

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»



УТВЕРЖДАЮ
Проректор по учебной работе
Д.С. Гуц / Д.С. Гуц /
«10» марта 2023 г.

ПРОГРАММА
кандидатского экзамена по научной специальности
1.5.8 Математическая биология, биоинформатика

Красноярск 2023

Содержание программы

1. Основы математического моделирования в биологии

1.1. Математическая формализация биологических систем

Переменные состояния, параметры, входы и выходы. Детерминированные и стохастические модели. Масштабы времени и пространства в биологии. Допущения и упрощения, калибровка модели.

1.2. Системы обыкновенных дифференциальных уравнений

Модели динамики концентраций и численности: уравнения баланса, закон действия масс. Фазовое пространство, стационарные точки, их устойчивость, бифуркации. Качественный анализ динамических систем, фазовые портреты.

1.3. Дискретные модели

Рекуррентные соотношения, логистическое отображение, хаотические режимы. Применение в популяционной динамике и эпидемиологии.

1.4. Чувствительность и идентифицируемость моделей

Локальный и глобальный анализ чувствительности. Оценка параметров по экспериментальным данным, переопределённые и недоопределённые модели.

2. Математические модели популяционной динамики и эпидемиологии

2.1. Классические модели популяционной динамики

Мальтузианский рост, логистическая модель Верхьюльста. Модели конкуренции и коэкзистенции видов. Модели «хищник–жертва» типа Лотки–Вольтерры. Возрастная и структурная популяционная динамика, матричные модели. Пищевые сети, модели с несколькими уровнями трофических связей, устойчивость и резилиентность экосистем.

2.3. Детерминированные эпидемиологические модели

Классы S, I, R. Модели типа SIR и их модификации (SIRD, SIRV, SIRVD, SIRVB, SEIR). Базовое репродуктивное число (R_0), пороговые явления, эффекты вакцинации.

2.4. Стохастические эпидемиологические модели

Цепи Маркова в непрерывном и дискретном времени, процесс рождения–смерти. Стохастические SIR-модели, распределения времени вымирания эпидемии, роль флуктуаций в малых популяциях.

3. Биохимическая кинетика и регуляторные сети

3.1. Закон действующих масс и простые реакции

Скорости реакций, последовательные и параллельные реакции, обратимые процессы.

3.2. Ферментативная кинетика

Модель Михаэлиса–Ментен, её вывод и допущения. Ингибирование, активация, кооперативность (уравнение Хилла).

3.3. Сети биохимических реакций

Стехиометрическая матрица, инварианты, консервация. Модели сигнальных путей и метаболических сетей. Анализ устойчивости стационарных состояний.

3.4. Генные регуляторные сети

Непрерывные модели (ODE-подход), логистические функции регуляции, переключатели и биологические часы. Булевы модели регуляторных сетей. Мультистабильность и клеточная дифференцировка.

4. Стохастические процессы в биологии

4.1. Элементы теории вероятностей для биологов

Случайные величины, распределения (биномиальное, Пуассона, нормальное, экспоненциальное, гамма-распределение) и их биологические интерпретации. Законы больших чисел и центральная предельная теорема.

4.2. Марковские цепи и процессы рождения–смерти

Основные определения и свойства. Время до поглощения, устойчивые распределения. Применения к демографии, популяционной динамике, внутриклеточным процессам.

4.3. Стохастические дифференциальные уравнения

Белый шум, броуновское движение. Модели с шумом в нейродинамике, популяционной биологии и биохимии.

4.4. Генетический дрейф и коалесцентная теория

Модели Райта–Фишера и Морана. Время до общего предка, эффективный размер популяции.

5. Математическая генетика и эволюционные модели

5.1. Популяционная генетика

Равновесие Харди–Вайнберга и его нарушения. Модели отбора, мутаций, миграций и дрейфа. Установившиеся частоты аллелей. Баланс мутаций и отбора.

5.2. Количественная генетика

Аддитивные и доминантные эффекты, наследуемость. Линейные смешанные модели в оценке племенной ценности и фенотипов.

5.3. Филогенетика

Модели замещения нуклеотидов (JC, K2P, HKY, GTR и др.). Методы построения филогенетических деревьев (дистанционные, максимального правдоподобия, байесовские). Оценка надёжности (bootstrap, posterior probabilities).

