier II,
Place Eugène Bataillon, 34095 Montpellier Cedex 05, France*

³*Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
(UMR 5506), CNRS-Université Montpellier II, 161 rue Ada, 34392
Montpellier Cedex 05, France*

⁴*Department of Biology, University of Washington, Seattle WA 98195, USA*

**These authors contributed equally to this work*

Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny

**Tiratha Raj Singh^{1*}, Georgia Tsagkogeorga^{2*}, Frédéric Delsuc², Samuel
Blanquart³, Noa Shenkar^{1,4}, Yossi Loya¹, Emmanuel J. P. Douzery^{2§} and
Dorothee Huchon^{1§}**

¹Department of Zoology, George S. Wise Faculty of Life Sciences, Tel-Aviv
University, Tel Aviv 69978, Israel

²Institut des Sciences de l'Evolution (UMR 5554), CNRS-Université Montpellier II,
Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

³Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier
(UMR 5506), CNRS-Université Montpellier II, 161 rue Ada, 34392 Montpellier
Cedex 05, France

⁴Department of Biology, University of Washington, Seattle WA 98195, USA

*These authors contributed equally to this work

§Corresponding authors

Email addresses:

TRS: trajsingh@rediffmail.com

GT: Georgia.Tsagkogeorga@univ-montp2.fr

FD: Frederic.Delsuc@univ-montp2.fr

SB: samuel.blanquart@gmail.com

NS: nshenkar@u.washington.edu

YL: yosiloya@post.tau.ac.il

EJPD: Emmanuel.Douzery@univ-montp2.fr

DH: huchond@post.tau.ac.il

Abstract

Background

Tunicates represent a key metazoan group as the sister-group of vertebrates within chordates. The six complete mitochondrial genomes available so far for tunicates have revealed distinctive features. Extensive gene rearrangements and particularly high evolutionary rates have been evidenced with regard to other chordates. This peculiar evolutionary dynamics has hampered the reconstruction of tunicate phylogenetic relationships within chordates based on mitogenomic data.

Results

In order to further understand the atypical evolutionary dynamics of the mitochondrial genome of tunicates, we determined the complete sequence of the solitary ascidian *Herdmania momus*. This genome from a stolidobranch ascidian presents the typical tunicate gene content with 13 protein-coding genes, 2 rRNAs and 24 tRNAs which are all encoded on the same strand. However, it also presents a novel gene arrangement, highlighting the extreme plasticity of gene order observed in tunicate mitochondrial genomes. Probabilistic phylogenetic inferences were conducted on the concatenation of the 13 mitochondrial protein-coding genes from representatives of major metazoan phyla. We show that whereas standard homogeneous amino acid models support an artefactual basal position of tunicates among bilaterians, the CAT and CAT+BP site- and time-heterogeneous mixture models place tunicates as the sister-group of vertebrates within monophyletic chordates. Moreover, the reference phylogeny indicates that tunicate mitochondrial genomes have experienced a drastic acceleration in their evolutionary rate that equally affects protein-coding and ribosomal-RNA genes.

Conclusions

This is the first mitogenomic study supporting the new chordate phylogeny revealed by recent phylogenomic analyses. It illustrates the beneficial effects of an increased taxon sampling coupled with the use of more realistic amino acid substitution models for the reconstruction of animal phylogeny.

Background

Mitochondrial genomes (mtDNAs) of bilaterian animals are short, circular DNA molecules of 14-16 kb in length, typically characterized by the absence of introns and the presence of only short intergenic regions, with the exception of the control region (CR), a non-coding region assumed to contain the elements for the regulation of replication and transcription of the genome [1, 2]. To date, about 1,700 complete metazoan mitochondrial genomes have been sequenced and used in comparative mitogenomics and phylogenetic studies on different taxonomic scales [3-9].

The mitochondrial gene content is highly conserved across the different bilaterian phyla, with typically 37 genes [1, 2, 10]. Among them, 13 genes encode for proteins [ATP synthase subunits 6 and 8 (*atp6* and *atp8*), cytochrome oxidase subunits (*cox1*, *cox2*, *cox3*), apocytochrome b (*cytb*), and dehydrogenase subunits (*nd1*, *nd2*, *nd3*, *nd4*, *nd5*, *nd6* and *nd4L*)]. The remaining genes encode two ribosomal subunits (*srRNA* and *lrRNA*) and basically 22 tRNAs. However, cases of duplication and loss of tRNAs have been reported within bilaterians [11, 12]. Tunicate mitochondrial genomes illustrate such exceptions and typically encode 24 tRNAs, apart from two *Phallusia* species which lack the tRNA-Asp [13], and *Halocynthia roretzi* which encode two tRNA-phe (Gissi and Pesole 2003). The two additional tRNAs present in tunicate mitochondrial genomes when compared to vertebrates are (i) the tRNA-Gly (for AGR codons), which is necessary for the translation due to the derived tunicate mitochondrial genetic code [14], and (ii) the tRNA-Met (for AUA codon), whose presence might reduce the conflict between translation initiation — which requires a tRNA-Met (for AUG) — and translation elongation that involves AUG codons [15].

The mitochondrial gene order is highly conserved within Deuterostomia [10, 16], and particularly in chordate genomes. Conversely, mitochondrial gene arrangement shows an important plasticity in some animal phyla, (*i.e.* molluscs and nematods [1, 2]), and especially in tunicates [2, 13, 17-21]. Tunicates, or Urochordates, are marine deuterostomes characterized by markedly diversified developmental and life history traits, and traditionally encompass three major classes: Ascidiacea (sea squirts), Thaliacea (salps) and Appendicularia (larvaceans). Ascidiacea, commonly referred to as ascidians, is the most speciose and widespread group. Several ascidian species have been identified as invasive species, such as *Styela clava* and *Pyura praeputialis* [22-25], and have a strong ecological impact on the invaded marine ecosystems. Some species are also widely used as model organisms in evo-devo studies like *Ciona intestinalis* and *Botryllus schlosseri* [26-30]. According to the traditional classification, the class Ascidiacea is subdivided into three major orders: Phlebobranchia, Aplousobranchia, and Stolidobranchia. In contrast with this taxonomic view, 18S rRNA-based phylogenies have shown that ascidians are in fact paraphyletic [31-33]. According to the 18S rRNA phylogenetic framework, Aplousobranchia, Phlebobranchia, and Thaliacea are closely related, whereas Stolidobranchia forms a distinct and monophyletic group, which might be close to Appendicularia, although the position of the latter is still debated [31, 32].

To date, complete mitochondrial genomes of tunicates are mainly available for a single representative of Thaliacea (*Doliolum nationalis*) and five ascidians [13, 17-21], including four phlebobranchians (*Ciona intestinalis* type A and B, *C. savignyi*, *Phallusia fumigata*, *P. mammillata*) and one stolidobranchian (*Halocynthia roretzi*). The available mitochondrial data suggest that several unique features characterize mitogenomic evolution in tunicates relative to other chordate phyla. Two main

peculiarities can be distinguished. The first refers to the highly variable gene order observed within the group, which implies that extensive gene rearrangements have occurred even at low taxonomic levels [13, 17, 18]. However, since most available tunicate complete mtDNA sequences belong to phylogenetically-related species (except *H. roretzi*) according to the 18S rRNA reference [31-33], it is not possible to evaluate whether mitochondrial gene rearrangements characterize the whole order or only the Aplousobranchia + Phlebobranchia + Thaliacea clade. The second specificity is that of an accelerated evolutionary rate of tunicates, as revealed by the long branches of the group in mitogenomic topologies [34-36] and the associated composition bias [36]. However, whether this accelerated substitution rate is restricted to protein coding genes as in snakes [37] or is a more general feature of the whole mtDNA of Tunicates, has yet to be investigated.

These two peculiar evolutionary features of tunicate mitochondrial genome evolution have hampered their reliable phylogenetic placement within metazoans. Analyses of mitochondrial protein-coding genes have almost systematically placed tunicates outside Bilateria [19, 34, 35]. This is in sharp contrast with recent nuclear-based phylogenomic studies that identified tunicates as the closest living relatives of vertebrates within chordates [38-41]. Only one recent mitogenomic study has found marginal support for chordate monophyly, using a site- and time-heterogeneous mixture model in Bayesian phylogenetic reconstructions [36]. Yet, this study grouped cephalochordates with vertebrates according to the traditional Euchordata hypothesis [36].

Here, we sequenced the complete mitochondrial genome of the solitary ascidian *Herdmania momus* (Ascidiacea: Stolidobranchia: Pyuridae), an Indo-Pacific species that was introduced into the Mediterranean Sea via the Suez Canal [42]. We

describe the structural and compositional features of *H. momus* mtDNA, discuss its evolutionary dynamics with respect to other tunicate and chordate mitochondrial genomes, and provide an updated metazoan phylogeny based on probabilistic analyses of the 13 mitochondrial proteins using site- and time-heterogeneous mixture models of amino acid substitutions.

Results and Discussion

General features of *H. momus* mtDNA

The mitochondrial genome of *H. momus* (Figure 1) accounts for 15,816 base pairs (bp) in length which falls within the typical range of other tunicate genomes; the smallest genome being 14,579 bp long in *P. fumigata* and the largest one 16,351 bp in *D. nationalis*. It presents the typical tunicate gene content with all 13 protein-coding genes of the mitochondrial respiratory apparatus, including *atp8*, which appears so divergent in Tunicates that it had been difficult to annotate in the initial assemblies of *C. savignyi* and *H. roretzi* [17, 20, 21]. It also encodes for the two mitochondrial ribosomal genes, *srRNA* and *lrRNA*, and a total of 24 tRNAs (Figure 2), among which are two distinct tRNAs for the Gly (AGR and GGN), Leu (UUR and CUN), and Ser (AGY and UCN) codons (Figure 2), which is consistent with the modified mitochondrial genetic code of tunicates. Finally, *H. momus* mtDNA contains an additional tRNA-Met (AUA), similar to all other tunicate genomes sequenced so far [13, 17-21].

In contrast to other deuterostome mtDNA genomes, all genes are encoded on the same DNA strand and thus mitochondrial gene transcription shows only one orientation in *H. momus*, as in all other tunicate genomes sequenced to date. In chordates, although the majority of genes is co-orientated on one major DNA strand

(H-strand), a minimum set of tRNAs and the *nd6* gene are encoded from the minor one (L-strand) [2, 16, 43]. The genome architecture of *H. momus* thus provides additional evidence in support of the hypothesis that gene arrangement in only one strand is a shared derived feature of tunicate mtDNAs [2].

With regard to genome organization, *H. momus* mtDNA seems substantially less compact than other available tunicate mitochondrial genomes. Adjacent genes overlap in four cases: *cox1* – tRNA-Val (9 bp), *nd4L* – tRNA-Leu (15 bp), *atp8* – tRNA-Asp (1 bp) and *nd2* – tRNA-Trp (2 bp). The total gene overlap thus appears slightly lower with respect to the average ~35 bp gene overlap of the other tunicate genomes [13]. Interestingly, there is no protein-protein gene overlap. Furthermore, *H. momus* shows a high percentage of non-coding (NC) sequence (ca. 10%).

Concerning base composition, the mitochondrial genome of *H. momus* is AT rich with a value of 63.5%. The overall base frequencies as estimated in the coding strand follow the decreasing order T, G, A and C. T is by far the most abundant nucleotide with 41.3% in the genome. Most tunicate mtDNAs present similar compositional AT-rich profiles varying from 61.1% in *D. nationalis* to 78.6% in *C. intestinalis*. The only exception concerns the mitochondrial genomes of species from the genus *Phallusia*, which present more balanced AT and GC compositions with 52.8% and 53.2% AT in *P. mammillata* and *P. fumigata*, respectively [13].

Protein-coding genes

The predicted protein-coding genes in the mtDNA of *H. momus* present lengths that are overall similar to their orthologues in the other tunicate genomes. Nevertheless, *H. momus* was not predicted to carry any deletion at the C-terminal ends of *nd4* and *nd5* genes similar to those reported in *H. roretzi* [20].

The base composition of protein-coding genes in *H. momus* was measured as being A+T rich, as is the profile of its entire genome. Consequently, the codon usage follows the same pattern, with C-rich codons showing lower frequencies than those that are T- or G- rich. Extreme cases are constituted by the CGC codon for Arginine, which is never used, and by the UUU codon for Phenylalanine and UUG codon for Leucine, which are the most frequently used. The CGC codon is also never used in the two *Ciona* genomes, whereas it is present in other tunicate genomes [17, 21].

The preferred start and stop codons in *H. momus* mitochondrial genes are GTG (6 genes) and TAG (7 genes), respectively. In the other Pyuridae (*H. roretzi*), GTG is the preferred start codon too, but TAA is the most frequent stop codon (8 genes). Like in other tunicate genomes, incomplete T-starting stop codons are predicted at the end of the three genes coding for *nd1*, *nd4* and *nd4L*. The complete TAA stop codon is probably created by polyadenylation.

rRNA genes

Because of high sequence divergence of tunicate rRNA sequences, the boundaries of all tunicate mitochondrial rRNA sequenced so far have been inferred from the flanking genes. Likewise, the *srRNA* gene in *H. momus* is estimated at most as 673 bp long, thus being slightly shorter than its acidian orthologues which range from 687 to 738 bp, and about 30 bp longer than the *srRNA* of *D. nationalis* mtDNA. On the genome map, it is found located between the tRNA-Ala and tRNA-Glu. The *lrRNA*, on the other hand, appears to be at most 1,159 bp long, a length similar to that of other tunicates, and is located between downstream of tRNA-Tyr and upstream of *cytb* (Figure 1).

Both lengths and locations of the two rRNA genes in the *H. momus* mitochondrial genome provide additional evidence for two unique features of mtDNA

evolution in tunicates. The first is that tunicates present the shortest mitochondrial rRNA genes among chordates, and more generally among deuterostomes, with lengths varying at most from 641 to 738 bp for the small ribosomal subunit and from 1,059 to 1,279 bp for the large ribosomal subunit [13]. The corresponding lengths for the *srRNA* and *lrRNA* genes in the other chordate lineages are estimated as higher than 844 bp and 1,367 bp, respectively. The second feature shared by all tunicates concerns the location of the two genes in the mitochondrial genome. In contrast to all other chordates, where the two rRNA coding genes are usually adjacent, *srRNA* and *lrRNA* are found rearranged and separated in all available tunicate mtDNAs [2]. The distance between the two genes in *H. momus* genome is about 5 kb. Distances of the same order of magnitude (5-7 kb) were also identified in the mtDNAs of Cnidaria, some Protostomia, and some Echinodermata [2].

tRNA genes

The predicted cloverleaf structures for the 24 tRNA genes of the *H. momus* mitochondrial genome are presented in Figure 2. As in *H. roretzi*, only two tRNAs are characterized by an unusual structure: tRNA-Ser (AGY) and tRNA-Asn. tRNA-Ser (AGY) appears to possess an 8 bp-long anticodon stem structure. Similar long stems (7-9 bp) have been reported in *Ciona* ssp., *D. nationalis*, and *H. roretzi* tRNA-Ser (AGY) [17, 19-21]. However, this structure is absent in *Phallusia* tRNA-Ser (AGY), which appears to have retained the classical cloverleaf conformation [13]. Concerning the tRNA-Asn gene, two alternative cloverleaf structures have been inferred, as also previously reported for *H. roretzi* [20]. The structure predicted by tRNA-scan (Figure 2) is characterized by a two-nucleotide spacer between the DHU and the anti-codon stem as well as a mismatch at the start of the T stem. The alternative conventional cloverleaf structure (Figure 2) appears less stable, in forming a mismatch at the start

of the D arm. Similar two-nucleotide spacers have also been found in the tRNA-Asn structures of *Phallusia* ssp., *H. roretzi*, and *D. nationalis*, but not in *Ciona* ssp [13, 17, 19-21].

Non-coding regions

Table 1 summarizes some attributes of the non-coding (NC) regions in *H. momus* and the other available tunicate mitochondrial genomes. In total, 28 NC regions have been identified in *H. momus* mtDNA, with a total size of 1,457 bp (9.21% of the total genome). When compared to other ascidians, the proportion of NC regions in *H. momus* appears higher, with previous estimates varying from 1.94 to 5.89%, in the mtDNAs of *P. mammillata* and *P. fumigata*, respectively. The proportion of NC sequences in the closely-related *H. roretzi* is three-fold lower (3.09%) than that of *H. momus*, yet the total number of NC fragments is almost identical in both species, and the two genomes differ by ca. 1 kb in length. Among tunicates, only *Doliolum nationalis* has more non-coding regions with 13.1% of the total genome length (Table 1).

Considering length distributions, 19 of the 28 NC regions present lengths over 20 bp, with the longest accounting for 197 bp and being located between the *nd6* and tRNA-Tyr genes (Figure 1). Three additional NC regions measuring more than 100 bp have been identified in *H. momus* mtDNA, involving the disjunction of the following gene pairs: *cytb* - *nd2*, *nd4* – tRNA-His, and tRNA-Lys – *nd3*. The remaining nine NC sequences are shorter intergenic spacers, mostly 3-10 bp long, distributed homogeneously through the genome.

The presence of palindrome sequences was checked by sequence similarity searches and has been detected within 12 NC regions, and in particular within the three longest NC regions of the *H. momus* mitochondrial genome. However, no stem-

loop structure similar to that of the control region of vertebrate mt-genomes has been identified in *H. momus*. Further studies are necessary to establish whether these regions are involved in the control of replication and transcription.

Sequence similarities searches have finally revealed the occurrence of a duplication event in the NC regions of the *H. momus* genome, similar to those previously determined in *Ciona intestinalis* and *Phallusia fumigata* mtDNAs [13]. Here, a fragment of 89 bp was found duplicated in two different locations of the genome: the first, situated in the non-coding region downstream of the *nd6* gene and upstream of the gene encoding for the tRNA-Tyr (NC = 197 bp); while the second involved the non-coding region downstream of the tRNA-Asn (NC = 88 bp) and 13 bp of the 5' region of the tRNA-Lys. The estimated distance between the repeated motifs is about 5 kb (Figure 1).

Gene order

The mtDNA of *H. momus* shows a novel gene arrangement with respect to other tunicates, and is radically different from that of its close relative *Halocynthia roretzi* (Figure 3). Surprisingly, only one block containing two consecutive genes is conserved between the two species: the pair tRNA-Trp – tRNA-Ile. When considering only the relative arrangement of protein-coding genes, an additional block appears to be shared by the two members of Stolidobranchia, consisting of the three genes *cytb* - *nd2* - *nd5*. In *H. roretzi*, *cytb* and *nd2* are separated by four tRNA-coding genes, whereas in *H. momus* the two genes are separated by a long non-coding region of 157 bp. Likewise, the number of tRNA genes separating *nd2* and *nd5* as well as their identities differ in each genome (two and five tRNAs separate these two genes in *H. momus* and *H. roretzi* respectively). Conversely, the pair *cox2*- *cytb*, previously considered as unique and shared by all other tunicates is absent in *H. momus*, as well

as the pair tRNA-Arg – tRNA-Gln which is conserved in most tunicate genomes [13, 17-21].

Thus, the mitochondrial genome organization in *H. momus* when compared to that of *H. roretzi* suggests that extensive gene rearrangements occur within Stolidobranchia, similar to those observed in Phlebobranchia [13]. Future work should determine whether co-generic species in Stolidobranchia also present high variations in gene order, as reported for the genera *Ciona* and *Phallusia* within Phlebobranchia [2].

Phylogenetic analyses

The mitochondrial genome has been repeatedly used in molecular phylogenetics of animals, principally due to several convenient features, such as its relatively small size, its cellular abundance, and its mostly uniparental mode of inheritance facilitating orthology assignment [5]. However, the phylogenetic resolving power of mtDNAs is often reduced by pervasive lineage-specific variations of base composition and substitution rate [44, 45]. Bilaterian evolutionary relationships obtained from previous phylogenetic reconstructions based on mitochondrial genes [34, 35] are in relatively good agreement with the currently accepted view of Protostomia phylogeny [46].

However, the situation is more complex concerning Deuterostomia, mainly because of the systematic placement of tunicates at the base of Bilateria [34-36]. Such placement outside chordates is in direct contradiction to results from recent large-scale phylogenomic studies, which strongly support the clade Olfactores, *i.e.*, the grouping of tunicates with craniates (jawed vertebrates + cyclostomes) to the exclusion of cephalochordates [38-41]. The basal position of tunicates within Bilateria is generally interpreted as a long-branch attraction artifact caused by the peculiarities of tunicate

mitogenomic evolution in terms of both lineage-specific evolutionary rate and amino-acid composition [35, 36].

In our mitogenomic dataset, the potential occurrence of compositional biases was explored through a Principal Component Analysis (PCA) of amino acid composition of the 54 taxa (Figure 4). This statistical analysis shows that tunicates have a markedly heterogeneous amino acid composition that is clearly distinct from most other sampled taxa. Moreover, tunicates appear extremely divergent from the other chordate representatives. Tunicates and jawed-vertebrates are located at the extreme opposite sides of the graph, with cephalochordates in an intermediate position. There is thus a strong compositional heterogeneity in our amino-acid dataset that might cause phylogenetic artefacts if it is not specifically accounted for in models of sequence evolution.

The maximum likelihood tree obtained on the 13 protein-coding genes concatenation using the site- and time-homogeneous mtREV+ Γ_4 model of amino-acid sequence evolution places tunicates as the sister group of the remaining Bilateria, as previously reported [34, 35]. However, the bootstrap support for this basal position is low (BP = 61). This result is presumably the outcome of two types of systematic biases: compositional heterogeneity of the tunicate lineage compared to other chordate and animal phyla (see Figure 4) combined with accelerated evolutionary rates [36]. ML phylogenetic analyses were also conducted under an empirical site-heterogeneous CAT mixture model [47] which has been shown to alleviate long-branch attraction artefacts in the context of animal phylogenomics [48]. The ML tree inferred under this model shows a dramatic topological shift in placing tunicates as the sister-group of craniates within monophyletic chordates and deuterostomes, in

agreement with phylogenomic studies [34, 39]. However, this phylogenetic placement is not statistically supported (BP < 50).

In order to further explore the effect of model misspecification on mitogenomics tree inference, Bayesian analyses have been conducted using a range of amino-acid sequence evolution models (Table 2). There is a clear distinction between the results obtained using site-homogeneous models (mtREV+ Γ_4 and GTR+ Γ_4) on the one side, and variants of the CAT+ Γ_4 site-heterogeneous model [49] and the site- and time-heterogeneous CAT+BP+ Γ_4 model [50], on the other side. Site-homogeneous models moderately support a sister-group relationship between tunicates and *Xenoturbella* (PP \geq 0.83), whereas the different CAT mixture models support the grouping of tunicates with craniates (*i.e.*, monophyly of Olfactores) with the highest Bayesian posterior probabilities obtained using the most complex and better-fitting models (PP \geq 0.94). These results exemplify the impact of the model of sequence evolution on mitogenomic inference of phylogeny by showing that the new chordate phylogeny [34, 38, 39] can be corroborated by using a model that accounts for both compositional and evolutionary rate heterogeneities.

The Bayesian consensus tree obtained under the site- and time-heterogeneous CAT+BP+ Γ_4 model is detailed in Figure 5. The overall phylogenetic picture is consistent with the common tripartite structure of Bilateria phylogeny [51], with Protostomia and Deuterostomia, and Protostomia being further divided into Lophotrochozoa and Ecdysozoa. Strong statistical support is obtained for the monophyly of Protostomia (PP = 1), and for both of its two major lineages Lophotrochozoa and Ecdysozoa (PP \geq 0.99). The monophyly of Deuterostomia is less strongly supported however (PP = 0.89), in agreement with recent mitogenomic [36] and phylogenomic results [39, 52]. Within Deuterostomia, Echinodermata and

Hemichordata were both retrieved as firmly monophyletic (PP = 1) and are strongly grouped into Ambulacraria, as suggested by early mitogenomic studies [53] and later confirmed by phylogenomics [34, 39, 41]. However, the position of *Xenoturbella* within Deuterostomia remains unresolved in our analysis (PP < 0.5) as also found in the latest mitogenomic analysis [36], whereas phylogenomics support its sister-group relationship with Ambulacraria into a clade named Xenambulacraria [34, 39]. Strong statistical support is obtained for the respective monophyly of the three chordate groups: Cephalochordata (PP = 1), Tunicata (PP = 1), and Craniata (PP = 0.99). The monophyly of Chordata is strongly supported (PP = 0.96) for the first time in a mitogenomic study, as is the monophyly of Olfactores (PP = 0.94). These results are fully congruent with the growing body of evidence coming from phylogenomic studies that revealed the unexpected sister-group relationship between tunicates and craniates [34, 38, 40, 41].

Finally, our mitogenomic data provided strong evidence for intra-tunicate relationships, with the overall phylogenetic scheme being well resolved despite the high lineage-specific evolutionary rate (Figure 5). The inferred relationships for the group were notably in concordance with nuclear-based phylogenies and morphological data. More precisely, *Herdmania* was unambiguously retrieved in a sister-group relationship with *Halocynthia* (PP = 1), in agreement with the systematic classification that places these two ascidian genera within the stolidobranch family Pyuridae [32]. Firm evidence was similarly obtained for the respective grouping of the congeneric species of *Ciona* and *Phallusia* (PP = 1), and high posterior probability is retrieved for a sister-group relationship of the two genera into Phlebobranchia (PP = 1). Finally, the thaliacean *Doliolum* is found to branch within ascidians with solid

statistical support (PP = 0.99), favoring the hypothesis of Ascidiacea paraphyly [19] as also supported by 18S rRNA phylogenies [31, 33, 54].

Compared evolutionary rates of rRNA- and protein-coding genes

A detailed analysis of lineage-specific evolutionary rate was performed by inferring branch lengths on the ML reference phylogeny (see Figure 5). The comparison of branch lengths obtained on the protein- and rRNA-coding mitochondrial partitions, first demonstrates that tunicates are clearly the fastest evolving species of the metazoan dataset on both partitions (Figure 6). It also shows that the rRNA-coding partition evolves on average about three times more slowly than the protein-coding gene partition (TBL ratio = $34.79 / 11.76 = 2.96$). However, in contrast to what has been observed in snake mitochondrial genomes [37, 55], there is very good correlation between root-to-tip distances inferred from the two partitions ($R^2 = 0.88$). These results reveal that tunicate mitochondrial genomes have experienced a drastic acceleration of evolutionary rate that affects both protein-coding genes and ribosomal-RNA genes. This latter observation seems therefore to exclude the occurrence of adaptive evolution episodes in protein-coding genes, recently reported for snake and agamid lizard mitochondrial genomes [56], as the ultimate cause behind the elevated substitution rates in tunicate mitochondrial genomes. Nevertheless, the lineage-specific rate acceleration of tunicates probably explains the difficulties previously encountered in reconstructing their phylogenetic position based on mitogenomic data.

Conclusions

The complete mitochondrial genome of the ascidian *H. momus* shares several features with the other tunicate mtDNAs so far sequenced: (i) all genes are encoded on the same strand; (ii) two additional tRNAs, when compared with vertebrates, are present

(one for Gly, one for Met); (iii) the gene order is highly rearranged, and differs from all other known tunicates, with the noticeable synteny disruption of the *cox2*—*cytb* gene block; and (iv) the evolutionary rate is elevated and involves both protein- and rRNA-coding genes. From the phylogenetic viewpoint, the analysis of the concatenated mitochondrial proteins recovers for the first time the tunicates + craniates clade within monophyletic chordates, provided that site- and time-heterogeneous mixture models of amino acid replacements are used.

Methods

Sampling and DNA extraction

The *H. momus* (Savigny, 1816) individual used for our study was collected in June 2006 on the Eilat-Ashkelon pipeline pier in Eilat, Red Sea 29°31' N 34°55' E at a depth of 12m [42]. Genomic DNA was isolated from gonads by placing ~0.5 cm³ of tissue in 1 ml of lysis buffer (100 mM Tris-HCl pH=7.8; 100 mM EDTA; 0.5% SDS; 0.2 mg/ml proteinase K) overnight at 55°C. The digestion was followed by Phenol:Chloroform:Isoamyl alcohol extraction (25:24:1) and isopropanol precipitation [57].

DNA sequencing and gene annotation

Primers were designed to match conserved regions of chordate genomes in the *cox1*, *cox2*, *cox3*, and *cytb* genes (Table S1). PCR amplifications were performed using the ExTaq polymerase (TaKaRa) in 25 µL reaction mixture according to manufacturer's instructions and were conducted in two steps. A first amplification was performed with external primers followed by a re-amplification of the initial PCR product using different sets of internal primers. Based on the partial sequences obtained for the

aforementioned genes, longer specific primers were subsequently designed. This allowed us to determine the *cox3-cox1* region and to redesign (long) primers in the *nd3* gene located in the same region.

The complete mitochondrial genome was finally amplified in two overlapping fragments using the Long and Accurate (LA) Taq polymerase (TaKaRa) in 25 µL reaction mixture according to manufacturer's instructions. The first fragment covered the *cox3-cytb* region (~7,700 bp) and was obtained by PCR amplification using the external primers

HMO_Cox3_Long_F1 5'-ACTGTTGTTTTACTTAGTTCGGGAGTTACTGTGAC-3' and HMO_Cob_Long_R1

5'-ACATATAAGCGACCCCCAACAAAAAGAAAC-3' followed by a re-amplification with the primers

HMO_Cox3_Long_F2 5'-AGTTTTGGTGGAGGGCTTATGGGATTAGTTTGGAC-3' and HMO_Cob_Long_R2

5'-GACACCAGAATAGGGCCAAAAATAAATCTTT-3'. The second fragment encompassed the *cytb-nd3* region (~8,900 bp) and was similarly acquired by a first PCR amplification using the external primers HMO_Cob_Long_F1

5'-AGGGGTGCTACTTTAACGCGGTTTTACAC-3' and

HMO_Cox1_Long_R2 5'-ACATTATACAACCTGCCCATCTCCAATTACCTG-3' and a subsequent re-amplification using the primers HMO_Cob_Long_F2

5'-AGTGGTTTTGTCTTTGGTGCATTTGGTTTTTC-3' and NADH3_R 5'-CTG AAT AGA ATA ACC CTC GCT ATC ACT-3'. The two fragments overlapped over ~ 150 bp at 5' end of the *cytb* and over ~ 780 bp at the *cox3-nd3* region. The sequencing of the long PCR products was performed by the company MACROGEN (Seoul, Korea) using shotgun sequencing (i.e., shearing of the PCR fragment,

bacterial library construction, and sequencing of the clones). Contigs were assembled using both ARACHNE [58] and Chromaspro (Technilysium Pty. Ltd.) programs. Regions of low coverage in the assembly (i.e., fewer than three clones) were confirmed by primer-walking sequencing (~ 4 kb of the genome was re-sequenced) with the use of specific primers and genomic DNA as template. The complete mtDNA sequence of *H. momus* was deposited under the EMBL accession number FN296153.

Protein coding genes were identified using the ORF-Finder tool of NCBI [59] and the DOGMA server [60]. Following Iannelli et al. [13], boundaries of protein coding gene were inferred in such a way that overlap between genes was minimized and similarity between tunicate proteins was maximized. ATG and non-standard initiation codons [61] were considered as reliable start codons.

Both the tRNAscan-SE 1.23 program and the DOGMA server were used to identify and infer the secondary structures of tRNA genes [62]. tRNAs that were not identified by the above tools were sought by folding all putative non-coding regions using the Mfold server [63]. All predicted cloverleaf secondary structures were manually checked and compared with other known ascidian tRNAs. The positions of the small (*srRNA*) and large (*lrRNA*) ribosomal RNA genes were identified by the DOGMA server and confirmed based on sequence similarity searches to orthologous genes in other ascidians. Finally, the boundaries of the rRNA genes were inferred from the flanking genes.

Dataset assembly

The phylogenetic dataset was built upon the taxon sampling used by Bourlat *et al.* [34]. We expanded the tunicate sampling to seven by including the newly-sequenced *H. momus* and two recently published genomes from the genus *Phallusia* [13]. The

snake representative (*Boa constrictor*) was excluded because snake mitochondrial genomes have been shown to be the subject of adaptive evolution that can severely bias phylogenetic inference [56]. The final dataset therefore comprises 54 taxa including 31 deuterostomes, 19 protostomes, and four non-bilaterian outgroups (*cf.* Additional file 1). The nucleotide sequences of all mitochondrial genes were retrieved from the Organellar Genome Retrieval (OGRe) database [64]. For the 13 protein coding genes, sequences were translated, and aligned at the amino-acid level using MAFFT with default parameters [65]. Ambiguously aligned sites were identified and removed from all individual genes separately, using the program Gblocks [66] with the following parameters: minimum number of sequences for a conserved position = 18; minimum number of sequences for a flanking position = 29; maximum number of contiguous nonconserved positions = 8; minimum length of a block = 2; allowed gap positions = with half. The concatenation of the 13 proteins yielded a phylogenetic dataset including 54 taxa and 2,136 unambiguously aligned amino-acid sites, of which 1687 were variable (Additional file 2).

Phylogenetic analyses

Phylogenetic analyses were conducted using Maximum Likelihood (ML) and Bayesian Inference (BI) reconstruction approaches on the concatenated amino acid dataset. Maximum Likelihood (ML) analyses were performed using the program PHYML 3 [67] under the models mtREV+ Γ_4 and CAT+ Γ_4 , with a 4-category Gamma (Γ_4) distribution of the among-site amino acid replacement rate heterogeneity [68], and with the number of CAT categories set to 20 (C20) as recommended by [47]. The heuristic ML searches were conducted by performing Subtree Pruning and Regrafting (SPR) moves on a Neighbor-Joining (NJ) starting tree. Statistical support was estimated by Bootstrap resampling with 100 pseudo-replicates generated by the

program SeqBoot of the PHYLIP package [69]. In all replicates, ML analyses were performed using PHYML through the same heuristic search strategy. Nodal bootstrap supports (BS) were obtained from the 50% majority rule consensus of the 100 reconstructed trees using the program TREEFINDER [70].

Bayesian inference was conducted using the program PhyloBayes 3.1 [71] under both homogeneous amino acid models (mtREV+ Γ_4 and GTR+ Γ_4) and variants of the site-heterogeneous CAT+ Γ_4 mixture model (CAT+Poisson+ Γ_4 , CAT+mtREV+ Γ_4 , and CAT+GTR+ Γ_4) [49]. The program nhPhyloBayes [72] was used to performed Bayesian analysis under a site- and time-heterogeneous model which combines the break-point approach (BP) in order to model variations of amino acid replacement rates along branches and the CAT mixture model in order to account for site-wise variations of these rates. In this CAT+BP+ Γ_4 analysis, the number of categories of the mixture component was fixed at 60, and the biochemical profiles were those inferred by Le et al. [47], rather than being estimated.

