

# РЕЦЕНЗИЯ

от

**доц. д-р Димитър Иванов Василев,**

катедра “Компютърна Информатика”, Факултет по Математика и Информатика,  
Софийски Университет “Св. Климент Охридски”

на

дисертационен труд на тема:

**Биоинформатични методи и софтуерни решения за *de novo* идентификация и анализ на растителни мРНК гени чрез данни от масово-паралелно секвениране**

автор

**Георги Иванов Минков,**

докторант на самостоятелна подготовка  
при катедра Физиология на Растенията и Молекулярна Биология, Биологически Факултет,  
Пловдивски Университет “Паисий Хилендарски”

за

**присъждане на образователната и научна степен 'доктор'**

в област на висше образование - 4. Природни науки, математика и информатика,  
професионално направление - 4.3. Биологически науки,  
докторска програма “Биоинформатика”.

научени ръководител:

**доц. д-р Веселин Петров Баев,**

катедра Физиология на Растенията и Молекулярна Биология, Биологически Факултет,  
Пловдивски Университет “Паисий Хилендарски”

Със заповед на ректора на Пловдивския университет „Паисий Хилендарски“ бях определен за член на научното жури и рецензент в процедурата за защита на дисертационен труд на тема „ Биоинформатични методи и софтуерни решения за *de novo* идентификация и анализ на растителни мРНК гени чрез данни от масово-паралелно секвениране” за придобиване на образователната и научна степен “доктор” от Георги Иванов Минков.

Представеният от Георги Минков комплект материали и документи на електронен носител е в съответствие с Чл.36 (1) от Правилника за развитие на академичния състав на Пловдивски Университет “Паисий Хилендарски”, като включва следните документи:

- автобиография в европейски формат;
- заповед за зачисляване в докторантура;
- заповед за отчисляване от докторантура;
- два протокола от катедрен съвети, свързан с докладване на готовност за откриване на процедура и с предварително обсъждане на дисертационния труд;

- дисертационен труд;
- автореферат;
- списък на научните публикации по темата на дисертацията;
- копия на научните публикации;
- декларация за оригиналност и достоверност на приложените документи;
- справка за спазване на специфичните изисквания на Биологически факултет на ПУ;

Докторантът е приложил 2 броя публикации, едната от които е публикувана в съавторство в реферираното престижно научно списание Genomics (Elsevier), а другата в Доклади на БАН, което има също, макар и малък импакт фактор

Георги Минков има бакалавърска степен по специалност „Приложна математика“, от Факултет по Математика и Информатика на Пловдивския университет „Паисий Хилендарски“ от 2008. Магистърската си степен получава през 2011 година по специалност “Биоинформатика” при катедра “Физиология на Растенията и Молекулярна Биология при Биологическия факултет на Пловдивския университет „Паисий Хилендарски“. През периода 2009-2015 година той работи като главен експерт по математика и информатика към катедра “Физиология на Растенията и Молекулярна Биология при Биологическия факултет на Пловдивския университет. Образователния ценз, знанията и уменията на Георги Минков, създават предпоставки и интерес у него за изследователска работа в областта на биоинформатиката, чрез разработване и използване на различен софтуер за анализ на молекулярни данни, главно от геномно секвениране. Като продължаване на образованието си той избира и разработването на алгоритмични подходи и софтуерни решения при анализ на секвенционни данни за малки РНК-и.

Представеният дисертационен труд от Георги Иванов Минков е написана на 133 страници, стандартно структурирана и включва: Литературен обзор, Цели и задачи; Материали и методи, Резултати и обсъждане, Изводи (5 бр.) и Заключение. Работата включва 38 фигури и 15 таблици. В Автореферата са представени три научни приноса към дисертацията. Списъкът на литературата включва 205 източника, като не малка част от тях са след 2000 година. Така представеният текст на дисертацията включва 40 фигури и 14 таблици, както и съкращения на използваните термини и понятия.

