

УДК: 004.6:57

БАЗЫ ДАННЫХ ДЛЯ БИОЛОГИИ

*Дидярова Е.В., студент 1 курса факультета ветеринарной
медицины и биотехнологии
Научный руководитель – Видеркер М.А., кандидат
биологических наук
ФГБОУ ВО Ульяновский ГАУ*

Ключевые слова: *биология, база данных, информационные технологии, информация, поисковая система.*

Работа посвящена описанию баз данных, которые используются специалистами, чья профессиональная деятельность связана с биологическими исследованиями. Рассмотрен ряд библиографических, молекулярно-биологических, экологических баз данных. Особенное внимание уделено таким базам данных и инструментам работы с ними, как MedLine, GenBank, Entrez, PubMed и некоторым другим.

Различные отрасли современной фундаментальной биологии (генетика, эволюционная биология, вычислительная биология и т.д.), а также другие области научной и практической деятельности, связанные с биологическими исследованиями (например, медицина, ветеринария, агрономия), имеют дело с большими объемами многообразных данных. Эти отрасли нуждаются в информационных технологиях для организации хранения и обработки массивов биологической информации.

Одной из таких технологий является технология баз данных. Базы данных представляют собой информационные модели, содержащие данные об объектах и их свойствах. Они хранят информацию о группах объектов с одинаковыми наборами свойств.

Современные базы данных по биологии можно классифицировать по содержащейся в них информации следующим образом: библиографические, молекулярно-биологические, экологические, прочие тематические.

В библиографических базах данных содержится информация о статьях, публикуемых в реферируемых журналах. Эта информация обычно включает в себя название статьи, имена авторов, название, год, выпуск и страницы журнала, ключевые слова, а также резюме статьи в

том виде, как оно было напечатано в журнале.

Примерами библиографических баз данных являются MedLine и pre-MedLine, Chemical Abstracts, AGRICOLA, AIDSLINE, CANCERLIT.

Для многих биологов первым и наиболее удобным источником библиографической информации является MedLine. В настоящее время она является полностью бесплатной. Medline разработана Национальной медицинской библиотекой. Охватывает около 75 % мировых медицинских изданий. Использует словарь MeSH. База содержит более 18,8 млн записей о публикациях с 1950 года и до настоящего времени.

К молекулярно-биологическим базам данных относятся: базы данных первичных последовательностей ДНК, РНК (EMBL, GenBank, DDBJ), аминокислотных последовательностей (UniProt, PIR, PRF, PDB), пространственной структуры молекул (MMDB, PDB, NDB), геномные базы данных (OMIA, GDB, MGD, ARKbd, ECDC).

К примеру, GenBank – база данных, находящаяся в открытом доступе, содержащая все аннотированные последовательности ДНК и РНК, а также последовательности закодированных в них белков. GenBank поддерживается Национальным центром биотехнологической информации США (NCBI) и доступен на бесплатной основе исследователям всего мира. GenBank получает и объединяет данные, полученные в разных лабораториях, для более чем 100 000 различных организмов [1].

Геномная база OMIA Online Mendelian Inheritance In Animals) – содержит информацию по геномам около полутора сотен видов животных. GDB (The Genome Database) является базой данных человеческого генома. В MGD (Mouse Genome Database) содержится информация о геноме мыши. ARKbd – семейство геномных баз данных сельскохозяйственных животных.

Основной тенденцией в развитии современных биологических баз данных является стремление к их интеграции, созданию перекрестных ссылок между ними (например, между библиографической ссылкой на статью, описывающую какой-то ген, и собственно нуклеотидной последовательностью этого гена). Такой подход позволяет быстро находить и получать исчерпывающую информацию по определенной теме.

Среди интерфейсов для работы с такими интегрированными базами данных наиболее известны, например, Entrez, PubMed, SRS.

WWW Entrez retrieval system – это поисковая система, созданная в Национальном центре биотехнологической информации и предназначенная для работы с интегрированными базами данных, находящимися

в нем. Network Entrez устанавливается на локальном компьютере и обеспечивает те же функции, что и WWW-версия.

PubMed – это поисковая система, также разработанная в NCBI. PubMed предоставляет бесплатный доступ к базам MedLine, NLM, где содержится более 19 миллионов библиографических ссылок и тезисов в сфере медицины, сестринского дела, стоматологии, ветеринарии, здравоохранения и доклинических дисциплин. Включает доступ к дополнительно отобранным журналам о естественных науках, не включенным в MedLine. Содержит ссылки на полные тексты статей в электронных журналах, на которые имеется подписка в Lane Library. Содержит ссылки на все базы данных NCBI, в том числе к новой базе «Images» («Изображения») [1].

Среди экологических баз данных можно назвать «Флору сосудистых растений Центральной России», «Качество среды обитания человека: атмосфера, вода, почва», HYSPLIT и другие [2].

Таким образом, в настоящее время для организации хранения и обработки массивов биологической информации разработаны базы данных, которые могут быть использованы для биологических исследований.

Библиографический список:

1. Мильчаков, К. С. PUBMED: описание системы и краткое руководство по пользованию / К. С. Мильчаков, Е. М. Леднев // Литобзор. – URL : https://lit-review.ru/materialy_po_litobzoram/pubmed-opisanie-i-kratkoe-rukovodstvo-po-ispolzovaniyu/.
2. Молородов, Ю. И. Электронные библиотеки и базы данных для задач экологии / Ю. И. Молородов, А. М. Федотов // Электронные библиотеки: перспективные методы и технологии, электронные коллекции : труды 9-й Всероссийской научной конференции. – Переславль-Залесский : Университет города Переславля, 2007. – Т. 2. – С. 14–16.

DATABASES FOR BIOLOGY

Didyarova E.V.

Key words: *biology, database, information technology, information, search engine.*

The work is devoted to the description of databases that are used by specialists whose professional activities are related to biological research. A number of bibliographic, molecular biological, environmental databases are considered. Particular attention is paid to databases and tools for working with them, such as MedLine, GenBank, Entrez, PubMed and some others.