

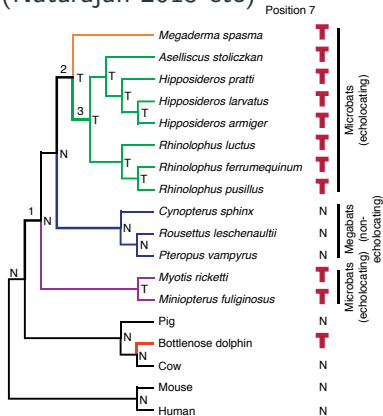
Genome-scale detection of interspecies convergent evolution: problems and prospects

Marie Sémon, Carine Rey

1^{er} juillet 2016

Do convergent evolution involve identical genetic pathways?

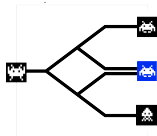
- Many cases of phenotypic convergent evolution have been described in nature
- Studies of the genetic bases of these events have mainly been based on candidate genes like Prestin (Li et al. 2010) and Hemoglobin (Natarajan 2015 etc)



Genomic screens are now proposed to search for convergent genomic changes.

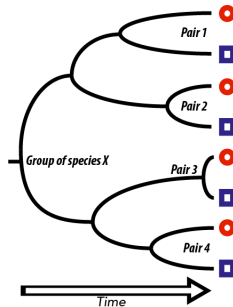
- Applied so far to systems with only two/three pairs of convergent/non-convergent species.
- Also true for convergent evolution at the level of expression (Gallant 2014; Pankey 2014)
- Heated discussions about the results (Parker 2013; Zou 2015; Thomas 2015; Zhou 2015).

- Convergenomix project



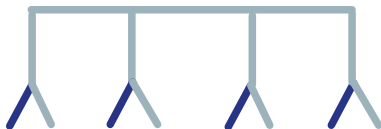
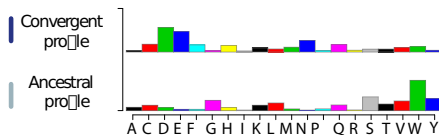
Genome-scale detection of interspecies convergent evolution

1. How many convergent events are required so that we can identify them ?
2. Can we detect convergent evolution in sequences and in expression levels/profiles ?
3. Development of new methods to analyse convergent evolution



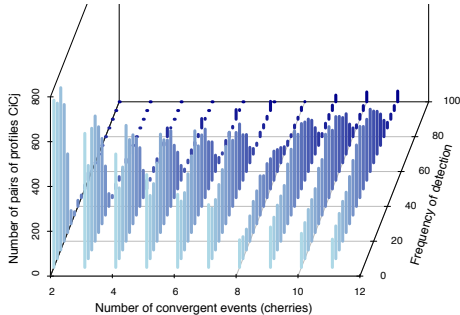
How many convergent events are required so that we can identify them ?

- Model of convergent sequence evolution that considers explicitly the phylogeny
- Convergent/ ancestral branches follow different CAT60 profiles
- Convergence detected if posterior probability > 0.8



The detection of convergent sequence evolution increases with the number of pairs

- Simulations for 2-11 pairs, and for all possible profiles
- Sufficient power for genome-scale detection of interspecies convergent evolution : about 5 pairs

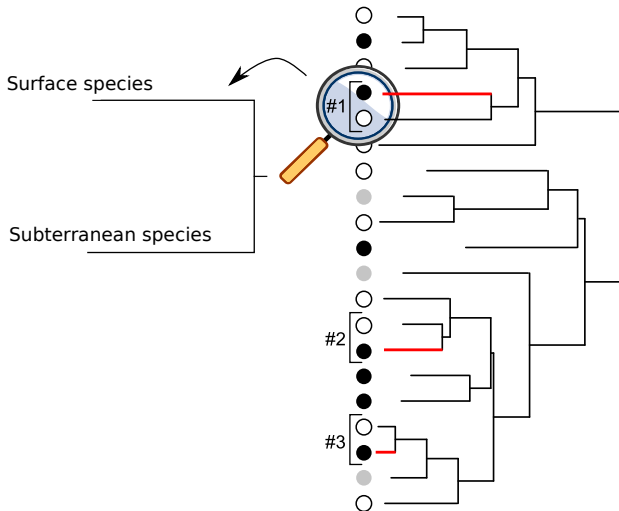


Can we detect convergent evolution in expression levels/profiles ?

