

# БИОИНФОРМАТИКА

**ПРОФ. ПЛАМЕНКА БОРОВСКА**

**КАТЕДРА КОМПЮТЪРНИ СИСТЕМИ  
ТЕХНИЧЕСКИ УНИВЕРСИТЕТ - СОФИЯ**

1



# НАУКИ ЗА ЖИВОТА (LIFE SCIENCE)

- Биоинформатиката принадлежи към “науките за живота” – “life science”
- Науките за живота обхващат изследване на живите организми, като растенията, животните и човешките същества.
- Докато биологията остава в центъра на науките за живота, технологичния напредък в молекулярната биология и биотехнологиите, е довел създаването на нови, често интердисциплинарни, области на изследвания.
- Науки за живота – напр. медикобиологични науки, биохимия, биокомпютри, биоуправление, биодинамика, биоинформатика, изчислителна (компютърна) неврологията, екология, науки за околната среда, еволюционна биология, еволюционна генетика, генетика, геномика, здравни науки, имуногенетика, имунология, имунотерапия, медицински прибори, медицински изображения, микробиология, молекулярна биология и мн. др.

# БИОИНФОРМАТИКА

- Биоинформатика се определя като *интердисциплинарна област*, включваща *биология, компютърни науки, математика и статистика* за анализ на биологични данни, съдържание и структура на геноми, и за предсказване на функциите и структурата на макромолекули.
- *С настъпването на ерата на геномната революция*, биоинформатиката придобива изключителна роля в биологичните и медицинските изследвания.

## Исторически аспекти

- Преди ерата на биоинформатиката съществуваха само два начина за биологични експерименти:
  1. *В жив организъм (in vivo) или*
  2. *В изкуствена среда (in vitro), от латински в стъкло).*
- *Като продължим с аналозите, може да се каже, че биоинформатиката представлява **in silico biology**, т.е. биологичните експерименти се рализират чрез компютърни модели, които се симулират в силиконови чипове.*

# БИОИНФОРМАТИКА

- *Биоинформатиката е интердисциплинарна наука и обхваща две припокриващи се изследователски области - информатиката и компютърната биология (computational biology).*
- Биоинформатиците участват в дейности, подпомагащи управлението на информацията (information management) в области като здравеопазване, фармакология, селското стопанство и други търговски приложения.

# КАКВО ПРЕДСТАВЛЯВА БИОИНФОРМАТИКАТА ?

- *Биоинформатика* се превърна в значима мултидисциплинарна наука, която интегрира развитието на информационните и компютърните технологии, които се прилагат при биотехнологиите и биологичните науки.
- Биоинформатика използва компютърни софтуерни инструменти за създаването на бази данни, управление на данните (data management), съхраняване на данни (data warehousing), data mining, и мрежови комуникационни технологии в световен мащаб.
- Биоинформатиката обхваща записа, аотиране, съхранение, анализ и търсене в нуклетидни секвенции (последователности - (ДНК и РНК), протеинови секвенции и структурна информация. Това включва базите данни на секвенциите и структурна информация, както и методи за достъп, търсене, визуализиране и анализ на болгична информация.

# КАКВО ПРЕДСТАВЛЯВА БИОИНФОРМАТИКАТА ?

- Биоинформатика се занимава със създаването и поддържането на бази данни биологична информация, като изследователите могат да осъществяват достъп и да използват съществуващата информация, както и да добавят нови данни.
- Научни области, в които биоинформатиката е интегрална част, са функционалната геномика, биомолекулярната структура, анализ на протеома (протеин+геном), клетъчния метаболизъм, биоразнообразието, химическото инженерство, дизайн на лекарствени средства и ваксини.

# КАКВО ПРЕДСТАВЛЯВА БИОИНФОРМАТИКАТА ?

- Биоинформатиката е използването на ИТ в областта на биотехнологиите за съхранение на данни, data warehousing, и анализ на ДНК секвенции.
- В биоинформатиката са необходими знания от много клонове на науката като биология, математика, компютърни науки, законите на физиката и химията, както и солидни знания в ИТ за да се анализират биотехнологичните данни.
- Биоинформатика не се ограничава само до обработката на компютърни модели на биологичните данни, но в действителност тя може да се използва за решаване на много биологични проблеми, вкл. и да се разбере същността на биологичните феномени.



