

Co można odczytać z sekwencji nukleotydowej? Bioinformatyka na przykładzie ryb akwariowych

Piotr Łapa¹

¹Towarzystwo Naukowe Branży Zoologicznej „Animalian”

Wiek XX był okresem dynamicznego postępu techniki w tym szczególnie spektakularnie rozwijały się technologie informatyczne. Obecny wiek XXI będzie najprawdopodobniej okresem, w którym prężnie rozwijać się będą nauki biologiczne z wyraźną dominacją biologii molekularnej. Stanie się to możliwe dzięki szerokiemu zastosowaniu metod obliczeniowych w biologii. Już dziś dzięki technikom informatycznym możliwe stało się pchnięcie tych nauk w jakościowo i ilościowo nowy zakres badań które wraz z rozwojem narzędzi analiz matematycznych znalazły szerokie zastosowanie, wyodrębniając się z czasem w samodzielną dyscyplinę naukową, znaną jako bioinformatyka. Zakres tematyczny zagadnień bioinformatycznych, jest bardzo rozległy, w niniejszym artykule przedstawione zostaną na rzeczywistych przykładach jedynie wybrane jej zastosowania. Obecnie największe zasługi biotechnologii to analizowanie sekwencji molekularnych tj. DNA, RNA i białek.

Czym byłyby obecnie genetyka gdyby nie techniki informatyczne? Trudno powiedzieć, ale pewnym jest, że bez ścisłej współpracy genetyków molekularnych i bioinformatyków nie możliwy stałby się rozwój genomiki. Możliwość poznania kompletnych sekwencji genetycznych to ogromny sukces genomiki, jednak archiwizowanie, udostępnianie i analizowanie coraz większych porcji informacji genetycznych już dawno przekroczyło możliwości percepcyjne ludzkiego umysłu. Do ogarnięcia wciąż rosnącej ilości informacji niezbędne jest oprogramowanie informatyczne.

W początkach bioinformatyki jej zadaniem było tworzenie i organizacja wyspecjalizowanych baz danych na temat sekwencji nukleotydowych i aminokwasowych, których liczba rosła w coraz szybszym tempie. Rozpowszechnienie się metod genetyki molekularnej w ośrodkach naukowych na całym świecie wywołało potrzebę szybkiej wymiany informacji między tymi placówkami. W efekcie powstawać zaczęły powszechnie dostępne internetowe bazy, z których można pobrać dane korzystając z dowolnego komputera mającego połączenie z Internetem. Ilość informacji molekularnych w tych bazach szybko

wzrastała jednak przełomowym momentem było poznanie pierwszych sekwencji nukleotydowych całych genomów. Era bioinformatyki klasycznej skończyła się wraz z poznaniem sekwencji nukleotydowej ludzkiego genomu. Bioinformatyka tak jak wiele innych dziedzin naukowych wkroczyła wtedy w nowy okres tzw. postgenomowy.

Obecnie dzięki powszechnej dostępności komputerów osobistych i narzędzi bioinformatycznych dynamicznie rozwija się genomika teoretyczna, w której swych sił może spróbować prawi każdy. Nie trzeba być w żaden sposób związanym z jakąkolwiek instytucją naukową, aby na własne oczy zobaczyć cały zapis genetyczny genomu np. *Takifugu rubripes* (Temminck & Schlegel, 1850) czy też *Tetraodon nigroviridis* Marion de Procé, 1822, wystarczy domowy komputer z dostępem do Internetu a odpowiednie sekwencje można ściągnąć na twardy dysk. Pojawia się wtedy pytanie. „Co z tym można zrobić?”. Jak w zapisie w postaci ciągu znaków oznaczających nukleotydy przedstawionym przy pomocy czterech liter; A (ang. Adenine), T (ang. Thymine), G (ang. Guanine), C (ang. Cytosine), odnajdywane są informacje, które fascynują miliony ludzi na całym świecie? By z zapisu genetycznego można było wyciągnąć naukowe wnioski przede wszystkim należy umiejętnie stawiać pytania a potem poszukiwać na nie odpowiedzi. Właśnie na tym drugim etapie nieodzowne staje się zastosowanie programów bioinformatycznych.

