

Fiabilité des clades et congruence taxinomique

Application à la phylogénie des téléostéens
acanthomorphes

Blaise Li

UMR 7138 CNRS IRD MNHN UPMC
Systématique, Adaptation, Évolution
Équipe Phylogénie

17 septembre 2008

Partie I : Fiabilité des clades et congruence taxinomique

Soutenance de
thèse

Blaise Li

Plan

Partie I : Fiabilité des
clades et congruence
taxinomique

Partie II : Application à la
phylogénie des
acanthomorphes

Les supports des clades

Sensibilité

Robustesse

Fiabilité

Partie I : Fiabilité des clades et congruence taxinomique

Plan

Partie I : Fiabilité des
clades et congruence
taxinomique

Partie II : Application à la
phylogénie des
acanthomorphes

Les supports des clades

Sensibilité

Robustesse

Fiabilité

Un indice de répétition des clades

Congruence taxinomique

Sommes et différences d'occurrences

Domaines de validité

Partie I : Fiabilité des clades et congruence taxinomique

Plan

Partie I : Fiabilité des
clades et congruence
taxinomique

Partie II : Application à la
phylogénie des
acanthomorphes

Les supports des clades

Sensibilité

Robustesse

Fiabilité

Un indice de répétition des clades

Congruence taxinomique

Sommes et différences d'occurrences

Domaines de validité

Synthétiser les résultats

Greedy summary tree

MRP

Combinaisons de relations élémentaires

Partie II : Application à la phylogénie des acanthomorphes

Plan

Partie I : Fiabilité des
clades et congruence
taxinomique

Partie II : Application à la
phylogénie des
acanthomorphes

Les téléostéens acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires récentes

Partie II : Application à la phylogénie des acanthomorphes

Plan

Partie I : Fiabilité des
clades et congruence
taxinomique

Partie II : Application à la
phylogénie des
acanthomorphes

Les téléostéens acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires récentes

Données et analyses

Données

Analyses primaires et secondaires

Partie II : Application à la phylogénie des acanthomorphes

Plan

Partie I : Fiabilité des
clades et congruence
taxinomique

Partie II : Application à la
phylogénie des
acanthomorphes

Les téléostéens acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires récentes

Données et analyses

Données

Analyses primaires et secondaires

Synthèse des synthèses

Méthode de résumé

Structure générale de l'arbre

Les principaux clades

Partie I

Fiabilité des clades et congruence taxinomique

Les supports des clades

Soutenance de
thèse

Blaise Li

Les supports des clades

Sensibilité

Robustesse

Fiabilité

Les supports des
clades

Sensibilité

Robustesse

Fiabilité

Un indice de
répétition des
clades

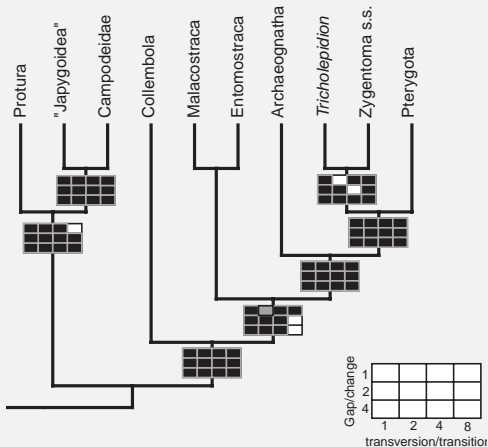
Synthétiser les
résultats

Un indice de répétition des clades

Synthétiser les résultats

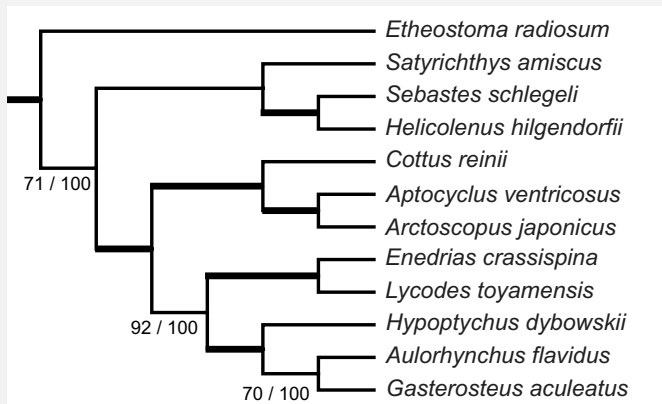
- ▶ Sensibilité : résister à des changements de méthode
 - ▶ Exemple : « tapis navajos »

- ▶ Sensibilité : résister à des changements de méthode
 - ▶ Exemple : « tapis navajos »



- ▶ Robustesse : résister à la perturbation des données
 - ▶ Exemples : bootstrap et jackknife

- ▶ Robustesse : résister à la perturbation des données
 - ▶ Exemples : bootstrap et jackknife



- ▶ Sensibilité et robustesse donnent une idée de la force du signal

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)
- ▶ Les relations obtenues sont-elles compatibles avec l'arbre réel ? Correspond-ce à la phylogénie des espèces ?

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)
- ▶ Les relations obtenues sont-elles compatibles avec l'arbre réel ? Correspond-ce à la phylogénie des espèces ?
- ▶ Problème majeur : on ne connaît pas l'arbre réel (en général, c'est lui qu'on cherche)

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)
- ▶ Les relations obtenues sont-elles compatibles avec l'arbre réel ? Correspond-ce à la phylogénie des espèces ?
- ▶ Problème majeur : on ne connaît pas l'arbre réel (en général, c'est lui qu'on cherche)
- ▶ Fiabilité interne : *total evidence* + support

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)
- ▶ Les relations obtenues sont-elles compatibles avec l'arbre réel ? Correspond-ce à la phylogénie des espèces ?
- ▶ Problème majeur : on ne connaît pas l'arbre réel (en général, c'est lui qu'on cherche)
- ▶ Fiabilité interne : *total evidence* + support
 - ▶ support « fait pour ça » : probabilités postérieures

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)
- ▶ Les relations obtenues sont-elles compatibles avec l'arbre réel ? Correspond-ce à la phylogénie des espèces ?
- ▶ Problème majeur : on ne connaît pas l'arbre réel (en général, c'est lui qu'on cherche)
- ▶ Fiabilité interne : *total evidence* + support
 - ▶ support « fait pour ça » : probabilités postérieures
- ▶ Fiabilité externe : congruence taxinomique

Congruence taxinomique

- ▶ Les différents jeux de données ont des signaux parfois contradictoires.

Soutenance de
thèse

Blaise Li

Les supports des
clades

Sensibilité

Robustesse

Fiabilité

Un indice de
répétition des
clades

Synthétiser les
résultats

Congruence taxinomique

- ▶ Les différents jeux de données ont des signaux parfois contradictoires.
- ▶ Dans une combinaison, on espère que le signal historique s'impose.

Soutenance de
thèse

Blaise Li

Les supports des
clades

Sensibilité

Robustesse

Fiabilité

Un indice de
répétition des
clades

Synthétiser les
résultats

Congruence taxinomique

Soutenance de
thèse

Blaise Li

Les supports des
clades

Sensibilité

Robustesse

Fiabilité

Un indice de
répétition des
clades

Synthétiser les
résultats

- ▶ Les différents jeux de données ont des signaux parfois contradictoires.
- ▶ Dans une combinaison, on espère que le signal historique s'impose.
- ▶ Mais parfois, des regroupements bien soutenus dans la combinaison ne correspondent pas aux relations de parenté réelles entre espèces.

Congruence taxinomique

Soutenance de
thèse

Blaise Li

Les supports des
clades

Sensibilité

Robustesse

Fiabilité

Un indice de
répétition des
clades

Synthétiser les
résultats

- ▶ Les différents jeux de données ont des signaux parfois contradictoires.
- ▶ Dans une combinaison, on espère que le signal historique s'impose.
- ▶ Mais parfois, des regroupements bien soutenus dans la combinaison ne correspondent pas aux relations de parenté réelles entre espèces.
- ▶ Les biais varient selon le jeu de données (sous certaines conditions).

Congruence taxinomique

Soutenance de
thèse

Blaise Li

Les supports des
clades

Sensibilité

Robustesse

Fiabilité

Un indice de
répétition des
clades

Synthétiser les
résultats

- ▶ Les différents jeux de données ont des signaux parfois contradictoires.
- ▶ Dans une combinaison, on espère que le signal historique s'impose.
- ▶ Mais parfois, des regroupements bien soutenus dans la combinaison ne correspondent pas aux relations de parenté réelles entre espèces.
- ▶ Les biais varient selon le jeu de données (sous certaines conditions).
- ▶ La part de signal commune indique des relations de parenté réelles.

Congruence taxinomique

- ▶ Les différents jeux de données ont des signaux parfois contradictoires.
- ▶ Dans une combinaison, on espère que le signal historique s'impose.
- ▶ Mais parfois, des regroupements bien soutenus dans la combinaison ne correspondent pas aux relations de parenté réelles entre espèces.
- ▶ Les biais varient selon le jeu de données (sous certaines conditions).
- ▶ La part de signal commune indique des relations de parenté réelles.
- ▶ La congruence entre les arbres est signe de fiabilité.

Un indice de répétition des clades

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Les supports des clades

Un indice de répétition des clades

Congruence taxinomique

Sommes et différences d'occurrences

Domaines de validité

Synthétiser les résultats

Un indice de répétition des clades

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Dettaï (2004, conclusion de la thèse) :

L'approche adoptée [...] est lourde à mettre en place [...]. Son automatisation serait intéressante.

Un indice de répétition des clades

Dettaï (2004, conclusion de la thèse) :

L'approche adoptée [...] est lourde à mettre en place [...]. Son automatisation serait intéressante.

Objectif méthodologique :

- ▶ Formaliser la congruence taxinomique pour pouvoir automatiser

Un indice de répétition des clades

Dettaï (2004, conclusion de la thèse) :

L'approche adoptée [...] est lourde à mettre en place [...]. Son automatisation serait intéressante.