In Bayesian analyses under mtREV+ Γ_4 , GTR+ Γ_4 , and under the different variants of CAT+ Γ_4 (CAT+Poisson+ Γ_4 , CAT+mtREV+ Γ_4 , and CAT+GTR+ Γ_4), three independent Markov Chain Monte Carlo (MCMC) were run in parallel, whereas five MCMC were run for analyses under CAT+BP+ Γ_4 . Each MCMC was launched from a random initialization and for a large number of cycles with parameters and trees saved every cycle. Priors were set to values as described in Blanquart and Lartillot [50] concerning the CAT+BP model, and in Lartillot et al. [71] concerning all other applied models. Convergence of MCMC was checked by monitoring the marginal likelihood through cycles. Bayesian Posterior Probabilities (PP) were obtained from the trees sampled during the stationary phase of the different MCMC.

Branch length analysis

Two concatenated nucleotide datasets were constructed to compare evolutionary rates of protein-coding and rRNA-coding mitochondrial genes. In both cases, the different genes were individually aligned using MAFFT, and ambiguously aligned sites were removed using Gblocks before building each concatenation. This led to a 13 protein-coding gene dataset containing 8,130 nucleotide sites (Additional file 3) and a rRNA dataset including 806 sites (Additional file 4). Using the rooted topology obtained from the analysis of the amino-acid sequences under the CAT+BP+ Γ_4 mixture model as a reference, ML branch lengths were optimized under the GTR+ Γ_8 model using PAUP* 4.0b10 [73] on both the protein-coding and rRNA-coding nucleotide concatenations. For each dataset, the root-to-tip distance was calculated for each of the 54 taxa by summing the branch lengths on the path going from the root of the tree to each of its terminals. The two sets of root-to-tip distances were then compared graphically to evaluate the lineage-specific rate correlation between the two mitochondrial partitions.

Authors' contributions

FD, EJPD and DH conceived and initiated the study. YL and NS collected and identified the biological specimens. TRS carried out DNA extraction, PCR amplifications, and sequencing. TRS and DH annotated the mitochondrial genome. GT, SB and FD performed the phylogenetic analyses. FD, EJPD, YL, and DH supervised the work in their respective labs. GT, FD, EJPD and DH wrote the manuscript. All other authors assisted in revising the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We would like to thank Tamar Feldstein and Isadora Cohen for their help at the bench, Benoit Nabholz for sharing scripts and Naomi Paz for editing the text. TRS was supported by a fellowship from the Israel National Committee for Planning and Budgeting Universities (VATAT). This work has been funded by the Research Networks Program in BIOINFORMATICS of the High Council for Scientific and Technological Cooperation between France and Israel to DH, EJPD and YL, and have benefited from the ISE-M bioinformatics cluster. This publication is contribution N° 2009-YYY of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 - CNRS).

References

1. Boore JL: Animal mitochondrial genomes. *Nucleic Acids Res* 1999, 27:1767-1780.
2. Gissi C, Iannelli F, Pesole G: Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 2008, 101:301-320.
3. Boore JL: The use of genome-level characters for phylogenetic reconstruction. *Trends Ecol Evol* 2006, 21:439-446.
4. Boore JL, Brown WM: Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev* 1998, 8:668-674.
5. Curole JP, Kocher TD: Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol Evol* 1999, 14:394-398.
6. Telford MJ, Herniou EA, Russell RB, Littlewood DT: Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc Natl Acad Sci U S A* 2000, 97:11359-11364.
7. Waugh J, Huynen L, Millar C, Lambert D: DNA barcoding of animal species-response to DeSalle. *Bioessays* 2008, 30:92-93.
8. Zink RM, Barrowclough GF: Mitochondrial DNA under siege in avian phylogeography. *Mol Ecol* 2008, 17:2107-2121.

9. Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A: Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* 1999, 238:195-209.
10. Saccone C, Gissi C, Reyes A, Larizza A, Sbisa E, Pesole G: Mitochondrial DNA in metazoa: degree of freedom in a frozen event. *Gene* 2002, 286:3-12.
11. Cameron SL, Downton M, Castro LR, Ruberu K, Whiting MF, Austin AD, Diement K, Stevens J: Mitochondrial genome organization and phylogeny of two vespid wasps. *Genome* 2008, 51:800-808.
12. Waeschenbach A, Telford MJ, Porter JS, Littlewood TJ: The complete mitochondrial genome of *Flustrellidra hispida* and the phylogenetic position of Bryozoa among the Metazoa. *Mol Phylogenet Evol* 2006, 40:195-207.
13. Iannelli F, Griggio F, Pesole G, Gissi C: The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. *BMC Evol Biol* 2007, 7:155.
14. Durrheim GA, Corfield VA, Harley EH, Ricketts MH: Nucleotide sequence of cytochrome oxidase (subunit III) from the mitochondrion of the tunicate *Pyura stolonifera*: evidence that AGR encodes glycine. *Nucl Acids Res* 1993, 21:3587-3588.
15. Xia X, Huang H, Carullo M, Betran E, Moriyama EN: Conflict between translation initiation and elongation in vertebrate mitochondrial genomes. *PLoS One* 2007, 2:e227.
16. Boore JL, Daehler LL, Brown WM: Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (Amphioxus). *Mol Biol Evol* 1999, 16:410-418.
17. Gissi C, Iannelli F, Pesole G: Complete mtDNA of *Ciona intestinalis* reveals extensive gene rearrangement and the presence of an atp8 and an extra trnM gene in ascidians. *J Mol Evol* 2004, 58:376-389.
18. Iannelli F, Pesole G, Sordino P, Gissi C: Mitogenomics reveals two cryptic species in *Ciona intestinalis*. *Trends Genet* 2007, 23:419-422.
19. Yokobori S, Oshima T, Wada H: Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. *Mol Phylogenet Evol* 2005, 34:273-283.
20. Yokobori S, Ueda T, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K: Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). *Genetics* 1999, 153:1851-1862.
21. Yokobori S, Watanabe Y, Oshima T: Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. *J Mol Evol* 2003, 57:574-587.

22. Castilla JC, Guinez R, Caro AU, Ortiz V: Invasion of a rocky intertidal shore by the tunicate *Pyura praeputialis* in the Bay of Antofagasta, Chile. *Proc Natl Acad Sci U S A* 2004, 101:8517-8524.
23. Locke A, Hanson JM, Ellis KM, Thompson J, Rochette R: Invasion of the southern Gulf of St. Lawrence by the clubbed tunicate (*Styela clava* Herdman): Potential mechanisms for invasions of Prince Edward Island estuaries. *J Exp Mar Biol Ecol* 2007, 342:69-77.
24. Ramsay A, Davidson J, Landry T, Arsenault G: Process of invasiveness among exotic tunicates in Prince Edward Island, Canada. *Biological Invasions* 2008, 10:1311-1316.
25. Therriault TW, Herborg LM: Predicting the potential distribution of the vase tunicate *Ciona intestinalis* in Canadian waters: informing a risk assessment. *ICES J Mar Sci* 2008, 65:788-794.
26. Cañestro C, Bassham S, Postlethwait JH: Seeing chordate evolution through the *Ciona* genome sequence. *Genome Biol* 2003, 4:208.
27. Cañestro C, Yokoi H, Postlethwait JH: Evolutionary developmental biology and genomics. *Nat Rev Genet* 2007, 8:932-942.
28. Davidson B: *Ciona intestinalis* as a model for cardiac development. *Semin Cell Dev Biol* 2007, 18:16-26.
29. Manni L, Zaniolo G, Cima F, Burighel P, Ballarin L: *Botryllus schlosseri*: a model ascidian for the study of asexual reproduction. *Dev Dyn* 2007, 236:335-352.
30. Godefroy N, Hoa C, Tsokanos F, Olofsson L, Le Goff E, Douzery EJP, Baghdiguian S, Martinand-Mari C: Identification of autophagy genes in *Ciona intestinalis*: A new experimental model to study autophagy mechanism. *Autophagy* 2009, 5:805-815.
31. Tsagkogeorga G, Turon X, Hopcroft RR, Tilak M-K, Feldstein T, Shenkar N, Loya Y, Huchon D, Douzery EJP, Delsuc F: An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evolutionary Biology* 2009, 9:187.
32. Zeng L, Jacobs M, Swalla B: Coloniality and sociality has evolved once in Stolidobranch ascidians. *Integr Comp Biol* 2006, 46:255-268.
33. Zeng L, Swalla B: Molecular phylogeny of the protochordates: chordate evolution. *Can J Zool* 2005, 83:24-33.
34. Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, et al: Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 2006, 444:85-88.

35. Boursat SJ, Nielsen C, Economou AD, Telford MJ: Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol* 2008, 49:23-31.
36. Boursat SJ, Rota-Stabelli O, Lanfear R, Telford MJ: The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. *BMC Evol Biol* 2009, 9:107.
37. Castoe TA, Jiang ZJ, Gu W, Wang ZO, Pollock DD: Adaptive evolution and functional redesign of core metabolic proteins in snakes. *PLoS ONE* 2008, 3:e2201.
38. Delsuc F, Brinkmann H, Chourrout D, Philippe H: Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 2006, 439:965-968.
39. Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H: Additional molecular support for the new chordate phylogeny. *Genesis* 2008, 46:592-604.
40. Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, et al: The amphioxus genome and the evolution of the chordate karyotype. *Nature* 2008, 453:1064-1071.
41. Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, et al: Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 2008, 452:745-749.
42. Shenkar N, Loya Y: The solitary ascidian *Herdmania momus*: native (Red Sea) versus non-indigenous (Mediterranean) populations. *Biol Invasions* 2008, 10:1431-1439.
43. Nohara M, Nishida M, Miya M, Nishikawa T: Evolution of the mitochondrial genome in Cephalochordata as inferred from complete nucleotide sequences from two *Epigonichthys* species. *J Mol Evol* 2005, 60:526-537.
44. Philippe H: Rodent monophyly: pitfalls of molecular phylogenies. *J Mol Evol* 1997, 45:712-715.
45. Phillips MJ, McLenachan PA, Down C, Gibb GC, Penny D: Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. *Syst Biol* 2006, 55:122-137.
46. Telford MJ: Animal phylogeny. *Curr Biol* 2006, 16:R981-985.
47. Le SQ, Gascuel O, Lartillot N: Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 2008, 24:2317-2323.
48. Lartillot N, Brinkmann H, Philippe H: Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* 2007, 7 Suppl 1:S4.

49. Lartillot N, Philippe H: A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 2004, 21:1095-1109.
50. Blanquart S, Lartillot N: A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* 2008, 25:842-858.
51. Philippe H, Lartillot N, Brinkmann H: Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* 2005, 22:1246-1253.
52. Lartillot N, Philippe H: Improvement of molecular phylogenetic inference and the phylogeny of Bilateria. *Philos Trans R Soc Lond B Biol Sci* 2008, 363:1463-1472.
53. Bromham LD, Degnan BM: Hemichordates and deuterostome evolution: robust molecular phylogenetic support for a hemichordate + echinoderm clade. *Evol Dev* 1999, 1:166-171.
54. Wada H: Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. *Mol Biol Evol* 1998, 15:1189-1194.
55. Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA, Parkinson CL, Pollock DD: Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. *BMC Evol Biol* 2007, 7:123.
56. Castoe TA, de Koning AP, Kim HM, Gu W, Noonan BP, Naylor G, Jiang ZJ, Parkinson CL, Pollock DD: Evidence for an ancient adaptive episode of convergent molecular evolution. *Proc Natl Acad Sci U S A* 2009.
57. Sambrook J, Fritsch EF, Maniatis T: *Molecular cloning: a laboratory manual (2nd edition)*. New York: Cold Spring Harbor Laboratory Press; 1989.
58. Batzoglou S, Jaffe DB, Stanley K, Butler J, Gnerre S, Mauceli E, Berger B, Mesirov JP, Lander ES: ARACHNE: a whole-genome shotgun assembler. *Genome Res* 2002, 12:177-189.
59. The NCBI ORF-Finder tool [<http://www.ncbi.nlm.nih.gov/projects/gorf/>]
60. Wyman SK, Jansen RK, Boore JL: Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 2004, 20:3252-3255.
61. Wolstenholme DR: Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 1992, 141:173-216.
62. Lowe TM, Eddy SR: tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997, 25:955-964.

63. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003, 31:3406-3415.
64. Jameson D, Gibson AP, Hudelot C, Higgs PG: OGRE: a relational database for comparative analysis of mitochondrial genomes. *Nucleic Acids Res* 2003, 31:202-206.
65. Katoh K, Kuma K, Toh H, Miyata T: MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 2005, 33:511-518.
66. Castresana J: Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 2000, 17:540-552.
67. Guindon S, Gascuel O: A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 2003, 52:696-704.
68. Yang Z: Among-site rate variation and its impact on phylogenetic analyses. *Trends Ecol Evol* 1996, 11:367-372.
69. Felsenstein J: PHYLIP (PHYLogeny Inference Package). . Version 3.66 edition. Seattle: Department of Genome Sciences. University of Washington; 2001.
70. Jobb G, von Haeseler A, Strimmer K: TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 2004, 4:18.
71. Lartillot N, Lepage T, Blanquart S: PhyloBayes 3. A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 2009, in press.
72. Blanquart S, Lartillot N: A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol Biol Evol* 2006, 23:2058-2071.
73. Swofford DL: *PAUP*: Phylogenetic Analysis Using Parsimony and other methods*. 4b10 edn: Sinauer, Sunderland, MA; 2002.

Figures

Figure 1 - *Herdmania momus* mitochondrial genome map.

Protein coding genes, rRNAs, and tRNAs are shown in red, green, and black, respectively. Gray numbers indicate the length of non coding region longer than 40bp, the location of the repeated sequences is indicated in blue.

Figure 2 - Putative cloverleaf structures for the 24 tRNA genes of the *Herdmania momus* mitochondrial genome.

Canonical and G-T base pairs are indicated by dots and asterisk respectively. The two boxed tRNA-Asn correspond to the tRNA-scan predicted structure (left) and a less stable alternative structure (right).

Figure 3 - Comparison of mitochondrial gene order in Stolidobranchia.

The gene orders in the mitochondrial genomes of the two stolidobranchs *Halocynthia roretzi* and the newly-obtained *Herdmania momus* are compared. To illustrate the extent of the gene order rearrangements, the same protein-coding genes are connected by black rectangles between the two stolidobranchian mitochondrial genomes. The only syntenic region between the two related genomes represented by a single gene pair, including tRNA-Trp – tRNA-Ile, is indicated by two connected white rectangles.

Figure 4 - Analysis of amino acid composition heterogeneity of mitochondrial proteins among metazoans.

Principal Component Analysis (PCA) of the amino acid composition of the 13 mitochondrial proteins from the 54-taxon metazoan dataset. Individuals are plotted in the first two principal components of the PCA which explain 53.1% and 15.3% of the total compositional variance, respectively. Points corresponding to Cephalochordata (red), Tunicata (purple) and Gnathostomata (blue) are circled.

Figure 5 - Phylogeny of Metazoa inferred from the concatenation of the 13 mitochondrial proteins.

Bayesian consensus tree of 5 individual MCMC obtained using the CAT+BP+ Γ_4 mixture model on the concatenation of the 13 proteins (54 taxa and 2,136 amino-acid sites). Values at nodes indicate Bayesian posterior probabilities (PP). Circles indicate strongly supported nodes with $PP \geq 0.95$. The scale bar represents the estimated number of substitutions per site.

Figure 6 - Comparison of lineage-specific evolutionary rates.

The graph is a plot of the 54 root-to-tip distances calculated from branch lengths estimated under the GTR+ Γ_8 model on the reference topology (see Figure 5) for the concatenations of rRNA-coding (y-axis) and protein-coding (x-axis) mitochondrial genes. Symbols corresponding to non-bilaterian outgroups (green squares), tunicates (red triangles), and all other bilaterians (blue diamonds) are indicated. The phylograms showing the branch lengths inferred from each dataset are presented with the highly evolving tunicates figured in red. The linear regression dotted line is figured with its associated equation and correlation coefficient. The $y = x$ line is also indicated to underline the difference in average evolutionary rate between the rRNA and protein mitochondrial partitions.

Tables

Table 1 — Statistics on non-coding (NC) sequences in tunicate mitochondrial genomes.

Species	% NC	Length (bp)	Longest NC region	
			bp	location
<i>Ciona intestinalis</i>	2.9	429	85	<i>nd1</i> - <i>lrRNA</i>
<i>Ciona savignyi</i>	2.9	428	194	<i>nd1</i> - tRNA Pro
<i>Phallusia fumigata</i>	5.89	915	134	<i>cox1</i> - tRNA Gly (GGN)
<i>Phallusia mammillata</i>	1.94	283	65	tRNA Cys - <i>nd4L</i>
<i>Doliolum nationalis</i>	14.71	2405	968	<i>cox3</i> - <i>cox2</i>
<i>Halocynthia roretzi</i>	3.09	456	112	<i>nd4</i> - tRNA Val
<i>Herdmania momus</i>	9.21	1457	197	<i>nd6</i> - tRNA Tyr

*including 1 bp-long spacers.

Table 2 - Bayesian posterior probabilities (PP) for alternative positions of tunicates using different models of amino-acid replacements.

PP were obtained from 3 independent MCMC runs for all models except

CAT+BP+ Γ_4 for which 5 MCMC were run. The number of trees used to compute PP

is indicated.

Model	Tunicata + Craniata	Tunicata + <i>Xenoturbella</i>	Tunicata basal among Bilateria	Tunicata + Euchordata*	Tunicata + Ambulacraria	Tunicata basal among Deuterostomia
mtREV+Γ_4 (12,000 trees)	0.14	0.85	0	<0.01	0	<0.01
GTR+Γ_4 (24,000 trees)	0	0.83	0.16	0	0	<0.01
CAT +Poisson+Γ_4 (240,000 trees)	0.71	0.08	<0.01	0.11	0.02	0.02
CAT +mtREV+Γ_4 (12,000 trees)	0.96	0	0	<0.01	0	0
CAT +GTR+Γ_4 (12,000 trees)	0.95	0	0	<0.01	<0.01	0
CAT +BP+Γ_4 (34,800 trees)	0.94	0.03	0	0.02	<0.01	<0.01

* Euchordata = Cephalochordata + Craniata

Additional files

Additional file 1

File format: XLS

Species sampling, taxonomy and sequence accession numbers.

Description: The table indicates the taxonomy and the species sampling used in the present study with associated sequence Accession Numbers for their complete mitochondrial genomes.

Additional file 2

File format: NEX

Protein sequence alignment. Concatenated amino-acid sequence alignment of the 13 mitochondrial protein-coding genes (in Nexus format) used in the phylogenetic analyses.

Additional file 3

File format: NEX

DNA sequence alignment. Concatenated nucleotid sequence alignment of the 13 mitochondrial protein-coding genes (in Nexus format) used in the branch length analysis.

Additional file 4

File format: NEX

DNA sequence alignment. Concatenated rRNA sequence alignment (in Nexus format) used in the branch length analysis.

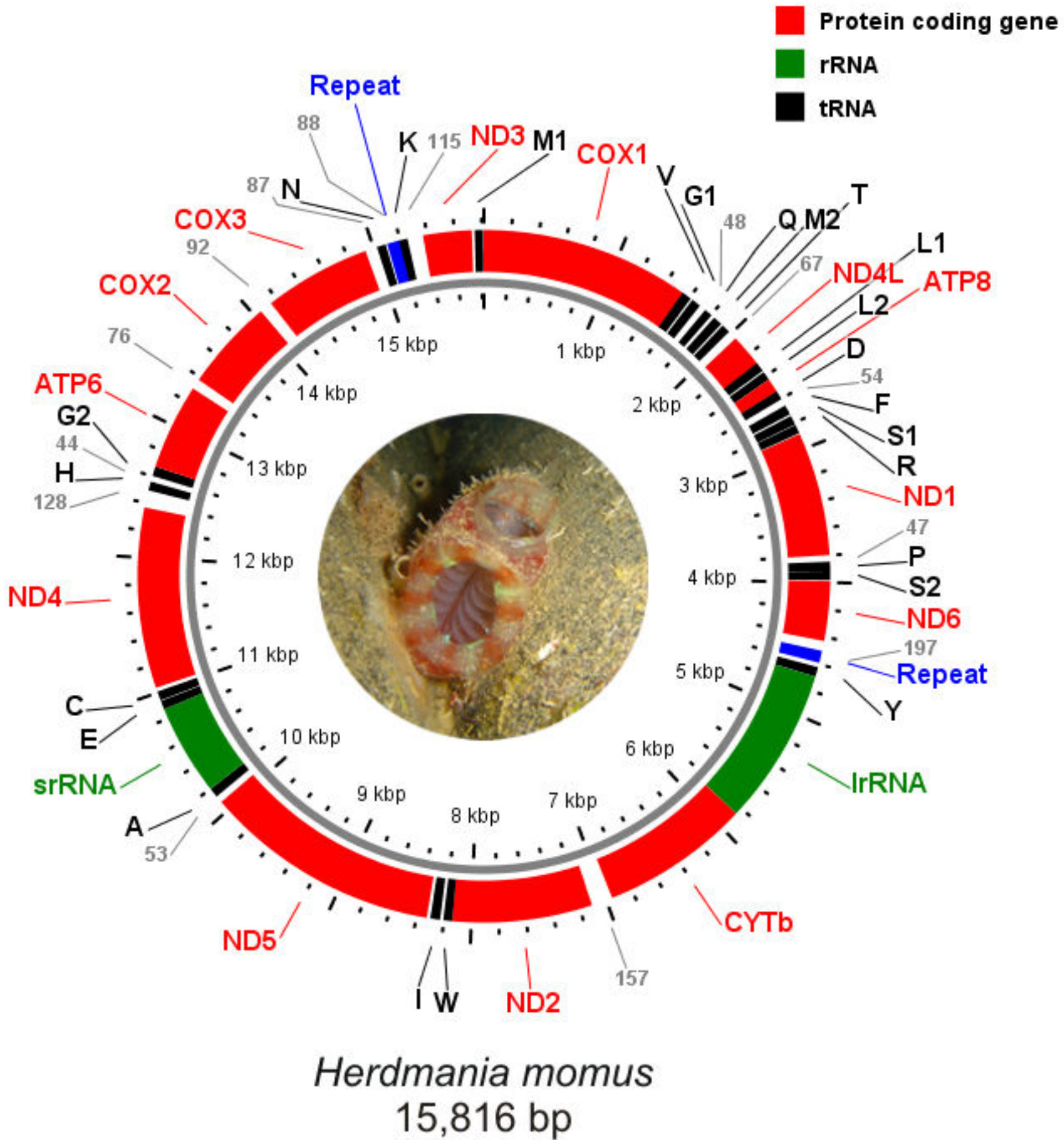


Figure 1

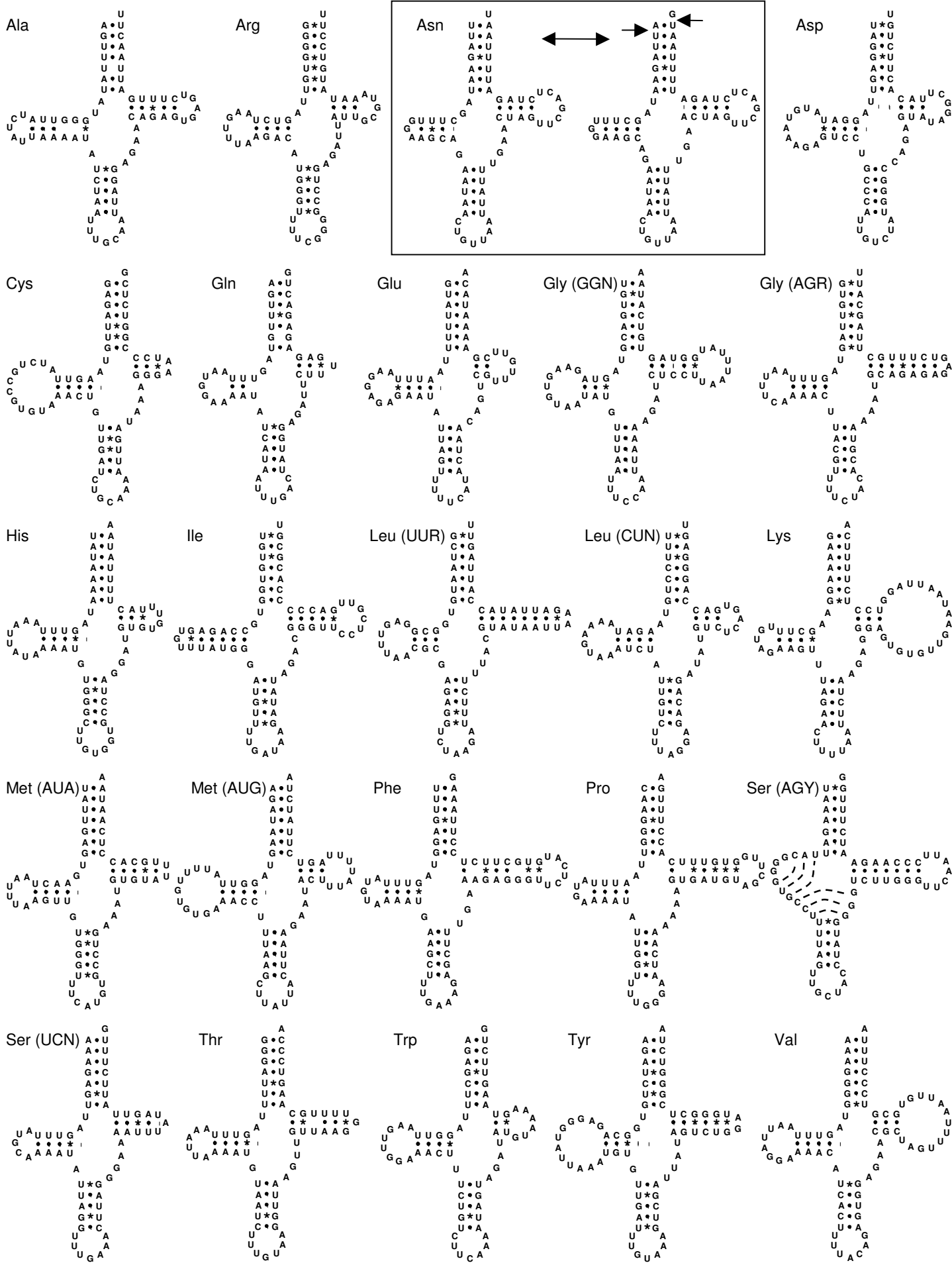
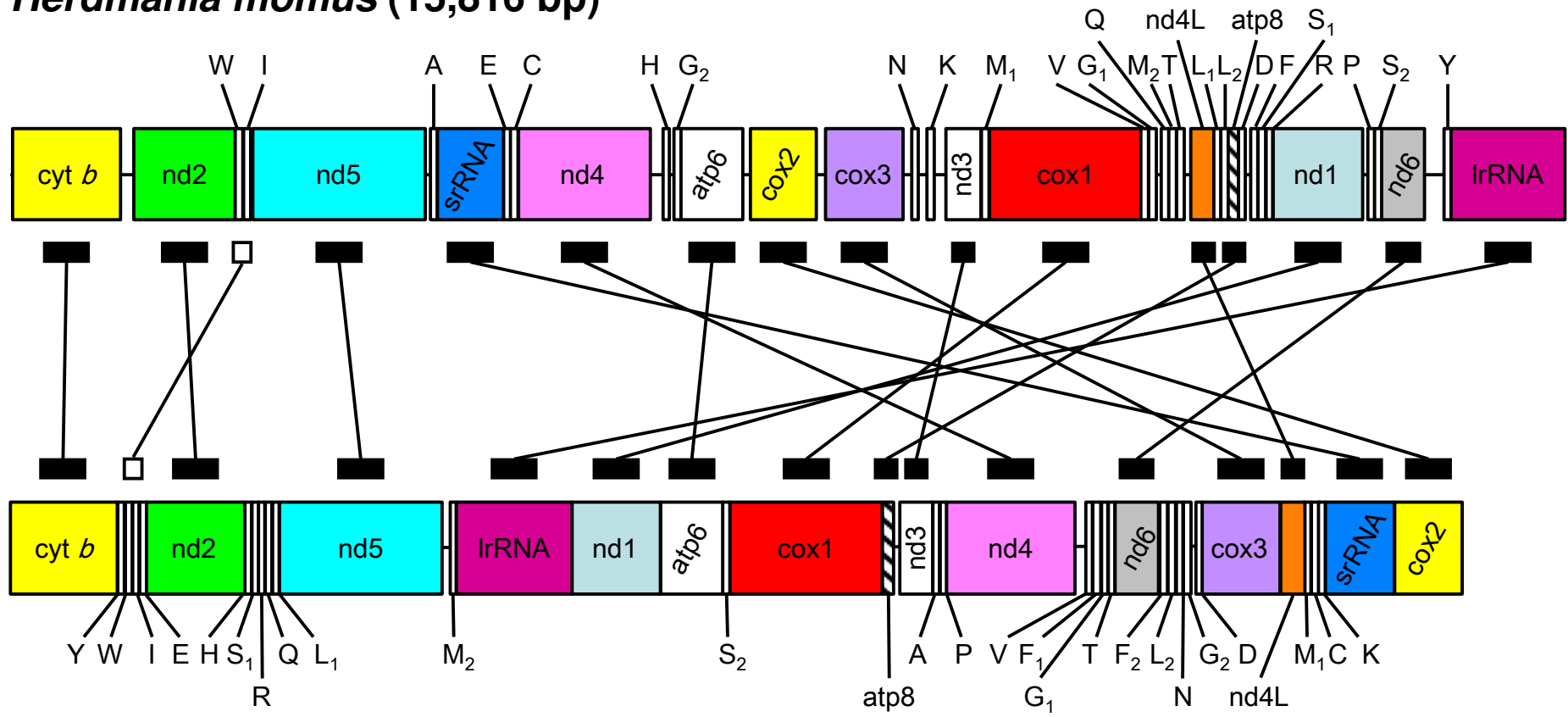