Pierwszymi metodami bioinformatycznego wnioskowania na podstawie sekwencji nukleotydowych były narzędzia pozwalające na poszukiwanie podobieństw pomiędzy różnymi odcinkami DNA lub RNA porównywanymi względem siebie. Zaobserwowanie identycznych fragmentów w sekwencjach pochodzących z genomów różnych gatunków świadczyć może o związkach ewolucyjnych obu gatunków, a jednocześnie ważnej roli kodującej takiego konserwatywnego odcinka. Jeśli porównywane sekwencje składają się z odcinków identycznych i odmiennych to wnioskować można o tym które fragmenty to eksony niosące informacje o białkach a które są intronami nie wykorzystywanymi do syntezy białek.

Analizy takie wykonywać można również samodzielnie co jest doskonałą formą samokształcenia i jest pomocne w procesie dydaktycznym. Podstawową informacją niezbędną przed przystąpieniem do obróbki sekwencji nukleotydowych jest zapoznanie się z formatem plików wsadowych obsługiwanym przez większość programów stosowanych w biologii molekularnej. Format ten to FASTA i charakteryzuje się tym że po znaku większości (>) określającym początek nowego dokumentu znajduje się identyfikator sekwencji czyli jej nazwa lub tytuł. W kolejnym wierszu jest już właściwa sekwencja nukleotydowa zapisana małymi bądź też dużymi literami. Plik taki można bardzo łatwo edytować w programie

tekstowym takim jak notatnik w który wyposażony jest system operacyjny Windows. Przykładowy zapis przedstawiono poniżej:

>Nazwa

```
CTCTGATGCAACTCTGTTTGTATGGTGGCATTGTTGTA
AGGTCTCAGGATTCTGTCCATAAATGAGCGTTTTTGATC
TGTTTCGTGGGTTTTTCGGGGTTCCCGGAGGTCATTATC
GAGAAGATGGCCGCAGGGACCCCTTCTTTGATGGAATGA
TTCATGAAGATGATGATGATGAAGACGAGGATGACTTTA
ACAGACCCCATCGGGATCCGTTTGACGATGCCTTCAGGT
TTGGATTCAGTTTCGGTCCAGGCGGGGCACGCTTTGAAG
AGCCTCAGATGTTTCGGCCAGATCTTCAGAGACATGGAGG
AGATGTTTGCTGGTCTGGGACGCTTTGATGAGCGACATG
GATTTGGACCGAGAGGTTTCCCGTCAATAGAAGCTCCAC
```

Najpopularniejszym narzędziem służącym do potwierdzania podobieństw sekwencji jest algorytm BLAST (ang. **B**asic **L**ocal **A**lignment **S**earch **T**ool). W programie tym ciąg znaków dostarczony przez użytkownika w zapytaniu, wykorzystywana jest do przeszukania całej dostępnej bazy danych w NCBI (ang. **N**ational **C**enter for **B**io**t**echnology **I**nformation). Wśród wszystkich sekwencji wyszukiwane są te, które w najwyższym stopniu podobne są do analizowanego fragmentu. Wśród zasobów łańcuchów nukleotydowych znajduje się wiele pochodzących z genomów ryb w tym również gatunków powszechnie hodowanych w akwariach. Szybki dostęp do algorytmu BLAST umożliwia strona internetowa NCBI pod adresem <http://www.ncbi.nih.gov/>. Jako opcje wyszukiwania należy wybrać „*Nucleotide*” gdyż poszukiwany ciąg nukleotydowy posłuży jako zapytanie do przeszukania całej bazy danych pod kątem sekwencji do niego podobnych. Gatunkiem który stanowił będzie obiekt poszukiwań jest gupik pawie oczko, więc jako kwerendę zapytanie należy wpisać łacińską nazwę tego gatunku „*Poecilia reticulata*” i nacisnąć „*Search*”.

