

Б.Н. Соболев: Известны ситуации, когда классы ингибиторной специфичности плохо соотносятся с филогенетическими таксонами. Насколько я знаю, у вас представляется возможность произвольного разбиения выровненных семейств на функциональные подгруппы независимо от филогенетических построений. Но и в этом случае возможна “переоценка позиций”, консервативных у более близких гомологов, попавших в некоторую подгруппу, которая содержит и более отдаленные гомологи. Предусмотрена ли возможность подавления такого “эволюционного следа», наподобие того, как это делается при взвешивании аминокислотных последовательностей при расчете НММ-профилей.

В.-Ю.К. Швядас: Буквально на днях было опубликовано обновление нашего метода поиска специфических позиций - <https://doi.org/10.1093/nar/gkaa276>. В нем предусмотрено три способа классификации белков на подсемейства с разными свойствами: (1) автоматическая, которая предлагает несколько вариантов "разбиения" выборки родственных белков на функциональные подсемейства на основе их кластеризации на разных уровнях сходства, которые выбираются таким образом, чтобы предложенные классификации принципиально различались и отражали разные уровни функционального разнообразия в суперсемействе; (2) автоматическая, на основе алгоритма кластеризации CD-HIT, т.е. посредством «разрезания» графа парной идентичности последовательностей на выбранном пользователем уровне; (3) вручную, т.е. классификация на подсемейства осуществляется пользователем, например, на основе экспериментальных/литературных данных по функциональным свойствам. Таким образом, отвечая непосредственно на Ваш вопрос, в тех случаях, когда функциональная классификация гомологов известна и не соответствует их филогении, имеет смысл применять способ 3. Это позволит использовать наш сервис для идентификации и приоритетного ранжирования ключевых позиций, ответственных за эти известные из литературы (или собственного эксперимента) функциональные различия. При этом будет доступен ключевой функционал веб-метода, который заключается в представлении результатов онлайн с использованием интерактивных инструментов на различных уровнях визуализации: посредством "сетей схожести последовательностей" (от англ. "sequence similarity network") для анализа классификаций; интерфейсов для анализа ключевых позиций с использованием информации о последовательностях и репрезентативной трехмерной структуры, и сопровождается подробной аннотацией белков, собранных из интегрированных баз данных с ссылками на внешние ресурсы (PDB, UniProt, BRENDA, BacDive).

Б.Н. Соболев: Для ваших методов очень важно подготовить качественное выравнивание со строгой фиксацией тех позиций, которые будут определены как функционально значимые. Но как быть в случае сильно дивергировавшего семейства, когда выравнивание буквально “расползается”. Как это может повлиять на распознавание аллостерических центров.

В.-Ю.К. Швядас: Для поиска гомологов в суперсемействе и подготовки выравнивания у нас есть специальный инструмент - Mustguseal (<https://doi.org/10.1093/bioinformatics/btx831>). Этот веб-сервер позволяет автоматически строить множественные выравнивания тысяч эволюционно родственных, но далеких белков, которые характеризуются функциональным разнообразием и, как правило, сильно различаются по аминокислотным последовательностям. Для точного выравнивания таких далеких гомологов Mustguseal в явном виде использует 3D-структурную информацию, поскольку изменение пространственной организации белков более консервативно в эволюции, чем аминокислотные последовательности.