Figure 2

***Herdmania momus* (15,816 bp)**



***Halocynthia roretzi* (14,771 bp)**

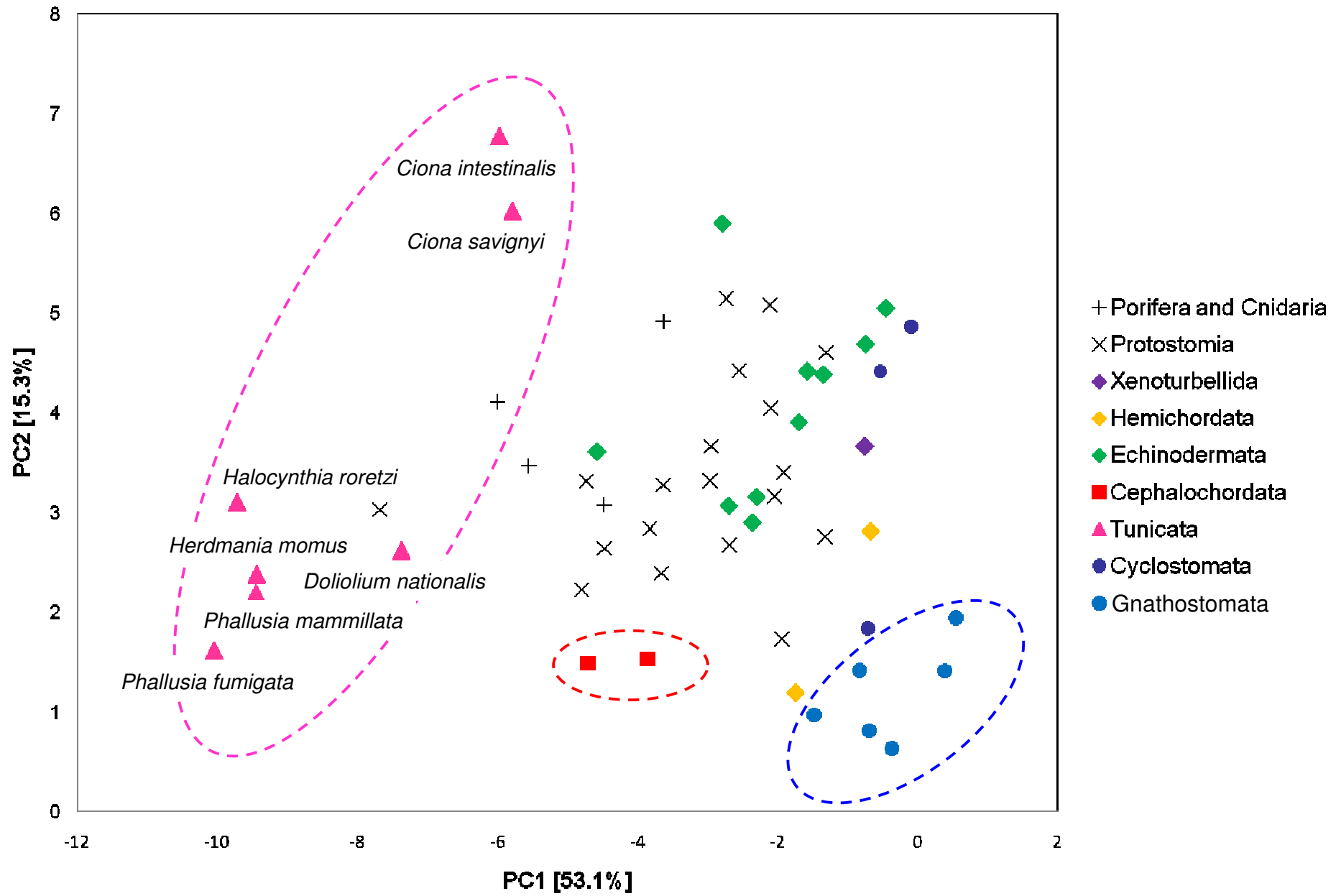


Figure 4

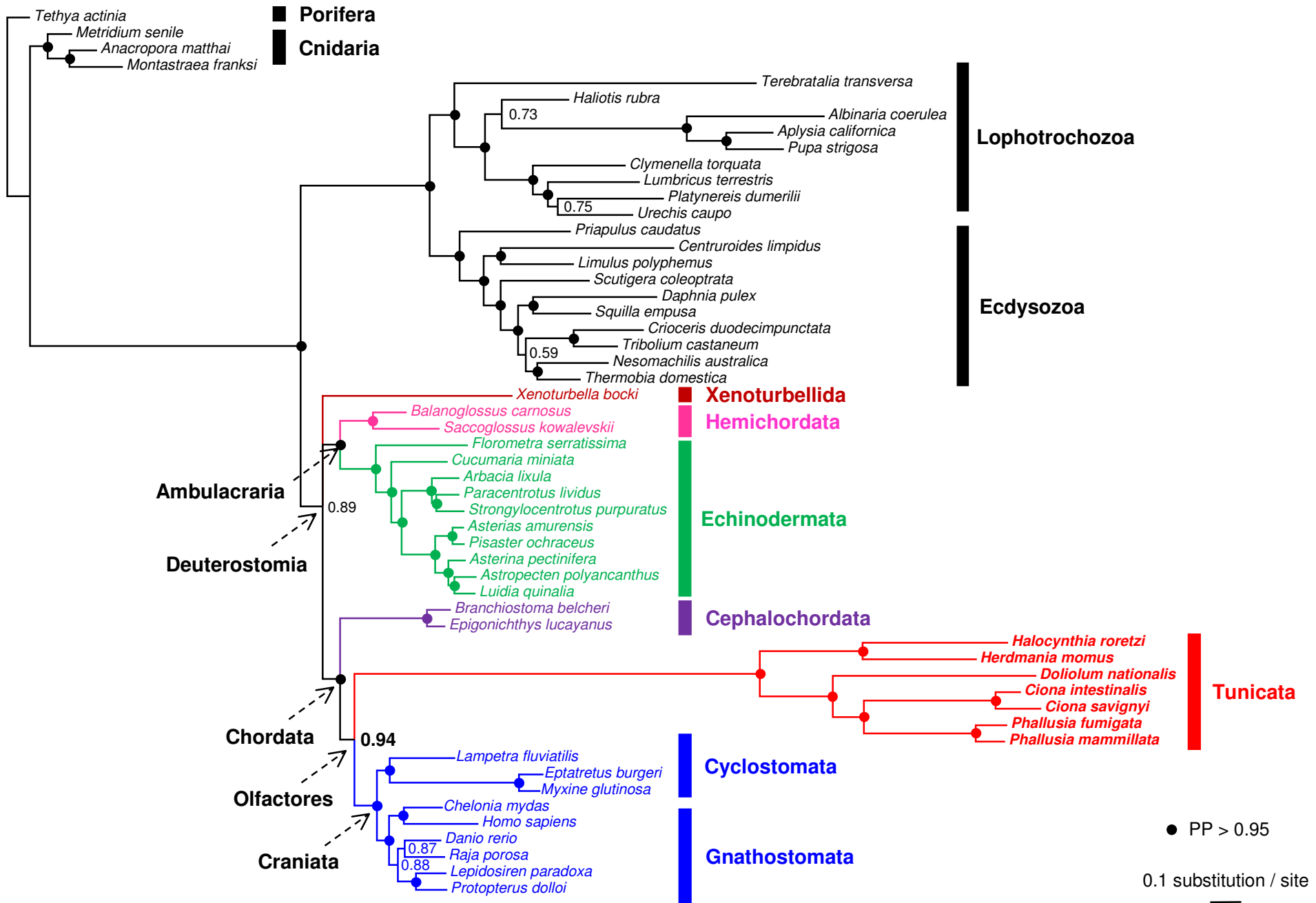


Figure 5

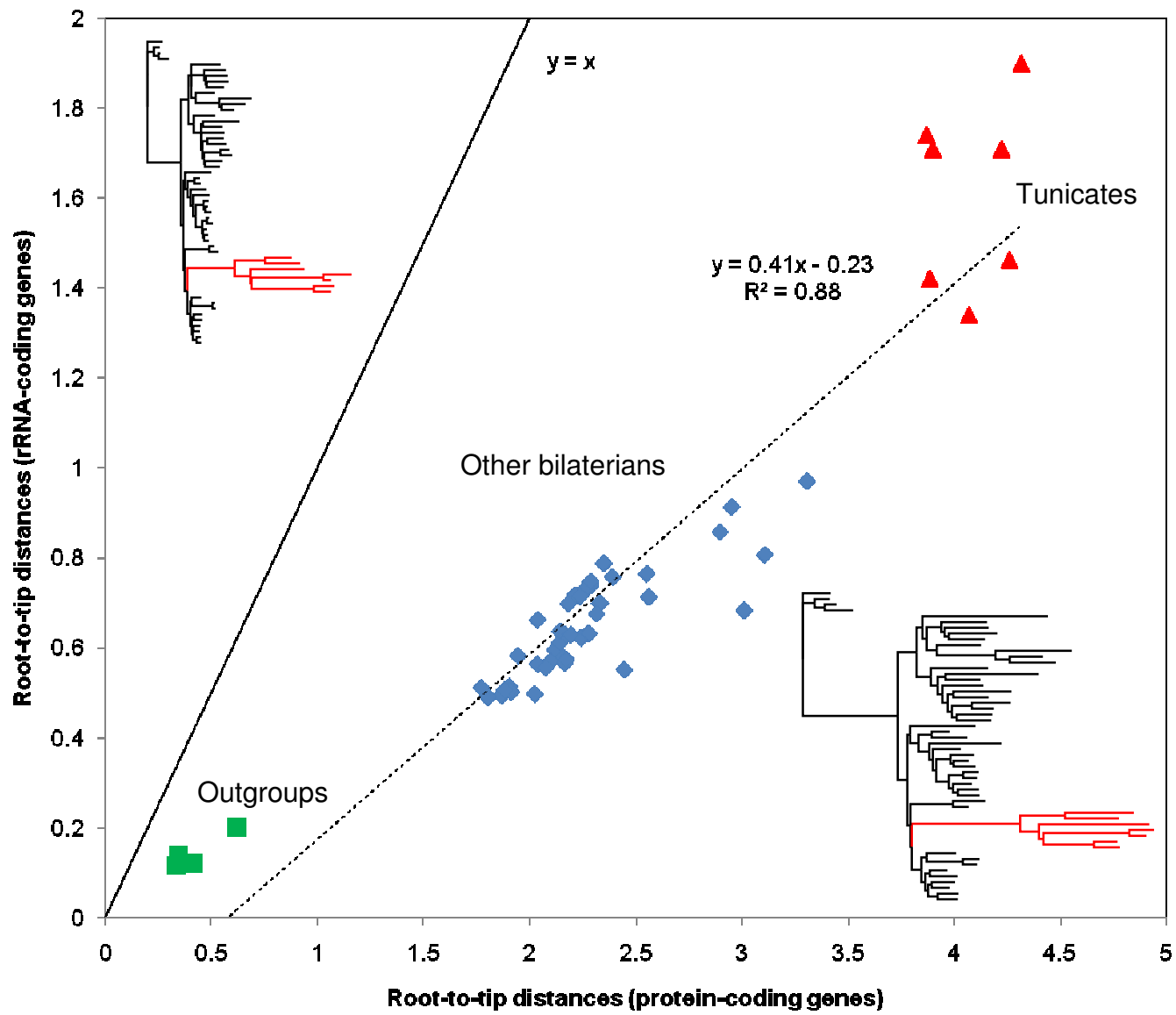


Figure 6

Additional files provided with this submission:

Additional file 1: additionalfile1_taxon_sampling.xls, 31K

<http://www.biomedcentral.com/imedia/1146026363302926/supp1.xls>

Additional file 2: additionalfile2_herdmania_aa_13proteins.nex, 114K

<http://www.biomedcentral.com/imedia/3711371143029269/supp2.nex>

Additional file 3: additionalfile3_herdmania_dna_13protein_genes.nex, 450K

<http://www.biomedcentral.com/imedia/8785381383029269/supp3.nex>

Additional file 4: additionalfile4_herdmania_rrna.nex, 46K

<http://www.biomedcentral.com/imedia/9848853130292696/supp4.nex>

Table S1: Taxon sampling

<i>Classification</i>	<i>Species</i>	<i>Accession Numbers</i>
Porifera	<i>Tethya actinia</i>	NC_006991
Cnidaria	<i>Anacropora matthai</i>	NC_006898
	<i>Metridium senile</i>	NC_000933
	<i>Montastraea franksi</i>	NC_007225
Protostomia		
Lophotrochozoa		
Brachiopoda	<i>Terebratalia transversa</i>	NC_003086
Annelida	<i>Clymenella torquata</i>	NC_006321
	<i>Lumbricus terrestris</i>	NC_001673
	<i>Platynereis dumerilii</i>	NC_000931
Echiura	<i>Urechis caupo</i>	NC_006379
Mollusca	<i>Albinaria coerulea</i>	NC_001761
	<i>Aplysia californica</i>	NC_005827
	<i>Haliotis rubra</i>	NC_005940
	<i>Pupa strigosa</i>	NC_002176
Ecdysozoa		
Priapulida	<i>Priapulidus caudatus</i>	NC_008557
Arthropoda	<i>Centruroides limpidus</i>	NC_006896
	<i>Limulus polyphemus</i>	NC_003057
	<i>Daphnia pulex</i>	NC_000844
	<i>Squilla empusa</i>	NC_007444
	<i>Crioceris duodecimpunctata</i>	NC_003372
	<i>Nesomachilis australica</i>	NC_006895
	<i>Thermobia domestica</i>	NC_006080
	<i>Tribolium castaneum</i>	NC_003081
	<i>Scutigera coleoptrata</i>	NC_005870
Deuterostomia		
Xenambulacraria		
Echinodermata	<i>Astropecten polyacanthus</i>	NC_006666
	<i>Asterina pectinifera</i>	NC_001627
	<i>Luidia quinalia</i>	NC_006664
	<i>Asterias amurensis</i>	NC_006665
	<i>Pisaster ochraceus</i>	NC_004610
	<i>Strongylocentrotus purpuratus</i>	NC_001453
	<i>Paracentrotus lividus</i>	NC_001572
	<i>Arbacia lixula</i>	NC_001770
	<i>Cucumaria miniata</i>	NC_005929
	<i>Florometra serratissima</i>	NC_001878
Hemichordata	<i>Balanoglossus carnosus</i>	NC_001887
	<i>Saccoglossus kowalevskii</i>	NC_007438
Xenoturbellida	<i>Xenoturbella bocki</i>	NC_008556
Chordata		
Cephalochordata	<i>Branchiostoma belcheri</i>	NC_004537
	<i>Epigonichthys lucayanus</i>	NC_006464
Tunicata	<i>Ciona intestinalis</i>	NC_004447
	<i>Ciona savignyi</i>	NC_004570
	<i>Phallusia fumigata</i>	NC_009834
	<i>Phallusia mammilata</i>	NC_009833
	<i>Doliolum nationalis</i>	NC_006627
	<i>Halocynthia roretzi</i>	NC_002177
	<i>Herdmania momus</i>	
Craniata	<i>Myxine glutinosa</i>	NC_002639
	<i>Eptatretus burgeri</i>	NC_002807
	<i>Protopterus dolloi</i>	NC_001708
	<i>Lepidosiren paradoxa</i>	NC_003342
	<i>Raja porosa</i>	NC_007173
	<i>Danio rerio</i>	NC_002333
	<i>Chelonia mydas</i>	NC_000886
	<i>Homo sapiens</i>	NC_012920
	<i>Lampetra fluviatilis</i>	NC_001131

CHAPITRE IV :
L'ACCES AUX DONNEES GENOMIQUES ET
L'EVOLUTION ACCELEREE DES TUNICIERS

1. L'accès aux données génomiques des Tuniciers

L'ensemble des études exposées aux chapitres précédents ont illustré les bénéfices de l'accès aux données génomiques pour la compréhension de l'histoire évolutive des Tuniciers. L'analyse de ces données par des outils phylogénétiques performants ont ainsi permis la corroboration du statut phylogénétique des Tuniciers au sein des Chordés en tant que groupe-frère des Vertébrés (Bourlat *et al* 2006; Delsuc *et al* 2006). A l'inverse, l'ARNr 18S malgré son pouvoir résolutif reconnu à large échelle apporta un signal phylogénétique relativement limité pour la reconstruction des relations de parenté des Tuniciers (Yokobori *et al* 2006; Zeng, Jacobs & Swalla 2006; Mallatt & Winchell 2007; Tsagkogeorga *et al* 2009). Il apparut ainsi nécessaire d'acquérir de nouvelles données phylogénomiques afin d'approfondir l'évolution moléculaire si particulière de ces animaux.

Le développement de marqueurs phylogénétiques chez les animaux est traditionnellement basé sur des approches d'amplification par PCR (« Polymerase Chain Reaction ») à partir d'ADN génomique de gènes cibles présents en simple copie dans les génomes (Rokas, Kruger & Carroll 2005; Paps, Baguna & Riutort 2009). Cependant, ces approches expérimentales classiques s'avérèrent très difficiles à mettre en œuvre chez les Tuniciers essentiellement à cause de la divergence nucléotidique extrême liée au taux d'évolution élevé caractérisant les gènes nucléaires du groupe. Les gènes protéiques de Tuniciers présentent une telle saturation substitutionnelle, notamment en troisième position du codon, qu'il est difficile de définir des amorces de PCR dès lors que l'on cible une diversité de taxons conséquente.

Par voie de conséquence, les données génomiques des Tuniciers disponibles ont été principalement issues de projets de séquençage des génomes complets pour *Ciona intestinalis*, *Ciona savignyi* et *Oikopleura dioica* (Seo *et al* 2001; Dehal *et al* 2002; Small *et al* 2007), d'une part, et d'autre part, de la construction de banques d'ADNc suivie par le séquençage aléatoire d'ESTs (« Expressed Sequence Tags »). Cette dernière approche a été utilisée notamment chez *Molgula tectiformis* où une large collection (106863 ESTs) a été séquencée (Gyoja *et al* 2007), *Halocynthia roretzi* (62501) (Kim *et al* 2008) et *Diplosoma listerianum* (2051) (Blaxter & Thomas 2004).

2. La révolution du séquençage massif à haut débit

Aux alentours de 2005, une série des nouvelles plateformes technologiques apparurent sur la scène du séquençage génomique, les plus connues étant Roche 454 GS System®, Illumina Genome Analyzer®, Applied Biosystems SOLiD® System, et Helicos Heliscope®. S'appuyant sur des innovations techniques permettant le séquençage massif et simultané de petits fragments d'ADN, ces approches promettent l'obtention de données d'une ampleur jamais égalée jusque là en des intervalles de temps particulièrement courts (Shendure & Ji 2008).

Le séquençage à haut débit est en train de véritablement révolutionner la recherche en biologie en transformant l'accès autrefois fastidieux aux données génomiques en un processus standard relativement peu coûteux. En réponse à ces développements, sous les auspices de l'« International Nucleotide Sequence Database Collaboration » (INSDC), les équipes dirigeant les bases de données NCBI, EBI et DDBJ ont récemment développé un système d'archivage pour la gestion de ces données sous la forme du « Sequence Read Archive » (SRA) (<http://www.ncbi.nlm.nih.gov/Traces/sra>). Moins de quatre ans après le développement de ces approches de séquençage nouvelle génération, le SRA dénombrait déjà au mois de Mars 2008 plus de 7 terabases (7000 gigabases) de données soumises, les entrées augmentant à une vitesse proprement hallucinante.

3. Projet pilote de l'application du séquençage 454 sur le transcriptome de l'ascidie invasive *Microcosmus squamiger*

Compte-tenu des nombreux obstacles pratiques rencontrés pour accéder à l'information génétique chez les Tuniciers par les méthodes classiques de PCR, les avancées technologiques du séquençage haut-débit semblèrent une piste expérimentale prometteuse pour le développement de nouveaux marqueurs phylogénétiques pour le groupe. Un projet pilote sur l'ascidie *Microcosmus squamiger* fut initié afin d'évaluer l'apport potentiel de l'application du pyroséquençage au groupe des Tuniciers. Cette espèce de la famille des Pyuridae est invasive en Méditerranée (Turon, Nishikawab & Riusa 2006) et elle appartient à un clade majeur de Tuniciers (Tsagkogeorga *et al* 2009) pour lequel les données génomiques disponibles sont les moins abondantes.

Une banque d'ADNc constituée d'environ $1,9 \times 10^7$ clones avec une taille moyenne d'insert de 2 Kb fut construite au départ dans l'optique d'un séquençage aléatoire d'ESTs. Le transcriptome partiel de *Microcosmus squamiger* contenu dans cette banque fut ensuite pyroséquencée par la compagnie GATC Biotech (<http://www.gatc-biotech.com/>) en un demi-run de 454 GS-FLX Standard équivalant à 50 Mb à l'époque. Le bilan qualitatif de cette expérience résulta en 211899 lectures simples d'une longueur moyenne de 241 pb. Parmi celles-ci, seules 7725 lectures furent mappées avec succès de manière automatique par GATC Biotech sur le génome distant de *Ciona intestinalis*, ce qui donne une idée du degré de divergence entre ces deux génomes d'espèces distantes dans la phylogénie (Tsagkogeorga *et al* 2009).

3. Analyse exploratoire des données de pyroséquençage de *Microcosmus squamiger*

Un pipeline bioinformatique a été développé afin de permettre l'analyse détaillée des résultats du pyroséquençage du point de vue du contenu en gènes, de leur degré de couverture, et de la redondance des lectures issues du run (Fig. 11). L'ensemble des lectures obtenues fut soumis à des recherches d'homologie par Blast contre trois bases de données principales composées de gènes orthologues définis à trois niveaux taxonomiques différents: I) Tuniciers, II) Deutérostomes, et III) Métazoaires.

I. Tuniciers

La première piste fut la comparaison du contenu en gènes de la banque de *Microcosmus* par rapport à ceux des ascidies *Ciona intestinalis* et *Ciona savignyi* dont les génomes complets sont disponibles. Ceci a été réalisé dans un premier temps en ciblant le transcriptome de *Ciona intestinalis* qui comprend 19858 séquences nucléotidiques. Puis, en ne considérant que les gènes communs aux deux Ciones, les données de *Microcosmus* ont été comparées à une base de données contenant 9520 gènes issu de l'intersection des transcrits orthologues entre *Ciona intestinalis* et *Ciona savignyi*.

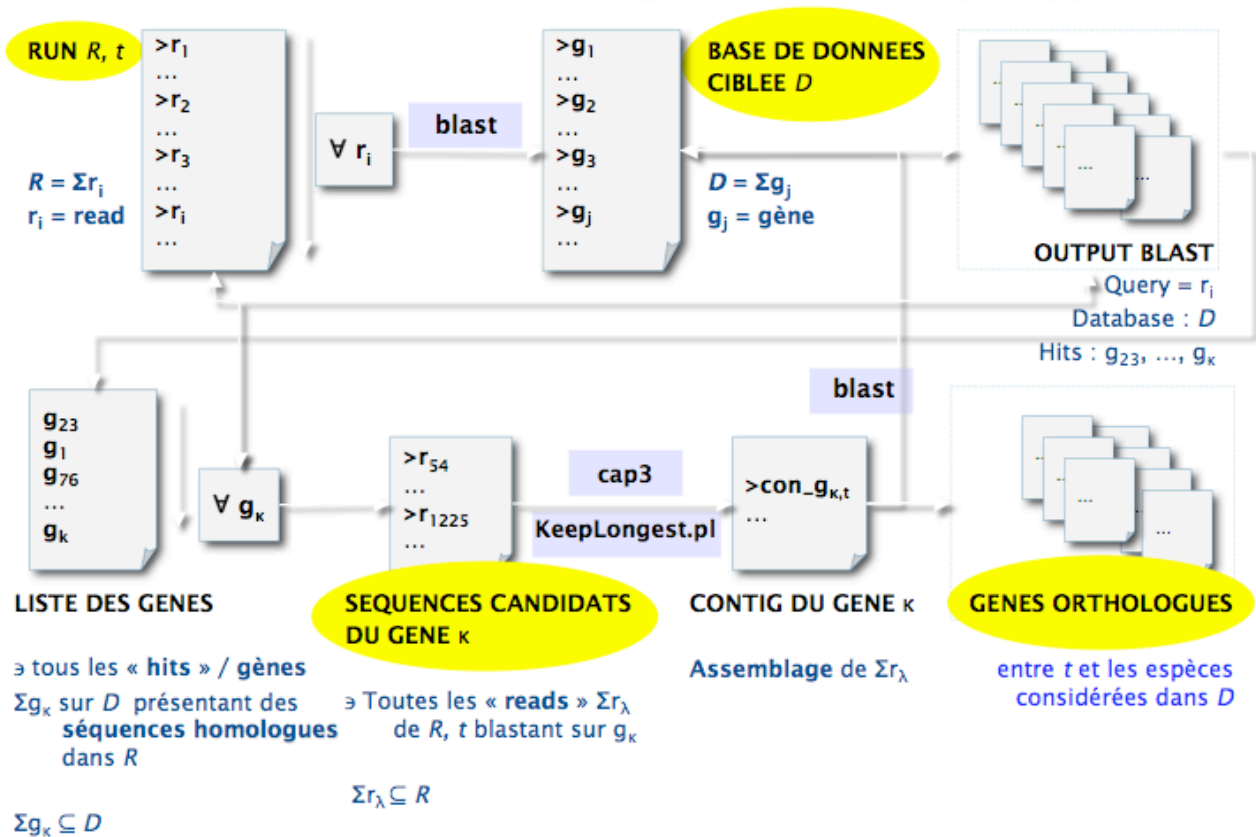


Figure 11 : Représentation schématique du pipeline

II. Deutérostomes

Afin d'identifier des marqueurs potentiels pour la phylogénie des Deutérostomiens, une seconde base de données a été assemblée via l'interrogation de la base de données *e* ! d'Ensembl (<http://www.ensembl.org/>). En bénéficiant d'une approche informatique développée dans l'équipe pour les Mammifères (Ranwez *et al* 2007), les génomes complets de *Ciona intestinalis* et *Ciona savignyi* furent comparés par rapport à celui d'*Homo sapiens*, l'objectif étant d'identifier les gènes orthologues présents en copie unique chez les trois Chordés et ayant des exons de longueur supérieure à 600 pb. Cette approche conduit à l'identification d'un premier ensemble de 350 gènes, sur la base desquels une deuxième recherche fut réalisée afin de récupérer les transcrits des 252 gènes qui parmi les précédents sont également présents chez trois Deutérostomes additionnels : *Tetraodon nigroviridis*,

Xenopus tropicalis et *Gallus gallus*, et chez deux groupes externes, *Caenorhabditis elegans* et *Drosophila melanogaster*.

III. Métazoaires

Enfin, pour évaluer le contenu du transcriptome partiel de *Microcosmus* vis-à-vis de gènes conservés à une plus large échelle évolutive, une troisième base de données a été constituée à partir du jeu de données phylogénomiques précédemment assemblé (Delsuc *et al* 2008). Incluant les séquences protéiques de 174 gènes pour 51 taxons, ce jeu de données sert à la recherche de marqueurs phylogénétiques adéquats pour la reconstruction phylogénétique à l'échelle des Métazoaires.

4. Contenu génique du transcriptome partiel de *Microcosmus squamiger*

Parmi les 211899 lectures, environ 50000 présentèrent un degré de similitude nucléotidique significatif par rapport à l'intersection des transcrits orthologues entre *Ciona intestinalis* et *Ciona savignyi* (9520 gènes), ou contre les ADNc de *Ciona intestinalis* (19858 transcrits représentant 15000 gènes). Malgré le fait que le nombre de séquences fut multiplié par un facteur 3 en considérant les séquences protéiques (tblastx), l'ensemble des séquences significativement similaires à des orthologues des deux Ciones se retreignit à 216 gènes (Tableau 4).

De manière analogue, les résultats du Blast mené à des niveaux taxonomiques supérieurs soulignèrent la divergence étonnante caractérisant les séquences de Tuniciers. Ces comparaisons aboutirent à l'identification d'un noyau de gènes conservés restreint à 83 à l'échelle des Deutérostomiens et 68 à l'échelle des Métazoaires, parmi les gènes représentés dans la banque de *Microcosmus* (Tableau 4). Ces gènes surexprimés chez *Microcosmus*, sont l'objet d'une énorme redondance dans les données issues du séquençage 454 puisqu'ils correspondent à environ 25000 lectures (Tableau 4).

Cette redondance était attendue étant donné l'identité ontologique de ces gènes qui sont impliqués pour la plupart dans les fonctions primordiales pour le fonctionnement de la cellule et le fait que la banque n'a pas été normalisée pour être sûr d'obtenir aux moins ces gènes « universels ». Une majeure partie de ceux-ci consiste en des gènes codant pour des protéines ribosomales, des protéines du cytosquelette, ou encore des enzymes impliquées dans

la régulation des processus du métabolisme cellulaire, de la transcription et de la traduction (Figure 12).

Tableau 4 : Analyses des séquences de *Microcosmus squamiger* issues du séquençage 454

ÉCHELLE EVOLUTIVE	# GENES INCLUS & TYPE DE DONNEES	ESPECES DE REFERENCE	OUTIL	#LECTURES	#GENES
TUNICIERS	19 858 transcrits	<i>Ciona intestinalis</i>	blastn	48 455	245
	9 520 gènes orthologues	<i>Ciona intestinalis</i> ,	blastn	46 235	216
		<i>Ciona savignyi</i>	tblastx	126 830	834
DEUTEROSTOMES	252 transcrits orthologues	<i>Ciona intestinalis</i> ,	blastn	1 171	33
		<i>Ciona savignyi</i> ,			
		<i>Gallus gallus</i> ,			
		<i>Homo sapiens</i> ,			
		<i>Tetraodon nigroviridis</i> ,			
<i>Xenopus tropicalis</i>					
	groupes externes :				
	<i>Caenorhabditis elegans</i> ,				
	<i>Drosophila melanogaster</i>				
METAZOAIRE	174 protéins orthologues	51 OTUs « Operational Taxonomic Units » (Delsuc <i>et al</i> 2008)	blastx	24 619	68

Annotation GO - fonction moléculaire

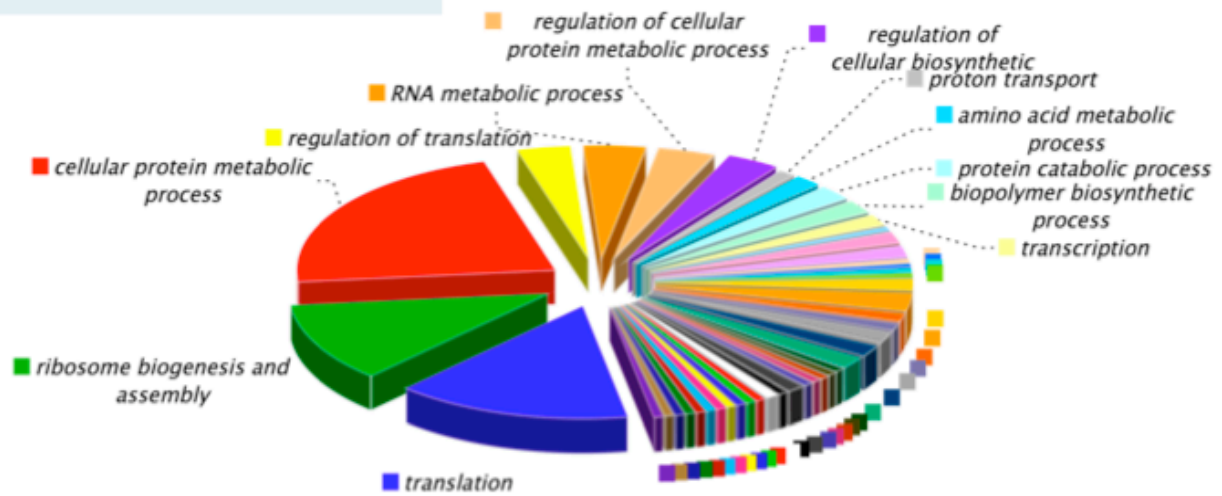


Figure 12 : Annotation GO « gene ontology » des 68 gènes potentiellement orthologues à l'échelle des Métazoaires identifiés chez *Microcosmus squamiger* à partir des données 454 (<http://www.geneontology.org/>).

5. Évolution accélérée des Tuniciers et l'hétérogénéité des taux d'évolution au sein des Chordés [Article 4]

L'image des données génomiques nucléaires et mitochondriales fournie par les phylogénies de gènes individuels (Turon & Lopez-Legentil 2004; Yokobori *et al* 2006; Tsagkogeorga *et al* 2009), de gènes multiples (Delsuc *et al* 2008) ou de familles de protéines (Gissi *et al* 2006; Goldstone *et al* 2007), la structure génomique altérée par les nombreux épisodes de pertes de gènes (Dehal *et al* 2002) et des remaniements géniques du génome mitochondrial (Gissi, Iannelli & Pesole 2008), constituent autant de preuves de la divergence génomique des Tuniciers par rapport aux autres lignées de Chordés (Putnam *et al* 2008). Cette divergence génomique a également été illustrée par l'analyse du contenu des données génomiques de *Microcosmus squamiger* issues ici de l'application pilote du séquençage 454 aux Tuniciers. L'ensemble de ces éléments d'observation sont à l'origine d'un projet qui s'attacha tout d'abord à l'identification de nouveaux marqueurs phylogénétiques permettant par la suite de mener une analyse comparative visant à quantifier l'évolution accélérée des Tuniciers au sein des Chordés.

Sur la base d'un jeu de données nucléotidique constitué de 35 gènes de ménage (marqueurs orthologues, universels, et essentiels de la machinerie cellulaire chez tous les animaux actuels), ce projet adressa trois questions fondamentales : 1) Le taux d'évolution moléculaire est-il systématiquement plus élevé chez les Tuniciers que chez les autres Chordés ? ; 2) Dans quel degré le taux d'évolution varie-t-il entre les trois lignées majeures de Chordés ? ; et 3) Quels sont les déterminants principaux de l'image contrastée des taux d'évolution dessinée par les Chordés ?

L'analyse des 35 marqueurs moléculaires sous des modèles d'horloges locales suivie d'une série de tests du ratio des vraisemblances (LRTs pour « Likelihood Ratio Tests ») a permis d'estimer la variation relative du taux d'évolution au sein des Chordés en utilisant le logiciel PAML (Yang 2007). L'occurrence de changements significatifs de ce taux d'évolution chez les Tuniciers a été détectée pour la majorité des gènes examinés. Dans leur ensemble, les Tuniciers montrent un taux d'évolution environ deux fois plus rapide en moyenne que celui des Vertébrés et Céphalochordés qui apparaît relativement similaire. Le degré de l'accélération varie néanmoins sensiblement selon les gènes. Une portion non négligeable des gènes présente toutefois une vitesse d'évolution relativement similaire entre les trois lignées de Chordés, mettant ainsi en évidence des régions codantes du génome des Tuniciers qui ont probablement échappées à l'accélération génomique générale.

Par une approche impliquant des modèles de taux d'évolution auto-corrélés le long des branches (Thorne, Kishino & Painter 1998), la variation du taux entre gènes et entre branches individuelles composant chaque lignée a pu être quantifiée de manière simultanée. Selon l'image obtenue, les représentants des groupes des Phlebobranchia et Stolidobranchia évoluent le plus lentement au sein des Tuniciers et présentent un taux d'évolution relativement homogène chez les espèces qui les composent, ainsi qu'entre les différents marqueurs. A l'opposé, les Aplousobranches et Appendiculaires, représentés par les espèces *Diplosoma listerianum* et *Oikopleura dioica*, respectivement, présentent des profils clairement plus divergents, avec notamment une remarquable variation quant à leur taux d'évolution entre les gènes. D'un point de vue plus global, les deux lignées s'avèrent conditionner le degré du contraste observé à l'échelle des Chordés, bien que la différence avec les autres lignées reste significative.

En ce qui concerne les déterminants de cette évolution accélérée des Tuniciers, une analyse préliminaire du ratio dN/dS basée sur des paires d'espèces au sein des Olfactores suggère que la composante principale de cette variation réside dans des changements du taux

de mutation plutôt que dans des modifications du régime sélectif. L'ensemble de ces résultats obtenus sur une fraction limitée, mais néanmoins essentielle, du répertoire génique des Chordés, fournissent des hypothèses à tester à plus large échelle afin de mieux quantifier le degré de la divergence moléculaire des Tuniciers et d'identifier les facteurs à l'origine de l'accélération de leur taux d'évolution génomique.

ARTICLE IV

ARTICLE EN PREPARATION

ACCELERATED EVOLUTIONARY RATE OF HOUSEKEEPING GENES IN TUNICATES

GEORGIA TSAGKOGEOGA^{1,2}, XAVIER TURON³,
EMMANUEL J. P. DOUZERY^{1,2} & FRÉDÉRIC DELSUC^{1,2}

¹Université Montpellier II, CC064, Place Eugène Bataillon, Montpellier Cedex 5, France

²CNRS, Institut des Sciences de l'Évolution (UMR5554), CC064, Place Eugène Bataillon,
Montpellier Cedex 5, France

³Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala S. Francesc 14, 17300
Blanes (Girona), Spain

Accelerated evolutionary rate of housekeeping genes in tunicates

Georgia TSAGKOGEOGA^{a,b}, Xavier TURON^c, Emmanuel J. P. DOUZERY^{a,b} and Frédéric DELSUC^{a,b*}

^a Université Montpellier 2

^b CNRS, Institut des Sciences de l'Evolution (UMR 5554), CC064, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France

^c Centre d'Estudis Avançats de Blanes (CEAB, CSIC), Accés Cala S. Francesc 14, 17300 Blanes (Girona), Spain

Running title: Tunicate accelerated evolutionary rates

Keywords: Tunicates – Urochordates – Chordates – Phylogenomics – 454 sequencing – Molecular evolution – Evolutionary rate

* Corresponding author. Tel.: +33 4 67 14 39 64. Fax: +33 4 67 14 56 10.

E-mail address: Frederic.Delsuc@univ-montp2.fr

Abstract

Phylogenomics has recently revealed that tunicates represent the sister-group of vertebrates, forming with them a monophyletic clade called Olfactores. However, phylogenomic and comparative genomic studies have also provided strong evidence for tunicate elevated rate of molecular evolution and their high degree of genomic divergence. Despite the recurrent interest in the group, the picture of tunicate peculiar evolutionary dynamics is still fragmentary, as it mainly lies in studies focusing on only few model species.