Pierwszą sekwencją wśród wyników wyszukiwania która pokazała się autorowi w dniu pisania tego tekstu jest cząsteczka mRNA o numerze dostępu DQ337476, kodująca insulino-podobny czynnik wzrostu 2 – IGF2 (ang. **I**nsulin-like **G**rowth **F**actor **2**). Czytelnikom jako pierwsze na liście pojawią się zapewne inne wyniki. Po naciśnięciu na hiperłącze (link) wyświetlany jest plik GenBanku. Zmieniając w polu „*Display Settings*” format wyświetlania na „*FASTA*” otrzymuje się plik składający się z wiersza nazwy i sekwencji nukleotydowej

którą należy skopiować bez nazwy. By dokonać porównania tej cząsteczki ze wszystkimi innymi znajdującymi się w bazie danych należy otworzyć stronę <http://www.ncbi.nlm.nih.gov/BLAST/> i wybrać opcje pozwalającą na porównanie genu IGF2 gupika do innych sekwencji nukleotydowych czyli „*Nucleotide blast*”. Skopiowany wcześniej zapis genu IGF2 pochodzący z genomu gupika należy wkleić w okno dialogowe „*Enter Query Sequence*” i uruchomić algorytm wciskając „*BLAST*” a następnie „*Format*”. W nowym oknie przeglądarki internetowej pojawią się wyniki wyszukiwania w kolejności od sekwencji najbardziej do najmniej podobnej w stosunku do tej z zapytania. Pierwszy z wyników jest identyczny z poszukiwanym fragmentem gdyż jest to ten sam gen. Kolejny na liście wynik jest podobny do wyszukiwanej cząsteczki w 99% i jest to sekwencja tego samego genu ale występująca u innego gatunku tj. *Pseudopoecilia festae* (Boulenger, 1898). Obie cząsteczki różnią się od siebie tym że wyjściowa pochodząca z genomu gupika jest dłuższa o 3 nukleotydy oraz w dwóch miejscach sekwencje te różnią się pojedynczymi nukleotydami. Kolejne sekwencje nukleotydowe wśród wyświetlonych wyników coraz bardziej odbiegają od wyjściowej nasuwając wnioski o coraz większej rozbieżności ewolucyjnej o której przekonać się można wybierając funkcje „*Tree View*”.

Analizy porównawcze obu cząsteczek nukleotydowych w zapisie liniowym prowadzić można również przy pomocy oprogramowania bioinformatycznego zainstalowanego na twardym dysku komputera. Po otwarciu w Gen Banku obu sekwencji nukleotydowych można je zapisać na twardym dysku wybierając w okienku „*Display*” format „*FASTA*” a w okienku „*Send to*” funkcje „*File*”. W tym momencie rozpocznie się zapisywanie na komputerze pliku z wytypowanymi sekwencjami, w wybranej przez użytkownika lokalizacji na dysku.

Do przeprowadzania procesu dopasowania (ang. Alignment) użyć można programu ClustalX lub korzystającego z tego samego algorytmu programu BioEdit. Po uruchomieniu programu ClustalX należy otworzyć plik z zapisaną sekwencją mRNA genu IGF2 gupika „*File/Load Sequences*” a następnie dodać komendą „*File/Append Sequence*” sekwencje mRNA tego samego genu pochodzącego od *Pseudopoecilia*. Barwna wizualizacja pozwala zorientować się że w początkowym odcinku oba ciągi znaków są identyczne aż do pozycji 83 gdzie dochodzi do przesunięcia sekwencji o 3 nukleotydy których jak gdyby brakuje w zapisie pochodzącym od *Pseudopoecilia festae*. W celu prawidłowego uliniowania obu sekwencji stosuje się funkcję „*Alignment/Do Complete Alignment*”. Brakujące nukleotydy wypełnione zostaną znakiem interpunkcji (–) od pozycji 83 do 85. Dzięki temu pozostała część zapisu znów się pokrywa za wyjątkiem dwóch pozycji o odmiennych nukleotydach.