# ПОД-ДИСЦИПЛИНИ В РАМКИТЕ НА БИОИНФОРМАТИКАТА

Има три важни под-дисциплини в рамките на биоинформатиката, включващи изчислителната (компютърната) биология:

1. Синтеза на нови алгоритми и статистика, с които да се оценят взаимоотношенията между компонентите на големи множества биологични данни.
2. Анализ и интерпретация на различни типове данни вкл. нуклеотидни и протеинови секвенции, домейни в протеините, и протеинови структури
3. Създаването и имплементирането на инструменти, които осигуряват ефективен достъп и мениджмънт на различни типове информация

# ДЕЙНОСТИ В БИОИНФОРМАТИКАТА

Дейностите могат да се разделят ба две големи области:

(1) Организация, и (2) анализ на биологични данни

## *Организационни дейности в биоинформатиката*

- Създаване на бази данни от биологична информация
- Поддръжката на тези бази данни

## *Дейности по анализа в биоинформатиката*

- Разработване на методи за предсказване на структурата и/или функциите на новооткрити протеини и структурирани РНК секвенции.
- Клъстериране на протеинови секвенции в семейства от свързани секвенции и разработването на протеинови модели.
- Изравняване (Aligning) на подобни протеини и генериране на филогенетични дървета за изследване на еволюционни взаимосвързки

# ЦЕЛИ НА БИОИНФОРМАТИКАТА

- Организация на биологични данни по-такъв начин, че да се осигури лесен достъп до биологичните данни, както и лесно да се добавят нови данни.
- Организацията и съхраняването на биологични данни е важна задача, но тя се обезмисля, ако информацията, съхранена в биологичните бази данни не може да бъде анализирана.

# ЦЕЛИ НА БИОИНФОРМАТИКАТА

- *Разработване на инструменти и създаването на ресурси, които да помагат в анализа на биологични данни.*
- Напр., ако даден протеин е секвениран, секвенцията на този протеин е съхранена в базата данни. Простото търсене на тази протеинова секвенция в базата данни не е достатъчно. Съществуват програми като FASTA и BLAST, които могат да определят подобие на протеини в бази данни от протеинови секвенции.
- Създаването на такива ресурси изисква дълбоки познания по компютърни науки, както и добро познаване на биологията.
- Софтуерните инструменти дават възможност за задълбочен анализ на индивидуалните системи, както и взаимотношенията им с други такива.



# ОБХВАТА НА БИОИНФОРМАТИКАТА

- Биоинформатика се занимава със създаването на база данни, анализ на тези данни и моделиране.
- Базите данни в биологията като цяло са в мултимедийна форма, организирани в релационни база данни.
- Моделиране се извършва не само на една-единствена биологична молекула, но също и на множество взаимодействащи сис системи, като по този начин, изискват използването на високи производителни изчислителни системи.

# ОБХВАТА НА БИОИНФОРМАТИКАТА

- Потенциалът на биоинформатика в идентифициране на полезни гени, водещи до създаването на нови генни продукти и за откриване на лекарства и на ваксини е довело до промяна на парадигмата в биологията и биотехнологиите - тези области изискват все по- интензивни и интелигентни изчисления.
- *Новата съвременна парадигма е, че всички известни гени ще бъдат съхранявани в електронен вид, в бази данни, като отправната точка за биологични изследвания ще бъде теоретично изследване, като се започне със създаването на хипотеза, която впоследствие се верифицира експериментално.*
- С придобиването на много по-дълбоко разбиране на биологичните процеси на молекулярно ниво, биоинформатиците са създали нови техники за анализ на гените в индустриален мащаб, в резултат на което се създава нова научна област, наречена *геномика*.



# ПРИЛОЖЕНИЯ НА БИОИНФОРМАТИКАТА

- Изследванията в областите на протеомиката и геномиката ще продължат със сложни и интелигентни софтуерни инструменти и компютърни мрежи, фокусирани върху съхраняването и трансфера на данни от *мултимедийни бази данни*.
- Понастоящем съществува необходимост от създаването на софтуерни инструменти за data mining, анализ и моделиране, както и поточно програмиране.
- Сигурността на данните, трансфера на данните, компресията и автоматичната проверка на коректността на биологичните данни, са важни аспекти на биоинформатиката.

# ПРИЛОЖЕНИЯ НА БИОИНФОРМАТИКАТА

- *Използването на виртуална реалност при синтеза на лекарствени средства (*virtual reality in drug design*), дизайна на метаболитния път, дизайна на едноклетъчни организми, синтеза и промяната на многоклетъчни организми, са предизвикателствата, които биоинформатиците трябва да посрещнат.*
- Понастоящем е всеобщо признат фактът, че биоинформатиката е ключов фактор за развитието на молекулярната биология през 21 век
- *Bioinformatics is the key to the new grand data-intensive molecular biology in the 21<sup>st</sup> century.*



# БИОИНФОРМАТИЧНА ИНДУСТРИЯ

- Биоинформатичната индустрия се развива бурно паралелно и в синхрон с информационния бум, като темповете на “експлозия” са 25-50% годишно.
- Според оценката на копманията за изследване на пазара Oscar Gruss (САЩ) стойността на индустрията, свързана с биоинформатиката, се оценява на повече от 2 милиарда долара с устойчива тенденция за нарастване.
- Биоинформатиката е нова гореща тема, след софтуера.
- Съществува голямо търсене на специалисти по биоинформатика в секторите на биотехнологиите, фармацевтичната и биомедицинската индустрия.