In the goal to expand the available genomic data for the group, here we introduced the high-throughput 454 technology to sequence the partial transcriptome of a previously unsampled tunicate, *Microcosmus squamiger*. This allowed us to get further insights on tunicate accelerated evolution by carrying out a comparative analysis based on 35 pertinent phylogenetic markers, consisting of a core of universal and highly conserved housekeeping genes across bilaterians.

Our results showed that tunicates have experienced an overall two-fold faster evolution compared to the other chordates, yet the degree of the predicted acceleration varied extensively upon gene and upon lineage. Appendicularia and Aplousobranchia were detected as the most divergent groups, marked also by highly heterogeneous rates across genes. Finally, a preliminary estimation of the d_N/d_S for a subset of three pairs of closely related taxa within Olfactores resulted in a roughly homogenous image of ω for tunicate and vertebrate lineages, which indicates that for the given set of genes tunicate accelerated evolution was plausibly due to an elevated mutation rate.

Introduction

Understanding chordate origins and evolution has been a preeminent challenge for scientists over the last two centuries (Swalla and Xavier-Neto 2008). The phylum Chordata consists of three distinct lineages: cephalochordates (lancelets), tunicates or urochordates (ascidians or sea squirts, salps and appendicularians) and vertebrates *sensu lato* (*i.e.*, including Cyclostomes). Among these, tunicates embrace a diversity of more than 3,000 filter-feeding and mostly hermaphrodite marine animals, characterized by both sessile (ascidians) and planktonic (salps and appendicularians) life styles, associated with a unique developmental program (Lambert 2005).

In sharp contrast with vertebrates and lancelets, the vast majority of tunicate adults lack the hallmarks of the typical chordate body plan, namely a dorsal neural tube and a notochord. Since it has been pointed out that ascidian larvae do possess these chordate features (Kowalevski 1868), tunicates have been considered as the sister group of Euchordata, *i.e.*, cephalochordates plus vertebrates (Cameron et al. 2000; Mallatt and Winchell 2007; Swalla et al. 2000; Winchell et al. 2002). Though widely accepted during the pre-genomic era, this euchordate view has been substantially impacted by the landmark genome projects of two ascidians *Ciona intestinalis* (Dehal et al. 2002) and *Ciona savignyi* (Small et al. 2007a), as well as the draft genome of the appendicularian *Oikopleura dioica* (Seo et al. 2001).

These genomic data allowed including tunicates in phylogenomic studies and accumulating evidence soon indicated that the evolutionary positions of cephalochordates and tunicates should be reversed, erecting a new clade called Olfactores that links tunicates with vertebrates (Bourlat et al. 2006; Delsuc et al. 2006; Delsuc et al. 2008; Dunn et al. 2008; Singh et al. in press). Although debated at first, the Olfactores hypothesis of chordate

evolution has now gained acceptance as it was corroborated by analyses of the complete genome sequence of the amphioxus (Putnam et al. 2008).

Comparative genomics also revealed that tunicates possess the basic developmental genetic toolkit of vertebrates in unexpectedly condensed genomes that have not undergone the subsequent vertebrate-specific duplication events (Canestro et al. 2003; Dehal and Boore 2005; Dehal et al. 2002). This prompted a growing interest in tunicates as useful experimental systems to decipher the developmental mechanisms underlying chordate origins (Corbo et al. 2001; Davidson 2007; Holland and Gibson-Brown 2003; Satoh 2003).

Despite this renewed interest, our current understanding of tunicate evolution remains fragmentary since phylogenomic and especially comparative genomic studies are taxonomically biased towards model species with completed genomes (Bourlat et al. 2006; Delsuc et al. 2006; Delsuc et al. 2008; Donmez et al. 2009; Imai et al. 2006; Kim et al. 2007; Satoh 2003; Sierro et al. 2006; Yandell et al. 2006). Complementary EST data are also available for a few additional species like *Molgula tectiformis* (Gyoja et al. 2007; Kim et al. 2008) and *Halocynthia roretzi* (Gyoja et al. 2007; Kim et al. 2008). However, the aforementioned species are far from covering the impressive biodiversity of tunicates, yielding to large amounts of molecular data to be disproportionately distributed among their major phylogenetics lineages. In this respect and as additionally indicated by the open controversies regarding the evolutionary history of the group (Yokobori et al. 2006; Zeng et al. 2006; Zeng and Swalla 2005), tunicates are currently marked by a prominent paucity of molecular phylogenetic markers.

One reason invoked to explain this situation involves the high levels of divergence of tunicate lineages due to their accelerated evolution (Zeng et al. 2006). Several recent genome based phylogenies inferred from either nuclear or mitochondrial genome data have provided clear-cut evidence for the particularly high rates distinguishing tunicate evolution, as

illustrated by the persisting long branches of the group in the reconstructed trees (Bourlat et al. 2008; Delsuc et al. 2008; Gissi et al. 2008; Singh et al. in press). Similar conclusions have been drawn by studies focusing on individual genes (*e.g.*, 18S rDNA, *cox1*, Huntingtin, P transposase) or chordate gene families (*e.g.*, *CYP 1*) suggesting that tunicate sequence divergence may increase up to 30% between species of the same genus (Gissi et al. 2006; Goldstone et al. 2007; Kimbacher et al. 2009; Turon and Lopez-Legentil 2004; Yokobori et al. 2006). Likewise, mitogenomics have recently unveiled episodes of cryptic speciation in *C. intestinalis* (Iannelli et al. 2007), whereas the *C. savignyi* genome was found to exhibit extreme levels of structural and single nucleotide polymorphism for a multicellular animal (Seo et al. 2004; Small et al. 2007b). As far as genomic features are concerned, it has been revealed that beyond genome contraction, tunicates have also undergone dramatic genomic rearrangements associated to numerous gene losses. Breaking prominent paradigms, the *Hox* cluster is disintegrated in *Ciona*, and the retinoic acid signaling genes are lacking from *Oikopleura dioica* genome (Canestro and Postlethwait 2007; Canestro et al. 2007; Edvardsen et al. 2004; Gee 2008). All afore elements suggest a high degree of tunicate genomic divergence from other chordate lineages and provide compelling evidence that tunicates exhibit high rates of molecular evolution. It is hence plausible that the current scarcity of molecular data for tunicates results from an extreme sequence divergence, hampering the *in vitro* acquisition of novel sequence data, for instance by standard PCR approaches.

Given these puzzling issues, here, (1) we used the recently developed high throughput sequencing 454 technology to acquire *in vitro* transcriptomic data from a previously unsampled and invasive Mediterranean tunicate species, *Microcosmus squamiger* (Asciidiacea: Stolidobranchia, Pyuridae); and (2) we shed more light on the patterns of evolutionary rate variation within chordates by carrying out a comparative analysis based on 35 *in silico* identified orthologous housekeeping genes, highly conserved across Bilateria.

More precisely, to test the hypothesis of tunicate accelerated evolution and assess the degree of their divergence from other chordate lineages, we addressed the following questions: (1) Are evolutionary rates of amino acid replacement consistently higher in tunicates than in vertebrates and to what extent? (2) Do chordates exhibit homogenous within-lineage rates and if not, which genes and/or which species contribute the most to the observed heterogeneity? And finally (3) what appears to be the underlying cause(s) of molecular evolutionary rate contrasts within Olfactores?

Materials and Methods

Sampling and 454 sequencing of the ascidian *Microcosmus squamiger*

Microcosmus squamiger individuals were collected at [to be completed]. Muscle and gonad tissues were dissected and rapidly frozen in liquid nitrogen. Given the species small size, tissue material from 8 individuals was mixed and used in standard cDNA library construction in pCMV Sport 6.1 vectors (Invitrogen ; Carlsbad, CA, USA). The obtained library was estimated to contain 1.9×10^7 cDNA clones, with inserts of about 2 kb in length. Inserts of the cDNA library of *Microcosmus* were subsequently sequenced using the standard Roche GS FLX system for ultra-high-throughput DNA sequencing. The 454 sequencing yielded 50 million high quality bases in total, which corresponded to approximately 217,000 sequences with a mean length of 250 bp. The 454 generated cDNA sequences of *Microcosmus squamiger* were deposited under the accession numbers [to be completed] at the EMBL databank. (A detailed description of the gene content will be discussed elsewhere.)

Identification of orthologs

First, we formatted a reference protein database using a previous phylogenomic dataset of 179 orthologous markers for 51 metazoans (Delsuc et al. 2008). We then scanned the 454 generated data by conducting similarity searches against this database, and *vice versa*, using the programs blastx and tblastn respectively, with a cut-off *e*-value of 10^{-6} (<http://blast.ncbi.nlm.nih.gov/>). Positive blast results were obtained for about 25,000 sequences of *Microcosmus*, corresponding to 66 of the initial 179 nuclear protein-coding genes.

Contigs of the matching sequences were assembled using the program cap3 (Huang and Madan 1999). Orthology assignment was controlled in two ways: (i) by reciprocal blast searches, *i.e.*, by checking that the best hit for a given contig corresponded indeed to the initially assumed gene, and (ii) by gene phylogenetic tree reconstruction. Use of the above criteria allowed us to determine the orthologous sequences of *Microcosmus squamiger* for 35 metazoan-conserved housekeeping genes (Table 1).

Dataset assembly

Individual gene datasets for the 35 housekeeping genes were built upon *in silico* available transcript data for 27 additional bilaterian species. Nucleotide sequences were retrieved from multiple on-line sources, such as the Ensembl Genome Browser (<http://www.ensembl.org/>), the EST Database (<http://www.ncbi.nlm.nih.gov/dbEST/>) and Trace Archive (<http://www.ncbi.nlm.nih.gov/Traces/>) of GenBank at the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>), as well as other eukaryote genome databases hosted at the Department of Energy Joint Genome Institute (<http://www.jgi.doe.gov/>).

For each of the 35 nuclear markers, amino acid alignments were initially generated using MAFFT version 5 (Kato et al. 2005) and converted to nucleotide alignments that were subsequently refined manually using MEGA 4 (Tamura et al. 2007). Ambiguously aligned sites were identified at amino acid level and separately removed from all individual gene datasets, using the program Gblocks (Castresana 2000), according to the following parameters: a minimum of 14 sequences for conserved and flanking positions, a maximum of 8 contiguous nonconserved positions, a minimum of 5 positions for the length of a block, a maximum of 50% of gaps per position, and similarity matrices were used. The resulting blocks of each amino-acid alignment were then extended at the nucleotide level.

The concatenation of the 35 phylogenetic markers yielded two 28-taxon datasets, the first including 25,026 unambiguously aligned nucleotide sites, and the second 8,342 amino acid sites. All datasets are available upon request.

Phylogenetic analyses

Probabilistic phylogenetic analyses of the two 28-taxon concatenated datasets were conducted following Bayesian Inference (BI) and Maximum Likelihood (ML) reconstruction approaches. As far as the nucleotide dataset is concerned, additional Bayesian analyses were performed on a reduced dataset including only sites at the first and second codon positions.

Bayesian inferences were conducted under the site-heterogeneous CAT mixture model (Lartillot and Philippe 2004) combined to a 4-category Gamma (G) distribution, *i.e.*, CAT + G for amino-acid and CAT + GTR + G for nucleotide alignments. All analyses were run under the Phylobayes 3.1 software (Lartillot et al. 2009), with deletion of constant sites due to computing time limitations. For each dataset, two independent Markov Chains Monte Carlo (MCMC) were launched for [to be completed] cycles, with parameters and trees sampled every [to be completed] cycles. Priors were set to default values and convergence of the two

chains was checked by monitoring the marginal likelihoods through cycles. Bayesian Posterior Probabilities (PP) were obtained from the 50% majority rule consensus of the trees sampled during the stationary phase, after a burnin of [to be completed].

ML analyses of the nucleotide dataset were conducted using PAUP* 4.0b10 (Swofford 2002). Under the selected GTR + G model and starting from a neighbour-joining (NJ) tree, model parameters were estimated using the likelihood criterion and further kept fixed for heuristic searches with Tree Bisection Reconnection (TBR) branch swapping. For the amino acid dataset, ML analyses were performed using the program PHYML 3 (Guindon and Gascuel 2003) under the CAT + G model with the number of CAT categories set to 20 (Quang le et al. 2008). The heuristic ML searches were conducted by performing Subtree Pruning and Regrafting (SPR) moves on a Neighbor-Joining (NJ) starting tree.

For both nucleotide and amino-acid ML trees, statistical support was estimated by Bootstrap resampling with 100 pseudo-replicates generated by the program SEQBOOT of the PHYLIP package (Felsenstein 2005). In all replicates, ML analyses were performed as described above for the original data sets. Bootstrap Percentages (BP) were obtained from the 50% majority rule consensus of the 100 reconstructed trees using the program TREEFINDER (Jobb 2007).

Compositional heterogeneity and saturation

To control for potential systematic errors, we checked for the occurrence of compositional biases and for the extent of substitutional saturation in the data. All analyses were realized using the R statistical package (Team 2004) and carried out for all types of concatenated datasets, nucleotide and amino-acid, as well as for the first, second, and third codon positions. We used Principal Component Analyses (PCA) to control for compositional heterogeneity, whereas the amount of saturation was estimated by plotting the observed

pairwise distances on the alignments against the tree-inferred distances. The latter path-length distances were calculated from CAT + Gbranch lengths estimated under a fixed topology.

Testing relative rate variations between tunicates and vertebrates using local clocks.

Given the deep divergences of our phylogeny, all following analyses were conducted at the amino-acid level in an effort to minimize the misleading effects of saturation. The tree topology previously obtained from the Bayesian CAT + G analyses of the concatenated amino-acid dataset was set as a reference, and the constancy or variation of rates of evolution was tested between tunicate and vertebrate lineages using the program CODEML of the PAML package version 4.2 (Yang 2007).

We first specified two local clock models: (1) a two-rate model, where all branches in our phylogeny conform with the clock assumption and have a default rate ($r_0 = 1$) to the exception of Olfactores clade, encompassing tunicate and vertebrate branches, that have a different rate r_1 ; and (2) a three-rate model in which all tunicate and all vertebrate branches were assigned distinct rates, r_1 and r_2 .

Maximum likelihood (ML) analyses were performed under the two afore-described clock models on the concatenated dataset as well as on each individual gene alignment separately. Given that the first model is hierarchically nested into the second, log-likelihoods of each model were compared using likelihood ratio tests (LRTs) (Felsenstein 1981), with degrees of freedom equal to one. The associated p -values were corrected for multiple testing, by being multiplied by the total number of genes tested in parallel (Bonferroni correction) (Boorsma et al. 2005). In these tests, the two-ratio model (1) represented the null hypothesis that the rate of evolution is homogenous between tunicate and vertebrate clades.

When the null hypothesis was not rejected, constancy of the protein evolutionary rates between the two lineages was assumed, since considering that tunicate and vertebrate rates

differ from each other did not provide any significant statistical improvement to the given dataset. In the opposite case, the model 2 had a significantly better fit on the data and thus the evolutionary rate was considered to vary between the two lineages.

Protein evolutionary rate estimation under autocorrelated models of rate evolution

All amino-acid alignments for the 35 orthologous markers were slightly modified with regards to the taxon sampling in order to prevent the occurrence of missing data for outgroup taxa. In this respect, *Nematostella* sequences were removed from all gene alignments and the missing sequences of *Reniera* (genes *grc5* and *rla2A*) were replaced by the corresponding orthologs of its relative *Suberites domuncula*, both species belonging to the phylum of Sponges.

For each of the 35 resulting protein datasets, branch-specific rates of amino acid evolution were estimated using the program MULTIDIVTIME (Thorne et al. 1998) under the JTT model. We used the previously defined reference topology after pruning *Nematostella*, and we set the age of the root at 700 Mya after Douzery *et al* (Douzery et al. 2004) with no additional constraints on nodes. Two independent MCMC were launched and run for 1 million generations, sampling every 100 generations and discarding the first 100,000 as the burnin. The rates assigned to terminal branches were taken as species-specific evolutionary rates expressed as the number of amino-acid replacements per 100 sites per Myr.

Since outgroups are not taken into account in the node rate calculations, the presence of chimerical sequences for the sponges is not expected to have had any influence on the resulted rate estimates. Finally, analyses were also performed with only the priors [to be completed] and after modifying the age of the root and the prior distribution of the autocorrelation parameter (mean and variance) respectively.

Pairwise estimation of synonymous and nonsynonymous substitution rates for tunicates, teleosts and mammals

To estimate the $w = d_N / d_S$ ratio for tunicates, teleosts and mammals, we first sought in our phylogeny pairs of species within each lineage that exhibit overall similar and as short as possible pairwise path-length distances. Under this criterion, *Ciona intestinalis* and *Ciona savignyi* were selected among tunicates, and the pairs *Danio rerio* – *Tetraodon nigroviridis* and *Monodelphis domestica* – *Homo sapiens* for teleosts and mammals, respectively.

From the 35 previously assembled individual gene datasets, we extracted the sequences of the six afore-mentioned taxa upon which refined nucleotide alignments were built using the program Gblocks (Castresana 2000) and excluding all codons containing gaps (option: no gap positions allowed). Five markers (*psmb-J*, *rla2-A*, *rpl33a*, *rpl39* and *sap40*) had to be discarded through this process because of too much occurrence of missing data. Thirty final individual gene datasets were thus analyzed, encompassing in total 19 083 bp of codon sequences for the three selected pairs of tunicate, teleost and mammal species.

Numbers of synonymous substitutions per synonymous sites (K_s) and nonsynonymous substitutions per nonsynonymous sites (K_a) were estimated for every pair of species and all thirty housekeeping genes, using the Maximum Likelihood (ML) of Goldman and Yang (Goldman and Yang 1994), implemented in CODEML of the PAML package (options in the codeml.ctl files: seqtype = 1, runmode = -1 and CodonFreq = 2). Similar results were also obtained when analyses were repeated using the counting method of Yang and Nielsen (Yang et al. 2000), implemented in YN00 of PAML package (Yang 2007). Finally, the distributions of the synonymous and nonsynonymous substitution rate (d_S and d_N) estimates for tunicates, teleosts and mammals were calculated using the R statistical package (Team 2004).

Results

Phylogenetic analyses on 35 highly conserved genes at both amino acid and nucleotide levels

We built a phylogenetic dataset upon 35 nuclear housekeeping genes, selected under two criteria: (1) their pertinence as phylogenetic markers, questing for previously determined orthologous and highly conserved genes across animals (Delsuc et al. 2008); and (2) their co-occurrence in the *Microcosmus squamiger* sequenced cDNA library, *i.e.* the corresponding ETSs were unambiguously identified within the pool of the 454 generated data.

With a primary focus on chordates, we sampled the amphioxus as the representative species for cephalochordates, as well as seven vertebrate and seven tunicate species for Olfactores. All major metazoan phyla were also represented by species chosen upon their *in silico* availability of sequence data. This yielded a dense matrix of 35 concatenated nuclear genes, encompassing 28 taxa and 25,026 bp (or 8,342 aa) in total, the percentage of missing data being as low as 10.7%.

Phylogenetic reconstructions were conducted following both ML and BI approaches, using standard and mixture models of sequence evolution. All methods yielded almost identical trees at both amino acid and nucleotide levels (Figure 1). The resulted phylogenetic picture conforms to the widely accepted view of animal evolution (Telford 2006) according to which Bilateria can be grouped into Protostomia and Deuterostomia, the former embracing two reciprocally monophyletic lineages; Lophotrochozoa (PP1 = 1, BP1 = 98, PP2 = 0.98, BP2 = 98) and Ecdysozoa (PP1 = 1, BP1 = 99, PP2 = 1, BP2 = 100). All analyses were highly congruent as far as the monophyletic origin of Protostomia was concerned (PP1 = 0.99, BP1 = 93, PP2 = 1, BP2 = 81), yet Deuterostomia relationships appeared more difficult to resolve, particularly in ML trees (PP1 = 0.89, BP1 = -, PP2 = 0.87, BP2 = -). In detail, ML analyses

yielded two topological alternatives regarding BI trees and Deuterostomia evolution, as it is currently assumed: (1) the first involved *Xenoturbella bocki*, which was retrieved into multifurcation with other animal clades yielding the disruption of Deuterostomia; (2) the second involved a spurious embranchment of *Branchiostoma floridae* with Ambulacraria that disrupted chordate monophyly in both amino acid and nucleotide ML trees. Nonetheless, no significant nodal support was obtained for neither of conformations, and similar results were previously shown to reflect to represent phylogenetic reconstruction artifacts (Bourlat et al. 2006; Delsuc et al. 2008), issued from an either important percentage of missing data (*Xenoturbella* 47% in the present case) or a limited taxonomic sampling (*Branchiostoma* and Ambulacraria).

Within Chordata, a sister-group relationship of tunicates with vertebrates was unequivocally recovered by all methods or type of dataset used (PP1 = 1, BP1 = 94, PP2 = 1, BP2 = 95). Of note, the Olfactores clade is retrieved from nucleotide-based reconstructions, accrediting furthestmost the robustness of the new chordate phylogeny (Delsuc et al. 2008) and suggesting that the small core of 35 housekeeping genes carries a very strong phylogenetic signal, comparable to that of larger phylogenomic datasets (Delsuc et al. 2006; Delsuc et al. 2008; Dunn et al. 2008; Putnam et al. 2008).

Similarly, the 35 phylogenetic markers appeared informative enough to provide an also clear-cut phylogenetic view within each group. The newly sequenced *Microcosmus squamiger* was found firmly branched with *Halocynthia* (PP1 = 1, BP1 = 100, PP2 = 1, BP2 = 100), a relationship consistent with the traditional classification, both species belonging to the same order and family (Stolidobranchia: Puyridae). Likewise, the resulted relationships for the rest of sampled tunicate species appeared highly concordant to recent phylogenies inferred from either mitochondrial or RNA data (Tsagkogeorga et al. 2009; Turon and Lopez-Legentil 2004; Zeng et al. 2006).

For the majority of the species, branch length estimates obtained from phylogenetic reconstructions on amino acid data were quite similar to those inferred from nucleotide sequences, albeit some moderate deviations were observed for the fast-evolving groups, *i.e.*, Ecdysozoa among Protostomia, and Tunicata among Deuterostomia. *Oikopleura dioica* constitutes however an exception, since its DNA branch length is 1.6 times longer than its amino acid one. Moreover, *Oikopleura dioica* was found to exhibit a highly derivative compositional profile in PCA analyses, occupying an outlier position to the other tunicate and deuterostome taxa, and being closely located to *Caenorhabditis elegans*, the second fastest-evolving species in our phylogeny (Figure 2).

Estimated individual gene rate differences between tunicates and vertebrates under local clocks

In all phylogenetic trees inferred from the full dataset of the 35 proteins, tunicates have displayed the longest branches within Chordata, which indicates they feature an overall faster evolutionary rate than vertebrates and cephalochordates. In order to test whether this acceleration affects all genes to a similar extent, we followed a ML approach to obtain relative rate estimates for the two Olfactores lineages and for each orthologous gene. Log-likelihood comparisons through LRTs between a constant-rate model and a variable-rate model allowed us afterwards to test for statistical significance in rate differences between the two groups. Our results indicated that the relative rate of amino acid replacement of tunicates and vertebrates differs significantly between the two groups in 20 genes (Table 2). With regards to the direction of the contrast, the tunicate rate was consistently found to be superior to the one of vertebrates for all the afore-mentioned genes, upholding the assumption of a tunicate rapid evolution for the housekeeping proteins here considered. Moreover, the rate-

ratio estimates ranged from 1.4 to 7.5 depending on the marker, which implies that the evolutionary shift in tunicate rates has not affected all genes to a similar extent (Figure 3).

Unexpectedly, for the remaining 15 genes (Table 2), the LRTs have not significantly rejected the hypothesis of equality of rates between the two groups. Thereby, in these cases the evolutionary rate within Olfactores may be considered roughly constant, a result providing strong evidence that tunicates do carry genes that have escaped the prevalent genomic acceleration.

The outline of the estimated rate variation within Olfactores is shown in figure 4, illustrating the markedly distinct relative rate distributions of tunicates and vertebrates across the 35 phylogenetic markers. When the same analytical protocol was applied to the concatenated dataset, a rate ratio of 1.9 in orders of magnitude was obtained. Therefore, it can be thus suggested that (1) the accelerated evolution affects the majority of the 35 housekeeping genes to a highly dissimilar extent and (2) that tunicates exhibit a nearly twofold faster rate of protein evolution compared to vertebrates.

Within-group rate variation as estimated under an autocorrelated model of rate evolution

Given the extensive variation upon gene, we next asked whether similar fluctuations can also be observed in species rates, and if yes, how they are linked to the afore described contrasts: do all tunicates contribute equally to the estimated rate differences among chordates, or the large range of among-gene variation is rather driven by only few species?

To answer this question, we estimated the rate of amino acid replacement in all branches of our phylogeny over the 35 housekeeping genes using an autocorrelated model of rate evolution. This allowed us to simultaneously explore two aspects of rate variation within Chordata, *i.e.*, across genes, and across branches of its three clades. Figure 5 shows the

branch-specific distributions for tunicate, vertebrate and the cephalochordate amphioxus rates, as estimated across the 35 molecular markers.

In overview, all branches of the tunicate clade were characterized by an accelerated evolutionary rate as compared to the other chordate groups, whereas the lineage-specific estimates were found to vary extensively across the sampled tunicate species. This yielded a particularly heterogeneous picture for tunicate rates, that contrasted sharply with the highly overlapping rate distributions of vertebrates.

More precisely, the two phlebobranch ascidians *Ciona* seemed to exhibit the lowest rates among tunicates, with the stolidobranch *Halocynthia*, *Microcosmus* and *Molgula* following next, conforming to previously reported phylogenetic results (Tsagkogeorga et al. 2009; Yokobori et al. 2006; Zeng et al. 2006). Interestingly, the evolutionary-rate heterogeneity appeared radically lessened when considering species of the same order, as best exemplified by the overlap of the resulted distributions within Stolidobranchia.

Notably, the aplousobranch *Diplosoma*, and particularly the appendicularian *Oikopleura* were detected as highly divergent, with rates deviating greatly from the average values of all other sampled tunicate in particular, and chordates in general. These two representatives were also marked by the largest variance in estimates across genes, illustrated by the quartiles of their rate distributions.

Considering the previously described local-clock analyses, we finally examined the among-branch variation over the 15 genes for which LRTs have not rejected the hypothesis of a homogeneous rate among Olfactores. As expected, rate estimates have also yielded a weaker disparity in rates across Chordata (Figure 6). Interestingly, this rather uniform picture among chordate lineages appeared to lie mainly in the decrease of *Diplosoma listerianum* and *Oikopleura dioica* rates. The two *Ciona*, and the three stolidobranch ascidians (*Halocynthia*, *Microcosmus* and *Molgula*) were characterized by more smooth changes, as their rate profile

across the 15 genes was almost identical to that issued from the full dataset. Thereby, it could be deduced that rate variation in the two most divergent species, *Oikopleura* and *Diplosoma*, contributed the most to the extent of rate discrepancies among tunicates, vertebrates and cephalochordates in our analyses.

d_N/d_S ratio variation between lineages

Comparing synonymous (d_S) and nonsynonymous (d_N) rates in genes from different lineages constitutes a prevailing approach for understanding the mechanisms of molecular change (Kryazhimskiy and Plotkin 2008). Nevertheless, a reliable estimation of these parameters for our full dataset was precluded by the high levels of saturation (figure 7). Resulted from the deep divergences or/and the presence of rapidly evolving taxa in a dataset (see above), saturation has been previously reported to yield a general underestimation of the d_S (Nabholz et al. 2008). Endeavoring to get insights on the determinants of the rate contrast between tunicates and the other chordates, we thus chose to compare d_N/d_S for a subset of taxa representing three lineages within Olfactores. With respect to tunicates, we estimated the d_N/d_S for the two congeneric species of *Ciona*, being the less distant pair of tunicates in our phylogeny. In parallel, we compared the tunicate estimates to the pairs *Danio rerio* - *Tetraodon nigroviridis* and *Monodelphis domestica* - *Homo sapiens*, respectively. Five genes had to be excluded from the analysis because of missing data, whereas genes for which we obtained a d_S greater than 10 are not shown, since they undoubtedly present unreliable estimates.

The synonymous and nonsynonymous rate ratios (d_N/d_S), as estimated for the three lineages across a subset of 26 genes are shown in figure 8. All estimates were clearly inferior to 1, which clearly suggests that the ensemble of the examined genes is under strong purifying

selection for all the represented lineages. This is consistent with the fact that every gene of our analysis is conserved across a notably large taxonomic scale, that of animals, and deserves a fundamental, housekeeping function in the cell.

Overall, the ω values ranged from 0.001 to 0.14 depending on the lineage and the gene, yet none of the lineages presented a systematically higher d_N/d_S as compared to the others across genes, ruling out the hypothesis that evolutionary rate variation is due to a relaxation of selective constraints in one lineage. Instead, the resulted picture of ω across the three Olfactores groups was rather uniform, as illustrated by the highly similar distributions of d_N/d_S for tunicates, telosts and mammals (figure 9). In conclusion, for the set of genes here analyzed, it can be suggested that the underlying cause for tunicate accelerated evolution among chordates seems to be an elevated mutation rate, rather than changes in selective pressures.

Discussion

Determining new phylogenetic markers for tunicates using 454 sequencing

Over the past four years, massively parallel sequencing platforms have become available, enabling the generation of large amounts of genomic data at extremely short times and accessible costs (Shendure and Ji 2008). Using such technologies for tunicates seemed to us a promising perspective from a phylogenetic and comparative genomics standpoint, mainly because this lineage is far from satisfying the prerequisites for the successful development of phylogenetic markers through traditional PCR approaches.

Evolving at elevated rates (Tsagkogeorga et al. 2009; Yokobori et al. 2006; Zeng et al. 2006; Zeng and Swalla 2005), tunicate protein-coding genes exhibit high levels of sequence divergence leading to large amounts of saturation, even when considering closely related

species (ex. *Ciona*). Thereby, the identification of conserved regions in a gene, suitable for primer designation, appears to be a particularly difficult task, as also can be deduced by the fact that besides whole-genome projects, currently available genomic information for tunicate representatives comes from EST projects that do not technically entail a minimum of *a priori* information and low sequence divergence.

Here, we first applied the high-throughput 454 technology (<http://www.454.com/>) to sequence the partial transcriptome of an ascidian tunicate, *Microcosmus squamiger*, a lessepsian migrant species widespread in Mediterranean sea and reported to have a great invasive potential in other parts of the world (Turon et al. 2006). We sought to identify orthologous sequences for a set of 179 previously determined phylogenetic markers, pertaining for large-scale comparative analyses (Delsuc et al. 2008). At the time this work was conceived, the 454 sequencing system generated an average of 100 megabases per run, at short 250 base-reads. Still this approach permitted us to determine *in silico* 66 putatively homologous genes, among which 35 were unambiguously identified as orthologous sequences to our query genes.

Although this might seem a small number compared to a 50-million-base of resulted data, several factors argue for the great potential of 454 sequencing approach in tunicate phylogenetics and genomics. First, the sequencing of *Microcosmus* transcriptome was conducted on a cDNA library that was not normalized. This yielded an important level of coverage or/and redundancy for highly expressed genes in our data, and the 35 housekeeping genes corresponded to a non-negligible amount of ~20,000 sequences. Second, albeit high-throughput sequencing alleviates bench problems, still the fast-evolving nature of tunicate species has a substantial impact on the post-data analyses, by hindering the reliable assignment of orthology between highly divergent sequences. Genome analyses of *Ciona* genome have demonstrated that in a transcriptome composed by 16,000 genes, only a small

fraction (5%) matched to vertebrate genomes, whereas about 60% of genes is shared with protostome animals (Pennisi 2002). In this respect, the layout of 454 sequencing at the time, has literally reinforced the problem of divergence, since short sequences are in general hard to treat through blast searches and in subsequent gene assemblies.

In conclusion, considering (1) tunicate scarcity in genomic data and available phylogenetic markers, (2) the results from the pilot application of a high-throughput sequencing approach to the group, and (3) the ongoing advances of these technologies in terms of layout, it can be suggested that the 454 sequencing approach provides a compelling tool for the construction of large comparative datasets allowing for more comprehensive studies on tunicate phylogenetics and molecular evolution.

Tunicates evolve twice as fast as vertebrates, but not all genes nor all lineages are equal determinants of this contrast

To quantify the tunicate molecular divergence in a comparative chordate framework, we assessed evolutionary rate variation within Chordata on the basis of 35 housekeeping genes. Overall, our results suggested that tunicates have experienced a nearly twofold faster evolution compared to their vertebrate counterparts within Olfactores (figure 4), whereas the cephalochordate rate pattern appeared fairly similar to that of vertebrates (figure 5).

From a gene standpoint, shifts in tunicate rates have been detected in the majority of the examined genes, yet the degree of the predicted acceleration varied extensively upon marker. Outlier estimates indicated an 1.4- to an uppermost 8- fold accelerated rate for tunicates versus vertebrates, clearly suggesting that not all genes are equal determinants of rate discrepancies within Chordata (figure 3).

In a large proportion of genes (15 out of 35 genes), tunicate evolutionary rates were recovered unexpectedly similar to those of vertebrates (Table 2). This result unequivocally

implies that despite the generally striking genomic divergence (Dehal et al. 2002; Seo et al. 2001; Small et al. 2007a), at least some coding parts of tunicate genomes have escaped the prevalent acceleration bursts.

Considering lineage-specific rate profiles, Phlebobranchia and Stolidobranchia ascidians exhibited lower among-gene variation, whereas the inclusion of *Microcosmus* allowed expanding taxonomic coverage of Stolidobranchia, providing clues for a rather uniform rate at the order taxonomic level. Finally, Aplousobranchia, and particularly Appendicularia, were characterized instead by the least homogenous rates across the examined markers, though the limited taxonomic sampling to a sole representative species per group precluded further insights into lineage-specific rate variation.

In overview, tunicates presented a propensity for extensive rate variation across genes and lineages but certainly our results cannot be generalized to a genome scale, since 35 housekeeping genes constitute a very small portion of the chordate gene repository. Due to the necessity of accurate orthology assessment among compared sequences, studies in comparative genomics are frequently based on highly conserved molecular markers being most of times restricted to only few functional classes of genes. The main challenge for forthcoming studies would be to counterbalance the trade-off between fast-evolving sequences and orthology assessment, in order to explore tunicate, and thereby chordate, rate variation at a larger scale. This would allow testing the validity of two prefatory hypotheses issued from this work: (1) The high heterogeneity of across-gene rates of protein evolution and (2) the preponderant impact of Appendicularia and Aplousobranchia deceleration / acceleration on the whole image of chordate evolution.

