

РЕЦЕНЗИЯ

от Акад. проф. Иван Георгиев Иванов, дбн
Институт по молекулярна биология, БАН

на дисертационен труд за присъждане на образователната и научна степен „Доктор“ в
Област на висше образование: 4. Природни науки, математика и информатика,
Професионално направление: 4.3. Биологични науки, Докторантска програма:
„Биоинформатика“

Автор: Георги Иванов Минков

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

Научен ръководител: доц. д-р Веселин Баев, Катедра „Физиология на растенията и молекулярна биология“, Биологически факултет, ПУ „П. Хилендарски“.

1. Общо описание на представените материали

Със заповед № Р33-5150/30.10.2017 г. на Ректора на Пловдивския университет „Паисий Хилендарски“ (ПУ) съм определен за член на научното жури за осигуряване на процедура 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“ в Област на висше образование 4. Природни науки, математика и информатика, Професионално направление 4.3. Биологични науки, Докторантска програма „Биоинформатика“.

Автор на дисертационния труд е Георги Иванов Минков – докторант на самостоятелна подготовка към катедра „Физиология на растенията и молекулярна биология“ с научен ръководител доц. д-р Веселин Баев от ПУ.

Представеният от Георги Иванов Минков комплект материали на хартиен носител е в съответствие с Чл.36 (1) от Правилника за развитие на академичния състав на ПУ и включва следните документи:

- Молба до Ректора на ПУ за разкриване на процедура за защита
- Автобиография в европейски формат
- Копия от дипломи за висше образование (ОКС „магистър“)
- Заповед за зачисляване в докторантура
- Заповед за провеждане на изпит по специалността и протокол от изпита (отл. 6)
- Протоколи от катедрени съвети свързани с откриване на процедура за защита
- Дисертационен труд, съдържащ 133 страници, 38 фигури, 12 таблици и 220 литературни източници
- Автореферат на дисертационния труд обхващащ 36 страници
- Списък и копия от научните публикации свързани с дисертацията (2 броя).
- Декларация за оригиналност на получените резултати

Прегледът на документите показва, че процедурата по зачисляване, провеждане на обучението и отчисляване от докторантура е проведена съгласно изискванията на ЗРАСРБ и вътрешния правилник на ПУ „Паисий Хилендарски“.

2. Кратки биографични данни за докторанта

Георги Иванов Минков е роден на 29.09.1984 г. Той е възпитаник на Математическия факултет на ПУ, откъдето през 2008 г. придобива магистърска степен по „Приложна математика“, а по-късно и по „Биоинформатика“ (от Биологическия факултет). След дипломирането си постъпва на работа като Математик-информатик (Главен експерт) в Пловдивски университет „П. Хилендарски“. По време на своята докторантура е бил на шестмесечна специализация в Улвърхамтънския университет, Великобритания.

3. Актуалност на тематиката и целесъобразност на поставените цели и задачи

Дисертацията на Георги Минков е посветена на идентифицирането на микро-РНК (миРНК) в растителни геноми. Те са подобни на малките интерфериращи РНК, но механизмът им на действие е доста по-различен. Двата класа малки РНК бяха открити преди малко повече от едно десетилетие в нематода *Caenorhabditis elegans*, след което стана ясно, че те са характерни за всички еукариотни клетки, включително растителните. Най-нови изследвания сочат, че функциите на миРНК включват контрол върху клетъчната пролиферация, апоптозата (клетъчната смърт), развитието на нервната тъкан при животните, развитието на листата и цветовете при растенията и много други. Множествените биологични функции на миРНК предизвикаха оправдан интерес към тях и за кратък период от време бяха проведени хиляди проучвания върху тяхната биосинтеза и механизъм на действие. Започна усилено търсене на миРНК в секвенирани еукариотни геноми и дори се заговори за „миРНК геномика“. Поради малкия размер на тези РНК молекули (21-24 нуклеотида) обаче, идентифицирането им се оказва съвсем нелека задача за геномната биоинформатика. Разработени бяха различни подходи и софтуерни продукти за тяхното откриване, но всеки от тях показва своите предимства и недостатъци. От тази гледна точка актуалността на дисертационната тема е извън всякакво съмнение.

Целта на дисертацията е да се разработи софтуер за идентификация и анализ на растителни миРНК гени, базиран на интеграцията на данни от масово паралелно секвениране на малки РНК. Адекватни на поставената цел са формулирани пет задачи свързани както с разработването на новия алгоритъм, така и с тестването му върху два моделни растителни генома - *Brachypodium distachyon* и *Solanum lycopersicum*.