# БИОИНФОРМАТИЧНА ИНДУСТРИЯ

- Според “The Tribuen” "В глобален мащаб, биотехнологичният компютърен сектор е достигнал невероятните \$ 30 млрд. през 2003 г. и \$ 60 млрд. през 2005г.“
- Сега има голямо търсене на специалисти по биоинформатика в световен мащаб.

# ПРОФИЛ НА ПРОФЕСИЯТА БИОИНФОРМАТИК

- Понастоящем търсенето на специалисти по биоинформатика е свързано главно с проектирането и имплементацията на софтуерни системи (Bioinformatics Systems) за data ware housing и анализ на геномни секвенции, протеинови структури и др.
- Професията биоинформатик обхваща Data Mining, ДВА и развитие на системи за Diagnostic kits, Bioinformatics software, Протеомика (структури и функции на протеините) & Геномика (представяне и функции на гените), публикуване на биотехнологични данни & научни статии и изследвания в web.

# ВАЖНИ ТЕМИ В БИОИНФОРМАТИКАТА

Биологичен аспект - Молекулярна биология

- Централна догма на молекулярната биология (Central Dogma of molecular biology)
- Опит с един или повече софтуерни пакети за молекулярната биология
- Използване на софтуер за анализ на биологични секвенции и молекулярно моделиране
- Софтуерни пакети за молекулярна биология BLAST, FASTA, и др.
- Знания и умения за работа под Linux (Free open source) – екстензивно се използва в биотехнологиите заради robustness и наличните инструменти& софтуер за тази платформа
- Езици за програмиране като C/C++, Perl или Python, Java и HTML
- Database Management Systems
- Oracle и MySQL (Free Database Server) – екстензивно (широко) се използват за съхраняване на гигабайти биологични данни за по-нататъшен анализ



# БИОИНФОРМАТИКАТА И КОМПЮТЪРНАТА БИОЛОГИЯ

- Биоинформатиката и компютърната биология (computational biology) поддържат тесни взаимодействия с науките за живота за да реализират пълния си потенциал.
- Биоинформатиката прилага принципите на информационните науки и технологии за да направи по-разбираеми и по-полезни широкомащабните, разнообразни и сложни данни от науките за живота.
- Компютърната биология използва математически и изчислителни подходи за справяне с теоретичните и експерименталните въпроси в биологията. Въпреки че, биоинформатиката и изчислителната биология са различни, има и значително припокриване и дейности в интерфейса им.

# БИОИНФОРМАТИКАТА СЕ ИЗПОЛЗВА В СЛЕДНИТЕ ОБЛАСТИ:

- Молекулярна медицина
- Персонализирана медицина
  - Превантивна медицина
    - Генна терапия
  - Създаване на лекарства
- Приложения на микробните геноми
- Алтернативни източници на енергия
- Изследвания на измененията на климата
  - Биотехнологии
- Резистентност на антибиотици
- Съдебномедицински анализ на микроби
  - Създаване на биологични оръжия
    - Изследване на еволюцията
    - Резистентност на насекоми
      - Качество на храненето
  - Сортове, резистентни на суша
    - Ветеринарна наука

# BIOCOMPUTING

- Biocomputing често се използва като всеобхващащ термин за обхвата на областите на припокриване на биологията и изчислителната наука (Computation)
- Различават се (не-непресичащи се) области:
- *Bioinformatics* – обхваща мениджмънт на биологични бази данни, data mining и моделиране на данни, като и IT-инструменти за визуализация
- ✓ *Computational Biology* – методи и подходи за решаване на биологични проблеми с изчислителни инструменти (като моделиране, алгоритми, евристика)
- ✓ *DNA computing and nano-engineering* – обхващат модели и експерименти за използване на ДНК (и други) молекули за осъществяване на изчисления
- ✓ *Computations in living organisms* – конструиране на изчислителни компоненти в живи клетки, както и изучаване на изчислителните процеси, които се извършват ежедневно в живите организми