From the opposite angle, it would also be interesting to attempt lower taxonomic scale analyses, under the goal to decrease the evolutionary distance separating sampled species. The number of shared orthologous genes between two species expected to be negatively correlated

to their evolutionary distance, more closely relatives leave a much broader potential in terms of number of molecular markers to be examined.

Insights into the underlying causes of tunicate accelerated evolution

Shifts in rates of molecular evolution in genes as well as among and within lineages could result from different factors, such as changes on selective constraints and positive selection, or changes in mutation rates, *i.e.* a general acceleration of nucleotide substitutions (Nabholz et al. 2008). In a preliminary effort to understand the causes that drove tunicate acceleration, we compared synonymous (d_S) and nonsynonymous (d_N) rates of *Ciona intestinalis* and *Ciona savignyi* versus that of teleosts and mammals.

The d_N/d_S ratio was roughly similar for the three lineages, albeit some pronounced differences observed for some genes (figure 8). Tunicates were not consistently retrieved with a higher d_N/d_S , and no signature of positive selection was traced in our data. On the contrary, all examined genes seemed to be under strong purifying or negative selection, which in conjunction with the rather uniform picture of w (figure 9) implies that tunicate rate shift with regards to vertebrates, and also chordates, resulted from an increase of their mutation rate.

Yet a generalization of our findings would be precarious, obviously because of the limited gene sampling but also because our estimations suffer from the uncertainty due to putative confounding effects of saturation (Kryazhimskiy and Plotkin 2008). Still, as a first insight into the causes of tunicate acceleration our results encourage for a thorough research in determining mutation rate variation within tunicates. Variability in mutation rate has been so far attributed to a number of intrinsic factors, such as generation time and metabolic rate, sexuality and more recently longevity (Nabholz et al. 2008). Undoubtedly, the highly diversified life history traits of the group (Lambert 2005) provide a challenging ground for

identifying links between biologic, environmental and functional factors, and changes in mutation rates.

Conclusions

Tunicates (ascidians, pyrosomes and salps, and appendicularians) have fairly been established as an unconventional chordate lineage, that have witnessed a unique evolutionary course compared to vertebrates and cephalochordates. In brief, tunicate evolution is today viewed through an oversimplification of chordate body plans, the relaxation of developmental constraints, the whole-genome contractions, the experience of dramatic gene losses and rearrangements, and a strikingly rapid rate of molecular evolution at both nuclear- and mitochondrial-genome levels.

Less advanced than expected, comprehensive analyses of tunicate divergent evolution are so far hampered by a prominent paucity in molecular markers and the disproportionably distributed genomic data to only few representatives. Here, we introduced the recently developed 454 high-throughput sequencing technology to the goal of the *in vitro* acquisition of genomic data for the group, targeting a previously unsampled invasive tunicate, *Microcosmus squamiger*. Within the pool of the 454 generated data, we *in silico* pinpointed 35 orthologous phylogenetic markers, consisting of a core of universal and highly conserved housekeeping genes across bilaterians. On the basis of these markers, we sought to better understand tunicate rapid evolution, by assessing the extent and determining the components of rate contrasts within chordates.

Despite the limited power of 35 genes to provide a conclusive image of the within-chordate rate variation, our study bears a first round of elements, worth to be tested in larger-scale analyses. More precisely, our results plead for (1) a two-fold faster evolution of tunicate

group within chordates; (2) a marked asymmetry and heterogeneity in tunicate rates, both across genes and across major lineages; (3) the prevailing effects of Appendicularia and Aplousobranchia rate variation to the extent of rate contrasts within Chordata and finally; (4) the predominant role of change in mutation rate as the underlying cause of tunicate accelerated evolution.

Acknowledgments

This work has been supported by the Research Networks Program in BIOINFORMATICS of the High Council for Scientific and Technological Cooperation between France and Israël, and the bioinformatics cluster of the ISE-M. This publication is contribution N° 2009-YYY of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 - CNRS).

References

- Boorsma A, Foat BC, Vis D, Klis F, Bussemaker HJ (2005) T-profiler: scoring the activity of predefined groups of genes using gene expression data. *Nucleic Acids Res* 33:W592-5
- Bourlat SJ, Juliusdottir T, Lowe CJ, Freeman R, Aronowicz J, Kirschner M, Lander ES, Thorndyke M, Nakano H, Kohn AB, Heyland A, Moroz LL, Copley RR, Telford MJ (2006) Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* 444:85-8
- Bourlat SJ, Nielsen C, Economou AD, Telford MJ (2008) Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol* 49:23-31
- Cameron CB, Garey JR, Swalla BJ (2000) Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc Natl Acad Sci U S A* 97:4469-74
- Canestro C, Bassham S, Postlethwait JH (2003) Seeing chordate evolution through the *Ciona* genome sequence. *Genome Biol* 4:208
- Canestro C, Postlethwait JH (2007) Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol* 305:522-38
- Canestro C, Yokoi H, Postlethwait JH (2007) Evolutionary developmental biology and genomics. *Nat Rev Genet* 8:932-42
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540-52
- Corbo JC, Di Gregorio A, Levine M (2001) The ascidian as a model organism in developmental and evolutionary biology. *Cell* 106:535-8
- Davidson B (2007) *Ciona intestinalis* as a model for cardiac development. *Semin Cell Dev Biol* 18:16-26
- Dehal P, Boore JL (2005) Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 3:e314
- Dehal P, Satou Y, Campbell RK, Chapman J, Degnan B, De Tomaso A, Davidson B, Di Gregorio A, Gelpke M, Goodstein DM, Harafuji N, Hastings KE, Ho I, Hotta K, Huang W, Kawashima T, Lemaire P, Martinez D, Meinertzhagen IA, Nacula S, Nonaka M, Putnam N, Rash S, Saiga H, Satake M, Terry A, Yamada L, Wang HG, Awazu S, Azumi K, Boore J, Branno M, Chin-Bow S, DeSantis R, Doyle S, Francino P, Keys DN, Haga S, Hayashi H, Hino K, Imai KS, Inaba K, Kano S, Kobayashi K, Kobayashi M, Lee BI, Makabe KW, Manohar C, Matassi G, Medina M, Mochizuki Y, Mount S, Morishita T, Miura S, Nakayama A, Nishizaka S, Nomoto H, Ohta F, Oishi K, Rigoutsos I, Sano M, Sasaki A, Sasakura Y, Shoguchi E, Shin-i T, Spagnuolo A, Stainier D, Suzuki MM, Tassy O, Takatori N, Tokuoka M, Yagi K, Yoshizaki F,

- Wada S, Zhang C, Hyatt PD, Larimer F, Detter C, Doggett N, Glavina T, Hawkins T, Richardson P, Lucas S, Kohara Y, Levine M, Satoh N, Rokhsar DS (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298:2157-67
- Delsuc F, Brinkmann H, Chourrout D, Philippe H (2006) Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* 439:965-8
- Delsuc F, Tsagkogeorga G, Lartillot N, Philippe H (2008) Additional molecular support for the new chordate phylogeny. *Genesis* 46:592-604
- Donmez N, Bazykin GA, Brudno M, Kondrashov AS (2009) Polymorphism due to multiple amino acid substitutions at a codon site within *Ciona savignyi*. *Genetics* 181:685-90
- Douzery EJ, Snell EA, Baptiste E, Delsuc F, Philippe H (2004) The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci U S A* 101:15386-91
- Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, Seaver E, Rouse GW, Obst M, Edgecombe GD, Sorensen MV, Haddock SH, Schmidt-Rhaesa A, Okusu A, Kristensen RM, Wheeler WC, Martindale MQ, Giribet G (2008) Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* 452:745-9
- Edwardsen RB, Lerat E, Maeland AD, Flat M, Tewari R, Jensen MF, Lehrach H, Reinhardt R, Seo HC, Chourrout D (2004) Hypervariable and highly divergent intron-exon organizations in the chordate *Oikopleura dioica*. *J Mol Evol* 59:448-57
- Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *J Mol Evol* 17:368-76
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) In: Department of Genome Sciences UoW (ed), Seattle
- Gee H (2008) Evolutionary biology: the amphioxus unleashed. *Nature* 453:999-1000
- Gissi C, Iannelli F, Pesole G (2008) Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* 101:301-20
- Gissi C, Pesole G, Cattaneo E, Tartari M (2006) Huntingtin gene evolution in Chordata and its peculiar features in the ascidian *Ciona* genus. *BMC Genomics* 7:288
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11:725-36
- Goldstone JV, Goldstone HM, Morrison AM, Tarrant A, Kern SE, Woodin BR, Stegeman JJ (2007) Cytochrome P450 1 genes in early deuterostomes (tunicates and sea urchins) and vertebrates (chicken and frog): origin and diversification of the CYP1 gene family. *Mol Biol Evol* 24:2619-31
- Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* 52:696-704

- Gyoja F, Satou Y, Shin-i T, Kohara Y, Swalla BJ, Satoh N (2007) Analysis of large scale expression sequenced tags (ESTs) from the anural ascidian, *Molgula tectiformis*. *Dev Biol* 307:460-82
- Holland LZ, Gibson-Brown JJ (2003) The *Ciona intestinalis* genome: when the constraints are off. *Bioessays* 25:529-32
- Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9:868-77
- Iannelli F, Pesole G, Sordino P, Gissi C (2007) Mitogenomics reveals two cryptic species in *Ciona intestinalis*. *Trends Genet* 23:419-22
- Imai KS, Levine M, Satoh N, Satou Y (2006) Regulatory blueprint for a chordate embryo. *Science* 312:1183-7
- Jobb G (2007) TREEFINDER Munich, Germany
- Katoh K, Kuma K, Toh H, Miyata T (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33:511-8
- Kim JH, Waterman MS, Li LM (2007) Diploid genome reconstruction of *Ciona intestinalis* and comparative analysis with *Ciona savignyi*. *Genome Res* 17:1101-10
- Kim YO, Cho HK, Park EM, Nam BH, Hur YB, Lee SJ, Cheong J (2008) Generation of expressed sequence tags for immune gene discovery and marker development in the sea squirt, *Halocynthia roretzi*. *J Microbiol Biotechnol* 18:1510-7
- Kimbacher S, Gerstl I, Velimirov B, Hagemann S (2009) *Drosophila* P transposons of the urochordata *Ciona intestinalis*. *Mol Genet Genomics*
- Kowalevski A (1868) Beiträge zur Entwicklungsgeschichte der Tunicaten. *Nachrichten Gesellschaft Wissenschaften Göttingen* 19:401-415
- Kryazhimskiy S, Plotkin JB (2008) The population genetics of dN/dS. *PLoS Genet* 4:e1000304
- Lambert C (2005) Historical introduction, overview, and reproductive biology of the protochordates. *Can J Zool* 83:1-7
- Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3. A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286-2288
- Lartillot N, Philippe H (2004) A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21:1095-109
- Mallatt J, Winchell CJ (2007) Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol Phylogenet Evol* 43:1005-22
- Nabholz B, Glemin S, Galtier N (2008) Strong variations of mitochondrial mutation rate across mammals--the longevity hypothesis. *Mol Biol Evol* 25:120-30

- Pennisi E (2002) Comparative genomics. Tunicate genome shows a little backbone. *Science* 298:2111-2
- Putnam NH, Butts T, Ferrier DE, Furlong RF, Hellsten U, Kawashima T, Robinson-Rechavi M, Shoguchi E, Terry A, Yu JK, Benito-Gutierrez EL, Dubchak I, Garcia-Fernandez J, Gibson-Brown JJ, Grigoriev IV, Horton AC, de Jong PJ, Jurka J, Kapitonov VV, Kohara Y, Kuroki Y, Lindquist E, Lucas S, Osoegawa K, Pennacchio LA, Salamov AA, Satou Y, Sauka-Spengler T, Schmutz J, Shin IT, Toyoda A, Bronner-Fraser M, Fujiyama A, Holland LZ, Holland PW, Satoh N, Rokhsar DS (2008) The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453:1064-71
- Quang le S, Gascuel O, Lartillot N (2008) Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* 24:2317-23
- Satoh N (2003) The ascidian tadpole larva: comparative molecular development and genomics. *Nat Rev Genet* 4:285-95
- Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaot M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D (2004) Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* 431:67-71
- Seo HC, Kube M, Edvardsen RB, Jensen MF, Beck A, Spriet E, Gorsky G, Thompson EM, Lehrach H, Reinhardt R, Chourrout D (2001) Miniature genome in the marine chordate *Oikopleura dioica*. *Science* 294:2506
- Shendure J, Ji H (2008) Next-generation DNA sequencing. *Nat Biotechnol* 26:1135-45
- Sierra N, Kusakabe T, Park KJ, Yamashita R, Kinoshita K, Nakai K (2006) DBTGR: a database of tunicate promoters and their regulatory elements. *Nucleic Acids Res* 34:D552-5
- Singh TR, Tsagkogeorga G, Delsuc F, Blanquart S, Shenkar N, Loya Y, Douzery EJP, Huchon D (in press) Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics*
- Small KS, Brudno M, Hill MM, Sidow A (2007a) A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. *Genome Biol* 8:R41
- Small KS, Brudno M, Hill MM, Sidow A (2007b) Extreme genomic variation in a natural population. *Proc Natl Acad Sci U S A* 104:5698-703
- Swalla BJ, Cameron CB, Corley LS, Garey JR (2000) Urochordates are monophyletic within the deuterostomes. *Syst Biol* 49:52-64
- Swalla BJ, Xavier-Neto J (2008) Chordate origins and evolution. *Genesis* 46:575-9
- Swofford D (2002) PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates Sunderland, Massachusetts

- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24:1596-9
- Team RDC (2004) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna(Austria)
- Telford MJ (2006) Animal phylogeny. *Curr Biol* 16:R981-5
- Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15:1647-57
- Tsagkogeorga G, Turon X, Hopcroft RR, Tilak MK, Feldstein T, Shenkar N, Loya Y, Huchon D, Douzery EJ, Delsuc F (2009) An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol Biol* 9:187
- Turon X, Lopez-Legentil S (2004) Ascidian molecular phylogeny inferred from mtDNA data with emphasis on the Aplousobranchiata. *Molecular Phylogenetics and Evolution* 33:309-320
- Turon X, Nishikawab T, Riuser M (2006) Spread of *Microcosmus squamiger* (Ascidiacea: Pyuridae) in the Mediterranean Sea and adjacent waters
J Exp Mar Biol Ecol 342:185-188
- Winchell CJ, Sullivan J, Cameron CB, Swalla BJ, Mallatt J (2002) Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol Biol Evol* 19:762-76
- Yandell M, Mungall CJ, Smith C, Prochnik S, Kaminker J, Hartzell G, Lewis S, Rubin GM (2006) Large-scale trends in the evolution of gene structures within 11 animal genomes. *PLoS Comput Biol* 2:e15
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586-91
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-49
- Yokobori S, Kurabayashi A, Neilan BA, Maruyama T, Hirose E (2006) Multiple origins of the ascidian-Prochloron symbiosis: molecular phylogeny of photosymbiotic and non-symbiotic colonial ascidians inferred from 18S rDNA sequences. *Mol Phylogenet Evol* 40:8-19
- Zeng L, Jacobs M, Swalla B (2006) Coloniality and sociality has evolved once in Stolidobranch ascidians. *Integr. Comp. Biol.* 46:255-268
- Zeng L, Swalla B (2005) Molecular phylogeny of the protochordates: chordate evolution. *Can. J. Zool* 83:24-33

FIGURES

Figure 1. — Superimposition of the consensus Bayesian trees inferred from amino acid sequences and from first and second codon positions of nucleotide sequences under the CAT site-heterogeneous mixture model. Support values for the nodes from the ML analyses on the corresponding datasets are also shown.

Figure 2. — Principal Component Analyses (PCA) to control for compositional heterogeneity among sequences. Principal component analysis of nucleotide (A) and amino acid (B) composition of the 35-gene concatenated data set. The graphs represent the projection of individuals on the first two axes explaining most of the total variance. In both graphs, *Oikopleura dioica* exhibits a highly derivative compositional profile by occupying an outlier position to other sampled tunicate taxa.

Figure 3. — Contrast of local molecular clock estimates among tunicates and vertebrates for the 35 housekeeping proteins. The graph shows the tunicate/vertebrate rate ratio estimates for each of the 35 genes. Red stars indicate genes for which rate differences between the two groups were not found significant after correcting for multiple testing (LTR test, critical e -value=0.05).

Figure 4. — Average contrast of rate variation in amino acid replacements within Olfactors. The figure illustrates tunicate (grey) and vertebrate (white) distributions of the evolutionary rate across the 35 markers. Horizontal bars give the median of rate distributions; boxes give the quartiles; whiskers extend to 1.5 times the interquartile range; and circles are for outliers.

Figure 5. — Branchwise distributions of tunicate, vertebrate and cephalochordate rates of amino acid replacements as estimated under an auto-correlated rate model across 35 proteins. Tunicate distributions are shown in dark grey, vertebrate ones in white and that of the cephalochordate amphioxus in light grey. Horizontal bars give the median of rate distributions; boxes give the quartiles; whiskers extend to 1.5 times the interquartile range; and circles are for outliers.

Figure 6. — Branchwise distributions of tunicate, vertebrate and cephalochordate rates of amino acid replacements as estimated under an auto-correlated rate model across all 35 proteins, the 20 faster and the 15 slower genes respectively. All-gene distributions are shown in white, the 20 fast-evolving-gene distributions in dark grey and the slow-evolving-gene ones in light grey. Horizontal bars give the median of distributions; boxes give the quartiles; whiskers extend to 1.5 times the interquartile range; and circles are for outliers. Red boxes highlight the highly deviating rate distributions of the aplousobranch *Diplosoma listerianum* and the appendicularian *Oikopleura dioica*.

Figure 7. — Levels of saturation inferred for amino acids, first + second codon positions, and third codon-positions. All three plots represent the relationship between tree-inferred distances and observed pairwise distances at amino acid and nucleotide level, respectively. Dotted lines represent the hypothesis of absence of saturation on the data, where estimated and observed distances are equal. Solid lines indicate the regression of the linear model fitted on the data.

Figure 8. — Estimation of the synonymous / nonsynonymous rate ratios (d_N/d_S) according to 26 housekeeping genes and to three pairs of closely related taxa. The graph shows the synonymous and nonsynonymous substitution rates as estimated using the ML method of Yang and Nielsen (codeml) between *C. intestinalis* - *C. savignyi*, *D. rerio* - *T.nigroviridis* and *M. domestica* - *H. sapiens* for each of the 26 genes, with no missing data and a $d_S \leq 10$.

Figure 9. — Overall estimation of the synonymous / nonsynonymous rate ratios in Olfactores. ω ou d_N/d_S distributions for the pairs of *C. intestinalis* - *C. savignyi*, *D. rerio* - *T.nigroviridis* and *M. domestica* - *H. sapiens* across the subset of 26 housekeeping genes.

TABLES

Table 1 — List of the 35 metazoan-conserved housekeeping genes for which orthologous sequences have been identified in the transcriptome of *Microcosmus squamiger*. These markers correspond to a subsample of the ones used in a recent phylogenomic study (Delsuc et al. 2008).

Table 2 — Likelihood Ratio Tests (LRTs) and relative rate variation between tunicates and vertebrates using local clocks.