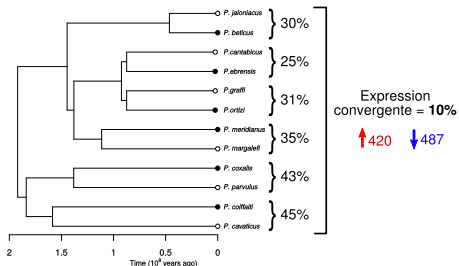
- surface/subterranean Asellidae isopods



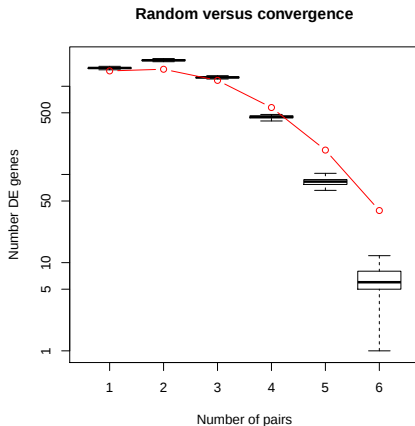
crédit photo: Le Pennec



Detecting of convergent evolution of expression also necessitates multiple pairs

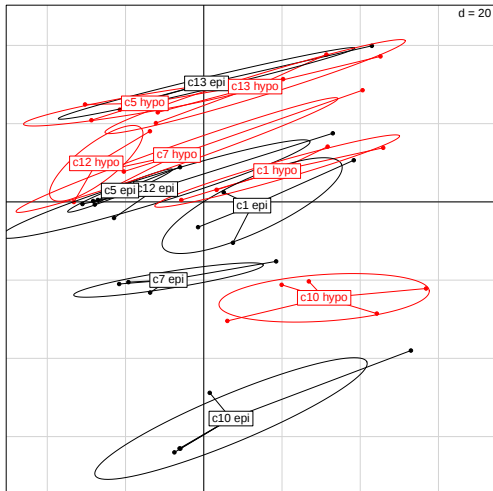


DE genes, $p < 0.1$ & FoldChange > 2



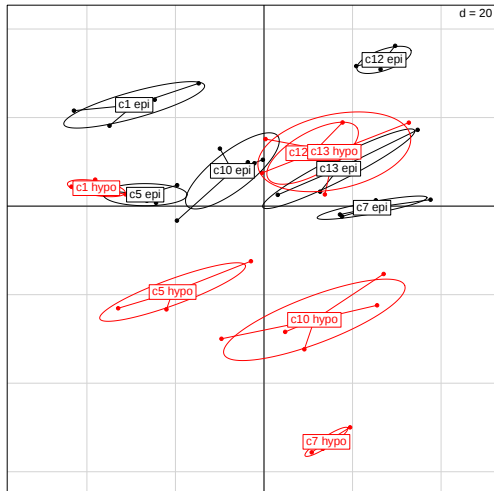
convergent / random

Footprint of convergent evolution in PCA plots



- Axis 1 (13%) : Molting (“cuticle proteins”)
- Axis 2 (12%) : Phylogeny

Footprint of convergent evolution in PCA plots



- Axe3 (9%) : Phylogeny
- Axe4 (9%) : Ecology

Interaction between divergence and ecology ?