Objectif méthodologique :

- ▶ Formaliser la congruence taxinomique pour pouvoir automatiser
- ▶ Sous forme d'un indice pour une meilleure lisibilité

Un indice de répétition des clades

Dettaï (2004, conclusion de la thèse) :

L'approche adoptée [...] est lourde à mettre en place [...]. Son automatisation serait intéressante.

Objectif méthodologique :

- ▶ Formaliser la congruence taxinomique pour pouvoir automatiser
- ▶ Sous forme d'un indice pour une meilleure lisibilité
- ▶ Au risque de perdre de l'information

Congruence taxinomique

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Obtenir des séries d'arbres à comparer

Congruence taxinomique

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Obtenir des séries d'arbres à comparer
- ▶ Il faut que les données soient **indépendantes**

Congruence taxinomique

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Obtenir des séries d'arbres à comparer
- ▶ Il faut que les données soient **indépendantes**
- ▶ Mais combiner évite certains biais, d'où la technique des combinaisons partielles

Indépendance

- ▶ La question n'est pas « Ai-je le droit de combiner les marqueurs ? » (test ILD) mais au contraire « Quelles sont les parties minimales de mes données que je peux garder séparées ? ».

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Indépendance

- ▶ La question n'est pas « Ai-je le droit de combiner les marqueurs ? » (test ILD) mais au contraire « Quelles sont les parties minimales de mes données que je peux garder séparées ? ».
- ▶ Si on analyse séparément des parties non-indépendantes, on risque de voir des résultats biaisés répétés.

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

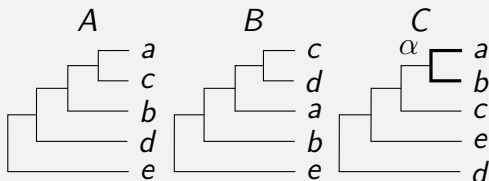
Domaines de validité

Synthétiser les
résultats

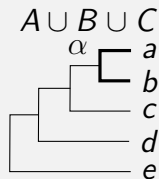
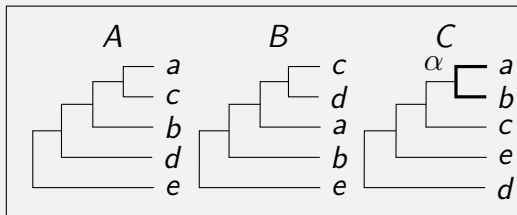
- ▶ La question n'est pas « Ai-je le droit de combiner les marqueurs ? » (test ILD) mais au contraire « Quelles sont les parties minimales de mes données que je peux garder séparées ? ».
- ▶ Si on analyse séparément des parties non-indépendantes, on risque de voir des résultats biaisés répétés.
- ▶ Ne pas séparer des gènes physiquement liés (transferts horizontaux, perte tardive du polymorphisme ancestral)

- ▶ La question n'est pas « Ai-je le droit de combiner les marqueurs ? » (test ILD) mais au contraire « Quelles sont les parties minimales de mes données que je peux garder séparées ? ».
- ▶ Si on analyse séparément des parties non-indépendantes, on risque de voir des résultats biaisés répétés.
- ▶ Ne pas séparer des gènes physiquement liés (transferts horizontaux, perte tardive du polymorphisme ancestral)
- ▶ Ne pas séparer des gènes codant des produits en interaction moléculaire directe (co-évolution donc peut-être co-convergence fonctionnelle)

Combinaisons partielles



Combinaisons partielles



Les supports des clades

Un indice de répétition des clades

Congruence taxinomique

Indépendance

Combinaisons partielles

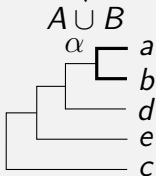
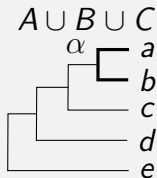
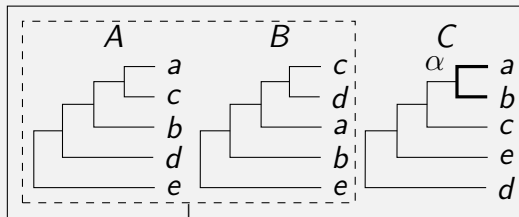
Schémas de partitionnement

Sommes et différences d'occurrences

Domaines de validité

Synthétiser les résultats

Combinaisons partielles



Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique
Indépendance

Combinaisons partielles

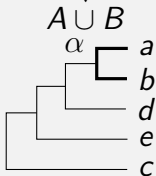
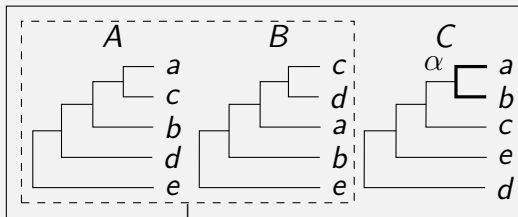
Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Combinaisons partielles



jeux de données	nombre d'occurrences pour α
A, B, C	1
$A \cup B \cup C$	1
$A \cup B, C$	2

Schémas de partitionnement

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Jeux de données :

Schémas de partitionnement

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique
Indépendance
Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences
Domaines de validité

Synthétiser les
résultats

Jeux de données :

- ▶ Jeux de données élémentaires : A , B et C

Schémas de partitionnement

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Jeux de données :

- ▶ Jeux de données élémentaires : A , B et C
- ▶ Combinaison totale : $A \cup B \cup C$

Schémas de partitionnement

Jeux de données :

- ▶ Jeux de données élémentaires : A , B et C
- ▶ Combinaison totale : $A \cup B \cup C$
- ▶ Combinaisons partielles : $B \cup C$, $A \cup C$ et $A \cup B$

Schémas de partitionnement

Jeux de données :

- ▶ Jeux de données élémentaires : A , B et C
- ▶ Combinaison totale : $A \cup B \cup C$
- ▶ Combinaisons partielles : $B \cup C$, $A \cup C$ et $A \cup B$

On analyse, puis on compare les arbres indépendants
entre eux :

Schémas de partitionnement

Jeux de données :

- ▶ Jeux de données élémentaires : A , B et C
- ▶ Combinaison totale : $A \cup B \cup C$
- ▶ Combinaisons partielles : $B \cup C$, $A \cup C$ et $A \cup B$

On analyse, puis on compare les arbres indépendants entre eux :

- ▶ Schémas de partitionnement : (A, B, C) , $(A, B \cup C)$, $(B, A \cup C)$, $(C, A \cup B)$ et $(A \cup B \cup C)$

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique
Indépendance

Combinaisons partielles
Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (extraction optimale du signal)

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (extraction optimale du signal)
- ▶ Somme d'occurrences brutes ou de supports du clade α :

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (extraction optimale du signal)
- ▶ Somme d'occurrences brutes ou de supports du clade α :

$$N(\alpha) = \max_{Schp \in SP} \left(\sum_{j \in Schp} s_j(\alpha) \right)$$

$s_j(\alpha)$: support du clade α pour le jeu de données j

j : jeu de données (élémentaire ou combiné)

$Schp$: schéma de partitionnement

SP : ensemble des schémas de partitionnement

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (extraction optimale du signal)
- ▶ **Somme d'occurrences brutes ou de supports du clade α** :

$$N(\alpha) = \max_{Schp \in SP} \left(\sum_{j \in Schp} s_j(\alpha) \right)$$

$s_j(\alpha)$: support du clade α pour le jeu de données j

j : jeu de données (élémentaire ou combiné)

$Schp$: schéma de partitionnement

Somme d'occurrences

- ▶ Plus un clade est répété de façon **indépendante**, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (extraction optimale du signal)
- ▶ Somme d'occurrences brutes ou de supports du clade α :

$$N(\alpha) = \max_{Schp \in SP} \left(\sum_{j \in Schp} s_j(\alpha) \right)$$

j : jeu de données (élémentaire ou combiné)

$Schp$: schéma de partitionnement

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (**extraction optimale du signal**)
- ▶ Somme d'occurrences brutes ou de supports du clade α :

$$N(\alpha) = \max_{Schp \in SP} \left(\sum_{j \in Schp} s_j(\alpha) \right)$$

$Schp$: schéma de partitionnement

SP : ensemble des schémas de partitionnement

Somme d'occurrences

- ▶ Plus un clade est répété de façon indépendante, plus il est fiable
→ somme d'occurrences
- ▶ On choisit le schéma de partitionnement offrant la meilleure somme (extraction optimale du signal)
- ▶ Somme d'occurrences brutes ou de supports du clade α :

$$N(\alpha) = \max_{Schp \in SP} \left(\sum_{j \in Schp} s_j(\alpha) \right)$$

$s_j(\alpha)$: support du clade α pour le jeu de données j

j : jeu de données (élémentaire ou combiné)

$Schp$: schéma de partitionnement

SP : ensemble des schémas de partitionnement

Contradiction

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Plus un clade est contredit, moins il est fiable

Contradiction

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Plus un clade est contredit, moins il est fiable
- ▶ Soustraire le nombre d'occurrences du meilleur contradicteur β :

- ▶ Plus un clade est contredit, moins il est fiable
- ▶ Soustraire le nombre d'occurrences du meilleur contradicteur β :

Indice de répétition :

$$R(\alpha) = N(\alpha) - N(\beta)$$

- ▶ Plus un clade est contredit, moins il est fiable
- ▶ Soustraire le nombre d'occurrences du meilleur contradicteur β :

Indice de répétition :

$$R(\alpha) = N(\alpha) - N(\beta)$$

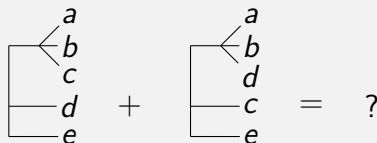
- ▶ Problème : quel est le meilleur contradicteur ?