4. Познаване на проблема

Литературният обзор обхваща 30 страници и се позовава на повече от 200 литературни източници. Той е съставен от два основни подраздела, първият от които е посветени на структурата и организацията на миРНК, а вторият – на масовото паралелно секвениране (NGS) и разпознаване на миРНК гени в секвенирани геноми. В първият подраздел докторантът разглежда биосинтезата на миРНК, хипотезите относно техния произход, структурната им организация, тяхната еволюция, функция и проблеми при идентификацията им в секвенирани геноми. Във втория подраздел са разгледани особеностите и спецификата на паралелното геномно секвениране, проблемите при обработката и анализа на първичните данни, картирането на рийдове и разпознаване/предвиждане на миРНК гени в секвенираните геноми. Критичният и аналитичен литературен анализ е позволил на докторанта да намери своята изследователска ниша и да формулира точно целите и задачите на своята дисертация (виж по-горе).

5. Методика на изследването

Методичният раздел включва подробно описание на използваните бази данни, софтуерни продукти и математични инструменти. Секвенционните (РНК/ДНК) бази данни са общодостъпни и по важните от тях са SOL Genomics Network, miRBase (специализирана миРНК база данни), данни от масово-паралелно секвениране на малки РНК и др. Докторантът използва и голям набор от търговски софтуерни продукти като Bowtie (за картиране на секвенции), fastaFromBed, Vienna RNA (за вторична структура на РНК), Galaxy (за контрол на работния процес), Eclipse и др., както и разнообразни математични инструменти (програмните езици Java, Perl, теория на обектно-ориентираното програмиране, теория на математическите графи и др.). Използвайки модерна методология, Г. Минков създава нов оригинален програмен продукт, който описва подробно в раздел Резултати и обсъждане (виж по-долу).

6. Характеристика и оценка на дисертационния труд

Резултатите от изследванията са представени на около 50 страници в раздел „Резултати и обсъждане”, който е богато илюстриран с фигури и таблици. Логиката и хронологията на проведените изследвания са в съответствие с предварително поставените цели и задачи.

Познавайки недостатъците на съществуващите методи за откриване на малки РНКи Г. Минков си поставя за задача да разшири техните възможности като създаде нови програмни продукти базирани на особености в структурата на миРНК и на специфичната организация на техните гени в растителните геноми. Така той успява да разработи оригинален разширен геномен подход за компютърно идентифициране на миРНКи базиран на интегрирането на пълната геномна секвенция със специализирани библиотеки за малки РНКи (sRNA). Подходът е тримодулен, като първият модул е клъстериращ (clustering module). Той открива области в генома съдържащи два еднопосочни sRNA клъстера. Новият метод е апробиран върху генома на *Brachypodium distachyon*, където са предсказани 102 миРНКи, от които 58 са нови за този геном. 56 от миРНКите са консервативни и 46 неконсервативни. Разработката е описана в детайли в първата част на раздел Резултати и обсъждане, а резултатите са публикувани в престижното научно списание *Genomics* (с ИФ 2.8). Чрез интегриране на два типа данни от масово-паралелно секвениране – малки РНК секвенции и деградомни секвенции, пък е разработен модул за клъстеризация и предвиждане на миРНК гени, както и за идентифициране на реално експресирани миРНК локуси.

Втората значима разработка свързана с докторската теза на Г. Минков е посветена на създаването на нов софтуерен продукт (Java модул) за изучаване на вторичната структура на РНК, наречен от авторите „LoopRNA“. Той не може да функционира самостоятелно, но се интегрира лесно в други програми за анализ на вторични структури на РНК. Модулът е базиран на теорията на графите и на обектно-ориентираното програмиране, използвано главно за писане на интерфейсни програми (прозорци, бутони, графики, текстови полета и т.н.). Повечето биоинформатични програми използват традиционното линейно програмиране, т. е. алгоритми съставени от последователни команди. „LoopRNA“ модулът създава отделен независим логически обект за всяка РНК структура съставена от стебла и бримки, което прави програмата "по-умна" и ѝ позволява да извършва сложни изчисления с прост програмен код. Основната идея на „LoopRNA“ не е да нагъва РНК (тя изисква вече-нагъната молекула във Vienna формат) и тя не е

предназначена да представя данни на потребителя. Целта е модулът да представя зададените данни на програмата в която е вграден така, както ние хората я възприемаме поглеждайки нагънатата РНК. Ако човек погледне нагънатата РНК, той вижда само стебла и бримки, а компютърът - съвкупност от нуклеотиди. „LoopRNA“ дава възможност на програмата да "мисли" като човек, т.е. вместо да обработва цялата РНК секвенция, тя може да „запита“ модула за разположението на всички стебла и бримки или колко пъти се "разклонява" една секвенция, с което ѝ спестява обработката на цялата нуклеотидна секвенция отново и отново. Дори при директна обработка на секвенцията, „LoopRNA“ намалява размера на обработвания регион. С други думи, „LoopRNA“ позволява на използвания софтуер лесно да намира отговори на неясни въпроси без да изисква дълго, сложно и многократно обработване на РНК секвенцията. Възможностите на модула са проверени анализирайки вторичната структура на миРНК от *Arabidopsis thaliana*. Разработеният нов софтуерен продукт е предоставен за свободно ползване на сайта <https://github.com/orgs/plantgene/>. Той запълва една празнина в компютърното изучаване и онагледяване на вторична структура на РНК.