# COMPUTATIONAL BIOLOGY

- Компютърната биология включва приложение на фундаменталните технологии на компютърните науки (алгоритми, изкуствен интелект, бази данни, и др.) за решаването на биологични проблеми.
- Биологичните проблеми са достатъчно сложни и широко мащабни за да мотивират разработката на ефективни алгоритми и като устойчива тенденция, изискванията на биологията към изчислителната наука нарастват непрекъснато
- Най-належащите задачи в биоинформатиката са свързани с анализа на секвенции, който е основен обект на компютърната биология
- Компютърната биология обхваща следните дейности:
  - ✓ Откриване на гени в геномните секвенции на различни организми
  - ✓ Разработка на методи за предсказване на структурата и/или функцията на новооткрити протеини и структурни РНК секвенции
  - ✓ Клъстериране на протеиновите секвенции в семейства (фамилии) от свързани секвенции и разработването на протеинови модели.
  - ✓ Подравняване (aligning) на подобни протеини и генериране на филогенетични дървета за изследване на еволюционни взаимозависимости



# HUMAN GENOME PROJECT

[HTTP://WWW.ORNL.GOV/SCI/TECHRESOURCES/HUMAN GENOME/HOME.SHTML](http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml)

- Human Genome Project доведе до промяна на широк спектър парадигми и един от главните резултати е, че доказва, че биологията е информационна наука (to the view that biology is an informational science).
- В резултат на проекта беше създаден списък на генетични фрагменти, който показва, че човешкият геном е изграден от приблизително 25,000 гена, техните регулаторни участъци, лексикон от мотиви, които са изграждащите градивни блокове на протеините и гените, достъп до човешката вариабилност (изменчивост), която прави всеки от нас да бъде строго индивидуален и различен от другите индивиди.

# HUMAN GENOME PROJECT

- Финализиран през 2003г., Human Genome Project (HGP) е 13-годишен проект, координиран от U.S. Department of Energy и National Institutes of Health.
- В началото на проекта HGP, Wellcome Trust (U.K.) става главен партньор;
- Допълнителен принос от Япония, Франция, Германия, Китай, и др.

# HUMAN GENOME PROJECT

## *Цели на проекта*

- *Да се идентифицират приблизително 20,000-25,000 гена в човешката ДНК,*
- *Да се определят секвенциите на 3 милиарда химически базови двойки, изграждащи човешката ДНК,*
- *Да се съхрани тази информация в бази данни,*
- *Да се подобрят инструментите за анализ на тези данни,*
- *Да се осъществи трансфер на тези технологии в частния сектор, и*
- *Да се разгледат етичните, правните, и социални аспекти на резултатите от проекта*
- **Въпреки, че проектът HGP е финализиран, анализът на получените данни ще продължи много години**
- **Трансферът на релевантните технологии към частни компании и предоставянето на грантове за иновативни изследвания в рамките на проекта стимулира *multibillion-dollar U.S. biotechnology industry* и насърчава развитието на нови медицински приложения.**

# Биологическа информация

○ Биологическата информация е от два типа:

1. *Секвенции*
2. *Съдържание, получено от геноми и структури - функционален анализ на генни продукти, получени чрез експерименти*

# Биологическа информация

- В случая на човешкия геном – събиране на информация за биологическата функция на 25 000 човешки гени с цел да се открият тези от тях, които играят съществена роля при дадена болест.
- Използването на съвременни технологии, като микроматрици на гени, генетични манипулации на гени в клетки и организми, както и бързият структурен и функционален анализ на протеини, ни дават много нови и ценни данни за функциите на гените.

# Компютърната биология и биоинформатиката

- Тясно свързани области
- Компютърната биология фокусира върху създаването на нови и ефективни алгоритми за решаването на сложни проблеми, като множествено подравняване на секвенции или асемблиране на геном от фрагменти
- Биоинформатиката фокусира повече върху разработката на практически инструменти за data management и анализ – представяне и визуализация на геномни данни и анализ на секвенции, но с по-малко ударение върху ефективност и доказана точност

# Модерната биоинформатика

- Поддържа широк спектър от изследвания върху определянето на биологичната значимост на данните
- Дава експертиза за организиране на биологични данни
- Разработва практически изчислителни средства to mine биологически данни
- Съдейства с практически приложения, като например идентифициране на нови протеинови таргети (цели) за лекарствена терапия

# Генезис на биоинформатиката

- Със създаването на първите компютри (с огромни габаритуи и главна памет 8К RAM!), първите компютърни биолози започват да създават алгоритми.
- Тази практика е била иновативна— никой преди това не е обработвал и анализирал молекулярни секвенции като текст.
- Повечето методи е трябвало да бъдат създавани от нулата, в процеса на изследванията, и така се създава нова област на научните изследвания - анализ на протеинови секвенции, при които се използват компютри.



КРАЙ