Le meilleur contradicteur

- ▶ contradicteur = clade incompatible
- ▶ α et β incompatibles
 - $\Leftrightarrow \alpha$ et β ne peuvent pas être ensemble dans le même arbre
 - $\Leftrightarrow \alpha$ et β ont au moins un taxon en commun et chacun au moins un taxon propre

Le meilleur contradicteur

- ▶ contradicteur = clade incompatible
- ▶ α et β incompatibles
 - $\Leftrightarrow \alpha$ et β ne peuvent pas être ensemble dans le même arbre
 - $\Leftrightarrow \alpha$ et β ont au moins un taxon en commun et chacun au moins un taxon propre



Le meilleur contradicteur

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Le meilleur contradicteur = le plus fiable

Le meilleur contradicteur

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Le meilleur contradicteur = le plus fiable
- ▶ Pour trouver le plus fiable, il faut connaître R , or, c'est ce qu'on cherche

Le meilleur contradicteur

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

- ▶ Le meilleur contradicteur = le plus fiable
- ▶ Pour trouver le plus fiable, il faut connaître R , or, c'est ce qu'on cherche
- ▶ Ajustements successifs à partir de N : soit c'est stable, soit c'est cyclique

Domaines de validité

- ▶ Comment comparer des relations obtenues sur des ensembles de taxons différents ?

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

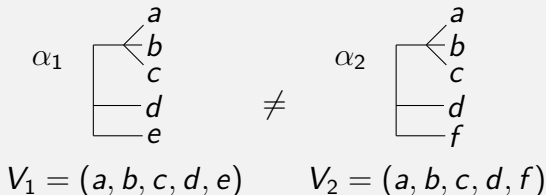
Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Domaines de validité

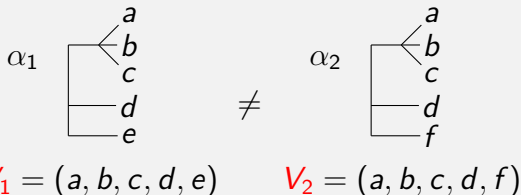
- ▶ Comment comparer des relations obtenues sur des ensembles de taxons différents ?



- ▶ Les clades sont en fait des **bipartitions** orientées (l'extérieur compte aussi).

Domaines de validité

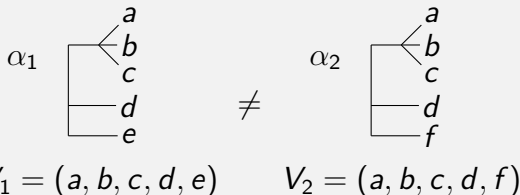
- ▶ Comment comparer des relations obtenues sur des ensembles de taxons différents ?



- ▶ Les clades sont en fait des bipartitions orientées (l'extérieur compte aussi).
- ▶ Les comparaisons ne peuvent se faire que dans l'intersection des **domaines de validité** des arbres comparés : $V' = V_1 \cap V_2 = (a, b, c, d)$

Domaines de validité

- ▶ Comment comparer des relations obtenues sur des ensembles de taxons différents ?



- ▶ Les clades sont en fait des bipartitions orientées (l'extérieur compte aussi).
- ▶ Les comparaisons ne peuvent se faire que dans l'intersection des domaines de validité des arbres comparés : $V' = V_1 \cap V_2 = (a, b, c, d)$
- ▶ $\alpha_{1|_{V'}} = \alpha_{2|_{V'}} = \alpha' = (d, (a, b, c))$

Domaines de validité

niveau 1

$$\begin{array}{l} T_A/V_A \\ T_B/V_B \\ T_C/V_C \end{array}$$

$$\begin{array}{l} T_{BUC}/V_{BUC} \\ T_{AUC}/V_{AUC} \\ T_{AUB}/V_{AUB} \\ T_{AUBUC}/V_{AUBUC} \end{array}$$

1. arbres-source $\rightarrow \alpha$

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

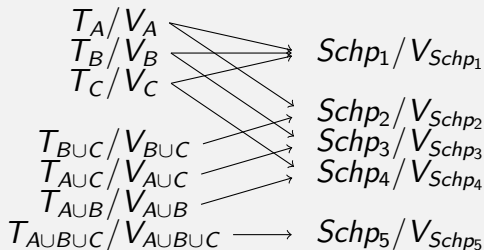
Domaines de validité

Synthétiser les
résultats

Domaines de validité

niveau 1

niveau 2



1. arbres-source $\rightarrow \alpha$
2. sommes de supports $\rightarrow N(\alpha')$
($\alpha' = \alpha|_{V_{Schp}}$)

Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

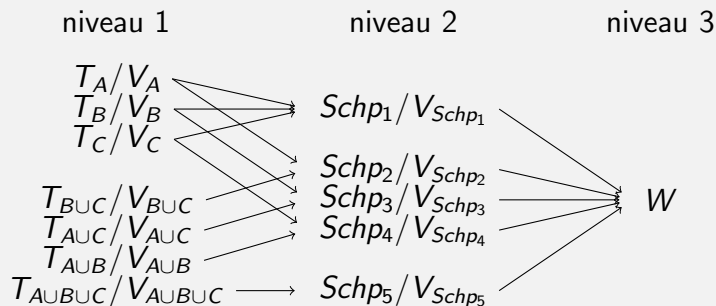
Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

Domaines de validité



Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

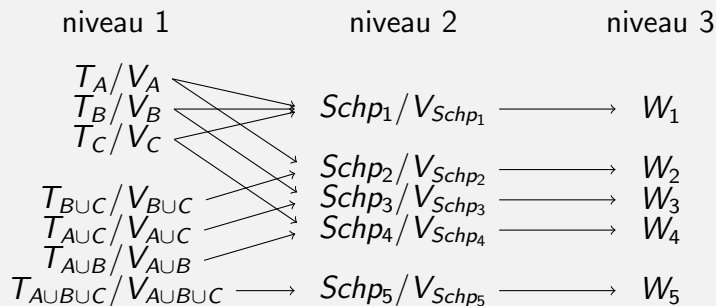
Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

1. arbres-source $\rightarrow \alpha$
2. sommes de supports $\rightarrow N(\alpha')$
($\alpha' = \alpha|_{V_{Schp}}$)
3. différences de sommes de supports $\rightarrow R(\alpha'')$
($\alpha'' = \alpha'|_W$)

Domaines de validité



Les supports des
clades

Un indice de
répétition des
clades

Congruence taxinomique

Indépendance

Combinaisons partielles

Schémas de
partitionnement

Sommes et différences
d'occurrences

Domaines de validité

Synthétiser les
résultats

1. arbres-source $\rightarrow \alpha$
2. sommes de supports $\rightarrow N(\alpha')$
($\alpha' = \alpha|_{V_{Schp}}$)
3. différences de sommes de supports $\rightarrow R(\alpha'')$
($\alpha'' = \alpha'|_W$)

Synthétiser les résultats

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

Greedy summary tree

MRP

Combinaisons de relations
élémentaires

Les supports des clades

Un indice de répétition des clades

Synthétiser les résultats

Greedy summary tree

MRP

Combinaisons de relations élémentaires

Synthétiser les résultats

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

**Synthétiser les
résultats**

Greedy summary tree

MRP

**Combinaisons de relations
élémentaires**

- ▶ But : lisibilité des résultats

Synthétiser les résultats

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

Greedy summary tree

MRP

Combinaisons de relations
élémentaires

- ▶ But : lisibilité des résultats
- ▶ Dans l'idéal : un arbre avec les indices sur les branches

Synthétiser les résultats

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

Greedy summary tree

MRP

Combinaisons de relations
élémentaires

- ▶ But : lisibilité des résultats
- ▶ Dans l'idéal : un arbre avec les indices sur les branches
- ▶ Problèmes : clades incompatibles, domaines de validité différents

Greedy summary tree

- ▶ On classe les clades du plus fiable au moins fiable.

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

Greedy summary tree

MRP

Combinaisons de relations
élémentaires

Greedy summary tree

- ▶ On classe les clades du plus fiable au moins fiable.
- ▶ On sélectionne les plus fiables compatibles avec ceux déjà retenus.

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

Greedy summary tree

MRP

Combinaisons de relations
élémentaires

Greedy summary tree

- ▶ On classe les clades du plus fiable au moins fiable.
- ▶ On sélectionne les plus fiables compatibles avec ceux déjà retenus.
- ▶ On assemble l'arbre à partir des clades retenus.

Greedy summary tree

- ▶ On classe les clades du plus fiable au moins fiable.
- ▶ On sélectionne les plus fiables compatibles avec ceux déjà retenus.
- ▶ On assemble l'arbre à partir des clades retenus.
- ▶ On peut directement placer les indices de répétition sur les branches.