- Effect of ecology but phylogenetic distance matters : Many species needed !
- Refine the definition of gene families + missing genes

Genome-scale detection of interspecies convergent evolution

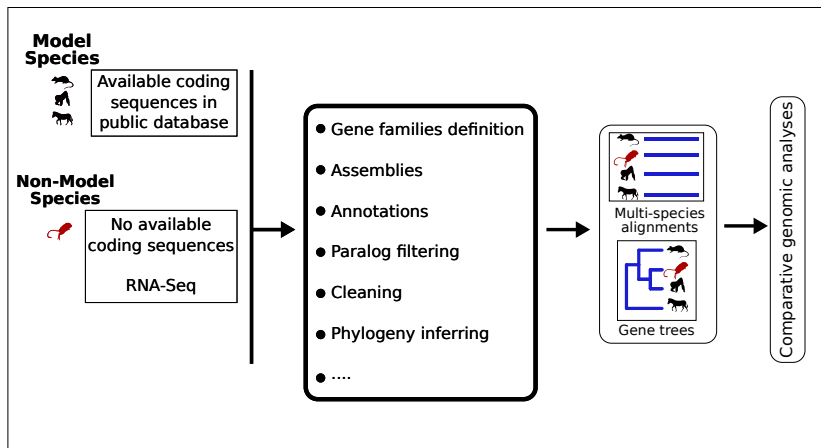
1. How many convergent events are required so that we can identify them? → ideal : at least 5 events
2. Can we detect convergent evolution in expression levels/profiles? → ongoing analysis : yes in proaselles
3. Development of new methods to analyse convergent evolution
→ Need a tool to create comparable reference transcriptomes for many species

Amalgam: an automated tool to annotate RNA-seq data using gene family alignments from other species and combine them

Carine Rey, Philippe Veber, Marie Sémon & Bastien Boussau

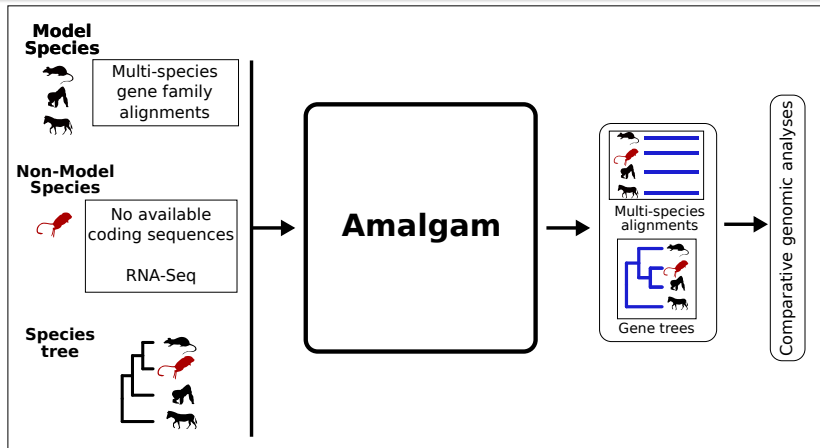
July 1, 2016

Dataset building in Comparative genomics



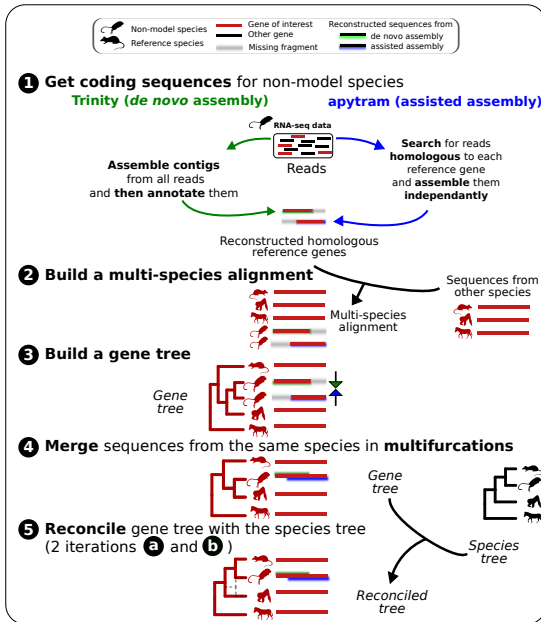
- Primordial step for subsequent analyses
- Difficult to set up and reproduce the results
- Mandatory to increase data set size