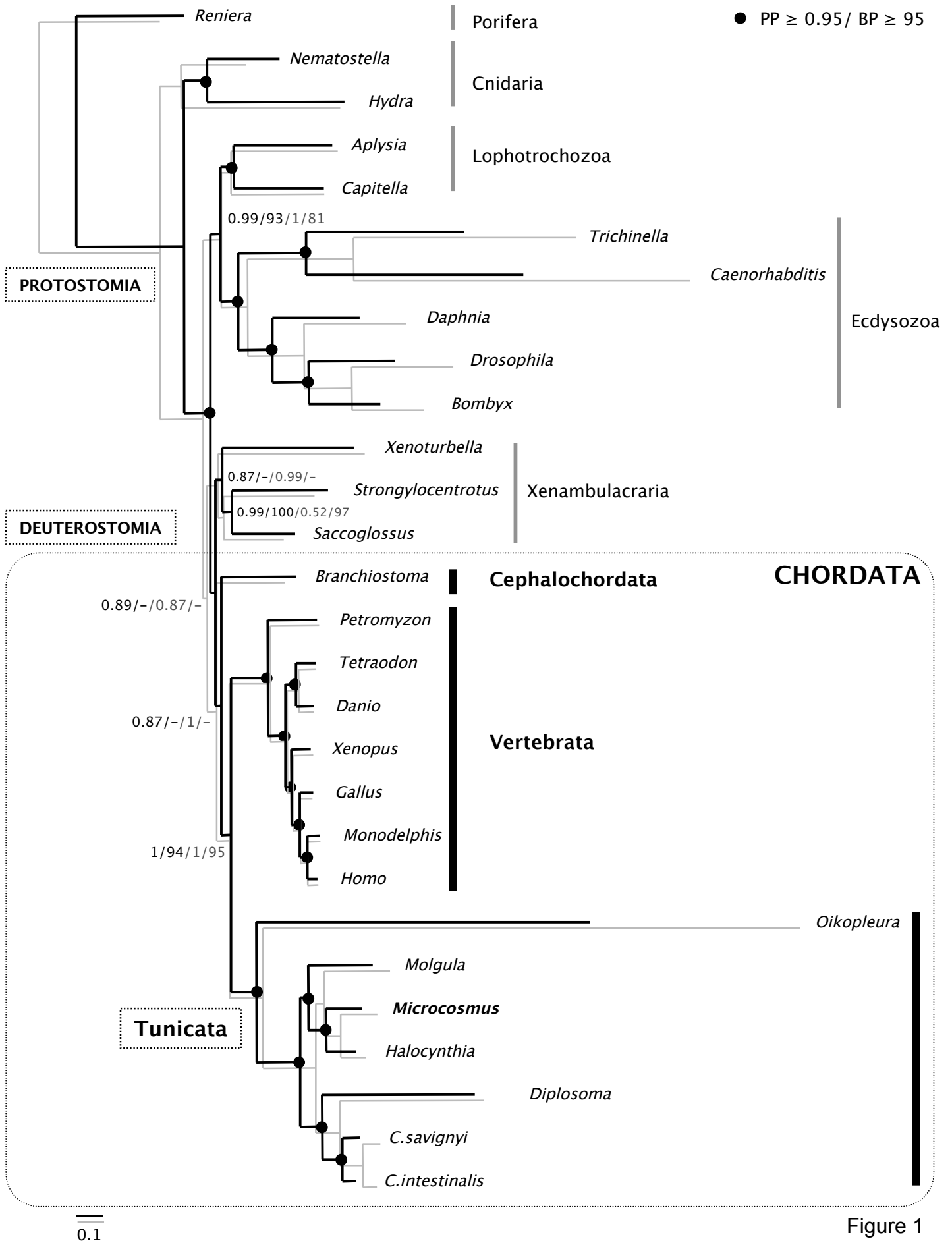


Figure 1

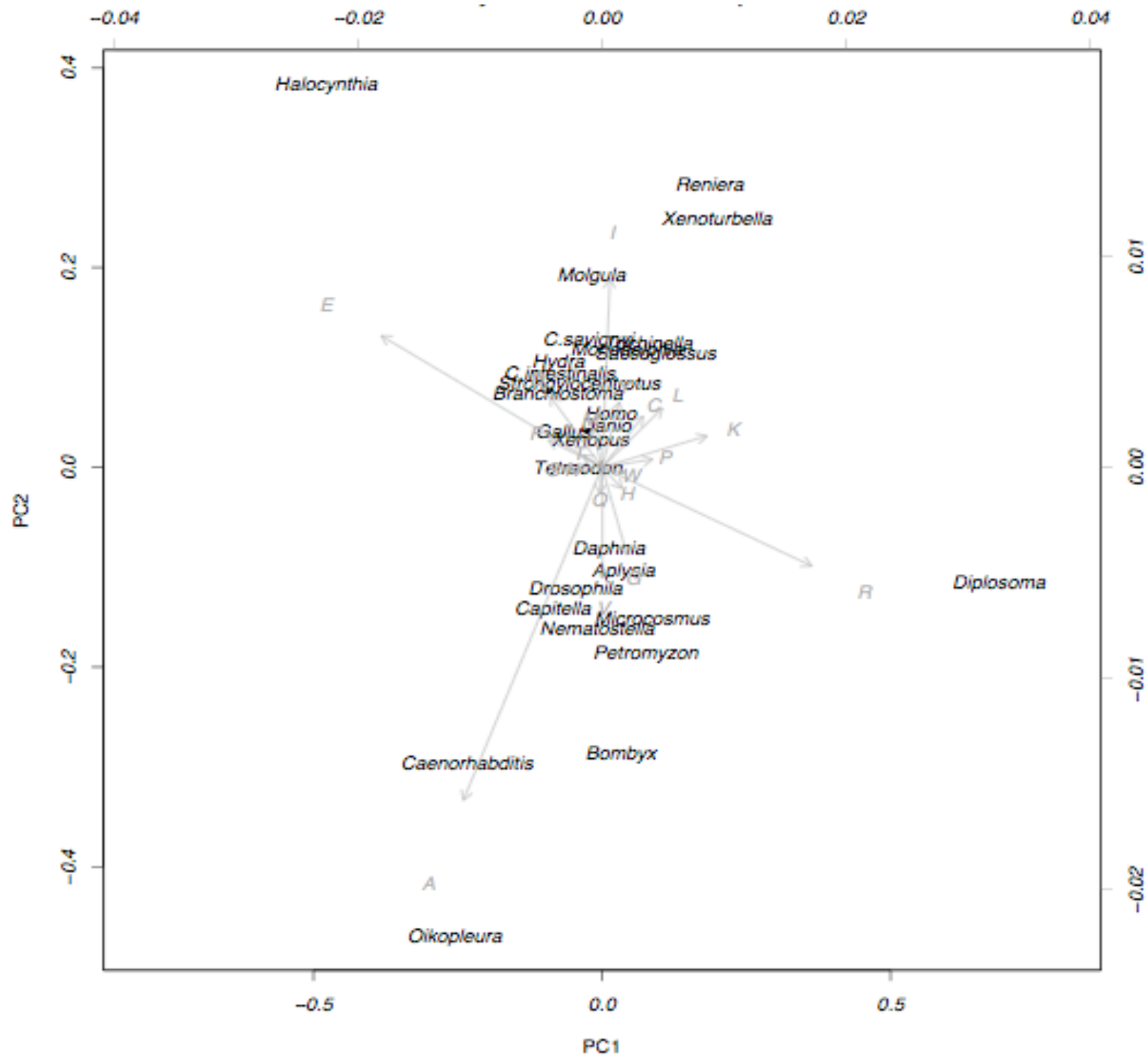


Figure 2A

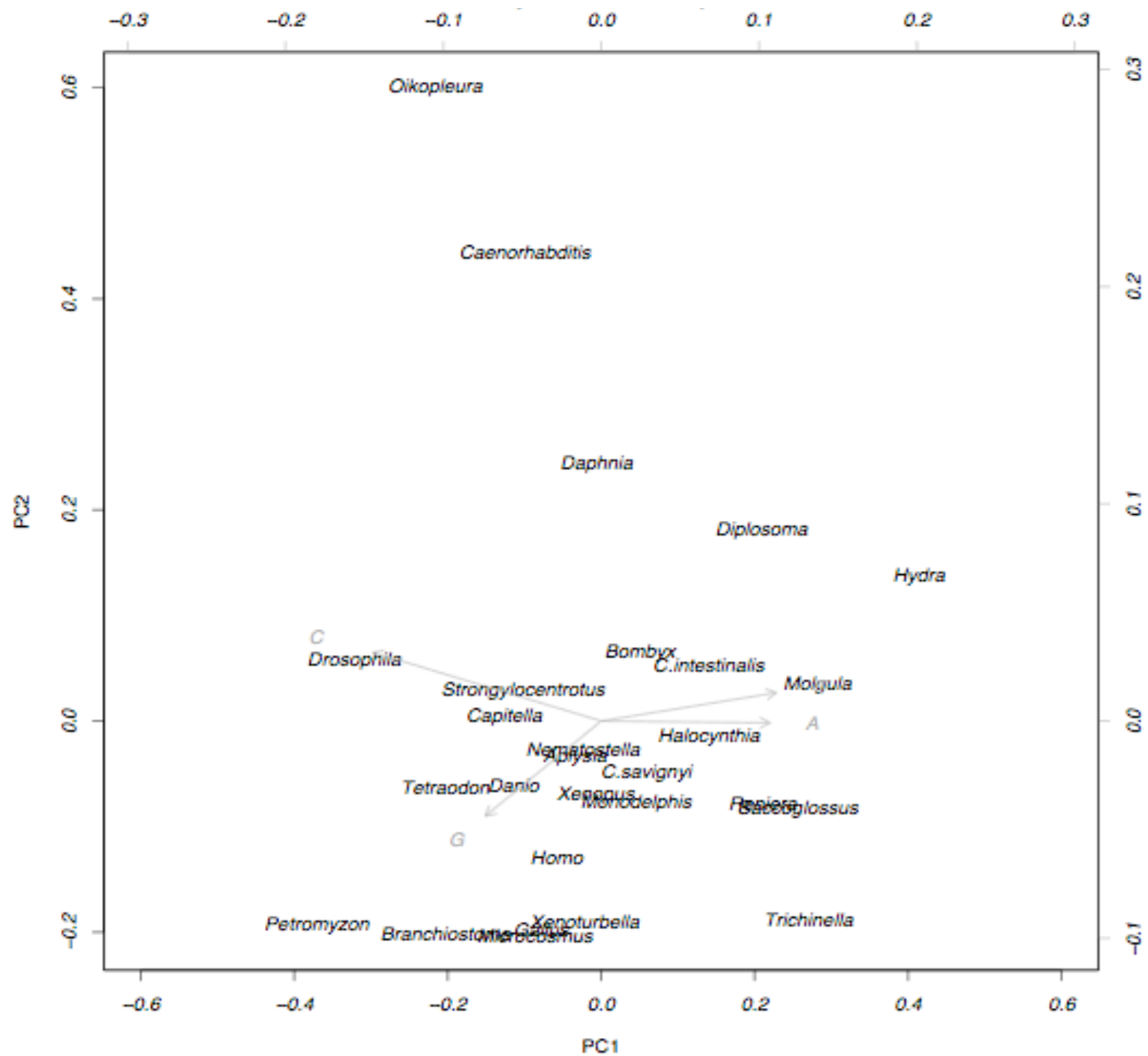


Figure 2B

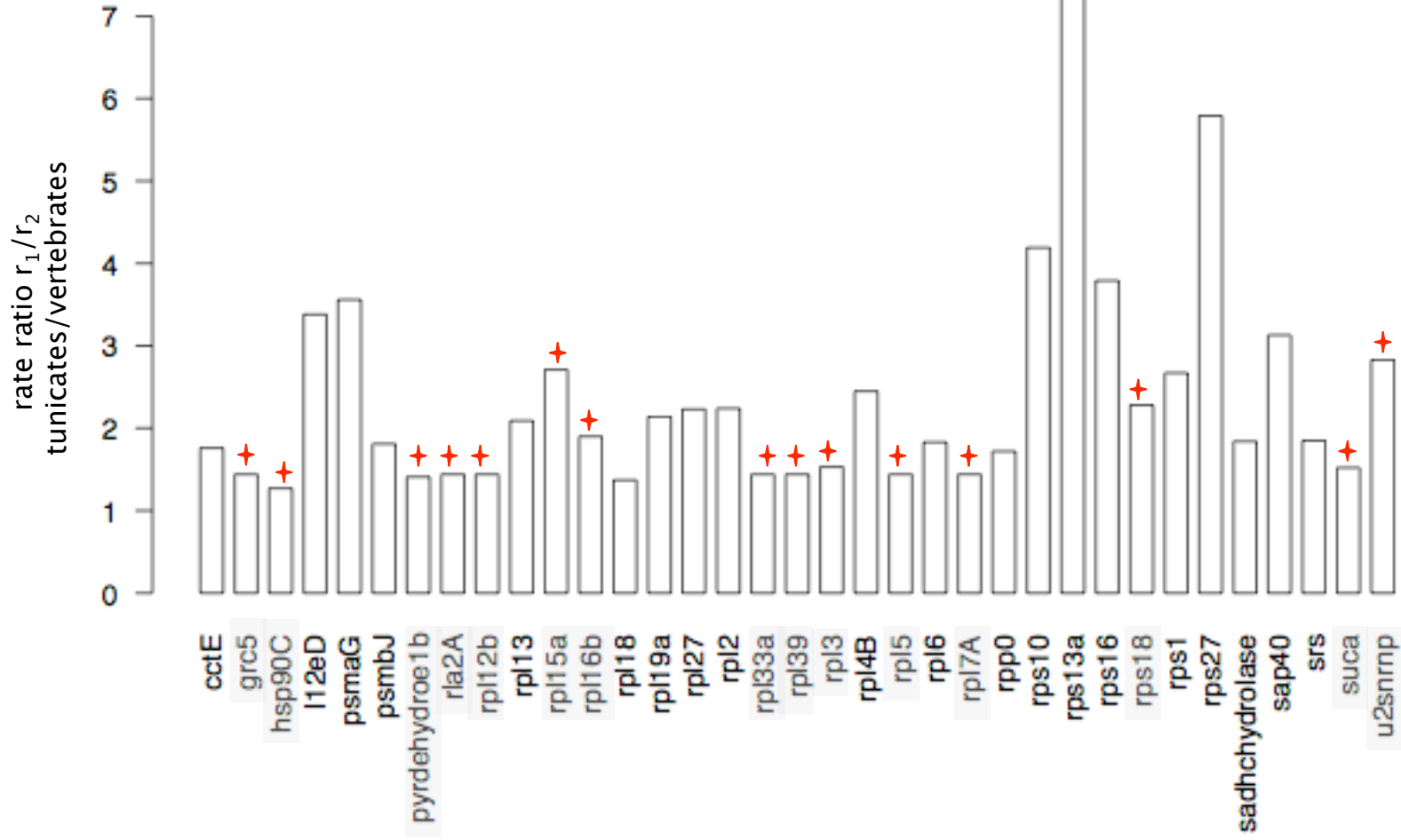


Figure 3

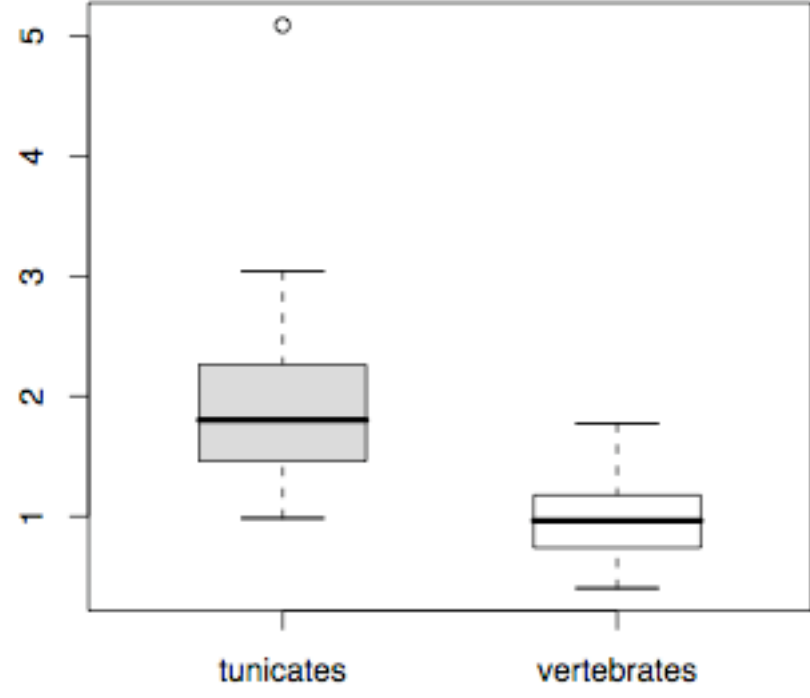


Figure 4

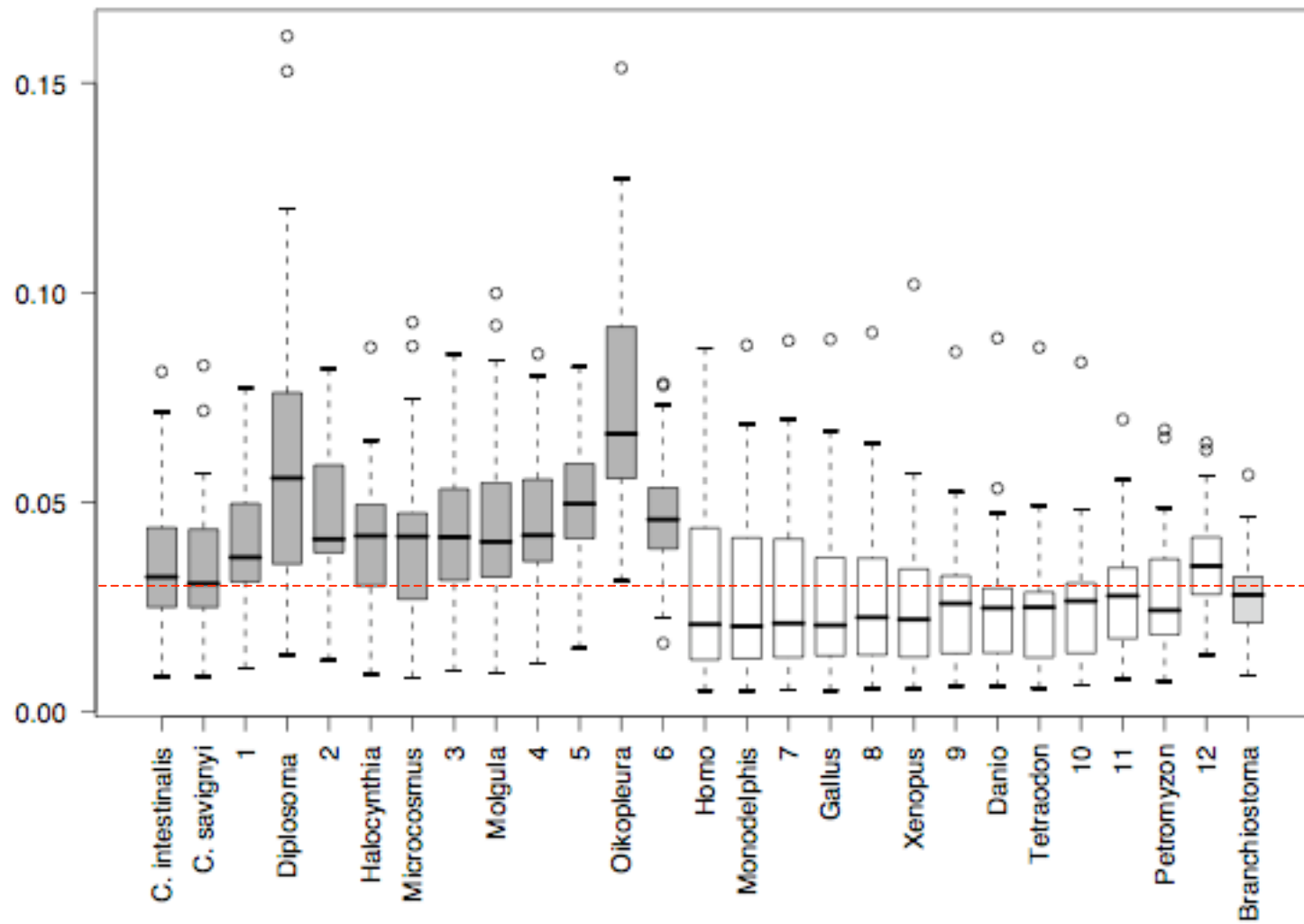
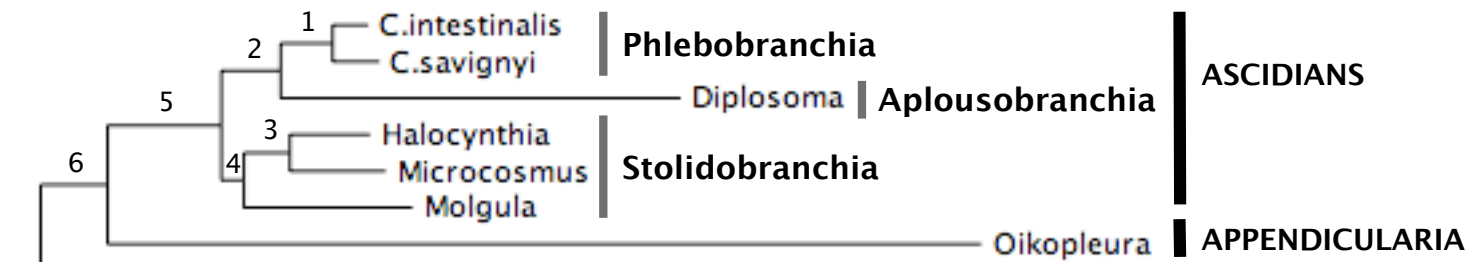


Figure 5

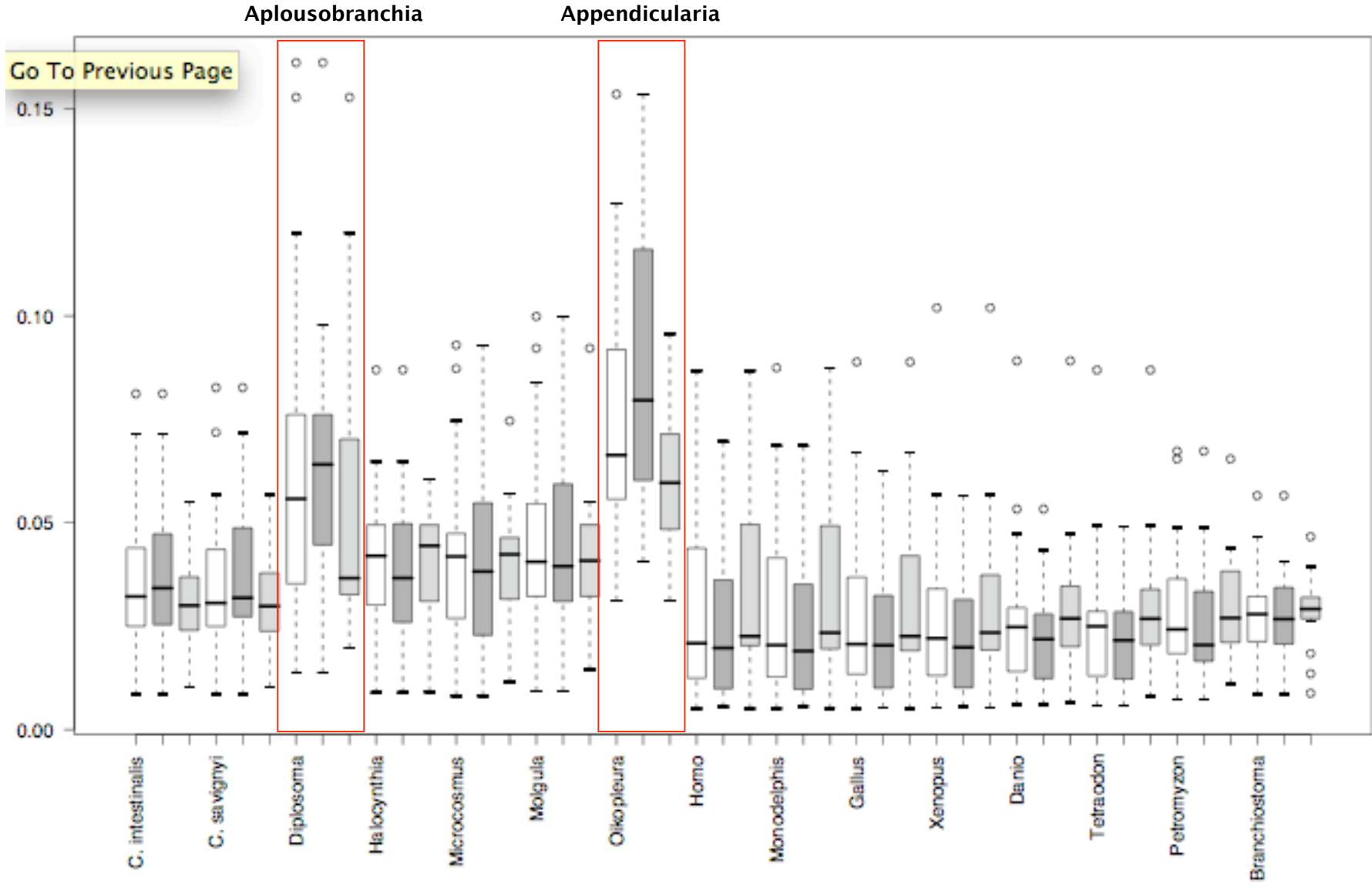


Figure 6

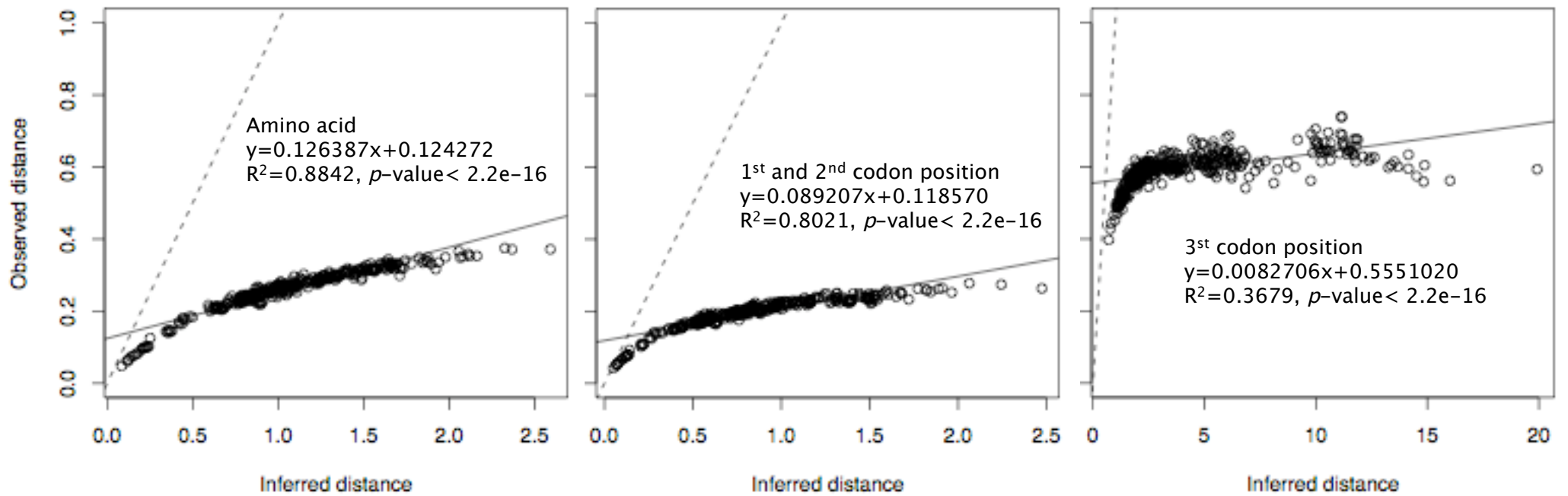


Figure 7

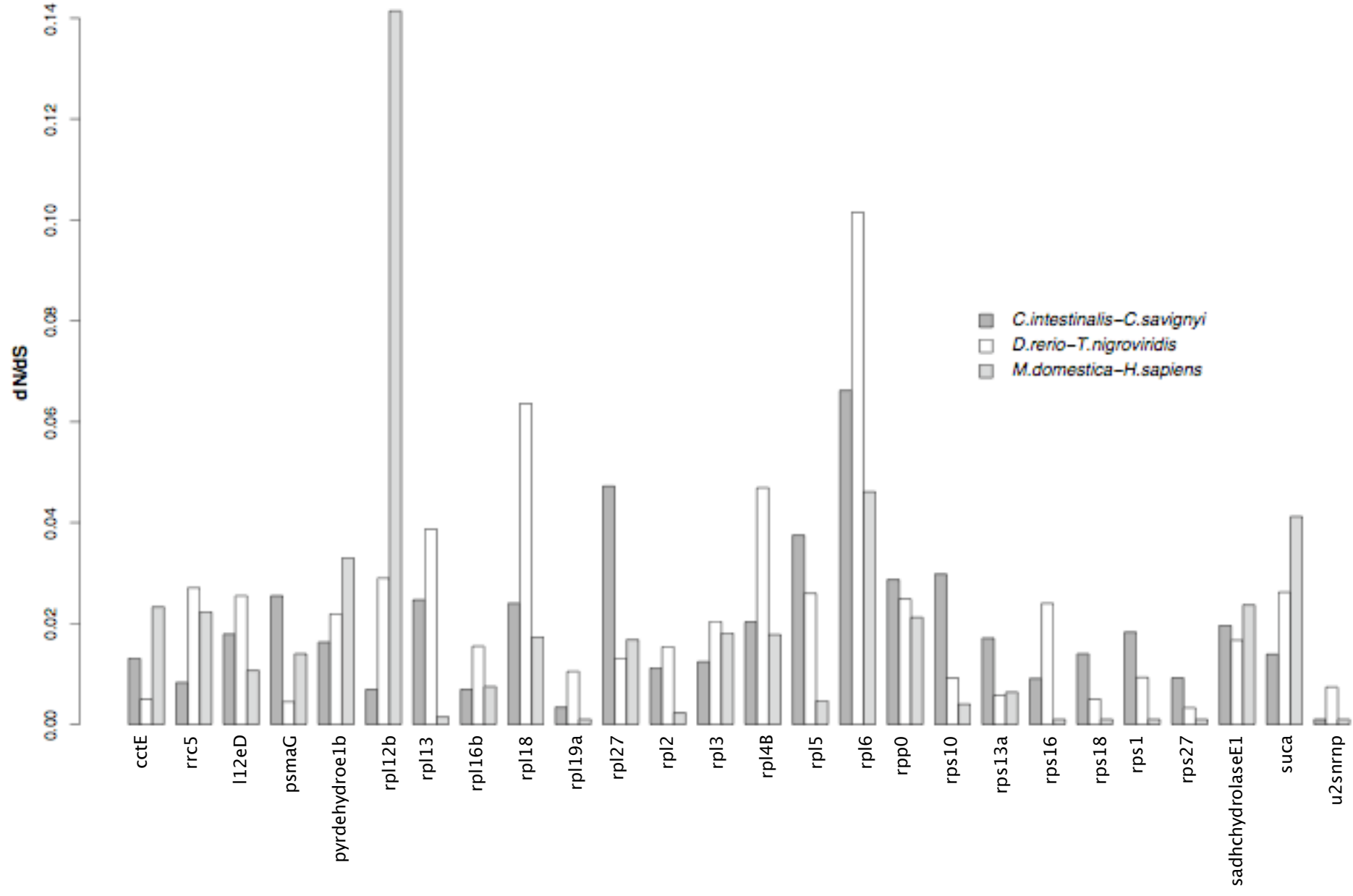


Figure 8

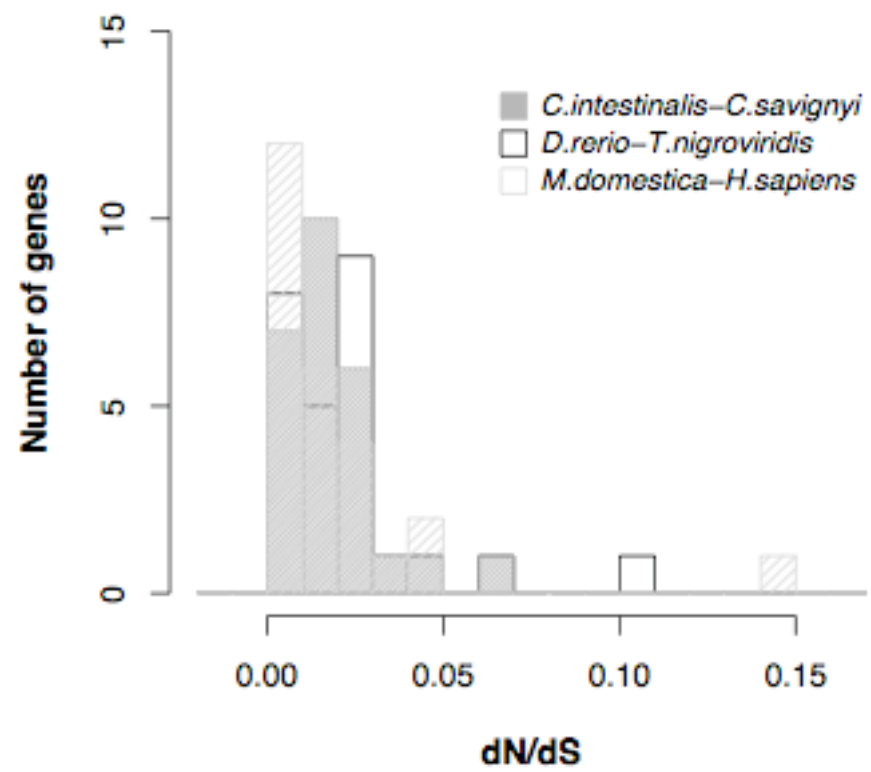


Figure 9

Table 1

Annotation		
<i>Human gene</i>		
Gene	<i>accessions in Ensembl</i>	Name
1	ENSG00000150753	<i>cct-E</i>
2	ENSG00000165496	<i>grc5</i>
3	ENSG00000096384	<i>hsp90-C</i>
4	ENSG00000148303	<i>l12e-D</i>
5	ENSG00000100902	<i>psma-G</i>
6	ENSG00000008018	<i>psmb-J</i>
7	ENSG00000168291	<i>pyrdehydroe1b-mt</i>
8	ENSG00000177600	<i>rla2-A</i>
9	ENSG00000197958	<i>rpl12b</i>
10	ENSG00000167526	<i>rpl13</i>
11	ENSG00000174748	<i>rpl15a</i>
12	ENSG00000142541	<i>rpl16b</i>
13	ENSG00000063177	<i>rpl18</i>
14	ENSG00000108298	<i>rpl19a</i>
15	ENSG00000166441	<i>rpl27</i>
16	ENSG00000161016	<i>rpl2</i>
17	ENSG00000182899	<i>rpl33a</i>
18	ENSG00000163923	<i>rpl39</i>
19	ENSG00000100316	<i>rpl3</i>
20	ENSG00000174444	<i>rpl4B</i>
21	ENSG00000122406	<i>rpl5</i>
22	ENSG00000089009	<i>rpl6</i>
23	ENSG00000147604	<i>rpl7-A</i>
24	ENSG00000089157	<i>rpp0</i>
25	ENSG00000124614	<i>rps10</i>
26	ENSG00000110700	<i>rps13a</i>
27	ENSG00000105193	<i>rps16</i>
28	ENSG00000182498	<i>rps18</i>
29	ENSG00000145425	<i>rps1</i>
30	ENSG00000130041	<i>rps27</i>
31	ENSG00000101444	<i>sadhchydrolase-E1</i>
32	ENSG00000168028	<i>sap40</i>
33	ENSG00000031698	<i>srs</i>
34	ENSG00000163541	<i>suca</i>
35	ENSG00000160201	<i>u2snrnp</i>

Table 2

Gene#	Annotation	Clock type	lnL	rates for branches		χ^2 , df = 1	p-value	Bonferroni correction	r1/r2
				r1	r2	LR = $2^{*(\ln L_1 - \ln L_2)}$		e-value	
Gene01	<i>cct-E</i>	local 1	-8737.53782	1.673					
		local 2	-8727.52147	2.244	1.273	20.03	7.61291E-06	0.000266452	1.763
Gene02	<i>grc5</i>	local 1	-3221.484596	1.248					
		local 2	-3219.932123	1.462	1.014	3.10	0.078054837	2.731919292	1.442 *
Gene03	<i>hsp90-C</i>	local 1	-9778.939133	1.030					
		local 2	-9777.153968	1.150	0.904	3.57	0.05882063	2.058722053	1.273 *
Gene04	<i>l12e-D</i>	local 1	-5193.415857	1.425					
		local 2	-5173.216839	2.207	0.653	40.40	2.07149E-10	7.25022E-09	3.380
Gene05	<i>psma-G</i>	local 1	-5252.068803	1.753					
		local 2	-5222.504286	2.525	0.709	59.13	1.47664E-14	5.16825E-13	3.559
Gene06	<i>psmb-J</i>	local 1	-5029.452127	1.745					
		local 2	-5021.856492	2.244	1.242	15.19	9.71515E-05	0.003400303	1.807
Gene07	<i>pyrdehydroe1b-mt</i>	local 1	-6921.459568	1.352					
		local 2	-6918.762476	1.558	1.105	5.39	0.020203972	0.707139033	1.410 *
Gene08	<i>rla2-A</i>	local 1	-1985.447535	1.098					
		local 2	-1985.175235	0.987	1.174	0.54	0.460532568	16.11863988	0.841 *
Gene09	<i>rpl12b</i>	local 1	-3206.479701	1.328					
		local 2	-3206.29454	1.261	1.441	0.37	0.542828803	18.99900811	0.875 *
Gene10	<i>rpl13</i>	local 1	-5357.891769	1.430					
		local 2	-5348.206588	1.762	0.843	19.37	1.07665E-05	0.000376828	2.090
Gene11	<i>rpl15a</i>	local 1	-3796.498676	1.194					
		local 2	-3786.845562	1.522	0.561	19.31	1.11343E-05	0.000389699	2.712
Gene12	<i>rpl16b</i>	local 1	-4266.442994	1.218					
		local 2	-4262.35874	1.412	0.745	8.17	0.004262396	0.149183847	1.896 *
Gene13	<i>rpl18</i>	local 1	-4271.806991	1.271					
		local 2	-4269.852113	1.460	1.066	3.91	0.048006575	1.680230126	1.370 *
Gene14	<i>rpl19a</i>	local 1	-3206.83912	1.734					
		local 2	-3201.611347	2.63	1.232	10.46	0.001222818	0.042798619	2.135
Gene15	<i>rpl27</i>	local 1	-2531.771729	1.448					
		local 2	-2526.401159	1.941	0.871	10.74	0.001047798	0.036672919	2.228
Gene16	<i>rpl2</i>	local 1	-4049.002087	1.614					
		local 2	-4041.837353	2.268	1.013	14.33	0.000153444	0.005370541	2.240
Gene17	<i>rpl33a</i>	local 1	-2378.39913	1.797					
		local 2	-2377.279869	2.623	1.773	2.24	0.13460965	4.711337743	1.480 *
Gene18	<i>rpl39</i>	local 1	-807.209281	1.376					
		local 2	-807.208368	1.683	1.718	0.00	0.965915393	33.80703875	0.979 *
Gene19	<i>rpl3</i>	local 1	-6541.419085	1.421					
		local 2	-6536.76581	1.781	1.168	9.31	0.002283361	0.079917618	1.525 *

Gene20	<i>rpl4B</i>	local 1	-6888.118878	1.312					
		local 2	-6873.79961	1.804	0.737	28.64	8.72289E-08	3.05301E-06	2.447
Gene21	<i>rpl5</i>	local 1	-5970.453288	0.968					
		local 2	-5969.401307	1.047	0.832	2.10	0.14691807	5.142132456	1.258 *
Gene22	<i>rpl6</i>	local 1	-5566.162622	0.954					
		local 2	-5560.65724	1.184	0.647	11.01	0.000905843	0.031704506	1.830
Gene23	<i>rpl7-A</i>	local 1	-5851.74447	1.341					
		local 2	-5850.902421	1.466	1.180	1.68	0.19438092	6.803332196	1.242 *
Gene24	<i>rpp0</i>	local 1	-6074.223047	1.704					
		local 2	-6068.446849	2.264	1.320	11.55	0.00067662	0.023681702	1.715
Gene25	<i>rps10</i>	local 1	-2730.646022	2.741					
		local 2	-2716.763742	5.089	1.216	27.76	1.37013E-07	4.79546E-06	4.186
Gene26	<i>rps13a</i>	local 1	-1969.953798	1.779					
		local 2	-1958.988861	3.041	0.404	21.93	2.82797E-06	0.000098979	7.531
Gene27	<i>rps16</i>	local 1	-2109.809843	1.190					
		local 2	-2099.769786	1.631	0.430	20.08	7.42649E-06	0.000259927	3.791
Gene28	<i>rps18</i>	local 1	-2145.708034	1.854					
		local 2	-2141.20657	2.377	1.045	9.00	0.002695474	0.094341596	2.275 *
Gene29	<i>rps1</i>	local 1	-4504.350374	1.562					
		local 2	-4492.168655	2.206	0.828	24.36	7.97683E-07	2.79189E-05	2.665
Gene30	<i>rps27</i>	local 1	-992.823738	1.657					
		local 2	-984.86231	2.523	0.436	15.92	6.59772E-05	0.002309201	5.787
Gene31	<i>sadhchydrolase-E1</i>	local 1	-7858.767207	1.522					
		local 2	-7848.730973	1.958	1.065	20.07	7.45624E-06	0.000260968	1.838
Gene32	<i>sap40</i>	local 1	-4057.689683	1.395					
		local 2	-4042.971676	1.915	0.611	29.44	5.77936E-08	2.02278E-06	3.133
Gene33	<i>srs</i>	local 1	-9841.590984	1.116					
		local 2	-9828.361653	1.434	0.777	26.46	2.69238E-07	9.42335E-06	1.845
Gene34	<i>suca</i>	local 1	-5627.561714	0.996					
		local 2	-5624.462152	1.144	0.751	6.20	0.012781357	0.447347493	1.524 *
Gene35	<i>u2snrnp</i>	local 1	-1968.049648	2.302					
		local 2	-1963.463305	2.737	0.966	9.17	0.002456538	0.085978821	2.833 *

* non significant

SYNTHESE :
CONCLUSIONS ET PERSPECTIVES

1. La position des Tuniciers dans l'arbre de la vie

A l'époque où la hiérarchisation de la diversité biologique se basait sur la ressemblance morphologique entre espèces, les Tuniciers (Ascidies, Thaliacés et Appendiculaires) ont été classifiés dans l'embranchement des Mollusques au sein du règne animal [(Lamarck 1809, Cuvier 1828) cité dans (Grassé 1948)]. Il a fallu attendre le début du XX^{ème} siècle pour que la systématique traditionnelle, à la lueur des travaux d'embryologie comparée de Kowalevsky (1866), Willey (1893) et Conklin (1905) [cités dans (Katz 1983)], et la théorie de l'origine néoténique des Vertébrés proposée par Garstang (1928) (Garstang 1928), établissent définitivement les Tuniciers comme membres des animaux deutérostomiens dans le phylum Chordata en tant que *Protochordata*, les Vertébrés avec les Céphalochordés constituant les *Euchordata*.

Près d'un siècle plus tard, les premières études de phylogénie moléculaire tendent à renforcer cette vision (Turbeville, Schulz & Raff 1994; Wada & Satoh 1994). Les phylogénies inférées à partir des marqueurs ribosomiques d'ARNr 18S et 28S retrouvent en effet une proche parenté entre Vertébrés et Céphalochordés, les Tuniciers apparaissant comme la lignée basale de l'embranchement des Chordés (Cameron, Garey & Swalla 2000; Swalla *et al* 2000; Winchell *et al* 2002; Mallatt & Winchell 2007). Cette hypothèse soutenue à la fois par la systématique traditionnelle et la phylogénie moléculaire figure dans tous les traités de zoologie et manuels d'enseignement.

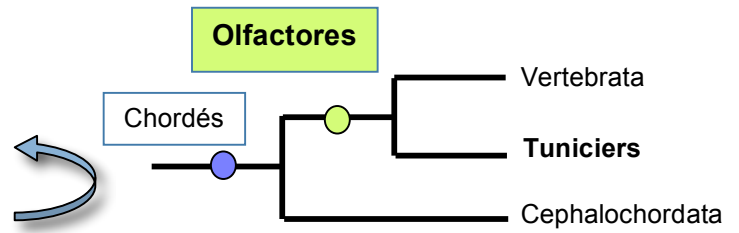
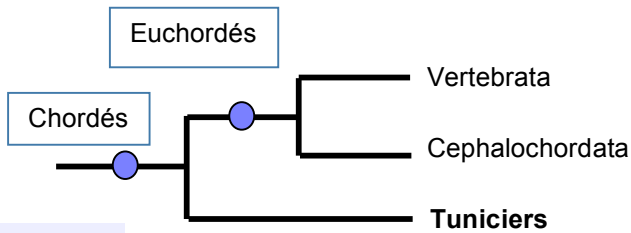
L'émergence de la phylogénomique au milieu des années 2000 (Delsuc, Brinkmann & Philippe 2005) réfute cependant ce schéma traditionnel (Delsuc *et al* 2006). En dépit du dogme des Euchordés, les analyses des données génomiques à large échelle fournissent des preuves tangibles pour l'inversion des positions respectives des Tuniciers et des Céphalochordés au sein des Chordés en favorisant l'hypothèse *Olfactores* qui proposent les Tuniciers et les Vertébrés comme groupe-frères (Bourlat *et al* 2006; Delsuc *et al* 2006; Delsuc *et al* 2008; Dunn *et al* 2008; Putnam *et al* 2008).

Entre temps, les études mitogénomiques apparaissent marginales en proposant un scénario phylogénétique éloignés des deux précédents, d'après lequel les Tuniciers se situent à la base des tous animaux bilatériens (Bourlat *et al* 2006; Bourlat *et al* 2008). Quant aux données paléontologiques, leur apport à la position phylogénétique du groupe a été limitée en raison de la rareté des fossiles, leur mauvaise préservation, et l'ambiguïté quant à leur

SYSTEMATIQUE TRADITIONNELLE &

PHYLOGENOMIQUE

MARQUEURS RIBOSOMIQUES



LES CARACTERES MORPHOLOGIQUES,
ANATOMIQUES ET EMBRYOLOGIQUES

LES CARACTERES MOLECULAIRES

PHYLOGENOMIQUE

LE CLADE OLFACTORES :
SUPPORT ADDITIONNEL

176 gènes nucléaires &
18S + 28S ARNr

(F Delsuc, G Tsagkogeorga
N Lartillot et H Philippe, 2006)

146 gènes nucléaires
Olfactores hypothesis
(Delsuc et al 2006)

170 gènes nucléaires et
mitochondriaux
Monophylie des *Chordata*
Xenambulacraria hypothesis
(Bourlat et al 2006)

13 gènes mitochondriaux
(TR Singh*, G Tsagkogeorga
F Delsuc, S Blanquart, N
Shenkar, Y Loya, EJP.
Douzery et D Huchon, *soumis*)

18S + 28S ARNr

Ambulacraria hypothesis, *Euchordata* hypothesis
(Cameron et al 2000, Wintchell 2002,
Mallat et al 2006)

19^{ème} - 20^{ème}
siècles

1924

1994 2000

2002

2006

2008

Génome miniature
d'*Oikopleura dioica*

(Seo et al 2001)

Génomes des

Ciona intestinalis et *Ciona savignyi*
(Dehal et al 2002, Small et al 2007)

Génome de

Strongylocentrotus purpuratus

(Sea Urchin Genome

Sequencing Consortium 2006)

Projet EST de l'hémichordé

Saccoglossus kowalewskii
(Freeman et al 2008)

Génome d'amphioxus,

Branchiostoma floridae
(Putnam et al 2008)

LES DONNEES GENOMIQUES

Figure 12 : Les différentes hypothèses phylogénétiques proposées pour l'évolution des Chordés et points de repère temporels dans l'étude systématique du groupe.

interprétation, à l'exemple des fossiles *Cheungkongella*, *Vetulicola*, et des Carpoïdes (Jefferies 1986; Shu *et al* 2001a; Shu *et al* 2001b; Chen *et al* 2003; Dominguez & Jefferies 2003) (cf. Introduction, Chapitre I§2).

En récapitulant ainsi le parcours gravi depuis le début du XIX^{ème} siècle jusqu'aux alentours de 2006, l'étude systématique (traditionnelle et moléculaire) des Chordés a proposé deux fronts opposés quant à la position des Tuniciers dans l'arbre de la vie : l'hypothèse des *Euchordata* contre l'hypothèse des *Olfactores* (Fig. 12). Présentement, la communauté scientifique converge vers un consensus selon lequel les Tuniciers sont les plus proches parents actuels des Crâniates (Cyclostomes + Vertébrés), les Céphalochordés représentant la lignée basale au sein des Chordés. L'acceptation de l'hypothèse *Olfactores* s'est progressivement imposée grâce notamment à deux principaux facteurs :

i. Le surcroît des données génomiques.