Greedy summary tree

- ▶ On classe les clades du plus fiable au moins fiable.
- ▶ On sélectionne les plus fiables compatibles avec ceux déjà retenus.
- ▶ On assemble l'arbre à partir des clades retenus.
- ▶ On peut directement placer les indices de répétition sur les branches.
- ▶ Ne marche qu'avec des clades définis sur le même ensemble de taxons

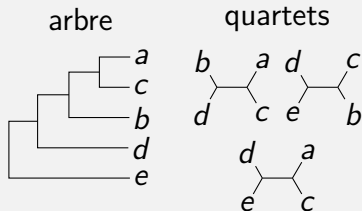
- ▶ La méthode « classique » de *supertree*

- ▶ La méthode « classique » de *supertree*
- ▶ Chaque clade est un pseudo-caractère dans une matrice taxons x caractères
- ▶ Pour chaque taxon : 0, 1 ou ? (accepte les différences de domaines de validité)
(*Matrix Representation*)

- ▶ La méthode « classique » de *supertree*
- ▶ Chaque clade est un pseudo-caractère dans une matrice taxons x caractères
- ▶ Pour chaque taxon : 0, 1 ou ? (accepte les différences de domaines de validité)
(*Matrix Representation*)
- ▶ On pondère les clades par leur indice de répétition

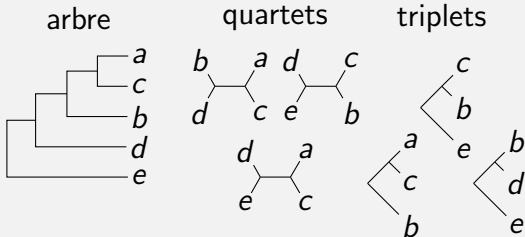
- ▶ La méthode « classique » de *supertree*
- ▶ Chaque clade est un pseudo-caractère dans une matrice taxons x caractères
- ▶ Pour chaque taxon : 0, 1 ou ? (accepte les différences de domaines de validité)
(*Matrix Representation*)
- ▶ On pondère les clades par leur indice de répétition
- ▶ On cherche l'arbre le plus parcimonieux
(*with Parsimony*)

Combinaisons de relations élémentaires



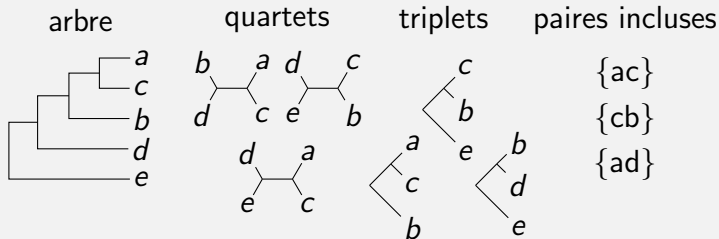
- ▶ Les arbres (et les clades) peuvent être décomposés en relations élémentaires.

Combinaisons de relations élémentaires



- ▶ Les arbres (et les clades) peuvent être décomposés en relations élémentaires.

Combinaisons de relations élémentaires



Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

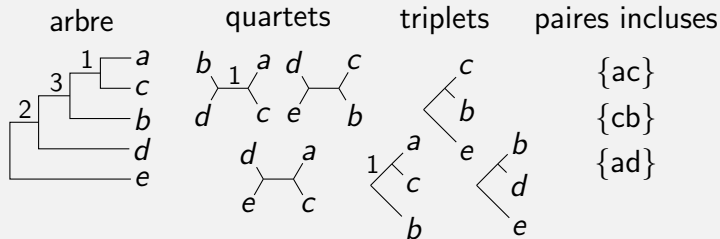
Greedy summary tree

MRP

Combinaisons de relations
élémentaires

- ▶ Les arbres (et les clades) peuvent être décomposés en relations élémentaires.

Combinaisons de relations élémentaires



- ▶ Les arbres (et les clades) peuvent être décomposés en relations élémentaires.
- ▶ L'indice d'une relation élémentaire est celui du meilleur clade l'induisant.

Combinaisons de relations élémentaires

Soutenance de
thèse

Blaise Li

Les supports des
clades

Un indice de
répétition des
clades

Synthétiser les
résultats

Greedy summary tree

MRP

Combinaisons de relations
élémentaires

- ▶ Onetree (Aho et al., 1981) : trouve la solution s'il y a compatibilité globale

Combinaisons de relations élémentaires

- ▶ Onetree (Aho et al., 1981) : trouve la solution s'il y a compatibilité globale
- ▶ MinCut (Semple et Steel, 2000; Page, 2002) : s'il n'existe pas de solution, résoud les incompatibilités en supprimant certaines paires incluses

Combinaisons de relations élémentaires

- ▶ Onetree (Aho et al., 1981) : trouve la solution s'il y a compatibilité globale
- ▶ MinCut (Semple et Steel, 2000; Page, 2002) : s'il n'existe pas de solution, résoud les incompatibilités en supprimant certaines paires incluses
- ▶ Dans cette thèse : tentatives d'adaptation de cette méthode en utilisant l'indice de fiabilité pour écarter certaines paires incluses

Partie II

Application à la phylogénie des acanthomorphes

Les téléostéens acanthomorphes

Soutenance de
thèse

Blaise Li

Les téléostéens acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires
récentes

Données et
analyses

Synthèse des
synthèses

Les téléostéens acanthomorphes

Position phylogénétique

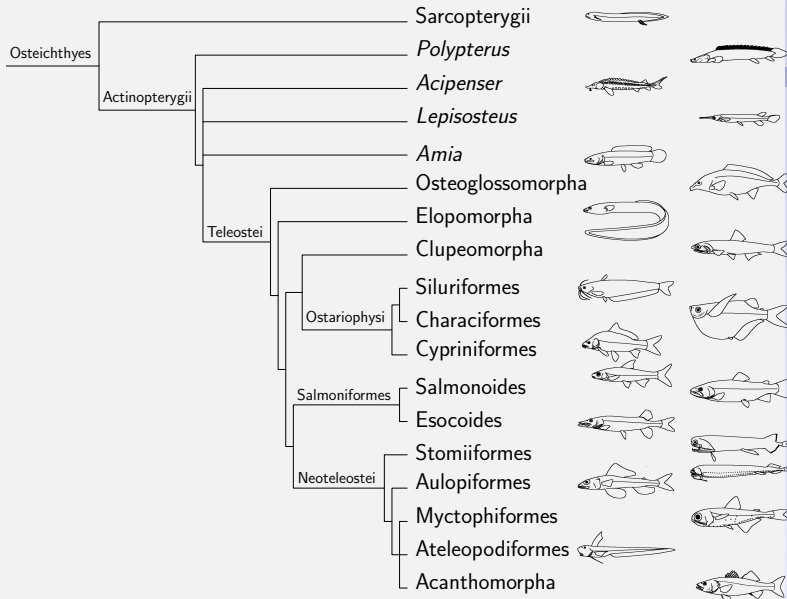
Diversité

Phylogénies moléculaires récentes

Données et analyses

Synthèse des synthèses

Position phylogénétique



Soutenance de thèse

Blaise Li

Les téléostéens acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires récentes

Données et analyses

Synthèse des synthèses

Diversité

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

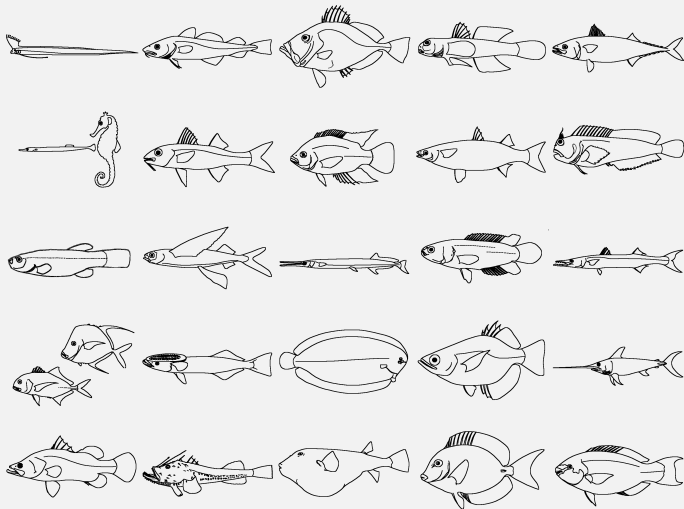
Position phylogénétique

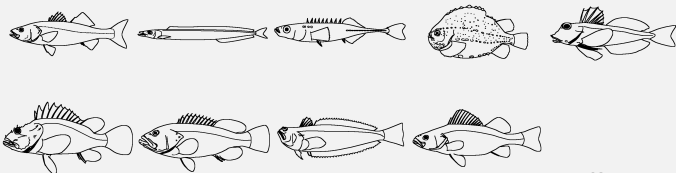
Diversité

Phylogénies moléculaires
récentes

Données et
analyses

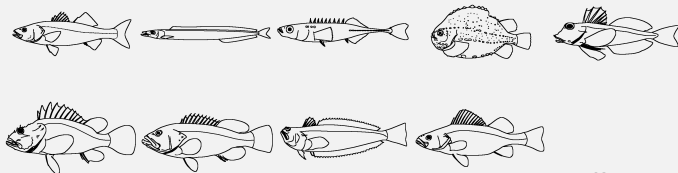
Synthèse des
synthèses





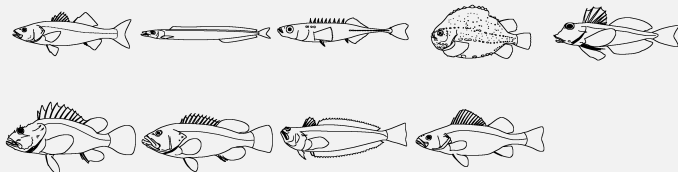
et d'autres...

- ▶ Une grande diversité (plus de 300 familles)



et d'autres...

- ▶ Une grande diversité (plus de 300 familles)
- ▶ Difficile de faire des rapprochements morphologique à large échelle



et d'autres...

- ▶ Une grande diversité (plus de 300 familles)
- ▶ Difficile de faire des rapprochements morphologique à large échelle
- ▶ Les phylogénies moléculaires apportent des résultats inattendus.

