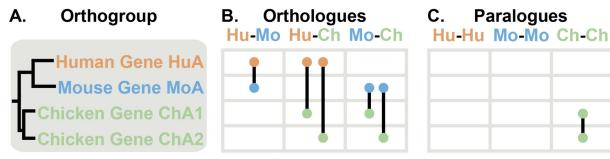
Orthogroups, Orthologs&Paralogs

Othologs 成对基因,来源两个物种距离最近的共同祖先(LCA)的一个基因。Orthogroup为orthology概 念拓展到一组物种,而不单是两个物种,为来源于一组物种距离最近的共同祖先的(LCA)一个基因的一组 基因。

例如下图中的包含三个物种的orthogroup,人/鼠/鸡。在orthogroup中人和数分呗拥有一个基因,HuA 和MoA, 然而鸡拥有两个基因, ChA1和 ChA2。人和鼠的基因(HuA/MoA)为来源两个物种最近的共同 祖先的单个基因,因此这两个基因为orthologs,从而这两个基因存在one-to-one的orthology关系。

两个鸡的基因(ChA1/ChA2),为鸡上面发生的基因复制事件。由于基因复制事件带来的paralogs, ChA1/ChA2两基因互为paralogs。然而,这两鸡的基因来源于三个物种最近的祖先的单个基因,因此, 鸡的这两个基因都和人,鼠的基因为orthologs关系,有时这些复杂的关系也指的是co-orthologs(e.g. ChA1/ChA2为HuA的co-orthologs)。在鸡和人的基因之间存在很多many-to-one的orthology关系。 这里仅存在三种orthology关系,one-to-one/many-to-one/many-to-many。所有关系都可通过 orthofinder识别。

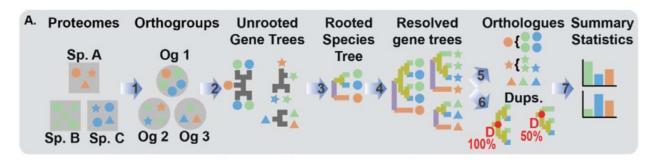


from single gene in LCA of group of species

Group of genes descended Pairs of genes descended from single gene in LCA of pair of species

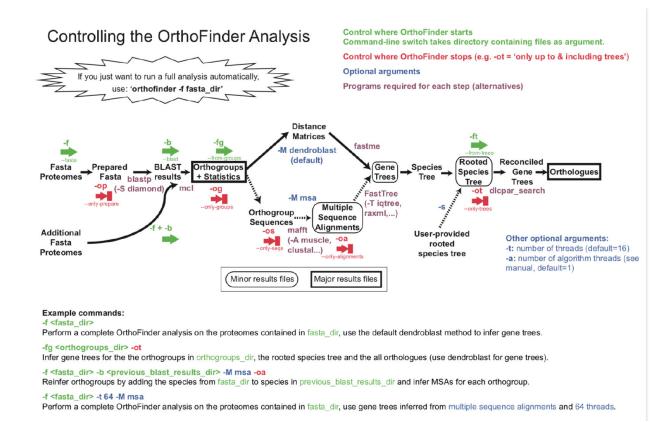
Pairs of genes descended from gene duplication event

OrthoFinder用于比较基因组研究,发现同源基因,同源基因群体,推导同源基因进化树,识别基因拷 贝事件。使用简单,输入单个物种的一系列蛋白fasta序列文件即可。



第一步,orthogroup检出,提供物种的fasta格式文件,每个文件包含了来自该物种的蛋白序列,通过 原始的orthofinder算法推导出orhtogroup;第二步,推导rooted species和gene trees,首先是用 STAG算法从unrooted gene trees中推导出species tree,然后使用STRIDE算法推导出rooted species tree,然后使用rooted species tree推导rooted gene trees;第三步,从rooted gene trees推导 orhologs和gene duplication事件,统计最终结果(speciation and duplication events are inferred from the rooted gene trees by a restriction of the Duplication-Loss-Coalescent model to apply to the problem of ortholog inference using the 'overlap' method).

Instruction



运行目录dir中的fasta格式蛋白

orthofinder.py [options] -f <dir>

增加目录中新的物种到之前运行到中的物种,然后重新运行

orthofinder.py [options] -f <dirl> -b <dir2>

从之前分析中去除物种,在之前分析目录"WorkingDirectory/"中的"SpeciesIDs.txt"文件中,将需要去除的物种前加上"#",表示去除,然后运行

orthofinder.py -b previous_orthofinder_dirctory

以上两步骤可同时进行,添加同时删除,满足以上准备步骤

orthofinder.py -b previous_orthofinder_directory -f new_fasta_dirctory

略!!!