6. Биоинформатика последовательностей и структур

6.1. Базы данных биологических последовательностей

Основные форматы и ресурсы: геномные и транскриптомные базы, базы белков и структур. Аннотация последовательностей, идентификаторы, версии.

6.2. Выравнивание последовательностей

Глобальное и локальное выравнивание, динамическое программирование (алгоритмы Нидлмана–Вунша и Смита–Ватермана). Матрицы замен (PAM, BLOSUM), штрафы за пропуски, оценка статистической значимости.

6.3. Множественное выравнивание и профили

Прогрессивные и итеративные алгоритмы. Профиль-НММ, поиск гомологов, предсказание доменной структуры.

6.4. Поиск мотивов и регуляторных элементов

Позиционно-весовые матрицы, алгоритмы типа Gibbs sampling и EM. Применения к промоторам, сайтам связывания факторов транскрипции.

6.5. Сборка и аннотация геномов

Графовые модели (оверлап-графы, графы де Брёйна), особенности коротких и длинных чтений. Оценка качества сборки. Функциональная аннотация генома.

7. Анализ «омиксных»-данных и статистические методы биоинформатики

7.1. Геномные и транскриптомные данные

Массивы и RNA-seq: нормализация, удаление batch-эффектов, оценка дифференциальной экспрессии генов (основные подходы).

7.2. Анализ вариаций генома

Типы вариаций (SNP, indel, CNV). Основные принципы GWAS, многофакторный анализ, коррекция на множественные сравнения.

7.3. Снижение размерности и кластеризация

PCA, t-SNE, UMAP. Иерархическая кластеризация, k-means, оценка качества кластеров (силуэт и др.). Применения к данным экспрессии, метаболомики, протеомики.

7.4. Сетевой анализ

Коэкспрессионные сети (WGCNA и др.), функциональные сети, центральность, модули, оценка биологических путей.

8. Машинное обучение в биологии и медицине

8.1. Общие понятия

Супервизируемое и несупервизируемое обучение. Обучающая, валидационная и тестовая выборки. Переобучение, регуляризация, кросс-валидация.

8.2. Классические модели

Линейная и логистическая регрессия, методы опорных векторов, деревья решений, случайные леса, градиентный бустинг.

8.3. Нейросетевые методы

Полносвязные и сверточные сети, основы рекуррентных сетей. Примеры применения к анализу изображений, последовательностей, структур.

8.4. Интерпретация моделей и оценка качества

ROC-кривые, AUC, точность, полнота, F-мера. Методы интерпретации (feature importance, SHAP, LIME — обзорно). Этические и практические аспекты применения алгоритмов в медицине.

9. Методология биологических исследований и воспроизводимость

9.1. Дизайн биологических и биомедицинских экспериментов

Генерация гипотез, формулировка целей и конечных точек. Рандомизация, ослепление, контрольные группы, репликация.

9.2. Статистический анализ и отчётность

Мощность эксперимента, ошибки I и II рода, р-значения и доверительные интервалы. Коррекция множественных сравнений.

9.3. Репродуцируемые вычислимые исследования

Документирование анализа, версии данных и кода, рабочие протоколы, использование систем контроля версий и ноутбуков. Принципы FAIR-данных и открытой науки.