Dataset building in Comparative genomics

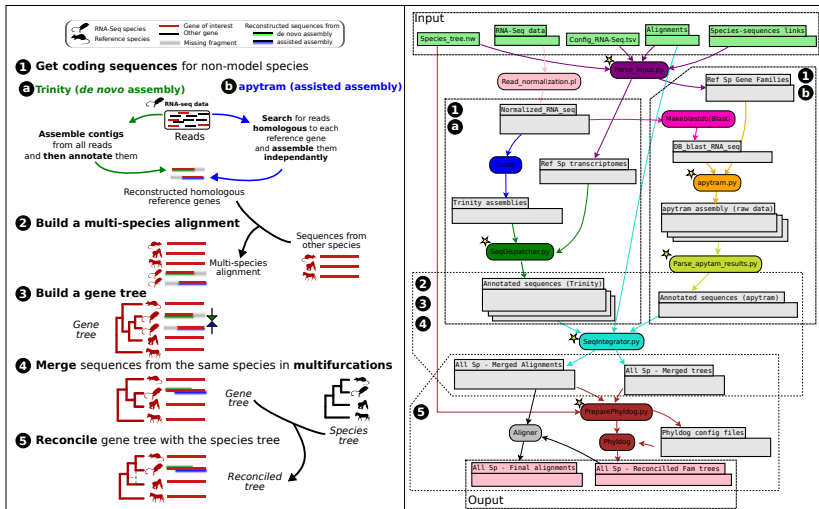


- Automated and reproducible pipeline
- Easy to use
- Relevant for large or small datasets
- Can reconstruct accurate sequences efficiently

How does Amalgam work ?



How does Amalgam work ?



a lot of modules ⇒ a lot of dependencies

Use of an unique conductor script

- manage internal dependencies between modules
- manage multi-threading and memory
- manage recovery upon error

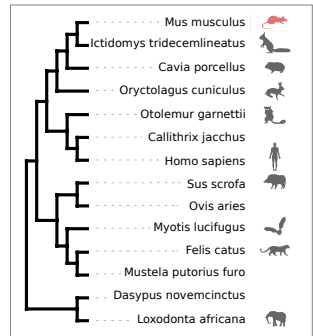
Installation is facilitated by a docker container

- no need to install external dependencies (Blast, Trinity ...)
- use on a cluster is easier
- (local installation is also possible)

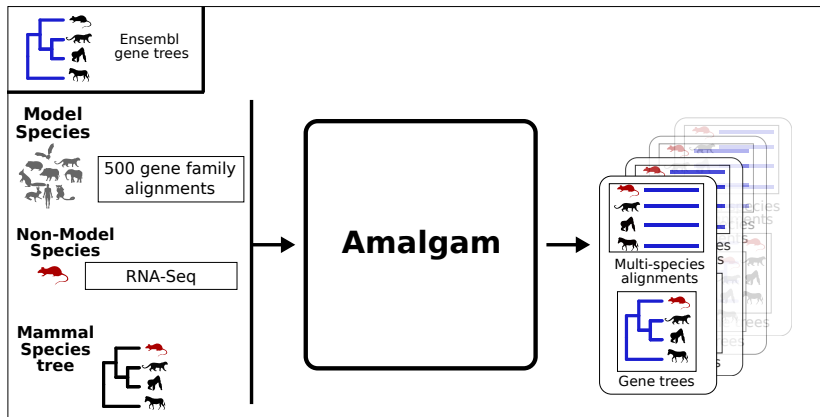


Preliminary test

- 500 random multi-species alignments from Ensembl (Compara data set)
 - ▶ 13 representative mammals
 - ▶ with Human
 - ▶ without Mouse
- RNA-seq data from an adult mouse (Kidney)
- Species tree of all selected mammals
- A reference species: Human
- (Keep aside Ensembl gene family trees with mouse)