Dopasowanie sekwencji jest procesem wskazującym te odcinki które wzajemnie sobie odpowiadają. Może to być konsekwencją konserwatywności ewolucyjnej tych odcinków, podobieństwa strukturalnego i funkcjonalnego lub jedynie efektem losowym. Zbieżność sekwencji jest wielkością obserwowalną którą można wyrazić jako np. procent identyczności który w powyższym przykładzie wynosi 99%. Na podstawie widocznej zgodności wnioskować można o homologii obu fragmentów, czyli o wspólnym pochodzeniu dwóch porównywanych genów. Począwszy od wspólnego przodka obu taksonów odbywa się proces dywergencji tzn. różnicowania się w toku ewolucji. W wyniku tego, powstają zmiany w zapisie genetycznym polegające głównie na substytucjach, insercjach i delecjach sekwencji DNA. Trójnukleotydowy fragment porównywanych genów IGF2 zlokalizowany od pozycji 83 do 85 to sytuacja którą można interpretować jako insercję (indele) w jednej albo delecję (przerwy) w drugiej sekwencji (ang. indels i gaps). Przerwy takie są zwykle przedstawiane w dopasowywanych łańcuchach jako rząd kresek.

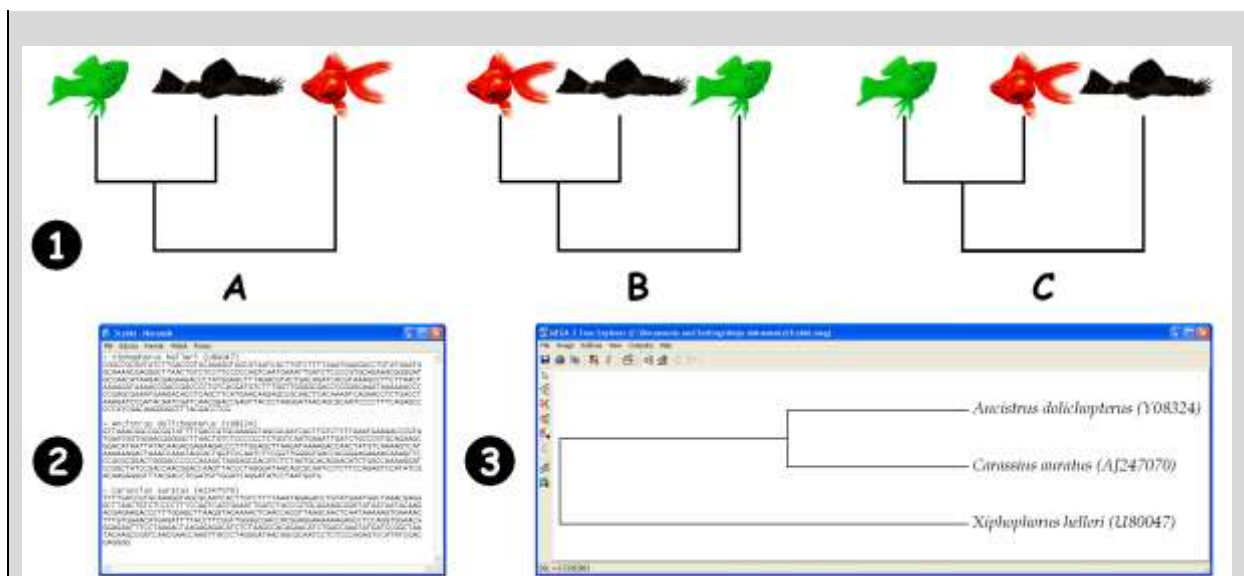
Substytucja to jedna z możliwych mutacji genetycznych polegająca na zmianie nukleotydu na inny. Mutacja polegająca na zamianie zasady purynowej (A lub G) na inną purynę (A lub G) bądź pirymidynowej (C lub T) na inną pirymidynę (C lub T) zwana jest tranzycją. Natomiast zamiana puryny (A lub G) na pirymidynę (C lub T) lub pirymidyny na purynę zwana jest transwersją. Substytucja występuje dwukrotnie podczas porównywania sekwencji mRNA genu IGF2 pochodzącego z genomu gupika i *Pseudopocilia*. W pozycji 156 u gupika znajduje się A natomiast u *Pseudopocilia festae* pozycję tą zajmuje G, a więc jest to tranzycja. Podobna sytuacja występuje też w pozycji 196 gdzie u gupika jest T natomiast u *Pseudopocilia festae* pozycję tą zajmuje C.