Литература

Основная

1. Мюррей, Джеймс. Математическая биология = Mathematical Biology : перевод с английского / Д. Мюррей ; под науч. ред. Г. Ю. Ризниченко. - Москва : Регулярная и хаотическая динамика ; Ижевск : Институт компьютерных исследований, 2009. - Текст : непосредственный. Том 1,2.
2. Таганов, Рейм Таганович. Системный и исторический методы в биологии : [монография] / Р. Т. Таганов. - Москва : Высшая школа, 1989. - 135 с. - ISBN 5-06-001556-4 : 1.7 р. - Текст : непосредственный.
3. Введение в молекулярную биологию = Introduction to Molecular Biology : перевод с английского : [курс лекций] / Д. Г. Хаггис, Д. Миши [и др.] ; под ред., предисл. В. О. Шпикитер. - Москва : Мир, 1967. - 434 с. : ил. - 2.16 р. - Текст : непосредственный.
4. Рис, Энтони. От клеток к атомам: иллюстрированное введение в молекулярную биологию = From Cells to Atoms. An Illustrated Introduction to Molecular Biology : перевод с английского / Э. Рис, М. Стернберг ; под ред.: Ю. С. Лазуркина, В. А. Ткачука. - Москва : Мир, 1988. - 142 с. : ил. - Список лит.: с.136; Предм. указ.: с.137-142. - 45.00 р. - Текст : непосредственный.
5. Анализ биологических последовательностей. Вероятностные модели белков и нуклеиновых кислот = Biological sequence Analysis. probabilistic Models of proteins and Nucleic Acids : перевод с английского / Р. Дурбин, Ш. Эдди [и др.]. - МоскваМосква : Институт компьютерных исследований ; Ижевск : [R&C Dynamics] Регулярная и хаотическая динамика [РХД], 2006. - 479 с. : ил. - Список лит.: с.441-468. Предм. указ.: с.469-479. - ISBN 5-939725-59-7 : 1007.18 р. - Текст : непосредственный.
6. Введение в статистическое обучение с примерами на языке R / Тибширани Р., Мاستицкий С. Э., Джеймс Г. [и др.]. - 2-е изд., испр. - МОСКВА : ДМК Пресс, 2017. - 456 с. - ISBN: 978-5-97060-495-3.
7. Стефанов, Василий Евгеньевич. Биоинформатика : учебник для студентов вузов, обучающихся по техническим и естественнонаучным направлениям / В. Е. Стефанов, А. А. Тулуб, Г. Р. Маврупуло-Столяренко. - Москва : Юрайт, 2020. - 252 с. : рис., табл. - (Высшее образование). - Библиогр. в конце глав. - ISBN 978-5-534-00860-9 (в пер.) : 691.90 р. - Текст : непосредственный.

Дополнительная

1. Elgin M. Mathematical Models, Explanation, Laws, and Evolutionary Biology // History and Philosophy of the Life Sciences. - 2010. - Vol. 32. - No. 4. - DOI: 10.2307/23335085.
2. Dale, Mark R. T... Spatial Analysis: A Guide For Ecologists / M. R. T. Dale, M.-J. Fortin. - 2th ed. - Electronic text data. - Cambridge : Cambridge University Press, 2014. - Загл. с титул. экрана. - Электрон. версия печ. публикации. - ISBN 978-0-521-88679-6. - ISBN 978-0-521-14350-9 : Б. ц. - Текст : электронный.

3. Ли, Ч. . Введение в популяционную генетику = First Course in Population Genetics : перевод с английского / Ч. Ли ; под ред.: Ю. П. Алтухов, Л. А. Животовский. - Москва : Мир, 1978. - 555 с. : ил. - Список лит.: с.527-547; Предм.указ.: с.548-549. - 6000 р. - Текст : непосредственный.
4. Sarkar, Sahotra. Cut-and-paste genetics : a CRISPR revolution / / Sahotra, Sarkar. - Электрон. текстовые дан. - [S. l. : s. n.], 2021. - 1 online resource (xvii, 205 pages) : il. - Загл. с титул. экрана. - Includes bibliographical references and index. - ISBN 9781786614391. - ISBN 1786614391 : Б. ц. - Текст : электронный. Примечания о происхождении: Коллекция цифровых книг Ebsco ebook (централизованная подписка 2023 г., бессрочный доступ). НБ СФУ.
5. Antao, Tiago. Bioinformatics with Python cookbook : use modern Python libraries and applications to solve real-world computational biology problems / Tiago Antao. - Third edition. - Электрон. текстовые дан. - [S. l. : s. n.], 2022. - 1 online resource : il. - Загл. с титул. экрана. - ISBN 9781803247724. - ISBN 180324772X : Б. ц. - Текст : электронный. Includes index Перевод заглавия: Биоинформатика с помощью Python cookbook: применение современных библиотек и приложений на Python для решения реальных задач вычислительной биологии Примечания о происхождении: Коллекция цифровых книг Ebsco ebook (централизованная подписка 2023 г., бессрочный доступ). НБ СФУ.
6. MacLean, Dan. R Bioinformatics Cookbook : Use R and Bioconductor to Perform RNAseq, Genomics, Data Visualization, and Bioinformatic Analysis / Dan MacLean. - Электрон. текстовые дан. - Birmingham : Packt Publishing, Limited, 2019. - 1 online resource (307 pages). - Загл. с титул. экрана. - ISBN 9781789955590. - ISBN 1789955599 : Б. ц. - Текст : электронный. Перевод заглавия: Кулинарная книга по биоинформатике R: Используйте R и Bioconductor для выполнения RNAseq, геномики, визуализации данных и биоинформатического анализа Примечания о происхождении: Коллекция цифровых книг Ebsco ebook (централизованная подписка 2023 г., бессрочный доступ), НБ СФУ.
7. Дополнительные современные статьи по тематике диссертационного исследования аспиранта (по рекомендации научного руководителя).