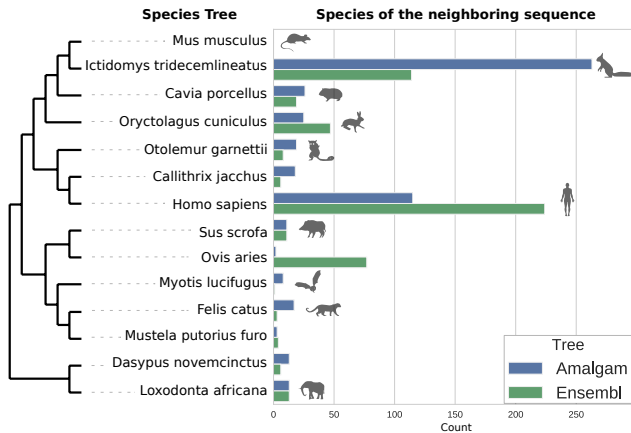
Preliminary test



The first totally automated test

- run in the docker container on an "empty" computer (8 threads, 128G RAM, Linux system)
- run in 1 day

A majority of sequences reconstructed by Amalgam are well placed in the phylogeny



First check, positions of reconstructed sequences in phylogenies

- Coherent positions according to the species tree
- Ensembl trees may not be good reference

- Amalgam increases data set size by adding new species with RNA-seq data
 - Amalgam aims to reconstruct reliable phylogenies and alignments for a large number of genes or gene families
 - Amalgam will be available in a format packaged to be easily distributed and tested.
-
- Work in progress, but close to a final stage.

Acknowledgements



Equipe CIGOGNE

Marie Sémon
Sophie Pantalacci
Coraline Petit
Luke Hayden
Lucas Michon
Marion Mouginot



Abderrahman Khila
Aidamalia Vargas
David Armisen
Thibault Lorin

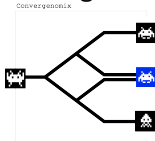


Bastien Boussau
Philippe Veber
Laurent Guéguen
Pole Info



Tristan Lefébure
Christophe Douady
Nathanaëlle Saclier

Convergenomix



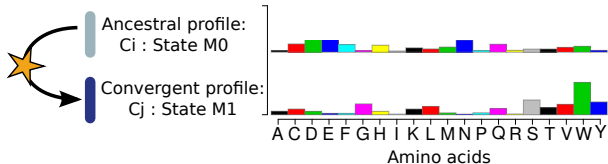
Computational facilities:



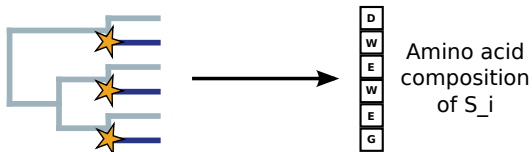
And You!

Simulations

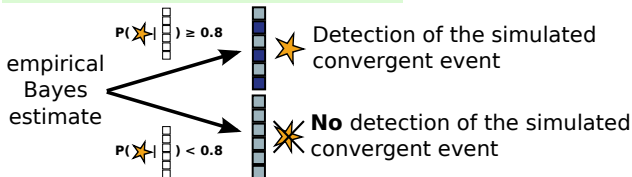
Convergent event at a site S_i :