Dziedzina zdominowaną obecnie przez bioinformatykę jest filogenetyka czyli nauka o relacjach ewolucyjnych. Celem analizy filogenetycznej jest wysuwanie i szacowanie istotności wniosków na ich temat. Odtwarzana dzięki niej historia ewolucyjna przedstawiana jest zazwyczaj w postaci diagramów przypominających wyglądem rozgałęziające się drzewo.

Dane przeznaczone do analiz filogenetycznych są zwykle zestawieniami sekwencji dopasowanych (ang. Multiple Sequence Alignment). Typowa procedura dopasowania sekwencji prowadzona jest za pomocą takich programów jak Clustal (W lub X) oraz przed przesłaniem danych do programu tworzącego drzewa filogenetyczne wymaga czasami dodatkowego ręcznego poprawienia uzyskanych wyników. Drzewa filogenetyczne uzyskać można programem ClustalX jednak nie są to analizy wiarygodne gdyż program ten nie służy do ich wyznaczania. Jeśli jednak istnieje taka potrzeba można je poddać obróbce przy pomocy

innych odpowiednich narzędzi np. TreeView lub szczególnie polecanym ze względu na łatwość programem MEGA.

Wiarygodne analizy filogenetyczne wymagają przyswojenia pewnej ilości wiedzy teoretycznej by prawidłowo konstruować, oceniać oraz interpretować uzyskane wyniki. Niebezpieczeństwo uzyskania błędnych wyników jest w przypadku filogenetyki obliczeniowej znacznie wyższe niż w wielu innych dziedzinach nauki. Jednak nawet bez szczegółowej wiedzy na ten temat, samodzielne próby wykreślenia drzew filogenetycznych są kształcące gdyż są doskonałym impulsem do dalszego zgłębienia tej fascynującej nauki.



Dzięki dostępowi do Internetu wykonać można analizę relacji filogenetycznych gatunków ryb posiadanych w akwarium. Przykładem będzie typowe proste akwarium w którym żyje kilka dorodnych mieczyków hellera (*Xiphophorus helleri* Heckel, 1848) wraz ze zbrojnikiem niebieskim (*Ancistrus dolichopterus* Kner, 1854) oraz nieprzemysłany prezent z ostatnich urodzin tj. welon (*Carassius auratus auratus* (Linnaeus, 1758)). W czasie obserwacji ryb może się nasunąć następujące pytanie, jakie relacje ewolucyjne wiążą te gatunki? Czy wyraźnie odmienny morfologicznie zbrojnik przyczepiony do szyby dzięki przysawce z przekształconego otworu gębowego może być bliżej spokrewniony z karykaturalną złotą rybką niż z mieczykiem? Może jednak mieczyk i welon mają więcej wspólnego ze sobą niż z „glonojadem”? Istnieje również możliwość że mieczyk i zbrojnik jako ryby z Ameryki są bliżej spokrewnione ze sobą niż z euroazjatyckim welonem. By się tego dowiedzieć można sięgnąć do literatury ale o wiele ciekawsza i zajmująca jest próba samodzielnego odkrycia prawdy. Wykorzystując Internet oraz dostępne darmowe zasoby danych i oprogramowania można samemu wygenerować odpowiedź popartą dowodami

naukowymi.