Phylogénies moléculaires récentes

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires
récentes

Données et
analyses

Synthèse des
synthèses

- ▶ Miya et al. (2001, 2003), Mabuchi et al. (2007) et Kawahara et al. (2008) : génomes mitochondriaux
- ▶ Chen et al. (2003) et Dettai et Lecointre (2004, 2005, 2008) : marqueurs mitochondriaux et nucléaires
- ▶ Smith et Wheeler (2004, 2006) et Smith et Craig (2007) : autres marqueurs mitochondriaux et nucléaires

Phylogénies moléculaires récentes

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Position phylogénétique

Diversité

Phylogénies moléculaires
récentes

Données et
analyses

Synthèse des
synthèses

- ▶ Miya et al. (2001, 2003), Mabuchi et al. (2007) et Kawahara et al. (2008) : génomes mitochondriaux
- ▶ Chen et al. (2003) et Dettai et Lecointre (2004, 2005, 2008) : marqueurs mitochondriaux et nucléaires
- ▶ Smith et Wheeler (2004, 2006) et Smith et Craig (2007) : autres marqueurs mitochondriaux et nucléaires

Les clades fiables de Chen (2001) et Dettai (2004) sont nommés par des lettres

→ on en retrouvera la plupart dans la synthèse

Les téléostéens acanthomorphes

Données et analyses

Données

Analyses primaires et secondaires

Synthèse des synthèses

4 marqueurs nucléaires :

	Séquences utilisées	Dont nouvelles
Rhodopsine (Chen, 2001)	190	92
MLL4 (Dettaï, 2004)	165	77
IRBP (Dettaï, 2004)	161	66
RNF213 (Nouveau)	118	114

4 marqueurs nucléaires :

	Séquences utilisées	Dont nouvelles
Rhodopsine (Chen, 2001)	190	92
MLL4 (Dettaï, 2004)	165	77
IRBP (Dettaï, 2004)	161	66
RNF213 (Nouveau)	118	114

- ▶ Dans une combinaison, seuls ceux qui ont plus de la moitié des marqueurs séquencés sont inclus

4 marqueurs nucléaires :

	Séquences utilisées	Dont nouvelles
Rhodopsine (Chen, 2001)	190	92
MLL4 (Dettaï, 2004)	165	77
IRBP (Dettaï, 2004)	161	66
RNF213 (Nouveau)	118	114

- ▶ Dans une combinaison, seuls ceux qui ont plus de la moitié des marqueurs séquencés sont inclus
- ▶ Assemblage semi-automatisé des matrices à analyser

Analyses primaires et secondaires

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Données

Analyses primaires et
secondaires

Synthèse des
synthèses

- ▶ MrBayes, modèle GTR + I + Γ pour toutes les combinaisons, partitionné

Analyses primaires et secondaires

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Données

Analyses primaires et
secondaires

Synthèse des
synthèses

- ▶ MrBayes, modèle GTR + I + Γ pour toutes les combinaisons, partitionné
- ▶ Dans tous les domaines de validité (W) possibles, calcul de R

Analyses primaires et secondaires

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Données

Analyses primaires et
secondaires

Synthèse des
synthèses

- ▶ MrBayes, modèle GTR + I + Γ pour toutes les combinaisons, partitionné
- ▶ Dans tous les domaines de validité (W) possibles, calcul de R
- ▶ Assemblage d'un greedy summary tree dans le plus petit W

Analyses primaires et secondaires

- ▶ MrBayes, modèle GTR + I + Γ pour toutes les combinaisons, partitionné
- ▶ Dans tous les domaines de validité (W) possibles, calcul de R
- ▶ Assemblage d'un greedy summary tree dans le plus petit W
- ▶ MRP et méthode dérivée du MinCut pour avoir des arbres avec tous les taxons

Synthèse des synthèses

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Synthèse des
synthèses

Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé

Structure générale de l'arbre

Les principaux clades

Méthode de résumé

Structure générale de
l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

Clade N restreint

Clade X

Synthèse des synthèses

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Synthèse des
synthèses

Méthode de résumé

Structure générale de
l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

Clade N restreint

Clade X

- ▶ Plusieurs arbres de synthèse

Synthèse des synthèses

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Synthèse des
synthèses

- ▶ Plusieurs arbres de synthèse
- ▶ Un avec indices mais peu de taxons

Méthode de résumé

Structure générale de
l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

Clade N restreint

Clade X

Synthèse des synthèses

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Synthèse des
synthèses

- ▶ Plusieurs arbres de synthèse
- ▶ Un avec indices mais peu de taxons
- ▶ Deux avec tous les taxons mais sans indices. . .

Méthode de résumé

Structure générale de
l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

Clade N restreint

Clade X

Synthèse des synthèses

Soutenance de
thèse

Blaise Li

Les téléostéens
acanthomorphes

Données et
analyses

Synthèse des
synthèses

- ▶ Plusieurs arbres de synthèse
- ▶ Un avec indices mais peu de taxons
- ▶ Deux avec tous les taxons mais sans indices...
- ▶ ... et des artefacts manifestes

Méthode de résumé

Structure générale de
l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

Clade N restreint

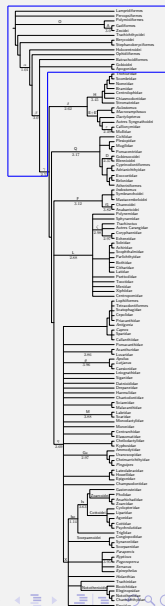
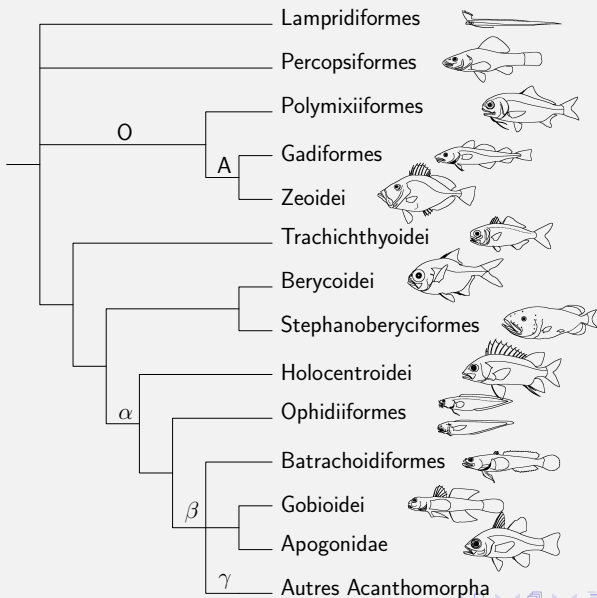
Clade X

Méthode de résumé

Clades retenus :

- ▶ présents dans le *greedy summary tree* sur l'échantillonnage réduit et ayant un indice de répétition supérieur à 1 (nouveaux avec lettres grecques)
- ▶ compatibles avec cet arbre et présents dans les deux arbres de synthèse construits sur l'échantillonnage étendu

Base de l'arbre



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

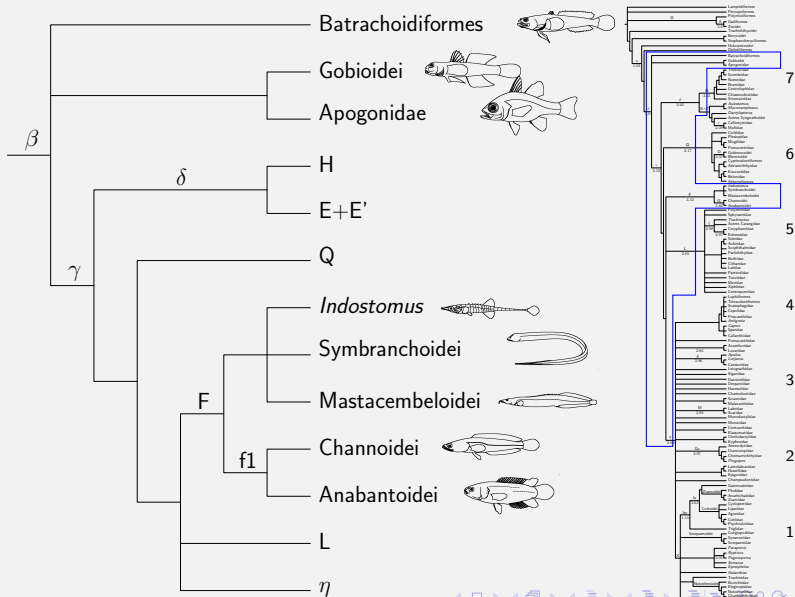
Clade L

Clade η

Clade N restreint

Clade X

Clade β



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre
Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

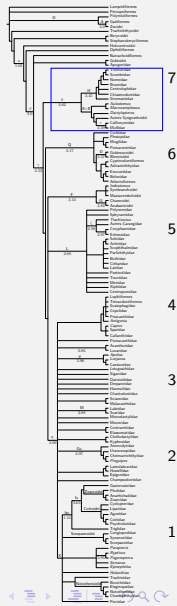
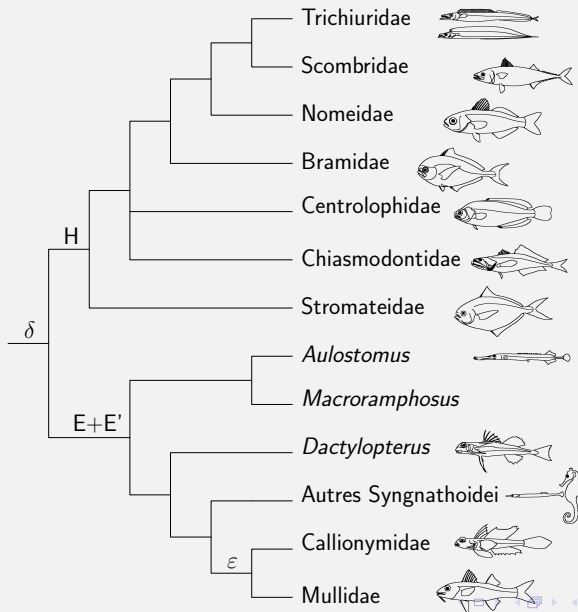
Clade N restreint