Перечень вопросов к экзамену

1. Понятие математической модели в биологии. Типы моделей и их классификация.
2. Системы обыкновенных дифференциальных уравнений в биологии: примеры, фазовый анализ.
3. Дискретные динамические системы и их применение в моделировании популяций.
4. Детерминированные модели популяционной динамики (Мальтус, Верхульст, Лотка–Вольтерра).
5. Модели взаимодействующих видов и устойчивость биологических сообществ.
6. Основные детерминированные эпидемиологические модели (SIR, SEIR), базовое репродуктивное число.
7. Стохастические эпидемиологические модели и роль флуктуаций.
8. Закон действия масс и кинетика простых биохимических реакций.
9. Кинетика ферментативных реакций: модель Михаэлиса–Ментен, ингибирование и кооперативность.
10. Модели сигнальных и метаболических сетей, анализ устойчивости.
11. Генные регуляторные сети и их математическое описание.
12. Основные понятия теории вероятностей, используемые в математической биологии.
13. Марковские цепи и процессы рождения–смерти в приложении к биологии.
14. Стохастические дифференциальные уравнения и их биологические приложения.
15. Генетический дрейф, модели Райта–Фишера и Морана.
16. Равновесие Харди–Вайнберга и факторы, его нарушающие.
17. Модели отбора, мутаций, миграций в популяционной генетике.
18. Основы количественной генетики и наследуемости признаков.
19. Модели замещения нуклеотидов и построение филогенетических деревьев.
20. Основные биологические базы данных. Идентификаторы и форматы последовательностей.
21. Алгоритмы глобального и локального выравнивания последовательностей.
22. Множественное выравнивание и профили. Скрытые марковские модели.
23. Методы поиска мотивов и регуляторных элементов в ДНК и белках.
24. Принципы сборки геномов и оценка качества сборки.
25. Анализ данных экспрессии: нормализация, дифференциальная экспрессия.
26. Анализ вариаций генома и основы GWAS.
27. Методы снижения размерности и кластеризации в анализе «омикс»-данных.
28. Сетевой анализ биологических данных: коэкспрессионные и функциональные сети.
29. Основные задачи и методы машинного обучения в биологии.
30. Линейные модели и методы регуляризации в биоинформатике.
31. Деревья решений, случайные леса и градиентный бустинг: принципы и примеры применения.
32. Базовые архитектуры нейронных сетей и их использование в биомедицинских задачах.
33. Метрики качества и методы кросс-валидации в задачах классификации и регрессии.
34. Планирование биологических экспериментов, рандомизация, контроль и репликация.
35. Элементы байесовского подхода в анализе биологических данных.

36. Роль воспроизводимости и открытой науки в математической биологии и биоинформатике.
37. Примеры комплексных моделей в системной биологии (интеграция нескольких уровней организации).
38. Этические и практические аспекты использования математических и вычислительных методов в медицине.

Программа соответствует паспорту номенклатуры специальности 1.5.8 «Математическая биология, биоинформатика» и может уточняться и дополняться с учётом развития направления и актуальной тематики научных исследований.

Составитель программы:
канд. физ.-мат. наук, доцент,
зав. базовой кафедрой медико-
биологических систем и комплексов



А.Н. Шуваев