Bardzo często do analiz filogenetycznych wykorzystuje się sekwencje nukleotydowe mitochondrialnego DNA – mtDNA (ang. **mitochondrial DNA**) kodujące cząsteczki rybosomalnego RNA – rRNA (ang. **ribosomal RNA**). Powszechność badań wykorzystujących te geny np. jedną z podjednostek rybosomu tj. 16S daje podstawy by sądzić, że są one równie znane w przypadku interesujących nas gatunków. By wejść w posiadanie interesujących sekwencji najprościej jest skorzystać z bazy danych genetycznych GenBank dostępnej na stronie <http://www.ncbi.nlm.nih.gov/> tworzonej przez NCBI (ang. **National Center for Biotechnology Information**). Wyszukując informacje w bazie tej jako zapytanie wpisujemy łacińską nazwę gatunku z dodatkową informacją „16S rRNA”. Dla mieczyków interesująca nas sekwencja zapisana jest pod numerem dostępu U80047, dla zbrojnika pod numerem dostępu Y08324 a dla złotej rybki numer dostępu to AJ247070. Po otwarciu pliku z interesującymi nas wynikami widocznych jest wiele informacji z których na tym etapie nie będziemy korzystać. By sprawnie i bezbłędnie skopiować sekwencje na twardy dysk osobistego komputera w okienku „*Display Setting*” wybieramy format „*FASTA*” a w okienku „*Send*” wybieramy „*File*” i w tym momencie rozpocznie się zapis. Gdy posiadamy już trzy interesujące nas sekwencje dla usprawnienia dalszej pracy tworzymy jeden zbiorczy plik notatnika upraszczając nagłówki np. do nazw gatunkowych. Plik zapisujemy z rozszerzeniem *.fas*.

Przystąpić teraz można do właściwej analizy i konstrukcji drzewa filogenetycznego dla tych gatunków na podstawie sekwencji nukleotydowej genu warunkującego podjednostkę rybosomu 16S. Dla tego celu istnieje wiele możliwych programów komputerowych ale wydaje się ze optymalnym rozwiązaniem dla domowego użytku jest program MEGA 3.1 (ang. **Molecular Evolutionary Genetics Analysis**) dostępny na stronie internetowej pod adresem <http://www.megasoftware.net/>. Przy jego pomocy otwieramy plik z porównywanymi sekwencjami z rozszerzeniem *.fas*. Z paska wybieramy funkcję „*Alignment*” a po rozwinięciu menu „*Align by ClustalW*”. Jeśli wcześniej nie wybrano żadnych sekwencji program zapyta czy zaznaczyć wszystkie, odpowiadamy naciskając „*Yes*”. W tym momencie pojawia się okno z parametrami algorytmu uliniowienia ale nic na tym etapie nie zmieniamy i potwierdzamy naciskając „*OK*”. W tym momencie program wykonuje operację uliniowienia sekwencji po której możemy zamknąć plik zapisując otrzymane wyniki. Prócz dotychczasowego pliku *.fas* na dysku zapisane zostaną dwa dodatkowe o tej samej nazwie lecz innych rozszerzeniach tj. *.mas* i *.meg*. Pierwszy to wynik uliniowienia a drugi pozwala na

dalszą obróbkę sekwencji w tym na tworzenie drzew filogenetycznych.

W tym momencie warto przeglądnąć otrzymane wyniki uliniowanie w celu wprowadzenia ewentualnych poprawek. W tym przypadku rzucającą się w oczy cechą jest znaczna różnica w długości porównywanych sekwencji, na co wskazuje bardzo nie równy początek i koniec tego dopasowania. By przekonać się, dlaczego tak się dzieje wystarczy przyjrzeć się jeszcze raz dokładnie informacjom z plików GenBanku. Okazuje się że o ile dla zbrojnika zsekwencjonowany został cały gen to w przypadku mieczyka i welona znany jest tylko jego fragment (ang. partial sequence). By uniknąć błędu związanego niedopasowaniem sekwencji początkowej i końcowej z powodu braku tych fragmentów należy wyrównać wszystkie sekwencje odcinając nadmiarowe odcinki. Usuwa się wszystkie nukleotydy od początku aż do pozycji, w której nukleotydy znane są we wszystkich trzech sekwencjach, czyli do pozycji 17 włącznie oraz od pozycji 449 do końca. Wprowadzone zmiany należy zapisać.