Allant du séquençage du génome de l'oursin *Strongylocentrotus purpuratus* (Sodergren *et al* 2006), jusqu'au point culminant du génome de l'amphioxus *Branchiostoma floridae* (Putnam *et al* 2008), la disponibilité de données génomiques supplémentaires autorise désormais des analyses plus étendues en termes de nombre ou encore de type de caractères (ex. séquences codants, disposition d'introns, liaisons géniques) utilisés pour la reconstruction phylogénétique (Delsuc *et al* 2008; Dunn *et al* 2008; Putnam *et al* 2008).

ii. Le raffinement des méthodes d'inférence phylogénétique.

Un des notables écueils de la phylogénomique à ses débuts fut le traitement des données multigéniques de façon quasi-homogène. Pour remédier à ce problème, des modèles d'évolution des séquences plus élaborés ont été développés au cours de ces dernières années, permettant d'atténuer les erreurs systématiques issues de la non-prise en compte de l'hétérogénéité des processus évolutifs entre molécules, taxons et sites (Lartillot & Philippe 2004; Blanquart & Lartillot 2006; Lartillot, Brinkmann & Philippe 2007; Blanquart & Lartillot 2008).

L'utilisation de ces nouveaux modèles d'évolution joua un rôle majeur dans l'abolition des conflits suscités par la confrontation des résultats la phylogénie moléculaire conventionnelle, la phylogénomique et la mitogénomique. Ceci est clairement illustré par la forte dépendance du patron phylogénétique des Chordés inféré vis-à-vis de l'approche méthodologique utilisée (Mallatt & Winchell 2007; Delsuc *et al* 2008; Bourlat *et al* 2009).

Par exemple, dans le cas du premier jeu de données phylogénomique, l'analyse phylogénétique sous des modèles standards aboutit au regroupement de l'amphioxus avec l'oursin conduisant à paraphylie des Chordés (Delsuc *et al* 2006). Ce résultat inattendu généra un certain scepticisme sur la validité du groupe des Olfactores qui constitue par ailleurs une vision révolutionnaire à l'époque. La ré-analyse de ce jeu de données sous le modèle CAT (Lartillot & Philippe 2004), qui accommode l'hétérogénéité site-spécifique, a permis de retrouver la monophylie des Chordés (Delsuc *et al* 2008). Dans la même étude (Delsuc *et al* 2008), l'analyse du jeu des données des ARNr 18S et 28S avec un codage RY révéla le rôle joué par les biais de composition nucléotidiques dans ces données qui positionnement généralement les Tuniciers à l'extérieur des Chordés favorisant ainsi le groupement des Vertébrés avec les Céphalochordés (Cameron, Garey & Swalla 2000; Swalla *et al* 2000; Winchell *et al* 2002; Mallatt & Winchell 2007).

L'exemple le plus illustratif de l'importance du choix du modèle pour l'interprétation de manière plus réaliste du signal phylogénétique se tire sans doute de la mitogénomique. En effet, l'étude phylogénétique des Tuniciers à partir des données mitochondriales fut très équivoque. Les jeux de données mitogénomiques, malgré une relativement bonne représentation taxonomique, se sont heurtés à des biais systématiques débouchant sur le positionnement artéfactuel des Tuniciers à la base des Bilatériens (Telford 2006; Bourlat *et al* 2008; Bourlat *et al* 2009). En appliquant des modèles CAT et CAT-BP (Lartillot & Philippe 2004; Blanquart & Lartillot 2008) sur un jeu de données mitogénomiques enrichi par les données génomiques obtenues pour *Herdmania momus*, et prenant ainsi en compte les variations majeures affectant la composition en acides aminés à la fois le long des séquences et au cours du temps, un soutien statistique solide pour la pertinence de l'hypothèse Olfactores a été obtenu pour la première fois (Singh *et al* En révision).

En résumé, les travaux réalisés durant la période 2006-2009, ayant bénéficiés de l'accès aux données génomiques et des avancées méthodologiques de la phylogénomique, ont atténué les précédents conflits phylogénétiques aboutissant à une vision non-ambiguë de la position des Tuniciers dans l'arbre de la vie en tant que groupe-frère des Crâniates.

2. Réinterprétation de l'histoire évolutive des Chordés : Disparités morphologique, génétique, et génomique des Olfactores

« *My sister is a sea squirt?* »

Peter W. Holland, 2006.

L'étroite affinité évolutive des Tuniciers avec les Vertébrés, dévoilée par la phylogénomique, a révélé le groupe des Olfactores comme un clade où existent d'importantes disparités entre deux taxons frères. L'inversion des positions des Tuniciers et Céphalochordés dans l'arbre de la vie requiert en effet la réinterprétation des origines des Chordés de par la repolarisation des pertes/acquisitions de traits au cours de l'évolution du phylum (Cañestro, Yokoi & Postlethwait 2007). Ceci met notamment en évidence de nombreuses différences morphologiques - anatomiques, génétiques ainsi que génomiques entre ces deux lignées.

I. Morphologie - anatomie

La nouvelle phylogénie des Chordés souligne tout d'abord que les traits morpho-anatomiques considérés auparavant comme innovations partagées par les Euchordés, ont été vraisemblablement présents chez le plus proche ancêtre commun des Chordés LCAC (« last common ancestor of chordates »). Dans ce cadre, les Tuniciers semblent avoir subi des régressions secondaires massives sur le plan morphologique (Delsuc *et al* 2006; Holland 2006). En particulier, le LCAC pourrait avoir été doté d'un ensemble des caractères morphologiques qui sont absents chez les Tuniciers. Parmi les plus connus, figurent le mésoderme segmenté en somites le long de l'axe antéro-postérieur de l'embryon de part et d'autre de la chorde et du tube neural, les myomères, le diverticule intestinal, ainsi que les fibres musculaires rouges (Ruppert 2005; Stach 2008).

Au même titre, l'hypothèse de Garstang (Garstang 1928) qui prévoyait l'émergence des Vertébrés à partir d'un ancêtre sessile suite à une série d'évènements de néoténie étant réfutée par le nouveau statut phylogénétique des Tuniciers, les Chordés tirent vraisemblablement leurs origines d'un ancêtre planctonique de type hémichordé enteropneuste (Putnam *et al* 2008).. En étant probablement plus proche de l'amphioxus morphologiquement, le LCAC possédait sans doute un mode de vie nageur (Putnam *et al* 2008; Swalla & Xavier-Neto 2008). À l'inverse, l'apparition du mode de vie fixé chez les

Ascidies ne représente pas un caractère ancestral, au contraire il s'agit d'un caractère dérivé qui est apparu plus tard au cours de l'histoire évolutive des Tuniciers.

II. Génétique du développement embryonnaire

Le développement embryonnaire des Tuniciers est remarquable à bien des égards, et contraste fortement avec celui des autres Chordés. Les Tuniciers montrent (i) une segmentation holoblastique bilatérale, mode de développement qui ne se rencontre guère en dehors de ce groupe (Gilbert 2006); (ii) une différenciation cellulaire autonome, le sort de chaque cellule étant défini sans ambiguïté par les déterminants cytoplasmiques qui lui ont été attribués (Lemaire 2009); (iii) la spécification des blastomères est autonome au niveau de l'endoderme et du tube digestif, du mésoderme musculaire, et de l'endoderme cutané (Gilbert 2006).

Les échanges importants entre la génétique et l'embryologie qui ont eu lieu durant cette dernière décennie procurent aujourd'hui des preuves supplémentaires pour la divergence du programme embryonnaire des Tuniciers (Seo *et al* 2004; Passamaneck & Di Gregorio 2005; Imai *et al* 2006; Satoh *et al* 2006; Sierro *et al* 2006; Cañestro & Postlethwait 2007; Cañestro, Yokoi & Postlethwait 2007; Putnam *et al* 2008). À la lumière de la nouvelle phylogénie, cette divergence se traduit en tant que voies ré-innovatrices conduisant à la formation du plan typique des Chordés. Les Tuniciers exhibent en effet de modalités alternatives qui régissent le développement des différents tissus de la larve de type « têtard » (Lemaire 2006; Lemaire 2009).

En ce qui concerne par exemple le répertoire de gènes Hox (cf. Introduction§7.1) chez les représentants de trois lignées des Chordés, le séquençage du génome de l'amphioxus permet de confirmer que les Céphalochordés préservent le « cluster » des gènes Hox typique des deutérostomiens intact et ordonné, en contenant 15 gènes décrits à ce jour (Putnam *et al* 2008). Ce cluster est en partie scindé chez *Ciona* (Ikuta *et al* 2004), et complètement désintégré chez *Oikopleura* (Seo *et al* 2004; Edvardsen *et al* 2005). La même image est fournie par l'ensemble des gènes étant pourvus de la « boîte » homéotique provenant de duplications anciennes. Le nombre de gènes homéotiques annotés chez l'homme et l'amphioxus atteint présentement 235 et 133 respectivement, alors que le génome de *Ciona intestinalis* n'en porte que 83 (Wada *et al* 2003; Putnam *et al* 2008).

Certains écarts relatifs à la génétique du développement chez les Tuniciers peuvent même se retracer jusqu'aux réseaux régulateurs des gènes du développement embryonnaire. Ces réseaux sont définis en tant que diagrammes ou circuits d'interconnexions fonctionnels entre facteurs de transcription, composants de la signalisation cellulaire et modules régulateurs *cis*-éléments (Lemaire 2009). Les réseaux régulateurs contrôlent l'expression précise des gènes durant la différenciation, conférant leurs différents attributs à chaque cellule embryonnaire. Ils sont ainsi sous-jacents aux processus morphogénétiques complexes et fondamentaux du développement des Chordés à la manière du fameux exemple du centre de signalisation connu comme l'organisateur de Spemann «Spemann organizer» des amphibiens (Gilbert 2006).

Des réseaux provisoires récemment cartographiés chez *Ciona intestinalis* ont permis de mettre en évidence qu'en fort contraste avec les Vertébrés le contrôle de transcription d'une grande portion des gènes du développement est simplifié, n'étant régi que par un petit nombre des molécules signal (Fgf, Nodal) et des facteurs de transcriptions (ZicL, FoxD, FoxA-a, Otx) (Imai *et al* 2006). En effet, 53 facteurs de transcription zygotiques paraissent suffisants pour assigner la diversité cellulaire observée à l'état de gastrula. De plus, le développement de *Ciona intestinalis* semble majoritairement régulé par des boucles d'autorégulation négatives contre très peu de positives (Lemaire 2006; Lemaire 2009).

III. Génomique

Constituant les premières ébauches génomiques des Chordés en dehors des Vertébrés, les projets de séquençage des espèces *Ciona intestinalis* (Dehal *et al* 2002), *Ciona savignyi* (Small *et al* 2007) et *Oikopleura dioica* (Seo *et al* 2001) ont avant tout révélé que les Tuniciers disposent de génomes très compacts dépourvus des duplications génomiques spécifiques aux Vertébrés (Dehal & Boore 2005). Le séquençage du génome complet de l'amphioxus *Branchiostoma floridae* (Putnam *et al* 2008) vient encore de renforcer l'image de disparité génomique des Tuniciers. Il autorise en effet la comparaison des caractères structuraux et fonctionnels des génomes des trois lignées de Chordés.

Du point de vue structural, le séquençage de *Branchiostoma floridae* a révélé un génome de Céphalochordés d'une taille d'environ 520 megabases organisé en 19 paires chromosomiques (Putnam *et al* 2008). Les génomes de *Ciona intestinalis* et *Ciona savignyi* montrent en revanche une répartition spatiale en 14 chromosomes et avec une taille estimée à

150 – 160 megabases, ces génomes sont environ 20 fois moins gros que le génome humain (~3,2 Gb) (Dehal *et al* 2002; Satoh *et al* 2006; Ranwez *et al* 2007). Le génome de *Ciona intestinalis* possède en particulier une densité génique d'environ un gène par 7,5 kb, une valeur très proche de chez la Drosophile (9kb), mais beaucoup plus élevée que chez l'homme (100kb) (Dehal *et al* 2002). Pour sa part, le génome d'*Oikopleura dioica* s'annonce d'une taille inférieure à 70 Mb ce qui représente jusqu'à présent le plus petit génome séquencé à l'échelle des animaux bilatériens (Seo *et al* 2001).

En ce qui concerne le contenu en gènes, l'image fournie par l'analyse comparée de génomes des Chordés apparaît également très contrastée, avec notamment la prédiction de 16000 locus codants pour des protéines chez *Ciona intestinalis*, contre 22000 annotés chez l'amphioxus, l'homme en comportant environ 30000. Plus précisément, les gènes considérés comme étant Chordés-spécifiques chez *Ciona* constituent seulement un sixième du contenu génique (2750) alors que plus de 60% des séquences codantes possèdent au moins un homologue chez la drosophile. Encore plus intéressant fut la mise en évidence de 3400 gènes « orphelins » qui n'ont pas d'orthologues détectés ni chez les Protostomiens ni chez les Deutérostomiens, et qui sont donc considérés comme des innovations spécifiques aux Tuniciers (Dehal *et al* 2002).

Très loin de cette image et notamment plus proche des Vertébrés, le génome de l'amphioxus ne comporte que 137 familles protéiques qui ne montre aucune similitude significative avec les autres Chordés séquencés, 8400 groupes de gènes sont effectivement partagés avec l'homme, ce qui réduit à environ 13000 gènes potentiellement déjà présents dans le génome ancestral des Chordés (Putnam *et al* 2008). La macro-synténie étant très conservée entre les deux lignées, Putnam *et al* (2008) ont inféré que le génome du LCAL était vraisemblablement pourvu de 17 proto-chromosomes (CLGs, « chordate linkage groups »). Selon cette estimation, la génomique comparée suggère donc la perte d'environ 2250 gènes chez *Ciona intestinalis*, dont le génome semble avoir été l'objet d'intenses remaniements chromosomiques.

L'amphioxus partage de plus une large fraction de ses régions introniques avec les Vertébrés à la fois en terme de position et de phase, 85% de ses séquences d'introns peuvent en effet être alignées avec celles de l'homme. La taille et la distribution des exons/introns chez *Ciona intestinalis* est beaucoup plus variable, néanmoins les analyses phylogénétiques basées sur la structure en exons/introns confirment le groupement des Tuniciers avec les Vertébrés en tant que groupe-frères au sein des Chordés (Putnam *et al* 2008).

3. Phylogénie moléculaire des Tuniciers

Au même titre que les Chordés, l'histoire évolutive des Tuniciers commença à être étudiée sur le plan moléculaire dans les années 1990 sur la seule base de l'ARNr 18S (Wada & Satoh 1994). Étant un marqueur phylogénétique au pouvoir résolutif reconnu, l'ARNr 18S a fourni un cadre général relativement bien défini pour l'évolution des lignées majeures de Tuniciers qui contraste fortement la vision systématique traditionnelle (Swalla *et al* 2000; Yokobori *et al* 2006; Zeng, Jacobs & Swalla 2006). Les Ascidies (Aplousobranchia, Phlebobranchia, Stolidobranchia), considérées jadis une Classe d'origine commune, se révélèrent être un groupe polyphylétique, l'histoire évolutive des Tuniciers distinguant en fait trois grands clades : (i) Phlebobranchia + Aplousobranchia + Thaliacea, (ii) Appendicularia, et (iii) Stolidobranchia (Tsagkogeorga *et al* 2009).

La phylogénie des Tuniciers reste cependant incertaine quant à la position de deux lignées majeures : les Appendicularia et les Aplousobranchia. La position des Appendiculaires constitue sans doute le point le plus critique puisqu'ils revêtent un intérêt primordial à plus d'un titre. Premièrement, en étant les seuls Tuniciers retenant une queue mobile tout au long de leur cycle de vie, ils occupent un rôle central pour la compréhension de l'émergence du plan d'organisation des Chordés (Garstang 1928; Wada 1998; Swalla *et al* 2000; Stach & Turbeville 2002; Putnam *et al* 2008). Du point de vue génomique, ils possèdent des génomes excessivement condensés et remaniés (Seo *et al* 2001; Seo *et al* 2004; Edvardsen *et al* 2005) et figurent du point de vue écologique au deuxième rang des animaux les plus abondants dans le plancton des mers du globe (Robison, Reisenbichler & Sherlock 2005).

Les études phylogénomiques menées jusqu'à ce jour placent l'appendiculaire *Oikopleura dioica* à la base des autres Tuniciers (Bourlat *et al* 2006; Delsuc *et al* 2006; Delsuc *et al* 2008; Dunn *et al* 2008; Putnam *et al* 2008). De manière conflictuelle, les analyses les plus récentes de l'ARNr 18S suggèrent une position plus interne des Appendiculaires au sein des Tuniciers en tant que groupe-frère des Stolidobranches (Zeng & Swalla 2005; Zeng, Jacobs & Swalla 2006; Tsagkogeorga *et al* 2009). Le positionnement fiable des Appendiculaires se heurte cependant à deux principaux facteurs : (i) un trop faible échantillonnage taxonomique restreint aux Oikopleuridae, et (ii) leur taux d'évolution extrêmement rapide matérialisée par les longues branches observées dans les arbres phylogénétiques obtenus (Tsagkogeorga *et al* 2009).

L'autre incertitude majeure concerne le statut de l'Ordre des Aplousobranchia. Définis à l'origine par la constitution particulière de leur branchie (cf. Introduction§3), les espèces de Tuniciers appartenant à cet ordre présentent un type d'organisation exclusivement colonial, les individus étant constitués de minuscules zooïdes clonaux (Lahille 1886; Lahille 1887; Lahille 1890). Les Aplousobranchia représentent à l'heure actuelle l'un des groupes cibles en pharmacologie au vu des nombreux composés chimiques qui ont été caractérisés chez des représentants du groupe (ex. la didemnine B chez *Trididemnum solidium*, l'aplidine chez *Aplidium albicans*, cf. Introduction§6 tableau 3) (Rinehart 2000; Blunt *et al* 2004; Blunt *et al* 2006).

D'après le dernier consensus issu de la systématique traditionnelle, les Aplousobranchia peuvent se diviser en cinq familles, *sensu* Kott : les Clavelinidae, Didemnidae, Polycitoridae, Polyclinidae et Holozoidae (Kott 2005). Les Didemnidae constitue actuellement la famille la plus connue du grand public, en raison de l'introduction involontaire des espèces invasives du genre *Didemnum* aux Etats-Unis, au Canada, aux Pays-Bas et en France (Gittenberger 2009; Locke & Carman 2009).

Le statut phylogénétique des Aplousobranchia a été problématique dès les premières études moléculaires puisque les premières séquences d'ARNr 18S obtenues pour le groupe (Stach & Turbeville 2002) ont été invalidées par la suite en tant que contaminations par de protistes symbiotiques (Stach & Turbeville 2002; Yokobori *et al* 2006). La première incorporation du groupe dans un contexte phylogénétique est une analyse portant sur le gène mitochondrial *cox1* dans laquelle une proche affinité des Aplousobranchia avec les Cionidae (Phlebobranchia) a été proposée (Turon & Lopez-Legentil 2004). Grâce aux travaux de Yokobori *et al* (2006) des séquences d'ARNr 18S authentiques ont été finalement obtenues pour des espèces représentant les familles Didemnidae, Polycitoridae et Polyclinidae. Contre toute attente, les Aplousobranchia échantillonnés montrèrent une énorme divergence moléculaire, illustrée par des longueurs de branches extrêmes dans les arbres phylogénétiques (Yokobori *et al* 2006).

La première véritable ébauche obtenue à partir des données de l'ARNr 18S favorisa l'hypothèse d'une étroite affinité du groupe avec les Thaliacés (Yokobori *et al* 2006). Cependant, l'acquisition de nouvelles séquences d'ARNr 18S pour la famille des Clavelinidae conduit actuellement à la dissociation des deux groupes (Tsagkogeorga *et al* 2009). La proche affinité des Aplousobranchia avec les Thaliacés semble ainsi être imputable à un phénomène d'attraction des longues branches, les analyses phylogénétiques sous des modèles d'évolution

des séquences les plus sophistiqués, soutiennent la monophylie des Aplousobranchia (Tsagkogeorga *et al* 2009). Le positionnement exact du groupe dans le clade I avec les Phlebobranches et les Thaliacés reste néanmoins irrésolu.

4. Divergence moléculaire et évolution accélérée des Tuniciers

« ...although tunicates are now generally agreed to be the sister group of vertebrates, their exceptionally rapid evolution has effaced much information that might have been useful for suggesting how the vertebrates evolved from chordate ancestors. Instead, they are excellent for understanding what evolution can do. »

Linda Z. Holland, 2007

Dans le but d'éclaircir l'histoire évolutive des Chordés, et notamment des Tuniciers, la génomique comparée et la phylogénétique ont fourni au cours de ces dernières années une somme d'évidences pour une divergence génomique surprenante des Tuniciers et leur évolution accélérée (Seo *et al*, 2001; Dehal *et al*, 2002; Patel, 2004; Bourlat *et al*, 2006; Delsuc *et al*, 2006; Imai *et al*, 2006; Holland *et al*, 2008; Putnam *et al*, 2008; Tsagkogeorga *et al*, 2009).

Sur le plan génomique, Hill *et al.* (2008) ont réalisé une analyse comparée des cartes chromosomiques de *Ciona savignyi* et *Ciona intestinalis* dans laquelle ils démontrèrent que malgré l'appartenance des deux espèces au même genre, la colinéarité spatiale des gènes est rompue par l'occurrence de remaniements intra-chromosomiques intensifs débouchant sur des patrons d'ordre de gènes notablement différents de ceux des Vertébrés (Hill *et al* 2008). Au même titre, les analyses des génomes mitochondriaux ont révélé une extrême plasticité de l'ordre de gènes au sein des Tuniciers (Gissi, Iannelli & Pesole 2004; Yokobori, Oshima & Wada 2005; Iannelli *et al* 2007a; Gissi, Iannelli & Pesole 2008), une observation renforcée ici par l'étude du génome d'*Herdmania momus* (cf. Chapitre III§3) (Singh *et al* En révision). Du point de vue phylogénétique, des preuves indubitables sont actuellement accumulées pour le taux d'évolution rapide des Tuniciers, comme illustré par les longueurs de branches du groupe dans l'ensemble des arbres reconstruits, et ce, quels que soient les marqueurs utilisés (*i.e.* ribosomiques, nucléaires et mitochondriaux).

Les analyses de l'exploration du contenu en gènes du transcriptome partiel de *Microcosmus squamiger* qui a été pyroséquencé au cours des présents travaux portent

également l’empreinte de l’évolution différentielle du groupe, marquée par un taux d’évolution moléculaire accéléré (cf. Chapitre IV).

Considérant la similitude entre séquences nucléotidiques, les recherches d’homologie par Blast ont démontré que seul $\frac{1}{4}$ des séquences de *Microcosmus* (50000 sur 210000) présente une similitude significative contre l’intersection des transcrits orthologues entre *Ciona intestinalis* et *Ciona savignyi*. En termes de nombre de gènes, celles-ci ne correspondent qu’à une fraction de 200 gènes parmi les 9520 gènes communs aux deux ciones (cf. Chapitre IV§4). Malgré le biais initial introduit par le fait que les ADNc séquencés ne couvraient pas l’intégralité du transcriptome de *Microcosmus sqamiger*, le chiffre de 200 gènes même au niveau intra-Tuniciers apparaît très faible.

Par ailleurs, une étude portant sur l’identification de 20000 ESTs issues des premiers stades du développement de *Molgula tectiformis* a démontré la surexpression de gènes protéiques qui ne possèdent pas d’homologues connus chez *Ciona* ou aucun autre organisme (Gyoja *et al* 2007). Une image similaire a été récemment obtenue pour l’analyse du contenu génique du transcriptome d’*Halocynthia roretzi*, dont le pourcentage de gènes orphelins atteint les 11% (75 ESTs parmi les 687 identifiées) (Kim *et al* 2008). De même, Small *et al*, en étudiant la variabilité génomique de *Ciona savignyi*, ont démontré un taux extrême de celle-ci par rapport à *Ciona intestinalis* sur les plans des inversions (1,96%), des insertions/délétions (16,6%), ainsi que de l’hétérozygotie SNP (4,5%) (Small *et al* 2007). Les deux ciones, malgré le fait qu’elles puissent être hybridées en laboratoire, présentent une divergence moléculaire comparable à celle des génomes de l’homme et du poulet (Small *et al* 2007). En particulier, en utilisant un seuil de conservation de longueur ≥ 100 bp et d’identité $\geq 70\%$, *Ciona intestinalis* et *Ciona savignyi* ne partagent que 1835 éléments conservés, ce qui ne correspond qu’à une fraction de 0,7% du génome complet de *Ciona intestinalis* (Kim *et al* 2008). Enfin, la découverte d’un ancien épisode de spéciation allopatrique au sein de l’espèce *Ciona intestinalis* est venue confirmer l’occurrence de divergences cryptiques chez les Tuniciers en révélant l’existence de deux populations génétiquement très distinctes (Caputi *et al* 2007; Iannelli *et al* 2007a).

Malgré tout, en ciblant à des niveaux taxonomiques supérieurs, *i.e.* Deutérostomes et Métazoaires, nous avons pu néanmoins identifier un noyau des gènes conservés et surexprimés correspondant à des « gènes de ménage » dans les données de séquençage 454 de *Microcosmus*. Ceci nous a permis de déterminer *in silico* 35 marqueurs moléculaires orthologues à l’échelle des animaux en nous autorisant par la suite une première tentative

d'appréhension de la vitesse d'évolution accélérée des Tuniciers vis-à-vis des autres Chordés (cf. Chapitre IV§4) (Tsagkogeorga *et al* En préparation). L'étude de ce lot de gènes fournit une quantification préliminaire du taux d'évolution génomique du groupe qui est estimé comme étant deux fois plus élevé en moyenne par rapport aux deux autres lignées de Chordés. Les premières analyses suggèrent que le déterminant principal de cette accélération est une hausse du taux de mutation chez les Tuniciers plutôt qu'un changement dans les contraintes sélectives portant sur ces gènes. L'ensemble des gènes examinés porte en effet la signature d'une sélection purificatrice ou négative confirmant les récentes estimations quant au régime de sélection des génomes de *Ciona intestinalis* et *Ciona savignyi* (prévalence de la sélection négative estimée dans > 78% des séquences d'exons) (Small *et al* 2007). L'étendue de cette accélération qui présente un profil hétérogène selon les espèces et les gènes apparaît particulièrement intéressante. En accord avec la phylogénie de l'ARNr 18S, les Appendicularia et Aplousobranchia s'avèrent les plus divergents au sein des Tuniciers, et de manière globale, l'image contrastée du taux d'évolution chez les Chordés semble en grande partie dépendante du degré d'accélération de ce deux groupes (Tsagkogeorga *et al* En préparation).

5. Séquençage à haut débit et Phylogénomique des Tuniciers

De l'ensemble des travaux présentés dans les points précédents de dégage le corollaire que le statut évolutif des Tuniciers est particulièrement difficile à aborder expérimentalement et à décrypter analytiquement. En effet, le signal phylogénétique dans la plupart de marqueurs est difficile à extraire, à la fois à cause d'une forte saturation induite par le taux d'évolution rapide du groupe, et à d'autres facteurs associés à l'hétérogénéité des processus évolutifs le long des séquences ou au cours du temps (Delsuc *et al* 2008; Tsagkogeorga *et al* 2009). Ceci a abouti par le passé à des phylogénies du groupe qui comportaient de nombreuses erreurs systématiques (Cameron, Garey & Swalla 2000; Swalla *et al* 2000; Winchell *et al* 2002; Mallatt & Winchell 2007).

Les projets de séquençage des génomes complets de *Ciona intestinalis* (Dehal *et al* 2002), *Ciona savignyi* (Small *et al* 2007) et *Oikopleura dioica* (Seo *et al* 2001) ont apporté une contribution fondamentale à la caractérisation de l'évolution génomique des Tuniciers et à la reconstruction de l'histoire évolutive des Chordés par la phylogénomique (Bourlat *et al* 2006; Delsuc *et al* 2006; Delsuc *et al* 2008; Dunn *et al* 2008; Putnam *et al* 2008). Par contre,

au niveau intra-tuniciers, l'ARNr 18S, qui est l'unique marqueur phylogénétique dont on dispose pour un nombre conséquent de représentants, apparaît d'un pouvoir résolutif limité pour répondre aux controverses persistantes sur le statut systématique du groupe (Tsagkogeorga *et al* 2009). Ainsi, l'éclaircissement de l'histoire phylogénétique et de la dynamique évolutive des Tuniciers exige impérativement le développement d'un jeu de données phylogénomique plus étendu à la fois en termes de taxons et de gènes.

À l'heure actuelle, nous assistons à une croissance continue des données génomiques grâce au développement récent des méthodes de séquençage de deuxième génération (cf. Chapitre IV§2). En se basant sur le principe de miniaturisation et de parallélisations des réactions de séquences, les techniques innovantes du séquençage à haut débit ouvrent l'accès à l'information génétique à grande échelle, dans des délais de plus en plus raccourcis, et à des coûts amoindris par nucléotide (Shendure & Ji 2008). Ainsi, de la fin de l'année 2005 jusqu'à Octobre 2009, l'approche du pyroséquençage ou séquençage 454 fut à elle seule à l'origine de 537 publications (<http://www.454.com/>), couvrant un large éventail des domaines de recherche, comme la génétique et la génomique, l'évolution et l'écologie, la microbiologie, la métagénomique, la biotechnologie et la biologie des systèmes (Rothberg & Leamon 2008).

À ce titre, l'application du séquençage 454 GS-FLX Standard sur *Microcosmus squamiger* dans les présents travaux nous a permis d'obtenir 35 marqueurs moléculaires orthologues à l'échelle des animaux. Ceci a été réalisé malgré la limitation imposée par la divergence moléculaire importante des Tuniciers, et notamment dans une période où la technique procurait des données de l'ordre de 100Mb par « run » à une longueur de séquence qui ne dépassait pas les 300 pb. En Avril 2009, la compagnie 454 Roche-Life Sciences a développé une nouvelle série de kits « Titanium series reagents » pour le pyroséquençage, portant la capacité de lecture à 400 à 600 pb et garantissant la livraison d'un million de séquences de haut qualité par run.

Au vu de ces avancées et considérant le potentiel de l'approche 454 après nos résultats pilotes obtenus sur *Microcosmus squamiger*, l'utilisation de la technologie 454 pour le séquençage *de novo* de transcriptomes de Tuniciers paraissait être une suite logique aux efforts consentis pour l'étude de l'évolution du groupe. Ceci consiste à l'acquisition *in vitro* de données génomiques pour un nombre important d'espèces avec pour objectif d'obtenir un nombre conséquent de marqueurs orthologues. Un projet a été initié en ciblant neuf espèces représentant les grandes lignées de Tuniciers d'après l'image phylogénétique obtenues à partir des données de l'ARNr 18S (Tsagkogeorga *et al* 2009). Le séquençage 454 GS-FLX

Titanium a été réalisé sur les transcriptomes des ascidies sélectionnées, après le marquage différentiel des ADNc amplifiés à partir des ARNm contenus dans le tissu musculaire et les gonades de chaque espèce, et le pyroséquençage de l'ensemble des transcriptomes en un seul « run » par la compagnie GATC Biotech.

Le rendement du séquençage 454 des neuf nouveaux Tuniciers atteint approximativement les 370 millions de paires de bases pour l'ensemble des transcriptomes, ce qui correspond en moyenne à 130000 séquences par espèce pyroséquencée selon la qualité des ADNc d'origine, la longueur des lectures fluctuant entre 270 et 335 bp (Tableau 5). Un premier aperçu des données génomiques acquises montre que l'intersection des gènes présents chez tous les Tuniciers séquencés atteint les 250 transcrits. Quant à la réunion des neuf transcriptomes, environ 10000 gènes parmi les 13896 transcrits orthologues entre *Ciona intestinalis* et *Ciona savignyi* ont pu être identifiés au moins dans un parmi les neuf représentants du groupe. Enfin, l'analyse *in silico* des données acquises a permis la l'identification de 153 marqueurs phylogénétiques conservés à l'échelle des animaux bilatériens, à partir desquels un cadre phylogénétique robuste a été reconstruit (Figure 1).

Bien que préliminaire, l'ébauche phylogénomique présentée pour les Tuniciers apparaît très prometteuse et servira de base pour la réalisation d'analyses phylogénétiques et d'évolution moléculaire plus détaillées. Afin d'éclaircir les incertitudes qui persistent quant au statut évolutif du groupe, les prochains axes de recherche à aborder incluent entre autres :

- (i) L'inférence des dates de divergence des Tuniciers et de leurs grandes lignées ;
- (ii) L'étude comparative de la dynamique évolutive par gène et espèce, avec notamment une emphase sur les profils compositionnels par gène et taxon, la présence / absence d'hétérotachie, et l'estimation des contraintes sélectives par molécule, taxon et site ;
- (iii) La quantification du taux d'évolution des Tuniciers et du degré de la variabilité de celui-ci au sein des taxons et gènes, dans le but de tester les hypothèses préliminaires expliquant ce taux variable à une plus large échelle (cf. Chapitre IV) ;
- (iv) L'identification des déterminants à l'origine des profils évolutifs hétérogènes des Tuniciers vis-à-vis des Chordés et également en leur sein (*i.e.* évaluation de l'hypothèse d'augmentation du taux de mutation des Tuniciers *versus* changement du régime sélectifs ; cf. Chapitre IV).