Clade X

Clade δ

Soutenance de thèse

Blaise Li



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre
Clade β

Les principaux clades

Clade δ

Clade Q

Clade L

Clade η

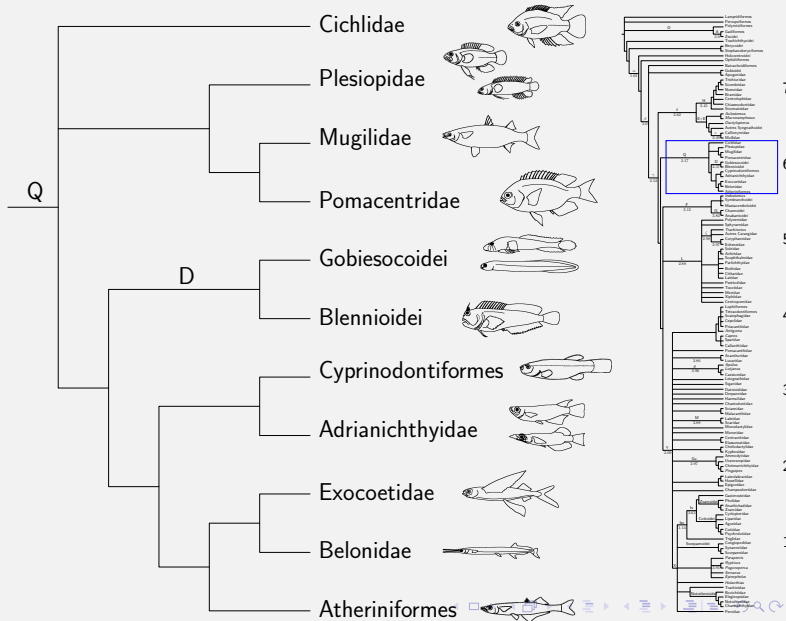
Clade N restreint

Clade X

Clade Q

Soutenance de thèse

Blaise Li



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

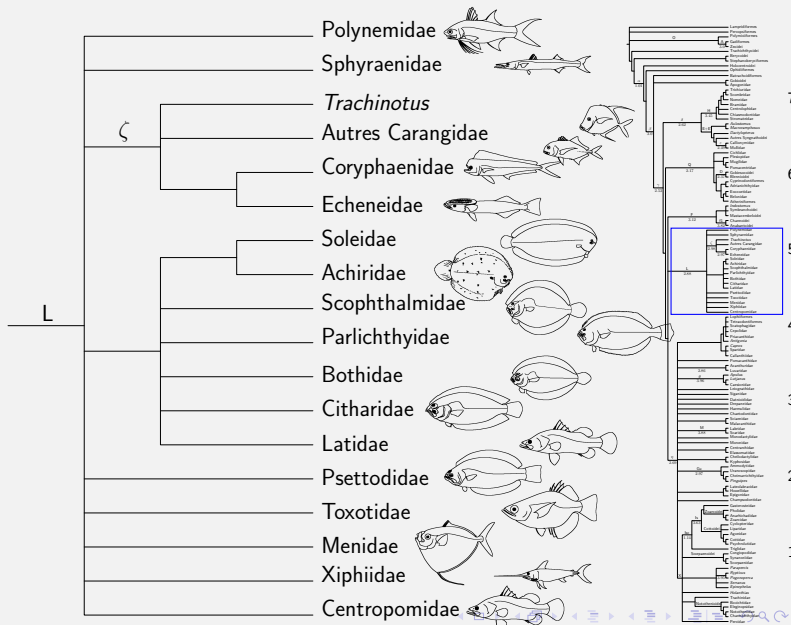
Clade L

Clade η

Clade N restreint

Clade X

Clade L



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre
Clade β

Les principaux clades
Clade δ
Clade Q

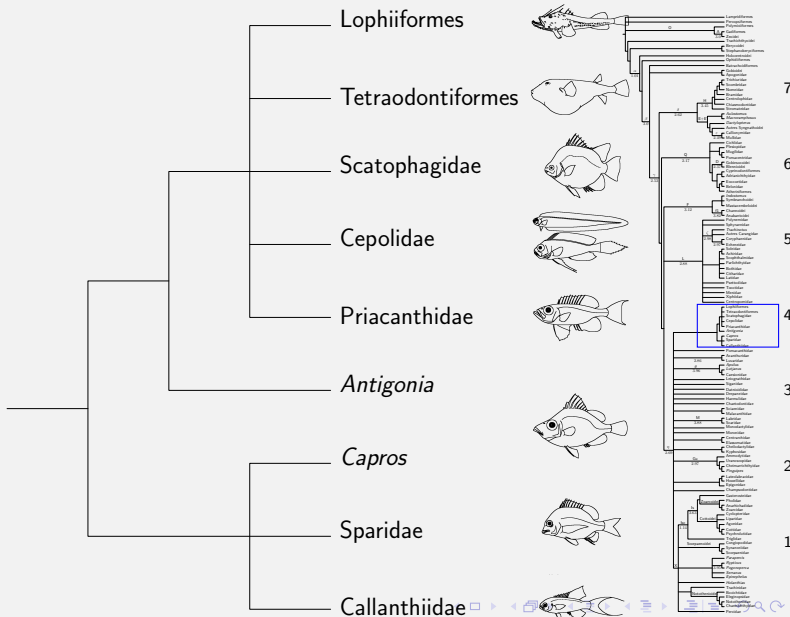
Clade L

Clade η
Clade N restreint
Clade X

Clade N restreint

Soutenance de thèse

Blaise Li



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre

Clade β

Les principaux clades

Clade δ

Clade Q

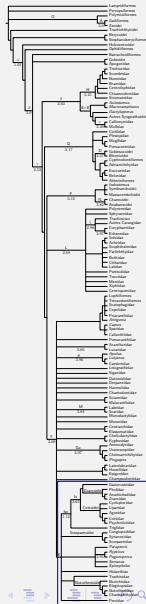
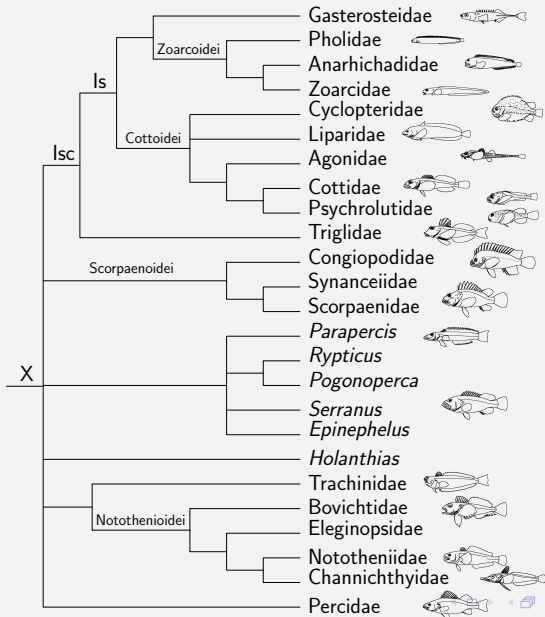
Clade L

Clade η

Clade N restreint

Clade X

Clade X



Les téléostéens acanthomorphes

Données et analyses

Synthèse des synthèses

Méthode de résumé
Structure générale de l'arbre

Base de l'arbre
Clade β
Les principaux clades

Clade δ
Clade Q
Clade L
Clade η
Clade N restreint
Clade X

Conclusion

- ▶ Indice de répétition : confirme la plupart de clades déjà estimés fiables
- ▶ Calcul automatisé (mais pas encore très commode à utiliser)
- ▶ *Greedy summary tree* : très satisfaisant dans l'échantillonnage réduit
- ▶ Autres méthodes de synthèse : quelques biais
- ▶ Manque encore une méthode pour attribuer des indices dans l'échantillonnage total

- ▶ Jeu de données étendu : plus de taxons, un nouveau gène prometteur
- ▶ Nombreuses relations confirmées (F, (Gobioidei, Apogonidae), (*Epigonus*, *Howella*, *Lateolabrax*))
- ▶ Nouveaux taxons placés
- ▶ Amélioration de la résolution dans l'arbre (β , δ , η)
- ▶ Nombreux trous à boucher
- ▶ Quelques taxons à échantillonner (Percophidae, Lactariidae, Kurtidae)
- ▶ Quelques gènes en plus ne feraient pas de mal

Merci pour votre attention

Annexe

Rappels sur la phylogénie

Soutenance de
thèse

Blaise Li

Rappels sur la phylogénie

Principe

Méthodes

Les supports des
clades

Acanthomorphes

Rappels sur la phylogénie

Principe

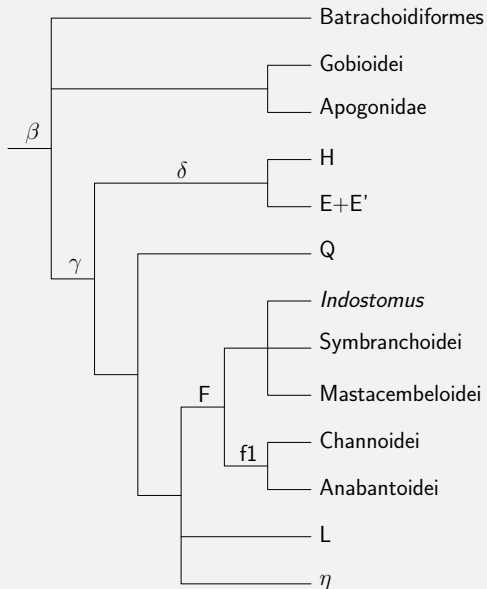
Méthodes

Les supports des clades

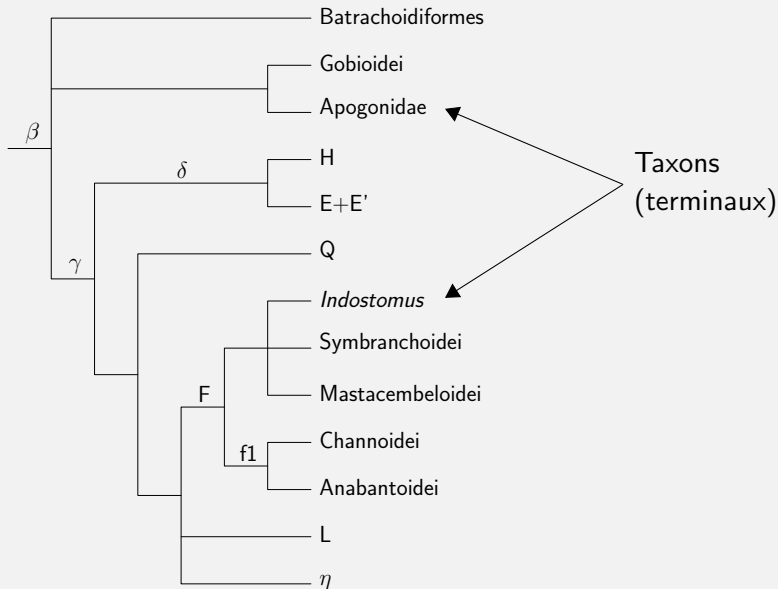
Acanthomorphes

- ▶ Les espèces sont plus ou moins apparentées.
- ▶ On cherche à savoir qui est plus proche de qui.
- ▶ On compare les caractères des espèces et on en tire des arbres.