Дисертационният труд е посветен на създаването и използването на биоинформатични подходи и софтуерните им реализации за анализ на рстителни миРНК при данни от паралелно секвениране. Темата на работата е без съмнение модерна, актуална и е в безусловно съзвучие с темите на много дисертации от подобен род по света. Паралелното секвениране или както звучи най-масово използваното название – next generation sequencing

(NGS) създаде големи предпоставки за повишаване актуалността на биоинформатиката и като нови методи, алгоритми, процедури и като огромно разнообразие от софтуерни реализации, свързани както с тези методи на анализ, съхранение, визуализация, интегриране и протекция на данните, така и с използването на нови области от информатиката, приложната математика и класическите математически дисциплини като изкуствен интелект и машинно самообучение, теория на графи, дискретна математика, логика, обектно-ориентирано програмиране и други.

Обектът на изследване – малките РНК-и е и актуална, и бързо развиваща се област, свързан пряко с изследването на генната експресия и е едно от основните направления, с които се занимавава и работи катедрата по “Физиология на Растенията и Молекулярна Биология”, където докторантът е реализирал своята теза за дисертационен труд. МикроРНК (миРНК) са клас малки, ендогенни, некодирани РНК, които имат важна роля в почти всички биологични процеси, най-главно при регулация на генната експресия особено що се отнася до важни фази в развитието като развитие, толерантност към стрес. Преди всичко изследването на малките миРНК е започнато при растенията. От биологична гледна точка ролята на биоинформатиката в тези изследвания е изключително голяма, особено при определяне структурата и функциите на генома, съпоставянето в контекста на еволюцията, предсказване взаимодействието на миРНК с различни таргетни гени, съпричастността им към биологичното разнообразие. Именно на изследване на откриването, подредбата, структурата, функцията, съпоставянето, аотирането и предсказването на тези миРНК в контекста на данни от де ново секвенирани растителни геноми е работата на докторанта.

Основната теза на дисертацията се заключава, че много от процесите на изследване на миРНК могат да се извършат *in silico*, особено що се отнася до идентификацията и анализ на функциите на миРНК от данни получени посредством високо-производително секвениране на малки РНК от растителни геноми.

Като цяло Георги Минков се е справил със поставената цел и съпътстващите я задачи, разработвайки определна методика и превърщайки я в реално работещи софтуерни програми за целите на работата си като анализ основана на клъстериране и на малки РНК и идентифициране на миРНК, предсказване структурата на миРНК, интегриране на данните от паралелно секвениране с данни от други секвенционни технологии за идентифициране на миРНК и като най-интересна част, поне за мен – аотиране на малки новооткрити миРНК чрез методи на обектно-ориентираното програмиране. Анализите са основани на секвенционни данни на *Brachypodium distachyon* и *Solanum lycopersicum*.

Основното достойнство на дисертацията са софтуерните реализации за анализ на миРНК от секвенционни данни:

Clustering module - предназначена за идентификация на растителни миРНК чрез изследване на резултати от масово-паралелно секвениране (на малки РНК) и представлява предшественик на програмата clusterSD.

ClusterSD, miRNA2D и miCompare са три отделни програми, събрани заедно в общия софтуерен пакет miRDEG за идентификация на растителни миРНК чрез анализиране на резултати от масово-паралелно секвениране като се интегрират два типа данни – малки РНК и доградомни секвенции.

LoopRNA е самостоятелен Java модул, предназначен да моделира вторичната структура на РНК секвенции в паметта чрез теориите на обектно-ориентираното програмиране и математическите графи, нужен за изследването на 2D РНК структурата на пред- полагаемите миРНК прекурсори и представлява част от програмата miRNA2D.

Литературният преглед на работата е обстоен, носещ повече биологическа и методична информация, структуриран е добре и е съразмерно дълъг по сравнение с общия обем на дисертацията. Основно внимание е отделено на всичко свързано с обекта на изследването – миРНК, както и въпроси засягащи определени аспекти, свързани с паралелното секвениране.

Раздел “Цели и задачи” е написан, кратко и ясно. Има известни стилови неточности – като например наличието на една цел, а заглавието да е насочено към повече от една цел. Тук би могло да се иска за малко по-добра композитивност на задачите като биологическа необходимост и *in silico* същност и реализация.