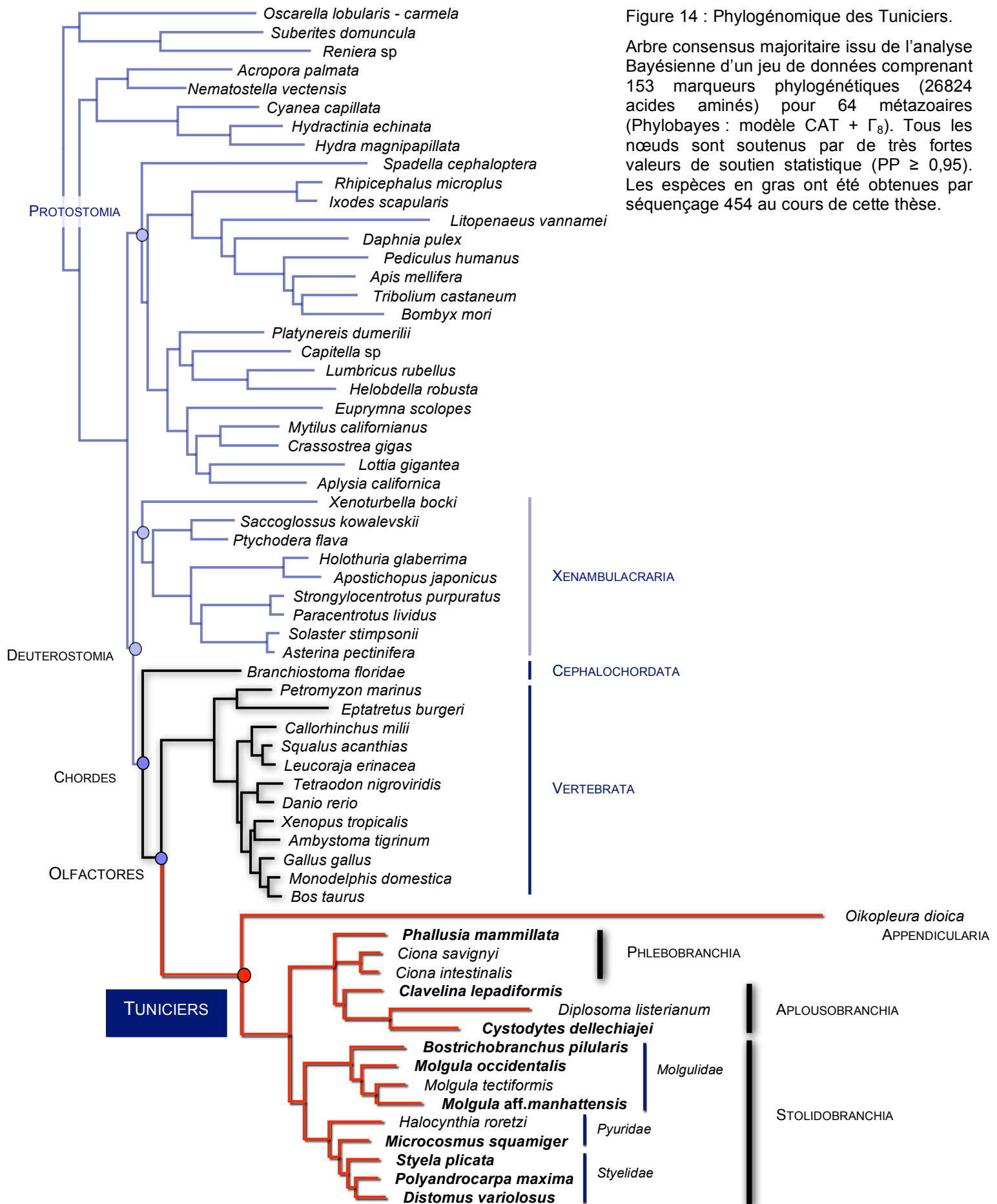


Figure 14 : Phylogénomique des Tuniciers.

Arbre consensus majoritaire issu de l'analyse Bayésienne d'un jeu de données comprenant 153 marqueurs phylogénétiques (26824 acides aminés) pour 64 métazoaires (Phylobayes : modèle CAT + Γ_8). Tous les nœuds sont soutenus par de très fortes valeurs de soutien statistique (PP \geq 0,95). Les espèces en gras ont été obtenues par séquençage 454 au cours de cette thèse.

Tableau 5 : Statistiques et résultats préliminaires du séquençage des neuf transcriptomes partiels des Tunciers

ESPECES	ORDRE (FAMILLE)	#PAIRES DE BASES (BP)	LONGUEUR		%AT	%GC	LECTURES		GENES*	
			#SEQUENCES	MOYENNE (BP)			> 350 BP	HITS*	GENES*	>15 LEC
<i>Clavelina lepadiformis</i> (Müller, 1776)	APLOUSOBRANCHIA (CLAVELINIDAE)	66 867 789	208 758	320	58,33	41,67	103 037	59 275	7 568	645
<i>Cystodytes dellechiaiei</i> (Della Valle, 1877)	APLOUSOBRANCHIA (POLYCITORIDAE)	45 752 032	139 938	326	56,41	43,59	73 580	50 620	3 905	258
<i>Phallusia mammillata</i> (Cuvier, 1815)	PHLEBOBRANCHIA (ASCIDIIDAE)	30 941 072	92 135	335	53,76	46,24	51 531	40 670	2 581	194
<i>Polyandrocarpa maxima</i> (Sluiter, 1904)	PHLEBOBRANCHIA (PEROPHORIDAE)	23 720 074	75 565	313	53,10	46,90	37 280	21 268	1 354	132
<i>Molgula occidentalis</i> (Traustedt, 1883)	STOLIDOBRANCHIA (MOLGULIDAE)	26 026 718	96 039	271	64,27	35,73	31 354	15 265	2 841	165
<i>Bostrichobranchus pilularis</i> (Verrill, 1871)	STOLIDOBRANCHIA (MOLGULIDAE)	44 781 807	146 593	305	60,12	39,88	62 234	36 592	5 636	359
<i>Molgula aff. manhattensis</i> (De Kay, 1843)	STOLIDOBRANCHIA (MOLGULIDAE)	35 990 178	112 856	318	60,34	39,66	55 465	33 018	5 147	343
<i>Styela plicata</i> (Lesueur, 1823)	STOLIDOBRANCHIA (STYELIDAE)	33 530 842	109 227	306	62,46	37,54	48 763	26 982	2 426	130
<i>Distomus variolosus</i> (Gaertner in Pallas, 1774)	STOLIDOBRANCHIA (STYELIDAE)	63 770 991	198 602	321	56,18	43,82	100 659	50 884	6 223	592

*tBlastx contre les transcrits orthologues partagés entre *C. intestinalis* \cap *C. savignyi* (seuil *e*-value = 1e-06)

REFERENCES BIBLIOGRAPHIQUES

- Adoutte, A., G. Balavoine, N. Lartillot, O. Lespinet, B. Prud'homme & R. de Rosa. 2000. The new animal phylogeny: reliability and implications. *Proc Natl Acad Sci U S A* **97**:4453-4456.
- Blanquart, S. & N. Lartillot. 2006. A Bayesian compound stochastic process for modeling nonstationary and nonhomogeneous sequence evolution. *Mol Biol Evol* **23**:2058-2071.
- Blanquart, S. & N. Lartillot. 2008. A site- and time-heterogeneous model of amino acid replacement. *Mol Biol Evol* **25**:842-858.
- Blaxter, M. & J. Thomas. 2004. A survey of genes expressed in the seasquirt *Diplosoma listerianum*. Unpublished.
- Blunt, J. W., B. R. Copp, M. H. Munro, P. T. Northcote & M. R. Prinsep. 2006. Marine natural products. *Nat Prod Rep* **23**:26-78.
- Blunt, J. W., B. R. Copp, M. H. Munro, P. T. Northcote & M. R. Prinsep. 2004. Marine natural products. *Nat Prod Rep* **21**:1-49.
- Boore, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res* **27**:1767-1780.
- Boore, J. L. & W. M. Brown. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Dev* **8**:668-674.
- Boore, J. L., L. L. Daehler & W. M. Brown. 1999. Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (Amphioxus). *Mol Biol Evol* **16**:410-418.
- Bourlat, S. J., T. Juliusdottir, C. J. Lowe, R. Freeman, J. Aronowicz, M. Kirschner, E. S. Lander, M. Thorndyke, H. Nakano, A. B. Kohn, A. Heyland, L. L. Moroz, R. R. Copley & M. J. Telford. 2006. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**:85-88.
- Bourlat, S. J., C. Nielsen, A. D. Economou & M. J. Telford. 2008. Testing the new animal phylogeny: a phylum level molecular analysis of the animal kingdom. *Mol Phylogenet Evol* **49**:23-31.
- Bourlat, S. J., O. Rota-Stabelli, R. Lanfear & M. J. Telford. 2009. The mitochondrial genome structure of *Xenoturbella bocki* (phylum Xenoturbellida) is ancestral within the deuterostomes. *BMC Evol Biol* **9**:107.
- Cameron, C. B., J. R. Garey & B. J. Swalla. 2000. Evolution of the chordate body plan: new insights from phylogenetic analyses of deuterostome phyla. *Proc Natl Acad Sci U S A* **97**:4469-4474.

- Cañestro, C., S. Bassham & J. H. Postlethwait. 2003. Seeing chordate evolution through the *Ciona* genome sequence. *Genome Biol* **4**:208.
- Cañestro, C. & J. H. Postlethwait. 2007. Development of a chordate anterior-posterior axis without classical retinoic acid signaling. *Dev Biol* **305**:522-538.
- Cañestro, C., H. Yokoi & J. H. Postlethwait. 2007. Evolutionary developmental biology and genomics. *Nat Rev Genet* **8**:932-942.
- Caputi, L., N. Andreakis, F. Mastrototaro, P. Cirino, M. Vassillo & P. Sordino. 2007. Cryptic speciation in a model invertebrate chordate. *Proc Natl Acad Sci U S A* **104**:9364-9369.
- Chabry, L. 1887. *Emryologie normale et tératologique des Ascidies*. Faculté des Sciences de Paris, Paris.
- Chen, J. Y., D. Y. Huang, Q. Q. Peng, H. M. Chi, X. Q. Wang & M. Feng. 2003. The first tunicate from the Early Cambrian of South China. *Proc Natl Acad Sci U S A* **100**:8314-8318.
- Corbo, J. C., A. Di Gregorio & M. Levine. 2001. The ascidian as a model organism in developmental and evolutionary biology. *Cell* **106**:535-538.
- Curole, J. P. & T. D. Kocher. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends Ecol Evol* **14**:394-398.
- Dehal, P. & J. L. Boore. 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* **3**:e314.
- Dehal, P., Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Nacula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H. G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K. S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B. I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y.

- Kohara, M., Levine, N., Satoh & D. S. Rokhsar. 2002. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**:2157-2167.
- Delsuc, F., H. Brinkmann, D. Chourrout & H. Philippe. 2006. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**:965-968.
- Delsuc, F., H. Brinkmann & H. Philippe. 2005. Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet* **6**:361-375.
- Delsuc, F., G. Tsagkogeorga, N. Lartillot & H. Philippe. 2008. Additional molecular support for the new chordate phylogeny. *Genesis* **46**:592-604.
- Dominguez, P. & R. Jefferies. 2003. Fossil evidence on the origin of appendicularians. International Urochordate Meeting 2003, Carry le Rouet, France.
- Dunn, C. W., A. Hejnol, D. Q. Matus, K. Pang, W. E. Browne, S. A. Smith, E. Seaver, G. W. Rouse, M. Obst, G. D. Edgecombe, M. V. Sorensen, S. H. Haddock, A. Schmidt-Rhaesa, A. Okusu, R. M. Kristensen, W. C. Wheeler, M. Q. Martindale & G. Giribet. 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**:745-749.
- Edwardsen, R. B., H. C. Seo, M. F. Jensen, A. Mialon, J. Mikhaleva, M. Bjordal, J. Cartry, R. Reinhardt, J. Weissenbach, P. Wincker & D. Chourrout. 2005. Remodelling of the homeobox gene complement in the tunicate *Oikopleura dioica*. *Curr Biol* **15**:R12-13.
- Garstang, W. 1928. The morphology of the Tunicata, and its bearings on the phylogeny of the Chordata. *Quart J Microsc Sci* **72**:51-187.
- Gilbert, S. F. 2006. *Developmental Biology*. Sinauer Associates Inc., Sunderland, Massachusetts USA.
- Gissi, C., F. Griggio & F. Iannelli. 2009. Evolutionary mitogenomics of Chordata: the strange case of ascidians and vertebrates. *Invertebrate Survival Journal* **6**:S21-S28.
- Gissi, C., F. Iannelli & G. Pesole. 2004. Complete mtDNA of *Ciona intestinalis* reveals extensive gene rearrangement and the presence of an atp8 and an extra trnM gene in ascidians. *J Mol Evol* **58**:376-389.
- Gissi, C., F. Iannelli & G. Pesole. 2008. Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity* **101**:301-320.
- Gissi, C., G. Pesole, E. Cattaneo & M. Tartari. 2006. Huntingtin gene evolution in Chordata and its peculiar features in the ascidian *Ciona* genus. *BMC Genomics* **7**:288.
- Gittenberger, A. 2009. Invasive tunicates on Zeeland and Prince Edward Island mussels, and management practices in The Netherlands. *Aquatic Invasions* **4**:279-281.

- Godeaux, J. 1998. The relationships and systematics of the Thaliacea, with keys for identification. Pp. 273-384 in Q. Bone, ed. The biology of pelagic tunicates. Oxford University Press, Oxford.
- Goldstone, J. V., H. M. Goldstone, A. M. Morrison, A. Tarrant, S. E. Kern, B. R. Woodin & J. J. Stegeman. 2007. Cytochrome P450 1 genes in early deuterostomes (tunicates and sea urchins) and vertebrates (chicken and frog): origin and diversification of the CYP1 gene family. *Mol Biol Evol* **24**:2619-2631.
- Grassé, P. 1948. *Traité de Zoologie - Anatomie, Systématique, Biologie* Tomes XI. Masson.
- Gyoja, F., Y. Satou, T. Shin-i, Y. Kohara, B. J. Swalla & N. Satoh. 2007. Analysis of large scale expression sequenced tags (ESTs) from the anural ascidian, *Molgula tectiformis*. *Dev Biol* **307**:460-482.
- Hill, M. M., K. W. Broman, E. Stupka, W. C. Smith, D. Jiang & A. Sidow. 2008. The *C. savignyi* genetic map and its integration with the reference sequence facilitates insights into chordate genome evolution. *Genome Res* **18**:1369-1379.
- Hinman, F. V. & M. B. Degnan. 2001. Homeobox genes, retinoic acid and the development and evolution of dual body plans in the ascidian *Herdmania curvata*. *Amer Zool* **41**:664-675.
- Holland, L. Z. 2007. Developmental biology: a chordate with a difference. *Nature* **447**:153-155.
- Holland, L. Z. 1990. Spermatogenesis in *Pyrosoma atlanticum* (Tunicata: Thaliacea: Pyrosomatida): Implications for tunicate phylogeny. *J Mar Biol* **105**:451-470.
- Holland, L. Z. 1999. Fine structure of spermatids and sperm of *Dolioletta gegenbauri* and *Doliolum nationalis* (Tunicata: Thaliacea): implications for tunicate phylogeny. *J Mar Biol* **110**:83-95.
- Holland, L. Z. & J. J. Gibson-Brown. 2003. The *Ciona intestinalis* genome: when the constraints are off. *Bioessays* **25**:529-532.
- Holland, P. 2006. Molecular phylogeny: My sister is a sea squirt? *Heredity* **96**:424-425.
- Iannelli, F., F. Griggio, G. Pesole & C. Gissi. 2007a. The mitochondrial genome of *Phallusia mammillata* and *Phallusia fumigata* (Tunicata, Ascidiacea): high genome plasticity at intra-genus level. *BMC Evol Biol* **7**:155.
- Iannelli, F., G. Pesole, P. Sordino & C. Gissi. 2007b. Mitogenomics reveals two cryptic species in *Ciona intestinalis*. *Trends Genet* **23**:419-422.

- Ikuta, T., N. Yoshida, N. Satoh & H. Saiga. 2004. *Ciona intestinalis* Hox gene cluster: Its dispersed structure and residual colinear expression in development. Proc Natl Acad Sci U S A **101**:15118-15123.
- Imai, K. S., M. Levine, N. Satoh & Y. Satou. 2006. Regulatory blueprint for a chordate embryo. Science **312**:1183-1187.
- Jefferies, R. 1986. The Ancestry of the Vertebrates. Cambridge Univ. Press, London.
- Jefferies, R. P. S. 1997. A defence of the calcichordates. Lethaia **30**:1-10.
- Jefferies, R. P. S. 1991. Two types of bilateral symmetry in the Metazoa: chordate and bilaterian. Pp. 94-127 in G. R. Bock & J. Marsh, eds. Biological Asymmetry and Handedness. Wiley, Chichester.
- Jeffery, W. R. 2007. Chordate ancestry of the neural crest: new insights from ascidians. Semin Cell Dev Biol **18**:481-491.
- Jeffery, W. R., A. G. Strickler & Y. Yamamoto. 2004. Migratory neural crest-like cells form body pigmentation in a urochordate embryo. Nature **431**:696-699.
- Katz, J. M. 1983. Comparative anatomy of the tunicate tadpole, *Ciona intestinalis*. Biol. Bull. **164**:1-27.
- Kim, Y. O., H. K. Cho, E. M. Park, B. H. Nam, Y. B. Hur, S. J. Lee & J. Cheong. 2008. Generation of expressed sequence tags for immune gene discovery and marker development in the sea squirt, *Halocynthia roretzi*. J Microbiol Biotechnol **18**:1510-1517.
- Kott, P. 2005. Catalogue of Tunicata in Australian Waters. Australian Biological Resources Study.
- Kowalevsky, A. 1868. Beiträge zur Entwicklungsgeschichte der Tunicaten. Nachrichten Gesellschaft Wissenschaften Göttingen **19**:401-415.
- Lacalli, T. C. 1999. Tunicate tails, stolons, and the origin of the vertebrate trunk. Biol Rev Camb Philos Soc **74**:177-198.
- Lahille, M. F. 1886. Sur la classification des Tuniciers. C. R. Acad. Sci. Paris **102**:1573-1575.
- Lahille, M. F. 1890. Recherche sur les Tuniciers. Faculté des Sciences de Paris, Toulouse.
- Lahille, M. F. 1887. Etude systématique des Tuniciers. C. R. Assoc. Fr. Adv. Sci. **16**:667-677.
- Lambert, C. C. 2005a. Historical introduction, overview, and reproductive biology of the protochordates. Can J Zool **83**:1-7.
- Lambert, G. 2005b. Ecology and natural history of protochordates. Can J Zool **83**:34-50.

- Lartillot, N., H. Brinkmann & H. Philippe. 2007. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol Biol* **7** **Suppl 1**:S4.
- Lartillot, N., T. Lepage & S. Blanquart. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**:2286-2288.
- Lartillot, N. & H. Philippe. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* **21**:1095-1109.
- Lemaire, P. 2006. Developmental biology. How many ways to make a chordate? *Science* **312**:1145-1146.
- Lemaire, P. 2009. Unfolding a chordate developmental program, one cell at a time: invariant cell lineages, short-range inductions and evolutionary plasticity in ascidians. *Dev Biol* **332**:48-60.
- Lemaire, P., W. C. Smith & H. Nishida. 2008. Ascidians and the plasticity of the chordate developmental program. *Curr Biol* **18**:R620-631.
- Locke, A. 2007. Invasion of the southern Gulf of St. Lawrence by the clubbed tunicate (*Styela clava* Herdman): Potential mechanisms for invasions of Prince Edward Island estuaries. *J. Exp. Mar. Biol. Ecol.* **342**:69-77.
- Locke, A. & M. Carman. 2009. An overview of the 2nd International Invasive Sea Squirt Conference: What we learned. *Aquatic Invasions* **4**:1-4.
- Mallatt, J. & C. J. Winchell. 2007. Ribosomal RNA genes and deuterostome phylogeny revisited: more cyclostomes, elasmobranchs, reptiles, and a brittle star. *Mol Phylogenet Evol* **43**:1005-1022.
- Marletaz, F., L. Z. Holland, V. Laudet & M. Schubert. 2006. Retinoic acid signaling and the evolution of chordates. *Int J Biol Sci* **2**:38-47.
- Nishino, A. & N. Satoh. 2001. The simple tail of chordates: phylogenetic significance of appendicularians. *Genesis* **29**:36-45.
- Paps, J., J. Baguna & M. Riutort. 2009. Bilaterian phylogeny: a broad sampling of 13 nuclear genes provides a new Lophotrochozoa phylogeny and supports a paraphyletic basal acoelomorpha. *Mol Biol Evol* **26**:2397-2406.
- Passamanek, Y. J. & A. Di Gregorio. 2005. *Ciona intestinalis*: chordate development made simple. *Dev Dyn* **233**:1-19.
- Pennisi, E. 2002a. Comparative genomics. Tunicate genome shows a little backbone. *Science* **298**:2111-2112.

- Pennisi, E. 2002b. Evolution and development. Comparative biology joins the molecular age. *Science* **296**:1792-1795.
- Peterson, K. J. 1995. A phylogenetic test of the calcichordate scenario. *Lethaia* **28**:25-38.
- Philippe, H., N. Lartillot & H. Brinkmann. 2005. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. *Mol Biol Evol* **22**:1246-1253.
- Putnam, N. H., T. Butts, D. E. Ferrier, R. F. Furlong, U. Hellsten, T. Kawashima, M. Robinson-Rechavi, E. Shoguchi, A. Terry, J. K. Yu, E. L. Benito-Gutierrez, I. Dubchak, J. Garcia-Fernandez, J. J. Gibson-Brown, I. V. Grigoriev, A. C. Horton, P. J. de Jong, J. Jurka, V. V. Kapitonov, Y. Kohara, Y. Kuroki, E. Lindquist, S. Lucas, K. Osoegawa, L. A. Pennacchio, A. A. Salamov, Y. Satou, T. Sauka-Spengler, J. Schmutz, I. T. Shin, A. Toyoda, M. Bronner-Fraser, A. Fujiyama, L. Z. Holland, P. W. Holland, N. Satoh & D. S. Rokhsar. 2008. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**:1064-1071.
- Ranwez, V., F. Delsuc, S. Ranwez, K. Belkhir, M. K. Tilak & E. J. Douzery. 2007. OrthoMaM: a database of orthologous genomic markers for placental mammal phylogenetics. *BMC Evol Biol* **7**:241.
- Rinehart, K. L. 2000. Antitumor compounds from tunicates. *Med Res Rev* **20**:1-27.
- Robison, B. H., K. R. Reisenbichler & R. E. Sherlock. 2005. Giant larvacean houses: rapid carbon transport to the deep sea floor. *Science* **308**:1609-1611.
- Rokas, A., D. Kruger & S. B. Carroll. 2005. Animal evolution and the molecular signature of radiations compressed in time. *Science* **310**:1933-1938.
- Rota-Stabelli, O., Z. Yang & M. J. Telford. 2009. MtZoa: a general mitochondrial amino acid substitutions model for animal evolutionary studies. *Mol Phylogenet Evol* **52**:268-272.
- Rothberg, J. M. & J. H. Leamon. 2008. The development and impact of 454 sequencing. *Nat Biotechnol* **26**:1117-1124.
- Rowe, T. 2004. Chordate phylogeny and development. Pp. 384-409 *in* J. Cracraft & M. J. Donoghue, eds. *Assembling the Tree of Life*. Oxford University Press, Oxford.
- Ruppert, E. 2005. Key characters uniting hemichordates and chordates: homologies or homoplasies? *Can J Zool* **83**:8-23.
- Saccone, C., C. De Giorgi, C. Gissi, G. Pesole & A. Reyes. 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. *Gene* **238**:195-209.

- Satoh, N. 2003. The ascidian tadpole larva: comparative molecular development and genomics. *Nat Rev Genet* **4**:285-295.
- Satoh, N., T. Kawashima, E. Shoguchi & Y. Satou. 2006. Urochordate genomes. *Genome Dyn* **2**:198-212.
- Savill, N. J., D. C. Hoyle & P. G. Higgs. 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. *Genetics* **157**:399-411.
- Seo, H. C., R. B. Edvardsen, A. D. Maeland, M. Bjordal, M. F. Jensen, A. Hansen, M. Flaatt, J. Weissenbach, H. Lehrach, P. Wincker, R. Reinhardt & D. Chourrout. 2004. Hox cluster disintegration with persistent anteroposterior order of expression in *Oikopleura dioica*. *Nature* **431**:67-71.
- Seo, H. C., M. Kube, R. B. Edvardsen, M. F. Jensen, A. Beck, E. Spriet, G. Gorsky, E. M. Thompson, H. Lehrach, R. Reinhardt & D. Chourrout. 2001. Miniature genome in the marine chordate *Oikopleura dioica*. *Science* **294**:2506.
- Shendure, J. & H. Ji. 2008. Next-generation DNA sequencing. *Nat Biotechnol* **26**:1135-1145.
- Shenkar, N. & Y. Loya. 2008. The solitary ascidian *Herdmania momus*: native (Red Sea) versus non-indigenous (Mediterranean) populations. *Biological Invasions* **10**:1431-1439.
- Shu, D. G., L. Chen, J. Han & X. L. Zhang. 2001a. An Early Cambrian tunicate from China. *Nature* **411**:472-473.
- Shu, D. G., S. C. Morris, J. Han, L. Chen, X. L. Zhang, Z. F. Zhang, H. Q. Liu, Y. Li & J. N. Liu. 2001b. Primitive deuterostomes from the Chengjiang Lagerstätte (Lower Cambrian, China). *Nature* **414**:419-424.
- Sierro, N., T. Kusakabe, K. J. Park, R. Yamashita, K. Kinoshita & K. Nakai. 2006. DBTGR: a database of tunicate promoters and their regulatory elements. *Nucleic Acids Res* **34**:D552-555.
- Singh, T. R., G. Tsagkogeorga, F. Delsuc, S. Blanquart, N. Shenkar, Y. Loya, E. J. P. Douzery & D. Huchon. En révision. Tunicate mitogenomics and phylogenetics: peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics*.
- Small, K. S., M. Brudno, M. M. Hill & A. Sidow. 2007. A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. *Genome Biol* **8**:R41.

Sodergren, E.G. M. WeinstockE. H. DavidsonR. A. CameronR. A. GibbsR. C. AngererL. M. AngererM. I. ArnoneD. R. BurgessR. D. BurkeJ. A. CoffmanM. DeanM. R. ElphickC. A. EttensohnK. R. FoltzA. HamdounR. O. HynesW. H. KleinW. MarzluffD. R. McClayR. L. MorrisA. MushegianJ. P. RastL. C. SmithM. C. ThorndykeV. D. VacquierG. M. WesselG. WrayL. ZhangC. G. ElsikO. ErmolaevaW. HlavinaG. HofmannP. KittsM. J. LandrumA. J. MackeyD. MaglottG. PanopoulouA. J. PoustkaK. PruittV. SapojnikovX. SongA. SouvorovV. SolovyevZ. WeiC. A. WhittakerK. WorleyK. J. DurbinY. ShenO. FedrigoD. GarfieldR. HaygoodA. PrimusR. SatijaT. SeversonM. L. Gonzalez-GarayA. R. JacksonA. MilosavljevicM. TongC. E. KillianB. T. LivingstonF. H. WiltN. AdamsR. BelleS. CarbonneauR. CheungP. CormierB. CossonJ. CroceA. Fernandez-GuerraA. M. GenevriereM. GoelH. KelkarJ. MoralesO. Mulner-LorillonA. J. RobertsonJ. V. GoldstoneB. ColeD. EpelB. GoldM. E. HahnM. Howard-AshbyM. ScallyJ. J. StegemanE. L. AllgoodJ. CoolK. M. JudkinsS. S. McCaffertyA. M. MusanteR. A. ObarA. P. RawsonB. J. RossettiI. R. GibbonsM. P. HoffmanA. LeoneS. IstrailS. C. MaternaM. P. SamantaV. StolcW. TongprasitQ. TuK. F. BergeronB. P. BrandhorstJ. WhittleK. BerneyD. J. BottjerC. CalestaniK. PetersonE. ChowQ. A. YuanE. ElhaikD. GraurJ. T. ReeseI. BosdetS. HeesunM. A. MarraJ. ScheinM. K. AndersonV. BrocktonK. M. BuckleyA. H. CohenS. D. FugmannT. HibinoM. Loza-CollaA. J. MajeskeC. MessierS. V. NairZ. PancerD. P. TerwilligerC. AgcaE. ArboledaN. ChenA. M. ChurcherF. HallbookG. W. HumphreyM. M. IdrisT. KiyamaS. LiangD. MellottX. MuG. MurrayR. P. OlinskiF. RaibleM. RoweJ. S. TaylorK. Tessmar-RaibleD. WangK. H. WilsonS. YaguchiT. GaasterlandB. E. GalindoH. J. GunaratneC. Julianom. KinukawaG. W. MoyA. T. NeillM. NomuraM. RaischA. ReadeM. M. RouxJ. L. SongY. H. SuI. K. TownleyE. VoroninaJ. L. WongG. AmoreM. BrannoE. R. BrownV. CavalieriV. DubocL. DuloquinC. FlytzanisC. GacheF. LaprazT. LepageA. LocascioP. MartinezG. MatassiV. MatrangaR. RangeF. RizzoE. RottingerW. BeaneC. BradhamC. ByrumT. GlennS. HussainG. ManningE. MirandaR. ThomasonK. WaltonA. WikramanayakeS. Y. WuR. XuC. T. BrownL. ChenR. F. GrayP. Y. LeeJ. NamP. OliveriJ. SmithD. MuznyS. BellJ. ChackoA. CreeS. CurryC. DavisH. DinhS. Dugan-RochaJ. FowlerR. GillC. HamiltonJ. HernandezS. HinesJ. Humel. JacksonA. JolivetC. KovarS. LeeL. LewisG. MinerM. MorganL. V. NazarethG. OkwuonuD. ParkerL. L. PuR. Thorn & R. Wright.

2006. The genome of the sea urchin *Strongylocentrotus purpuratus*. *Science* **314**:941-952.
- Stach, T. 2008. Chordate phylogeny and evolution: a not so simple three-taxon problem. *J Zool* **276**:117-141.
- Stach, T. & J. M. Turbeville. 2002. Phylogeny of Tunicata inferred from molecular and morphological characters. *Mol Phylogenet Evol* **25**:408-428.
- Swalla, B. J., C. B. Cameron, L. S. Corley & J. R. Garey. 2000. Urochordates are monophyletic within the deuterostomes. *Syst Biol* **49**:52-64.
- Swalla, B. J. & J. Xavier-Neto. 2008. Chordate origins and evolution. *Genesis* **46**:575-579.
- Telford, M. J. 2006. Animal phylogeny. *Curr Biol* **16**:R981-985.
- Telford, M. J., M. J. Wise & V. Gowri-Shankar. 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: examples from the bilateria. *Mol Biol Evol* **22**:1129-1136.
- Thorne, J. L., H. Kishino & I. S. Painter. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* **15**:1647-1657.
- Tsagkogeorga, G., X. Turon, E. J. P. Douzery & F. Delsuc. En préparation. Accelerated evolutionary rate of housekeeping genes in tunicates. *Journal of Molecular Evolution*.
- Tsagkogeorga, G., X. Turon, R. R. Hopcroft, M. K. Tilak, T. Feldstein, N. Shenkar, Y. Loya, D. Huchon, E. J. Douzery & F. Delsuc. 2009. An updated 18S rRNA phylogeny of tunicates based on mixture and secondary structure models. *BMC Evol Biol* **9**:187.
- Turbeville, J. M., J. R. Schulz & R. A. Raff. 1994. Deuterostome phylogeny and the sister group of the chordates: evidence from molecules and morphology. *Mol Biol Evol* **11**:648-655.
- Turon, X. & S. Lopez-Legentil. 2004. Ascidian molecular phylogeny inferred from mtDNA data with emphasis on the Aplousobranchiata. *Molecular Phylogenetics and Evolution* **33**:309-320.
- Turon, X., T. Nishikawa & M. Rius. 2006. Spread of *Microcosmus squamiger* (Asciacea: Pyuridae) in the Mediterranean Sea and adjacent waters. *J Exp Mar Biol Ecol* **342**:185-188.
- Wada, H. 1998. Evolutionary history of free-swimming and sessile lifestyles in urochordates as deduced from 18S rDNA molecular phylogeny. *Mol Biol Evol* **15**:1189-1194.

- Wada, H. & N. Satoh. 1994. Phylogenetic relationships among extant classes of echinoderms, as inferred from sequences of 18S rDNA, coincide with relationships deduced from the fossil record. *J Mol Evol* **38**:41-49.
- Wada, S., M. Tokuoka, E. Shoguchi, K. Kobayashi, A. Di Gregorio, A. Spagnuolo, M. Branno, Y. Kohara, D. Rokhsar, M. Levine, H. Saiga, N. Satoh & Y. Satou. 2003. A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. *Dev Genes Evol* **213**:222-234.
- Winchell, C. J., J. Sullivan, C. B. Cameron, B. J. Swalla & J. Mallatt. 2002. Evaluating hypotheses of deuterostome phylogeny and chordate evolution with new LSU and SSU ribosomal DNA data. *Mol Biol Evol* **19**:762-776.
- Wuyts, J., G. Perriere & Y. Van De Peer. 2004. The European ribosomal RNA database. *Nucleic Acids Res* **32**:D101-103.
- Xia, X., H. Huang, M. Carullo, E. Betran & E. N. Moriyama. 2007. Conflict between translation initiation and elongation in vertebrate mitochondrial genomes. *PLoS One* **2**:e227.
- Yang, Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**:1586-1591.
- Yoder, A. D. & Z. Yang. 2000. Estimation of primate speciation dates using local molecular clocks. *Mol Biol Evol* **17**:1081-1090.
- Yokobori, S., A. Kurabayashi, B. A. Neilan, T. Maruyama & E. Hirose. 2006. Multiple origins of the ascidian-*Prochloron* symbiosis: molecular phylogeny of photosymbiotic and non-symbiotic colonial ascidians inferred from 18S rDNA sequences. *Mol Phylogenet Evol* **40**:8-19.
- Yokobori, S., T. Oshima & H. Wada. 2005. Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. *Mol Phylogenet Evol* **34**:273-283.
- Yokobori, S., T. Ueda, G. Feldmaier-Fuchs, S. Paabo, R. Ueshima, A. Kondow, K. Nishikawa & K. Watanabe. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). *Genetics* **153**:1851-1862.
- Yokobori, S., Y. Watanabe & T. Oshima. 2003. Mitochondrial genome of *Ciona savignyi* (Urochordata, Ascidiacea, Enterogona): comparison of gene arrangement and tRNA genes with *Halocynthia roretzi* mitochondrial genome. *J Mol Evol* **57**:574-587.

- Zeng, L., M. Jacobs & B. Swalla. 2006. Coloniality and sociality has evolved once in Stolidobranch ascidians. *Integr. Comp. Biol.* **46**:255-268.
- Zeng, L. & B. Swalla. 2005. Molecular phylogeny of the protochordates: chordate evolution. *Can. J. Zool* **83**:24-33.