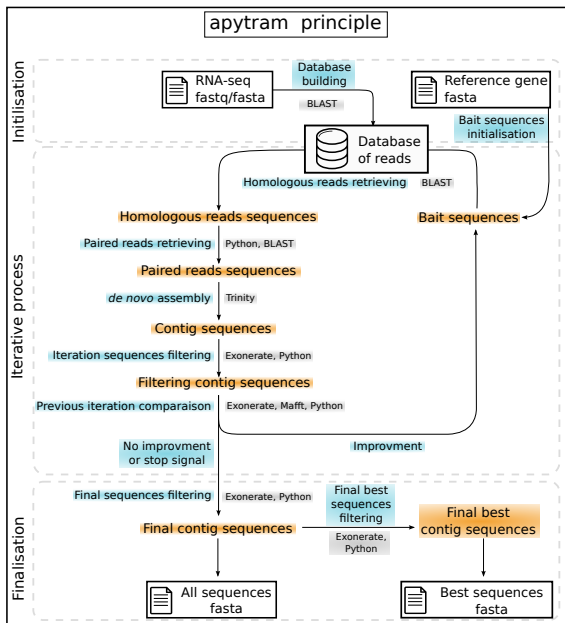
Simulation of convergent evolution for a site S_i :



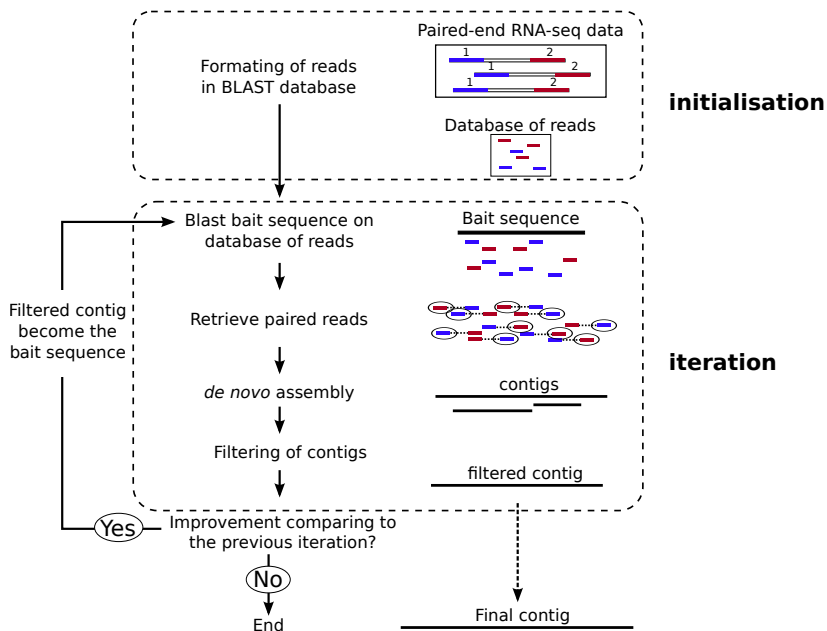
Statistical convergent detection test:



Overview - apytram



apytram iteration process



This repository

Pull requests Issues Gist

CarineRey / apytram

Unwatch 4 Star 0 Fork 1

Code Issues Pull requests **Wiki** Discourse Graphs Settings

Home

CarineRey edited this page 5 days ago · 12 revisions

apytram v1.0

Automated (PYTHON implemented) Target Restricted Assembly Method

Contact

Any question or suggestion on the program can be addressed to: carine.rey@ens-lyon.fr

Preamble

This software is inspired from aTRAM (see References) but was implemented in Python and internal strategies have been designed differently.

Why use apytram?

apytram allows the assembly of sequences from RNA-seq data (paired-end or single-end) using one or more reference homologous sequences. The reference sequences can come from the species of interest or from a different species.

Quick start

Installing apytram

To install apytram you need:

- First: to satisfy some dependencies -> [See here](#)
- Second: to download apytram -> [See here](#)

Pages

- apytram Wiki Home
- Installation
 - Get apytram
 - Check Dependencies
 - Dependencies
- How to use apytram?
 - Example usage
 - Advanced usage
 - Output files
- How does apytram work?
- Help
- apytram FAQ

Clone this wiki locally

<https://github.com/CarineRey/apytram>