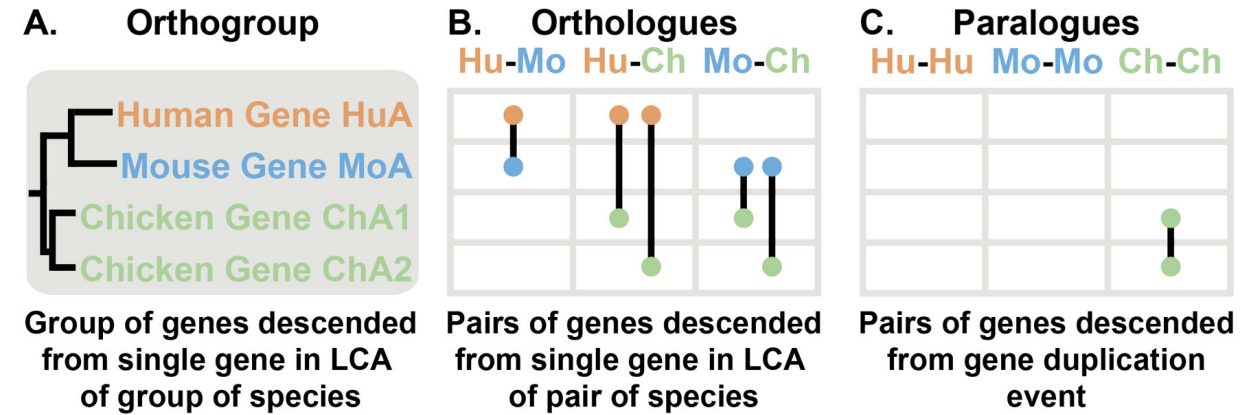


Orthogroups, Orthologs&Paralogs

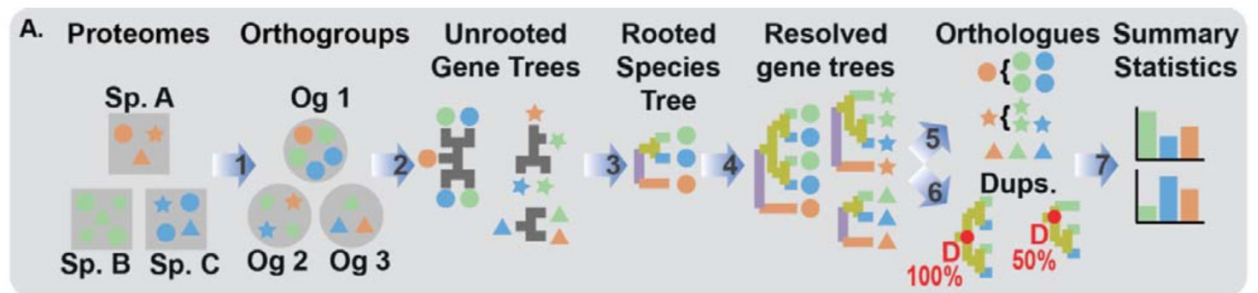
Othologs 成对基因，来源两个物种距离最近共同祖先(LCA)的一个基因。Orthogroup为orthology概念拓展到一组物种，而不单是两个物种，为来源于一组物种距离最近共同祖先(LCA)一个基因的一组基因。

例如下图中的包含三个物种的orthogroup，人/鼠/鸡。在orthogroup中人和鼠拥有一个基因，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。



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orthogroup；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orthologs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)。