Раздел “Материали и Методи” се отличава силно в дисертацията, като в него най-подробно са изложени, обяснени и интерпретирани всички проблеми свързани със създаването на софтуерните програми за целите на работата. Според мен е добре и интересно написан, с някои уговорки, които са изложени по-долу в рецензията, но тук за информатиката и всичко свързано с ней превалира в голяма степен над биологическия контекст на работата. Неоспорим е фактът, че разделът е написан много ясно и дори непредубеден биолог би разбрал без особени затруднения за какво точно се използват разглежданите методики и теоретични постановки.

Най-важната част от една дисертационна работа, тази обхващаща резултатите и обсъждането на пръв поглед като че ли са отнели малко от Материал и Методи по отношение на обяснението на функционалността на създадените софтуерни решения от докторанта, но това би могло да се пренебрегне и да се вгледаме по сериозно в резултатите, свързани с тези софтуери, както и тяхната функционалност. Би могло и да се продължи дискусията в тази

глава относно сравнимостта на предложените софтуерни продукти с някои подобни и разпространени такива.

Изводите отговарят на поставената цел и задачи на работата. Заключение, макар и лаконично е отражение на тезата на автора за разработването на дисертационния труд.

Във връзка с написаното, към представения за защита дисертационен труд от Георги Минков биха могли да се зададат някои въпроси като:

- 1) Семантичният смисъл на използването на обектно-ориентирано програмиране в моделирането, предсказването, оценката на мРНК. Какво то дава повече и какви са неговите големи преимущества ?
- 2) С цел повишаването на приложимостта на разработените от докторанта софтуерно решения, как биха могли да се използват в единен конвейър (пайплайн) и дали може да във връзка с това web-реализация ?
- 3) Анотирането като процес е най-проблемния в биоинформатиката. С какво и как се подобрява качеството на този процес на базата на развитите и изработените в работата подходи и софтуерни приложения? С какво се обогатява анотирането и на каква база се сравнява ?

Естествено е към представения за защита дисертационен труд да има и някои забележки :

#### Систематичен характер:

- 1) подредбата на определения за категориите на обект ана изследване, на методи и резултати – това не се наблюдава постоянно по най-добрия начин в работата на дисертанта. Ва тази връзка ми се струва, че фиг. 18 би могла да мине значително по напред като илюстрация на постановка на тезата на работата.
- 2) Малко фриволното обяснение на теория на графи и структури от данни в работата. Би могло да се направи по-добре и по ясно.
- 3) В раздел “Материал и методи” трябва да се подчертае особено силно какво е създадено – т.е. какви са авторските решения за постигане на целта в тезата на дисертационния труд.
- 4) Много място е отделено на филтрирането, а би могло да се предсатви по-кратко и по-ясно. А малко място е отделено на възможното *in silico* валидиране.
- 5) Би било много добре ако в началото на дисертацията имаше и нещо като тълковен речник на не малка част от разнообразието от понятия в дисертацията.

#### Синтактичен характер:

- 1) доста голямо ранообразие на термини за съпоставяне на нуклеотидни низове: “алайнмент”, “мапване”, “картиране” – според мен първите два не звучат добре на български. Тук забележката ми е и към научния ръководител.
- 2) Много от техническите термини според мен са непълно и неправилно употребени: “математически” графи, според мен “математически” е излишно, метадата – метаданни, интергенност, особено “машинно учене”...
- 3) Java е по-близо до C# отколкото до C++. Най-голямото достоинство на Perl е възможностите за писане на регулярни изрази, като най-просто подреждане изпълнението на различни десйтвия и команди
- 4) Някои от по-сложните изречения, тези които съдържат поне едно подчинено изречение се губят.

## **ЗАКЛЮЧЕНИЕ**

Независимо от направените забележки и препоръки, аз считам, че дисертантът е Георги Иванов Минков е представил една добре формулирана теза за дисертация, която е изпълнил с много старание и знания. Представеният дисертационен труд има определен приносен характер: интересен обект на изследване, нови методи за анализ, много нови и интересни резултати, основната част от работата е публикувана в добри списания. Всичко това, а и написаното по-горе от мен в рецензията към работата ми дава основание да подкрепя присъждането на Георги Иванов Минков на научната и образователна степен “доктор” в област на висше образование - 4. Природни науки, математика и информатика, професионално направление - 4.3. Биологически науки, докторска програма - Биоинформатика.

14.11.2017

София

**Изготвил рецензията:**

**Доц. д-р Димитър Ив. Василев**