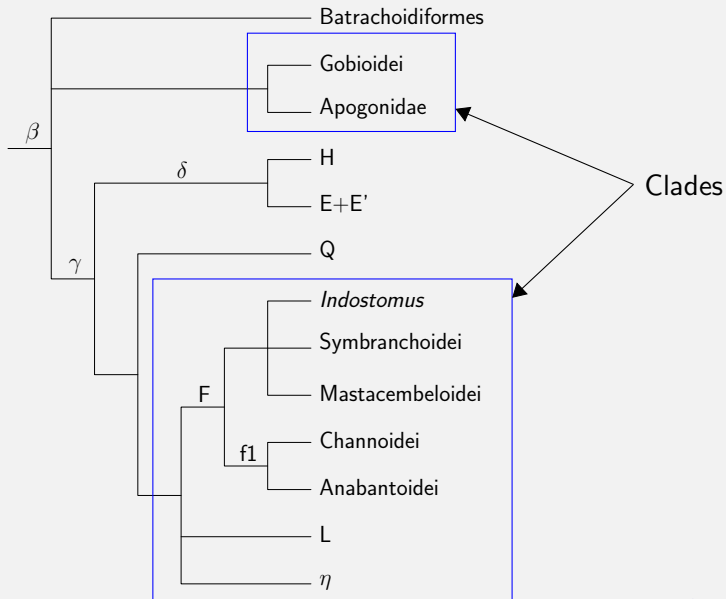
Principe



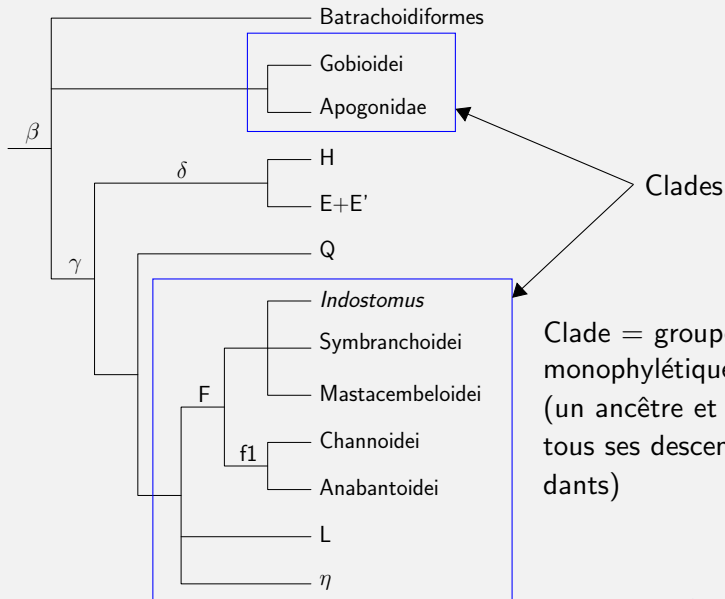
Principe



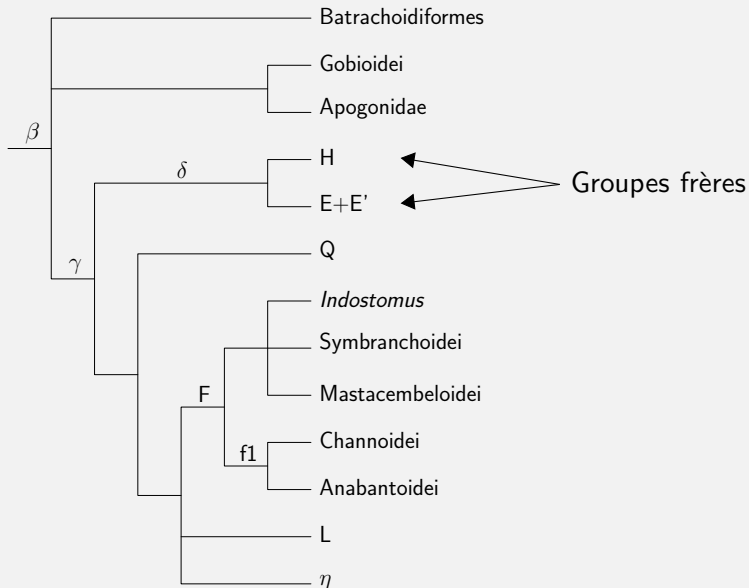
Principe



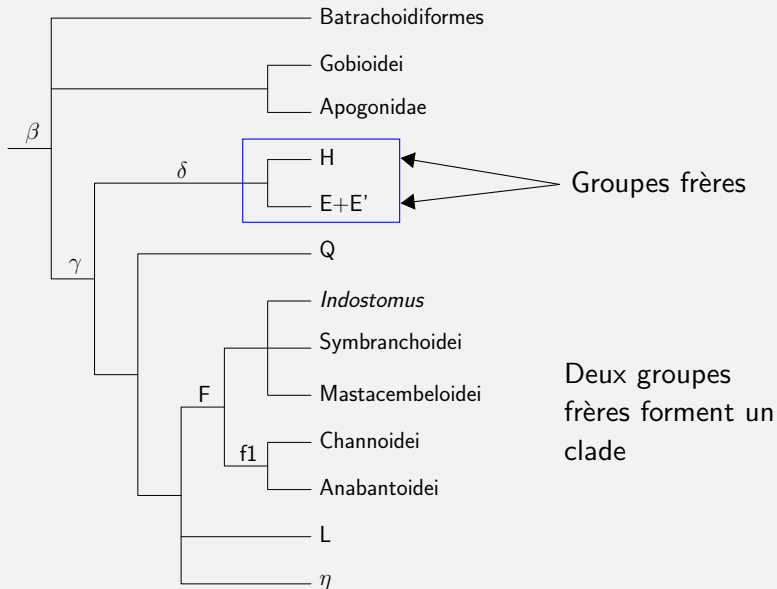
Principe



Principe



Principe



Comment passer des caractères à un arbre ?

Comment passer des caractères à un arbre ?

Distances Le moins de différences entre espèces proches

Comment passer des caractères à un arbre ?

Distances Le moins de différences entre espèces proches

Parcimonie Le moins de transformations des caractères
dans l'arbre

Comment passer des caractères à un arbre ?

Distances Le moins de différences entre espèces proches

Parcimonie Le moins de transformations des caractères dans l'arbre

Maximum de vraisemblance L'arbre offrant la plus grande probabilité des données d'après leurs modalités d'évolution supposées

Comment passer des caractères à un arbre ?

Distances Le moins de différences entre espèces proches

Parcimonie Le moins de transformations des caractères dans l'arbre

Maximum de vraisemblance L'arbre offrant la plus grande probabilité des données d'après leurs modalités d'évolution supposées

Méthode Bayésienne L'arbre le plus probable étant donné les caractères observés et leurs modalités d'évolution supposées

Les supports des clades

Soutenance de
thèse

Blaise Li

Rappels sur la
phylogénie

Les supports des
clades

Sensibilité

Robustesse

Probabilités postérieures

Acanthomorphes

Rappels sur la phylogénie

Les supports des clades

Sensibilité

Robustesse

Probabilités postérieures

Acanthomorphes

Les supports des clades

Kawahara et al.
(2008)

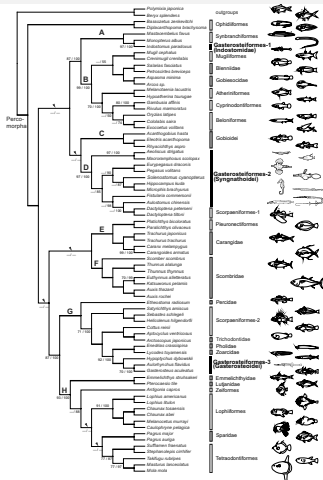


Fig. 3. Maximum likelihood tree using 121,874 data set. Topological incongruities between analyses (partitioned ML and Bayesian analyses) denoted by arrowheads. Numbers beside internal branches indicate bootstrap values (above 50%) from 100 replicates and Bayesian PP, respectively (shown as percentages). “—” indicates node not recovered in the other analysis. Thick branches supported by 100% bootstrap values and PP. Gasterosteiform orders indicated by black bars, perciform suborders by gray, non-perciform orders by white.

Rappels sur la
phylogénie

Les supports des
clades

Sensibilité

Robustesse

Probabilités postérieures

Acanthomorphes

Les supports des clades

Kawahara et al.
(2008)

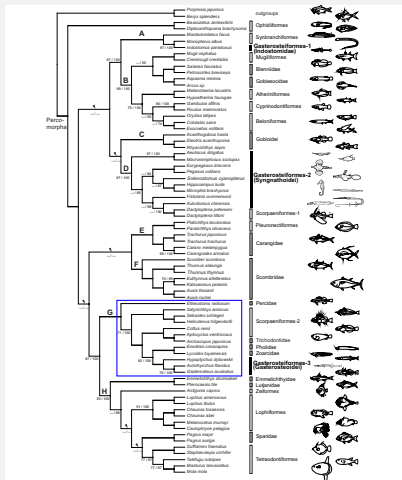


Fig. 3. Maximum likelihood tree using 121,874 data set. Topological incongruities between analyses (partitioned ML and Bayesian analyses) denoted by arrowheads. Numbers beside internal branches indicate bootstrap values (above 50%) from 100 replicates and Bayesian PP, respectively (above as percentages). “—” indicates node not recovered in the other analysis. Thick branches supported by 100% bootstrap values and PPs. Gasterosteiform fishes indicated by black bars, perciform suborders by gray, non-perciform orders by white.