Przy pomocy programu MEGA otwieramy plik z rozszerzeniem *.meg*. W głównym oknie wybieramy funkcję „*Phylogeny*” a z rozwiniętego menu „*Construct Phylogeny* → *Neighbor-Joining (NJ)*”. W otwartym oknie „*Analysis References*” naciskamy na zielony kwadracik w wierszu „*Model*” i wybieramy „*Nucleotide* → *p-distance*”. Wybór potwierdzamy naciskając na „*Compute*”. W tym momencie pojawia się okno „*Tree Explorer*” z graficzną prezentacją drzewa filogenetycznego rozpatrywanych gatunków.

Wynika z niego, że welony są bliżej spokrewnione ze zbrojnikami niż z mieczykami, choć każdy z tych gatunków należy w systematyce do innego rzędu. Identyczny wynik uzyskuje się porównując sekwencje nukleotydowe genu kodującego podjednostkę 12S rybosomalnego RNA.

Analogiczną analizę przeprowadzić można w odniesieniu do kompletnego genomu mitochondrialnego kilku gatunków ryb.

By analizy takie były wiarygodne naukowe należałoby zapoznać się z szeregiem pojęć i metod filogenetycznych. Jednak dla naszych „domowych” edukacyjnych potrzeb wiedza ta nie jest niezbędna. Być może jednak takie zabawy zaowocują w przyszłości poważnym studiowaniem zagadnień bioinformatycznych.

Kolejną ciekawą możliwością, jaką daje nam inne oprogramowanie bioinformatyczne jest przewidywanie rejonów kodujących tzn. eksonów i niekodujących intronów wchodzących w skład analizowanej sekwencji nukleotydowej. Do prognozowania rejonów kodujących użyć można np. oprogramowania „*on-line*”. Pierwszym z prezentowanych

programów jest FGENESH który odnaleźć można pod internetowym adresem <http://www.softberry.com/>. Uruchamia się go kierując kursor na obszar „Gene finding in Eucaryota / FGENESH”. W oknie dialogowym wkleić można np. zapis liniowy mRNA genu kodującego hormon wzrostu – GH (ang. **G**rowth **H**ormone) pochodzącą z genomu bojownika syjamskiego (*Betta splendens* Regan, 1910) o numerze dostępu AY873790. Z opisu zamieszonego w GenBanku wynika, że sekwencja ta składa się z 814 nukleotydów, ale odcinek kodujący ma długość 615 nukleotydów i zawarty jest między pozycją 29 a 643. Korzystając z programu FGENESH jako organizm wzorcowy należy wybrać ryby „fish”. Po naciśnięciu „Search” wyświetlany jest wynik, czyli w tym przypadku jest to jeden ekson od pozycji 29 do 643 o długości 615 nukleotydów.

Analizę tego samego genu można przeprowadzić przy pomocy programu GENSCAN na stronie <http://genes.mit.edu/GENSCAN.html>. Jako organizm należy wybrać „Vertebrate” czyli kręgowce. W oknie dialogowym należy wkleić sekwencje mRNA genu GH bojownika a uruchomienie analizy następuje po naciśnięciu na „Run GENSCAN”. W wyniku predykcji otrzymaliśmy identyczne wyniki, co w przypadku FGENESH to znaczy jeden ekson o długości 615 nukleotydów rozpoczynający się od 29 do 643 pozycji.

Prezentowane powyżej przykłady analiz nie wyczerpują tego zagadnienia a wręcz przeciwnie gdyż są jedynie drobnym fragmentem szerokiej palety możliwości analitycznych. Bioinformatyka to interdyscyplinarna dziedzina integrująca ze sobą m.in. biologię molekularną, genetykę, genomikę, informatykę i matematykę. Spojrzenie na akwarystyczne hobby przez pryzmat bioinformatyki jest podejściem nietypowym. Jednak efektem tych zainteresowań jest poszerzanie swojej wiedzy biologicznej na znacznie szerszym obszar niż „litraż swojego zbiornika”.



Literatura:

1. Kumar S., Tamura K., Nei M.: 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5:150-163.
2. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G.: 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, 24:4876-4882.
3. Page R.D.M.: 1996. TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, 12: 357-358.
4. Hall T.A.: 1999. Bio-Edit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41:95–8.