Чрез комбиниране на модулите ClusterSD и LoopRNA в софтуерния пакет miRDEG, докторантът потвърждава наличието на 25 миРНК в генома на *Solanum lycopersicum* и открива 29 нови локуса, от които 22 на известни и 7 на нови миРНК.

Като цяло, резултатите свързани с дисертационния труд са отлично документирани и добре интерпретирани в светлината на литературните данни, което показва умението на докторанта да борави творчески с литературните източници и да оценява обективно своя собствен принос в разработваната научна област.

7. Приноси и значимост на разработката за науката и практиката

Дисертационният труд на Георги Минков има приносен характер за геномиката и в частност за растителната геномика. Той съумява да разработи софтуерни модули за откриване на миРНК, както и за определяне на вторичната структура на малки РНК, които могат да бъдат интегрирани в други програмни пакети. Тяхната ефективност и висока разрешителна способност е проверена върху няколко секвенирани бактериални геноми и това е довело до разкриване на нови миРНК и нови структури на малки РНК молекули (виж по-горе).

8. Преценка на публикациите по дисертационния труд

Резултатите от дисертационния труд на Г. Минков са отразени в **2 научни статии** в реферирани списания, от които едната е публикувана в реномираното международно списание **Genomics** с **ИФ 2.8**, а другата в **Доклади на БАН (ИФ 0.233)**. В една от статиите докторантът е водещ автор.

9. Лично участие на докторанта

От лични наблюдения, от уверението на научния ръководител и водещото място на докторанта в една от приложените статии съдя, че разработката на дисертационния труд е негово лично дело.

10. Автореферат

Запознат съм с проекта за автореферат и намирам, че той отразява адекватно съдържанието и постиженията на дисертацията.

11. Критични бележки и препоръки

Дисертационният труд на Г. Минков е написан на точен професионален език и е оформен съгласно традиционния за докторски дисертации формат.

Към докторанта имам следния въпрос: Могат ли неговите софтуерни продукти да бъдат директно приложени за откриване и структурни изследвания на миРНК и други малки РНК в животински и други еукариотни геноми?

12. Препоръки за бъдещо използване на дисертационните приноси и резултати

Разглеждам настоящите изследвания на Г. Минков като дескриптивен период в неговата научна кариера. Препоръката ми е да използва своите знания и умения по биоинформатика за изучаване в бъдеще на функциите на миРНК в норма и патология при подходящи биологични модели (растителни, животински или най-добре човешки).

ЗАКЛЮЧЕНИЕ

Биоинформатиката е нова наука без която авангардните клонове на медико-биологичните науки като геномика, протеомика, молекулярна биология, молекулярна генетика, таксономия, структурна и еволюционна биология и др. вече не могат да се развиват. Биоинформатикът трябва да има добри познания по математика и информатика и определени биологични знания в областта в която работи. Точно такава професионална подготовка има докторантът Георги Минков, което е предопределило и успеха на неговата докторантура. В своя дисертационен труд той е разработил и доказал ефективността на оригинални софтуерни продукти за идентифициране и определяне на структурата на миРНК в секвенирани растителни геноми. Благодарение на тях са открити нови миРНК и РНК структури в геномите на *Brachypodium distachyon*, *Arabidopsis thaliana* и *Solanum lycopersicum* като по този начин са обогатени световните бази данни с нови секвенции и структури. Резултатите от проведените изследвания са публикувани в реномирани научни списания с ИФ, което е атестат за тяхната научна стойност. Така докторантът се представя пред научната общност като зрял и перспективен изследовател в областта на биоинформатиката, който напълно удовлетворява изискванията на Закона за развитие на академичния състав в Република България, Правилника за неговото прилагане и вътрешния Правилник на ПУ „Паисий Хилендарски“ за придобиване на ОНС „Доктор“. Всичко това ми дава основание **убедено да препоръчам** на уважаемото научно жури **да присъди на Георги Иванов Минков образователната и научна степен „Доктор“** по научната специалност **„Биоинформатика“**.

07.11.2017 г.

Рецензент:

/Акад. Иван Г. Иванов/