Rappels sur la
phylogénie

Les supports des
clades

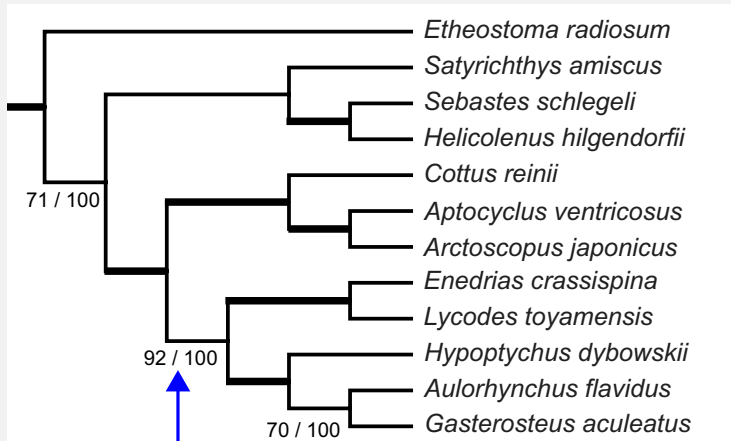
Sensibilité

Robustesse

Probabilités postérieures

Acanthomorphes

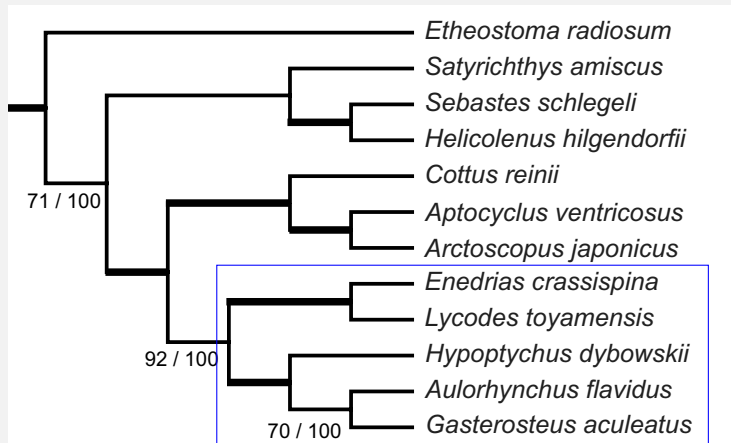
Les supports des clades



Il y a des nombres sous les branches.
Que signifient ces nombres ?

...

Les supports des clades



Le clade (Zoarcoidei, Gasterosteoidei) a une valeur de bootstrap de 92% et une probabilité postérieure de 100% : il a un bon **support**.

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)
 2. pour chaque combinaison de paramètres on voit si l'analyse donne un clade donné ou pas

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)
 2. pour chaque combinaison de paramètres on voit si l'analyse donne un clade donné ou pas
 3. 2 paramètres → petit tableau binaire (présence/absence) à placer près de la branche

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)
 2. pour chaque combinaison de paramètres on voit si l'analyse donne un clade donné ou pas
 3. 2 paramètres → petit tableau binaire (présence/absence) à placer près de la branche
- ▶ Autre exemple, l'indice de Bremer : variation du seuil d'acceptation des clades

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)
 2. pour chaque combinaison de paramètres on voit si l'analyse donne un clade donné ou pas
 3. 2 paramètres → petit tableau binaire (présence/absence) à placer près de la branche
- ▶ Autre exemple, l'indice de Bremer : variation du seuil d'acceptation des clades
 1. (en parcimonie) L : nombre de transformations dans l'arbre le plus court

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)
 2. pour chaque combinaison de paramètres on voit si l'analyse donne un clade donné ou pas
 3. 2 paramètres → petit tableau binaire (présence/absence) à placer près de la branche
- ▶ Autre exemple, l'indice de Bremer : variation du seuil d'acceptation des clades
 1. (en parcimonie) L : nombre de transformations dans l'arbre le plus court
 2. $I(\alpha)$: nombre de transformations dans le plus court arbre contredisant α

- ▶ Résister à des changements de méthode
- ▶ Exemple : « tapis navajos » :
 1. on fait varier certains paramètres (typiquement : coûts relatifs des transformations des caractères)
 2. pour chaque combinaison de paramètres on voit si l'analyse donne un clade donné ou pas
 3. 2 paramètres → petit tableau binaire (présence/absence) à placer près de la branche
- ▶ Autre exemple, l'indice de Bremer : variation du seuil d'acceptation des clades
 1. (en parcimonie) L : nombre de transformations dans l'arbre le plus court
 2. $I(\alpha)$: nombre de transformations dans le plus court arbre contredisant α
 3. indice de Bremer = $I(\alpha) - L$

- ▶ Résister à la perturbation des données
- ▶ Exemple : bootstrap et jackknife :

- ▶ Résister à la perturbation des données
- ▶ Exemple : bootstrap et jackknife :
 1. rééchantillonnage aléatoire des données

- ▶ Résister à la perturbation des données
- ▶ Exemple : bootstrap et jackknife :
 1. rééchantillonnage aléatoire des données
 2. analyses des données rééchantillonnées

- ▶ Résister à la perturbation des données
- ▶ Exemple : bootstrap et jackknife :
 1. rééchantillonnage aléatoire des données
 2. analyses des données rééchantillonnées
 3. on recommence N fois

- ▶ Résister à la perturbation des données
- ▶ Exemple : bootstrap et jackknife :
 1. rééchantillonnage aléatoire des données
 2. analyses des données rééchantillonnées
 3. on recommence N fois
 4. proportion des analyses donnant un clade donné?

- ▶ Résister à la perturbation des données
- ▶ Exemple : bootstrap et jackknife :
 1. rééchantillonnage aléatoire des données
 2. analyses des données rééchantillonnées
 3. on recommence N fois
 4. proportion des analyses donnant un clade donné?

- ▶ Sensibilité et robustesse donnent une idée de la force du signal

- ▶ Sensibilité et robustesse donnent une idée de la force du signal (quel que soit ce signal)

Probabilités postérieures

Méthode bayésienne

Soutenance de
thèse

Blaise Li

Rappels sur la
phylogénie

Les supports des
clades

Sensibilité

Robustesse

Probabilités postérieures

Acanthomorphes

Probabilités postérieures

Soutenance de
thèse

Blaise Li

Méthode bayésienne

1. Succession (chaîne de Markov) d'états :

- ▶ topologie
- ▶ longueurs de branches
- ▶ paramètres du modèle d'évolution des caractères

Rappels sur la
phylogénie

Les supports des
clades

Sensibilité
Robustesse

Probabilités postérieures

Acanthomorphes

Probabilités postérieures

Soutenance de
thèse

Blaise Li

Méthode bayésienne

1. Succession (chaîne de Markov) d'états :
 - ▶ topologie
 - ▶ longueurs de branches
 - ▶ paramètres du modèle d'évolution des caractères
2. Nouvel état proposé aléatoirement

Rappels sur la
phylogénie

Les supports des
clades

Sensibilité
Robustesse

Probabilités postérieures

Acanthomorphes

Méthode bayésienne

1. Succession (chaîne de Markov) d'états :
 - ▶ topologie
 - ▶ longueurs de branches
 - ▶ paramètres du modèle d'évolution des caractères
2. Nouvel état proposé aléatoirement
3. Acceptation d'un nouvel état :
 - ▶ oui si plus vraisemblable
 - ▶ sinon, tirage au sort selon le rapport des vraisemblances de l'ancien et du nouvel état

Méthode bayésienne

1. Succession (chaîne de Markov) d'états :
 - ▶ topologie
 - ▶ longueurs de branches
 - ▶ paramètres du modèle d'évolution des caractères
2. Nouvel état proposé aléatoirement
3. Acceptation d'un nouvel état :
 - ▶ oui si plus vraisemblable
 - ▶ sinon, tirage au sort selon le rapport des vraisemblances de l'ancien et du nouvel état
4. En théorie, à l'infini, la fréquence d'occurrence d'un clade parmi les états de la chaîne est sa probabilité :
 - ▶ sachant les données
 - ▶ d'après le modèle

Méthode bayésienne

1. Succession (chaîne de Markov) d'états :
 - ▶ topologie
 - ▶ longueurs de branches
 - ▶ paramètres du modèle d'évolution des caractères
2. Nouvel état proposé aléatoirement
3. Acceptation d'un nouvel état :
 - ▶ oui si plus vraisemblable
 - ▶ sinon, tirage au sort selon le rapport des vraisemblances de l'ancien et du nouvel état
4. En théorie, à l'infini, la fréquence d'occurrence d'un clade parmi les états de la chaîne est sa probabilité :
 - ▶ sachant les données
 - ▶ d'après le modèle

Nombreux problèmes pratiques (on ne va pas à l'infini)

Acanthomorphes

Soutenance de
thèse

Blaise Li

Rappels sur la
phylogénie

Les supports des
clades

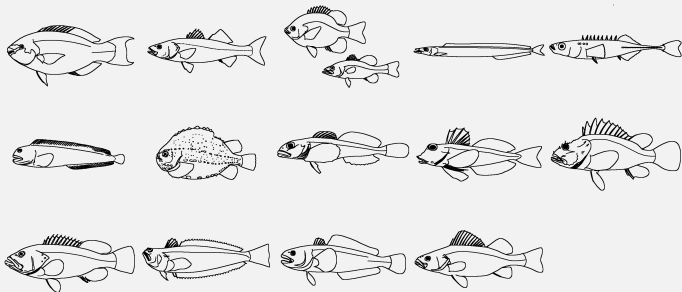
Acanthomorphes

Rappels sur la phylogénie

Les supports des clades

Acanthomorphes

Diversité



et d'autres...

- ▶ Une grande diversité

